Kybernetika

Mohamed Boukeloua; Sarra Leulmi; Ferial Saihi
Asymptotic normality of kernel estimators under censoring and association

Kybernetika, Vol. 62 (2026), No. 4, 643–674

Persistent URL: http://dml.cz/dmlcz/153778

Terms of use:

© Institute of Information Theory and Automation AS CR, 2026

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://dml.cz

KYBERNETIKA — VOLUME *62* (2026), NUMBER 4, PAGES 643–674

ASYMPTOTIC NORMALITY OF KERNEL ESTIMATORS UNDER CENSORING AND ASSOCIATION

Mohamed Boukeloua, Sarra Leulmi and Ferial Saihi

In this work, we study kernel estimators of the density and the failure rate in the context of association and twice censoring. Association is a very interesting type of dependence given the fact that it covers many useful random processes. In this context, we assume that we have at disposal a sample of strictly stationary and associated copies of the variable of interest. Furthermore, this variable is right censored by another variable, the minimum of the two variables is left censored and the three latent variables are independent. Under these assumptions, we establish the asymptotic normality of kernel estimators of the density and the failure rate. Then, we illustrate the performance and the Gaussian behavior of these estimators through a simulation study. The results of this study show the good quality of estimation and the good behavior of the studied asymptotic distributions with respect to the standard normal distribution. We also show the performance of our estimators using an application on a real data set.

Keywords: twice censoring, associated data, kernel estimators, density, failure rate, asymptotic normality

Classification: 62N02, 62G07, 62G20

1. INTRODUCTION

The notion of association takes a very important place in statistical problems dealing with dependent data. It includes a large class of popular random processes such as Gaussian processes with non negative covariances (see [19]), Bernoulli processes with non negative covariances and some autoregressive processes (see [12]). The association in a random process means that the covariance of coordinate-wise non decreasing functions of any parts of the process is always non negative. More precisely, we say that the sequence $(X_n)_{n\in\mathbb{N}^*}$ of real random variables (r.r.v.) is associated if for any finite set $I \subset \mathbb{N}^*$ and any bounded and coordinate-wise non decreasing functions $g_1, g_2 : \mathbb{R}^{|I|} \longrightarrow \mathbb{R}$ (where $|I|$ denotes the cardinal number of I), we have

$$\operatorname{Cov}\left(g_1\left(X_i, i \in I\right), g_2\left(X_i, i \in I\right)\right) \geq 0.$$

The non parametric estimation under association received a lot of interest in the literature for both complete and right censored data. We cite for example the works of

DOI: 10.14736/kyb-2026-4-0643

[1] and [24] who studied the strong consistency and the asymptotic normality of the empirical cumulative distribution function, in the presence of complete associated data. Within the same framework, the kernel density estimation was considered by [20] and [2] who studied the strong consistency. The asymptotic normality was established by [21]. Moreover, [10] proposed a general estimator of the density and they proved its weak consistency under association. Furthermore, [11] established Berry-Esseen type bounds for associated data and they applied them to study the mean integrated absolute error of the kernel density estimator. In the context of associated right censored data, [9] established the strong consistency and the weak convergence of the Kaplan-Meier estimator. Moreover, [14] proved the strong consistency of the kernel density estimator. The asymptotic normality was studied in [13].

In the present paper, we are interested in associated twice censored data. In this framework, the variable of interest X is right censored by another variable R, the minimum of the two variables is itself left censored by L and the three latent variables are independent. So, we observe $Z = \max(\min(X, R), L)$. In practice, associated twice censored may be encountered in reliability studies. For example, consider a system composed of n components $C_1, C_2, \ldots, C_n$. If these components are subject to the same environmental stress, then their respective lifetimes are associated (see [2]). Alternatively, if the lifetimes of the components follow a multivariate exponential distribution, then they are associated (see [17]). Concerning the twice censoring, it may occur when a component C_i is connected in series to a component D and the sub-system composed of C_i and D is connected in parallel to an other component E. In this situation, the lifetime of the component C_i is right censored by the lifetime of D and left censored by the lifetime of E (see [18]).

In the case of independent and identically distributed (i.i.d.) data, the non parametric estimation under twice censoring was considered by many authors such as [18] who studied the estimation of the cumulative distribution function and [16], [4] and [6] who studied the estimation of the density function. Other problems related to twice censored data were considered by [23], [7] and [5]. In the case of associated twice censored data, [22] proved the strong consistency of the estimator of Patilea and Rolin of the cumulative distribution function. They also proved the strong consistency of kernel estimators of the density, the failure rate and the mode. For our part, we establish the asymptotic normality of the kernel estimator $\widehat{f}_n$ of the density f of X. The asymptotic normality is a very important property since it allows to give an approximation of the distribution of $\widehat{f}_n$, when n becomes large. It also permits to construct an asymptotic confidence interval for f. To prove the asymptotic normality, we introduce a pseudo estimator where we replace, in $\widehat{f}_n$, the estimators of S_X, F_L and F_Z by these true functions (where F_U and S_U denote respectively the distribution function and the survival function of the r.r.v. U). The importance of this pseudo estimator lies in the fact that it is written as a sum of identically distributed, associated variables which is not the case for the estimator $\widehat{f}_n$. Then, we decompose the difference $\widehat{f}_n - f$ into three terms. The first one is a stochastic term which involves $\widehat{f}_n$ and the pseudo estimator. The second one is a centering of the pseudo estimator and the third one represents the bias of the pseudo estimator (see relation (1) below). This decomposition is the same used by [13] in the case of associated right censored data, where the pseudo estimator involves only the function S_R instead

of the functions S_X, F_L and F_Z. Therefore, the study of the asymptotic normality is more complicated in our context. In particular, the study of the stochastic term necessitates the properties of the estimators of the functions S_X, F_L and F_Z studied in [22]. Moreover, the study of the bias of the pseudo estimator involves the functions S_X, F_L and F_Z instead of the function S_R in the case of right censoring. Finally, the study of the second term, which is a centering of the pseudo estimator, is conducted in the same way as in the case of complete associated data (see [21]) since this term is a sum of identically distributed associated variables. The novelty in this study is the involvement of the functions S_X, F_L and F_Z which are not present in the case of complete data. Another novelty in this paper is the fact that we also establish the asymptotic normality of two kernel estimators of the failure rate of X. For each estimator, we use a specific decomposition to show the claimed result. Then, we carry out a simulation study in order to check the performance and the Gaussian behavior of our studied estimators for finite samples. We also apply our estimators on a real data set.

The rest is organized as follows. In Section 2, we present our assumptions and theoretical results. In Section 3, we present the results of the simulation study. Section 4 is devoted to the real data application. Some discussions and perspectives are given in Section 5. The proofs are postponed to Appendix A. Finally, the data set we use in the application is given in Appendix B.

2. MAIN RESULTS

Let X be a r.r.v. with a probability density function (p.d.f.) f. In the context of associated twice censored data, we observe a sample $(Z_i = \max(\min(X_i, R_i), L_i), A_i)_{1\leq i\leq n}$, where

$(X_i)_{1\leq i\leq n}$ is a sequence of strictly stationary and associated r.r.v. with the same distribution as X,

$(R_i)_{1\leq i\leq n}$ (resp. $(L_i)_{1\leq i\leq n}$) is a sequence of i.i.d. r.r.v. with the same distribution as R (resp. L). R is called the right censoring variable and L is called the left censoring one and

$$A_i = \begin{cases} 0 & \text{if } L_i < X_i \leq R_i, \\ 1 & \text{if } L_i < R_i < X_i, \\ 2 & \text{if } \min(X_i, R_i) \leq L_i. \end{cases}$$

The variables X, R and L are assumed to be non negative and independent.

In what follows, for any r.r.v. U, we denote by $I_U = \inf\{t \in \mathbb{R}, F_U(t) > 0\}$ the lower endpoint of the support of U and $T_U = \sup\{t \in \mathbb{R}, F_U(t) < 1\}$ the upper endpoint of the support of U. Moreover, for any right continuous function $g : \mathbb{R} \longrightarrow \mathbb{R}$, we set $g(t^-) = \lim_{\varepsilon \geq \to 0} g(t - \varepsilon)$ and $\Delta g(t) = g(t) - g(t^-)$, when the limit exists. We assume that the functions F_R and F_L are continuous.

In the present case of twice censored data, the kernel density estimator of f has been introduced by [16]. It is defined by

$$\widehat{f}_n(x) = \frac{1}{h_n} \int K\left(\frac{x - z}{h_n}\right) \mathrm{d}\widehat{F}_n(z),$$

where K is the kernel, h_n is the bandwidth and $\widehat{F}_n$ is the product-limit estimator of F_X (see [18]), defined by

$$\widehat{F}_n(x) = 1 - \prod_{j/Z'_j \leq x} \left(1 - \frac{D_{0j}}{nF_L^{(n)}(Z'^-_j) - nF_Z^{(n)}(Z'^-_j)}\right),$$

where $(Z'_j)_{1\leq j\leq m}$ $(m \leq n)$ are the distinct values in increasing order of $(Z_i)_{1\leq i\leq n}$,

$$D_{kj} = \sum_{i=1}^{n} 1_{\{Z_i = Z'_j, A_i = k\}}, \quad \text{for all } \ 1 \leq j \leq m \ \text{ and } k \in \{0,1,2\},$$

(1_E denotes the indicator function of the set E),

$$F_Z^{(n)}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{Z_i \leq x\}}$$

is the empirical version of F_Z and

$$F_L^{(n)}(x) = \prod_{j/Z'_j > x} \left(1 - \frac{D_{2j}}{F_Z^{(n)}(Z'_j)}\right)$$

is the product-limit estimator of F_L.

Remarking that

$$\Delta\widehat{F}_n(Z_i) = \frac{\widehat{S}_n\left(Z_i^-\right) 1_{\{A_i=0\}}}{nF_L^{(n)}\left(Z_i^-\right) - nF_Z^{(n)}\left(Z_i^-\right)},$$

where $\widehat{S}_n(x) = 1 - \widehat{F}_n(x)$, we get

$$\widehat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - Z_i}{h_n}\right) \frac{\widehat{S}_n\left(Z_i^-\right) 1_{\{A_i=0\}}}{F_L^{(n)}\left(Z_i^-\right) - F_Z^{(n)}\left(Z_i^-\right)}.$$

So, we introduce the following pseudo estimator of f, which will be helpful in the sequel.

$$\begin{aligned}\widetilde{f}_n(x) &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - Z_i}{h_n}\right) \frac{S_X(Z_i)\, 1_{\{A_i=0\}}}{F_L(Z_i) - F_Z(Z_i)} \\ &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - Z_i}{h_n}\right) \frac{1_{\{A_i=0\}}}{F_L(Z_i)\, S_R(Z_i)},\end{aligned}$$

because $F_L - F_Z = S_X F_L S_R$.

The strong consistency of $\widehat{f}_n$ in the presence of associated data, has been established by [22]. In the present paper, we are interested in its asymptotic normality at a fixed point $x \in]I_X, \min(T_X, T_R)[$. For that, we suppose that the following assumptions are satisfied for some $\beta > 0$.

H1 There exists $\alpha > 1$ and C_0 $(0 < C_0 < \infty)$ such that

$$\frac{\log n}{p_n c_n^2} \exp\left\{\left(\frac{\alpha n \log n}{2p_n}\right)^{1/2}\right\} \sum_{j=p_n+2}^{\infty} (\mathrm{Cov}(X_1, X_j))^{1/3} \leq C_0,$$

where $p_n = \lfloor (\log n)^\beta \rfloor$ and $c_n = (\log n)^{\beta/2}$ ($\lfloor t \rfloor$ denotes the integer part of the real t).

H2 $\max(I_L, I_R) < I_X$.

H3 The p.d.f. f is twice continuously differentiable at x.

H4 The covariance term satisfies: $\rho(r) := \sup_{j:|\ell-j|\geq r} \mathrm{Cov}\,(X_j, X_\ell)$ for all $\ell \geq 1$ and $r > 0$, where $\rho(r) \leq \gamma_0 e^{-\gamma r}$ for some positive constants γ_0 and γ.

H5 K is a Lipschitz p.d.f. with compact support and it satisfies $\int uK(u)\,\mathrm{d}u = 0$.

H6 $\frac{\log(n)}{nh_n} \longrightarrow 0$ and $nh_n^5 \longrightarrow 0$ as $n \to \infty$.

H7 There exist two sequences of positive integers $(a_n)_{n\in\mathbb{N}^*}$ and $(b_n)_{n\in\mathbb{N}^*}$ that tend to infinity and such that

(i) $k_n \longrightarrow \infty$ and $\frac{k_n b_n}{n} \longrightarrow 0$, as $n \to \infty$, where $k_n = \left\lfloor \frac{n}{a_n + b_n} \right\rfloor$.

(ii) $a_n h_n \longrightarrow 0$ and $\frac{a_n}{\sqrt{nh_n}} \longrightarrow 0$, as $n \to \infty$.

(iii) $\frac{1}{h_n^3} \sum_{j=b_n}^{\infty} |\mathrm{Cov}(X_1, X_{1+j})| \longrightarrow 0$, as $n \to \infty$.

Assumptions **H1**–**H5** have already been employed in [22]. Assumption **H6** is standard in the non parametric estimation setting. Note that it implies that $nh_n \longrightarrow \infty$, as $n \to \infty$. Moreover, the new assumption **H7** is needed to establish the asymptotic normality of $\widehat{f}_n$ and it has already been used in [21] and [13].

Theorem 2.1. Let $x \in]I_X, \min(T_X, T_R)[$ and let $\beta > 0$ for which assumption **H1** holds. Under assumptions **H2**–**H7**, we have

$$\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x)\right), \text{ as } n \to \infty,$$

where

$$\sigma^2(x) = \frac{f(x)}{F_L(x) S_R(x)} \int K^2(z)\,\mathrm{d}z$$

and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution.

In the case of i.i.d. twice censored data, the asymptotic normality of $\widehat{f}_n(x)$ has been proved by [6]. The same result has been proved by [13] in the case of associated right censored data. Theorem 2.1 above generalizes both these results.

A useful application, in survival analysis, of the density estimation is the estimation of the failure rate. The failure rate of X is defined by $\lambda(x) = \frac{f(x)}{S_X(x)}$, if $S_X(x) \neq 0$ and

$\lambda(x) = 0$ otherwise. [16] and [22] studied two kernel estimators of $\lambda(x)$, given respectively by

$$\widehat{\lambda}_n(x) = \frac{\widehat{f}_n(x)}{\widehat{S}_n(x) + u_n},$$

where $(u_n)_{n\in\mathbb{N}}$ is a sequence of positive real numbers going to zero, which allows to avoid division by zero and

$$\lambda_n^*(x) = \frac{1}{h_n} \int K\left(\frac{x-z}{h_n}\right) \mathrm{d}\widehat{\Lambda}_n(z),$$

where

$$\widehat{\Lambda}_n(x) = \int_0^x \frac{\mathrm{d}H_n^{(0)}(t)}{F_L^{(n)}(t^-) - F_Z^{(n)}(t^-)}, \text{ with } H_n^{(0)}(x) = \frac{1}{n}\sum_{i=1}^n 1_{\{Z_i \le x, A_i = 0\}}.$$

$\widehat{\Lambda}_n(x)$ is the estimator of the cumulative hazard function of X.

The asymptotic normality of $\widehat{\lambda}_n(x)$ and $\lambda_n^*(x)$ is given by the following theorem.

Theorem 2.2. Let $x \in]I_X, \min(T_X, T_R)[$ and let $\beta > 0$ for which assumption **H1** holds. Under assumptions **H2**–**H7**, we have

i) If $u_n = o\left(\frac{1}{\sqrt{nh_n}}\right)$, then

$$\sqrt{nh_n}\left(\widehat{\lambda}_n(x) - \lambda(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{f(x)}{S_X^2(x)F_L(x)S_R(x)} \int K^2(z)\,\mathrm{d}z\right), \text{ as } n \to \infty.$$

ii)

$$\sqrt{nh_n}\left(\lambda_n^*(x) - \lambda(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{f(x)}{S_X(x)F_L(x)S_R(x)} \int K^2(z)\,\mathrm{d}z\right), \text{ as } n \to \infty.$$

We remark that the estimator $\lambda_n^*(x)$ is better than $\widehat{\lambda}_n(x)$, given that its asymptotic variance is smaller.

3. SIMULATION STUDY

In this section, we present the results of a simulation study aiming to illustrate the performances and the Gaussian behavior of our studied estimators, for samples of finite size n. In order to show the influence of the strength of the dependence between the variables $(X_i)_{1\le i\le n}$, we use two theoretical models:

In model I, we generate $(n+2)$ i.i.d. random variables W_i from the standard normal distribution. Then, we generate n associated random variables $X_i = \exp\left((W_{i-1} + W_{i-2})/2\right)$. The variables $(X_i)_{1\le i\le n}$ have a log normal distribution with parameters $\mu = 0$ and $\sigma^2 = 1/2$ (denoted by $\log\mathcal{N}(0, 1/2)$). Concerning the censoring variables, we take $R = 100$, almost surely (a.s.) and L as a random variable having a Weibull distribution with scale parameter 0.78 and shape parameter 3 (denoted by $\mathcal{W}(0.78, 3)$). For this model, the rate of censoring is about 30%.

In model II, we generate $(n+3)$ i.i.d. random variables W_i from the standard normal distribution. Then, we generate n associated random variables

$$X_i = \exp\left(\left(W_{i-1} + W_{i-2} + W_{i-3}\right)/3\right).$$

The variables X_i follow the $\log\mathcal{N}(0, 1/3)$ distribution. Moreover, we generate the censoring variables R and L from the $\mathcal{W}(2.1, 3)$ and the $\mathcal{W}(0.5, 3)$ distributions, respectively. It is clear that for this model, the dependence between the $(X_i)_{1\leq i\leq n}$ is stronger than that in model I. As in the previous model, the rate of censoring is about 30%.

For each model, we draw the curves of the estimators $\widehat{f}_n(x)$, $\widehat{\lambda}_n(x)$ and $\lambda_n^*(x)$ in the same graphs with the corresponding theoretical curves. The computation of our estimators is based on samples of size $n = 300$. Moreover, we use the Epanechnikov kernel defined by $K(u) = \frac{3}{4}(1-u^2)1_{[-1,1]}(u)$ and the optimal bandwidth $h_n = \left(\frac{\log(n)}{n}\right)^{1/5}$, obtained in [22]. For $n = 300$, we get $h_n = 0.4526978$. Furthermore, for the estimator $\widehat{\lambda}_n(x)$, we take $u_n = 1/n$. The graphs we obtain are given in Figures 1 – 3. These figures show that our studied estimators have a good quality of estimation. We also remark that the results for model I are slightly better than those for model II and that the estimator $\lambda_n^*(x)$ is a little better than $\widehat{\lambda}_n(x)$.

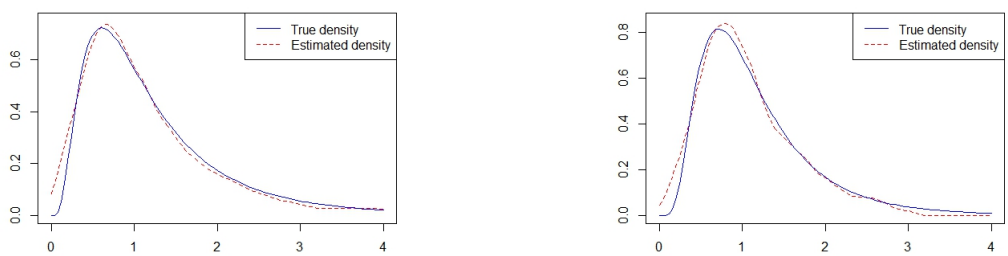


Fig. 1. Representation of $\widehat{f}_n(x)$ for model I (left graph) and model II (right graph).

To illustrate the Gaussian behavior of our estimators, we generate 1000 samples of size $n = 300$ of the variables X, R and L and we calculate the estimators $\widehat{f}_n(x)$, $\widehat{\lambda}_n(x)$ and $\lambda_n^*(x)$ corresponding to each sample, at a fixed value $x = 1$. Then, we cut out the true density (or failure rate) and we divide by the sample standard deviation, which gives a sample of standardized values. Finally, we plot the histogram of this sample and its kernel density estimator against the density of the standard Gaussian distribution. We also use the Q-Q plot graph as a further tool to check the normality. To compute the studied estimators, we use the Epanechnikov kernel and the same optimal bandwidth used previously. The graphs we obtain are presented in Figures 4 – 9. These figures show the good behavior of the studied asymptotic distributions with respect to the standard normal distribution. For the three estimators, the results in model I are slightly better than those in model II.

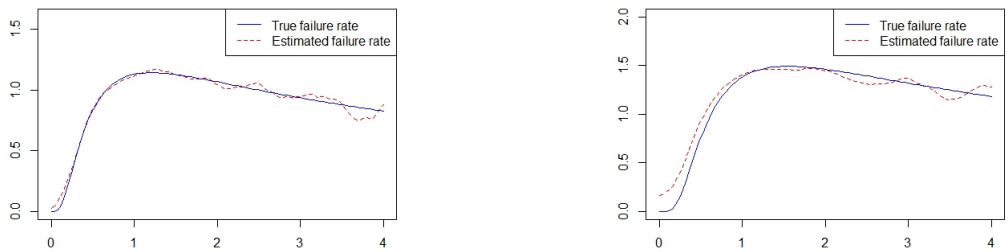


Fig. 2. Representation of $\widehat{\lambda}_n(x)$ for model I (left graph) and model II (right graph).

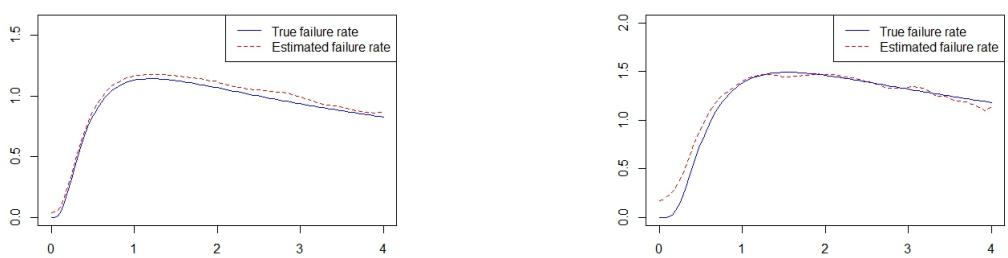


Fig. 3. Representation of $\lambda_n^*(x)$ for model I (left graph) and model II (right graph).

In order to confirm our results, we use the Pearson's χ^2 and the Lilliefors tests whose $p-$values are presented in Table 1. In almost all cases, the two tests accept the normality. For the estimator $\widehat{\lambda}_n(x)$, the p-value of the Lilliefors test is small in model I but it allows to accept the normality at the significance level of 5%.

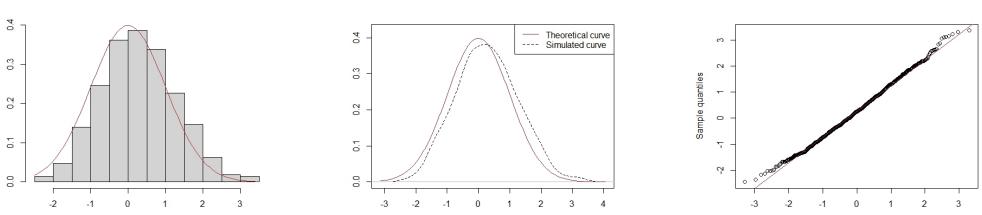


Fig. 4. The obtained results for $\widehat{f}_n(x)$ in model I.

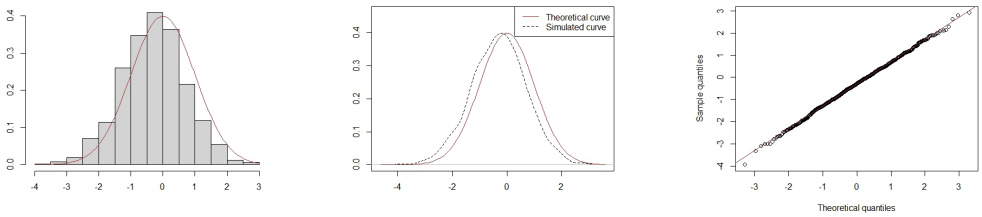


Fig. 5. The obtained results for $\widehat{f}_n(x)$ in model II.

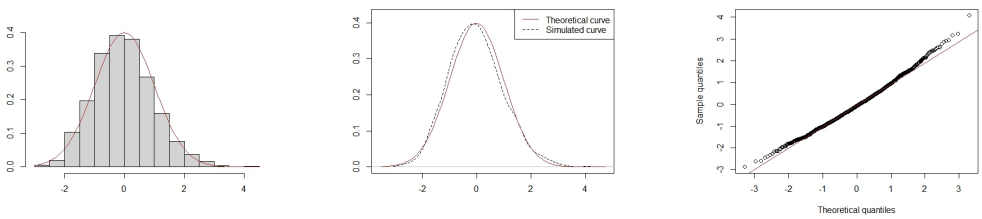


Fig. 6. The obtained results for $\widehat{\lambda}_n(x)$ in model I.

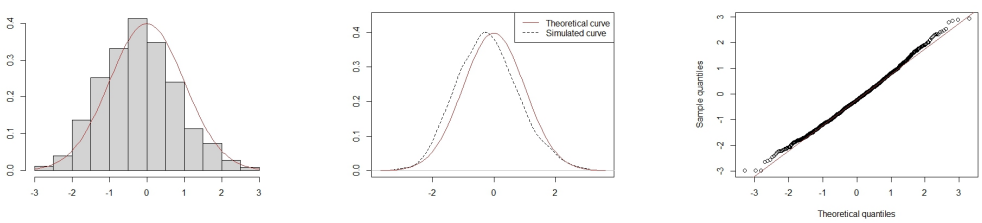


Fig. 7. The obtained results for $\widehat{\lambda}_n(x)$ in model II.

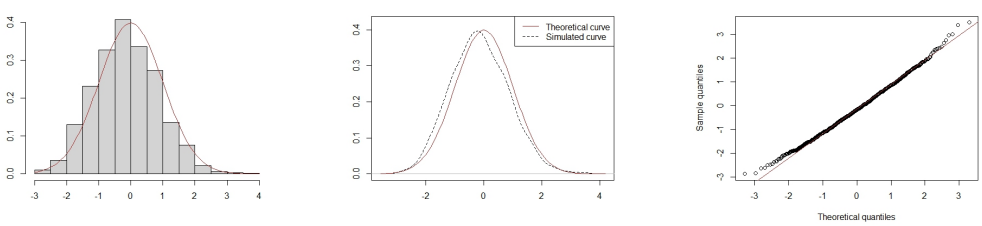


Fig. 8. The obtained results for $\lambda_n^*(x)$ in model I.

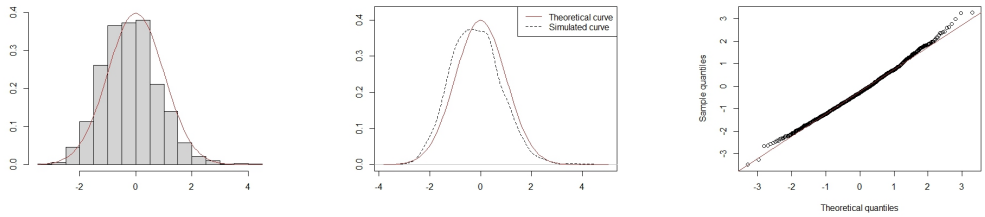


Fig. 9. The obtained results for $\lambda_n^*(x)$ in model II.

The estimator	The model	$p-$value of the Pearson's χ^2 test	$p-$value of the Lilliefors test
$\widehat{f}_n(x)$	Model I	0.3605	0.2025
	Model II	0.6023	0.5960
$\widehat{\lambda}_n(x)$	Model I	0.3696	0.0755
	Model II	0.1745	0.2476
$\lambda_n^*(x)$	Model I	0.7476	0.3394
	Model II	0.6263	0.1383

Tab. 1. $p-$values of the Pearson's χ^2 and the Lilliefors tests.

4. REAL DATA APPLICATION

In this section, we present the results of an application on a real data set. This data set concerns intravenous drug users who where infected with HIV due to the needle-sharing. A study was conducted in German Trias i Pujol Hospital in Badalona (Spain) where the elapsed time between the first exposed to HIV by injecting drugs and the AIDS diagnosis, was recorded, in days, for $n = 232$ patients. This time is subject to right and left censoring in the same time. Right censoring corresponds to the patients who still AIDS-free at the end of the study. Left censoring corresponds the patients who died with AIDS without registration of the time of AIDS diagnosis. This data set which has already been used in [15] contains $n_1 = 82$ complete data, $n_2 = 136$ right censored data and $n_3 = 14$ left censored data. For our part, we are interested in the estimation of the density and the failure rate of the mean time elapsed between the first exposed to HIV by injecting drugs and the AIDS diagnosis. Since the observed data are independent, we use the moving average technique to induce association. For that, we set

$$\begin{aligned}
\mathcal{T} &= \{T_i, \text{ such that:} 1 \leq i \leq n\}, \\
\mathcal{T}^c &= \{T_i, \text{ such that:} 1 \leq i \leq n \text{ and } A_i = 0\}, \\
\mathcal{T}^r &= \{T_i, \text{ such that:} 1 \leq i \leq n \text{ and } A_i = 1\}
\end{aligned}$$

and

$$\mathcal{T}^l = \{T_i, \text{ such that:} 1 \leq i \leq n \text{ and } A_i = 2\},$$

where T is the observed variable and A is the indicator of censoring.

The sets $\mathcal{T}$, $\mathcal{T}^c$, $\mathcal{T}^r$ and $\mathcal{T}^l$ represents respectively the set of all observations, the set of complete observations as well as the sets of right and left observations. Moreover, we denote by $(T_i^c)_{1\leq i\leq n_1}, (T_i^r)_{1\leq i\leq n_2}$ and $(T_i^l)_{1\leq i\leq n_3}$ the elements of $\mathcal{T}^c$, $\mathcal{T}^r$ and $\mathcal{T}^l$, respectively.

We consider three cases:

In case 1, we calculate the variables

$$Z_i^c = \frac{1}{2}\left(T_i^c + T_{i+1}^c\right), \text{ for } i \in \{1, \ldots, n_1 - 1\},$$

$$Z_i^r = \frac{1}{2}\left(T_i^r + T_{i+1}^r\right), \text{ for } i \in \{1, \ldots, n_2 - 1\}$$

and

$$Z_i^l = \frac{1}{2}\left(T_i^l + T_{i+1}^l\right), \text{ for } i \in \{1, \ldots, n_3 - 1\}.$$

The variables Z_i^c, Z_i^r and Z_i^l represent respectively the complete observations as well as the right censored and the left censored ones in the new sample. Concatenating these values, we get a sample of strictly stationary, associated r.r.v. (as coordinate-wise increasing functions applied on finite subsets of the i.i.d. family $\mathcal{T}$).

In case 2, we calculate the variables

$$Z_i^c = \frac{1}{3}\left(T_i^c + T_{i+1}^c + T_{i+2}^c\right), \text{ for } i \in \{1, \ldots, n_1 - 2\},$$

$$Z_i^r = \frac{1}{3}\left(T_i^r + T_{i+1}^r + T_{i+2}^r\right), \text{ for } i \in \{1, \ldots, n_2 - 2\}$$

and

$$Z_i^l = \frac{1}{3}\left(T_i^l + T_{i+1}^l + T_{i+2}^l\right), \text{ for } i \in \{1, \ldots, n_3 - 2\}.$$

In case 3, we calculate the variables

$$Z_i^c = \frac{1}{4}\left(T_i^c + T_{i+1}^c + T_{i+2}^c + T_{i+3}^c\right), \text{ for } i \in \{1, \ldots, n_1 - 3\},$$

$$Z_i^r = \frac{1}{4}\left(T_i^r + T_{i+1}^r + T_{i+2}^r + T_{i+3}^r\right), \text{ for } i \in \{1, \ldots, n_2 - 3\}$$

and

$$Z_i^l = \frac{1}{4}\left(T_i^l + T_{i+1}^l + T_{i+2}^l + T_{i+3}^l\right), \text{ for } i \in \{1, \ldots, n_3 - 3\}.$$

In each case, the original observed sample is modified. This is the price to pay in order to obtain a sample of associated twice censored data. Consequently, the density and the failure rate that we will estimate do not really correspond to the elapsed time

between the first exposed to HIV by injecting drugs and the AIDS diagnosis. Since the transformed samples are constructed using the moving average technique, they can give an idea about the density and the failure rate of the mean time elapsed between the two events.

We present the original data set as well as our moving averages for the three cases in Appendix B below. Obviously, the strongest dependence between the observations is reached in case 3. Moreover, the dependence in case 2 is stronger than that in case 1. Indeed, in case 3, the variables (Z_i) are constructed using four consecutive values of (T_i). In case 2, they are constructed using three consecutive values of (T_i) and in case 1, they are constructed using two consecutive values of (T_i). To be more precise, we will calculate the auto-correlation function of the series (Z_i^c) in each case. We denote this function by ρ^I (resp. ρ^{II}, ρ^{III}) in case 1 (resp. case 2, case 3). In case 1, we have

$$\rho^I(1) = \frac{\text{Cov}\left(Z_i^c, Z_{i+1}^c\right)}{Var\left(Z_i^c\right)} = \frac{\text{Cov}\left(\frac{1}{2}\left(T_i^c + T_{i+1}^c\right), \frac{1}{2}\left(T_{i+1}^c + T_{i+2}^c\right)\right)}{Var\left(\frac{1}{2}\left(T_i^c + T_{i+1}^c\right)\right)}.$$

Since the variables (T_i^c) are independent, we can write

$$\rho^I(1) = \frac{\frac{1}{4}Var\left(T_{i+1}^c\right)}{\frac{1}{4}\left(Var\left(T_i^c\right) + Var\left(T_{i+1}^c\right)\right)} = \frac{\sigma_T^2}{2\sigma_T^2} = \frac{1}{2},$$

where $\sigma_T^2 = Var(T)$.

Moreover, for $k \geq 2$, we have $Z_i^c = \frac{1}{2}\left(T_i^c + T_{i+1}^c\right)$ and $Z_{i+k}^c = \frac{1}{2}\left(T_{i+k}^c + T_{i+k+1}^c\right)$ are independent because they have no common term, so $\rho^I(k) = 0$. In conclusion

$$\rho^I(k) = \begin{cases} \frac{1}{2} & \text{if } k = 1 \\ 0 & \text{if } k \geq 2. \end{cases}$$

In case 2, we have

$$\begin{aligned}
\rho^{II}(1) &= \frac{\text{Cov}\left(Z_i^c, Z_{i+1}^c\right)}{Var\left(Z_i^c\right)} \\
&= \frac{\text{Cov}\left(\frac{1}{3}\left(T_i^c + T_{i+1}^c + T_{i+2}^c\right), \frac{1}{3}\left(T_{i+1}^c + T_{i+2}^c + T_{i+3}^c\right)\right)}{Var\left(\frac{1}{3}\left(T_i^c + T_{i+1}^c + T_{i+2}^c\right)\right)} \\
&= \frac{\frac{1}{9}\left(Var\left(T_{i+1}^c\right) + Var\left(T_{i+2}^c\right)\right)}{\frac{1}{9}\left(Var\left(T_i^c\right) + Var\left(T_{i+1}^c\right) + Var\left(T_{i+2}^c\right)\right)} \text{ (by independence of } (T_i^c)) \\
&= \frac{2\sigma_T^2}{3\sigma_T^2} \\
&= \frac{2}{3}.
\end{aligned}$$

In the same way, we get

$$\rho^{II}(2) = \frac{\text{Cov}\left(Z_i^c, Z_{i+2}^c\right)}{Var\left(Z_i^c\right)}$$

$$= \frac{\mathrm{Cov}\left(\frac{1}{3}\left(T_i^c + T_{i+1}^c + T_{i+2}^c\right), \frac{1}{3}\left(T_{i+2}^c + T_{i+3}^c + T_{i+4}^c\right)\right)}{Var\left(Z_i^c\right)}$$
$$= \frac{\frac{1}{9}Var\left(T_{i+2}^c\right)}{\frac{3}{9}\sigma_T^2}$$
$$= \frac{1}{3}.$$

Moreover, for $k \geq 3$, we have $Z_i^c = \frac{1}{3}\left(T_i^c + T_{i+1}^c + T_{i+2}^c\right)$ and $Z_{i+k}^c = \frac{1}{3}\left(T_{i+k}^c + T_{i+k+1}^c + T_{i+k+2}^c\right)$ are independent because they have no common term, so $\rho^{II}(k) = 0$. In conclusion

$$\rho^{II}(k) = \begin{cases} \frac{2}{3} & \text{if } k = 1 \\ \frac{1}{3} & \text{if } k = 2 \\ 0 & \text{if } k \geq 3. \end{cases}$$

In case 3, we can show as in the previous cases that

$$\rho^{III}(k) = \begin{cases} \frac{3}{4} & \text{if } k = 1 \\ \frac{1}{2} & \text{if } k = 2 \\ \frac{1}{4} & \text{if } k = 3 \\ 0 & \text{if } k \geq 4. \end{cases}$$

We remark that for all $k \in \mathbb{N}$, $\rho^{III}(k) \geq \rho^{II}(k) \geq \rho^{I}(k)$, so the dependence in case 3 is stronger than that of case 2 which is stronger than that of case 1.

In each case, we obtain a sample of strictly stationary and associated twice censored r.r.v. On the basis of this sample, we use the bootstrap technique with 1000 replications to evaluate the integrated standard error of the estimators $\widehat{f}_n$, $\widehat{\lambda}_n$ and λ_n^*. The integrated standard error is defined for $\widehat{f}_n$ by

$$\int \left(SE\left(\widehat{f}_n(x)\right)\right) \mathrm{d}x,$$

where $SE\left(\widehat{f}_n(x)\right)$ is the standard error of $\widehat{f}_n(x)$.

The integrated standard error of $\widehat{\lambda}_n$ and λ_n^* is defined in the same way. To calculate our studied estimators, we use the Epanechnikov kernel and the optimal bandwidth $h_n = \left(\frac{\log(n)}{n}\right)^{1/5}$, obtained in [22]. We get the results given in Table 2. These results show that the studied estimators have a good quality of estimation since the integrated standard error is relatively small. In all cases, the estimator $\widehat{f}_n(x)$ has the smallest integrated standard error. Moreover, the estimator $\lambda_n^*(x)$ is better than $\widehat{\lambda}_n(x)$. Furthermore, the integrated standard error increases when the dependence between the observations becomes stronger.

The estimator	Case 1	Case 2	Case 3
$\widehat{f_n}(x)$	3.123×10^{-4}	4.665×10^{-4}	5.851×10^{-4}
$\widehat{\lambda_n}(x)$	2.304×10^{-3}	3.086×10^{-3}	4.117×10^{-3}
$\lambda_n^*(x)$	2.066×10^{-3}	2.910×10^{-3}	3.441×10^{-3}

Tab. 2. The integrated standard error of the studied estimators.

5. DISCUSSION

In this work, we have studied kernel estimators of the density and the failure rate in the presence of associated twice censored data. In particular, we have established the asymptotic normality of theses estimators. We have also shown the performances of these estimators through a simulation study and a real data application. Our theoretical results extend those of [6] to the case of associated data. They also extend the results of [21] and [13] to the context of twice censored data. The limitation of our work lies in the fact that we prove our results under the condition that the kernel function is symmetric and compactly supported. In the future, it would be interesting to extend our results to other types of kernels. It would also be interesting to look at other types of censored data such as interval censored data or current status data. An other extension of our work could be the study of the asymptotic properties of the random processes $\left\{\sqrt{nh_n}\left(\widehat{f_n}(x) - f(x)\right),\ x \in]I_X, \min(T_X, T_R)[\right\}$, $\left\{\sqrt{nh_n}\left(\widehat{\lambda}_n(x) - \lambda(x)\right),\right.$ $\left.x \in]I_X, \min(T_X, T_R)[\right\}$ and $\left\{\sqrt{nh_n}\left(\lambda_n^*(x) - \lambda(x)\right),\ x \in]I_X, \min(T_X, T_R)[\right\}$ which generalizes our pointwise study and allows, in particular, to explore the mutual correlations between different points.

ACKNOWLEDGEMENT

The authors wish to thank the editor, the associate editor and the reviewers for their valuable suggestions and comments.

(Received November 1, 2025)

REFERENCES

[1] I. Bagai and B. L. S. Prakasa Rao: Estimation of the survival function for stationary associated processes. Statist. Probab. Lett. *12* (1991), 385–391. DOI:10.1016/0167-7152(91)90027-O

[2] I. Bagai and B. L. S. Prakasa Rao: Kernel-type density and failure rate estimation for associated sequences. Ann. Inst. Statist. Math. *47* (1995), 253–266. DOI:10.1007/BF00773461

[3] P. Billingsley: Probability and Measure. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons Inc., New York 1995.

[4] M. Boukeloua: Rates of mean square convergence of density and failure rate estimators under twice censoring. Statist. Probab. Lett. *106* (2015), 121–128. DOI:10.1016/j.na.2015.07.006

[5] M. Boukeloua: Testing local hypotheses with different types of incomplete data. TWMS Journal of Applied and Engineering Mathematics *14* (2024), 1374–1389.

[6] M. Boukeloua and F. Messaci: Asymptotic normality of kernel estimators based upon incomplete data. J. Nonparametr. Statist. *28* (2016), 469–486. DOI:10.1080/10485252.2016.1164312

[7] M. Boukeloua, S. Toumi, N. Idiou, and F. Benatia: Kernel estimation of the copula function under twice censoring. Stud. Engrg. Exact Sci. *5* (2024), e12036 DOI:10.54021/seesv5n2-763

[8] A. Bulinski and A. Shashkin: Limit Theorems for Associated Random Fields and Relayed Systems. World Scientific, Singapore 2007.

[9] Z. Cai and G. G. Roussas: Kaplan-Meier estimator under association. J. Multivar. Analysis *67* (1998), 318–348. DOI:10.1006/jmva.1998.1769

[10] I. Dewan and B. L. S. Prakasa Rao: A general method of density estimation for associated random variables. J. Nonparametr. Statist. *10* (1999), 405–420. hrefhttps://doi.org/10.1080/10485259908832769DOI:10.1080/10485259908832769

[11] I. Dewan and B. L. S. Prakasa Rao: Non-uniform and uniform Berry-Esseen type bounds for stationary associated sequences. J. Nonparametr. Statist. *17* (2005), 217–235. DOI:10.1080/1048525042000267734

[12] P. Doukhan and S. Louhichi: A new weak dependence condition and applications to moment inequalities. Stochast. Processes Appl. *84* (1999), 313–342. DOI:10.1016/S0304-4149(99)00055-1

[13] Y. Ferrani: Sur l'estimation non paramétrique de la densité et du mode dans les modèles de données incomplètes et associées. Ph.D. Thesis, Université du Littoral Côte d'Opale France; Université des Sciences et de la Technologie Houari-Boumediene Algeria 2014.

[14] Y. Ferrani, E. Ould Saïd, and A. Tatachak: On kernel density and mode estimates for associated and censored data. Commun. Statist. Theory Methods *45* (2016), 1853–1862. DOI:10.1080/03610926.2013.867996

[15] O. Julià and G. Gómez: Simultaneous marginal survival estimators when doubly censored data is present. Lifetime Data Analysis *17* (2011), 347–372. DOI:10.1007/s10985-010-9186-5

[16] A. Kitouni, M. Boukeloua, and F. Messaci: Rate of strong consistency for nonparametric estimators based on twice censored data. Statist. Probab. Lett. *96* (2015), 255–261. DOI:10.1016/j.spl.2014.10.005

[17] A. W. Marshall and I. Olkin: A multivariate exponential distribution. J. Amer. Statist. Assoc. *62* (1967), 30–44. DOI:10.1002/j.2330-9474.1967.tb03356.x

[18] V. Patilea and J. M. Rolin: Product-limit estimators of the survival function with twice censored data. Ann. Statist. *34* (2006), 925–938. DOI:10.1214/009053606000000065

[19] L. D. Pitt: Positively correlated normal variables are associated. Ann. Probab. *10* (1982), 496–499. DOI:10.1214/aop/1176993872

[20] G. G. Roussas: Kernel estimates under association: Strong uniform consistency. Statist. Prob. Lett. *12* (1991), 393–403. DOI:10.1016/0167-7152(91)90028-P

[21] G. G. Roussas: Asymptotic normality of the kernel estimate of a probability density function under association. Statist. Probab. Lett. *50* (2000), 1–12. DOI:10.1016/s0167-7152(00)00072-9

[22] F. Saihi, M. Boukeloua, and S. Leulmi: Non parametric estimation in the presence of associated and twice censored data. Commun.Statist. Theory Methods *54* (2025), 6536–6561. DOI:10.1080/03610926.2025.2458206

[23] S. Toumi, M. Boukeloua, N. Idiou, and F. Benatia: Nonparametric estimation of the copula function with bivariate twice censored data. Boletim da Sociedade Paranaense de Matemática *42* (2024), 1–22. DOI:10.5269/bspm.65601

[24] H. Yu: A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences. Probab. Theory Related Fields *95* (1993), 357–370. DOI:10.1007/BF01192169

A. PROOFS

Let $x \in]I_X, \min(T_X, T_R)[$. In all the sequel, C stands for a generic constant and ε stands for a positive real number such that $[x-\varepsilon, x+\varepsilon] \subset]I_X, \min(T_X, T_R)[$. Moreover, we denote by

$$Lip(K) = \sup_{u \neq v} \frac{|K(u) - K(v)|}{|u - v|}$$

the Lipschitz modulus of continuity of K.

P r o o f . [Proof of Theorem 2.1] We make use of the following decomposition

$$\begin{aligned}\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right) &= \sqrt{nh_n}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right) + \sqrt{nh_n}\left(\widetilde{f}_n(x) - E\left(\widetilde{f}_n(x)\right)\right) \\ &\quad + \sqrt{nh_n}\left(E\left(\widetilde{f}_n(x)\right) - f(x)\right). \end{aligned} \tag{1}$$

From the proof of Lemma 3.2 in [22], we can deduce that

$$\widehat{f}_n(x) - \widetilde{f}_n(x) = O_{a.s.}\left(\sqrt{\frac{(\log(n))^{1+2\beta}}{n}}\right).$$

So,

$$\sqrt{nh_n}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right) = O_{a.s.}\left(\sqrt{h_n(\log(n))^{1+2\beta}}\right) = o_{a.s.}(1), \tag{2}$$

because

$$\sqrt{h_n(\log(n))^{1+2\beta}} = \sqrt{n^{1/5}h_n \times \frac{(\log(n))^{1+2\beta}}{n^{1/5}}} = o(1),$$

in view of **H6** .
Moreover, from Lemma 3.3 in the same paper, we deduce that

$$\sqrt{nh_n}\left(E\left(\widetilde{f}_n(x)\right) - f(x)\right) = O\left(\sqrt{nh_n^5}\right) = o(1), \tag{3}$$

thanks to **H6**.
It remains to deal with the term $\sqrt{nh_n}\left(\widetilde{f}_n(x) - E\left(\widetilde{f}_n(x)\right)\right)$. For that, we proceed as in Theorem 2.1 of [21] and Theorem 5.1 of [13].
We have

$$\sqrt{nh_n}\left(\widetilde{f}_n(x) - E\left(\widetilde{f}_n(x)\right)\right) = \sum_{i=1}^{n} \xi_{i,n}(x),$$

where

$$\xi_{i,n}(x) \quad := \quad \frac{1}{\sqrt{nh_n}}\eta_{i,n}(x) := \frac{1}{\sqrt{nh_n}}\left(K\left(\frac{x - Z_i}{h_n}\right)\frac{1_{\{A_i=0\}}}{F_L\left(Z_i\right)S_R\left(Z_i\right)}\right.$$

$$-E\left(K\left(\frac{x-Z_i}{h_n}\right)\frac{1_{\{A_i=0\}}}{F_L\left(Z_i\right)S_R\left(Z_i\right)}\right)\Bigg).$$

Since $k_n = \left\lfloor \frac{n}{a_n+b_n} \right\rfloor$, we have $k_n(a_n+b_n) \leq n$, so we can split this sum as follows.

$$\sqrt{nh_n}\left(\widetilde{f}_n(x) - E\left(\widetilde{f}_n(x)\right)\right) = S_{1,n}(x) + S_{2,n}(x) + S_{3,n}(x), \tag{4}$$

where

$$S_{1,n}(x) = \sum_{m=1}^{k_n} V_m(x), \text{ with } V_m(x) = \sum_{j=(m-1)(a_n+b_n)+1}^{(m-1)(a_n+b_n)+a_n} \xi_{j,n}(x),$$

$$S_{2,n}(x) = \sum_{m=1}^{k_n} W_m(x), \text{ with } W_m(x) = \sum_{j=(m-1)(a_n+b_n)+a_n+1}^{m(a_n+b_n)} \xi_{j,n}(x)$$

and

$$S_{3,n}(x) = \sum_{l=k_n(a_n+b_n)+1}^{n} \xi_{l,n}(x).$$

We will deal with each sum in turn. For that, we need the following lemmas.

Lemma A.1. Under assumptions **H3, H5, H7**(i) and **H7**(ii), we have

a) $\lim_{n\to\infty} \frac{1}{h_n} Var\left(\eta_{1,n}(x)\right) = \sigma^2(x)$.

b) $\left|\mathrm{Cov}\left(\eta_{i,n}(x), \eta_{j,n}(x)\right)\right| \leq Ch_n^2$, for all $i, j \in \mathbb{N}^*$, $i \neq j$.

c) $\lim_{n\to\infty} \frac{k_n}{nh_n} \sum_{1\leq i<j\leq a_n+b_n} \left|\mathrm{Cov}\left(\eta_{i,n}(x), \eta_{j,n}(x)\right)\right| = 0$.

P r o o f . [Proof of Lemma A.1] For parts *a*) and *b*), we refer to the proof of Lemma 3.1 in [22]. They respectively correspond to the terms $V(x)$ and $J_{1,n}(x)$ in this paper.
Concerning part *c*), using part *b*) we get

$$\frac{k_n}{nh_n} \sum_{1\leq i<j\leq a_n+b_n} \left|\mathrm{Cov}\left(\eta_{i,n}(x), \eta_{j,n}(x)\right)\right| \leq \frac{k_n}{nh_n} \times Ch_n^2 (a_n+b_n)^2 = C \times \frac{k_n(a_n+b_n)}{n} \times (a_n+b_n)h_n. \tag{5}$$

Moreover, we have $k_n = \left\lfloor \frac{n}{a_n+b_n} \right\rfloor$, so

$$\lim_{n\to\infty} \frac{k_n(a_n+b_n)}{n} = 1. \tag{6}$$

Furthermore, we have

$$\begin{aligned} b_nh_n &= \frac{k_nb_n}{n} \times \frac{n}{k_n(a_n+b_n)} \times (a_n+b_n)h_n \\ \Longrightarrow &\left(1 - \frac{k_nb_n}{n} \times \frac{n}{k_n(a_n+b_n)}\right) b_nh_n = \frac{k_nb_n}{n} \times \frac{n}{k_n(a_n+b_n)} \times a_nh_n \\ \Longrightarrow\ & b_nh_n = \frac{\frac{k_nb_n}{n} \times \frac{n}{k_n(a_n+b_n)}}{1 - \frac{k_nb_n}{n} \times \frac{n}{k_n(a_n+b_n)}} \times a_nh_n \end{aligned}$$

Since $\lim_{n\to\infty} \frac{n}{k_n(a_n+b_n)} = 1$, $\lim_{n\to\infty} \frac{k_nb_n}{n} = 0$ and $\lim_{n\to\infty} a_nh_n = 0$ (thanks to assumptions **H7**(i) and **H7**(ii)), we deduce that

$$\lim_{n\to\infty} b_nh_n = 0. \tag{7}$$

Combining this with (5), (6) and assumption **H7**(ii), we obtain the claimed result. □

Lemma A.2. Under assumptions **H3, H5** and **H7**, we have

a) $\lim_{n\to\infty} k_n Var(V_1(x)) = \sigma^2(x)$.

b) For n large enough, we have

$$|\mathrm{Cov}\,(V_1(x), V_{l+1}(x))| \leq \left[\frac{Lip(K)}{F_L(x-\varepsilon)S_R(x+\varepsilon)}\right]^2 \frac{a_n}{nh_n^3} \sum_{r=l(a_n+b_n)-(a_n+1)}^{l(a_n+b_n)+(a_n-1)} |\mathrm{Cov}(X_1, X_{r+1})|\,,$$

for all $l \in \mathbb{N}^*$.

c) $\lim_{n\to\infty} \sum_{1\leq i<j\leq k_n} |\mathrm{Cov}\,(V_i(x), V_j(x))| = 0$.

d) $\lim_{n\to\infty} Var\,(S_{1,n}(x)) = \sigma^2(x)$.

P r o o f. [Proof of Lemma A.2]

a) The stationarity of the variables $\xi_{i,n}(x)$ allows to write

$$\begin{aligned} k_n Var(V_1(x)) &= k_n Var\left(\sum_{j=1}^{a_n} \xi_{j,n}(x)\right) \\ &= \frac{k_n}{nh_n} Var\left(\sum_{j=1}^{a_n} \eta_{j,n}(x)\right) \\ &= \frac{k_n a_n}{nh_n} Var(\eta_{1,n}(x)) + \frac{2k_n}{nh_n} \sum_{1\leq i<j\leq a_n} \mathrm{Cov}\,(\eta_{i,n}(x), \eta_{j,n}(x))\,. \end{aligned} \tag{8}$$

On the one hand, using assumption **H7**(i) and relation (6), we get

$$\lim_{n\to\infty} \frac{k_n a_n}{n} = \lim_{n\to\infty} \left(\frac{k_n(a_n+b_n)}{n} - \frac{k_n b_n}{n}\right) = 1. \tag{9}$$

On the other hand, Lemma A.1-b) gives

$$\begin{aligned} \frac{k_n}{nh_n} \sum_{1\leq i<j\leq a_n} |\mathrm{Cov}\,(\eta_{i,n}(x), \eta_{j,n}(x))| &\leq \frac{k_n a_n^2}{nh_n} \times Ch_n^2 \\ &\leq C \times \frac{k_n a_n}{n} \times a_n h_n \underset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

thanks to assumption **H7**(ii) and relation (9).
Combining this with Lemma A.1-a) and relations (8) and (9), we obtain the claimed result.

b) We have

$$\begin{aligned} |\mathrm{Cov}(V_1(x), V_{l+1}(x))| &= \left|\mathrm{Cov}\left(\sum_{i=1}^{a_n} \xi_{i,n}(x), \sum_{j=l(a_n+b_n)+1}^{l(a_n+b_n)+a_n} \xi_{j,n}(x)\right)\right| \\ &\leq \sum_{i=1}^{a_n} \sum_{j=l(a_n+b_n)+1}^{l(a_n+b_n)+a_n} |\mathrm{Cov}\,(\xi_{i,n}(x), \xi_{j,n}(x))| \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{a_n} \sum_{j=l(a_n+b_n)+1}^{l(a_n+b_n)+a_n} \left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{j-i+1,n}(x)\right)\right| \\
&\text{(thanks to the stationarity of the } (\xi_{i,n}(x))_{1\le i\le n}) \\
&= a_n \left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{l(a_n+b_n)+1,n}(x)\right)\right| \\
&+ (a_n-1)\left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{l(a_n+b_n)+2,n}(x)\right)\right| \\
&+ (a_n-2)\left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{l(a_n+b_n)+3,n}(x)\right)\right| \\
&+ \cdots + \left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{l(a_n+b_n)+a_n,n}(x)\right)\right| \\
&+ (a_n-1)\left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{l(a_n+b_n),n}(x)\right)\right| \\
&+ (a_n-2)\left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{l(a_n+b_n)-1,n}(x)\right)\right| \\
&+ \cdots + \left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{l(a_n+b_n)-a_n,n}(x)\right)\right| \\
&\le a_n \sum_{s=l(a_n+b_n)-a_n}^{l(a_n+b_n)+a_n} \left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{s,n}(x)\right)\right| \\
&= a_n \sum_{r=l(a_n+b_n)-(a_n+1)}^{l(a_n+b_n)+a_n-1} \left|\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{r+1,n}(x)\right)\right| \\
&\text{(using the change of indices } r = s-1) \\
&= \frac{a_n}{nh_n} \sum_{r=l(a_n+b_n)-(a_n+1)}^{l(a_n+b_n)+a_n-1} \left|\mathrm{Cov}\left(\eta_{1,n}(x), \eta_{r+1,n}(x)\right)\right|. \qquad (10)
\end{aligned}
$$

Moreover, we have

$$
\left|\mathrm{Cov}\left(\eta_{1,n}(x), \eta_{r+1,n}(x)\right)\right| = \left|\mathrm{Cov}\left(\widetilde{K}_1(X_1), \widetilde{K}_{r+1}(X_{r+1})\right)\right|,
$$

where

$$
\widetilde{K}_i(u) = K\left(\frac{x-u}{h_n}\right)\frac{1_{\{L_i<u\le R_i\}}}{F_L(u)S_R(u)}, i \in \{1,\ldots,n\}.
$$

So, Theorem 5.3 page 89 of [8] permits to write

$$
\left|\mathrm{Cov}\left(\eta_{1,n}(x), \eta_{r+1,n}(x)\right)\right| \le Lip(\widetilde{K}_1)Lip(\widetilde{K}_{r+1})\left|\mathrm{Cov}\left(X_1, X_{r+1}\right)\right|.
$$

Denote by $[-M, M]$ the support of K. We have $\widetilde{K}_i(u) = 0$ if $\frac{x-u}{h_n} \notin [-M, M]$, so

$$
-M \le \frac{x-u}{h_n} \le M \Longrightarrow x - Mh_n \le u \le x + Mh_n \Longrightarrow x-\varepsilon \le u \le x+\varepsilon,
$$

for n large enough.
Thus

$$
\begin{aligned}
&Lip(\widetilde{K}_i) \le \frac{Lip(K)}{F_L(x-\varepsilon)S_R(x+\varepsilon)h_n}, \forall i \in \{1,\ldots,n\} \\
&\Longrightarrow \left|\mathrm{Cov}\left(\eta_{1,n}(x), \eta_{r+1,n}(x)\right)\right| \le \left(\frac{Lip(K)}{F_L(x-\varepsilon)S_R(x+\varepsilon)h_n}\right)^2 \left|\mathrm{Cov}\left(X_1, X_{r+1}\right)\right|.
\end{aligned}
$$

Combining this with (10), we obtain the claimed result.

c) Using part b) and the stationarity of the variables $\xi_{i,n}(x)$, we get for n large enough

$$\begin{aligned}\sum_{1\le i<j\le k_n} |\mathrm{Cov}\,(V_i(x),V_j(x))| &= \sum_{l=1}^{k_n-1}(k_n-1)\,|\mathrm{Cov}\,(V_1(x),V_{l+1}(x))|\\ &\le k_n\sum_{l=1}^{k_n-1}|\mathrm{Cov}\,(V_1(x),V_{i+1}(x))|\\ &\le \left[\frac{Lip(K)}{F_L(x-\varepsilon)S_R(x+\varepsilon)}\right]^2\frac{k_n a_n}{nh_n^3}\times\\ &\sum_{l=1}^{k_n-10}\ \sum_{r=l(a_n+b_n)-(a_n+1)}^{l(a_n+b_n)+(a_n-1)}|\mathrm{Cov}(X_1,X_{r+1})|\\ &\le \left[\frac{Lip(K)}{F_L(x-\varepsilon)S_R(x+\varepsilon)}\right]^2\frac{k_n a_n}{nh_n^3}\sum_{r=b_n-1}^{\infty}|\mathrm{Cov}(X_1,X_{r+1})|\underset{n\to\infty}{\longrightarrow}0,\end{aligned}$$

thanks to assumption **H7**(iii) and relation (9).

d) Using parts a) and c), we get

$$\begin{aligned}\lim_{n\to\infty} Var\,(S_{1,n}(x)) &= \lim_{n\to\infty} Var\left(\sum_{m=1}^{k_n}V_m(x)\right)\\ &= \lim_{n\to\infty}\left(k_n Var(V_1(x)) + 2\sum_{1\le i<j\le n}\mathrm{Cov}(V_i(x),V_j(x))\right)\\ &= \sigma^2(x).\end{aligned}$$

□

Lemma A.3. Under assumptions **H3, H5** and **H7**, we have

a) $\lim_{n\to\infty} k_n Var\,(W_1(x)) = 0.$

b) For n large enough, we have

$$|\mathrm{Cov}\,(W_1(x),W_{l+1}(x))| \le \left[\frac{Lip(K)}{F_L(x-\varepsilon)S_R(x+\varepsilon)}\right]^2\frac{b_n}{nh_n^3}\sum_{r=l(a_n+b_n)-(b_n+1)}^{l(a_n+b_n)+(b_n-1)}|\mathrm{Cov}(X_1,X_{r+1})|,$$

for all $l\in\mathbb{N}^*$.

c) $\lim_{n\to\infty}\sum_{1\le i<j\le k_n}|\mathrm{Cov}\,(W_i(x),W_j(x))| = 0.$

d) $\lim_{n\to\infty} Var\,(S_{2,n}(x)) = 0.$

e) $\lim_{n\to\infty} Var\,(S_{3,n}(x)) = 0.$

P r o o f. [Proof of Lemma A.3]

a) Proceeding as in Lemma A.2-a), we get

$$k_n Var(W_1(x)) = \frac{k_n b_n}{n}\times\frac{1}{h_n}Var(\eta_{1,n}(x)) + \frac{2k_n}{nh_n}\sum_{1\le i<j\le b_n}\mathrm{Cov}\,(\eta_{i,n}(x),\eta_{j,n}(x))$$

$$\leq \frac{k_n b_n}{n} \times \frac{1}{h_n} Var(\eta_{1,n}(x)) + 2C \times \frac{k_n b_n}{n} \times b_n h_n.$$

This together with Lemma A.1-a), assumption **H7**(i) and relation (7), gives the result we seek.

b) We have

$$\begin{aligned}
|\mathrm{Cov}(W_1(x), W_{l+1}(x))| &= \left| \mathrm{Cov} \left(\sum_{i=a_n+1}^{a_n+b_n} \xi_{i,n}(x), \sum_{j=l(a_n+b_n)+a_n+1}^{(l+1)(a_n+b_n)} \xi_{j,n}(x) \right) \right| \\
&\leq \sum_{s=1}^{b_n} \sum_{j=l(a_n+b_n)+a_n+1}^{(l+1)(a_n+b_n)} |\mathrm{Cov}\left(\xi_{a_n+s,n}(x), \xi_{j,n}(x)\right)| \\
&\quad \text{(using the change of indices } s = i - a_n\text{)} \\
&= \sum_{s=1}^{b_n} \sum_{j=l(a_n+b_n)+a_n+1}^{(l+1)(a_n+b_n)} |\mathrm{Cov}\left(\xi_{1,n}(x), \xi_{j-a_n-s+1,n}(x)\right)|,
\end{aligned}$$

thanks to the stationarity of the variables $\xi_{i,n}(x)$.
So, proceeding as in Lemma A.2-b), we obtain for n large enough

$$\begin{aligned}
|\mathrm{Cov}(W_1(x), W_{l+1}(x))| &\leq \frac{b_n}{nh_n} \sum_{r=l(a_n+b_n)-(b_n+1)}^{l(a_n+b_n)+(b_n-1)} |\mathrm{Cov}\left(\eta_{1,n}(x), \eta_{r+1,n}(x)\right)| \\
&\leq \left[\frac{Lip(K)}{F_L(x-\varepsilon) S_R(x+\varepsilon)} \right]^2 \frac{b_n}{nh_n^3} \sum_{r=l(a_n+b_n)-(b_n+1)}^{l(a_n+b_n)+(b_n-1)} |\mathrm{Cov}(X_1, X_{r+1})|.
\end{aligned}$$

c) Using part b) and proceeding as in Lemma A.2-c), we obtain for n large enough

$$\sum_{1 \leq i < j \leq k_n} |\mathrm{Cov}\left(W_i(x), W_j(x)\right)| \leq \left[\frac{Lip(K)}{F_L(x-\varepsilon) S_R(x+\varepsilon)} \right]^2 \frac{k_n b_n}{nh_n^3} \sum_{r=a_n-1}^{\infty} |\mathrm{Cov}(X_1, X_{r+1})|.$$

On the other hand, we have in view of assumption **H7**(i) and relation (9)

$$\lim_{n\to\infty} \frac{b_n}{a_n} = \lim_{n\to\infty} \left[\frac{\frac{k_n b_n}{n}}{\frac{k_n a_n}{n}} \right] = 0 \Longrightarrow b_n = o(a_n).$$

So, for n large enough, we have

$$\sum_{1 \leq i < j \leq k_n} |\mathrm{Cov}\left(W_i(x), W_j(x)\right)| \leq \left[\frac{Lip(K)}{F_L(x-\varepsilon) S_R(x+\varepsilon)} \right]^2 \frac{k_n b_n}{nh_n^3} \sum_{r=b_n}^{\infty} |\mathrm{Cov}(X_1, X_{r+1})| \longrightarrow 0,$$

thanks to assumptions **H7**(i) and **H7**(iii).

d) Using parts a) and c), we get

$$\begin{aligned}
\lim_{n\to\infty} Var\left(S_{2,n}(x)\right) &= \lim_{n\to\infty} Var\left(\sum_{m=1}^{k_n} W_m(x) \right) \\
&= \lim_{n\to\infty} \left(k_n Var(W_1(x)) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(W_i(x), W_j(x)) \right) \\
&= 0.
\end{aligned}$$

e) We have

$$\begin{aligned}
Var\left(S_{3,n}(x)\right) &= Var\left(\sum_{l=k_n(a_n+b_n)+1}^{n}\xi_{l,n}(x)\right)\\
&= \frac{1}{nh_n}Var\left(\sum_{l=k_n(a_n+b_n)+1}^{n}\eta_{l,n}(x)\right)\\
&= \frac{n-k_n(a_n+b_n)}{nh_n}Var\left(\eta_{1,n}(x)\right)+\frac{2}{nh_n}\sum_{k_n(a_n+b_n)+1\le i<j\le n}\mathrm{Cov}\left(\eta_{i,n}(x),\eta_{j,n}(x)\right)\\
&= \frac{n-k_n(a_n+b_n)}{nh_n}Var\left(\eta_{1,n}(x)\right)+\frac{2}{nh_n}\sum_{1\le i<j\le n-k_n(a_n+b_n)}\mathrm{Cov}\left(\eta_{i,n}(x),\eta_{j,n}(x)\right),
\end{aligned}$$

thanks to the stationarity of the variables $\eta_{i,n}(x)$.
Moreover, since $k_n=\left\lfloor\frac{n}{a_n+b_n}\right\rfloor$, we have

$$\frac{n}{a_n+b_n}-1<k_n\le\frac{n}{a_n+b_n}\Longrightarrow n-(a_n+b_n)<k_n(a_n+b_n)\le n\Longrightarrow 0\le n-k_n(a_n+b_n)<a_n+b_n.$$

Thus

$$Var\left(S_{3,n}(x)\right)\le\frac{a_n+b_n}{n}\times\frac{1}{h_n}Var\left(\eta_{1,n}(x)\right)+\frac{1}{k_n}\times\frac{2k_n}{nh_n}\sum_{1\le i<j\le a_n+b_n}\left|\mathrm{Cov}\left(\eta_{i,n}(x),\eta_{j,n}(x)\right)\right|. \tag{11}$$

Furthermore, since $k_n\le\frac{n}{a_n+b_n}$ and $\lim_{n\to\infty}k_n=\infty$, we deduce that $\lim_{n\to\infty}\frac{n}{a_n+b_n}=\infty$, so $\lim_{n\to\infty}\frac{a_n+b_n}{n}=0$. This together with relation (11) and parts a) and c) of Lemma A.1, gives the result we seek.

□

Since the variables $\xi_{i,n}(x)$ are centered, we have $E\left(S_{2,n}(x)\right)=E\left(S_{3,n}(x)\right)=0$ and since $\lim_{n\to\infty}Var\left(S_{2,n}(x)\right)=\lim_{n\to\infty}Var\left(S_{3,n}(x)\right)=0$, we deduce that

$$S_{2,n}(x)\xrightarrow{P}0\ \text{ and }\ S_{3,n}(x)\xrightarrow{P}0,\ \text{ as }\ n\to\infty, \tag{12}$$

where $\xrightarrow{P}$ denotes the convergence in probability.
It remains to establish the asymptotic normality of $S_{1,n}(x)$.
Denote by φ_n (resp. φ_{V_m}) the characteristic function of $S_{1,n}(x)$ (resp. $V_m(x)$). For all $t\in\mathbb{R}$, we have

$$\begin{aligned}
I_{k_n}(t) &:= \left|\varphi_n(t)-\prod_{m=1}^{k_n}\varphi_{V_m}(t)\right|\\
&= \left|E\left[\exp\left\{it\sum_{m=1}^{k_n}V_m(x)\right\}\right]-\prod_{m=1}^{k_n}E\left[\exp\left\{itV_m(x)\right\}\right]\right|\\
&= \left|E\left[\exp\left\{it\sum_{m=1}^{k_n}V_m(x)\right\}\right]-\prod_{m=1}^{k_n-1}E\left[\exp\left\{itV_m(x)\right\}\right]\times E\left[e^{itV_{k_n}(x)}\right]\right|\\
&\le \left|E\left[\exp\left\{it\sum_{m=1}^{k_n}V_m(x)\right\}\right]-E\left[\exp\left\{it\sum_{m=1}^{k_n-1}V_m(x)\right\}\right]\times E\left[e^{itV_{k_n}(x)}\right]\right|
\end{aligned}$$

$$
\begin{aligned}
&+\left|E\left[\exp\left\{it\sum_{m=1}^{k_n-1}V_m(x)\right\}\right]\times E\left[e^{itV_{k_n}(x)}\right]-\prod_{m=1}^{k_n-1}E\left[\exp\left\{itV_m(x)\right\}\right]\times E\left[e^{itV_{k_n}(x)}\right]\right| \\
&\leq\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-1}V_m(x)\right\},e^{itV_{k_n}(x)}\right)\right| \\
&+\left|E\left[e^{itV_{k_n}(x)}\right]\right|\left|E\left[\exp\left\{it\sum_{m=1}^{k_n-1}V_m(x)\right\}\right]-\prod_{m=1}^{k_n-1}E\left[\exp\left\{itV_m(x)\right\}\right]\right| \\
&\leq I_{k_n-1}(t)+\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-1}V_m(x)\right\},e^{itV_{k_n}(x)}\right)\right|.
\end{aligned}
$$

In the same way, we get

$$
I_{k_n-1}(t)\leq I_{k_n-2}(t)+\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-2}V_m(x)\right\},e^{itV_{k_n-1}(x)}\right)\right|.
$$

So,

$$
\begin{aligned}
I_{k_n}(t)&\leq I_{k_n-2}(t)+\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-2}V_m(x)\right\},e^{itV_{k_n-1}(x)}\right)\right| \\
&\quad+\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-1}V_m(x)\right\},e^{itV_{k_n}(x)}\right)\right| \\
&\leq I_{k_n-3}(t)+\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-3}V_m(x)\right\},e^{itV_{k_n-2}(x)}\right)\right| \\
&\quad+\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-2}V_m(x)\right\},e^{itV_{k_n-1}(x)}\right)\right| \\
&\quad+\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{k_n-1}V_m(x)\right\},e^{itV_{k_n}(x)}\right)\right| \\
&\leq \\
&\vdots \\
&\leq I_1(t)+\sum_{j=2}^{k_n}\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{j-1}V_m(x)\right\},e^{itV_j(x)}\right)\right| \\
&=\sum_{j=2}^{k_n}\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{j-1}V_m(x)\right\},e^{itV_j(x)}\right)\right|, \qquad (13)
\end{aligned}
$$

because $I_1(t)=0$.
Using Corollary 5.5 page 90 of [8], we obtain for all $j\in\{2,\ldots,k_n\}$

$$
\begin{aligned}
\left|\mathrm{Cov}\left(\exp\left\{it\sum_{m=1}^{j-1}V_m(x)\right\},e^{itV_j(x)}\right)\right|&=\left|E\left[\exp\left\{it\sum_{m=1}^{j}V_m(x)\right\}\right]-E\left[\exp\left\{it\sum_{m=1}^{j-1}V_m(x)\right\}\right]\times\right. \\
&\left.\times E\left[e^{itV_j(x)}\right]\right|
\end{aligned}
$$

$$\begin{aligned}
&\leq 4t^2 \left| \mathrm{Cov}\left(\sum_{m=1}^{j-1} V_m(x), V_j(x) \right) \right| \\
&\leq 4t^2 \sum_{m=1}^{j-1} \left| \mathrm{Cov}\left(V_m(x), V_j(x) \right) \right| \\
&\leq 4t^2 \sum_{m=1}^{j-1} \left| \mathrm{Cov}\left(V_1(x), V_{j-m+1}(x) \right) \right| \\
&\text{(thanks to the stationarity of the } (V_j(x))_{1\leq j\leq k_n}\text{)} \\
&\leq 4t^2 \sum_{l=1}^{j-1} \left| \mathrm{Cov}\left(V_1(x), V_{l+1}(x) \right) \right|,
\end{aligned}$$

using the change of indices $l = j - m$.
This together with (13) entails that

$$\begin{aligned}
I_{k_n}(t) &\leq 4t^2 \sum_{j=2}^{k_n} \sum_{l=1}^{j-1} \left| \mathrm{Cov}\left(V_1(x), V_{l+1}(x) \right) \right| \\
&\leq 4t^2 \left[(k_n - 2) \left| \mathrm{Cov}\left(V_1(x), V_2(x) \right) \right| + (k_n - 1) \left| \mathrm{Cov}\left(V_1(x), V_3(x) \right) \right| + \cdots + \left| \mathrm{Cov}\left(V_1(x), V_{k_n}(x) \right) \right| \right] \\
&\leq 4t^2 k_n \sum_{m=2}^{k_n} \left| \mathrm{Cov}\left(V_1(x), V_m(x) \right) \right| \\
&\leq 4t^2 k_n \left[\frac{Lip(K)}{F_L(x-\varepsilon) S_R(x+\varepsilon)} \right]^2 \frac{a_n}{n h_n^3} \sum_{m=2}^{k_n} \sum_{r=(m-1)(a_n+b_n)-(a_n+1)}^{(m-1)(a_n+b_n)+(a_n-1)} \left| \mathrm{Cov}(X_1, X_{r+1}) \right| \\
&\text{(for } n \text{ large enough, thanks to Lemma A.2-b))} \\
&\leq 4t^2 \left[\frac{Lip(K)}{F_L(x-\varepsilon) S_R(x+\varepsilon)} \right]^2 \frac{k_n a_n}{n h_n^3} \sum_{r=b_n-1}^{\infty} \left| \mathrm{Cov}(X_1, X_{r+1}) \right| \underset{n\to\infty}{\longrightarrow} 0,
\end{aligned}$$

in view of assumption **H7**(iii) and relation (9).
Let $(U_n(x))_{1\leq m\leq k_n}$ be a sequence of i.i.d. r.r.v. having the same distribution as $V_1(x)$. From the previous relation and the independence of the variables $U_m(x)$, we have

$$\lim_{n\to\infty} \left| \varphi_n(t) - E\left[\exp\left\{ it \sum_{m=1}^{k_n} U_m(x) \right\} \right] \right| = \lim_{n\to\infty} \left| \varphi_n(t) - \prod_{m=1}^{k_n} E\left[\exp\left\{ itU_m(x) \right\} \right] \right| = 0, \ \forall t \in \mathbb{R}. \tag{14}$$

Now, we will show that the sum $\sum_{m=1}^{k_n} U_m(x)$ converges in distribution to the $\mathcal{N}\left(0, \sigma^2(x)\right)$ distribution.
Remark that the variables $U_m(x)$ are centered because the variables $\xi_{i,n}(x)$ are centered. Moreover, we set

$$s_n^2(x) := \sum_{m=1}^{k_n} Var\left(U_m(x) \right) = k_n Var\left(U_1(x) \right) = k_n Var\left(V_1(x) \right). \tag{15}$$

We will apply the Lindeberg theorem (see [3], Theorem 27.2 page 359) to $U_m(x)$. For that, we will show that for all $\delta > 0$

$$\lim_{n\to\infty} \frac{1}{s_n^2(x)} \sum_{m=1}^{k_n} E\left[U_m^2(x) 1_{\{|U_m(x)| > \delta s_n(x)\}} \right] = 0.$$

We have

$$\begin{aligned}|V_1(x)| &= \left|\sum_{j=1}^{a_n}\xi_{j,n}(x)\right| \\ &\leq \frac{1}{\sqrt{nh_n}}\sum_{j=1}^{a_n}|\eta_{j,n}(x)| \\ &\leq \frac{1}{\sqrt{nh_n}}\sum_{j=1}^{a_n}\left|K\left(\frac{x-Z_j}{h_n}\right)\frac{1_{\{A_j=0\}}}{F_L\left(Z_j\right)S_R\left(Z_j\right)} - E\left(K\left(\frac{x-Z_j}{h_n}\right)\frac{1_{\{A_j=0\}}}{F_L\left(Z_j\right)S_R\left(Z_j\right)}\right)\right|.\end{aligned}$$

Since the support of K is $[-M,M]$, we have

$$-M \leq \frac{x-Z_i}{h_n} \leq M \Longrightarrow x-\varepsilon \leq Z_i \leq x+\varepsilon,$$

for n large enough.
So,

$$|V_1(x)| \leq \frac{2\|K\|_\infty}{F_L(x-\varepsilon)S_R(x+\varepsilon)} \times \frac{a_n}{\sqrt{nh_n}},\ a.s.$$

where $\|K\|_\infty = \sup_{u\in\mathbb{R}}|K(u)|$.
Therefore

$$|U_m(x)| \leq \frac{2\|K\|_\infty}{F_L(x-\varepsilon)S_R(x+\varepsilon)} \times \frac{a_n}{\sqrt{nh_n}},\ a.s.$$

and

$$\begin{aligned}\frac{1}{s_n^2(x)}\sum_{m=1}^{k_n}E\left[U_m^2(x)1_{\{|U_m(x)|>\delta s_n(x)\}}\right] &\leq \frac{4\|K\|_\infty^2}{F_L^2(x-\varepsilon)S_R^2(x+\varepsilon)} \times \frac{a_n^2}{nh_ns_n^2(x)}\sum_{m=1}^{k_n}P\left(|U_m(x)|>\delta s_n(x)\right) \\ &= \frac{4\|K\|_\infty^2}{F_L^2(x-\varepsilon)S_R^2(x+\varepsilon)} \times \frac{k_na_n^2}{nh_ns_n^2(x)}P\left(|U_1(x)|>\delta s_n(x)\right) \\ &\leq \frac{4\|K\|_\infty^2}{F_L^2(x-\varepsilon)S_R^2(x+\varepsilon)} \times \frac{k_na_n^2}{nh_ns_n^2(x)} \times \frac{Var\left(U_1(x)\right)}{\delta^2 s_n^2(x)},\end{aligned}$$

thanks to the Chebyshev-Bienaymé inequality (see [3]).
Taking into account relation (15), we get

$$\begin{aligned}\frac{1}{s_n^2(x)}\sum_{m=1}^{k_n}E\left[U_m^2(x)1_{\{|U_m(x)|>\delta s_n(x)\}}\right] &\leq \frac{4\|K\|_\infty^2}{\delta^2F_L^2(x-\varepsilon)S_R^2(x+\varepsilon)} \times \frac{k_nVar(V_1(x))a_n^2}{nh_nk_nVar(V_1(x))} \times \frac{1}{k_nVar(V_1(x))} \\ &\leq \frac{4\|K\|_\infty^2}{\delta^2F_L^2(x-\varepsilon)S_R^2(x+\varepsilon)} \times \frac{a_n^2}{nh_n} \times \frac{1}{k_nVar(V_1(x))} \underset{n\to\infty}{\longrightarrow} 0,\end{aligned}$$

in view of Lemma A.2-a) and assumption **H7**(ii).
So, the Lindeberg condition is satisfied which implies that

$$\frac{1}{s_n(x)}\sum_{m=1}^{k_n}U_m(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),\ \text{ as } n\to\infty$$

and since

$$\lim_{n\to\infty}s_n^2(x) = \lim_{n\to\infty}k_nVar(V_1(x)) = \sigma^2(x)\ \ \text{(thanks to Lemma 2-a)},$$

we deduce that

$$\sum_{m=1}^{k_n} U_m(x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x)\right), \text{ as } n \to \infty. \tag{16}$$

Relations (14) and (16) imply that

$$S_{1,n}(x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x)\right), \text{ as } n \to \infty.$$

This together with relations (4) and (12) gives

$$\sqrt{nh_n}\left(\widetilde{f}_n(x) - E\left(\widetilde{f}_n(x)\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x)\right), \text{ as } n \to \infty.$$

Combining this with relations (1), (2) and (3), we obtain

$$\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x)\right), \text{ as } n \to \infty.$$

□

Proof. [Proof of Theorem 2.2]

i) We have

$$\sqrt{nh_n}\left(\widehat{\lambda}_n(x) - \lambda(x)\right) = \frac{\sqrt{nh_n}\left(\widehat{f}_n(x) - f(x)\right)}{\widehat{S}_n(x) + u_n} + \sqrt{nh_n}\left(S_X(x) - \widehat{S}_n(x) - u_n\right) \times \frac{f(x)}{S_X(x)\left(\widehat{S}_n(x) + u_n\right)}. \tag{17}$$

Moreover, from Theorem 2.1 of [22], we have

$$\begin{aligned}\sqrt{nh_n}\left(S_X(x) - \widehat{S}_n(x)\right) &= O_{a.s.}\left(\sqrt{nh_n} \times \sqrt{\frac{(\log(n))^{1+2\beta}}{n}}\right) \\ &= O_{a.s.}\left(\sqrt{n^{1/5}h_n \times \frac{(\log(n))^{1+2\beta}}{n^{1/5}}}\right) \\ &= o_{a.s.}(1),\end{aligned}$$

in view of **H6**.
Combining this with relation (17), Theorem 2.1 and the fact that $u_n = o\left(\frac{1}{\sqrt{nh_n}}\right)$, we obtain the claimed result.

ii) We have

$$\begin{aligned}\lambda_n^*(x) &= \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right) \mathrm{d}\widehat{\Lambda}_n(z) \\ &= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{x - Z_i}{h_n}\right) \Delta\widehat{\Lambda}_n(Z_i) \\ &= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - Z_i}{h_n}\right) \frac{1_{\{A_i=0\}}}{F_L^{(n)}\left(Z_i^-\right) - F_Z^{(n)}\left(Z_i^-\right)}.\end{aligned}$$

So, we set

$$\widetilde{\lambda}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\frac{1_{\{A_i=0\}}}{F_L\left(Z_i\right)-F_Z\left(Z_i\right)}$$
$$= \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-Z_i}{h_n}\right)\frac{1_{\{A_i=0\}}}{S_X\left(Z_i\right)F_L\left(Z_i\right)S_R\left(Z_i\right)}.$$

Using the decomposition

$$\sqrt{nh_n}\left(\lambda_n^*(x)-\lambda(x)\right) = \sqrt{nh_n}\left(\lambda_n^*(x)-\widetilde{\lambda}_n(x)\right)+\sqrt{nh_n}\left(\widetilde{\lambda}_n(x)-E\left(\widetilde{\lambda}_n(x)\right)\right)$$
$$+\sqrt{nh_n}\left(E\left(\widetilde{\lambda}_n(x)\right)-\lambda(x)\right),$$

this part can be proved following similar steps to those of Theorem 2.1 above.

□

B. THE AIDS DATA SET

The AIDS original data and our moving averages in cases 1,2 and 3.

Original data		Data in case 1		Data in case 2		Data in case 3	
Times	Censoring indicator	Z_i	A_i	Z_i	A_i	Z_i	A_i
3592	1	3498	1	4499.667	1	3955.25	1
3404	1	4953.5	1	4076.333	1	4872.25	1
6503	1	4412.5	1	5361.667	1	5312	1
2322	1	4791	1	4915	1	4657	1
7260	1	6211.5	1	5435.333	1	5420.5	1
5163	1	4523	1	4807.333	1	5190	1
3883	1	4629.5	1	5199	1	5345.5	1
5376	1	5857	1	5833	1	5830.5	1
6338	1	6061.5	1	5982	1	5372	1
5785	1	5804	1	5050	1	4425	1
5823	1	4682.5	1	3971.667	1	3387.25	1
3542	1	3046	1	2575.333	1	3531	1
2550	1	2092	1	3527.333	1	4151.75	1
1634	1	4016	1	4685.667	1	3974.75	1
6398	1	6211.5	1	4755	1	5446.5	1
6025	1	3933.5	1	5129.333	1	5558	1
1842	1	4681.5	1	5402.333	1	4719	1
7521	1	7182.5	1	5678	1	5543.75	1
6844	1	4756.5	1	4884.667	1	5191	1
2669	1	3905	1	4640	1	5334.25	1
5141	1	5625.5	1	6222.667	1	5515	1
6110	1	6763.5	1	5639.667	1	5938.75	1
7417	1	5404.5	1	5881.667	1	5207.75	1
3392	1	5114	1	4471.333	1	4247.5	1
6836	1	5011	1	4532.667	1	4487.25	1
3186	1	3381	1	3704.333	1	4333.5	1

Original data		Data in case 1		Data in case 2		Data in case 3	
Times	Censoring indicator	Z_i	A_i	Z_i	A_i	Z_i	A_i
3576	1	3963.5	1	4716	1	5156	1
4351	1	5286	1	5682.667	1	6020.75	1
6221	1	6348.5	1	6577.333	1	5293	1
6476	1	6755.5	1	4983.667	1	4789.25	1
7035	1	4237.5	1	4227	1	4731.5	1
1440	1	2823	1	3963.667	1	4445.5	1
4206	1	5225.5	1	5447.333	1	5034.25	1
6245	1	6068	1	5310.333	1	5238	1
5891	1	4843	1	4902.333	1	5156	1
3795	1	4408	1	4911	1	4599.25	1
5021	1	5469	1	4867.333	1	5261	1
5917	1	4790.5	1	5341	1	5628	1
3664	1	5053	1	5531.667	1	5575.25	1
6442	1	6465.5	1	6212.333	1	5889	1
6489	1	6097.5	1	5704.667	1	5883.75	1
5706	1	5312.5	1	5682	1	6173	1
4919	1	5670	1	6328.667	1	5479.5	1
6421	1	7033.5	1	5666.333	1	5718.75	1
7646	1	5289	1	5484.667	1	5663.25	1
2932	1	4404	1	5002.333	1	5104.75	1
5876	1	6037.5	1	5829	1	5787.5	1
6199	1	5805.5	1	5758	1	4823.5	1
5412	1	5537.5	1	4365	1	3955	1
5663	1	3841.5	1	3469.333	1	3981.5	1
2020	1	2372.5	1	3421	1	3977	1
2725	1	4121.5	1	4629.333	1	4965	1
5518	1	5581.5	1	5711.667	1	5855.5	1
5645	1	5808.5	1	5968	1	5617.25	1
5972	1	6129.5	1	5608	1	4625	1
6287	1	5426	1	4176	1	4396.5	1
4565	1	3120.5	1	3766.333	1	4270.5	1
1676	1	3367	1	4172.333	1	5016	1
5058	1	5420.5	1	6129.333	1	5392	1
5783	1	6665	1	5503.333	1	5548.5	1
7547	1	5363.5	1	5470.333	1	5493.5	1
3180	1	4432	1	4809	1	4094	1
5684	1	5623.5	1	4398.667	1	4656.5	1
5563	1	3756	1	4314	1	3900.75	1
1949	1	3689.5	1	3346.667	1	3263.25	1
5430	1	4045.5	1	3701.333	1	4059.75	1
2661	1	2837	1	3603	1	3746.25	1
3013	1	4074	1	4108	1	4277.25	1
5135	1	4655.5	1	4698.667	1	4913	1
4176	1	4480.5	1	4839	1	5106	1
4785	1	5170.5	1	5416	1	5830.5	1
5556	1	5731.5	1	6179	1	5776	1
5907	1	6490.5	1	5849.333	1	5629.5	1

Original data		Data in case 1		Data in case 2		Data in case 3	
Times	Censoring indicator	Z_i	A_i	Z_i	A_i	Z_i	A_i
7074	1	5820.5	1	5537	1	5265	1
4567	1	4768.5	1	4662	1	4816.75	1
4970	1	4709.5	1	4900	1	5201.5	1
4449	1	4865	1	5278.667	1	4410.5	1
5281	1	5693.5	1	4397.667	1	3523.5	1
6106	1	3956	1	2937.667	1	3348	1
1806	1	1353.5	1	2428.667	1	2242.5	1
901	1	2740	1	2388	1	2691.75	1
4579	1	3131.5	1	3288.667	1	4491	1
1684	1	2643.5	1	4461.667	1	4170.5	1
3603	1	5850.5	1	4999.333	1	4138	1
8098	1	5697.5	1	4316.333	1	4840.5	1
3297	1	2425.5	1	3754.667	1	3418.5	1
1554	1	3983.5	1	3459	1	3340.5	1
6413	1	4411.5	1	3936	1	3207.25	1
2410	1	2697.5	1	2138.667	1	2370	1
2985	1	2003	1	2356.667	1	2745.75	1
1021	1	2042.5	1	2666	1	3025	1
3064	1	3488.5	1	3693	1	4393.75	1
3913	1	4007.5	1	4837	1	4500.25	1
4102	1	5299	1	4696	1	4964.5	1
6496	1	4993	1	5252	1	4590	1
3490	1	4630	1	3954.667	1	4147	1
5770	1	4187	1	4366	1	3883	1
2604	1	3664	1	3254	1	3972.25	1
4724	1	3579	1	4428.333	1	4043	1
2434	1	4280.5	1	3816	1	3539.25	1
6127	1	4507	1	3907.667	1	4283	1
2887	1	2798	1	3668.333	1	4186.5	1
2709	1	4059	1	4619.667	1	4109.25	1
5409	1	5575	1	4576	1	3974	1
5741	1	4159.5	1	3495.667	1	4063.75	1
2578	1	2373	1	3504.667	1	4040.25	1
2168	1	3968	1	4527.667	1	4313.25	1
5768	1	5707.5	1	5028.333	1	4134.5	1
5647	1	4658.5	1	3590	1	3717.5	1
3670	1	2561.5	1	3074.333	1	3501	1
1453	1	2776.5	1	3444.667	1	3574.5	1
4100	1	4440.5	1	4281.667	1	4386.5	1
4781	1	4372.5	1	4482	1	4257.25	1
3964	1	4332.5	1	4082.667	1	4722.25	1
4701	1	4142	1	4975	1	5142.5	1
3583	1	5112	1	5289.667	1	4929.5	1
6641	1	6143	1	5378.333	1	5737	1
5645	1	4747	1	5435.667	1	5300.5	1
3849	1	5331	1	5185.667	1	5265	1
6813	1	5854	1	5737	1	4582.75	1

Original data		Data in case 1		Data in case 2		Data in case 3	
Times	Censoring indicator	Z_i	A_i	Z_i	A_i	Z_i	A_i
4895	1	5199	1	3839.333	1	3605.25	1
5503	1	3311.5	1	3175.333	1	3095	1
1120	1	2011.5	1	2292.333	1	2733.25	1
2903	1	2878.5	1	3271	1	2991.5	1
2854	1	3455	1	3021	1	3609.75	1
4056	1	3104.5	1	3861.667	1	3948.25	1
2153	1	3764.5	1	3912.333	1	4088.25	1
5376	1	4792	1	4733.333	1	4456.75	1
4208	1	4412	1	4150.333	1	4824.25	1
4616	1	4121.5	1	5029.667	1	5015.25	1
3627	1	5236.5	1	5148.333	1	4451.5	1
6846	1	5909	1	4726.333	1	4026	1
4972	1	3666.5	1	3086	1	3215.75	1
2361	1	2143	1	2630.333	1	3616.75	0
1925	1	2765	1	3690.333	0	3320.75	0
3605	1	4321	0	3495.667	0	2874.25	0
3980	0	3545.5	0	2873.667	0	3536.75	0
4662	0	2912.5	0	3022.667	0	3339.25	0
2429	0	3096	0	3583.667	0	3137.25	0
3396	0	2836	0	3520.333	0	3497.75	0
2796	0	3977.5	0	3224.333	0	2789.5	0
2876	0	3842.5	0	2970.667	0	3396	0
5079	0	2297	0	2850.667	0	4397.5	0
2606	0	3153	0	3865.333	0	4684	0
1988	0	3282	0	4424	0	4996.75	0
4318	0	3639	0	5496.667	0	4686	0
2246	0	5513	0	4985	0	3988.25	0
5032	0	5729	0	4250	0	3681.5	0
5994	0	4480.5	0	3496.333	0	3748.25	0
5464	0	3643	0	3743	0	4031	0
3497	0	3496	0	3734.667	0	4367	0
3789	0	3720	0	4307	0	4635.25	0
3203	0	4000.5	0	4410.333	0	4823.25	0
4237	0	4342	0	4925.667	0	5116.5	0
3764	0	4733.5	0	4791	0	5034.5	0
4920	0	4928.5	0	5306.333	0	4426.5	0
4547	0	4913	0	4942.667	0	3487.25	0
5310	0	5304.5	0	4396.667	0	3435	0
4516	0	5156	0	2618.667	0	3070.5	0
6093	0	3548.5	0	3173.667	0	3487	0
4219	0	1818.5	0	3134.667	0	4846.5	0
2878	0	3321.5	0	4396.333	0	4004.5	0
759	0	4322.5	0	4500.667	0	4755.25	0
5884	0	3652.5	0	4419	0	4200.25	0
2761	0	5370.5	0	4825.667	0	3972.75	0
4544	0	4356.5	0	3534.667	0	3896	0
6197	0	4140	0	4458.333	0	3605.5	0

Original data		Data in case 1		Data in case 2		Data in case 3	
Times	Censoring indicator	Z_i	A_i	Z_i	A_i	Z_i	A_i
2516	0	4044	0	3273.333	0	3375.5	0
5764	0	3805.5	0	4032.667	0	3143	0
2324	0	3748	0	2738.333	0	3278.5	0
5287	0	3405.5	0	3454.333	0	3542.75	0
2209	0	3003	0	2837.333	0	4040	0
4602	0	2880.5	0	4255.667	0	3831	0
1404	0	3554	0	3934.333	0	4072.25	0
4357	0	4205	0	4191	0	4031.25	0
2751	0	4526	0	3543.333	0	3763	0
5659	0	3457	0	4244	0	4044.25	0
3393	0	3618.5	0	3843.667	0	4354.5	0
3521	0	4605.5	0	4153.667	0	4113.25	0
3716	0	3907.5	0	3974.333	0	4637.25	0
5495	0	3483	0	4711	0	4456.75	0
2320	0	4801.5	0	4634.333	0	4253	0
4646	0	4743.5	0	4290	0	3693.75	0
4957	0	4473	0	4160.667	0	3491.5	0
4530	0	4170	0	3453	0	3236.75	0
4416	0	4033	0	3347.333	0	2952.75	0
3924	0	3217.5	0	2935	0	3151.75	0
4142	0	2950	0	3172.667	0	3228.5	0
2293	0	3256	0	3000	0	2945.25	0
3607	0	2955.5	0	3336.333	0	3002.5	0
2905	0	3047.5	0	2925	0	3578.75	0
3006	0	3501.5	0	2973.667	0	4100.5	0
3089	0	2843	0	3467	0	4148	0
3914	0	2503.5	0	4876.667	0	4748.25	0
1772	0	4314.5	0	4452.333	0	3845	0
3235	0	5697.5	0	4533	0	3186	0
5394	0	3981.5	0	3126.333	0	2812.25	0
6001	0	3799	0	3594	0	2106.5	0
1962	0	3708.5	0	1871	0	2739	0
5636	0	2573	0	2215	0	2791.25	0
1781	0	1916	0	2530.333	0	4221.75	0
3365	0	1640	0	3566	0	4695.75	0
467	0	3562	0	4691.333	0	4381.75	0
2813	0	3942.5	0	4824	0	4625.75	0
4311	0	4881.5	0	4651	0	3351.75	0
3574	0	5449	0	4104.667	0	2682.25	0
6189	0	3882	0	2899.333	0	2667.75	0
4709	0	3802.5	0	2558	0	2488	0
3055	0	2821.5	0	2040.3333	0	3333.5	0
4550	0	1562	0	2953	0	3861.5	0
1093	0	2514	0	3767.667	0	4115.75	0
2031	0	3414	0	4149.667	0	4096.25	0
2997	0	4153	0	4210.667	0	4250.5	2
3831	0	4309	0	3970	0	4398.5	2

Original data		Data in case 1		Data in case 2		Data in case 3	
Times	Censoring indicator	Z_i	A_i	Z_i	A_i	Z_i	A_i
4475	0	4078.5	0	4595.333	2	4334.5	2
4143	0	3883.5	0	4448.333	2	3822.5	2
4014	0	4359.5	2	4177.333	2	4638	2
3753	0	5064.5	2	4090.333	2	4782.75	2
3657	2	4141.5	2	4024.667	2	4009.75	2
5062	2	3732.5	2	4767.667	2	3740.25	2
5067	2	4527.5	2	4775	2	3190.5	2
3216	2	3912.5	2	4340	2	2531.75	2
4249	2	4748.5	2	2827.667	2	2758	2
4806	2	5653	2	2644.667	2		
3019	2	3271	2	2804.333	2		
6478	2	1827.5	2	3030.333	2		
4828	2	3110	2				
1714	2	3236	2				
1941	2	2406	2				
4279	2						
2193	2						
2619	2						

Tab. 3: The AIDS original data and our moving averages in cases 1,2,3.

Mohamed Boukeloua, Laboratoire de Génie des Procédés pour le Développement Durable et les Produits de Santé (LGPDDPS), Ecole Nationale Polytechnique de Constantine, Algeria and Laboratoire de Biostatistique, Bioinformatique et Méthodologie Mathématique Appliquées aux Sciences de la Santé (BIOSTIM), Faculté de Médecine, Université Salah Boubnider Constantine 3. Algeria.

e-mail: mohamed.boukeloua@enp-constantine.dz boukeloua.mohamed@gmail.com

Sarra Leulmi, Laboratoire de Mathématiques et Sciences de La Décision (LAMASD), Université Frères Mentouri Constantine 1. Algeria.

e-mail: sara.leulmi@umc.edu.dz

Ferial Saihi, Laboratoire Mathématiques Appliquées et Modélisation (MAM), Université Frères Mentouri Constantine 1. Algeria.

e-mail: ferial.saihi@doc.umc.edu.dz