REVSTAT – Statistical Journal
Volume 18, Number 3, July 2020, 337–355

ON KERNEL HAZARD RATE FUNCTION ESTIMATE FOR ASSOCIATED AND LEFT TRUNCATED DATA

Authors: Zohra Guessoum
– Laboratory MSTD, Department of Probability and Statistics,
University of Sciences and Technology Houari Boumediene (USTHB),
BP 32, El-Alia 16111, Algiers, Algeria
zguessoum@usthb.dz

Abdelkader Tatachak
– Laboratory MSTD, Department of Probability and Statistics,
University of Sciences and Technology Houari Boumediene (USTHB),
BP 32, El-Alia 16111, Algiers, Algeria
atatachak@usthb.dz

Received: September 2016 Revised: November 2017 Accepted: March 2017

Abstract:

- Let $\{X_N, N \geq 1\}$ be a sequence of strictly stationary associated random variables of interest, and $\{T_N, N \geq 1\}$ be a sequence of random truncating variables assumed to be independent from $\{X_N, N \geq 1\}$. In this paper, we establish the strong uniform consistency with a rate of a kernel hazard rate function estimator, when the variable of interest is subject to random left truncation under association condition. Simulation results are also provided to evaluate the finite-sample performances of the proposed estimator.

Key-Words:

- *associated data; hazard rate function; Lynden-Bell estimator; random left truncation; strong uniform consistency rate.*

AMS Subject Classification:

- 62G05, 62G07, 62G20.

1. INTRODUCTION AND MOTIVATION

First of all, let us recall that a set of random variables (rv's) $(X_1, X_2, ..., X_N)$ is said to be associated if for every pair of functions $g_1(\cdot)$ and $g_2(\cdot)$ from $\mathbb{R}^N$ to $\mathbb{R}$, which are non decreasing component-wise, $\text{cov}(g_1(\mathbf{X}), g_2(\mathbf{X})) \geq 0$, whenever the covariance is defined, where $\mathbf{X} = (X_1, X_2, ..., X_N)$. An infinite sequence $\{X_N,\, N \geq 1\}$ of rv's is said to be associated if every finite subset is associated. This definition was introduced by Esary *et al.* ([9]), mainly for the sake of applications. For instance, association occurs often in certain reliability theory problems, as well as in some important models employed in statistical mechanics. It is of interest to note that association and mixing define two distinct but not disjoint classes of processes (see, e.g. Doukhan and Louhichi ([7]), for examples of sequences that are associated but not mixing, associated and mixing, and mixing but not associated ones).

Let us now recall that a strong mixing condition refers more to σ-algebra than to rv's. On the one hand, a main inconvenience of mixing conditions is the difficulty of checking them. On the other hand, an important property of associated random rv's is that zero correlation implies independence. Also, large classes of examples of associated processes are deduced from the fact that any independent sequence is associated and that monotonic functions of independent sequences remain associated. So, the main advantage of the concept of association compared to mixing is that the conditions of limit theorems are easier to verify since, a covariance is much easier to compute than a mixing coefficient.

As examples of associated rv's, we recall that most often in reliability studies, the rv's which are generally lifetimes of components, are not independent but are associated. In fact, as an example, there are structures in which the components share the load so that failure of one component results in increased load on each of the remaining components. Thus, failure of one component will adversely effect the performance of all the minimal path structures containing it. In such a model, the random variables of interest are not independent but are associated. In addition, let $\{X_i,\, i \geq 1\}$ be independent and identically distributed (iid) rv's and Y be independent of $\{X_i,\, i \geq 1\}$. Then $\{Z_i = X_i + Y,\, i \geq 1\}$ are associated. Thus, if independent components of a system are subject to the same stress, then their lifetimes are associated. A variety of relevant examples and ample bibliographical references can be found in (Bulinski and Shashkin ([3])). In that book, the reader can find a number of new results and examples related to associated random sequences and random fields. For completeness on the subject in the complete data case we refer the reader to the monographs by Oliveira ([17]) and Prakasa Rao ([20]).

Survival analysis is the part of statistics, in which the variable of interest (lifetime) may often be interpreted as the time elapsed between two events and then, one may not be able to observe completely the variable under study. Such variables typically appear in a medical or an engineering life test studies. Among the different forms in which incomplete data appear, censoring and truncation are two common ones.

Left truncation in studies of developmental processes is not just of theoretical interest: It can cause substantial bias if ignored. An important example of such a model arises in the analysis of survival data of patients infected by the AIDS virus from contaminated blood transfusions (Chen *et al.* ([6])). Other examples in which a large fraction of potential observations are left truncated are rate of spontaneous abortion (Meister and Schaefer ([16])) and age at menopause transition stages (Harlow *et al.* ([12])).

Let $\{X_N,\, N\geq 1\}$ be a sequence of strictly stationary associated rv's of interest defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ with an unknown probability density function (pdf) $f=dF$. Let $\{T_N,\, N\geq 1\}$ be a sequence of stationary associated rv's of truncation with an unknown Lipschitz distribution function (df) G. In this paper we follow the same sampling scheme as that of (Woodroofe ([25])) whose observable sample of size n is a subset of N pairs $\{(X_1,T_1),...,(X_N,Y_N)\}$, where N is deterministic but unknown while n is random. As it was pointed out by the reviewer, one may consider another approach in which the sample size n is a non-random known value, and the observations are drawn from an infinite sequence of random vectors. In fact, such an approach was used by (He and Yang ([14])). However, our main motivation in following the first approach is computational since, in our simulation studies we use the ratio $\frac{n}{N}$ to estimate different values of the parameter α.

Under random left truncation scheme, only those pairs (X_i,T_i) satisfying $X_i\geq T_i$ are observed. In the sequel we assume that $\{X_N,\, N\geq 1\}$ is independent from $\{T_N,\, N\geq 1\}$ and $(X_1,T_1),...,(X_n,T_n)$ denotes the sequence which one actually observes within a sample (X_i,T_i); $1\leq i\leq N$. Obviously the observed sequence remains associated since any subset of associated rv's are associated (see Esary *et al.* ([9]), Property $\mathscr{P}_1$)). As a consequence of truncation, the sample size $n=\sum_{i=1}^{N}1_{\{X_i\geq T_i\}}$ is random, and from the strong law of large numbers, $n/N\to\alpha:=\mathbb{P}(X_i\geq T_i)$, almost surely (a.s.), as $N\to\infty$. Without further mention, we shall assume that $\alpha>0$ because, otherwise, no data will be available.

Throughout this study, all probability statements are to be interpreted as conditional probability statements, that is $\mathbf{P}(\cdot)=\mathbb{P}(\cdot\,|\,X\geq T)$. Likewise $\mathbf{E}$ and $\mathbb{E}$ will denote the expectation operators related to $\mathbf{P}$ and $\mathbb{P}$, respectively. Then conditionally on n, estimation results are stated considering $n\to\infty$ which hold true with respect to the probability $\mathbb{P}$ since $n\leq N$.

In what follows, the star notation $(^\star)$ relates to any characteristic of the actually observed data (conditionally on n). So, following Stute ([21]), the df's of X and T become

$$F^\star(x) := \alpha^{-1}\int_{-\infty}^{x} G(z)\,dF(z) \qquad \text{and} \qquad G^\star(x) := \alpha^{-1}\int_{-\infty}^{\infty} G(x\wedge z)\,dF(z)\,,$$

where $t\wedge z:=\min(x,z)$. Then, for any df W, let us define $a_W=\inf\{u\colon W(u)>0\}$ and $b_W=\sup\{u\colon W(u)<1\}$, as the endpoints of the W support. As pointed out in Woodroofe ([25]), the df's F and G can be completely estimated only if $a_G\leq a_F$, $b_G\leq b_F$ and $\int_{a_F}^{\infty}(G)^{-1}dF<\infty$.

Let $C(\cdot)$ be a function defined by

$$\begin{aligned} C(x) &:= \mathbf{P}\big(T\leq x\leq X\big) \\ &= G^\star(x)-F^\star(x) \\ &= \alpha^{-1}G(x)\left[1-F(x)\right], \qquad a_G<x<b_F\,. \end{aligned} \tag{1.1}$$

It is easily seen that $F^\star$, $G^\star$ and C are readily estimable through

$$F_n^\star(x)=n^{-1}\sum_{i=1}^{n}1_{\{X_i\leq x\}},\quad G_n^\star(t)=n^{-1}\sum_{i=1}^{n}1_{\{T_i\leq t\}}\quad\text{and}\quad C_n(x)=G_n^\star(x)-F_n^\star(x)\,.$$

The well-known estimates of F and G proposed by Lynden-Bell ([15]) are

$$F_n(x)=1-\prod_{X_i\leq x}\left[\frac{n\,C_n(X_i)-1}{n\,C_n(X_i)}\right]\quad\text{and}\quad G_n(t)=\prod_{T_i>t}\left[\frac{n\,C_n(T_i)-1}{n\,C_n(T_i)}\right], \tag{1.2}$$

respectively, assuming no ties among the rv's. Note that Stute and Wang ([22]) showed how to break ties without destroying the product limit structure. Therefore, throughout we shall assume without loss of generality that there are no ties.

For technical reasons, we need to introduce a pseudo-kernel estimate of f, which will be of a great importance later, defined by

$$\tilde{f}_n(x) := \frac{\alpha}{nh_n} \sum_{i=1}^{n} \frac{1}{G(X_i)}\, K\left(\frac{x - X_i}{h_n}\right), \tag{1.3}$$

where K is a smooth probability kernel and $h_n =: h$ is a sequence of bandwidths tending to 0 at appropriate rates. For an interesting overview of nonparametric curve estimation, we refer the reader to Cao *et al.* ([5]) and the references therein.

Note that in a real data situation or in simulation studies we shall, however, not dwell on (1.3) since G is unknown. And, as the original sample size N is unknown (although deterministic), the classical estimator $\hat{\alpha}_n = n/N$ for the rate of no truncation α cannot be used, and then, another estimator derived from (1.1) is required, namely

$$\alpha_n := \frac{G_n(x)\,\big[1 - F_n(x)\big]}{C_n(x)}\,,$$

for any x such that $C_n(x) > 0$. This estimator was proposed and studied in the iid case in (He and Yang ([13]) Theorem 2.2, p. 1014). These authors proved that α_n does not depend upon the argument x and its value can then be obtained for any x such that $C_n(x) \neq 0$. Furthermore, they showed (Corollary 2.5) that $\alpha_n \to \alpha$, a.s., as $n \to \infty$. Then, by plug-in method we can construct a feasible kernel estimate of f. Thus

$$\hat{f}_n(x) := \frac{\alpha_n}{nh} \sum_{i=1}^{n} \frac{1}{G_n(X_i)}\, K\left(\frac{x - X_i}{h}\right). \tag{1.4}$$

From now on, the sum in the latter formula is taken over the i's such that $G_n(X_i) \neq 0$. Recall that asymptotic results for (1.4), in both iid and strong mixing condition cases have been stated in (Ould Saïd and Tatachak ([18], [19]), Benrabah *et al.* ([2])).

It is well known that the cumulative hazard function $\Lambda(y) = -\log(1 - F(x))$, for any x such that $F(x) < 1$, and its corresponding hazard rate function $\lambda(x) := \Lambda'(x) = f(x)/(1 - F(x))$, are important in several fields of applied statistics (medicine, reliability, ...) for the assessment of risks in survival studies. Recall that the nonparametric hazard rate estimation was introduced in statistical literature by Watson and Leadbetter ([24]). Now, using (1.2) and (1.4), an estimate for $\lambda(x)$ for an n-sample, at risk of being truncated from the left, is defined by

$$\hat{\lambda}_n(x) = \frac{\hat{f}_n(x)}{1 - F_n(x)}\,. \tag{1.5}$$

As far as we know, in truncation and dependence setting, the only existing result dealing with hazard rate estimation is that of Sun and Zhou ([23]) stated under strong mixing condition, while in the complete associated data case (no truncation), Bagai and Prakasa Rao ([1]) stated strong uniform consistencies (with no rates) for kernel-type density and failure rate estimates. Hence, in this paper, we intend to extend the existing results to truncated and associated data.

The paper is organized as follows: In Section 2, an asymptotic analysis is presented together with the list of the assumptions under which the main results are stated. To support the main results, a simulation study illustrates the behaviour of the estimators as shown in Section 3. Proofs and some auxiliary results with their proofs are relegated to Section 4.

2. ASYMPTOTIC ANALYSIS

In the sequel, $\mathscr{D} := [a,b]$ such that $a_G \leq a_F < a < b < b_F$ will denote a compact set and the letter c is used indiscriminately as a generic positive constant. To state our asymptotic analysis, the following conditions are assumed:

A1. $\displaystyle\int \frac{dF(z)}{G^2(z)} < +\infty\,;$

A2. The covariance term satisfies: $\rho(s) := \sup_{j:|\ell-j|\geq s} \operatorname{cov}(X_j, X_\ell)$ for all $\ell \geq 1$ and $s > 0$, where $\rho(s) \leq \gamma_0\, e^{-\gamma s}$ for some positive constants γ_0 and γ ;

A3. K is a Lipschitz continuous pdf, compactly supported and $\displaystyle\int u\, K(u)\, du = 0\,;$

A4. f is twice continuously differentiable on $\mathscr{D}$ such that $\displaystyle\sup_{x\in\mathscr{D}} \big|f^{(2)}(x)\big| < +\infty\,;$

A5. The joint pdf $f^\star_{1,j}(\cdot,\cdot)$ of (X_1, X_{1+j}) satisfies: $\displaystyle\sup_{j>1}\,\sup_{u,v\in\mathscr{D}} \big|f^\star_{1,j}(u,v)\big| \leq c\,;$

A6. h satisfies: $h \to 0$ and $nh^{1+\delta} \to +\infty$ along with n, for any $0 < \delta < 1$.

Remark 2.1. Assumptions A1–A2 satisfy conditions H1–H3 in (Guessoum *et al.* ([11])). Furthermore, Assumption A1 was used in (Stute ([21])) and Assumption A2 quantifies a progressive tendency to asymptotic independence of "past" and "future". This latter condition was used in (Doukhan and Neumann ([8])) in order to state an exponential inequality which is needed to prove Proposition 2.1 hereinafter. Assumptions A3–A4 are frequently used in studying uniform consistency of estimates. Assumption A5 is often assumed in kernel estimation studies under dependence structure and allows to bound the covariance term. Finally, Assumption A6 is standard in nonparametric density estimation.

Proposition 2.1. *Under assumptions A1–A6, for large enough n we have*

$$\sup_{x\in\mathscr{D}} \Big|\tilde{f}_n(x) - \mathbf{E}\big(\tilde{f}_n(x)\big)\Big| = O\left(\sqrt{\frac{\log n}{nh}}\right) \quad \text{a.s.}$$

Theorem 2.1. *If assumptions A1–A6 hold true, then for large enough n we have*

$$\sup_{x\in\mathscr{D}} \Big|\hat{f}_n(x) - f(x)\Big| \;=\; O\left\{\sqrt{\frac{\log n}{nh}} + \left(\frac{\log\log n}{n}\right)^{\theta} + h^2\right\} \quad \text{a.s.}\,,$$

where $0 < \theta < \gamma/(2\gamma + \beta + 9)$ for any real $\beta > 0$ and γ is that in A2.

Theorem 2.2. *Under assumptions A1–A6, for large enough n we have*

$$\sup_{x\in\mathscr{D}} \Big|\hat{\lambda}_n(x) - \lambda(x)\Big| \;=\; O\left\{\sqrt{\frac{\log n}{nh}} + \left(\frac{\log\log n}{n}\right)^{\theta} + h^2\right\} \quad \text{a.s.}$$

Remark 2.2. The rates in Theorem 2.1 and Theorem 2.2 are still slower than those stated for complete data in the iid and mixing cases (see Estévez and Quintela ([10])), or under left truncation model (see Ould Saïd and Tatachak ([18], [19]), Sun and Zhou ([23])). Our rates depend upon the parameter θ which controls the covariance's decaying under association dependence as stated in (Cai and Roussas ([4])), whereas the iterated logarithm form is related to the truncation effect. Note that by setting $\gamma = 3(r-2)/2$, $r > 2$, we recognize the θ appearing in (Guessoum *et al.* ([11]), Theorem 3.1). Finally, we point out that for γ large enough, our rates approach the classical optimal ones as θ grows to its upper bound ($\theta = 1/2$).

3. SOME SIMULATION RESULTS

To examine the behaviour over finite samples of the estimators given in statements (1.4) and (1.5), respectively, we have conducted a numerical study via simulation. The log-normal distribution has been selected because of the shape of its hazard function which is flatter around of its maximum. In the computation of the estimators, we used the bi-weight kernel $\left(K(x) = \left(1 - |x|^2\right)^2 1_{|x|\leq 1}\right)$ which verifies our conditions in stating our main results. We also used optimal global and local bandwidths, that minimized the global mean square error (GMSE) and the simple mean square error (MSE) criteria, respectively. These bandwidths were selected in the grid of values $\mathscr{H} = \left\{h_k = 10^{-1} + 5(k-1)10^{-2},\ k = 1, 2, ..., 19\right\}$.

3.1. Models and procedure

- *Step 1.* The sequence $\left\{(X_k, T_k),\ k = 1, ..., n\right\}$ is generated as follows:
  For $i = 1, 2, ..., N$, we first generate $Z_i = (W_{i-1} + W_{i-2}/2)$, where $\{W_r,\ r = -1, 0, ..., N-1\}$ are iid rv's drawn from $\mathcal{N}(0,1)$ and put $X_i = \exp(Z_i)$, $i = 1, ..., N$. Hence, the sequence $\{X_k,\ k = 1, 2, ..., N\}$ is associated and follows a $\log\left(\mathcal{N}\left(0, \sqrt{1/2}\right)\right)$ distribution. At each iteration the X_i's are compared to the T_i's generated from $\exp(\mu)$ in order to keep only the pairs (X_i, T_i) satisfying $X_i \geq T_i$. The parameter μ is adjusted to get $\mathbb{P}(X \geq T) \approx \alpha$. Hence, a truncation sequence $\left\{(X_i, T_i),\ i = 1, ..., n\right\}$ is generated and the estimator $\lambda_n(\cdot)$ is computed using the bi-weight kernel and bandwidths $h \in \mathscr{H}$.

- *Step 2.* We repeat B simulation runs as described in *Step 1* for every fixed combination of size n and truncating rate (TR) $1 - \alpha$.
  For a given functional g and its estimate $\hat{g}_{n,h}$, the GMSE computed along $B = 200$ Monte Carlo trials and a grid of bandwidths $h \in \mathscr{H}$ is defined as

$$\mathrm{GMSE}(h) = \frac{1}{Bm} \sum_{k=1}^{B} \sum_{\ell=1}^{m} \left(\hat{g}_{n,h,k}(x_\ell) - g(x_\ell)\right)^2,$$

  where m is a number of equidistant points x_ℓ belonging to the range $]0, 4]$ and $\hat{g}_{n,h,k}(x_\ell)$ is the value of $\hat{g}_{n,h}(x_\ell)$ computed at iteration k. In computing the GMSE's, optimal global bandwidths (ogb) were used for both density and hazard rate function estimation. The values $\mathrm{GMSE} := \min_{h\in\mathscr{H}} \mathrm{GMSE}(h)$ and the corresponding global bandwidths $h_{\mathrm{opt}} := \arg\min_{h\in\mathscr{H}} \mathrm{GMSE}(h)$ are reported in Table 1 and Table 2.

The MSE's reported in Table 3 were evaluated by using optimal local bandwidths (olb) for hazard rate estimation. Furthermore, to display the quality of fit of the estimators, we first plotted the target density f together with its average and median estimates as illustrated in Figure 1 and Figure 2. Then, we plotted the target hazard rate λ with its average and median estimates for both global optimal bandwidths and local optimal ones as shown in Figure 3, Figure 4 and Figure 5.

Table 1: Density function with optimal global bandwidths.

$1-\alpha$ (TR)	n					
	50		100		200	
	h_{opt}	GMSE	h_{opt}	GMSE	h_{opt}	GMSE
0.05	0.575	0.0073	0.475	0.0048	0.375	0.0029
0.15	0.600	0.0090	0.475	0.0079	0.400	0.0073
0.25	0.575	0.0155	0.475	0.0122	0.400	0.0096

Table 2: Hazard rate function with optimal global bandwidths.

TR	n					
	50		100		200	
	h_{opt}	GMSE	h_{opt}	GMSE	h_{opt}	GMSE
0.05	0.825	0.2052	0.675	0.1404	0.600	0.0743
0.15	0.800	0.2574	0.725	0.1732	0.640	0.0908
0.25	0.750	0.2676	0.750	0.1800	0.650	0.1145

Table 3: Hazard rate function with optimal local bandwidths.

TR	n		
	30	50	100
	MSE	MSE	MSE
0.05	0.2247	0.1949	0.1286
0.15	0.2632	0.2104	0.1500
0.25	0.3384	0.2848	0.1848

3.2. Comments on the simulation results

As it can be seen from the tables and figures, the higher the sample size and smaller the TR, the better the quality of fit. This means that the errors tend to be negligible in each case when n increases. Likewise, the quality of fit deteriorates slightly for sufficiently high TR value but, it increases along with n and becomes better in any cases. Note also that, in particular, the estimation of the hazard rate function suffers from the well-known boundary effects that occur in nonparametric functional estimation. If the target functional has a support on $[0,\infty)$, the use of classical estimation methods with symmetric kernels yield a large bias on the zero boundary and leads to a bad quality of the estimates. This is the case here and is due to the fact that symmetric kernel estimators assign non-zero weight at the interval $(-\infty,0]$.

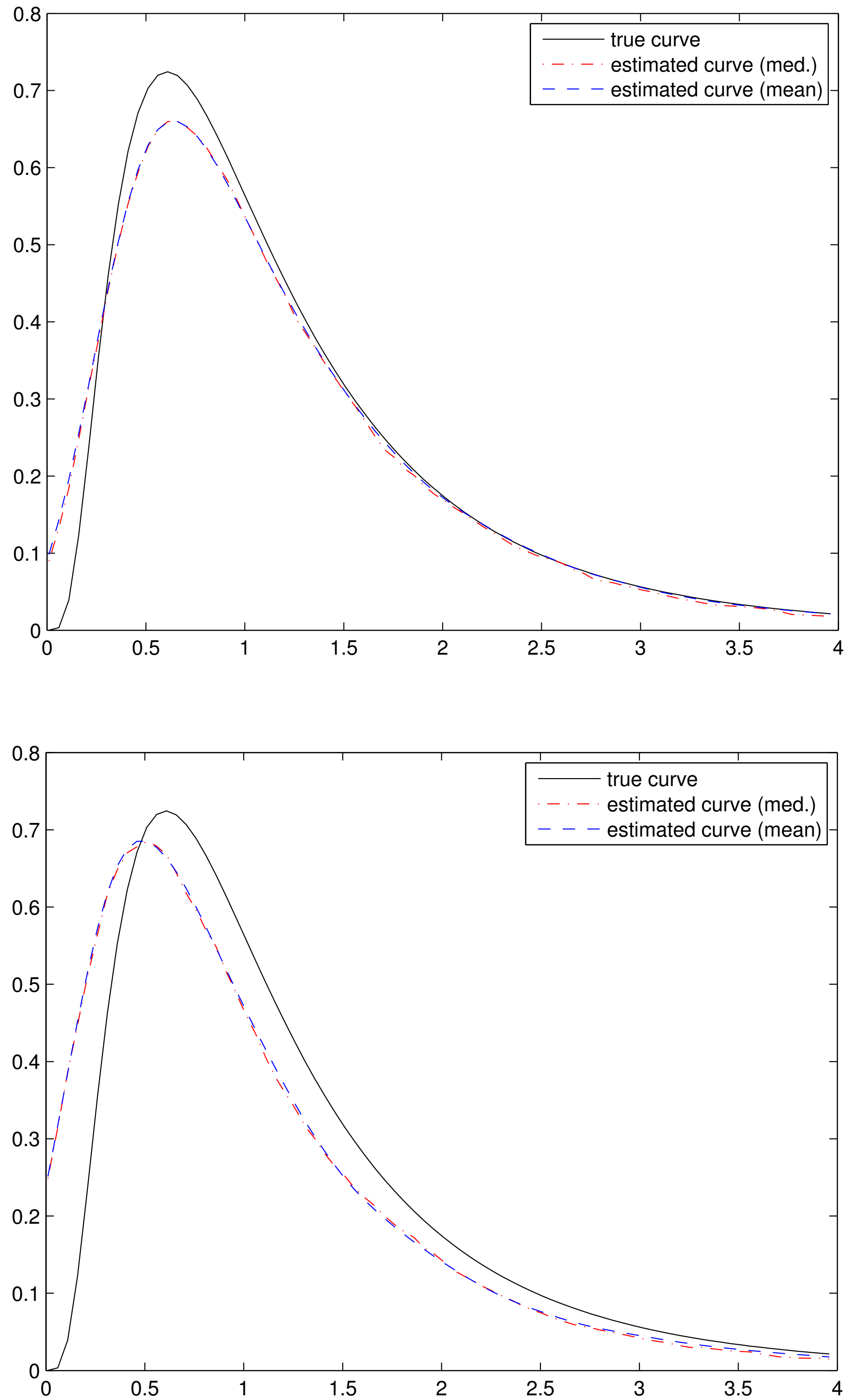
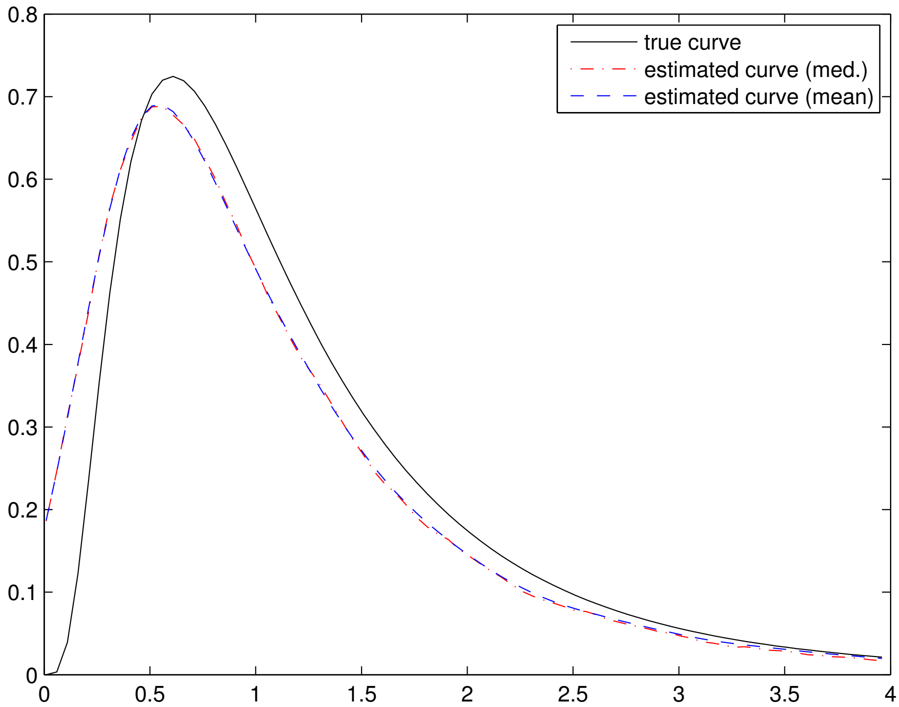


Figure 1: Density estimation (ogb): $n = 100$ and TR $\approx 0.05, 0.25$.



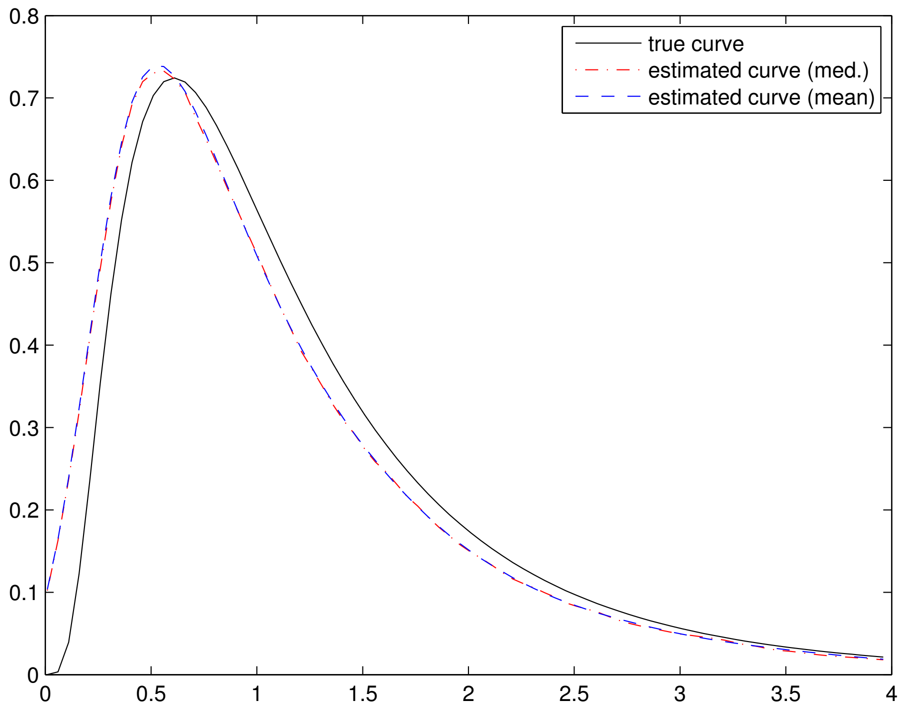


Figure 2: Density estimation (ogb): $n = 100, 500$ and TR ≈ 0.15.

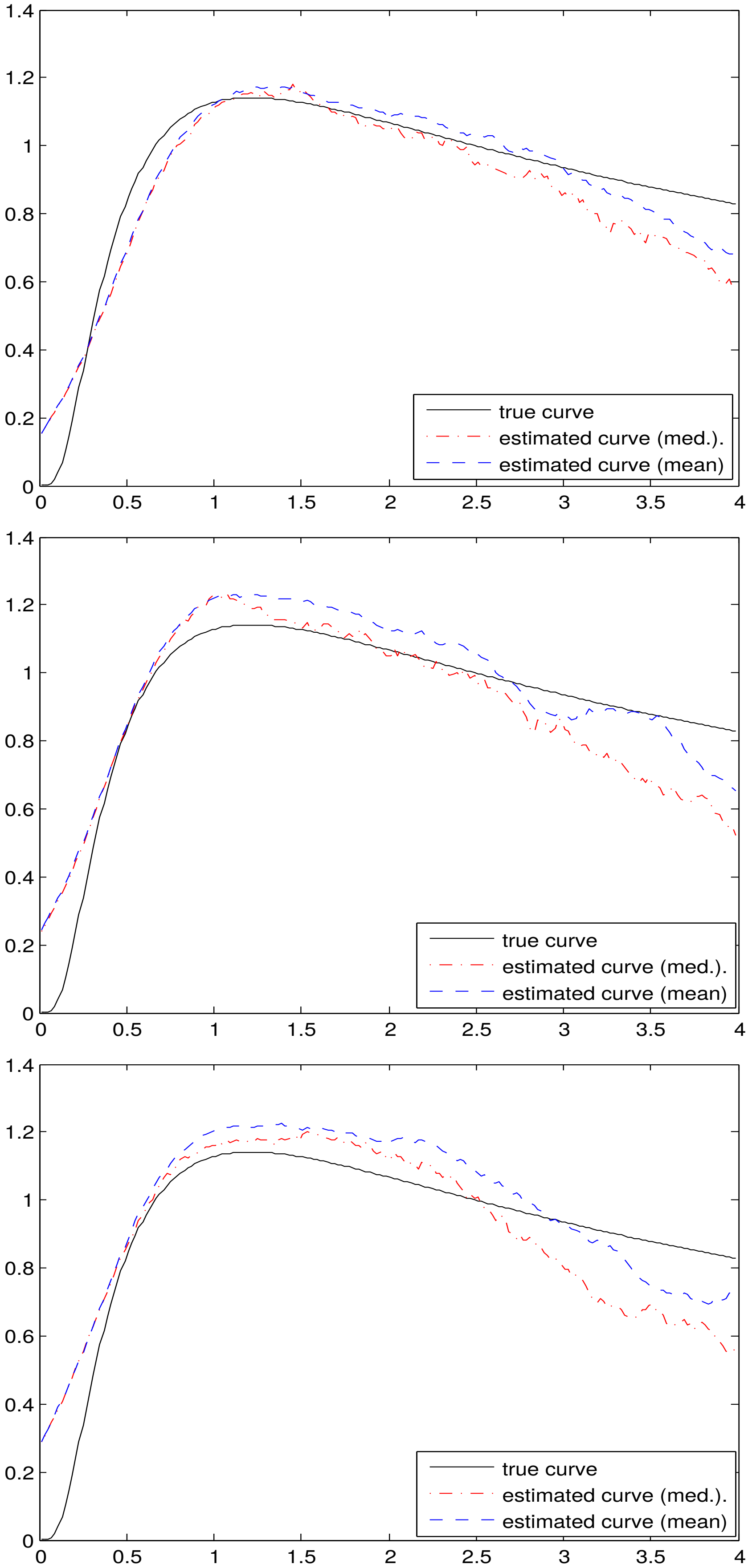


Figure 3: Hazard rate (ogb): $n = 100$ and TR $\approx 0.05, 0.15, 0.25$.

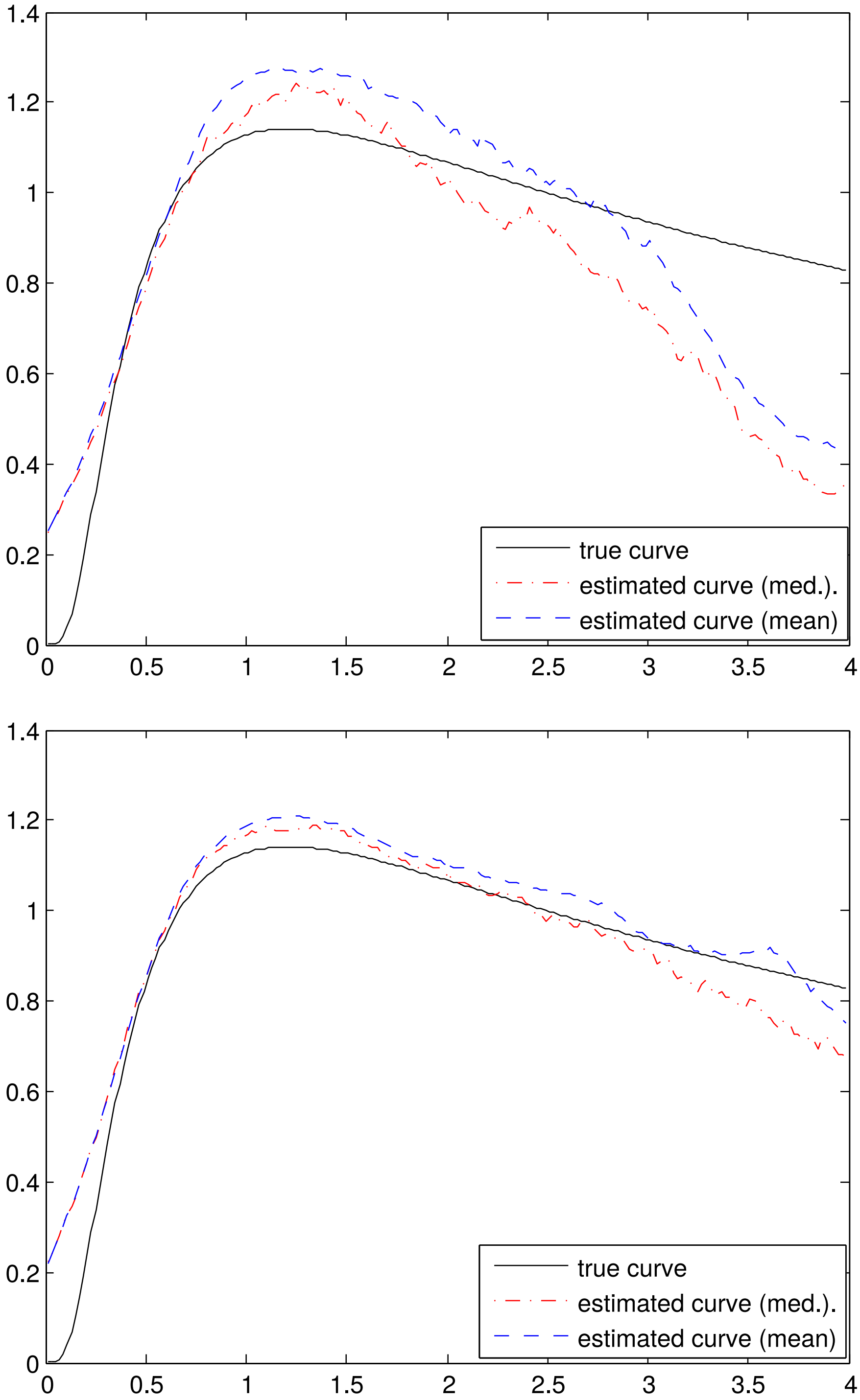


Figure 4: Hazard rate (ogb): TR ≈ 0.15 and $n = 50, 200$.

The graphs reveal this phenomenon when using optimal global bandwidths but, the bias effect is subsequently reduced and tends to disappear when optimal local bandwidths are used as shown in Figure 5. We point out that one may also select another approach to deal with the boundary bias effect which consists in using an asymmetric kernel as the Gamma kernel since it is non-negative and changes its shape depending on the position on the semi-axis. The inverse Gaussian kernel is also an interesting alternative.

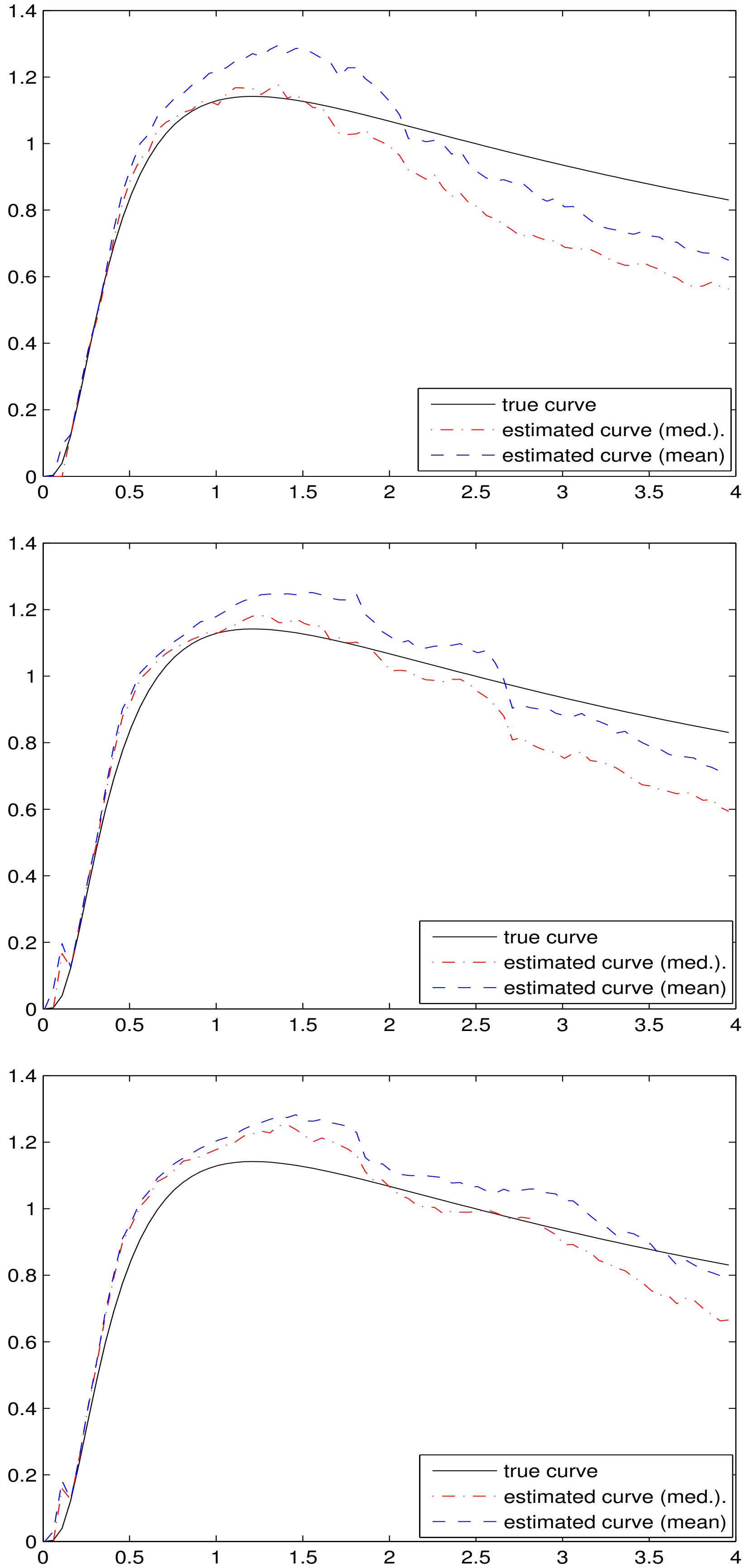


Figure 5: Hazard rate (olb): TR ≈ 0.15 and $n = 30, 50, 100$.

4. AUXILIARY RESULTS AND PROOFS

Before proving the main results, we briefly discuss the tools used here.

Remark 4.1. As it was mentioned above, there are processes which are associated but not mixing. In such cases, it would be interesting to have at disposal similar results as stated here.

It is noteworthy that for the proof of our results we use similar tools used in the α-mixing frameworks. The main difference here is that functional of associated rv's are not associated in general, which is the case when dealing with nonparametric kernel estimation. This is due to the fact that the random functions $K\big(\frac{x-X_i}{h}\big)$ are, in general not associated but remain α-mixing if the X_i's are, since K is a measurable function in general. To keep the association, we should apply only monotone transformations to the original variables, which is not the case with a general kernel. To overcome this problem, one may assume that the kernel K is of bounded variation. This condition permits to write $K = K_1 - K_2$, with K_1 and K_2 monotone functions. In this paper, we do not follow this procedure but we use results stated for weakly dependent models in the sense of Doukhan and Louhichi ([7]), since associated models are κ-weakly dependent. Note also that to treat the fluctuation part in Proposition 2.1, we use bounds for covariances in applying an exponential inequality stated by Doukhan and Neumann ([8]) for weakly dependent rv's. To this end, we use Theorem 5.3 in (Bulinski and Shashkin ([3])), and Proposition 8 in (Doukhan and Neumann ([8])).

Indeed, for any $x \in \mathscr{D}$, set $U_i(x,h) := \frac{\alpha}{G(X_i)}\, K\big(\frac{x-X_i}{h}\big) - \mathbf{E}\big(\frac{\alpha}{G(X_i)}\, K\big(\frac{x-X_i}{h}\big)\big)$. So, it follows that

$$\tilde{f}_n(x) - \mathbf{E}\big(\tilde{f}_n(x)\big) = \frac{1}{nh}\sum_{i=1}^{n} U_i(x,h)\,. \tag{4.1}$$

The proof of Proposition 2.1 is based on Lemma 4.1 and Lemma 4.2 hereafter.

Lemma 4.1. *Under the assumptions of Proposition 2.1, for all u-tuples $(s_1,...,s_u)$ and all v-tuples $(w_1,...,w_v)$ with $1 \le s_1 \le \cdots \le s_u \le w_1 \le \cdots \le w_v \le n$, we have*

$$\textbf{(i)}\quad \operatorname{cov}\left(\prod_{i=s_1}^{s_u} U_i(x,h)\,,\ \prod_{j=w_1}^{w_v} U_j(x,h)\right) =: \operatorname{cov}_1 \le c^{u+v} h^{-2}\, u\, v\, \rho(w_1 - s_u)\,,$$

$$\textbf{(ii)}\quad \operatorname{cov}\left(\prod_{i=s_1}^{s_u} U_i(x,h)\,,\ \prod_{j=w_1}^{w_v} U_j(x,h)\right) =: \operatorname{cov}_2 \le c^{u+v} h^{2}\,.$$

Proof of Lemma 4.1: Let $\operatorname{Lip}(\Phi)$ denote the Lipschitz modulus of continuity of Φ, that is

$$\operatorname{Lip}(\Phi) = \sup_{x\neq y} \frac{\big|\Phi(x) - \Phi(y)\big|}{|x-y|_1}\,, \qquad \text{where}\quad \big|(z_1,...,z_d)\big|_1 = |z_1| + \cdots + |z_d|\,.$$

To prove item (i), we use a result in (Bulinski and Shashkin ([3]), Theorem 5.3, p. 89) and then we have

$$\operatorname{cov}_1 \le \operatorname{Lip}\left(\prod_{i=s_1}^{s_u} U_i(x,h)\right) \operatorname{Lip}\left(\prod_{j=w_1}^{w_v} U_j(x,h)\right) \sum_{i=s_1}^{s_u} \sum_{j=w_1}^{w_v} \operatorname{cov}(X_i, X_j)\,.$$

Now since

$$\operatorname{Lip}\left(\prod_{i=1}^{k} U_i(x,h)\right) \le \frac{c}{h}\left(\frac{2}{G(a)}\right)^k \|K\|_\infty^{k-1}\,,$$

where $\|K\|_\infty := \sup_u K(u)$. Then by stationarity and Assumption A2, we get

$$\operatorname{cov}_1 \le \frac{c^2\, 2^{u+v}}{h^2\, G^{u+v}(a)}\, \|K\|_\infty^{u+v-2}\, u\, v\, \rho(w_1 - s_u)\,.$$

Thus result (i) holds. The result (ii) follows by simple algebra using assumptions A3–A5. The proof is finished. □

Lemma 4.2. *There exist constants $M, L_1, L_2 < +\infty$, $\mu, \lambda \ge 0$ and a non-increasing sequence of real numbers $\big(\phi(n)\big)_{n\ge1}$ such that*

(**a**) $\operatorname{cov}\left(\prod_{i=s_1}^{s_u} U_i(x,h)\,,\ \prod_{j=w_1}^{w_v} U_j(x,h)\right) =: \operatorname{cov} \le c^{u+v} h\, u\, v\, \phi(w_1 - s_u)\,,$

(**b**) $\sum_{t\ge0} (t+1)^{k_0}\phi(t) \le L_1 L_2^{k_0} \big(k_0!\big)^{\mu}, \quad \forall\, k_0 \ge 0\,,$

(**c**) $\mathbf{E}\Big(\big|U_i(x,h)\big|^{k_0}\Big) \le \big(k_0!\big)^{\lambda} M^{k_0}\,.$

The items in Lemma 4.2 are nearly the conditions of Theorem 1 in (Doukhan and Neumann ([8])). This latter will allow us to use their exponential inequality in proving Proposition 2.1.

Proof of Lemma 4.2: To prove item (a) we apply Lemma 4.1 by taking $\phi(\cdot) = \rho^{1/4}(\cdot)$ and writing $\operatorname{cov} = \operatorname{cov}_1^{1/4} \operatorname{cov}_2^{3/4}$. The proofs for (b) and (c) are similar to those in (Doukhan and Neumann ([8]), Proposition 8) by choosing $\lambda = 0$, $\mu = 1$ and $L_1 = L_2 = \frac{1}{1-e^{-\gamma/4}}$, and then we omit them. □

Proof of Proposition 2.1: The main tool used here to bound the fluctuation term in (4.1), is an exponential inequality due to Doukhan and Neumann ([8]), that is

$$\mathbf{P}\left(\sum_{i=1}^{n} U_i(x,h) \ge \varepsilon\right) \le \exp\left(-\frac{\varepsilon^2/2}{A_n + B_n^{1/(\mu+\lambda+2)}\,\varepsilon^{(2\mu+2\lambda+3)/(\mu+\lambda+2)}}\right), \tag{4.2}$$

where A_n can be chosen such that $A_n \le \sigma_n^2$ with

$$\sigma_n^2 := \operatorname{Var}\left(\sum_{i=1}^{n} U_i(x,h)\right),$$

and

$$B_n = 2\,c\,L_2\left(\frac{2^{4+\mu+\lambda}\,c\,n\,h\,L_1}{A_n} \vee 1\right).$$

For this purpose, let us calculate $\sigma_n^2 = (nh)^2\,\mathrm{Var}\big(\tilde{f}_n(x)\big)$. We have

$$\begin{aligned}(nh)^2\,\mathrm{Var}\big(\tilde{f}_n(x)\big) &= n\left\{\mathbf{E}\left[\frac{\alpha^2}{G^2(X_1)}\,K^2\left(\frac{x-X_1}{h}\right)\right] - \mathbf{E}^2\left[\frac{\alpha}{G(X_1)}\,K\left(\frac{x-X_1}{h}\right)\right]\right\}\\ &\quad + \sum_{i=1}^{n}\sum_{j\neq i, j=1}^{n}\mathrm{cov}\left(\frac{\alpha}{G(X_i)}\,K\left(\frac{x-X_i}{h}\right),\ \frac{\alpha}{G(X_j)}\,K\left(\frac{x-X_j}{h}\right)\right)\\ &=: V_1 + V_2\,.\end{aligned}$$

On the one hand, by assumptions A3–A4, a change of variable and the Dominated Convergence Theorem, we obtain $V_1 = O(nh)$.
On the other hand, from Lemma 4.1, we can write

$$\mathrm{cov}\left(\frac{\alpha}{G(X_i)}\,K\left(\frac{x-X_i}{h}\right),\ \frac{\alpha}{G(X_j)}\,K\left(\frac{x-X_j}{h}\right)\right) = O\big(h^2\big)\,. \tag{4.3}$$

And, let

$$\mathcal{B}_1 = \Big\{(i,j)\,/\ 1 \le |i-j| \le \eta_n\Big\} \quad \text{and} \quad \mathcal{B}_2 = \Big\{(i,j)\,/\ \eta_n + 1 \le |i-j| \le n-1\Big\},$$

where $\eta_n = o(n)$. Then

$$\begin{aligned}V_2 &= \sum_{i=1}^{n}\sum_{j\in\mathcal{B}_1}\mathrm{cov}\left(\frac{\alpha}{G(X_i)}\,K\left(\frac{x-X_i}{h}\right),\ \frac{\alpha}{G(X_j)}\,K\left(\frac{x-X_j}{h}\right)\right)\\ &\quad + \sum_{i=1}^{n}\sum_{j\in\mathcal{B}_2}\mathrm{cov}\left(\frac{\alpha}{G(X_i)}\,K\left(\frac{x-X_i}{h}\right),\ \frac{\alpha}{G(X_j)}\,K\left(\frac{x-X_j}{h}\right)\right)\\ &=: V_{21} + V_{22}\,.\end{aligned}$$

From (4.3) we have

$$V_{21} = O\big(\eta_n\,n h^2\big)\,, \tag{4.4}$$

then by Assumption A2 and Lemma 4.2 (a) we obtain

$$\frac{V_{22}}{nh} \le \frac{c}{nh}\sum_{i=1}^{n}\sum_{j\in\mathcal{B}_2} h\,e^{-\frac{\gamma|i-j|}{4}} \le c\int_{\eta_n}^{n} e^{-\frac{\gamma u}{4}}\,du = O\Big(e^{-\frac{\gamma\eta_n}{4}}\Big). \tag{4.5}$$

Choosing $\eta_n = O\big(h^{\delta-1}\big)$ with $0<\delta<1$ (δ may be the same as that in A6), the statements (4.4) and (4.5) give $V_{21} = o(nh)$ and $\frac{V_{22}}{nh} = o(1)$. Consequently

$$\sigma_n^2 = O\big(nh\big)\,.$$

Thus we choose $A_n = O(nh)$ and $B_n = O(1)$.
At this step we are able to apply (4.2). To end the proof of Proposition 2.1, we use a covering of the compact $\mathscr{D}$ by a finite number ℓ_n of intervals $\mathscr{D}_1, ..., \mathscr{D}_{\ell_n}$ of equal length

$a_n = O\big(n^{-1/2}h^{3/2}\big)$ and centered at points $x_1, ..., x_{\ell_n}$, respectively. Note that as $\mathscr{D}$ is bounded, there exists a constant $M_0 > 0$ such that $\ell_n \leq M_0\, a_n^{-1}$. Then observe that

$$\begin{aligned}\sup_{x\in\mathscr{D}}\left|\tilde{f}_n(x) - \mathbf{E}\big(\tilde{f}_n(x)\big)\right| &= \sup_{x\in\mathscr{D}}\frac{1}{nh}\left|\sum_{i=1}^{n}U_i(x,h)\right| \\ &\leq \max_{k=1,\ldots,\ell_n}\sup_{x\in\mathscr{D}_k}\frac{1}{nh}\sum_{i=1}^{n}\big|U_i(x,h) - U_i(x_k,h)\big| \\ &\quad + \max_{k=1,\ldots,\ell_n}\frac{1}{nh}\left|\sum_{i=1}^{n}U_i(x_k,h)\right|.\end{aligned}$$

First, since K is Lipschitz we have

$$\begin{aligned}\frac{1}{nh}\sum_{i=1}^{n}\big|U_i(x,h) - U_i(x_k,h)\big| &\leq \frac{1}{nh}\sum_{i=1}^{n}\frac{\alpha}{G(X_i)}\left|K\Big(\frac{x-X_i}{h}\Big) - K\Big(\frac{x_k-X_i}{h}\Big)\right| \\ &\quad + \frac{1}{h}\,\mathbf{E}\left(\frac{\alpha}{G(X_i)}\left|K\Big(\frac{x-X_i}{h}\Big) - K\Big(\frac{x_k-X_i}{h}\Big)\right|\right) \\ &\leq \frac{1}{h}\,\frac{2}{G(a)}\left|\frac{x-x_k}{h}\right| \\ &\leq \frac{c}{G(a)\sqrt{nh}} = O\Big(\frac{1}{\sqrt{nh}}\Big).\end{aligned} \tag{4.6}$$

Next, by Assumption A6, if we replace ε by $\varepsilon_0\sqrt{nh\,\log n} =: \varepsilon_n$ in (4.2), we then get

$$\begin{aligned}\mathbf{P}\left(\max_{k=1,\ldots,\ell_n}\frac{1}{nh}\left|\sum_{i=1}^{n}U_i(x_k,h)\right| > \varepsilon_0\sqrt{\frac{\log n}{nh}}\right) &\leq \sum_{k=1}^{\ell_n}\mathbf{P}\left(\left|\sum_{i=1}^{n}U_i(x_k,h)\right| > \varepsilon_n\right) \\ &\leq c\,a_n^{-1}\exp\left(\frac{-\big(\varepsilon_0^2\,\log n\big)/2}{c + \varepsilon_0^{5/3}\Big(\frac{\log^5 n}{nh}\Big)^{1/6}}\right) \\ &\leq \frac{c}{\big(nh^{1+\delta}\big)^{\frac{3}{2(1+\delta)}}}\; n^{-c\varepsilon_0^2 + \frac{4+\delta}{2(1+\delta)}}.\end{aligned} \tag{4.7}$$

For a suitable choice of ε_0, the right hand side term in (4.7) becomes the general term of a convergent series. Then Borel–Cantelli's lemma gives

$$\max_{k=1,\ldots,\ell_n}\frac{1}{nh}\left|\sum_{i=1}^{n}U_i(x_k,h)\right| = O\left(\sqrt{\frac{\log n}{nh}}\right).$$

This latter jointly with (4.6) allow us to conclude the desired result, that is

$$\sup_{x\in\mathscr{D}}\frac{1}{nh}\left|\sum_{i=1}^{n}U_i(x,h)\right| = O\left(\sqrt{\frac{\log n}{nh}}\right) = \sup_{x\in\mathscr{D}}\left|\tilde{f}_n(x) - \mathbf{E}\big(\tilde{f}_n(x)\big)\right|,$$

which ends the proof of Proposition 2.1. □

Now the proof of Theorem 2.1 is immediately established once the following lemmas (Lemma 4.3 and Lemma 4.4) are stated.

Lemma 4.3. *Under assumptions A1–A2, for n sufficiently large we have*

$$\sup_{x\in\mathscr{D}}\big|G_n(x)-G(x)\big| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \quad \text{a.s.}, \tag{4.8}$$

$$\big|\alpha_n-\alpha\big| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \quad \text{a.s.} \tag{4.9}$$

Proof of Lemma 4.3: To prove (4.8), it suffices to follow step by step the proof in (Guessoum *et al.* ([11]), Theorem 3.2). The result (4.9) ensues using the following decomposition

$$\begin{aligned}\big|\alpha_n-\alpha\big| &= \frac{1}{C_n(x)\,C(x)}\bigg|C(x)\Big(G_n(x)-G(x)\Big)\,(1-F_n(x))\\ &\quad + C(x)\,G(x)\Big(F(x)-F_n(x)\Big)+G(x)\Big(C_n(x)-C(x)\Big)\,(F(x)-1)\bigg|.\end{aligned}$$

Thus the result holds using (4.8) jointly with Theorem 3.2 and Lemma 4.2 in (Guessoum *et al.* ([11])). □

Lemma 4.4. *Under the hypotheses of Theorem 2.1, for large n enough we have*

$$\sup_{x\in\mathscr{D}}\Big|\Big(\hat f_n(x)-\tilde f_n(x)\Big)\Big| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \quad \text{a.s.}, \tag{4.10}$$

$$\sup_{x\in\mathscr{D}}\Big|\Big(\mathbf{E}\big(\tilde f_n(x)\big)-f(x)\Big)\Big| = O\big(h^2\big) \quad \text{a.s.} \tag{4.11}$$

Proof of Lemma 4.4: To get (4.10), remark that

$$\Big|\hat f_n(x)-\tilde f_n(x)\Big| = \frac{1}{nh}\sum_{i=1}^{n}\left|\frac{\alpha_n\Big(G(X_i)-G_n(X_i)\Big)+(\alpha_n-\alpha)\,G_n(X_i)}{G_n(X_i)\,G(X_i)}\right| K\left(\frac{x-X_i}{h}\right).$$

Then, Lemma 4.3 gives the result. For the bias term in statement (4.11), the result is obtained by using classical tools under assumptions A3 and A4. □

Proof of Theorem 2.1: The result holds by writing

$$\hat f_n(x)-f(x) = \Big(\hat f_n(x)-\tilde f_n(x)\Big)+\Big(\tilde f_n(x)-\mathbf{E}\big(\tilde f_n(x)\big)\Big)+\Big(\mathbf{E}\big(\tilde f_n(x)\big)-f(x)\Big)$$

and using Proposition 2.1 together with Lemma 4.4. □

Proof of Theorem 2.2: Let us consider the following decomposition

$$\hat\lambda_n(x)-\lambda(x) = \frac{\big(1-F(x)\big)^{-1}}{\big(1-F_n(x)\big)}\bigg(\big(1-F(x)\big)\Big(\hat f_n(x)-f(x)\Big)-f(x)\Big(F(x)-F_n(x)\Big)\bigg).$$

Then the proof follows from Theorem 3.2 in (Guessoum *et al.* ([11])) and Theorem 2.1. □

ACKNOWLEDGMENTS

We are grateful to an anonymous reviewer for his/her particularly careful reading, relevant remarks and constructive comments, which helped us to improve the quality and the presentation of an earlier version of this paper.

REFERENCES

[1] Bagai, I. and Prakasa Rao, B.L.S. (1995). Kernel-type density and failure rate estimation for associated sequences, *Ann. Inst. Stat. Math.*, **47**, 253–266.

[2] Benrabah, O,; Ould Saïd, E. and Tatachak, A. (2015). A kernel mode estimate under random left truncation and time series model: asymptotic normality, *Stat. Papers*, **56**, 887–910.

[3] Bulinski, A. and Shashkin, A. (2007). *Limit Theorems for Associated Random Fields and Related Systems*, World Scientific, Singapore.

[4] Cai, Z. and Roussas, G.G. (1998). Kaplan–Meier estimator under association, *J. Multivariate Anal.*, **67**, 318–348.

[5] Cao, R.; Delgado, M.A. and Gonzalez-Manteiga, W. (1997). Nonparametric curve estimation: an overview, *Investigaciones Economicas*, **XXI**, 209–252.

[6] Chen, K.; Chao, M.T. and Lo, S.H. (1995). On strong uniform consistency of the Lynden-Bell estimator for truncation data, *Ann. Statist.*, **23**, 440–449.

[7] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moments inequalities, *Stochastic Processes and their Applications*, **84**, 313–342.

[8] Doukhan, P. and Neumann, M. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications, *Stochastic Processes and their Applications*, **117**, 878–903.

[9] Esary, J.; Proschan, F. and Walkup, D. (1967). Association of random variables with applications, *Ann. Math. Statist.*, **38**, 1466–1476.

[10] Estévez, G. and Quintela, A. (1999). Nonparametric estimation of the hazard function under dependence conditions, *Com. Statist. Theory* & *Meth.*, **28**(10), 2294–2331.

[11] Guessoum, Z.; Ould Saïd, E.; Sadki, O. and Tatachak, A. (2012). A note on the Lynden-Bell estimator under association, *Statist. Probab. Lett.*, **82**, 1994–2000.

[12] Harlow, S.D.; Cain, K.; Crawford, S. *et al.* (2006). Evaluation of four proposed bleeding criteria for the onset of late menopausal transition, *J. Clin. Endocrinol. Metab.*, **91**, 3432–3438.

[13] He, S. and Yang, G. (1998). Estimation of the truncation probability in the random truncation model, *Ann. Statist.*, **26**, 1011–1027.

[14] He, S. and Yang, G. (2003). Estimation of regression parameters with left truncated data, *J. Statist. Plan. Inference*, **117**, 99–122.

[15] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars, *Monthly Notices Royal Astronomy Society*, **155**, 95–118.

[16] Meister, R. and Schaefer, C. (2008). Statistical methods for estimating the probability of spontaneous abortion in observational studies — Analyzing pregnancies exposed to coumarin derivatives, *Reprod. Toxicol.*, **26**, 31–35.

[17] Oliveira, P.E. (2012). *Asymptotics for Associated Random Variables*, Springer Verlag.

[18] Ould Saïd, E. and Tatachak, A. (2009a). On the nonparametric estimation of the simple mode under left-truncation model, *Romanian Journal of Pure and Applied Mathematics*, **54**, 243–266.

[19] Ould Saïd, E. and Tatachak, A. (2009b). Strong consistency rate for the kernel mode under strong mixing hypothesis and left truncation, *Com. Statist. Theory* & *Meth.*, **38**, 1154–1169.

[20] Prakasa Rao, B.L.S. (2012). *Associated Sequences, Demimartingales and Nonparametric Inference*, Probability and its Applications, Springer Basel AG.

[21] Stute, W. (1993). Almost sure representation of the product-limit estimator for truncated data, *Ann. Statist.*, **21**, 146–156.

[22] Stute, W. and Wang, J.L. (2008). The central limit theorem under random truncation, *Bernoulli*, **14**, 604–622.

[23] Sun, L. and Zhou, X. (2001). Survival function and density estimation for truncated dependent data, *Statist. Probab. Lett.*, **52**, 47–57.

[24] Watson, G.S. and Leadbetter, M.R. (1964). Hazard analysis I, *Biometrika*, **51**, 175–184.

[25] Woodroofe, M. (1985). Estimating a distribution function with truncated data, *Ann. Statist.*, **13**, 163–177.