ELSEVIER
Journal of Computational and Applied Mathematics 55 (1994) 165-182
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Some stochastics on monotone functions

J.A. Cuesta-Albertos[a,*,1], J.S. Domínguez-Menchero[a,2], C. Matrán-Bea[b,1]

[a] *Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Avda de Las Castros, s/n, E-39071 Santander, Spain*

[b] *Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Valladolid, Spain*

Received 25 March 1992; revised 30 April 1993

Abstract

A measure of departures of monotonicity of a given function, the L_r-DIP, $1 \leqslant r \leqslant \infty$, is introduced. Our analysis is performed to cover two different situations: When the function is known our interest is related to its behavior in a stochastic model. However, in most cases, the knowledge of the function is obtained through a preliminary estimation of the function. In both situations the aim focuses in the obtainment of strong consistency results.

Keywords: Monotone functions; Quantile function; Stochastic behavior; L_r-norm; L_r-DIP; L_r-best monotone approximation; Consistency; Nonparametric regression

1. Introduction

Probability theory and statistics are often concerned in a natural way with monotone functions. The distribution and quantile functions or the monotone density functions are basic tools in probability and provide examples of the use and importance of monotone functions in probability theory.

On the other hand, monotone functions, or, in a more general setting, isotonic functions, constitute the basic elements in inference under order restrictions which, since the pioneering works by Brunk, Ewing and others, plays a prominent role in statistics. Special interest has received the problem of isotonic regression (see e.g. [1, 4, 13–15, 19, 20] and references therein), where it is assumed that the regression function $m(x) = E(Y/X = x)$ is nondecreasing.

* Corresponding author. e-mail: jcuesta@ccucvx.unican.es.

[1] Partially supported by the Spanish DGICYT under grants PB91-0306-C02-00, 01 and 02.

[2] Present address: Departamento de Matemáticas, Universidad de Oviedo, Spain.

SSDI 0377-0427(93)E0183-M

In this paper we are concerned with monotone real functions in two different directions. However in both cases we pay special attention to obtain almost sure convergence results and the problem is related with measuring departures of monotonicity.

Very often in mathematics the measurement of some kind of precision or anomaly is based on the consideration of metrics related to the problem under study. The different L_r-metrics possess a wide range of properties which make them specially suitable in these situations: typically it is possible to get a good approximation for the study of the desired problem by considering some suitable L_r-metric. In probability and statistics, this leads in a natural fashion to the usual fact of making reference to optimal properties formulated in terms of L_r-approximations. There is a vast literature about the advantages of considering any one of the L_r-norms.

Therefore, in spite of the existence of some other possibilities to measure departures from the isotonicity (for instance one might use the Haussdorf metric, see [17]), we suggest to measure departures of isotonicity, with respect to the different L_r-metrics, by considering the distance to the class of nondecreasing functions. This is carried out through the introduction of the L_r-DIP, $1 \leqslant r \leqslant \infty$, in Section 2. Notice that in the L_∞ case the DIP was introduced in [7].

As it is stated above, such an approach can be useful in two different situations. When H is a known function, we can be interested in the estimation of how increasing is its behavior with respect to a stochastic model or parent population. In Section 2 we present the results of some simulations based on two examples which may well illustrate this situation.

The other direction of our study arises in the nonparametric regression context, where the knowledge of the function is limited to a random sample relating the joint distribution of a pair of random variables. Therefore the estimation of the L_r-DIP must be based on a preliminary estimation of that function. We study this problem, in Section 4, by considering the L_r-DIP measured over the estimation of the regression. When such an estimation is sampling L_r-consistent, we obtain strong consistency of estimators of the theoretical L_r-DIP (Theorem 4.1).

In Corollary 4.3 and Theorem 4.7 we particularize our study to the well-known Nadaraya–Watson estimator, obtaining strong consistency under the usual hypotheses in the literature to get L_r-consistency of the kernel estimator.

In Section 3 we provide the technical support for both estimation problems. We prove the strong consistency of the sample version of the L_r-DIP in Theorems 3.1.4 ($1 \leqslant r < \infty$) and 3.2.3 ($r = \infty$). This section also includes the proof of the consistency of the corresponding best monotone approximations (Theorems 3.1.5 and 3.2.2).

Finally, in the Appendix we show some results related with the quantile function and Skorohod a.s. representation theorem. For example, Theorem A.1 is a natural extension of the Skorohod result, even though, for us, it can be of independent interest. Notice that these results are needed only in our proof of Theorem 3.1.5.

Remark. For simplicity, our study of monotone functions will be carried out for the nondecreasing case, even though the results will generally remain valid when working with nonincreasing functions. Therefore, from now on, the term “monotone function” means nondecreasing function, while we reserve the term “nondecreasing function” to be used when the involved result or reasoning strongly depends on that property.

2. Measuring deviations from monotonicity

Our approach to the consideration of measurements of departures from isotonicity assumptions is based on some well-known aspects of approximation theory related to the L_r-spaces, $1 \leqslant r \leqslant \infty$. Useful references are [1, 5, 10].

In this and in subsequent sections, we will make use of a rich enough probability space of reference (Ω, σ, μ), where the r.r.v.'s $X, X_1, \ldots, X_n, \ldots$ will be defined.

Let X be a r.r.v. with probability distribution given by P, and denote by $L_r(P)$, $1 \leqslant r \leqslant \infty$, the L_r-space $L_r(\mathbb{R}, \mathrm{ß}, P, \mathbb{R})$ of (classes of) real Borel functions with finite L_r-norm (w.r.t. P). Also, let us consider the closed convex cone $C_r(P)$ in $L_r(P)$ of those classes which contain a nondecreasing Borel function. Notice that we will assume that all the representative functions in a class of $C_r(P)$ are nondecreasing.

We introduce the L_r-DIP of an arbitrary function H in $L_r(P)$ through the expression

$$L_r\text{-DIP}(H) \equiv D_r(P, H) = \inf_{G \in C_r(P)} \|H - G\|_r. \tag{1}$$

Note that the right-hand side equals $\inf_{G \in C} \|H - G\|_r$, where now G ranges over the set C of all nondecreasing real-valued Borel functions. From the definition we get general properties such as

(i) $D_r(P, H) = 0$ iff H coincides P-a.s. with a nondecreasing function.

(ii) $|D_r(P, H_1) - D_r(P, H_2)| \leqslant \|H_1 - H_2\|_r$ for any $H_1, H_2 \in L_r(P)$.

Moreover the infimum in Eq. (1) is attained for any r, $1 \leqslant r \leqslant \infty$. The cases $1 < r < \infty$ follow from smoothness properties of the L_r-norms (see e.g., [18, p. 369]), while the case $p = 1$ was treated in [9] and the case $p = \infty$ was studied in [5] in the general framework of conditioning on σ-lattices (recall that the study of isotonocity is related to the consideration of measurability with respect to σ-lattices (see e.g., [3] or [1]).

The set of functions in $C_r(P)$ where the minimum in (1) is attained, i.e., the L_r-best approximations to H by elements of $C_r(P)$, will be denoted by $P^r(H)$, and the same notation will be used for any of its elements when no confusion arises. In fact, for $1 < r < \infty$, it is well known that the L_r-spaces are uniformly convex so there is unicity (P-a.s.) for the "L_r-metric projection" defined through (1) and it becomes continuous.

When based on a random sample $X_1, \ldots, X_n$ of X, the sample version of the different L_r-DIP's will obviously be defined as the r.r.v. given by

$$D_r(P_{n,\omega}, H) = \inf_{G \in C_r(P_{n,\omega})} \|H - G\|_r^{n,\omega}, \tag{2}$$

where $P_{n,\omega}$ is the sample distribution based on $X_1(\omega), \ldots, X_n(\omega)$, and the superscripts in the norm mean that it is computed with respect to $P_{n,\omega}$.

Observe that $C_r(P_{n,\omega})$ contains exactly the Borel functions which are nondecreasing on $\{X_1(\omega), \ldots, X_n(\omega)\}$.

The comments after the definition of the L_r-DIP are in force, so we can also express Eq. (2) as

$$D_r(P_{n,\omega}, H) = \inf_{G \in C} \|H - G\|_r^{n,\omega}. \tag{3}$$

As an application of the preceding ideas we have carried out some simulations to show that the estimation of how increasing a known function is has some interest in statistics.

With this aim, let us consider the functions $f_a(x) = x - a[x]$, $a > 0$, where $[x]$ denotes the integer part of the real number x. These functions describe the profit in processes like the following two:

(1) A person has to arrange the composition of a train given the number of passengers. Let us suppose that each transported passenger gives a profit of $1, that each coach can carry 100 passengers, that each additional coach supposes a cost of 50$ and that he/she has to supply as much coaches as needed (i.e., with 100 passengers only one coach is needed, but for 101 passengers two coaches are needed).

(2) Let us suppose that in a country the personal income tax is fixed as follows: Your first $1000 are free. Then for each additional $1000 or fraction you must pay $100.

Let us assume that we are interested in knowing if the profit (resp. net income) really increases when the number of passengers (resp. gross income) increases.

With the simulations which follow we try to answer those questions by taking into account that a relevant fact is the distribution of the number of passengers or that of the gross income of the people.

In this way, the cases in which the distribution is very concentrated would describe situations in which either there is little variation in the number of passengers or we are interested in a fixed narrow category in the population (for instance, the salary of the beginning teachers).

On the contrary, if the distribution is very sparse, we would be in a situation in which either the number of passengers vary very much from a time to another or we are concerned with the total population in the country.

We have carried out 50 simulations of 250 pseudo-random numbers each. We show the mean of the $\mathrm{DIP}_2(P_{250,\omega}, f_a)$ on these simulations in Tables 1(A) and 1(B) depending on P. We have chosen this distribution both as uniform or Gaussian.

The pseudo-random numbers have been generated with ZBASIC and the Gaussian distributions have been simulated through the Box–Muller algorithm.

These data support the idea we have presented: the DIP of a given function ranges from 1 to 2 or 3 depending only on the parent distribution. However it is evident that $\mathrm{DIP}_2(P_{n,\omega}, f_a)$ depends on the scale we chose to represent our function and therefore these values are relative. Even when we will not consider it in the theoretical development, it seems reasonable to normalize the DIPs to get some feeling about them. We have considered the index $\mathrm{DIP}_2(P_{n,\omega}, f_a)$ over the empirical

Table 1(A)
Simulation results for normal distribution

Mean	Standard deviation	$\mathrm{DIP}_2(P_{n,\omega}, f_a)$			$\mathrm{DIP}_2(P_{n,\omega}, f_a)/S_n$		
		$a = 0.5$	$a = 0.25$	$a = 0.1$	$a = 0.5$	$a = 0.25$	$a = 0.1$
30	0.1	0.173	0.066	0.017	0.999	0.898	0.233
30	0.5	0.099	0.035	0.008	0.331	0.087	0.018
30	1.0	0.099	0.034	0.008	0.180	0.042	0.008
30	3.0	0.094	0.032	0.007	0.059	0.013	0.002
30	5.0	0.092	0.029	0.006	0.035	0.007	0.001
30	10.0	0.083	0.024	0.005	0.016	0.003	0.000

Table 1(B)
Simulation results for uniform distribution

Interval	$\mathrm{DIP}_2(P_{n,\omega}, f_a)$			$\mathrm{DIP}_2(P_{n,\omega}, f_a)/S_n$		
	$a = 0.5$	$a = 0.25$	$a = 0.1$	$a = 0.5$	$a = 0.25$	$a = 0.1$
(0.8, 1.3)	0.142	0.049	0.012	0.999	0.655	0.115
(0.3, 1.8)	0.080	0.028	0.007	0.314	0.083	0.018
(0.8, 4.3)	0.107	0.037	0.009	0.203	0.047	0.010
(0.3, 7.8)	0.094	0.032	0.008	0.085	0.020	0.004
(0.8, 12.3)	0.099	0.033	0.007	0.058	0.013	0.002
(0.3, 12.3)	0.094	0.032	0.006	0.051	0.012	0.002

standard deviation, S_n. We have chosen that index because, if f_a were a decreasing function, then the better approximant increasing function would be a constant one. Therefore $\mathrm{DIP}_2(P_{n,\omega}, f_a)/S_n$ has two important advantages: It is scale-free and it measures how decreasing is the considered function.

In Table 1 we also show the quotient of the mean of $\mathrm{DIP}_2(P_{250,\omega}, f_a)$ in the 50 simulations over the mean of S_{250} in the 50 simulations. This column reinforces our first interpretation. From it we conclude that the function is the same but depending on the parent distribution we must consider it as increasing, as decreasing or maybe as in an intermediate case.

3. Consistency of the L_r-best monotone approximation and the L_r-DIP, $1 \leqslant r \leqslant \infty$

This section is devoted to proving consistency of the L_r-DIP when $1 \leqslant r \leqslant \infty$. Since the approaches to prove consistency are completely different depending on whether r is finite or infinite, it is preferable to consider each case separately.

We use $\{X_n\}_n$ as a common notation to denote a sequence of independent r.r.v.'s equally distributed as X with probability distribution P. The sample probability distribution based on $X_1(\omega), \ldots, X_n(\omega)$ will be denoted by $P_{n,\omega}$.

3.1. The case $1 \leqslant r < \infty$

As an intermediate step for the proof of the consistency of the L_r-DIP we consider the problem for stepwise functions, for which a technique based on [8] is useful (even for $r = 1$, in spite of the fact that the L_1-best approximation is not necessarily unique).

Lemma 3.1.2 is a particular version of [8, Lemma 2.1], taking advantage of the simplicity given by the stepwise functions under consideration. The first lemma shows that if H is a stepwise function, $H = \sum_{1 \leqslant i \leqslant k} h_i I_{A_i}$, then $D_r(P, H)$ depends on P only through the values $P(A_i)$, $i = 1, \ldots, k$.

Lemma 3.1.1. *Let $H = \sum_{1 \leqslant i \leqslant k} h_i I_{A_i}$ be a stepwise function, with $A_1 = (-\infty, a_1]$, $A_2 = (a_1, a_2], \ldots, A_k = (a_{k-1}, \infty)$, and P be a probability measure in $(\mathbb{R}, \mathcal{B})$. Then for all r in $[1, \infty)$ there* *exists a version G_r of $P^r(H)$ which is constant in every A_i, $i = 1, \ldots, k$, and bounded by* $\sup_{1 \leqslant i \leqslant k} |h_i|$; *i.e., there exists a stepwise function $G_r = \sum_{1 \leqslant i \leqslant k} t_i^r I_{A_i}$ in $P^r(H)$ with* $\sup_{1 \leqslant i \leqslant k} |t_i^r| \leqslant \sup_{1 \leqslant i \leqslant k} |h_i|$.

Proof. The last assertion follows from the fact that if $|H| \leqslant c$ and h is an increasing function, the function $h_c = \inf\{c, \sup(h, -c)\}$ is increasing and a better approximation to H than h:

$$\int |H - h|^r \,\mathrm{d}P = \int_{\{|h| \leqslant c\}} |H - h|^r \,\mathrm{d}P + \int_{\{|h| > c\}} |H - h|^r \,\mathrm{d}P$$

$$\geqslant \int_{\{|h| \leqslant c\}} |H - h_c|^r \,\mathrm{d}P + \int_{\{|h| > c\}} |H - h_c|^r \,\mathrm{d}P = \int |H - h_c|^r \,\mathrm{d}P.$$

Therefore, it suffices to prove the existence of a stepwise function, $G_r = \sum_{i \leqslant k} t_i^r I_{A_i}$, in $P^r(H)$. For this, let $f_r \in P^r(H)$ and define t_i^r, $i = 1, \ldots, k$, through the relation

$$|h_i - t_i^r| = \inf_{x \in A_i} |h_i - f_r(x)|$$

with t_i^r in the closure of the set $\{f_r(x), x \in A_i\}$.

Then the function $G_r = \sum_{1 \leqslant i \leqslant k} t_i^r I_{A_i}$ provides an obvious version of $P^r(H)$. □

Lemma 3.1.2. *Let H be the stepwise function considered in the previous lemma and P, Q be two probability measures on $(\mathbb{R}, \mathcal{B})$. For all r in $[1, \infty)$ the following inequalities hold:*

$$\left(\int |Q^r(H) - H|^r \,\mathrm{d}Q\right)^{1/r} \leqslant \left(\int |P^r(H) - H|^r \,\mathrm{d}Q\right)^{1/r}$$

$$\leqslant \left(\int |P^r(H) - H|^r \,\mathrm{d}P\right)^{1/r} + 2\left(\sup_{i = 1, \ldots, k} |h_i|\right)\left(\sum_{i=1}^{k} |P(A_i) - Q(A_i)|\right)^{1/r}.$$

Proof. The first inequality is an immediate consequence of the definition of $Q^r(H)$.

From Lemma 3.1.1 we have that

$$|P^r(H) - H|^r \leqslant \left(2 \sup_{i = 1, \ldots, k} |h_i|\right)^r,$$

hence

$$\int |P^r(H) - H|^r \,\mathrm{d}P - \int |P^r(H) - H|^r \,\mathrm{d}Q \leqslant \left(2 \sup_{i = 1, \ldots, k} |h_i|\right)^r \sum_{i=1}^{k} |P(A_i) - Q(A_i)|.$$

Now, using that $|a^s - b^s| \leqslant |a - b|^s$ if $a, b \geqslant 0$ and $0 < s \leqslant 1$, we obtain, for $s = 1/r$:

$$\left|\left(\int |P^r(H) - H|^r \,\mathrm{d}P\right)^{1/r} - \left(\int |P^r(H) - H|^r \,\mathrm{d}Q\right)^{1/r}\right| \leqslant 2 \sup_{i = 1, \ldots, k} |h_i| \left(\sum_{i=1}^{k} |P(A_i) - Q(A_i)|\right)^{1/r},$$

which proves the second inequality. □

The previous lemmas permit us to prove the consistency of the L_r-DIP, $1 \leqslant r < \infty$, for stepwise functions in the following way.

Proposition 3.1.3. *If H is the stepwise function previously considered, then*

$$D_r(P_{n,\omega}, H) \to D_r(P, H) \quad \text{as } n \to \infty,$$

for μ-a.e. ω and every r in $[1, \infty)$.

Proof. An application of the Glivenko–Cantelli theorem gives

$$\sum_{i=1}^{k} |P_{n,\omega}(A_i) - P(A_i)| \to 0 \quad \text{as } n \to \infty, \text{ for } \mu\text{-a.e. } \omega,$$

which with the strong law of large numbers and making a double use of the inequalities obtained in Lemma 3.1.2, by alternating the role of $P_{n,\omega}$ and P, implies

$$D_r(P,H) = \left(\int |P^r(H) - H|^r \,\mathrm{d}P\right)^{1/r} = \lim_{n\to\infty} D_r(P_{n,\omega}, H) \quad \text{for } \mu\text{-a.e. } \omega.$$

This completes the proof. □

Theorem 3.1.4 (Consistency of the L_r-DIP, $1 \leqslant r < \infty$). *Let $\{X_n\}_n$ be a sequence of independent r.r.v.'s with probability distribution P and let H be a function in $L_r(P)$. Then $D_r(P_{n,\omega}, H) \to D_r(P,H)$ as $n \to \infty$ for μ-a.e. ω, where $P_{n,\omega}$ is the sample distribution based on $X_1(\omega), \ldots, X_n(\omega)$.*

Proof. First observe that

$$D_r(P_{n,\omega}, H) \leqslant \left(\int |P^r(H) - H|^r \,\mathrm{d}P_{n,\omega}\right)^{1/r}$$

and use the strong law of large numbers to obtain

$$\limsup_{n\to\infty} D_r(P_{n,\omega}, H) \leqslant D_r(P,H).$$

On the other hand, it is a well-known fact that the stepwise functions constitute a dense subset of $L_r(P)$ for $1 \leqslant r < \infty$, so we can choose a sequence $\{H_k\}_k$ of stepwise functions such that

$$\left(\int |H_k - H|^r \,\mathrm{d}P\right)^{1/r} \to 0 \quad \text{as } k \to \infty.$$

Now, considering the inequalities

$$\begin{aligned} D_r(P_{n,\omega}, H) &\geqslant \left(\int |H_k - P^r_{n,\omega}(H)|^r \,\mathrm{d}P_{n,\omega}\right)^{1/r} - \left(\int |H - H_k|^r \,\mathrm{d}P_{n,\omega}\right)^{1/r} \\ &\geqslant D_r(P_{n,\omega}, H_k) - \left(\int |H - H_k|^r \,\mathrm{d}P_{n,\omega}\right)^{1/r}, \end{aligned}$$

and letting $n \to \infty$, the previous proposition and the strong law of large numbers give

$$\liminf_{n\to\infty} D_r(P_{n,\omega}, H) \geqslant D_r(P, H_k) - \left(\int |H - H_k|^r \,\mathrm{d}P\right)^{1/r} \quad \text{for } \mu\text{-a.e. } \omega.$$

Finally, let $k \to \infty$ and take into account item (ii) in Section 2 to obtain

$$\liminf_{n \to \infty} D_r(P_{n,\omega}, H) \geqslant D_r(P, H) \quad \text{for } \mu\text{-a.e. } \omega,$$

which finalizes the proof. $\square$

The consistency of the L_r-DIP is used in our proof of consistency of the L_r-best monotone approximation. We also take advantage of two results included in the Appendix and related to the Skorohod a.s. representation theorem.

Note that the assumption of uniqueness of the theoretical L_r-best monotone approximation in the hypotheses is superfluous unless $r = 1$.

Theorem 3.1.5 (Consistency of monotone L_r-approximations). *Let $\{X_n\}_n$ be a sequence of independent r.r.v.'s with probability distribution P, and let H be a function in $L_r(P)$. Assume that $P^r(H) = \{h\}$, and let $h_{n,\omega} \in P^r_{n,\omega}(H)$ where $P_{n,\omega}$ is the sample distribution based on $X_1(\omega), \ldots, X_n(\omega)$. Then, for μ-a.e. ω in Ω, the sequence $\{h_{n,\omega}\}_n$ converges to h, P-a.s.*

Proof. From the elementary inequality

$$\int |h_{n,\omega}|^r \, \mathrm{d}P_{n,\omega} \leqslant C\left[\int |H - h_{n,\omega}|^r \, \mathrm{d}P_{n,\omega} + \int |H|^r \, \mathrm{d}P_{n,\omega}\right],$$

where C is a constant which only depends on r, we have, after the strong law of large numbers applied to the last term in the sum, and the previous theorem of consistency of the L_r-DIP, that

$$\sup_n \int |h_{n,\omega}|^r \, \mathrm{d}P_{n,\omega} < \infty \quad \text{for } \mu\text{-a.e. } \omega.$$

Now let $\varepsilon > 0$ be given. Then, for μ-a.e. ω, by Markov's inequality there exists an interval $[a, b]$ such that

$$P_{n,\omega}[h_{n,\omega} \in [a, b]] > 1 - \tfrac{1}{2}\varepsilon$$

for all n, where a and b depend on ω.

Also the Glivenko–Cantelli theorem implies that for μ-a.e. ω, for large enough n, $|P_{n,\omega}(I) - P(I)| < \frac{1}{2}\varepsilon$ for each interval. Hence, taking into account the monotone character of each $h_{n,\omega}$, the set $I_{n,\omega} = \{x\colon h_{n,\omega}(x) \in [a, b]\}$ is an interval and we have

$$\begin{aligned} P[h_{n,\omega} \in [a, b]] &= P[I_{n,\omega}] \geqslant P_{n,\omega}[I_{n,\omega}] - \tfrac{1}{2}\varepsilon \\ &= P_{n,\omega}[h_{n,\omega} \in [a, b]] - \tfrac{1}{2}\varepsilon > 1 - \tfrac{1}{2}\varepsilon - \tfrac{1}{2}\varepsilon = 1 - \varepsilon. \end{aligned}$$

Therefore, for μ-a.e. ω, the sequence of probability measures induced by the random variables $h_{n,\omega}$ defined on $(\mathbb{R}, \beta, P)$ is tight. Theorem A.1 in the Appendix implies that every subsequence has a P-a.s. convergent subsequence.

Let g be a monotone function which is the P-a.s. limit of the subsequence $\{h_{n_k,\omega}\}_{n_k}$ and let $C = \{y\colon h_{n_k,\omega}(y) \to g(y)\}$.

Let Ω_0 be the probability one set obtained in Proposition A.3, let $\omega \in \Omega_0$, and let $Y_0^\omega, Y_1^\omega, Y_2^\omega, \ldots$ be the sequence of r.r.v.'s defined there. We are going to prove that

$$\lambda\{t: h_{n_k,\omega}(Y_{n_k}^\omega(t)) \to g(Y_0^\omega(t))\} = 1. \tag{4}$$

First note that if t_0 is such that g is continuous on $Y_0^\omega(t_0)$, then a standard argument based on the increasing character of g gives us that, in this case, $h_{n_k,\omega}(Y_{n_k}^\omega(t_0)) \to g(Y_0^\omega(t_0))$.

Note that the set $\boldsymbol{D} = \{t: g \text{ is not continuous in } Y_0^\omega(t) \text{ and } P[Y_0^\omega(t)] = 0\}$ is denumerable because g is increasing (therefore it is continuous except in a denumerable set) and, if $P[Y_0^\omega(t)] = 0$, then, by definition of Y_0^ω, we have that $(Y_0^\omega)^{-1}[Y_0^\omega(t)] = \{t\}$.

Therefore the only problem remains of those points t_0, such that g is not continuous in $Y_0^\omega(t_0)$ and $P[Y_0^\omega(t_0)] > 0$.

Let us suppose that t_0 belongs to the interior of the set $(Y_0^\omega)^{-1}[Y_0^\omega(t_0)]$. Then, by Proposition A.3, there exists n_0, which depends on ω and t_0, such that if $n \geqslant n_0$ then $Y_n^\omega(t_0) = Y_0^\omega(t_0)$. On the other hand, $P[Y_0^\omega(t_0)] > \lambda\{(Y_0^\omega)^{-1}[Y_0^\omega(t_0)]\} > 0$ and, then $Y_0^\omega(t_0)$ belongs to $\boldsymbol{C}$. So, also in this case, $h_{n_k,\omega}(Y_{n_k}^\omega(t_0)) \to g(Y_0^\omega(t_0))$.

Therefore we have proved that $h_{n_k,\omega}(Y_{n_k}^\omega(t)) \to g(Y_0^\omega(t))$ with the exception of those points t, which either are in $\boldsymbol{D}$ or satisfy $P[Y_0^\omega(t)] > 0$ and t does not belong to the interior of $(Y_0^\omega)^{-1}[Y_0^\omega(t)]$ which is, at most, a denumerable set and (4) is proved.

Now, if H_ε is a continuous function, then, by Eq. (4)

$$\lambda\{t: H_\varepsilon(Y_{n_k}^\omega(t)) - h_{n_k,\omega}(Y_{n_k}^\omega(t)) \to H_\varepsilon(Y_0^\omega(t)) - g(Y_0^\omega(t))\} = 1 \tag{5}$$

and by the basic integration to the limit theorem

$$\int |H_\varepsilon - g|^r \,\mathrm{d}P \leqslant \liminf_{n_k \to \infty} \int |H_\varepsilon - h_{n_k,\omega}|^r \,\mathrm{d}P_{n_k,\omega}.$$

Therefore, by considering a continuous function H_ε which satisfies that $\int |H - H_\varepsilon|^r \,\mathrm{d}P < \varepsilon^r$ and a point ω in the probability one set where the law of large numbers $\int |H - H_\varepsilon|^r \,\mathrm{d}P_{n,\omega} \to \int |H - H_\varepsilon|^r \,\mathrm{d}P$ holds, we have

$$\begin{aligned}
\left(\int |H - g|^r \,\mathrm{d}P\right)^{1/r} - \varepsilon &\leqslant \left(\int |H_\varepsilon - g|^r \,\mathrm{d}P\right)^{1/r} - \left(\int |H - H_\varepsilon|^r \,\mathrm{d}P\right)^{1/r} \\
&\leqslant \liminf_{n_k} \left(\int |H_\varepsilon - h_{n_k,\omega}|^r \,\mathrm{d}P_{n_k,\omega}\right)^{1/r} - \liminf_{n_k} \left(\int |H - H_\varepsilon|^r \,\mathrm{d}P_{n_k,\omega}\right)^{1/r} \\
&\leqslant \liminf_{n_k} \left(\int |H - h_{n_k,\omega}|^r \,\mathrm{d}P_{n_k,\omega}\right)^{1/r} = \lim_{n_k} D_r(P_{n_k,\omega}, H).
\end{aligned}$$

Finally, the assumption of uniqueness of the theoretical L_r-best monotone approximation, and the consistency of the L_r-DIP provided by Theorem 3.1.4, give us the result. □

3.2. The case $r = \infty$

The double problem of consistency for both the L_∞-DIP and the L_∞-best monotone approximation of a P-a.s. bounded function H have an easy joint solution based on the paper [5].

First recall [5, Theorem 4.1.3] that, in general, the set $P^{\infty}(H)$ of all L_{∞}-best approximations to H by monotone essentially bounded functions is not a singleton and can be described as the set of functions $\{f$: monotone and such that $U - D_{\infty}(P,H) \leqslant f \leqslant L + D_{\infty}(P,H)$ P-a.s.$\}$. Here U (resp. L) denotes for the essential infimum (resp. supremum) of the monotone functions h, satisfying $h \geqslant H$ (resp. $h \leqslant H$) P-a.s.

Moreover the conditional midrange, defined as $M = \frac{1}{2}(U + L)$, is a distinguished element in $P^{\infty}(H)$, and our consistency result will be precisely the consistency of the conditional midrange.

Indeed, $D_{\infty}(P,H)$ can be also described in terms of L and U (see [5, Theorem 4.1.4]):

$$D_{\infty}(P,H) = \tfrac{1}{2}\|U - L\|_{\infty}.$$

Let us consider the nondecreasing functions U and L. It is easy to show that versions of U and L are given, respectively, by

$$U(x) = \begin{cases} P\text{-essup}\{H(y),\ y \leqslant x\}, & \text{if } P[(-\infty, x]] > 0,\\ P\text{-esinf}\{H(y),\ y \in \mathbb{R}\}, & \text{if } P[(-\infty, x]] = 0, \end{cases}$$

and

$$L(x) = \begin{cases} P\text{-esinf}\{H(y),\ y \geqslant x\}, & \text{if } P[[x, \infty)] > 0,\\ P\text{-essup}\{H(y),\ y \in \mathbb{R}\}, & \text{if } P[[x, \infty)] = 0. \end{cases}$$

Therefore the a.s. consistency of the conditional midrange, $M_{n,\omega}$, is an obvious consequence of that one of the conditional bounds, $U_{n,\omega}$ and $L_{n,\omega}$, which follow from the next lemma.

Lemma 3.2.1. *Let $\{X_n\}$ be a sequence of independent r.r.v.'s with probability distribution P, and let $P_{n,\omega}$ denote the empirical probability distribution based on $X_1(\omega), \ldots, X_n(\omega)$. For every bounded nonempty Borel set A in $\mathbb{R}$ with $P(A) > 0$ we have that*

$$P_{n,\omega}\text{-essup}\,A \to P\text{-essup}\,A \quad \textit{for } \mu\textit{-a.e. } \omega,$$

$$P_{n,\omega}\text{-esinf}\,A \to P\text{-esinf}\,A \quad \textit{for } \mu\textit{-a.e. } \omega.$$

Proof. The proofs of both convergences are similar, so we will only consider the first.

Let $\pi = P\text{-essup}\,A$. Then $\mu\{X \in A,\ X > \pi\} = 0$, so

$$\mu\{\omega\colon X_i(\omega) \in A \text{ and } X_i(\omega) > \pi \text{ for some } i\} = 0,$$

i.e., $P_{n,\omega}\text{-essup}\,A \leqslant \pi$ for μ-a.e. ω.

On the other hand, for every $\varepsilon > 0$ we have that $\mu\{X \in A, X > \pi - \varepsilon\} > 0$, hence $\mu\{\omega\colon X_i(\omega) \in A$ and $X_i(\omega) > \pi - \varepsilon$ for some $i\} = 1$, which proves that $P_{n,\omega}\text{-essup}\,A > \pi - \varepsilon$ from an index on for μ-a.e. ω. □

Now we have the announced consistencies of the conditional bounds and midrange via the standard argument of considering a μ-probability one set, where the convergences $U_{n,\omega}(x_k) \to U(x_k)$ and $L_{n,\omega}(x_k) \to L(x_k)$, $k = 1, 2, \ldots$, hold for a convergence-determining denumerable set.

Theorem 3.2.2 (Consistency of monotone L_∞-approximations). *Let H be a function in $L_\infty(P)$. The sequence of sample L_∞-best approximations of H by monotone functions given by the sample conditional midranges, $M_{n,\omega}$, converges to the theoretical one, M, for μ-a.e. ω.*

The previous considerations about $D_\infty(P,H)$ relate their consistency to that of the conditional bounds and make possible a simple proof of the consistency of the L_∞-DIP.

Theorem 3.2.3 (Consistency of the L_∞-DIP). *Let $\{X_n\}$ be a sequence of independent r.r.v.'s with probability distribution P and let $P_{n,\omega}$ be the sample probability distribution based on $X_1(\omega),\ldots,X_n(\omega)$. Then for every function H in $L_\infty(P)$ we have $D_\infty(P_{n,\omega},H)\to D_\infty(P,\ H)$ as $n\to\infty$, for μ-a.e. ω.*

Proof. First observe that for any function G, the inequality $|G|\leqslant\|G\|_\infty$ holds P-a.s., whence $\mu\{\omega$: $G(X_i(\omega))>\|G\|_\infty$ for some $i\}=0$ and $\mu\{\omega$: $\|G\|_\infty^{n,\omega}\leqslant\|G\|_\infty$ for every $n\}=1$.

Therefore we have $\|H-P_{n,\omega}^\infty(H)\|_\infty^{n,\omega}\leqslant\|H-P^\infty(H)\|_\infty^{n,\omega}\leqslant\|H-P^\infty(H)\|_\infty$, and, then

$$\limsup_{n\to\infty} D_\infty(P_{n,\omega},H)\leqslant D_\infty(P,H)\quad \text{for } \mu\text{-a.e. } \omega.$$

So, if $D_\infty(P,H)=0$, the proof is finished. In the other case, let us denote $d=\|U-L\|_\infty$ and let ε be such that $d>\varepsilon>0$. Then $P[|U-L|>d-\varepsilon]>0$, and the monotone character of U and L implies that there exists an interval (possibly degenerate) $[a,b]$ such that

$$0<\mu[X\in[a,b]]\quad\text{and}\quad [a,b]\subset\{|U-L|>d-\varepsilon\}.$$

Therefore for μ-a.e. ω there exists some n_0 $(=n_0(\omega))$ such that $X_{n_0}(\omega)\in[a,b]$.

Finally, by taking into account that $L_{n,\omega}\to L$ and $U_{n,\omega}\to U$ for μ-a.e. ω as $n\to\infty$, we have also

$$|U_{n,\omega}(X_{n_0}(\omega))-L_{n,\omega}(X_{n_0}(\omega))|>d-\varepsilon$$

from an index on for μ-a.e. ω, then $P_{n,\omega}\{|U_{n,\omega}-L_{n,\omega}|>d-\varepsilon\}>0$ from that index on and $\liminf_{n\to\infty}\|U_{n,\omega}-L_{n,\omega}\|_\infty^{n,\omega}>d-\varepsilon$ for μ-a.e. ω.

Whence $D_\infty(P_{n,\omega},H)=\frac{1}{2}\|U_{n,\omega}-L_{n,\omega}\|_\infty^{n,\omega}\to\frac{1}{2}\|U-L\|_\infty=D_\infty(P,H)$ for μ-a.e. ω. □

4. Consistency of the L_r-DIP in nonparametric regression

In the framework of nonparametric regression estimation, the purpose is to estimate the regression function $m(x)=E(Y/X=x)$ from a sample of independent random vectors $\{(X_n,Y_n)\}$ with the same distribution as (X,Y).

Let m_n be a nonparametric estimator of the function m. We begin by stating a basic result in which we propose a simple condition under what the sample L_r-DIP of m_n provides a strongly consistent estimation of the L_r-DIP of m. Then we analyze the question of whether the Nadaraya–Watson estimator satisfies this condition.

Note that in this result we do not distinguish between the cases in which r is finite or infinite.

Theorem 4.1. *Let $\{(X_n, Y_n)\}_n$ be a sequence of independent identically distributed two-dimensional random vectors and let $m(x) = E(Y/X = x)$. Let us assume that $\|Y\|_r < \infty$, $1 \leqslant r \leqslant \infty$. If $\{m_n\}_n$ is a sequence of nonparametric estimators of m such that $\|m_n - m\|_r^{n,\omega} \to 0$, μ-a.e. as $n \to \infty$, then we have that*

$$D_r(P_{n,\omega}, m_n) \to D_r(P, m), \quad \mu\text{-a.e.} \quad \text{as } n \to \infty.$$

Proof. Taking into account inequality (ii) in Section 2, we have that

$$|D_r(P_{n,\omega}, m_n) - D_r(P, m)| \leqslant \|m_n - m\|_r^{n,\omega} + |D_r(P_{n,\omega}, m) - D_r(P, m)|.$$

By hypothesis, the first term on the right-hand side converges to zero μ-a.e.

With respect to the second one, we have that the hypothesis on the L_r-norm of Y implies that $\|m(X)\|_r < \infty$ and to get the result we can apply either Theorem 3.1.4 if r is finite or Theorem 3.2.3 if r is infinite. □

Next we apply the preceding result to the well-known Nadaraya–Watson estimators which are defined by

$$m_n(x) = \sum_{i=1}^{n} \frac{Y_i K[(X_i - x)/h_n]}{\sum_{j=1}^{n} K[(X_j - x)/h_n]}, \quad \text{if } \textstyle\sum_{j=1}^{n} K[(X_j - x)/h_n] \neq 0, \text{ and 0 otherwise.}$$

The hypotheses about the real random variables X, Y, or about the involved kernel, K, and the bandwidths will be of a different nature as those in the available literature in the treatment of the consistency of kernel estimators in the L_r, $1 \leqslant r < \infty$, and the uniform senses.

According to Theorem 4.1 we only need to check whether the condition $\|m_n - m\|_r^{n,\omega} \to 0$, μ-a.e. as $n \to \infty$ holds or not. We analyze separately the cases $1 \leqslant r < \infty$ and $r = \infty$.

The first case is solved in the next result. Its proof will be given in a future paper and it is quoted here for the sake of completeness.

Proposition 4.2. *Let us assume the following hypotheses.*

(1) *On the kernel:*

(i) *There exists $0 < r < r' < \infty$ and $0 < \beta < b < \infty$ such that*

$$\beta I_{B(0,r)} \leqslant K \leqslant b I_{B(0,r')},$$

where $B(0,s)$ means the closed ball with center in 0 and radius s.

(ii) $\int K(u)\,\mathrm{d}u = 1$.

(2) *On the sequence of bandwidths:*

$$\sum_n \mathrm{e}^{-n \cdot \alpha \cdot h_n} < \infty \quad \textit{for every } \alpha > 0.$$

(3) *On the distribution of the random variables:*

(i) $E|Y|^r < \infty$, $r \geqslant 1$.

(ii) *The distribution of X admits a continuous density function with respect to the Lebesgue measure.*

Then

$$\|m_n - m\|_r^{n,\omega} \to 0, \ \mu\text{-a.e.} \quad \text{as } n \to \infty.$$

Therefore we have proved the following corollary.

Corollary 4.3. *Let $\{m_n\}$ be a sequence of kernel estimators of the regression function m. Under the hypotheses in Proposition* 4.2 *we have that*

$$D_r(P_{n,\omega}, m_n) \to D_r(P, m), \ \mu\text{-a.e.} \quad \text{as } n \to \infty.$$

More involved hypotheses are necessary in our study of the case $r = \infty$ for which we give a complete analysis of it.

Since we will occasionally be concerned with the L_∞-DIP measured over subsets of the line, let us briefly mention some terminology and facts related to this peculiarity.

First note that the definition of the norm $\|\_\|_\infty$ or that of the L_∞-DIP can be changed to cover any positive finite measure with the same properties as those obtained before. Therefore, given a Borel set, A, let us consider the measures $P_{n|A}(B) = P_{n,\omega}(A \cap B)$, and $P_{|A}(B) = P(A \cap B)$, and let us denote the L_∞-seminorm associated with $P_{n|A}$ by $\|\_\|_A^n$ (note that $\|H\|_A^n = \sup_{x \in \{X_1, \dots, X_n\} \cap A} |H(x)|$).

The L_∞-DIP measured with respect to the measures $P_{|A}$ or $P_{n|A}$ will be denoted by D_A and D_A^n, respectively. Obviously, the basic inequality in Theorem 4.1 holds also in this case and we have that

$$|D_A^n(m_n) - D_A(m)| \leqslant \|m_n - m\|_A^n + |D_A^n(m) - D_A(m)|. \tag{6}$$

Now a slight modification of Theorem 3.2.3 gives us that $|D_A^n(m) - D_A(m)| \to 0$ μ-a.e.

On the other hand, if A is a bounded set, conditions are known (see e.g., [6, 11, 12]) under which the first term in Eq. (6) converges to zero μ-a.s. So we have the consistency of the kernel-based estimation of the L_∞-DIP measured over compact sets.

Next we are going to prove that it is possible to estimate consistently the L_∞-DIP over $\mathbb{R}$ by using a kernel estimation of m. The idea is to continue using the decomposition (6) but over a sequence of empirically chosen sets $\{A_n\}$. The problem here is that if we choose too large sets, then $\|m_n - m\|_{A_n}^n \to 0$ fails, but if we choose them too small, the convergence $D_{A_n}^n(m) \to D_\infty(P, m)$ fails. Having this in mind, in Theorem 4.7 we propose an empirical election criteria which handles both problems.

As usual in the related literature, it is useful to consider the following functions:

$$u_n(x) = \frac{1}{n h_n} \sum_{i=1}^{n} Y_i K[(X_i - x)/h_n],$$

$$g_n(x) = \frac{1}{n h_n} \sum_{i=1}^{n} K[(X_i - x)/h_n],$$

so g_n is the well-known kernel estimator for the marginal density, g, of X. With this notation the estimator m_n becomes

$$m_n = \frac{u_n}{g_n}, \ \text{if } g_n \neq 0, \quad \text{and 0 otherwise.}$$

Our hypotheses (see below) are essentially those in [12].

We begin with two important properties for which we need the following hypotheses: The random vector (X, Y) has a joint density $f(x, y)$. The marginal density of X is denoted by g; therefore, if $u(x) = \int yf(x, y)\,dy$, then u/g is a version of the regression function, m. The assumptions on the kernel K are:

(AK)
(a) K is continuous and of bounded variation,
(b) K is absolutely integrable w.r.t. Lebesgue measure on the line,
(c) $\int |x \log|x|^{1/2}||dK(x)| < \infty$.

Then, if we denote

$$\theta_n = \left|\frac{1}{n \cdot h_n}\log\frac{1}{h_n}\right|^{1/2}$$

we have (see [12]) the following result.

Proposition 4.4. *Suppose that* $E|Y|^s < \infty$ *for some* $s \geqslant 2$ *and that* $\sup_x \int |y|^s f(x, y)\,dy < \infty$. *Let us assume that* K *satisfies* (AK), *that* $\{h_n\}_n$ *are such that* $\sum_n h_n^\lambda < \infty$ *and* $n^{2\eta-1}h_n \to \infty$ *for some* $\eta < 1 - s^{-1}$ *and* $\lambda > 0$.

Then $\sup_x |u_n(x) - E[u_n(x)]| = O(\theta_n)$, *μ-a.e.*

Proposition 4.5. *If* K *satisfies* (AK), g *is continuous and* $\{h_n\}_n$ *satisfies that* $\sum_n h_n^\lambda < \infty$ *and that* $n^{1-\varepsilon}h_n \to \infty$ *for some* $\lambda > 0$ *and* $\varepsilon > 0$, *then*

$$\sup_x |g_n(x) - E[g_n(x)]| = O(\theta_n),\ \mu\text{-a.e.}$$

Additional requirements for the distribution of (X, Y), bandwidths and the kernel are the following:

(RD)
(a) there exists $s \geqslant 2$ with $E|Y|^s < \infty$ and $\sup_x \int |y|^s f(x, y)\,dy < \infty$,
(b) g and u have bounded second derivatives,
(c) there exists H such that $u/g \leqslant H$ on the set $\{g > 0\}$,

(RB)
(a) $n^{2\eta-1}h_n \to \infty$ for some $\eta < 1 - s^{-1}$, where s satisfies (a) in (RD),
(b) $\sum_{n=1}^{\infty} h_n^\lambda < \infty$ for some $\lambda > 0$,
(c) $h_n^2 = O(\theta_n)$,

(RK)
(a) $\int K(x)\,dx = 1$,
(b) $\int xK(x)\,dx = 0$,
(c) $\int x^2|K(x)|\,dx < \infty$.

These assumptions permit us to bound the bias in the next proposition. Note that assumptions (RB) imply that $n^{1-\varepsilon}h_n \to \infty$ for some $\varepsilon > 0$, and they hold when $s > \frac{5}{2}$ and $h_n \in [n^{-\alpha}, n^{-\beta}]$, with $1 - 2/s > \alpha > \beta > \frac{1}{5}$.

Proposition 4.6. *If assumptions* (AK), (RD), (RB) *and* (RK) *hold then*

$$\max\{\sup_x |g_n(x) - g(x)|, \sup_x |u_n(x) - u(x)|\} = O(\theta_n),\ \mu\text{-a.e.}$$

Proof. After Propositions 4.4 and 4.5 it suffices to show that

$$\max\{\sup_x |E[g_n(x)] - g(x)|, \sup_x |E[u_n(x)] - u(x)|\} = \mathrm{O}(\theta_n)$$

but we will only prove that

$$\sup_x |E[g_n(x)] - g(x)| = \mathrm{O}(\theta_n),$$

because the other asymptotic bound is proved analogously.

Let n be a positive integer and let $x \in \mathbb{R}$. For every t in $\mathbb{R}$ we can choose $\xi_{x,t}$ in the interval $[x, x + th_n]$ in such a way that

$$\begin{aligned} E[g_n(x)] &= \int K\left(\frac{v - x}{h_n}\right) g(v)\,\mathrm{d}v = \int K(t) g(x + th_n)\,\mathrm{d}t \\ &= \int K(t)[g(x) + th_n g'(x) + t^2 h_n^2 g''(\xi_{x,t})]\,\mathrm{d}t \end{aligned}$$

and the result follows from (RD) (b), (RB) (c) and (RK) (c). □

The announced estimator of $D_\infty(P, m)$ and its strong consistency appear in the following theorem.

Theorem 4.7. *Let* (X, Y), k *and* $\{h_n\}_n$ *be such that they satisfy assumptions* (AK), (RD), (RB) *and* (RK). *Let* $A_n = \{X_i\colon 1 \leqslant i \leqslant n \text{ and } g_n(X_i) \geqslant c_n\}$, *where* $\{c_n\}_n$ *is a sequence of positive numbers such that* $c_n \to 0$, *and* $\theta_n = \mathrm{o}(c_n)$ *as* $n \to \infty$. *Then*

$$D^n_{A_n}(m_n) \to D_\infty(P, m), \ \mu\text{-a.e.}$$

Proof. Since $P\{x, g(x) > 0\} = 1$ we have $g(X_i) > 0$ for all i, μ-a.e. The definition of A_n implies

$$\|m_n - m\|^n_{A_n} = \left\| \frac{u_n}{g_n} - \frac{u}{g} \right\|^n_{A_n} = \left\| \frac{u_n - u}{g_n} - \frac{u}{g}\frac{g - g_n}{g_n} \right\|^n_{A_n} \quad \text{for every } n,\ \mu\text{-a.e.}$$

Now the assumptions (RB) imply

$$\|m_n - m\|^n_{A_n} < \frac{1}{c_n}\left[\sup_x |u_n(x) - u(x)| + H \sup_x |g_n(x) - g(x)| \right],$$

which from Proposition 4.6 and the choice of $\{c_n\}_n$ gives

$$\|m_n - m\|^n_{A_n} \to 0, \ \mu\text{-a.e.}$$

Therefore, the inequality

$$|D^n_{A_n}(m_n) - D_\infty(P, m)| \leqslant \|m_n - m\|^n_{A_n} + |D^n_{A_n}(m) - D_\infty(P, m)|$$

shows that the theorem would be proved if the μ-a.e. convergence $D^n_{A_n}(m) \to D_\infty(P, m)$ holds.

To show this, let $M_\infty \in P^\infty(m)$ (for instance, the conditional midrange), then

$$D^n_{A_n}(m) \leqslant \|m - M_\infty\|^n_{A_n} \leqslant \|m - M_\infty\|^n_\infty \quad \text{for all } n,$$

so $\limsup_n D^n_{A_n}(m) \leqslant \lim_n \|m - M_\infty\|^n_\infty = \|m - M_\infty\|_\infty = D_\infty(P,m)$, μ-a.e.

On the other hand, the consistency of the L_∞-DIP, Proposition 4.6 and the fact that $P\{x: g(x) > 0\} = 1$, imply that:

(a) $D_\infty(P_n, m) \to D_\infty(P,m)$, μ-a.e,

(b) $g(X_n) > 0$ for all n, μ-a.e,

(c) for every $\varepsilon > 0$,

$$\{g > \varepsilon\} \subset \liminf_n \{g_n > \tfrac{1}{2}\varepsilon\} \subset \liminf_n \{g_n \geqslant c_n\}, \quad \mu\text{-a.e},$$

hence

(c′) $\{g > 0\} \subset \liminf_n \{g_n \geqslant c_n\}$, μ-a.e.

Therefore (b), (c′) and the definition of A_n imply that $D_\infty(P_k, m) \leqslant \liminf_n D^n_{A_n}(m)$ μ-a.e. for every k, and from (a):

$$\liminf_n D^n_{A_n}(m) \geqslant D_\infty(P,m) \quad \mu\text{-a.e.},$$

which finalizes the proof. □

Acknowledgements

The authors wish to thank Prof. S.T. Rachev and an anonymous referee for some valuable comments which have improved considerably the paper.

Appendix

The following theorem can be considered as a natural extension of the Skorohod a.s. representation theorem for the weak convergence of probability measures on the line (see [2, p. 343]).

Theorem A.1. *Let $\{f_n\}_n$ be a sequence of monotone functions on the line, and let X be a r.r.v. defined on the probability space (Ω, σ, μ). Then the sequence $\{f_n(X)\}_n$ converges in distribution if and only if it converges μ-almost surely. Moreover, in this case, the limit r.r.v. can be written as $f(X)$ where $f = \lim_n f_n$ P_X-a.s.*

Proof. A direct proof is possible, but the following one is a very simple consequence of the Skorohod theorem.

First note the easy fact that if we denote by $F_n^{-1}(U)$ the quantile function associated with the distribution function F_n of the r.v. $f_n(X)$, and U is a uniformly distributed in $(0,1)$ r.v., then $(f_1(X), \ldots, f_n(X))$ and $(F_1^{-1}(U), \ldots, F_n^{-1}(U))$ have the same distribution for every n. Therefore the whole sequences $\{f_n(X)\}_n$ and $\{F_n^{-1}(U)\}_n$ are equally distributed, so the μ-a.s. convergence of $\{f_n(X)\}_n$ is equivalent to the a.s. convergence of the quantile sequence $\{F_n^{-1}(U)\}_n$, which, after the Skorohod theorem, is equivalent to the weak convergence of the sequence of D.F.'s $\{F_n\}_n$, i.e., it is equivalent to the convergence in distribution of $\{f_n(X)\}_n$.

The second part is trivial from the first one. $\square$

Some immediate consequences which arise from the given proof are collected in the following corollary.

Corollary A.2. *Let f_1 and f_2 be nondecreasing functions on the line, and let X be an r.v. defined on the probability space (Ω, σ, μ).*

(i) *If $f_1(X)$ and $f_2(X)$ are equally distributed, then $f_1(X) = f_2(X)$, μ-a.s.*

(ii) *If Y and Z are r.r.v.'s (possibly defined on a different probability space) such that $f_1(X) =^d Y$ and $f_2(X) =^d Z$, then*

$$E|f_1(X) - f_2(X)|^r \leqslant E|Y - Z|^r, \quad 1 \leqslant r < \infty.$$

(iii) *The quantile function F^{-1} is the (a.s.) only nondecreasing function on the unit interval whose D.F. is F (obviously the a.s. statement here is related to the Lebesgue measure on the unit interval).*

Proof. Statements (i) and (ii) are immediate consequences of the fact that $(f_1(X), f_2(X)) =^d (F_1^{-1}(U), F_2^{-1}(U))$ and known results for the quantile functions (see e.g., [16]), while (iii) is trivial from (i). $\square$

When the involved probabilities are the sample ones, we can improve on the Skorohod theorem through the following proposition.

Proposition A.3. *For μ-a.e. ω in Ω:*

(1) *the empirical distribution functions converge, uniformly, to the theoretical one.*

Moreover there exists a sequence $Y_0^\omega, Y_1^\omega, Y_2^\omega, \ldots$ of r.r.v.'s defined in the probability space $((0,1), \beta, \lambda)$, where λ is the Lebesgue measure, such that:

(2) *$Y_n^\omega \to Y_0^\omega$, λ-a.s.,*

(3) *the distribution law of Y_0^ω is P and that one of Y_n^ω is $P_{n,\omega}$, $n = 1, 2, \ldots,$*

(4) *let a be in $\mathbb{R}$ and t in the interior of the set $(Y_0^\omega)^{-1}(a)$. Then there exists n_0 $(= n_0(\omega))$ such that if $n \geqslant n_0$, then $Y_n^\omega(t) = a$.*

Proof. Let Ω_0 be the set in which the Glivenko–Cantelli theorem is satisfied and let us denote by F_n^{-1} $(= (F_{n,\omega})^{-1})$, $n = 1, 2, \ldots$, and F_0^{-1} the quantile function for the empirical and theoretical distributions, respectively. It is well known that if we take ω in Ω_0 and

$$Y_n^\omega = F_n^{-1}, \quad n = 0, 1, 2, \ldots,$$

then (2) and (3) are verified (see, for instance, [2, p. 190]).

With respect to (4), let a in $\mathbb{R}$ be such that the interior of $(Y_0^\omega)^{-1}(a)$ is not empty. (Note that this is equivalent to $P\{a\} > 0$.) Let t_0, t_1, t_2 be three points in the interior of the set $(Y_0^\omega)^{-1}(a)$ such that $t_1 < t_0 < t_2$. Then, by definition of F_0^{-1}, it must be

$$P(-\infty, a) < t_1 < t_0 < t_2 < P(-\infty, a].$$

Now taking into account that in this ω the empirical functions converge uniformly to the theoretical one, there exists n_0 $(= n_0(\omega))$ such that if $n \geqslant n_0$ then

$$P_{n,\omega}(-\infty, a) < t_1 < t_0 < t_2 < P_{n,\omega}(-\infty, a].$$

Therefore, if $n \geqslant n_0$, by definition of Y_n^{ω}, we have that $Y_n^{\omega}(t_0) = a$. $\square$

References

[1] R.E. Barlow, D.J. Bartholomew, J.M. Bremner and H.D. Brunk, *Statistical Inference Under Order Restrictions* (Wiley, New York, 1972).

[2] P. Billingsley, *Probability and Measure* (Wiley, New York, 1986).

[3] H.D. Brunk, *Estimation of Isotonic Regression. Nonparametric Techniques in Statistical Inference* (Cambridge Univ. Press, New York, 1970).

[4] H.D. Brunk, G.M. Ewing and W.R. Utz, Minimizing integrals in certain classes of monotone functions, *Pacific J. Math.* **7** (1957) 833–847.

[5] J.A. Cuesta and C. Matrán, Conditional bounds and best L_∞-approximations in probability spaces, *J. Approx. Theory* **56** (1989) 1–12.

[6] L.P. Devroye, The uniform convergence of the Nadaraya–Watson regression function estimate, *Canad. J. Statist.* **6** (2) (1978) 179–191.

[7] J.A. Hartigan and P.M. Hartigan, The dip test of unimodality, *Ann. Statist.* **13** (1) (1985) 70–84.

[8] D. Landers and L. Rogge, Convergence of conditional P-means given a σ-lattice, *Ann. Probab.* **4** (1976) 145–150.

[9] D. Landers and L. Rogge, Best approximants in L_Φ-spaces, *Z. Wahrsch. Verw. Gebiete* **51** (1980) 215–237.

[10] D. Landers and L. Rogge, Isotonic approximation in L_s, *J. Approx. Theory* **31** (1981) 199–223.

[11] H. Liero, Strong uniform consistency of nonparametric regression function estimates, *Probab. Theory Related Fields* **82** (1989) 587–614.

[12] Y.P. Mack and B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, *Z. Wahrsch. Verw. Gebiete* **61** (1982) 405–415.

[13] G.G. Makowski, Laws of the iterated logarithm for permuted random variables and regression applications, *Ann. Statist.* **5** (1973) 872–887.

[14] E. Mammen, Estimating a smooth monotone regression function, *Ann. Statist.* **19** (1991) 724–740.

[15] H. Mukherjee, Monotone nonparametric regression, *Ann. Statist.* **16** (2) (1988) 741–750.

[16] S.T. Rachev, The Monge–Kantorovich mass transference problem and its stochastic applications, *Theory Probab. Appl.* **29** (1984) 647–676.

[17] B. Sendov, Some problems of the theory of approximations of functions and sets in the Hausdorf metrics, *Uspekhi Mat. Nauk* **24** (5) (1969) 141–173.

[18] I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces* (Springer, Berlin, 1970).

[19] F.T. Wright, The asymptotic behavior of monotone regression estimates, *Ann. Statist.* **9** (1981) 443–448.

[20] F.T. Wright, Monotone regression estimates for grouped observations, *Ann. Statist.* **10** (1982) 278–286.