Statistics and Probability Letters ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A note on the universal consistency of the kernel distribution function estimator

José E. Chacón [a,*], Alberto Rodríguez-Casal [b]

[a] *Departamento de Matemáticas, Universidad de Extremadura, Badajoz, Spain*

[b] *Departamento de Estatística e Investigación Operativa, Universidade de Santiago de Compostela, Santiago de Compostela, Spain*

ARTICLE INFO

Article history:
Received 12 January 2010
Received in revised form 26 April 2010
Accepted 9 May 2010
Available online xxxx

Keywords:
Data-dependent bandwidth
Distribution function
Kernel estimator
Minimal smoothness assumptions
Uniform in bandwidth consistency

ABSTRACT

The problem of universal consistency of data driven bandwidth selectors for the kernel distribution estimator is analyzed. We provide a uniform in bandwidth result for the kernel estimate of a continuous distribution function. Our smoothness assumption is minimal in the sense that if the true distribution function has some discontinuity then the kernel estimate is no longer consistent.

1. Introduction

There is an increasing interest in obtaining so-called uniform in bandwidth (UiB) results for nonparametric estimators depending on a bandwidth sequence. Although these kinds of arguments had been used sparsely in the nonparametric literature before (see for instance Chapter 6 in Devroye and Györfi, 1985; Marron and Härdle, 1986, or Devroye, 1989), the first paper that focused on obtaining UiB results for the kernel density estimator making use of empirical processes techniques was Einmahl and Mason (2005). After this seminal paper many other works studied these UiB problems in different contexts, for example for local polynomial regression (Dony et al., 2006), for the local uniform empirical process (Varron, 2006), for conditional U-statistics (Dony and Mason, 2008), for the estimation of integral functionals of the density (Giné and Mason, 2008) and for kernel distribution function estimators and the smoothed empirical process (Mason and Swanepoel, in press). See also Chapter 2 in Dony (2008) for a recent detailed review on the subject.

In this paper we concentrate on kernel distribution function estimators. If $X_1, \ldots, X_n$ is a random sample drawn from a distribution function F, the kernel estimate of F is given by

$$F_{nh}(x) = \frac{1}{n}\sum_{i=1}^{n} L\left(\frac{x - X_i}{h}\right),$$

where L is a fixed distribution function and $h > 0$ is the bandwidth. This estimator was considered for the first time in Nadaraya (1964), and it is constructed by integrating out the Parzen–Rosenblatt kernel density estimate (Parzen, 1962; Rosenblatt, 1956). A recent paper on the subject is Giné and Nickl (2009). The almost sure uniform consistency of F_{nh} was established in Yamato (1973) with the only smoothness condition that F be continuous. In Section 2 we show that this

* Corresponding author.
E-mail addresses: jechacon@unex.es (J.E. Chacón), alberto.rodriguez.casal@usc.es (A. Rodríguez-Casal).

0167-7152/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2010.05.007

smoothness condition is minimal, in the sense that F_{nh} with $h > 0$ is not consistent if F is discontinuous at some point. Moreover, below we show a stronger version of Yamato's result: we prove that all modes of convergence (in probability, almost sure, complete) of the kernel distribution estimator are equivalent with respect to the uniform distance. This peculiar behaviour is shared with the kernel density estimator with respect to the L_1 distance (see Devroye, 1983).

Our main contribution (Theorem 3 below) provides an almost sure UiB consistency theorem for F_{nh}. This result improves on the existing ones in several ways: we only impose the minimal continuity condition on F and no conditions on the distribution function L, our result is uniform over a wider range of bandwidths and we obtain complete consistency instead of strong consistency; see Remark 3. We should mention, however, that our proof is not so closely linked to empirical processes theory. Indeed, our method of proof is closer to that of Devroye and Penrod (1984) for the case of the kernel density estimate.

2. Main results

2.1. An equivalence theorem

Let us denote by F_n the empirical distribution function, that is

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \Delta(x - X_i),$$

where $\Delta(x) = 1$ if $x \geq 0$ and $\Delta(x) = 0$ if $x < 0$. Using the empirical distribution function we can also write the kernel distribution estimator as

$$F_{nh}(x) = \int L\left(\frac{x-y}{h}\right) \mathrm{d}F_n(y) = \int F_n(x - hz)\mathrm{d}L(z), \tag{1}$$

with the last equality due to the fact that

$$\int \Delta(x - hz - X_i)\mathrm{d}L(z) = \int_{\{z \leq (x-X_i)/h\}} \mathrm{d}L(z) = L\left(\frac{x - X_i}{h}\right).$$

It is obvious that when $L = \Delta$ we obtain $F_{nh} = F_n$ for all values of h. In that case the smoothing parameter h does not play any role, so henceforth we will only consider proper kernel estimates, in the sense that $L \neq \Delta$. Notice however that, even if $L \neq \Delta$, we can still recover the empirical estimator since the last expression in (1) in fact makes sense for all $h \geq 0$, giving $F_{nh} = F_n$ for $h = 0$. This means that for $L \neq \Delta$ the class of kernel estimators indexed by the smoothing parameter, $\{F_{nh}: h \geq 0\}$, contains the empirical one (as well as infinitely many others).

We will measure the performance of F_{nh} using the uniform distance $\|F_{nh} - F\| = \sup_{x\in\mathbb{R}} |F_{nh}(x) - F(x)|$, which is well-defined for any distribution F. As regards this uniform distance, Yamato (1973, Thm. 3) shows two statements about the consistency of a proper kernel distribution estimate:

(Y1) For any distribution function F, if $\|F_{nh_n} - F\| \to 0$ almost surely then $h_n \to 0$.
(Y2) If F is continuous and $h_n \to 0$ then $\|F_{nh_n} - F\| \to 0$ almost surely.

In fact, in Yamato's results the condition $h_n \to 0$ is stated as $L(x/h_n) \to \Delta(x)$ as $n \to \infty$ for all $x \neq 0$, but it is not hard to check that this is equivalent to the fact that either $L = \Delta$ or $h_n \to 0$, and this reduces to $h_n \to 0$ for proper kernel estimates.

From (Y1) and (Y2) it follows that almost sure uniform consistency of F_{nh_n} is equivalent to $h_n \to 0$ when F is continuous. Next we show that this smoothness condition on F is minimal, in the sense that F_{nh_n} is not consistent for a discontinuous distribution F, unless it coincides ultimately with the empirical estimator.

Theorem 1. *Let F be a discontinuous distribution function. Then $\|F_{nh_n} - F\| \to 0$ almost surely if and only if $F_{nh_n} = F_n$ eventually, that is, if there exists n_0 such that $F_{nh_n} = F_n$ for all $n \geq n_0$.*

Remark 1. Theorem 1 represents an improvement over Proposition 2 in Giné and Nickl (2009), providing a more precise characterization of the problem of kernel estimation of a discontinuous distribution function. Furthermore, in giving a very different proof, the assumptions in Proposition 2 in Giné and Nickl (2009) are relaxed, since it is no longer necessary that L has a density.

Although we have focused on almost sure consistency so far, the distance $\|F_{nh} - F\|$ is sharply concentrated around its mean. To see this, let us denote by F^*_{nh} the kernel estimate in which one of the X_i is changed to another value, with the remaining $n-1$ data points fixed; then $\|F_{nh} - F^*_{nh}\| \leq 1/n$. Therefore by the bounded difference inequality (McDiarmid, 1989; Devroye and Lugosi, 2001, p. 8) we have

$$\mathbb{P}\left(|\|F_{nh} - F\| - \mathbb{E}\|F_{nh} - F\|| > t\right) \leq 2e^{-2nt^2}, \quad t > 0. \tag{2}$$

Consequently, by the Borel–Cantelli lemma we obtain that $\mathbb{E}\|F_{nh_n} - F\| \to 0$ is equivalent to $\|F_{nh_n} - F\| \to 0$ almost surely or in probability (since $\|F_{nh} - F\| \leq 1$). The following result, which is a stronger version of Theorem 3 in Yamato (1973), adds the complete convergence to this equivalence. Let us denote by $\mathcal{F}$ the class of all continuous distribution functions.

Theorem 2. *The following statements are equivalent:*

(i) $\|F_{nh_n} - F\| \to 0$ *in probability for some* $F \in \mathcal{F}$.
(ii) $\|F_{nh_n} - F\| \to 0$ *in probability for all* $F \in \mathcal{F}$.
(iii) $\|F_{nh_n} - F\| \to 0$ *almost surely for all* $F \in \mathcal{F}$.
(iv) *For every* $F \in \mathcal{F}$ *and every* $\varepsilon > 0$ *there exists* $n_0 = n_0(\varepsilon, F)$ *such that*

$$\mathbb{P}(\|F_{nh_n} - F\| > \varepsilon) \le 2e^{-n\varepsilon^2/2}, \quad \forall n \ge n_0,$$

that is, $\|F_{nh_n} - F\| \to 0$ *completely.*
(v) $h_n \to 0$.

As noted above, this theorem resembles the equivalence theorem for the convergence of the L_1 error of the kernel density estimate given by Devroye (1983). All modes of consistency of F_{nh} are equivalent. We also note that there is no intermediate situation for the kernel distribution function estimate: it is consistent either for all continuous distribution functions or for none of them.

However, from Theorem 2 we see the differences between kernel distribution estimation and density estimation. In density estimation we need two conditions on the bandwidth for consistency, $h_n \to 0$ and $nh_n \to \infty$. They are needed to show that the bias and the variation terms go to zero, respectively; see Devroye and Györfi (1985, Ch. 3).

Here, the uniform error of the kernel distribution estimate can also be bounded by the sum of a bias term and a variation term:

$$\|F_{nh} - F\| \le \|\mathbb{E}F_{nh} - F\| + \|F_{nh} - \mathbb{E}F_{nh}\| \tag{3}$$

(respectively). The key to understanding why we just need the condition $h_n \to 0$ in the previous theorem is that the variation term can be bounded by the uniform error of the empirical distribution function: from (1) we can write

$$\mathbb{E}F_{nh}(x) = \int L\left(\frac{x-y}{h}\right) \mathrm{d}F(y) = \int F(x - hz)\mathrm{d}L(z) \tag{4}$$

and so

$$\|F_{nh} - \mathbb{E}F_{nh}\| \le \sup_{x\in\mathbb{R}} \int |F_n(x - hz) - F(x - hz)|\mathrm{d}L(z) \le \|F_n - F\|, \tag{5}$$

and so, due to the properties of F_n, the variation term converges to zero with no conditions on the bandwidth sequence.

Remark 2. Inequality (5) resembles the situation described in Fernholz (1997), where it is shown that smoothing reduces the variance of empirical estimators of some statistical functionals.

2.2. A UiB result

Our main result, which we state next, provides a UiB theorem for kernel distribution estimators. We remark that we only impose the minimal conditions on F, L and the bandwidth sequence h_n.

Theorem 3. *If* F *is a continuous distribution function and* $h_n \to 0$ *then, for any* $\varepsilon > 0$, *there is* $n_0 = n_0(\varepsilon, F) \in \mathbb{N}$ *such that*

$$\mathbb{P}\left(\sup_{0\le h\le h_n} \|F_{nh} - F\| > \varepsilon\right) \le 2e^{-n\varepsilon^2/2} \quad \text{for all } n \ge n_0.$$

Therefore, $\sup_{0\le h\le h_n} \|F_{nh} - F\| \to 0$ *completely.*

Remark 3. Recently, using empirical processes techniques, Mason and Swanepoel (in press) obtain a very general UiB theorem for a wide class of kernel-type estimators. They apply this general theorem to the particular case of kernel distribution function estimation, proving that

$$\sup_{cn^{-1}\log n \le h \le b_n} \|F_{nh} - F\| \to 0 \quad \text{almost surely}$$

for all $c > 0$ and $0 < b_n < 1$ such that $b_n \to 0$ and $b_n \ge 1/\log n$. For this result they need F to satisfy the Lipschitz condition and $\int |z|\mathrm{d}L(z) < \infty$. Theorem 3 above improves on this by relaxing the assumption on F, allowing a wider range of bandwidths and showing complete convergence.

Remark 4. Theorem 1 in Giné and Nickl (2009) gives, for ε large enough, an exponential bound for the distance between F_{nh} and F_n under some restrictions on the distribution function F. They use their result for deriving a Dvoretzky–Kiefer–Wolfowitz-type (DKW-type) of exponential bound for the distance between F_{nh} and F. Theorem 3 above provides a more general DKW-type exponential bound, uniform on the broadest possible range of parameters h, with no restrictions on the distribution function F (apart from the necessary continuity condition) and valid for all values of ε.

Remark 5. With some additional conditions on L, Zielińsky (2007) shows that the inequality in our Theorem 3 cannot hold uniformly over the class of all continuous distribution functions, in the sense that for any sequence (h_n) there exist $\varepsilon > 0$ and $0 < \eta < 1$ such that for every $n \in \mathbb{N}$ one can find a continuous distribution function F such that

$$\mathbb{P}\left(\|F_{nh_n} - F\| > \varepsilon\right) \geq \eta.$$

Therefore it follows that, contrary to the situation with the DKW inequality for the empirical distribution function, the value of n_0 in Theorem 3 must necessarily depend on ε and the distribution F.

As in Mason and Swanepoel (in press) we remark that Theorem 3 is useful for showing consistency of kernel estimators with data-dependent bandwidths, as shown in the following corollary.

Corollary 1. *Let F be any continuous distribution function and assume that $\hat{h}_n$ is a positive measurable function of $X_1, \ldots, X_n$. If $\hat{h}_n \to 0$ completely then $\|F_{n\hat{h}_n} - F\| \to 0$ completely. The result also holds if we replace 'completely' for 'almost surely' or 'in probability'.*

Remark 6. The conditions on $\hat{h}_n$ cannot be improved, since $h_n \to 0$ is necessary for the non-random case, as shown in Theorem 2 above. In fact, it can be proved, following the same lines as in Yamato (1973), that almost sure consistency of $F_{n\hat{h}_n}$ is equivalent to $\hat{h}_n \to 0$ almost surely.

3. Proofs

Proof of Theorem 1. If F_{nh_n} is eventually F_n then consistency follows from the well-known properties of F_n. Conversely, we will see that consistency implies $L = \Delta$ or $h_n = 0$ eventually which obviously ensures that $F_{nh_n} = F_n$, eventually. If this were not the case we would have that $L \neq \Delta$ and $h_n > 0$ along a subsequence (which we will denote in the same way). If $\|F_{nh_n} - F\| \to 0$ almost surely then by (Y1) in Section 2 above we know that $h_n \to 0$. On the other hand, since $\|F_{nh_n} - F\|$ is uniformly bounded, the almost sure convergence of $\|F_{nh_n} - F\|$ implies the convergence in mean, that is, $\mathbb{E}\|F_{nh_n} - F\| \to 0$. Therefore, $\mathbb{E}F_{nh_n}(x) \to F(x)$ for all $x \in \mathbb{R}$.

Since h_n is positive and $h_n \to 0$, it follows that $F(x - h_n z) \to F(x)$ for all $z \leq 0$ whereas $F(x - h_n z) \to F(x^-)$ for $z > 0$, where $F(x^-)$ denotes the limit from the left at x. Therefore, for all $x \in \mathbb{R}$, using (4) we get

$$\lim_{n\to\infty} \mathbb{E}F_{nh_n}(x) = L(0)F(x) + (1 - L(0))F(x^-). \tag{6}$$

Let $x = x_D$ be a discontinuity point of F. Since $\mathbb{E}F_{nh_n}(x_D) \to F(x_D)$, (6) implies that

$$(1 - L(0))(F(x_D^-) - F(x_D)) = 0,$$

which is only possible if $L(0) = 1$. Hence, $L(x) = 1$ for all $x \geq 0$. This allows us to write $F_{nh_n}(x) = F_n(x) + R_{nh_n}(x)$, where

$$R_{nh_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i > x) L\left(\frac{x - X_i}{h_n}\right),$$

with $\mathbb{I}(A)$ standing for the indicator function of the set A.

Let us assume that there exists $u < 0$ such that $L(u) > 0$. In this case, let $y_n = x_D + uh_n$. Note that $y_n < x_D$ and, therefore, with probability 1,

$$R_{nh_n}(y_n) \geq \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i = x_D) L\left(\frac{y_n - x_D}{h_n}\right) = L(u)\frac{\sum_{i=1}^{n}\mathbb{I}(X_i = x_D)}{n} \to pL(u) > 0, \tag{7}$$

where $p = F(x_D) - F(x_D^-)$. On the other hand

$$R_{nh_n}(y_n) = F_{nh_n}(y_n) - F(y_n) + F(y_n) - F_n(y_n).$$

So,

$$0 \leq R_{nh_n}(y_n) \leq \|F_{nh_n} - F\| + \|F - F_n\| \to 0,$$

almost surely. This leads to a contradiction to (7). Therefore $L(u) = 0$ for all $u < 0$, so $L = \Delta$, which is itself a contradiction to the fact that $L \neq \Delta$. □

Proof of Theorem 2. Clearly it suffices to show (i) ⇒ (v) and (v) ⇒ (iv). For the first claim, using the bounded difference inequality (2), it can be easily shown that the convergence in probability implies $\mathbb{E}\|F_{nh_n} - F\| \to 0$ and therefore $\mathbb{E}F_{nh_n}(x) \to F(x)$ for all x. This leads to (v) by using Lemma 1 in Yamato (1973). To show the second implication notice that $\overline{F_n}(x) = \mathbb{E}F_{nh_n}(x) = \int F(x - h_n z)\mathrm{d}L(z)$ is a continuous distribution function (because F is continuous) such that

$\overline{F_n}(x) \to F(x)$ for all x (again by Lemma 1 in Yamato, 1973). Using the continuity of F and proceeding as in the proof of the Glivenko–Cantelli theorem (see for instance Billingsley, 1986) it follows that $\|\mathbb{E}F_{nh_n} - F\| \to 0$. So there is $n_0 = n_0(\varepsilon, F)$ such that $\|\mathbb{E}F_{nh_n} - F\| < \varepsilon/2$, for all $n \geq n_0$. This, together with (3) and (5) and the Dvoretzky–Kiefer–Wolfowitz inequality (see Massart, 1990) implies that

$$\mathbb{P}(\|F_{nh_n} - F\| > \varepsilon) \leq \mathbb{P}(\|F_{nh_n} - \mathbb{E}F_{nh_n}\| > \varepsilon/2) \leq \mathbb{P}(\|F_n - F\| > \varepsilon/2) \leq 2\mathrm{e}^{-2n(\varepsilon/2)^2},$$

as desired. □

Proof of Theorem 3. We start by showing that when F is continuous and $h_n \to 0$ we have

$$\sup_{0\leq h\leq h_n} \|\mathbb{E}F_{nh} - F\| \to 0. \tag{8}$$

This can be proved as follows: for a given $\varepsilon > 0$, we choose $M > 0$ large enough that $\int_{\{|z|>M\}} \mathrm{d}L(z) \leq \varepsilon/4$. Besides this, as any continuous distribution function is uniformly continuous and $h_n \to 0$, there is $n_0 \in \mathbb{N}$ such that $\sup_{|x-y|\leq Mh_n} |F(x) - F(y)| \leq \varepsilon/2$ whenever $n \geq n_0$. Then

$$\begin{aligned}
|\mathbb{E}F_{nh}(x) - F(x)| &\leq \int |F(x-hz) - F(x)|\,\mathrm{d}L(z) \\
&= \int_{\{|z|\leq M\}} |F(x-hz) - F(x)|\,\mathrm{d}L(z) + \int_{\{|z|>M\}} |F(x-hz) - F(x)|\,\mathrm{d}L(z) \\
&\leq \sup_{|z|\leq M} |F(x-hz) - F(x)| + 2\int_{\{|z|>M\}} \mathrm{d}L(z).
\end{aligned}$$

Therefore, for $n \geq n_0$,

$$\begin{aligned}
\sup_{0\leq h\leq h_n} \|\mathbb{E}F_{nh} - F\| &\leq \sup_{0\leq h\leq h_n} \sup_{x\in\mathbb{R}} \sup_{|z|\leq M} |F(x-hz) - F(x)| + \varepsilon/2 \\
&\leq \sup_{|x-y|\leq Mh_n} |F(x) - F(y)| + \varepsilon/2 \leq \varepsilon,
\end{aligned}$$

which concludes the proof of (8).

Next we use (8) to prove the statement of the theorem: we know that for a given $\varepsilon > 0$ there is $n_0 = n_0(\varepsilon, F) \in \mathbb{N}$ such that $\sup_{0\leq h\leq h_n} \|\mathbb{E}F_{nh} - F\| < \varepsilon/2$ whenever $n \geq n_0$. Besides this, combining (3) and (5) we have $\|F_{nh} - F\| \leq \|\mathbb{E}F_{nh} - F\| + \|F_n - F\|$ so, for $n \geq n_0$,

$$\mathbb{P}\left(\sup_{0\leq h\leq h_n} \|F_{nh} - F\| > \varepsilon\right) \leq \mathbb{P}\left(\|F_n - F\| > \varepsilon/2\right)$$

and we conclude by using the Dvoretzky–Kiefer–Wolfowitz inequality. □

Proof of Corollary 1. As $\hat{h}_n \to 0$ completely, we may find a sequence of positive numbers $(h_n)_{n\geq 1}$ such that $h_n \to 0$ and $\sum_{n=1}^{\infty} \mathbb{P}(\hat{h}_n > h_n) < \infty$. But for any $\varepsilon > 0$ we have

$$\begin{aligned}
\mathbb{P}(\|F_{n\hat{h}_n} - F\| > \varepsilon) &= \mathbb{P}(\|F_{n\hat{h}_n} - F\| > \varepsilon,\ \hat{h}_n > h_n) + \mathbb{P}(\|F_{n\hat{h}_n} - F\| > \varepsilon,\ \hat{h}_n \leq h_n) \\
&\leq \mathbb{P}(\hat{h}_n > h_n) + \mathbb{P}\left(\sup_{0\leq h\leq h_n} \|F_{nh} - F\| > \varepsilon\right)
\end{aligned}$$

and the right hand side is summable over n due to the assumption on $\hat{h}_n$ and Theorem 3, so $\sum_{n=1}^{\infty} \mathbb{P}(\|F_{n\hat{h}_n} - F\| > \varepsilon) < \infty$, as desired.

The proofs for almost sure convergence and convergence in probability are entirely similar. □

Acknowledgements

This work was partially supported by the grants MTM2009-07302 (J. E. Chacón) and MTM2008-03010 (A. Rodríguez-Casal) from the Spanish Ministry of Science and Innovation. The second author was supported by the IAP research network: grant No. P6/03 from the Belgian government.

References

Billingsley, P., 1986. Probability and Measure, 2nd ed. Wiley, New York.
Devroye, L., 1983. The equivalence of weak, strong and complete convergence in l_1 for kernel density estimates. Ann. Statist. 11, 896–904.
Devroye, L., 1989. The double kernel method in density estimation. Ann. Henri Poincaré 25, 533–580.

Devroye, L., Györfi, L., 1985. Nonparametric Density Estimation: The L_1 View. Wiley, New York.

Devroye, L., Lugosi, G., 2001. Combinatorial Methods in Density Estimation. Springer, New York.

Devroye, L., Penrod, C.S., 1984. The consistency of automatic kernel density estimates. Ann. Statist. 12, 1231–1249.

Dony, J., 2008. Nonparametric regression estimation: an empirical processes approach to uniform in bandwidth consistency of kernel-type estimators and conditional u-statistics. Ph.D. Thesis, Vrije Universiteit, Brussels.

Dony, J., Einmahl, U., Mason, D.M., 2006. Uniform in bandwidth consistency of local polynomial regression function estimators. Austr. J. Stat. 35, 105–120.

Dony, J., Mason, D.M., 2008. Uniform in bandwidth consistency of conditional U-statistics. Bernoulli 14, 1108–1133.

Einmahl, U., Mason, D.M., 2005. Uniform in bandwidth consistency of kernel-type function estimators. Ann. Statist. 33, 1380–1403.

Fernholz, L.T., 1997. Reducing the variance by smoothing. J. Statist. Plann. Inference 57, 29–38.

Giné, E., Mason, D.M., 2008. Uniform in bandwidth estimation of integral functionals of the density function. Scand. J. Statist. 35, 739–761.

Giné, E., Nickl, R., 2009. An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. Probab. Theory Related Fields 143, 569–596.

Marron, J.S., Härdle, W., 1986. Random approximations to some measures of accuracy in nonparametric curve estimation. J. Multivariate Anal. 20, 91–113.

Mason, D.M., Swanepoel, J., 2010. A general result on the uniform in bandwidth consistency of kernel-type function estimator. Test, in press (doi:10.1007/s11749-010-0188-0).

Massart, P., 1990. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. Ann. Probab. 18, 1269–1283.

McDiarmid, C., 1989. On the method of bounded differences. In: Surveys in Combinatorics 1989. In: London Mathematical Society Lecture Notes Series, vol. 141. Cambridge University Press, Cambridge, pp. 148–188.

Nadaraya, E.A., 1964. Some new estimates for distribution functions. Theory Probab. Appl. 15, 497–500.

Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065–1076.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27, 832–837.

Varron, D., 2006. Some uniform in bandwidth functional results for the tail uniform empirical and quantile processes. Ann. I.S.U.P. 50, 83–102.

Yamato, H., 1973. Uniform convergence of an estimator of a distribution function. Bull. Math. Statist. 15, 69–78.

Zieliński, R., 2007. Kernel estimators and the Dvoretzky–Kiefer–Wolfowitz inequality. Appl. Math. 34, 401–404.

Please cite this article in press as: Chacón, J.E., Rodríguez-Casal, A., A note on the universal consistency of the kernel distribution function estimator. Statistics and Probability Letters (2010), doi:10.1016/j.spl.2010.05.007