



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 Estadística 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 27 May 2010

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.

© 2010 Elsevier B.V. All rights reserved.

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 $X_1, \dots, X_n$ 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).

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

$$F_{nh}(x) = \int L\left(\frac{x-y}{h}\right) dF_n(y) = \int F_n(x-hz) dL(z), \quad (1)$$

with the last equality due to the fact that

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

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 $\|F_{nh_n} - F\| \rightarrow 0$ almost surely then $h_n \rightarrow 0$.

(Y2) If F is continuous and $h_n \rightarrow 0$ then $\|F_{nh_n} - F\| \rightarrow 0$ almost surely.

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

From (Y1) and (Y2) it follows that almost sure uniform consistency of F_{nh_n} is equivalent to $h_n \rightarrow 0$ 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 $\|F_{nh_n} - F\| \rightarrow 0$ 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

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

Consequently, by the Borel–Cantelli lemma we obtain that $\mathbb{E}\|F_{nh_n} - F\| \rightarrow 0$ is equivalent to $\|F_{nh_n} - F\| \rightarrow 0$ 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.

Download English Version:

<https://daneshyari.com/en/article/1152950>

Download Persian Version:

<https://daneshyari.com/article/1152950>

[Daneshyari.com](https://daneshyari.com)