Uniform in bandwidth consistency of kernel regression estimators at a fixed point

Julia Dony

Department of Mathematics,
Free University of Brussels (VUB)
Pleinlaan 2
B–1050 Brussels, Belgium.
(e-mail: jdony@vub.ac.be)

Abstract. We start by considering a kernel estimator $\varphi_{n,h}(t)$ for the regression function $m_\varphi(t) := \mathbb{E}[\varphi(Y)|X = t]$, where $t \in \mathbb{R}^d$ is fixed and $\varphi : \mathbb{R} \to \mathbb{R}$ is a measurable function with finite second moment. If $h \equiv h_n$ is a deterministic sequence satisfying $h_n \to 0$ and $nh_n^d / \log\log n \to \infty$, as well as a condition depending on the moment of φ, it is well–known that $\varphi_{n,h_n}(t)$ estimates consistently $m_\varphi(t) f_X(t)$, where f_X is the density function of X. As an extension, we present a result in which additional assumptions are imposed to make $\varphi_{n,h}(t)$ a strongly consistent estimator, uniformly for a certain range of bandwidths h. As an application, we consider random variables $Y_1, \ldots, Y_n$ for which the common distribution function F has regularly varying upper tails of exponent $-1/\tau < 0$, and study the asymptotic behavior of a kernel–based version of the Hill estimator for τ. The results that we obtain for the process $\varphi_{n,h}(t)$ are used to establish the weak consistency of this estimator, uniformly for a certain range of bandwidths tending to zero at particular rates.

Keywords. Hill estimator, extreme values, nonparametric regression, kernel estimators, consistency, empirical processes, uniform in bandwidth.

1 Introduction and statement of the main results

Let $(X_i, Y_i), i \geq 1$ be i.i.d. random vectors in $\mathbb{R}^d \times \mathbb{R}$ and let $\mathcal{F}$ be a class of measurable functions $\varphi : \mathbb{R} \to \mathbb{R}$ for which $\mathbb{E}\varphi^2(Y) < \infty$. Then for $t \in \mathbb{R}^d$ fixed, we denote the regression function by $m_\varphi(t) := \mathbb{E}[\varphi(Y)|X = t]$. For a uniformly bounded kernel function K with support contained in $[-1/2, 1/2]^d$, and bandwidth $0 < h < 1$, define the estimator

$$\varphi_{n,h}(t) := \frac{1}{nh^d} \sum_{i=1}^{n} \varphi(Y_i) K\Big(\frac{t - X_i}{h}\Big). \tag{1}$$

Likewise, let $f_{n,h}(t)$ be the kernel density estimator of the marginal density $f_X(t)$ of X with bandwidth h, which corresponds to the choice $\varphi(y) \equiv 1$ in the above formula. If h_n is a deterministic sequence of positive numbers going to zero and such that $nh_n^d / \log\log n \to \infty$, as well as a condition depending on the moment of φ, it is well–known that under regularity conditions

$\varphi_{n,h_n}(t)/f_{n,h_n}(t)$ is a (strongly) consistent estimator for $m_\varphi(t)$. See for instance, Einmahl and Mason (1998). Moreover, it was shown in Einmahl and Mason (2005) that under some additional assumptions, the consistency of $\varphi_{n,h}(t)$ is preserved uniformly in $a_n \leq h \leq b_n$ for appropriate positive sequences a_n and b_n converging to zero and uniformly for $t \in I$ where $I \subset \mathbb{R}^d$ is compact.

The uniformity in h makes it possible to choose the bandwidth h_n depending on the data and/or location. If the function class $\mathcal{F}$ is uniformly bounded, they show that one can choose h uniformly from an interval of the form $[a_n, b_n]$, where $a_n = c_n \log n/n$ with $c_n \to \infty$ and $b_n \to 0$. Assuming in the unbounded case that the envelope function F of $\mathcal{F}$ satisfies the condition,

$$\text{(F.}p\text{)} \qquad \mu_p := \sup_{x \in J} \mathbb{E}\left[F^p(Y)|X = x\right] < \infty, \quad \text{for some } p > 2,$$

where $J = I^\epsilon$ for some $\epsilon > 0$, these authors have shown that the above result remains valid in this more general case if one chooses $a_n \geq c(\log n/n)^\gamma$, where $\gamma = 1 - 2/p$. In addition, in both cases, the corresponding convergence rates of $\varphi_{n,h}(t)$ to $m_\varphi(t) f_X(t)$ have been obtained as well.

One of the main purposes of this note is to describe uniform in bandwidth consistency results for $\varphi_{n,h}(t)$ at fixed points $t \in \mathbb{R}^d$, i.e. pointwise and not uniformly over compact subsets. This will allow us to achieve the unifomity in h on larger intervals than in the previous "compact" case. In Section 2 we shall apply our "pointwise" uniform in bandwidth consistency results to establish the uniform in bandwidth consistency of a class of kernel tail index estimators.

Towards establishing these consistency results, we impose some additional conditions. In particular, we consider classes $\mathcal{F}$ such that

(F.i) $\mathcal{F}$ is a pointwise measurable class,
(F.ii) $\mathcal{F}$ has a measurable envelope function $F(y) \geq \sup_{\varphi \in \mathcal{F}} |\varphi(y)|, y \in \mathbb{R}$,
(F.iii) $\mathcal{F}$ is a VC class of functions,

and we let $a_n, n \geq 1$ be a sequence of non–random numbers satisfying

(H.i) $a_n \searrow 0$,
(H.ii) $a_n^d \log\log n \searrow$ and $na_n^d/\log\log n \nearrow$,

where "$\nearrow, \searrow$" denote non–decreasing and non–increasing respectively. For convenience, we recall the assumptions on the kernel function $K : \mathbb{R}^d \to \mathbb{R}$.

(K.i) $\sup_{x \in \mathbb{R}^d} |K(x)| < \infty$ and $\int K(x)dx = 1$,
(K.ii) K has support contained in $[-1/2, 1/2]^d$.

Our first uniform in h consistency result for $\varphi_{n,h}(t)$ with $t \in \mathbb{R}^d$ fixed, holds when the moment generating function of the envelope function is finite.

Clearly, this is more general than having a bounded function class (see Einmahl and Mason (1997)), and this extension seems to be new also for fixed bandwidth sequences.

Theorem 1. *Suppose that the envelope function of $\mathcal{F}$ has a finite moment generating function in a neighborhood of 0. Then if f_X is bounded on a neighborhood of t, and $b_0 < 1$ is a positive constant, it follows from the above mentioned assumptions on $\mathcal{F}$ and K that*

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_0} \sup_{\varphi\in\mathcal{F}} \frac{\sqrt{nh^d}|\varphi_{n,h}(t) - \mathbb{E}\varphi_{n,h}(t)|}{\sqrt{\log\log n}} < \infty, \quad a.s.,$$

for all non–increasing sequences $a_n, n \ge 1$ that go to zero and such that $na_n^d \ge c \log\log n$, for some $c > 0$.

It is well known that the assumption "$na_n^d \ge c\log\log n$, for some $c > 0$" is optimal in the bounded case, and this shows that there is no difference in terms of range of bandwidths between the bounded case and the case where the moment generating function of F is finite.

Next, we consider the case where one has a moment–type condition on the envelope function like (F.p).

Theorem 2. *Suppose that f_X is bounded on a neighborhood of t, and let $b_0 < 1$ be a positive constant. Then it follows from the above assumptions on $\mathcal{F}$ and K that if (F.p) is satisfied with J being a neighborhood of t,*

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_0} \sup_{\varphi\in\mathcal{F}} \frac{\sqrt{nh^d}|\varphi_{n,h}(t) - \mathbb{E}\varphi_{n,h}(t)|}{\sqrt{\log\log n}} < \infty, \quad a.s.,$$

where a_n satisfies the assumptions stated in (H), and is such that

$$\text{(H.p)} \qquad na_n^d \ge (\log n)^{2/(p-2)}.$$

Our uniform in bandwidth results are important in establishing the consistency of kernel regression estimators using a data dependent bandwidth sequence $\hat{h}_n := H_n(X_1, \ldots, X_n), n \ge 1$. Namely, to show that almost surely or in probability,

$$\sup_{\varphi\in\mathcal{F}} |m_{n,\hat{h}_n,\varphi}(t) - m_\varphi(t)| \longrightarrow 0,$$

where $m_{n,\hat{h}_n,\varphi}(t)$ is a Nadaraya–Watson estimator for $m_\varphi(t)$. This application is investigated in detail in Dony (2008).

The range of $a_n, n \ge 1$ provided by (H.p) for fixed t is less restrictive than the one obtained by Einmahl and Mason (2005) in the uniform in t case, namely $a_n^d = O((\log n/n)^{1-2/p})$. Moreover, any a_n as in (H.p) always fulfills (H.ii). For a related result in case of a fixed bandwidth sequence h_n, refer to Theorem 4 of Einmahl and Mason (1998).

Detailed proofs of these results are provided in Dony (2008). It turns out that $\varphi_{n,h}(t)$ has an empirical process representation, where the index class is a class of functions that depends upon the sample size $n \geq 1$. This permits us to use in our proofs some moment and exponential deviation inequalities for empirical processes. To achieve the consistency uniformly in bandwidth, a blocking argument is applied and the interval $[a_n, b_0]$ is split into several smaller intervals.

In the next section, we illustrate how these results have proved to be particularly useful to establish the uniform in bandwidth consistency of a class of kernel estimators for the tail index of a Pareto distribution.

2 Application to kernel estimators of the tail index

Let $Y_1, \ldots, Y_n$ be independent real–valued random variables with mean one and denote their common distribution function by F. Then set $t_{i,n} := i/(n+1)$ and define the following fixed design process :

$$\beta_{n,h} := \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{t_{i,n}}{h}\right), \quad 0 < h < 1.$$

Under some particular assumptions on the kernel function K, the asymptotic behavior of $\sup_{a_n \leq h \leq b_0} |\beta_{n,h} - \mathbb{E}\beta_{n,h}|$ can be inferred from the asymptotic behavior of the corresponding random design process $\varphi_{n,h}(0)$, defined in (1). The following theorem is a consequence of a result proved in Dony (2008).

Theorem 3. *Let $Y_1, \ldots, Y_n$ be independent bounded random variables with mean one, and suppose that K is a Hölder–continuous kernel function with exponent $0 < \alpha \leq 1$. Then it holds that $\sup_{a_n \leq h \leq b_0} |\beta_{n,h} - \mathbb{E}\beta_{n,h}| = o_{\mathbb{P}}(1)$, provided α and $a_n \to 0$ are such that $n^{\alpha/2} a_n^{\alpha+1} \to \infty$.*

This result remains valid under a wide variety of assumptions on the variables Y_i as well. For instance, the weak consistency of $\beta_{n,h}$ remains valid when, instead of being bounded, the variables are unbounded, but have a finite p–the moment, $p > 2$. In this case, the result is a consequence of Theorem 2, and the sequence a_n has to be chosen appropriately, taking into account that the choice in (H.p) does not satisfy the assumption in Theorem 3.

We shall next describe an application of our results to extreme value statistics. Assume that F has regularly varying upper tails of exponent $-1/\tau < 0$, i.e. there exists a $\tau > 0$ such that,

$$(F.\tau) \qquad \lim_{y \to \infty} \frac{1 - F(\lambda y)}{1 - F(y)} = \lambda^{-1/\tau}, \quad \lambda > 0.$$

Such a distribution function is also said to be of Pareto–type with parameter $1/\tau$. We call $1/\tau$ the *tail index*. A well–known estimator for τ is the so–called Hill estimator, introduced by Hill (1975), and defined as

$$\hat{a}_{n,k} := k^{-1} \sum_{i=1}^{k} \log Y_{n-i+1:n} - \log Y_{n-k:n},$$

where $Y_{i:n}, 1 \leq i \leq n$ denote the order statistics of $Y_1, \ldots, Y_n$. The consistency of $\hat{a}_{n,k}$ as an estimator of τ was established by Mason (1982), who provided necessary and sufficient conditions for $\hat{a}_{n,k} \to_{\mathbb{P}} \tau$ for all sequences $k = k_n \to \infty$ and $k_n/n \to 0$. Somewhat later, Csörgő et al. (1985) proposed the following kernel–based estimator for τ,

$$\hat{\tau}_{n,h} := \frac{\sum_{j=1}^{n} \frac{j}{nh} K\left(\frac{j}{nh}\right) \{\log Y_{n-j+1:n} - \log Y_{n-j:n}\}}{\frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{j}{nh}\right)} =: \frac{\phi_{n,h}}{\kappa_{n,h}}, \tag{2}$$

where K is a kernel function and $0 < h < 1$ is a bandwidth. We are interested in the asymptotic behavior of $\hat{\tau}_{n,h}$, uniformly in h. To begin, note that $(F.\tau)$ implies that $Y_i \overset{d}{=} (1-U_i)^{-\tau} L((1-U_i)^{-1})$, where $U_1, \ldots, U_n$ are independent uniformly distributed variables over $]0,1[$, and L is a slowly varying function at infinity. By Karamata's representation, it holds that

$$\begin{aligned} &\log Y_{n-j+1:n} - \log Y_{n-j:n} \\ &\qquad \overset{d}{=} \quad -\tau\{\log(1 - U_{n-j+1:n}) - \log(1 - U_{n-j:n})\} \\ &\qquad\qquad + \quad \int_{(1-U_{n-j:n})^{-1}}^{(1-U_{n-j+1:n})^{-1}} \frac{b(u)}{u} du \\ &\qquad\qquad + \quad \log \frac{c((1 - U_{n-j+1:n})^{-1})}{c((1 - U_{n-j:n})^{-1})}, \end{aligned}$$

where $c(y) \to c_0$ and $b(y) \to 0$ as $y \to \infty$. Consequently, $\phi_{n,h}$ can be decomposed into three processes as follows :

$$\phi_{n,h} \overset{d}{=} \sum_{j=1}^{n} \frac{j}{nh} K\left(\frac{j}{nh}\right) \{\tau A_{n,j}^{(1)} + A_{n,j}^{(2)} + A_{n,j}^{(3)}\} =: \tau\phi_{n,h}^{(1)} + \phi_{n,h}^{(2)} + \phi_{n,h}^{(3)}.$$

In order to derive the consistency results for $\phi_{n,h}$ and $\hat{\tau}_{n,h}$, the following additional assumptions on the representation of $L(y)$ and on the kernel K need to be made :

(L.i) $c(y)$ is constant,
(K.ii)$'$ K has support contained in $[0, M]$ for some $M < \infty$,
(K.iii) K is non–increasing and non–negative on its positivity set,
(K.iv) K is a Hölder–continuous kernel with exponent $0 < \alpha \leq 1$.

Theorem 4. *For any right–continuous kernel function K satisfying (K.i)–(K.iii), and any non–increasing sequences $a_n \leq b_n, n \geq 1$ such that $a_n \to 0$, $b_n \to 0$ and $na_n \to \infty$, it holds that*

(i) $\sup_{a_n \leq h \leq b_n} |\kappa_{n,h} - 1| \to 0,$

(ii) $\sup_{a_n \leq h \leq b_n} |\phi^{(3)}_{n,h}| = o_{\mathbb{P}}(1).$

If moreover (K.iv) holds with $0 < \alpha \leq 1$ such that $n^{\alpha/2} a_n^{\alpha+1} \to \infty$, we have also that

(iii) $\sup_{a_n \leq h \leq b_n} |\phi^{(1)}_{n,h} - 1| \to 0$ *in probability,*

(iv) $\sup_{a_n \leq h \leq b_n} |\phi^{(2)}_{n,h}| = o_{\mathbb{P}}(1).$

As a direct consequence, we obtain readily that $\sup_{a_n \leq h \leq b_n} |\hat{\tau}_{n,h} - \tau| = o_{\mathbb{P}}(1)$, establishing the (weak) uniform in bandwidth consistency of the kernel–based Hill estimator $\hat{\tau}_{n,h}$ to τ. Our previous discussion of the application of our results to data dependent bandwidth sequences carries directly over to kernel estimators of the tail index. Refer to Section 4.7 of Beirlant et al. (2004) for a description of relevant adaptive selectors of the tail fraction.

Acknowledgments. The author thanks Sándor Csörgő for suggesting the problem that led to this paper.

Julia Dony's research is financed by a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).

References

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004) : Statistics of extremes. Theory and Applications. *Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester.*

Csörgő, S., Deheuvels, P. and Mason, D.M. (1985) : Kernel estimates of the tail index of a distribution, *Ann. Stat. 13,* 1050–1077.

Deheuvels, P. and Mason, D. M. (2004) : General asymptotic confidence bands based on kernel–type functional estimators. *Stat. Infer. Stoch. Proc. 7,* 225–277.

Dony, J. (2008) : *Nonparametric regression estimation : An empirical process approach to the uniform in bandwidth consistency of kernel estimators and conditional U–statistics.* Ph.D. Thesis, Vrije Universiteit Brussel, Belgium.

Einmahl, U. and Mason, D. M. (1997) : Gaussian approximation of local empirical processes indexed by functions. *Prob. Th. Rel. Fields 107,* 283–311.

Einmahl, U. and Mason, D. M. (1998) : Strong approximations to the local empirical process. *Progr. Probab. 43,* 75–92.

Einmahl, U. and Mason, D. M. (2005) : Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat. 33,* 1380–1403.

Hill, B.M. (1975) : A simple general approach to inference about the tail of a distribution. *Ann. Stat. 3,* 1163–1174.

Mason, D.M. (1982) : Laws of large numbers for sums of extreme values. *Ann. Probab. 3,* 754–764.