

A note on the estimation of the Hölder constant

Guillermo Henry, Daniela Rodriguez and Mariela Sued
FCEyN, Universidad de Buenos Aires and CONICET, Argentina

Abstract

In this paper we develop a nonparametric estimator for the Hölder constant of a density function. We consider a simulation study to evaluate the performance of the proposal and construct smooth bootstrap confidence intervals. Also, we give a brief review over the impossibility to decide whether a density function is Hölder.

Key words and phrases: Discernibility, Hölder density, Nonparametric estimation.

1 Introduction

Modern statistics assumes many hypothesis concerning the underlying distribution of the data in order to obtain good estimators. It is often assumed that the underlying distribution has a density f that belongs to certain class \mathcal{F} . To decide whether f belongs to \mathcal{F} based on the data is not always possible. When it is, we say that the class \mathcal{F} is discernible. Devroy and Lugossi (2002) discuss this problem in detail. Between many examples considered in their work, they show that there is no rule able to decide whether a density is Lipschitz or not. We generalize this fact proving that the class of γ -Hölder densities is not discernable. Given these circumstances, we limit ourselves to estimate the Hölder constant of a density, including the Lipschitz case, whenever it is finite. In other words, assuming that the density is γ -Hölder, we will make inference on its constant. Carando, *et al.* (2009) studied a maximum likelihood non parametric density estimator under the assumption that the density is Lipschitz with a known constant: our proposal can be plugged in their method.

This note is organized as follows. In Section 2, we discuss the discernibility of the class of Hölder functions. The proposed estimator of the Hölder constant and a simulation study to evaluate its performance are included in Section 2 and 3, respectively. Finally, smoothed bootstrap confidence intervals are defined in Section 4.

2 Hölder densities are not discernible

According to Devroy and Lugossi (2002), a class of densities \mathcal{F} is said to be discernible if for any density f we can decide whether it belongs to \mathcal{F} or not, based on sample X_1, X_2, \dots of i.i.d. random

variables with common density f . The decision is made based on a classification rule, which is a sequence $\{T_n\}$ of functions $T_n : \mathbb{R}^n \rightarrow \{0, 1\}$ that satisfies

$$\lim_{n \rightarrow \infty} T_n(X_1, \dots, X_n) = \begin{cases} 1 \text{ a.s.} & \text{if } X_i \text{ has density } f \in \mathcal{F}, \\ 0 \text{ a.s.} & \text{if } X_i \text{ has density } f \notin \mathcal{F}. \end{cases} \quad (1)$$

The authors also show that the class \mathcal{F} of densities supported in $[0, 1]$, such that each $f \in \mathcal{F}$ has a Lipschitz version, is not discernible. We will prove that γ -Hölder densities are not discernible, including the Lipschitz case with $\gamma = 1$. Recall that, for $\gamma \in (0, 1]$, f is γ -Hölder if

$$|f(x) - f(y)| \leq w_\gamma(f) |x - y|^\gamma \quad (2)$$

with finite $w_\gamma(f)$, where

$$w_\gamma(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\gamma}. \quad (3)$$

We follow the approach by Fraiman and Meloche (1999) used to prove that the class of densities with finitely many modes is not discernible. Their presentation is based on the results obtained by Le Cam and Scharz (1960), where necessary and sufficient conditions for the existence of consistent estimates are discussed.

We denote by $L^1[0, 1]$ the set of integrable Lebesgue functions on $[0, 1]$, and denote the L^1 norm with $\|f\|_1 = \int_0^1 f(x) dx$. Let $\mathcal{B} \subset L^1[0, 1]$ be the set of densities in the interval $[0, 1]$. Note that $(\mathcal{B}, \|\cdot\|_1)$ is a closed subset of $L^1[0, 1]$. Let \mathcal{G} be a subset of \mathcal{B} such that both \mathcal{G} and \mathcal{G}^c are dense in \mathcal{B} . By Lemma 1.1 in Fraiman and Meloche (1999) we can conclude that such \mathcal{G} is not discernible. As we mentioned, we will take advantage of this fact to prove that \mathcal{G}_γ , the class of densities supported in $[0, 1]$ such that each $f \in \mathcal{G}_\gamma$ has a γ -Hölder version, is not discernible. We will see that both \mathcal{G}_γ and \mathcal{G}_γ^c are dense in \mathcal{B} , as follows:

1. To prove that \mathcal{G}_γ is dense in \mathcal{B} , we use the fact that the convolution preserves regularity in order to approximate any density f by a sequence of densities in \mathcal{G}_γ . To be more precise, let K be a kernel in \mathcal{G}_γ and consider $K_h(y) = h^{-1}K(y/h)$. Given $f \in \mathcal{B}$, consider

$$f_h(x) = \frac{f * K_h(x) I_{[0,1]}(x)}{C_h(f)}, \quad \text{where } C_h(f) = \int_0^1 f * K_h(x) dx.$$

Note that $f * K_h$ converges to f in $L^1[R]$ and so, since f is supported in $[0, 1]$, we get that $C_h \rightarrow 1$. This guarantees that f_h converges to f in $L^1[0, 1]$. On the other hand, $f * K_h$ is as regular as K , meaning that if $K \in \mathcal{G}_\gamma$, $f * K_h$ is γ -Hölder and by consequence $f_h \in \mathcal{G}_\gamma$.

2. In order to prove that \mathcal{G}_γ^c is dense in $L^1[0, 1]$, we need to prove that any element $f \in \mathcal{G}_\gamma$ can be approximated by a sequence in \mathcal{G}_γ^c . Take $g \in \mathcal{G}_\gamma^c$ and define

$$f_h = (1 - h)f + hg.$$

Then, $f_h \in \mathcal{G}_\gamma^c$ and converges to f in $L^1[0, 1]$, when $h \downarrow 0$.

3 Estimation of the Hölder constant.

Let X be a random variable with density f γ -Hölder; i.e. f satisfies condition (2). We want to remark that this condition has a nice geometric interpretation. In the Lipschitz case, i.e. *Hölder function with $\gamma = 1$* , there is a double cone whose vertex can be translated along the graph in such a way that the graph of the density remains always entirely inside the cone. When the density is γ -Hölder, the situation is analogous, but considering a double parabolic cone instead of linear cone, as we can see in Figure 1. Note that given a double cone with slope m , any cone with slope $m' \geq m$ satisfies also this property. The Hölder constant agree with the smallest of these slopes.

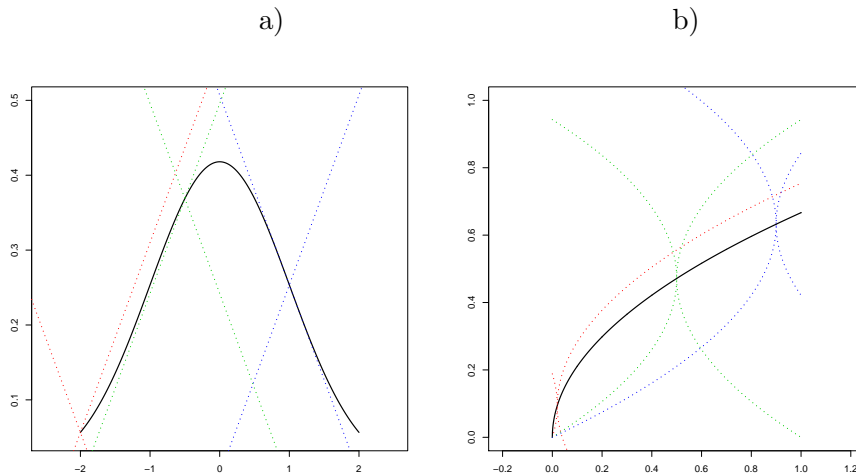


Figure 1: a) Lipschitz's cone. b) Hölder's cone.

Let X_1, \dots, X_n be independent observations of a random variable X with unknown probability density function f γ -Hölder, with known γ . In order to define a nonparametric estimator of the Hölder constant $w_\gamma(f)$, we consider the following result whose proof can be obtained straightforwardly by Borel–Cantelli's Lemma.

Lemma: Let $G \subset \mathbb{R}^k$ be an open set and $h : G \rightarrow \mathbb{R}$ a bounded and continuous function. Let $(Y_i)_{i \geq 1}$ be independent and identically distributed random variables taking values on \mathbb{R}^k , with density function f , such that $f(y) > 0$ for all $y \in G$ and $P(Y_1 \in G) = 1$. Then

$$M_n = \max_{1 \leq i \leq n} h(Y_i) \rightarrow \sup_{y \in G} h(y) = M \text{ a.s.} \quad (4)$$

Note that if we apply the Lemma to the function $h(x, y) = \frac{|f(x) - f(y)|}{|x - y|}$ and the variables $W_j = (X_{2j-1}, X_{2j})$, for $1 \leq j$, in the open set that excludes the diagonal, then a natural approach is to consider

$$\hat{w}_\gamma(f) = \max_{1 \leq i, j \leq n; i \neq j} \frac{|f(X_i) - f(X_j)|}{|X_i - X_j|^\gamma}.$$

But, since f is unknown, we will estimate f through a nonparametric estimator f_n . Therefore the nonparametric estimator of the Hölder constant $w_\gamma(f)$ is defined as $\hat{w}_\gamma(f_n)$.

In this work, we consider the class of kernel type estimators of f which has been widely studied since the work of Rosenblatt (1956) and Parzen (1962). These estimators are defined as

$$f_n(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right), \quad (5)$$

where $K(u)$ is a bounded density on \mathbb{R} and h is a sequence of positive number such that $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$.

We note that the estimator $\hat{w}_\gamma(f_n)$ has also a geometric interpretation. If we consider the Lipschitz case, the estimator can be defined as follow: for each point of the random sample of size n , for example X_1 , we consider the $n-1$ segments that join the points $(X_1, f_n(X_1))$ and $(X_i, f_n(X_i))$ for $i = 2, \dots, n$, and denote by $m_{X_1}(X_i)$ the slopes of these segments. Figure 2 a) shows some slopes for two points on the random sample. Therefore for each point of the sample, we can estimate a local double cone as the cone given by the maximum of these slopes. Finally, we consider a estimator of the Lipschitz constant as the maximum slope of the local cones given the sample.

In the Hölder case, the procedure is the same but instead of consider straights we have to consider curves defined by the graph of x^γ . In Figure 2 b), we show this situation.

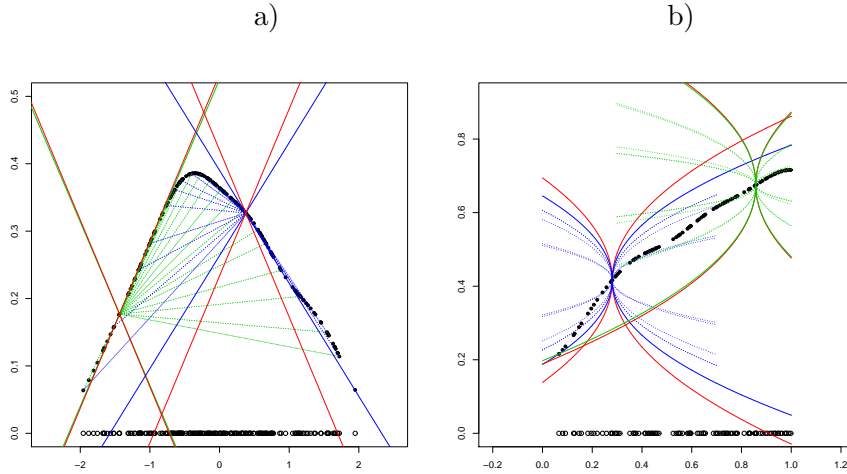


Figure 2: a) Lipschitz's cone. The lines in red correspond to the cone with the estimated constant. The solid lines in blue correspond to the local cone estimator for a particular sample point and the dashed lines correspond to the straight lines for this point with some of the others points in the sample. In green the same kind of plots are shown, constructed at a different point of the sample. b) Hölder's cone. The figure is analogous to a) but with an appropriate curve.

4 MonteCarlo study

This section contains the results of a simulation study designed to evaluate the performance of the estimator defined in Section 3. To perform this study, we considered four models. For each model, we generated X_1, \dots, X_n independent samples of size $n = 100, 500, 1000$ according with a density

f . In each model, we performed $N = 1000$ replications. The different models were generated from the following density functions.

Model 1 : $f_X(x) = \frac{1}{\sqrt{2\pi(2\phi(2)-1)}} \exp\{-\frac{1}{2}x^2\}I_{(-2,2)}(x)$ in this case the Lipschitz constant $w_1(f) = e^{-1/2}(\sqrt{2\pi(2\phi(2)-1)})^{-1} = 0.2535053$ where ϕ denotes the standar normal distribution function.

Model 2 : $f_X(x) = (x+1)I_{(-1,0)}(x) + (1-x)I_{(0,1)}$ in this case the Lipschitz constant $w_1(f) = 1$.

Model 3 : $f_X(x) = \frac{1}{2}((x+1)I_{(-1,0)}(x) + (1-x)I_{(0,1)}(x) + (x-1)I_{(1,2)}(x) + (3-x)I_{(2,3)}(x))$ in this case the Lipschitz constant $w_1(f) = \frac{1}{2}$.

Model 4 : $f_X(x) = \frac{2}{3}\sqrt{x}I_{(0,1)}(x)$ in this case the Hölder constant $w_{\frac{1}{2}}(f) = \frac{2}{3}$.

Figure 1 shows the densities considered in models 1 and 4 while Figure 3 shows the density functions in the models 2 and 3.

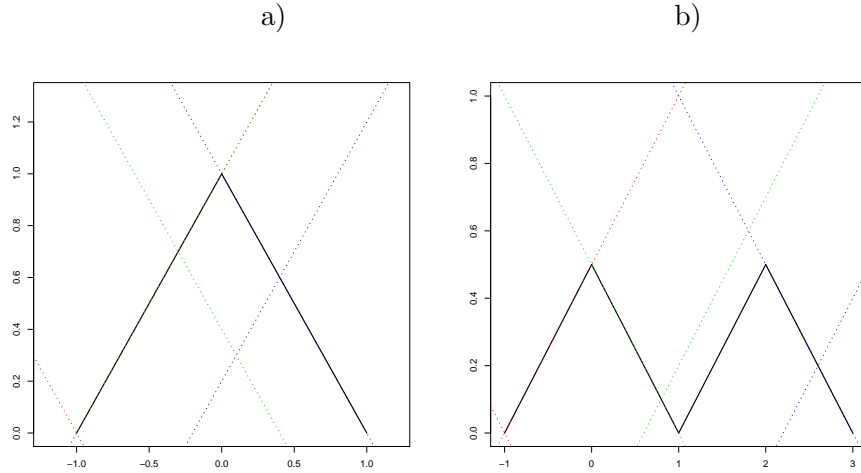


Figure 3: a) Density function of Model 2. b) Density function of Model 3.

In all cases, for smoothing procedure, we take the gaussian kernel and we consider an automatic bandwidth selector given by Scott (1992). To study the performance of the estimators of $w_\gamma(f)$, denoted by $\hat{w}_\gamma(f_n)$, we have considered the mean square error (MSE) calculated over the N replications, *i.e.*

$$\text{MSE}(\hat{w}_\gamma(f_n)) = \frac{1}{N} \sum_{i=1}^N [\hat{w}_\gamma(f_n)_i - w_\gamma(f)]^2$$

where $\hat{w}_\gamma(f_n)_i$ denotes the estimator of the constant given by the replication i .

Table 1 gives the values of the MSE of $\hat{w}_\gamma(f_n)$ for the four considered models. The simulation study confirms the good behavior of constant estimators, under the different models. Model 1 corresponds to a infinitely differential function and as we expected the estimator of the constant

has a good performance. The densities considered in Models 2 and 3 correspond to non differentiable functions however the behavior is comparable. Finally, Model 4 is an example of the Hölder case where we also obtained a good performance.

	Model 1	Model 2	Model 3	Model 4
$n = 100$	0.0101	0.0113	0.0281	0.0174
$n = 500$	0.0050	0.0084	0.0051	0.0062
$n = 1000$	0.0037	0.0096	0.0016	0.0047

Table 1: MSE for the different models.

5 Smoothed bootstrap confidence intervals

In this section, we develop a bootstrap procedure to implement confidence intervals for the Hölder constant. One of the main motivations to construct confidence intervals is to be able to answer questions about the true value of the Holder constant. For example, if a $100(1 - \alpha)\%$ confidence intervals for $w_\gamma(f)$ does not contain the value c_0 , this is evidence against the null hypothesis that $w_\gamma(f)$ is c_0 . In this way, as usual, the confidence intervals can provide a test for such null hypothesis. Moreover, confidence intervals give an idea about the global variability of the estimator.

The bootstrap method introduced by Efron (1979) is a very useful tool used primarily to estimate the sampling distribution of some statistic of interest. Given a set of independent and identically distributed observations X_1, \dots, X_n , the estimator of the Hölder constant $\widehat{w}_\gamma(f_n)$ can be defined as some function $T(X_1, \dots, X_n)$. The smoothed bootstrap technic (Efron (1979, 1982); Silverman and Young, (1987)) estimates the sampling distribution of interest by that of $T(X_1^*, \dots, X_n^*)$, for bootstrap samples X_1^*, \dots, X_n^* of size n obtained by independent sampling from the distribution F_n with density f_n defined in (5).

In order to construct bootstrap confidence intervals, we consider the smoothed bootstrap method. Thus, the bootstrap confidence intervals can be described as follows.

Step 1 Based on X_1, \dots, X_n , we estimate f through $f_n(x) = (nh)^{-1} \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right)$ and compute $\widehat{w}_\gamma(f_n)$.

Step 2 A bootstrap sample X_1^*, \dots, X_n^* is generated as $X_j^* = X_{i_j} + \varepsilon_j$, with i_1, \dots, i_n independent and uniformly distributed integers from $\{1, \dots, n\}$ and $\varepsilon_1, \dots, \varepsilon_n$ independent and identically distributed with density $K_h(t) = h^{-1}K(t/h)$. The bootstrap estimate $\widehat{w}_\gamma^*(f_n^*)$ of $w_\gamma(f)$ is calculated based on X_1^*, \dots, X_n^* .

Step 3 Repeat Step 2, B times, to get B values of $\widehat{w}_\gamma^*(f_n^*)_r$ for $1 \leq r \leq B$.

Step 4 A bootstrap estimator of the variance of $\widehat{w}_\gamma(f_n)$ is computed as

$$\text{var}_B(\widehat{w}_\gamma(f_n)) = \frac{1}{B-1} \sum_{r=1}^B (\widehat{w}_\gamma^*(f_n^*)_r - \overline{\widehat{w}_\gamma^*(f_n^*)})^2$$

where $\overline{\widehat{w}_\gamma^*(f_n^*)} = \frac{1}{B} \sum_{r=1}^B \widehat{w}_\gamma^*(f_n^*)_r$.

Step 5 The $1 - \alpha$ confidence interval of $w_\gamma(f)$ is defined as

$$IC(w_\gamma(f)) = \left(\widehat{w}_\gamma(f_n) - q^* \left(1 - \frac{\alpha}{2} \right) \sqrt{\text{var}_B(\widehat{w}_\gamma(f_n))}, \widehat{w}_\gamma(f_n) - q^* \left(\frac{\alpha}{2} \right) \sqrt{\text{var}_B(\widehat{w}_\gamma(f_n))} \right)$$

where $q^*(\alpha)$ is the α -quantile of the empirical bootstrap distribution of $T_r^* = \frac{\widehat{w}_\gamma^*(f_n^*)_r - \widehat{w}_\gamma(f_n)}{\sqrt{\text{var}_B(\widehat{w}_\gamma(f_n))}}$ for $1 \leq r \leq B$.

To evaluate the performance of the bootstrap intervals described above, we implement this procedure to the models considered in the previous Section. The following Table summarizes the results of the empirical coverage, defined as the proportion of the bootstrap intervals containing the true parameter over the $N = 1000$ replications. For each model, we consider $B = 1000$ bootstrap replications and in all cases, we compute confidence intervals with level 0.95. Note that, as expected, the bootstrap intervals have a good performance since the empirical coverages are near to the nominal level 0.95.

	Model 1	Model 2	Model 3	Model 4
$n = 100$	94.1	95	97	96.5

Table 2: Empirical coverage of nominal 95% confidence intervals. The empirical coverages are in percent.

Acknowledgments

We wish to thank the anonymous referee for his/her valuable comments which led to an improved version of the original article. We would like to thank to Lucio Pantazis for his careful reading of the manuscript and for useful suggestions. This research was partially supported by Grants PIP 112-200801-00216 and 0592 from CONICET and UBACYT 20020100300057 from the Universidad de Buenos Aires.

References

- [1] Carando, D.; Fraiman, R. and Groisman, P. (2009). Nonparametric likelihood based estimation for a multivariate Lipschitz density. *Journal of Multivariate Analysis*. **100**. 981-992.
- [2] Devroye, L and Lugosi, G. (2002). Almost sure classification of densities. *Journal of Nonparametric Statistics*. **14**, 675-698.
- [3] Efron, B. (1979) Bootstrap methods-another look at the jack-knife. *Ann. Statist.*, **7**, 1-26.

- [4] Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- [5] Fraiman, R. and Meloche, J. (1999). Counting Bumps. *Annals of the Institute of Statistical Mathematics*, **51**, 1–29.
- [6] Le Cam, L. and Schwartz, L. (1960) A necessary and sufficient condition for the existence of consistent estimates. *Annals of Statistics*. **31**, 140–150.
- [7] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.
- [8] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837.
- [9] Scott, D. W. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley.
- [10] Silverman, B. W. and Young, G. A. (1987) The bootstrap: to smooth or not to smooth? *Biometrika*, 74, 469–479.