ELSEVIER

Statistics & Probability Letters 44 (1999) 299–308

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

On the Hilbert kernel density estimate

Luc Devroye[a,*,1], Adam Krzyżak[b,2]

[a] School of Computer Science, McGill University, Montreal, Quebec, Canada H3A 2K6
[b] Department of Computer Science, Concordia University, Montreal, Canada H3G 1M8

Received January 1998; received in revised form November 1998

Abstract

Let X be an $\mathbb{R}^d$-valued random variable with unknown density f. Let $X_1,\dots,X_n$ be i.i.d. random variables drawn from f. We study the pointwise convergence of a new class of density estimates, of which the most striking member is the Hilbert kernel estimate

$$\frac{1}{V_d n \log n}\sum_{i=1}^{n}\frac{1}{\|x-X_i\|^d},$$

where V_d is the volume of the unit ball in $\mathbb{R}^d$. This is particularly interesting as this density estimate is basically of the format of the kernel estimate (except for the $\log n$ factor in front) and the kernel estimate does not have a smoothing parameter.

MSC: Primary 62G05

Keywords: Density estimation; Kernel estimate; Convergence; Bandwidth selection; Nearest-neighbor estimate; Nonparametric estimation

1. Introduction

Let $X,\dots,X_n$ be independent observations of an $\mathbb{R}^d$-valued random vector X with unknown density f. The classical kernel estimate of f is

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

* Corresponding author.

[1] Work supported by NSERC Grant A3456 and by FCAR Grant 90-ER-0291.

[2] Work supported by a grant from the Humboldt Foundation and by NSERC grant OGP0000270.

0167-7152/99/$ - see front matter
PII: S0167-7152(99)00021-8

where $h>0$ is a smoothing factor depending upon n, K is an absolutely integrable function (the kernel), and $K_h(x)=(1/h^d)K(x/h)$ (Akaike, 1954; Parzen, 1962; Rosenblatt, 1956). Observe that for the kernel $K(u)=1/\|u\|^d$, the smoothing factor h is cancelled and we obtain

$$f_n(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\|x-X_i\|^d}.$$

One may wonder what happens in this situation, now that the smoothing factor is absent. Unfortunately, in its unaltered form, we have $f_n(x)\to\infty$ in probability at almost all x (f). But a mere renormalization yields a consistent estimate of f. We formulate this as our introductory theorem:

Theorem 1. *The Hilbert estimate*

$$f_n(x)=\frac{1}{V_d n\log n}\sum_{i=1}^{n}\frac{1}{\|x-X_i\|^d},$$

where V_d is the volume of the unit ball in $\mathbb{R}^d$, is weakly consistent at almost all x, that is, $f_n(x)\to f(x)$ in probability at almost all x.

We use the name Hilbert estimate because of the related Hilbert integral with a similar kernel. There is also a Hilbert kernel regression function estimate, introduced and studied by Devroye et al. (1998). The first part of this paper is used to prove the consistency theorem. The shape of the Hilbert estimate is not useful for visualization purposes, as there are infinite peaks at all X_i's. Furthermore, $\int f_n=\infty$ for all n, so that the density estimate itself is not a density. For these reasons, modifications are proposed that have fewer disadvantages.

Connection with the nearest-neighbor estimates. The k-nearest-neighbor density estimate of Fix and Hodges (1951) and Loftsgaarden and Quesenberry (1965) is

$$g_{k,n}(x)\stackrel{\text{def}}{=}\frac{k}{nV_d\|x-X_{(k,x)}\|^d}$$

where $X_{(k,x)}$ is the kth nearest neighbor of x among $X_1,\ldots,X_n$. Its properties are well-understood (Moore and Yackel, 1977; Devroye and Wagner, 1977; Mack, 1980; Bhattacharya and Mack, 1987; Mack and Rosenblatt, 1979). For example, at almost all x, we have $g_{k,n}(x)\to f(x)$ as $n\to\infty$ if $k=\mathrm{o}(n)$ and $k\to\infty$. If we replace $\log n$ by the harmonic number $H_n=\sum_{i=1}^{n}1/i$ in the definition of f_n, then we have

$$f_n(x)\equiv\sum_{k=1}^{n}\frac{1}{kH_n}g_{k,n}(x).$$

Thus, the Hilbert kernel estimate is a harmonically weighted nearest-neighbor estimate! One could deduce the weak convergence results for f_n from those of $g_{k,n}$, but this is not the route we will follow. We will also not deal with the generalized weighted nearest neighbor-estimate

$$\sum_{k=1}^{n}w_{nk}g_{k,n}(x),$$

where w_{nk}, $1\leqslant k\leqslant n$, is a probability vector. The kth nearest-neighbor estimate corresponds to $w_{nj}=I_{j=k}$. There is indeed a problem of the selection of the best weight vector, but that will not be dealt with here.

Relation with variable kernel estimates. Through the thesis of Udina (1998), we came across an interesting and useful relationship with ordinary kernel estimates. Assume the kernel is the uniform density on $[-1,1]$, and the variable kernel estimate is given by

$$g_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{I_{|x-X_i|\leqslant h_i/2}}{h_i},$$

where the h_i's are positive bandwidths. Regardless of how these bandwidths are chosen – a difficult problem indeed – , we have

$$\sup_{h_1,\ldots,h_n} g_n(x) \leqslant \frac{1}{2n}\sum_{i=1}^{n}\frac{1}{|x-X_i|} = \log n\, f_n(x),$$

where f_n is the Hilbert kernel estimate. Modulo the logarithmic factor, our estimate is a uniform overbound of all variable kernel estimates. In other words, regardless how the bandwidths are picked individually and as a function of n, our results will imply that $g_n(x) = \mathrm{O}(\log n)$ in probability at almost all x, so catastrophic oscillations are unlikely.

Smart estimates. One of the reasons we considered Hilbert estimates, was that we thought that it might be a good candidate for a universally consistent density estimate whose expected performance (L_1, L_∞, pointwise) is monotone in n for all densities. Unfortunately, this is not the case. To date, we have not found one such density estimate!

2. The behavior of the estimate

It should be noted that the Hilbert density estimate scales perfectly. That is, no special adjustments are needed when all data are transformed by a linear transformation. For the multivariate kernel estimate to have a similar property, one should have a data-dependent smoothing factor that is somehow sensitive to such linear transformations. It is also noted that while f_n is not integrable, we can construct modified estimates (see below) that are in L_p for all $p > 1$.

3. A new class of density estimates

The estimates we propose to avoid the infinite peaks at the data points are symmetric in the data (for otherwise they would be suboptimal). We let $\|\,.\,\|$ denote the L_2 metric on $\mathbb{R}^d$. Furthermore, we pick a fixed integer $k > 0$. Then define the Hilbert density estimate of order k by

$$f_n(x) = \left[\frac{2^k}{V_d^k k V_k \binom{n}{k}\log n}\sum_{A\subseteq\{1,\ldots,n\}:|A|=k}\frac{1}{(\sum_{i\in A}\|x-X_i\|^{2d})^{k/2}}\right]^{1/k},$$

where V_d and V_k are the volumes of the unit balls of $\mathbb{R}^d$ and $\mathbb{R}^k$. For $k\geqslant 2$, the density estimate is almost surely bounded (for fixed n), and $\int f_n^s < \infty$ for all $s > 1$. In other words, $f_n \in L_s$ for all $s > 1$. As these nice properties hold for all $k > 1$, the simplest case $k = 2$ looms as the most important one. For $k = 1$, we

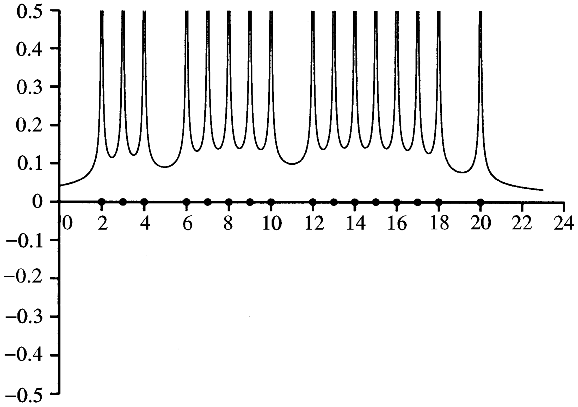

Fig. 1. The Hilbert kernel density estimate with $k=1$.

will simply speak of the Hilbert density estimate. For $k=2$, the estimate reduces to

$$f_n(x)=\sqrt{\frac{4}{V_d^2\pi n(n-1)\log n}\sum_{1\leqslant i<j\leqslant n}\frac{1}{\|x-X_i\|^{2d}+\|x-X_j\|^{2d}}}.$$

Because of the similarity with the multivariate Cauchy density, this will be dubbed the Cauchy density estimate.

Other metrics. We may analyze estimates similar to the Hilbert density estimate of order k by replacing $\|\,.\,\|$ by the L_p norm $\|\,.\,\|_p$ on $\mathbb{R}^d$, and by using an L_q norm for summing, thus obtaining

$$f_n(x)=\left[\frac{C(k,d,p,q)}{\binom{n}{k}\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k}\frac{1}{(\sum_{i\in A}\|x-X_i\|_p^{qd})^{k/q}}\right]^{1/k},$$

where C is a certain function of the parameters. However, this generalization would unnecessarily clutter the paper without adding substantial new information, and so we will not study it here.

Bootstrapped estimates. Note that the computation of the density estimate grows at least as n^k unless special data structures are used to reduce the complexity. Fortunately, we may bootstrap the estimates by picking a subclass B of all the sets A of cardinality k. This sort of estimate will not be studied here.

4. Shape of the estimates

In Figs. 1 and 2, we show the Hilbert density estimate for $d=1$, $k=1$ and $d=1$, $k=2$ with the same 16-point data set.

5. Weak pointwise consistency

In the next few sections, we prove the main consistency theorem:

Theorem 2. *For any $k\geqslant 1$, we have at almost all x, $f_n(x)\to f(x)$ in probability.*

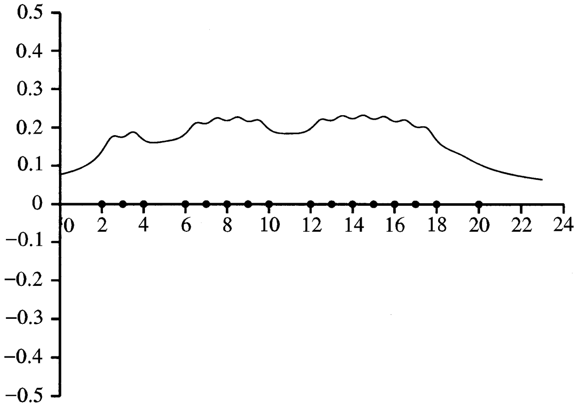

Fig. 2. The Hilbert kernel density estimate with $k = 2$.

6. A basic lemma

Lemma. *Let $U_1, \ldots, U_n$ be i.i.d. uniform* $[0,1]$ *random variables, let $k \in \mathcal{N}$, and let $V_k = \pi^{k/2}/\Gamma(k/2+1)$ denote the volume of the unit ball in $\mathbb{R}^k$. Then*

$$\frac{1}{\binom{n}{k}\log n} \sum_{A \subseteq \{1,\ldots,n\}:|A|=k} \frac{1}{(\sum_{i\in A} U_i^2)^{k/2}} \to \frac{kV_k}{2^k} \quad \textit{in probability.}$$

The proof of the Lemma requires a bit of care, as it does not suffice to apply standard results on U-statistics. Indeed, the summands have such large tails that the extra normalization factor $1/\log n$ is needed in front of the average. We may proceed in a variety of ways, but the one that we will follow uses the order statistics of the U_i's, denoted by $U_{(1)} < \cdots < U_{(n)}$. We may use a well-known connection between uniform samples and Poisson point processes. If $E_1, E_2, \ldots$ are i.i.d. standard exponential random variables, then

$$(U_{(1)}, \ldots, U_{(n)}) \stackrel{\mathscr{L}}{=} \left(\frac{\sum_{i=1}^{1} E_i}{\sum_{i=1}^{n+1} E_i}, \ldots, \frac{\sum_{i=1}^{n} E_i}{\sum_{i=1}^{n+1} E_i} \right)$$

(see, e.g., Ch. 8 of Shorack and Wellner, 1986). Let $\varepsilon \in (0,1)$. For integer $l \geqslant k$, define the event

$$B_l = \left[\bigcup_{i=l}^{\infty} \left[\left| \frac{E_1 + \cdots + E_i}{i} - 1 \right| \geqslant \varepsilon \right] \right].$$

By the Hájek–Rényi inequality,

$$\boldsymbol{P}\{B_l\} \leqslant \frac{2\mathrm{Var}\{E_1 + \cdots + E_l\}}{l^2 \varepsilon^2} = \frac{2}{l\varepsilon^2}.$$

Now, the complement B_l^c implies A_l, which is the event

$$A_l \stackrel{\text{def}}{=} \left[\bigcap_{i=l}^{n} \left[\frac{(n+1)U_{(i)}}{i} \in \left[\frac{1-\varepsilon}{1+\varepsilon}, \frac{1+\varepsilon}{1-\varepsilon} \right] \right] \right].$$

The subsets $A\subseteq\{1,\dots,n\}$ of size $|A|=k$ may be partitioned according to the cardinality of $\{1,\dots,l\}\cap A$, which we denote by c_A. Now note that on the event A_l,

$$\frac{1}{\binom{n}{k}\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k}\frac{1}{(\sum_{i\in A}U_i^2)^{k/2}}$$

$$\leqslant\frac{1}{\binom{n}{k}\log n}\sum_{j=0}^{k}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;c_A=j}\frac{1}{(jU_{(1)}^2+((1-\varepsilon/1+\varepsilon))^2\sum_{i\in A,\,i\geqslant l}(i/(n+1))^2)^{k/2}}$$

$$\leqslant\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^k\frac{1}{\binom{n}{k}\log n}\sum_{j=0}^{k-1}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\,c_A=j}\frac{1}{(\sum_{i\in A,\,i\geqslant l}(i/(n+1))^2)^{k/2}}$$

$$+\frac{1}{\binom{n}{k}\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\,c_A=k}\frac{1}{k^{k/2}U_{(1)}^k}$$

and

$$\frac{1}{\binom{n}{k}\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k}\frac{1}{(\sum_{i\in A}U_i^2)^{k/2}}\geqslant\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^k\frac{1}{\binom{n}{k}\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\;c_A=0}\frac{1}{(\sum_{i\in A}(i/(n+1))^2)^{k/2}}.$$

As $\boldsymbol{P}\{A_l\}>1-2/(l\varepsilon^2)$, we can for every ε make A_l as likely as desired by the choice of $l\geqslant k$. As ε is arbitrary, it remains to prove the following facts:

(A) $1/\binom{n}{k}\log n\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\,c_A=0}1/(\sum_{i\in A}(i/(n+1))^2)^{k/2}\to kV_k/2^k$;

(B) $1/\binom{n}{k}\log n\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\,c_A=k}1/k^{k/2}U_{(1)}^k\to 0$ in probability;

(C) $1/\binom{n}{k}\log n\sum_{j=1}^{k-1}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\,c_A=j}1/(\sum_{i\in A,\,i\geqslant l}(i/(n+1))^2)^{k/2}\to 0$.

Proof of part B. Note that $nU_{(1)}\xrightarrow{\mathscr{L}}E$ where $\xrightarrow{\mathscr{L}}$ denotes convergence in distribution and E is an exponential random variable. But $|\{A\subseteq\{1,\dots,n\}:|A|=k;\;c_A=k\}|=\binom{l}{k}$, and therefore,

$$\frac{1}{\binom{n}{k}\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;c_A=k}\frac{1}{k^{k/2}U_{(1)}^k}\leqslant\frac{\mathrm{O}(1)}{\log n\,(nU_{(1)})^k}\to 0\text{ in probability.}$$

Proof of part A. As k is fixed, it suffices to show the following:

$$\frac{k!}{\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;c_A=0}\frac{1}{(\sum_{i\in A}i^2)^{k/2}}\to\frac{kV_k}{2^k}.$$

Let $\mathscr{A}_{k,n}$ be the class of vectors $(i_1,\dots,i_k)$ of k elements drawn from $\{1,\dots,n\}$ (with duplicates allowed). This class has n^k members. Then, if θ denotes an arbitrary number between 0 and k^k,

$$\frac{1}{\log n}\sum_{(i_1,\dots,i_k)\in\mathscr{A}_{k,n}:\min_{1\leqslant m\leqslant k}i_m>l}\frac{1}{(\sum_{m=1}^k i_m^2)^{k/2}}$$

$$=\frac{k!}{\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\,c_A=0}\frac{1}{(\sum_{i\in A}i^2)^{k/2}}+\frac{\theta}{\log n}\sum_{j=1}^{k}\binom{k}{j}\sum_{(i_1,\dots,i_j)\in\mathscr{A}_{j,n}:\min_{1\leqslant m\leqslant j}i_m>l}\frac{1}{(\sum_{m=1}^j i_m^2)^{k/2}}.$$

In view of part C (which will be proved below), it suffices to show that

$$L\stackrel{\text{def}}{=}\frac{1}{\log n}\sum_{(i_1,\dots,i_k)\in\mathscr{A}_{k,n}:\min_{1\leqslant m\leqslant k}i_m>l}\frac{1}{(\sum_{m=1}^k i_m^2)^{k/2}}\to\frac{kV_k}{2^k}.$$

Consider the continuous approximant given by

$$L' \stackrel{\text{def}}{=} \frac{1}{\log n} \int_{[l,\,n]^k} \|x\|^{-k} \,\mathrm{d}x,$$

where x denotes a vector of $\mathbb{R}^k$. Assume $k \geqslant 2$. Clearly,

$$\begin{aligned}
0 &\leqslant L' - L \\
&\leqslant \frac{1}{\log n} \int_{[l,\,n]^k} (\|x\|^{-k} - \|x+1\|^{-k}) \,\mathrm{d}x \\
&\qquad \text{(where 1 is a vector of } k \text{ ones)} \\
&= \frac{1}{\log n} \int_{[l,\,n]^k} \left(\frac{\|x+1\|^k - \|x\|^k}{\|x\|^k \|x+1\|^k} \right) \mathrm{d}x \\
&\leqslant \frac{1}{\log n} \int_{[l,\,n]^k} \left(\frac{(k/2)\|x+1\|^{k-2}(\|x+1\|^2 - \|x\|^2)}{\|x\|^k \|x+1\|^k} \right) \mathrm{d}x \\
&= \frac{1}{\log n} \int_{[l,\,n]^k} \left(\frac{(k/2)(2\sum_{i=1}^k x_i + k)}{\|x\|^k \|x+1\|^2} \right) \mathrm{d}x \\
&\leqslant \frac{k}{\log n} \int_{[l,\,n]^k} \left(\frac{1}{\|x\|^{k+1}} \right) \mathrm{d}x + \frac{k^2}{2\log n} \int_{[l,\,n]^k} \left(\frac{1}{\|x\|^{k+2}} \right) \mathrm{d}x \\
&\leqslant \frac{k}{\log n} \int_{r \geqslant l} \left(\frac{kV_k r^{k-1}}{r^{k+1}} \right) \mathrm{d}r + \frac{k^2}{2\log n} \int_{r \geqslant l} \left(\frac{kV_k r^{k-1}}{r^{k+2}} \right) \mathrm{d}r \\
&\qquad \text{(by polar coordinate transforms)} \\
&= \mathrm{O}\left(\frac{1}{\log n} \right).
\end{aligned}$$

It is trivial to verify $L' - L = \mathrm{O}(1/\log n)$ for $k = 1$ as well. As $L' - L \to 0$, it suffices to study L'. By simple bounding and polar coordinate transformation, we have

$$\begin{aligned}
L' &\leqslant \frac{1}{2^k \log n} \int_{n\sqrt{k} \geqslant r \geqslant l} \left(\frac{kV_k r^{k-1}}{r^k} \right) \mathrm{d}r = \frac{kV_k}{2^k \log n} \int_{n\sqrt{k} \geqslant r \geqslant l} \left(\frac{1}{r} \right) \mathrm{d}r \\
&= \frac{kV_k \log(n\sqrt{k}/l)}{2^k \log n} \to \frac{kV_k}{2^k}.
\end{aligned}$$

By similar methods, a lower bound may be obtained, and we conclude that $L' \to kV_k/2^k$. This concludes the proof of part A.

Proof of part C. Fix $j \in \{1,\dots,k-1\}$, and assume $k>1$. Thus, by arguments as in the proof of part A,

$$\frac{1}{\binom{n}{k}\log n}\sum_{A\subseteq\{1,\dots,n\}:|A|=k;\,c_A=j}\frac{1}{(\sum_{i\in A,\,i>l}(i/(n+1))^2)^{k/2}}$$

$$=\frac{1}{\binom{n}{k}\log n}\binom{l}{j}\sum_{A\subseteq\{l+1,\dots,n\}:|A|=k-j}\frac{1}{(\sum_{i\in A}(i/(n+1))^2)^{k/2}}$$

$$\sim\frac{k!\binom{l}{j}}{\log n}\sum_{A\subseteq\{l+1,\dots,n\}:|A|=k-j}\frac{1}{(\sum_{i\in A}i^2)^{k/2}}$$

$$\leqslant\frac{k!\,l^j}{\log n}\int_{[l,\,n]^{k-j}}\|x\|^{-k}\,\mathrm{d}x$$

$$\leqslant\frac{k!\,l^j}{\log n}\int_{r\geqslant l}\frac{(k-j)V_{k-j}r^{k-j-1}}{r^k}\,\mathrm{d}r$$

$$=\frac{k!\,l^j(k-j)V_{k-j}}{\log n}\int_{r\geqslant l}\frac{1}{r^{j+1}}\,\mathrm{d}r$$

$$=\mathrm{O}\left(\frac{1}{\log n}\right).$$

This concludes the proof of part C and thus of the Lemma. □

7. Proof of Theorem 2

Let $S(x,r)$ denote the closed ball in $\mathbb{R}^d$ of radius r centered at x. We will show the convergence result for all Lebesgue points of f, that is, for all x for which $f(x)>0$ and for which at the same time

$$\lim_{r\downarrow 0}\frac{\int_{S(x,r)}f(y)\,\mathrm{d}y}{\int_{S(x,r)}\mathrm{d}y}=f(x).$$

As f is a density, we know that almost all x satisfy the properties given above (Wheeden and Zygmund, 1977, p. 189; see also Devroye, 1981, Lemma 1.1). Let x be such a point.

Fix $\varepsilon\in(0,1)$ and find $\delta>0$ such that

$$\sup_{0<r\leqslant\delta}\left|\frac{\int_{S(x,r)}f(y)\,\mathrm{d}y}{\int_{S(x,r)}\mathrm{d}y}-f(x)\right|\leqslant\varepsilon f(x).$$

Define $p=\int_{S(x,\delta)}f$. Let F be the univariate distribution function of $W\stackrel{\text{def}}{=}\|x-X\|^dV_d$. Note that F has a density and that if $u\leqslant V_d\delta^d$,

$$F(u)=\boldsymbol{P}\{V_d\|x-X\|^d\leqslant u\}=\boldsymbol{P}\{X\in S(x,(u/V_d)^{1/d})\}$$

$$=\int_{S(x,(u/V_d)^{1/d})}f(y)\,\mathrm{d}y\in[(1-\varepsilon)f(x)u,(1+\varepsilon)f(x)u].$$

Define $W_i=V_d\|x-X_i\|^d$, $1\leqslant i\leqslant n$, and let $W_{(1)}<\cdots<W_{(n)}$ be the order statistics for $W_1,\dots,W_n$. If $U_{(1)}<\cdots<U_{(n)}$ are uniform order statistics, we have in fact the representation

$$U_{(i)}\stackrel{\mathscr{L}}{=}F(W_{(i)}),\quad W_{(i)}\stackrel{\mathscr{L}}{=}F^{\mathrm{inv}}(U_{(i)})$$

jointly for all i. Thus,

$$(1-\varepsilon)f(x)W_{(i)} \leqslant U_{(i)} \leqslant (1+\varepsilon)f(x)W_{(i)}$$

provided $W_{(i)} \leqslant V_d\delta^d$. Put differently, under the latter condition,

$$\frac{U_{(i)}}{(1+\varepsilon)f(x)} \leqslant W_{(i)} \leqslant \frac{U_{(i)}}{(1-\varepsilon)f(x)}.$$

The Hilbert estimate $f_n^k(x)$ may be written as follows:

$$f_n^k(x) = \frac{2^k}{V_{k-1}\binom{n}{k}\log n} \sum_{A\subseteq\{1,\ldots,n\}:|A|=k} \frac{1}{(\sum_{i\in A} W_i^2)^{k/2}}.$$

Let B be the set of indices $i \leqslant n$ with $W_i \leqslant V_d\delta^d$, and let $B^c = \{1,\ldots,n\} - B$. Then, if θ denotes an arbitrary random variable with values in $[(1-\varepsilon)^k,(1+\varepsilon)^k]$, and η denotes a random variable with value in $[0,1]$, then

$$f_n^k(x) = \frac{2^k\theta(f(x))^k}{V_{k-1}\binom{n}{k}\log n} \sum_{A\subseteq B:|A|=k} \frac{1}{\left(\sum_{i\in A} U_i^2\right)^{k/2}} + \frac{2^k\eta}{V_{k-1}\binom{n}{k}\log n} \sum_{A\subseteq\{1,\ldots,n\}:|A|=k,|A\cap B^c|>0} \frac{1}{(\delta^{2d})^{k/2}}.$$

The last term is obviously $O(1/\log n)$ and thus does not matter. We thus have $f_n^k(x) \to f^k(x)$ if

$$\frac{2^k}{V_{k-1}\binom{n}{k}\log n} \sum_{A\subseteq B:|A|=k} \frac{1}{(\sum_{i\in A} U_i^2)^{k/2}} \to 1 \text{ in probability.}$$

Set $p = F(V_d\delta^d)$, and note that given $U_i < p$, U_i is distributed as $Z_i p$, where Z_i is uniform $[0,1]$. Thus, conditional on $N = \sum_{i=1}^n I_{U_i<p}$, we have

$$\frac{2^k}{V_{k-1}\binom{n}{k}\log n} \sum_{A\subseteq B:|A|=k} \frac{1}{(\sum_{i\in A} U_i^2)^{k/2}} = \frac{2^k}{V_{k-1}p^k\binom{n}{k}\log n} \sum_{A\subseteq\{1,\ldots,N\}:|A|=k} \frac{1}{(\sum_{i\in A} Z_i^2)^{k/2}}.$$

If $N \to \infty$ in probability, we have by the lemma,

$$\frac{2^k}{V_{k-1}\binom{N}{k}\log N} \sum_{A\subseteq\{1,\ldots,N\}:|A|=k} \frac{1}{(\sum_{i\in A} Z_i^2)^{k/2}} \to 1 \text{ in probability.}$$

Thus, it remains to show that $N \to \infty$ in probability and $N/(pn) \to 1$ in probability, so that we may conclude. But clearly, N is binomial (n,p), and $p > 0$ (as x is a Lebesgue point), so that the Theorem is proved. □

8. Lack of strong convergence

For all f, it is true that at almost all x with $f(x) > 0$, the Hilbert estimates of order k cannot possibly converge to f in a strong sense. Rather than to prove the full-blown universal theorem, we restrict ourselves to the uniform density and show the following result.

Theorem 3. *Let f be the uniform density on $[0,1]^d$ and let $k \geqslant 1$ be arbitrary. Then, for any $x \in [0,1]^d$, $\boldsymbol{P}\{f_n(x) \geqslant (\log\log n)^{1/k} \text{ i.o.}\} = 1$, so that there is no strong convergence at any point in the support.*

Proof. We cut the sequence of indices into sections, with the ith section consisting of $\{2^{i-1}+1,\ldots,2^i\}$. Let $n = 2^i$. Then, for fixed $x \in [0,1]^d$, we say that X_j is near x if

$$V_d\|x - X_j\|^d \leqslant \frac{1}{n(a\log n\log\log n)^{1/k}},$$

where $a>0$ is an arbitrary number. In the ith section, the number of X_j's near x is binomial $(n/2, p_n)$ with

$$\frac{1}{n(a\log n\log\log n)^{1/k}}\, p_n \geqslant \frac{1}{2^d n(a\log n\log\log n)^{1/k}}.$$

In particular, the probability that this section has exactly k points near x is at least

$$\binom{n/2}{k} p_n^k (1-p_n)^{n/2-k} \sim \frac{(np_n)^k}{2^k k!} \quad (\text{as } np_n \to 0)$$
$$\geqslant \frac{1}{2^{dk}2^k k! a\log n\log\log n} \sim \frac{1}{2^{dk}2^k k! ai\log i\log 2}.$$

By the Borel–Cantelli lemma and independence of the sections, the number of sections with k points near x at times $n=2^i$ is infinite with probability one if and only if the probabilities of these events sum to infinity, which is the case here as $\sum_i 1/(i\log i)=\infty$. Now, if a section has k points near x at time $n=2^i$, then, if C is a constant depending upon k and d only (as in Theorem 2), then

$$f_n(x) \geqslant \left[\frac{C}{\binom{n}{k}\log n} \frac{1}{k^{k/2}\left(V_d^k n^k a\log n\log\log n\right)^{-1}} \right]^{1/k}.$$

Adjust a to conclude that

$$\boldsymbol{P}\{f_n(x) > (\log\log n)^{1/k} \text{ i.o.}\} = 1.$$

This concludes the proof of Theorem 3. □

Rate of convergence. The poor rate of convergence of the estimate is best seen by considering points outside the support of f. If x is at least distance s away from the support of f, then $f_n(x)\geqslant c/(s^d(\log n)^{1/k})$ for some constant c only depending upon k and d.

References

Akaike, H., 1954. An approximation to the density function. Ann. Inst. Statist. Math. 6, 127–132.

Bhattacharya, P.K., Mack, Y.P., 1987. Weak convergence of k-NN density and regression estimators with varying k and applications. Ann. Statist. 15, 976–994.

Devroye, L., 1981. On the almost everywhere convergence of nonparametric regression function estimates. Ann. Statist. 9, 1310–1319.

Devroye, L., Györfi, L., Krzyżak, A., 1998. The Hilbert kernel regression estimate. J. Multivariate Anal. 65, 209–227.

Devroye, L.P., Wagner, T.J., 1977. The strong uniform consistency of nearest neighbor density estimates. Ann. Statist. 5, 536–540.

Fix, E., Hodges, J.L., 1951. Discriminatory analysis, nonparametric discrimination, consistency properties. Rep. 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

Loftsgaarden, D.O., Quesenberry, C.P., 1965. A nonparametric estimate of a multivariate density function. Ann. Math. Statist. 36, 1049–1051.

Mack, Y.P., 1980. Asymptotic normality of multivariate k-NN density estimates. Sankhya A42, 53–63.

Mack, Y.P., Rosenblatt, M., 1979. Multivariate k-nearest neighbor density estimates. J. Multivariate Anal. 9, 1–15.

Moore, D.S., Yackel, J.W., 1977. Consistency properties of nearest neighbor density function estimators. Ann. Statist. 5, 143–154.

Parzen, E., 1962. On the estimation of a probability density function and the 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.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Udina, F., 1998. Mètodes no paramètrics d'estimació de densitat i de regressió: aspects computacionals i d'implementació. Ph.D. Dissertation, EIO, Universitat Politecnica de Catalunya.

Wheeden, R.L., Zygmund, A., 1977. Measure and Integral. Marcel Dekker, New York, 1977.