The Annals of Statistics
1994, Vol. 22, No. 2, 886–889

ON GOOD DETERMINISTIC SMOOTHING SEQUENCES FOR KERNEL DENSITY ESTIMATES[1]

BY LUC DEVROYE

McGill University

We use the probabilistic method to show that if f_{nh} is the standard kernel estimate with smoothing factor h, then there exists a deterministic sequence h_n such that, for all densities,

$$\liminf_{n\to\infty} \frac{\mathbf{E}\int |f_{nh_n} - f|}{\inf_h \mathbf{E}\int |f_{nh} - f|} = 1.$$

1. Introduction. Let $X_1, \ldots, X_n$ be i.i.d. random variables with common density f on the real line. We consider the kernel estimate

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

where $K_h(x) = (1/h)K(x/h)$, $h > 0$, is the smoothing factor depending upon n only, and K, the *kernel*, is a given function integrating to 1 [Akaike (1954), Rosenblatt (1956) and Parzen (1962)]. Sometimes we will write f_{nh} to make the dependence upon h explicit. We assume throughout that K is L_1-Lipschitz, that is, that there exists a constant C such that

$$\int |K_u(x) - K_v(x)|\,dx \le \frac{C|u - v|}{\max(u, v)}.$$

Furthermore, we require that the smallest symmetric unimodal majorant of $|K|$ be in L_1 and L_4. (Both conditions are satisfied for all kernels of general interest.) The L_1-error given by

$$J_{nh} = \int |f_{nh} - f|$$

measures in many situations the quality of the estimate f_n.

THEOREM 1. *There exists a deterministic sequence h_n such that, for all densities,*

$$\liminf_{n\to\infty} \frac{\mathbf{E}\int |f_{nh_n} - f|}{\inf_h \mathbf{E}\int |f_{nh} - f|} = 1.$$

Received January 1993; revised September 1993.
[1]Supported by NSERC Grant A3456 and FCAR Grant 90-ER-0291.
AMS 1991 *subject classifications.* Primary 62G07; secondary 62G05, 62F12, 60F25.
Key words and phrases. Density estimation, kernel estimate, probabilistic method, nonparametric methods, smoothing.

This theorem shows that there is a deterministic sequence of smoothing factors that is asymptotically optimal for any density in the world, at least along a subsequence. What is interesting is that there is an uncountable continuum of possible rates to zero for $\inf_h \mathbf{E}\int |f_{nh} - f|$. (To see this, play a bit with the unsmoothness or the tails.) Yet, our sequence has only countably many values.

One would think that data-based smoothing sequences should do better than this. Maybe we might even suspect that there exists a sequence of functions H_n: $\mathbb{R}^n \to (0, \infty)$ (called data-based smoothing factors) such that, for example,

$$\frac{\int |f_{nH_n} - f|}{\inf_h \int |f_{nh} - f|} \to 1$$

almost surely for all densities, where $H_n = H_n(X_1, \ldots, X_n)$. However, such a rule has not been exhibited to date. In fact, it is probably futile to look for one. Theorem 1 simply says that if we are going to prove that any data-based smoothing sequence is poor, it can only be provably poor along subsequences.

The proof is nonconstructive. However, with probability 1, an i.i.d. exponential sequence will do. While almost every exponential random sequence has the optimality property stated in the theorem, no data-based smoothing factor published in the literature shares this property, as all methods I am aware of are asymptotically suboptimal on given subclasses of densities.

2. Proof.

We introduce two real number sequences, α_n and β_n linked by the relation

$$\mathbf{E}\int |f_{n\beta_n} - f| = \inf_h \mathbf{E}\int |f_{nh} - f| = \alpha_n.$$

The existence of β_n follows from the continuity of the L_1 criterion with respect to h. From Devroye and Györfi [(1985), page 12], we have $n\beta_n \to \infty$. If the kernel K has a characteristic function that is not identically 1 in an open neighborhood of the origin, or if f has a characteristic function of unbounded support, then $\beta_n \to 0$ as well [Devroye (1989), Lemma S1]. For now, we assume such a situation. Also, $\alpha_n \to 0$ for all densities. For fixed $\varepsilon > 0$, we further note that if $h \in (\beta_n - \varepsilon\alpha_n\beta_n, \beta_n + \varepsilon\alpha_n\beta_n)$, then

$$\begin{aligned}
\mathbf{E}\int |f_{nh} - f| &\le \mathbf{E}\int |f_{n\beta_n} - f| + \mathbf{E}\int |f_{n\beta_n} - f_{nh}| \\
&\le \alpha_n + \int |K_{\beta_n} - K_h| \\
&\le \alpha_n + \frac{C|\beta_n - h|}{\max(\beta_n, h)} \\
&\le \alpha_n(1 + C\varepsilon).
\end{aligned}$$

If we take a random sequence of i.i.d. exponential random variables H_n as smoothing factors and make sure that the sequence is also independent of the

data, then

$$\mathbf{P}\{|H_n - \beta_n| \le \varepsilon\alpha_n\beta_n\} = \big(2 + o(1)\big)\varepsilon\alpha_n\beta_n.$$

Let A_n be the event that

$$\frac{\mathbf{E}\{J_{nH_n} \mid H_n\}}{\inf_h \mathbf{E}J_{nh}} \le 1 + C\varepsilon.$$

Then

$$\mathbf{P}\{A_n\} \ge \big(2 + o(1)\big)\varepsilon\alpha_n\beta_n.$$

As the A_n's are independent, we see that

$$\mathbf{P}\{A_n \text{ i.o.}\} = 1$$

when

$$\sum_{n=1}^{\infty} \alpha_n\beta_n = \infty.$$

By Devroye and Györfi [(1985), page 139],

$$\alpha_n \ge \frac{1}{2}\mathbf{E}\int |f_{n\beta_n} - K_{\beta_n} * f|.$$

Next, by Devroye [(1988), Lemma 5] and the fact that $\beta_n \to 0$,

$$\liminf_{n\to\infty} \sqrt{n\beta_n}\,\alpha_n \ge \frac{\sqrt{\int K^2}\int\sqrt{f}}{4}.$$

We therefore need only verify that

$$\sum_{n=1}^{\infty} \sqrt{\frac{\beta_n}{n}} = \infty;$$

but this is a simple consequence of the fact that $n\beta_n \to \infty$. We have shown that, for our random sequence,

$$\mathbf{P}\{\forall\, \varepsilon > 0\colon A_n \text{ i.o.}\} = 1.$$

Therefore, there exists at least one deterministic sequence $\{h_n\}$ such that, for all $\varepsilon > 0$,

$$\frac{\mathbf{E}\{J_{nh_n}\}}{\inf_h \mathbf{E}J_{nh}} \le 1 + C\varepsilon$$

for infinitely many n. For more examples of existence proofs through randomization, we refer to the literature on the so-called probabilistic method [see Alon, Spencer and Erdös (1992)].

When the kernel has a characteristic function that is identically 1 in an open neighborhood of the origin, and the characteristic function of f vanishes outside a compact set, then β_n tends to a constant $\beta > 0$. It is easy to modify the proof to handle this case as well. Note that this is the reason why we need a density with full support on $[0, \infty)$ such as the exponential density instead of, say, the uniform $[0, 1]$ density, as we want to ensure that β is in the support of the H_n sequence. □

Acknowledgments. The comments by a referee and an Editor are greatly appreciated.

REFERENCES

AKAIKE, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math.* **6** 127–132.

ALON, N., SPENCER, J. and ERDÖS, P. (1992). *The Probabilistic Method.* Wiley, New York.

DEVROYE, L. (1988). Asymptotic performance bounds for the kernel estimate. *Ann. Statist.* **16** 1162–1179.

DEVROYE, L. (1989). The double kernel method in density estimation. *Ann. Inst. H. Poincaré* **25** 533–580.

DEVROYE, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20** 2037–2056.

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

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.

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
3480 UNIVERSITY STREET
MONTREAL, QUEBEC
CANADA H3A 2A7