Statistics & Probability Letters 8 (1989) 419–423
North-Holland
October 1989

A UNIVERSAL LOWER BOUND FOR THE KERNEL ESTIMATE

Luc DEVROYE
School of Computer Science, McGill University, Montreal, Canada H3A 2K6

Received May 1988
Revised September 1988

Abstract: Let f_n be the kernel density estimate with arbitrary smoothing factor h and arbitrary (absolutely integrable) kernel K, based upon an i.i.d. sample of size n drawn from a density f. It is shown that

$$\inf_{h,K,f} E\left(\int |f_n - f|\right) \geqslant \frac{1}{\sqrt{528n}},$$

and that

$$\liminf_{n\to\infty} \sqrt{n} \inf_{h,K,f} E\left(\int |f_n - f|\right) \geqslant \tfrac{1}{16}.$$

Keywords: density estimation, L_1 error, inequalities, empirical characteristic function, kernel estimate.

1. Introduction

In this paper, we give a short proof of a lower bound for the expected L_1 error for the kernel density estimate

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

where K is an absolutely integrable function (the kernel), $h > 0$ is a smoothing factor, $K_h(x) \triangleq (1/h)K(x/h)$, and $X_1, \ldots, X_n$ are i.i.d. random variables with common density f on the real line (Rosenblatt, 1956; Parzen, 1962). The main result is

Theorem 1.

$$\inf_{K,h,f} E\left(\int |f_n - f|\right) \geqslant \frac{1}{\sqrt{512n}\sqrt{1 + 1/(32n)}} \geqslant \frac{1}{\sqrt{528n}}.$$

Theorem 1 states that even if we are allowed to choose f, K and h, we can't possibly have an expected error that is roughly speaking smaller than about $1/(\sqrt{512n})$. The lower bound is the price we have to pay for the use of the kernel estimate. This result could be used to determine if n is large enough for someone to be able to use the kernel estimate. It also states that under no circumstances can the kernel estimate compare favorably with an estimate which, for a given f, has an expected error $o(1/\sqrt{n})$. The latter estimates are usually "parametric".

Research of the author was supported by NSERC grant A3456 and FCAR grant EQ-1679.

There are two kinds of universal lower bounds one can study. First, one might consider lower bounds for

$$\inf_{K,h} E\left(\int |f_n - f|\right).$$

The bound depends upon n and f only, and is in the spirit of a celebrated L_2 lower bound obtained by Watson and Leadbetter (1963) (see also Davis, 1975, 1977). Unfortunately, if one takes the infimum over all f of the Watson–Leadbetter lower bound, one obtains zero, since the L_2 criterion is not scale invariant. Furthermore, lower bounds that depend upon f are less useful for determining sample sizes since one does not know f in the first place.

For another density-free lower bound for the kernel estimate, see for example Devroye and Penrod (1984) (or Devroye and Györfi, 1985):

$$\inf_{K\in \boldsymbol{K},h,f} E\left(\int |f_n - f|\right) \geqslant (0.86 + \mathrm{o}(1))n^{-2/5}.$$

Here the class $\boldsymbol{K}$ is the class of all symmetric densities. It is well known that by allowing negative-valued K with vanishing positive moments, better rates than $n^{-2/5}$ are achievable for very smooth densities f. In fact, the rate that one can achieve is limited by the smallest s for which $\int x^s K \neq 0$. The rate $1/\sqrt{n}$ can be obtained for densities f with bounded spectrum (i.e., whose characteristic function has compact support), provided that a kernel is used whose Fourier transform is flat in an open neighborhood of the origin (see Devroye and Györfi, 1985, Section V.11, for a discussion; see also Ibragimov and Khasminskii, 1982).

Note also the contrast with minimax lower bounds, which are valid for all estimates (not just the kernel estimate), but only tell us about the worst density in a given target class of densities.

2. Proof of Theorem 1

At a crucial junction, we will need the following lower bound (see, e.g., Devroye and Györfi, 1985).

Lemma 1. *Let $X_1,\ldots, X_n$ be i.i.d. zero mean random variables with finite first absolute moment. Then*

$$E\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right|\right) \geqslant \frac{1}{\sqrt{8}}E(|X_1|). \qquad \square$$

The proof of Theorem 1 is based upon Fourier transforms. Let ϕ be the characteristic function for f, and let ψ be the Fourier transform of K_h defined by $\psi(t) = \int \mathrm{e}^{\mathrm{i}tx} K_h(x)\,\mathrm{d}x$. Observe that the convolution $f * K_h$ has Fourier transform $\phi(t)\psi(t)$. The dependence upon h is absorbed in ψ. From simple inequalities (see, e.g., Devroye and Györfi, 1985, p. 139), we have

$$E\left(\int |f - f_n|\right) \geqslant \int |f - f * K_h| \geqslant \sup_t |\phi(t) - \phi(t)\psi(t)| = \sup |\phi|\,|1-\psi|.$$

Also,

$$E\left(\int |f - f_n|\right) \geqslant \tfrac{1}{2}E\left(\int |f_n - f * K_h|\right) \geqslant \tfrac{1}{2}E\left(\sup_t |\phi_n - \phi|\,|\psi|\right) \geqslant \tfrac{1}{2}\sup_t E(|\phi_n - \phi|\,|\psi|)$$

where $\phi_n = (1/n)\sum_{j=1}^{n} \mathrm{e}^{\mathrm{i}tX_j}$ is the empirical characteristic function. Note that we used the fact that the

Fourier transform of f_n is $\phi_n \psi$. Now, we can finally use Lemma 1. Indeed, $|\phi_n - \phi| \geqslant |\mathrm{Re}(\phi_n) - \mathrm{Re}(\phi)|$. A similar inequality holds true for the imaginary part. But

$$\mathrm{Re}(\phi_n) - \mathrm{Re}(\phi) = \frac{1}{n} \sum_{j=1}^{n} \left(\cos(tX_j) - \mathrm{Re}(\phi(t))\right)$$

is an average of n i.i.d. zero mean random variables, whose absolute value does not exceed 2. By Lemma 1,

$$\begin{aligned} \sqrt{n}\, E\left(|\mathrm{Re}(\phi_n) - \mathrm{Re}(\phi)|\right) &\geqslant (1/\sqrt{8})\, E\left(|\cos(tX_1) - \mathrm{Re}(\phi(t))|\right) \\ &\geqslant (1/\sqrt{32})\, E\left((\cos(tX_1) - \mathrm{Re}(\phi(t)))^2\right) \\ &= (1/\sqrt{32})\left(\tfrac{1}{2} + \tfrac{1}{2}\mathrm{Re}(\phi(2t)) - \mathrm{Re}^2(\phi(t))\right). \end{aligned}$$

A similar argument on the imaginary part can be used to show that

$$\sqrt{n}\, E\left(|\mathrm{Im}(\phi_n) - \mathrm{Im}(\phi)|\right) \geqslant (1/\sqrt{32})\left(\tfrac{1}{2} - \tfrac{1}{2}\mathrm{Re}(\phi(2t)) - \mathrm{Im}^2(\phi(t))\right).$$

Averaging the two bounds yields

$$\sqrt{n}\, E(|\phi_n - \phi|) \geqslant (1/\sqrt{128})(1 - |\phi|^2).$$

Collecting all the bounds shows that

$$\begin{aligned} E\left(\int |f_n - f|\right) &\geqslant \max\left(\sup_t |1-\psi|\,|\phi|,\ \sup_t |\psi|\, n^{-1/2}(1/\sqrt{128})(1 - |\phi|^2)\right) \\ &\geqslant \sup_t \max\left(|1-\psi|\,|\phi|,\ |\psi|\, n^{-1/2}(1/\sqrt{128})(1 - |\phi|^2)\right) \\ &\geqslant \tfrac{1}{2} \sup_t \min\left(|\phi|,\ n^{-1/2}(1/\sqrt{128})(1 - |\phi|^2)\right). \end{aligned}$$

The minimum is maximal when both operands are equal, which occurs for

$$\begin{aligned} |\phi| &= \tfrac{1}{2}\left(-\sqrt{128n} + \sqrt{128n+4}\right) = \tfrac{1}{2}\sqrt{128n}\left(-1 + \sqrt{1 + 4/(128n)}\right) \\ &\geqslant \tfrac{1}{2}\sqrt{128n}\left(-1 + 1 + \frac{4}{256n\sqrt{1 + 4/(128n)}}\right) \quad \left(\text{use } \sqrt{1+\xi} \geqslant 1 + \frac{\xi}{2\sqrt{1+\xi}}\right) \\ &= \frac{1}{\sqrt{128n}\sqrt{1 + 1/(32n)}}. \end{aligned}$$

The value for $|\phi|$ is between zero and one, so that by the continuity of ϕ, it is attained for some t. The sought lower bound is $\frac{1}{2}$ times the given value of $|\phi|$. □

Table 1 gives some constants.

Table 1

Sample size n	Lower bound	Sample size n	Lower bound
1	0.043519	10000	0.000441
10	0.013953	100000	0.000139
100	0.004418	1000000	0.000044
1000	0.001397	10000000	0.000013

3. An asymptotic bound

We can give another inequality which is slightly better (by a factor of $1/\sqrt{2}$) than the bound of Theorem 1 for very large n. It is not as good for small and moderate values of n. We will only indicate where the proofs require modification.

Theorem 2.

$$\inf_{K,h,f} E\left(\int |f_n - f|\right) \geqslant \frac{1}{16\sqrt{n}}\left(\frac{1-8/(4n)^{1/4}}{1+1/(4\sqrt{n})}\right).$$

In particular,

$$\liminf_{n\to\infty} \inf_{K,h,f} \sqrt{n}\, E\left(\int |f_n - f|\right) \geqslant \tfrac{1}{16}.$$

At a crucial junction, we will need the following lower bound (see Devroye, 1988, Lemma 2).

Lemma 2. *Let $X_1, \ldots, X_n$ be i.i.d. zero mean random variables with finite first fourth absolute moment. Then*

$$E\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right|\right) \geqslant \tfrac{1}{2}\sqrt{E(X_1^2)} - (1/4n)^{1/4} E^{1/4}(X_1^4). \quad \square$$

Proof of Theorem 2. We proceed as in the proof of Theorem 1. By Lemma 2,

$$\begin{aligned}\sqrt{n}\, E(|\mathrm{Re}(\phi_n) - \mathrm{Re}(\phi)|) &\geqslant \tfrac{1}{2}\sqrt{E\left((\cos(tX_1) - \mathrm{Re}(\phi(t)))^2\right)} - 2(1/4n)^{1/4}\\ &\geqslant \tfrac{1}{2}\sqrt{E\left(\cos^2(tX_1) - \mathrm{Re}^2(\phi(t))\right)} - 2(1/4n)^{1/4}\\ &= \tfrac{1}{2}\sqrt{\tfrac{1}{2} + \tfrac{1}{2}\mathrm{Re}(\phi(2t)) - \mathrm{Re}^2(\phi(t))} - 2(1/4n)^{1/4}.\end{aligned}$$

A similar argument on the imaginary part can be used to show that

$$\sqrt{n}\, E\left(|\mathrm{Im}(\phi_n) - \mathrm{Im}(\phi)|\right) \geqslant \tfrac{1}{2}\sqrt{\tfrac{1}{2} - \tfrac{1}{2}\mathrm{Re}(\phi(2t)) - \mathrm{Im}^2(\phi(t))} - 2(1/4n)^{1/4}.$$

Averaging the two bounds, and using the inequality $\sqrt{u+v} \leqslant \sqrt{u} + \sqrt{v}$ yields

$$\sqrt{n}\, E(|\phi_n - \phi|) \geqslant \tfrac{1}{4}\sqrt{1 - |\phi(t)|^2} - 2(1/4n)^{1/4} \geqslant \tfrac{1}{4}(1 - 2|\phi|) - (4/n)^{1/4}.$$

Collecting all the bounds shows that

$$\begin{aligned}E\left(\int |f_n - f|\right) &\geqslant \max\left(\sup_t |1-\psi|\,|\phi|,\ \sup_t |\psi|\, n^{-1/2}\left(\tfrac{1}{8}(1-2|\phi|) - (1/4n)^{1/4}\right)\right)\\ &\geqslant \sup_t \max\left(|1-\psi|\,|\phi|,\ |\psi|\, n^{-1/2}\left(\tfrac{1}{8}(1-2|\phi|) - (1/4n)^{1/4}\right)\right)\\ &\geqslant \tfrac{1}{2}\sup_t \min\left(|\phi|,\ n^{-1/2}\left(\tfrac{1}{8}(1-2|\phi|) - (1/4n)^{1/4}\right)\right).\end{aligned}$$

The minimum is maximal when both operands are equal, which occurs for

$$|\phi| = \frac{1/(8\sqrt{n}) - 1/(4n^3)^{1/4}}{1 + 1/(4\sqrt{n})}.$$

The lower bound in question is half this value of $|\phi|$. $\square$

References

Davis, K.B. (1975), Mean square error properties of density estimates, *Ann. Statist.* **5**, 1025–1030.

Davis, K.B. (1977), Mean integrated square error properties of density estimates, *Ann. Statist.* **5**, 530–535.

Devroye, L. (1987), *A Course in Density Estimation* (Birkhauser, Basel).

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

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

Devroye, L. and C.S. Penrod (1984), Distribution-free lower bounds in density estimation, *Ann. Statist.* **12**, 1250–1262.

Ibragimov, I.A. and R.Z. Khasminskii (1982), Estimation of distribution density belonging to a class of entire functions, *Theory Probab. Appl.* **27**, 551–562.

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.

Scheffe, H. (1947), A useful convergence theorem for probability distributions, *Ann. Math. Statist.* **18**, 434–458.

Watson, G.S. and M.R. Leadbetter (1963), On the estimation of the probability density, *Ann. Math. Statist.* **34**, 480–491.