A NOTE ON THE USEFULNESS OF SUPERKERNELS IN DENSITY ESTIMATION

Luc Devroye
School of Computer Science
McGill University

ABSTRACT. We consider the Akaike-Parzen-Rosenblatt density estimate f_{nh} based upon any superkernel L (i.e. an absolutely integrable function with $\int L = 1$, whose characteristic function is 1 on $[-1, 1]$), and compare it with a kernel estimate g_{nh} based upon an arbitrary kernel K. We show that for a given subclass of analytic densities,

$$\inf_L \sup_K \limsup_{n\to\infty} \frac{\inf_h \mathbf{E} \int |f_{nh} - f|}{\inf_h \mathbf{E} \int |g_{nh} - f|} = 1 \ ,$$

where $h > 0$ is the smoothing factor. Thus, asymptotically, the class of superkernels is as good as any other class of kernels when certain analytic densities are estimated. We also obtain exact asymptotic expressions for the expected L_1 error of the kernel estimate when superkernels are used.

KEYWORDS AND PHRASES. Density function, nonparametric estimation, rate of convergence, kernel estimate, consistency, smoothing methods.

1991 MATHEMATICS SUBJECT CLASSIFICATIONS: Primary 62G05.

ACKNOWLEDGMENT. The comments of both referees and the Associate Editor are gratefully acknowledged.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

1. Introduction.

In this paper, 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, K, the *kernel*, is an absolutely integrable function with $\int K = 1$, and $X_1, \ldots, X_n$ are i.i.d. random variables with common density f on the real line (Akaike, 1954; Rosenblatt, 1956; Parzen, 1962). Sometimes we will write f_{nh} to make the dependence upon h explicit. The expected L_1 error $\mathbf{E}\int |f_n - f|$ is a function of n,f,h and K. Of these factors, the user can only choose K and h. The choices of h and K have led to extensive discussions, especially data-dependent choices for h for fixed K. Where the choice of K is concerned, we basically have little to go by when a small expected L_1 error is desired. What is known is that for class s kernels (i.e., symmetric kernels K for which $\int K = 1$, $\int x^i K(x)\,dx = 0$ for $1 \le i < s$, $\int |x|^s |K(x)|\,dx < \infty$, and $\int x^s K(x)\,dx \neq 0$) with even positive s,

$$\inf_f \liminf_{n\to\infty} n^{s/(2s+1)} \inf_h \mathbf{E}\int |f_{nh} - f| \ge c(K) > 0$$

where $c(K)$ depends upon K only (Devroye, 1988). This implies that our choice of K implicitly limits the performance to about $n^{-s/(2s+1)}$, regardless of how smooth f is. If we don't want implicit limitations, we should really consider kernels with $\int x^i K(x)\,dx = 0$ for all $i > 0$. These too create implicit limitations, but now they are of the order of $g(n)/\sqrt{n}$, where $g(n)$ is a slowly varying function. An important subclass is the class of all superkernels L, which are defined as symmetric absolutely integrable functions with $\int L = 1$, and whose characteristic function is 1 on an open neighborhood of the origin (see e.g. Devroye and Györfi (1985) or Devroye (1988)). Superkernels do not produce any performance limits, as we will see below. In L_2 studies, such kernels have been known for a long time; see e.g. Watson and Leadbetter (1963), Konakov (1972), Davis (1975, 1977) and Ibragimov and Khasminskii (1982). Watson and Leadbetter obtained lower bounds for the expected square error as a function of n and the density f only, uniformly over all kernels. Davis showed that for a large class of densities, the Watson-Leadbetter lower bounds can be attained in the limit up to a multiplicative constant by using the kernel $S(x) = \sin(x)/\pi x$. Thus, for L_2 performance, S is virtually asymptotically optimal.

Our main result is Theorem 5, which is stated in rough form in the abstract. Theorem 5 implies that if we had an infallible oracle that would give us for every n, K and f the best possible h, then we should look no further than superkernels in our search for a kernel that is asymptotically optimal. There are however three caveats. First of all, the optimality is only for all densities that are smooth enough; note that the characteristic function φ for f has to decrease basically at an exponential power rate. All those densities are necessarily analytic. In particular, φ has to be absolutely integrable, and hence f is at the very least absolutely continuous. Secondly, the present analysis does not provide us with a data-based method for selecting nearly optimal h for particular superkernels. Thirdly, choosing a superkernel for small sample sizes may be the wrong thing to do; somehow, we need more information about when the asymptotics really start to count.

2. Smooth densities.

Consider extremely smooth densities with exponentially decaying characteristic functions. These densities are necessarily analytic. Furthermore, it is possible to choose a kernel K such that the expected L_1 error is $O(\sqrt{\log n/n})$ (Devroye, 1987). In fact, with superkernels the best possible rate of convergence is solely determined by the density; unlike with class s kernels, the performance is thus no longer limited by the kernel.

In section 3, we derive simple asymptotic upper bounds for the expected L_1 error. These allow us to obtain information about the best possible choice of h. With an additional order restriction on the characteristic function, we show that the bounds are in fact asymptotically of the right magnitude. The optimal h is at least of the order $1/\log n$, but despite its huge size, the optimal performance of the kernel estimate is rather sensitive to its choice. This is best seen as follows. The variation term in the error is of the order of $1/\sqrt{nh}$, while the bias term is of the order of $\exp(-1/h)$ or smaller. Naively summing the two terms gives us an upper bound for the expected L_1 error. Interestingly, if we put $h = 2/\log n$, then this sum is $O(\sqrt{\log n/n}) + O(1/\sqrt{n})$, and in fact, Bias $= o(\text{Variation})$. However, if we put $h = (2+\epsilon)/\log n$ for arbitrary small fixed $\epsilon > 0$, then the sum is $O(n^{-1/(2+\epsilon)})$, with Variation $= o(\text{Bias})$. If anything, it is better to underestimate h than to overestimate it.

When studying the bias of the error, it becomes somewhat difficult to use the Taylor series expansion technique, while inequalities linking distances between functions with distances between their Fourier transforms seem to be of some help. However, something is lost in this transition, and the bounds may be loose in some circumstances. Nevertheless, in section 4, for smooth densities, we will be able to obtain exact expressions for

$$\inf_h \mathbf{E} \int |f_{nh} - f|$$

when a superkernel is used.

In section 6, we look at other kernels, and show that for the same subclass of densities alluded to above, any kernel leads to an expected L_1 error that is asymptotically not better than that for the superkernels studied here. Thus, barring small sample situations, it is probably not interesting to look any further for better kernels.

3. Upper bounds for the performance.

In what follows, we say that a function g is in L_1 if it is absolutely integrable. It is in M when there exists a $p > 1$ such that $\int (1 + |x|^p)|g(x)|\, dx < \infty$. It is easy to verify that $M \subseteq L_1$, and that for $g \in M$, $\int \sqrt{|g|} < \infty$. Also, for $f, L \in M$, and $f, L \geq 0$, we have

$$\lim_{h \to 0} \int \sqrt{f * L_h} = \int \sqrt{f} \sqrt{\int L}$$

where from here on $*$ denotes the convolution operator, and $f * L_h$ denotes $f * (L_h)$ (see exercise 7.8 on p.130 of Devroye, 1987). Also, the left-hand-side integral is not less than the right-hand-side (Devroye, 1987, p.115).

We define a *superkernel* L by the condition that it is a kernel (hence, $\int |L| < \infty$ and $\int L = 1$) with absolutely integrable Fourier transform ψ (hence, L is bounded), and with the properties that $\psi(t) \equiv 1$ on $[-1, 1]$ and $|\psi| < 1$ off $[-1, 1]$. Note that the interval $[-1, 1]$ is chosen for convenience only. Superkernels necessarily have infinite support.

Often, our kernels require an additional regularity condition, i.e they should be strong approximate identities: a function $K \in L_1$ is a strong approximate identity if for all $f \in L_1$, $\lim_{h \to 0} K_h * f = f \int K$ at almost all x. By an extension of the Lebesgue density theorem (Stein, 1970, pp.62–63) it suffices that $|K|$ is bounded from above by a unimodal integrable function. Virtually all candidate kernels are thus strong approximate identities.

EXAMPLE 1. SUPERKERNELS. Define the Fourier transform $\psi_r(t) = I_{[-r,r]}(t)$, which has as inverse the function $rS(rx)$, where $r > 0$ is a parameter. For $b > 0$, the Fourier transform defined by $\psi \stackrel{\text{def}}{=} \psi_{1+2b} * \psi_b * \psi_b$ has as inverse $b^2(1 + 2b)S^2(bx)S((1 + 2b)x)$. This inverse is 1 on $[-1, 1]$, and is $O(x^{-3})$ as $x \to \infty$. Thus, we have a superkernel. Also, it is in M. Inverses that decrease at an exponential power rate can be obtained by convolving ψ_{1+2b} with extremely smooth transforms. $\square$

To bound the L_1 error, we compute bounds for the bias and the variation terms separately in Lemmata 1 and 2.

LEMMA 1. *Define $\Phi(u) = \int_{|t| \geq u} |\varphi(t)|dt$, and let $\alpha \in (0, 1]$ be fixed such that $\int f^\alpha < \infty$. If L is a superkernel, then*

$$\int |f * L_h - f| \leq 2 \int \min\left(f(x), \frac{1}{\pi}, \Phi(1/h)\right)\, dx \leq 2 \left(\frac{1}{\pi}\Phi(1/h)\right)^{1-\alpha} \int f^\alpha \,.$$

PROOF.

$$\int |f * L_h - f| = 2\int (f * L_h - f)_+ \le 2\int \min\left(f(x), \sup_y |f(y) - f * L_h(y)|\right)\,dx$$
$$\le 2\int \min\left(f(x), \frac{1}{2\pi}\int |\varphi(t)||\psi(th) - 1|\right)\,dx$$
$$\le 2\int \min\left(f(x), \frac{2}{2\pi}\Phi(1/h)\right)\,dx\ .$$

Combine this with the fact that for any positive constant a,

$$\int \min(f(x), a)\,dx \le a^{1-\alpha}\int f^\alpha\ . \square$$

LEMMA 2. *Let $f \in M$ be a given density, and let $L \in M$ be a bounded kernel. Then, if f_n is the kernel estimate with kernel L and smoothing parameter $h \to 0$ as $n \to \infty$, we have*

$$\sqrt{nh}\,\mathbf{E}\int |f_n - f * L_h| \le \int \sqrt{f * (L^2)_h} \le \int \sqrt{f}\sqrt{\int L^2}.$$

If also $nh \to \infty$ and both L and L^2 are strong approximate identities, then

$$\lim_{n\to\infty} \sqrt{nh}\,\mathbf{E}\int |f_n - f * L_h| = \sqrt{\frac{2}{\pi}}\int \sqrt{f}\sqrt{\int L^2}.$$

PROOF. By boundedness, $L^2 \in M$ when $L \in M$. Thus, a fact stated at the top of this section proves the first part of the Lemma. The second part begins with the pointwise estimate

$$\sqrt{nh}\,\mathbf{E}|f_n - f * L_h| \le \sqrt{f * (L^2)_h}$$

and the upper bound tends to $\sqrt{f}\sqrt{\int L^2}$ at almost all x if L^2 is a strong approximate identity. Since we also have integral convergence, we are in a position to apply an extended version of the Lebesgue dominated convergence theorem. We can conclude Lemma 2 if we can show that for almost all x,

$$\sqrt{nh}\,\mathbf{E}|f_n - f * L_h| \to \sqrt{2/\pi}\sqrt{f * (L^2)_h}.$$

For this, we use a technique borrowed from Devroye and Györfi (1985, p.90-93). Define $\sigma_n^2(x) = \sigma_n^2 = \mathbf{E}(f_n - f * L_h)^2$, let $c > 0$ be a universal positive constant, and let L^* be an upper bound on the size of L. Then by Lemmata 8 and 9 on pages 90-91 of Devroye and Györfi (1985), applied with $a = 0$, we note that

$$\left|\mathbf{E}|f_n - f * L_h| - \sqrt{\frac{2}{\pi}}\sigma_n\right| \le \frac{cL^*}{nh}.$$

Next, at almost all x, since L and L^2 are strong approximate identities,

$$\begin{aligned}\sigma_n^2 &= \frac{1}{n}(f*(L_h)^2 - (f*L_h)^2) \\ &= \frac{1}{nh}\left(f\int L^2 + o(1) - h(f+o(1))^2\right) \sim \frac{f\int L^2}{nh}.\end{aligned}$$

Since $nh \to \infty$, we can conclude that for almost all x,

$$\mathbf{E}|f_n - f*L_h| \sim \sqrt{\frac{2}{\pi}}\frac{\sqrt{f\int L^2}}{\sqrt{nh}}. \square$$

Remark 1. Lemma 2 is applicable to superkernels since they are necessarily bounded (recall that $\int|\psi| < \infty$). $\square$

Armed with the two lemmata given above, we can now give the first result, an upper bound on the asymptotic performance with superkernels.

Theorem 1. *Let f_n be a kernel estimate with superkernel L. Assume furthermore that L and L^2 are strong approximate identities and that $f, L \in M$. Define the (possibly infinite) constant*

$$T \stackrel{\text{def}}{=} \sup\{|t| : |\varphi(t)| \neq 0\} \ .$$

If $T < \infty$, then

$$\inf_h \mathbf{E}\int|f_{nh} - f| \leq \frac{1}{\sqrt{n}}\int\sqrt{Tf*(L^2)_{1/T}} \ .$$

If $T = \infty$ (hence $\Phi(u) > 0$ for all $u > 0$), $\int f^\alpha < \infty$ for some $0 < \alpha < 1$, φ is absolutely integrable (hence $\Phi(u) < \infty$ for all $u > 0$), and $1/\Phi$ is rapidly varying at infinity, i.e. $\Phi(cu)/\Phi(u) \to 0$ as $u \to \infty$ for all $c > 1$, then

$$\inf_h \mathbf{E}\int|f_{nh} - f| \leq (1+o(1))\frac{\sqrt{2/\pi}\int\sqrt{f}\sqrt{\int L^2}}{\sqrt{nh_{\alpha,n}}} \ ,$$

where $h_{\alpha,n}$ is the unique solution of the equation

$$\frac{\sqrt{2/\pi}\int\sqrt{f}\sqrt{\int L^2}}{\sqrt{nh}} = 2\int f^\alpha\left(\frac{1}{\pi}\Phi(1/h)\right)^{1-\alpha} \ .$$

Without the rapid variation condition, we have

$$\inf_h \mathbf{E}\int|f_{nh} - f| \leq (2\sqrt{2/\pi} + o(1))\frac{\int\sqrt{f}\sqrt{\int L^2}}{\sqrt{nh_{\alpha,n}}} \ .$$

PROOF. We begin with the second part. Let $h^* = (1-\epsilon)h_{\alpha,n}$, where $\epsilon > 0$ is arbitrary. Then, by Lemma 2,

$$\mathbf{E}\int |f_{nh^*} - f * L_{h^*}| \le \frac{\sqrt{2/\pi}\int\sqrt{f}\sqrt{\int L^2} + o(1)}{\sqrt{nh_{\alpha,n}(1-\epsilon)}}.$$

By Lemma 1,

$$\begin{aligned}\int |f - f * L_{h^*}| &\le 2\int f^\alpha \left(\frac{1}{\pi}\Phi(1/h^*)\right)^{1-\alpha} \\ &= o\left((\Phi(1/(h_{\alpha,n})))^{1-\alpha}\right) \quad \text{(rapid variation of } 1/\Phi) \\ &= o\left(1/\sqrt{nh_{\alpha,n}}\right) \quad \text{(definition of } h_{\alpha,n}).\end{aligned}$$

The proof of the second part of Theorem 1 now follows from the triangle inequality applied to $\mathbf{E}\int |f_{nh^*} - f|$, and the arbitrary nature of ϵ. Without the rapid variation condition, we note that $\Phi(1/h^*) \le \Phi(1/h_{\alpha,n})$, and note that the upper bounds for the bias and variation are asymptotically of the same size. For the first part, we note that $\int |f - f * L_{1/T}| = 0$ since $\varphi(t) \equiv \varphi(t)\psi(t/T)$ for all t. By Lemma 2, $\inf_h \mathbf{E}\int |f_{nh} - f| \le \sqrt{f * (L^2)_{1/T}}/\sqrt{n/T}$. $\square$

REMARK 2. It is a simple exercise to prove that $h_{\alpha,n}$ is uniquely defined, and that $h_{\alpha,n} \to 0$ as $n \to \infty$ and $nh_{\alpha,n} \to \infty$ as $n \to \infty$. $\square$

REMARK 3. The heaviness of the tail of f is represented by the factor $\int f^\alpha$ in the bound. Assume for example that $\int (1+|x|^p)f(x)\,dx < \infty$. Then

$$\begin{aligned}\int f^\alpha &= \int (1+|x|^p)^\alpha f^\alpha \times (1+|x|^p)^{-\alpha}\,dx \\ &\le \left(\int (1+|x|^p)f\right)^\alpha \times \left(\int (1+|x|^p)^{-\alpha/(1-\alpha)}\right)^{1-\alpha} \quad \text{(Hölder's inequality)} \\ &< \infty\end{aligned}$$

when $\alpha > 1/(1+p)$. In particular, if $f \in M$, then $\int \sqrt{f} < \infty$. $\square$

REMARK 4. THE STABLE DENSITIES. Consider the stable densities with characteristic function $\varphi(t) = \exp(-|t|^a)$ for $a \in (0,2]$. We note that for $a > 1$ these densities are in M, and the rapid variation condition of Theorem 1 is satisfied. Furthermore, $\Phi(u) \sim 2/(au^{a-1}e^{-u^a})$ as $u \to \infty$. We note that for $a \ne 2$, the tails of $f(x)$ drop off as $|x|^{-(1+a)}$. Thus, $\int f^\alpha < \infty$ when $\alpha > 1/(1+a)$. For such a choice of α, simple computations show that $h_{\alpha,n} \sim (2(1-\alpha)/\log n)^{1/a}$, and $\inf_h \mathbf{E}\int |f_{nh} - f| = O(\sqrt{\log^{1/a} n\,/n})$. These rates are much better than those usually obtained in kernel density estimation with "standard" kernels. $\square$

REMARK 5. The rapid variation condition does not imply however that Φ has to decrease at an exponential rate or faster. Consider for example

$$\Phi(u) \sim \exp(-\log^p u)$$

as $u \to \infty$, where $p > 1$ is a constant. Verify that $h_{1/2,n} = \exp(-(1-(1/p)+o(1))^{1/p}\log^{1/p} n)$ for $p > 1$. Also, the conditions of Theorem 1 are satisfied. Hence, $\inf_h \mathbf{E} \int |f_{nh} - f| = O(\sqrt{\exp(\log^{1/p} n)/n})$. This shows that Theorem 1 gives a continuum of rates from of the order of $1/\sqrt{n}$ times a function which increases at any rate slower than polynomial. $\square$

REMARK 6. ANALYTIC DENSITIES. Analytic densities are characterized by characteristic functions that are $O(\exp(-c|t|))$ as $|t| \to \infty$ for some constant c (Kawata, 1972). Theorem 1 implies that if in addition the rapid variation condition is satisfied, then the estimate of Theorem 1 holds, and that in any case (i.e without the condition), we have $\mathbf{E} \int |f_{nh} - f| = O(\sqrt{\log n/n})$. The rapid variation condition only allows us to explicitly pin down a good constant in the upper bound, but it is by no means necessary for calculating rates. For other estimates with analytic densities, we refer to Koronacki (1987), Gajek (1987) or Devroye (1987,p.132; 1988). Minimax results for analytic densities can be found in Ibragimov and Khasminskii (1982). $\square$

REMARK 7. The upper bound shows that it is of interest to choose a superkernel with small value of $\int L^2 = (1/(2\pi)) \int \psi^2$. By our scale restriction on ψ, we always have $\int L^2 \geq \frac{1}{\pi}$. Thus, it is to our advantage to pick ψ as close as possible to the rectangular function on $[-1,1]$ with height one. However, small smooth tails are required so as to force L to be in M. $\square$

REMARK 8. The values $h_{\alpha,n}$ represent nearly optimal choices of h for a given density f. Obviously, since α can vary, they differ among each other; yet, as we will see, the variation in the values of $h_{\alpha,n}$ is modest with respect to α. The values are greatly affected by the smoothness of f; from the definition, it is apparent that $h_{\alpha,n}$ is related to the inverse of Φ, which in turn is a measure of the smoothness of the density. See also the example of the stable family mentioned above. For $\alpha = 1/2$, the value of $h_{\alpha,n}$ is solely determined by Φ, and does not depend upon any tail factor such as $\int \sqrt{f}$ or $\int f^\alpha$. $\square$

4. Rates of convergence with superkernels.

In Theorem 1, we obtained upper bounds for the expected L_1 error when a superkernel is used. They can be employed whenever $\int \sqrt{f} < \infty$ and $\int |\varphi| < \infty$. What happens when one of these conditions fails to hold would lead us astray; after all, the benefits of superkernels are only seen when f is smooth. For an important subclass of densities, we can in fact get exact asymptotics that show that the upper bounds of Theorem 1 are tight.

We recall the definitions of T and $h_{\alpha,n}$ from Theorem 1. We need yet another parameter related to the distribution,

$$\xi \stackrel{\text{def}}{=} \inf_u \frac{\sup_{|t|\geq u} |\varphi(t)|}{\int_{|t|\geq u} |\varphi(t)| dt} ,$$

This number will be required to be positive. The results of this section are summarized in the following Theorems:

THEOREM 2. *Let $L \in M$ be a superkernel, and let f_{nh} be the kernel estimate with kernel L and smoothing parameter h. We assume that $f \in M$. If $T < \infty$, then*

$$\inf_h \mathbf{E} \int |f_{nh} - f| \sim \sqrt{\frac{2}{\pi n}} \inf_{h \leq 1/T} \int \sqrt{(L_h)^2 * f - (L_h * f)^2} .$$

THEOREM 3. *Let $L \in M$ be a superkernel, let L and L^2 be strong approximate identities, let $\xi > 0$, let f_{nh} be the kernel estimate with kernel L and smoothing parameter h, and assume that $f \in M$. If $T = \infty$, and $\log(1/\Phi)$ is rapidly varying at ∞, then*

$$\inf_h \mathbf{E} \int |f_{nh} - f| \sim \sqrt{\frac{2}{\pi}} \frac{\sqrt{\int L^2} \int \sqrt{f}}{\sqrt{n h_{1/2,n}}} .$$

THEOREM 4. *Let $L \in M$ be a superkernel, let L and L^2 be strong approximate identities, let $\xi > 0$, let f_{nh} be the kernel estimate with kernel L and smoothing parameter h, and assume that $f \in M$. If $T = \infty$, $\int f^\alpha < \infty$ for <u>fixed</u> $\alpha \in (0, 1/2]$, and $\log(1/\Phi)$ is regularly varying at ∞ of order $\theta > 0$, then*

$$(1-\alpha)^{1/(2\theta)} + o(1) \leq \frac{\sqrt{\pi n h_{\alpha,n}/2}}{\sqrt{\int L^2} \int \sqrt{f}} \inf_h \mathbf{E} \int |f_{nh} - f| \leq 1 + o(1) .$$

If $T = \infty$, $\int f^\alpha < \infty$ for <u>all</u> $\alpha \in (0, 1/2]$, and $\log(1/\Phi)$ is regularly varying at ∞ of order $\theta > 0$, then

$$\inf_h \mathbf{E} \int |f_{nh} - f| \sim \sqrt{\frac{2}{\pi}} \frac{\int \sqrt{f} \sqrt{\int L^2}}{\sqrt{n h_{0,n}}} ,$$

where $h_{0,n}$ is the unique solution of the equation

$$\frac{1}{\sqrt{nh}} = \Phi(1/h) .$$

Remark 9. The conditions of rapid and regular variation of Φ can be replaced by weaker conditions that are, however, somewhat more difficult to formulate.

Remark 10. The relative difficulty with Theorems 2–4 is due to the fact that the bias in the error has to be sandwiched between matching expressions that involve characteristic functions. However, in L_1, we do not have some sort of equivalent of Bessel's inequality. This forces us to opt for additional conditions on f related to the regular or rapid variation of $1/\Phi$ and/or $\log(1/\Phi)$. Another approach, based on an exponential tail condition for φ'', is outlined in Devroye (1987, p. 132).

Remark 11. If $\log(1/\Phi)$ is regularly varying at ∞ of order $\theta > 0$, then by well-known representation theorems (see e.g. Seneta, 1976), we find that

$$\Phi(t) = e^{-t^{\theta} \exp(o(\log t))+O(1)} .$$

This implies that $1/\Phi$ is rapidly varying at ∞.

Remark 12. The ξ condition virtually implies that $|\varphi|$ should decrease exponentially quickly in the tails. Take for example a monotone $|\varphi|$. Since Φ is absolutely continuous, we note that $|\Phi'(u)|/\Phi(u) \geq \xi$. This implies that $\Phi(u) \leq \Phi(0)e^{-\xi u}$. This, and the monotonicity of $|\varphi|$, implies that $|\varphi(t)| = O(e^{-\xi t/2})$ as $t \to \infty$, and hence that our density f is analytic. These densities are so smooth that they cannot have compact support. In fact (Kawata, 1972, p. 288 and p. 435) $\limsup_{t\to\infty}|\varphi(t)| \exp(|t|/\log|t|) = \infty$ for all compact support densities.

Remark 13. For the stable distribution ($\varphi(t) = \exp(-|t|^a)$) we have $\theta = a$ in Theorem 4, while $\xi = 0$ for $0 < a < 1$, $\xi = 1/2$ for $a = 1$, and $\xi = \infty$ for $a > 1$.

Remark 14. The most interesting special case of Theorem 4 occurs when $\int f^{\alpha} < \infty$ for all $\alpha \in (0, 1)$. This is for example satisfied when there exist positive a, b, c such that $f(x) \leq a \exp(-b|x|^c)$. Note however that we can't in this case draw the same conclusion as in Theorem 3 since $h_{\alpha,n}$ varies with α.

REMARK 15. SMALL-TAILED CHARACTERISTIC FUNCTIONS. Consider the characteristic function $(1-|t|)_+$, which corresponds to the Fejèr-de la Vallée-Poussin density

$$f(x) = (2\pi)^{-1} \left(\frac{\sin(x/2)}{x/2} \right)^2 .$$

Let Z be a random variable with this density, and let Y be an independent random variable with distribution function F on $(0,\infty)$. Then the characteristic function of Z/Y is given by $\varphi(t) = \mathbf{E}(1-|\frac{t}{Y}|)_+$. Thus, φ inherits the tail of F in the t-domain!. To see this, note that

$$|\varphi(|t|)| \le 1 - F(t).$$

For example, with $F(x) = 1-\exp(1-e^x)$, we obtain a characteristic function whose tail is asymptotically not greater than $\exp(1-e^{|t|})$. With a superkernel, for this example, we can obtain rates of convergence that are $O(\sqrt{\log\log n/n})$. $\square$

REMARK 16. THE ULTIMATE KERNEL. If one shape of kernel is to be used for all n, then the best kernel depends upon the density f. For densities with compact support characteristic function, the kernel L should be chosen such that $\inf_{h\le 1/T} \int \sqrt{(L_h)^2 * f - (L_h * f)^2}$ is minimal. The best pair (L,f) is that for which this minimum is minimal. We recall from Devroye (1988) that $\inf_h \mathbf{E}\int |f_{nh} - f| \ge (1+o(1))/(8\sqrt{n})$ for any (L,f) pair however. $\square$

REMARK 17. THE NORMAL DENSITY. The last part of theorem 4 is applicable to the normal density. To determine $h_{0,n}$, we note first that $\Phi(1/h) \sim 2h\exp(-1/(2h^2))$ as $h \downarrow 0$. It is easy to verify then that $h_{0,n} \sim 1/\sqrt{\log n}$. We conclude that

$$\inf_h \mathbf{E} \int |f_{nh} - f| \sim \sqrt{\frac{2\sqrt{\log n}}{\pi n}} \int \sqrt{f} \sqrt{\int L^2} .$$

5. Proofs of Theorems 2–4.

PROOF OF THEOREM 2. We begin with the upper bound. We write Z_i for $L_h(x-X_i) - \mathbf{E}L_h(x-X_i)$, so that $\sqrt{n}|f_{nh} - f * L_h| = |\sum_{i=1}^n Z_i/\sqrt{n}| \stackrel{\text{def}}{=} Y_n$. Define $\sigma_h^2(x) = \text{Var}\{Z_1\}$. In other words, $\sigma_h^2(x) = (L_h)^2 * f - (L_h * f)^2$. By the central limit theorem, for fixed h, Y_n tends in distribution to $\sigma_h(x)|N|$, where N is a standard normal random variate. Following Devroye and Györfi (1985, p.90), we see that there is a universal constant C such that

$$\begin{aligned}\left|\mathbf{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\right| - \sqrt{\frac{2}{\pi}}\sigma_h(x)\right| &\le C\frac{E|Z_1|^3}{\sigma_h^3(x)\sqrt{n}} \\ &\le \frac{C \operatorname{ess\,sup}|Z_1|}{\sigma_h(x)\sqrt{n}} \\ &\le \frac{C\|L\|_\infty}{h\sigma_h(x)\sqrt{n}} \ .\end{aligned}$$

Also,

$$\mathbf{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\right| \le \sqrt{\text{Var}\{Z_1\}} = \sigma_h(x) \ .$$

For any $h \in (0, 1/T]$, $\int |f - f * L_h| = 0$ because the characteristic functions of f and $f * L_h$ coincide ($\varphi(t) \equiv \varphi(t)\psi(th)$ for $|t| \le T$ since $h \le 1/T$). Thus, for constant $h \le 1/T$,

$$\begin{aligned}\mathbf{E}\left\{\sqrt{n}\int |f_{nh} - f|\right\} &\le \int \sqrt{2/\pi}\sigma_h(x)\,dx + \int \min\left(\sigma_h(x), \frac{C\|L\|_\infty}{h\sigma_h(x)\sqrt{n}}\right)\,dx \\ &= \int \sqrt{2/\pi}\sigma_h(x)\,dx + o(1) \ ,\end{aligned}$$

where we used the Lebesgue dominated convergence theorem, the fact that h is constant and that $\int \sigma_h(x)\,dx < \infty$. The last fact requires the condition $f \in M$. Since a fixed h is certainly not better than the optimal h, we see that

$$\limsup_{n\to\infty} \sqrt{n} \inf_h \mathbf{E}\int |f_{nh} - f| \le \inf_{0<h\le 1/T} \int \sqrt{2/\pi}\sigma_h(x)\,dx \ .$$

Let us now turn to the lower bound. The proof takes 9 steps.

Step 1. Define h^* as a given sequence satisfying

$$\inf_h \mathbf{E}\int |f_{nh} - f| \sim \mathbf{E}\int |f_{nh^*} - f| \ .$$

Step 2. If $h \equiv c$, where $c \in (0, 1/T]$, then

$$\limsup_{n\to\infty} \sqrt{n}\mathbf{E}\int |f_{nh} - f| \le \int \sqrt{2/\pi}\sigma_c(x)\,dx < \infty \ .$$

This was shown in the first part of the proof. We will use this later to exclude the possibility that h^* has a subsequence that converges to $c \notin (0, 1/T]$, or that diverges.

Step 3. If $h^* \to \infty$ or $h^* \to 0$ along a subsequence, then along this subsequence,

$$\sqrt{n}\mathbf{E}\int |f_{nh^*} - f| \to \infty \ .$$

Consider first $h^* \to \infty$. We have

$$\liminf_{n\to\infty} \mathbf{E} \int |f_{nh^*} - f| \geq \int |f - f * L_{h^*}| \geq \liminf_{n\to\infty} \sup_t |\varphi(t)| = 1 \ .$$

Consider next $h^* \to 0$. Then $\sqrt{n}\mathbf{E} \int |f_{nh} - f| \to \infty$ by Theorem 16 on p.136 of Devroye and Györfi (1985).

Step 4. If $h^* \to c > 1/T$ along a subsequence, then along this subsequence,

$$\sqrt{n}\mathbf{E} \int |f_{nh^*} - f| \to \infty \ .$$

To see this, note that

$$\mathbf{E} \int |f_{nh^*} - f| \geq \int |f - f * L_{h^*}| \geq \sup_t |\varphi(t)||\psi(th^*) - 1| \to \sup_t |\varphi(t)||\psi(tc) - 1| > 0$$

by the uniform continuity of ψ. Note that if the lower bound were zero, then we would have $\psi(t) \equiv 1$ for all $|t| \leq cT$, which contradicts the definition of a superkernel.

Step 5. Steps 2 through 4 together imply that

$$0 < \liminf_{n\to\infty} h^* \leq \limsup_{n\to\infty} h^* \leq 1/T \ .$$

Step 6. If $h \to c > 0$, then

$$\sqrt{n}\mathbf{E} \int |f_{nh} - f| \sim \sqrt{n}\mathbf{E} \int |f_{nc} - f| \geq \frac{1 + o(1)}{8} \ .$$

The last inequality follows from Remark 16. The asymptotic equivalence can be obtained by introducing the function $g(x) = L_h(x) - L_c(x)$ and observing that

$$\begin{aligned}
\left|\mathbf{E} \int |f_{nh} - f| - \mathbf{E} \int |f_{nc} - f|\right| &\leq \mathbf{E} \int |f_{nh} - f_{nc}| \\
&\leq n^{-1/2} \int \sqrt{f * (g^2)} \\
&\leq n^{-1/2} \int \sqrt{f} \sqrt{\int g^2} \\
&= o(1/\sqrt{n}) \ ,
\end{aligned}$$

where we used a bound similar to Lemma 2, and the fact that $\int g^2 \to 0$ as $h \to c$.

Step 7. Consider a subsequence of h^* converging to $c \in (0, 1/T]$. Then

$$\liminf_{n\to\infty} \sqrt{n}\mathbf{E} \int |f_{nh^*} - f| \geq \liminf_{n\to\infty} \sqrt{n}\mathbf{E} \int |f_{nc} - f| \ .$$

Step 8. Assume that

$$\sqrt{n}\mathbf{E} \int |f_{nh^*} - f| \to \alpha < \inf_{0<c\leq 1/T} \liminf_{n\to\infty} \sqrt{n}\mathbf{E} \int |f_{nc} - f|$$

along a given subsequence. Since $\liminf h^* > 0$ and $\limsup h^* \leq 1/T$ along this subsequence, the subsequence must have a further subsequence for which $h^* \to c \in (0, 1/T]$, and along this subsequence, the lower bound of Step 7 applies. This leads to a contradiction, and we must conclude that

$$\liminf_{n\to\infty} \sqrt{n} \inf_h \mathbf{E} \int |f_{nh} - f| = \liminf_{n\to\infty} \sqrt{n}\mathbf{E} \int |f_{nh^*} - f| \geq \inf_{0<c\leq 1/T} \liminf_{n\to\infty} \sqrt{n}\mathbf{E} \int |f_{nc} - f| \ .$$

Step 9. The lower bound can be made more precise as follows. Fix $c \in (0, 1/T]$. We note that $\liminf_{n\to\infty} \mathbf{E}|Y_n| \geq \sigma_c(x)\mathbf{E}|N|$. Thus, by Fatou's lemma,

$$\liminf_{n\to\infty} \sqrt{n}\mathbf{E} \int |f_{nc} - f| = \liminf_{n\to\infty} \sqrt{n} \int \mathbf{E}|f_{nh} - f * L_h| \geq \mathbf{E}|N| \int \sigma_c(x)\, dx \ .$$

Combining this with the bound of Step 8 shows that

$$\liminf_{n\to\infty} \sqrt{n} \inf_h \mathbf{E} \int |f_{nh} - f| \geq \inf_{0<c\leq 1/T} \sqrt{2/\pi} \int \sigma_c(x)\, dx \ .$$

This concludes the proof of Theorem 2. □

Let us consider some general properties of $h_{\alpha,n}$. To do so, for $a > 0$, $0 < b \leq 1$, we define the quantities $R_{a,b}(n)$ as follows:

$$R_{a,b}(n) \stackrel{\text{def}}{=} \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a\Phi^b(1/h)\right) \ .$$

The relationship with $h_{\alpha,n}$ is obvious from its definition in Theorem 1.

LEMMA 3. *Assume that φ is absolutely integrable. If $\log(1/\Phi)$ is rapidly varying at ∞, then, for all $a, a' > 0$, and all $0 < b \leq b' \leq 1$,*

$$\lim_{n\to\infty} \frac{R_{a,b}(n)}{R_{a',b'}(n)} = 1 \ .$$

If $\log \Phi$ is regularly varying of order $\theta > 0$ at ∞, then

$$1 \leq \liminf_{n\to\infty} \frac{R_{a,b}(n)}{R_{a',b'}(n)} \leq \limsup_{n\to\infty} \frac{R_{a,b}(n)}{R_{a',b'}(n)} \leq \left(\frac{b'}{b}\right)^{1/(2\theta)} \ .$$

PROOF. The minimizing h in the definition of $R_{a,b}(n)$ is such that $h \to 0$ as $n \to \infty$. The function $1/\Phi$ is rapidly varying at ∞ (remark 11). Let $\epsilon > 0$ and $\delta > 0$ be arbitrary small positive numbers. We have

$$
\begin{aligned}
R_{a,b}(n) &= \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a\Phi^b(1/h)\right) \\
&= \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a\Phi^b(0)(\Phi(1/h)/\Phi(0))^b\right) \\
&\geq \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a\Phi^b(0)(\Phi(1/h)/\Phi(0))^{b'}\right) \\
&= \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a\Phi^{b-b'}(0)\Phi^{b'}(1/h)\right) \\
&\geq \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a'\Phi^{b'}((1+\epsilon)/h)\right) \quad (n \text{ large enough}) \\
&\geq \frac{1}{\sqrt{1+\epsilon}} \inf_{h>0} \max\left(\frac{\sqrt{1+\epsilon}}{\sqrt{nh}}, a'\Phi^{b'}((1+\epsilon)/h)\right) \\
&= \frac{1}{\sqrt{1+\epsilon}} R_{a',b'}(n) \ .
\end{aligned}
$$

In this inequality, we only used the rapid variation and monotonicity of $1/\Phi$. For the second half of the proof, we assume the regular variation of $\log(1/\Phi)$—the argument for the rapid variation case follows trivially from this, and will be omitted. Also, for $\epsilon > 0$ as above, $c \overset{\text{def}}{=} (b'/(b(1-\epsilon)))^{1/\theta}$. We have

$$
\begin{aligned}
R_{a,b}(n) &= \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, ae^{b \log \Phi(1/h)}\right) \\
&\leq \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, ae^{bc^{\theta}(1-\epsilon) \log \Phi(1/ch)}\right) \quad (\text{for all } n \text{ large enough}) \\
&= \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a\Phi^{b'}(1/ch)\right) \quad (\text{defnition of } c) \\
&\leq \inf_{h>0} \max\left(\frac{1}{\sqrt{nh}}, a'\Phi^{b'}((1-\delta)/ch)\right) \quad (\text{for all } n \text{ large enough}) \\
&\leq \max\left(1, \sqrt{c/(1-\delta)}\right) \inf_{h>0} \max\left(\sqrt{\frac{1-\delta}{cnh}}, a'\Phi^{b'}((1-\delta)/ch)\right) \\
&= \max\left(1, \sqrt{c/(1-\delta)}\right) R_{a',b'}(n) \ .
\end{aligned}
$$

The coefficient in the upper bound is arbitrarily close to $(b'/b)^{1/(2\theta)}$ by our choice of δ and ϵ. □

LEMMA 4. *Let f_{nh} be the kernel estimate with bounded kernel L, and assume that L and L^2 are strong approximate identities. Then if the density f is arbitrary, $h \to 0$ and $nh \to \infty$ as $n \to \infty$,*

$$\liminf_{n\to\infty} \sqrt{nh} \int \mathbf{E}|f_{nh} - f| \geq \sqrt{\frac{2}{\pi}} \int \sqrt{f} \sqrt{\int L^2} \ .$$

PROOF. Assume first that $h \to 0$ and $nh \to \infty$. We introduce the notation $B_n = f * L_h - f$ and

$$\sigma_n^2 = \mathbf{E}\left(f_{nh} - f * L_h\right)^2 = \frac{1}{n}\left(f * (L_h)^2 - (f * L_h)^2\right) \ .$$

Both the bias B_n and the variance σ_n^2 depend upon x. At almost all x, we have $f * L_h \to f$ and $f * (L^2)_h \to f \int L^2$ since L and L^2 are strong approximate identities. For such x, we see that $\sigma_n^2 \sim f * (L^2)_h/(nh) \sim f \int L^2/(nh)$. By Lemma 9 on p. 91 of Devroye and Györfi (1985), there exists a universal Berry-Esseen type constant C such that

$$\Big|\mathbf{E}|f_{nh} - f| - \sigma_n \rho(|B_n|/\sigma_n)\Big| \leq \frac{C\|L\|_\infty}{nh} \ ,$$

where

$$\rho(u) \stackrel{\text{def}}{=} |u|\mathbf{P}\{|N| \leq |u|\} + \sqrt{\frac{2}{\pi}} e^{-u^2/2} = \mathbf{E}|N - u| \geq \sqrt{\frac{2}{\pi}} \ ,$$

and N is a standard normal random variable. At almost all x, we have

$$\liminf_{n\to\infty} \sqrt{nh}\, \mathbf{E}|f_{nh} - f| \geq \int \sqrt{f} \sqrt{\int L^2} \liminf_{n\to\infty} \rho(|B_n|/\sigma_n) \geq \sqrt{2/\pi} \int \sqrt{f} \sqrt{\int L^2} \ .$$

Thus, by Fatou's Lemma,

$$\liminf_{n\to\infty} \sqrt{nh}\, \mathbf{E} \int |f_{nh} - f| \geq \sqrt{2/\pi} \int \sqrt{f} \sqrt{\int L^2} \ . \ \square$$

PROOF OF THEOREM 3. From Theorem 1, we recall that, with $A \stackrel{\text{def}}{=} \sqrt{2/\pi} \int \sqrt{f} \sqrt{\int L^2}$, $a \stackrel{\text{def}}{=} 2 \int \sqrt{f}/(A\sqrt{\pi})$,

$$\begin{aligned} \inf_h \mathbf{E} \int |f_{nh} - f| &\leq (1 + o(1)) \inf_h \max\left(\frac{A}{\sqrt{nh}}, 2\int \sqrt{f}\sqrt{\Phi(1/h)/\pi}\right) \\ &= (A + o(1))\, R_{a,1/2}(n) \ , \end{aligned}$$

in the notation of Lemma 3. This result uses the fact that $1/\Phi$ is rapidly varying at ∞. If $h \to 0$ and $nh \to \infty$, then the lower bound of Lemma 4 is applicable:

$$\mathbf{E} \int |f_{nh} - f| \geq (A + o(1))/\sqrt{nh} \ .$$

Let $\eta > 0$ be so small that $\zeta \stackrel{\text{def}}{=} \inf_{|t|\geq 1+\eta} |\psi(t) - 1| > 0$. Then

$$\begin{aligned}
\mathbf{E}\int |f_{nh} - f| &\geq \int |f * L_h - f| \quad \text{(Jensen's inequality)} \\
&\geq \sup_t |\varphi(t)||\psi(th) - 1| \\
&\geq \zeta \sup_{|t|\geq(1+\eta)/h} |\varphi(t)| \\
&\geq \xi\zeta \int_{|t|\geq(1+\eta)/h} |\varphi(t)|\, dt \\
&= \xi\zeta\Phi((1+\eta)/h)\ .
\end{aligned}$$

If h remains bounded away from zero, then the last bound shows that $\mathbf{E}\int |f_{nh} - f|$ too remains bounded away from zero. The same is true if nh remains bounded away from ∞ (see e.g., Devroye (1987), pp. 37–38). Let h^* be such that

$$\mathbf{E}\int |f_{nh^*} - f| \sim \inf_h \mathbf{E}\int |f_{nh} - f|\ .$$

Clearly, by what we noted above, $h^* \to 0$ and $nh^* \to \infty$. With $a' \stackrel{\text{def}}{=} \xi\zeta\sqrt{1+\eta}/A$,

$$\begin{aligned}
\inf_h \mathbf{E}\int |f_{nh} - f| &\sim \mathbf{E}\int |f_{nh^*} - f| \\
&\geq (1+o(1)) \max\left(\frac{A}{\sqrt{nh^*}}, \xi\zeta\Phi((1+\eta)/h^*)\right) \\
&\geq (1+o(1)) \inf_h \max\left(\frac{A}{\sqrt{nh}}, \xi\zeta\Phi((1+\eta)/h)\right) \\
&\geq (A/\sqrt{1+\eta} + o(1)) \inf_h \max\left(\sqrt{\frac{1+\eta}{nh}}, \frac{\xi\zeta\sqrt{1+\eta}\,\Phi((1+\eta)/h)}{A}\right) \\
&= (A/\sqrt{1+\eta} + o(1))\, R_{a',1}(n)\ .
\end{aligned}$$

The upper bound for $\inf_h \mathbf{E}\int |f_{nh} - f|$ divided by the lower bound does not exceed

$$\sqrt{1+\eta}\frac{R_{a,1/2}(n)}{R_{a',1}(n)} \to \sqrt{1+\eta}\ ,$$

where we used Lemma 3. This is as close to one as desired by our choice of η. $\square$

PROOF OF THEOREM 4. From Theorem 1, we have, with A as in the proof of Theorem 3, $a \stackrel{\text{def}}{=} 2\int f^\alpha/(A\pi^{1-\alpha})$,

$$\inf_h \mathbf{E}\int |f_{nh}-f| \le (A+o(1))\, R_{a,1-\alpha}(n)\ .$$

With a' and η as in the proof of Theorem 3, we also have

$$\inf_h \mathbf{E}\int |f_{nh}-f| \ge (A/\sqrt{1+\eta}+o(1))\, R_{a',1}(n)\ .$$

The limit supremum of the upper bound divided by the lower bound does not exceed

$$\sqrt{1+\eta}\limsup_{n\to\infty}\frac{R_{a,1-\alpha}(n)}{R_{a',1}(n)} \le \sqrt{1+\eta}\left(\frac{1}{1-\alpha}\right)^{1/(2\theta)}\ ,$$

where we invoked Lemma 3. Now let η decrease to zero. The last statement of Theorem 4 is proved in a similar manner. By taking both η and α small enough, we see that

$$\inf_h \mathbf{E}\int |f_{nh}-f| \sim AR_{a',1}(n)\ .$$

By Lemma 3, this is also asymptotic to $AR_{a,1}(n)$ for any positive a. $\square$

6. Optimality of superkernels.

The main result of this section establishes the optimality of superkernels over all other kernels for a subclass of densities whose characteristic function satisfies a monotonicity condition in the tails. When we wish to compare the performance of superkernels in general with that of other kernels, there are several issues. A realistic comparison should be based upon non-asymptotic results such as explicit inequalities and the like. Without many additional regularity assumptions on f and φ, this is a rather messy undertaking. Instead, we proceed with an asymptotic comparison, which should be interpreted with the usual warnings. Two comparisons can be made, one between superkernels and other kernels, with the understanding that the kernel does not vary with n, and one in which the superkernel is fixed but the competitor is allowed to use any kernel of his choice for any n. It turns out that there is not much difference between these methodologies for the extremely smooth densities considered in this note. Recently, Hall and Marron (1988) showed that for very smooth densities, nothing is gained by allowing K to vary with n (within a rich family of kernels) when the performance is compared with that of the estimate based on $S(x)$ in an L_2 setting. The same conclusion can be drawn from the work of Davis (1975, 1977). However, there is a slight gain over the kernel S for less smooth densities, a case we won't consider here. See however, Cline (1990).

When we compare superkernels among each other, Theorem 5 below partially indicates that we should try to minimize $\int L^2$ (or $\int \psi^2$). The bound given in Theorem 5 depends upon the factor $\pi\int L^2$. By Parseval's identity, this is strictly greater than one for all superkernels. Equality is only attained in the limit, for $L\equiv S$, but unfortunately, the limit kernel, S, is not in L_1. This seems to be a case in which there is no best element, since the winner is a limit of elements from the family of superkernels. The recommendation is to take ψ rectangular with two smooth tails added on so as to make the tails of L small. The size of these tails has to be determined from non-asymptotic considerations, perhaps via some data-based rule.

THEOREM 5. *Let $L \in M$ be a superkernel, let L and L^2 be strong approximate identities, let $\xi > 0$, let f_{nh} be the kernel estimate with kernel L and smoothing parameter h, and assume that $f \in M$, that $|\varphi|$ is monotonically $\downarrow$, and that $T = \infty$, $\int f^\alpha < \infty$ for some $\alpha \in (0, 1/2]$, and* $\log(1/\Phi)$ *is regularly varying at ∞ of order $\theta > 0$. Assume that g_{nh} is the kernel estimate with arbitrary bounded kernel K such that K and K^2 are strong approximate identities. Then*

$$\limsup_{n\to\infty} \frac{\inf_h \mathbf{E} \int |f_{nh} - f|}{\inf_h \mathbf{E} \int |g_{nh} - f|} \le (1-\alpha)^{-1/(2\theta)} \sqrt{\pi \int L^2} \, .$$

If $\log(1/\Phi)$ *is rapidly varying, or if* $\log(1/\Phi)$ *is regularly varying of order $\theta > 0$, while at the same time $\int f^\alpha < \infty$ for all $\alpha \in (0, 1/2]$, then*

$$\limsup_{n\to\infty} \frac{\inf_h \mathbf{E} \int |f_{nh} - f|}{\inf_h \mathbf{E} \int |g_{nh} - f|} \le \sqrt{\pi \int L^2} \, .$$

Proof. From Theorem 1,

$$\limsup_{n\to\infty} \sqrt{nh_{\alpha,n}}\, \inf_h \mathbf{E} \int |f_{nh} - f| \le \sqrt{\frac{2}{\pi}} \int \sqrt{f} \sqrt{\int L^2} \stackrel{\text{def}}{=} A \ .$$

In other words,

$$\inf_h \mathbf{E} \int |f_{nh} - f| \le (A + o(1))\, R_{a,1-\alpha}(n) \ ,$$

where $a \stackrel{\text{def}}{=} 2 \int f^\alpha / (A\pi^{1-\alpha})$, and R is as defined in Lemma 3. Consider next the estimate g_{nh}. Observe that if $v \in (0,1)$, and $\sup_{|t|\le z} |1 - \psi(t)| = v$, then

$$\int K^2 = \frac{1}{2\pi} \int \psi^2 \ge \frac{1}{2\pi} \int_{-z}^{z} \psi^2 \ge \frac{1}{\pi} z (1-v)^2 \ .$$

We have

$$\begin{aligned}
\mathbf{E} \int |f_{nh} - f| &\ge \sup_t |\varphi(t)| |1 - \psi(th)| \quad (\psi \text{ is the Fourier transform of } K) \\
&\ge |\varphi(z/h)| \sup_{|t|\le z/h} |1 - \psi(th)| \\
&\ge |\varphi(z/h)| \sup_{|t|\le z} |1 - \psi(t)| \\
&\stackrel{\text{def}}{=} v |\varphi(z/h)| \\
&\ge v |\varphi(\pi \int K^2 / (1-v)^2 h)| \\
&\ge v \xi \Phi(\pi \int K^2 / (1-v)^2 h) \ .
\end{aligned}$$

Recalling the lower bound of Lemma 4, and using arguments as in the proof of Theorem 3, we see that, if $a' \stackrel{\text{def}}{=} v\xi \sqrt{\pi \int K^2} / (A(1-v))$, $B \stackrel{\text{def}}{=} \sqrt{2/\pi} \int \sqrt{f} \sqrt{\int K^2}$,

$$\begin{aligned}
\inf_h \mathbf{E} \int |g_{nh} - f| &\ge \inf_h \max \left(\frac{B + o(1)}{\sqrt{nh}}, v\xi \Phi(\pi \int K^2 / (1-v)^2 h) \right) \\
&\ge (B + o(1)) \inf_h \max \left(\frac{1}{\sqrt{nh}}, v\xi \Phi(\pi \int K^2 / (1-v)^2 h) / B \right) \\
&= (B + o(1)) \inf_h \max \left(\frac{(1-v)}{\sqrt{nh\pi \int K^2}}, v\xi \Phi(1/h) / B \right) \\
&\sim \frac{B(1-v)}{\sqrt{\pi \int K^2}} \inf_h \max \left(\frac{1}{\sqrt{nh}}, a' \Phi(1/h) \right) \\
&= \frac{\sqrt{2} \int \sqrt{f} (1-v)}{\pi} R_{a',1}(n) \ ,
\end{aligned}$$

where we used the notation of Lemma 3. The Theorem is proved if we can bound the limit supremum of the ratio

$$\frac{A\,R_{a,1-\alpha}(n)}{\sqrt{2}\int\sqrt{f}((1-v)/\pi)R_{a',1}(n)} = \frac{\sqrt{\pi\int L^2}\,R_{a,1-\alpha}(n)}{(1-v)R_{a',1}(n)}\ .$$

But by Lemma 3, this does not exceed

$$\frac{\sqrt{\pi\int L^2}}{(1-v)}\,(1-\alpha)^{-1/(2\theta)}\ .$$

The first part of Theorem 5 follows from our arbitrary choice of v. The last part is obtained from this either by letting α tend to zero, or by letting θ tend to ∞. □

7. Data-based choice for h.

With superkernels, it seems particularly challenging to derive a data-based method for the selection of a nearly optimal h. This question is even more pressing here since, as we have seen, the estimate is extremely sensitive to overestimation of h. Furthermore, relative stability inequalities that are so handy in the study of the behavior of data-based smoothing factors (Wand, 1989; Hall and Wand, 1988; Devroye, 1987) tell us that $\int|f_n-f|$ oscillates about its mean with deviations that are of the order of $1/\sqrt{n}$; in fact, the probability that the difference exceeds ϵ is not more than $2\exp(-n\epsilon^2/(32\int^2|K|))$ for any kernel K, any f,h and n (Devroye, 1988). To make this less than a negative power of n, it is necessary to take ϵ of the order of $\sqrt{\log n/n}$, which is larger than the errors that we are working with for extremely smooth densities. In an L_2 context, Hall and Marron (1988) encountered similar problems, and had to exclude extremely smooth densities from their analysis of the asymptotic optimality of the L_2 cross-validation method for picking h.

8. References.

H. Akaike, "An approximation to the density function," *Annals of the Institute of Statistical Mathematics*, vol. 6, pp. 127–132, 1954.

D. B. H. Cline, "Optimal kernel estimation of densities," *Annals of the Institute of Statistical Mathematics*, vol. 42, pp. 287–303, 1990.

K. B. Davis, "Mean square error properties of density estimates," *Annals of Statistics*, vol. 5, pp. 1025–1030, 1975.

K. B. Davis, "Mean integrated square error properties of density estimates," *Annals of Statistics*, vol. 5, pp. 530–535, 1977.

L. Devroye, *A Course in Density Estimation*, Birkhauser, Boston, 1987.

L. Devroye, "The kernel estimate is relatively stable," *Probability Theory and Related Fields*, vol. 77, pp. 521–536, 1988.

L. Devroye, "Asymptotic performance bounds for the kernel estimate," *Annals of Statistics*, vol. 16, pp. 1162–1179, 1988.

L. Devroye, "The double kernel method in density estimation," *Annales de l'Institut Henri Poincaré*, vol. 25, pp. 533–580, 1989.

L. Devroye and L. Györfi, *Nonparametric Density Estimation: The L1 View*, John Wiley, New York, 1985.

L. Gajek, "Estimating a density and its derivatives via the minumum distance method," *Probability Theory and Related Fields*, vol. 80, pp. 601–617, 1989.

P. Hall and J. S. Marron, "Choice of kernel order in density estimation," *Annals of Statistics*, vol. 16, pp. 161–173, 1988.

P. Hall and M. P. Wand, "Minimizing L1 distance in nonparametric density estimation," *Journal of Multivariate Analysis*, vol. 26, pp. 59–88, 1988.

I. A. Ibragimov and R. Z. Khasminskii, "Estimation of distribution density belonging to a class of entire function," *Theory of Probability and its Applications*, vol. 27, pp. 551–562, 1982.

T. Kawata, *Fourier Analysis in Probability Theory*, Academic Press, New York, 1972.

V. D. Konakov, "Non-parametric estimation of density functions," *Theory of Probability and its Applications*, vol. 17, pp. 361–362, 1972.

J. Koronacki, "Kernel estimation of smooth densities using Fabian's approach," *Statistics*, vol. 18, pp. 37–47, 1987.

E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Annals of Mathematical Statistics*, vol. 27, pp. 832–837, 1956.

E. Seneta, "Regularly Varying Functions," Lecture Notes in Mathematics 508, Springer-Verlag, Berlin, 1975.

E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, New Jersey, 1970.

M. P. Wand, “On nonparametric curve estimation and discrimination,” Ph.D. Thesis, Australian National University, Canberra, Australia, 1989.

G. S. Watson and M. R. Leadbetter, “On the estimation of the probability density,” *Annals of Mathematical Statistics*, vol. 34, pp. 480–491, 1963.