VARIABLE KERNEL ESTIMATES: ON THE IMPOSSIBILITY OF TUNING THE PARAMETERS

Luc Devroye
School of Computer Science
McGill University
Montreal, Canada H3A 2A7

Gábor Lugosi
Department of Economics
Pompeu Fabra University
Ramon Trias Fargas, 25-27
08005 Barcelona, Spain

ABSTRACT. For the standard kernel density estimate, it is known that one can tune the bandwidth such that the expected L1 error is within a constant factor of the optimal L1 error (obtained when one is allowed to choose the bandwidth with knowledge of the density). In this paper, we pose the same problem for variable bandwidth kernel estimates where the bandwidths are allowed to depend upon the location. We show in particular that for positive kernels on the real line, for any data-based bandwidth, there exists a density for which the ratio of expected L1 error over optimal L1 error tends to infinity. Thus, the problem of tuning the variable bandwidth in an optimal manner is "too hard". Moreover, from the class of counterexamples exhibited in the paper, it appears that placing conditions on the densities (monotonicity, convexity, smoothness) does not help.

KEYWORDS AND PHRASES. Density estimation, variable kernel estimate, convergence, smoothing factor, minimax lower bounds, asymptotic optimality.

1991 MATHEMATICS SUBJECT CLASSIFICATIONS: Primary 62G05.

RUNNING HEAD: VARIABLE KERNEL ESTIMATES

The first author's work was supported by NSERC Grant A3456 and by FCAR Grant 90-ER-0291. The second author's work was supported by DIGES Grant PB96-0300.

1. Introduction.

We are given an i.i.d. sample $X_1, \ldots, X_n$ drawn from an unknown density f on $\mathbb{R}$. A density estimate $f_n(x) = f_n(x, X_1, \ldots, X_n)$ is a real-valued measurable function of its arguments. One of the most popular estimates is

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

where $K : \mathbb{R} \to \mathbb{R}$ is a fixed kernel with $\int K = 1$, $K_h(x) = (1/h^d)K(x/h)$, and $h > 0$ is the smoothing factor (Akaike, 1954; Parzen, 1962; Rosenblatt, 1956). Many data-dependent choices for h have been proposed in the literature. The question of whether h can be tuned in an optimal manner has been answered in the affirmative by Devroye and Lugosi (1996, 1997), where data-dependent smoothing factors H are introduced for which

$$\sup_f \limsup_{n \to \infty} \frac{\mathbf{E} \int |f_{nH} - f|}{\inf_h \mathbf{E} \int |f_{nh} - f|} \leq 3 \ ,$$

whenever the kernel K is nonnegative, Lipschitz, and of a compact support.

The variable kernel estimate (Breiman, Meisel and Purcell, 1977; Raatgever and Duin, 1978; Habbema, Hermans and Remme, 1978) allows h to depend upon either i or x. In the BMP (Breiman-Meisel-Purcell) form

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

f_n is still a density. Various ways of letting the data pick the h_i's were proposed by Breiman, Meisel and Purcell (1977) and others later. As each h_i should be tuned to what is happening near X_i, the behavior of the BMP estimate should not be radically different from that of the raw variable kernel estimate in which $h = h(x)$. In a data-based form, we have $H = H(x; X_1, \ldots, X_n)$. It is this form that will be studied in this paper.

We note here some general references on variable kernel estimates, such as Devroye (1985), Devroye and Penrod (1986), Jones (1990), Terrell and Scott (1992), Hall (1992), and Marron, Hall and Hu (1995), who all deal with convergence issues. A particularly influential paper was that of Abramson (1982), who showed how variable bandwidths with positive kernels can nevertheless induce convergence rates usually attainable with fixed bandwidths and fourth order kernels (see also Hall and Marron, 1988). This sets variable bandwidth kernels apart from fixed bandwidth kernels. Particular variable bandwidths include the cross-validation methods of Hall and Marron (1988), Hall and Schucany (1989) and Mielniczuk, Sarda and Vieu (1989), Sheather's solve-the-equation bandwidth (1983, 1986) (Thombs and Sheather, 1992), Sain's (1994) and Sain and Scott's bootstrap bandwidth (1996), and Hazelton's (1996) and Farmen's (1996) smoothed bootstrap bandwidths. None of the cited papers addresses the question posed here. Our work was inspired by interesting observations by David Scott at presentations in Louvain La Neuve and Montreal in 1997, and which are summarized in Sain and Scott (1997).

To introduce our main result, we fix the notation as follows: the raw variable kernel estimate is

$$f_{n,h(x)}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h(x)}(x - X_i),$$

and the data-based variable kernel estimate is

$$f_{n,H(x)}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{H(x)}(x - X_i),$$

where it is understood that $H(x) = H(x; X_1, \ldots, X_n)$. Let $\mathcal{F}_B$ be the class of nondecreasing, convex-shaped densities f on $[t, t+s]$ with $s \sup_{(t,t+s)} f(x) \leq B$, where $t \in \mathbb{R}$ and $s > 0$ are arbitrary. (Note that t is just a translation constant and that if $f \in \mathcal{F}_B$, then so are all rescaled and translated versions of f.) One of the results we show is the following: if $K \geq 0$ is a symmetric square-integrable kernel on $[-1, 1]$, then there exists a positive constant C not depending upon n such that

$$\inf_{H:\mathbb{R}^{n+1} \to (0,\infty)} \sup_{f \in \mathcal{F}_B} \frac{\mathbf{E} \int |f_{n,H(x)}(x) - f(x)| dx}{\inf_{h:\mathbb{R} \to (0,\infty)} \mathbf{E} \int |f_{n,h(x)}(x) - f(x)| dx} \geq C n^{\frac{1}{10}}.$$

So, even with the knowledge that $f \in \mathcal{F}_B$, one cannot efficiently design a variable bandwidth. Generalizations and extensions of this result round out the paper.

It is generally accepted that some versions of variable bandwidths outperform fixed bandwidth estimates, and our results do not contradict this. However, we only say that the class of variable bandwidth kernel estimates is too large to be optimized. Devroye, Lugosi and Udina (1998) basically describe how far one can go in the optimization: they partition the real line into k intervals and use a different (fixed) bandwidth on each interval. Simultaneous optimization of the k intervals and the k bandwidths is possible in the sense that there exists a data-based choice of a piecewise constant function $H(x)$ with k pieces such that for all densities

$$\mathbf{E} \int |f_{n,H(x)}(x) - f(x)| dx \leq c_0 \inf \mathbf{E} \int |f_{n,h(x)}(x) - f(x)| dx + c_1 \sqrt{\frac{k \log n}{n}},$$

where c_0 and c_1 are absolute constants, and the infimum is taken over all variable kernel estimates where $h(x)$ is piecewise constant with k pieces. However, as shown in this paper, optimization becomes impossible when $k = n$. As a happy by-product, any heuristic for variable bandwidths is suboptimal, and therefore, no claim of superiority can be made for any method.

2. The main result.

We state our main result for the simplest possible kernels and will generalize later. No attempt is made to optimize the constants in the bounds. We introduce a shape parameter for nonnegative symmetric kernels K:

$$\rho \stackrel{\text{def}}{=} \text{support}(K) \times \int K^2 \ .$$

Note that by the Cauchy-Schwarz inequality, $\rho \geq 1$, and equality is reached for the uniform kernel. In any case, ρ is a scale-invariant parameter that will appear in our bounds.

THEOREM 1. *Let K be a symmetric nonnegative square-integrable kernel with shape parameter ρ. Then, for $n \geq 24$,*

$$\inf_{H:\mathbb{R}^{n+1}\to(0,\infty)} \sup_{f\in\mathcal{F}_{7/3}} \frac{\mathbf{E}\int |f_{n,H(x)}(x)-f(x)|dx}{\inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E}\int |f_{n,h(x)}(x)-f(x)|dx} \geq \frac{n^{\frac{1}{10}}}{324\sqrt{\rho}}.$$

The choice of the constant $B = 7/3$ is irrelevant, and is motivated by convenience. To prove Theorem 1, we combine two results, an estimate of the L1 error for raw variable kernel estimates inspired by recent work of Sain and Scott (1997), and a minimax lower bound for $\mathcal{F}_B$. Sain and Scott remarked that for most locations x, the kernel estimate could be made unbiased by taking $h(x)$ fixed and positive (not depending upon n). A similar but less explicit remark may also be found in Hazelton (1996, p. 223). This prompted us to consider $f \in \mathcal{F}_B$, where every x has this property. In particular, the rate of convergence could be $O(1/\sqrt{n})$ if we knew f and thus the optimal map $h : \mathbb{R} \to (0,\infty)$.

LEMMA 1. *Let K be a symmetric nonnegative square-integrable kernel on the real line with shape parameter ρ. Then*

$$\sup_{f\in\mathcal{F}_B} \inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E}\int |f_{n,h(x)}(x)-f(x)|dx \leq \sqrt{\frac{4B\rho}{n}}.$$

PROOF. Fix $f \in \mathcal{F}_B$. By a linear transformation, we may and will assume that f is supported on $[0,1]$ and bounded by B. Let the support of K be $[-c,c]$. Note that we may set $h(x) = |x|$ for $x < 0$ and $h(x) = x - 1$ for $x > 1$, so that since f has support on $[0,1]$ and K on $[-c,c]$, we have $f_{n,h(x)} \equiv f(x) \equiv 0$ outside $[0,1]$. Thus, pick $x \in (0,1)$. We use the notation $*$ for the convolution operator, and drop the argument of $h(x)$. Note that as long as $h = h(x)$ is less than $\min(x, 1-x)/c$,

$$\begin{aligned}
\mathbf{E} f_{n,h}(x) &= f * K_h(x) \\
&= (1/h) \int f(y) K((x-y)/h) dy \\
&= \int f(x+hw) K(w) dw \\
&= \mathbf{E} f(x + hW) \quad \text{(where } W \text{ has density } K\text{)} \\
&\geq f(x + h\mathbf{E}\{W\}) \quad \text{(by Jensen's inequality)} \\
&= f(x).
\end{aligned}$$

Also, it is trivial to check that $\lim_{h\to\infty} f * K_h(x) = 0$. Thus, since $f * K_h(x)$, considered as a function of h, is continuous in h, we see that necessarily $f(x) = f * K_h(x)$ for some $h > 0$, and, in fact, there exists at least one such h with $h \geq \min(x, 1-x)/c$. For $x \in (0,1)$, we define

$$h(x) = \sup\{z > 0 : f * K_z(x) = f(x)\} ,$$

and note for further use that $h(x) \geq \min(x, 1-x)/c$. From here on, h denotes this choice. Then, by routine calculations,

$$\begin{aligned}
\mathbf{E}\left\{|f_{n,h}(x) - f(x)|\right\} &= \mathbf{E}\left\{|f_{n,h}(x) - \mathbf{E} f_{n,h}(x)|\right\} \\
&\leq \sqrt{\mathbf{E}\left\{(f_{n,h}(x) - \mathbf{E} f_{n,h}(x))^2\right\}} \\
&= \sqrt{\frac{\mathrm{Var}\left\{K_h(x - X_1)\right\}}{n}} \\
&\leq \sqrt{\frac{\mathbf{E}\left\{(K_h(x - X_1))^2\right\}}{n}} \\
&= \sqrt{\frac{\mathbf{E}\left\{\int K^2((x-y)/h) h^{-2} f(y) dy\right\}}{n}} \\
&\leq \sqrt{\frac{B \int K^2}{nh}} \, .
\end{aligned}$$

This is true for all $x \in (0,1)$, and integration yields

$$\mathbf{E}\left\{\int |f_{n,h}(x) - f(x)| dx\right\} \leq \sqrt{\frac{B\int K^2}{n}} \int_0^1 \frac{1}{\sqrt{h(x)}} dx$$

$$\leq 2\sqrt{\frac{Bc\int K^2}{n}}\int_0^{1/2}\frac{1}{\sqrt{x}}dx$$
$$= \sqrt{\frac{4B\rho}{n}}$$

as required. □

It is of course impossible to estimate all f's in $\mathcal{F}_B$ at the rate $1/\sqrt{n}$ with any density estimate, let alone the data-based variable kernel density estimate. To support this, we merely require a minimax lower bound. For general results in density estimation, this was done by Birgé (1984, 1985, 1986) and Devroye (1987), and details for monotone densities are worked out in Birgé (1987a, 1987b) and Yang (1996). For the family $\mathcal{F}_B$, we merely require a good non-asymptotic bound with the right dependence on n but not B, so a bound is derived here which is optimal in n but suboptimal in B.

LEMMA 2. *For $n \geq 24$, we have*

$$\inf_{f_n}\sup_{f\in\mathcal{F}_{7/3}} \mathbf{E}\int |f_n - f| \geq \frac{1}{106.02476\ldots n^{2/5}},$$

where the infimum is over all density estimates.

PROOF. Let k be a suitable positive integer to be determined further on. We will construct a family of 2^k densities contained in F_B, where $B = 7/3$. The interval $[0,1]$ is partitioned into k intervals denoted by $A_1, \ldots, A_k$. Thus, $A_i = [(i-1)/k, i/k)$. On $[(i-1)/k, i/k]$, we consider two piecewise linear functions f_i and g_i with the following properties:

A. $f_i = g_i$ at the ends of the interval, and $f_{i-1}((i-1)/k) = f_i((i-1)/k)$;

B. $\int_{A_i} f_i = \int_{A_i} g_i$;

C. $\sup_{x\in((i-2)/k,(i-1)/k)} \max(f'_{i-1}(x), g'_{i-1}(x)) \leq \inf_{x\in((i-1)/k,i/k)} \min(f'_i(x), g'_i(x))$;

D. $\sup_{x\in A_{i-1}} \max(f_{i-1}(x), g_{i-1}(x)) \leq \inf_{x\in A_i} \min(f_i(x), g_i(x))$;

E. $f_1(0) = g_1(0) = A > 0$, where $A = 1/3$;

F. $f_k(1) = g_k(1) \leq 7/3$;

G. $\sum_{i=1}^k \int_{A_i} f_i = 1$.

If we piece together a function f by choosing either f_i or g_i on A_i, then we obtain a bona fide density (by B, D, E and G). Furthermore, f is increasing (by D), continuous on (0,1) (by A) and convex (by A and C). There are 2^k such possible functions, and we may parametrize the family by a bit vector $b = (b_1, \ldots, b_k)$, where $b_i = 1$ if the corresponding f picks f_i on A_i, and $b_i = 0$ otherwise. We will denote this f by f_b and apologize for using the same notation as in f_i.

Define $d_i = 4ai$ for some constant a. We will pick the f_i's such that on A_i, both f_i' and g_i' are between d_{i-1} and d_i. This will then insure nondecreasing derivatives, and thus convexity for f. The positivity is also insured. On A_i, we set

$$f_i' = \begin{cases} d_{i-1} + a & x \in [(i-1)/k, (i-1/3)/k) \\ d_i & x \in [(i-1)/k + 2/(3k), i/k) \end{cases}$$

and

$$g_i' = \begin{cases} d_{i-1} & x \in [(i-1)/k, (i-2/3)/k) \\ d_i - a = d_{i-1} + 3a & x \in [(i-1)/k + 1/(3k), i/k) \end{cases}$$

It takes a moment to verify that $\int_{A_i} f_i' = \int_{A_i} g_i'$ so that f_i and g_i make equal jumps. Another routine computation shows that $\int_{A_i} f_i = \int_{A_i} g_i$ as well. Thus, the functions will suit us, provided that the total integral is one. To do this, we observe that

$$\int_{A_i} f_i' = \frac{2}{3k}(4a(i-1) + a) + \frac{4ai}{3k} = \frac{4ai - 2a}{k},$$

and similarly for $\int_{A_i} g_i'$. Hence, at i/k, the value of any f in our class is

$$A + \sum_{j=1}^{i} \frac{4aj - 2a}{k} = A + \frac{2ai^2}{k}.$$

For future reference, we note that $f_k(1) = g_k(1) = A + 2ak$. Next, we compute the integrals:

$$\begin{aligned} \int_{A_i} f_i &= \int_0^{2/3k} (f_i((i-1)/k) + (4(i-1)a + a)x)dx + \int_0^{1/3k} (f_i(i/k) - 4aix)dx \\ &= \frac{A}{k} + \frac{a(18i^2 - 18i + 22)}{9k^2}. \end{aligned}$$

Thus, the integral of any f in our class is the sum over the k individual pieces, which is

$$A + \frac{a(6k^2 + 16)}{9k}.$$

As $A = 1/3$ and the integral must be one, we must take

$$a = \frac{6k}{6k^2 + 16}.$$

We verify quickly that $a \le 1/k$, and that for $k \ge 2$, $a \ge 4/7k$. Finally, we insure that each f is bounded by 7/3:

$$f_k(1) = \frac{1}{3} + \frac{12k^2}{6k^2+16} = \frac{42k^2+16}{18k^2+48} < \frac{7}{3} \; .$$

This ends the construction of our parametric family with 2^k members.

We now apply a minimax lower bound method pioneered by Assouad (1983), in the form given in Devroye (1987, p. 60): we need to compute two lower bounds α and β, where α is a uniform lower bound on $\int_{A_i} |f_i - g_i|$, and β is a uniform lower bound on $\int \sqrt{fg}$, where f and g are two of the 2^k functions that differ only on one interval. The required uniformity is with respect to i, the interval index. Clearly,

$$\int_{A_i} |f_i - g_i| = 2\left(\int_0^{1/3k} axdx + \int_0^{1/6k} 2axdx \right) = \frac{a}{6k^2}.$$

Thus, we may set

$$\alpha = \frac{a}{6k^2}.$$

Next, pick i, and let $I = A_i$, $O = [0,1] - A_i$. Let f and g be two of our functions that are equal on all intervals except A_i. Note that on A_i, $|f_i - g_i|$ is maximal at $1/3k$ from the left boundary, and the value there is $a/(3k)$. Then

$$\begin{aligned}
\int \sqrt{fg} &= \int_O f + \int_I \sqrt{fg} \\
&= \int_O f + \int_I \sqrt{\left(\frac{f+g}{2}\right)^2 - \left(\frac{f-g}{2}\right)^2} \\
&= \int_O f + \int_I \frac{f+g}{2} - \int_I \frac{f+g}{2}\left(1 - \sqrt{1 - \left(\frac{f-g}{f+g}\right)^2}\right) \\
&= \int_O f + \int_I f - \int_I \frac{f+g}{2}\left(1 - \sqrt{1 - \left(\frac{f-g}{f+g}\right)^2}\right) \\
&\ge 1 - \int_I \frac{(f+g)(f-g)^2}{2(f+g)^2} \\
&\ge 1 - \frac{a^2}{18k^2}\int_I (f+g)^{-1}dx \\
&\ge 1 - \frac{3a^2}{18k^3} \\
&= 1 - \frac{a^2}{6k^3}
\end{aligned}$$

$$\geq 1 - \frac{1}{6k^5}$$

We set $1-\beta = 1/(6k^5)$. By Assouad's theorem (as in Devroye, 1987, p. 60),

$$\begin{aligned}
\inf_{f_n} \sup_{f \in \mathcal{F}_{7/3}} \mathbf{E} \int |f_n - f| &\geq \frac{k\alpha}{2}\left(1 - \sqrt{2n(1-\beta)}\right) \\
&= \frac{ka}{12k^2}\left(1 - \sqrt{\frac{n}{3k^5}}\right) \\
&\geq \frac{1}{21k^2}\left(1 - \sqrt{\frac{1}{4}}\right) \\
&\quad \text{(provided } k \geq 2 \text{ and } k^5 \geq 4n/3) \\
&= \frac{1}{42k^2} \\
&= \frac{1}{42\lceil (4n/3)^{1/5} \rceil^2} \\
&\quad \text{(upon taking } k = \lceil (4n/3)^{1/5} \rceil) \\
&\geq \frac{1}{42(1 + (4n/3)^{1/5})^2}.
\end{aligned}$$

Note that the condition $k \geq 2$ holds if $n \geq 24$. In that case, $1 \leq (4n/3)^{1/5}/2$, and the lower bound is thus at least

$$\frac{1}{42((3/2)(4n/3)^{1/5})^2} \geq \frac{1}{106.02476\ldots n^{2/5}}$$

which was to be shown. $\square$

We may now combine Lemmas 1 and 2 to prove Theorem 1. Indeed,

$$\begin{aligned}
\inf_{H:\mathbb{R}^{n+1} \to (0,\infty)} \sup_{f \in \mathcal{F}_{7/3}} \frac{\mathbf{E} \int |f_{n,H(x)}(x) - f(x)| dx}{\inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E} \int |f_{n,h(x)}(x) - f(x)| dx} &\geq \frac{\inf_{f_n} \sup_{f \in \mathcal{F}_{7/3}} \mathbf{E} \int |f_{n,H(x)}(x) - f(x)| dx}{\sup_{f \in \mathcal{F}_{7/3}} \inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E} \int |f_{n,h(x)}(x) - f(x)| dx} \\
&\geq \frac{\frac{1}{106.02476\ldots n^{2/5}}}{\sqrt{\frac{28\rho}{3n}}} \\
&\geq \frac{n^{\frac{1}{10}}}{324\sqrt{\rho}}.
\end{aligned}$$

This concludes the proof of Theorem 1.

From looking at Lemma 1, one is tempted to try to find data-based variable bandwidths that achieve $O(1/\sqrt{n})$ error rates over the given class of convex densities, but

Theorem 1 shows that in a uniform sense, no such rate is possible. It does not imply that there always exists one convex density for which the rate is worse. To prove that this is the case—and thus, that the "bad density" does not change with n, as in minimax results—, we need additional work, which is presented in the next section.

Sain and Scott (1997) give a detailed and lucid account of the zero-bias bandwidth for locally convex densities. They consider pointwise L2 errors (or MSE) and observe $O(1/n)$ pointwise rates if $h(x)$ were known. They try to estimate $h(x)$ by cross-validation. Unfortunately, they report that in practice the $O(1/n)$ rate (which corresponds to $O(1/\sqrt{n})$ L1 errors) is not achievable, which in view of our results is to be expected. We would like to make one fundamental remark however about the MISE $\mathbf{E}\int(f_n - f)^2$. If f_n is the kernel estimate with optimal $h(x)$ and $f \in \mathcal{F}_B$, then the MISE is in general infinite (while the L1 error by Lemma 1 is $O(1/\sqrt{n})$. Indeed, due to the squaring, the variance is $O(\int_0^1 1/(nh(x))dx)$, and as $h(x)$ near 0 varies about like x, we see that the integrated variance blows up.

3. Bad densities for the entire sequence.

Assume that we are given an entire sequence of local data-based bandwidths, where the n-th mapping is $H_n : \mathbb{R}^{n+1} \to (0,\infty)$. We write H_n also for $H_n(x; X_1, \ldots, X_n)$. In this section, we show (Theorem 2 below) that if $a_n \downarrow 0$ arbitrarily slowly, then there exists a monotone bounded piecewise convex density on $[0,1]$ for which the ratio

$$\frac{\mathbf{E}\int |f_{n,H_n(x)}(x) - f(x)|dx}{\inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E}\int |f_{n,h(x)}(x) - f(x)dx}$$

is infinitely often more than $a_n n^{1/10}$. Thus, the same density can be used as a counterexample no matter how large the sample size is. In our proof, we follow to some extent the lead of Birgé (1986).

LEMMA 3. *Let K be a symmetric nonnegative square-integrable kernel on the real line with shape parameter ρ. Partition the line into an infinite number of intervals of length l_i and weight p_i. On the i-th interval, let f be a nondecreasing convex nonnegative function of integral p_i and taking maximal value m_i. Assume that $\sum_i p_i = 1$. Then*

$$\inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E}\int |f_{n,h(x)}(x) - f(x)|dx \le \sum_i \min\left(\sqrt{\frac{4l_i m_i \rho}{n}}, 2p_i\right) .$$

PROOF. We refer to the proof of Lemma 1, and only modify a few bounds. Consider the i-th interval, and assume without loss of generality that it is $[0, l_i]$. Define $h(x)$ as in the proof of Lemma 1, and note that $h(x)$ is at least equal to the minimum distance from x to the border of its interval. As intervals can thus be treated separately, we may argue as on one interval, as in Lemma 1. For $x \in [0, l_i]$, define $\delta(x) = \min(x, l_i - x)$. Note that arguing as in Lemma 1, $h(x) \geq \delta(x)/c$, where $[-c, c]$ is the support of K.

We also note that by the choice of $h(x)$, and positivity of K, we have

$$\mathbf{E}\{|f_{n,h}(x) - f(x)|\} \leq f(x) + \mathbf{E} f_{n,h}(x) = 2f(x) \ .$$

Thus, for all x in the i-th interval,

$$\begin{aligned}
\mathbf{E}\{|f_{n,h}(x) - f(x)|\} &\leq \min\left(\sqrt{\frac{m_i \int K^2}{nh(x)}}, 2f(x)\right) \\
&\leq \min\left(\sqrt{\frac{m_i c \int K^2}{n\delta(x)}}, 2f(x)\right) \\
&= \min\left(\sqrt{\frac{m_i \rho}{2n\delta(x)}}, 2f(x)\right)
\end{aligned}$$

so that, taking integrals then yields

$$\begin{aligned}
\mathbf{E}\int |f_{n,h}(x) - f(x)|\, dx &\leq \sum_i \int_{i\text{-th interval}} \min\left(\sqrt{\frac{m_i \rho}{2n\delta(x)}}, 2f(x)\right)\, dx \\
&\leq \sum_i \min\left(\int_{i\text{-th interval}} \sqrt{\frac{m_i \rho}{2n\delta(x)}}\, dx, \int_{i\text{-th interval}} 2f(x)\, dx\right) \\
&= \sum_i \min\left(\sqrt{\frac{m_i \rho}{2n}} 2\int_0^{l_i/2} 1/\sqrt{x} dx, 2p_i\right) \\
&= \sum_i \min\left(\sqrt{\frac{4 l_i m_i \rho}{n}}, 2p_i\right) \ . \square
\end{aligned}$$

Lemma 3 provides us with a rich enough family of densities from which to draw examples: it is applicable to basically all bounded piecewise convex densities. The freedom in the choice of the Lemma's parameters will be useful further on. The following Lemma allows us to use minimax lower bounds for densities on a fixed interval for mixtures of densities.

LEMMA 4. *Let g be a fixed density supported outside $[0,1]$, and let $\mathcal{F}_B$ be as in Lemma 2. Let $p \in (0,1)$ and let $\mathcal{G}$ be the class of densities of the form $pf + (1-p)g : f \in \mathcal{F}_{7/3}$. For $n \geq 24$, we have*

$$\inf_{f_n} \sup_{f \in \mathcal{G}} \mathbf{E} \int |f_n - f| \geq \frac{p}{106.02476\ldots n^{2/5}},$$

where the infimum is over all density estimates.

PROOF. We mimic the proof of Lemma 2, and make only changes where appropriate. The 2^k-member subclass construction is as in Lemma 2, except for the multiplicative factor p. This leads to the choice $1 - \beta = p/6k^5$. By Assouad's theorem,

$$\begin{aligned}
\inf_{f_n} \sup_{f \in \mathcal{G}} \mathbf{E} \int |f_n - f| &\geq \frac{kp\alpha}{2} \left(1 - \sqrt{2n(1-\beta)}\right) \\
&= \frac{kpa}{12k^2} \left(1 - \sqrt{\frac{np}{3k^5}}\right) \\
&\geq \frac{p}{21k^2} \left(1 - \sqrt{\frac{p}{4}}\right) \\
&\quad \text{(provided } k \geq 2 \text{ and } k^5 \geq 4n/3\text{)} \\
&\geq \frac{p}{42k^2} \\
&= \frac{p}{42\lceil (4n/3)^{1/5} \rceil^2} \\
&\quad \text{(upon taking } k = \lceil (4n/3)^{1/5} \rceil\text{)} \\
&\geq \frac{p}{42(1 + (4n/3)^{1/5})^2}.
\end{aligned}$$

Note that the condition $k \geq 2$ holds if $n \geq 24$. In that case, $1 \leq (4n/3)^{1/5}/2$, and the lower bound is thus at least

$$\frac{p}{42((3/2)(4n/3)^{1/5})^2} \geq \frac{p}{106.02476\ldots n^{2/5}}$$

which was to be shown. □

THEOREM 2. *Let a_n be a strictly decreasing sequence of positive numbers with zero limit.*

A. *Let f_n be any density estimate. Then there exists a piecewise convex nondecreasing density f on $[-1,0]$ bounded by 5 and a subsequence n_j such that along this subsequence*
$$\mathbf{E}\int |f - f_{n_j}| \geq a_{n_j} n_j^{-2/5}.$$

B. *Let $f_{n,H_n(x)}$ denote any variable kernel estimate with kernel K and local bandwidth H_n. If K is symmetric, nonnegative, and has shape parameter $\rho < \infty$, then there exists a piecewise convex nondecreasing density f on $[-1,0]$ bounded by 5 and a subsequence n_j such that along this subsequence*
$$\frac{\mathbf{E}\int |f_{n_j,H_{n_j}(x)} - f|}{\inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E}\int |f_{n_j,h(x)}(x) - f(x)|dx} \geq a_{n_j} n_j^{1/10}\ .$$

PROOF. We consider an infinite subfamily of densities within the conditions of Lemma 3. Let $p_1 > p_2 > \ldots$ be a probability vector, and let $B_i = [-1+1/2^i, -1+2/2^i)$ for $i \geq 1$. On B_i, define a class of densities as in the proof of Lemma 2, parametrized by k_i, the number of partitions, and scale each density by $l_i = 1/2^i$. Each density is characterized by a bit vector b_i with k_i bits. The density for b_i on B_i is denoted by f_{i,b_i}. If $b = (b_1, b_2, \ldots)$, define the density
$$f_b = p_1 f_{1,b_1} + p_2 f_{2,b_2} + \cdots$$
and consider the class $\mathcal{F}$ of all these densities. Note that on B_i, each f_b takes values in $[p_i/3l_i, (7/3)p_i/l_i]$, is nondecreasing and piecewise convex. Formally, in Lemma 3, we have $m_i = 7p_i/3l_i$, and $l_i = 1/2^i$. Each f_b is supported on $[-1,0]$, and is nondecreasing provided that
$$\frac{7p_{i+1}}{3l_{i+1}} \leq \frac{p_i}{3l_i}$$
for all i. This translates in the condition
$$\frac{p_{i+1}}{p_i} \leq \frac{1}{14}\ .$$
Finally, note that each f_b is bounded by $(7/3)p_1/l_1 \leq 14/3$. Our class is thus contained in the class of bounded support bounded monotone densities. (With some extra effort, we could have made the densities continuous as well within the support.) By Lemma 3, for all b,
$$\inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E}\int |f_{n,h(x)}(x) - f_b(x)|dx \leq \frac{\sum_i \sqrt{4l_i m_i \rho}}{\sqrt{n}} = \frac{\sqrt{\frac{28\rho}{3}}\sum_i \sqrt{p_i}}{\sqrt{n}}\ .$$

It remains to show how to replace Lemma 2. Let $\{f_n\}$ be any sequence of density estimates. Then, let $24 < n_1 < n_2 < \cdots$ be a specially selected sequence of sample sizes. The interval B_i is targeted when $n = n_i$, and we do not care about sample sizes n outside the collection of n_i's. The observations in the sample are all determined from b and the i.i.d. pairs $(Z_1, U_1), (Z_2, U_2), \ldots$, where the Z_i's are discrete and take the value j with probability p_j, $j \geq 1$, and the U_i's are uniform $[0,1]$. Z_i picks the mixture component for X_i, and U_i is used in the probability integral transform to obtain X_i from the density f_{Z_i, b_i}. Note in particular that with this embedding, if $N_j = 0$ (where N_j is the number of X_i's in interval B_j), then the sample is unaffected by changes of b_j. The existence of a bit vector b will be established inductively. Assume that we have fixed $b_1, \ldots, b_{j-1}$. We have $n = n_j$, by assumption. Let E denote the event that $N = \sum_{k>j} N_k = 0$. Let $\mathcal{F}$ denote the family of densities f_b with the first $j-1$ components of b fixed as above and with the components b_k, $k > j$ all zero. The family has 2^{k_j} members distinguished by all possible values for b_j. Then at $n = n_j$,

$$\sup_{b_j} \mathbf{E} \int |f_n - f_b| \geq \sup_{b_j} \mathbf{E} \int_{B_j} |f_n - p_j f_{b_j}| \geq p_j n^{-2/5}/107$$

if $n_j \geq 24$ by Lemma 4. Pick b_j so that $\mathbf{E} \int |f_{n_j} - f_b| \geq p_j n_j^{-2/5}/107$. Continue in this fashion, and find $b = (b_1, b_2, \ldots)$. Then, at $n = n_j$, we have, setting $b^* = b$ in its first j components, but zero for the other components,

$$\begin{aligned}
\mathbf{E} \int |f_n - f_b| &\geq \mathbf{E} \left\{ I_E \int |f_n - f_{b^*}| \right\} \\
&= \mathbf{E} \left\{ \int |f_n - f_{b^*}| |E \right\} \mathbf{P}\{E\} \\
&\geq \frac{p_j}{p_1 + \cdots + p_j} n_j^{-2/5} \mathbf{P}\{E\}/107 \\
&\geq p_j n_j^{-2/5} \mathbf{P}\{E\}/107 \ .
\end{aligned}$$

Here the ratio comes from the observation that conditional on E, the sample may be thought to be drawn from mixture densities with only the first j components. Now, $\mathbf{P}\{E\} \geq 1 - n_j \sum_{k>j} p_k \geq 1/2$ by our choice of p_j and n_j. As a_j is decreasing, we will make $p_j n_j^{-2/5}/214$ greater than or equal to $n_j^{-2/5} a_{n_j}$ for all j. From the sequence a_n, pick a subsequence a_{n_j}, with the property that $\sum_j a_{n_j} < 1/214$, $a_{n_2} \leq 1/428$, and such that $a_{n_{j+1}} < a_{n_j}/(2n_j)$. Clearly, $\sum_{k>j} a_{n_k} < a_{n_j}/n_j$ for all j. Set $p_j = 214 a_{n_j}$ so that we have the desired inequality, and the p_j's sum to less than one. Give the remaining mass to p_1, to make a probability vector. Note that

$$n_j \sum_{k>j} p_k = n_j \sum_{k>j} 214 a_{n_j} \leq 214 a_{n_j} \leq \frac{1}{2}$$

for $j > 1$ by our choice of a subsequence. Thus, $\mathbf{P}\{E\} \geq 1/2$ as required. Also, $p_{j+1}/p_j = a_{n_{j+1}}/a_{n_j} \leq 1/2n_j \leq 1/48 < 1/14$ as required for monotonicity. Therefore, for our recursively constructed b, we have at all $j > 1$,

$$\mathbf{E} \int |f_{n_j} - f_b| \geq n_j^{-2/5} a_{n_j}.$$

This concludes the proof of part A of the Theorem. For part B, we use the same subsequence, but note an earlier bound for the denominator in the ratio and obtain for all $j > 1$,

$$\begin{aligned}
\frac{\mathbf{E} \int |f_{n_j, H_{n_j}(x)} - f_b|}{\inf_{h:\mathbb{R}\to(0,\infty)} \mathbf{E} \int |f_{n,h(x)}(x) - f_b(x)| dx} &\geq \frac{n_j^{-2/5} a_{n_j}}{\frac{\sqrt{\frac{28\rho}{3}} \sum_i \sqrt{p_i}}{\sqrt{n_j}}} \\
&= \frac{n_j^{1/10} a_{n_j}}{\sqrt{\frac{28\rho}{3}} \sum_i \sqrt{p_i}} \\
&\geq \frac{n_j^{1/10} a_{n_j}}{\sqrt{\rho a_{n_1}}} \times \frac{\sqrt{2} - 1}{\sqrt{28 \times 428/3}}
\end{aligned}$$

because

$$\begin{aligned}
\sum_i \sqrt{p_i} &= \sum_i \sqrt{214 a_{n_i}} \\
&\leq \sqrt{214} \left(\sqrt{a_{n_1}} + \sqrt{a_{n_1}/(2n_1)} + \sqrt{a_{n_1}/(2^2 n_1 n_2)} + \sqrt{a_{n_1}/(2^3 n_1 n_2 n_3)} + \cdots \right) \\
&\leq \sqrt{214 a_{n_1}} \sum_{i=0}^{\infty} \frac{1}{(\sqrt{2})^i} \\
&= \sqrt{428 a_{n_1}} \, \frac{1}{\sqrt{2} - 1} \, .
\end{aligned}$$

By replacing the a_n's by $\sqrt{a_n}$ at the outset of the proof, it is trivial to see that the ratio studied in part B of the Theorem is infinitely often greater than $a_n n^{1/10}$, as required. This concludes the proof of Theorem 2. □

4. Extensions and generalizations.

The proofs show that Lemma 1 and Theorems 1 and 2, with suitable changes of the constants, remain valid for piecewise convex densities, and in particular, for any piecewise linear density. However, for piecewise linear densities with a finite number of breakpoints, it may be possible to design a variable kernel density estimate with L1 error rate $O(1/\sqrt{n})$. Indeed, this is possible for the uniform density on $[0,1]$ and for trapezoidal densities. And since it is possible to estimate the breakpoints efficiently, so that the class of piecewise linear densities with $\leq k$ breakpoints is truly a parametric class.

In $\mathbb{R}^d$, if we use bounded support positive symmetric product kernels, the same methodology would work on the class of all convex densities on $[0,1]^d$ bounded by a constant B. There is no problem with the generalization of Lemma 1: the rate $O(1/\sqrt{n})$ would be achievable for the best $h(x)$ (assuming the same h is used in all dimensions). The minimax rate for the new class needs to be determined however. Clearly, it is going to be worse than in the one-dimensional case, so that even if we pick h identical in all directions (an easier problem than picking it separately for each dimension), a phenomenon similar to that described by Theorems 1 and 2 should occur.

A more subtle situation occurs when the kernel K is of order $2s$ for some positive integer s, that is, K is symmetric, $\int K = 1$, $\int x^i K(x)dx = 0$ for $1 \leq i < 2s$ and $\int x^{2s}K(x)dx = S \neq 0$. Even here, there is an analog to Theorem 1. The variance bound as described in Lemma 1 remains obviously valid uniformly over all bounded densities on $[0,1]$ that have for each $x \in (0,1)$ a bandwidth $h(x)$ that makes the bias zero. What one needs is the property that at every x, locally, $f * K_h$ increases initially when h increases from 0. By Taylor's series expansion with remainder, this is easy to establish if at every $x \in (0,1)$, f^{2s} is of the same sign as S, and if K vanishes off $[-1,1]$. The minimax lower bound over such densities on $[0,1]$ bounded by a constant B is of the order of $n^{-2s/(4s+1)}$ (we could not find a reference for this though), and therefore, the factor $n^{1/10}$ in Theorem 1 should be replaced by $n^{1/(8s+2)}$, which still tends to infinity, albeit more slowly.

5. Remarks on unbiasedness.

We showed that it is futile to look for the zero-bias choice $h(x)$ even for convex densities. Of course, we knew since Rosenblatt (1956) that no universally unbiased non-negative density estimate f_n exists:

$$\inf_{f_n : f_n \geq 0} \sup_f \int |\mathbf{E} f_n - f| > 0 \ .$$

On the other hand, for small classes $(\mathcal{F})$ of densities, unbiased density estimates exist:

$$\inf_{f_n} \sup_{f \in \mathcal{F}} \int |\mathbf{E} f_n - f| = 0 \ .$$

An example is the class of all normal densities with unknown mean and variance, for which an unbiased estimate was found by Basu (1964)—see also exercise 7.14 of Devroye (1987). A second example is the class $A_{T,s,C}$ of Devroye and Györfi (1985, page 142), which roughly speaking is a subclass of all densities with bounded support characteristic function and absolute s-th integrated derivative of the characteristic function bounded by C. Here the unbiased estimate is the ordinary kernel estimate with a superkernel and a bandwidth less than a constant depending upon T only. For more information on unbiasedness, see Lumelskii and Sapozhnikov (1969), Wertz (1975), Guttmann and Wertz (1976), and Seheult and Quesenberry (1971).

One may ask then where the boundary is for $\mathcal{F}$? Which classes are too large to find unbiased density estimates for all members in the class? We offer the following result, which relates the non-existence of unbiased estimates to the richness of the class of densities under consideration.

THEOREM 3. *Let $\mathcal{F}$ be a class of uniformly bounded densities, and let $R_m(\mathcal{F})$ denote its minimax risk for sample size m:*

$$R_m(\mathcal{F}) = \inf_{f_m} \sup_{f \in \mathcal{F}} \mathbf{E} \int |f_m - f| \ .$$

Let f_n be a density estimate, where n is a fixed integer. If $\lim_{m \to \infty} \sqrt{m} R_m(\mathcal{F}) = \infty$, then either f_n is not unbiased ($\sup_{f \in \mathcal{F}} \int |\mathbf{E} f_n - f| > 0$) or

$$\sup_{f \in \mathcal{F}} \int \sqrt{\mathbf{E}\{f_n^2\}} = \infty.$$

Let us illustrate this Theorem. The fact that the minimax risk increases faster than $1/\sqrt{n}$ usually is accepted as an indication that $\mathcal{F}$ is "nonparametric". An example of a rich class is the class of all densities on $[0,1]$ with 25 continuous derivatives on the real line, each of which is bounded in absolute value by B for a sufficiently large constant B. With the knowledge that f is in this class, one would be tempted to construct density estimates bounded by B. But Theorem 3 then says that f_n cannot be unbiased for all f in the class! The price to pay for unbiasedness is unboundedness in expectation as in the last condition of Theorem 3. In particular, for such rich nonparametric classes, no bounded and compactly supported unbiased density estimate exists, even if we know a uniform bound on the densities and the support in the class.

While Theorem 3 does not supersede Rosenblatt's result, it complements it by addressing the question of the size of the classes. The two examples cited earlier of course had minimax risks of the order of $1/\sqrt{n}$. Note also that Theorem 3 does not say that $\int |f_n| = \infty$ with probability one: indeed, we could have $\int |f_n| < \infty$ with probability one.

PROOF. Assume that f_n is unbiased for all $f \in \mathcal{F}$. We then construct the following density estimate for sample size mn:

$$f_{mn}(x) = \frac{1}{m} \sum_{j=1}^{m} f_n(x; X_{(j-1)n+1}, \ldots, X_{(j-1)n+n})$$

which is unbiased, and a sum of n independent summands. Therefore,

$$\mathrm{Var}\{f_{mn}(x)\} = \frac{1}{m} \mathrm{Var}\{f_n(x)\}.$$

But then

$$\begin{aligned} \int \mathbf{E}\{|f_{nm} - f|\} &= \int \mathbf{E}\{|f_{nm} - \mathbf{E} f_{nm}|\} \\ &\leq \int \sqrt{\mathrm{Var}\{f_{nm} - \mathbf{E} f_{nm}\}} \\ &= \frac{1}{\sqrt{m}} \int \sqrt{\mathrm{Var}\{f_n\}}\ . \end{aligned}$$

Taking supremums shows that

$$\sup_{f \in \mathcal{F}} \sqrt{nm} \mathbf{E} \int |f_{nm} - f| \leq \sup_{f \in \mathcal{F}} \frac{\sqrt{nm}}{\sqrt{m}} \int \sqrt{\mathrm{Var}\{f_n\}}.$$

as $m \to \infty$, the left-hand side tends to ∞ by assumption. As n is fixed, this shows that

$$\sup_{f \in \mathcal{F}} \int \sqrt{\mathrm{Var}\{f_n\}} = \infty.$$

By the uniform boundeness of f, this implies that

$$\sup_{f \in \mathcal{F}} \int \sqrt{\mathbf{E}\{f_n^2\}} = \infty. \ \square$$

6. References

I. Abramson, "On bandwidth variation in kernel estimates—a square root law," *Annals of Statistics*, vol. 10, pp. 1217–1223, 1982.

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

P. Assouad, "Deux remarques sur l'estimation," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 296, pp. 1021–1024, 1983.

A. P. Basu, "Estimates of reliability for some distribution useful in life testing," *Technometrics*, vol. 6, pp. 215–219, 1964.

L. Birgé, "Non-asymptotic minimax risk for Hellinger balls," *Probability and Mathematical Statistics*, vol. 5, pp. 21–29, 1985.

L. Birgé, "On estimating a density using Hellinger distance and some other strange facts," *Probability Theory and Related Fields*, vol. 71, pp. 271–291, 1986.

L. Birgé, "On the risk of histograms for estimating decreasing densities," *Annals of Statistics*, vol. 15, pp. 1013–1022, 1987a.

L. Birgé, "Estimating a density under order restrictions: nonasymptotic minimax risk," *Annals of Statistics*, vol. 15, pp. 995–1012, 1987b.

L. Birgé, "The Grenander estimator: a nonasymptotic approach," *Annals of Statistics*, vol. 17, pp. 1532–1549, 1989.

L. Breiman, W. Meisel, and E. Purcell, "Variable kernel estimates of multivariate densities," *Technometrics*, vol. 19, pp. 135–144, 1977.

L. Devroye, "A note on the L1 consistency of variable kernel estimates," *Annals of Statistics*, vol. 13, pp. 1041–1049, 1985.

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

L. Devroye, "Another proof of a slow convergence result of Birgé," *Statistics and Probability Letters*, vol. 23, pp. 63–67, 1995.

L. Devroye, "Universal smoothing factor selection in density estimation: theory and practice," *Test*, vol. 6, pp. 223–320, 1997.

L. Devroye and G. Lugosi, "A universally acceptable smoothing factor for kernel density estimation," *Annals of Statistics*, vol. 24, pp. 2499–2512, 1996.

L. Devroye and G. Lugosi, "Non-asymptotic universal smoothing factors, kernel complexity and Yatracos classes," *Annals of Statistics*, vol. 25, pp. 2626-2637, 1997.

L. Devroye, G. Lugosi, and F. Udina, "Inequalities for a new data-based method for selecting nonparametric density estimates," Technical Report, Facultat de Ciencies Economiques, Universitat Pompeu Fabra, Barcelona, 1998.

L. Devroye and C. S. Penrod, "The strong uniform convergence of multivariate variable kernel estimates," *Canadian Journal of Statistics*, vol. 14, pp. 211–219, 1986.

M. Farmen, "The smoothed bootstrap for variable bandwidth selection and some results in nonparametric logistic regression" Ph.D. Dissertation, Department of Statistics, University of North Carolina, Chapel Hill, 1996.

H. Guttmann and W. Wertz, "Note on estimating normal densities," *Sankhya, Series B*, vol. 38, pp. 231–236, 1976.

J. D. F. Habbema, J. Hermans, and J. Remme, "Variable kernel density estimation in discriminant analysis," in: *COMPSTAT 1978: Proceedings*, ed. L. C. A. Corsten and J. Hermans, pp. 0–0, Birkhauser, Basel, 1978.

P. Hall, "On global properties of variable bandwidth density estimators," *Annals of Statistics*, vol. 20, pp. 762–778, 1992.

P. Hall and J. S. Marron, "Variable window width kernel estimates," *Probability Theory and related Fields*, vol. 80, pp. 37–49, 1988.

P. Hall and W. R. Schucany, "A local cross-validation algorithm," *Statistics and Probability Letters*, vol. 8, pp. 109–117, 1989.

M. Hazelton, "Bandwidth selection for local density estimation," *Scandinavian Journal of Statistics*, vol. 23, pp. 221–232, 1996.

M. C. Jones, "Variable kernel density estimates and variable kernel density estimates," *Australian Journal of Statistics*, vol. 32, pp. 361–371, 1990.

Ya. P. Lumelskii and P. N. Sapozhnikov, "Unbiased estimates of density functions," *Theory of Probability and its Applications*, vol. 14, pp. 357–364, 1969.

J. S. Marron, P. Hall, and T. C. Hu, "Improved variable window estimators of probability densities ," *Annals of Statistics*, vol. 23, pp. 1–10, 1995.

J. Mielniczuk, P. Sarda, and P. Vieu, "Local data-driven bandwidth choice for density estimation," *Journal of Statistical Planning and Inference*, vol. 23, pp. 53–69, 1989.

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

J. W. Raatgever and R. P. W. Duin, "On the variable kernel model for multivariate nonparametric density estimation," in: *COMPSTAT 1978: Proceedings*, ed. L. C. A. Corsten and J. Hermans, pp. 0–0, Birkhauser, Basel, 1978.

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

S. R. Sain, "Adaptive kernel density estimation," Ph.D. Dissertation, Department of Statistics, Rice University, Houston, 1994.

S. R. Sain and D. W. Scott, "On locally adaptive density estimation," *Journal of the American Statistical Association*, vol. 91, pp. 1525–1534, 1996.

S. R. Sain and D. W. Scott, "Zero-bias locally adaptive density estimators," Technical Report, Rice University, Houston, 1997.

A. H. Seheult and C. P. Quesenberry, "On unbiased estimation of density functions," *Annals of Mathematical Statistics*, vol. 42, pp. 1434–1438, 1971.

S. J. Sheather, "A data-based algorithm for choosing the window width when estimating the density at a point," *Computational Statistics and Data Analysis*, vol. 1, pp. 229–239, 1983.

S. J. Sheather, "An improved data-based algorithm for choosing the window width when estimating the density at a point," *Computational Statistics and Data Analysis*, vol. 4, pp. 61–65, 1986.

G. R. Terrell and D. W. Scott, "Variable kernel density estimation," *Annals of Statistics*, vol. 20, pp. 1236–1265, 1992.

L. A. Thombs and S. J. Sheather, "Local bandwidth selection for density estimation," in: *Proceedings of the 22nd Symposium on the Interface*, pp. 111–116, Springer-Verlag, New York, 1992.

W. Wertz, "On unbiased density estimation," *An. Acad. Brasil. Cienc.*, vol. 47, pp. 65–72, 1975.

Y. Yang, "Minimax Optimal Density Estimation," Ph.D. Dissertation, Yale University, 1996.