The double kernel method in density estimation


<http://www.numdam.org/item?id=AIHPB_1989__25_4_533_0>
The double kernel method in density estimation

by

Luc DEVROYE

School of Computer Science,
McGill University

ABSTRACT. — Let $f_{nh}$ be the Parzen-Rosenblatt kernel estimate of a density $f$ on the real line, based upon a sample of $n$ iid random variables drawn from $f$, and with smoothing factor $h$. Let $g_{nh}$ be another kernel estimate based upon the same data, but with a different kernel. We choose the smoothing factor $H$ so as to minimize $\int |f_{nh} - g_{nh}|$, and study the properties of $f_{nH}$ and $g_{nH}$. It is shown that the estimates are consistent for all densities provided that the characteristic functions of the two kernels do not coincide in an open neighborhood of the origin. Also, for some pairs of kernels, and all densities in the saturation class of the first kernel, we show that

$$\limsup_{n \to \infty} \inf_{h} \frac{\mathbb{E} \left( \int |f_{nH} - f| \right)}{\mathbb{E} \left( \int |f_{nh} - f| \right)} \leq C,$$

where $C$ is a constant depending upon the pair of kernels only. This constant can be arbitrarily close to one.

Key words: Density estimation, asymptotic optimality, nonparametric estimation, strong convergence, kernel estimate, automatic choice of the smoothing factor.

RESUMÉ. — Soit $f_{nh}$ l’estimateur du noyau de Parzen-Rosenblatt d’une densité $f$ sur la droite réelle, à partir d’un échantillon de $n$ variables...
aléatoires indépendantes équidistribuées de densité $f$ et avec facteur régularisant $h$. Soit $g_{nh}$ un autre estimateur de noyaux basé sur les mêmes données mais avec un noyau différent, nous choisissons le facteur régularisant $H$ afin de minimiser $\int |f_{nh} - g_{nh}|$, et nous étudions les propriétés de $f_{nH}$ et $g_{nH}$. Nous montrons que les estimateurs sont conséquents pour toutes les densités pourvu que les fonctions caractéristiques des deux noyaux ne coïncident pas sur un voisinage ouvert de l’origine. De plus, pour certains couples de noyaux, pour toutes les densités de la classe de saturation du premier noyau nous montrons que

$$\limsup_{n \to \infty} \frac{E\left(\int |f_{nh} - f|\right)}{\inf_h E\left(\int |f_{nh} - f|\right)} \leq C,$$

où $C$ est une constante qui dépend seulement du couple de noyaux. Cette constante peut être rendue arbitrairement proche de 1.

---

**TABLE OF CONTENTS**

1. Introduction.
2. Consistency.
   2.1 The purpose.
   2.2 The decoupling device.
   2.3 Uniform convergence with respect to $h$.
   2.4 Behavior of the minimizing integral.
   2.5 Necessary conditions of convergence.
   2.6 Proof of Theorem Cl.
3. C-optimality.
   3.1 The main result.
   3.2 A better estimate.
   3.3 Complete convergence of the L1 error.
   3.4 Relative stability of the estimate.
   3.5 Proof of Theorems O1 and O2.
   3.6 Remarks and further work.
   3.7 The final series of Lemmas.
4. Properties of the optimal smoothing factor.
1. INTRODUCTION

We consider the standard problem of estimating a density \( f \) on \( \mathbb{R}^1 \) from an iid sample \( X_1, \ldots, X_n \) drawn from \( f \). The density estimates considered in this note are the well-known kernel estimates

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

where \( h > 0 \) is a smoothing factor, \( K \) is an absolutely integrable function called the kernel, \( \int K = 1 \), and \( K_h(x) = (1/h) K(x/h) \) (Parzen, 1962; Rosenblatt, 1956). We are particularly interested in smoothing factors that are functions of the data (which are denoted by \( H \) to reflect that they are random variables). Most proposals for functions \( H \) found in the literature minimize some criterion; for example, many attempt to keep \( \int (f_n H - f)^2 \) as small as possible. Extending Stone (1984), we say that \( f_n \) is asymptotically optimal for \( f \) if \( H \) is such that

\[
\frac{\mathbb{E} \left( \int |f_{nH} - f| \right)}{\inf_{h} \mathbb{E} \left( |f_{nH} - f| \right)} \to 1
\]

as \( n \to \infty \). Stone defined this notion with \( L_2 \) errors instead of \( L_1 \) errors, without the expected values, and with almost sure convergence to one for the ratio. In Theorem 5.1, we will show that in general \( \mathbb{E} \inf_{h} \) and \( \inf_{h} \mathbb{E} \) can be used interchangeably, and that

\[
\int |f_{nH} - f| / \mathbb{E} \int |f_{nH} - f| \to 1 \text{ almost surely, so that all definitions are equivalent for our method.}
\]

It should be noted that \( f_n \) may be asymptotically optimal for some \( f \) and \( K \) and not for other choices. A perhaps too trivial example is that in which \( K \) is the uniform density on \( [-1, 1] \), and \( H = cn^{-1/5} \) where \( c \) is known to be optimal for the normal \((0,1)\) density and the given \( K \). Obviously, such a choice leads in general to asymptotic optimality for the normal \((0,1)\) density, and to suboptimality in nearly all other cases. It is
clearly of interest to the practitioner to make the class of densities on which asymptotic optimality is obtained as large as possible. Remarkably, in the $L_2$ work of Stone (1984), asymptotic optimality was established for all bounded densities and all bounded kernels of compact support if $H$ is chosen by the $L_2$ cross-validation method of Rudemo (1982) and Bowman (1984). This $H$ is of little use in $L_1$, and the method is dangerous: for some unbounded densities, we have $\lim \inf E \left( \int \left| f_{nH} - f \right| \right) \geq 1$ (Devroye, 1988). Since data-based techniques for choosing $H$ are supposed to be automated and inserted into software packages, it is important that the method be consistent.

It is perhaps useful to reflect on the possible strategies. Hall and Wand (1988) have proposed a plug-in adaptive method, in which unknown quantities in the theoretical formula for the asymptotically optimal $h$ are estimated from the data using other nonparametric estimates, and then plugged back in the formula to obtain $H$. Similar strategies have worked in the past for $L_2$ [see e.g. Woodroofe (1970), Nadaraya (1974) and Bretagnolle and Huber (1979)]. The advantages of this approach are obvious: the designer clearly understands what is going on, and the problem is conceptually cut in clearly identifiable subproblems. On the other hand, how does one choose the smoothing parameters needed for the secondary nonparametric estimates? And, assuming that the conditions for the theoretical formula for $h$ are not fulfilled, isn't it possible to obtain inconsistent density estimates? To avoid the latter drawback, it is imperative to go back to first principles. Cross-validated maximum likelihood products have been studied by many: Duin (1976) and Habema, Hermans and Vandenbroek (1974) proposed the method, and Chow, Geman and Wu (1983), Schuster and Gregory (1981), Hall (1982), Devroye and Györfi (1985), Marron (1985) and Broniatowski, Deheuvels and Devroye (1988) studied the consistency and rate of convergence. Unfortunately, the maximum likelihood methods for choosing $h$ pertain to the Kullback-Leibler distances between densities, and bear little relation to the $L_1$ criterion under investigation here. The $L_2$ cross-validation method proposed by Rudemo (1982) and Bowman (1984) has no straightforward extension to $L_1$. Its properties in $L_2$ are now well understood, see e.g. Hall (1983, 1985), Stone (1984), Burman (1985), Scott and Terrell (1987) and Hall and Marron (1987). This seems to leave us empty-handed were it not for the versatility of the kernel estimate itself. Indeed, the method we are about to propose does not easily generalize beyond the class of kernel estimates.
The estimator proposed below has two advantages:

A. It is consistent for all \( f \), i.e. \( \lim_{n \to \infty} \mathbb{E} \left( \int |f_{nH} - f| \right) = 0 \) for all \( f \).

B. For a large family of "nice" densities, we have C-optimality, i.e. there exists a constant \( C \) such that for all \( f \) in the class,

\[
\lim_{n \to \infty} \sup_h \frac{\mathbb{E} \left( \int |f_{nH} - f| \right)}{\inf_h \mathbb{E} \left( \int |f_{nh} - f| \right)} \leq C.
\]

The constant \( C \) can be as close to one as desired. We define our \( H \) simply as the \( h \) that minimizes \( \int |f_{nh} - g_{nh}| \), where \( g_{nh} \) is the kernel estimate based upon the same data, but with kernel \( L \) instead of kernel \( K \). The key idea is that most kernels considered in practice have built-in limitations, including the class of all kernels with compact support. For any such kernel \( K \), it is fairly easy to construct another kernel \( L \) whose bias is asymptotically superior in the sense that

\[
\lim_{h \to 0} \frac{\int |f \ast L_h - f|}{\int |f \ast K_h - f|} = 0,
\]

where \( \ast \) is the convolution operator. The class of densities \( f \) for which this happens coincides roughly speaking with the class of densities for which \( \int |f \ast K_h - f| \) tends to zero at the best possible rate (or: saturation rate) for the given \( K \). These classes are rich, but they won't satisfy everyone. What the improved kernel can do for us is simple: it is very likely that \( g_{nh} \), the kernel estimate with \( K \) replaced by \( L \), is much closer to \( f \) than \( f_{nh} \), and thus that \( \int |f_{nh} - g_{nh}| \) is of the order of magnitude of \( \int |f_{nh} - f| \).

We won't worry here about the numerical details. First of all, if \( K \) and \( L \) are polynomial and of compact support (as they often are), then the integral to be minimized can be rewritten conveniently as finite sum with \( O(n) \) terms, by considering that each kernel estimate is piecewise polynomial with \( O(n) \) pieces at most. The minimization with respect to \( h \) is a bit harder to do. Observing that the function to be optimized is uniformly

Vol. 25, n° 4-1989.
continuous on any interval \((a, b) \subseteq [0, \infty)\), we see that the minimum exists and is a random variable. For more general non-polynomial \(K\) and \(L\), under some smoothness conditions, we still have
\[
\left| \int |f_{nh} - f| - \int |f_{nh'} - f| \right| \leq c \frac{|h - h'|}{h} \tag{2}
\]
for some \(c > 0\). For \(h, h'\) close enough, this can be made smaller than \(1/n\), which is much smaller than the smallest possible \(L_1\) error (which is \(1/\sqrt{528n}\) by Devroye, 1988). Thus, the minimization can be carried out over a grid of points, and in any case, it is possible to define a random variable \(H\) with the property that
\[
\int |f_{nH} - g_{nH}| \sim \inf_h \int |f_{nh} - g_{nh}|.
\]

There is another interesting by-product of this method, namely that we end up with two kernel estimates \(f_{nH}\) and \(g_{nH}\), where for the class of densities under consideration, \(g_{nH}\) is probably better than \(f_{nH}\). Interestingly, \(f_{nH}\) is asymptotically optimal for \(K\) but \(g_{nH}\) is usually \textit{not} asymptotically optimal for \(L\). One way of looking at our method is as a technique for creating a better estimate \((g_{nh})\) \textit{without} imposing additional smoothness conditions on the densities. Another by product of the method is that
\[
\int |f_{nH} - g_{nH}| \tag{3}
\]
is a rough estimate of the actual error \(\int |f_{nH} - f|\). Unfortunately, if one decides to use \(g_{nH}\) instead of \(f_{nH}\), then \(\int |f_{nH} - g_{nH}|\) provides little information about the actual error obtained with \(g_{nH}\).

Not all pairs \((K, L)\) are useful. The most important property needed to be fulfilled is that the characteristic functions of \(K\) and \(L\) do not coincide in an open neighborhood of the origin. This often forces \(K\) and \(L\) to be kernels of a different order. In addition, we will see that the constant \(C\) can be chosen equal to \((1 + u)/(1 - u)\), where \(u = 4 \sqrt{L^2 / K^2}\).

The length of the paper is partially explained by the fact that we wanted to state as many properties of the estimate as possible in a “density-free” manner. This also renders the results more useful for future work on the same topic. Among the density-free results, we cite:

The consistency (Theorem C1).
The complete convergence (Lemma O2).
The strong relative stability of the estimate (Lemma O5).
Bounds on the error that are uniform over all \(h\) (Theorem C2 and Lemma O1).

Necessary conditions of convergence (Lemmas C1 and C2).
A universal lower bound for the expected error (Lemma O5).
Universal lower bounds for the variation in the error (Lemmas O7, O8).
While it is good to know that the estimate always converges and is C-optimal for virtually all densities in the saturation class of a kernel $K$, it is informative to find out what we have not been able to achieve. First of all, the kernels $K$ considered here for C-optimality are class $s$ kernels, i.e. all their moments up to but not including the $s$-th moment vanish. This implies, as we will see, that the expected error can go to zero no faster than a constant times $n^{-s/(2s+1)}$ for any density. In this respect, we are severely limited, since it is well-known that for very smooth densities kernels can be found that yield error rates that are $O(1/\sqrt{n})$ or come close to it [see e.g. Watson and Leadbetter (1963) for an equivalent statement in the $L_2$ setting]. We can exhibit constants $C$ and $D$ such that for $n$ large enough,

$$
E\left(\int |f_{nH} - f|\right) \leq C \inf_h E\left(\int |f_{nh} - f|\right) + D \frac{\log n}{n}.
$$

This inequality implies that C-optimality can only be hoped for when the best possible error rate for the present $K$ and $f$ is at least $\sqrt{\log n/n}$. Unfortunately, this would exclude such interesting densities as the normal density, for which we can get $O(\log^{1/4}n/\sqrt{n})$ (Devroye, 1988). In particular, it seems that for analytic densities in general, the techniques presented here need some strengthening.

But perhaps the biggest untackled question is what happens to the expected error for densities $f$ that are not in the saturation class of $K$; these are usually densities that are not smooth enough or not small-tailed enough to attain the rate $n^{-s/(2s+1)}$. Despite this, it may still be possible to apply the present minimization technique to obtain good asymptotic performance for most of them. All that is needed is to verify the fact that for the pair $(K, L)$, $\int |f \ast L_h - f| = o\left(\int |f \ast K_h - f|\right)$. On the other hand, it is also possible that a general result such as the one obtained by Stone (1984) for $L_2$ errors does not exist in the $L_1$ setting.

2. CONSISTENCY

2.1. The purpose

The purpose of this section is to prove the following:

**Theorem C1.** — Let $K$ and $L$ be absolutely integrable kernels such that their (generalized) characteristic functions do not coincide on any open
neighborhood of the origin, and let \( f \) be an arbitrary density. Then
\[
\mathbb{E} \left| f_{nH} - f \right| \quad \text{and} \quad \mathbb{E} \left| g_{nH} - f \right|
\]
tend to zero as \( n \to \infty \).

One of the difficulties with this sort of Theorem is that it needs to be shown for all densities \( f \), even those \( f \) for which the procedure for selecting \( H \) is not specifically designed. Furthermore, the \( L_1 \) errors are not easily decomposed into bias and variance terms, since \( H \) depends upon the data, so that conditional on \( H \), the summands in the definition of the kernel estimate are not independent. We will provide a mechanism for "decoupling" \( H \) and the data. In the final analysis, the proof of Theorem 1 rests on an exponential inequality of Devroye (1988) and some other properties of the random function \( \int |f_{nh} - f| \) (considered as a function of \( h \)). The proof will be cut into many lemmas, some of which will be useful outside this paper and in other sections.

The condition on \( K \) and \( L \) implies that \( \int |K - L| > 0 \). It is possible to have consistency even if the characteristic functions of \( K \) and \( L \) coincide on some open neighborhood of the origin, but such consistency would not be universal; it would apply only to densities whose characteristic function vanishes off a compact set. The details for such cases can be deduced from the proof. We have also unveiled where it is possible to go wrong: it suffices to have the said coincidence of the two characteristic functions, while \( f \) has a characteristic function with an infinite tail. In those cases, the \( H \) may actually end up tending to a positive constant as \( n \to \infty \). Unfortunately, as is well known, for such densities, it is impossible to have consistency unless \( H \to 0 \). From this, we retain that the behavior of the characteristic function of \( K - L \) near the origin is somehow a measure of the discriminatory power of the method. Usually, we take a standard nonnegative kernel for \( K \) whose characteristic function varies as \( 1 - at^2 \) near \( t = 0 \), whereas for \( L \) we can take a kernel whose characteristic function is flatter near the origin, behaving possibly as \( 1 - bt^4 \) or even identically 1 on an open neighborhood of the origin.

### 2.2. The decoupling device

We have seen that \( X_1, \ldots, X_n \) are iid random variables with density \( f \), and that \( H = H(n) \) is a sequence of random variables where \( H(n) \) is measurable with respect to the \( \sigma \)-algebra generated by \( X_1, \ldots, X_n \). i. e. it is a function of \( X_1, \ldots, X_n \). Consider now independent identically distributed copies of the two sequences, denoted by \( \tilde{X}_1, \ldots, \tilde{X}_n \) and \( \tilde{H} = \tilde{H}(n) \) respectively. Density estimates based upon the former data are denoted...
by \( f_{nh}, g_{nh}, f_{nH}, \) and \( g_{nH} \) typically, while for the latter data, we will write \( \tilde{f}_{nh} \) and so forth.

In our decoupling, we will show that \( \int |f_{nH} - f| \) and \( \int |\tilde{f}_{nH} - f| \) are close in a very strong sense. Note that the second error is that committed if \( H \) is used in a density estimate constructed with a new data set. The independence thus introduced will make the ensuing analysis more manageable. To keep the notation simple, we will write \( E_n \) for the conditional expectation given \( X_1, \ldots, X_n \), and \( \bar{E}_n \), for the conditional expectation given \( X_1, \ldots, X_n \). With this notation, note that \( E_n \int |\tilde{f}_{nH} - f| \) is distributed as \( \bar{E}_n \int |f_{nH} - f| \), and thus that \( E \int |\tilde{f}_{nH} - f| = E \int |f_{nH} - f| \).

2.3. Uniform convergence with respect to \( h \)

The first auxiliary result is so crucial that we are permitting ourselves to elevate it to a Theorem:

**THEOREM C2.** Let \( M \) be an arbitrary absolutely integrable function, and define \( m_{nh} = \frac{1}{n} \sum_{i=1}^{n} M_h(x - X_i) \) where \( X_1, \ldots, X_n \) are iid random variables with an arbitrary density \( f \), and \( h > 0 \) is a real number. Then

\[
\sup_h \left| \int |m_{nh}| - E \int |m_{nh}| \right| \to 0 \quad \text{almost surely as } n \to \infty.
\]

In fact, for every \( \varepsilon > 0 \), there exists a constant \( \gamma > 0 \) possibly depending upon \( f, M \) and \( \varepsilon \), such that for all \( n \) large enough,

\[
P\left( \sup_h \left| \int |m_{nh}| - E \int |m_{nh}| \right| > \varepsilon \right) \leq e^{-\gamma n}.
\]

To see how Theorem C2 exactly provides us with the required decoupling between \( H \) and the data, consider the following

**COROLLARIES OF THEOREM C2.** Let \( f_{nh}, g_{nh} \) be kernel estimates with kernels \( K \) and \( L \) respectively, and define \( J_{nh} = \int |f_{nh} - g_{nh}| \) and

\[
J_{nh} = \int |\tilde{f}_{nh} - \tilde{g}_{nh}|.
\]

Then \( A. \sup_{h > 0} |J_{nh} - EJ_{nh}| \to 0 \) almost surely as \( n \to \infty \).
B. For any random variable $H$ (possibly not independent of the data), $J_{nH} - E_nJ_{nH} \to 0$ almost surely as $n \to \infty$.

C. For any random variable $H$ (possibly not independent of the data), $J_{nH} \to 0$ in probability implies $EJ_{nH} \to 0$, $E_j J_{nH} \to 0$ in probability, $E_j J_{nH} \to 0$, and $EJ_{nH} \to 0$.

Proof of Theorem C2. – Note first that 
\[ \int |m_{nh} - u_{nh}| \leq \int |M - M'| \]
when $u_{nh}$ is the kernel estimate with kernel $M'$. The fact that the bound does not depend upon $h$ and that $M'$ is arbitrary means that we need only show the Theorem for all $M$ that are continuous and of compact support (since the latter collection is dense in the space of $L_1$ functions).

The first auxiliary result is the following inequality, valid for all fixed $h$, $n$, $M$ and $f$:
\[
P\left( \int |m_{nh} - E m_{nh}| > \varepsilon \right) \leq 2 e^{- (n^2 M') / 2} \int |M|
\]
(Devroye, 1988). It is this inequality that will be extended to an interval of $h$'s using a rather standard grid technique. Set \[ \Delta(h) = \int |m_{nh} - E m_{nh}|, \]
and \[ \Delta(a, b) = \sup_{h, \ h' \in [a, b]} |\Delta(h) - \Delta(h')|. \]
Then the following inclusion of events is valid:
\[
\left\{ \sup_{a \leq h \leq b} \Delta(h) > \varepsilon \right\} \subseteq \bigcup_{i=0}^{k} \left\{ \Delta(ac^i) > \varepsilon / 2 \right\} \bigcup_{i=1}^{k} \left\{ \Delta(ac^{i-1}, ac^i) > \varepsilon / 2 \right\}
\]
where $k$ is an integer so large that $ac^k \geq b$, and $c > 1$ is such that
\[
\sup_{1 \leq h \leq \varepsilon} \int |M_1 - M_h| \leq \frac{\varepsilon}{4}.
\]
Such a $c$ can indeed be found, since for all absolutely integrable $M$,
\[
\lim_{h \to 1} \int |M_1 - M_h| = 0 \ (\text{see e.g. Devroye, 1987, pp. 38-39}).
\]
The second union in the inclusion inequality is a union of empty events since
\[
\Delta(ac^{i-1}, ac^i) \leq \sup_{ac^{i-1} \leq h, h' < ac^i} \left| \int m_{nh} - \int m_{nh'} \right| + \sup_{ac^{i-1} \leq h, h' < ac^i} E \left| m_{nh} - E \int m_{nh'} \right| \leq 2 \sup_{ac^{i-1} \leq h, h' < ac^i} \int |M_h - M_{h'}| = 2 \sup_{1 \leq h \leq \varepsilon} \int |M_1 - M_h| < \frac{\varepsilon}{2}.
\]
Thus,
\[ P\left( \sup_{a \leq h \leq b} \Delta(h) > \varepsilon \right) \leq \sum_{i=0}^{k} P\left( \Delta(ac^i) > \frac{\varepsilon}{2} \right) \leq 2(k+1)e^{-\frac{(nx^2/128)}{2|\mathcal{M}|}}. \]

This tends to 0 when \( k \) increases at a polynomial rate in \( n \). To see how large \( k \) just is, note that \( ac^k \geq b \), so that \( k \geq \log(b/a)/\log(c) \). Since \( c \) is a constant depending upon \( \varepsilon \) and \( M \) only, it suffices to have limits \( a \) and \( b \) that are such that \( b/a \leq d^n \) for some constant \( d \). We will only need the present inequality with \( b/a = O(n) \), so that \( k \) can be taken smaller than \( d + \log(n) \) for some constant \( d \). Recapping, we need only establish that for some sequences \( a = a(n) \) and \( b = b(n) \) with \( b/a = O(n) \) that
\[ \lim_{n \to \infty} \left( P\left( \sup_{h < a} \Delta(h) > \varepsilon \right) + P\left( \sup_{h > b} \Delta(h) > \varepsilon \right) \right) = 0. \]

Assume that \( M \) vanishes off \([-1, 1]\). Take \( a = \frac{\delta}{2n} \) where \( \delta > 0 \) is a constant to be picked further on. Note that
\[ 1 \geq \int |m_{nh}| - \int |M| \geq \frac{n-N}{n} \]
where \( N \) is the number of \( X_i \)'s for which \([X_i - 2a, X_i + 2a]\) has at least one \( X_j \) with \( j \neq i \). The inequality is uniform over \( h \leq a \). We have
\[
P\left( \sup_{h < a} \Delta(h) > \varepsilon \right) \leq P\left( \sup_{h < a} \left| \int d_u \right| - \int |M| \right) > \frac{\varepsilon}{2} \right) 
= P\left( \sum_{i=0}^{k} P\left( \Delta(ac^i) > \frac{\varepsilon}{2} \right) \leq 2(k+1)e^{-\frac{(nx^2/128)}{2|\mathcal{M}|}}. \]

By Markov's inequality, this can be made smaller than a given small constant \( \varepsilon' \) if
\[ \frac{EN}{n} \leq \min\left( 1, \varepsilon' \right) \frac{\varepsilon}{2 \int |M|}. \]

But \( EN/n \leq \int f(x) \min\left( 1, n \int _{-\delta/n}^{+\delta/n} f(y) dy \right) dx \) and the right-hand-side tends to \( \int f(x) \min\left( 1, 2 \delta f(x) \right) dx \) by the Lebesgue dominated convergence theorem and the Lebesgue density theorem (see e.g. Wheeden and Zygmund, 1977). It can be made as small as desired by our choice of \( \delta \). We
conclude that for any $\varepsilon, \varepsilon' > 0$, we can find $\delta > 0$ small enough such that

$$\limsup_{n \to \infty} \mathbb{P}(\sup_{h < a} \Delta(h) > \varepsilon) \leq \varepsilon'.$$

We should note here that if Markov's inequality is replaced by an exponential bounding method, then it should be obvious that for $\delta$ small enough the probability in question can be bounded by $e^{-dn}$ for some constant $d > 0$.

We finally proceed to show that

$$\mathbb{P}(\sup_{h > b} \Delta(h) > \varepsilon)$$

can be made arbitrarily small by choosing a large enough constant $b$. This would conclude the proof of the Theorem since $b/a = O(n)$ as required.

We have

$$\int \left| m_{nh} \right| \geq \frac{n-N}{n} \int_{|x| \leq Th} |M_h(x)| \, dx$$

$$- \int \sup_{|x| \leq Th, |y| \leq T} |M_h(x) - M_h(x-y)| \, dx - \frac{N}{n} \int |M|,$$

where $T$ is a large constant, and $N$ is the number of $X_i$'s with $|X_i| > T$. Let $\omega$ be the modulus of continuity of $M$ defined by $\omega(u) = \sup_{x, y} |M(x) - M(x+y)|$. By our assumptions on $M$, $\omega(u) \to 0$ as $u \downarrow 0$. Then

$$\int \sup_{|x| \leq Th, |y| \leq T} |M_h(x) - M_h(x-y)| \, dx$$

$$\leq \int \frac{1}{h} \omega \left( \frac{T}{h} \right) \, dx \leq 2T \omega \left( \frac{T}{b} \right) \leq 2T \omega \left( \frac{T}{b} \right),$$

when $h \geq b$. Furthermore, by Hoeffding's inequality (Hoeffding, 1963),

$$\mathbb{P} \left( \frac{N}{n} > 2 \int_{|y| > T} f \right) \leq e^{-2n \int_{|y| > T} f}$$

so that, combining all this,

$$\mathbb{P} \left( \sup_{h \geq b} \left| \int \left| m_{nh} \right| - \int |M| \right| > 2T \omega \left( \frac{T}{b} \right) + 4 \int |M| \int_{|y| \geq T} f \right) \leq e^{-2n \int_{|y| > T} f}.$$
From this, we have since \( \int |m_{nh}| \leq \int |M|, \)

\[
\sup_{h \geq b} \left| E \int |m_{nh}| - \int |M| \right| 
\leq 2 \int_{1/y \geq T} f + \int |M| e^{-2n \int_{1/y \geq T} f}.
\]

For fixed \( \varepsilon > 0 \), we choose \( T \) and \( b \) so large that the terms on the right hand side are \( \frac{\varepsilon}{6}, \frac{\varepsilon}{6} \) and \( o(1) \) respectively. Then

\[
P(\sup_{h \geq b} \Delta(h) > \varepsilon) \leq P \left( \sup_{h \geq b} \left| \int |m_{nh}| - \int |M| \right| > \frac{\varepsilon}{2} \right) + P \left( \sup_{h \geq b} \left| E \int |m_{nh}| - \int |M| \right| > \frac{\varepsilon}{2} \right) \leq e^{-2n \int_{1/y \geq T} f}
\]

for all \( n \) large enough. This concludes the proof of Theorem C2. \( \square \)

2.4. Behavior of the minimizing integral

Although this seems rather obvious, we will nevertheless state and prove the following property:

**Theorem C3.** - Let \( H \) minimize \( \int |f_{nh} - g_{nh}| \). For all \( f \) and all absolutely integrable \( K \) and \( L \), \( \int |f_{nH} - g_{nH}| \) tends to 0 almost surely and in the mean.

Also, \( E_n \int |\tilde{f}_{nH} - \tilde{g}_{nH}| \to 0 \) almost surely, and \( E \int |f_{nH} - g_{nH}| \to 0. \)

**Proof of Theorem C3.** - Let the sequence \( h^* = h^*(n) \) be such that \( E \int |f_{nh^*} - g_{nh^*}| \sim \inf_h E \int |f_{nh} - g_{nh}|. \) We know that whenever \( h \to 0 \) and \( nh \to \infty \), it follows that \( \int |f_{nh} - f| \to 0 \) almost surely and in the mean (see e.g. Devroye, 1983). In particular, \( \inf_h \int |f_{nh} - f| \to 0 \) almost surely, and \( E \int |f_{nh^*} - g_{nh^*}| \to 0 \) as \( n \to \infty \). Assume that \( n \) is so large that
Then, since \( \int |f_{nH} - g_{nH}| \leq \int |f_{nh} - g_{nh}| \) by definition, we see that for such \( n \),

\[
P \left( \int |f_{nH} - g_{nH}| > \varepsilon \right) 
\leq P \left( \int |f_{nh} - g_{nh}| - E \int |f_{nh} - g_{nh}| > \frac{\varepsilon}{2} \right) 
\leq 2e^{-\left(\frac{\varepsilon^2}{2} \int |K - L| \right)}.
\]

We can now apply the corollaries of Theorem C2, to conclude that

\[
E_n \int |f_{nH} - g_{nH}| \to 0 \text{ almost surely, and } E \int |f_{nH} - g_{nH}| \to 0.
\]

The decoupling necessary for the proof of Theorem C1 is now complete.

### 2.5. Necessary conditions of convergence

**Lemma C1.** Let \( K \) and \( L \) be absolutely integrable kernels such that their (generalized) characteristic functions do not coincide on any open neighborhood of the origin, and let \( f \) be an arbitrary density. If

\[
E \int |f_{nH} - g_{nH}| \to 0,
\]

then \( H \to 0 \) in probability (and thus \( H \to 0 \) in probability) as \( n \to \infty \).

Also, for every \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} \inf_{h \geq \varepsilon} \inf E \int |f_{nh} - g_{nh}| > 0.
\]

**Proof of Lemma C1.** By the conditional form of Jensen’s inequality,

\[
E \left| E_n f_{nH} - E_n g_{nH} \right| \to 0
\]
as \( n \to \infty \). By the independence of \( H \) and the data, this implies that

\[
\lim_{n \to \infty} E \int |f*K_H - f*L_H| = 0.
\]

We can now replace \( H \) by \( H \) since they are identically distributed. Assume that \( f \) and \( K - L \) have characteristic functions (suitably generalized when \( K - L \) is not nonnegative) \( \phi \) and \( \psi \) respectively. By the \( L_1 - L_\infty \) inequality for functions and their Fourier transforms, we have

\[
E \left( \int |f*K_H - f*L_H| \right) \geq P(H \geq \varepsilon) \inf_{h \geq \varepsilon} \sup_{t} |\phi(t)| \cdot |\psi(th)|.
\]
where $\varepsilon > 0$ is arbitrary. If the right-hand-side tends to zero, then we deduce that $P(H \geq \varepsilon) \to 0$ if we can show that

$$\inf_{h \geq \varepsilon} \sup_t |\varphi(t)| \cdot |\psi(th)| > 0.$$  

We note that $\int K = \int L$, so that $\psi(0) = 0$. Furthermore, the absolute integrability of $K - L$ implies that $\psi$ is uniformly continuous. Let $b > 0$ and $c > 0$ be such that $|\varphi(t)| \geq b > 0$ on $[-c, c]$. Then,

$$\inf_{h \geq \varepsilon} \sup_t |\varphi(t)| \cdot |\psi(th)| \geq b \inf_{h \geq \varepsilon} \sup_{t \in [-c, c]} |\psi(th)| = b \sup_{t \in [-\infty, \infty]} |\psi(t)|.$$  

Since $\psi$ is not identically zero on every open neighborhood of the origin, the last lower bound is positive for every $\varepsilon > 0$. This concludes the first part of Lemma C1.

For the second statement of the Lemma, define $\bar{H} = \bar{H}(n) \geq \varepsilon$ in such a way that $\bar{H}$ depends upon $n$ only, and

$$E \left( \int |f_n \bar{H} - g_n \bar{H}| \right) \sim \inf_{h \geq \varepsilon} E \left( \int |f_n h - g_nh| \right).$$  

If this infimum tends to zero, then, by the first part of the Lemma, $\bar{H} \to 0$ deterministically. But this is impossible, since $\bar{H} \geq \varepsilon$. This concludes the proof of Lemma C1.

**Lemma C2.** Let $K$ and $L$ be absolutely integrable kernels such that $\int |K - L| > 0$, and let $f$ be an arbitrary density. If $E \int |f_n \bar{H} - g_n \bar{H}| \to 0$, then $n \bar{H} \to \infty$ in probability (and thus $n H \to \infty$ in probability) as $n \to \infty$. Also, for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \inf_{h \leq \varepsilon/n} E \int |f_n h - g_nh| > 0.$$  

**Proof of Lemma C2.** We can assume without loss of generality that $\bar{H} \to 0$ in probability, because for any subsequence on which $P(\bar{H} \geq \varepsilon) \to 1$, we have $n \bar{H} \to \infty$ in probability. Hence it suffices to consider only subsequences along which $\bar{H} \to 0$ in probability. Observe that with $m_{nh} = f_n h - g_nh$,

$$E \left( \int |m_n \bar{H} - \bar{E}_n m_n \bar{H}| \right) \leq E \left( \int |m_n \bar{H}| \right) + E \int |\bar{E}_n m_n \bar{H}| \leq o(1) + E \int |\bar{E}_n f_n \bar{H} - f| + E \int |\bar{E}_n g_n \bar{H} - f| = o(1)$$  

Vol. 25, n° 4-1989.
by the fact that $\bar{H} \rightarrow 0$ in probability. Here we applied Theorem 6.1 of Devroye and Györfi (1985).

Assume that $n \bar{H} \rightarrow 0$ in probability, i.e., for all $n$ in an infinite subsequence, we have $P(n \bar{H} \leq b) \geq a$ for some $a > 0$, $b < \infty$. For notational convenience, we will assume that the subsequence is the full sequence. We will then show that this implies that

$$\liminf_{n \rightarrow \infty} E \left( \int |m_n \bar{H} - \bar{E}_n m_n \bar{H}| \right) > 0,$$

which is a contradiction. The proof is in two parts. First we show that there exists a positive constant $d$ such that

$$\liminf_{n \rightarrow \infty} \inf_{0 < h \leq d/n} E \left( \int |m_{nh} - E m_{nh}| \right) > 0.$$

To do so, let $M = K - L$, choose a large constant $c$, and let $p = \int_{-c/2}^{c/2} |M| \, dx$ and $q = \int |M| - p$. Let $N$ be the number of $X_i$'s for which $[X_i - ch, X_i + ch]$ contains no $X_j$ with $j \neq i$. We have

$$\inf_{h \leq d/n} \mathbb{E} \left[ |m_{nh}| \right] \geq \inf_{h \leq d/n} \mathbb{E} \left[ \frac{p N}{n} - \frac{q (n-N)}{n} \right] = \inf_{h \leq d/n} \mathbb{E} \left[ \frac{N}{n} - q \right] = \inf_{h \leq d/n} \int f(x) \left( 1 - \int_{x-ch}^{x+ch} f \right)^{n-1} dx \int |M| - q = \int f(x) \left( 1 - \int_{x-ch}^{x+ch} f \right)^{n-1} dx \int |M| - q.$$

By the Lebesgue density theorem and Fatou's lemma, the limit infimum of the lower bound is at least equal to

$$\int f e^{-2cd} \int |M| - q.$$

This is positive if we first choose $c$ so large that $q / \int |M| < 1/2$ and then choose $d$ small enough. For such choices of $c$ and $d$,

$$\liminf_{n \rightarrow \infty} \inf_{h \leq d/n} \mathbb{E} \left( \int |m_{nh} - E m_{nh}| \right) \geq \liminf_{n \rightarrow \infty} \inf_{h \leq d/n} \mathbb{E} \left( \int |m_{nh}| \right) - \limsup_{n \rightarrow \infty} \sup_{h \leq d/n} \int |E m_{nh}| > 0,$$

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
where we used the fact that the last term in the lower bound is $o(1)$ since $E m_{nh} = f \ast M_h$ and, as $h \downarrow 0$, $\int |f \ast M_h - f \ast M| \to 0$.

For the second part of the proof, let $h^* = h^*(n)$ be a sequence with $d/n \leq h^* \leq \max(b, d)/n$ and

$$E \left( \int |m_{nh^*} - E m_{nh^*}| \right) \sim \inf_{d/n < h \leq \max(b, d)/n} E \left( \int |m_{nh} - E m_{nh}| \right).$$

We have

$$\lim \inf_{n \to \infty} E \left( \int |m_{nh} - E m_{nh}| \right) \geq \lim \inf_{n \to \infty} a \inf_{0 < h \leq \max(b, d)/n} E \int |m_{nh} - E m_{nh}|$$

$$\geq \min \left( a \lim \inf_{n \to \infty} E \int |m_{nh^*} - E m_{nh^*}|, \right)$$

$$a \lim \inf_{n \to \infty} \inf_{h \leq d/n} E \left( \int |m_{nh} - E m_{nh}| \right).$$

We have already shown that the second term in the minimum is positive. So it suffices to consider the positiveness of the first term. By Fatou's lemma, it is at least equal to

$$a \int \lim \inf_{n \to \infty} E \left( |m_{nh^*} - E m_{nh^*}| \right) dx.$$

Let us now use the following lower bound for $E \left( \left| \sum_{i=1}^{n} Z_i \right| \right)$ where $Z_1, \ldots, Z_n$ are iid zero mean random variables:

$$E \left( \left| \sum_{i=1}^{n} Z_i \right| \right) \geq \frac{u(nr)^{3/2}}{4(4+nr)}$$

where $u > 0$, $r = P \left( |Z_1 - Z_1| \geq u \right)$ and $Z_1, Z_1'$ are iid [see e.g. Devroye and Györfi (1985, p. 138); the inequality can be obtained with some work from Szarek's inequality (Szarek, 1976) and Cantelli's form of Chebysev's inequality]. We will apply this with $Z_i = \frac{1}{n} M_h(x - X_i)$, and $u = c/(nh)$ for some constant $c > 0$. Now,

$$r = P \left( |Z_1 - Z_1'| \geq \frac{c}{nh^*} \right) \geq P(X_1 \in x + h^* C) P(X_1 \in x + h^* D)$$
where $C = \{ v : |M(v)| \geq 2c \}$ and $D = \{ v : |M(v)| < c \}$. By the Lebesgue density theorem, applied first to compact restrictions of $C$, and then used by removing the compactness, we have for almost all $x$,

$$\liminf_{n \to \infty} \frac{P(X_1 \in x + h^* C)}{h^*} \geq f(x) \lambda(C)$$

where $\lambda(C)$ is the Lebesgue measure for $C$. A similar formula holds for $P(X_1 \in x + h^* D^c)$ where $D^c$ is the complement of $D$. From this, we have

$$\liminf_{n \to \infty} \frac{r}{h^*} \geq f(x) \lambda(C)$$

for almost all $x$. Recapping, we have

$$a \int \liminf_{n \to \infty} E\left( |m_{nh^*} - E m_{nh^*}| \right) dx$$

$$\geq a \int \liminf_{n \to \infty} \frac{c(nr)^{3/2}}{4nh^* (4+nr)} dx$$

$$\geq a \int \liminf_{n \to \infty} \frac{c(nh^*)^{1/2} (r/h^*)^{3/2}}{4(4+nh^* (r/h^*))} dx$$

$$\geq \int \frac{ac (\liminf nh^*)^{1/2} (f(x) \lambda(C))^{3/2}}{4(4 + \sup nh^* f(x) \lambda(C))} dx$$

which is positive when $\lambda(C) > 0$ (which can be assured by choosing $c$ small enough, since $\int |M| > 0$).

For the second statement of the Lemma, define $\bar{H} = \bar{H}_n$ in such a way that $\bar{H}$ depends upon $n$ only, $n \bar{H} \leq \varepsilon$, and

$$E \int |f_n \bar{H} - g_n \bar{H}| \sim \inf_{0 < h \leq \varepsilon/n} E \left( \int |f_{nh} - g_{nh}| \right).$$

If this infimum tends to zero, then, by the first part of the Lemma, $n \bar{H} \to \infty$ deterministically. But this is impossible, since $n \bar{H} \leq \varepsilon$. This concludes the proof of Lemma C2. ■
2.6. Proof of Theorem C1

From Theorem C3, we see that \( E |f_{nH} - g_{nH}| \rightarrow 0 \). From Lemmas C1 and C2 we retain that \( H \rightarrow 0 \) and \( nH \rightarrow \infty \) in probability as \( n \rightarrow \infty \). Since \( H \) and \( H \) are identically distributed, the same statement is true for \( H \). By Theorem 6.1 of Devroye and Györfi (1985) or Theorem 3.3 of Devroye (1987), this implies that \( \int |f_{nH} - f| \rightarrow 0 \) in probability for all \( f \) and all absolutely integrable \( K \), and similarly for \( \int |g_{nH} - f| \). This in turn implies convergence in the mean of both quantities.

3. C-OPTIMALITY

3.1. The main result

This is the main body of the paper, even though it is concerned only with specific subclasses of densities. The kernel estimates considered here are class \( s \)-estimates (\( s \) is an even positive integer), i.e. estimates based upon class \( s \)-kernels, which are kernels \( K \) having the following properties:

A. \( \int (1 + |x|^s) |K(x)| \, dx < \infty \).

B. \( \int K = 1 \), \( \int x^i K(x) \, dx = 0 \) for \( 0 < i < s \), and \( \int x^s K(x) \, dx \neq 0 \).

C. \( K \) is symmetric.

D. \( \int K^2 < \infty \).

Note that nonnegative kernels are at best class 2 kernels. It is worth recalling that for every density \( f \), no matter how \( h \) is picked as a function of \( n \),

\[
\lim_{n \to \infty} \inf n^{s/(2s+1)} E \int |f_n - f| \geq c
\]

where \( c > 0 \) is a constant depending upon \( K \) only [for \( s = 2 \) and \( K \geq 0 \), see Devroye and Penrod (1984), and for general \( s \), see Devroye, 1988]. This lower bound is not achievable for many densities. The rate \( n^{-s/(2s+1)} \) can
however be attained for densities with \( s-1 \) absolutely continuous derivatives (i.e., \( f, f^{(1)}, \ldots, f^{(s-1)} \) all exist and are absolutely continuous) satisfying the tail condition \( \int \sqrt{f} < \infty \) [see e.g. Rosenblatt (1979), Abou-Jaoude (1977), or Devroye and Györfi (1985)]. The class of such densities will be called \( \mathcal{F}_s \) (or \( \mathcal{F} \) when no confusion is possible).

To handle the tails of \( f \) satisfactorily, it is necessary to introduce a minor tail condition, slightly stronger than \( \int \sqrt{f} < \infty \): we let \( \mathcal{W} \) be the class of all \( f \) for which \( \int |x|^{1+\varepsilon} f(x) \, dx < \infty \) for some \( \varepsilon > 0 \), and let \( \mathcal{V} \) be the class of all \( f \) for which \( \int u_f(x) \, dx < \infty \), where

\[
\frac{\Delta}{\delta} f(x) = \sup_{|y| \leq 1} f(x+y).
\]

Devroye and Györfi (1985) noted that \( \int \sqrt{f} < \infty \) is virtually equivalent to \( \int |x| f(x) \, dx < \infty \) although there are exceptions both ways. Thus, \( \mathcal{F}_s \cap \mathcal{W} \) is not much smaller than \( \mathcal{F}_s \). The same is true for \( \mathcal{V} \), since \( \int \sqrt{f} < \infty \) implies \( \int \sqrt{u_f} < \infty \) for most smooth densities (e.g. it is always implied when \( f \) is monotone in the tails).

Next, we will impose a weak smoothness condition on an absolutely integrable function \( M \): \( M \) is said to be smooth if there exists a constant \( C \) such that

\[
\sup_{1 \leq h \leq c} \int |M - M_h| \leq C (c-1)
\]

for all \( c > 1 \). Smoothness of a kernel implies that small changes in \( h \) induce proportionally small changes on \( f_{nh} \) with regard to the \( L_1 \) distance. It seems vital to control these changes for any method that is based upon the minimization of a criterion involving \( h \). Consider now the problem of picking the “smoothness constant” \( C \). For example, if \( M \geq 0 \) is unimodal at the origin and \( \int M = 1 \), then we can always take \( C = 2 \) [see Devroye and Györfi (1985, p. 187)]. However, this is not interesting for us, since we need to have smoothness for the difference function \( K - L \), which takes on negative values. When \( M \) is absolutely continuous, we can take

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
The claim is now obtained by considering $\int_Q$ as well. Finally, a kernel $K$ is said to be regular if it is bounded, and if there exists a symmetric unimodal integrable nonnegative function $M$ such that $|K| \leq M$.

Our main result can now be announced as follows:

**THEOREM 01.** Let $K$ be a smooth regular class $s$ kernel. Assume that $f \in F_s \cap W$, and that $L$ is chosen such that:

A. $\int (1 + |x|^s) |L(x)| \, dx < \infty$.

B. $\int L = 1$, $\int x^i L(x) \, dx = 0$ for $0 < i \leq s$.

C. $L$ is symmetric, smooth, and regular.

D. The generalized characteristic functions of $K$ and $L$ do not coincide on any open neighborhood of the origin. (This implies that $\int |K - L| > 0$.)

E. $\int L^2 < \int K^2/16$.

Then, the kernel estimate with smoothing factor $H$ minimizing $\int |f_{nh} - g_{nh}|$, is $C$-optimal where $C = (1 + u)/(1 - u)$ and $u = 4 \sqrt{\int L^2/\int K^2}$. When $f \in F_s \cap V \cap W^r$, the same is true, provided that, additionally, $L$ has compact support.

A pair $(K, L)$ is chosen as a function of $s$ and the constant $C$ only. Rescaling $L$ changes $\int L^2$, and can thus be used to push the value of $C$ as close to one as desired.
Example 1. — Let us illustrate the choice of $L$ on a simple but important example with $s=2$ and nonnegative kernels $K$. There is plenty of evidence in favor of choosing Bartlett’s kernel $K(x) = \frac{3}{4} (1-x^2)_+$ in those cases, see e.g. Bartlett (1963), Epanechnikov (1969) and Devroye and Penrod (1984). This 2-kernel is smooth, regular, absolutely integrable and of compact support. It is easy to find 4-kernels with the same properties. A little algebra shows that we can take, for example, the continuous kernel 

$$L(x) = \frac{75}{16} (1-x^2)_+ - \frac{105}{32} (1-x^4)_+.$$ 

Interestingly, this coincides with an optimal 4-kernel described e.g. in Gasser, Müller and Mammitzsch (1985, Table 1).

Example 2. — It is interesting to note that the functional form of $L$ can be fixed for all $s$ and $K$ once and for all. It suffices for example to consider bounded smooth symmetric kernels $L$ whose characteristic function is zero in an open neighborhood of the origin satisfying the moment condition A of Theorem 1 for all $s$. This can be done by defining the characteristic function of $L$ as the convolution of the uniform function on $[-1,1]$ with a symmetric bounded infinitely many times continuously differentiable function with support on $[-1/2, 1/2]$.

Example 3. — There is a systematic way of creating higher order kernels. Stuetzle and Mittal (1979) pointed out that $2K - K \ast K$ is a class 2 $s$ kernel whenever $K$ is a class $s$ kernel. This can be iterated at will. There are other tricks. For example, if $M$ is a class $r$ kernel, and $K$ is a class $s$ kernel, the $K + M - K \ast M$ is a class $s + r$ kernel. We can also use higher convolutions for creating better kernels. One can verify that $3K - 3K \ast K + K \ast K \ast K$ is a class 3 $s$ kernel whenever $K$ is a class $s$ kernel. Good sources of possible definitions of families of kernels that are optimal in certain senses are Müller (1984), Gasser, Müller and Mammitzsch (1985), Su-Wong, Prasad and Singh (1982) and Singh (1979).

3.2. A better estimate

We have mentioned that $g_{nh}$ is probably preferable over $f_{nh}$, even though it is not asymptotically optimal in any sense for its kernel $L$, it has a smaller error than that of $f_{nh}$, basically because the class $s$ kernel used in $f_{nh}$ limits its performance. It is interesting to observe that for any absolutely integrable compact support kernel $K$, we must have $\int x^s K(x) \neq 0$ for some finite $s$ [see e.g. Theorem 22 of Hardy and Rogosinski (1962)]. Thus, for
all such kernels we have the saturation phenomenon, and a judiciously picked \( L \) for \( g_{nh} \) is potentially better.

We offer the following

**Theorem O2.** — Let \( L \) and \( K \) be as in Theorem O1, and assume that all the conditions of Theorem O1 are satisfied. Then

\[
\limsup_{n \to \infty} \frac{E \int |g_{nh} - f|}{E \int |f_{nh} - f|} \leq c = 4 \sqrt{\frac{\int L^2}{\int K^2}},
\]

and

\[
\limsup_{n \to \infty} \frac{E \int |g_{nh} - f|}{\inf_{h} E \int |f_{nh} - f|} \leq \frac{c(1+c)}{1-c}.
\]

Theorem O2 basically states that since \( c \) can be chosen arbitrarily small, the asymptotic performance of \( g_{nh} \) can be made any desired fraction of \( \inf_{h} E \int |f_{nh} - f| \). The argument that \( g_{nh} \) itself is not asymptotically optimal for \( L \) can be countered with the observation that no smoothing parameters have to be chosen for \( g_{nh} \) either. Unless, of course, one considers the "spread" of \( L \) (measured by \( \int L^2 \)) as a hidden smoothing factor of sorts.

### 3.3. Complete convergence of the L1 error

The first result is related to Theorem C2, but differs in that it is more specific in its error estimates.

**Lemma O1.** — Let \( f \) be an arbitrary density and let \( K \) and \( L \) be smooth absolutely integrable kernels with \( \int |K - L| > 0 \). Then, for arbitrary fixed \( \epsilon, \eta > 0 \),

\[
P \left( \sup_{\frac{\epsilon}{\eta} \leq h \leq \frac{1}{\eta}} \left| \int f_{nh} - g_{nh} \right| - E \int f_{nh} - g_{nh} \right| \leq \frac{\sqrt{128} \int |K - L| \sqrt{(1+u) \log(n)}}{\sqrt{n} \cdot \frac{C}{\sqrt{n} \log(n)} n^{-1+u}} \right)
\]

\[
\leq (1 + o(1)) \frac{\sqrt{2} \int |K - L| \sqrt{1+u}}{n^{1+u}}
\]

Vol. 25, n° 4-1989.
where $C$ is the smoothness constant for $K - L$ (see definition of smoothness above).

Proof of Lemma O1. — From the proof of Theorem C2, we recall the following inequality:

$$
P \left( \sup_{t/n \leq h \leq 1/t} \left| \int f_{nh} - g_{nh} \right| - E \int f_{nh} - g_{nh} \right| \geq t \right) \leq 2 \left( \frac{\log (n/\varepsilon^2)}{\log (c)} + 2 \right) e^{-\left( n^2/128 \right) \left( \int (K - L) \right)}$$

valid for all $t > 0$. Here $c$ depends upon $t$ and $K - L$ in the following manner: it is so small that

$$\sup_{1 \leq h \leq c} \int \left| (K - L) - (K_h - L_h) \right| \leq \frac{t}{4}.$$ 

Now, upon replacing $t$ by $\sqrt{128} \left| K - L \right| \sqrt{1 + u} \log (n)/n$, we obtain as upper bound

$$\frac{\log (n/\varepsilon^2)}{\log (c)} + 2 \left( \frac{8 C + t}{2 t} \right) n^{-1 + u}.$$ 

We are left with the sole problem of choosing $c$. From our assumption on $K - L$, we see that it suffices to take $c = 1 + t/(4 C)$. Using the fact that $\log (c) \geq 2 t/(8 C + t)$, the upper bound becomes

$$\frac{C \sqrt{n \log (n)}}{\sqrt{2} \left| K - L \right| \sqrt{1 + u}} n^{-1 + u}$$

when $u$ and $\varepsilon$ are held fixed and $n \to \infty$. □

Lemma O2. — Let $f$ be an arbitrary density. Let $K$ and $L$ be smooth absolutely integrable kernels with $\int |K - L| > 0$, and let $H$ minimize

$$\int |f_{nh} - g_{nh}|.$$ 

Assume also that the generalized characteristic functions of $K$ and $L$ do not coincide on any open neighborhood of the origin. Then, for arbitrary fixed $\varepsilon > 0$,

$$P \left( H \notin \left[ \frac{1}{n \varepsilon} \right] , \varepsilon \right) < \frac{1}{n^2}$$

for all $n$ large enough. Furthermore, $H \to 0$ and $n H \to \infty$ completely. Finally, $\int |f_{nH} - f| \to 0$ completely.
**Proof of Lemma O2.** We will inherit the notation of Lemma O1. Define
\[
 t = \frac{\sqrt{128} \int |K - L| \sqrt{3 \log(n)}}{\sqrt{n}}.
\]
Then, by Lemma O1,
\[
P \left( \sup_{1/(n\epsilon) \leq h \leq \epsilon} \left| \int f_{nh} - g_{nh} - E \int f_{nh} - g_{nh} \right| \geq t \right)
\leq (1 + o(1)) \frac{C \sqrt{\log(n)}}{\sqrt{6} \int |K - L|} n^{-5/2}.
\]
This will be combined with the fact that
\[
\liminf_{n \to \infty} \inf_{h \notin [1/(n\epsilon), \epsilon]} E \left( \left| \int f_{nh} - g_{nh} \right| \right) > 0
\]
(Lemmas C1 and C2), and with the observation that for fixed $u > 0$,
\[
P \left( \sup_{h} \left| \int f_{nh} - g_{nh} - E \int f_{nh} - g_{nh} \right| > u \right) \leq e^{-u}
\]
where $\gamma = \gamma(u) > 0$ (Theorem C2). Let $A$ be the set $[1/(n\epsilon), \epsilon]$. For any $\delta > 0$, we have the following inclusion of events:
\[
[H \notin A] \subseteq \left[ \sup_{h \in A} \left| \int f_{nh} - g_{nh} - E \int f_{nh} - g_{nh} \right| \geq t \right]
\cup \left[ \inf_{h} E \left| f_{nh} - g_{nh} \right| + t \geq \delta \right]
\cup \left[ \inf_{h \notin A} E \left| f_{nh} - g_{nh} \right| \leq 2\delta \right]
\cup \left[ \inf_{h \notin A} \left| f_{nh} - g_{nh} - E \int f_{nh} - g_{nh} \right| \leq -\delta \right].
\]
Since $t \to 0$, the second event on the right-hand-side is vacuous for large enough $n$. Also, for $\delta$ small enough and $n$ large enough, the third event is vacuous as we have pointed out above. Hence, for such $\delta$ and such large $n$,
\[
P (H \notin A) \leq (1 + o(1)) \frac{C \sqrt{\log(n)}}{\sqrt{6} \int |K - L| n^{-5/2} + e^{-\gamma(\delta)n} \leq \frac{1}{n^2}}
\]
Vol. 25, n° 4-1989.
for $n$ large enough. The last statement of Lemma O2 follows from Theorem 6.1 of Devroye and Györfi (1985).

3.4. Relative stability of the estimate

**Lemma O3.** Let $f$ be an arbitrary density. Let $K$ and $L$ be smooth absolutely integrable kernels with $\int |K - L| > 0$, and let $H$ minimize $\int |f_{nH} - g_{nH}|$. Assume also that the generalized characteristic functions of $K$ and $L$ do not coincide on any open neighborhood of the origin.

$$P\left( \left| \int f_{nH} - g_{nH} - E_n \int f_{nH} - g_{nH} \right| \leq \frac{\sqrt{128} \int |K - L| \sqrt{3 \log(n)}}{\sqrt{n}} \right) \leq \frac{2}{n^2}$$

for all $n$ large enough. Also,

$$E \left| \int f_{nH} - g_{nH} - E \int f_{nH} - g_{nH} \right| \leq E \left| \int f_{nH} - g_{nH} - E_n \int f_{nH} - g_{nH} \right| \leq \frac{\sqrt{129} \int |K - L| \sqrt{3 \log(n)}}{\sqrt{n}}$$

for all $n$ large enough.

**Proof of Lemma O3.** Define $t = \frac{\sqrt{128} \int |K - L| \sqrt{3 \log(n)}}{\sqrt{n}}$. Let $A$ be as in the proof of Lemma O2 for arbitrary $\varepsilon > 0$. Define the random variable $H_A$ as the projection to $A$ of $H$. We have

$$P\left( \left| \int f_{nH} - g_{nH} - E_n \int f_{nH} - g_{nH} \right| \geq t \right) \leq P\left( \left| \int f_{nH_A} - g_{nH_A} - E_n \int f_{nH_A} - g_{nH_A} \right| \geq t \right) + P(H \notin A) < \frac{1}{n^2} + P(H \notin A) < \frac{2}{n^2}$$
for all $n$ large enough, where we used estimates from Lemma O2. Also,

$$\mathbb{E} \left| \int f_{nH} - g_{nH} - E_n \int f_{nH} - g_{nH} \right|$$

$$\leq t + \int |K - L| P \left( \left| \int f_{nH} - g_{nH} - E_n \int f_{nH} - g_{nH} \right| \geq t \right)$$

$$\leq \frac{\sqrt{129} \int |K - L| \sqrt{3 \log(n)}}{\sqrt{n}}$$

for all $n$ large enough.

**Lemma O4.** Let $f$ be an arbitrary density. Let $K$ and $L$ be smooth absolutely integrable kernels with $\int |K - L| > 0$, and let $H$ minimize $\int |f_{nh} - g_{nh}|$. Assume also that the generalized characteristic functions of $K$ and $L$ do not coincide on any open neighborhood of the origin.

$$P \left( \left| \int f_{nH} - f - E_n \int f_{nH} - f \right| \geq \frac{\sqrt{128} \int |K| \sqrt{3 \log(n)}}{\sqrt{n}} \right) \leq \frac{2}{n^2}$$

for all $n$ large enough. Also,

$$\mathbb{E} \left| \int f_{nH} - f - E \int f_{nH} - f \right| \leq \mathbb{E} \left| \int f_{nH} - f - E_n \int f_{nH} - f \right|$$

$$\leq \frac{\sqrt{129} \int |K| \sqrt{3 \log(n)}}{\sqrt{n}}$$

for all $n$ large enough.

**Proof of Lemma O4.** We mimick the proof of Lemma O1 first, replacing $g_{nh}$ throughout by $f$, and $K - L$ by $K$. From this, we conclude that for arbitrary fixed $\varepsilon, u > 0$,

$$P \left( \sup_{\varepsilon/n \leq h \leq 1/e} \left| \int f_{nh} - f - E \int f_{nh} - f \right| \right.$$

$$\geq \frac{\sqrt{128} \int |K| \sqrt{(1 + u) \log(n)}}{\sqrt{n}}$$

$$\leq (1 + o(1)) \frac{C \sqrt{n \log(n)}}{\sqrt{2} \int |K| \sqrt{1 + u}} n^{-(1 + u)}$$
where $C$ is the smoothness constant for $K$ (see definition of smoothness just before Lemma O1). Then turn to the proof of Lemma O3, replacing $K-L$ in the definition of $t$ by $K$. Furthermore, replace all the references to $g_{nH}$ and $g_{nH_A}$ by $f$, and note that $\int |f_{nh}-f| \leq 1 + \int |K|$. This concludes the proof of Lemma O4.

**Lemma O5.** Let $f_{nh}$ be a kernel estimate with class $s$ kernel $K$. Then there exists a constant $c = c(K) > 0$ such that for any $f$ and for any sequence $h = h(n)$,

$$\liminf_{n \to \infty} n^{s/2s+1} \mathbb{E} \int |f_{nh} - f| \geq c > 0,$$

The same bound is valid if $f_{nh}$ is replaced by $f_{nH}$, where $H = H(n)$ is any sequence of positive random variables independent of the data sequence.

If $H$ is obtained by minimizing $\int |f_{nh} - g_{nh}|$ and $K$ and $L$ are smooth absolutely integrable kernels with $\int |K - L| > 0$, such that the generalized characteristic functions of $K$ and $L$ do not coincide on any open neighborhood of the origin, then the asymptotic bound is also valid. In that case, we also have

$$\mathbb{E}_n \int |f_{nH} - f| \to 1 \text{ almost surely as } n \to \infty$$

for all densities $f$.

**Proof of Lemma O5.** The asymptotic bound for deterministic $h(n)$ is obtained in Devroye (1988). It is clear that for any sequence of random variables $\{H = H(n)\}$, that

$$\mathbb{E} \int |f_{nH} - f| \liminf_{n \to \infty} \frac{\mathbb{E} \int |f_{nH} - f|}{\mathbb{E} \int |f_{nh^*} - f|} \geq 1,$$

where $h^* = h^*(n)$ is such that $\mathbb{E} \int |f_{nh^*} - f| \leq \inf_{{h}} \mathbb{E} \int |f_{nh} - f|$. Since the asymptotic lower bound is valid for $f_{nh^*}$, it must be valid for $f_{nH}$.
Let $H$ now be found by minimization as indicated in the statement of Lemma 05. Then, as we have seen in Lemma 04,

$$E \int |f_nH - f| - E \int |f_nH - f| = O\left(\frac{\log n}{n}\right).$$

Thus, the asymptotic bound also applies to $E \int |f_nH - f|$. The last statement of the Lemma is obtained from the probability bound of Lemma 04, the asymptotic lower bound of Lemma 05, and the Borel-Cantelli lemma (the sequence $2/n^2$ is summable in $n$).

It is perhaps worthwhile to pause here to see what Lemma 05 implies for us. First of all, $f_nH$ is strongly relatively stable, as shown in the last statement of the Lemma. Thus, the random variable $\int |f_nH - f|$ is very close to its mean. This is true for all densities $f$. For general theorems on the relative stability of $f_nH$, with arbitrary $f$, $K$ and $H$, see e.g. Devroye (1988). What this means for us is that $\int |f_nH - g_nH|$, a known quantity, is probably close to its mean, which, as we shall see below, is not too far away from $E \int |f_nH - f|$. By relative stability again, the last quantity is about equal to $\int |f_nH - f|$. In other words, we have another useful by-product of the minimization, i.e. a rough estimate of the actual $L_1$ error $\int |f_nH - f|$.

### 3.5. Proof of Theorems O1 and O2

Let $h^* = h^*(n)$ be such that

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

We note the following:

$$E \int |f_nH - f| \leq E \int |f_nH - g_nH| + E \int |g_nH - f|$$

$$= E \left(\inf_h \int |f_{nh} - g_{nh}|\right) + E \int |g_nH - f|$$

Vol. 25, n° 4-1989.
for all $n$ large enough (Lemma 04, applied to $g_{nh}$). So far, everything is valid for all densities.

Under the conditions of Theorem 01, with $L$ as suggested in the statement of the Theorem, it is possible to show that (1) through (6) are satisfied with $c_1 = c_2 = 4 \sqrt{\int L^2/\int K^2}$ (Lemmas O6, O11 and O12):

\[
\int |E_{g_{nh^*}} - f| = O\left(\int |Ef_{nh^*} - f|\right), \quad (1)
\]
\[
\int |g_{nh^*} - f| = O\left(\int |f_{nh^*} - f|\right), \quad (2)
\]
\[
\int g_{nh^*} - f \leq (c_1 + o(1)) \int |f_{nh^*} - f|, \quad (3)
\]
\[
\int f \ast L_{\tilde{h}} - f = O\left(\int |f \ast K_{\tilde{h}} - f|\right), \quad (4)
\]
\[
\int g_{n\tilde{h}} - f \ast L_{\tilde{h}} \leq (c_2 + o(1)) \int |f_{n\tilde{h}} - f|, \quad (5)
\]
\[
\int g_{n\tilde{h}} - f \leq (c_2 + o(1)) \int |f_{n\tilde{h}} - f|. \quad (6)
\]

We may conclude from (3) and (6) that

\[
\int |f_{n\tilde{h}} - f| \leq (1 + o(1))(1 + c_1 + o(1)) \int |f_{nh^*} - f|\]
\[
+ (c_2 + o(1)) \int |f_{n\tilde{h}} - f| + \frac{\sqrt{129} \int |L| \sqrt{3 \log(n)}}{\sqrt{n}}
\]

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
THE DOUBLE KERNEL METHOD

563

by Lemma 04. By Lemma 05, we see that the last two terms in the upper
bound are asymptotically negligible with respect to the first term. Thus,
we can conclude that

\[
E \int | f_{nH} - f | \leq \frac{1 + c_1 + o(1)}{1 - c_2 + o(1)} E \left( \int | f_{nh^*} - f | \right).
\]

The right hand side can be made smaller than \(1 + \varepsilon + o(1)\) for any \(\varepsilon > 0\)
by the appropriate choice of \(L\), since \(c_1 = c_2 = 4 \sqrt{L^2 / K^2}\). This conclu-
des the proof of Theorem O1.

Recall the \(n^{-s(2s+1)}\) lower bound for \(E \int | f_{nH} - f |\) and \(\inf_h E \int | f_{nh} - f |\)
(Lemma O5). Theorem O2 follows from this fact, (6), Theorem O1, and
the fact that \(E \int | f_{nH} - f | - E \int | f_{nH} - f |\) is \(O(\sqrt{\log n/n})\), and similarly for
\(g_{nh}\) (Lemma O4).

3.6. Remarks and further work

First we observe that the condition that \(f \in F_s\) is too strong. Theorem O1
holds for a much larger class of densities. It suffices to note that the
crucial asymptotic result used in the proof of Lemma O6 remains valid,
in the case \(s = 2\), \(K \geq 0\), when \(f\) is such that it has a finite functional

\[
D(f) \equiv \lim_{a \downarrow 0} \int \left| (f * \varphi_a)^{(2)} \right|,
\]

where \(\varphi\) is a mollifier, i.e. a kernel with \(\int \varphi = 1\), \(\varphi \geq 0\), \(\varphi = 0\) outside
\([-1,1]\), and \(\varphi\) has infinitely many continuous derivatives on the real line.
This functional coincides with \(\int |f^{(2)}|\) when \(f\) and \(f'\) are absolutely con-
tinuous, and is well-defined (possibly \(\infty\)) and independent of the choice of
\(\varphi\) for all \(f\). For the proofs of this, see e.g. Devroye (1987), pp. 108-111.
To illustrate this, consider the triangular density. It does not have an

Vol. 25, n° 4-1989.
absolutely continuous derivative, yet $D(f) < \infty$. For smooth regular symmetric non-negative $K$ with finite second moment, Theorem 01 is valid for all $f$ in $W$ or $V$ for which $\int \sqrt{f} < \infty$ and $D(f) < \infty$.

It is possible to get asymptotic optimality for a proper subclass of $F_s$ by choosing $L$ in such a way that $L$ varies with $n$ by a scale factor only, i.e., $L_n$ is replaced throughout by $L_{a_n}$ where $a_n$ tends very slowly to $\infty$ so as not to upset properties (1) and (4). This will allow us to formally take $c_1 = c_2 = 0$, and obtain the asymptotic optimality. The proper subclass of $F_s$ is determined by the rate of divergence of $a_n$. This will not be pursued any further here.

In Theorem 01, $K$ is a class $s$ kernel, so that the best possible rate of convergence is $n^{-s(2s+1)}$ (Lemma 05). If we know that $f$ is very smooth, then this could be an unwelcome restriction. One might wonder if there is nothing that can be said if we employ a class $\infty$ kernel. We have the following general result, which can be proved along the lines of the proof of Theorem 1, provided that Lemma 06 is replaced by a (trivial) counterpart.

**Theorem 03.** Let $K$ and $L$ be symmetric smooth regular kernels, with $\int L^2 < \int K^2/16$. Assume furthermore that the generalized characteristic functions of $K$ and $L$ do not coincide on any open neighborhood of the origin. (This implies that $\int |K - L| > 0$.) Let $f \in W$ be such that $\int |f \ast L_n - f|/\int |f \ast K_n - f| \to 0$ as $h \downarrow 0$. Then, the kernel estimate in which $H$ is defined by $\int |f_n H - g_n H| \sim \inf_h \int |f_n h - g_n h|$, satisfies the following inequality:

$$E\left(\int |f_n H - f|\right) \leq (1 + o(1)) \frac{1 + u}{1 - u} \inf_h \left(\int |f_n h - f|\right)$$

$$+ (1 + o(1)) \frac{\sqrt{387 \log n}}{1 - u} \sqrt{\frac{|K| + \int |L|}{n}},$$

where $u = 4 \sqrt{\int L^2/\int K^2}$. When $f \in V \cap W^c$, the same is true, provided that, additionally, $L$ has compact support.

It is easy to see that the $H$ obtained with the pair $(K, L) = (K, 2K - K \ast K)$ is indistinguishable from the $H$ obtained by the
pair \((K, K \ast K)\). This has an interesting interpretation: indeed, the kernel estimate \(f_{nh}\) can formally be considered as \(\mu_n \ast K_h\) where \(\mu_n\) is the standard empirical measure. With \(L = K \ast K\), the estimate \(g_{nh}\) is nothing but \(\mu_n \ast K_h \ast K_h = f_{nh} \ast K_h\). Minimizing \(\int |f_{nh} - g_{nh}|\) is like asking that the operation \(" \ast K_h\)" yields a stable point (doesn't change things too much); if \(h\) is really good, then \(\mu_n \ast K_h\) should be close to \(f\). But then applying the same operator again should not yield a very different curve, so \(\mu_n \ast K_h \ast K_h\) should be close to \(\mu_n \ast K_h\). So, what are the properties of the double kernel estimate with the pair \((K, K \ast K)\)?

3.7. The final series of lemmas

**Lemma 06.** Assume that \(f \in \mathcal{F}_s\). Let \(K\) and \(L\) be smooth absolutely integrable kernels whose generalized characteristic functions do not coincide on any open neighborhood of the origin, and let \(K\) be a class \(s\) kernel. Then facts (1) and (4) are valid provided that \(L\) is picked such that \(L\) is symmetric,

\[
\begin{align*}
\int L &= 1, \\
\int |L| < \infty, \\
\int x^i L(x) \, dx &= 0 \quad \text{for all } 0 < i \leq s, \quad \text{and} \\
\int |x|^\nu |L(x)| \, dx &< \infty.
\end{align*}
\]

Proof of Lemma 06. We recall that \(h^* \to 0\) as \(n \to \infty\) (see the proof of Theorem C3 together with Lemmas C1 and C2). Under the conditions of Theorem O1, we have for \(f \in \mathcal{F}_s\), as \(h \downarrow 0\),

\[
\left| \int E f_{nh} - f \right| = \left| \int f \ast K_h - f \right| \sim h^s \left| \int \frac{x^s}{s!} K(x) \, dx \right| \int |f^{(s)}|,
\]

[see e.g. Devroye and Györfi (1985, p. 209) or Devroye (1987, p. 110)]. Also, if \(\left| f^{(s)} \right| < \infty\), and if \(L\) is such that \(\int L = 1\), \(\int |L| < \infty\), \(\int x^i L(x) \, dx = 0 \quad \text{for all } 0 < i \leq s\), and \(\int |x|^\nu |L(x)| \, dx < \infty\), then, from Devroye (1987, p. 110) we retain that

\[
\left| \int E g_{nh} - f \right| = \left| \int f \ast L_h - f \right| = o(h^s).
\]

This establishes (1). For the proof of (4), we note that \(H \to 0\) in probability (from Theorem C3 and Lemma C1). Thus, if \(\mu\) is the probability measure for \(H\), and \(F(h)\) and \(G(h)\) denote the biases \(\int |f \ast K_h - f|\) and

Vol. 25, n° 4-1989.
respectively, then, for $\varepsilon > 0$,

\[
\frac{\mathbb{E} \int |f \ast L_h - f|}{\mathbb{E} \int |f \ast K_{H} - f|} = \frac{\int G(h) \mu(\text{d}h)}{\int F(h) \mu(\text{d}h)} \leq \frac{\int_{0}^{\varepsilon} G(h) \mu(\text{d}h) + \left(1 + \int |L| \right) \mathbb{P}(H > \varepsilon)}{\int F(h) \mu(\text{d}h)} \leq \sup_{h < \varepsilon} \frac{G(h)}{F(h)} + \left(1 + \int |L| \right) \mathbb{P}(H \in [1/(n \varepsilon), \varepsilon]) \inf_{1/(n \varepsilon) \leq h \leq \varepsilon} \frac{\mathbb{P}(H > \varepsilon)}{F(h)}.
\]

The first term in the upper bound tends to zero as $\varepsilon \downarrow 0$ [by (1)], while for fixed $\varepsilon$, the denominator of the second term is at least equal to a constant times $n^{-s}$ (by Lemma 02 and an estimate for the bias used above). Its numerator is not greater than $n^{-1(s+1)}$ (say) for $n$ large enough by a suitable generalization of the bound of Lemma 04 (it suffices to replace the constant 3 in the definition of $t$ there by a larger constant). This proves (4).

The following lemma provides a uniform lower bound for the expected variation of any kernel estimate with a regular kernel.

**Lemma 07.** Let $f_{nh}$ be the kernel estimate with regular kernel $K$. Then, for all $f$, and for all sequences $a_n$ and $b_n$ with $b_n \to 0$, $n a_n \to \infty$, $a_n \leq b_n$,

\[
\liminf_{n \to \infty} \inf_{a_n \leq h \leq b_n} \frac{4nh}{K^2} \mathbb{E} \int \left| f_{nh} - K_h \ast f \right| \geq \sqrt{f}.
\]

**Proof of Lemma 07.** Let $h^* = h^*(n)$ be a sequence of positive numbers with $a_n \leq h^*(n) \leq b_n$ such that

\[
\mathbb{E} \sqrt{nh^*} \int \left| f_{nh^*} - K_{h^*} \ast f \right| \sim \inf_{a_n \leq h \leq b_n} \mathbb{E} \sqrt{nh} \int \left| f_{nh} - K_h \ast f \right|.
\]

Then, since $h^* \to 0$ and $nh^* \to \infty$, we know that

\[
\liminf_{n \to \infty} \left[ \frac{4nh^*}{K^2} \right]^{1/2} \mathbb{E} \int \left| f_{nh^*} - K_{h^*} \ast f \right| \geq \sqrt{f}.
\]

(Devroye, 1987, Lemma 5). This proves Lemma 07.
**Lemma 08.** Let $f$ be an arbitrary density. Let $K$ be regular. Let $H = H(n)$ have an arbitrary sequence of distributions. Then, for any sequences $a_n \leq b_n$ with $b_n \to 0$ and $n a_n \to \infty$, and for any $\varepsilon$, we have for $n$ large enough (where the definition of "large enough" does not depend upon the distribution of the sequence $\{H = H(n)\}$).

$$
\frac{\mathbb{E} \left| \frac{f_{nH} - K_h * f}{\mathbb{E} \left[ (1/\sqrt{nH}) I_{a_n \leq H \leq b_n} \right] } \right|}{\sqrt{\frac{K^2}{2}}} \geq \begin{cases} \frac{1}{\varepsilon} & \text{if } \int \sqrt{f} = \infty \\ \sqrt{\frac{K^2}{2}} \int \sqrt{f} - \varepsilon & \text{if } \int \sqrt{f} < \infty \\
\end{cases}
$$

**Proof of Lemma 08.** Let us write $\Delta(n, h)$ for $\mathbb{E} \left| f_{nH} - K_h * f \right|$. Then, since $f$ and $H$ are independent,

$$
\mathbb{E} \left| f_{nH} - K_h * f \right| = \mathbb{E} \Delta(n, H) \mathbb{E} \left( \Delta(n, H) I_{a_n \leq H \leq b_n} \right) \\
\geq \mathbb{E} \left[ \sqrt{nH} \Delta(n, H) \frac{1}{\sqrt{nH}} I_{a_n \leq H \leq b_n} \right] \\
\geq \inf_{a_n \leq h \leq b_n} \sqrt{nh} \Delta(n, h) \mathbb{E} \left[ \frac{1}{\sqrt{nH}} I_{a_n \leq H \leq b_n} \right].
$$

Now apply Lemma 07.

**Lemma 09.** Assume that $\mathbb{E} \left| f \right| < \infty$. Let $f_{nh}$ be the kernel estimate with regular kernel $K$. Then, for all sequences $a_n$ and $b_n$ with $b_n \to 0$, $n a_n \to \infty$, $a_n \leq b_n$,

$$
\limsup_{n \to \infty} \sup_{a_n \leq h \leq b_n} \left( \frac{nh}{\int K^2} \right)^{1/2} \mathbb{E} \left| f_{nh} - K_h * f \right| \leq \int \sqrt{f},
$$

when either $f \in \mathcal{W}$ and $K$ has finite second moment, or when $f \in \mathcal{V}$ and $K$ has compact support.

**Proof of Lemma 09.** Note first that every regular kernel is square integrable. Now apply a standard bound that can be obtained directly via the Cauchy-Schwarz inequality:

$$
\mathbb{E} \left| f_{nh} - f * K_h \right| \leq \frac{\int \sqrt{f} (K^2) \right]_h}{\sqrt{nh}}
$$
(see e.g. Devroye, 1987, p. 113). This upper bound is

\[
(1 + o(1)) \frac{\sqrt{\int K^2} \int \sqrt{f}}{\sqrt{nh}}
\]

when \( h \to 0 \) as \( n \to \infty \) when both \( f \) and \( K^2 \) have finite absolute \( 1 + \varepsilon \) moments for some \( \varepsilon > 0 \) [see e.g. exercise 7.8 of Devroye (1987)]. The latter condition on \( K \) is satisfied if \( K \) has finite second moment and is regular. For \( f \), the condition is implied when \( f \in W \). The same asymptotic result is valid if \( K \) has compact support and is bounded, and \( f \in V \) (Devroye and Györfi, 1985, Lemma 5.26).

Let \( h^* = h^*(n) \) be a sequence of positive numbers with \( a_n \leq h^*(n) \leq b_n \) such that

\[
E \frac{1}{nh^*} \int |f_{nh^*} - K_h \ast f| \sim \sup_{a_n \leq h \leq b_n} E \frac{1}{nh} \int |f_{nh} - K_h \ast f|.
\]

Then, since \( h^* \to \infty \) and \( nh^* \to \infty \), we know that

\[
\limsup_{n \to \infty} \left( \frac{nh^*}{\int K^2} \right)^{1/2} E \frac{1}{nh} \int |f_{nh} - K_h \ast f| \leq \int \sqrt{f}.
\]

This proves Lemma O9. ■

**Lemma O10.** — Let \( f \) and \( K \) be as in Lemma O9. Then, for any sequences \( a_n \leq b_n \) with \( b_n \to 0 \) and \( na_n \to \infty \), and for any definition of the random variables \( H = H(n) \),

\[
E \int |f_{nh} - K_h \ast f| \limsup_{n \to \infty} \frac{E[(1/\sqrt{nh}) 1_{a_n \leq H \leq b_n}]}{E[1/\sqrt{nH} 1_{a_n \leq H \leq b_n}]}
\]

\[
\leq \sqrt{\int K^2} \int \sqrt{f} + \limsup_{n \to \infty} \frac{2\sqrt{nb_n} \int K |P(H \notin [a_n, b_n])|}{P(H \in [a_n, b_n])}.
\]
Proof of Lemma 010. − Let us write $\Delta(n, h)$ for $E \int |f_{nh} - K_h * f|$. Then, since $\tilde{f}_{nh}$ and $H$ are independent,

$$E \int |\tilde{f}_{nh} - K_H * f| = E \Delta(n, H) \leq E(\Delta(n, N) I_{a_n \leq H \leq b_n}) + 2 \int K |P(H \notin [a_n, b_n])|$$

$$\leq E \left[ \frac{1}{\sqrt{nH}} \Delta(n, H) \frac{1}{\sqrt{nH}} I_{a_n \leq H \leq b_n} \right] + 2 \int K |P(H \notin [a_n, b_n])|$$

$$\leq \sup_{a_n \leq h \leq b_n} \sqrt{nh} \Delta(n, h) E \left[ \frac{1}{\sqrt{nH}} I_{a_n \leq H \leq b_n} \right] + 2 \int K |P(H \notin [a_n, b_n])|.$$

Now apply Lemma 09. □

**Lemma 011.** − Let $f_{nh}, g_{nh}$ and $H$ be as in Theorem 01 and assume that $\int \sqrt{f} < \infty$. Assume that $K$ is a smooth regular kernel. Then (2) and (5) hold if we choose a smooth regular $L$ in such a way that the generalized characteristic functions of $K$ and $L$ do not coincide on any open neighborhood of the origin, and that either $f \in W$ and $L$ has finite second moment, or $f \in V$ and $L$ has compact support.

Also, we can take $c_1 = c_2 = 4 \sqrt{\int L^2 / \int K^2}$.

Proof of Lemma 011. − By Lemmas C1 and C2, we have $h^* \to 0$ and $nh^* \to \infty$ when $K$ and $L$ are absolutely integrable kernels whose generalized characteristic function do not coincide on any open neighborhood of the origin. Thus, from Lemma 07 applied to $f_{nh}$ and $K$ (which requires that $K$ be regular) and Lemma 09 applied to $g_{nh}$ and $L$ (which requires that $\int \sqrt{f} < \infty$, that $L$ be regular, and that either $f \in W$ and $L$ has finite second moment, or $f \in V$ and $L$ has compact support),

$$\lim sup_{n \to \infty} \sqrt{nh^*} E |g_{nh^*} - E g_{nh^*}| \leq \frac{1}{2} \sqrt{\int L^2} \frac{c_1}{2} \sqrt{\int K^2} \int \sqrt{f}$$

$$\leq \lim inf_{n \to \infty} \frac{c_1}{2} \sqrt{nh^*} E |f_{nh^*} - E f_{nh^*}|$$

$$\leq \lim inf_{n \to \infty} c_1 \sqrt{nh^*} E \int |f_{nh^*} - f|,$$

Vol. 25, n° 4-1989.
by the triangle inequality and Jensen's inequality, where
c_1 = 4 \sqrt{\int L^2} \int K^2. We will see that we can take \( c_2 = c_1 \). This concludes
the proof of (2).

To prove (5), let \( a_n \) and \( b_n \) be \( 3/n^2 \) and \( 1 - 3/n^2 \) quantiles of \( H \) respectively. We show first that \( b_n \to 0 \) and \( na_n \to \infty \). Take \( \varepsilon > 0 \) arbitrary. Assume
for example that for an infinite subsequence, we have \( b_n > \varepsilon \). Then, on
that subsequence,
\[
P(\varepsilon \leq H \leq b_n) = P(H \geq b_n) - P(H \geq \varepsilon)
\geq \frac{3}{n^2} - \frac{1}{n^2} \quad \text{(Lemma 02, all \( n \) large enough) > \frac{1}{n^2},}
\]
which is a contradiction, since \( P(H \geq \varepsilon) \leq 1/n^2 \) for \( n \) large enough. Hence
\( b_n < \varepsilon \) for all \( n \) large enough, and, by symmetry, \( na_n > 1/\varepsilon \) for all \( n \) large
enough. Thus, \( b_n \to 0 \) and \( na_n \to \infty \) as required. Note that Lemma 02
required that \( K \) and \( L \) both be smooth absolutely integrable kernels
whose generalized characteristic functions do not coincide on any open
neighborhood of the origin.

Since \( \int \sqrt{f} < \infty \) and \( K \) is regular, we can employ Lemma 08, and since
\[
\int \sqrt{f} < \infty, \quad L \text{ is regular, and either } f \in \mathcal{W} \text{ and } L \text{ has finite second moment,}
\]
or \( f \in \mathcal{V} \) and \( L \) has compact support, we can use Lemma O10 to conclude
that
\[
\limsup_{n \to \infty} \frac{2}{\sqrt{n}} \frac{n^2}{\int L \left( |P(H \notin [a_n, \ b_n])| \right.}
\leq \sqrt{\int L^2} \int \sqrt{f} + \limsup_{n \to \infty} \frac{2}{\sqrt{n}} \int L \left. \left| P(H \notin [a_n, \ b_n]) \right|\right)
\leq \sqrt{\int L^2} \int \sqrt{f} + \limsup_{n \to \infty} \frac{2}{\sqrt{n}} \int L \left( |6/n^2\right)
\leq \frac{c_2}{2} \sqrt{\int K^2} \int \sqrt{f} \left( \frac{1 - 6/n^2}{2}\right)
\leq \frac{c_2}{2} \liminf_{n \to \infty} \frac{\int |f_{n} H - K_{H} * f|}{\int L \left( |(1/\sqrt{n} H) I_{a_n \leq H \leq b_n} |\right)}
\].

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
where we once again used the fact that
\[ E \int \left| \mathcal{F}_{nH} - f \right| \leq c_2 \liminf_{n \to \infty} \frac{E \left[ (1/\sqrt{nH}) \mathbf{1}_{u_n \leq H \leq b_n} \right]}{E \left[ (1/\sqrt{nH}) \right]} \]

Facts (5) follows from the chain of inequalities derived above and the observation that for sequences of positive numbers \( u_n, v_n, w_n \),
\[ \limsup \frac{u_n}{v_n} \leq \liminf \frac{v_n}{w_n}. \]

**Lemma O12.** *In the proof of Theorem O1, (1) and (2) together imply (3), and (4) and (5) together imply (6).*

**Proof of Lemma O12.** We will use the facts that
\[ \int \left| E f_{nh*} - f \right| \leq E \int \left| f_{nh*} - f \right| \] and that
\[ E \int \left| f * K_H - f \right| = E \int \left| E_n f_{nh*} - f \right| \leq E \int \left| f_{nh*} - f \right| = E \int \left| f_{nh*} - f \right|. \]

Now, \( E \int \left| g_{nh*} - f \right| \) does not exceed the sum of the left-hand-sides of (1) and (2), from which the claim about (3) follows. Similarly, \( E \int \left| g_{nh} - f \right| \) does not exceed the sum of the left-hand-sides of (4) and (5), from which the claim about (5) follows. ■

**4. PROPERTIES OF THE OPTIMAL SMOOTHING FACTOR**

In this section, we intend to show the following

**Theorem S1.** Let \( f \) be an arbitrary density. Let \( f_{nh} \) be a kernel estimate with smooth absolutely integrable class \( s \) kernel \( K \). (Note: its characteristic function does not coincide with 1 on any open neighborhood of the origin.) Let \( H = H(n) \) be any sequence of random variables for which
\[ \int \left| f_{nH} - f \right| \sim \inf_{h} \int \left| f_{nh} - f \right|. \]
Then

$$\lim_{n \to \infty} \frac{E \int |f_{nH} - f|}{\inf_h E \int |f_{nh} - f|} = 1,$$

and

$$\lim_{n \to \infty} \frac{E \int |f_{nH} - f|}{E \int |f_{nH} - f|} = 1 \text{ almost surely.}$$

Theorem S1 reassures us that it is irrelevant whether we study $\inf_E E \int |f_{nh} - f|$ or $E \left( \inf_h \int |f_{nh} - f| \right)$, since both are asymptotically equal for all densities $f$ when $K$ is a class $s$ kernel. The study of the previous section was with respect to the former quantity. We now see that in the definition of C-optimality, it would have been possible to replace the denominator by $E \left( \inf_h \int |f_{nh} - f| \right)$.

**Lemma S1.** – Let $f$ be an arbitrary density. Let $K$ be an absolutely integrable kernel whose characteristic function does not coincide with 1 on any open neighborhood of the origin. Then, for all $\varepsilon > 0$,

$$\lim \inf \inf_{n \to \infty, \varepsilon \in [1/(n \varepsilon), \varepsilon]} E \left( \int |f_{nh} - f| \right) > 0.$$

**Proof of Lemma S1.** – The first statement is obtained by mimicking the proofs of Lemmas C1 and C2. In Lemma C1, it suffices to replace $f \ast L_h$ throughout by $f$ (which formally corresponds to taking $L$ with characteristic function identical to one). Hence the need to introduce the condition that $K$ not coincide with 1 on any open neighborhood of the origin. Lemma C2 remains valid with little change, provided that $M$ in that proof is replaced by $K$. It is necessary there to reverify that

$$\lim \inf \inf_{n \to \infty, 0 < h \leq d/n} E \left( \int |m_{nh} - E m_{nh}| \right) > 0,$$
where \( m_{nh} \) is in the notation of Lemma C2; it is the \( f_{nh} \) of the present Lemma. Still in the notation of Lemma C2, we have

\[
\inf_{h \leq d/n} E \left| m_{nh} - E m_{nh} \right| \geq \inf_{h \leq d/n} E \left( \frac{pN}{n} - \frac{q(n-N)}{n} - \int B \left| E m_{nh} \right| - E \int B \left| E m_{nh} \right| \right)
\]

where \( B \) is the collection of all sets \([X_i - ch, X_i + ch]\) that capture no data point besides \( X_i \). Let us consider the last term. It is surely not greater than \( \sup_{h \leq d/n} \left( \int \left| E m_{nh} - f \right| + E \int f \right) \). This is \( o(1) + \sup_{h \leq d/n} E \int f \) by the convergence of the bias (Theorem 2.1 of Devroye and Györfi, 1985). The last term is in turn not larger than \( E(\omega(\lambda(B'))) \) where \( B' \) is defined as \( B \), but with the interval lengths \( 2ch \) replaced by the larger values \( 2cd/\lambda \) replaced by the larger values \( 2cd/n \), \( \lambda \) is Lebesgue measure, and \( \omega(u) = \sup_{A : \lambda(A) \leq u} \int_A f \) (which \( \rightarrow 0 \) as \( u \downarrow 0 \)). We can bound the term by \( 2cdN/n \leq 2cd \). Combined with the lower bounds of Lemma C2, we can conclude that

\[
\inf_{h \leq d/n} E \left( \int \left| m_{nh} - E m_{nh} \right| \right) \geq \int e^{-2cdf} \left| M \right| - q - \omega(2cd) > 0,
\]

for \( n \) large enough, \( c \) large enough and \( d \) small enough. The second part of the proof of Lemma C2 requires no modifying. \( \blacksquare \)

**LEMMA S2.** Let \( f \) be an arbitrary density. Let \( K \) be an absolutely integrable kernel. For fixed \( u > 0 \),

\[
P\left( \sup_h \left| f_{nh} - f \right| - E \int f_{nh} - f \right| > u \right) \leq e^{-\gamma n}
\]

where \( \gamma = \gamma(u) > 0 \). As a consequence, with \( J_{nh} = \int \left| f_{nh} - f \right| \), we have the following:

A. \( \sup_{h > 0} \left| J_{nh} - E J_{nh} \right| \rightarrow 0 \) almost surely as \( n \rightarrow \infty \).

B. For any random variable \( H \) (possibly not independent of the data), \( J_{nh} - E J_{nh} \rightarrow 0 \) almost surely as \( n \rightarrow \infty \).

C. For any random variable \( H \) (possibly not independent of the data), \( J_{nh} \rightarrow 0 \) in probability implies \( E J_{nh} \rightarrow 0 \), \( E J_{nh} \rightarrow 0 \) in probability, \( E J_{nh} \rightarrow 0 \) and \( E J_{nh} \rightarrow 0 \).

**Proof of Lemma S2.** We extend the proof of Theorem C2. Note first that \( \left| f_{nh} - u_{nh} \right| \leq \int \left| K - K' \right| \) when \( u_{nh} \) is the kernel estimate with kernel \( K' \).
The fact that the bound does not depend upon $h$ and that $K'$ is arbitrary means that we need only show the Theorem for all $K$ that are continuous and of compact support (since the latter collection is dense in the space of $L_1$ functions).

The following inequality is valid for all fixed $h$, $n$, $K$ and $f$:

$$P\left( \left| \int f_{nh} - f - \mathbb{E} \int f_{nh} - f \right| > \varepsilon \right) \leq 2e^{-\left(\frac{\log 2}{2}\right) \left( \int |K| \right)^2}$$

(Devroye, 1988). We employ the grid technique of Theorem C2 again. Set

$$\Delta(h) = \left| \int f_{nh} - f - \mathbb{E} \int f_{nh} - f \right|,$$

and $\Delta(a, b) = \sup_{h, h' \in [a, b]} |\Delta(h) - \Delta(h')|$. Then let $c > 1$ be such that

$$\sup_{1 \leq h \leq c} \int |K_1 - K_h| \leq \frac{\varepsilon}{4}.$$  

Noting that

$$\left| \int f_{nh} - f - \int f_{nh'} - f \right| \leq \int |f_{nh} - f_{nh'}| \leq \int |K_h - K_{h'}|,$$

we see that as in the proof of Theorem C2,

$$P\left( \sup_{a \leq h \leq b} \Delta(h) > \varepsilon \right) \leq \sum_{i=0}^{k} P\left( \Delta(ac^i) > \frac{\varepsilon}{2} \right) \leq 2 \left( 1 + \frac{\log (b/a)}{\log c} \right) e^{-\left(\frac{\log 2}{2}\right) \left( \int |K| \right)^2}.$$  

It suffices to have limits $a$ and $b$ that are such that $b/a = O(n)$, for this upper bound to tend to zero with $n$ at an exponential rate. We need only establish that for some sequences $a = a(n)$ and $b = b(n)$ with $b/a = O(n)$ that

$$\lim_{n \to \infty} \left( P\left( \sup_{h < a} \Delta(h) > \varepsilon \right) + P\left( \sup_{h > b} \Delta(h) > \varepsilon \right) \right) = 0.$$  

Assume that $K$ vanishes off $[-1, 1]$. Take $a = \frac{\delta}{2n}$ where $\delta > 0$ is a constant to be picked further on. Let $N$ be the number of $X_i$'s for which $[X_i - 2a, X_i + 2a]$ has at least one $X_j$ with $j \neq i$, and let $A$ be the union of the sets $[X_i - a, X_i + a]$ for those $X_i$ not counted in $N$. Note that

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
THE DOUBLE KERNEL METHOD

\[ 1 \geq \int_A |f_{nh}| \mathcal{L} K \geq \frac{n-N}{n}, \text{ uniformly over } h \leq a. \text{ We have} \]

\[ P \left( \sup_{h < a} \left( \int |f_{nh} - f| - \left( 1 + \int |K| \right) \right) > \varepsilon \right) \]

\[ + P \left( \sup_{h < a} \left( \mathbb{E} \int |f_{nh} - f| - \left( 1 + \int |K| \right) \right) > \frac{\varepsilon}{2} \right). \]

We claim that

\[ P \left( \sup_{h < a} \left( \int |f_{nh} - f| - \left( 1 + \int |K| \right) \right) > \frac{\varepsilon}{2} \right) \leq P \left( \sup_{h < a} \int |f_{nh}| \leq \int |K| - \frac{\varepsilon}{6} \right) \]

if \( \delta < \rho(f, \varepsilon) \) for some positive-valued function \( \rho \). Indeed, since

\[ \int |f_{nh} - f| \leq 1 + \int |K|, \] it suffices to consider only one kind of signed difference. Take \( \delta \) so small that uniformly over all sets \( B \) with \( \lambda(B) < \delta \),

\[ \int f < \varepsilon/12 \] where \( \lambda \) is Lebesgue measure (this is always possible). Note in passing that \( \lambda(A) \leq \delta \). It suffices to show that

\[ \sup_{B : \lambda(B) < \delta} \int |f_{nh}| \geq \int |K| - \varepsilon/6 \] implies that

\[ \int |f_{nh} - f| \geq 1 + \int |K| - \varepsilon/2. \] We have

\[ \int |f_{nh} - f| \geq \int |K| - \varepsilon/6 - \int f, \int |f_{nh} - f| \geq \int f - \int |f_{nh}|, \] which is at least \( 1 - \int f - \varepsilon/6 \). Summing this and noting that \( \int f \leq \varepsilon/12 \) shows that

\[ \int |f_{nh} - f| \geq 1 + \int |K| - \varepsilon/2. \]

The previous facts can now be combined to conclude that for \( \delta \) small enough,

\[ P \left( \sup_{h < a} \Delta(h) > \varepsilon \right) \leq P \left( \frac{N}{n} > \frac{\varepsilon}{6} \right) + P \left( \frac{EN}{n} > \frac{\varepsilon}{6} \right). \]

This is handled precisely as in the proof of Theorem C2. Thus, \( \lim_{n \to \infty} P \left( \sup_{h < a} \Delta(h) > \varepsilon \right) = 0 \) for \( a = \delta/(2n) \) and \( \delta \) small enough. In fact, the said probability does not exceed \( e^{-dn} \) for some constant \( d > 0 \) depending upon \( \varepsilon \).

We finally proceed to show that

\[ P \left( \sup_{h > b} \Delta(h) > \varepsilon \right) \]
can be made exponentially small in $n$ by choosing a large enough constant $b$. This would conclude the proof of the Theorem since $b/a = O(n)$ as required. Let $\omega$ be the modulus of continuity of $K$ defined by $\omega(u) = \sup_x \sup_y |K(x) - K(x + y)|$. By our assumptions on $K$, $\omega(u) \to 0$ as $u \downarrow 0$. Let $t$ and $T > t$ be positive numbers chosen in such a way that $\int_{|x| \leq T} |K| \geq \int |K| - \frac{6}{8}$ and $\sup_z \int_{z-t}^{z+t} |K| \leq \frac{\epsilon}{8}$. Also, $T$ should be so large that $\int_{|x| \geq T} f < \epsilon \left(12 \int |K|\right)$. This fixes $t$ and $T$ once and for all.

Let $N$ be the number of $X_i$'s with $|X_i| \geq T$. We have the following inequality:

$$\int_{|x| \leq T} |f_{nh} - K_h| \leq (T-t) \omega\left(\frac{T}{b}\right) + \frac{N}{n} \int |K|.$$ 

This can best be seen by noting that

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

$$\leq \frac{1}{n} \sum_{i : |X_i| \geq T} \sup_{|y| \leq T} |K_h(x - y) - K_h(x)| + \frac{1}{n} \sum_{i : |X_i| > T} |K_h(x - X_i)|$$

$$\leq \frac{1}{h} \omega\left(\frac{T}{h}\right) + \frac{1}{n} \sum_{i : |X_i| > T} |K_h(x - X_i)|.$$

Now, integrating over the given interval and noting that $h \geq b$ yields the result. For $h \geq b$,

$$\int |f_{nh} - f| \geq \int_{|x| \leq T} |f_{nh} - f| \geq \int_{|x| \leq T} |K_h - f| + \int_{|x| \leq T} |f - f_{nh}|$$

$$\geq \int_{|x| \leq T} |K_h| - \int_{|x| \leq T} |f_{nh} - K_h| - \int_{|x| \leq T} f$$

$$\geq \int_{|x| \leq T} f_{nh} - \frac{1}{n} \sum_{i=1}^{n} \int_{|x| \leq T} |K_h(x - X_i)|$$

$$\geq \left(\int |K| - \frac{\epsilon}{8} - (T-t) \omega\left(\frac{T}{h}\right) - \int |K| \frac{N}{n} + 1 - 2 \int_{|x| \leq T} f_{nh} \sup_{|z| \leq T} z + \int_{|x| \leq T} |K| \frac{N}{n} + 1 - 2 \int_{|x| \leq T} f - \frac{\epsilon}{8} \right)$$

$$\geq \int |K| + 1 - \frac{\epsilon}{3} - \int |K| \frac{N}{n}.$$
THE DOUBLE KERNEL METHOD

if $b$ is so large that $(T-t)\omega(T/b) + 2\int_{t+h \leq |x|} f \leq \varepsilon/12$. Thus,

$$
P\left( \inf_{h \geq b} \int |f_{nh} - f| \leq 1 + \int |K| - \frac{\varepsilon}{2} \right) \leq P\left( \frac{N}{n} \geq \frac{\varepsilon}{6 \int |K|} \right)
$$

and

$$
\inf_{h \geq b} E \int |f_{nh} - f| \geq 1 + \int |K| - \frac{\varepsilon}{2}
$$

when $EN/n \leq \varepsilon/(6 \int |K|)$. Now, $EN/n = \int_{|x| \geq T} f < \varepsilon/(12 \int |K|)$ by our choice of $T$. By Hoeffding’s inequality (Hoeffding, 1963),

$$
P\left( \frac{N}{n} \geq \frac{\varepsilon}{6 \int |K|} \right) \leq P\left( \frac{N - EN}{n} \geq \frac{\varepsilon}{12 \int |K|} \right) \leq e^{-\left(2mc^2/144 \int |K| \right)}.
$$

In conclusion, for our choice of $t$, $T$ and $b$,

$$
P\left( \sup_{h \geq b} \Delta(h) > \varepsilon \right) \leq P\left( \sup_{h \geq b} \left| \int |f_{nh} - f| - \left( 1 + \int |K| \right) \right| > \frac{\varepsilon}{2} \right)
$$

$$
+ P\left( \sup_{h \geq b} E \int |f_{nh} - f| - \left( 1 + \int |K| \right) > \frac{\varepsilon}{2} \right) \leq e^{-\left(2mc^2/144 \int |K| \right)}.
$$

This concludes the proof of Lemma S2. ■

Proof of Theorem S1. — Let us first try to prove that for arbitrary fixed $\varepsilon > 0$,

$$
P\left( H \notin \left[ \frac{1}{n\varepsilon}, \varepsilon \right] \right) < \frac{1}{n^2}
$$

for all $n$ large enough, and that $H \to 0$ and $n H \to \infty$ completely. This statement parallels that of Lemma O2. It suffices to replace $g_{nh}$ throughout by $f$ and $K - L$ by $K$. Also, the constant $C$ now becomes the smoothness constant for $K$. It is easy to see then that we need only two facts at this stage:

$$
\liminf_{n \to \infty} \inf_{h \notin \{1/(nm), \varepsilon\}} E\left( \int |f_{nh} - f| \right) > 0
$$

and for fixed $u > 0$,

$$
P\left( \sup_{h} \left| \int |f_{nh} - f| - E\int |f_{nh} - f| \right| > u \right) \leq e^{-\gamma n}
$$

Vol. 25, n° 4-1989.
where $\gamma = \gamma(u) > 0$. These were proved in Lemmas S1 and S2.

We note now that the inequalities of Lemma O4 apply without change to the present $H$. In particular,

$$E \int |f_{nH} - f| - E \int |f_{nH} - f| = O\left(\sqrt{\frac{\log n}{n}}\right),$$

where $H$ is distributed as $H$ but independent of the data stream. From Lemma O5 we retain that for any $f$

$$\lim \inf_{n \to \infty} n^{q/2s+1} E \int |f_{nH} - f| \geq c > 0,$$

for some constant $c > 0$. Hence, this bound also applies if $H$ is replaced by $H$. In fact, we have

$$\lim_{n \to \infty} \frac{E \int |f_{nH} - f|}{E \int |f_{nH} - f|} = 1$$

for all densities $f$. Thus,

$$E \int |f_{nH} - f| \sim E \int |f_{nH} - f| \leq (1 + O(1)) \inf_h E \int |f_{nh} - f| \leq (1 + O(1)) E \int |f_{nH} - f|,$$

which shows the first part of the Theorem. The strong convergence is obtained from the probability bound of Lemma O4 generalized above, the asymptotic lower bound of Lemma O5 (also generalized above), and the Borel-Cantelli lemma (the sequence $2/n^2$ is summable in $n$).

ACKNOWLEDGEMENTS

This research was sponsored by N.S.E.R.C. Grant A3456 and F.C.A.C. Grant EQ-1678.

REFERENCES

S. ABOU-JAOUDE, La convergence L1 et L infini de certains estimateurs d'une densité de


(Manuscript received November 8th, 1988)  
(accepted June 19th, 1989.)