# RANDOM APPROXIMATIONS TO SOME MEASURES

# OF ACCURACY IN NONPARAMETRIC CURVE ESTIMATION

by

James Stephen Marron[1]

University of North Carolina
Chapel Hill, N.C.   27514


and


Wolfgang Härdle[2]

Universität Heidelberg
Sonderforschungsbereich 123,
Heidelberg, West Germany

## SUMMARY

   This paper deals with a quite general nonparametric statistical
curve estimation setting.  Special cases include estimation of prob-
ability density functions, regression functions and hazard functions.
The class of "fractional delta sequence estimators" is defined and
treated here.  This class includes the familiar kernel, orthogonal
series and histogram methods.  It is seen that, under some mild
assumptions, both the Average Square Error and Integrated Square
Error provide reasonable (random) approximations to the Mean Integrated
Square Error.  This is important for two reasons.  First, it provides
theoretical backing to a practice that has been employed in several
simulation studies.  Second, it provides a vital tool for proving
theorems about selecting smoothing parameters for several different
nonparametric curve estimators.

## 1. Introduction

Let $X, X_1, \cdots, X_n$ be a random sample of d-dimensional random vector having density function $f(x)$ and cumulative distribution function $F(x)$. Suppose we are interested in a certain functional $g(x)$, $x \in \mathbb{R}^d$ of the distribution of X. The problem of estimating the curve $g(x)$ from the random sample is called <u>nonparametric curve estimation</u>.

Some special cases of nonparametric curve estimation are:

<u>D- density estimation</u>: where g is taken to be f.

<u>H- hazard function estimation</u>: where g is given by

$$g(x) = \frac{f(x)}{1-F(x)} .$$

<u>R- Regression estimation</u>: where g is the regression curve of Y on X',

$$g(x) = g(x') = E[Y|X' = x'] ,$$

using the notation:

$$d' = d-1 ,$$

$$x' = (x^{(1)}, \ldots, x^{(d')}) ,$$

(1.1)

$$x = (x^{(1)}, \ldots, x^{(d')}, y) ,$$

$$X = (X^{(1)}, \ldots, X^{(d')}, Y) .$$

This list of examples is meant to be representative, not exhaustive. Other possibilities include derivatives of the above functions.

Quite a number of different estimators have been proposed for each of the curves given above. For comparison of these estimators, several measures of accuracy have been considered. A very common measure is the Mean Integrated Square Error,

$$\text{MISE} = E \int [\hat{g}(x)-g(x)]^2 w(x)dx ,$$

with some nonnegative weight function w(x).

While MISE is theoretically pleasing as a distance between $\hat{g}$ and g, it is often hard to compute. The literature contains two different ways of overcoming this difficulty. The first is to study the asymptotic (as $n \to \infty$) behavior of MISE. The second is to consider Monte Carlo (and hence random) approximations to MISE. In this paper it is seen that, for many estimators, these two approaches give quite similar results for large values of n.

Stochastic (ie: random) distances that have been considered include the Integrated Square Error, given by

$$ISE = \int [\hat{g}(x) - g(x)]^2 w(x) dx,$$

and the Average Square Error, given by

$$ASE = n^{-1} \sum_{i=1}^{n} [\hat{g}(X_i) - g(X_i)]^2 \tilde{w}(X_i) .$$

Wegman (1972), argued in the setting of density estimation that if $\tilde{w}(x)$ is defined by

(1.2) $$\tilde{w}(x) f(x) = w(x) ,$$

then, for n large, ASE should be a good approximation of MISE.

He then used ASE as a distance measure for a Monte Carlo comparison of several density estimators. ASE has also been employed for this purpose by Fryer (1977) and Wahba (1977). Breiman, Meisel and Purcell (1977) and Raatgever and Duin (1978) used a "normalized version" of ASE in their Monte Carlo studies. The distance ISE also has been attractive to several authors, see for example Rust and Tsokos (1981)

Scott and Factor (1981), Bean and Tsokos (1982) and Bowman (1982).

In the regression setting, Stone (1976) has used a "leave-one-out"

version of ASE and Engle, Granger, Rice and Weiss (1983) and Silver-

man (1984) have used ASE to study cross-validated estimators. In

the hazard function setting, Tanner and Wong (1982) have compared

two estimators by computing the difference of their ASE's.

The use of ASE and ISE as measures of accuracy was criticized by

Steele (1978) who gave an example in which, asymptotically as $n \to \infty$,

ASE behaved very differently from ISE (hence, at least one is a poor

approximation to MISE). In reply to this objection, Hall (1982)

showed that Steele's example was somewhat artificial by showing that,

in the case d=1, if $\hat{g}(x)$ is a kernel density estimator, then under

some reasonable assumptions, as $n \to \infty$,

$$(1.3) \qquad \text{ASE} = \text{MISE} + o_p(\text{MISE}) ,$$

$$(1.4) \qquad \text{ISE} = \text{MISE} + o_p(\text{MISE}) ,$$

and if $\hat{g}(x)$ is a trigonometric series density estimator (1.4) holds.

The object of this paper is two-fold. First, Hall's results are

extended to a wider class of estimators and to a variety of nonpar-

ametric curve estimation settings. This demonstrates that the objections

of Steele (1978) need cause no concern in the case of many commonly

considered estimators. Second, the results of this paper provide

an important tool for use in analyzing curve estimators with data-based

smoothing parameter selection. In particular, asymptotic optimality

results can be derived from suitable uniform versions of (1.3) and (1.4).

Special cases of this may be seen, either explicitly or implicitly, in

the results of Hall (1983),

Stone (1984a), Stone (1984b), Burman (1984) and Marron (1984) in the density estimation setting, and in the results of Rice (1984) and Marron and Härdle (1984), in the regression setting.

Section 2 introduces the class of "fractional delta sequence estimators" and makes evident that many of the most widely studied nonparametric estimators are contained in this class. Section 3 contains theorems which give sufficient conditions for (1.3) and (1.4) for a subset of these estimators. Section 4 contains theorems which extend the results of Section 3 to all fractional delta sequence estimators. Section 5 contains examples for illustration of these theorems. The proofs of the theorems are in section 6. The convention (1.2) will be used throughout this paper.

## 2. Fractional Delta Sequence Estimators

The class of _fractional delta sequence estimators_ is defined to consist of all estimators of the form

$$(2.1) \qquad \hat{g}(x) = \frac{\sum\limits_{i=1}^{n} \delta_\lambda(x, X_i)}{\sum\limits_{i=1}^{n} \delta_\lambda{}'(x, X_i)}$$

where $\delta_\lambda$ and $\delta_\lambda{}'$ are measurable functions on $\mathbb{R}^d \times \mathbb{R}^d$, which are indexed by a "smoothing parameter" $\lambda = \lambda(n) \in \mathbb{R}^+$. The special case $\delta_\lambda{}'(x, X_i) \equiv 1$ gives the delta sequence estimators studied by Watson and Leadbetter (1964b), Földes and Revesz (1974), and Walter and Blum (1979), among others.

In the setting of density estimation, some well known estimators of this type are:

D-1 kernel estimators: Introduced by Rosenblatt (1956) and Parzen (1962), given a "kernel function," $K: \mathbb{R}^d \to \mathbb{R}$, and the smoothing parameter, $\lambda \in \mathbb{R}^+$, define

$$(2.2) \qquad \delta_\lambda(x, X_i) = \lambda K(\lambda^{1/d}(x - X_i)),$$

$$\delta_\lambda{}'(x, X_i) \equiv 1.$$

D-2 histogram estimators: Let B denote a bounded subset of $\mathbb{R}^d$, and suppose that (given the smoothing parameter $\lambda \in \mathbb{Z}^+$) $A_1, \ldots, A_\lambda$ form a partition of B. For $\ell = 1, \ldots, \lambda$ let $1_\ell(x)$ denote the indicator of $A_\ell$ and let $\mu_\ell$ denote the Lebesque measure of $A_\ell$.

Define

$$(2.3) \qquad \delta_\lambda(x, X_i) = \sum\limits_{\ell=1}^{\lambda} \mu_\ell^{-1} 1_\ell(x) 1_\ell(X_i)$$

$$\delta_\lambda{}'(x, X_i) \equiv 1.$$

<u>D-3 orthogonal series estimators</u>:  Introduced by Cencov (1962).

Suppose $\{\psi_\ell(x)\}$ is a sequence of functions which is orthonormal

and complete with respect to the inner product

(2.4)        $\int \psi_\ell(x)\psi_{\ell'}(x)w(x)dx$ .

Given the smoothing parameter $\lambda \in \mathbb{Z}^+$ , define

$$\delta_\lambda(x,X_i) = \sum_{\ell=1}^{\lambda} \psi_\ell(x)\psi_\ell(X_i)\tilde{w}(X_i) ,$$

(2.5)

$$\delta_\lambda'(x,X_i) \equiv 1 .$$

Further examples of delta sequence density estimators may be found

in Walter and Blum (1979) and Susarla and Walter (1981).  Some examples

of fractional delta sequence estimators in the regression setting are:

<u>R-1 kernel estimators</u>:  Introduced by Nadaraya (1964) and Watson (1964).

Given a kernel function, $K(x')$ and a smoothing parameter, $\lambda$, using the

notation (1.1), define

$$\delta_\lambda(x,X_i) = \lambda K(\lambda^{1/d'}(x'-X_i'))Y_i$$

$$\delta_\lambda'(x,X_i) = \lambda K(\lambda^{1/d'}(x'-X_i')) .$$

Note that, $\hat{g}(x)$ is a weighted average of the $Y_i$ .

<u>R-2 known-marginal kernel estimators</u>:  Studied by Johnston (1982).

Let $f_M(x')$ denote the marginal density of $X_i'$ and define

$$\delta_\lambda(x,X_i) = \lambda K(\lambda^{1/d'}(x'-X_i'))Y_i$$

$$\delta_\lambda'(x,X_i) = f_M(x') .$$

To see the idea behind this estimator, note that when the denominator of R-2 is properly normalized, it becomes the estimate D-1 of the marginal density, $f_M(x')$ .

R-3 delta sequence estimators: A generalization of R-1, discussed in Collomb (1981), define $\tilde{\delta}_\lambda(x',X_i')$ as for any of the density estimators and let

$$\delta_\lambda(x,X_i) = \tilde{\delta}_\lambda(x',X_i')Y_i \, ,$$

$$\delta_\lambda'(x,X_i) = \tilde{\delta}_\lambda(x',X_i') \, .$$

Note that the regressogram of Tukey (1961) is a special case where $\tilde{\delta}_\lambda$ is defined as for D-2.

In the setting of hazard function estimation, Watson and Leadbetter (1964a) have introduced the following fractional delta sequence estimators.

H-1 kernel estimators: Given a kernel function, K(x), and a smoothing parameter, $\lambda$, define

(2.6)
$$\delta_\lambda(x,X_i) = \lambda K(\lambda(x-X_i)) \, ,$$

$$\delta_\lambda'(x,X_i) = 1- \int_{-\infty}^x \lambda K(\lambda(t-X_i))dt \, .$$

H-2 delta sequence estimators: A straightforward generalization of H-1, define $\delta_\lambda(x,X_i)$ as in any of the density estimators and let

$$\delta_\lambda'(x,X_i) = 1- \int_{-\infty}^x \delta_\lambda(t,X_i)dt \, .$$

## 3. Approximation Theorems for Delta Sequence Estimators

This section gives sufficient conditions for (1.3) and (1.4) in

the special case of delta sequence estimators, which are of the form

$$(3.1) \qquad \hat{g}(x) = n^{-1} \sum_{i=1}^{n} \delta_\lambda(x, X_i) = \int \delta_\lambda(x, x_1) dF_n(x_1) .$$

Assume that $\lambda$ ranges over a finite set $\Lambda_n$, whose cardinality is bounded by

$$(3.2) \qquad \#(\Lambda_n) \le C n^\rho , \quad \rho > 0$$

(ie: is increasing at most algebraically fast.)

For estimators with a continuous smoothing parameter, such as the kernel

estimators, the results of this paper can be easily extended to $\Lambda_n$ an

interval by a continuity argument (compare Marron (1984) and Härdle and

Marron (1984).)

For ease of presentation, it will be assumed that there are constants

$C$ and $\epsilon > 0$ so that, for each n, and for all $\lambda \in \Lambda_n$ ,

$$(3.3) \qquad C^{-1} n^\epsilon \le \lambda \le C n^{1-\epsilon} .$$

The next assumptions are rather technical in nature, but are stated

in this form because these are the common properties which make all of

the diverse estimators of section 2 satisfy (1.3) and (1.4). In section

5 it is seen that quite different methods are required to verify these

for the different estimators.

For k=1,2,... assume there is a constant $C_k$ so that for any m=2,

..., 2k and $\lambda \ge 1$ ,

$$(3.4) \quad \left| \int \cdots \int \left[ \prod_{i,i'=1}^{m} \delta_\lambda(x_i, x_{i'})^{\alpha_{ii'}} \right] \left[ \prod_{i=1}^{m} \tilde{w}(x_i)^{\beta_i} \right] dF(x_1) \cdots dF(x_m) \right| \le C_k \lambda^{k-m/2} ,$$

where $\alpha_{ii'} = 0,\ldots,k$ subject to

$$\sum_{i,i'=1}^{m} \alpha_{ii'} = k$$

and the restriction that for each $i=1,\ldots,m$, there is an $i' \neq i$ so that either $\alpha_{ii'}$ or $\alpha_{i'i}$ is nonzero, and where $\beta_i = 0,1$ with $\beta_i = 1$ any time an $\alpha_{i,i'} \geq 1$.

Assume there is a constant $C$ so that

$$(3.5) \qquad \iint \delta_\lambda(x_1,x_2)^2 \, w(x_1)dx_1 dF(x_2) \geq C\lambda .$$

Assume that the quantity

$$(3.6) \qquad \tilde{\delta}_\lambda(x_1,x_2) = \int \delta_\lambda(x_3,x_1)\delta_\lambda(x_3,x_2)w(x_3)dx_3$$

satisfies the assumption (3.4), with each $\beta_i = 0$, and that there is a constant $C$ so that

$$(3.7) \qquad \iint \tilde{\delta}_\lambda(x_1,x_2)dF(x_1)dF(x_2) \leq C .$$

Another assumption is that there is a constant $\epsilon > 0$, so that for $k=1,2,\ldots$ there is a constant $C_k$ such that

$$(3.8) \qquad \int B(x)^{2k}w(x)dx \leq C_k b(\lambda)\lambda^{(k-1)(1-\epsilon)} ,$$

where $B(x)$ denotes the bias and $b(\lambda)$ denotes the integrated squared bias of the estimator $\hat{g}$ given by

$$B(x) = E[\hat{g}(x)]-g(x) = \int \delta_\lambda(x,x_2)dF(x_2)-g(x) ,$$

$$(3.9)$$

$$b(\lambda) = \int B(x)^2 w(x)dx .$$

Finally assume that for k=1,2,... there is a constant $C_k$ so that

$$(3.10) \qquad \int [\delta_\lambda(x,x)]^{2k} w(x) dx \le C_k \lambda^{2k} .$$

Theorem 1: <u>Under the assumptions</u> (3.1)-(3.7),

$$\lim_{\substack{n \to \infty \\ \lambda \in \Lambda_n}} \sup \left| \frac{ISE(\lambda) - MISE(\lambda)}{MISE(\lambda)} \right| = 0 \qquad a.s.$$

Theorem 2: <u>Under the assumptions (3.1)-(3.10), and</u> $\tilde{w}$ <u>bounded</u>,

$$\lim_{\substack{n \to \infty \\ \lambda \in \Lambda_n}} \sup \left| \frac{ASE(\lambda) - MISE(\lambda)}{MISE(\lambda)} \right| = 0 \qquad a.s.$$

<u>Remark</u>: We believe that the proofs of these approximations can be extended to the case of $\lambda$ a vector or even a matrix, but additional messy notation and tedious work are required for this.

## 4. Approximation Theorems for Fractional Delta Sequence Estimators

This section extends Theorems 1 and 2 to include fractional delta sequence estimators. Since these estimators have denominators containing random variables, they are technically more difficult to work with. In fact, for the estimator R-1, if the kernel function, K, is allowed to take on negative values, then the moments of $\hat{g}(x)$ may not exist (see Rosenblatt (1969) and Härdle and Marron (1983)) so MISE is not a reasonable distance.

To overcome these difficulties, assume there is a function $D(x)$ and a set $S \subseteq \mathbb{R}^d$ so that, uniformly over $x \in S$, $\lambda \in \Lambda_n$

$$(4.1) \qquad n^{-1} \sum_{i=1}^{n} \delta_\lambda'(x, X_i) \to D(x) \qquad \text{a.s.}$$

and assume that

$$(4.2) \qquad \inf_{x \in S} D(x) > 0 .$$

Then, uniformly over $x \in S$, $\lambda \in \Lambda_n$,

$$\hat{g}(x) - g(x) = n^{-1} \sum_i \left[ \frac{\delta_\lambda(x, X_i) - \delta_\lambda'(x, X_i) g(x)}{D(x)} \right]$$

$$+ \frac{[D(x) - n^{-1}\sum_i \delta_\lambda'(x, X_i)] n^{-1} \sum_i [\delta_\lambda(x, X_i) - \delta_\lambda'(x, X_i) g(x)]}{D(x) n^{-1} \sum_i \delta_\lambda'(x, X_i)}$$

$$= n^{-1} \sum_i \delta_\lambda^*(x, X_i) + o(n^{-1} \sum_i \delta_\lambda^*(x, X_i)) ,$$

where

$$(4.3) \qquad \delta_\lambda^*(x, X_i) = [\delta_\lambda(x, X_i) - \delta_\lambda'(x, X_i) g(x)] / D(x) .$$

Thus, for $w(x)$ supported inside S, it makes sense to replace MISE by

$$(4.4) \qquad \text{MISE*} = E \int [n^{-1} \sum_{i=1}^{n} \delta_\lambda^*(x,X_i)]^2 w(x) dx .$$

Similarly, ISE and ASE may be replaced with

$$\text{ISE*} = \int [n^{-1} \sum_{i=1}^{n} \delta_\lambda^*(x,X_i)]^2 w(x) dx ,$$

$$(4.5)$$

$$\text{ASE*} = n^{-1} \sum_{j=1}^{n} [n^{-1} \sum_{i=1}^{n} \delta_\lambda^*(X_j,X_i)]^2 \tilde{w}(X_j) ,$$

Before the theorems are stated, note that MISE* may be considered to be an assessment of how accurately the delta sequence estimator $\hat{g}^*(x)$, defined by

$$\hat{g}^*(x) = n^{-1} \sum_{i=1}^{n} \delta_\lambda^*(x,X_i) ,$$

estimates the function $g^*(x)$, defined by

$$g^*(x) \equiv 0 .$$

Similarly ISE* and ASE* are the ISE and ASE for this new estimation problem. This observation allows immediate application of Theorems 1 and 2.

Theorem 3: If $\delta_\lambda^*$ satisfies the assumptions (3.1)-(3.7) then

$$\limsup_{\substack{n \to \infty \\ \lambda \in \Lambda_n}} \left| \frac{\text{ISE*}(\lambda) - \text{MISE*}(\lambda)}{\text{MISE*}(\lambda)} \right| = 0 \qquad \text{a.s.}$$

Corollary: If, in addition, (4.1) holds, then

$$\lim_{\substack{n \to \infty \\ \lambda \in \Lambda_n}} \sup \left| \frac{ISE(\lambda) - MISE^*(\lambda)}{MISE^*(\lambda)} \right| = 0 \qquad a.s.$$

Theorem 4: If $\delta_\lambda^*$ satisfies the assumptions (3.1)-(3.10) and $\tilde{w}$ is is bounded, then

$$\lim_{\substack{n \to \infty \\ \lambda \in \Lambda_n}} \sup \left| \frac{ASE^*(\lambda) - MISE^*(\lambda)}{MISE^*(\lambda)} \right| = 0 \qquad a.s.$$

Corollary: If, in addition, (4.1) holds, then

$$\lim_{\substack{n \to \infty \\ \lambda \in \Lambda_n}} \sup \left| \frac{ASE(\lambda) - MISE^*(\lambda)}{MISE^*(\lambda)} \right| = 0 \qquad a.s.$$

To see how Theorems 1 and 2 are intimately related to Theorems 3 and 4, note that in the special case where $\hat{g}(x)$ is a delta sequence estimator (ie: $\delta_\lambda'(x, X_i) \equiv 1$), conditions (4.1) and (4.2) hold trivially and the quantities MISE*, ISE* and ASE* are the same as their unasterisked counterparts. Thus Theorems 1 and 2 are special cases of Theorems 3 and 4. On the otherhand, using the viewpoint given above, Theorems 3 and 4 are consequences of Theorems 1 and 2.

## 5. Examples

In this section it is seen how the fractional delta sequence estimators of section 2 satify the conditions of sections 3 and 4.

**D-1 kernel estimators:** Conditions (3.4)-(3.7) follow easily from integration by substitution and the assumptions that f, w and K are bounded with $\int K(x)\, dx = 1$ and f, w not mutually singular. Condition (3.8) is also easily satisfied with $\epsilon = 1$. Condition (3.10) requires the additional assumption that w be integrable. Thus the results of Marron (1985) and Theorems 1 and 2 of Hall (1982) are special cases of the results of this paper.

**D-2 histogram estimators:** In this setting, a reasonable assumption is that there is a constant $C$ so that, for $\lambda = 1, 2, \ldots,\ \ell = 1, \ldots, \lambda$,

$$(5.1) \qquad C\, \lambda^{-1} \le \mu_\ell$$

From this it follows that

$$\sup_{x_1, x_2} \delta_\lambda(x_1, x_2) \le \sup_\ell \mu_\ell^{-1} \le C^{-1}\lambda$$

and

$$\sup_{x_2} \int \delta_\lambda(x_1, x_2)\, dx_1 = \sup_{x_2} \sum_{\ell=1}^{\lambda} 1_\ell(x_2) = 1 .$$

(and similarly for $\delta_\lambda(x_2, x_2)$). Condition (3.4) is an easy consequence of this and the assumption of w and f bounded. To check (3.5) note that

$$\iint \delta_\lambda(x_1, x_2)^2 w(x_1)\, dx_1\, dF(x_2) = \sum_{\ell=1}^{\lambda} \mu_\ell^{-2} \int_{A_\ell} \int_{A_\ell} w(x_1)\, dx_1\, dF(x_2) ,$$

so it is enough to assume that w and f are

not mutually singular.   Next observe that

$$\tilde{\delta}_\lambda(x_1,x_2) = \sum_{\ell=1}^{\lambda} \mu_\ell^{-2} \int_{A_\ell} w(x)\,dx\, 1_\ell(x_1) 1_\ell(x_2) \; ,$$

so (3.6) and (3.7) follow from the above assumptions.   Condition

(3.8) is also easily satisfied with $\epsilon=1$ , and (3.10) follows from

(5.1).

D-3 orthogonal series estimators:   The assumptions needed to verify

(3.4) are summarized in

lemma 1:   If, for k=1,2,... there is a constant $C_k$ so that for

$\ell_1,\dots,\ell_k=1,2,\dots$ and for r=1,...,k,

(5.2) $$\int \psi_{\ell_1}^2(x) \dots \psi_{\ell_k}^2(x)\tilde{w}(x)^r dF(x) \leq C_k^2 \; ,$$

then (3.4) holds.

The proof of this lemma is in section 7.   Note that (5.2) is easily

satisfied for either the familiar trigonometric or Hermite series.   Next

observe that

$$\iint \delta_\lambda(x_1,x_2)^2 w(x_1)\,dx_1\,dF(x_2) = \int \sum_{\ell=1}^{\lambda} \psi_\ell(x_2)^2 \tilde{w}(x_2)^2 dF(x_2) \; ,$$

so (3.5) is easily satisfied.   Condition (3.6) follows from

(5.3) $$\tilde{\delta}_\lambda(x_1,x_2) = \delta_\lambda(x_1,x_2)\tilde{w}(x_1) \; ,$$

and the assumption that $\tilde{w}$ is bounded.   Condition (3.7) follows from

(5.3) together with the Schwartz Inequality.   Verification of (3.8)

follows easily from

$$\sup_{x_1} \; [\int \delta_\lambda(x_1,x_2)dF(x_2) - f(x_1)]^2 \; w(x_1) \le C \lambda^{(1-\epsilon)} \; ,$$

which is easy to check in the Hermite series case, but, using the computations of Hall (1981), requires the additional assumption of f'' bounded in the case of trigonometric series. Condition (3.10) is obvious under the above assumptions for either the trigonometric or Hermite series. Theorem 3 of Hall (1982) is a special case of this.

R-1 kernel estimators: The conditions (3.4)-(3.10) are easily verified under the same assumptions as D-1, above, together with the assumption that for k=1,2,... there is a constant $C_k$ so that, for x' in the support of w' ,

$$E[Y^k|X'=x'] \le C_k \; .$$

The verification of (4.1) is easy, in view of Lemma 1 of Härdle and Marron (1984), under the additional assumption that $f_M$ is Hölder continuous.

R-2 known marginal kernel estimators

This case is similar to R-1 except that (4.1) is not required. R-1 and R-2 contain the results of Härdle (1984) as special cases.

H-1 kernel estimators: The conditions (3.4)-(3.10) are easily checked when it is assumed that

$$\int K(x)\,dx = 1 \; ,$$

and K, f and w are bounded, together with the assumption that 1-F is bounded above 0 on the support of w.

## 6. Proofs of Theorems 1 and 2

Note that, by (3.2) and the Chebyshev Inequality, for $\epsilon > 0$, $k=1,2,\ldots$

$$P\left[\sup_{\lambda \in \Lambda_n} \left| \frac{ISE(\lambda)-MISE(\lambda)}{MISE(\lambda)} \right| > \epsilon \right] \leq$$

$$Cn^\rho \sup_{\lambda \in \Lambda_n} E\left[\frac{ISE(\lambda)-MISE(\lambda)}{MISE(\lambda)}\right]^{2k} .$$

Thus, by the Borel-Cantelli Lemma, the proof of Theorem 1 will be complete when it is seen that there is a constant $\gamma > 0$, so that for $k=1,2,\ldots$, there are constants $C_k$ so that

$$(6.1) \qquad E\left[\frac{ISE(\lambda)-MISE(\lambda)}{MISE(\lambda)}\right]^{2k} \leq C_k n^{-\gamma k} .$$

Theorem 2 will be established by the same technique when it is shown that

$$(6.2) \qquad E\left[\frac{ASE(\lambda)-MISE(\lambda)}{MISE(\lambda)}\right]^{2k} \leq C_k n^{-\gamma k} ,$$

The distance ISE can be decomposed as:

$$ISE = R(\lambda) + 2S(\lambda) + b(\lambda)$$

where $b(\lambda)$ is defined in (3.9) and

$$R(\lambda) = \iiint \delta_\lambda(x_1,x_2)\, \delta_\lambda(x_1,x_3)\, w(x_1)\, dF(x_1)\, d(F_n-F)(x_2)\, d(F_n-F)(x_3) ,$$

$$S(\lambda) = \iint \delta_\lambda(x_1,x_2)\, B(x_1)\, w(x_1)\, dx_1\, d(F_n-F)(x_2) .$$

The first term may be further split into

$$R(\lambda) = R_1(\lambda) + R_2(\lambda) + R_3(\lambda)$$

where, using the notation (3.6),

$$R_1(\lambda) = \iint_{\{x_2 \neq x_3\}} \tilde{\delta}_\lambda(x_2, x_3) \, d(F_n - F)(x_2) \, d(F_n - F)(x_3) \ ,$$

$$R_2(\lambda) = n^{-1} \int \tilde{\delta}_\lambda(x_2, x_2) \, d(F_n - F)(x_2) \ ,$$

$$R_3(\lambda) = n^{-1} \int \tilde{\delta}_\lambda(x_2, x_2) \, dF(x_2) \ ,$$

To finish the proof of (6.1) it is enough to show that

$$(6.3) \qquad \left[ \frac{R_3(\lambda) + b(\lambda) - \text{MISE}(\lambda)}{\text{MISE}(\lambda)} \right]^{2k} \leq C_k \, n^{-\gamma k} \ ,$$

and for "term" denoting $R_1$, $R_2$, or $S$:

$$(6.4) \qquad E\left[ \frac{\text{term}}{\text{MISE}(\lambda)} \right]^{2k} \leq C_k \, n^{-\gamma k}$$

Write

$$ASE = ISE + T(\lambda) \ .$$

As above, $T(\lambda)$ admits the decomposition

$$T = T_1 + T_2 + T_3 + 2T_4 + 2T_5 + T_6 + T_7 + 2U_1 + 2U_2 + 2U_3 + V$$

where

$$T_1(\lambda) = \iint_{\{x_1 \neq x_2 \neq x_3 \neq x_1\}} \delta_\lambda(x_1, x_2) \, \delta_\lambda(x_1, x_3) \, \tilde{w}(x_1) \, d(F_n - F)(x_1)$$

$$d(F_n - F)(x_2) \, d(F_n - F)(x_3)$$

$$T_2(\lambda) = n^{-1} \iint\limits_{\{x_1 \neq x_2\}} \delta_\lambda(x_1,x_2)^2 \, \tilde{w}(x_1) \, d(F_n-F)(x_1) \, d(F_n-F)(x_2) \; ,$$

$$T_3(\lambda) = n^{-1} \iint \delta_\lambda(x_1,x_2)^2 \, \tilde{w}(x_1) \, d(F_n-F)(x_1) \, dF(x_2) \; ,$$

$$T_4(\lambda) = n^{-1} \iint\limits_{\{x_1 \neq x_2\}} \delta_\lambda(x_1,x_2) \, \delta_\lambda(x_1,x_1) \, \tilde{w}(x_1) \, d(F_n-F)(x_1) \, d(F_n-F)(x_2) \; ,$$

$$T_5(\lambda) = n^{-1} \iint \delta_\lambda(x_1,x_2) \, \delta_\lambda(x_1,x_1) \, w(x_1) dx_1 \, d(F_n-F)(x_2) \; ,$$

$$T_6(\lambda) = n^{-2} \int \delta_\lambda(x,x)^2 \, \tilde{w}(x) \, d(F_n-F)(x) \; ,$$

$$T_7(\lambda) = n^{-2} \int \delta_\lambda(x,x)^2 \, w(x) dx \; ,$$

$$U_1(\lambda) = \iint\limits_{\{x_1 \neq x_2\}} \delta_\lambda(x_1,x_2) \, B(x_1) \, \tilde{w}(x_1) \, d(F_n-F)(x_1) \, d(F_n-F)(x_2) \; ,$$

$$U_2(\lambda) = n^{-1} \int \delta_\lambda(x_1,x_1) \, B(x_1) \, \tilde{w}(x_1) \, d(F_n-F)(x_1) \; ,$$

$$U_3(\lambda) = n^{-1} \int \delta_\lambda(x_1,x_1) \, B(x_1) \, w(x_1) dx \; ,$$

$$V(\lambda) = \int B(x_1)^2 \, \tilde{w}(x_1) \, d(F_n-F)(x_1) \; .$$

Thus, (6.2) will be established when (6.4) is verified for each of the above terms as well.

To check (6.3), note that by the familiar variance $-$ bias$^2$ decomposition (see, for example, Rosenblatt (1971)), using the notation (3.9),

$$\text{MISE} = R_3(\lambda) - r(\lambda) + b(\lambda) \; ,$$

where, using the notation (3.6),

$$r(\lambda) = n^{-1} \iint \tilde{\delta}_\lambda(x_2,x_3) \, dF(x_2) \, dF(x_3) \; ,$$

The inequality (6.3) follows from this and (3.3), (3.5) and (3.7).

The verification of (6.4) will now be done term by term, starting with those which do not involve $d(F_n-F)$:

Term $T_7$:  Using (3.10)

$$\left[\frac{n^{-2}\int \delta_\lambda(x,x)^2 w(x)dx}{\text{MISE}(\lambda)}\right]^{2k} \le C_k\left[\frac{n^{-2}\lambda^2}{n^{-1}\lambda}\right]^{2k} \le C_k' \, n^{-2k\epsilon} \ .$$

Term $U_3$:  As above, using the Schwartz Inequality,

$$\left[\frac{n^{-1}\int \delta_\lambda(x_1,x_1) \, B(x_1) \, w(x_1)dx_1}{\text{MISE}(\lambda)}\right]^{2k} \le$$

$$\left[\frac{n^{-1}\left[\int \delta_\lambda(x_1,x_1)^2 w(x_1) \, dx_1\right]^{\frac{1}{2}} b(\lambda)^{\frac{1}{2}}}{\text{MISE}(\lambda)}\right]^{2k} \le$$

$$C_k\left[\frac{n^{-1}\lambda\cdot b(\lambda)^{\frac{1}{2}}}{(n^{-1}\lambda)^{\frac{1}{2}} b(\lambda)^{\frac{1}{2}}}\right]^{2k} \le C_k(n^{-1}\lambda)^k \le C_k' n^{-k\epsilon}$$

The remaining terms all have at least one $d(F_n-F)$, and so have mean 0.  Thus to check (6.4), by the cumulant expansion of the $2k$-th moment, it is enough to check that, for $k=2,3,\ldots$, there is a constant $C_k$ so that

(6.5)  $$\left|\text{cum}_k\left(\frac{\text{term}}{\text{MISE}}\right)\right| \le C_k \, n^{-\gamma k} \ ,$$

where $\text{cum}_k(\cdot)$ denotes the $k$-th order cumulant, for which each argument is the same.

To verify (6.5) in the case of those terms having only one $d(F_n - F)$, note that they may be written

$$n^{-1} \sum_{i=1}^{n} W(X_i) .$$

Thus, using the independence property and linearity of cumulants, it is enough to show that

$$n^{-k+1} \text{ MISE}^{-k} \mid E[W(X_1)]^k \mid \leq C_k n^{-\gamma k} .$$

Term $R_2$: Note that here

$$W(X_2) = n^{-1} \left[ \int \delta_\lambda (x_1, x_2)^2 w(x_1) dx_1 - \iint \delta_\lambda (x_1, x_2)^2 w(x_1) dx_1 \, dF(x_2) \right] .$$

So by the Binomial Theorem and repeated application of (3.4),

$$n^{-k+1} \text{ MISE}^{-k} \mid E\left[W(X_2)\right]^k \mid \leq C_k n^{-2k+1} (n^{-1} \lambda)^{-k} \lambda^{2k - \frac{k+1}{2}} \leq$$

$$\leq C_k' n^{-k/4} .$$

Term $T_3$: Similar to $R_2$.

Term $T_5$: Similar to $R_2$.

Term $T_6$: Note that here

$$W(X_1) = n^{-2} \left[ \delta_\lambda (X_1, X_1)^2 \tilde{w}(X_1) - \int \delta_\lambda (x_1, x_1)^2 w(x_1) dx_1 \right] .$$

So by (3.10)

$$n^{-k+1} \text{ MISE}^{-k} \mid E\left[W(X_1)\right]^k \mid \leq C_k n^{-3k+1} (n^{-1} \lambda)^{-k} \lambda^{2k} \leq C_k' n^{-k/2} .$$

<u>Term V</u>: Note that here

$$W(X_1) = B(X_1)^2 \tilde{w}(X_1) - b(\lambda) \ .$$

Thus, by (3.8),

$$n^{-k+1} \text{MISE}^{-k} \mid E\left[W(X_1)\right]^k \mid \leq C_k \, n^{-k+1} \Big( b(\lambda)^k [b(\lambda)]^{-k} +$$

$$+ \sum_{j=1}^{k} \left[ b(\lambda) \lambda^{(j-1)(1-\epsilon)} \right] b(\lambda)^{k-j} \Big[ (n^{-1}\lambda)^{j-1} b(\lambda)^{k-j+1} \Big]^{-1} \Big) \leq$$

$$\leq C_k' n^{-\gamma k} \ .$$

<u>Term $U_2$</u>: Note that here

$$W(X_1) = n^{-1} \Big[ \delta_\lambda(X_1,X_1)B(X_1) \tilde{w}(X_1) - \int \delta_\lambda(x_1,x_1) B(x_1) w(x_1)dx_1 \Big] \ .$$

By (3.8), (3.10) and the Schwartz Inequality, for $j=1,2,\ldots$, there is $C_j$ so that

$$\mid \int [\delta_\lambda(x_1,x_1) B(x_1) \tilde{w}(x_1)]^j d F(x_1) \mid \leq$$

$$\leq [\int \delta_\lambda(x_1,x_1)^{2j} \tilde{w}(x_1)^{2j-1} dF(x_1)]^{\frac{1}{2}} [\int B(x_1)^{2j} w(x_1)dx_1]^{\frac{1}{2}} \leq$$

$$\leq C_j \lambda^{(3j-1)/2} b(\lambda)^{\frac{1}{2}} \ .$$

Hence,

$$n^{-k+1} \text{MISE}^{-k} \mid E[W(X_1)]^k \mid \leq C_k n^{-2k+1} [(n^{-1}\lambda)^{-k+\frac{1}{2}} b(\lambda)^{-\frac{1}{2}}] \lambda^{(3k-1)/2} b(\lambda)^{\frac{1}{2}} \leq$$

$$\leq C_k^1 n^{-k/4} \ .$$

<u>Term S</u>: Note that here

$$W(X_2) = \int \delta_\lambda(x_1,X_2) B(x_1) w(x_1)dx_1 - \iint \delta_\lambda(x_1,x_2) B(x_1) w(x_1)dx_1 \, dF(x_2) \ .$$

It follows from the Schwartz Inequality that,

$$\left| \int [ \int \delta_\lambda(x_1,x_2) B(x_1) w(x_1)dx_1]^j dF(x_2) \right| \le$$

(6.6)

$$\le \int [ \int \delta_\lambda(x_1,x_2)^2 w(x_1)dx_1]^{j/2} b(\lambda)^{j/2} dF(x_2) .$$

So, by (3.4), for j even, there is a constant $C_j$ such that (6.6) is bounded by

$$C_j b(\lambda)^{j/2} \lambda^{j-(j/2+1)/2} = C_j \lambda^{3j/4-1/2} b(\lambda)^{j/2} .$$

And by the moment inequality, for j odd, there is a constant $C_j$ such that (6.6) is bounded by

$$b(\lambda)^{j/2} [ \int ( \int \delta_\lambda(x_1,x_2)^2 w(x_1)dx_1)^{(j+1)/2} dF(x_2)]^{j/(j+1)} \le$$

$$\le C_j b(\lambda)^{j/2} \lambda^{3j/4-j/2(j+1)} .$$

Thus, for k=3,4,...

$$n^{-k+1} MISE^{-k} | EW(X_2)^k | \le C_k n^{-k+1} (n^{-1}\lambda)^{-k/2} b(\lambda)^{-k/2} b(\lambda)^{k/2} \lambda^{3k/4-3/8} =$$

$$= C_k n^{-k/2+1} \lambda^{k/4-3/8} = C_k (n^{-1}\lambda)^{k/2-1} \lambda^{-k/4+5/8} \le$$

$$\le C_k' n^{-\epsilon k/4} .$$

More precise computations are required in the case k=2. By (3.6)

$$E(W(X_2)^2 \le E[ \int \delta_\lambda(x_1,X_2) B(x_1) w(x_1)dx_1]^2 =$$

$$= \int [ \int ( \int \delta_\lambda(x_1,x_2) \delta_\lambda(x_1',x_2)dF(x_2)) B(x_1')w(x_1')dx_1'] B(x_1) w(x_1)dx_1 \le$$

$$\le ( \int [ \int ( \int \delta_\lambda(x_1,x_2) \delta_\lambda(x_1',x_2)dF(x_2) B(x_1') w(x_1')dx_1']^2 w(x_1)dx_1)^{\frac{1}{2}} b(\lambda)^{\frac{1}{2}} \le$$

$$\leq \left( \int [ \int ( \int \delta_\lambda (x_1,x_2) \, \delta_\lambda (x_1',x_2) \, dF(x_2) )^2 \, w(x_1') dx_1' )^{\frac{1}{2}} b(\lambda)^{\frac{1}{2}} ]^2 \, w(x_1) dx_1 )^{\frac{1}{2}} b(\lambda)^{\frac{1}{2}} = .$$

$$= b(\lambda) ( \iint \tilde{\delta}_\lambda (x_2,x_2')^2 \, dF(x_2) \, dF(x_2') )^{\frac{1}{2}} \leq b(\lambda) C(\lambda^{2-2/2})^{\frac{1}{2}} .$$

Thus,

$$n^{-1} \, \text{MISE}(\lambda)^{-2} \, EW(X_2)^2 \leq C \, n^{-1} (n^{-1}\lambda \, b(\lambda))^{-1} \, b(\lambda) \lambda^{\frac{1}{2}} \leq C' \, n^{-\epsilon/2} .$$

It remains to verify (6.5) for the terms containing two or three $d(F_n-F)$'s. The terms containing 2 may all be written in the form:

$$n^{-1} \sum_{\substack{i,i'=1 \\ i \neq i'}}^{n} W(X_i, X_{i'}) ,$$

where

$$E \, W(X_i, X_{i'}) = 0, \quad i \neq i' .$$

So, using the linearity property of cumulants, (6.5) will be established in this case when it is seen that there is a constant $\gamma > 0$, so that for $k=2,3,\dots$, there are constants $C_k$ such that

$$\left| n^{-2k} \text{MISE}^{-k} \sum_{i_1,i_1',\dots,i_k,i_k'} \text{cum}_k (W(X_{i_1},X_{i_1'}),\dots,W(X_{i_k},X_{i_k'})) \right| \leq C_k \, n^{-\gamma k} ,$$

where, by a moment expansion of $\text{cum}_k$, it may be assumed that each of $i_1,i_1',\dots,i_k,i_k'$ appears at least twice. In each case, it will be convenient to let m denote the number of $i_1,i_1',\dots,i_k,i_k'$ that are unique. Note that, for $m=2,3,\dots,k$, the number of $\text{cum}_k$ with m distinct indices is bounded by $C_k \, n^m$.

Term $T_2$:   Note that here

$$W(X_i, X_{i'}) = n^{-1}[\delta_\lambda(X_i, X_{i'})^2 \, \tilde{w}(X_i) - \int \delta_\lambda(X_i, x_2)^2 \, \tilde{w}(X_i) \, dF(x_2) - $$

$$- \int \delta_\lambda(x_1, X_{i'})^2 \, w(x_1) dx_1 + \iint \delta_\lambda(x_1, x_2)^2 \, w(x_1) dx_1 \, dF(x_2)] \; .$$

So, by (3.4)

$$\left| n^{-2k} \, \text{MISE}^{-k} \, \Sigma \, \text{cum}_k (W(X_{i_1}, X_{i_1'}), \ldots, W(X_{i_k}, X_{i_k'})) \right| \leq$$

$$\leq n^{-2k}(n^{-1}\lambda)^{-k} n^{-k} C_k \sum_{m=2}^{k} n^m \lambda^{2k-m/2} \leq C_k' \, n^{-k/2} \; .$$

Term $T_4$:   Similar to $T_2$.

Term $R_1$:   Here

$$W(X_i, X_{i'}) = \tilde{\delta}_\lambda(X_i, X_{i'}) - \int \tilde{\delta}_\lambda(X_i, x_2) \, dF(x_2) - \int \tilde{\delta}_\lambda(x_1, X_{i'}) \, dF(x_1) + $$

$$+ \iint \tilde{\delta}_\lambda(x_1, x_2) \, dF(x_1) \, dF(x_2) \; .$$

Thus,

$$\left| n^{-2k} \, \text{MISE}^{-k} \, \Sigma \, \text{cum}_k (W(X_{i_1}, X_{i_1'}), \ldots, W(X_{j_k}, X_{i_k'})) \right| \leq$$

$$\leq n^{-2k}(n^{-1}\lambda)^{-k} C_k \sum_{m=2}^{k} n^m \lambda^{k-m/2} \leq C_k' \, n^{-\epsilon k/2} \; .$$

Term $U_1$:   Here

$$W(X_i, X_{i'}) = \delta_\lambda(X_i, X_{i'}) \, B(X_i) \, \tilde{w}(X_i) - \int \delta_\lambda(X_i, x_2) \, B(X_i) \, \tilde{w}(X_i) \, dF(x_2) - $$

$$- \int \delta_\lambda(x_1, x_{i'}) \, B(X_1) \, w(x_1) dx_1 + \iint \delta_\lambda(x_1, x_2) \, B(x_1) \, w(x_1) dx_1 \, dF(x_2) \; .$$

This term is handled by means quite similar to those used on Term $T_2$ above, except that (3.4) is augmented by the Schwartz Inequality and (3.8). The result is, for k=2,3,...

$$\left| n^{-2k} \text{MISE}^{-k} \Sigma \text{cum}_k (W(X_{i_1}, X_{i_1'}), \ldots, W(X_{i_k}, X_{i_k'})) \right| \le$$

$$\le n^{-2k} ((n^{-1}\lambda)^{k-1/2} b(\lambda)^{1/2})^{-1} C_k \sum_{m=2}^{k} n^m \lambda^{(2k-m)/2} b(\lambda)^{1/2} \lambda^{(k-1)(1-\epsilon)/2}$$

$$\le C_k' n^{-\epsilon^2 k/4} .$$

It remains to verify (6.5) for:

Term $T_1$:  This term may be handled by methods similar to those used on term $T_2$.

This completes the proof of Theorems 1 and 2.

## 7. Proof of Lemma 1

Using the definition of $\delta_\lambda(x,y)$, write

$$| \int \cdots \int [\prod_{i,i'} \delta_\lambda(x_i,x_{i'})^{\alpha_{ii'}}][\prod_i \tilde{w}(x_i)^{\beta_i}] \, dF(x_1) \cdots dF(x_m) | =$$

(7.1)

$$= | \sum_{\ell_1=1}^{\lambda} \cdots \sum_{\ell_k=1}^{\lambda} \int \cdots \int \psi_{\ell_1} \psi_{\ell_1} \tilde{w} \cdots \psi_{\ell_k} \psi_{\ell_k} \tilde{w}[\pi w^{\beta_i}] \, dF(x_1) \cdots dF(x_m) |$$

The multiple integral on the right hand side may now be factored to give an expression of the form

(7.2) $\qquad | \sum_{\ell_1} \cdots \sum_{\ell_k} [\int \quad d F(x_1)] \cdots [\int \quad d F(x_m)] | .$

Consider the set of $\alpha_{ii'}$ which have $i \neq i'$ and are positive. Find a subset, A, which has the property that each of $1,\ldots,m$ appears at least once as an i or i', and suppose that this subset is minimal in the sense that if any $\alpha_{ii'}$ is removed, then $1,\ldots,m$ no longer all appear as the index of an $\alpha$ .

Group $1,\ldots,m$ into two subsets, I and I', by the following rules:

1) any of $1,\ldots,m$ that appear twice (or more) as an index of an $\alpha$ in
   A goes into I.

2) if i is in I, and $\alpha_{ii'}$ (or $\alpha_{i'i}$) is in A, put i' into I' .

3) For the remaining $\alpha_{ii'}$ in A, put i in I and i' in I' .

The above rules partition $\{1,\ldots m\}$ into I and I' .

Observe that for each $\alpha_{ii'}$ in A, there is an $\ell$ so that $\psi_\ell(x_i) \psi_\ell(x_{i'})$ appears in the integrand on the right side of (7.1). Suppose, without loss of generality, that $\ell_1,\ldots,\ell_L$ each correspond in this manner to a different element of A, where L denotes the cardinality of A. Also

assume, without loss of generality, that $\ell_1,\ldots\ell_b$ correspond to those $\alpha$ in A which have an index appearing more than once in A.

By the Schwartz Inequality, (7.2) may be written as

$$\left| \sum_{\ell_1} \cdots \sum_{\ell_k} \prod_{i=1}^{m} \int [\quad] dF(x_i) \right| =$$

$$= \left| \sum_{\ell_{L+1}} \cdots \sum_{\ell_k} \left[ \sum_{\ell_1} \cdots \sum_{\ell_L} \left( \prod_{i\in I} \int [\quad] dF(x_i) \right) \left( \prod_{i\in I'} \int [\quad] dF(x_i) \right) \right] \right| \leq$$

$$\leq \sum_{\ell_{L+1}} \cdots \sum_{\ell_k} \left[ \sum_{\ell_1} \cdots \sum_{\ell_L} \left( \prod_{i\in I} \int [\quad] dF(x_i) \right)^2 \right]^{\frac{1}{2}}$$

$$\left[ \sum_{\ell_1} \cdots \sum_{\ell_L} \left( \prod_{i\in I'} \int [\quad] dF(x_i) \right)^2 \right]^{\frac{1}{2}}.$$

Suppose, without loss of generality, that $I' = \{1,\ldots,L\}$. Then,

$$\sum_{\ell_1} \cdots \sum_{\ell_L} \left( \prod_{i\in I'} \int [\quad] dF(x_i) \right)^2 = \prod_{i=1}^{L} \sum_{\ell_i} \left( \int [\quad] dF(x_i) \right)^2 \leq \prod_{i=1}^{L} C_k \,,$$

where the last inequality follows from (5.2) and the Bessel inequality, because $\int [\quad] dF(x_i)$ is the $\ell_i$th Fourier coefficient of a function whose norm is bounded in (5.2). Similar techniques give

$$\sum_{\ell_1} \cdots \sum_{\ell_L} \left( \prod_{i\in I} \int [\quad] dF(x_i) \right)^2 \leq \lambda^{b-1} C_k^{L-b+1} \,.$$

It follows from the above that there is a constant $C_k$ so that (7.1) is bounded by

$$C_k \lambda^{k-L} \lambda^{(b-1)/2} \,.$$

To put this in more useful terms, note that

$$2L - b \geq m - 1$$

and so

$$-L + b/2 - \tfrac{1}{2} \leq -m/2$$

It follows that (7.1) is bounded by

$$c_k \lambda^{k-m/2} .$$

This completes the proof of Lemma 1.

## References

BEAN, S. and TSOKOS, C.P. (1982). Bandwidth selection procedures for kernel density estimates, Comm. Statist. A11, 1045-1069.

BURMAN, P. (1984). A data dependent approach to density estimation, unpublished manuscript.

BOWMAN, A.W. (1982). A comparative study of some kernel-based non-parametric density estimators, Research Report, Statistical Laboratory, University of Manchester.

BREIMAN, L., MEISEL, W., and PURCELL, E. (1977). Variable kernel estimates of multivariate densities, Technometrics, 19, 135-144.

CENCOV, N.N. (1962). Evaluation of an unknown distribution density from observations. Soviet Math. 3, 1559-1562.

COLLOMB, G. (1981). Estimation non paramétrique de la régression: Revue Bibliographique, Int. Statist. Rev. 49, 75-93.

ENGLE, R.F., GRANGER, C.W.J., RICE, J. and WEISS, A. (1983). Non-parametric estimates of the relation between weather and elasticity of demand. Discussion paper #83-17, Dept. of Economics, Univ. of California at San Diego.

FÖLDES, A. and RÉVÉSZ, P. (1974). A general method for density estimation, Studia Sci. Math. Hungar, 9, 81-92.

FRYER, M.J. (1977). Review of some non-parametric methods of density estimation. J. Inst. Math Appl. 20, 335-354.

HALL, P. (1981). On trigonometric series estimates of densities. Ann. Statist. 9, 683-685.

HALL, P. (1982). Limit theorems for stochastic measures of the accuracy of density estimators. Stoch. Processes and Their Appln. 13, 11-25.

HALL, P. (1983). Large sample optimality of least squares cross-validation in density estimation. Ann. Statist., 11, 1156-1174.

HÄRDLE, W. (1984). Approximations to the mean integrated squared error with applications to optimal bandwidth selection for nonparametric regression function estimators, to appear in Stoch. Processes and Their Appln.

HÄRDLE, W. and MARRON, J.S. (1983). The nonexistence of moments of some kernel regression estimators. North Carolina Institute of Statistics, Mimeo Series #1537.

HÄRDLE, W. and MARRON, J.S. (1984). Optimal bandwidth selection
    in nonparametric regression function estimation. North Carolina
    Institute of Statistics, Mimeo Series #1546.

JOHNSON, G.J. (1982). Properties of maximal deviations for non-
    parametric regression function estimates. J. Multivar. Anal.
    12, 402-414.

MARRON, J.S. (1984). An asymptotically efficient solution to the
    bandwidth problem of kernel density estimation. North Carolina
    Institute of Statistics, Mimeo Series #1545.

MARRON, J.S. (1985). Convergence properties of an empirical error
    criterion for multivariate density estimation, to appear in
    J. Multivar. Anal.

NADARAYA, E.A. (1964). On estimating regression, Theor. Prob. Appl.
    9, 141-142.

PARZEN, E. (1962). On estimation of a probability density function and
    mode, Ann. Statist. 33, 1056-1076.

RAATGEVER, J.W. and DUIN, R.P.W. (1978). On the variable kernel
    model for multivariate nonparametric density estimation. Compstat
    1978. eds. L.C.A. Corsten and J. Hermans.

RICE, J. (1982). Bandwidth choice for nonparametric kernel regression,
    unpublished manuscript.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of
    of a density function, Ann. Math. Statist. 27, 832-837.

ROSENBLATT, M. (1969). Conditional probability density and regression
    estimators, Multivariate Analysis II, ed. P.R. Krishnaiah, 25-31.

ROSENBLATT, M. (1971). Curve estimates, Ann. Math. Statist. 42,
    1815-1842.

RUST, A.E. and TSOKOS, C.P. (1981). On the convergence of kernel
    estimators of probability density functions. Ann. Inst. Statist.
    Math. 33, 233-246.

SCOTT, D.W. and FACTOR, L.E. (1981). Monte Carlo study of three
    data-based nonparametric probability density estimators, J. Amer.
    Statist. Assoc. 76, 9-15.

SILVERMAN, B.W. (1984). A fast and efficient cross-validation method
    for smoothing parameter choice in spline regression. J. Amer.
    Statist. Assoc. (to appear).

STEELE, J.M. (1978). Invalidity of average squared error criterion
    in density estimation. Canad. J. Statist. 6, 193-200.

STONE, C.J. (1976). Nearest neighbor estimators of a nonlinear regression function. Proc. Comp. Sci. Statist. 8th Annual Symposium on the Interface. Health Sciences Computing Facility, U.C.L.A. 413-418.

STONE, C.J. (1984a). An asymptotically efficient histogram selection rule. Proceedings of the Neyman-Kiefer Meeting, to appear.

STONE, C.J. (1984b). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist. (to appear).

SUSARLA, V. and WALTER, G. (1981). Estimation of a multivariate density function using delta sequences, Ann. Statist. 9, 347-355.

TANNER, M.A. and WONG, W.H. (1982). Data based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis, Tech. Report No. 137, Dept. of Statist., Univ. of Chicago.

TUKEY, J.W. (1961). Curves as parameters, and touch estimation, Proc. 4th Berkeley Symp. 681-694.

WAHBA, G. (1977). Optimal smoothing of density estimates, Classification and Clustering. ed. J. van Ryzin, 423-458.

WALTER, G. (1977). Properties of Hermite series estimation of probability density, Ann. Statist. 5, 1258-1264.

WALTER, G. and BLUM, J. (1979). Probability density estimation using delta sequences, Ann. Statist. 7, 328-340.

WATSON, G.S. (1964). Smooth regression analysis, Sankhyā, A26, 359-372.

WATSON, G.S. and LEADBETTER, M.R. (1964a). Hazard analysis I, Biometrika 51, 175-184.

WATSON, G.S. and LEADBETTER, M.R. (1964b). Hazard analysis II, Sankhyā A26, 101-116.

WEGMAN, E.J. (1972). Nonparametric probability density estimation: II. A comparison of density estimation methods, J. Statist. Comp. and Simul. 1, 225-245.