BEST POSSIBLE CONSTANT FOR BANDWIDTH SELECTION

Jianqing Fan James S. Marron [1]

Department of Statistics

University of North Carolina

Chapel Hill, N.C. 27599-3260

October 15, 1990

Abstract

For the data based choice of the bandwidth of a kernel density estimator, several methods have recently been proposed which have a very fast asymptotic rate of convergence to the optimal bandwidth. In the particular the relative rate of convergence is the square root of the sample size, which is known to be the possible. The point of this paper is to show how semiparametric arguments can be employed to calculate the best possible constant coefficient, i.e. an analog of the usual Fisher Information, in this convergence. This establishes an important benchmark as to how well bandwidth selection methods can ever hope to perform. It is seen that some methods attain the bound, others do not.

[1]Supported by NSF Grant DMS-8902973.

Abbreviated title. Efficient Bandwidth Selection.

AMS 1980 subject classification. Primary 62G20. Secondary 62G05.

Key words and phrases. Bandwidth selection, best constant, efficient bounds, kernel density estimator, semiparametric methods.

1 Introduction

Nonparametric curve estimation provides a useful tool for understanding the structure of a data set. See Silverman (1986), Eubank (1988), Müller (1988), Härdle (1990) and Wahba (1990) for many examples of this, and good introductions to the general subject area. The most important practical hurdle, in applications of this methodology, is choice of the smoothing parameter.

A large amount of recent progress has been made on data based smoothing parameter selection, see the survey paper by Marron (1988). Because it provides a simple context in which to study the problem (hence allowing deeper results), much of this progress has come in the case of kernel density estimation. Hence that setting is discussed here as well.

A useful asymptotic means of assessing performance of a data driven smoothing parameter, i.e. bandwidth, is through the relative rate of convergence to the bandwidth that minimizes the Mean Integrated Squared Error.

Hall *et al.* (1990), Jones, Marron and Park (1990) and Chiu (1991) have all proposed methods for which this rate of convergence is extremely fast. In particular, it goes down as $O(n^{-1/2})$, where n denotes sample size, which is unusually fast in nonparametric settings. This rate of convergence has been shown to be the best possible, in an important minimax sense, by Hall and Marron (1990). But the fact that there are competing selectors motivates deeper analysis.

A natural step in this direction is to consider not only the exponent in the rate of convergence, but also the constant coefficient. This type of question is frequently addressed in semiparametric analysis, which is an extension of the classical Fisher Information ideas. See Bickel *et al.* (1990), and van de Vaart (1988) for details. In this paper a straightforward application of these methods is used to calculate the best possible constant in our setting of bandwidth selection for kernel density estimation. It turns out that the problem of bandwidth selection is closely related to the problem of estimating some specific kind of quadratic functionals, which is studied by Hall and Marron (1987), Bickel and Ritov (1988) and Jones and Sheather (1990) in density estimation models, and by Fan (1990) and Donoho and Nussbaum (1990) in Gaussian white noise models. The knowledge gained there is also very useful to bandwidth selection.

Chiu (1991) proposes two $n^{-1/2}$ bandwidth selectors, and shows that for both, the relative error is asymptotically normal. It is a simple calculation to show that his asymptotic variance is the same as the best possible constant coefficient calculated here. This provides a sense in which our lower bound is very informative. With more work, the selector of Hall *et al.* (1990) can be shown to have the same limiting distribution. However the $n^{-1/2}$ method of Jones, Marron and Park (1990) has a larger constant, and thus is not optimal in this sense.

Section 2 gives a precise formulation, and discussion, of the main results. Proofs are in section 3.

2 Main Results

To describe the problem mathematically, assume that $X_1, \cdots, X_n$ are i.i.d. from an unknown density f. Let $K(\cdot)$ denote a kernel function, h_n be a bandwidth. A kernel density estimator is defined by

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_1^n K\left(\frac{x - X_j}{h_n}\right), \tag{2.1}$$

whose performance is typically measured by MISE

$$M(h_n) = E \int_{-\infty}^{\infty} \left(\hat{f}_n(x) - f(x)\right)^2 dx. \tag{2.2}$$

The optimal bandwidth $h_n(f)$ is the one that minimizes the MISE (2.2).

For convenience, denote a class of density having $(k+\alpha)$-derivatives:

$$\mathcal{F} = \{g : |g^{(k)}(x) - g^{(k)}(y)| \le M|x-y|^\alpha, |g^{(4)}(x)| \le g_0(x)\},$$

where $g_0(x)$ is bounded continuous and integrable. Let $||\cdot||_2$ denote the usual L_2-norm, and let

$$H_n(f, C) = \left\{g \in \mathcal{F} : ||\sqrt{g} - \sqrt{f}||_2 \le C/\sqrt{n}\right\} \tag{2.3}$$

be a Hellinger ball in the neighborhood of f.

The following Theorem shows that the relative error of *any bandwidth selection procedure* can not be smaller than $B(f)n^{-1/2}$, where

$$B^2(f) = \frac{4}{25}\left(\frac{\int_{-\infty}^{\infty}\left[f^{(4)}(x)\right]^2 f(x)dx}{\left[\int_{-\infty}^{\infty}(f''(x))^2dx\right]^2} - 1\right), \tag{2.4}$$

Theorem 1. *Let K be a continuous second order kernel with $\int_{-\infty}^{\infty}|x|^6|K(x)|dx < \infty$. Assume that $f \in \mathcal{F}$, and $k + \alpha > 4$. Then, for any bandwidth selection procedure $\hat{h}_n$,*

$$\lim_{C\to\infty}\liminf_{n\to\infty}\inf_{\hat{h}_n}\sup_{g\in H_n(f,C)} nE_g\left(\frac{\hat{h}_n - h_n(g)}{h_n(g)}\right)^2 \geq B^2(f). \tag{2.5}$$

As discussed in the introduction, the bound in (2.5) is the best attainable one, when $k + \alpha \geq 4.25$. Note that the bound (2.5) does not depend on the kernel function K, even though the optimal bandwidth $h_n(f)$ does.

The following theorem gives the result on the lower bound of the relative error of MISE.

Theorem 2. *Under the assumptions of Theorem 1, for any bandwidth selector $\hat{h}_n$,*

$$\lim_{C\to\infty}\liminf_{n\to\infty}\inf_{\hat{h}_n}\sup_{g\in H_n(f,C)} n^2E_g\left(\frac{M(\hat{h}_n) - M(h_n(g))}{M(h_n(g))}\right)^2 \geq 4B^4(f). \tag{2.6}$$

The last result indicates that for any bandwidth selector, the relative error of MISE can not be smaller than $2n^{-1}B^2(f)$. Thus, the quantity $B(f)$ plays an important role to the relative error of bandwidth selection, measured in either way: the larger $B(f)$ is, the harder the problem will be. In other words, $B(f)$ measures the difficulty of bandwidth selection problems.

Note that $B(f)$ is both location and scale invariance: for any $\sigma > 0$ and μ, $B(f_{\mu,\sigma}) = B(f)$, where

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right).$$

This is expected, because for example, estimating a density of $N(0,1)$ is as difficult as estimating a density of $N(2,4)$: plots of two estimates should look the same except the scales on x-axis and y-axis are marked differently. In this normal case,

$$B(f) = \frac{2}{5}\sqrt{\frac{4864}{3^{5.5}} - 1} = 1.300.$$

(*Put Figure 1 about here.*)

Table 1 shows the values of $B(f)$ for the 15 normal mixture densities in Figure 1. See Marron and Wand (1990) for the parameters and for the discussion of these densities.

Table 1. Constant Factors in the Lower Bounds

Density number	$B(f)$	Density number	$B(f)$	Density number	$B(f)$
1	1.300	2	1.771	3	4.973
4	2.638	5	1.388	6	1.868
7	1.286	8	3.390	9	4.742
10	2.125	11	19.394	12	9.635
13	25.587	14	9.408	15	3.515

Remark 1. A direct consequence of Theorem 1 is for any open neighborhood V of f (in L_2-topology), we have

$$\liminf_{n\to\infty} \inf_{\hat{h}_n} \sup_{g \in V \cap \mathcal{F}} nE_g \left(\frac{\hat{h}_n - h_n(g)}{h_n(g)}\right)^2 \geq B^2(f).$$

A similar formula holds for MISE.

Remark 2. Note that $B^2(f)$ plays a role analogous to the classical Fisher information. Thus, given any bandwidth selector (past or future) $\hat{h}_n$, its efficiency can be defined by

$$\left(\frac{B^2(f)}{nE_f\left([\hat{h}_n - h_n(f)]/h_n(f)\right)^2} \right)^{1/2}.$$

Remark 3. On the Hellinger ball $H_n(f, C)$, we have

$$\lim_{n\to\infty} \sup_{g\in H_n(f,C)} \left| \frac{h_n(g)}{h_n(f)} - 1 \right| = 0. \tag{2.7}$$

Moreover,

$$\lim_{n\to\infty} \sup_{g\in H_n(f,C)} |g^{(4)}(x) - f^{(4)}(x)| = 0, \forall x \tag{2.8}$$

and

$$\lim_{n\to\infty} \sup_{g\in H_n(f,C)} \left| \int_{-\infty}^{\infty} [g''(x)]^2 dx - \int_{-\infty}^{\infty} [f''(x)]^2 dx \right| = 0. \tag{2.9}$$

In other words, the Hellinger neighborhood is so small that the important characteristics of g are very close to those of f. These conclusions are proved in Lemma 5 of section 3, by using statistical ideas in the proof of the mathematical results, which are not easy to prove by conventional methods.

3 Proofs

3.1 Lemmas

The idea of the proof of Theorem 1 is to relate the problem of estimating $h_n(f)$ with that of estimating $\theta_2^{-1/5}(f)$ defined by (3.1), via a series of lemmas. To this end, denote

$$\theta_j(f) = \int_{-\infty}^{\infty} \left[f^{(j)}(x)\right]^2 dx, \quad j = 2, 3. \tag{3.1}$$

It will be seen that the optimal bandwidth $h_n(f)$ is approximated by

$$\phi_n(f) = c_1\theta_2^{-1/5}n^{-1/5} + c_2\theta_3\theta_2^{-8/5}n^{-3/5}, \tag{3.2}$$

where

$$c_1 = \left(\int_{-\infty}^{\infty} K^2(x)dx \left(\int_{-\infty}^{\infty} z^2 K(z)dz \right)^{-2} \right)^{1/5},$$

and

$$c_2 = \frac{1}{20} \int_{-\infty}^{\infty} x^4 K(x)dx \left(\int_{-\infty}^{\infty} K^2(x)dx \right)^{3/5} \left(\int_{-\infty}^{\infty} z^2 K(z)dz \right)^{-11/5}.$$

In the following discussions, we will suppress the dependence of θ_j, whenever its argument is g, a density in the Hellinger neighborhood of f. Recall that f is fixed throughout our arguments.

Lemma 1. *The optimal bandwidth $h_n(f)$ satisfies*

$$\sup_{g \in H_n(f,C)} \frac{h_n(g) - \phi_n(g)}{\phi_n(g)} = o\left(n^{-1/2}\right) \tag{3.3}$$

Proof. The proof follows the same argument as in the section 2 of Hall *et al.* (1990).

Thus, it is intuitively clear that the problem of estimating $h_n(f)$ is equivalent to that of estimating $\phi_n(f)$. The following lemma gives a lower bound for estimating $\theta_2^{-1/5}(f)$.

Lemma 2. *Let $R_{n,C,1}(f)$ be the minimax risk for estimating $\theta_2^{-1/5}(f)$:*

$$R_{n,C,1}(f) = \inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} E_g \left(\hat{h}_n - \theta_2^{-1/5}(g) \right)^2.$$

Then,

$$\lim_{C \to \infty} \liminf_{n \to \infty} n R_{n,C,1}(f) \geq \theta_2^{-2/5}(f) B^2(f), \tag{3.4}$$

where $B(f)$ was defined by (2.4).

Proof. It is shown in the proof of Theorem 2 (i) of Bickel and Ritov (1988) that $\theta_2(f)$ is pathwise differentiable along paths

$$\{f_\nu : \|\sqrt{f_\nu} - \sqrt{f}\|_2 \longrightarrow 0, \text{ and } \|(f_\nu^{(4)} - f^{(4)})\sqrt{f}\|_2 \longrightarrow 0\}$$

with the derivative function

$$4\left[f^{(4)}(x) - \theta_2(f)\right]\sqrt{f}.$$

Thus, $\theta_2^{-1/5}(f)$ is also pathwise differentiable along such paths with derivative function

$$\left(-\frac{1}{5}\theta_2^{-6/5}\right) 4 \left[f^{(4)}(x) - \theta_2(f)\right] \sqrt{f}.$$

As at the end of the proof of Theorem 2(i) of Bickel and Ritov (1988), the information bound for $\theta_2^{-1/5}(f)$ is

$$\begin{aligned} & \| - \frac{2}{5}\theta_2^{-6/5} \left[f^{(4)}(x) - \theta_2(f)\right] \sqrt{f} \|_2^2 \\ = & \frac{4}{25}\theta_2^{-12/5} \int_{-\infty}^{\infty} \left(f^{(4)}(x) - \theta_2\right)^2 f(x)dx \\ = & \frac{4}{25}\theta_2^{-12/5} \left[\int_{-\infty}^{\infty} \left[f^{(4)}(x)\right]^2 f(x)dx - \theta_2^2\right], \end{aligned}$$

by using the fact that $\theta_2 = \int_{-\infty}^{\infty} f^{(4)}(x)f(x)dx$. The result follows by the standard semi-parametric theory (e.g. Theorem 2.10, van der Vaart (1988)).

In order to show that the second term of $\phi_n(f)$ is not important to Theorem 1, the following lemma gives an estimate of $\theta(f) = \theta_3(f)\theta_2^{-8/5}(f)$.

Lemma 3. *There exists an estimator* $\hat{\delta}_n$ *such that*

$$\sup_{g \in H_n(f,C)} E_g(\hat{\delta}_n - \theta(g))^2 = O\left(n^{-32/85}\right), \tag{3.5}$$

Proof. Note that for $g \in \mathcal{F}$, $g^{(4)}(x)$ is bounded by $g_0(x) \in L_1 \cap L_\infty$. By the construction of Bickel and Ritov (1988) (see Hall and Marron (1990) for a simpler estimator which can also be used), there exist estimators $\hat{\theta}_2 \geq 0$ and $\hat{\theta}_3$ such that

$$\sup_{g \in H_n(f,C)} E\left(\hat{\theta}_3 - \theta_3\right)^4 = O\left(n^{-4 \times \frac{4}{17}}\right), \tag{3.6}$$

and

$$\sup_{g \in H_n(f,C)} E\left(\hat{\theta}_2 - \theta_2\right)^4 = O\left(n^{-4 \times \frac{8}{17}}\right). \tag{3.7}$$

To avoid zero denominator problem, choose

$$\hat{\delta}_n = \frac{\hat{\theta}_3}{\hat{\theta}_2^{8/5} + n^{-4/17}}.$$

Then,

$$E\left(\hat{\delta}_n - \theta\right)^2 = E\frac{\left(\theta_2^{8/5}\hat{\theta}_3 - \hat{\theta}_2^{8/5}\theta_3 - n^{-4/17}\theta_3\right)^2}{\left(\hat{\theta}_2^{8/5} + n^{-4/17}\right)^2 \theta_2^{16/5}} = I_1 + I_2, \tag{3.8}$$

where

$$I_1 = E\frac{\left(\theta_2^{8/5}\hat{\theta}_3 - \hat{\theta}_2^{8/5}\theta_3 - n^{-4/17}\theta_3\right)^2}{\left(\hat{\theta}_2^{8/5} + n^{-4/17}\right)^2 \theta_2^{16/5}} 1_{\{|\hat{\theta}_2 - \theta_2| > \frac{\theta_2}{2}\}}$$

and I_2 is defined similarly.

Since I_2 is integrated over the range $|\hat{\theta}_2 - \theta_2| \le \frac{\theta_2}{2}$, we have $\hat{\theta}_2 \ge \frac{\theta_2}{2}$, and

$$I_2 = O\left(E(\hat{\theta}_3 - \theta_3)^2 + E(\hat{\theta}_2^{\,8/5} - \theta_2^{8/5})^2 + n^{-8/17}\right) = O\left(n^{-8/17}\right).$$

Now, let's consider I_1. By the fact that $\hat{\theta}_2 \ge 0$, we have

$$I_1 = O\left(n^{8/17} E\left(\theta_2^{8/5}\hat{\theta}_3 - \hat{\theta}_2^{8/5}\theta_3 - n^{-4/17}\theta_3\right)^2 1_{\{|\hat{\theta}_2 - \theta_2| > \theta_2/2\}}\right).$$

By Hölder's inequality with $p = 5/4$ and $q = 5$,

$$\begin{aligned} I_1 &= O\left(n^{8/17}\left[E\left(\theta_2^{8/5}\hat{\theta}_3 - \hat{\theta}_2^{8/5}\theta_3 - n^{-4/17}\theta_3\right)^{10/4}\right]^{4/5} \right. \\ &\qquad \left. \times \left[E1_{|\hat{\theta}_2 - \theta_2| > \theta_2/2}\right]^{1/5}\right) \\ &= O\left(n^{8/17} n^{-8/17} n^{-32/85}\right) \\ &= O\left(n^{-32/85}\right) \end{aligned}$$

where the inequality

$$P\left(|\hat{\theta}_2 - \theta_2| > \frac{\theta_2}{2}\right) \le \left(\frac{2}{\theta_2}\right)^4 E|\hat{\theta}_2 - \theta_2|^4 = O(n^{-32/17})$$

was used. This completes the proof.

The following lemma shows that the minimax risk lower bound for $\phi_n(f)$ is equivalent to that of $n^{-1/5} c_1 \theta_2^{-1/5}$, i.e. the second term of $\phi_n(f)$ is indeed negligible.

Lemma 4. *Let $R_{n,C,2}(f)$ be the minimax risk for estimating $\phi_n(f)$:*

$$R_{n,C,2}(f) = \inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} E_g \left(\hat{h}_n - \phi_n(g)\right)^2 .$$

Then,

$$R_{n,C,2}(f) \geq n^{-2/5} c_1^2 R_{n,C,1}(f)(1 + o(1)),$$

where $\xi_{n,C} = o(1)$ means that $\lim_{C \to \infty} \lim_{n \to \infty} \xi_{n,C} = 0$.

Proof. Recall that $\theta = \theta_3 \theta_2^{-8/5}$ and that $\phi_n(g)$ is defined by (3.2). Let $\hat{\delta}_n$ be the estimator defined by Lemma 3 and $c_3 = c_2/c_1$. Then by making the change of variable $\hat{h}_n \to n^{-1/5} c_1(\hat{h}_n + n^{-2/5} c_3 \hat{\delta})$,

$$\begin{aligned} R_{n,C,2}(f) &= n^{-2/5} c_1^2 \inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} E\left[\hat{h}_n - \theta_2^{-1/5} + n^{-2/5} c_3(\hat{\delta} - \theta)\right]^2 \\ &\geq n^{-2/5} c_1^2 \inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} \left[E(\hat{h}_n - \theta_2^{-1/5})^2 - a_n \sqrt{E(\hat{h}_n - \theta_2^{-1/5})^2}\,\right], \end{aligned}$$

where

$$a_n = 2c_3 n^{-2/5} \sqrt{E(\hat{\delta}_n - \theta)^2} = O\left(n^{-2/5-16/85}\right) = o\left(n^{-1/2}\right).$$

Thus,

$$R_{n,C,2}(f) \geq n^{-2/5} c_1^2 \inf_{\hat{h}_n} \left[q^2(\hat{h}_n) - a_n q(\hat{h}_n)\right], \tag{3.9}$$

where

$$q(\hat{h}_n) = \sqrt{\sup_{g \in H_n(f,C)} E(\hat{h}_n - \theta_2^{-1/5})^2}.$$

By Lemma 2, for any estimator $\hat{h}_n$, $q(\hat{h}_n) \geq R_{n,C,1}^{1/2} \geq d n^{-1/2}$ for some constant $d > 0$, when n and C are large. Since $a_n = o(n^{-1/2})$, the quadratic $x^2 - a_n x$ is increasing for $x > a_n/2$, and $R_{n,C,1}^{1/2} = \inf_{\hat{h}_n} q(\hat{h}_n)$, we arrive at

$$\inf_{\hat{h}_n} \left[g^2(\hat{h}_n) - a_n g(\hat{h}_n)\right] = R_{n,C,1} - a_n \sqrt{R_{n,C,1}} = R_{n,C,1}(1 + o(1)). \tag{3.10}$$

The conclusion follows from (3.9) and (3.10).

Lemma 5. *On the set of Hellinger ball $H_n(f,C)$, we have*

$$\lim_{n \to \infty} \sup_{g \in H_n(f,C)} \left| \frac{h_n(g)}{h_n(f)} - 1 \right| = 0.$$

Proof. By Lemma 1, we need only to show that (2.9) holds. By a useful statistical lower bound [e.g., page 18 of Fan (1989)], for any estimator $\hat{T}_n$, we have

$$\sup_{g\in H_n(f,C)} E|\hat{T}_n - g^{(j)}(x)|^2 \geq \frac{1-\sqrt{1-e^{-2C}}}{2} \sup_{g\in H_n(f,C)} |g^{(j)}(x) - f^{(j)}(x)|^2. \quad (3.11)$$

Since g has more than 4 derivatives, there exist estimators [e.g. kernel density estimators (2.1)] such that $g^{(j)}(x)$ $(j = 0, \cdots, 4)$ can be estimated consistently, i.e. such that the right hand side of (3.11) converges to 0. Thus,

$$\sup_{g\in H_n(f,C)} |g^{(j)}(x) - f^{(j)}(x)| \to 0, \quad \text{for } j = 0, \cdots, 4.$$

Now, by the dominate convergence theorem,

$$\begin{aligned} & \sup_{g\in H_n(f,C)} \left| \int_{-\infty}^{\infty} [g''(x)]^2 dx - \int_{-\infty}^{\infty} [f''(x)]^2 dx \right| \\ = & \sup_{g\in H_n(f,C)} \left| \int_{-\infty}^{\infty} g^{(4)} g - \int_{-\infty}^{\infty} f^{(4)} f \right| \\ \leq & \int_{-\infty}^{\infty} f \sup_{g\in H_n(f,C)} |g^{(4)} - f^{(4)}| + \int_{-\infty}^{\infty} g_0 \sup_{g\in H_n(f,C)} |g - f| \\ \longrightarrow & \; 0, \end{aligned}$$

where $|g^{(4)}| \leq g_0$ (see the definition of $\mathcal{F}$) was used in the inequality above. This completes the proof.

3.2 Proof of Theorem 1

Write $h_n(g) = \phi_n(g) + \xi_n(g)$, where by Lemma 1,

$$\sup_{g\in H_n(f,C)} \xi_n(g) = o(n^{-\frac{1}{2}-\frac{1}{5}}).$$

Now by using Lemma 5,

$$\begin{aligned} & \inf_{\hat{h}_n} \sup_{g\in H_n(f,C)} E_g \left(\frac{\hat{h}_n - h_n(g)}{h_n(g)} \right)^2 \\ \geq & \inf_{\hat{h}_n} \left[\sup_{g\in H_n(f,C)} E_g \left(\hat{h}_n - h_n(g) \right)^2 / \sup_{g\in H_n(f,C)} h_n^2(g) \right] \\ = & \inf_{\hat{h}_n} \sup_{g\in H_n(f,C)} E_g \left(\hat{h}_n - \phi_n(g) - \xi_n(g) \right)^2 n^{2/5} c_1^{-2} \theta_2^{2/5}(f)(1 + o(1)). \end{aligned}$$

By using the same argument as in the proof of Lemma 4, we can show that $\xi_n(g)$ is indeed negligible and conclude that

$$\inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} E_g \left(\frac{\hat{h}_n - h_n(g)}{h_n(g)} \right)^2 \geq n^{2/5} c_1^{-2} \theta_2^{2/5}(f) R_{n,c,2}(f)(1 + o(1)).$$

The conclusion follows directly from Lemmas 4 & 2.

3.3 Proof of Theorem 2

Denote

$$r = \int_{-\infty}^{\infty} K^2(x)dx, \quad \text{and} \quad \mu = \int_{-\infty}^{\infty} x^2 K(x)dx.$$

By using the fact that $M'(h_n(g)) = 0$, we have

$$M(\hat{h}_n) - M(h_n(g)) = \frac{1}{2} M''(\tilde{h})(\hat{h}_n - h_n(g))^2, \tag{3.12}$$

where $\tilde{h}$ lies between $\hat{h}_n$ and $h_n(g)$. Note that [see Hall *et al.* (1990)]

$$\begin{aligned} M''(h) &= 2rn^{-1}h^{-3} + 3h^2\mu^2\theta_2 + O(n^{-1} + h^4) \\ &\geq 5r^{2/5}\mu^{6/5}\theta_2^{3/5}n^{-2/5}(1 + o(1)) \end{aligned} \tag{3.13}$$

and

$$M(h_n(g)) = \frac{5}{4} r^{4/5} \mu^{2/5} \theta_2^{1/5} n^{-4/5}(1 + o(1)) \tag{3.14}$$

By Lemma 1, (3.13) and (3.14), we have

$$\frac{M''(\tilde{h})h_n^2(g)}{2M(h_n(g))} \geq 2 + o(1).$$

By the last display and (3.12), we arrive at

$$\begin{aligned} &\inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} n^2 E_g \left(\frac{M(\hat{h}_n) - M(h_n(g))}{M(h_n(g))} \right)^2 \\ = &\inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} n^2 E_g \left[\frac{M''(\tilde{h})h_n^2(g)}{2M(h_n(g))} \right]^2 \left(\frac{\hat{h}_n - h_n(g)}{h_n(g)} \right)^4 \\ \geq &\inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} n^2 E_g \left(\frac{\hat{h}_n - h_n(g)}{h_n(g)} \right)^4 (4 + o(1)) \\ \geq &\inf_{\hat{h}_n} \sup_{g \in H_n(f,C)} n^2 \left[E_g \left(\frac{\hat{h}_n - h_n(g)}{h_n(g)} \right)^2 \right]^2 (4 + o(1)). \end{aligned}$$

The conclusion follows directly from Theorem 1.

References

[1] Bickel, P.J., Klassen, C.A.J., Ritov, Y. and Wellner, J.A. (1990). *Efficient and Adaptive Inference in Semi-parametric Models.* Forthcoming monograph, Johns Hopkins University Press, Baltimore.

[2] Bickel, P.J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā* Ser. A, **50**, 381-393.

[3] Chiu, S.T. (1991). Bandwidth selection for kernel density estimation, *Ann. Statist.*, to appear.

[4] Donoho, D.L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *Tech. Report. 236*, Dept. of Statist., Univ. of California, Berkeley.

[5] Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* Dekker, New York.

[6] Fan, J. (1989). Contributions to the estimation of nonregular functionals. *Dissertation*, University of California, Berkeley.

[7] Fan, J. (1990). One the estimation of quadratic functionals. *Ann. Statist.*, to appear.

[8] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Boston.

[9] Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statistics and Probability Letters*, **6**, 109-115.

[10] Hall, P. and Marron, J.S. (1990). Lower bounds for bandwidth selection in density estimation. *Probability Theory and Related Fields*, to appear.

[11] Hall, P., Sheather, S. J., Jones, M.C. and Marron, J.S. (1990). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, to appear.

[12] Jones, M. C., Marron, J. S. and Park, B. U. (1990). A simple root n bandwidth selector. *Annals of Statistics*, to appear.

[13] Jones, M. C. and Sheather, S. J. (1990). Using nonstochastic terms to advantage in kernel based estimation of integrated squared density derivatives. *Unpublished manuscript.*

[14] Marron, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics*, **13**, 187-208.

[15] Marron, J.S. and Wand, M.P. (1990). Exact mean integrated squared error. *Manuscript.*

[16] Müller, H.-G. (1988). *Nonparametric Analysis of Longitudinal Data.* Springer Verlag, Berlin.

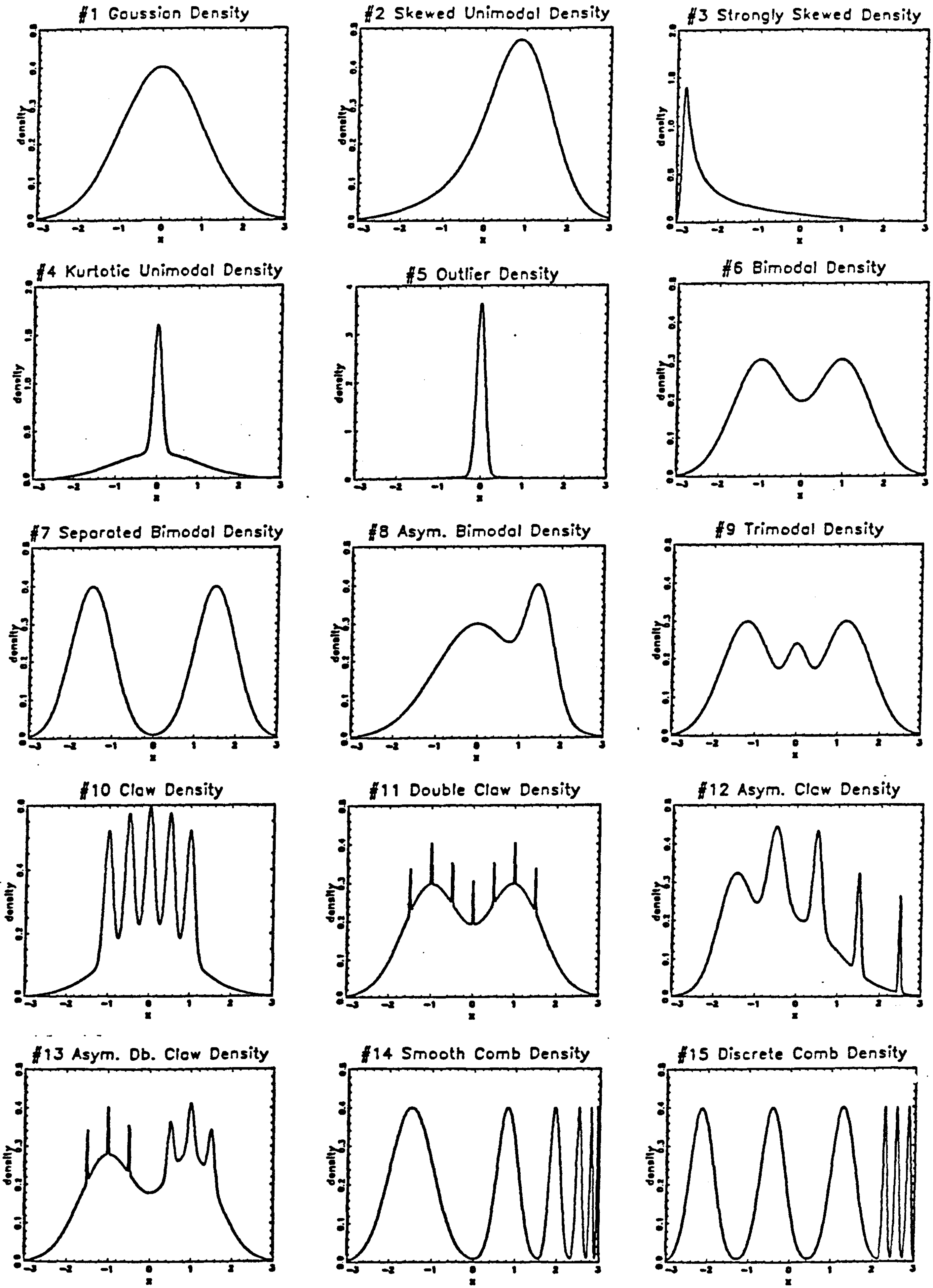
[17] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

[18] van der Vaart, A.W.(1988). *Statistical Estimation in Large Parameter Spaces.* CWI tract 44, Centre for Mathematics and Computer Science, Amsterdam.

[19] Wahba, G. (1990) *Spline Models for Observational Data.* SIAM, Philadelphia.

CAPTIONS

Figure 1. Normal mixture densities.


#1 Gaussian Density
#2 Skewed Unimodal Density
#3 Strongly Skewed Density
#4 Kurtotic Unimodal Density
#5 Outlier Density
#6 Bimodal Density
#7 Separated Bimodal Density
#8 Asym. Bimodal Density
#9 Trimodal Density
#10 Claw Density
#11 Double Claw Density
#12 Asym. Claw Density
#13 Asym. Db. Claw Density
#14 Smooth Comb Density
#15 Discrete Comb Density
density
x