EFFECT OF DIMENSION IN MULTIVARIATE DECONVOLUTION PROBLEMS

by

Xianglian Tang
Purdue University

Technical Report #94-8

Department of Statistics
Purdue University

April 1994

*

EFFECT OF DIMENSION IN MULTIVARIATE DECONVOLUTION PROBLEMS [1]

XIANGLIAN TANG

Purdue University

Suppose we observe n i.i.d. random vectors $Y_1, \cdots, Y_n$ from $Y = X + \varepsilon$, where ε is a noise component with a known distribution, and X has an unknown density f which belongs to a smooth function class. This paper focuses on the relationship between the dimension of Y to that of the minimax rate in which $f(\mathbf{x})$ can be estimated using $Y_1, \ldots, Y_n$ under squared error loss. The results obtained reveal that the effect of dimension on the minimax rate depends crucially on the smoothness of the error distribution of ε. In particular, the minimax rate is independent of the dimension in extremely smooth cases.

1 Introduction

We begin our discussion with the model $Y = X + \varepsilon$ where X and ε are independent p-dimensional random vectors with densities f and f_ε respectively. Here we assume that f_ε is known. In practice, X usually represents a true value and ε is a measurement error. The observation Y can then be thought of as the convolution of X and ε. Suppose we observe a random sample $Y_1, Y_2, ..., Y_n$ according to the above model, and we want to estimate the unknown density $f(\mathbf{x})$ at $X = \mathbf{x}$.

In the univariate case ($p = 1$), this problem has been widely studied. Many different estimators have been proposed and their consistency as well as convergence rates under different loss functions have been obtained. Some recent works include Stefanski (1990), Devroye (1989), Zhang (1990) among others. It is well known that the convergence rate depends heavily on the smoothness of the distribution of ε [see for example, Carroll and Hall (1988), Stefanski and Carroll (1990), Zhang (1990) and Fan (1991a, b, 1993)]. In particular, Fan exhibited a class of mixing density kernel estimators that achieves the optimal convergence rates for a number of commonly encountered error distributions.

[1] *AMS 1991 subject classifications.* Primary 62H12; secondary 62G05, 62G99.

Key words and phrases. Deconvolution, multivariate kernel estimation, optimal convergence rate, characteristic function.

In this paper, we consider the analogous problem of multivariate deconvolution under squared error loss. Section 2 gives a brief review of some multivariate distribution theory as well as a discussion of the kind of multivariate error distributions that we are concerned with here. In particular, we define three classes of multivariate probability density functions, namely classes A, B and C (see Definitions 1 to 3), based on the rate of decay of their characteristic functions. This is illustrated by a number of examples. Section 3 proposes a class of multivariate mixing density kernel estimators for $f(\mathbf{x})$. This class is a generalization of the 1-dimensional mixing density kernel estimators of Fan (1991a). Upper bounds for the convergence rates of these estimators are established for various error distributions and f belonging to a number of smooth function classes. In particular, Theorems 1, 2 and 3 give upper bounds for the convergence rates when ε belongs to classes A, B and C respectively. Corresponding lower bounds for the optimal convergence rates are also obtained in Section 4 (see Theorems 6, 7 and 8). The results of Sections 3 and 4 indicate that this class of multivariate mixing density kernel estimators achieves the optimal convergence rates for error distributions which belong to classes A and C. In particular, it is interesting to note that if ε belongs to class A (i.e. extremely smooth), the optimal convergence rate does not depend on the dimension p.

2 Definitions and examples

Throughout this paper, we shall use capital letters X, Y, etc., to denote p-dimensional random vectors, bold faced letters $\mathbf{x}$, $\mathbf{y}$, $\mathbf{t}$, $\mathbf{a}$, etc., to denote constant vectors in R^p. $||\cdot||$ refers to the usual Euclidean norm on R^p.

We first define three classes of density functions by the rate of decay of their characteristic functions. Although these three classes do not exhaust all possible multivariate densities, most commonly occurring densities are included in these classes.

Let $\phi_\varepsilon(\mathbf{t}) = E\exp(i\mathbf{t}'\varepsilon)$ denote the characteristic function of ε, and $\phi_\varepsilon^{\mathbf{a}}(s) = E\exp(is\mathbf{a}'\varepsilon)$, where s is a scaling constant and $\mathbf{a}$ is a fixed unit vector (i.e. $||\mathbf{a}|| = 1$).

Definition 1 *A random vector ε is said to be in Class A of order β if its characteristic function ϕ_ε satisfies*

$$d_0|s|^{k_0}\exp(-\frac{|s|^\beta}{\gamma_0}) \leq |\phi_\varepsilon^{\mathbf{a}}(s)|$$

as $|s| \to \infty$, for all unit vectors $\mathbf{a}$*, and if there exists a unit hypercube (e.g. $[1,2]^p$) which does not contain the origin such that for all lines through the origin (having direction* $\mathbf{a}$*) which goes through this hypercube, we have*

$$|\phi_\varepsilon^{\mathbf{a}}(s)| \le d_1 |s|^{k_1} \exp(-\frac{|s|^\beta}{\gamma_1})$$

as $|s| \to \infty$. Here $\gamma_0 < \gamma_1$ are positive numbers and k_0, k_1, d_0, d_1 are constants.

REMARK. The above definition implies that, in most directions, the projection of ε has an exponentially decayed characteristic function of order β. The unit hypercube can be replaced by any volume with a positive Lebesgue measure. However for convenience we shall use the unit hypercube and denote $\mathbf{A}(\varepsilon)$ to be the set of unit vectors $\mathbf{a}$ which passes through this hypercube. Note that $\mathbf{A}(\varepsilon)$ is a connected area on the surface of unit sphere which may depend on ε.

Definition 2 *A random vector ε is said to be in Class B of order β (> 1) if there exist constants $\gamma_0 < \gamma_1$, d_0, d_1, k_0, k_1 such that*

$$d_0 |s|^{k_0} \exp(-(\log|s|)^\beta/\gamma_0) \le |\phi_\varepsilon^{\mathbf{a}}(s)|$$

for all unit vectors $\mathbf{a}$ *and*

$$|\phi_\varepsilon^{\mathbf{a}}(s)| \le d_1 |s|^{k_1} \exp(-(\log|s|)^\beta/\gamma_1), \quad \forall \mathbf{a} \in \mathbf{A}(\varepsilon),$$

as $|s| \to \infty$.

Definition 3 *A random vector ε is said to be in Class C of order β if there exist constants $d_0 \le d_1$ such that*

$$d_0 |s|^{-\beta} \le |\phi_\varepsilon^{\mathbf{a}}(s)|$$

for all unit vectors $\mathbf{a}$ *and*

$$|\phi_\varepsilon^{\mathbf{a}}(s)| \le d_1 |s|^{-\beta}, \quad \forall \mathbf{a} \in \mathbf{A}(\varepsilon),$$

as $|s| \to \infty$.

REMARK. Classes A and C are essentially generalizations of Fan's (1991a) supersmooth and ordinary smooth classes. However Class A is larger than the supersmooth class in the 1-dimensional case. For example, $\phi(t) = \exp(-t^2/2 - |t|)$ belongs to Class A but not to Fan's supersmooth class. Also we note that Class C can be viewed as a continuous extension of Class B for $\beta = 1$. This is significant because as we will see later, this is where the optimal convergence rate stops depending on p.

Class B appears to be somewhat small but is nevertheless non-empty. The lemma below proves this assertion.

Lemma 1 *Let $\beta > 1$ and $\gamma > 0$. Then there exists a characteristic function with decay rate*

$$\exp(-(\log|t|)^\beta/\gamma), \qquad \text{as } |t| \to \infty.$$

PROOF. Without loss of generality, let $\gamma = 1$. Let $g(t) = c\exp(-(\log t)^\beta)$. We observe that

$$g^{(1)}(t) = -c\beta\frac{(\log t)^{\beta-1}}{t}\exp(-(\log t)^\beta)$$

and

$$g^{(2)}(t) = c\beta(\log t)^{\beta-2}\exp(-(\log t)^\beta)\frac{\beta(\log t)^\beta + \log t - (\beta-1)}{t^2}.$$

Since $g^{(2)}(t) > 0$ for t not too small, and $g^{(1)}(t) < 0$, we can construct a function $f(t)$ such that

(a) $f(t) = g(t)$ for t not too small,

(b) $f(0) = 1$ and that $f(t)$ is convex on $(0, \infty)$.

Now extend $f(t)$ to the whole real line by symmetrization. By Pólya's sufficient condition (see Lukacs (1970)), this is a characteristic function with the desired decay rate. □

We shall now briefly look at a number of examples.

EXAMPLE 1. The characteristic function $\phi(t) = \exp(-t^2/2 - |t|)$ is in Class A with order 2. This corresponds to the convolution of a normal and a Cauchy distribution.

EXAMPLE 2. Let $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_p)'$, where $\varepsilon_1, \cdots, \varepsilon_p$ are independent random variables. If at least one of them belongs to Class A, then $\phi_\varepsilon(\mathbf{t})$ is in Class A with order matching the highest Class A order among all these $\varepsilon_i^{'}$s; if all $\varepsilon_1, \cdots, \varepsilon_p$ are in class C, then ε belongs to Class C with order matching the sum of the orders of the ε_i's.

EXAMPLE 3. (Multivariate Gamma distribution of type I)

Let $x_0, x_1, \cdots, x_m$ be independent Gamma(θ_j) distributions. Denote $y_j = x_j + x_0, j = 1, 2, \cdots, m$. The distribution of $(y_1, \cdots, y_m)'$ is called the type I Gamma-Distribution in Johnson and Kotz (1972). Its characteristic function is

$$\phi(\mathbf{t}) = (1 - \sum_{j=1}^{m} it_j)^{-\theta_0} \prod_{j=1}^{m} (1 - it_j)^{-\theta_j},$$

which implies that it belongs to Class C of order $\sum_{j=0}^{m} \theta_j$.

EXAMPLE 4. We say that X possesses a symmetric Kotz type distribution if it has density

$$C_n |\Sigma|^{-1/2} [(X - \mu)' \Sigma^{-1} (X - \mu)]^{N-1} \exp\{-r[(X - \mu)' \Sigma^{-1} (X - \mu)]^s\}$$

with $r, s > 0, 2N + n > 2$. This is a very useful class from both practical and theoretical viewpoints [see Fang, Kotz and Ng (1990)]. The characteristic function of this distribution is

$$\phi(\mathbf{t}) = \exp(i\mathbf{t}'\mu) \psi_{n,N}(\mathbf{t}'\Sigma\mathbf{t}; r)$$

where

$$\psi_{n,N}(u; r) = \exp(-u/4r) \sum_{m=0}^{N-1} \binom{N-1}{m} \frac{\Gamma(n/2)}{\Gamma(n/2 + m)} (-\frac{u}{4r})^m.$$

This distribution belongs to Class A of order 2.

EXAMPLE 5. Another useful class of distributions is the symmetric Pearson type VII distributions. With a density generator $g(t) = C_n(1 + t/m)^{-N}$ where $N > n/2, m > 0$, it has density

$$C_n |\Sigma|^{-1/2} (1 + (X - \mu)' \Sigma^{-1} (X - \mu)/m)^{-N}.$$

When $N = (n + m)/2$, it reduces to the multivariate t-distribution; For $N = (n + 1)/2$ and $m = 1$, we have the multivariate Cauchy distribution. The characteristic function is $\phi(\mathbf{t}) = \exp(i\mathbf{t}'\mu)\psi(\mathbf{t}'\Sigma\mathbf{t})$ where

$$\psi(u^2) = \frac{\Gamma(N - (n-1)/2)}{\pi^{1/2}\Gamma(N - n/2)} \int_{-\infty}^{\infty} \cos(m^{1/2} tu)(1 + t^2)^{-N + (n-1)/2} dt.$$

For general N and m, the decay rate is complicated. For the above two special cases, we have the following results.

(1) Multivariate t-distribution : $N = (m+n)/2$. For the odd degree of freedom m, we have

$$\begin{aligned} \psi(\mathbf{t}) &= \frac{\pi^{1/2}\Gamma((m+1)/2)}{2^{m-1}\Gamma(m/2)} \exp(i\mathbf{t}'\mu - (\mathbf{t}'\Sigma\mathbf{t})^{1/2}) \\ &\quad \times \sum_{r=1}^{s} [\binom{2s-r-1}{s-r} \frac{(2m^{1/2}\mathbf{t}'\Sigma\mathbf{t})^{r-1}}{(r-1)!}] \end{aligned}$$

where $s = (m+1)/2$. It is in Class A of order 1.

(2) Multivariate Cauchy. Now we have $\psi(u) = \exp(-u^{1/2})$, and hence is in Class A of order 1.

3 Mixing density kernel estimation

We shall first derive a class of multivariate mixing density kernel estimators for $f(\mathbf{x})$. From the model $Y = X + \varepsilon$, we have $\phi_Y(\mathbf{t}) = \phi_X(\mathbf{t})\phi_\varepsilon(\mathbf{t})$, where ϕ_Y and ϕ_X denote the characteristic function of Y and X respectively. Suppose $\phi_\varepsilon(\mathbf{t}) \neq 0$ for all $\mathbf{t}$ and that $\int |\phi_Y(\mathbf{t})/\phi_\varepsilon(\mathbf{t})| d\mathbf{t} < \infty$, then the inverse Fourier transform gives

$$f(\mathbf{x}) = (2\pi)^{-p} \int \exp(-i\mathbf{t}'\mathbf{x})\phi_Y(\mathbf{t})[\phi_\varepsilon(\mathbf{t})]^{-1} d\mathbf{t}. \tag{1}$$

Let $K : R^p \to R$ be a kernel satisfying

1. $K(\mathbf{y}) = K(-\mathbf{y})$
2. $\int K(\mathbf{y}) d\mathbf{y} = 1$
3. $\int x_i^\nu K(\mathbf{x}) dx_i = 0$ for all $1 \leq i \leq p$, $1 \leq \nu \leq m$ and $\mathbf{x} \in R^p$.

Writing

$$f_n^*(\mathbf{y}) = (nh^p)^{-1} \sum_{j=1}^{n} K(\frac{\mathbf{y} - Y_j}{h}),$$

we get the following estimate for ϕ_Y,

$$\begin{aligned} \hat{\phi}_Y(\mathbf{t}) &= \int \exp(i\mathbf{t}'\mathbf{y}) f_n^*(\mathbf{y}) d\mathbf{y} \\ &= \phi_K(h\mathbf{t})\hat{\phi}_n(\mathbf{t}) \end{aligned}$$

where $\hat{\phi}_n(\mathbf{t}) = (1/n)\sum_{j=1}^n \exp(i\mathbf{t}'Y_j)$ and ϕ_K is the Fourier transform of K. Substituting

$$\phi_X(\mathbf{t}) = \phi_K(h\mathbf{t})\hat{\phi}_n(\mathbf{t})[\phi_\varepsilon(\mathbf{t})]^{-1}$$

into (1), we have

$$\begin{aligned} \hat{f}_n(\mathbf{x}) &= (2\pi)^{-p}\int \exp(-i\mathbf{t}'\mathbf{x})\phi_K(h\mathbf{t})\hat{\phi}_n(\mathbf{t})[\phi_\varepsilon(\mathbf{t})]^{-1}d\mathbf{t} \\ &= \frac{1}{nh^p}\sum_{j=1}^n K_c(\mathbf{x}; h, Y_j), \end{aligned} \tag{2}$$

where

$$K_c(\mathbf{x}; h, Y) = (2\pi)^{-p}\int \exp(-i\mathbf{t}'(\mathbf{x}-Y)/h)\phi_K(\mathbf{t})[\phi_\varepsilon(\mathbf{t}/h)]^{-1}d\mathbf{t}. \tag{3}$$

We propose using $\hat{f}_n(\mathbf{x})$ as an estimator for $f(\mathbf{x})$. Let m be a non-negative integer, $0 < \alpha \le 1$ and $B \ge 0$. We shall investigate the convergence rate of $\hat{f}_n(\mathbf{x})$ with respect to the function class $\mathcal{F}_{m,\alpha,B}$ where $\mathcal{F}_{m,\alpha,B}$ denotes the class of p-dimensional densities whose elements satisfy

$$|\frac{\partial^m}{\partial x_i^m}f(\mathbf{y}) - \frac{\partial^m}{\partial x_i^m}f(\mathbf{x})| \le B||\mathbf{y}-\mathbf{x}||^\alpha, \quad \forall \mathbf{x}, \mathbf{y} \in R^p, 1 \le i \le p. \tag{4}$$

For $f \in \mathcal{F}_{m,\alpha,B}$, using Taylor's formula we have

$$\begin{aligned} f(\mathbf{y}+\mathbf{x}) - f(\mathbf{x}) &= \sum_{i=1}^p \frac{\partial f(\mathbf{x})}{\partial x_i}y_i + \frac{1}{2}\sum_{i,j}^p \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}y_i y_j + \cdots \\ &+ \frac{1}{l!}\sum_{\alpha_1+\cdots+\alpha_p=l} \frac{\partial^l f(\mathbf{x})}{\partial x_1^{\alpha_1}\cdots\partial x_p^{\alpha_p}}y_1^{\alpha_1}\cdots y_p^{\alpha_p} + \cdots \\ &+ \frac{1}{m!}\sum_{\alpha_1+\cdots+\alpha_p=m} \frac{\partial^m f(\xi)}{\partial x_1^{\alpha_1}\cdots\partial x_p^{\alpha_p}}y_1^{\alpha_1}\cdots y_p^{\alpha_p}. \end{aligned}$$

Hence by the conditions stated above, we have

$$\begin{aligned} & E\hat{f}_n(\mathbf{x}) - f(\mathbf{x}) \\ = \; & E\frac{1}{nh^p}\sum_{j=1}^n K_c(\mathbf{x}; h, Y_j) - f(\mathbf{x}) \\ = \; & \int (f(\mathbf{x}-\mathbf{y}) - f(\mathbf{x}))\frac{1}{h^p}K(\mathbf{y}/h)d\mathbf{y} \\ = \; & \frac{1}{m!h^p}\int K(\frac{\mathbf{y}-\mathbf{x}}{h})[\sum_{i=1}^p (y_i - x_i)^m(\frac{\partial^m}{\partial x_i^m}f(\xi) - \frac{\partial^m}{\partial x_i^m}f(\mathbf{x}))]d\mathbf{y}. \end{aligned}$$

The last equation holds because that $\int K((\mathbf{y}-\mathbf{x})/h)\sum(y_i - x_i)^m d\mathbf{y} = 0$. Consequently, we have

$$
\begin{aligned}
|E\hat{f}_n(\mathbf{x}) - f(\mathbf{x})| &\leq \frac{B}{h^p}\int |K(\frac{\mathbf{y}-\mathbf{x}}{h})| \times ||\mathbf{y}-\mathbf{x}||^{\alpha}\frac{\sum_{i=1}^{p}|y_i - x_i|^m}{m!}d\mathbf{y} \\
&\leq \frac{Bp}{m!h^p}\int |K(\frac{\mathbf{y}-\mathbf{x}}{h})| \times ||\mathbf{y}-\mathbf{x}||^{\alpha+m}d\mathbf{y} \\
&= h^{\alpha+m}\frac{\Delta}{m!}, \quad \text{say.} \qquad (5)
\end{aligned}
$$

This gives the bound for the bias term. For the variance term, we have

$$
\begin{aligned}
Var(\hat{f}_n(\mathbf{x})) &= \frac{1}{n^2h^{2p}}\sum_{j=1}^{n} Var(K_c(\mathbf{x}; h, Y_j)) \\
&\leq \frac{1}{nh^{2p}}EK_c^2(\frac{\mathbf{x}-Y}{h}) \\
&= \frac{1}{n(2\pi h)^{2p}}E|\int e^{-i\mathbf{t}'(\mathbf{x}-Y)/h}\frac{\phi_K(\mathbf{t})}{\phi_\varepsilon(\mathbf{t}/h)}d\mathbf{t}|^2 \\
&\leq \frac{1}{n(2\pi h)^{2p}}(\int \frac{|\phi_K(\mathbf{t})|}{|\phi_\varepsilon(\mathbf{t}/h)|}d\mathbf{t})^2. \qquad (6)
\end{aligned}
$$

Theorem 1 *Let the kernel K satisfy conditions 1 to 3 and that $\phi_K(\mathbf{t}) = 0$ for $\mathbf{t} \notin [-1,1]^p$; Also suppose that*

(a) $\phi_\varepsilon(\mathbf{t}) \neq 0$ for any $\mathbf{t}$;

(b) $|\phi_\varepsilon^{\mathbf{a}}(t)||t|^{-\beta_0}\exp(|t|^\beta/\gamma_0) \geq d_0$ for all unit vectors $\mathbf{a}$ as $|t| \to \infty$.

Then

$$\sup_{f\in\mathcal{F}_{m,\alpha,B}} E(\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))^2 = O((\log n)^{-2(m+\alpha)/\beta}), \qquad (7)$$

with the bandwidth $h = (4/\gamma_0)^{1/\beta}(\log n)^{-1/\beta}$.

PROOF. According to (5) and (6), we have

$$
\begin{aligned}
\sup_{f\in\mathcal{F}_{m,\alpha,B}} E(\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))^2 &= \sup_{f\in\mathcal{F}_{m,\alpha,B}} (E\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))^2 + Var(\hat{f}_n(\mathbf{x})) \\
&\leq h^{2(m+\alpha)}\frac{\Delta^2}{(m!)^2} + \frac{1}{nh^a}O(\exp(2h^{-\beta}/\gamma_0))
\end{aligned}
$$

where $a = 2p$ if $\beta_0 > 0$; otherwise $a = 2p - 2\beta_0$. The proof is completed by choosing $h = (4/\gamma_0)^{1/\beta}(\log n)^{-1/\beta}$. □

REMARK. It is interesting to note that the right hand side of (7) does not depend on the dimension p. As we will see in the next section, this rate cannot be improved upon (up to a multiplying constant) by any estimate based only on the n i.i.d. observations. This implies that the kernel estimator $\hat{f}(\mathbf{x})$ as given in (2) is optimal.

For ε in the Class B, we have the following result.

Theorem 2 *Let the kernel K satisfy conditions 1 to 3 and that $\phi_K(\mathbf{t}) = 0$ for $\mathbf{t} \notin [-1,1]^p$; Also suppose that*

(a) $\phi_\varepsilon(\mathbf{t}) \neq 0$ for any $\mathbf{t}$;

(b) $d_0|t|^{k_0} \exp(-(\log|t|)^\beta/\gamma_0) \leq |\phi_\varepsilon^{\mathbf{a}}(t)|$ for all unit vector $\mathbf{a}$ as $|t| \to \infty$. Then

$$\sup_{f \in \mathcal{F}_{m,\alpha,B}} E(\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))^2 \leq C \exp(-2(\gamma_0/2)^{1/\beta}(m+\alpha)(\log n)^{1/\beta}),$$

with bandwidth $h = \exp(-(\gamma_0/c)^{1/\beta}(\log n)^{1/\beta})$.

PROOF. We observe that

$$\sup_{f \in \mathcal{F}_{m,\alpha,B}} E(\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))^2 \leq h^{2(m+\alpha)} \frac{\Delta^2}{(m!)^2} + \sup_{f \in \mathcal{F}_{m,\alpha,B}} Var(\hat{f}_n(\mathbf{x})).$$

by virtue of Theorem 1. ¿From (6), we have

$$\begin{aligned} Var(\hat{f}_n(\mathbf{x})) &\leq \frac{1}{n(2\pi h)^{2p}} \left(\int \frac{|\phi_K(\mathbf{t})|}{|\phi_\varepsilon(\mathbf{t}/h)|} d\mathbf{t}\right)^2 \\ &\leq \frac{C}{n(2\pi h)^{2p}} h^{2k_0} \exp(2(|\log h|)^\beta/\gamma_0). \end{aligned}$$

Let $c > 2$ be a constant. Then by taking $h = \exp(-(\gamma_0/c)^{1/\beta}(\log n)^{1/\beta})$, we have

$$\sup_{f \in \mathcal{F}_{m,\alpha,B}} E(\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))^2 \leq O(1) \exp(-2(m+\alpha)(\gamma_0/c)^{1/\beta}(\log n)^{1/\beta}).$$

□

Next we consider the case when ε belongs to Class C.

Theorem 3 *Let the kernel K satisfy conditions 1 to 3 and that $\phi_K(\mathbf{t}) = 0$ for $\mathbf{t} \notin [-1,1]^p$; Also suppose that*

(a) $\phi_\varepsilon(\mathbf{t}) \neq 0$ for any $\mathbf{t}$;

(b) $|\phi_\varepsilon^{\mathbf{a}}(t)||t|^\beta \geq d_0$ for all unit vectors $\mathbf{a}$ as $|t| \to \infty$.

Then by choosing the bandwidth $h = dn^{-1/(2(k+\beta)+p)}$, we have

$$\sup_{f \in \mathcal{F}_{m,\alpha,B}} E(\hat{f}_n(\mathbf{x}) - f(\mathbf{x}))^2 = O(n^{-2k/(2k+2\beta+p)}),$$

where $k = m + \alpha$.

PROOF. Since $\hat{f}_n(\mathbf{x}) = (1/n)\sum_{j=1}^n K_c((\mathbf{x}-Y_j)/h)$, with K_c as defined in (3), we have

$$\begin{aligned} EK_c^2(\frac{\mathbf{x}-Y_1}{h}) &= \int K_c^2(\frac{\mathbf{x}-\mathbf{y}}{h})f_Y(\mathbf{y})d\mathbf{y} \\ &= \int K_c^2(\mathbf{y})f_Y(\mathbf{x}-h\mathbf{y})h^p d\mathbf{y} \\ &\leq Ch^p\int K_c^2(\mathbf{y})d\mathbf{y} \\ &= Ch^p\int \frac{|\phi_K(\mathbf{t})|^2}{|\phi_\varepsilon(\mathbf{t}/h)|^2}d\mathbf{t}. \end{aligned}$$

The inequality holds because that $\sup_{f\in\mathcal{F}}|f(\mathbf{x})| \leq C$ [Similar to that in Bickel and Ritov (1988)]. Because $\phi_\varepsilon(\mathbf{t}) \neq 0$, hence for fixed M, $\exists\ C_0$ such that $|\phi_\varepsilon(\mathbf{t}/h)| \geq C_0$ for $\mathbf{t}/h \in [-M,M]^p$. Observing that

$$\int_{R^p\setminus[-hM,hM]^p}\frac{|\phi_K(\mathbf{t})|^2}{|\phi_\varepsilon(\mathbf{t}/h)|^2}d\mathbf{t} \leq \int_{R^p\setminus[-hM,hM]^p}\frac{|\phi_K(\mathbf{t})|^2}{d_0h^{2\beta}}d\mathbf{t} \leq C_1h^{-2\beta},$$

we get $\sup_{f\in\mathcal{F}_{\alpha,mB}} Var(\hat{f}_n(x)) \leq D/(nh^{p+2\beta})$. Combining with the bias term, we complete the proof by taking $h = dn^{-1/(2(k+\beta)+p)}$. Here C_0, C_1,and D are constants. □

4 Lower bounds on the optimal convergence rate

In this section we are concerned with the establishment of lower bounds for the optimal rate of convergence, i.e. $\inf_{\hat{T}_n}\sup_{f\in\mathcal{F}_{\alpha,mB}} E(\hat{T}_n - f(\mathbf{x}))^2$, where the infimum is over all estimators $\hat{T}_n$ based on the n i.i.d. observations. We start with a brief review of the method for getting a lower bound for a general nonparametric estimation problem, A detailed discussion can be found in Donoho, Liu (1987), (1991a) and (1991b).

Let $F_{1,n}$ be a sequence of distributions that converges to a distribution F_0 in a certain sense. Let $P_{1,n}$ be the product measure with marginal distribution $F_{1,n}$, and $P_{0,n}$ be the product measure with marginal distribution F_0. We denote the likelihood ratio as $L_n = dP_{1,n}/dP_{0,n}$.

For a measurable set S, define

$$\pi(P_{0,n}, P_{1,n}) = \inf_S(P_{0,n}(S) + P_{1,n}(S^C))$$

In the setting of testing whether a set of random variables comes from F_0 or $F_{1,n}$, π is the minimum of the sum of type I and type II errors over all possible tests.

For sequences of product measures $(P_{0,n})$ and $(P_{1,n})$, we say that there exists no perfect test between $\{P_{0,n}\}$ and $\{P_{1,n}\}$ if there exists $\alpha > 0$ such that

$$\pi(P_{0,n}, P_{1,n}) > \alpha, \quad \forall n > n_0.$$

Furthermore we say that there is a good test between $(P_{0,n})$ and $(P_{1,n})$ if there is a $\beta > 0$ such that

$$1 - \beta > \pi(P_{0,n}, P_{1,n}), \quad \forall n > n_0.$$

Note that this bounds the sum of errors uniformly in n.

These concepts are important for deriving the lower bound, because from them, we can derive a lower bound for a certain very useful probability inequality given below. Let $T(\cdot)$, a functional defined on a space of probability distributions, be the quantity we want to estimate and $\hat{T}_n$ is any estimator based only on n i.i.d. observations distributed according to the true (but unknown) distribution.

Theorem 4 *If there is no perfect test between $(P_{0,n})$ and $(P_{1,n})$, then*

$$\inf_{\hat{T}_n} \max_{F \in \{F_0, F_{1,n}\}} P_F\left(|\hat{T}_n - T(F)| > \Delta_n\right) > \alpha/2$$

where $\Delta_n = |T(F_{1,n}) - T(F_0)|/2$. Hence $\sup_F E_F(\hat{T}_n - T(F))^2 > (\alpha^2/4)\Delta_n^2$.

PROOF. Let $S = \{\hat{T}_n - T(F_0) < \Delta_n\}$. Since

$$\{\hat{T}_n - T(F_1) > -\Delta_n\} \supset \{\hat{T}_n - T(F_0) > \Delta_n\},$$

we have

$$\begin{aligned} & P_{F_0}\{\hat{T}_n - T(F_0) < \Delta_n\} + P_{F_1}\{\hat{T}_n - T(F_1) > -\Delta_n\} \\ \geq\ & P_{F_0}(S) + P_{F_1}(S^C) \geq \pi(P_0, P_{1,n}) > \alpha. \end{aligned}$$

Using the inequality $\max(a, b) \geq \text{ave}(a, b) = (a + b)/2$, we conclude that

$$\begin{aligned} & \inf \max_{F \in \{F_0, F_{1,n}\}} P_F\{|\hat{T}_n - T(F)| > \Delta_n\} \\ \geq\ & \inf \text{ave}_{F \in \{F_0, F_{1,n}\}} P_F\{|\hat{T}_n - T(F)| > \Delta_n\} > \alpha/2. \end{aligned}$$

□

There are now many known sufficient conditions that would ensure the non-existence of a perfect test. The theorem below is due to Farrell (1972). Similar results can also be found in Donoho and Liu (1987).

Theorem 5 *Let* $\Delta_n = |T(F_{1,n}) - T(F_0)|/2$. *If* $\{F_{1,n}\}$ *satisfies the constraint* $E_{P_{0,n}} L_n^2 < c_0 < \infty$, *then there is no perfect test between* $(P_{0,n})$ *and* $(P_{1,n})$.

There are several equivalent conditions to $E_{P_{0,n}} L_n^2 < c_0 < \infty$. In Stone (1980), the condition $E_{P_{0,n}} |\log L_n| < c$ was used; Hasminskii (1979) used $\log L_n \rightarrow_D N(1/2c_0, c_0^2)$ to derive the lower bound. Here we will use the χ^2 distance.

Let the distributions F_0 and F_1 have densities f_0 and f_1 respectively. The χ^2 distance between F_1 and F_0 is defined to be

$$\chi^2(F_0, F_1) = \int (f_1 - f_0)^2 / f_0.$$

Lemma 2 *With the above notation,*

$$1 + \chi^2(P_{0,n}, P_{1,n}) = E_{P_{0,n}} L_n^2$$

and

$$\chi^2((P_{0,n}, P_{1,n}) = (\chi^2(F_{1,n}, F_0) + 1)^n - 1.$$

Hence $\chi^2(F_{1,n}, F_0) \leq c/n$ *implies that* $E_{P_{0,n}} L_n^2 \leq (1 + c/n)^n < e^c$.

PROOF. We observe that

$$\begin{aligned} E_{P_{0,n}} L_n^2 &= \int (\prod_{j=1}^n f_1(x_j))^2 / \prod_{j=1}^n f_0(x_j) \\ &= \int (\frac{\prod_{j=1}^n f_1(x_j) - \prod_{j=1}^n f_0(x_j)}{\prod_{j=1}^n f_0(x_j)} + 1)^2 \prod_{j=1}^n f_0(x_j) \\ &= \int (\frac{\prod_{j=1}^n f_1(x_j) - \prod_{j=1}^n f_0(x_j)}{\prod_{j=1}^n f_0(x_j)})^2 \prod_{j=1}^n f_0(x_j) + 1 \\ &= \chi^2(P_{0,n}, P_{1,n}) + 1, \end{aligned}$$

and

$$\begin{aligned}\chi^2(P_{1,n}, P_{0,n}) &= \int \frac{(\prod_{j=1}^n f_1(x_j) - \prod_{j=1}^n f_0(x_j))^2}{\prod_{j=1}^n f_0(x_j)} \\ &= \int \frac{\prod_{j=1}^n f_1^2(x_j)}{\prod_{j=1}^n f_0(x_j)} - 1 \\ &= (\int \frac{f_1^2(x_1)}{f_0(x_1)} dx_1 - 1 + 1)^n - 1 \\ &= (\chi^2(F_{1,n}, F_0) + 1)^n - 1.\end{aligned}$$

This proves the lemma. □

Suppose now we observe a random sample $Y_1, Y_2, \cdots, Y_n$ from the model $Y = X + \varepsilon$ where ε is a noise component with known distribution, and X has an unknown density f which belongs to the smooth function class $\mathcal{F}_{m,\alpha,B}$. We are interested in estimating $T(f) = f(\mathbf{x})$. Without loss of generality, we assume that $\mathbf{x} = 0$ and hence $T(f) = f(0)$.

Select a $f_0 \in \mathcal{F}_{m,\alpha,B}$ and define

$$f_n(\mathbf{x}) = f_0(\mathbf{x}) + \delta_n^k H(\mathbf{x}/\delta_n), \qquad \forall \mathbf{x} \in R^p,$$

where $k = m + \alpha$ and $H : R^p \to R$ is a function which satisfies that $H(0) \neq 0$, $\int H(\mathbf{x}) d\mathbf{x} = 0$, and that the mth derivatives of H satisfy the Lipschitz's condition of order α, that is, H satisfies (4). By a further suitable choice of the tails of H and f_0, the function f_n will also be a density in $\mathcal{F}_{m,\alpha,B}$ for small δ_n. We will choose δ_n such that $\chi^2(f_{Y_n}, f_{Y_0}) \leq c/n$, where f_{Y_n} is the convolution of ε and f_n, and f_{Y_0} is the convolution of ε and f_0.

If $\chi^2(f_{Y_0}, f_{Y_n}) \leq c/n$, we have

$$\inf_{\hat{T}_n} \sup_{f \in \{f_n, f_0\}} P_f(|\hat{T}_n - T(f)| > |T(f_n) - T(f_0)|/2) > d_1,$$

and hence by Tchebycheff's inequality,

$$\sup_{f \in \mathcal{F}_{m,\alpha,B}} E_f(\hat{T}_n - T(f))^2 > d_1^2 |T(f_0) - T(f_n)|^2/4 = \delta_n^{2k} |H(0)| d_1^2/4.$$

This gives us a lower bound δ_n^{2k}. Now we apply this argument to the case where ε belongs to Class A.

Theorem 6 *Suppose that the tail of ϕ_ε satisfies*

$$|\phi_\varepsilon^{\mathbf{a}}(t)||t|^{-\beta_1} \exp(|t|^\beta/\gamma) \leq d_1 \tag{8}$$

as $|t| \to \infty$ *for all* $\mathbf{a} \in \mathbf{A}(\varepsilon)$ *,where* $\beta, \gamma > 0, d_1 \geq 0$ *and* β_1 *are all constants, and*

$$P\{||\varepsilon - \mathbf{x}|| \leq ||\mathbf{x}||^{\alpha_0}\} = O(||\mathbf{x}||^{-(a-\alpha_0)}) \tag{9}$$

as $||\mathbf{x}|| \to \infty$ *for some* $0 < \alpha_0 < 1$*,* $a > p + \alpha_0$*. Then*

$$\inf_{\hat{T}_n} \sup_{f \in \mathcal{F}_{m,\alpha,B}} E_f(\hat{T}_n - T(f))^2 \geq d(\log n)^{-2(m+\alpha)/\beta}$$

for some constant $d > 0$*.*

Without loss of generality we shall assume that $\mathbf{A}(\varepsilon)$ corresponds to the hypercube $[1,2]^p$. To prove Theorem 6, we need the following two lemmas.

Lemma 3 *Suppose that* F *is a p-dimensional distribution function. Then by choosing* r *such that* $2r > p$*,*

$$g_0(\mathbf{x}) = \int \frac{C_r}{(1 + ||\mathbf{x} - \mathbf{y}||^2)^r} dF(\mathbf{y})$$

satisfies $g_0(\mathbf{x}) \geq D||\mathbf{x}||^{-2r}$ *as* $||\mathbf{x}|| \to \infty$ *for some constant* $D > 0$*, where* C_r *is the constant such that* $C_r(1 + ||\mathbf{x}||^2)^{-r}$ *is a density function.*

PROOF. To ensure that $C_r(1 + ||\mathbf{x}||^2)^{-r}$ is a density, we must have

$$\int \frac{1}{(1 + ||\mathbf{x}||^2)^r} d\mathbf{x} < \infty.$$

Under L^2 norm, this is equivalent to

$$\int_0^\infty \frac{\rho^{p-1}}{(1 + \rho^2)^r} d\rho < \infty,$$

This can be true only if $2r > p$. By the triangular inequality we have

$$\begin{aligned} \frac{1}{(1 + ||\mathbf{x} - \mathbf{y}||^2)^r} &\geq \frac{1}{(1 + (||\mathbf{x}|| + ||\mathbf{y}||)^2)^r} \\ &\geq \frac{1}{(1 + 2||\mathbf{x}||^2 + 2||\mathbf{y}||^2)^r}. \end{aligned}$$

Let $||\mathbf{x}|| > 1$, we then have

$$\begin{aligned} \int \frac{1}{(1 + ||\mathbf{x} - \mathbf{y}||^2)^r} dF(\mathbf{y}) &\geq \int \frac{1}{(1 + 2||\mathbf{x}||^2 + 2||\mathbf{y}||^2)^r} dF(\mathbf{y}) \\ &\geq \int \frac{1}{||\mathbf{x}||^{2r}(2 + ||\mathbf{x}||^{-2} + 2||\mathbf{x}||^{-2}||\mathbf{y}||^{-2})^r} dF(\mathbf{y}) \\ &\geq ||\mathbf{x}||^{-2r} \int \frac{1}{(3 + 2||\mathbf{y}||^2)^r} dF(\mathbf{y}) \\ &= D||\mathbf{x}||^{-2r}. \end{aligned}$$

This proves Lemma 3. □

Lemma 4 *Suppose* $P\{||\varepsilon - \mathbf{x}|| \leq ||\mathbf{x}||^{\alpha_0}\} = O(||\mathbf{x}||^{-(a-\alpha_0)})$ *as* $||\mathbf{x}|| \to \infty$ *for some* $0 < \alpha_0 < 1$ *and* $a > p + \alpha_0$. *Suppose* $H : R^p \to R$ *be a bounded function satisfying* $H(\mathbf{x}) = O(||\mathbf{x}||^{-m_0})$. *Then there exist constants* M *and* C *such that whenever* $||\delta\mathbf{x}|| > M$, *we have*

$$\int H(\mathbf{x}-\mathbf{y})dF_\varepsilon(\delta\mathbf{y}) \leq C(||\delta\mathbf{x}||)^{-a+\alpha_0}, \quad \forall\delta \leq 1, \tag{10}$$

provided $(m_0 + 1)\alpha_0 > a$.

PROOF. We observe that

$$\begin{aligned}\int H(\mathbf{x}-\mathbf{y})dF_\varepsilon(\delta\mathbf{y}) &= \int_{||\mathbf{x}-\mathbf{y}/\delta|| \leq ||\mathbf{x}||^{\alpha_0}} H(\mathbf{x}-\mathbf{y}/\delta)dF_\varepsilon(\mathbf{y}) \\ &\quad + \int_{||\mathbf{x}-\mathbf{y}/\delta|| > ||\mathbf{x}||^{\alpha_0}} H(\mathbf{x}-\mathbf{y}/\delta)dF_\varepsilon(\mathbf{y}) \\ &\leq C_1 P\{||\mathbf{x}-\mathbf{y}/\delta|| \leq ||\mathbf{x}||^{\alpha_0}\} + O(||\mathbf{x}||^{-\alpha_0 m_0}) \\ &\leq C_1 P\{||\delta\mathbf{x}-\mathbf{y}|| \leq \delta^{1-\alpha_0}||\delta\mathbf{x}||^{\alpha_0}\} + O(||\mathbf{x}||^{-\alpha_0 m_0}) \\ &\leq C_2||\delta\mathbf{x}||^{-(a-\alpha_0)} + O(||\mathbf{x}||^{-m_0\alpha_0}).\end{aligned}$$

Since $(m_0 + 1)\alpha_0 > a$, we have

$$\int H(\mathbf{x}-\mathbf{y})dF_\varepsilon(\delta\mathbf{y}) \leq C||\delta\mathbf{x}||^{-(a-\alpha_0)}$$

whenever $||\delta\mathbf{x}|| > M$. □

REMARK. We observe that (10) says the tail probability of ε goes to 0 at a polynomial rate of order greater than p.

PROOF OF THEOREM 6. First we construct a function $H : R^p \to R$ which satisfies

(i) $H(\mathbf{x}) = O(||\mathbf{x}||^{-m_0})$, for some given m_0 such that$(m_0 + 1)\alpha_0 > a$;

(ii) $\int H(\mathbf{x})d\mathbf{x} = 0$;

(iii) $\phi_H(\mathbf{t}) = 0$ whenever $\mathbf{t} \notin [1, 2]^p$, where ϕ_H is the Fourier transform of H;

(iv) $H(\mathbf{x})$ have all bounded continuous derivatives.

The existence of such a function was proved in Fan (1991a) when $p = 1$. For the sake of completeness, we shall now construct such a H.

Take a symmetric bounded function $\phi(t)$ which vanishes outside [1,2] for $t > 0$, and let ϕ has continuous first m_0 derivatives.

Let $g(x)$ be the Fourier inversion of $\phi(t)$ defined by

$$g(x) = \frac{1}{\pi}\int_1^2 \cos(tx)\phi(t)dt. \tag{11}$$

Define $h(x) = g(x) - g(x+1)$, then its Fourier transform $\phi_h(t) = (1 - e^{-it})\phi(t) = 0$ for t outside [1,2]. Also such a h has the following properties:

1. $\int h(x)dx = 0$ because that $\phi_h(0) = 0$,

2. $g(x)$ have all bounded derivatives. This follows from (11) and that ϕ is bounded,

3. That $\phi(t)$ has m_0 continuous derivatives implies that $\phi^{(i)}(1) = \phi^{(i)}(2) = 0$ for $i = 1, \cdots, m_0$. Thus

$$\begin{aligned} g(x) &= \frac{1}{\pi}\int_1^2 \cos(tx)\phi(t)dt \\ &= \frac{1}{\pi x}\int_1^2 \phi(t)d\sin(tx) \\ &= -\frac{1}{\pi x}\int_1^2 \sin(tx)\phi^{(1)}(t)dx \\ &= -\frac{1}{\pi x^2}\int_1^2 \cos(tx)\phi^{(2)}(t)dx \\ &= \text{etc.,} \end{aligned}$$

which implies that $g(x) = O(x^{-m_0})$ when $|x| \to \infty$. Now define

$$H(\mathbf{x}) = \prod_{i=1}^{p} h(x_i).$$

It is straightforward to verify (ii) to (iv). For (i), notice that

$$\begin{aligned} \lim_{||\mathbf{x}|| \to \infty} H(\mathbf{x}) &= \lim_{\max|x_i| \to \infty} \prod_{i=1}^{p} h(x_i) \\ &= O((\max|x_i|)^{-m_0}) \\ &= O(||\mathbf{x}||^{-m_0}). \end{aligned}$$

The last equality holds because all norms in R^p are equivalent. Now consider the pair of densities

$$f_0(\mathbf{x}) = C_r/(1+||\mathbf{x}||^2)^r, \quad \forall \mathbf{x} \in R^p,$$

and

$$f_n(\mathbf{x}) = f_0(\mathbf{x}) + c\delta_n^k H(\mathbf{x}/\delta_n), \quad \forall \mathbf{x} \in R^p,$$

where $0.5p < r < \min[p, a - \alpha_0 - 0.5p]$. We observe that $f_0 \in \mathcal{F}_{m,\alpha,B}$. By choosing c properly and making δ_n small, we can ensure that $f_n \in \mathcal{F}_{m,\alpha,B}$ provided that $k \geq m + \alpha$.

Let $f_{Y_0} = g_0 = f_0 * F_\varepsilon$ and $f_{Y_n} = f_n * F_\varepsilon$. Then

$$\begin{aligned} \chi^2(f_{Y_0}, f_{Y_n}) &= \int (f_0 * F_\varepsilon - f_n * F_\varepsilon)^2 g_0^{-1} \\ &= \delta_n^{2k} \int (\int H((\mathbf{x}-\mathbf{y})/\delta_n) dF_\varepsilon(\mathbf{y}))^2 g_0^{-1}(\mathbf{x}) d\mathbf{x} \\ &= \delta_n^{2k+p} \int (\int H(\mathbf{x}-\mathbf{y}) dF_\varepsilon(\delta_n \mathbf{y}))^2 g_0^{-1}(\delta_n \mathbf{x}) d\mathbf{x}. \end{aligned}$$

Using Parseval's identity, we have

$$\begin{aligned} \int \left(\int H(\mathbf{x}-\mathbf{y}) dF_\varepsilon(\delta_n \mathbf{y}) \right)^2 d\mathbf{x} &= (2\pi)^{-p} \int |\phi_H(\mathbf{t}) \phi_\varepsilon(\mathbf{t}/\delta_n)|^2 d\mathbf{t} \\ &= (2\pi)^{-p} \int_{[1,2]^p} |\phi_H(\mathbf{t}) \phi_\varepsilon(\mathbf{t}/\delta_n)|^2 d\mathbf{t} \\ &\leq O(\delta_n^{2\beta_1} \exp(-2\delta_n^{-\beta}/\gamma)). \end{aligned}$$

The last inequality holds because every $\mathbf{t}$ passes through $\mathbf{A}(\varepsilon)$. Consequently we observe from Lemmas 3 and 4 that

$$\begin{aligned} & \int \left(\int H(\mathbf{x}-\mathbf{y}) dF_\varepsilon(\delta_n \mathbf{y}) \right)^2 g_0^{-1}(\delta_n \mathbf{x}) d\mathbf{x} \\ \leq\ & \int_{|\delta_n \mathbf{x}| \leq M_n} + \int_{|\delta_n \mathbf{x}| > M_n} (H(\mathbf{x}-\mathbf{y}) dF_\varepsilon(\delta_n \mathbf{y}))^2 g_0^{-1}(\delta_n \mathbf{x}) d\mathbf{x} \\ \leq\ & M_n^{2r}/D \int \left(\int H(\mathbf{x}-\mathbf{y}) dF_\varepsilon(\delta_n \mathbf{y}) \right)^2 d\mathbf{x} + \int_{|\delta_n \mathbf{x}| > M_n} \frac{C^2 |\delta_n \mathbf{x}|^{-2a+2\alpha_0}}{D |\delta_n \mathbf{x}|^{-2r}} d\mathbf{x} \\ =\ & O(\delta_n^{-2\beta_1} M_n^{2r} \exp(-2\delta_n^{-\beta}/\gamma) + \delta_n^{-p} M_n^{-\varepsilon_0}) \\ \equiv\ & O(\Delta), \end{aligned}$$

where $\varepsilon_0 = 2(a - \alpha_0 - r) - p$. Now by taking $M_n = \exp(\xi\delta_n^{-\beta}/\gamma)$ with $\xi < 1/r$, we get

$$\begin{aligned} \Delta &= \delta_n^{-2\beta_1}\exp(-2\delta_n^{-\beta}/\gamma + 2r\xi\delta_n^{-\beta}/\gamma) + \delta_n^{-p}\exp(-\varepsilon_0\xi\delta_n^{-\beta}/\gamma) \\ &= o(\exp(-\varepsilon_1\delta_n^{-\beta})), \end{aligned}$$

where $\varepsilon_1 = \min((1 - r\xi)/\gamma, \varepsilon_0\xi/\gamma)$. Hence

$$\int (f_{Y_1} - f_{Y_0})^2 (f_{Y_0})^{-1} d\mathbf{x} = o(\delta_n^{2k+p}\exp(-\varepsilon_1\delta_n^{-\beta})).$$

Taking $\delta_n = \varepsilon_1^{1/\beta}(\log n)^{-1/\beta}$, we conclude that the χ^2 distance is of order $o(1/n)$, no matter what k is. However in order to make $f_n \in \mathcal{F}_{m,\alpha,B}$, we must have $k \geq m + \alpha$. Thus

$$|f_n(0) - f_0(0)| = O(\delta_n^k) = O((\log n)^{-(m+\alpha)/\beta}).$$

This completes the proof of Theorem 6. □

Next we consider the case when ε is in Class B.

Theorem 7 *Suppose (9) holds and*

$$|\phi_\varepsilon^{\mathbf{a}}(t)| \leq d_1|t|^{k_1}\exp(-(\log|t|)^\beta/\gamma_1)$$

for all $\mathbf{a} \in \mathbf{A}(\varepsilon)$. *Then there exists a constant* $d > 0$ *such that*

$$\inf_{\hat{T}_n} \sup_{f \in \mathcal{F}_{m,\alpha,B}} E_f(\hat{T}_n - T(f))^2 \geq d\exp(-2\varepsilon_1^{-1/\beta}(m+\alpha)(\log n)^{1/\beta}), \qquad (12)$$

where $\varepsilon_1 = \min((1 - r\xi)/\gamma_1, \varepsilon_0\xi/2\gamma_1)$, r, ξ *and* ε_0 *are as in the proof of Theorem 6.*

Proof. The proof is similar to that of Theorem 6. By choosing

$$M_n = \exp((|\log\delta|)^{-\beta}\xi/\gamma_1),$$

we have

$$\chi^2(f_{Y_0}, f_{Y_n}) \leq o(\exp(-\varepsilon_1(|\log\delta|)^{-\beta}).$$

Then $\chi^2(f_{Y_0}, f_{Y_n}) < 1/n$ if we take $\delta = \exp(-\varepsilon_1^{1/\beta}(\log n)^{-1/\beta})$. □

Remark. We observe from Theorem 2 that the convergence rate for the mixing density kernel estimator is $\exp(-2(\gamma_0/c)^{1/\beta}(m+\alpha)(\log n)^{1/\beta})$. It follows from $c > 2$, $\gamma_0 \leq \gamma_1$ and the definition of ε_1 that $\gamma_0/c < 1/\gamma_1$. This implies that the right hand side of (12) is asymptotically strictly smaller than the upper bound of Theorem 2.

When ε is in class C, we have the following result.

Theorem 8 *Suppose that the characteristic function ϕ_ε satisfies*

$$|t|^\beta |\phi_\varepsilon^{\mathbf{a}}(t)| \leq d_1, \quad \forall \mathbf{a} \in \mathbf{A}(\varepsilon), \tag{13}$$

for some constant d_1 and also

$$|\delta^{-\beta-k} \frac{\partial^k}{\partial t_1^{\alpha_1} \cdots \partial t_p^{\alpha_p}} \phi_\varepsilon(\mathbf{t}/\delta)| < C_k \tag{14}$$

for all $\mathbf{t} \in [1,2]^p$, and $1 \leq k \leq 2p$. Here $\alpha_1 + \cdots + \alpha_p = k$, and C_k are constants. Then there exists a constant $d > 0$ such that

$$\inf_{\hat{T}_n} \sup_{f \in \mathcal{F}_{m,\alpha,B}} E_f(\hat{T}_n - T(f))^2 > dn^{-2(m+\alpha)/(2m+2\alpha+2\beta+p)}.$$

Proof. We only need to estimate

$$\chi^2(f_{Y_0}, f_{Y_n}) = \delta_n^{2k+p} \int (\int H(\mathbf{x} - \mathbf{y}) dF_\varepsilon(\delta_n \mathbf{y}))^2 g_0^{-1}(\delta_n \mathbf{x}) d\mathbf{x}. \tag{15}$$

We shall use the same notation as in the proof of Theorem 6 with the exception of g_0. Define

$$g_0(\mathbf{x}) = \int C_r \{\prod_{i=1}^p [1 + (y_i - x_i)^2]\}^{-1} dF_\varepsilon(\mathbf{y}). \tag{16}$$

As in the proof of Lemma 3, we observe that for $\mathbf{x} \notin [-M, M]^p$ as $M \to \infty$,

$$g_0(\mathbf{x}) \geq D|x_1|^2 \cdots |x_p|^2,$$

for some constant D. By the Fourier inversion formula, we have

$$\int H(\mathbf{x} - \mathbf{y}) dF_\varepsilon(\delta \mathbf{y}) = \frac{1}{(2\pi)^p} \int e^{-i\mathbf{t}'\mathbf{x}} \phi_H(\mathbf{t}) \phi_\varepsilon(\mathbf{t}/\delta) d\mathbf{t}.$$

Define $\phi^\delta(\mathbf{t}) = (1/2\pi)^p \phi_H(\mathbf{t}) \phi_\varepsilon(\mathbf{t}/\delta)$, Note that ϕ_H posesses very good properties, hence condition (14) can be passed to ϕ_H. Also note that ϕ_H and its derivatives are 0 outside the hypercube $[1,2]^p$. Hence whenever $\mathbf{x} \notin [-1,1]^p$, we have

$$
\begin{aligned}
& \int H(\mathbf{x}-\mathbf{y})dF_\varepsilon(\delta\mathbf{y}) \\
= & \int_{[1,2]^p} e^{-i\mathbf{t}'\mathbf{x}}\phi^\delta(\mathbf{t})d\mathbf{t} \\
= & \int e^{-i(t_2x_2+\cdots+t_px_p)}(-1/(2\pi x_1^2)e^{-it_1x_1}\frac{\partial^2}{\partial t_1^2}\phi^\delta(\mathbf{t}))d\mathbf{t} \\
= & \cdots \\
= & (-\frac{1}{2\pi})^p\frac{1}{x_1^2\cdots x_p^2}\int e^{-i\mathbf{t}'\mathbf{x}}\frac{\partial^{2p}\phi^\delta(\mathbf{t})}{\partial t_1^2\cdots\partial t_p^2}d\mathbf{t}.
\end{aligned}
$$

Now we divide the right hand side of (15) into two terms. For the first term we get

$$
\begin{aligned}
TERM_1 & \equiv \int_{[-1,1]^p}(\int H(\mathbf{x}-\mathbf{y})dF_\varepsilon(\delta\mathbf{y}))^2 g_0^{-1}(\delta\mathbf{x})d\mathbf{x} \\
& \leq C_1\int(\int H(\mathbf{x}-\mathbf{y})dF_\varepsilon(\delta\mathbf{y}))^2 d\mathbf{x} \\
& = C_1\int|\phi_H(\mathbf{t})\phi_\varepsilon(\mathbf{t}/\delta)|^2 d\mathbf{t} \\
& = O(\delta^{2\beta}).
\end{aligned}
$$

The second last and final equalities use Parseval's identity and (13) respectively, where C_1 is the maximum of $g_0^{-1}(\mathbf{x})$ over $[-1,1]^p$.

For the second term, we have

$$
\begin{aligned}
& TERM_2 \\
\equiv & \int_{R^p\setminus[-1,1]^p}(\int H(\mathbf{x}-\mathbf{y})dF_\varepsilon(\delta\mathbf{y}))^2 g_0^{-1}(\delta\mathbf{x})d\mathbf{x} \\
= & \int_{R^p\setminus[-1,1]^p}(2\pi)^{-2p}(x_1^2\cdots x_p^2)^{-2}|\int_{[1,2]^p}\frac{\partial^{2p}\phi^\delta(\mathbf{t})}{\partial t_1^2\cdots\partial t_p^2}d\mathbf{t}|^2 g_0^{-1}(\delta\mathbf{x})d\mathbf{x} \\
= & O\{\delta^{2\beta}\int_{R^p\setminus[-1,1]^p}(2\pi)^{-2p}(x_1^2\cdots x_p^2)^{-2}g_0^{-1}(\delta\mathbf{x})d\mathbf{x}\} \\
= & O(\delta^{2\beta}).
\end{aligned}
$$

Consequently, the χ^2 distance in (15) is of order

$$\delta^{2k+p}(TERM_1+TERM_2)=O(n^{-1})$$

if we take $\delta = n^{-1/(2(m+\alpha)+p+2\beta)}$. In this case a lower bound will be

$$O(n^{-2(m+\alpha)/(2(m+\alpha)+2\beta+p)}).$$

This proves Theorem 8. □

(14) is not an unnatural requirement. Because of (13), we expect that for $\mathbf{t}$ far away from the origin, $\phi_\varepsilon(\mathbf{t}) \approx D/(t_1^{l_1} \cdots t_p^{l_p})$, and taking a partial derivative of $\phi_\varepsilon(\mathbf{t})$ will decrease the power of the corresponding t_i by 1. Hence (14) holds. The following example illustrates this.

EXAMPLE. Let $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_p)'$ where $\varepsilon_1, \cdots, \varepsilon_p$ are i.i.d. exponential random variables with mean λ. Thus

$$\phi_\varepsilon(\mathbf{t}) = \frac{1}{(1 - i\lambda t_1) \cdots (1 - i\lambda t_p)}.$$

In this instance we have $\phi_\varepsilon^{\mathbf{a}}(s) \approx s^{-p}$. Hence $\beta = p$,

$$\begin{aligned} \frac{\partial \phi_\varepsilon(\mathbf{t})}{\partial t_1} &= \frac{-i}{(1 - i\lambda t_1)^2 \cdots (1 - i\lambda t_p)}, \\ \frac{\partial^2 \phi_\varepsilon(\mathbf{t})}{\partial t_1 \partial t_2} &= \frac{-i}{(1 - i\lambda t_1)^2 (1 - i\lambda t_2)^2 \cdots (1 - i\lambda t_p)}, \end{aligned}$$

etc., and we conclude that (14) holds.

5 Acknowledgments

This paper forms a part of my Ph.D. dissertation at Purdue University. I would like to thank Professor Wei-Liem Loh for introducing this problem to me and for his patient reading of the manuscript.

References

[1] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order convergence estimates. *Sankhyā Ser. A* **50** 381-393.

[2] CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184-1186.

[3] DEVROYE, P. L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235-239.

[4] DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence, I. Tech. Report. 137a, Dept. Statist., Univ. California, Berkeley.

[5] DONOHO, D. L. and LIU, R. C. (1991a). Geometrizing rates of convergence, II. *Ann. Statist.* **19** 633-667.

[6] DONOHO, D. L. and LIU, R. C. (1991b). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668-701.

[7] FAN, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1255-1272.

[8] FAN, J. (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** 541-551.

[9] FAN, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21** 600-610.

[10] FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London.

[11] FARRELL, R. H. (1972). On the best obtainable asymptotic rates of estimators of a density function at a point. *Ann. Math. Statist.* **43** 170–180.

[12] JOHNSON, N. L. and KOTZ, S. (1972). *Distribution in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

[13] KOTZ, S. (1975). Multivariate distributions at a cross-road. In *Statistical Distributions in Scientific Work*, **1**. (eds. G. P. Patil, S. Kotz and J. K. Ord). D. Reidel Pub. Co., Boston.

[14] LUKCAS, E. (1970). *Characteristic Functions.* Hafner Pub. Co., New York.

[15] MENDELSOHN, J. and Rice, J. (1982). Deconvolution of microfluorometric histograms with B splines. *J. Amer. Statist. Assoc.* **77** 748-753.

[16] STEFANSKI, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229-235.

[17] STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimates. *Statist.* **21** 169-184.

[18] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators *Ann. Statist.* **8** 1348-1360.

[19] ZHANG, C. H. (1990). Fourier methods for estimation mixing densities and distributions. *Ann. Statist.* **18** 806-830.

Department of Statistics
Mathematical Sciences Building
Purdue University
West Lafayette, IN 47906 – 1399