Statistica Sinica 2(1992), 553-575

ON DECONVOLUTION USING TIME OF FLIGHT INFORMATION IN POSITRON EMISSION TOMOGRAPHY

Hang Paul Zhang

Stanford University

Abstract: We study an estimation problem in PET when the *time-of-flight* information is available. The continuous idealisation of the PET reconstruction problem, formulated by Johnstone and Silverman (1990) as a special case of bivariate density estimation based on indirect observations, is used. A *Modified Deconvoluting Kernel Density Estimator* (MDK) is proposed. For densities with mth derivatives satisfying α *Lipschitz* condition in L_2 norm and in L_∞ norm, the convergence rates of *mean integrated square error* and *maximum mean square error* are shown to be $O(n^{-\frac{2(m+\alpha)}{2(m+\alpha)+3}})$ where n is the number of counts. These rates are optimal. By comparing our results with those in the literature where *no* time-of-flight is considered, it is shown that although the time of flight does not yield better convergence rates in this model, it can yield better constants when the noise is small.

key words and phrases: Density estimation, deconvoluting kernel estimator, minimax, Radon transform, tomography.

1. Introduction

Positron emission tomography (PET) is a medical technology used to reconstruct the internal structure of an organ of interest by detecting the particles emitted from injected radioactive material. Since it poses interesting statistical reconstruction problems involving incomplete data, there have been several papers recently on this subject in the statistical literature. See Johnstone and Silverman (1990), Shepp and Vardi (1982), Vardi, Shepp and Kaufman (1985) for references.

The formulation of the PET problem we shall consider is basically the idealised version of that given by Johnstone and Silverman (1990).

1.1. The basic setup of PET

As described in Johnstone and Silverman, we consider a particular PET experiment, where the patient is injected with a quantity of radioactively tagged glucose or other metabolite. Emissions are recorded in one or more rings of detectors put around the patient's head. The ring of detectors defines a slice of the head: we shall regard the slice as a plane or cross-section and consider an essentially two dimensional problem in which (see Figure 1) emissions take place in the plane according to the density of glucose within the slice. We take the detector ring be the unit circle. Thus we are estimating a two dimensional density supported within the unit disc. An emission at P gives rise to a photon pair whose directions of flight lie in the plane along a line l through P with random, uniformly distributed orientation. We assume that the points B and C of the intersection of l with the detector circle are observed exactly.

For the PET problem formulated in this way, Johnstone and Silverman established the exact minimax rate of convergence of estimation, for all possible estimators, over suitable smoothness classes of functions. For densities in a class corresponding to bounded square integrable pth derivatives relative to a suitable weight function, the rate is $n^{-\frac{p}{p+2}}$.

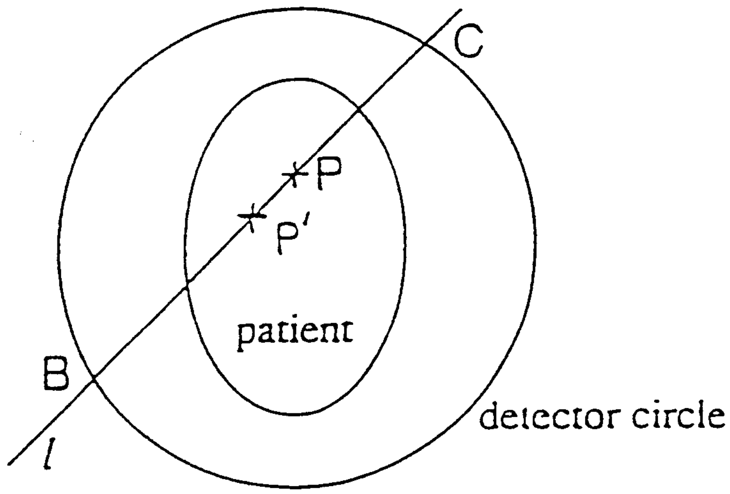

Figure 1. The patient and the detector circle

1.2. New problem and model

Technological advances now present the option of using photo-electric devices that are so sensitive that they can record the different times of incidence of the two photons that come from the original positron-electron annihilation. The difference in time allows the approximate position along the line at which the annihilation occured to be determined (by simple distance-time calculations, since the speed of photons is the same as that of light). For more information, see Snyder et al. (1981). Thus, we can detect not only the line where the emission must have lain, but also the approximate position on the line. See Figure 1, where P' is the approximate position of P we detected (cf. Snyder et al.). There are various sources of error involved, however, so the position can only be found approximately.

Given this extra information, is better reconstruction possible? We use an asymptotic approach to show that the answer is positive.

Model:

$$Y = X + \epsilon\vec{a}, \quad X, Y \in R^2 \tag{1.1}$$

Here Y is the approximate emission position (P') detected, X is the original position (P) of the emission. We assume:

1. X has density g, with support on the *unit disc*.
2. $\epsilon \sim N(0, \sigma^2), \sigma^2$ is known and small.
3. $\vec{a} = (\cos\theta, \sin\theta)$ where θ is uniformly distributed on $[0, 2\pi]$.
4. $X, \epsilon, \vec{a}$ are independent.

Therefore, the joint density of Y and $\vec{a}$ relative to *Lebesgue* measure is:

$$L \circ g(y, \vec{a}) = f(y, \vec{a}) = \frac{1}{(2\pi)^{3/2}\sigma}\int_{-\infty}^{\infty} g(y - t\vec{a}) \exp\left(-\frac{t^2}{2\sigma^2}\right) dt. \tag{1.2}$$

Denote

$$T \circ f = g$$

Problem: Determine g, the density of X, when we can only observe Y and $\vec{a}$.

This paper is organized as follows: Section 2 proposes an estimator called the *Modified Deconvoluting Kernel Density Estimator.* Section 3 establishes some asymptotic properties of the estimator and shows that the estimator is optimal in the sense of achieving the optimal convergence rates for MISE and MMSE. Section 4 compares our results with other recent work and shows that in some cases even though time-of-flight does not give us better convergence rate, it does give us better (smaller) coefficients in the leading error term of the error expansion. In Section 5, the theorems in Sections 3 and 4 are proved.

2. Estimator

Apparently our problem is just a particular kind of *deconvolution problem.* We will therefore, begin with some generalities on the deconvolution problem, and with a particular *deconvoluting kernel density estimator* introduced by Stefanski and Carroll (1990). Their estimator is modified for our problem in which use is made of the extra time of flight information.

2.1. Deconvolution problem

Let U and Z be independent random variables with probability density functions g and h respectively. Then the random variable $X = U + Z$ has the density

$f = g \star h$ where "$\star$" denotes convolution. Assuming h is known, we consider estimation of g from a set of independent observations $\{x_1, \ldots, x_n\}$ having the common density f.

2.2. Deconvoluting kernel density estimator

Let K be a kernel function and $\hat{f}$ be the ordinary kernel density estimator of f based on the kernel K, i.e.,

$$\hat{f} = (n\lambda)^{-1} \sum_{j=1}^{n} K((x_j - x)/\lambda). \tag{2.1}$$

Denote by $\phi_{\hat{f}}, \phi_h, \phi_g, \phi_K$ the *Fourier transformations* of $\hat{f}, h, g, K$, respectively.

Let $\hat{\phi}_n$ be the empirical characteristic function of $x_1, \ldots, x_n$, that is

$$\hat{\phi}_n(t) = \frac{1}{n} \sum_{i=1}^{n} e^{itx_i}.$$

Now we have

$$\phi_{\hat{f}} = \phi_K \hat{\phi}_n.$$

Assume that

$$|\phi_h(t)| > 0, \ \ \forall\, t \in R \tag{2.2}$$

and

$$\sup_t |\phi_K(t)/\phi_h(t/\lambda)| < \infty; \ \int |\phi_K(t)/\phi_h(t/\lambda)| dt < \infty, \ \ \forall \lambda > 0. \tag{2.3}$$

The *deconvoluting kernel density estimator is then defined as*

$$\begin{aligned} \hat{g}(t) &= (2\pi)^{-1} \int e^{-itx} \phi_{\hat{f}}(t)/\phi_h(t) dt \\ &= (2\pi)^{-1} \int e^{-itx} \phi_K(\lambda t) \hat{\phi}_n(t)/\phi_h(t) dt. \end{aligned} \tag{2.4}$$

It is straightforward to extend the estimator to the multivariate case.

Like the ordinary kernel density estimator, the *deconvoluting kernel density estimator* has some advantages. It is computationally feasible, has good asymptotic properties, and, in a broad range of error distributions, is optimal, in the sense of rates of convergence. See Carroll and Hall (1988), Stefanski and Carroll (1990), Stefanski (1990), Fan (1991).

Although our problem is a deconvolution problem, the estimator (2.4) should not be used directly, because we know not only the error density, but also the direction along which the error occurs. In fact, it can be shown that estimator

(2.4) (by treating $\epsilon\vec{a}$ as error) does not give the best convergence rate. The question now is how to use the extra information.

2.3. Modified Deconvoluting Kernel density estimator (MDK)

Intuitively, the *Deconvoluting Kernel Estimator* of Stefanski and Carroll simply divides the empirical characteristic function of f by that of h, and regarding that as an estimator of the characteristic function of g, it then applies the *inverse Fourier transformation.* Our strategy is to get a better estimator of the characteristic function of g with the extra information.

Let the sample be $(\boldsymbol{y}_1, \boldsymbol{x}_1, \epsilon_1, \vec{\boldsymbol{a}}_1), \ldots, (\boldsymbol{y}_n, \boldsymbol{x}_n, \epsilon_n, \vec{\boldsymbol{a}}_n)$, of which we only observe $\boldsymbol{y}_j$'s and $\vec{\boldsymbol{a}}_j$'s.

Our idea is based on the following simple observations:

- If $\vec{\boldsymbol{a}}_j \perp \boldsymbol{t}$, then

$$\boldsymbol{t}^\tau \boldsymbol{y}_j = \boldsymbol{t}^\tau \boldsymbol{x}_j + \epsilon_j \boldsymbol{t}^\tau \vec{\boldsymbol{a}}_j = \boldsymbol{t}^\tau \boldsymbol{x}_j \Longrightarrow \exp(i\boldsymbol{t}^\tau \boldsymbol{y}_j) = \exp(i\boldsymbol{t}^\tau \boldsymbol{x}_j).$$

  So, in this case $\boldsymbol{y}_j$ gives as much information about $\phi_g(\boldsymbol{t})$ as $\boldsymbol{x}_j$, as if there were no noise!

- If $\vec{\boldsymbol{a}}_j \; // \; \boldsymbol{t}$, then

$$\exp(i\boldsymbol{t}^\tau \boldsymbol{y}_j) = \exp(i\boldsymbol{t}^\tau \boldsymbol{x}_j \pm |\epsilon_j|),$$

  and $\boldsymbol{y}_j$ gives the least accurate information about $\boldsymbol{t}^\tau \boldsymbol{x}_j$ and hence $\phi_g(\boldsymbol{t})$.

This suggests weighting $\boldsymbol{y}_j$: giving larger weight to those y_j, such that $< \vec{\boldsymbol{a}}_j, \boldsymbol{t} >$ is about $\pi/2$, and smaller weight to those $\boldsymbol{y}_j$, such that $< \vec{\boldsymbol{a}}_j, \boldsymbol{t} >$ is about 0, or π. Here, $< \vec{\boldsymbol{a}}_j, \boldsymbol{t} >$ is the angle between $\vec{\boldsymbol{a}}_j$ and $\boldsymbol{t}$, and $0 \leq < \vec{\boldsymbol{a}}_j, \boldsymbol{t} > \leq \pi$. That suggests the *Weighted Empirical Characteristic Function*,

$$\hat{\phi}_n^w(\boldsymbol{t}) = \frac{1}{n}\sum_{j=1}^{n} w(\vec{\boldsymbol{a}}_j, \boldsymbol{t}) \exp(i\boldsymbol{t}^\tau \boldsymbol{y}_j)$$

as an estimator of $\phi_g(\boldsymbol{t})$. Here $w(\boldsymbol{t}, \vec{\boldsymbol{a}})$ is assumed to be a positive bounded weight function, such that w is nonincreasing with $|\boldsymbol{t}^\tau \vec{\boldsymbol{a}}|$. To make it unbiased, divide it by $\phi_\epsilon^w(\boldsymbol{t})$ such that

$$E(\phi_n^w(\boldsymbol{t})/\phi_\epsilon^w(\boldsymbol{t})) = \phi_g(\boldsymbol{t}).$$

Thus

$$\phi_\epsilon^w(\boldsymbol{t}) = E(\exp(i\epsilon \boldsymbol{t}^\tau \vec{\boldsymbol{a}}) w(\vec{\boldsymbol{a}}, \boldsymbol{t})). \tag{2.5}$$

Finally, *Modified Deconvoluting Kernel Density Estimator* (MDK) is defined as

$$\hat{g}_{n,h}^w(\boldsymbol{x}) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-i\boldsymbol{t}^\tau \boldsymbol{x})\phi_K(\boldsymbol{t}h)\frac{\hat{\phi}_n^w(\boldsymbol{t})}{\phi_\epsilon^w(\boldsymbol{t})} dt_1 dt_2 \tag{2.6}$$

where K is a two-dimensional kernel.

This estimator can be expressed as a sum of *i.i.d.* random variables:

$$\hat{g}^w_{n,h}(\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-i\boldsymbol{t}^\tau\boldsymbol{x})\phi_K(\boldsymbol{t}h)\frac{w(\vec{\boldsymbol{a}}_j,\boldsymbol{t})\exp(i\boldsymbol{t}^\tau\boldsymbol{y}_j)}{\phi^w_\epsilon(\boldsymbol{t})}dt_1dt_2. \quad (2.7)$$

Selection of w. From now on, use

$$\begin{aligned} w_0(\vec{\boldsymbol{a}},\boldsymbol{t}) &= E(\exp(i\epsilon\boldsymbol{t}^\tau\vec{\boldsymbol{a}})) \\ &= \exp\left(-\frac{(\boldsymbol{t}^\tau\vec{\boldsymbol{a}})^2\sigma^2}{2}\right) \end{aligned}$$

as weight function. (2.5) can be rewritten as

$$\phi^w_\epsilon(\boldsymbol{t}) = E(w_0(\vec{\boldsymbol{a}},\boldsymbol{t})w(\vec{\boldsymbol{a}},\boldsymbol{t})). \quad (2.8)$$

We shall show that w_0 is optimal in the sense of minimizing the global variance of the estimator.

Remark. In model (1.1), the normality assumption on ϵ is not crucial. For any symmetric error ϵ, let ϕ denote its characteristic function. In estimator (2.6), choosing

$$w_0(\vec{\boldsymbol{a}},\boldsymbol{t}) = \phi(\boldsymbol{t}^\tau\vec{\boldsymbol{a}}),$$

we can do the same analysis and get the same results.

For simplicity, use $\hat{g}_{n,h}$ for $\hat{g}^{w_0}_{n,h}$, ϕ_ϵ for $\phi^{w_0}_\epsilon$, $\hat{\phi}_n$ for $\hat{\phi}^{w_0}_n$, and when there is no confusion about h, use $\hat{g}_n$ for $\hat{g}_{n,h}$.

3. Asymptotic Properties of the Estimator

The *Mean Integrated Square Error* of $\hat{g}_n$ is defined as

$$\text{MISE}_n(h) = E\int(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x}))^2 d\boldsymbol{x}.$$

For a fixed point $\boldsymbol{x}_0$, the *Mean Square Error* is defined as

$$\text{MSE}_n(h) = E\left|\hat{g}_n(\boldsymbol{x}_0) - g(\boldsymbol{x}_0)\right|^2$$

and for a class C of densities, the *Maximum Mean Square Error* is defined as

$$\text{MMSE}^C_n(h) = \sup_{g\in C}\text{MSE}_n(h).$$

When C is clear, we simply use $\text{MMSE}_n(h)$ for $\text{MMSE}^C_n(h)$.

In this section, the properties of the sequences $\{\mathrm{MISE}_n(h_n)\}$ and $\{\mathrm{MMSE}_n(h_n)\}$ are studied. In particular, if g and its kth derivatives satisfy some smoothness conditions, then using a proper kernel and choosing h_n to be $n^{-1/(2k+3)}$, we shall show that $\mathrm{MISE}_n(h_n)$ and $\mathrm{MMSE}_n(h_n)$ are of the order $n^{-\frac{k}{2k+3}}$.

3.1. Mean integrated square error

We say that $g \in C_{m,\alpha,B}$ if

1. g is a density function on $\{\boldsymbol{x} : \boldsymbol{x} \in R^2, \|\boldsymbol{x}\| < 1\}$,
2. g has up to mth order derivatives,
3. $\forall\, 0 \le i \le m$, $\forall\, \Delta \in R^2$,

$$\left\| \frac{\partial^m g}{\partial^i x_1 \partial^{m-i} x_2}(\boldsymbol{x}) - \frac{\partial^m g}{\partial^i x_1 \partial^{m-i} x_2}(\boldsymbol{x}+\Delta) \right\|_2 \le B\,\|\Delta\|^\alpha.$$

Here, m is a positive integer, $0 \le \alpha < 1$, $B > 0$, and $\|\ \|_2$ is the L_2 norm.

Theorem 1. *Let $p = m + \alpha$ and $h_n = n^{-\frac{1}{2p+3}}$. Then, as $n \to \infty$,*

$$\sup_{g \in C_{m,\alpha,B}} E\|\hat{g}_n - g\|_2^2 = O\left(n^{-\frac{2p}{2p+3}}\right). \tag{3.1}$$

Theorem 2.

$$\lim_{n\to\infty} \left(n^{\frac{2p}{2p+3}}\right) \inf_{\hat{T}_n} \sup_{g \in C_{m,\alpha,B}} E\|\hat{T}_n - g\|_2^2 > 0, \tag{3.2}$$

where the infimum is over all estimators $\hat{T}_n$.

3.2. Maximum mean square error

We say that $g \in C'_{m,\alpha,B}$ if

1. g is a density function on $\{\boldsymbol{x} : \boldsymbol{x} \in R^2, \|\boldsymbol{x}\| < 1\}$,
2. g has mth order derivatives, and $|g| \le B$,
3. $\forall\, 0 \le i \le m$, $\forall\, \Delta \in R^2$,

$$\left| \frac{\partial^m g}{\partial^i x_1 \partial^{m-i} x_2}(\boldsymbol{x}) - \frac{\partial^m g}{\partial^i x_1 \partial^{m-i} x_2}(\boldsymbol{x}+\Delta) \right| \le B\|\Delta\|^\alpha.$$

Here, m is a positive integer, $0 \le \alpha < 1$, $B > \frac{1}{\pi}$.

Theorem 3. *Let $p = m + \alpha$ and $h_n = n^{-\frac{1}{2p+3}}$. Then*

$$\sup_{\|\boldsymbol{x}_0\| \le 1} \sup_{g \in C'_{m,\alpha,B}} E\,|\hat{g}_n(\boldsymbol{x}_0) - g(\boldsymbol{x}_0)|^2 = O(n^{-\frac{2p}{2p+3}}). \tag{3.3}$$

And

Theorem 4. *For every $\boldsymbol{x}_0$ inside the unit disc,*

$$\left(n^{\frac{2p}{2p+3}}\right) \inf_{\hat{T}_n} \sup_{g\in C'_{m,\alpha,B}} E\left|\hat{g}_n(\boldsymbol{x}_0) - g(\boldsymbol{x}_0)\right|^2 > 0. \tag{3.4}$$

By the theorems, we know that the MDK is optimal in the sense of *Stone* (1982).

Remark. For our problem, estimator (2.4) has convergence rate $n^{\frac{2p}{2p+4}}$ for both cases. Also note that $n^{\frac{2p}{2p+3}}$ is slower than $n^{\frac{2p}{2p+2}}$, the convergence rate when we have direct data (see *Stone*).

4. How Is Time-of-Flight Helpful?

In this section, we try to compare our results in *Section 3* with those in the literature in order to understand the effect of *time-of-flight* information.

4.1. Johnstone and Silverman

The first effort along this direction of research was made by Johnstone and Silverman (1990). They studied the problem when there is no time-of-flight available. They put a different kind of smoothness condition on their underlying function class. We say a function $g \in B_{p,M}$, if g has pth order weak derivatives which are square-integrable with respect to $d\mu_{p+1}(x) \equiv (p+1)(1-\|x\|^2)^{p+1}dx$, and

$$\int_D \left(g^{(p)}(x)\right)^2 d\mu_{p+1}(x) \le M.$$

Here D is the *unit disc.*

Their main result is that for $B_{p,M}$, the minimax MISE convergence rates are:

(i) $(n/\log n)^{-\frac{2p}{2p+2}}$ for direct case, that is, when we can observe the emission location exactly.

(ii) $n^{-\frac{2p}{2p+4}}$ for the indirect case, that is, when we can observe only the line of the coincidence pairs.

Note that

$$C_{m,\alpha,M} \subset B_{p,M}$$

and since $(p+1)(1-\|x\|^2)^{p+1}$ vanishes near the edge of the unit disc, $C_{m,\alpha,M}$ is essentially a smoother class. But how much broader is $B_{p,M}$? By comparing the direct case with the results of Stone (1982), it seems that as far as convergence rates are concerned, the two function classes differ only by a factor of $\log n$. Therefore, we previously thought that *time-of-flight* might give us a better convergence rate until the recent work of Bickel and Ritov (1990) became known.

4.2. Bickel and Ritov

In the case of no *time-of-flight*, Bickel and Ritov studied the problem of estimating the linear functionals of densities on the unit disc. They obtained an estimator of the density at a point by estimating a series of linear functionals which converge, as $n \to \infty$, to the value of the density at the point. More explicitly, instead of estimating g, estimate

$$g_h = g \star K_h$$

where $K_h(\boldsymbol{x}) = K(\boldsymbol{x}/h)/h^2$, K is a kernel. By *Parseval's* theorem

$$g_h(\boldsymbol{x}) = \int_0^{\pi} \int_{-\infty}^{\infty} \Phi_h(\psi, t; \boldsymbol{x})\lambda(\psi, t)d\psi dt. \tag{4.1}$$

Here, λ is the *Radon transform* of g (see Figure 2, in which $\lambda(\phi, s)$ equals the integral of g along the line BC divided by π) and

$$\Phi_h(\psi, t; \boldsymbol{x}) = \rho_h(x_1 \cos\psi + x_2 \sin\psi - t, \psi),$$

where

$$\begin{aligned} \rho_h(u, \psi) &\stackrel{\text{def}}{=} \frac{1}{4\pi}\int e^{iuw}|w|\,\tilde{K}(\psi, wh)dw \\ \tilde{K}(\psi, w) &\stackrel{\text{def}}{=} \int e^{iw(x_1\cos\psi + x_2\sin\psi)}K(\boldsymbol{x})d\boldsymbol{x}. \end{aligned} \tag{4.2}$$

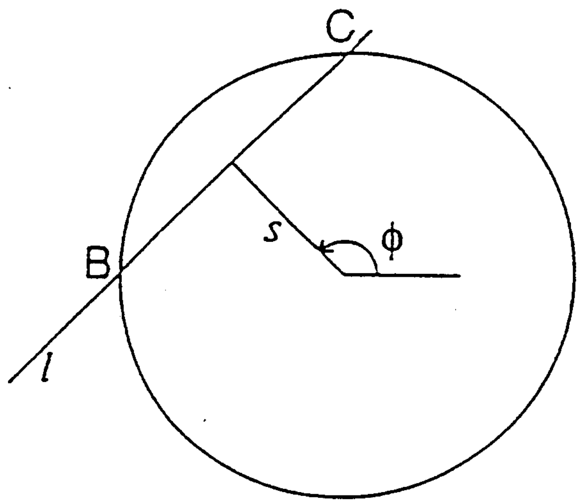

Figure 2. Radon transformation

Note that

$$g_h(\boldsymbol{x}) = E(\Phi_h(\psi, t; \boldsymbol{x})) \tag{4.3}$$

by (4.1) where the expectation is taken with respect to (ψ, t) according to λ.

Suppose we have the sample $(\psi_1, t_1), \ldots, (\psi_n, t_n)$. Then an unbiased estimator of $g_h(\boldsymbol{x})$ is

$$\hat{g}_h(\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{n} \Phi_h(\psi_j, t_j; \boldsymbol{x}).$$

By letting $h \to 0$ as $n \to \infty$, $g_h \to g$, Bickel and Ritov proved that the MMSE of $\hat{g}_h$ is of the order $n^{-\frac{2p}{2p+3}}$ for $C'_{m,\alpha,M}$ and MISE is of the order $n^{-\frac{2p}{2p+3}}$ for $C_{m,\alpha,M}$.

By the results in *Section 3*, we know that Bickel and Ritov's estimator is optimal. Thus time-of-flight is not helpful in getting better *convergence rate*. One might ask if we can get better *constants* by using time-of-flight. We show next that it is true if σ, presenting the noise, is small.

4.3. Constant calculation

We shall use the same *kernel* K for both estimators. For technical reasons, we assume that K is radially symmetric. It is not an unreasonable assumption, since it is well known that the best kernel in some cases is radially symmetric, like the *Epanechnikov kernel* (cf. Silverman (1986)). Then we have

Theorem 5. *For large* n,

$$\frac{\min_h \mathrm{MISE}_h(\mathrm{MDK})}{\min_h \mathrm{MISE}_h(\mathrm{BR})} \le \left[\frac{\sigma\pi^{3/2}}{2}\right]^{\frac{2p}{2p+3}} \tag{4.4}$$

where equality holds iff g puts all its mass on the edge of the unit disc.

Therefore, we know that when σ is small, we can estimate the true density (image) better.

In Theorem 5, fix σ, and now let σ vary to get different results in the following.

Theorem 6. *Let* $\sigma = n^{-\beta}$. *Then for the* MDK, *we have*

(i) *if* $\beta \le \frac{1}{2p+2}$,

$$\sup_{g\in C_{m,\alpha,M}} E\|\hat{g}_n - g\|_2^2 = O\left(n^{-(1+\beta)\frac{2p}{2p+3}}\right),$$

$$\sup_{\|\boldsymbol{x}_0\|\le 1}\ \sup_{g\in C'_{m,\alpha,B}} E\left|\hat{g}_n(\boldsymbol{x}_0) - g(\boldsymbol{x}_0)\right|^2 = O\left(n^{-(1+\beta)\frac{2p}{2p+3}}\right);$$

(ii) *if* $\beta \ge \frac{1}{2p+2}$,

$$\sup_{g\in C_{m,\alpha,M}} E\|\hat{g}_n - g\|_2^2 = O\left(n^{-\frac{2p}{2p+2}}\right),$$

$$\sup_{\|x_0\|\le 1} \sup_{g\in C'_{m,\alpha,B}} E\left|\hat{g}_n(x_0) - g(x_0)\right|^2 = O\left(n^{-\frac{2p}{2p+2}}\right).$$

4.4. Some remarks

1. Snyder et al. (1981) took account of the physical size of the detectors. They modeled this factor as a zero-mean normal random noise in the transverse direction. (See Figure 3, in which an annihilation occuring at x is measured as the point (u, θ), where $u = x + \epsilon(\theta)$ and $\epsilon(\theta)$ is a two-dimensional measurement-error vector with a component ϵ_e parallel to the center line of the detector pair and a component ϵ_b transverse to this line.) Assume that the measurement error ϵ associated with an annihilation at x is independent of x as well as the locations and measurement errors of all other annihilations.

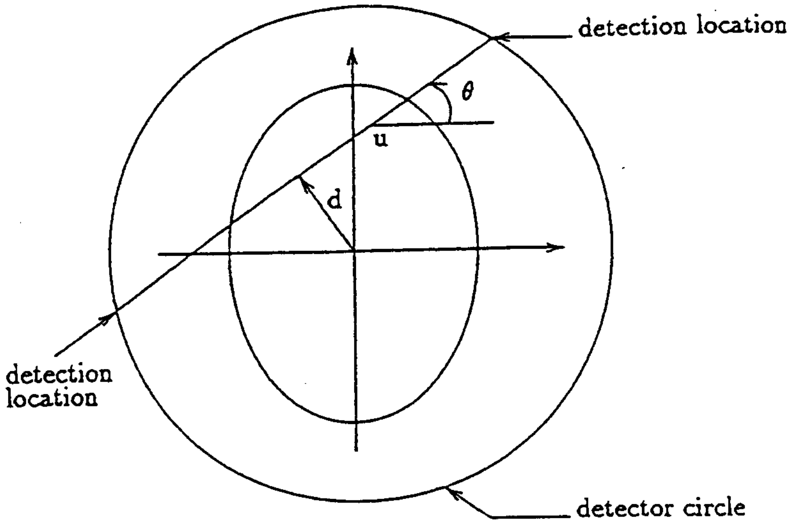

Figure 3. Each measurement is a point with coordinates (u, θ).

Then

$$\epsilon(\theta) \sim T(\theta)(\epsilon_e, \epsilon_b)^\tau \tag{4.5}$$

where

$$\begin{pmatrix} \epsilon_e \\ \epsilon_b \end{pmatrix} \sim N\begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix},$$

$$T(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

When $\sigma_b \neq 0$ (usually $\frac{\sigma_e}{\sigma_b} \approx 3$), and because the directional information of θ does not help us much, we can simply use the *deconvoluting estimator*, and show that the MISE (MMSE) is $(\log n)^{-(p+2)}$ for $C_{m,\alpha,B}(C'_{m,\alpha,B})$. Fan (1991) proved that it is the optimal rate for both MISE and MMSE. In this case, the best convergence rate is very slow. For the case in which only the approximate lines where the emissions lie can be detected, we do not know how much time of flight information helps us.

2. O'Sullivan (1989) studied the model proposed by Snyder et al., and in his practical implementation, the target is $g \star w_b$ instead of g itself. Here w_b is a two-dimensional spherically symmetric Gaussian density with mean zero and standard deviation σ_b. Therefore, he actually used model (1.1), except that $g \star w_b$ has derivatives of any order. Moreover his *Confidence-Weighted Backprojection* estimator is exactly the same as our estimator (2.6).
3. It would be interesting to know other applications of the model (1.1).

5. Proofs of the Results

5.1. Convergence rate for MDK

Proof of Theorem 1

First we prove a lemma.

Lemma 5.1.

$$\phi_\epsilon(t_1,t_2) = \frac{1}{2\pi}\int_0^{2\pi} \exp(-\sigma^2(t_1^2+t_2^2)\sin^2\theta)d\theta.$$

Proof. By the definition, since X and ϵ are independent,

$$\begin{aligned}
\phi_\epsilon(t) &= E\left(\exp(-it^\tau Y)\exp\left(-\frac{\sigma^2}{2}(t^\tau\vec{a})^2\right)/\phi_g(t)\right)\\
&= E\left(\exp(-it^\tau(X+\epsilon\vec{a})\exp\left(-\frac{\sigma^2}{2}(t^\tau\vec{a})^2\right)/\phi_g(t)\right)\\
&= \left(\frac{E\exp(-it^\tau X)}{\phi_g(t)}\right)E\left(\exp(-it^\tau\epsilon\vec{a})\exp\left(-\frac{\sigma^2}{2}(t^\tau\vec{a})^2\right)\right)\\
&= E(\exp(-\sigma^2(t^\tau\vec{a})^2))\\
&= \frac{1}{2\pi}\int_0^{2\pi}\exp(-\sigma^2(t_1^2+t_2^2)\sin^2\theta)d\theta.
\end{aligned}$$

Since $\frac{|\sin\theta|}{|\theta-k\pi|}\to 1$ as $\theta\to k\pi$, it is easy to show that

$$\begin{aligned}
&\sqrt{t_1^2+t_2^2}\int_0^{2\pi}\exp(-\sigma^2(t_1^2+t_2^2)\sin^2\theta)d\theta\\
\approx\ & 2\sqrt{t_1^2+t_2^2}\int_{-\infty}^{\infty}\exp(-\sigma^2(t_1^2+t_2^2)\theta^2)d\theta\\
&\text{as } \sqrt{t_1^2+t_2^2}\to\infty.
\end{aligned}$$

Therefore, we have

Corollary 5.1.

$$\left(\sigma\sqrt{\pi}\sqrt{t_1^2+t_2^2}\right)\phi_\epsilon(t)\to 1, \quad \text{as } \sqrt{t_1^2+t_2^2}\to\infty. \tag{5.1}$$

Next, choose K to satisfy the following conditions:

(a) K has finite support and is bounded.

(b) $\int K(t)dt = 1$.

(c) $\int\int t_1^i t_2^j K(t)dt = 0$, $\forall\ 1\le i+j\le m$.

By some standard methods (see Prakasa Rao (1983)), we have $\forall\ g\in C_{m,\alpha,B}$,

$$E(\hat{g}_n^w(\boldsymbol{x})) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\boldsymbol{x}-h\boldsymbol{y})K(\boldsymbol{y})dy_1dy_2, \tag{5.2}$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|E(\hat{g}_n^w(\boldsymbol{x}))-g(\boldsymbol{x})|^2dx_1dx_2 = C_K(g)(h^{2p})(1+o(1)). \tag{5.3}$$

Here, $C_K(g)$ is a function of g depending on K and satisfies

$$\sup_{g\in C_{m,\alpha,B}} C_K(g)<\infty.$$

Note that by (5.2), $E(\hat{g}_n^w(\boldsymbol{x}))$ does not depend on w. Denote

$$C_n(g) = \frac{1}{(2\pi)^4 n}\int E^2(\hat{g}_n^w(\boldsymbol{x}))d\boldsymbol{x}.$$

By (2.7),

$$\begin{aligned}
&\int \mathrm{var}(\hat{g}_n^w(\boldsymbol{x}))d\boldsymbol{x}\\
=\ & \frac{1}{(2\pi)^4 n}\int \mathrm{var}\left(\int e^{-i\boldsymbol{t}^\tau\boldsymbol{x}}\frac{\phi_K(\boldsymbol{t}h)}{\phi_\epsilon^w(\boldsymbol{t})}e^{i\boldsymbol{t}^\tau\boldsymbol{Y}}w(\vec{\boldsymbol{a}},\boldsymbol{t})d\boldsymbol{t}\right)d\boldsymbol{x}\\
=\ & \frac{1}{(2\pi)^4 n}\int\left[E\left|\int e^{-i\boldsymbol{t}^\tau\boldsymbol{x}}\frac{\phi_K(\boldsymbol{t}h)}{\phi_\epsilon^w(\boldsymbol{t})}e^{i\boldsymbol{t}^\tau\boldsymbol{Y}}w(\vec{\boldsymbol{a}},\boldsymbol{t})d\boldsymbol{t}\right|^2 - E^2(\hat{g}_n^w(\boldsymbol{x}))\right]d\boldsymbol{x}\\
=\ & \frac{1}{(2\pi)^4 n}E\left(\int\left|\int e^{-i\boldsymbol{t}^\tau\boldsymbol{x}}\frac{\phi_K(\boldsymbol{t}h)}{\phi_\epsilon^w(\boldsymbol{t})}e^{i\boldsymbol{t}^\tau\boldsymbol{Y}}w(\vec{\boldsymbol{a}},\boldsymbol{t})d\boldsymbol{t}\right|^2d\boldsymbol{x}\right) - C_n\\
=\ & \frac{1}{(2\pi)^4 n}E\left((2\pi)^2\int\left|\frac{\phi_K(\boldsymbol{t}h)}{\phi_\epsilon^w(\boldsymbol{t})}e^{i\boldsymbol{t}^\tau\boldsymbol{Y}}w(\vec{\boldsymbol{a}},\boldsymbol{t})\right|^2d\boldsymbol{t}\right) - C_n \qquad (5.4)\\
=\ & \frac{1}{(2\pi)^2 n}\int\frac{|\phi_K(\boldsymbol{t}h)|^2}{|\phi_\epsilon^w(\boldsymbol{t})|^2}E(w^2(\vec{\boldsymbol{a}},\boldsymbol{t}))d\boldsymbol{t} - C_n\\
\ge\ & \frac{1}{(2\pi)^2 n}\int\frac{|\phi_K(\boldsymbol{t}h)|^2}{|\phi_\epsilon(\boldsymbol{t})|}d\boldsymbol{t} - C_n, \qquad (5.5)
\end{aligned}$$

where (5.4) holds by the *Parseval Equality.* To prove (5.5), recall that

$$w_0(\vec{a}, t) = E\exp(i\epsilon t^\tau \vec{a})$$

and

$$E(w_0^2(\vec{a}, t)) = \phi_\epsilon(t).$$

Thus

$$\begin{aligned} |\phi_\epsilon^w|^2 &= |E(\exp(i\epsilon t^\tau \vec{a}) w(t, \vec{a}))|^2 \\ &= |E(w_0(t, \vec{a}) w(t, \vec{a}))|^2 \\ &\leq E(w_0^2(t, \vec{a})) E(w^2(t, \vec{a})) \\ &\quad \text{(by Cauchy's inequality)} \\ &= \phi_\epsilon(t) E^2(w(t, \vec{a})). \end{aligned}$$

The equality holds *iff* $w \propto w_0$, so in fact we have shown that w_0 is best in the sense that $\hat{g}_n$ has minimum global variability.

By *Corrollary 5.1*, we have

$$\int \frac{|\phi_K(th)|^2}{|\phi_\epsilon(t)|} dt = (1 + o(1)) \frac{\sigma\sqrt{\pi}}{h^3} \int |\phi_K(t)|^2 \|t\| \, dt. \tag{5.6}$$

Since

$$\sup_{g \in C_{m,\alpha,B}} C_n(g) = O\left(\frac{1}{n}\right),$$

it follows that, when h is small,

$$\sup_{g \in C_{m,\alpha,B}} \int \operatorname{var}(\hat{g}_n(x)) dx = (1 + o(1)) \frac{\sigma\sqrt{\pi}}{4\pi^2 n h^3} \int |\phi_K(t)|^2 \|t\| \, dt. \tag{5.7}$$

Therefore by (5.3) and (5.7),

$$\mathrm{MISE}_n(h) = O(h^{2p}) + O\left(\frac{1}{nh^3}\right). \tag{5.8}$$

Choosing $h_n = n^{-\frac{1}{2p+3}}$ yields the best convergence rate $n^{-\frac{2p}{2p+3}}$ for $\mathrm{MISE}_n(h)$.

Proof of Theorem 3. By the *Taylor expansion* (see Prakasa Rao (1983) for details), $\forall\, x_0$ in the *unit disc*,

$$|E(\hat{g}_n(x_0)) - g(x_0)| \leq Dh^p. \tag{5.9}$$

Let $f_{\vec{a}}$ be the density of $\boldsymbol{y}$ given $\vec{a}$. Since $\forall\, g \in C'_{m,\alpha,B}$, $|g| < B$, we have

$$\begin{aligned}\forall\, \boldsymbol{y},\ f_{\vec{a}}(\boldsymbol{y}) &= \int g(\boldsymbol{y} - \vec{a}\boldsymbol{t})\phi_\epsilon(\boldsymbol{t})dt \\ &< \int B\phi_\epsilon(\boldsymbol{t})dt = B,\end{aligned}$$

where ϕ_ϵ is the density of ϵ, it follows that

$$\begin{aligned}&\mathrm{var}(\hat{g}_n(\boldsymbol{x}_0)) \\ \leq\ & \frac{1}{(2\pi)^4 n} E\left|\int e^{-i\boldsymbol{t}^\tau \boldsymbol{x}_0}\frac{\phi_K(\boldsymbol{t}h)}{\phi_\epsilon(\boldsymbol{t})} e^{i\boldsymbol{t}^\tau \boldsymbol{Y}} w_0(\vec{a},\boldsymbol{t})dt\right|^2 \\ <\ & \frac{B}{(2\pi)^4 n} E\left(\int\left|\int e^{-i\boldsymbol{t}^\tau \boldsymbol{x}_0}\frac{\phi_K(\boldsymbol{t}h)}{\phi_\epsilon(\boldsymbol{t})} e^{i\boldsymbol{t}^\tau \boldsymbol{y}} w_0(\vec{a},\boldsymbol{t})dt\right|^2 dy\right) \\ =\ & \frac{B}{(2\pi)^4 n} E\left((2\pi)^2\int\left|\frac{\phi_K(\boldsymbol{t}h)}{\phi_\epsilon^w(\boldsymbol{t})} e^{-i\boldsymbol{t}^\tau \boldsymbol{x}_0} w(\vec{a},\boldsymbol{t})\right|^2 dt\right) \\ =\ & \frac{B}{(2\pi)^2 n}\int \frac{|\phi_K(\boldsymbol{t}h)|^2}{|\phi_\epsilon(\boldsymbol{t})|}dt. \qquad (5.10)\end{aligned}$$

So by the same argument as in proving *Theorem 1*,

$$\mathrm{MMSE}_n(h) = O(h^{2p}) + O\left(\frac{1}{nh^3}\right). \qquad (5.11)$$

Choosing $h_n = n^{-\frac{1}{2p+3}}$ yields the best convergence rate $n^{-\frac{2p}{2p+3}}$ for $\mathrm{MMSE}_n(h)$.

5.2. Lower bound on the convergence rate

Recently, there have been several results about the lower bounds in density estimation problems, (see Donoho and Liu (1991a,b)). For our problem, we are going to adopt the *cubical lower bound* idea of Fan (1992).

Let m be an even integer. Denote

$$\boldsymbol{t}_{i,j}^{(n)} = (-1/2, -1/2) + (i/m, j/m),\ \ i,j = 1,\ldots,m.$$

Let $\{\boldsymbol{x}_k^{(n)}\}_{1\leq k\leq m^2}$ be a sequence of points in $[-1/2,1/2]^2$ satisfying

$$\boldsymbol{x}_{(i-1)m+j}^{(n)} = \boldsymbol{t}_{i,j}^{(n)},\ \forall\, 1 \leq i,\, j \leq m.$$

Let $H : R^2 \to R$, satisfy the following conditions:

1. $\int H(\boldsymbol{x})d\boldsymbol{x} = 0$.
2. H is $m+1$ differentiable.

3. $H(0,0) > 0$ and $\|H\|_\infty < \min(\frac{1}{2\pi}, B - \frac{1}{\pi})$.
4. H vanishes outside $\{\boldsymbol{x} : \|\boldsymbol{x}\| \le 1/2\} \subset [-\frac{1}{2}, \frac{1}{2}]^2$.

Remark. The condition $H(0,0) > 0$ is for proving the lower bound result for MMSE at $\boldsymbol{x}_0 = (0,0)$. It should be obvious how to adjust it for a general point $\boldsymbol{x}_0$.

Let g_0 be the uniform distribution on the unit disc, that is

$$g_0(\boldsymbol{x}) = \frac{1}{\pi}, \ \forall \ \|\boldsymbol{x}\| < 1.$$

Let $0 < \delta < 1$, and take $m = [\frac{1}{2\delta}]$. Then $\forall \ \vec{\theta}_m = (\theta_1, \ldots, \theta_{m^2})$, define

$$g_{\vec{\theta}_m}(\boldsymbol{x}) = g_0(\boldsymbol{x}) + \delta^p \sum_{i=1}^{m} \sum_{j=1}^{m} \theta_{(i-1)m+j} H\left(\left(\boldsymbol{x} - \boldsymbol{x}^{(n)}_{(i-1)m+j}\right)/\delta\right) \tag{5.12}$$

$$f_{\vec{\theta}_m} = L \circ g_{\vec{\theta}_m}.$$

Denote

$$\mathcal{G}_n = \{g_{\vec{\theta}_m} : \vec{\theta}_m \text{ is a sequence of 0's and 1's}\},$$
$$\mathcal{F}_n = \{f_{\vec{\theta}_m} : \vec{\theta}_m \text{ is a sequence of 0's and 1's}\}.$$

It is not hard to show that $\mathcal{G}_n \subset C'_{m,\alpha,B}$. By choosing H smooth enough, $\mathcal{G}_n \subset C_{m,\alpha,B}$, $\forall \ \vec{\theta}_m$. Define

$$\vec{\theta}_{j_0} = (\theta_1, \ldots, \theta_{j-1}, 0, \theta_{j+1}, \ldots, \theta_{m^2}), \ \vec{\theta}_{j_1} = (\theta_1, \ldots, \theta_{j-1}, 1, \theta_{j+1}, \ldots, \theta_{m^2}).$$

Let $f_{\vec{\theta}_{j_0}} = L \circ g_{\vec{\theta}_{j_0}}$, $f_{\vec{\theta}_{j_1}} = L \circ g_{\vec{\theta}_{j_1}}$. Suppose $\mathcal{G}_n$ is such that for some c,

$$\max_{1 \le j \le m^2} \max_{\vec{\theta}_m \in \{0,1\}^{m^2}} \chi^2(f_{\vec{\theta}_{j_0}}, f_{\vec{\theta}_{j_1}}) \le \frac{c}{n}, \tag{5.13}$$

where

$$\chi^2(f_{\vec{\theta}_{j_0}}, f_{\vec{\theta}_{j_1}}) = \oint \int \frac{\left(f_{\vec{\theta}_{j_1}}(\boldsymbol{y}, \vec{\boldsymbol{a}}) - f_{\vec{\theta}_{j_0}}(\boldsymbol{y}, \vec{\boldsymbol{a}})\right)^2}{\left(f_{\vec{\theta}_{j_0}}(\boldsymbol{y}, \vec{\boldsymbol{a}})\right)} d\boldsymbol{y} d\vec{\boldsymbol{a}}.$$

Here "$\oint$" represents the line integral along the unit circle. Then we have

Lemma 5.2. *Suppose that* (5.13) *holds. Set* $\delta = n^{-\frac{1}{2p+3}}$. *Then for any estimator* $\hat{T}_n(\boldsymbol{x})$ *of* $g(\boldsymbol{x}) = T \circ f(\boldsymbol{x})$ *based on the* n *i.i.d. samples from an unknown density* f, *we have*

(i) *for* $f \in C_{m,\alpha,B}$,

$$\inf_{\hat{T}_n} \sup_{f\in\mathcal{F}_n} E_f \int_{\text{unit disc}} |\hat{T}_n(\boldsymbol{x}) - g(\boldsymbol{x})|^2 d\boldsymbol{x} \geq \frac{1-\sqrt{1-e^{-c}}}{8}\left[\int_{\text{unit disc}} |H(\boldsymbol{x})|^2 d\boldsymbol{x}\right] n^{-\frac{2p}{2p+3}}, \tag{5.14}$$

(ii) *for* $f \in C'_{m,\alpha,B}$ *and* $\boldsymbol{x}_0 = (0,0)$

$$\inf_{\hat{T}_n} \sup_{f\in\{f_{1n},f_{2n}\}} E_f(|\hat{T}_n(\boldsymbol{x}_0) - T\circ f(\boldsymbol{x}_0)|^2) \geq \frac{1-\sqrt{1-e^{-c}}}{8} H^2(0,0) n^{-\frac{2p}{2p+3}}, \tag{5.15}$$

where $f_{jn} = L\circ g_{jn}$, $j=1,2$, $g_{1n} = g_0$, $g_{2n}(\boldsymbol{x}) = g_0(\boldsymbol{x}) + \delta^p H(\boldsymbol{x}/\delta)$.

The proof of the lemma is basically the same as in Fan (1992). We omit it here.

To prove Theorems 2 and 4 by the lemma, we only need to verify condition (5.13). By the construction of $g_{\vec{\theta}_m}$, we know that for any given $\boldsymbol{x}$, at most one of the $H((\boldsymbol{x}-\boldsymbol{x}_j^{(n)})/\delta)$ is not zero. By *condition 3* on H,

$$g_{\vec{\theta}_m}(\boldsymbol{x}) \geq \frac{1}{2\pi}, \ \forall\, \boldsymbol{x} \in \text{unit disc},$$

$$\begin{aligned}
&\chi^2(f_{\vec{\theta}_{j_0}}, f_{\vec{\theta}_{j_1}}) \\
&= \oint\int \frac{\left(f_{\vec{\theta}_{j_1}}(\boldsymbol{y},\vec{\boldsymbol{a}}) - f_{\vec{\theta}_{j_0}}(\boldsymbol{y},\vec{\boldsymbol{a}})\right)^2}{f_{\vec{\theta}_{j_0}}(\boldsymbol{y},\vec{\boldsymbol{a}})} d\boldsymbol{y}\, d\vec{\boldsymbol{a}} \\
&= \delta^{2p}\oint\int \frac{\left(\frac{1}{(2\pi)^{3/2}\sigma}\int_{-\infty}^{\infty} H\left(\frac{\boldsymbol{y}-\boldsymbol{x}_j^{(n)}-t\vec{\boldsymbol{a}}}{\delta}\right)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right)^2}{f_{\vec{\theta}_{j_0}}(\boldsymbol{y},\vec{\boldsymbol{a}})} d\boldsymbol{y}\, d\vec{\boldsymbol{a}} \\
&\leq 2\delta^{2p}\oint\int \frac{\left(\frac{1}{(2\pi)^{3/2}\sigma}\int_{-\infty}^{\infty} H\left(\frac{\boldsymbol{y}-\boldsymbol{x}_j^{(n)}-t\vec{\boldsymbol{a}}}{\delta}\right)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right)^2}{f_0(\boldsymbol{y},\vec{\boldsymbol{a}})} d\boldsymbol{y}\, d\vec{\boldsymbol{a}},
\end{aligned} \tag{5.16}$$

where $f_0 = L\circ g_0$. To bound this distance, we establish, next,

Lemma 5.3. *There exists* $C_1 > 0$ *such that for all* j, $\boldsymbol{y}$, $\vec{\boldsymbol{a}}$, *and* $0 < \delta < \frac{1}{3}$,

$$\frac{\left|\frac{1}{(2\pi)^{3/2}\sigma}\int_{-\infty}^{\infty} H\left(\frac{\boldsymbol{y}-\boldsymbol{x}_j^{(n)}-t\vec{\boldsymbol{a}}}{\delta}\right)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right|}{f_0(\boldsymbol{y},\vec{\boldsymbol{a}})} \leq C_1\delta. \tag{5.17}$$

Proof. First, assume $x_j^{(n)} = 0$. Denote

$$H^{\delta}_{y,\vec{a}}(t) = H\left(\frac{y - t\vec{a}}{\delta}\right).$$

From the definition,

$$\frac{\left|\frac{1}{(2\pi)^{3/2}\sigma}\int_{-\infty}^{\infty} H\left(\frac{y - t\vec{a}}{\delta}\right)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right|}{f_0(y,\vec{a})}$$

$$= \frac{\left|\int_{-\infty}^{\infty} H^{\delta}_{y,\vec{a}}(t)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right|}{\int_{-\infty}^{\infty} g_0(y - t\vec{a})\exp\left(-\frac{t^2}{2\sigma^2}\right)dt}.$$

Clearly,

$$f_0(y,\vec{a}) > 0.$$

By definition, H vanishes outside the disc centered at 0 and of radius $1/2$. So, if

$$\|y\|^2 - |\vec{a}^{\tau} y|^2 > \delta^2/4,$$

then

$$H^{\delta}_{y,\vec{a}}(t) \equiv 0.$$

In this case, (5.17) is trivially true. Now suppose

$$\|y\|^2 - |\vec{a}^{\tau} y|^2 \leq \delta^2/4.$$

Let

$$y_{\vec{a}} = (\vec{a}^{\tau} y)\vec{a}, \quad \text{and} \quad t(y,\vec{a}) = \vec{a}^{\tau} y.$$

Then $H^{\delta}_{y,\vec{a}}(t)$ vanishes outside $[t(y,\vec{a}) - \delta/2, t(y,\vec{a}) + \delta/2]$. Since

$$\|y - t\vec{a}\| \leq 1, \text{ when } t \in \left[t(y,\vec{a}) - \delta/2 - \frac{4}{5}, t(y,\vec{a}) + \delta/2 + \frac{4}{5}\right],$$

and $\|H\|_{\infty} < \frac{1}{\pi}$,

$$\frac{\left|\int_{-\infty}^{\infty} H^{\delta}_{y,\vec{a}}(t)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right|}{\int_{-\infty}^{\infty} g_0(y - t\vec{a})\exp\left(-\frac{t^2}{2\sigma^2}\right)dt}$$

$$
\begin{aligned}
&\leq \frac{\int_{-\infty}^{\infty}\left|H^{\delta}_{\boldsymbol{y},\vec{\boldsymbol{a}}}(t)\right|\exp\left(-\frac{t^2}{2\sigma^2}\right)dt}{\int_{-\infty}^{\infty} g_0(\boldsymbol{y}-t\vec{\boldsymbol{a}})\exp\left(-\frac{t^2}{2\sigma^2}\right)dt} \\
&\leq 2\frac{\int_{t(\boldsymbol{y},\vec{\boldsymbol{a}})-\delta/2}^{t(\boldsymbol{y},\vec{\boldsymbol{a}})+\delta/2}\exp\left(-\frac{t^2}{2\sigma^2}\right)dt}{\int_{t(\boldsymbol{y},\vec{\boldsymbol{a}})-\delta/2-\frac{4}{5}}^{t(\boldsymbol{y},\vec{\boldsymbol{a}})+\delta/2+\frac{4}{5}}\exp\left(-\frac{t^2}{2\sigma^2}\right)dt} \\
&= 2\frac{P(\epsilon\in[t(\boldsymbol{y},\vec{\boldsymbol{a}})-\delta/2,\, t(\boldsymbol{y},\vec{\boldsymbol{a}})+\delta/2])}{P\left(\epsilon\in\left[t(\boldsymbol{y},\vec{\boldsymbol{a}})-\delta/2-\frac{4}{5},\, t(\boldsymbol{y},\vec{\boldsymbol{a}})+\delta/2+\frac{4}{5}\right]\right)} \qquad (5.18)\\
&\leq C_1\delta. \qquad (5.19)
\end{aligned}
$$

(5.19) can be simply derived from the unimodal property of the *normal* density function. For general $\boldsymbol{x}_j^{(n)}$,

$$
\begin{aligned}
&\frac{\left|\frac{1}{(2\pi)^{3/2}\sigma}\int_{-\infty}^{\infty}H\left(\frac{\boldsymbol{y}-\boldsymbol{x}_j^{(n)}-t\vec{\boldsymbol{a}}}{\delta}\right)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right|}{f_0(\boldsymbol{y},\vec{\boldsymbol{a}})} \\
&\leq 2\frac{P\left(\epsilon\in\left[t\left(\boldsymbol{y}-\boldsymbol{x}_j^{(n)},\vec{\boldsymbol{a}}\right)-\delta/2,\, t\left(\boldsymbol{y}-\boldsymbol{x}_j^{(n)},\vec{\boldsymbol{a}}\right)+\delta/2\right]\right)}{P\left(\epsilon\in\left[t(\boldsymbol{y},\vec{\boldsymbol{a}})-\delta/2-\frac{4}{5},\, t(\boldsymbol{y},\vec{\boldsymbol{a}})+\delta/2+\frac{4}{5}\right]\right)}. \qquad (5.20)
\end{aligned}
$$

Since $\|\boldsymbol{x}_j^{(n)}\|\leq\sqrt{2}/2<4/5$,

$$\left[t\left(\boldsymbol{y}-\boldsymbol{x}_j^{(n)},\vec{\boldsymbol{a}}\right)-\delta/2,\, t\left(\boldsymbol{y}-\boldsymbol{x}_j^{(n)},\vec{\boldsymbol{a}}\right)+\delta/2\right]\subset\left[t(\boldsymbol{y},\vec{\boldsymbol{a}})-\delta/2-\frac{4}{5},\, t(\boldsymbol{y},\vec{\boldsymbol{a}})+\delta/2+\frac{4}{5}\right].$$

Thus, (5.19) still holds, probably with a different constant. This proves the lemma.

By the lemma, we have

$$
\begin{aligned}
&\chi^2(f_{\vec{\theta}_{j_0}}, f_{\vec{\theta}_{j_1}}) \\
&\leq 2C_1\delta^{2p+1}\oint\int\frac{1}{(2\pi)^{3/2}\sigma}\left|\int_{-\infty}^{\infty}H\left(\frac{\boldsymbol{y}-\boldsymbol{x}_j^{(n)}-t\vec{\boldsymbol{a}}}{\delta}\right)\exp\left(-\frac{t^2}{2\sigma^2}\right)dt\right|d\boldsymbol{y}d\vec{\boldsymbol{a}} \\
&\leq \frac{2C_1\delta^{2p+3}}{(2\pi)^{3/2}\sigma}\oint\int_{-\infty}^{\infty}\left(\int\left|H\left(\boldsymbol{y}-\boldsymbol{x}_j^{(n)}-\frac{t\vec{\boldsymbol{a}}}{\delta}\right)\right|d\boldsymbol{y}\right)\exp\left(-\frac{t^2}{2\sigma^2}\right)dtd\vec{\boldsymbol{a}} \\
&\leq \frac{C_1\delta^{2p+3}}{(2\pi)^{3/2}\sigma}\oint\int_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2\sigma^2}\right)dtd\vec{\boldsymbol{a}} \qquad (5.21)\\
&= C_1\delta^{2p+3}. \qquad (5.22)
\end{aligned}
$$

(5.21) holds because H vanishes outside the unit disc and $H < \frac{1}{\pi}$. Since j is arbitrary, we know (5.13) holds.

5.3. Comparison of estimators

Proof of Theorem 5

Since we use the same kernel K, we have the same *bias* terms,

$$C_K(g)h^{2p}(1 + o(1)). \tag{5.23}$$

For the Bickel-Ritov estimator, standard computation shows

$$\begin{aligned}
&\operatorname{var}(\hat{g}_h(\boldsymbol{x})) \\
= &\ \frac{1+o(1)}{n} \iint \rho_h^2(x_1 \cos\psi + x_2 \sin\psi - t)\lambda(\psi, t)dtd\psi \\
= &\ \frac{1+o(1)}{nh^3} \iint \rho_1^2(t,\psi)\lambda(\psi, (x_1 \cos\psi + x_2 \sin\psi + ht))dtd\psi \\
= &\ \frac{(1+o(1))}{nh^3} \iint \rho_1^2(t,\psi)\lambda(\psi, x_1 \cos\psi + x_2 \sin\psi)dtd\psi \\
= &\ \frac{(1+o(1))}{8\pi nh^3} \iint w^2|\phi_K^2(w\cos\psi, w\sin\psi)|\lambda(\psi, x_1\cos\psi + x_2 \sin\psi)dwd\psi \\
&\ \text{(by Parseval's equality and (4.2))} \\
= &\ \frac{(1+o(1))}{8\pi^2 nh^3} \left[\iint |\phi_K^2(\boldsymbol{t})|\ \|\boldsymbol{t}\|dt\right] \left[\int_0^\pi \lambda(\psi, x_1 \cos\psi + x_2 \sin\psi)d\psi\right] \\
&\ (\phi_K^2(w\cos\psi, w\sin\psi) \text{ does not depend on } \psi).
\end{aligned}$$

Globally, we have

$$\int_{\|\boldsymbol{x}\|\le 1} \operatorname{var}(\hat{g}_h(\boldsymbol{x}))d\boldsymbol{x} = \frac{(1+o(1))}{8\pi^2 nh^3} \left[\iint |\phi_K^2(\boldsymbol{t})|\ \|\boldsymbol{t}\|\ dt\right] B(g), \tag{5.24}$$

where

$$B(g) = \int_{\|\boldsymbol{x}\|\le 1} \int_0^\pi \lambda(\psi, x_1\cos\psi + x_2\sin\psi)d\psi d\boldsymbol{x},$$

and λ is the *Radon transform* of g. For $B(g)$, we have

Lemma 5.4. *For any density g on the* unit disc

$$B(g) \ge \frac{4}{\pi}$$

with equality iff g is concentrated on the boundary of the disc.

Proof. Under a change of coordinates:

$$\begin{aligned} t_1 &= x_1 \cos\psi + x_2 \sin\psi, \\ t_2 &= -x_1 \sin\psi + x_2 \cos\psi. \end{aligned} \tag{5.25}$$

$$\begin{aligned} B(g) &= \int_{\|\boldsymbol{t}\|\le 1} \int_0^{\pi} \lambda(\psi, t_1) d\psi d\boldsymbol{t} \\ &= 2\int_0^{\pi}\int_{-1}^{1} \sqrt{1-t_1^2}\, \lambda(\psi, t_1) d\psi dt_1. \end{aligned}$$

We know that

$$\lambda(\psi, t_1) = \frac{1}{\pi}\int g(t_1\cos\psi - t_2\sin\psi, t_1\sin\psi + t_2\cos\psi)dt_2.$$

Thus,

$$\begin{aligned} B(g) &= \frac{2}{\pi}\int_{\|\boldsymbol{t}\|\le 1}\int_0^{\pi} \sqrt{1-t_1^2}\, g(t_1\cos\psi - t_2\sin\psi,\ t_1\sin\psi + t_2\cos\psi)d\psi d\boldsymbol{t} \\ &= \frac{2}{\pi}\int_{\|\boldsymbol{x}\|\le 1}\int_0^{\pi} \sqrt{1-(x_1\cos\psi + x_2\sin\psi)^2}\, g(x_1,x_2)d\psi d\boldsymbol{x} \\ &= \frac{2}{\pi}\int_{\|\boldsymbol{x}\|\le 1} g(x_1,x_2)\left[\int_0^{\pi}\sqrt{1-\|\boldsymbol{x}\|^2\cos\psi^2}\, d\psi\right] d\boldsymbol{x} \\ &\ge \frac{2}{\pi}\int_{\|\boldsymbol{x}\|\le 1} g(x_1,x_2)\left[\int_0^{\pi}\sqrt{1-\cos\psi^2}\, d\psi\right] d\boldsymbol{x} \\ &= \frac{4}{\pi}. \end{aligned}$$

This proves the lemma.

Therefore, for the Bickel-Ritov estimator, the MISE at g is

$$\text{MISE}_h(g) \ge C_K(g)h^{2p}(1+o(1)) + \frac{1}{2\pi^3 n h^3}\left[\iint |\phi_K^2(\boldsymbol{t})|\ \|\boldsymbol{t}\|\, d\boldsymbol{t}\right](1+o(1)). \tag{5.26}$$

By (5.7), for MDK, the MISE is

$$\text{MISE}_h(g) \ge C_K(g)h^{2p}(1+o(1)) + \frac{\sigma\sqrt{\pi}}{4\pi^2 n h^3}\int |\phi_K(\boldsymbol{t})|^2\|\boldsymbol{t}\|\, d\boldsymbol{t}(1+o(1)). \tag{5.27}$$

Choosing the optimal bandwidths h for both estimators, we have

$$\frac{\min_h \text{MISE}_h(\text{MDK})}{\min_h \text{MISE}_h(\text{BR})} \le \left[\frac{\sigma\pi^{3/2}}{2}\right]^{\frac{2p}{2p+3}}$$

This proves *Theorem 5.*

Proof of Theorem 6

By *Corrollary 5.1*, if σ goes to zero not as fast as h does, then (5.6) is still valid and, therefore,

$$\mathrm{MISE}_n(h) = O(h^{2p}) + O\left(\frac{n^{-\beta}}{nh^3}\right). \tag{5.28}$$

Choosing $h_n = n^{-\frac{1+\beta}{2p+3}}$ yields the best convergence rate $n^{-(1+\beta)\frac{2p}{2p+3}}$. Note that if $\beta < \frac{1}{2p+2}$, then $\sigma \gg h_n$. This proves the first assertion.

In (5.6), if $\sigma/h \to 0$, we can treat $\phi_\epsilon(t/h)$ as 1. So in this case, (5.6) turns out to be

$$\int \frac{|\phi_K(th)|^2}{|\phi_\epsilon(t)|} dt = (1+o(1))\frac{\sigma\sqrt{\pi}}{h^2}\int |\phi_K(t)|^2 \|t\|\, dt. \tag{5.29}$$

Choosing $h_n = n^{-\frac{1}{2p+2}}$ yields the best convergence rate $n^{-\frac{2p}{2p+2}}$. Note that if $\beta > \frac{1}{2p+2}$, then $\sigma \ll h_n$. This proves the second assertion.

Acknowledgement

I am pleased to thank Professor Iain Johnstone for leading me to this interesting problem and for many helpful discussions and suggestions. I am also grateful to two referees whose help greatly improved the presentation.

References

Bickel, P. J. and Ritov, Y. (1990). Estimating linear functionals of a PET image. Technical Report, Department of Statistics, University of California, Berkeley.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184–1186.

Donoho, D. L. and Liu, R. C. (1991a). Geometrizing rate of convergence, II. *Ann. Statist.* **19**, 633–667.

Donoho, D. L. and Liu, R. C. (1991b). Geometrizing rate of convergence, III. *Ann. Statist.* **19**, 668–701.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.

Fan, J. (1992). Adaptively local 1-dimensional subproblems. *Ann. Statist.*, to appear.

Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* **18**, 251–280.

O'Sullivan, F. (1989). Penalized likelihood-type estimation in PET and more general inverse problems involving linear operators. Technical Report, Department of Biostatistics, University of Washington.

Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, Orlando, Florida.

Shepp, L. A. and Vardi, Y. (1982). Maximum likelihood reconstruction in positron emission tomography. *IEEE Trans. Medical Imaging* **1**, 113–122.

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

Snyder, D. L., Thomas, L. J., Je. and Ter-Pogossian, M. M. (1981). A mathematical model for positron-emission tomography systems having time-of-flight measurements. *IEEE Trans. Nuclear Sci.* **NS-28**, 3575–3583.

Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 169–184.

Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9**, 229–235.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040–1053.

Vardi, Y., Shepp, L. A. and Kaufman, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80**, 8–37.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received September 1990; accepted February 1992)