Deconvolution with Supersmooth Distributions

Jianqing Fan

Department of Statistics

University of North Carolina

Chapel Hill, N.C. 27514

July 15, 1990

Abstract

The desire to recover the unknown density when data are contaminated with errors leads to nonparametric deconvolution problems. Optimal global rates of convergence are found under the weighted L_p-loss ($1 \leq p \leq \infty$). It appears that the optimal rates of convergence are extremely slow for supersmooth error distributions. To overcome the difficulty, we examine how large the noise level can be for deconvolution to be feasible, and for the deconvolution estimate to be as good as the ordinary density estimate. It is shown that if noise level is not too large, nonparametric Gaussian deconvolution can still be practical. Several simulation studies are also presented.

---

[0] *Abbreviated title.* Supersmooth Deconvolution.

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

Key words and phrases. Deconvolution, Fourier transforms, kernel density estimates, L_p-norm, global rates of convergence, minimax risks.

Section 4 examines how the theory works for moderate sample sizes via simulation studies. Futher remarks are given in section 5. Proofs are deferred in section 6.

2. Optimal Global Rates

Let's give a global lower bound on rates for supersmooth error distributions. Let's assume that the second half inequality of (1.4) holds:

$$|\phi_\varepsilon(t)||t|^{-\beta_1}\exp(|t|^\beta/\gamma) \le d_1 \qquad (\text{as } t \to \infty), \tag{2.1}$$

for some constants $\beta, \gamma > 0$, $d_1 \ge 0$, and β_1, and that

$$P\{|\varepsilon - x| \le |x|^{\alpha_0}\} = O\left(|x|^{-(a-\alpha_0)}\right), \qquad (\text{as } x \to \pm\infty), \tag{2.2}$$

for some $0 < \alpha_0 < 1$ and $a > 1 + \alpha_0$.

Theorem 1. *Suppose that the distribution of error variable ε satisfies (2.1) and (2.2) and $f \in C_{m,B}$. Then, no estimator can estimate $f^{(l)}(x)$ faster than the rate $O\left((\log n)^{-(m-l)/\beta}\right)$ in the sense that for any estimator $\hat{T}_n(x)$,*

$$\liminf_{n\to\infty} \sup_{f \in C_{m,B}} (\log n)^{(m-l)/\beta} E_f \|\hat{T}_n(\cdot) - f_X^{(l)}(\cdot)\|_{wp} > C_{p,l}, \tag{2.3}$$

for all $1 \le p \le \infty$, provided that the weight function $w(\cdot)$ is positive continuous on some interval, where $C_{p,l}$ is a positive constant independent of the estimator.

In terms of technical argument of Theorem 1, we will use the technique of adaptively local one-dimensional subproblems developed by Fan (1989), and then reduce the global problem to a pointwise estimation problem so that the existing lower bound (Fan (1990)) on pointwise rates of convergence can be used. To our knowledge, the technical argument appears to be new!

Now, let's show that the rate above is indeed attained by the deconvolution kernel estimator (1.3), and hence it is optimal. Some assumptions on kernel function $K(\cdot)$ are

Condition 1:

- $K(\cdot)$ is bounded continuous, and $\int_{-\infty}^{+\infty} |y|^m |K(y)| dy < \infty$.

- The Fourier transform ϕ_K of K has a bounded support $|t| \le M_0$. Moreover, $\phi_K(t) = 1 + O(|t|^m)$.

Theorem 2. *Assume that $\phi_\varepsilon(t) \neq 0$ for* any t, *and that*

$$|\phi_\varepsilon(t)||t|^{-\beta_2} \exp(|t|^\beta/\gamma) \ge d_2, \tag{2.4}$$

for some positive constants β, γ, d_2 and constant β_2. If the kernel function K satisfies Condition 1, then for $h_n = cM_0(2/\gamma)^{1/\beta}(\log n)^{-1/\beta}$ with $c > 1$,

$$\sup_{f \in \mathcal{C}_{m,B}} E\|\hat{f}_n^{(l)}(\cdot) - f_X^{(l)}(\cdot)\|_{wp} = O\left((\log n)^{-(m-l)/\beta}\right) \tag{2.5}$$

for all $0 \le p \le \infty$, *provided that the weight function is integrable.*

In light of the bandwidth given by Theorem 2, there is no much room for bandwidth selection. If $c > 1$, then the variance converges to 0 much faster than the bias does, while if $c < 1$, the variance goes to infinity. Thus, practical selection of bandwidth would select a constant c close to 1 in Theorem 2.

The distributions satisfying conditions (2.1), (2.2), and (2.4) include normal, mixture normal, and Cauchy distributions. For these supersmooth error distributions, nonparametric deconvolution is extremely hard: the *optimal* rate of convergence is only of order $(\log n)^{-m/\beta}$. One way of resolving this difficulty will be discussed in the next section.

Some special global results (basically $p = m = 2, l = 0$, ε normal or Cauchy) are obtained independently by Zhang (1990) under different formulation. The results in Theorem 1 & 2 provide better insights: it shows that both lower and upper bounds depend on F_ε only through the tail of ϕ_ε, and the dependence is explicitly addressed.

Remark 1. In an early version of the proof of Theorem 1 (see Fan (1988), for which the results in this section are based), a 1-dimensional subproblem is hard enough to capture the difficulty of the full global deconvolution problem. In contrast with the ordinary density estimation (Stone (1982)), in order to construct an attainable lower bound under the global and

$$\theta_{j1} = (\theta_1, \cdots, \theta_{j-1}, 1, \theta_{j+1}, \cdots, \theta_{m_n}).$$

Let F_ε be the distribution of ε, and $\chi^2(f,g) = \int (f-g)^2/f dx$ be the χ^2-distance. By Theorem 1 of Fan (1989), if

$$\max_{1\le j\le m_n} \max_{\theta\in\{0,1\}^{m_n}} \chi^2\left(f_{\theta_{j0}} * F_\varepsilon, f_{\theta_{j1}} * F_\varepsilon\right) \le c_1/n, \tag{6.3}$$

then

$$\begin{aligned} &\inf_{\hat{T}_n(x)} \sup_{f\in\mathcal{C}_{m,B}} E_f \int_0^1 |\hat{T}_n(x) - f_X^{(l)}(x)|^p w(x)dx \\ \ge\ & \frac{1-\sqrt{1-\exp(-c_1)}}{2^{p+1}} \int_0^1 w(x)dx \int_0^1 |H^{(l)}(x)|^p dx (m_n^{-(m-l)})^p. \end{aligned} \tag{6.4}$$

Thus, $m_n^{-(m-l)}$ is the global lower rate.

Let's determine m_n from (6.3). Note that there exists a positive constant c_2 such that $f_0(x) > c_2 f_0(x + j/m_n)$ $(1 \le j \le m_n)$. By (6.1) and (6.2) with a change of variable, we have

$$\begin{aligned} \max_{1\le j\le m_n} \max_{\theta\in\{0,1\}^{m_n}} \chi^2\left(f_{\theta_{j0}} * F_\varepsilon, f_{\theta_{j1}} * F_\varepsilon\right) &\le 2\delta^2 m_n^{-2m} \int_{-\infty}^{+\infty} \frac{[H(m_n(\cdot)) * F_\varepsilon]^2}{f_0(\cdot + x_j) * F_\varepsilon} dx. \\ &\le 2\frac{\delta^2 m_n^{-2m}}{c_2} \int_{-\infty}^{+\infty} \frac{[H(m_n(\cdot)) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dx. \end{aligned} \tag{6.5}$$

To construct a pointwise minimax lower bound, one has also to select m_n such that (6.5) is of order $O(1/n)$, which is determined by Fan (1990) to be $m_n = c_3(\log n)^{1/\beta}$, for some constant $c_3 > 0$. Consequently, the global rate is of order $m_n^{-(m-l)} = c_3^{-(m-l)}(\log n)^{-(m-l)/\beta}$. The conclusion follows.

6.2. Proof of Theorem 2

We need only to prove the result for $p = \infty$; the other result follows from

$$E\|\hat{f}_n^{(l)}(\cdot) - f_X^{(l)}(\cdot)\|_{wp} \le E\|\hat{f}_n^{(l)}(\cdot) - f_X^{(l)}(\cdot)\|_\infty,$$

by assuming that $\int_{-\infty}^{+\infty} w(x)dx = 1$.

Note that

$$Ef_n^{(l)}(x) = \int_{-\infty}^{+\infty} f_X^{(l)}(x - h_n y)K(y)dy,$$

which is independent of the error distribution F_ϵ. Thus, by the results in the ordinary density estimation, or Taylor's expansion

$$\sup_{f \in \mathcal{C}_{m,B}} \|E\hat{f}_n^{(l)}(\cdot) - f_X^{(l)}(\cdot)\|_\infty \le \frac{B}{(m-l)!} \int_{-\infty}^{+\infty} |y|^{(m-l)}|K(y)|dy h_n^{(m-l)}. \tag{6.6}$$

Thus, we need only to verify that

$$\sup_{f \in \mathcal{C}_{m,B}} E\|\hat{f}_n^{(l)}(\cdot) - E\hat{f}_n^{(l)}(\cdot)\|_\infty = O\left((\log n)^{-(m-l)/\beta}\right).$$

Note that by (1.3),

$$\begin{aligned} \|\hat{f}_n^{(l)}(\cdot) - E\hat{f}_n^{(l)}(\cdot)\|_\infty &\le \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\phi_K(th_n)||t|^l \frac{E|\hat{\phi}_n(t) - \phi_Y(t)|}{|\phi_\epsilon(t)|} dt \\ &\le \frac{\max|\phi_K|}{2\pi n^{1/2}} \int_{|t| \le M_0/h_n} \frac{|\phi_K(th_n)||t|^l}{|\phi_\epsilon(t)|} dt \\ &\le \frac{M_0^l \max|\phi_K|}{2\pi n^{1/2} h_n^{l+1}} \int_{|t| \le M_0} \frac{1}{|\phi_\epsilon(t/h_n)|} dt, \end{aligned} \tag{6.7}$$

by using the fact that ϕ_K has a support $[-M_0, M_0]$, and that

$$E|\hat{\phi}_n(t) - \phi_Y(t)| \le \left(E|\hat{\phi}_n(t) - \phi_Y(t)|^2\right)^{1/2} \le n^{-1/2}.$$

By (2.4), there exists a constant t_0 such that when $|t| \ge t_0$,

$$|\phi_\epsilon(t)||t|^{-\beta_2} \exp(|t|^\beta/\gamma) \ge d_2/2.$$

Consequently, by (6.7) and the fact that $\min_{|t| \le t_0} |\phi_\epsilon(t)| > 0$, we have

$$\|\hat{f}_n^{(l)}(\cdot) - E\hat{f}_n^{(l)}(\cdot)\|_\infty = O\left(h_n^{-\beta_2 - l - 1} n^{-1/2} \exp(|M_0/h_n|^\beta/\gamma)\right).$$

With the bandwidth given by Theorem 2, the last display is of order $o((\log n)^{-d})$, for any positive constant d. This completes the proof.

6.5. Proof of Theorem 5

First, using the integration by parts twice, it is easy to see that K_n defined by (3.5) is bounded by

$$|K_n(x)| \le \frac{C}{1+x^2} \quad \text{(for some constant } C\text{)},$$

i.e. $K_n(x)$ is bounded and decays at the rate $|x|^{-2}$ as $|x| \to \infty$. Thus, $\hat{F}_n^*$ is well defined, and can be expressed as

$$\hat{F}_n^*(x) = \frac{1}{n}\sum_1^n K^*\left(\frac{x - Y_j}{h_n}\right)$$

with $K^*(x) = \int_{-\infty}^{x} K_n(y)dy$. Note that

$$\sup_x |E\hat{F}_n^*(x) - F(x)| = \sup_x |\int_{-\infty}^{+\infty} F_X(x - h_n y)K(y)dy - F(x)| = O(h_n^3) = O(n^{-1/2}).$$

Thus, we need to prove that

$$E\|\hat{F}_n^*(\cdot) - E\hat{F}_n^*(\cdot)\|_{wp} = O(n^{-1/2}),$$

which follows from Marcinkiewicz-Zugmund's inequality (Chow and Teicher (1988), p356) or direct expansion by assuming $p = 2j$ as Theorem 4,

$$E\left|\frac{1}{n}\sum_1^n \left[K^*\left(\frac{x - Y_j}{h_n}\right) - EK^*\left(\frac{x - Y_j}{h_n}\right)\right]\right|^p \le Dn^{-p/2},$$

for some constant D, as had to be shown.

References

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

[2] Chow, Y.S. and Teicher, H. (1988). *Probability Theory*, 2^{nd} edition. Springer-Verlag, New York.

[3] Devroye (1989). Consistent deconvolution in density estimation. *Canadian J. Statist.*, **17**, 235-239.

[4] Fan, J. (1988). Optimal global rates of convergence for nonparametric deconvolution problems. *Tech. Report 162*, Dept. of Statist., Univ. of California, Berkeley.

[5] Fan, J. (1989). Adaptively local 1-dimensional subproblems. *Institute of Statistics Mimeo Series #2010*, Univ. of North Carolina, Chapel Hill.

[6] Fan, J. (1990). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.*, to appear.

[7] Fan, J. and Truong, Y. (1990). Nonparametric regression with errors-in-variables. *Institute of Statistics Mimeo Series #2023*, Univ. of North Carolina, Chapel Hill.

[8] Fuller, W. A. (1987). *Measurement error models.* Wiley, New York.

[9] Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canadian J. Statist.*, **17**, 427-438.

[10] Masry, E. and Rice, J.A. (1990). Gaussian deconvolution via differentiation. *Manuscript.*

[11] Mendelson, J. and Rice, J.A. (1982). Deconvolution of microfluorometric histograms with B splines. *J. Amer. Stat. Assoc.*, **77**, 748–753.

[12] Stefanski, L.A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Letters*, **9**, 229-235.

[13] Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, to appear.

[14] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040-1053.

[15] Wise, G. L., Traganitis, A. D., and Thomas, J. B. (1977). The estimation of a probability density from measurements corrupted by Poisson noise. *IEEE Trans. Inform. Theory*, **23**, 764–766.