Pointwise deconvolution with unknown error distribution

Fabienne Comte[a], Claire Lacour[b]

[a] *MAP5, UMR 8145, Université Paris Descartes, 45 rue des Saints-Pères, 75006 Paris, France*
[b] *Laboratoire de mathématiques, Université Paris-Sud, 91405 Orsay cedex, France*

Abstract

This note presents rates of convergence for the pointwise mean squared error in the deconvolution problem with estimated characteristic function of the errors.

Résumé

Déconvolution ponctuelle avec distribution de l'erreur inconnue. Cette note présente les vitesses de convergence pour le risque quadratique ponctuel dans le problème de déconvolution avec fonction caractéristique des erreurs estimée.

1. Introduction

Let us consider the following model:

$$Y_j = X_j + \varepsilon_j \qquad j = 1, \ldots, n \tag{1}$$

where $(X_j)_{1\le j\le n}$ and $(\varepsilon_j)_{1\le j\le n}$ are independent sequences of i.i.d. variables. We denote by f the density of X_j and by f_ε the density of ε_j. The aim is to estimate f when only $Y_1, \ldots, Y_n$ are observed. Contrary to the classical convolution model, we do not assume that the density of the error is known, but that we additionally observe $\varepsilon_{-1}, \ldots, \varepsilon_{-M}$, a noise sample with distribution f_ε, independent of $(Y_1, \ldots, Y_n)$. Note that the availability of two distinct samples makes the problem identifiable.

Altough there exists a huge literature concerning the estimation of f when f_ε is known, this problem without the knowledge of f_ε has been less studied. One can cite Efromovich (1997) in a context of circular data and Diggle and Hall (1993) who examine the case $M \ge n$. Neumann (1997) and Johannes (2009) give bounds for the integrated risk.

The contribution of this paper is to give rates of convergence for the pointwise squared error depending on M and n.

Notations

For z a complex number, $\bar{z}$ denotes its conjugate and $|z|$ its modulus. For a function $t : \mathbb{R} \mapsto \mathbb{R}$ belonging to $\mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R})$, we denote by $\|t\|$ the $\mathbb{L}^2$ norm of t and by $\|t\|_1$ the $\mathbb{L}^1$ norm of t. The Fourier transform t^* of t is defined by $t^*(u) = \int e^{-ixu} t(x)dx$.

2. Estimation procedure

It follows easily from Model (1) and independence assumptions that, if f_Y denotes the common density of the Y_j's, then $f_Y = f * f_\varepsilon$ and thus $f_Y^* = f^* f_\varepsilon^*$. Therefore, under the classical assumption:

Email addresses: fabienne.comte@parisdescartes.fr (Fabienne Comte), claire.lacour@math.u-psud.fr (Claire Lacour)

(A1) $\forall x \in \mathbb{R},\ f_\varepsilon^*(x) \neq 0,$

the equality $f^* = f_Y^*/f_\varepsilon^*$ yields an estimator of f^* by considering the following estimate of f_Y^*:

$$\hat{f_Y^*}(u) = \frac{1}{n}\sum_{j=1}^{n} e^{-iuY_j}.$$

Indeed, if f_ε^* is known, we can use the following estimate of f^*: $\hat{f_Y^*}/f_\varepsilon^*$. Then, we should use inverse Fourier transform to get an estimate of f. As $1/f_\varepsilon^*$ is in general not integrable (think of a Gaussian density for instance), this inverse Fourier transform does not exist, and a cutoff is used. The final estimator for known f_ε can thus be written: $(2\pi)^{-1}\int_{|u|\leq \pi m} e^{iux}\hat{f_Y^*}(u)/f_\varepsilon^*(u)du$. This estimator is classical in the sense that it corresponds both to a kernel estimator built with the sinc kernel (see Butucea (2004)) or to a projection type estimator as in Comte et al. (2006).

Now, f_ε^* is unknown and we have to estimate it. Therefore, we use the preliminary sample and we define the natural estimator of f_ε^*: $\hat{f_\varepsilon^*}(x) = \frac{1}{M}\sum_{j=1}^{M} e^{-ix\varepsilon_{-j}}$. Next, we introduce as in Neumann (1997) the truncated estimator:

$$\frac{1}{\tilde{f_\varepsilon^*}(x)} = \frac{\mathbb{1}_{\{|\hat{f_\varepsilon^*}(x)|\geq M^{-1/2}\}}}{\hat{f_\varepsilon^*}(x)} = \frac{1}{\hat{f_\varepsilon^*}(x)} \text{ if } |\hat{f_\varepsilon^*}(x)| \geq M^{-1/2} \text{ and } \frac{1}{\tilde{f_\varepsilon^*}(x)} = 0 \text{ otherwise.}$$

Then our estimator is

$$\hat{f}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f_Y^*}(u)}{\tilde{f_\varepsilon^*}(u)}du. \tag{2}$$

3. Study of the pointwise mean squared error

We introduce the notations

$$\Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2}du, \quad \Delta^0(m) = \frac{1}{2\pi}\left(\int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-1}du\right)^2, \quad \Delta_f^0(m) = \frac{1}{2\pi}\left(\int_{-\pi m}^{\pi m} \frac{|f^*(u)|}{|f_\varepsilon^*(u)|}du\right)^2.$$

Proposition 3.1. *Consider model (1) under* **(A1)***, then there exist constants $C, C' > 0$ such that*

$$\mathbb{E}[(\hat{f}_m(x) - f(x))^2] \leq 2\left(\frac{1}{2\pi}\int_{|t|\geq \pi m} |f^*(t)|dt\right)^2 + \frac{C}{n}\min(\|f_Y^*\|_1\Delta(m), \Delta^0(m)) + C'\frac{\Delta_f^0(m)}{M}$$

Assumption **(A1)** is generally strengthened by the following description of the rate of decrease of f_ε^*:

(A2) There exist $s \geq 0, b > 0$, $\gamma \in \mathbb{R}$ ($\gamma > 0$ if $s = 0$) and $k_0, k_1 > 0$ such that

$$k_0(x^2+1)^{-\gamma/2}\exp(-b|x|^s) \leq |f_\varepsilon^*(x)| \leq k_1(x^2+1)^{-\gamma/2}\exp(-b|x|^s)$$

Moreover, the density function f to estimate generally belongs to the following type of smoothness spaces:

$$\mathcal{A}_{\delta,r,a}(l) = \{f \text{ density on } \mathbb{R} \text{ and } \int |f^*(x)|^2(x^2+1)^\delta \exp(2a|x|^r)dx \leq l\} \tag{3}$$

with $r \geq 0, a > 0$, $\delta \in \mathbb{R}$ and $\delta > 1/2$ if $r = 0$, $l > 0$.
When $r > 0$ (respectively $s > 0$), the function f (respectively f_ε) is known as supersmooth, and as ordinary smooth otherwise. The spaces of ordinary smooth functions correspond to classic Sobolev classes, while supersmooth functions are infinitely differentiable. It includes for example normal ($r = 2$) and Cauchy ($r = 1$) densities.

Corollary 3.2. *If f_ε^* satisfies* **(A2)** *and if $f \in \mathcal{A}_{\delta,r,a}(l)$, the rates of convergence for the Mean Squared Error $\mathbb{E}[(\hat{f}_m(x) - f(x))^2]$ are given in Table 1.*

Indeed, if $f \in \mathcal{A}_{\delta,r,a}(l)$, the bias term can be bounded in the following way

$$2\left(\frac{1}{2\pi}\int_{|t|\geq \pi m}|f^*(t)|dt\right)^2 \leq K_1(\pi m)^{-2\delta+1-r}e^{-2a(\pi m)^r}.$$

and straightforward computation gives $\Delta(m) \leq K_2(\pi m)^{2\gamma+1-s}e^{2b(\pi m)^s}$, $\Delta^0(m) \leq K_3(\pi m)^{2\gamma+2-2s}e^{2b(\pi m)^s}$ and, with $v = 2\gamma+1-s$,

$$\begin{aligned}\Delta_f^0(m)K_4^{-1} \quad &\leq \quad (\pi m)^{(2\gamma+1-2\delta)_+}(\log(m))^{\mathbb{1}_{\delta=\gamma+1/2}}\mathbb{1}_{\{r=s=0\}} + (\pi m)^{v-\max(2\delta,s-1)}e^{2b(\pi m)^s}\mathbb{1}_{\{s>r\}}\\ &+(\pi m)^{v-2\delta}e^{2(b-a)(\pi m)^s}\mathbb{1}_{\{r=s,b\geq a\}} + \mathbb{1}_{\{r>s\}\cup\{r=s,b<a\}}\end{aligned}$$

where K_1, K_2, K_3, K_4 are positive constants. Then the rates of Table 1 are established choosing a dequate m_0 depending on n, M and the smoothness indices (for example, in the case $r = s = 0$, m_0 is the integer part of $n^{1/(2\delta+2\gamma)} \wedge M^{1/\max(2\gamma,2\delta-1)}$).

	$s = 0$	$s > 0$
$r = 0$	$n^{-\frac{2\delta-1}{2\delta+2\gamma}} + M^{-[\min(1,\frac{2\delta-1}{2\gamma})]}(\log M)^{\mathbb{1}_{\delta=\gamma+1/2}}$	$(\log n)^{-\frac{2\delta-1}{s}} + (\log M)^{-\frac{2\delta-1}{s}}$
$r > 0$	$\dfrac{(\log n)^{\frac{2\gamma+1}{r}}}{n} + \dfrac{1}{M}$	See Lacour (2006) and comment below.

Table 1: Rates of convergence for the MSE if f_ε^* satisfies **(A2)** and $f \in \mathcal{A}_{\delta,r,a}(l)$.

For the case ($r > 0, s > 0$), the rules for the compromise between supersmooth terms in both squared bias and variance are given in Lacour (2006) and are very tedious to write; moreover, this case contains several sub-cases.

The rates in term of n are known to be the optimal one for the deconvolution with known error (see Fan (1991) and Butucea (2004)). They are recovered as soon as $M \geq n$. Extending the proof of Neumann (1997) we can prove the optimality of the rate M^{-1} in the cases where f is smoother than f_ε and $r \leq 1$.

4. Proof of Proposition 3.1

First, let us denote $f_m(x) = (2\pi)^{-1}\int_{-\pi m}^{\pi m} e^{ixu}f^*(u)du$ and $R(x) = \left(\frac{1}{\tilde{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)}\right)$. Then we have the following decomposition:

$$\begin{aligned}&\mathbb{E}[(\hat{f}_m(x) - f(x))^2] \leq 2(f_m(x) - f(x))^2 + 2\mathbb{E}[(\hat{f}_m(x) - f_m(x))^2]\\ \leq \quad & 2(f_m(x)-f(x))^2 + 4\mathrm{Var}\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(-u)}du\right) + 4\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\hat{f}_Y^*(u)R(u)du\right)^2\right] \quad (4)\end{aligned}$$

Since $(f - f_m)(x) = (1/2\pi)(f^* - f_m^*)^*(-x)$, we can bound the biais term in the following way

$$(f_m(x) - f(x))^2 \leq \left(\frac{1}{2\pi}\int_{|t|\geq \pi m}|f^*(t)|dt\right)^2. \quad (5)$$

The second term of the right-hand-side of (4) is the variance term when f_ε^* is known and has already been studied: it follows from Butucea and Comte (2009) that

$$\mathrm{Var}\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(-u)}du\right) \leq \frac{1}{2\pi n}\min(\|f_Y^*\|_1\Delta(m), \Delta^0(m)). \quad (6)$$

For the remaining term in (4), we write first:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\hat{f_Y^*}(u)R(u)du\right)^2\right] &\leq 2\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}(\hat{f_Y^*}(u)-f_Y^*(u))R(u)du\right)^2\right]\\ &\quad +2\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}f_Y^*(u)R(u)du\right)^2\right] := 2T_1+2T_2.\end{aligned}$$

Neumann (1997) proved that there exists a positive constant C_1 such that

$$\mathbb{E}|[R(u)|^2] = \mathbb{E}\left(\left|\frac{1}{\tilde{f_\varepsilon^*}(u)} - \frac{1}{f_\varepsilon^*(u)}\right|^2\right) \leq C_1 \min\left(\frac{1}{|f_\varepsilon^*(u)|^2}, \frac{1}{M|f_\varepsilon^*(u)|^4}\right).$$

Then we find

$$\begin{aligned}T_1 &= \frac{1}{4\pi^2}\iint e^{ix(u-v)}\mathrm{Cov}(\hat{f_Y^*}(u),\hat{f_Y^*}(v))\mathbb{E}(R(u)\bar{R}(v))dudv\\ &\leq \frac{1}{4\pi^2 n}\iint |f_Y^*(u-v)|\sqrt{\mathbb{E}(|R(u)|^2)\mathbb{E}(|R(v)|^2)}dudv \leq \frac{C_1}{4\pi^2 n}\iint \frac{|f_Y^*(u-v)|}{|f_\varepsilon^*(u)f_\varepsilon^*(v)|}dudv.\end{aligned}$$

This term is clearly bounded by $C_1(2\pi n)^{-1}\Delta^0(m)$. Moreover writing it as

$$\frac{C_1}{4\pi^2 n}\iint \frac{\sqrt{|f_Y^*(u-v)|}}{|f_\varepsilon^*(u)|}\frac{\sqrt{|f_Y^*(u-v)|}}{|f_\varepsilon^*(v)|}dudv$$

and using first the Schwarz Inequality, and second the Fubini Theorem yields the bound $C_1(2\pi n)^{-1}\|f_Y^*\|_1\Delta(m)$. Therefore

$$\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}(\hat{f_Y^*}(u)-f_Y^*(u))R(u)du\right)^2\right] \leq \frac{C_1}{2\pi n}\min(\|f_Y^*\|_1\Delta(m), \Delta^0(m)), \tag{7}$$

and thus it has the same order as the usual variance term. Lastly,

$$\begin{aligned}T_2 &\leq \frac{1}{4\pi^2}\iint_{|u|,|v|\leq \pi m} |f_Y^*(u)f_Y^*(v)|\sqrt{\mathbb{E}(|R(u)|^2)\mathbb{E}(|R(v)|^2)}dudv\\ &\leq \frac{1}{4\pi^2}\left(\int_{-\pi m}^{\pi m}|f_Y^*(u)|\sqrt{\mathbb{E}(|R(u)|^2)}du\right)^2 \leq \frac{C_1}{4\pi^2 M}\left(\int_{-\pi m}^{\pi m}\frac{|f_Y^*(u)|}{|f_\varepsilon^*(u)|^2}du\right)^2 = C_1\frac{\Delta_f^0(m)}{2\pi M}.\end{aligned} \tag{8}$$

Inserting the bounds (5) to (8) in Inequality (4), we obtain the result of Proposition 3.1. □

References

Butucea, C., 2004. Deconvolution of supersmooth densities with smooth noise. Canad. J. Statist. 32 (2), 181–192.

Butucea, C., Comte, F., 2009. Adaptive estimation of linear functionals in the convolution model and applications. Bernoulli 15 (1), 69–98.

Comte, F., Rozenholc, Y., Taupin, M.-L., 2006. Penalized contrast estimator for adaptive density deconvolution. Canad. J. Statist. 34 (3), 431–452.

Diggle, P. J., Hall, P., 1993. A Fourier approach to nonparametric deconvolution of a density estimate. J. Roy. Statist. Soc. Ser. B 55 (2), 523–531.

Efromovich, S., 1997. Density estimation for the case of supersmooth measurement error. J. Amer. Statist. Assoc. 92 (438), 526–535.

Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 19 (3), 1257–1272.

Johannes, J., 2009. Deconvolution with unknown error distribution. Ann. Statist.

Lacour, C., 2006. Rates of convergence for nonparametric deconvolution. C. R. Math. Acad. Sci. Paris 342 (11), 877–882.

Neumann, M. H., 1997. On the effect of estimating the error density in nonparametric deconvolution. J. Nonparametr. Statist. 7 (4), 307–330.