C. R. Acad. Sci. Paris, Ser. I 348 (2010) 323–326

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

Pointwise deconvolution with unknown error distribution

Déconvolution ponctuelle avec distribution de l'erreur inconnue

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*

ARTICLE INFO

Article history:
Received 6 July 2009
Accepted after revision 10 February 2010
Available online 26 February 2010

Presented by Paul Deheuvels

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

Cette Note présente les vitesses de convergence pour le risque quadratique ponctuel dans le problème de déconvolution avec fonction caractéristique des erreurs estimée.
© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

Let us consider the following model:

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n \tag{1}$$

where $(X_j)_{1\leqslant j\leqslant n}$ and $(\varepsilon_j)_{1\leqslant j\leqslant n}$ 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 $Y_1, \dots, Y_n$ 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 $\varepsilon_{-1}, \dots, \varepsilon_{-M}$, a noise sample with distribution f_ε, independent of $(Y_1, \dots, Y_n)$. Note that the availability of two distinct samples makes the problem identifiable.

Although 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 [6] in a context of circular data and [5] who examine the case $M \geqslant n$. [10] gives an upper bound and a lower bound for the integrated risk in the case where both f and f_ε are ordinary smooth, and [8] gives upper bounds for the integrated risk in a larger context of regularities. An other practical issue to the considered problem is the study of the model of repeated observations, see [4].

The contribution of this Note is to provide a class of estimators and compute upper bounds for their pointwise rates of convergence depending on M and n in a general setting.

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 $t^*(u) = \int e^{-ixu} t(x)\,dx$.

E-mail addresses: fabienne.comte@parisdescartes.fr (F. Comte), claire.lacour@math.u-psud.fr (C. Lacour).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2010.02.012

2. Estimation procedure

It easily follows 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:

(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^*: $\hat{f}_Y^*(u) = n^{-1}\sum_{j=1}^n e^{-iuY_j}$. Indeed, if f_ε^* is known, we can use the 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: $(2\pi)^{-1}\int_{|u|\leqslant \pi m} e^{iux}\hat{f}_Y^*(u)/f_\varepsilon^*(u)\,\mathrm{d}u$. Here m is a real positive bandwidth parameter. This estimator is classical in the sense that it corresponds both to a kernel estimator built with the sinc kernel (see [1]) or to a projection type estimator as in [3].

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 [10] the truncated estimator:

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

Then our estimator is

$$\hat{f}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)}\,\mathrm{d}u. \tag{2}$$

3. Study of the pointwise mean squared error

We introduce the notations

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

Proposition 3.1. *Consider model* (1) *under* (A1)*, then there exist constants $C, C' > 0$ such that for all positive real m and all positive integers n, M,*

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

Note that the result of Proposition 3.1 holds for any fixed and independent integers M and n.

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

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

$$\forall x \in \mathbb{R} \quad k_0\left(x^2+1\right)^{-\gamma/2}\exp\left(-b|x|^s\right) \leqslant \left|f_\varepsilon^*(x)\right| \leqslant k_1\left(x^2+1\right)^{-\gamma/2}\exp\left(-b|x|^s\right).$$

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

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

with $r \geqslant 0$, $a > 0$, $\delta \in \mathbb{R}$ 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. For example normal ($r = 2$) and Cauchy ($r = 1$) densities are supersmooth.

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 $\mathbb{E}[(\hat{f}_{m_0}(x) - f(x))^2]$ are given in Table* 1 (*which also contains the chosen m_0*).

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

	$s = 0$	$s > 0$
$r = 0$	$n^{-\frac{2\delta-1}{2\delta+2\gamma}} + M^{-[\min(1,\frac{2\delta-1}{2\gamma})]}(\log M)^{\mathbb{1}_{\delta=\gamma+1/2}}$ for $m_0 = \min(n^{1/(2\delta+2\gamma)}, M^{1/\max(2\gamma,2\delta-1)})$	$(\log n)^{-(2\delta-1)/s} + (\log M)^{-(2\delta-1)/s}$ for $m_0 = \pi^{-1}(\log(\min(n, M))/(2b+1))^{1/s}$
$r > 0$	$\frac{(\log n)^{(2\gamma+1)/r}}{n} + \frac{1}{M}$ for $m_0 = \pi^{-1}[(\log(n) - (1 + 2(\delta+\gamma)/r)\log\log(n))/(2a)]^{1/r}$	See comment in text.

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

$$2\left(\frac{1}{2\pi}\int_{|t|\geqslant \pi m}\left|f^*(t)\right|\mathrm{d}t\right)^2 \leqslant K_1(\pi m)^{-2\delta+1-r}\exp\left(-2a(\pi m)^r\right)$$

and straightforward computation gives $\Delta(m) \leqslant K_2(\pi m)^{2\gamma+1-s}\exp(2b(\pi m)^s)$ and $\Delta^0(m) \leqslant K_3(\pi m)^{2\gamma+2-2s}\exp(2b(\pi m)^s)$; lastly, denoting by $v = 2\gamma+1-s$, we have

$$\Delta_f^0(m)K_4^{-1} \leqslant (\pi m)^{(2\gamma+1-2\delta)_+}\left(\log(m)\right)^{\mathbb{1}_{\delta=\gamma+1/2}}\mathbb{1}_{\{r=s=0\}} + (\pi m)^{v-\max(2\delta,s-1)}\exp\left(2b(\pi m)^s\right)\mathbb{1}_{\{s>r\}}$$
$$+ (\pi m)^{v-2\delta}\exp\left(2(b-a)(\pi m)^s\right)\mathbb{1}_{\{r=s,b\geqslant a\}} + \mathbb{1}_{\{r>s\}\cup\{r=s,b<a\}}$$

where K_1, K_2, K_3, K_4 are positive constants. Then the rates of Table 1 are obtained by choosing adequate m_0 depending on n, M and the smoothness indices.

For the case $(r > 0, s > 0)$, the rules for the compromise between supersmooth terms in both squared bias and variance are given in [9] in the case of a known noise. The computations are similar for the present study. As this case is very tedious to write and contains several sub-cases, we omit the precise rates: it is sufficient to know that they decrease faster than any logarithmic functions, both in M and n.

The rates in term of n are known to be the optimal one for the deconvolution with known error (see [7] and [1]). They are recovered as soon as $M \geqslant n$. Extending the proof of [10], we can prove the optimality of the rate M^{-1} in the cases where f is smoother than f_ε and $r \leqslant 1$. Note that even for $M \geqslant n$, automatic selection of m should be performed in the spirit of [2], but none of the quoted works proves theoretical results about it.

Notice that Corollary 3.2 has not only a theoretical importance but also provides an answer to practical problems of noised observations by studying in detail the effect of preliminary measurements.

4. Proof of Proposition 3.1

First, let us denote $f_m(x) = (2\pi)^{-1}\int_{-\pi m}^{\pi m} e^{ixu} f^*(u)\,\mathrm{d}u$ and $R(x) = ((\tilde{f}_\varepsilon^*(x))^{-1} - (f_\varepsilon^*(x))^{-1})$. Then

$$\mathbb{E}\left[\left(\hat{f}_m(x) - f(x)\right)^2\right] \leqslant 2\left(f_m(x) - f(x)\right)^2 + 2\mathbb{E}\left[\left(\hat{f}_m(x) - f_m(x)\right)^2\right]$$
$$\leqslant 2\left(f_m(x) - f(x)\right)^2 + 4\mathrm{Var}\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(-u)}\,\mathrm{d}u\right)$$
$$+ 4\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\hat{f}_Y^*(u)R(u)\,\mathrm{d}u\right)^2\right]. \tag{4}$$

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

$$\left(f_m(x) - f(x)\right)^2 \leqslant \left(\frac{1}{2\pi}\int_{|t|\geqslant\pi m}\left|f^*(t)\right|\mathrm{d}t\right)^2. \tag{5}$$

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 [2] that

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

For the last remaining term in the right-hand side of (4), we bound it by

$$2\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\big(\hat{f}_Y^*(u)-f_Y^*(u)\big)R(u)\,\mathrm{d}u\right)^2\right]+2\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}f_Y^*(u)R(u)\,\mathrm{d}u\right)^2\right]:=2T_1+2T_2.$$

Neumann [10] has proved that there exists a positive constant C_1 such that

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

Then we find

$$T_1=\frac{1}{4\pi^2}\iint e^{ix(u-v)}\mathrm{Cov}\big(\hat{f}_Y^*(u),\hat{f}_Y^*(v)\big)\mathbb{E}\big(R(u)\bar{R}(v)\big)\,\mathrm{d}u\,\mathrm{d}v$$
$$\leqslant\frac{1}{4\pi^2 n}\iint\big|f_Y^*(u-v)\big|\sqrt{\mathbb{E}\big(\big|R(u)\big|^2\big)\mathbb{E}\big(\big|R(v)\big|^2\big)}\,\mathrm{d}u\,\mathrm{d}v\leqslant\frac{C_1}{4\pi^2 n}\iint\frac{|f_Y^*(u-v)|}{|f_\varepsilon^*(u)f_\varepsilon^*(v)|}\,\mathrm{d}u\,\mathrm{d}v.$$

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

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

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

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

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

$$T_2\leqslant\frac{1}{4\pi^2}\iint_{|u|,|v|\leqslant\pi m}\big|f_Y^*(u)f_Y^*(v)\big|\sqrt{\mathbb{E}\big(\big|R(u)\big|^2\big)\mathbb{E}\big(\big|R(v)\big|^2\big)}\,\mathrm{d}u\,\mathrm{d}v$$
$$\leqslant\frac{1}{4\pi^2}\left(\int_{-\pi m}^{\pi m}\big|f_Y^*(u)\big|\sqrt{\mathbb{E}\big(\big|R(u)\big|^2\big)}\,\mathrm{d}u\right)^2\leqslant\frac{C_1}{4\pi^2 M}\left(\int_{-\pi m}^{\pi m}\frac{|f_Y^*(u)|}{|f_\varepsilon^*(u)|^2}\,\mathrm{d}u\right)^2=C_1\frac{\Delta_f^0(m)}{2\pi M}.\tag{8}$$

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

References

[1] C. Butucea, Deconvolution of supersmooth densities with smooth noise, Canad. J. Statist. 32 (2) (2004) 181–192.
[2] C. Butucea, F. Comte, Adaptive estimation of linear functionals in the convolution model and applications, Bernoulli 15 (1) (2009) 69–98.
[3] F. Comte, Y. Rozenholc, M.-L. Taupin, Penalized contrast estimator for adaptive density deconvolution, Canad. J. Statist. 34 (3) (2006) 431–452.
[4] A. Delaigle, P. Hall, A. Meister, On deconvolution with repeated measurements, Ann. Statist. 36 (2) (2008) 665–685.
[5] P.J. Diggle, P. Hall, A Fourier approach to nonparametric deconvolution of a density estimate, J. Roy. Statist. Soc. Ser. B 55 (2) (1993) 523–531.
[6] S. Efromovich, Density estimation for the case of supersmooth measurement error, J. Amer. Statist. Assoc. 92 (438) (1997) 526–535.
[7] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist. 19 (3) (1991) 1257–1272.
[8] J. Johannes, Deconvolution with unknown error distribution, Ann. Statist. 37 (5A) (2009) 2301–2323.
[9] C. Lacour, Rates of convergence for nonparametric deconvolution, C. R. Acad. Sci. Paris, Ser. I 342 (11) (2006) 877–882.
[10] M.H. Neumann, On the effect of estimating the error density in nonparametric deconvolution, J. Nonparametr. Stat. 7 (4) (1997) 307–330.