Pointwise deconvolution with unknown error distribution

Fabienne Comte, Claire Lacour

To cite this version:


HAL Id: hal-00533627
https://hal.archives-ouvertes.fr/hal-00533627
Submitted on 5 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Pointwise deconvolution with unknown error distribution

Fabienne Comte\textsuperscript{a}, Claire Lacour\textsuperscript{b}

\textsuperscript{a}MAP5, UMR 8145, Universit\'e Paris Descartes, 45 rue des Saints-P\`eres, 75006 Paris, France
\textsuperscript{b}Laboratoire de math\'ematiques, Universit\'e 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:

\[ Y_j = X_j + \epsilon_j \quad j = 1, \ldots, n \]  

(1)

where \((X_j)_{1 \leq j \leq n}\) and \((\epsilon_j)_{1 \leq j \leq n}\) are independent sequences of i.i.d. variables. We denote by \(f\) the density of \(X_j\) and by \(f_\epsilon\) the density of \(\epsilon_j\). 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 \(\epsilon_{-1}, \ldots, \epsilon_{-M}\), a noise sample with distribution \(f_\epsilon\), independent of \((Y_1, \ldots, Y_n)\). Note that the availability of two distinct samples makes the problem identifiable.

Although there exists a huge literature concerning the estimation of \(f\) when \(f_\epsilon\) is known, this problem without the knowledge of \(f_\epsilon\) has been less studied. One can cite Efromovich (1997) in a context of circular data and Diggle and Hall (1993) who examine the case \(M \geq n\). Neumann (1997) gives an upper bound and a lower bound for the integrated risk in the case where both \(f\) and \(f_\epsilon\) are ordinary smooth, and Johannes (2009) 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 Delaigle et al (2008).

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 \(L^1 \cap L^2(\mathbb{R})\), we denote by \|t\| \(L^2\) norm of \(t\) and by \(\|t\|_1\) \(L^1\) norm of \(t\). The Fourier transform \(\hat{t}\) of \(t\) is defined by \(\hat{t}(u) = \int e^{-ixu}t(x)dx\).

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 \(f_Y = f * f_\epsilon\) and thus \(f_Y^\ast = f^\ast f_\epsilon^\ast\). Therefore, under the classical assumption:
the equality \( f^* = f_1^*/f_2^* \) yields an estimator of \( f^* \) by considering the following estimate of \( f_1^* \): 
\[
\hat{f}_1^*(u) = n^{-1} \sum_{j=1}^n e^{-i\pi y_j}.
\]
Indeed, if \( f_2^* \) is known, we can use the estimate of \( f^* \): 
\[
\hat{f}_2^*(u) = \hat{f}_1^*(u)/f_2^*(u).
\]
Then, we should use inverse Fourier transform to get an estimate of \( f \). As \( 1/f^* \) 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 
\[
M = \frac{1}{n}\sum_{j=1}^n e^{-i\pi y_j}
\]
is a real positive bandwidth. 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, \( f_2^* \) is unknown and we have to estimate it. Therefore, we use the preliminary sample and we define the natural estimator of \( f_2^* \): 
\[
\hat{f}_2^*(x) = \frac{1}{M} \sum_{j=1}^M e^{-i\pi x_j}.
\]
Next, we introduce as in Neumann (1997) the truncated estimator:
\[
\frac{1}{f_2^*(x)} = \frac{1}{f_2^*(x)} \frac{|f_2^*(x)| \geq M^{-1/2}}{f_2^*(x)} = \begin{cases} 
\frac{1}{f_2^*(x)} & \text{if } |f_2^*(x)| \geq M^{-1/2} \\
0 & \text{otherwise.}
\end{cases}
\]
Then our estimator is
\[
\hat{f}_m(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\hat{f}_1^*(u)}{f_2^*(u)} \, du.
\]

3. Study of the pointwise mean squared error

We introduce the notations
\[
\Delta(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_2^*(u)|^{-2} \, du, \quad \Delta^0(m) = \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} |f_2^*(u)|^{-1} \, du \right)^2, \quad \Delta^0(m) = \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} \left| f_2^*(u) \right| \, du \right)^2.
\]

**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 \( n, M \),
\[
E[(\hat{f}_m(x) - f(x))^2] \leq 2 \left( \frac{1}{2\pi} \int_{|t| \geq \pi} |f^*(t)| \, dt \right)^2 + \frac{C}{n} \min(\|f_2^*\|_1, \Delta(m), \Delta^0(m)) + C' \frac{\Delta^0(m)}{M}.
\]

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 \( f_2^* \):

(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
\[
\forall x \in \mathbb{R} \quad k_0 (x^2 + 1)^{-\gamma/2} \exp(-b|x|^\gamma) \leq |f_2^*(x)| \leq k_1 (x^2 + 1)^{-\gamma/2} \exp(-b|x|^\gamma).
\]

Moreover, the density function \( f \) to estimate generally belongs to the following type of smoothness spaces:
\[
A_{\delta,r,a}(l) = \{ f \text{ density on } \mathbb{R} \text{ and } \int |f^*(x)|^2 (x^2 + 1)^{l} \exp(2a|x|^r) \, dx \leq l \}
\]
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 \( f_2 \)) 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 \( f_2^* \) satisfies (A2) and if \( f \in A_{\delta,r,a}(l) \), the rates of convergence for the Mean Squared Error \( E[(\hat{f}_m(x) - f(x))^2] \) are given in Table 1 (which also contains the chosen \( m_0 \)).
Indeed, if \( f \in \mathcal{A}_{\delta,r,a}(l) \), the bias term can be bounded in the following way

\[
2 \left( \frac{1}{2\pi} \int_{|t| \geq \pi m} |f^*(t)| dt \right)^2 \leq K_1(\pi m)^{-25+1-r} \exp(-2a(\pi m)^r).
\]

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

\[
\Delta_\delta^0(m) K_4^{-1} \leq (\pi m)^{2\gamma+1-2\delta} \log(min(n, M)) \text{ for } m_0 = \min(n^{1/(2\gamma+2)} \max(2\gamma, 2\delta-1), M^{1/(\max(2\gamma, 2\delta-1))}).
\]

Indeed, if \( f^* \) 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 bias and variance are given in Lacour (2006) 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 the smoothness indices.

<table>
<thead>
<tr>
<th>( s )</th>
<th>( r = 0 )</th>
<th>( s &gt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r = 0 )</td>
<td>( n^{-\delta/r+1} + M^{-\min(1, 2\delta/r-1)} \log M { \log (n-M) }^{1/2} ) for ( m_0 = \min(n^{1/(2\gamma+2)} \max(2\gamma, 2\delta-1), M^{1/(\max(2\gamma, 2\delta-1))}) )</td>
<td>( (\log n)^{-\delta/(2\gamma-1)/r} + \log M ) for ( m_0 = \pi^{-1}\log(n-M)/(2b+1) )</td>
</tr>
<tr>
<td>( r &gt; 0 )</td>
<td>( \frac{(\log n)^{2\gamma+1/r}}{n} + \frac{1}{M} ) for ( m_0 = \pi^{-1}\log(n-M)/(2b+1) )</td>
<td>See comment below.</td>
</tr>
</tbody>
</table>

Table 1: Rates of convergence for the MSE if \( f^* \) 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 bias and variance are given in Lacour (2006) 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 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 \( f_x \) and \( r \leq 1 \). Note that even for \( M \geq n \), automatic selection of \( m \) should be performed in the spirit of Butucea and Comte (2009), 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 \( f_m(x) = (2\pi)^{-1} \int_{-\pi m}^{\pi m} e^{ixu} f^*(u) du \) and \( R(x) = (f^*_x(x))^{-1} - (f^*_x(x))^{-1} \). Then

\[
E[(f_m(x) - f(x))^2] \leq 2(f_m(x) - f(x))^2 + 2E[(f_m(x) - f(x))]^2
\]

\[
\leq 2(f_m(x) - f(x))^2 + 4 \left[ \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{ixu} \frac{f^*_x(u)}{f_x^*(-u)} du \right] + 4E \left[ \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{ixu} f^*_x(u) R(u) du \right] \quad (4)
\]

Since \( (f - f_m)(x) = (1/2\pi)(f^* - f_m^*)^*(-x) \), we can bound the bias 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 $f'_c$ is known and has already been studied: it follows from Butucea and Comte (2009) that
\[
\text{Var}\left(\frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{iux} \hat{f}'_c(u) du\right) \leq \frac{1}{2\pi n} \min(\|f'_c\|_1 \Delta(m), \Delta^0(m)). \tag{6}
\]
For the last remaining term in the right-hand-side of (4), we bound it by
\[
2\mathbb{E}\left[\left(\frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{iux} (\hat{f}'_c(u) - f'_c(u)) R(u) du\right)^2\right] + 2\mathbb{E}\left[\left(\frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{iux} f'_c(u) R(u) du\right)^2\right] := 2T_1 + 2T_2.
\]
Neumann (1997) proved that there exists a positive constant $C_1$ such that
\[
\mathbb{E}[|R(u)|^2] = \mathbb{E}\left(\left|\frac{1}{f'_c(u)} - \frac{1}{f'_c(u)}\right|^2\right) \leq C_1 \min\left(\frac{1}{|f'_c(u)|^2}, \frac{1}{M|f'_c(u)|^2}\right).
\]
Then we find
\[
T_1 = \frac{1}{4\pi^2} \int e^{ix(u-v)} \text{Cov}(f'_c(u), f'_c(v)) \mathbb{E}(R(u)R(v)) dudv 
\leq \frac{1}{4\pi^2n} \int |f'_c(u-v)| \sqrt{\mathbb{E}[|R(u)|^2]} \mathbb{E}[|R(v)|^2] dudv \leq \frac{C_1}{4\pi^2n} \int |f'_c(u-v)| \mathbb{E}(R(u)R(v)) dudv.
\]
This term is clearly bounded by $C_1 (2\pi n)^{-1} \|f'_c\|_1 \Delta^0(m)$. Moreover writing it as
\[
\frac{C_1}{4\pi^2n} \int \sqrt{|f'_c(u-v)|} \sqrt{|f'_c(u-v)|} dudv
\]
and using the Schwarz Inequality, and the Fubini Theorem yields the bound $C_1 (2\pi n)^{-1} \|f'_c\|_1 \Delta^0(m)$. Therefore
\[
\mathbb{E}\left[\left(\frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{iux} (\hat{f}'_c(u) - f'_c(u)) R(u) du\right)^2\right] \leq \frac{C_1}{2\pi n} \min(\|f'_c\|_1 \Delta(m), \Delta^0(m)), \tag{7}
\]
and thus it has the same order as the usual variance term. Lastly,
\[
T_2 \leq \frac{1}{4\pi^2} \int_{|u|, |v| \leq \pi m} |f'_c(u)f'_c(v)| \sqrt{\mathbb{E}[|R(u)|^2]} \mathbb{E}[|R(v)|^2] dudv 
\leq \frac{1}{4\pi^2} \int_{|u| \leq \pi m} |f'_c(u)| \sqrt{\mathbb{E}[|R(u)|^2]} du \leq \frac{C_1}{4\pi^2M} \left(\int_{-\pi m}^{\pi m} |f'_c(u)|^2 du\right)^2 = C_1 \frac{\Delta^0(m)}{2\pi M}. \tag{8}
\]
Inserting the bounds (5) to (8) in Inequality (4), we obtain the result of Proposition 3.1.

\[
\square
\]

References