Nonparametric methods for the convolution model.

JOURNÉES MAS — 27-29 AOÛT 2008 — RENNES.

by F. Comte*.

* MAP5, UMR 8145, Université Paris Descartes.
MODEL.

\[ Z_i = X_i + \varepsilon_i, \quad i = 1, \ldots, n. \]

Generally:
- \((\varepsilon_i)_{1 \leq i \leq n}\) are i.i.d., density \(f_{\varepsilon}\),
- \((X_i)_{1 \leq i \leq n}\) are i.i.d., density \(f_X\),
- The sequences \((\varepsilon_i)\) and \((X_i)\) are independent.

Aim:

**Estimate** \(f_X\) when only \(Z_1, \ldots, Z_n\) is observed (not \(X_1, \ldots, X_n\)).

Identifiability: \(f_{\varepsilon}\) is known.
Plan of the talk.

I. In the model: A. Estimators (kernel, projection)  
                              B. Rates and optimality  
                              C. Adaptive strategy and adaptive rates.

II. Extensions of the model:

A. Convolution models with dependence structure  
B. The density $f_\varepsilon$ is:  
   * not entirely known (semiparametric models)  
   * unknown $\rightarrow$ repeated observations  
      $\rightarrow$ preliminary observation of the noise.
Why convolution model?

Because what is easy to estimate is $f_Z = f_X \ast f_{\epsilon}$.

**Notations:**

$u \ast v(x) = \int u(y)v(x - y)dy,$

$u^*(x) = \int e^{ixt}u(t)dt, \|u\|^2 = \int |u(x)|^2dx, \langle u, v \rangle = \int u(x)\overline{v(x)}dx.$

Recall that for $u, v \in L^1 \cap L^2(\mathbb{R})$,

$(u \ast v)^* = u^*v^*, \langle u, v \rangle = \frac{1}{2\pi}\langle u^*, v^* \rangle, (u^*)^*(x) = 2\pi u(-x).$
Therefore \( f_Z^* = f_X^* f_\varepsilon^* \Rightarrow f_X(x) = \frac{1}{2\pi} \left( \frac{f_Z^*}{f_\varepsilon^*} \right)^* (\varepsilon - x). \)

That is \( f_X(x) = \frac{1}{2\pi} \int e^{-ixt} \frac{f_Z^*(t)}{f_\varepsilon^*(t)} dt. \)

where \( f_\varepsilon \) is known and \( f_Z^* \) is easy to estimate:

\[
\hat{f}_Z^*(t) = \frac{1}{n} \sum_{k=1}^{n} e^{itZ_k}.
\]

But now, think of \( f_\varepsilon^*(x) = e^{-x^2/2} \) in the denominator and solve the problem of integrability.
Kernel estimator with bandwidth $h_n$:

$$\hat{f}_n(x) = \frac{1}{2\pi} \int e^{-itx} K^*(th_n) \frac{\hat{f}_Z^*(t)}{f_\varepsilon^*(t)} dt.$$

Other formulation:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} L_n \left( \frac{x - Z_j}{h_n} \right), \quad L_n(x) = \frac{1}{2\pi} \int e^{itx} \frac{K^*(t)}{f_\varepsilon^*(t/h_n)} dt.$$
References of the “first wave”


Take $K(x) = \frac{\sin(\pi x)}{\pi x}$, $K^*(x) = \mathbb{1}_{|x| \leq 1}$.

Write the decomposition:

$$2\pi (f_X - \hat{f}_n)(x) = \int e^{-itx} f_X^*(t) dt - \int e^{-itx} \mathbb{1}_{|t| \leq 1/h_n} \frac{\hat{f}^*_Z(t)}{f^*_\varepsilon(t)} dt$$

$$= \int_{|t| \geq 1/h_n} e^{-itx} f_X^*(t) dt + \int_{|t| \leq 1/h_n} e^{-itx} \frac{\hat{f}^*_Z(t) - \hat{f}^*_Z(t)}{f^*_\varepsilon(t)} dt$$

Two quantities determine the rate of convergence:

* the rate of decrease of $f_X^*$ for the bias,
* the rate of decrease of $f^*_\varepsilon$ for the variance.
Nonstandard case ⇒ nonstandard rates.

Assumptions on $f^*_\varepsilon$:

$\exists s \geq 0, b \geq 0, \gamma \in \mathbb{R}$ ($\gamma > 0$ if $s = 0$) and $k_0, k_1 > 0$:

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

$f_X$ belongs to

$$A_{\delta,r,a}(L) = \{ f \text{ density on } \mathbb{R} \text{ and } \int |f^*(x)|^2(x^2 + 1)^\delta \exp(2a|x|^r) \leq L \}$$

with $r \geq 0, a \geq 0, \delta \in \mathbb{R}$ ($\delta > 1/2$ if $r = 0$), $L > 0$.

$r > 0$ : function supersmooth, $r = 0$ : ordinary smooth.
First study supersmooth $f_X$:


\[
\text{MSE} = \mathbb{E}[(\hat{f}_n(x) - f_X(x))^2],
\]

<table>
<thead>
<tr>
<th>$s = 0$</th>
<th>$s &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 0$</td>
<td>$\frac{1-2\delta}{n^{2\delta+2\gamma}}$</td>
</tr>
<tr>
<td>$r &gt; 0$</td>
<td>$(\ln n)^{\frac{2\gamma+1}{r}}$</td>
</tr>
</tbody>
</table>

Rates: upper/lower bounds previous authors for OS $f_X$, Fan (1991)
MISE = \mathbb{E}\|\hat{f}_n - f_X\|^2 = \mathbb{E}\left(\int |\hat{f}_n(x) - f_X(x)|^2 \, dx\right).

\begin{array}{ccc}
\hline
& s = 0 & s > 0 \\
\hline
r = 0 & n^{-\frac{2\delta}{2\delta + 2\gamma + 1}} & (\ln n)^{-\frac{2\delta}{s}} \\
r > 0 & (\ln n)^{\frac{2\gamma + 1}{r}} & \text{see below} \\
\hline
\end{array}

Rates of convergence for the MISE.
The rates of convergence in the case \((r > 0, s > 0)\) are difficult.

Lower bound on implicit definitions: Butucea and Tsybakov (2007)


Rates depend on the integer \(k\) such that \(r/s\) or \(r/s \in \left( \frac{k}{k+1}, \frac{k+1}{k+2} \right)\):

Assume \(r > 0\) and \(s > 0\). Let \(k \in \mathbb{N}\) and \(\lambda = \mu^{-1} = r/s\). Then

- if \(r = s\), \(MISE = O \left( n^{-\frac{a}{a+b}} (\ln n)^{-\frac{2\delta b + (s-2\gamma-1)a}{(a+b)s}} \right)\);

\[
MSE = O \left( n^{-\frac{a}{a+b}} (\ln n)^{-\frac{2\delta - (1-s)_+ b + (s-2\gamma-1)a}{(a+b)s}} \right)
\]
• if \( r < s \) and \( \frac{k}{k+1} < \lambda \leq \frac{k+1}{k+2} \), there exist reals \( b_i \) such that

\[
MISE = O \left( (\ln n)^{-\frac{2\delta}{s}} \exp\left[ \sum_{i=0}^{k} b_i (\ln n)^{(i+1)\lambda-i} \right] \right);
\]

\[
MSE = O \left( (\ln n)^{-\frac{2\delta-r+1}{s}} \exp\left[ \sum_{i=0}^{k} b_i (\ln n)^{(i+1)\lambda-i} \right] \right)
\]

• if \( r > s \) and \( \frac{k}{k+1} < \mu \leq \frac{k+1}{k+2} \) ...
Bandwidth selection.


Projection estimator and model selection
(Comte, Rozenholc, Taupin (2006, 2007)).

\[ \varphi(x) = \frac{\sin(\pi x)}{(\pi x)}, \]

and let \( \varphi_{m,j}(x) = \sqrt{m} \varphi(mx - j), m \in \mathcal{M}_n = \{1, \ldots, m_n\}. \)

\[ S_m = \text{span}(\varphi_{m,j}, j \in \mathbb{Z}) = \{ f, \text{supp}(f^*) \subset [-\pi m, \pi m] \} \]

\( (S_m)_{m \in \mathcal{M}_n} \) collection of linear spaces

Let \( t \in S_m : \)

\[ \gamma_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left[ \|t\|^2 - 2u_\gamma^*(Z_i) \right], \text{ with } u_t(x) = \frac{1}{2\pi} \left( \frac{t^*(-x)}{f_\epsilon^*(x)} \right). \]
\[ \hat{f}_m = \arg \min_{t \in S_m^{(n)}} \gamma_n(t). \]

\[ \hat{f}_m = \sum_{|j| \in \mathbb{Z}} \hat{a}_{m,j} \varphi_{m,j} \text{ where } \hat{a}_{m,j} = \frac{1}{n} \sum_{i=1}^{n} u_{\varphi_{m,j}}^*(Z_i) \]

and \( \mathbb{E}(\hat{f}_m) = f_m \) the orthogonal projection of \( f \) on \( S_m \).

Note that \( f_m^* = f \mathbb{I}_{[-\pi m, \pi m]} \).

Same MISE as the kernel estimator. \( m \leftrightarrow h_n^{-1}. \)
Define \( \Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\epsilon^*(x)|^2 \, dx \),

\[
\text{pen}(m) = K(\pi m)[s-(1-s)+/2] + \frac{\Delta(m)}{n},
\]

and \( \hat{m} = \arg \min_{m \in \mathcal{M}_n} \gamma_n(\hat{f}_m) + \text{pen}(m) \).

Then

\[
\mathbb{E}(\|f_X - \hat{f}_{\hat{m}}\|^2) \leq C \inf_{m \in \mathcal{M}_n} \left\{ \|f_X - f_m\|^2 + \text{pen}(m) \right\} + \frac{C}{n}.
\]

Rates: a loss occurs only if \( s > 1/3 \) but it is negligible w.r.t. the rate.
Butucea and Comte (2007) study an adaptive pointwise strategy and complete the study of optimality when some loss occur. Pointwise model selection is rather different.
Why should we consider dependent contexts?

Case of particular Hidden Markov Models, when the noise is additive, and \((X_i)\) is a \(\beta\)-mixing Markov process.

Famous example. The discrete time Stochastic Volatility Model:

\[ r_t = \sigma_t \eta_t \Rightarrow \log(r_t^2) = \log(\sigma_t^2) + \varepsilon_t, \quad \varepsilon_t = \log(\eta_t^2). \]

**Dependence Assumptions, Case 1.**

The sequences \((X_i)\) and \((\varepsilon_i)\) are independent and the \(\varepsilon_i\) are i.i.d.

**The sequence \((X_i)\) is strongly stationary and \(\beta\)-mixing.** (or \(\alpha\), or \(\tau\)-mixing).
ARCH models. Let \((\eta_i)\) be an i.i.d. noise sequence.

\[ Y_i = \sigma_i \eta_i \text{ with } \sigma_i = F(\eta_{i-1}, \eta_{i-2}, \ldots), \]

for some measurable functions \(F\), or

\[ Y_i = \sigma_i \eta_i \text{ with } \sigma_i = F(\sigma_{i-1}, \eta_{i-1}) \text{ and } \sigma_0 \text{ independent of } (\eta_i)_{i \geq 0}. \]

**Dependence Assumption, case 2.**
The \(\varepsilon_i\) are i.i.d and for any given \(i\), \(X_i\) and \(\varepsilon_i\) are independent (but the sequences \((X_i)\) and \((\varepsilon_i)\) are not independent).

The sequence \((Z_i, X_i)_{i \in \mathbb{Z}}\) is strongly stationary and \(\beta\)-mixing.


Density estimation:

\[ \text{pen}(m) = \kappa \frac{D_m}{n} \quad \rightarrow \quad (\text{dependence}) \text{pen}(m) = \kappa \sum_{k \in \mathbb{N}} \beta_k \frac{D_m}{n} \]

A problem which is avoided in density \textit{deconvolution}. 
No longer knowing $f_\varepsilon^*$, new directions.

Semi-parametric strategies.


\[
Z_i = X_i + \sigma \varepsilon_i, \quad k_0 \exp(-|x|^s) \leq |f_\varepsilon^*(x)| \leq k_1 \exp(-|x|^s).
\]

\[
\Rightarrow F(\tau, x) = f_Z^*(x)e^{(\tau x)^s},
\]

\[
|F(\tau, x)| = O(1)|f_X^*(x)|e^{(\tau^s-\sigma^s)x^s} \begin{cases} 
0 & \text{if } \tau \leq \sigma, \\
+\infty & \text{if } \tau > \sigma.
\end{cases}
\]
$$\hat{\sigma}_n = \inf \left\{ \tau, \tau > 0, |\hat{F}(\tau, x_n)| \geq 1 \right\}, \quad \hat{F}(\tau, x) = \hat{f}_Z^*(x)e^{(\tau u)^s}.$$ 

$x_n$ is a well chosen positive sequence. $f_X$ is OS or SS with $r < s$.

Logarithmic but **optimal** rates for the estimation of $\sigma$, which also impose the rate of the plug-in estimator of $f_X$. 

\[ f_\varepsilon^*(x) = \exp(-|x|^s), \ s \in [\underline{s}, \bar{s}], \ |f_X^*(x)| \geq A|x|^{-\beta'}. \]

Their estimator is deduced from

\[ A|x|^{-\beta'} \exp(-|x|^s) \leq |f_Z^*(x)| \leq \exp(-|x|^s). \]

Then applied knowing that the true regularity of \( f_X \) is in \([\beta, \bar{\beta}]\).

Rates of estimation of \( s \) (exponential), of the plug-in estimator of \( f_X \) \( \Rightarrow \) optimal (but logarithmic).
Nonparametric strategy: repeated measurements.


Repeated observations: \(Z_{i,j} = X_i + \varepsilon_{i,j}, j = 1, \ldots, K\)

Idea: \(K = 2, Z_{i,1} - Z_{i,2} = \varepsilon_{i,1} - \varepsilon_{i,2} \Rightarrow \) if \(f_\varepsilon^*(t) > 0,\)

\[
\hat{f}_\varepsilon^*(t) = \left| \frac{1}{n} \sum_{i=1}^{n} \frac{1}{K(K - 1)/2} \sum_{j \neq l} \cos(t(Z_{i,j} - Z_{i,l})) \right|^{1/2}
\]

And plug-in the kernel estimator.
**Nonparametric strategy : preliminary noise observation.**

$\varepsilon_{-M}, \ldots, \varepsilon_{-1}, \varepsilon_0$ to estimate $f_{\varepsilon}^*$

Plug in the estimator (Kernel or projection)

$\rightarrow$ New tables of rates taking $M$ into account.

<table>
<thead>
<tr>
<th>$s = 0$</th>
<th>$s &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 0$</td>
<td>$\frac{2^{\delta-1}}{2^{\delta+2\gamma}} + M^{-[\text{1}(2^{\delta-1}/(2\gamma))] \log(M)^u}$</td>
</tr>
<tr>
<td>$u = 1_{\delta = \gamma + 1/2}$</td>
<td></td>
</tr>
</tbody>
</table>

$r > 0$ | $\frac{(\log n)^{2\gamma+1}}{n} + \frac{1}{M}$ | see the discussion below. |

Rates of convergence for the MSE.
Previous rates preserved if $M \geq n$.

**Study of smaller choices of $M$**

Problem: Model selection is difficult.

The collection of models must be such that

\[
\frac{1}{M} \int_{-\pi m}^{\pi m} \frac{1}{f_\varepsilon^*(x)} 2 dx
\]

is bounded.


Comte, F. and Lacour, C. Deconvolution with estimated characteristic function of the errors. Preprint MAP5 2008-14