Nonparametric methods for the convolution model.

JOURNÉES MAS — 27-29 AOÛT 2008 – RENNES.

by F. Comte*.

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

MODEL.

$$Z_i = X_i + \varepsilon_i, \;\; i = 1, \dots, n.$$

Generally :

- $(\varepsilon_i)_{1 \le i \le n}$ are i.i.d., density f_ε,
- $(X_i)_{1 \le i \le n}$ are i.i.d., density f_X,
- The sequences (ε_i) and (X_i) are independent.

Aim :

Estimate $\boldsymbol{f_X}$ when only $\boldsymbol{Z_1, \dots, Z_n}$ is observed (not $X_1, \dots, X_n$).

Identifiability : f_ε 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** $\boldsymbol{f_\varepsilon}$ is :

* not entirely known (semiparametric models)

* unknown $\longrightarrow$ repeated observations

$\longrightarrow$ preliminary observation of the noise.

Why convolution model ?

Because what is easy to estimate is $\boldsymbol{f_Z = f_X \star f_\varepsilon}$.

Notations : $u \star v(x) = \int u(y)v(x-y)dy$,

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

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

$$(u \star v)^* = u^* v^*, \langle u, \ \ v\rangle = \frac{1}{2\pi}\langle u^*, v^*\rangle, \ (u^*)^*(x) = 2\pi u(-x).$$

Therefore $\boldsymbol{f_Z^* = f_X^* f_\varepsilon^*} \Rightarrow \boldsymbol{f_X(x) = \frac{1}{2\pi}\left(\frac{f_Z^*}{f_\varepsilon^*}\right)^*(-x).}$

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

where f_ε is known and f_Z^* is easy to estimate :

$$\widehat{f_Z^*}(t) = \frac{1}{n}\sum_{k=1}^{n} e^{itZ_k}.$$

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

Kernel estimator with bandwidth h_n :

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

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"

R.J. Carroll, P. Hall (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83, 1184–1186

L. Devroye (1989). Consistent deconvolution in density estimation. *The Canadian Journal of Statistics*, 17, 235–239.

J. Fan (1991a). On the optimal rates of convergence for nonparametric deconvolution problem. *The Annals of Statistics*, 19, 1257–1272.

J. Fan (1991b). Asymptotic normality for deconvolution kernel estimators, *Sankhya Series A*, 53, 97–110.

M.C. Liu, R.L. Taylor (1989). A consistent nonparametric density estimator for the deconvolution problem. *The Canadian Journal of Statistics*, 17, 427–438.

L. Stefanski, R.J. Carroll (1990). Deconvoluting kernel density estimators. *Statistics*, 21, 1696-184.

C.H. Zhang (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics*, 18, 806-831.

Take $K(x) = \dfrac{\sin(\pi x)}{\pi x}$, $K^*(x) = 1\!\!{\rm I}_{|x|\le 1}$.

Write the decomposition :

$$2\pi(f_X - \hat{f}_n)(x) = \int e^{-itx} f_X^*(t)dt - \int e^{-itx} 1\!\!{\rm I}_{|t|\le 1/h_n} \frac{\widehat{f_Z^*}(t)}{f_\varepsilon^*(t)} dt$$

$$= \underbrace{\int_{|t|\ge 1/h_n} e^{-itx} f_X^*(t)dt}_{\text{bias}} + \underbrace{\int_{|t|\le 1/h_n} e^{-itx} \frac{f_Z^*(t) - \widehat{f_Z^*}(t)}{f_\varepsilon^*(t)} dt}_{\text{variance.}}$$

Two quantities determine the rate of convergence :
* the rate of decrease of f_X^* for the bias,
* the rate of decrease of f_ε^* for the variance.

Nonstandard case $\Rightarrow$ nonstandard rates.

Assumptions on f_ε^* :
$\exists s \geq 0, b \geq 0, \gamma \in \mathbb{R}$ ($\gamma > 0$ 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

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

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 :

M. Pensky & B. Vidakovic (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27, 2033–2053.

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

	$s = 0$	$s > 0$
$r = 0$	$n^{\frac{1-2\delta}{2\delta+2\gamma}}$	$(\ln n)^{\frac{1-2\delta}{s}}$
$r > 0$	$\frac{(\ln n)^{\frac{2\gamma+1}{r}}}{n}$	see below

Rates : upper/lower bounds previous authors for OS f_X, Fan (1991)
For SS f_X : Butucea (2004), Butucea and Tsybakov (2008).

$$\text{MISE} = \mathbb{E}\|\hat{f}_n - f_X\|^2 = \mathbb{E}\left(\int |\hat{f}_n(x) - f_X(x)|^2 dx\right).$$

	$s = 0$	$s > 0$
$r = 0$	$n^{-\frac{2\delta}{2\delta+2\gamma+1}}$	$(\ln n)^{-\frac{2\delta}{s}}$
$r > 0$	$\frac{(\ln n)^{\frac{2\gamma+1}{r}}}{n}$	see below

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)

Upper bounds and explicit studies :
Lacour (2006), Comte, Rozenholc, Taupin (2006).

Rates depend on the integer k such that
r/s or $r/s \in (k/(k+1), (k+1)/(k+2)]$:
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

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

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

- if $r > s$ and $\frac{k}{k+1} < \mu \leq \frac{k+1}{k+2}$ ...

Bandwidth selection .

A. Delaigle, I. Gijbels (2004a). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics and Data Analysis*, 45, 249–267.

A. Delaigle,I. Gijbels (2004b). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56, 19–47.

C. Hesse (1999) Data-driven deconvolution.
Journal of Nonparametric Statistics, 10, 343–373.

Projection estimator and model selection

(Comte, Rozenholc, Taupin (2006, 2007)).

$$\varphi(x) = \sin(\pi x)/(\pi x),$$

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

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

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

Let $t \in S_m$:

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\|t\|^2 - 2u_t^*(Z_i)\right], \quad \text{with} \quad u_t(x) = \frac{1}{2\pi}\left(\frac{t^*(-x)}{f_\varepsilon^*(x)}\right).$$

$$\hat{f}_m = \arg \min_{t \in S_m^{(n)}} \gamma_n(t).$$

$$\hat{\boldsymbol{f}}_{\boldsymbol{m}} = \sum_{|\boldsymbol{j}| \in \mathbb{Z}} \hat{\boldsymbol{a}}_{\boldsymbol{m},\boldsymbol{j}} \boldsymbol{\varphi}_{\boldsymbol{m},\boldsymbol{j}} \quad \text{where} \quad \hat{\boldsymbol{a}}_{\boldsymbol{m},\boldsymbol{j}} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{u}^*_{\boldsymbol{\varphi}_{\boldsymbol{m},\boldsymbol{j}}}(\boldsymbol{Z}_{\boldsymbol{i}})$$

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

Note that $f_m^* = f \mathbb{1}_{[-\pi m, \pi m]}$.

Same MISE as the kernel estimator. $m \leftrightarrow h_n^{-1}$.

Define $\Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{dx}{|f_\varepsilon^*(x)|^2}$,

$$\mathbf{pen}(\boldsymbol{m}) = \boldsymbol{K}(\boldsymbol{\pi m})^{[\boldsymbol{s}-(\mathbf{1}-\boldsymbol{s})_+/\mathbf{2}]_+}\frac{\boldsymbol{\Delta}(\boldsymbol{m})}{\boldsymbol{n}}$$

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

Then

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

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 β-mixing Markov process.

Famous example. The discrete time Stochastic Volatility Model :

$$r_t = \sigma_t \eta_t \Rightarrow \log(r_t^2) = \log(\sigma_t^2) + \varepsilon_t, \;\; \varepsilon_t = \log(\eta_t^2).$$

Dependence Assumptions, Case 1.
The sequences (X_i) and (ε_i) are independent
and the ε_i are i.i.d.
The sequence (X_i) is strongly stationary and β-mixing.
(or α, or τ-mixing).

ARCH models. Let (η_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 ε_i are i.i.d and for any given i, X_i and ε_i are independent

(but the sequences (X_i) and (ε_i) are not independent).

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

E. Masry (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory*, 37, 1105–1115.

B. Van Es, P. Spreij, and H. van Zanten. Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat.*, 17(2) :237–251, 2005.

F. Comte, J. Dedecker, and M.-L. Taupin. Adaptive density estimation for general arch models. To appear in *Econometric Theory*, 2008.

F. Comte, J. Dedecker, and M.-L. Taupin. Adaptive density deconvolution for dependent inputs with measurement errors. To appear in *Mathematical methods of Statistics*, 2008.

Density estimation :

$$\text{pen}(m) = \kappa \frac{D_m}{n} \longrightarrow (dependence)\text{pen}(m) = \kappa \sum_{k \in \mathbb{N}} \beta_k \frac{D_m}{n}$$

A problem which is avoided in density **deconvolution**.

No longer knowing f_ε^*, new directions.

Semi-parametric strategies.

C. Butucea, C. Matias, Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model, *Bernoulli* **11**(2), 309-340, 2005.

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

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

$$|F(\tau, x)| = O(1)|f_X^*(x)| e^{(\tau^s - \sigma^s)x^s} \longrightarrow \begin{cases} 0 & \text{if } \tau \le \sigma, \\ +\infty & \text{if } \tau > \sigma. \end{cases}$$

$$\hat{\sigma}_n = \inf\left\{\tau, \tau > 0, |\hat{F}(\tau, x_n)| \geq 1\right\}, \ \hat{F}(\tau, x) = \widehat{f_Z^*}(x)e^{(\tau u)^s}.$$

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 σ,

which also impose the rate of the plug-in estimator of f_X.

C. Butucea, C. Matias and C. Pouet, Adaptive procedures in convolution models with known or partially known noise distribution. Working paper, 2007.

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

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 $[\underline{\beta}, \bar{\beta}]$.
Rates of estimation of s (exponential) , of the plug-in estimator of f_X
$\Rightarrow$ optimal (but logarithmic).

Nonparametric strategy : repeated measurements.

Delaigle, A., Hall, P and Meister, A. On deconvolution with repeated measurements. *Ann. Statist.* 36(2) (2008), 665-685.

Neumann, M. H. Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* 98 (2007), no. 10, 1955–1968.

Carroll, R. J., Eltinge, J. L. and Ruppert, D. (1993). Robust linear regression in replicated measurement error models. *Statist. Probab. Lett.* 19, 169175.

Susko, E. and Nadon, R. (2002). Estimation of a residual distribution with small numbers of repeated measurements. *Canad. J. Statist.* 30, 383400.

Repeated observations : $Z_{i,j} = X_i + \varepsilon_{i,j}$, $j = 1, \dots, K$
Idea : $K = 2$, $Z_{i,1} - Z_{i,2} = \varepsilon_{i,1} - \varepsilon_{i,2} \Rightarrow$ if $f_\varepsilon^*(t) > 0$,

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

And plug-in the kernel estimator.

Nonparametric strategy : preliminary noise observation.

$\varepsilon_{-M}, \dots, \varepsilon_{-1}, \varepsilon_0$ to estimate f^*_ε
Plug in the estimator (Kernel or projection)
$\longrightarrow$ New tables of rates taking M into account.

	$s = 0$	$s > 0$
$r = 0$	$n^{-\frac{2\delta-1}{2\delta+2\gamma}} +$ $M^{-[1\wedge(2\delta-1/(2\gamma))]}\log(M)^u$ $u = \mathbb{1}_{\delta=\gamma+1/2}$	$(\log n)^{-\frac{2\delta-1}{s}} + (\log M)^{-\frac{2\delta-1}{s}}$
$r > 0$	$\frac{(\log n)^{\frac{2\gamma+1}{r}}}{n} + \frac{1}{M}$	see the discussion below.

Rates of convergence for the MSE.

Previous rates preserved if $\boldsymbol{M} \geq n$.

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.

Diggle, P. J. and 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.

Johannes, J. (2007). Deconvolution with unknown error distribution. *ArXiv*.

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

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