Workshop on inverse problems

Laguerre deconvolution
with unknown error distribution

Fabienne Comte[1]
Joint work with **Gwennaëlle Mabon**[2]

(1) *MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, France*

(2) *Humboldt-Universität zu Berlin
Institut für Mathematik , Berlin, Germany*

October 29th 2016

Model and questions

Convolution Model :

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n \tag{1}$$

- X_i and Y_i **nonnegative**, i.i.d.
- $(X_i)_{1\leqslant i\leqslant n}$ and $(Y_i)_{1\leqslant i\leqslant n}$ independent,
- X_i with **unknown** density f,
- Y_i with density g

Observations: $Z_1, \dots, Z_n$.

Question: estimate the density of the X_i's.

- $\hookrightarrow$ when g is known, Mabon (2015)
- $\hookrightarrow$ when g is unknown, additional sample $Y'_1, \dots, Y'_{n_0}$ i.i.d. with density g.

Bibliography

- Known noise density, book by Meister (2009).
  - Fourier and kernel methods: Carroll and Hall (1988), Fan (1991)
  - Model/bandwidth selection: Pensky and Vidakovic (1999), Comte, Taupin and Rozenholc (2006).
  - Rates and optimality: Fan (1991) on Hölder spaces, Butucea (2004), Butucea and Tsybakov (2008) on Sobolev spaces.
  - Multivariate setting: Comte and Lacour (2013), Rebelles (2015).
- Unknown noise density
  - Neumann (1997), Johannes (2009), Comte and Lacour (2011)
  - Kappus and Mabon (2014).

Problem: regularity on $\mathbb{R}$ of $\mathbb{R}^+$-supported functions.

Convolution on $\mathbb{R}^+$

Specifically here, X and Y are **nonnegative processes** (or lower bounded with known lower bound)

Denote by h the density of the observations,

$$\begin{array}{ccccc} Z & = & X & + & Y \\ \downarrow & & \downarrow & & \downarrow \\ h & & f & & g \end{array} \quad \Rightarrow \quad h(x) = \int_0^x f(u)g(x-u)du.$$

Principle:

- All functions are assumed to be square-integrable on $\mathbb{R}^+$ and developed in a relevant basis.
- Estimating f = estimating (part of) the coefficients of the development (projection estimator).

The Laguerre basis (1)

Specific basis : Laguerre functions $\forall k \in \mathbb{N}, \forall x \geqslant 0$,

$$\varphi_k(x) = \sqrt{2} L_k(2x) e^{-x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \frac{x^j}{j!}. \tag{2}$$

$(\varphi_k)_{k \geqslant 0}$ orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$.

Bounded basis: $\|\varphi_k\|_\infty \leqslant \sqrt{2}$.
For a function p in $\mathbb{L}^2(\mathbb{R}^+)$,

$$p(x) = \sum_{k \geqslant 0} a_k(p) \varphi_k(x) \quad \text{where} \quad a_k(p) = \int_{\mathbb{R}^+} p(u) \varphi_k(u) du.$$

Orthogonal projection of p on $\mathcal{S}_m = \mathrm{Span}\{\varphi_0, \ldots, \varphi_{m-1}\}$ is

$$p_m(x) = \sum_{k=0}^{m-1} a_k(p) \varphi_k(x).$$

The Laguerre basis (2)

Key property (Abramowitz-Stegun):

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u)\varphi_j(x-u)du = 2^{-1/2}\left(\varphi_{k+j}(x) - \varphi_{k+j+1}(x)\right) \qquad (3)$$

where $\star$ stands for the convolution product.

Model (1) implies that $h = f \star g$ and Formula (3) yields

$$\begin{aligned}\sum_{k=0}^{\infty} a_k(h)\varphi_k(x) &= \sum_{j=0}^{+\infty}\sum_{k=0}^{+\infty} a_j(f)a_k(g)\varphi_j \star \varphi_k(x) \\ &= \sum_{k=0}^{\infty} \varphi_k(x) \underbrace{\sum_{\ell=0}^{k} 2^{-1/2}\left(a_{k-\ell}(g) - a_{k-\ell-1}(g)\right)a_\ell(f)}_{}\end{aligned}$$

For any integer m, $\vec{h}_m = \mathbf{G}_m \vec{f}_m$, and

$$\vec{h}_\infty = \mathbf{G}_\infty \vec{f}_\infty,$$

with

$$\vec{h}_m = \begin{pmatrix} a_0(h) \\ \vdots \\ a_{m-1}(h) \end{pmatrix}, \vec{f}_m = \begin{pmatrix} a_0(f) \\ \vdots \\ a_{m-1}(f) \end{pmatrix},$$

and

$$[\mathbf{G}_m]_{i,j} = \begin{cases} 2^{-1/2} a_0(g) & \text{if } i = j, \\ 2^{-1/2} \left(a_{i-j}(g) - a_{i-j-1}(g)\right) & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

$\hookrightarrow$ Infinite linear **triangular** Toeplitz system

$$\Rightarrow \quad \forall m, \quad \vec{h}_m = \mathbf{G}_m \vec{f}_m.$$

Remark. $\mathbf{G}_m$ lower triangular with diagonal elements $a_0(g)/\sqrt{2}$,

$$a_0(g) = \int_{\mathbb{R}^+} g(u)\varphi_0(u)du = \sqrt{2}\int_{\mathbb{R}^+} g(u)e^{-u}du = \sqrt{2}\mathbb{E}[e^{-Y}] > 0.$$

$$\Rightarrow \quad \mathbf{G}_m \text{ is invertible and } \mathbf{G}_m^{-1}\vec{h}_m = \vec{f}_m.$$

Finally $a_k(h) = \mathbb{E}[\varphi_k(Z_1)]$, the projection of f on $\mathcal{S}_m$ can be estimated by

$$\hat{f}_m(x) = \sum_{k=0}^{m-1} \hat{a}_k\varphi_k(x) \quad \text{with} \quad \hat{\vec{f}}_m = \mathbf{G}_m^{-1}\hat{\vec{h}}_m, \tag{5}$$

$$\hat{\vec{f}}_m = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_{m-1} \end{pmatrix} \text{ and } \hat{\vec{h}}_m = \begin{pmatrix} \hat{a}_0(Z) \\ \hat{a}_1(Z) \\ \vdots \\ \hat{a}_{m-1}(Z) \end{pmatrix}, \quad \text{and} \quad \hat{a}_k(Z) = \frac{1}{n}\sum_{i=1}^{n}\varphi_k(Z_i). \tag{6}$$

Previous results, g known.

- **Risk bound** on $\hat{f}_m$ (Mabon 2015)

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \leqslant \|f - f_m\|^2 + \frac{1}{n}\left[(2m\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2) \wedge (\|h\|_\infty \|\mathbf{G}_m^{-1}\|_F^2)\right]$$

where $\|\mathbf{A}\|_{\mathrm{op}}^2 = \lambda_{\max}({}^t\mathbf{A}\mathbf{A})$ **(spectral norm)**, and $\|\mathbf{A}\|_F^2 = \sum_{i,j} a_{i,j}^2$ **(Frobenius or Trace norm)**.
The spectral norm of $\mathbf{G}_m^{-1}$ grows with the dimension m (Mabon (2015)).

- **Model selection**

(A1) $\mathcal{M}_n^{(0)} = \left\{1 \leqslant m \leqslant d_1,\ m\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \leqslant \frac{n}{\log(n)}\right\}$ with $d_1 < n$
(A2) $\forall b > 0, \quad \sum_{m \in \mathcal{M}_n^{(0)}} \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 e^{-bm} < C(b) < +\infty$

$$\mathrm{pen}_0(m) = \frac{\kappa_1}{n}\left[(2m\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2) \wedge \log(n)(\|h\|_\infty \vee 1)\|\mathbf{G}_m^{-1}\|_F^2)\right]$$

Risk bound for the data driven estimator

Select $\hat{m}_0 = \arg \min_{m \in \mathcal{M}_n^{(0)}} \left\{ -\|\hat{f}_m\|^2 + \text{pen}_0(m) \right\}$

$$\mathbb{E}(\|\hat{f}_{\hat{m}_0} - f\|^2) \leqslant 4 \min_{m \in \mathcal{M}_n^{(0)}} \left\{ \|f - f_m\|^2 + \text{pen}_0(m) \right\} + \frac{C}{n}$$

where C depends on $\|f\|$ and $\|g\|$ (but not on n).
$\hookrightarrow$ Bias : $\|f - f_m\|^2$ - variance compromise

- For $G_m = I_m$ (identity matrix), $m\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 = m$ and $\|\mathbf{G}_m^{-1}\|_F^2 = m$, and variance has order m/n.
- General Inequality

$$\frac{1}{\sqrt{m}} \|\mathbf{G}_m^{-1}\|_F \leqslant \|\mathbf{G}_m^{-1}\|_{\text{op}} \leqslant \|\mathbf{G}_m^{-1}\|_F.$$

- For example for $Y \sim \gamma(q, \theta)$, variance term has order m^{2q}/n

Order of the bias term?

Bongioanni, B. and Torrea, J. L *What is a Sobolev space for the Laguerre function systems?*

$$s > 0, \quad W^s(\mathbb{R}^+, L) := \{f : \mathbb{R}^+ \to \mathbb{R}, f \in \mathbb{L}^2(\mathbb{R}^+), \sum_{j \geqslant 0} j^s \langle f, \varphi_j \rangle^2 \leqslant L < +\infty\},$$

Then:

$$\|f_m - f\|^2 = \sum_{j=m}^{\infty} a_j^2(f) = \sum_{j=m}^{\infty} a_j^2(f) j^s j^{-s} \leqslant L m^{-s}.$$

Example of $\gamma(q, \theta)$ noise: choose $m = m_{\text{opt}}$ which minimizes $Lm^{-s} + c_2 m^{2q}/n$,

$$m_{\text{opt}} = C n^{1/(s+2q)} \text{ with } C := C(s, q, L)$$

which implies

$$\mathbb{E}[\|\hat{f}_{m_{\text{opt}}} - f\|^2] \leqslant O(n^{-s/(s+2q)})$$

Optimality? In progress... in the exp-noise case.

Case g **unknown**

We have a preliminary sample of the noise distribution $Y_1', \dots, Y_{n_0}'$, independent of $(Z_i)_{1\leqslant i\leqslant n}$.

Estimate of $\mathbf{G}_m$:

$$[\widehat{\mathbf{G}}_m]_{i,j} = \begin{cases} 2^{-1/2}\hat{a}_0(Y') & \text{if } i = j, \\ 2^{-1/2}\left(\hat{a}_{i-j}(Y') - \hat{a}_{i-j-1}(Y')\right) & \text{if } j < i, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where

$$\hat{a}_k(Y') = \frac{1}{n_0}\sum_{\ell=1}^{n_0} \varphi_k(Y_\ell') \quad \to \mathbb{E}[\widehat{\mathbf{G}}_m] = \mathbf{G}_m.$$

For the same reason as $\mathbf{G}_m$, the matrix $\widehat{\mathbf{G}}_m$ is **invertible** (lower triangular, nonzero diagonal elements).

But we have to introduce a **cutoff.**

Inversion of $\widehat{\mathbf{G}}_m$ with cutoff

Define an inverse of $\widehat{\mathbf{G}}_m$:

$$\widetilde{\mathbf{G}}_m^{-1} = \begin{cases} \widehat{\mathbf{G}}_m^{-1} & \text{if} \quad \|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}} \leqslant \sqrt{\dfrac{n_0}{m \log m}} \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We can see that the invertibility condition is defined with $\|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}$.

Remark.

$$\frac{\sqrt{2}}{|a_0(g)|} = \mathrm{spr}(\mathbf{G}_m^{-1}) \leqslant \|\mathbf{G}_m^{-1}\|_{\mathrm{op}} \tag{9}$$

$\mathrm{spr}(A)$ is the spectral radius (largest eigenvalue in absolute value) of A.

Estimator with g unknown

Finally, we estimate the projection f_m of f on the space $\mathcal{S}_m$ by

$$\tilde{f}_m(x) = \sum_{k=0}^{m-1} \tilde{a}_k \varphi_k(x) \quad \text{with} \quad \vec{\tilde{f}}_m = \widetilde{\mathbf{G}}_m^{-1} \hat{\vec{h}}_m \tag{10}$$

with $\hat{\vec{h}}_m$ defined by Equation (5).
We get

$$\mathbb{E}\|f - \tilde{f}_m\|^2 \leqslant \underbrace{\|f - f_m\|^2 + 2\mathbb{E}\|\mathbf{G}_m^{-1}(\vec{h}_m - \hat{\vec{h}}_m)\|_{2,m}^2}_{\text{terms for known } g} + 2\underbrace{\mathbb{E}\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\hat{\vec{h}}_m\|_{2,m}^2}_{g \text{ unknown}}.$$

- First two terms correspond to the squared bias term and the variance term already studied in Mabon (2015) for known $\mathbf{G}_m$,
- **Difficulty**: bound the last term, matrix concentration inequalities.

Application of matricial Bernstein by Tropp (2012)

Lemma

For $\widetilde{\mathbf{G}}_m^{-1}$ defined by Equation (8), $\|g\|_\infty < \infty$ and $m \log m \leqslant n_0$, then for any integer p there exists a positive constant $\mathfrak{C}_{\mathrm{op},p}$ such that

$$\mathbb{E}\left[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^{2p}\right] \leqslant \mathfrak{C}_{\mathrm{op},p}\left(\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \wedge \log m \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^4 \frac{m}{n_0}\right)^p. \quad (11)$$

Corollary

Under the Assumptions of Lemma 1, there exists a positive constant $\mathfrak{C}_{\mathrm{F}}$ such that

$$\mathbb{E}\left[\|\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^{2}\right] \leqslant \mathfrak{C}_{\mathrm{F}}\left(\|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2 \wedge \log m \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2 \frac{m}{n_0}\right). \quad (12)$$

Corollary

i) Under the Assumptions of Lemma 1, there exists a positive constant $\mathfrak{C}_{\mathrm{E}}$ such that

$$\mathbb{E}\left[\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2\right] \leqslant \mathfrak{C}_{\mathrm{op},1}\|f\|^2 \left(1 \wedge \log m \frac{m}{n_0}\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2\right).$$

ii) Under Assumptions of Lemma 1 and $\|f\|_{\ell_1} = \sum_{k\geqslant 0}|a_k(f)| < \infty$, there exists a positive constant $\mathfrak{C}'_{\mathfrak{e}}$ such that

$$\mathbb{E}\left[\|(\mathbf{G}_m^{-1} - \widetilde{\mathbf{G}}_m^{-1})\vec{h}_m\|_{2,m}^2\right] \leqslant \mathfrak{C}'_{\mathrm{E}}(\|g\|_\infty, \|f\|_{\ell_1}^2, \|f\|^2)\frac{1}{n_0}\left(m\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \wedge \|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2\right)$$

$$+ 2^{2p+1}\mathfrak{C}_{2p}\left(\frac{2m\log(m)\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}}{n_0}\right)^p.$$

General risk bound

Proposition

If f and g belong to $\mathbb{L}^2(\mathbb{R}^+)$, $\|g\|_\infty < \infty$, for $\tilde{f}_m$ defined by (10) the following result holds

$$\begin{aligned}\mathbb{E}\|f - \tilde{f}_m\|^2 \quad &\leqslant \quad \|f - f_m\|^2 + \frac{\mathfrak{C}}{n}\left(2m\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_{\text{F}}^2\right) \\ &\qquad + 2\mathfrak{C}_{\text{E}} \log(m) \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2\end{aligned}$$

with $\mathfrak{C} = 2 + \mathfrak{C}_{\text{op},1} + \mathfrak{C}_{\text{F}}$.

- First two terms correspond to the upper bound on the mean integrated risk when the matrix $\mathbf{G}_m^{-1}$ is known.
- Third term is due to the estimation of the matrix $\mathbf{G}_m^{-1}$.

Proposition

If f and g belong to $\mathbb{L}^2(\mathbb{R}^+)$, $\|g\|_\infty < \infty$,

$$\|f\|_{\ell_1} := \sum_j |a_j(f)| < \infty,$$

$$\mathbf{m} \log(m)\|\mathbf{G}_m^{-1}\|_{\text{op}}/n_0 < 1/4,$$

then $\tilde{f}_m$ defined by (10) satisfies, for any $p \geqslant 2$

$$\begin{aligned}
\mathbb{E}\|f - \tilde{f}_m\|^2 \quad \leqslant \quad & \|f - f_m\|^2 + \frac{\mathfrak{C}}{n}\left(2m\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_{\text{F}}^2\right) \\
& + \frac{2\mathfrak{C}'(\|g\|_\infty, \|f\|, \|f\|_{\ell_1})}{n_0}(m\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \|\mathbf{G}_m^{-1}\|_{\text{F}}^2) \\
& + 2^{2p+1}\mathfrak{C}_{2p}\left(\frac{2m\log(m)\|\mathbf{G}_m^{-1}\|_{\text{op}}}{n_0}\right)^p .
\end{aligned}$$

Upper rates

$$W^s(\mathbb{R}^+, L) = \left\{ p : \mathbb{R}^+ \to \mathbb{R},\, p \in \mathbb{L}^2(\mathbb{R}^+),\, \sum_{k \geqslant 0} k^s a_k^2(p) \leqslant L < +\infty \right\} \quad \text{with} \quad s \geqslant 0$$

where $a_k(p) = \langle p, \varphi_k \rangle$. For $f \in W^s(\mathbb{R}^+, L)$,

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} a_k^2(f) = \sum_{k=m}^{\infty} a_k^2(f) k^s k^{-s} \leqslant L m^{-s}.$$

So, we have the order of the bias-term.

Variance term?

Order of variance term

Order of $\|\mathbf{G}_m^{-1}\|_{\text{op}}^2$ and $\|\mathbf{G}_m^{-1}\|_{\text{F}}^2$.
Define $r \geqslant 1$ such that

$$\frac{d^j}{dx^j} g(x)|_{x=0} = \begin{cases} 0 & \text{if } j = 0, 1, \dots, r-2 \\ B_r \neq 0 & \text{if } j = r-1. \end{cases}$$

Comte, Cuenod, Pensky, Rozenholc (2015) show that if conditions

(C1) $g \in \mathbb{L}^1(\mathbb{R}^+)$ is r times differentiable and $g^{(r)} \in \mathbb{L}^1(\mathbb{R}^+)$,

(C2) The Laplace transform of g has no zero with non negative real parts except for the zeros of the form $\infty + ib$,

hold then $\|\mathbf{G}_m^{-1}\|_{\text{F}}^2 \asymp \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \asymp m^{2r}$.

Rates

Proposition

*If f belongs to $W^s(\mathbb{R}^+, L)$ and g satisfies Assumptions **(C1)-(C2)**, for $\tilde{f}_m$ defined by (10) there exists a choice m_{opt} such that*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}(\|f - \tilde{f}_{m_{\text{opt}}}\|^2) \lesssim n^{-s/s+2r} + (n_0/\log(n_0))^{-s/s+2r+1}.$$

If in addition $n_0 \geqslant n^{3/2}$, $\quad \sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}(\|f - \tilde{f}_{m_{\text{opt}}}\|^2) \lesssim n^{-s/s+2r}$.

Corollary

*If f belong to $W^s(\mathbb{R}^+, L)$ and $\|f\|_{\ell_1} < \infty$, g satisfies Assumptions **(C1)-(C2)** for $\tilde{f}_m$ defined by (10) there exists a choice m_{opt} such that*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}(\|f - \tilde{f}_{m_{\text{opt}}}\|^2) \lesssim (n \vee n_0)^{-s/s+2r}.$$

Penalization

(A1). $\widehat{\mathcal{M}} = \left\{1 \leqslant m \leqslant C\lfloor n/\log n\rfloor,\ m\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2 \leqslant n \wedge n_0\right\}$ with C a positive constant.
$\hookrightarrow$ **random collection of models**

and is theoretical counterpart:
$\mathcal{M} = \{1 \leqslant m \leqslant C\lfloor n/\log n\rfloor,\ m\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \leqslant n \wedge n_0\}$.

We define the two parts of the penalty

$$\widehat{\mathrm{pen}}_1(m) := \kappa_1 \log n \mathfrak{C}\left(\frac{2m\|h\|_\infty}{n}\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2 \wedge \frac{(\|h\|_\infty \vee 1)}{n}\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{F}}^2\right)$$
$$\widehat{\mathrm{pen}}_2(m) := 2\kappa_2 \mathfrak{C}_{\mathrm{E}}(\|g\|_\infty \vee 1)\log n_0 \frac{m}{n_0}\|\widetilde{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2.$$

and set the random penalty as

$$\widehat{\mathrm{pen}}(m) := \widehat{\mathrm{pen}}_1(m) + \widehat{\mathrm{pen}}_2(m) \tag{13}$$

Theoretical counterpart:

$$\begin{aligned}\mathrm{pen}(m) &:= \mathrm{pen}_1(m) + \mathrm{pen}_2(m)\\ &= \kappa_1 \log n \left(\frac{2m\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \wedge \frac{(\|h\|_\infty \vee 1)}{n} \|\mathbf{G}_m^{-1}\|_{\mathrm{F}}^2 \right)\\ &+ 2\kappa_2 \mathfrak{C}_{\mathrm{E}}(\|g\|_\infty \vee 1) \log n_0 \frac{m}{n_0} \|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2. \end{aligned} \tag{14}$$

where κ_1 and κ_2 are numerical constants.

Model selection result

Theorem

If f and $g \in \mathbb{L}^2(\mathbb{R}^+)$, $\|g\|_\infty < \infty$, let us suppose that ***(A)*** *and* ***(A1)*** *are true. Let $\hat{f}_{\widetilde{m}}$ be defined by* (10) *and*

$$\widetilde{m} = \arg\min_{m \in \widehat{\mathcal{M}}} \left\{ -\|\tilde{f}_m\|^2 + \widehat{\mathrm{pen}}(m) \right\}$$

with $\widehat{\mathrm{pen}}$ defined by (13), *then there exists a positive numerical constant κ_1 such that*

$$\mathbb{E}(\|f - \tilde{f}_{\widetilde{m}}\|^2) \leqslant 4 \inf_{m \in \mathcal{M}} \left\{ \|f - f_m\|^2 + \mathrm{pen}(m) \right\} + \frac{C}{n \wedge n_0}, \tag{15}$$

where C depends on $\|f\|$ and $\|g\|$, pen *is defined by* (14).

$$\begin{aligned}\widetilde{\text{pen}}(m) &:= \widetilde{\text{pen}}_1(m) + \widetilde{\text{pen}}_2(m)\\ &= \kappa_1 \log n(\|\hat{h}_D\|_\infty \vee 1)\left(\frac{2m\|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2}{n} \wedge \frac{\|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{F}}^2}{n}\right)\\ &+ 2\kappa_2(\|\hat{g}_D\|_\infty \vee 1)\frac{m}{n_0}\|\widetilde{\mathbf{G}}_m^{-1}\|_{\text{op}}^2.\end{aligned}$$

which is the penalty indeed implemented

Simulation experiments

Four densities

(i) A Gamma density, $\gamma(3, 1/2)$, variance 0.75,

(ii) A mixed Gamma $(5/8)X$ with $X \sim 0.4\gamma(2, 1/2) + 0.6\gamma(16, 1/4)$, variance 1.16,

(iii) A beta density $5X$ with $X \sim \beta(4, 5)$, variance 0.62,

(iv) A Rayleigh density $X \sim f$ with $f(x) = (x/\sigma^2)\exp(-x^2/(2\sigma^2))$ with $\sigma^2 = 2/(4-\pi)$, variance 1.

All with mainly support $[0, 5]$.

Calibration of constants in penalties.

Two types of nuisance process Y with same variance $1/4$,

- an exponential density $\mathcal{E}(\lambda)$ with $\lambda = 2$,
- a gamma density $\gamma(2, 1/\lambda')$ with $\lambda' = 2\sqrt{2}$.

We can compute analytically the matrix $\mathbf{G}_m$:

- For $Y \sim \mathcal{E}(\lambda)$, we have $[\mathbf{G}_m]_{i,i} = \lambda/(1+\lambda)$ and

$$[\mathbf{G}_m]_{i,j} = -2\lambda \frac{(\lambda-1)^{i-j-1}}{(\lambda+1)^{(i-j+1)}} \text{ if } j < i \tag{16}$$

  and $[\mathbf{G}_m]_{i,j} = 0$ otherwise.
- For $Y \sim \gamma(2, \mu)$, we have

$$[\mathbf{G}_m]_{i,i} = (\mu/(1+\mu))^2, \quad [\mathbf{G}_m]_{i+1,i} = -4\mu^2/(1+\mu))^3 \text{ and}$$

$$[\mathbf{G}_m]_{i,j} = 4(i-j-\mu)\mu^2 \frac{(\mu-1)^{i-j-2}}{(\mu+1)^{(i-j+2)}} \text{ if } i > j+1 \tag{17}$$

  and $[\mathbf{G}_m]_{i,j} = 0$ otherwise.

Simulation results

		$n = 400$ direct	$n = 400$ Known noise	$n = 400$ Noise sample $n_0 = 400$	$n = 2000$ direct	$n = 2000$ Known noise	$n = 2000$ Noise sample $n_0 = 400$	$n = 2000$ Noise sample $n_0 = 2000$
(i)	MISE	3.9	7.4	5.4	0.6	1.4	1.4	1.2
	(std)	(0.2)	(10.3)	(4.0)	(0.5)	(1.7)	(1.2)	(1.4)
	Oracles	1.6	3.1	3.3	0.4	0.7	1.1	0.8
	(std)	(1.4)	(2.2)	(2.3)	(0.3)	(0.5)	(0.9)	(0.6)
(ii)	MISE	6.4	22.5	12.8	1.4	5.2	4.6	4.6
	(std)	(3.0)	(32.8)	(7.3)	(0.7)	(5.0)	(2.1)	(3.3)
	Oracles	3.7	8.2	8.9	0.9	2.8	4.0	2.9
	(std)	(2.0)	(4.9)	(5.3)	(0.6)	(1.6)	(1.5)	(1.6)
(iii)	MISE	2.3	3.3	3.5	0.6	1.2	1.6	1.3
	(std)	(1.3)	(3.3)	(3.5)	(0.4)	(0.7)	(0.7)	(0.7)
	Oracles	1.5	2.1	2.2	0.4	0.7	0.9	0.7
	(std)	(1.2)	(1.5)	(1.6)	(0.3)	(0.5)	(0.6)	(0.5)
(iv)	MISE	1.5	4.0	3.3	0.5	1.0	1.0	0.9
	(std)	(0.9)	(6.0)	(3.5)	(0.3)	(1.7)	(0.6)	(0.8)
	Oracles	1.3	1.8	1.9	0.3	0.5	0.6	0.5
	(std)	(1.1)	(1.5)	(1.7)	(0.2)	(0.3)	(0.4)	(0.4)

Table: Results after 200 iterations of simulations of density (i) to (iv). First two lines: MISE$\times$ 1000 with (std $\times$ 1000) in parenthesis. Third and fourth lines: mean with std in parenthesis of oracles. The noise is $\mathcal{E}(\lambda)$ with $\lambda = 2$ (mean 1/2). Mean ration to oracles 1.85 for $n = 400$, 1.65 for $n = 2000$.

$\bar{m} = 13.9$ (2.4) $\bar{m} = 12.0$ (3.0) $\bar{m} = 11.5$ (2.3)

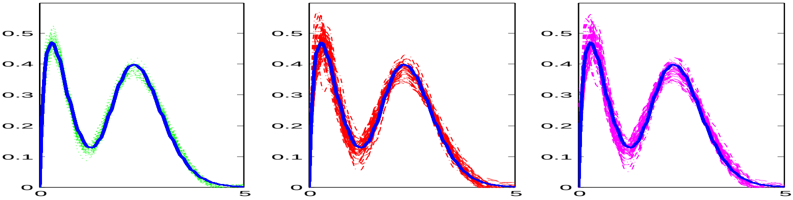

Figure: True density f of Model (ii) (Mixed Gamma distribution) with exponential noise. Left: 50 estimators of f from direct observation of X in dotted green. Middle: 50 estimators of f from observation of Z with known noise density (G_m is known), in dotted red. Right: 50 estimators of f from observation of Z with estimated G_m from a n_0-sample of noise, in dotted magenta. Above each plot, $\bar{m}$ is the mean of the selected dimensions with standard deviation in parenthesis.

$\bar{m} = 5.7\ (1.0)$ $\bar{m} = 4.1\ (0.3)$ $\bar{m} = 4.12\ (0.3)$

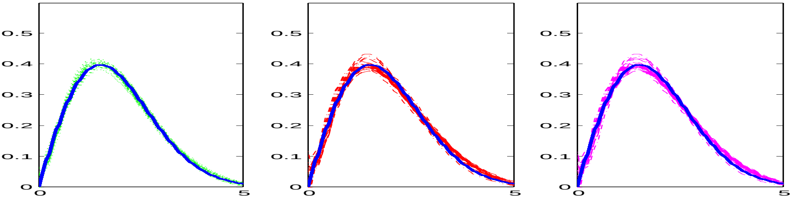

Figure: True density f of Model (iv) (Rayleigh distribution) with Gamma noise with $n = n_0 = 2000$. Left: 50 estimators of f from direct observation of X in dotted green. Middle: 50 estimators of f from observation of Z with known noise density (G_m is known), in dotted red. Right: 50 estimators of f from observation of Z with estimated G_m from a n_0-sample of noise, in dotted magenta. Above each plot, $\bar{m}$ is the mean of the selected dimensions with standard deviation in parenthesis.

$\bar{m} = 4.2\ (0.6)$ $\bar{m} = 3.6\ (0.8)$ $\bar{m} = 3.8\ (0.6)$

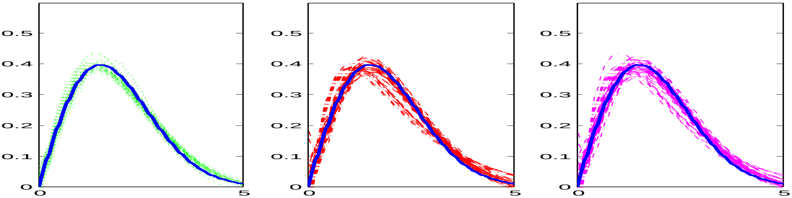

Figure: True density f of Model (iv) (Rayleigh distribution) with Gamma noise with $n = n_0 = 400$. Left: 50 estimators of f from direct observation of X in dotted green. Middle: 50 estimators of f from observation of Z with known noise density (G_m is known), in dotted red. Right: 50 estimators of f from observation of Z with estimated G_m from a n_0-sample of noise, in dotted magenta. Above each plot, $\bar{m}$ is the mean of the selected dimensions with standard deviation in parenthesis.

Thank you for your attention !