Ruprecht-Karls-Universität Heidelberg

Optimal aggregation in circular deconvolution by a frequentist Bayesian approach

StatMathAppli
Fréjus

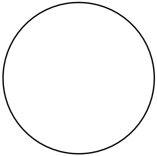
Xavier Loizeau

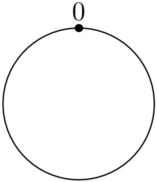
4^{th} *of September 2017*

In collaboration with
Fabienne Comte and Jan Johannes

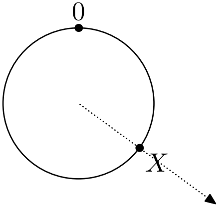
Contents

Considered model
Circular deconvolution model

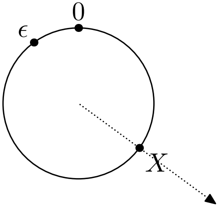




0

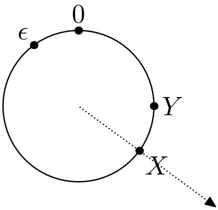


- Object of interest : density of a circular r.v. $X \sim f^X$;



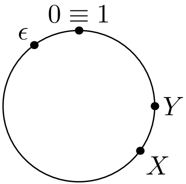
- Object of interest : density of a circular r.v. $X \sim f^X$;
- Noise : circular r.v. $\epsilon \sim f^\epsilon$;

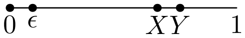
Considered model
Circular deconvolution model



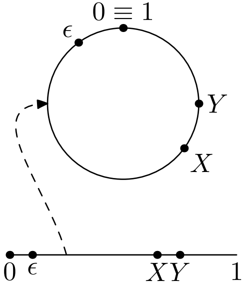
- Object of interest : density of a circular r.v. $X \sim f^X$;
- Noise : circular r.v. $\epsilon \sim f^\epsilon$;
- Observation : noisy circular r.v. $Y = X + \epsilon \sim f^Y = f^X \star f^\epsilon$

Considered model
Circular deconvolution model





- Object of interest : density of a circular r.v. $X \sim f^X$;
- Noise : circular r.v. $\epsilon \sim f^\epsilon$;
- Observation : noisy circular r.v. $Y = X + \epsilon \sim f^Y = f^X \star f^\epsilon$

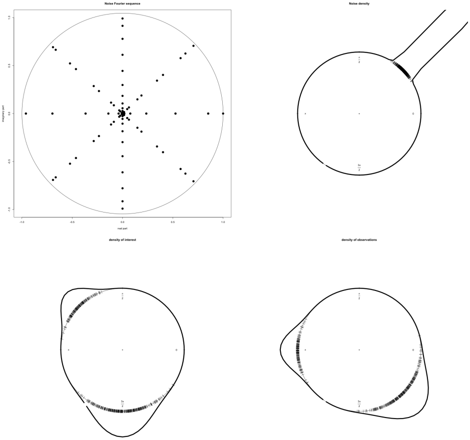


- Object of interest : density of a circular r.v. $X \sim f^X$;
- Noise : circular r.v. $\epsilon \sim f^\epsilon$;
- Observation : noisy circular r.v. $Y = X + \epsilon \sim f^Y = f^X \star f^\epsilon$

As all our variables have values in $[0,1]$, their densities can be extended in Fourier series. Hence we give the following definitions :

- $\forall j \in \mathbb{Z}, x \in [0,1], \quad e_j(x) = \exp[-2\pi \cdot i \cdot j \cdot x]$, the Fourier basis ;
- $\left[f^X\right]_j = \mathbb{E}_{f^X}\left[e_j(-X_1)\right]$, Fourier coefficients of f^X ;
- $[f^\epsilon]_j = \mathbb{E}_{f^\epsilon}\left[e_j(-\epsilon_1)\right]$, the Fourier coefficients of f^ϵ ;
- $\left[f^Y\right]_j = \mathbb{E}_{f^Y}\left[e_j(-Y_1)\right] = \left[f^X\right]_j \cdot [f^\epsilon]_j$, the Fourier coefficients of f^Y ;

Considered model
Illustration



The problem is here treated from a Bayesian point of view :

- the parameter $[\boldsymbol{f}]$ is a random variable with prior $\mathbb{P}_{[\boldsymbol{f}]}$ with density $\pi_{[\boldsymbol{f}]}$ with respect to some measure ;

The problem is here treated from a Bayesian point of view :

- the parameter $[\boldsymbol{f}]$ is a random variable with prior $\mathbb{P}_{[\boldsymbol{f}]}$ with density $\pi_{[\boldsymbol{f}]}$ with respect to some measure ;
- given $[\boldsymbol{f}]$, the likelihood of Y is $\mathbb{P}^n_{Y|[\boldsymbol{f}]}$ with density $\pi^n_{Y|[\boldsymbol{f}]}(y_1, \dots, y_n, [f]) = \prod_{k \in [\![1,n]\!]} \sum_{j \in \mathbb{Z}} [\boldsymbol{f}]_j [f^\epsilon]_j e_j(y_k)$;

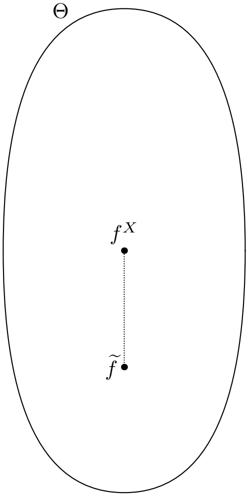
The problem is here treated from a Bayesian point of view :

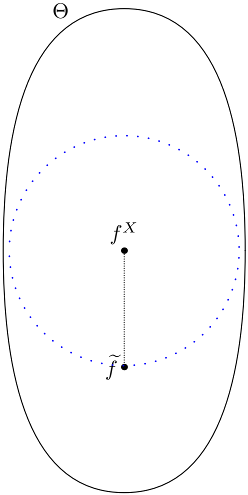
- the parameter $[\boldsymbol{f}]$ is a random variable with prior $\mathbb{P}_{[\boldsymbol{f}]}$ with density $\pi_{[\boldsymbol{f}]}$ with respect to some measure ;
- given $[\boldsymbol{f}]$, the likelihood of Y is $\mathbb{P}^n_{Y|[\boldsymbol{f}]}$ with density $\pi^n_{Y|[\boldsymbol{f}]}(y_1, \dots, y_n, [f]) = \prod_{k \in [\![1,n]\!]} \sum_{j \in \mathbb{Z}} [\boldsymbol{f}]_j [f^\epsilon]_j e_j(y_k)$;
- the marginal distribution of Y is $\mathbb{P}^n_Y$;

The problem is here treated from a Bayesian point of view :

- the parameter $[\boldsymbol{f}]$ is a random variable with prior $\mathbb{P}_{[\boldsymbol{f}]}$ with density $\pi_{[\boldsymbol{f}]}$ with respect to some measure ;
- given $[\boldsymbol{f}]$, the likelihood of Y is $\mathbb{P}^n_{Y|[\boldsymbol{f}]}$ with density $\pi^n_{Y|[\boldsymbol{f}]}(y_1, \dots, y_n, [f]) = \prod_{k \in [\![1,n]\!]} \sum_{j \in \mathbb{Z}} [\boldsymbol{f}]_j [f^\epsilon]_j e_j(y_k)$;
- the marginal distribution of Y is $\mathbb{P}^n_Y$;
- we are interested in the posterior distribution $\mathbb{P}^n_{[\boldsymbol{f}]|Y}$ with density $\pi^n_{[\boldsymbol{f}]|Y}([f], y) \propto \pi^n_{Y|[\boldsymbol{f}]}(y, [f]) \cdot \pi_{[\boldsymbol{f}]}([f])$.

Formulation of optimality
Frequentist case

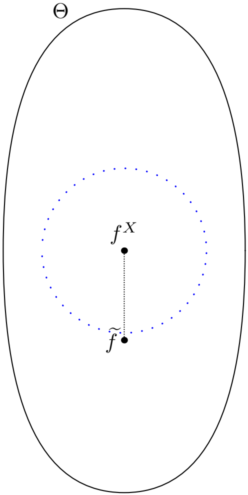




- For each frequentist estimator, measure the performance by its risk

$$\mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right].$$

Formulation of optimality
Frequentist case



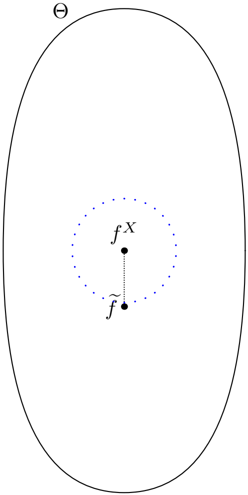
- For each frequentist estimator, measure the performance by its risk

$$\mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{f}} \mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right] \geq \mathscr{R}^{\circ}_n\left(f^X\right)$$

over a family of estimators.



- For each frequentist estimator, measure the performance by its risk

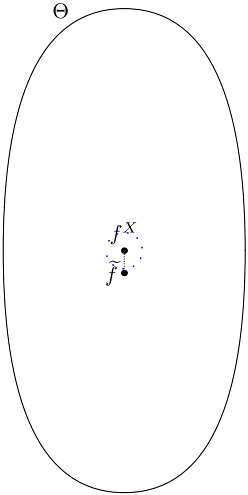
$$\mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{f}} \mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right] \geq \mathscr{R}^{\circ}_n\left(f^X\right)$$

over a family of estimators.

Formulation of optimality
Frequentist case



- For each frequentist estimator, measure the performance by its risk

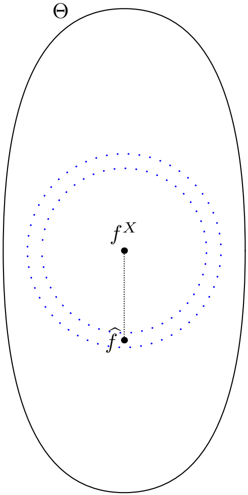
$$\mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{f}} \mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right] \geq \mathscr{R}^{\circ}_n\left(f^X\right)$$

over a family of estimators.

Formulation of optimality
Frequentist case



- For each frequentist estimator, measure the performance by its risk

$$\mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{f}} \mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right] \geq \mathscr{R}^\circ_n\left(f^X\right)$$

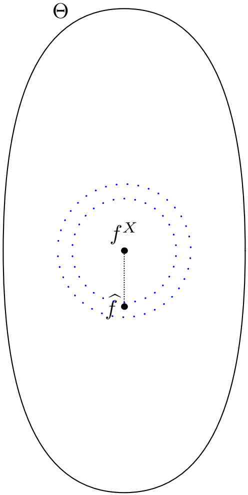
over a family of estimators.

- Finding $\widehat{f}$ in this family such that

$$\mathbb{E}^n_{f^X}\left[d\left(\widehat{f}, f^X\right)^2\right] \leq K^\circ \cdot \mathscr{R}^\circ_n\left(f^X\right),$$

it is then oracle rate optimal and adaptive if $\widehat{f}$ does not depend on f^X.

Formulation of optimality
Frequentist case



- For each frequentist estimator, measure the performance by its risk

$$\mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right].$$

- Goal : derive a lower bound for this risk

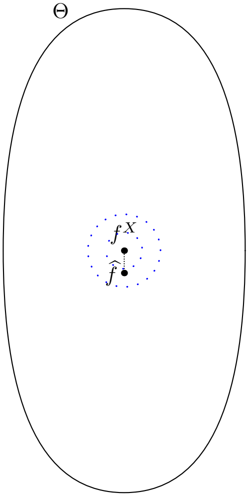
$$\inf_{\widetilde{f}} \mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right] \geq \mathscr{R}^\circ_n\left(f^X\right)$$

over a family of estimators.

- Finding $\widehat{f}$ in this family such that

$$\mathbb{E}^n_{f^X}\left[d\left(\widehat{f}, f^X\right)^2\right] \leq K^\circ \cdot \mathscr{R}^\circ_n\left(f^X\right),$$

it is then oracle rate optimal and adaptive if $\widehat{f}$ does not depend on f^X.



- For each frequentist estimator, measure the performance by its risk

$$\mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{f}} \mathbb{E}^n_{f^X}\left[d\left(\widetilde{f}, f^X\right)^2\right] \geq \mathscr{R}^\circ_n\left(f^X\right)$$

over a family of estimators.

- Finding $\widehat{f}$ in this family such that

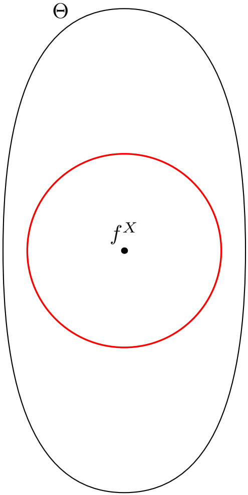
$$\mathbb{E}^n_{f^X}\left[d\left(\widehat{f}, f^X\right)^2\right] \leq K^\circ \cdot \mathscr{R}^\circ_n\left(f^X\right),$$

it is then oracle rate optimal and adaptive if $\widehat{f}$ does not depend on f^X.

How to transfer this in a Bayesian point of view ?

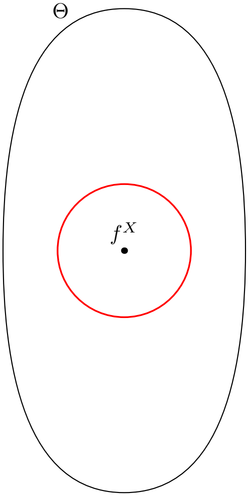
Taking a "pragmatic Bayesian" point of view :

- f^X the true parameter.
- Is $\mathbb{P}^n_{\boldsymbol{f}|Y}$ shrinking around f^X as n tends to ∞ ?
- How fast ?



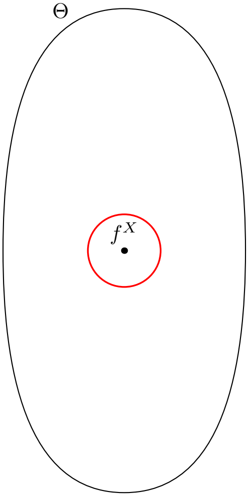
- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n\right)\right] = 0.$$



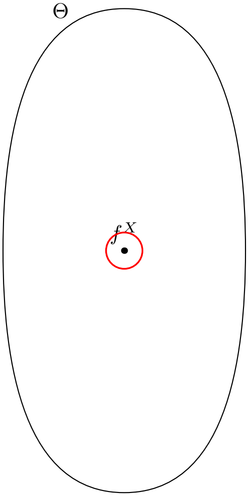
- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n\right)\right] = 0.$$



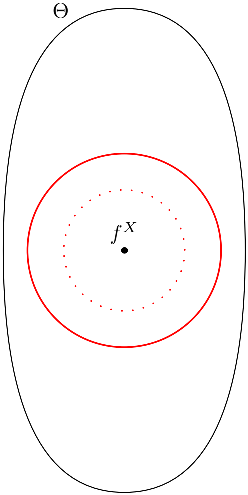
- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n\right)\right] = 0.$$



- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n\right)\right] = 0.$$

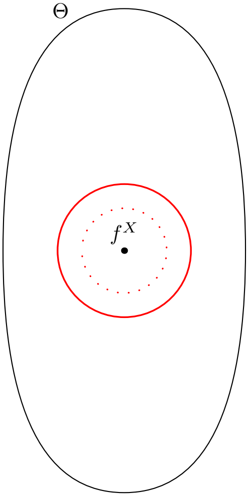


- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n\right)\right] = 0.$$

- Exact concentration rate $(\phi_n)_{n\in\mathbb{N}}$ if in addition

$$\lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \leq c^{-1}\,\phi_n\right)\right] = 0.$$

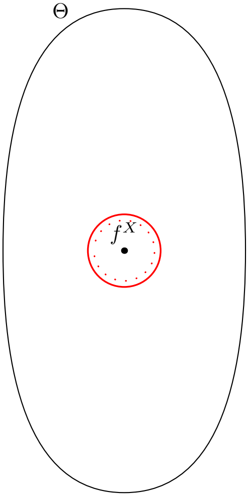


- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n\right)\right] = 0.$$

- Exact concentration rate $(\phi_n)_{n\in\mathbb{N}}$ if in addition

$$\lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \leq c^{-1}\,\phi_n\right)\right] = 0.$$

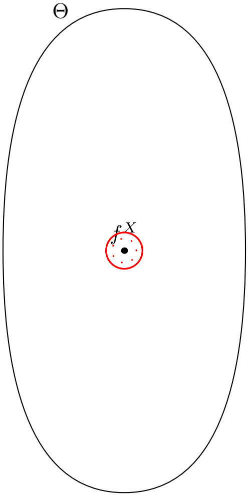


- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n\right)\right] = 0.$$

- Exact concentration rate $(\phi_n)_{n\in\mathbb{N}}$ if in addition

$$\lim_{n\to\infty} \mathbb{E}^n_{f^X}\left[\mathbb{P}^n_{\boldsymbol{f}|Y}\left(d\left(\boldsymbol{f}, f^X\right)^2 \leq c^{-1}\,\phi_n\right)\right] = 0.$$



- Concentration rate $(\phi_n)_{n \in \mathbb{N}}$

$$\exists c \in \mathbb{R}_+, \quad \lim_{n \to \infty} \mathbb{E}^n_{f^X} \left[\mathbb{P}^n_{\boldsymbol{f}|Y} \left(d\left(\boldsymbol{f}, f^X\right)^2 \geq c\,\phi_n \right) \right] = 0.$$

- Exact concentration rate $(\phi_n)_{n \in \mathbb{N}}$ if in addition

$$\lim_{n \to \infty} \mathbb{E}^n_{f^X} \left[\mathbb{P}^n_{\boldsymbol{f}|Y} \left(d\left(\boldsymbol{f}, f^X\right)^2 \leq c^{-1}\,\phi_n \right) \right] = 0.$$

A popular frequentist method
Projection estimators

From a frequentist point of view, a natural method to answer this problem is :

- define $\widehat{[f^Y]}_j = \frac{1}{n}\sum_{k=1}^{n} e_j(-Y_k)$, an unbiased estimator for $[f^Y]_j = \mathbb{E}_{f^Y}[e_j(-Y_1)]$;

From a frequentist point of view, a natural method to answer this problem is :

- define $\widehat{[f^Y]}_j = \frac{1}{n}\sum_{k=1}^{n} e_j(-Y_k)$, an unbiased estimator for $[f^Y]_j = \mathbb{E}_{f^Y}[e_j(-Y_1)]$;
- consider $\widehat{[f]}_j = \frac{\widehat{[f^Y]}_j}{[f^\epsilon]_j}$, an unbiased estimator for $[f^X]_j$;

A popular frequentist method
Projection estimators

From a frequentist point of view, a natural method to answer this problem is :

- define $\widehat{[f^Y]}_j = \frac{1}{n}\sum_{k=1}^{n} e_j(-Y_k)$, an unbiased estimator for $[f^Y]_j = \mathbb{E}_{f^Y}[e_j(-Y_1)]$;
- consider $\widehat{[f]}_j = \frac{\widehat{[f^Y]}_j}{[f^\epsilon]_j}$, an unbiased estimator for $[f^X]_j$;
- use an element from the family of projection estimators : $\left\{\left(\widehat{[f]}_j^m\right)_{j\in\mathbb{Z}} : \forall j \in \mathbb{Z}, \quad \widehat{[f]}_j^m = \widehat{[f]}_j \mathbb{1}_{\{0<|j|\leq m\}} + \mathbb{1}_{\{j=0\}}\right\}$;

A popular frequentist method
Projection estimators

From a frequentist point of view, a natural method to answer this problem is :

- define $\widehat{[f^Y]}_j = \frac{1}{n}\sum_{k=1}^{n} e_j(-Y_k)$, an unbiased estimator for $[f^Y]_j = \mathbb{E}_{f^Y}[e_j(-Y_1)]$;
- consider $\widehat{[f]}_j = \frac{\widehat{[f^Y]}_j}{[f^\epsilon]_j}$, an unbiased estimator for $[f^X]_j$;
- use an element from the family of projection estimators : $\left\{\left(\widehat{[f]}_j^m\right)_{j\in\mathbb{Z}} : \forall j \in \mathbb{Z}, \quad \widehat{[f]}_j^m = \widehat{[f]}_j \mathbb{1}_{\{0<|j|\leq m\}} + \mathbb{1}_{\{j=0\}}\right\}$;
- define a method to select m in a satisfactory way (e.g. Massart [2003]).

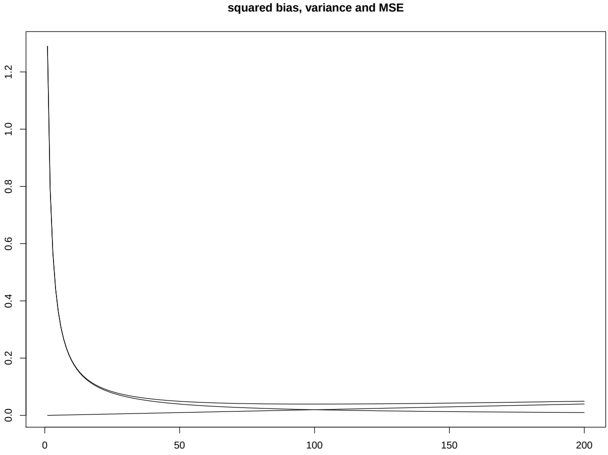


Figure: MSE of projection estimators in the direct case

Illustration
Direct case

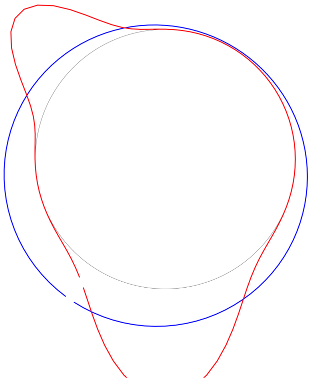


Illustration
Inverse case

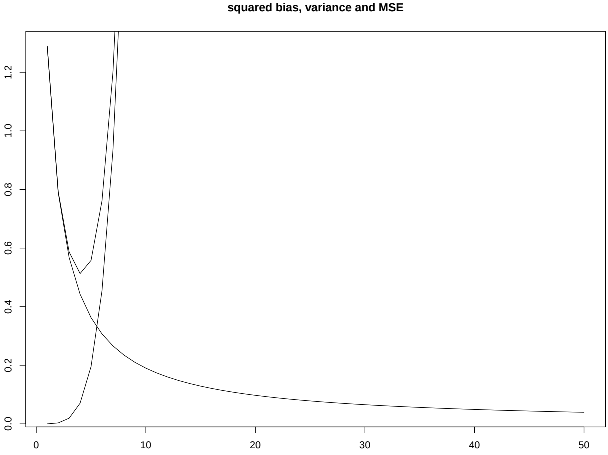
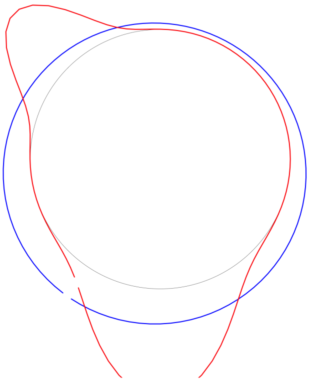


Figure: MSE of projection estimators in the mildly ill-posed case case

Considered model
Illustration



Contents

We first introduce the so called sieve priors with deterministic threshold parameter m :

$$\pi_{[\boldsymbol{f}]}([f]) = \prod_{|j|>m} \delta_0\left([f]_j\right) \cdot \exp\left[\Delta_1([f])\right].$$

The posterior is then of the form

$$\pi_{[\boldsymbol{f}]|Y^n}([f], y^n) = \prod_{|j|>m} \delta_0\left([f]_j\right) \cdot$$

$$\exp\left[\underbrace{\Delta_1([f]) + \sum_{k=1}^{n} \log\left(1 + \sum_{0<|j|\leq m} [f]_j \cdot [f^\epsilon]_j \cdot e_j(y_k)\right) + \Delta_2(y^n)}_{R([f],y^n)}\right].$$

Open question : By using a Taylor expansion of the log around 1, it seems like this distribution could be approximated by a Gaussian, indeed :

$$R([f], y^n) = -\sum_{0<|j|\leq m} \frac{\left|[f]_j - \widehat{[f]}_j\right|^2}{2\cdot\sigma_j}$$

$$+\sum_{l=2}^{\infty}\sum_{k=1}^{n}\frac{(-1)^{l+1}}{l}\left(\sum_{0<|j|\leq m}[f]_j\cdot[f^\epsilon]_j\cdot e_j(y_k)\right)^l + \delta_1([f]) + \delta_2(y^n);$$

where $\sigma_j = \frac{1}{1+n\left|[f^\epsilon]_j\right|^2}$ and $\widehat{[f]}_j = \frac{[f^\epsilon]_j \cdot \sum_{k=1}^{n} e_j(y_k)}{1+n\left|[f^\epsilon]_j\right|^2}$

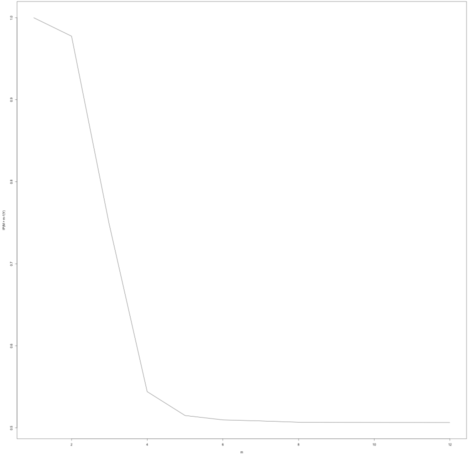
- Consider a random hyper-parameter M, with values in $[\![1, n]\!]$, acting like a threshold :

$$\forall j \in \mathbb{Z} : |j| > m, \quad \mathbb{P}_{[\boldsymbol{f}]_j | M=m} = \delta_0$$

- Write the density of M $\pi_M(m) \propto \exp[pen(m)]$ with respect to the counting measure, then, the posterior can be written

$$\begin{aligned}
\pi_{M|Y^n}(m, y^n) &= \exp[pen(m) - 2\Delta_2(y^n)] \\
\pi_{[\boldsymbol{f}]|Y^n} &= \sum_{m \in \mathbb{N}} \pi_{[\boldsymbol{f}]|M=m, Y^n} \cdot \pi_{M=m|Y^n}; \\
\widetilde{[\boldsymbol{f}]} &= \left(\mathbb{E}_{[\boldsymbol{f}]|M \geq j, Y^n}\left[[\boldsymbol{f}]_j\right] \cdot \mathbb{P}_{M|Y^n}(M \geq j)\right)_{j \in \mathbb{N}}.
\end{aligned}$$

Hierarchical prior
Graphical representation



Contents

We here note : $\Phi_n^\circ = \Phi_n^\circ([f^X])$ the optimal convergence rate for the family of projection estimators for each $[f^X]$; this objects has analytical formulation (see for example Comte and Taupin [2003]).

Optimality of the Bayes estimate

Theorem : oracle optimality J & C & L [2017]

For each $[f^X]$, there exist C° such that

$$\forall n, \in \mathbb{N}^\star, \quad \mathbb{E}^n_{[f^X]}\left[\left\| \widetilde{[\boldsymbol{f}]} - \left[f^X\right] \right\|^2\right] \leq C^\circ \Phi^\circ_n.$$

Theorem J & C & L [2017]

Φ_n° can be considered as upper bound for the family of sieve priors :

$$\lim_{n\to\infty} \inf_{\mathbb{Q}_{[f]}} \mathbb{E}^n_{[f^X]} \left[\mathbb{Q}_{[\boldsymbol{f}]|Y^n} \left(\left\| [\boldsymbol{f}] - \left[f^X \right] \right\|^2 \geq \Phi_n^\circ \right) \right] = 1,$$

where $\inf_{\mathbb{Q}_{[f]}}$ is taken over the family of sieve priors with Gaussian margins.

This gives a purely Bayesian formulation of oracle optimality.

Optimality of the posterior distribution

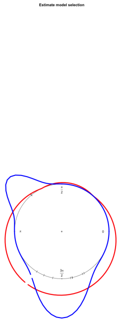
Theorem : oracle concentration J & C & L [2017]

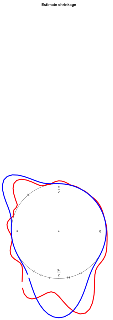
For all $[f^X]$ in Θ°, there exist K° such that

$$\lim_{n\to\infty} \mathbb{E}^n_{[f^X]}\left[\mathbb{P}_{[\boldsymbol{f}],M|Y^n}\left((K^\circ)^{-1}\Phi_n^\circ \leq \left\|[\boldsymbol{f}] - \left[f^X\right]\right\|^2 \leq K^\circ\Phi_n^\circ\right)\right] = 1.$$

Contents

Direct case : visualisation of the estimates





Direct case : sampling from the approximated posterior

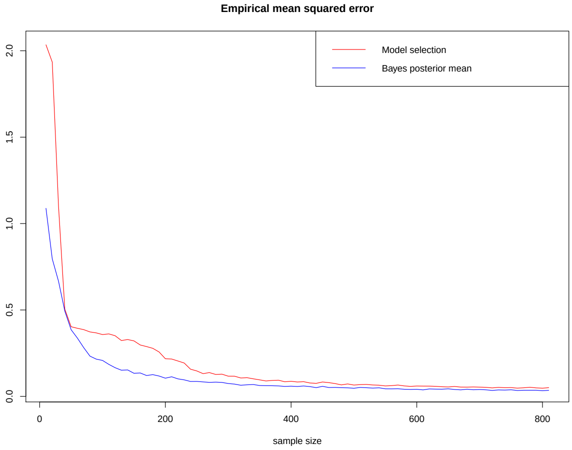


Figure: Evolution of the empirical mean squared error of the two estimates with respect to the number of observations in the direct case

Direct case : empirical mean square errors of the estimates

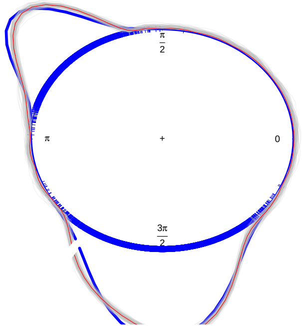


Figure: Sample from the posterior distribution and the posterior mean in the direct case

Summary

- Bayesian method for circular deconvolution with interpretable prior ;
- providing a fully data driven estimator, fulfilling oracle optimality ;
- giving a Bayesian formulation of Oracle optimality...

And much more !

- Also fulfills minimax optimality (no log term !)
- actually provides a complete family of optimal estimators using iterated posterior method
- works with β-mixing dependance
- suitable for severely ill-posed problems
- can be adapted to real line deconvolution