

Optimal aggregation in circular deconvolution by a frequentist Bayesian approach

EMS 2017

Helsinki

Xavier Loizeau

July 2017

In collaboration with
Jan Johannes and Fabienne Comte

Contents



Introduction

Suggested method

Optimality results

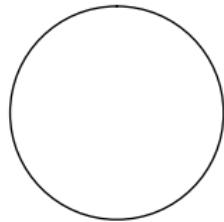
Simulations

Considered model

Circular deconvolution model

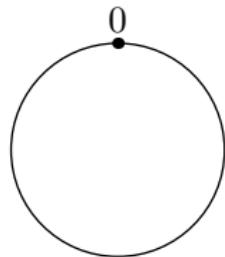


1



Considered model

Circular deconvolution model

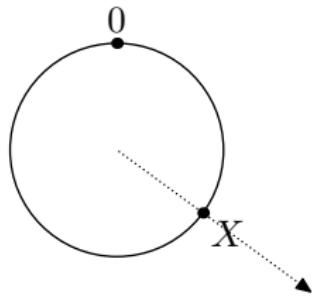


Considered model

Circular deconvolution model



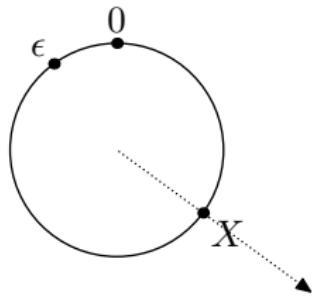
1



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

Considered model

Circular deconvolution model



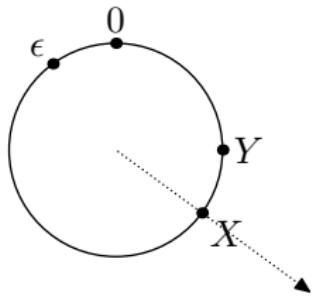
- ▶ Object of interest: density of a circular r.v. $X \sim f^X$;
- ▶ Noise: circular r.v. $\epsilon \sim f^\epsilon$;

Considered model

Circular deconvolution model



1



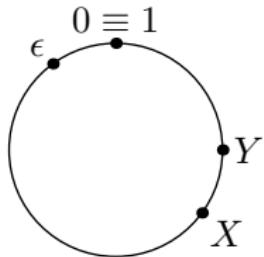
- ▶ Object of interest: density of a circular r.v. $X \sim f^X$;
- ▶ Noise: circular r.v. $\epsilon \sim f^\epsilon$;
- ▶ Observation: noisy circular r.v. $Y = X + \epsilon \sim f^Y = f^X \star f^\epsilon$

Considered model

Circular deconvolution model



1



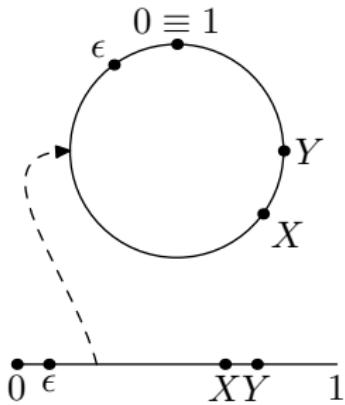
- ▶ Object of interest: density of a circular r.v. $X \sim f^X$;
- ▶ Noise: circular r.v. $\epsilon \sim f^\epsilon$;
- ▶ Observation: noisy circular r.v. $Y = X + \epsilon \sim f^Y = f^X \star f^\epsilon$

Considered model

Circular deconvolution model



1



- ▶ 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

Frequency domain



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;
- ▶ $[f^X]_j = \mathbb{E}_{f^X} [e_j(-X_1)]$, Fourier coefficients of f^X ;
- ▶ $[f^\epsilon]_j = \mathbb{E}_{f^\epsilon} [e_j(-\epsilon_1)]$, the Fourier coefficients of f^ϵ ;
- ▶ $[f^Y]_j = \mathbb{E}_{f^Y} [e_j(-Y_1)] = [f^X]_j \cdot [f^\epsilon]_j$, the Fourier coefficients of f^Y ;

Considered model

Illustration



Bayesian point of view

Bayesian fundamental paradigm



4

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

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

Bayesian point of view

Bayesian fundamental paradigm



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

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

Bayesian point of view

Bayesian fundamental paradigm



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

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

Bayesian point of view

Bayesian fundamental paradigm



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

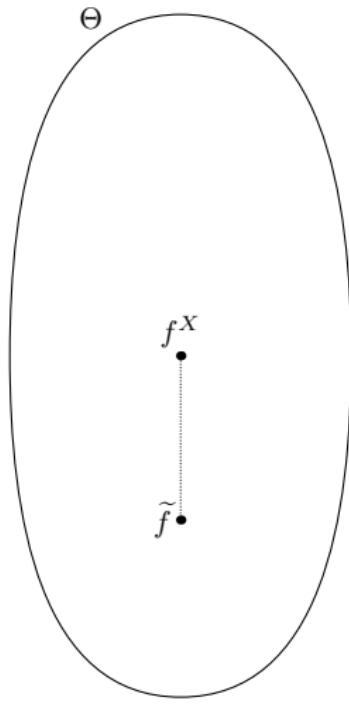
- ▶ the parameter $[f]$ is a random variable with **prior** $\mathbb{P}_{[f]}$ with density $\pi_{[f]}$ with respect to some measure;
- ▶ given $[f]$, the **likelihood** of Y is $\mathbb{P}_{Y|[f]}^n$ with density
$$\pi_{Y|[f]}^n(y_1, \dots, y_n, [f]) = \prod_{k \in [1, n]} \sum_{j \in \mathbb{Z}} [f]_j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

Formulation of optimality

Frequentist case



5



Formulation of optimality

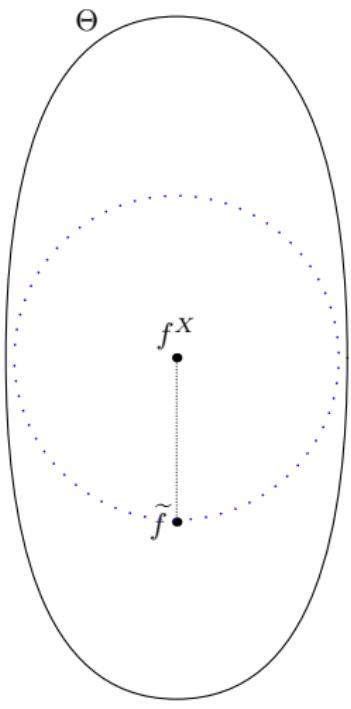
Frequentist case



5

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

$$\mathbb{E}_{f^X}^n \left[d\left(\tilde{f}, f^X\right)^2 \right].$$



Formulation of optimality

Frequentist case



5

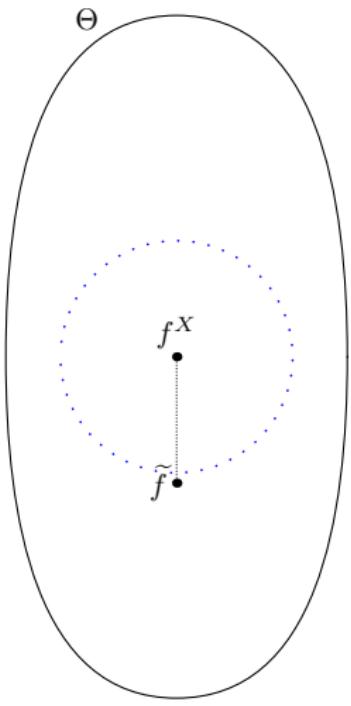
- ▶ For each frequentist estimator, measure the performance by its risk

$$\mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right].$$

- ▶ Goal: derive a lower bound for this risk

$$\inf_{\tilde{f}} \mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right] \geq \mathcal{R}_n^{\circ}(f^X)$$

over a family of estimators.

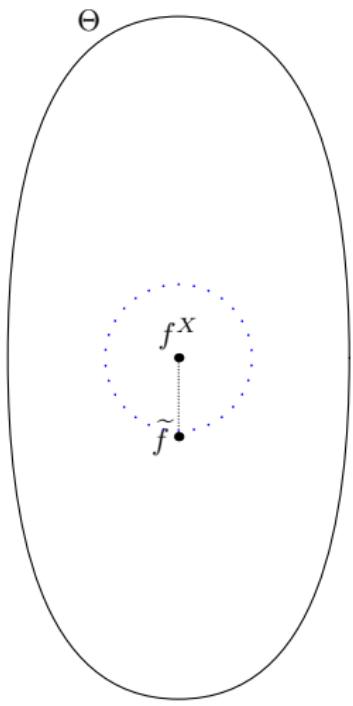


Formulation of optimality

Frequentist case



5



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

$$\mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right].$$

- ▶ Goal: derive a lower bound for this risk

$$\inf_{\tilde{f}} \mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right] \geq \mathcal{R}_n^\circ(f^X)$$

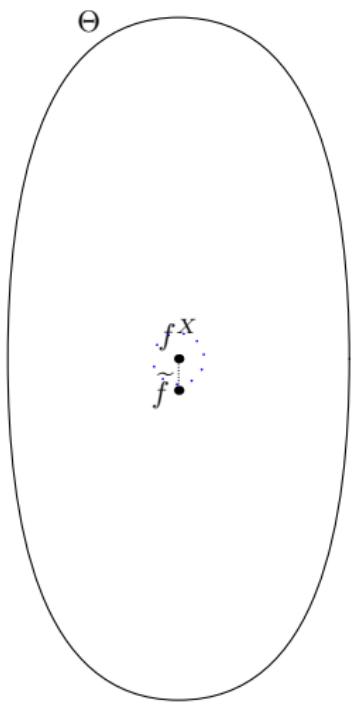
over a family of estimators.

Formulation of optimality

Frequentist case



5



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

$$\mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right].$$

- ▶ Goal: derive a lower bound for this risk

$$\inf_{\tilde{f}} \mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right] \geq \mathcal{R}_n^\circ(f^X)$$

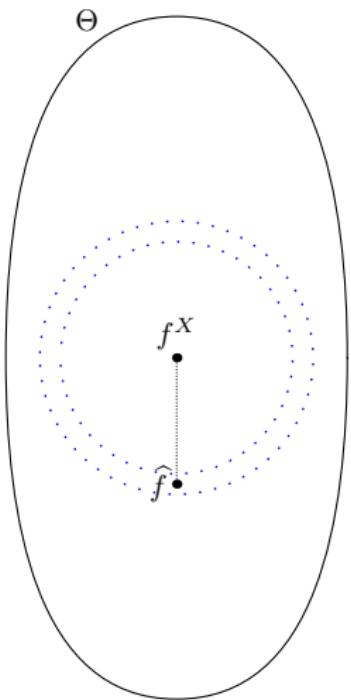
over a family of estimators.

Formulation of optimality

Frequentist case



5



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

$$\mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right].$$

- ▶ Goal: derive a lower bound for this risk

$$\inf_{\tilde{f}} \mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right] \geq \mathcal{R}_n^\circ(f^X)$$

over a family of estimators.

- ▶ Finding \hat{f} in this family such that

$$\mathbb{E}_{f^X}^n \left[d(\hat{f}, f^X)^2 \right] \leq K^\circ \cdot \mathcal{R}_n^\circ(f^X),$$

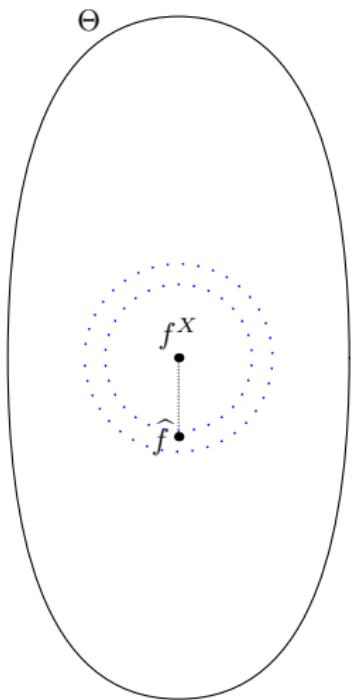
it is then **oracle rate optimal** and **adaptive** if \hat{f} does not depend on f^X .

Formulation of optimality

Frequentist case



5



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

$$\mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right].$$

- ▶ Goal: derive a lower bound for this risk

$$\inf_{\tilde{f}} \mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right] \geq \mathcal{R}_n^\circ(f^X)$$

over a family of estimators.

- ▶ Finding \hat{f} in this family such that

$$\mathbb{E}_{f^X}^n \left[d(\hat{f}, f^X)^2 \right] \leq K^\circ \cdot \mathcal{R}_n^\circ(f^X),$$

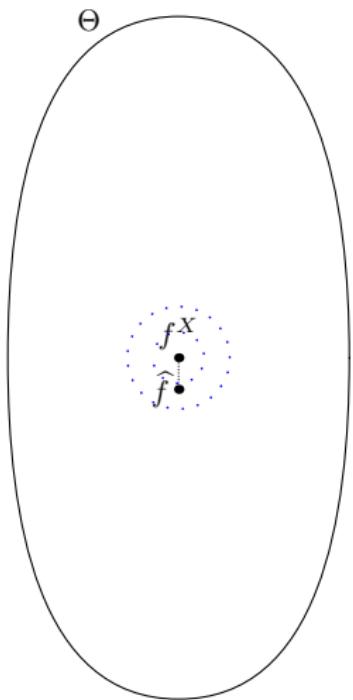
it is then **oracle rate optimal** and **adaptive** if \hat{f} does not depend on f^X .

Formulation of optimality

Frequentist case



5



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

$$\mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right].$$

- Goal: derive a lower bound for this risk

$$\inf_{\tilde{f}} \mathbb{E}_{f^X}^n \left[d(\tilde{f}, f^X)^2 \right] \geq \mathcal{R}_n^\circ(f^X)$$

over a family of estimators.

- Finding \hat{f} in this family such that

$$\mathbb{E}_{f^X}^n \left[d(\hat{f}, f^X)^2 \right] \leq K^\circ \cdot \mathcal{R}_n^\circ(f^X),$$

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

Formulation of optimality

Pragmatic Bayesian paradigm



6

How to transfer this in a Bayesian point of view ?

Taking a "pragmatic Bayesian" point of view :

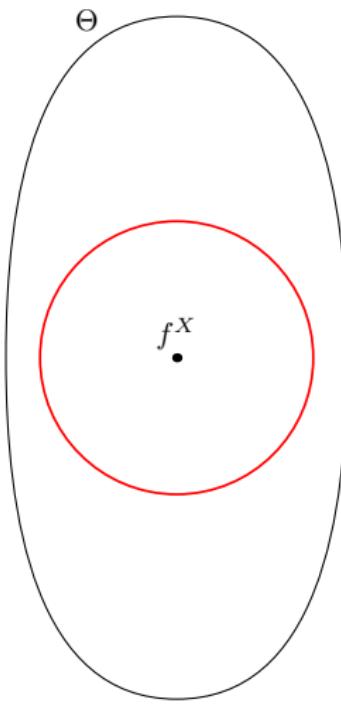
- ▶ f^X the **true parameter**.
- ▶ Is $\mathbb{P}_{f|Y}^n$ **shrinking** around f^X as n tends to ∞ ?
- ▶ How fast ?

Formulation of optimality

Pragmatic Bayesian formulation of optimality



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



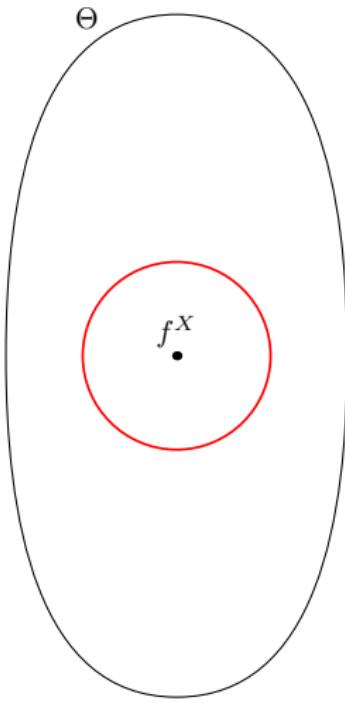
$$\exists c \in \mathbb{R}_+, \quad \lim_{n \rightarrow \infty} \mathbb{E}_{f^X}^n \left[\mathbb{P}_{f|Y}^n \left(d(f, f^X)^2 \geq c \phi_n \right) \right] = 0.$$

Formulation of optimality

Pragmatic Bayesian formulation of optimality



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



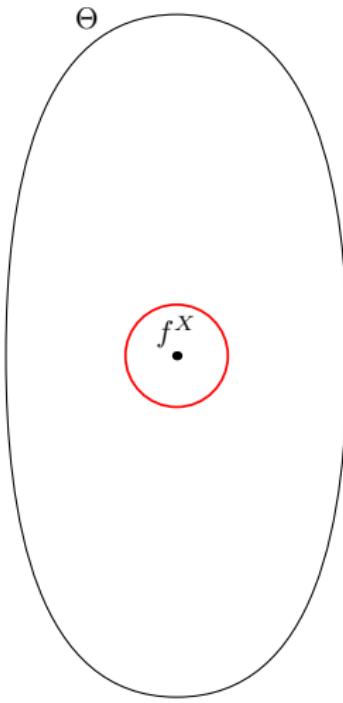
$$\exists c \in \mathbb{R}_+, \quad \lim_{n \rightarrow \infty} \mathbb{E}_{f|Y}^n \left[\mathbb{P}_{f|Y}^n \left(d(f, f^X)^2 \geq c \phi_n \right) \right] = 0.$$

Formulation of optimality

Pragmatic Bayesian formulation of optimality



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



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

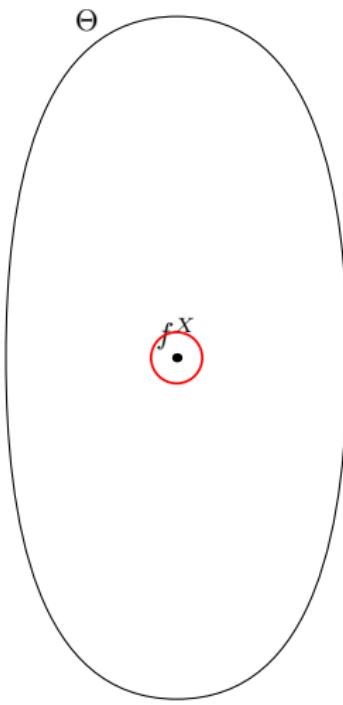
Formulation of optimality

Pragmatic Bayesian formulation of optimality



7

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



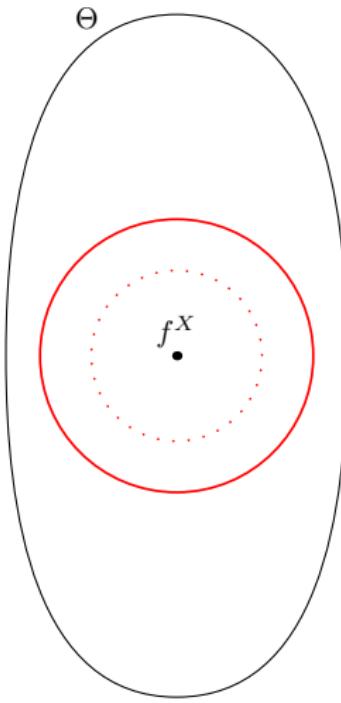
$$\exists c \in \mathbb{R}_+, \quad \lim_{n \rightarrow \infty} \mathbb{E}_{f^X}^n \left[\mathbb{P}_{f|Y}^n \left(d(f, f^X)^2 \geq c \phi_n \right) \right] = 0.$$

Formulation of optimality

Pragmatic Bayesian formulation of optimality



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



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

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

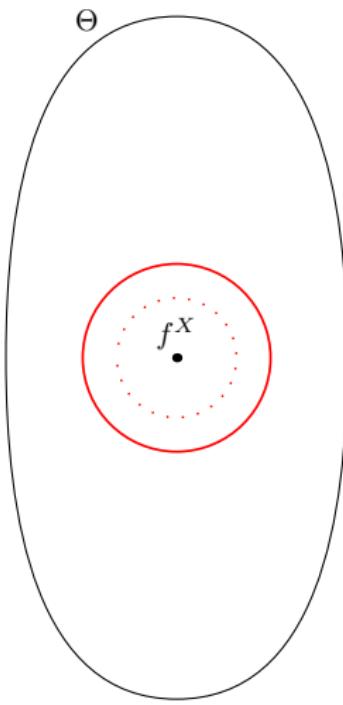
$$\lim_{n \rightarrow \infty} \mathbb{E}_{f^X}^n \left[\mathbb{P}_{f|Y}^n \left(d(f, f^X)^2 \leq c^{-1} \phi_n \right) \right] = 0.$$

Formulation of optimality

Pragmatic Bayesian formulation of optimality



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



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

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

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

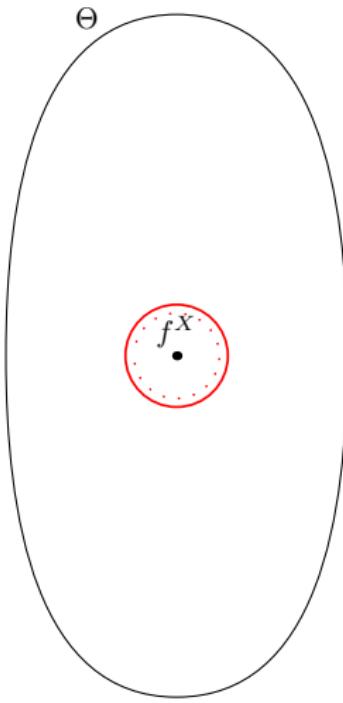
Formulation of optimality

Pragmatic Bayesian formulation of optimality



7

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



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

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

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

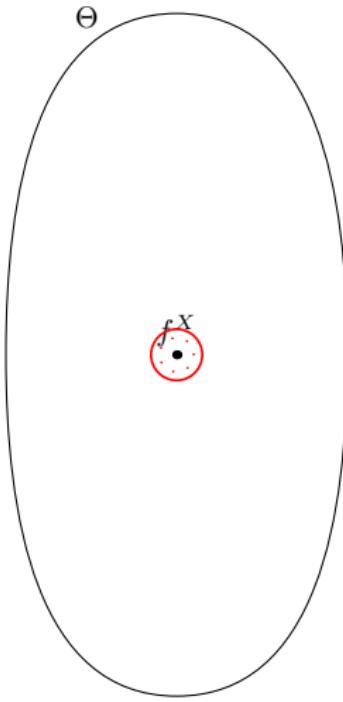
Formulation of optimality

Pragmatic Bayesian formulation of optimality



7

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



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

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

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

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 $[f^Y]_j = \mathbb{E}_{f^Y} [e_j(-Y_1)]$;

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 $\widehat{[f]}_j = \frac{\widehat{[f^Y]}_j}{\widehat{[f^e]}_j}$, 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 $\widehat{[f]}_j = \frac{\widehat{[f^Y]}_j}{\widehat{[f^e]}_j}$, 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 $\widehat{[f]}_j = \frac{\widehat{[f^Y]}_j}{\widehat{[f^e]}_j}$, 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. ?).

Illustration

Direct case



squared bias, variance and MSE

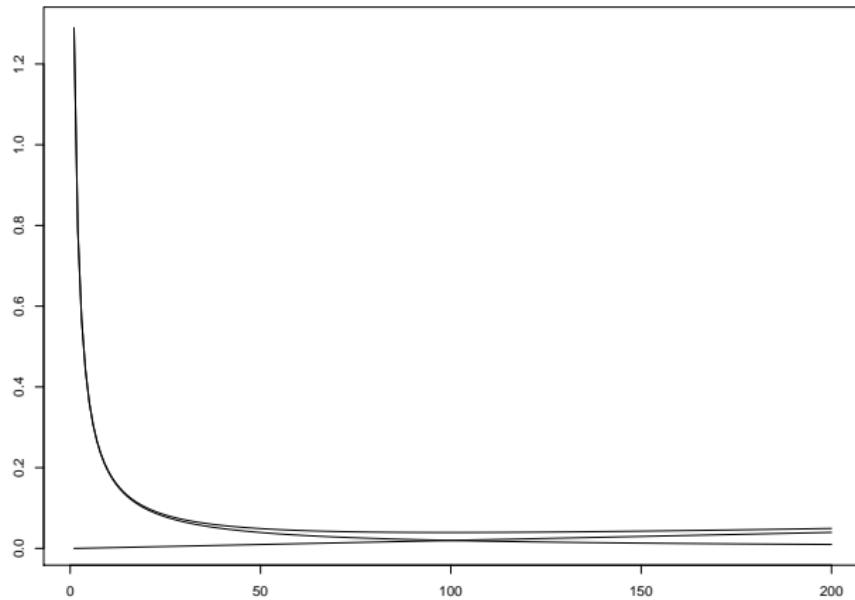


Figure: MSE of projection estimators in the direct case

Illustration

Direct case



10

Illustration

Inverse case



11

squared bias, variance and MSE

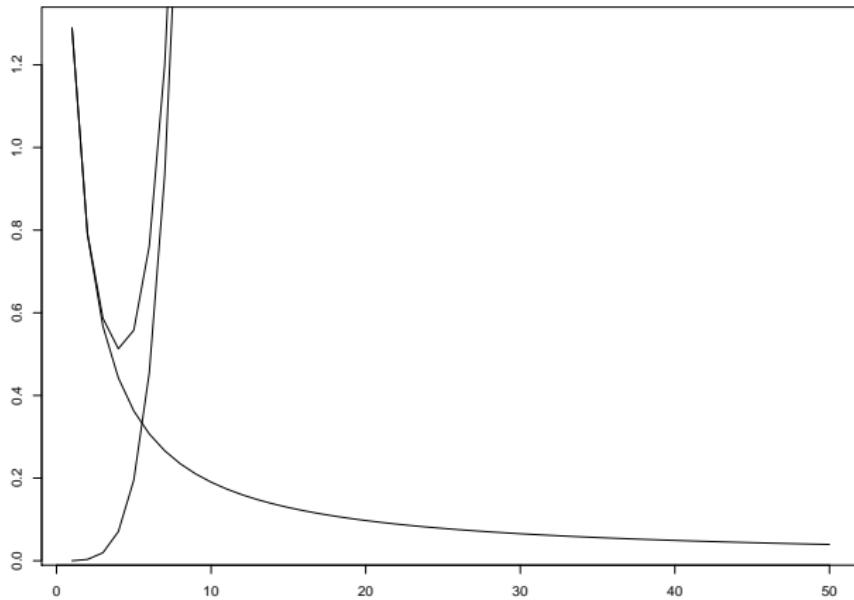


Figure: MSE of projection estimators in the mildly ill-posed case case

Considered model

Illustration



Contents



12

Introduction

Suggested method

Optimality results

Simulations



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

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



The posterior is then of the form

$$\pi_{[f]|Y^n}([f], y^n) = \prod_{|j|>m} \delta_0([f]_j) \cdot \\ \exp \left[\underbrace{\Delta_1([f]) + \sum_{k=1}^n \log \left(1 + \sum_{0<|j|\leq m} [f]_j \cdot [f^\epsilon]_j \cdot e_j(y_k) \right) + \Delta_2(y^n)}_{\exp[R([f], y^n)]} \right].$$



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

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

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

Looking for a prior

Hierarchical prior



- ▶ Consider a **random hyper-parameter M** , with values in $\llbracket 1, n \rrbracket$, acting like a threshold:

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

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

$$\pi_{M|Y^n}(m, y^n) = \exp[\text{pen}(m) - 2\Delta_2(y^n)]$$

$$\pi_{[\mathbf{f}]|Y^n} = \sum_{m \in \mathbb{N}} \pi_{[\mathbf{f}]|M=m, Y^n} \cdot \pi_{M=m|Y^n};$$

$$\widetilde{[\mathbf{f}]} = \left(\mathbb{E}_{[\mathbf{f}]|M \geq j, Y^n} \left[[\mathbf{f}]_j \right] \cdot \mathbb{P}_{M|Y^n}(M \geq j) \right)_{j \in \mathbb{N}}.$$

Hierarchical prior

Graphical representation



Contents



Introduction

Suggested method

Optimality results

Simulations

Optimality results

Definitions



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 ?).



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

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

$$\forall n \in \mathbb{N}^*, \quad \mathbb{E}_{[f^X]}^n \left[\left\| [\widetilde{f}] - [f^X] \right\|^2 \right] \leq C^\circ \Phi_n^\circ.$$



Theorem J & C & L [2017]

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

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

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.



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

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

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

Contents



21

Introduction

Suggested method

Optimality results

Simulations

Direct case : visualisation of the estimates



Direct case : sampling from the approximated posterior



23

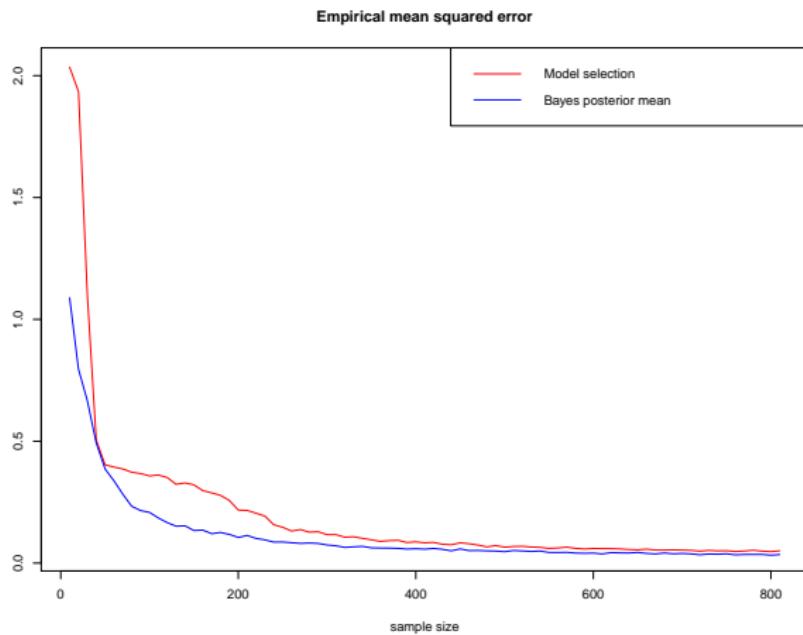


Figure: Evolution of the empirical mean squared error of the two estimates with respect to the number of observations in the direct case

Direct case : empirical mean square errors of the estimates



24

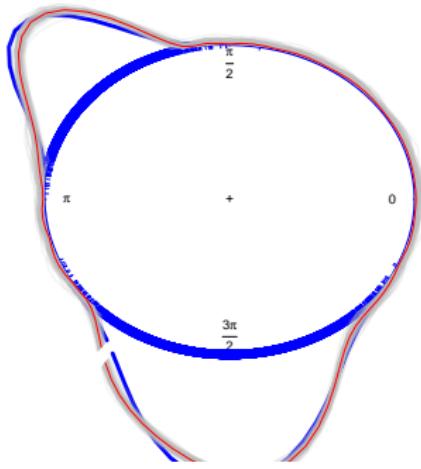


Figure: Sample from the posterior distribution and the posterior mean in the direct case



- ▶ 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