Ruprecht-Karls-Universität Heidelberg

Bayesian minimax and oracle optimality in an inverse Gaussian sequence space model

Xavier Loizeau

9^{th} *of March 2016*

In collaboration with Jan Johannes

Contents

Consider an indirect Gaussian sequence space model consisting of :

$$Y = \left(\theta^\circ_j \cdot \lambda_j + \sqrt{n}^{-1} \cdot \xi_j\right)_{j\in\mathbb{N}}, \quad (\xi_j)_{j\in\mathbb{N}} \sim_{iid} \mathcal{N}(0,1).$$

The goal is to recover θ° and prove asymptotic optimality.

Consider an indirect Gaussian sequence space model consisting of :

- an unknown parameter of interest $\left(\theta_j^\circ\right)_{j\in\mathbb{N}} = \theta^\circ$;

$$Y = \left(\theta_j^\circ \cdot \lambda_j + \sqrt{n}^{-1} \cdot \xi_j\right)_{j\in\mathbb{N}}, \quad (\xi_j)_{j\in\mathbb{N}} \sim_{iid} \mathcal{N}(0,1).$$

The goal is to recover θ° and prove asymptotic optimality.

Consider an indirect Gaussian sequence space model consisting of :

- an unknown parameter of interest $\left(\theta_j^\circ\right)_{j\in\mathbb{N}} = \theta^\circ$;
- a polynomially decreasing multiplicative sequence $(\lambda_j)_{j\in\mathbb{N}} = \lambda$ converging to 0 ;

$$Y = \left(\theta_j^\circ \cdot \lambda_j + \sqrt{n}^{-1} \cdot \xi_j\right)_{j\in\mathbb{N}}, \quad (\xi_j)_{j\in\mathbb{N}} \sim_{iid} \mathcal{N}(0,1).$$

The goal is to recover θ° and prove asymptotic optimality.

Consider an indirect Gaussian sequence space model consisting of :

- an unknown parameter of interest $\left(\theta_j^\circ\right)_{j\in\mathbb{N}} = \theta^\circ$;
- a polynomially decreasing multiplicative sequence $(\lambda_j)_{j\in\mathbb{N}} = \lambda$ converging to 0 ;
- observations $(Y_j)_{j\in\mathbb{N}} = Y$, contaminated by an additive independent centered Gaussian noise with variance n^{-1} ;

$$Y = \left(\theta_j^\circ \cdot \lambda_j + \sqrt{n}^{-1} \cdot \xi_j\right)_{j\in\mathbb{N}}, \quad (\xi_j)_{j\in\mathbb{N}} \sim_{iid} \mathcal{N}(0,1).$$

The goal is to recover θ° and prove asymptotic optimality.

Consider a function

$$
\begin{aligned}
u : [0,\pi] \times \mathbb{R}_+ &\to \mathbb{R} \\
(x, \quad t) &\mapsto u(x,t),
\end{aligned}
$$

such that

$$
\begin{aligned}
&\forall t \in \mathbb{R}_+, && u(0,t) = u(\pi,t) = 0; \\
&\forall (x,t) \in [0,1] \times \mathbb{R}_+, && \frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t) \quad \text{(heat equation);} \\
&\forall x \in [0,1], && u(x,0) = f(x) \quad \text{(function of interest);} \\
&\forall x \in [0,1], && u(x,T) + \frac{1}{n}\xi(x) = g(x) \quad \text{(observed function).}
\end{aligned}
$$

Considered model
Background

Define a function basis for $\mathscr{L}^2([0,\pi])$ and decompositions of objects of interest in this basis :

$$\forall j \in \mathbb{N}, \forall x \in [0,\pi] \quad \phi_j(x) := \sqrt{\frac{2}{\pi}} \sin(j \cdot x);$$

$$\forall j \in \mathbb{N} \quad \theta_j^\circ := \sqrt{\frac{2}{\pi}} \int_0^\pi f(\tau) \sin(j \cdot \tau) d\tau$$

$$\forall x \in [0,\pi] \quad f(x) = \sum_{j=0}^{\infty} \theta_j^\circ \phi_j(x);$$

$$\forall j \in \mathbb{N}, \forall t \in]0,T] \quad a_j(t) := \sqrt{\frac{2}{\pi}} \int_0^\pi u(\tau,t) \sin(j \cdot \tau) d\tau$$

$$\forall x \in [0,\pi], \forall t \in]0,T] \quad u(x,t) = \sum_{j=0}^{\infty} a_j(t) \phi_j(x);$$

Using the differential equation fulfilled by u, the fact that $\{\phi_j\}$ is orthogonal composed of eigen vectors for differentiation, we get

$$\begin{aligned}&\forall t \in [0,T], \forall j \in \mathbb{N}, \quad a_j'(t) = j^2 a_j(t);\\&a_j(0) = \theta_j^\circ;\\&a_j(T) = \theta_j^\circ \exp\left[-j^2 T\right] = \theta_j^\circ \cdot \lambda_j.\end{aligned}$$

Because our observation is noised by a white noise we can actually only compute $(Y_i)_{i\in\mathbb{N}}$ with :

$$\forall j \in \mathbb{N}, \quad Y_j = a_j(T) + \frac{1}{n}\xi_j = \theta_j^\circ \cdot \lambda_j + \frac{1}{j}\cdot \xi_j,$$

where the ξ_j are *iid* with distribution $\mathcal{N}(0,1)$.

$$\sum_{j\geq 1} \theta_j^\circ \lambda_j^t \cdot \psi_j(x);$$

$$\sum_{j\geq 1} Y_j \cdot \psi_j(x);$$

$$\lambda_j^t = \exp\left[-(j+1)^2 \cdot t\right]$$

$$\lambda_j^t = \exp\left[-2 \cdot \log(j+1) \cdot t\right]$$

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

- for any j in $\mathbb{N}$, consider an unbiased estimator $\widetilde{\theta}_j = \frac{Y_j}{\lambda_j} = \theta^\circ + \frac{1}{\sqrt{n}\cdot\lambda_j}\xi_j$;

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

- for any j in $\mathbb{N}$, consider an unbiased estimator $\widetilde{\theta}_j = \frac{Y_j}{\lambda_j} = \theta^\circ + \frac{1}{\sqrt{n} \cdot \lambda_j} \xi_j$;
- family of projection estimators $\left\{ \left(\widetilde{\theta}_j^m \right)_{j \in \mathbb{N}} : \forall j \in \mathbb{N}, \quad \widetilde{\theta}_j^m = \widetilde{\theta}_j \mathbb{1}_{\{j \leq m\}} \right\}$;

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

- for any j in $\mathbb{N}$, consider an unbiased estimator $\widetilde{\theta}_j = \frac{Y_j}{\lambda_j} = \theta^\circ + \frac{1}{\sqrt{n}\cdot\lambda_j}\xi_j$;
- family of projection estimators $\left\{\left(\widetilde{\theta}^m_j\right)_{j\in\mathbb{N}} : \forall j \in \mathbb{N}, \quad \widetilde{\theta}^m_j = \widetilde{\theta}_j \mathbb{1}_{\{j\leq m\}}\right\}$;
- define a method to select m in a satisfactory way.

A popular frequentist method
Model selection

To do so, use model selection by penalized contrast :

- select $\widehat{m} := \underset{m \in [\![1,n]\!]}{\arg\min} \left\{ pen(m) + \gamma(m) \right\}$;

Thus an estimator is defined by : $\widehat{\theta} = \left(\widetilde{\theta}_j^{\widehat{m}} \right)_{j \in \mathbb{N}}$.

? gives an overview of this method.

A popular frequentist method
Model selection

To do so, use model selection by penalized contrast :

- select $\widehat{m} := \underset{m \in [\![1,n]\!]}{\arg\min} \{pen(m) + \gamma(m)\}$;
- for any m in $[\![1,n]\!]$ define $pen(m) := 3 \cdot m$;

Thus an estimator is defined by : $\widehat{\theta} = \left(\widetilde{\theta}_j^{\widehat{m}}\right)_{j \in \mathbb{N}}$.

? gives an overview of this method.

To do so, use model selection by penalized contrast :

- select $\widehat{m} := \underset{m \in [\![1,n]\!]}{\arg\min} \{pen(m) + \gamma(m)\}$;
- for any m in $[\![1,n]\!]$ define $pen(m) := 3 \cdot m$;
- for any m in $[\![1,n]\!]$ define $\gamma(m) := -\sum_{j=1}^{m} Y_j^2$.

Thus an estimator is defined by : $\widehat{\theta} = \left(\widetilde{\theta}_j^{\widehat{m}}\right)_{j \in \mathbb{N}}$.

? gives an overview of this method.

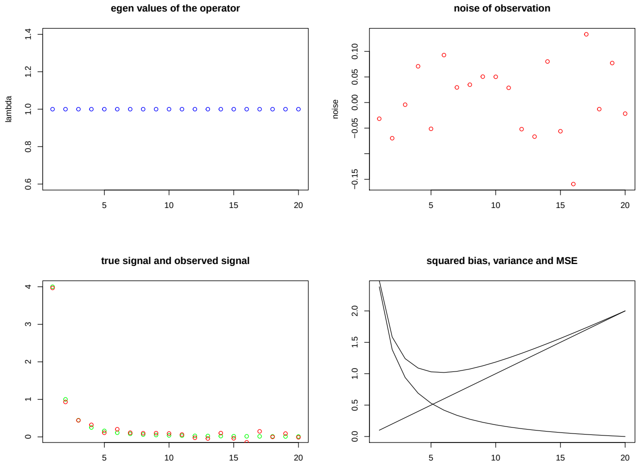


Figure: MSE of projection estimators in the direct case

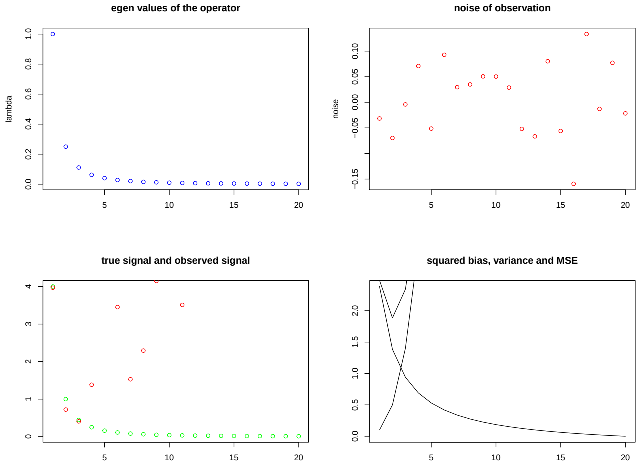


Figure: MSE of projection estimators in the mildly ill-posed case

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

- the parameter $\boldsymbol{\theta}$ is a random variable with prior $\mathbb{P}_{\boldsymbol{\theta}}$;

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

- the parameter $\boldsymbol{\theta}$ is a random variable with prior $\mathbb{P}_{\boldsymbol{\theta}}$;

- given $\boldsymbol{\theta}$, the likelihood of Y is $\mathbb{P}^{n}_{Y|\boldsymbol{\theta}} = \mathcal{N}\left(\boldsymbol{\theta} \cdot \lambda, n^{-1}\mathscr{I}\right)$;

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

- the parameter $\boldsymbol{\theta}$ is a random variable with prior $\mathbb{P}_{\boldsymbol{\theta}}$;
- given $\boldsymbol{\theta}$, the likelihood of Y is $\mathbb{P}^n_{Y|\boldsymbol{\theta}} = \mathcal{N}\left(\boldsymbol{\theta}\cdot\lambda, n^{-1}\mathscr{I}\right)$;
- we are interested in the posterior distribution $\mathbb{P}^n_{\boldsymbol{\theta}|Y} \propto \mathbb{P}^n_{Y|\boldsymbol{\theta}} \cdot \mathbb{P}_{\boldsymbol{\theta}}$.

In the spirit of **?**, we then generate a posterior family by introducing an iteration parameter η :

- for $\eta = 1$, the prior distribution is $\mathbb{P}_{\boldsymbol{\theta}^1} = \mathbb{P}_{\boldsymbol{\theta}}$, the likelihood $\mathbb{P}^n_{Y^1|\boldsymbol{\theta}^1} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and the posterior distribution is $\mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1} = \mathbb{P}^n_{\boldsymbol{\theta}|Y}$;

In the spirit of **?**, we then generate a posterior family by introducing an iteration parameter η :

- for $\eta = 1$, the prior distribution is $\mathbb{P}_{\boldsymbol{\theta}^1} = \mathbb{P}_{\boldsymbol{\theta}}$, the likelihood $\mathbb{P}^n_{Y^1|\boldsymbol{\theta}^1} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and the posterior distribution is $\mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1} = \mathbb{P}^n_{\boldsymbol{\theta}|Y}$;
- for $\eta = 2$, we take the posterior for $\eta = 1$ as prior, hence, the prior is $\mathbb{P}^n_{\boldsymbol{\theta}^2} = \mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1}$, the likelihood does not change $\mathbb{P}^n_{Y^2|\boldsymbol{\theta}^2} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and we compute the posterior with the same observations Y, which we note $\mathbb{P}^n_{\boldsymbol{\theta}^2|Y^2}$;

In the spirit of **?**, we then generate a posterior family by introducing an iteration parameter η :

- for $\eta = 1$, the prior distribution is $\mathbb{P}_{\boldsymbol{\theta}^1} = \mathbb{P}_{\boldsymbol{\theta}}$, the likelihood $\mathbb{P}^n_{Y^1|\boldsymbol{\theta}^1} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and the posterior distribution is $\mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1} = \mathbb{P}^n_{\boldsymbol{\theta}|Y}$;
- for $\eta = 2$, we take the posterior for $\eta = 1$ as prior, hence, the prior is $\mathbb{P}^n_{\boldsymbol{\theta}^2} = \mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1}$, the likelihood does not change $\mathbb{P}^n_{Y^2|\boldsymbol{\theta}^2} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and we compute the posterior with the same observations Y, which we note $\mathbb{P}^n_{\boldsymbol{\theta}^2|Y^2}$;
- ...

In the spirit of **?**, we then generate a posterior family by introducing an iteration parameter η :

- for $\eta = 1$, the prior distribution is $\mathbb{P}_{\boldsymbol{\theta}^1} = \mathbb{P}_{\boldsymbol{\theta}}$, the likelihood $\mathbb{P}^n_{Y^1|\boldsymbol{\theta}^1} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and the posterior distribution is $\mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1} = \mathbb{P}^n_{\boldsymbol{\theta}|Y}$;
- for $\eta = 2$, we take the posterior for $\eta = 1$ as prior, hence, the prior is $\mathbb{P}^n_{\boldsymbol{\theta}^2} = \mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1}$, the likelihood does not change $\mathbb{P}^n_{Y^2|\boldsymbol{\theta}^2} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and we compute the posterior with the same observations Y, which we note $\mathbb{P}^n_{\boldsymbol{\theta}^2|Y^2}$;
- ...
- for any value of $\eta > 1$, the prior is $\mathbb{P}^n_{\boldsymbol{\theta}^\eta} = \mathbb{P}_{\boldsymbol{\theta}^{\eta-1}|Y^{\eta-1}}$ and we compute the posterior with the same likelihood $\mathbb{P}^n_{Y^\eta|\boldsymbol{\theta}^\eta} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and same observations Y which gives $\mathbb{P}^n_{\boldsymbol{\theta}^\eta|Y^\eta}$.

Considered questions

- Interpretation : giving more and more weight to observations, making prior knowledge vanish ;
- generating a family of Bayes estimates $\widehat{\theta}^{\eta} := \mathbb{E}^{n}_{\boldsymbol{\theta}^{\eta}|Y^{\eta}}[\boldsymbol{\theta}]$;
- for any η, study the behavior of $\mathbb{P}^{n}_{\boldsymbol{\theta}^{\eta}|Y^{\eta}}$ and $\widehat{\theta}^{\eta}$ as $n \to \infty$;
- give particular attention to the "selfinformative limit" $\lim_{\eta\to\infty} \widehat{\theta}^{\eta}$ and the "selfinformative Bayes carrier" $\lim_{\eta\to\infty} \mathbb{P}^{n}_{\boldsymbol{\theta}^{\eta}|Y^{\eta}}$.

Questions : optimality as $n \to \infty$? Which formulation ? For the posterior mean ? For the posterior distribution ? As $\eta \to \infty$?

Important remark

We distinguish from here two asymptotes which are NOT equivalent :

- $n \to \infty$ and η fixed ;
- $\eta \to \infty$ and n fixed.

Contents

Consider a threshold parameter m :

$$\forall j > m, \quad \mathbb{P}_{\boldsymbol{\theta}_j} = \delta_0;$$
$$\forall j \leq m, \quad \mathbb{P}_{\boldsymbol{\theta}_j} = \mathcal{N}(0,1).$$

Problem : m has to be chosen.

Hierarchical prior

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

$$\forall j > m, \quad \mathbb{P}_{\boldsymbol{\theta}_j | M=m} = \delta_0;$$
$$\forall j \leq m, \quad \mathbb{P}_{\boldsymbol{\theta}_j | M=m} = \mathcal{N}(0,1).$$

- If we denote $\mathbb{P}_M$ the distribution of M, then, the iterated posterior can be written

$$\mathbb{P}^n_{\boldsymbol{\theta}^\eta | Y^\eta} = \sum_{m \in \mathbb{N}} \mathbb{P}^n_{\boldsymbol{\theta}^\eta | M^\eta = m, Y^\eta} \cdot \mathbb{P}^n_{M^\eta = m | Y^\eta};$$
$$\widehat{\theta}^{(\eta)} = \left(\mathbb{E}^n_{\boldsymbol{\theta}^\eta | M^\eta \geq j, Y^\eta} [\boldsymbol{\theta}_j] \cdot \mathbb{P}^n_{M^\eta | Y^\eta} (M^\eta \geq j) \right)_{j \in \mathbb{N}}.$$

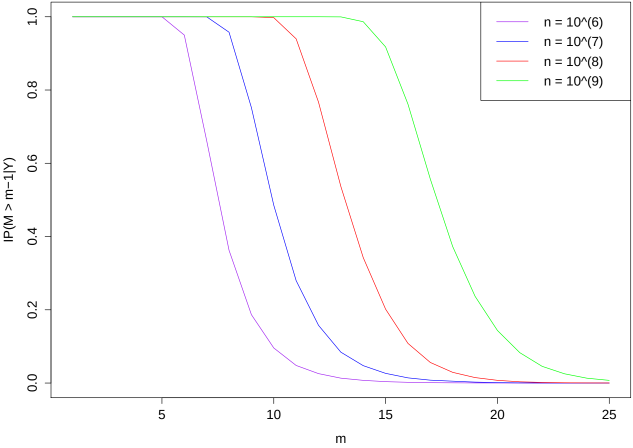


Figure: Posterior survival function of M for different values of n

Contents

Self-informative carrier/limit

Proposition : Self informative limit J & L [2016]

As η tends to ∞, the posterior mean $\widehat{\theta}^{(\eta)}$ converges almost surely towards the projection estimator given by the model selection $\widetilde{\theta}^{\widehat{m}}$.

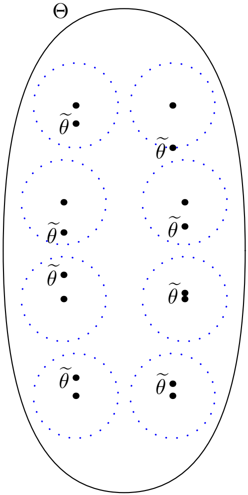
Proposition : Self informative Bayes carrier J & L [2016]

As η tends to ∞, the posterior distribution $\mathbb{P}^{n}_{\boldsymbol{\theta}^{\eta}|Y^{\eta}}$ converges towards the degenerated distribution on the projection estimator given by the model selection $\delta_{\widetilde{\theta}^{\widehat{m}}}$.

Hence, η introduces a form of continuum between the Bayes method and frequentist estimation.

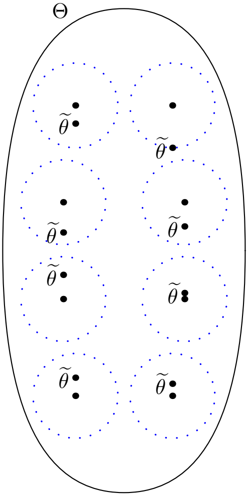
Contents

Formulation of optimality
Frequentist case



Formulation of optimality

Frequentist case

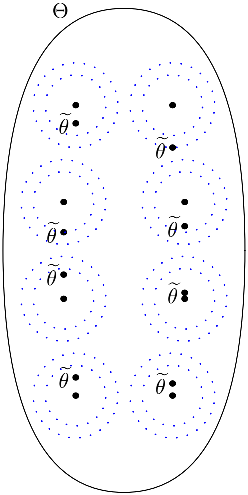


- For each frequentist estimator, consider the maximal risk over a class Θ° of parameters

$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ} \left[d\left(\widetilde{\theta}, \theta^\circ\right)^2 \right].$$

Formulation of optimality

Frequentist case



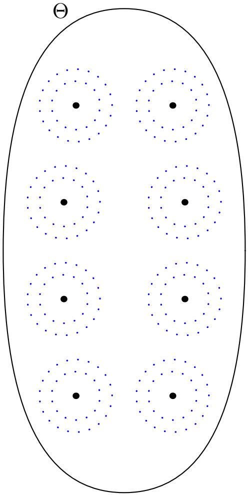
- For each frequentist estimator, consider the maximal risk over a class Θ° of parameters

$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right] \geq \mathscr{R}^\star_n\left(\Theta^\circ\right).$$

Formulation of optimality
Frequentist case

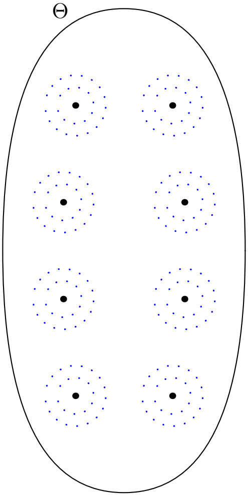


- For each frequentist estimator, consider the maximal risk over a class Θ° of parameters

$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta}, \theta^\circ\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta}, \theta^\circ\right)^2\right] \geq \mathcal{R}^\star_n\left(\Theta^\circ\right).$$

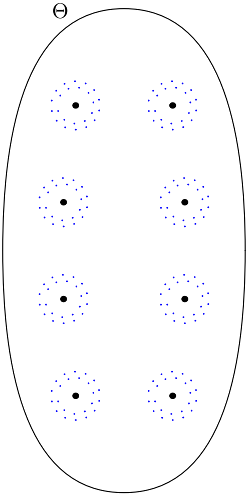


- For each frequentist estimator, consider the maximal risk over a class Θ° of parameters

$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta}, \theta^\circ\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta}, \theta^\circ\right)^2\right] \geq \mathscr{R}^\star_n\left(\Theta^\circ\right).$$

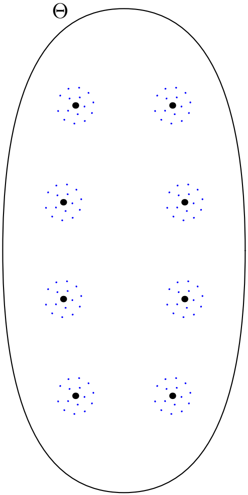


- For each frequentist estimator, consider the maximal risk over a class Θ° of parameters

$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right] \geq \mathscr{R}^\star_n\left(\Theta^\circ\right).$$



- For each frequentist estimator, consider the maximal risk over a class Θ° of parameters

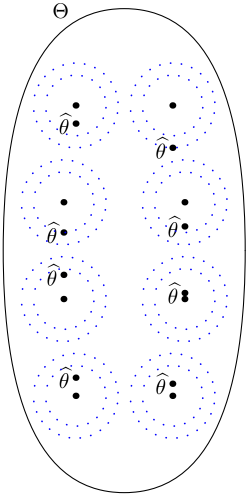
$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right] \geq \mathscr{R}^\star_n\left(\Theta^\circ\right).$$

Formulation of optimality

Frequentist case



- For each frequentist estimator, consider the maximal risk over a class Θ° of parameters

$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right].$$

- Goal : derive a lower bound for this risk

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widetilde{\theta},\theta^\circ\right)^2\right] \geq \mathscr{R}^\star_n(\Theta^\circ).$$

- Finding $\widehat{\theta}$ such that

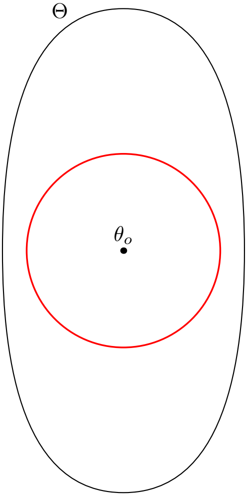
$$\sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[d\left(\widehat{\theta},\theta^\circ\right)^2\right] \leq K^\star \cdot \mathscr{R}^\star_n(\Theta^\circ),$$

it is then minimax rate optimal and adaptive if $\widehat{\theta}$ does not depend on Θ°.

How to transfer this in a Bayesian point of view ?

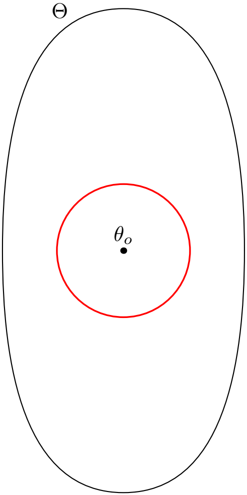
Taking a "pragmatic Bayesian" point of view :

- θ° the true parameter.
- Is $\mathbb{P}^{n}_{\boldsymbol{\theta}|Y}$ shrinking around θ° as n tends to ∞ ?
- How fast ?



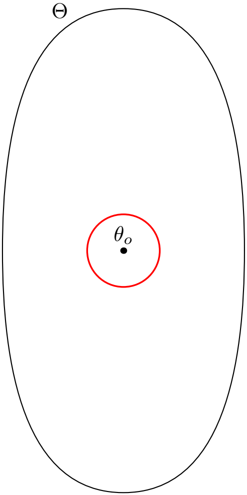
- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

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



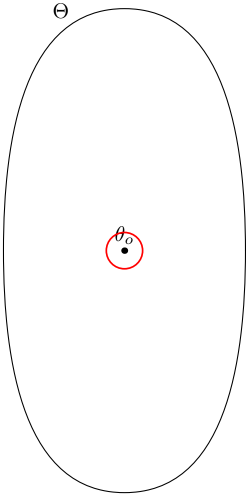
- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

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



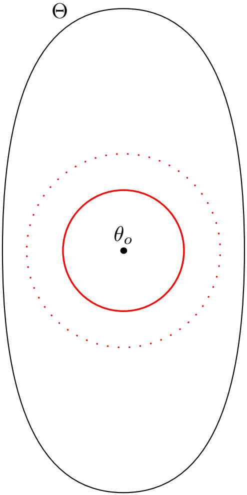
- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

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



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

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

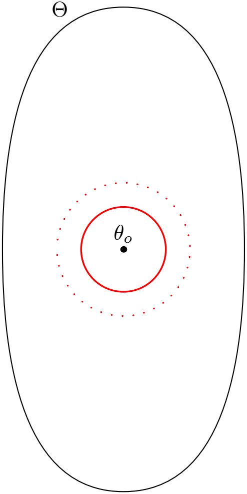


- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

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

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

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

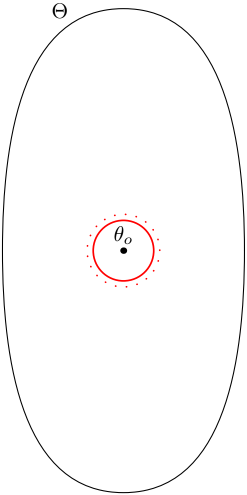


- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

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

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

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

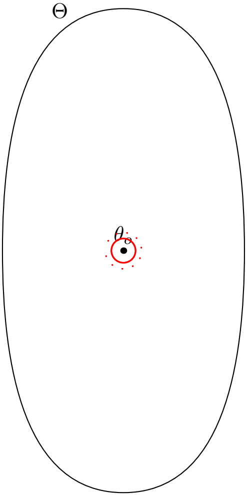


- Concentration rate $(\phi_n)_{n\in\mathbb{N}}$

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

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

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



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

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

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

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

- Uniform concentration rate $(\phi_n)_{n\in\mathbb{N}}$ over a subset on $\Theta^\circ \subset \Theta$

$$\lim_{n\to\infty} \sup_{\theta^\circ\in\Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\boldsymbol{\theta}|Y}\left(d\left(\boldsymbol{\theta},\theta^\circ\right)^2 \geq c\,\phi_n\right)\right] = 0.$$

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

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

We here note :

- $\Phi_n^{\circ} = \Phi_n^{\circ}(\theta^{\circ})$ the optimal convergence rate for the family of projection estimators for each θ° ;
- $\Phi_n^{\star} = \Phi_n^{\star}(\Theta^{\circ})$ the minimax optimal convergence rate over some Sobolev ellipsoid Θ°.

Those objects have analytical formulation (see for example **?**).

Theorem : oracle optimality J & L [2016]

For all θ° in Θ° and all fixed η in $\mathbb{N}^\star \cup \infty$, there exist C° such that

$$\forall n, \in \mathbb{N}^\star, \quad \mathbb{E}^n_{\theta^\circ}\left[\|\widehat{\theta}^{(\eta)} - \theta^\circ\|^2\right] \leq C^\circ \Phi^\circ_n;$$

The case $\eta = 1$ was already studied in **?**.

Theorem : oracle optimality J & L [2016]

For all θ° in Θ° and all fixed η in $\mathbb{N}^\star \cup \infty$, there exist C° such that

$$\forall n, \in \mathbb{N}^\star, \quad \mathbb{E}^n_{\theta^\circ}\left[\|\widehat{\theta}^{(\eta)} - \theta^\circ\|^2\right] \leq C^\circ \Phi^\circ_n;$$

Theorem : minimax optimality J & L [2016]

For all fixed η in $\mathbb{N}^\star \cup \infty$, there exist $C^\star$ such that

$$\forall n \in \mathbb{N}^\star, \quad \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[\|\widehat{\theta}^{(\eta)} - \theta^\circ\|^2\right] \leq C^\star \Phi^\star_n.$$

The case $\eta = 1$ was already studied in **?**.

Theorem J & L [2016]

First, we underline that Φ_n° and $\Phi_n^\star$ can be considered as upper bounds for the family of sieve priors :

- $\lim_{n\to\infty} \inf_{\mathbb{Q}_{\boldsymbol{\theta}}} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{Q}_{\boldsymbol{\theta}|Y}^n \left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \geq \Phi_n^\circ \right) \right] = 1,$
- $\lim_{n\to\infty} \inf_{\mathbb{Q}_{\boldsymbol{\theta}}} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}_{\theta^\circ}^n \left[\mathbb{Q}_{\boldsymbol{\theta}|Y}^n \left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \geq \Phi_n^\star \right) \right] = 1,$

where $\inf_{\mathbb{Q}_{\boldsymbol{\theta}}}$ is taken over the family of sieve priors.

The first expression gives a purely Bayesian formulation of oracle optimality.
Could the second expression be a first step towards a purely Bayesian formulation of minimax optimality ?

Theorem : oracle concentration J & L [2016]

For all θ° in Θ° and all fixed η in $\mathbb{N}^\star \cup \infty$, there exist K° such that

$$\lim_{n\to\infty} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\boldsymbol{\theta}^\eta, M^\eta | Y^\eta}\left(\left(K^\circ\right)^{-1}\Phi^\circ_n \leq \|\boldsymbol{\theta} - \theta^\circ\|^2 \leq K^\circ \Phi^\circ_n\right)\right] = 1;$$

For $\eta \to \infty$, this gives convergence in probability.
The case $\eta = 1$ was already studied in **?**.

Theorem : oracle concentration J & L [2016]

For all θ° in Θ° and all fixed η in $\mathbb{N}^\star \cup \infty$, there exist K° such that

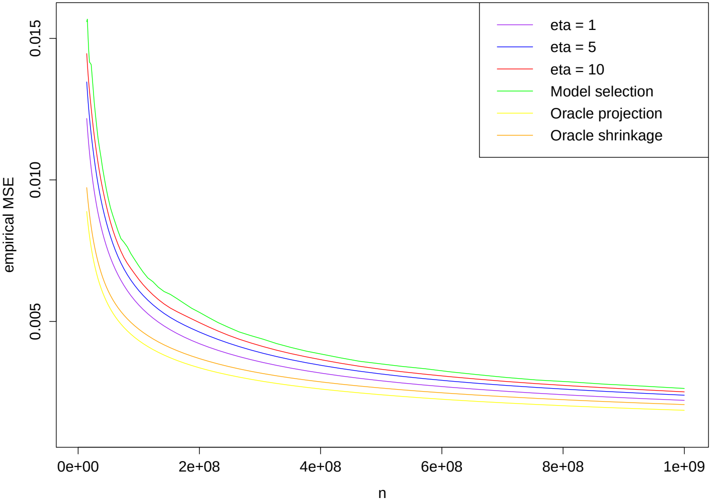
$$\lim_{n\to\infty} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\boldsymbol{\theta}^\eta, M^\eta | Y^\eta}\left(\left(K^\circ\right)^{-1}\Phi^\circ_n \leq \|\boldsymbol{\theta} - \theta^\circ\|^2 \leq K^\circ \Phi^\circ_n\right)\right] = 1;$$

Theorem : minimax concentration J & L [2016]

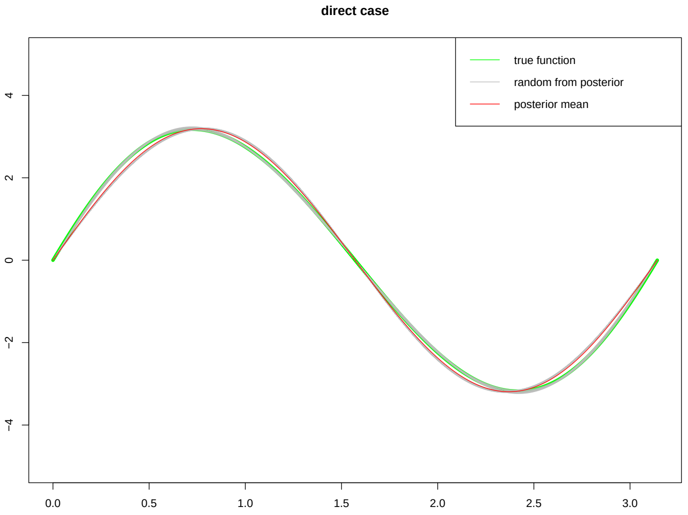
for all fixed η in $\mathbb{N}^\star \cup \infty$, and any unbounded sequence K_n, we have

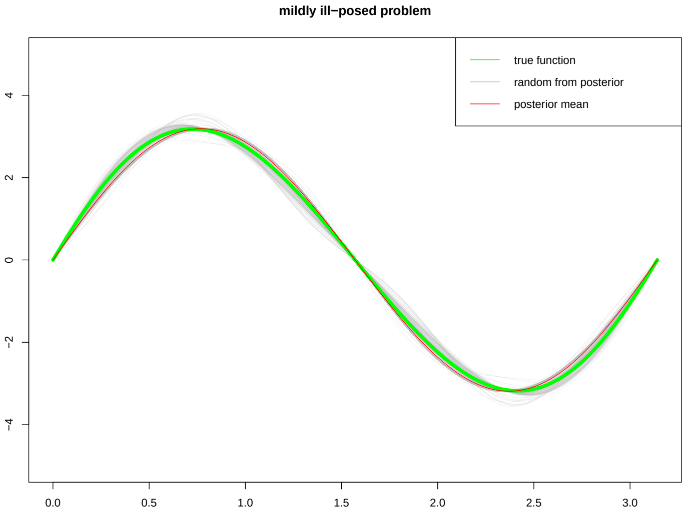
$$\lim_{n\to\infty} \sup_{\theta^\circ \in \Theta^\circ} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\boldsymbol{\theta}^\eta, M^\eta | Y^\eta}\left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \leq K_n \Phi^\star_n\right)\right] = 1.$$

For $\eta \to \infty$, this gives convergence in probability.
The case $\eta = 1$ was already studied in **?**.



eta = 1
eta = 5
eta = 10
Model selection
Oracle projection
Oracle shrinkage
0.015
0.010
0.005
empirical MSE
0e+00
2e+08
4e+08
6e+08
8e+08
1e+09
n


direct case
true function
random from posterior
posterior mean
−4
−2
0
2
4
0.0
0.5
1.0
1.5
2.0
2.5
3.0



mildly ill−posed problem
true function
random from posterior
posterior mean
4
2
0
−2
−4
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Summary

- Family of Bayesian methods indexed by an iteration parameter ;
- frequentist "model selection" method is a limit case ;
- optimality of the estimators given by the posterior means ;
- optimal concentration rate of the posterior distributions ;
- first step towards a Bayesian formulation of minimax optimality ?