Adaptive Bayesian estimation in indirect Gaussian sequence space models

Jan Johannes

Ruprecht-Karls-Universität Heidelberg

"Mathematical statistics and inverse problems"
February 2016, CIRM, Marseille

joint work with Anna Simoni and Rudolf Schenk

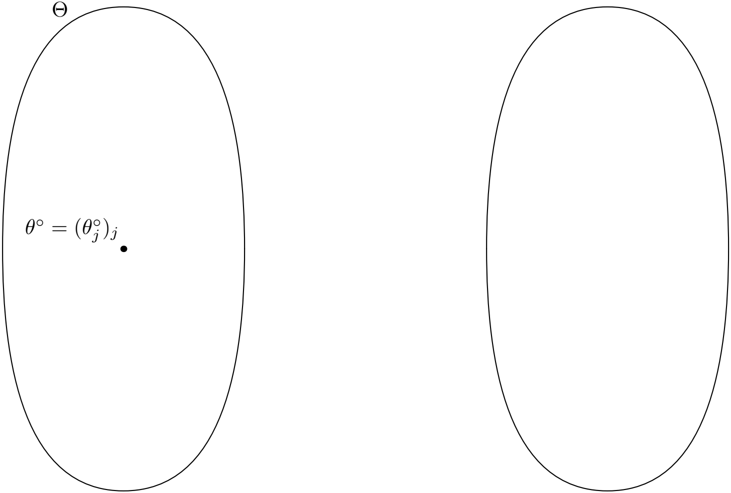
Outline

- Introduction
- Frequentist perspective reviewed
- Posterior concentration
- Adaptive posterior concentration
- Adaptive Bayes estimator

Outline

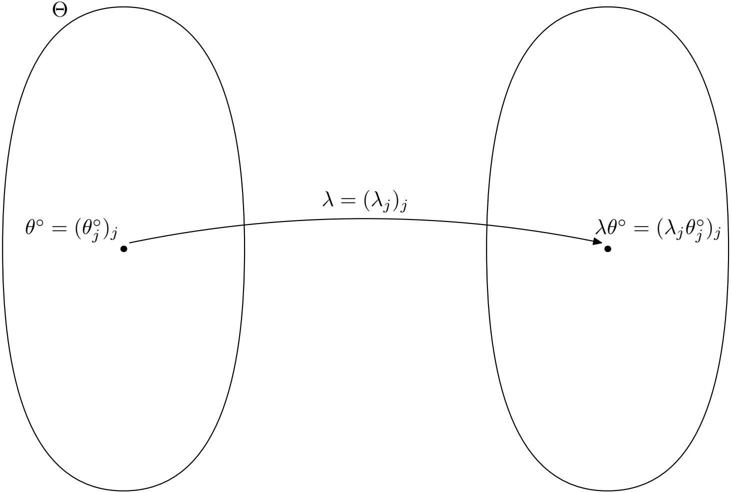
- Introduction
- Frequentist perspective reviewed
- Posterior concentration
- Adaptive posterior concentration
- Adaptive Bayes estimator

A glimpse to the essential:



Parameter of interest   Transformed parameter

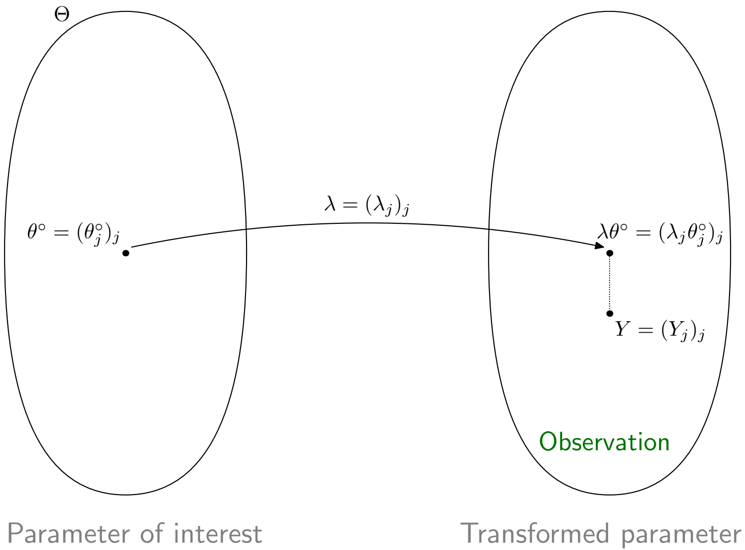
A glimpse to the essential:



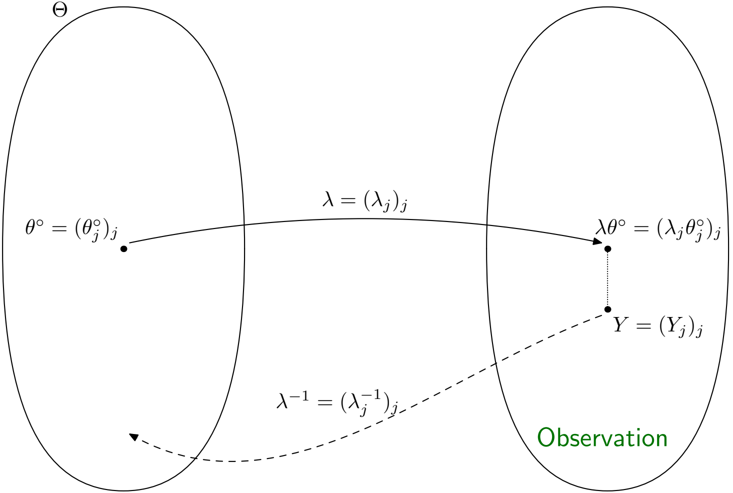
Parameter of interest

Transformed parameter

A glimpse to the essential:



A glimpse to the essential: statistical inverse problem



Parameter of interest

Transformed parameter

Background: analytical inverse problem

Consider $(\mathcal{F}, \langle\cdot,\cdot\rangle)$, $(\mathcal{G}, \langle\cdot,\cdot\rangle)$ and $T : \mathcal{F} \to \mathcal{G}$ linear

$$f \xmapsto{T} g$$

Background: analytical inverse problem

Consider $(\mathcal{F}, \langle\cdot,\cdot\rangle)$, $(\mathcal{G}, \langle\cdot,\cdot\rangle)$ and $T : \mathcal{F} \to \mathcal{G}$ linear

$$f \xmapsto{T} g$$

special case T permits a singular value decomposition

- Singular values $\lambda := (\lambda_j)_j$
- Eigenfunctions $\{u_j\}_j$ and $\{v_j\}_j$ ONB of $\mathcal{F}$ and $\mathcal{G}$, resp.

Background: analytical inverse problem

Consider $(\mathcal{F}, \langle\cdot,\cdot\rangle)$, $(\mathcal{G}, \langle\cdot,\cdot\rangle)$ and $T : \mathcal{F} \to \mathcal{G}$ linear

$$\begin{array}{ccc} f & \overset{T}{\longmapsto} & g \\ \updownarrow & & \\ (\langle f, u_j\rangle)_j & & \end{array}$$

special case T permits a singular value decomposition

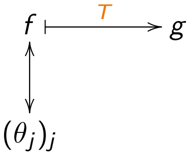
- Singular values $\lambda := (\lambda_j)_j$
- Eigenfunctions $\{u_j\}_j$ and $\{v_j\}_j$ ONB of $\mathcal{F}$ and $\mathcal{G}$, resp.

Representation

- $f \in \mathcal{F}$

Background: analytical inverse problem

Consider $(\mathcal{F}, \langle\cdot,\cdot\rangle)$, $(\mathcal{G}, \langle\cdot,\cdot\rangle)$ and $T : \mathcal{F} \to \mathcal{G}$ linear



special case T permits a singular value decomposition

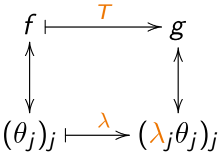
- Singular values $\lambda := (\lambda_j)_j$
- Eigenfunctions $\{u_j\}_j$ and $\{v_j\}_j$ ONB of $\mathcal{F}$ and $\mathcal{G}$, resp.

Representation

- $f \in \mathcal{F} \leftrightarrow \theta \in \Theta := \ell^2$ via $\theta_j = \langle f, u_j \rangle$

Background: analytical inverse problem

Consider $(\mathcal{F}, \langle\cdot,\cdot\rangle)$, $(\mathcal{G}, \langle\cdot,\cdot\rangle)$ and $T : \mathcal{F} \to \mathcal{G}$ linear



special case T permits a singular value decomposition

- Singular values $\lambda := (\lambda_j)_j$
- Eigenfunctions $\{u_j\}_j$ and $\{v_j\}_j$ ONB of $\mathcal{F}$ and $\mathcal{G}$, resp.

Representation

- $f \in \mathcal{F} \leftrightarrow \theta \in \Theta := \ell^2$ via $\theta_j = \langle f, u_j \rangle$
- Operator $T \leftrightarrow$ Multiplication with λ

Framework: Gaussian sequence space models

indirect GSSM

$$Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j,$$

- $\{Z_j\}_j \overset{iid}{\sim} \mathcal{N}(0, 1)$,

Framework: Gaussian sequence space models

indirect GSSM

$$Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j,$$

- $\{Z_j\}_j \overset{iid}{\sim} \mathcal{N}(0,1)$,
- $\lambda := (\lambda_j)_j \to 0$.

Framework: Gaussian sequence space models

indirect GSSM

$$Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j,$$

- $\{Z_j\}_j \overset{iid}{\sim} \mathcal{N}(0,1)$,
- $\lambda := (\lambda_j)_j \to 0$.

direct GSSM

$$\lambda_j = 1$$

Framework: Gaussian sequence space models

indirect GSSM

$$Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j,$$

- $\{Z_j\}_j \overset{iid}{\sim} \mathcal{N}(0,1)$,
- $\lambda := (\lambda_j)_j \to 0$.

direct GSSM

$$\lambda_j = 1$$

indirect Gaussian regression

$$dY = Tf + \sqrt{\varepsilon} dW$$

Framework: Gaussian sequence space models

indirect GSSM

$$Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j,$$

- $\{Z_j\}_j \overset{iid}{\sim} \mathcal{N}(0,1)$,
- $\lambda := (\lambda_j)_j \to 0$.

direct GSSM

$$\lambda_j = 1$$

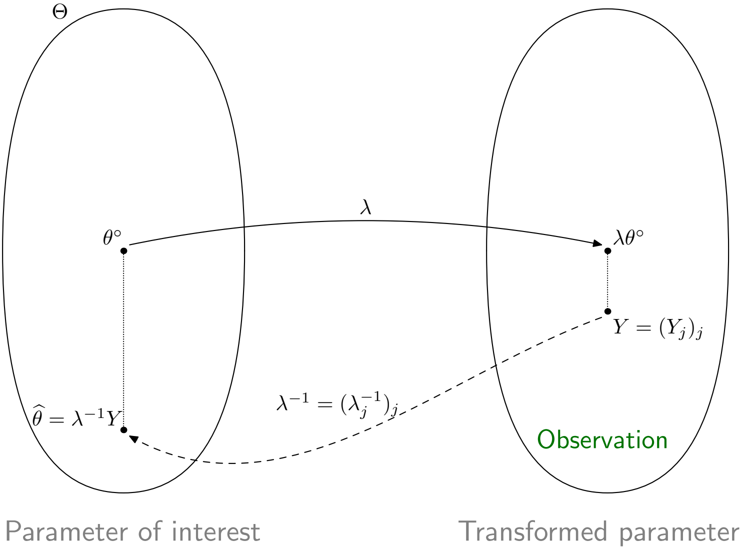
indirect Gaussian regression

$$dY = Tf + \sqrt{\varepsilon} dW$$

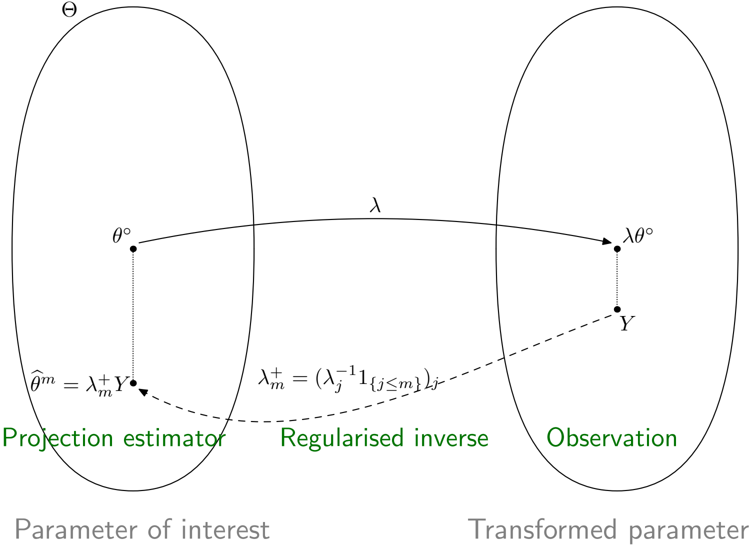
direct Gaussian regression

$$T = id$$

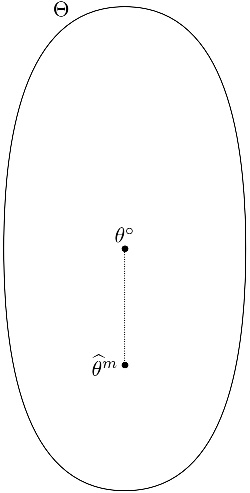
Frequentist point of view: statistical inverse problem



Frequentist point of view: regularised estimation

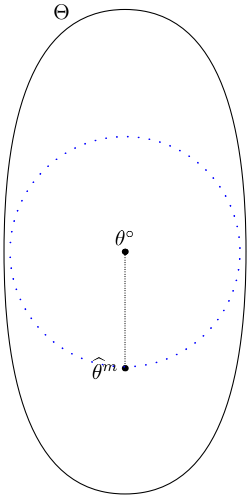


Frequentist point of view: oracle optimality



Parameter of interest

Frequentist point of view: oracle optimality

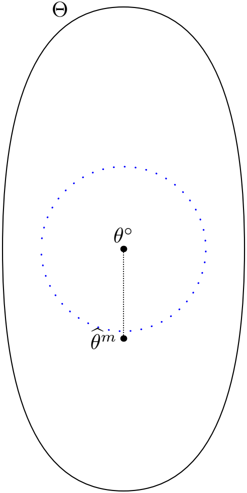


Parameter of interest

- Measure the performance – risk

$$\mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta}^m - \theta^\circ\|^2\right]$$

Frequentist point of view: oracle optimality



Parameter of interest

- Measure the performance – risk

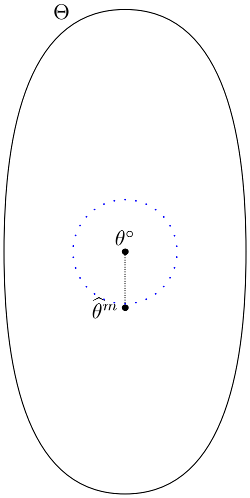
$$\mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta}^m - \theta^\circ\|^2\right]$$

- Complexity – lower bound

$$\inf_m \mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta}^m - \theta^\circ\|^2\right] \gtrsim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

over a family $\{\widehat{\theta}^m\}$ of estimators

Frequentist point of view: oracle optimality



Parameter of interest

- Measure the performance – risk

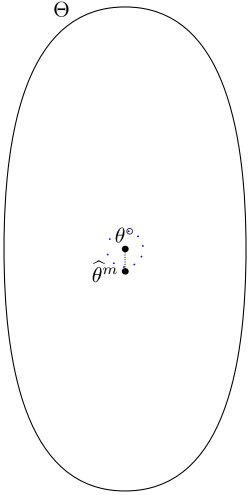
$$\mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta}^m - \theta^\circ\|^2\right]$$

- Complexity – lower bound

$$\inf_m \mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta}^m - \theta^\circ\|^2\right] \gtrsim \mathcal{R}_\varepsilon^\circ(\theta^\circ)$$

over a family $\{\widehat{\theta}^m\}$ of estimators

Frequentist point of view: oracle optimality



Parameter of interest

- Measure the performance – risk

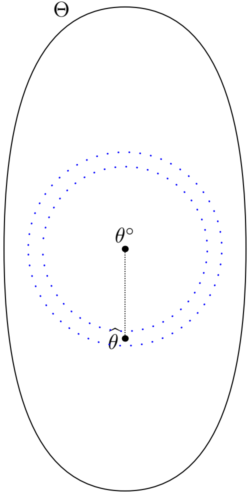
$$\mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big]$$

- Complexity – lower bound

$$\inf_m \mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big] \gtrsim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

over a family $\{\widehat{\theta}^m\}$ of estimators

Frequentist point of view: oracle optimality



Parameter of interest

- Measure the performance – risk

$$\mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big]$$

- Complexity – lower bound

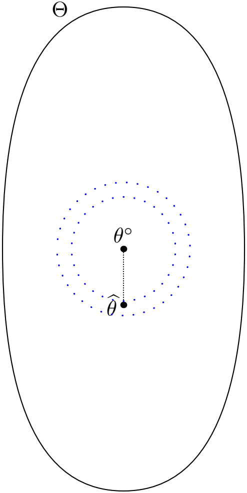
$$\inf_m \mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big] \gtrsim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

  over a family $\{\widehat{\theta}^m\}$ of estimators

- $\widehat{\theta} \in \{\widehat{\theta}^m\}$ is called an oracle if

$$\mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta} - \theta^\circ\|^2\big] \lesssim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

Frequentist point of view: oracle optimality



Parameter of interest

- Measure the performance – risk

$$\mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta}^m - \theta^\circ\|^2\right]$$

- Complexity – lower bound

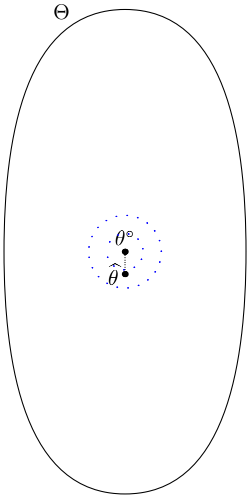
$$\inf_m \mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta}^m - \theta^\circ\|^2\right] \gtrsim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

over a family $\{\widehat{\theta}^m\}$ of estimators

- $\widehat{\theta} \in \{\widehat{\theta}^m\}$ is called an oracle if

$$\mathbb{E}_{\theta^\circ}\left[\,\|\widehat{\theta} - \theta^\circ\|^2\right] \lesssim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

Frequentist point of view: oracle optimality



Parameter of interest

- Measure the performance – risk

$$\mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big]$$

- Complexity – lower bound

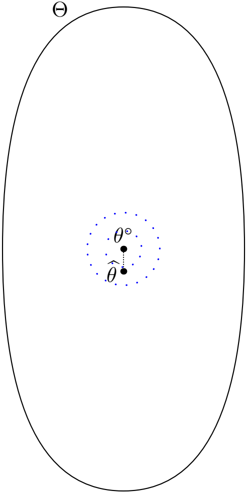
$$\inf_m \mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big] \gtrsim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

over a family $\{\widehat{\theta}^m\}$ of estimators

- $\widehat{\theta} \in \{\widehat{\theta}^m\}$ is called an oracle if

$$\mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta} - \theta^\circ\|^2\big] \lesssim \mathcal{R}^\circ_\varepsilon(\theta^\circ)$$

Frequentist point of view: oracle optimality



Parameter of interest

- Measure the performance – risk

$$\mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big]$$

- Complexity – lower bound

$$\inf_m \mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta}^m - \theta^\circ\|^2\big] \gtrsim \mathcal{R}_\varepsilon^\circ(\theta^\circ)$$

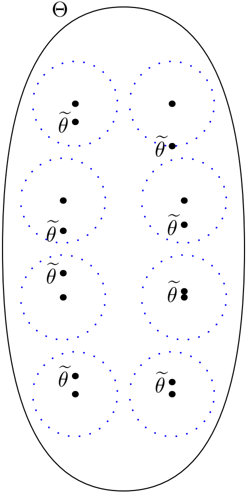
over a family $\{\widehat{\theta}^m\}$ of estimators

- $\widehat{\theta} \in \{\widehat{\theta}^m\}$ is called an oracle if

$$\mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta} - \theta^\circ\|^2\big] \lesssim \mathcal{R}_\varepsilon^\circ(\theta^\circ)$$

and adaptive if $\widehat{\theta}$ does not depend on θ°.

Frequentist point of view: minimax optimality



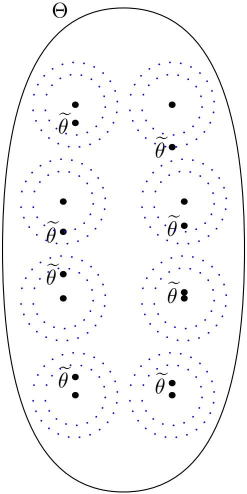
Parameter of interest

- Measure the performance – maximal risk

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\, \|\widetilde{\theta} - \theta^\circ\|^2 \right]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

Frequentist point of view: minimax optimality



Parameter of interest

- Measure the performance – maximal risk

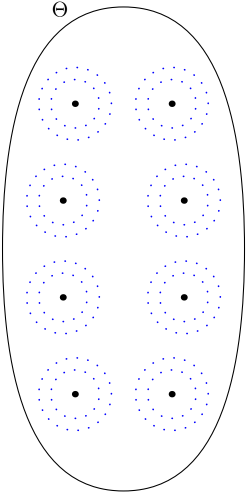
$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2\right]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

- Complexity – lower bound

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2\right] \gtrsim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

Frequentist point of view: minimax optimality



Parameter of interest

- Measure the performance – maximal risk

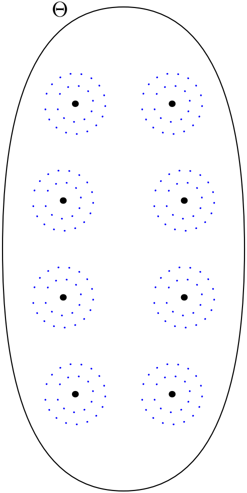
$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

- Complexity – lower bound

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right] \gtrsim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

Frequentist point of view: minimax optimality



Parameter of interest

- Measure the performance – maximal risk

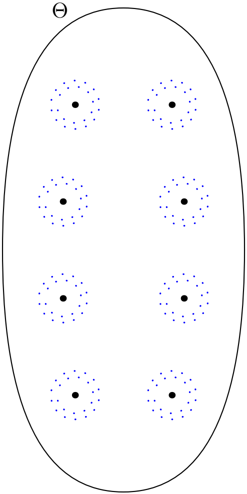
$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

- Complexity – lower bound

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right] \gtrsim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

Frequentist point of view: minimax optimality



Parameter of interest

- Measure the performance – maximal risk

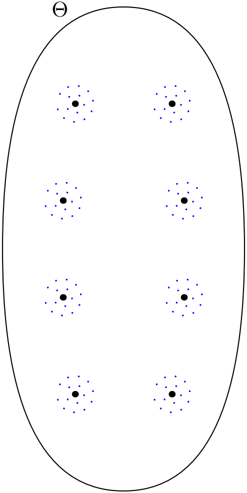
$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

- Complexity – lower bound

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right] \gtrsim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

Frequentist point of view: minimax optimality



Parameter of interest

- Measure the performance – maximal risk

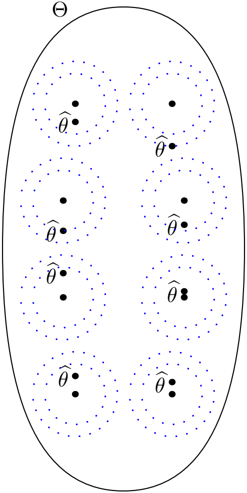
$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\, \|\widetilde{\theta} - \theta^\circ\|^2 \right]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

- Complexity – lower bound

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\, \|\widetilde{\theta} - \theta^\circ\|^2 \right] \gtrsim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

Frequentist point of view: minimax optimality



Parameter of interest

- Measure the performance – maximal risk

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

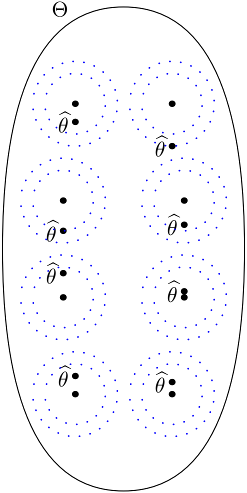
- Complexity – lower bound

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widetilde{\theta} - \theta^\circ\|^2 \right] \gtrsim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

- $\widehat{\theta}$ is called minimax rate optimal if

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\left[\|\widehat{\theta} - \theta^\circ\|^2 \right] \lesssim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

Frequentist point of view: minimax optimality



Parameter of interest

- Measure the performance – maximal risk

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\big[\,\|\widetilde{\theta} - \theta^\circ\|^2\big]$$

over a class $\Theta_{\mathfrak{a}}$ of paramters

- Complexity – lower bound

$$\inf_{\widetilde{\theta}} \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\big[\,\|\widetilde{\theta} - \theta^\circ\|^2\big] \gtrsim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

- $\widehat{\theta}$ is called minimax rate optimal if

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ}\big[\,\|\widehat{\theta} - \theta^\circ\|^2\big] \lesssim \mathcal{R}^\star_\varepsilon(\mathfrak{a})$$

and adaptive if $\widehat{\theta}$ does not depend on $\mathfrak{a}$.

A glimpse to the essential: Bayesian perspective

- θ outcome of Θ-valued r.v. $\boldsymbol{\vartheta}$

A glimpse to the essential: Bayesian perspective

- θ outcome of Θ-valued r.v. $\boldsymbol{\vartheta}$
- likelihood: $P_{Y|\boldsymbol{\vartheta}}$ with density $p_{Y|\boldsymbol{\vartheta}}$

A glimpse to the essential: Bayesian perspective

- θ outcome of Θ-valued r.v. $\boldsymbol{\vartheta}$
- likelihood: $P_{Y|\boldsymbol{\vartheta}}$ with density $p_{Y|\boldsymbol{\vartheta}}$
- prior distribution: $P_{\boldsymbol{\vartheta}}$ on Θ with density $p_{\boldsymbol{\vartheta}}$

A glimpse to the essential: Bayesian perspective

- θ outcome of Θ-valued r.v. $\boldsymbol{\vartheta}$
- likelihood: $P_{Y|\boldsymbol{\vartheta}}$ with density $p_{Y|\boldsymbol{\vartheta}}$
- prior distribution: $P_{\boldsymbol{\vartheta}}$ on Θ with density $p_{\boldsymbol{\vartheta}}$
- posterior distribution $P_{\boldsymbol{\vartheta}|Y}$ with density:

$$p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \frac{p_{Y|\boldsymbol{\vartheta}}(y|\theta)p_{\boldsymbol{\vartheta}}(\theta)}{\int_{\Theta} p_{Y|\boldsymbol{\vartheta}}(y|\theta)p_{\boldsymbol{\vartheta}}(\theta)d\theta}$$

A glimpse to the essential: Bayesian perspective

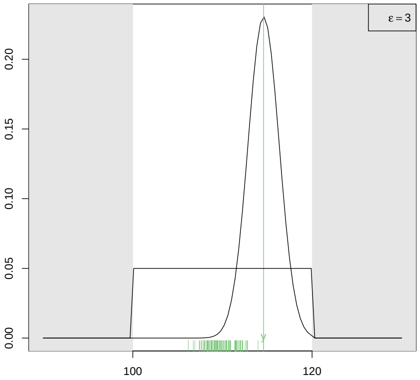
- θ outcome of Θ-valued r.v. $\boldsymbol{\vartheta}$
- likelihood: $P_{Y|\boldsymbol{\vartheta}}$ with density $p_{Y|\boldsymbol{\vartheta}}$
- prior distribution: $P_{\boldsymbol{\vartheta}}$ on Θ with density $p_{\boldsymbol{\vartheta}}$
- posterior distribution $P_{\boldsymbol{\vartheta}|Y}$ with density:

$$p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \frac{p_{Y|\boldsymbol{\vartheta}}(y|\theta)p_{\boldsymbol{\vartheta}}(\theta)}{\int_{\Theta} p_{Y|\boldsymbol{\vartheta}}(y|\theta)p_{\boldsymbol{\vartheta}}(\theta)d\theta}$$

- "posterior is proportional to likelihood times prior"

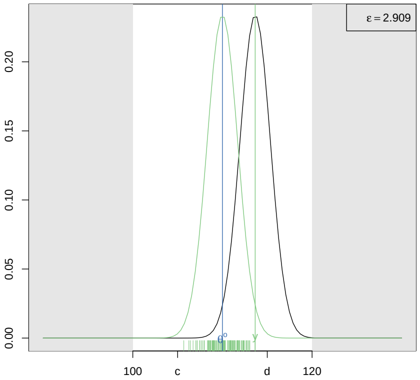
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



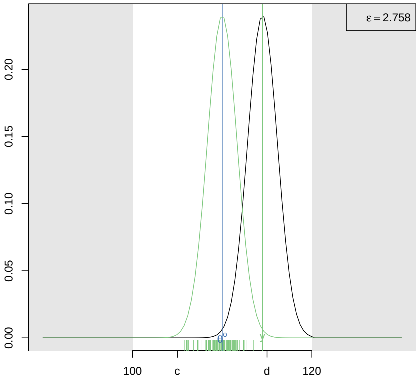
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \frac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)}\mathbb{1}_{[a,b]}(\theta)$



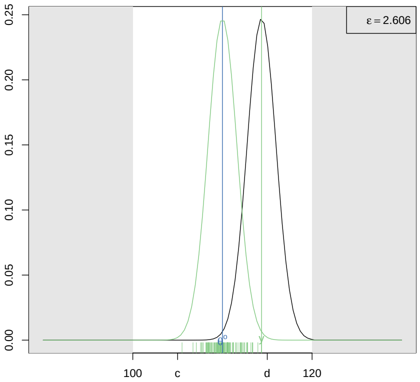
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)}\mathbb{1}_{[a,b]}(\theta)$



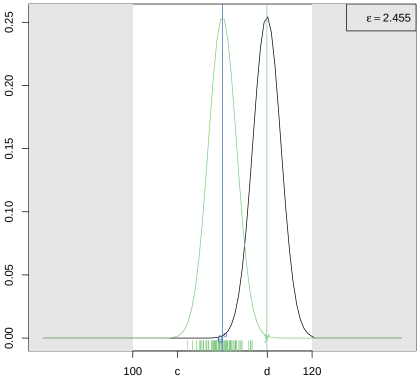
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \frac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



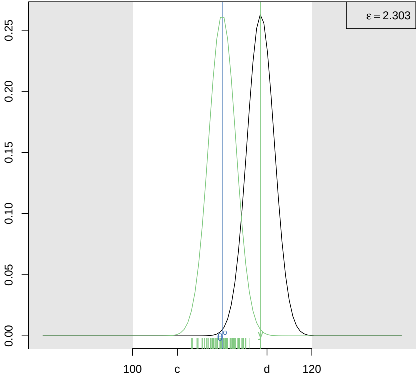
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



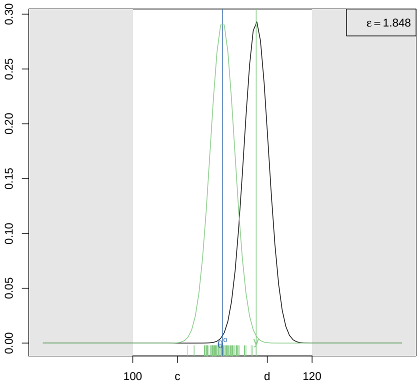
Illustration

- likelihood $Y | \boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta} | Y}(\theta | y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



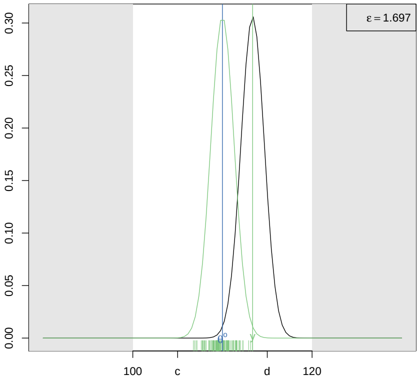
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \frac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)}\mathbb{1}_{[a,b]}(\theta)$



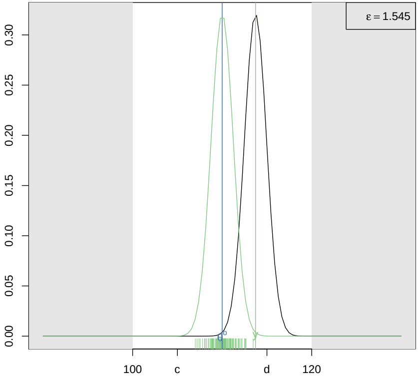
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)}\mathbb{1}_{[a,b]}(\theta)$



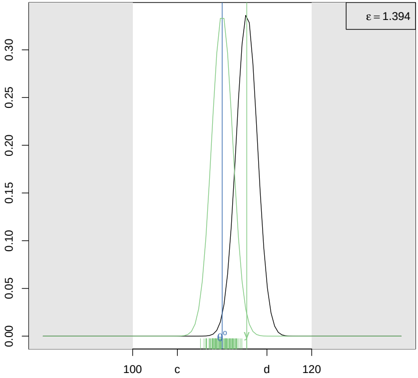
Illustration

- likelihood $Y|\,\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}\,|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



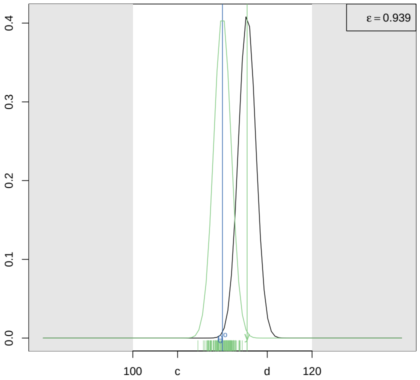
Illustration

- likelihood $Y | \boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



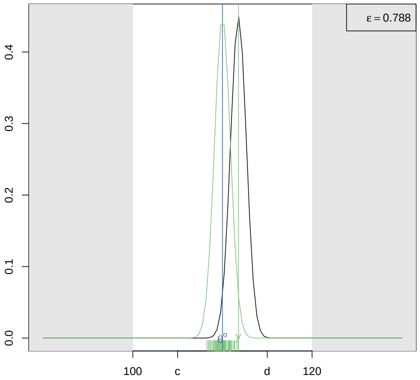
Illustration

- likelihood $Y|\,\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}\,|\,Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)}\mathbb{1}_{[a,b]}(\theta)$



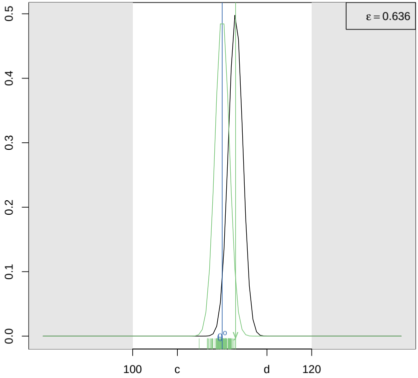
Illustration

- likelihood $Y | \boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \frac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



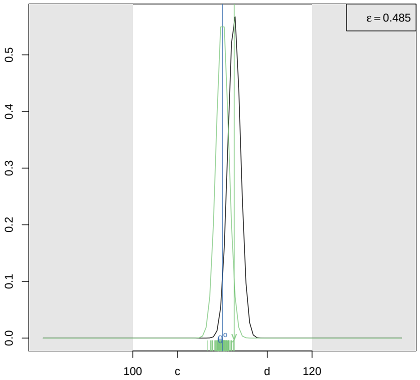
Illustration

- likelihood $Y | \boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} \mathbb{1}_{[a,b]}(\theta)$



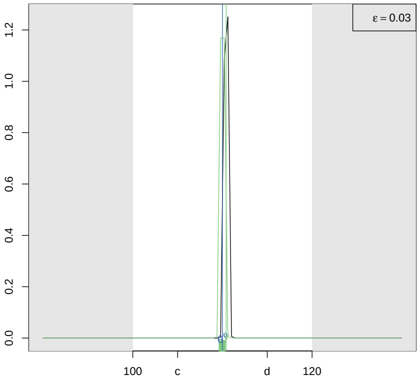
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \frac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)}\mathbb{1}_{[a,b]}(\theta)$



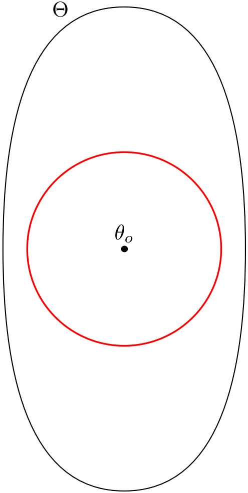
Illustration

- likelihood $Y|\boldsymbol{\vartheta} = \theta \sim \mathcal{N}(\theta, \varepsilon)$
- prior $\boldsymbol{\vartheta} \sim \mathcal{U}[a, b]$
- posterior density: $p_{\boldsymbol{\vartheta}|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)}\mathbb{1}_{[a,b]}(\theta)$



A glimpse to the essential: exact posterior concentration

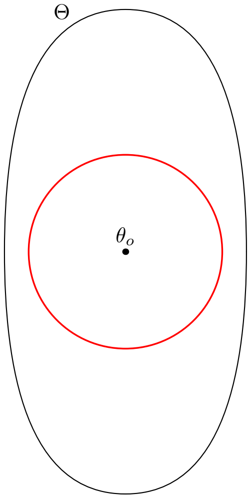
Given prior P_{ϑ} find $\mathcal{R}_{\varepsilon}$



A glimpse to the essential: exact posterior concentration

Given prior $P_{\boldsymbol{\vartheta}}$ find $\mathcal{R}_{\varepsilon} \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\boldsymbol{\vartheta}=\theta^\circ}$

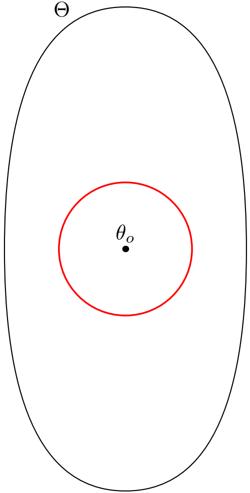
- $P_{\boldsymbol{\vartheta}|Y}(\|\boldsymbol{\vartheta} - \theta^\circ\| \lesssim \mathcal{R}_{\varepsilon}) \overset{P_{\theta^\circ}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior P_{ϑ} find $\mathcal{R}_{\varepsilon} \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\,\vartheta=\theta^\circ}$

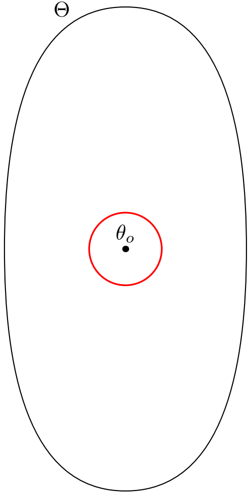
- $P_{\vartheta\,|Y}(\|\vartheta - \theta^\circ\| \lesssim \mathcal{R}_{\varepsilon}) \overset{P_{\theta^\circ}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior $P_{\boldsymbol{\vartheta}}$ find $\mathcal{R}_\varepsilon \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\boldsymbol{\vartheta}=\theta^\circ}$

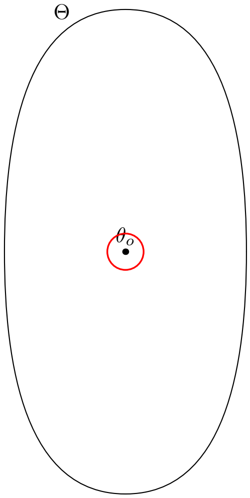
- $P_{\boldsymbol{\vartheta}|Y}(\|\boldsymbol{\vartheta} - \theta^\circ\| \lesssim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior $P_{\boldsymbol{\vartheta}}$ find $\mathcal{R}_{\varepsilon} \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\boldsymbol{\vartheta}=\theta^\circ}$

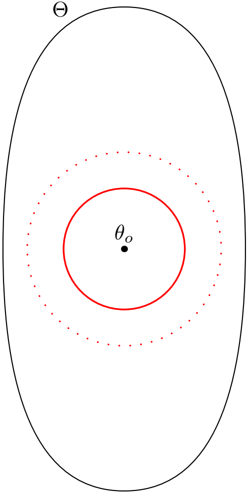
- $P_{\boldsymbol{\vartheta}|Y}(\|\boldsymbol{\vartheta} - \theta^\circ\| \lesssim \mathcal{R}_{\varepsilon}) \overset{P_{\theta^\circ}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior P_{ϑ} find $\mathcal{R}_{\varepsilon} \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^{\circ}} = P_{Y|\vartheta=\theta^{\circ}}$

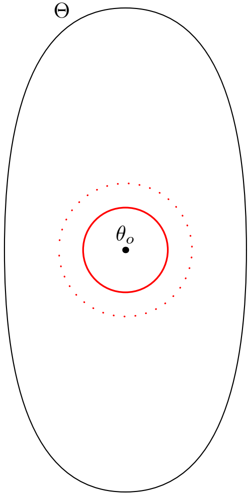
- $P_{\vartheta|Y}(\|\vartheta - \theta^{\circ}\| \lesssim \mathcal{R}_{\varepsilon}) \overset{P_{\theta^{\circ}}}{\to} 1$
- $P_{\vartheta|Y}(\|\vartheta - \theta^{\circ}\| \gtrsim \mathcal{R}_{\varepsilon}) \overset{P_{\theta^{\circ}}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior $P_{\boldsymbol{\vartheta}}$ find $\mathcal{R}_\varepsilon \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\boldsymbol{\vartheta}=\theta^\circ}$

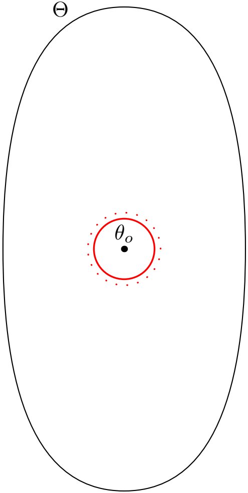
- $P_{\boldsymbol{\vartheta}|Y}(\|\boldsymbol{\vartheta} - \theta^\circ\| \lesssim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$
- $P_{\boldsymbol{\vartheta}|Y}(\|\boldsymbol{\vartheta} - \theta^\circ\| \gtrsim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior P_{ϑ} find $\mathcal{R}_{\varepsilon} \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\vartheta=\theta^\circ}$

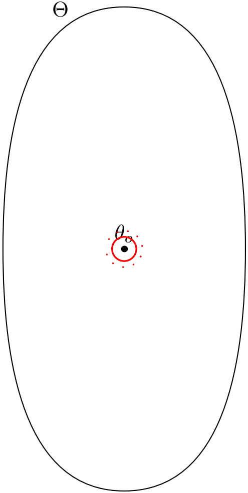
- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \lesssim \mathcal{R}_{\varepsilon}) \overset{P_{\theta^\circ}}{\to} 1$
- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \gtrsim \mathcal{R}_{\varepsilon}) \overset{P_{\theta^\circ}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior P_{ϑ} find $\mathcal{R}_\varepsilon \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\vartheta=\theta^\circ}$

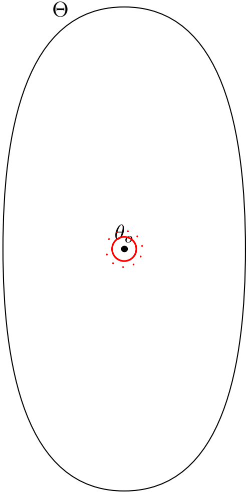
- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \lesssim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$
- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \gtrsim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$



A glimpse to the essential: exact posterior concentration

Given prior P_{ϑ} find $\mathcal{R}_\varepsilon \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\vartheta=\theta^\circ}$

- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \lesssim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$
- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \gtrsim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$



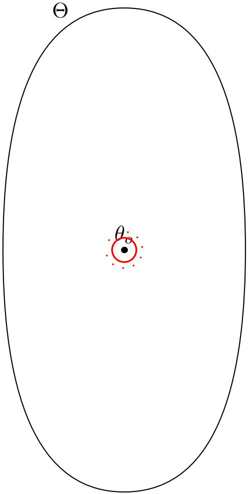
Remarks:

- $\mathcal{R}_\varepsilon$ depends on the prior

A glimpse to the essential: exact posterior concentration

Given prior P_{ϑ} find $\mathcal{R}_\varepsilon \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\vartheta=\theta^\circ}$

- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \lesssim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$
- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \gtrsim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$



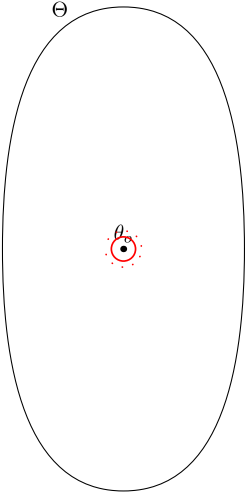
Remarks:

- $\mathcal{R}_\varepsilon$ depends on the prior
- $\mathcal{R}_\varepsilon$ might be arbitrarily slow

A glimpse to the essential: exact posterior concentration

Given prior P_{ϑ} find $\mathcal{R}_\varepsilon \to 0$ as $\varepsilon \to 0$ such that for $P_{\theta^\circ} = P_{Y|\vartheta=\theta^\circ}$

- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \lesssim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$
- $P_{\vartheta|Y}(\|\vartheta - \theta^\circ\| \gtrsim \mathcal{R}_\varepsilon) \overset{P_{\theta^\circ}}{\to} 1$



Remarks:

- $\mathcal{R}_\varepsilon$ depends on the prior
- $\mathcal{R}_\varepsilon$ might be arbitrarily slow
- consistency could fail

S. Ghosal, J.K. Ghosh and A.W. Van Der Vaart (2000) *Convergence rates of posterior distributions.* The Annals of Statistics 28(2):500–531

X. Shen and L. Wasserman. (2001) *Rates of convergence of posterior distributions.* The Annals of Statistics, 29:687–714

B. Knapik, A. Van der Vaart, and J. Van Zanten (2011) *Bayesian inverse problems with gaussian priors.* The Annals of Statistics, 39:2626–2657

J. Arbel, G. Gayraud and J. Rousseau (2013) *Bayesian optimal adaptive estimation using a sieve prior.* Scandinavian Jounral of Statistics, 40:549–570

K. Ray (2013) *Bayesian inverse problems with non-conjugate priors.* Electronic Journal of Statistics, 7: 2516–2549

M. Hoffmann, J. Rousseau and J. Schmidt-Hieber (2015) *On adpative posterior concentration rates.* The Annals of Statistics, 43:5,2259–2295

I. Castillo (2008) *Lower bounds for posterior rates with Gaussian process priors.* Electronic Journal of Statistics, 2:1281-1299

Objective: exact posterior concentration

- Construct a family $\{P_{\boldsymbol{\vartheta}^m}\}_m$ of prior distributions with exact posterior concentration $\mathcal{R}_\varepsilon^{m_\varepsilon} := \mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ)$ for $\theta^\circ \in \Theta$, i.e.,

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} | Y}\Big(\mathcal{R}_\varepsilon^{m_\varepsilon} \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^{m_\varepsilon}\Big) = 1;$$

Objective: exact posterior concentration

- Construct a family $\{P_{\vartheta^m}\}_m$ of prior distributions with exact posterior concentration $\mathcal{R}_\varepsilon^{m_\varepsilon} := \mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ)$ for $\theta^\circ \in \Theta$, i.e.,

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\vartheta^{m_\varepsilon}|Y}\Big(\mathcal{R}_\varepsilon^{m_\varepsilon} \lesssim \|\vartheta^{m_\varepsilon} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^{m_\varepsilon}\Big) = 1;$$

  - Consider for a given $\theta^\circ \in \Theta$ the oracle rate

$$\mathcal{R}_\varepsilon^\circ := \mathcal{R}_\varepsilon^\circ(\theta^\circ) := \inf_m \mathcal{R}_\varepsilon^m(\theta^\circ).$$

Objective: exact posterior concentration

- Construct a family $\{P_{\boldsymbol{\vartheta}^m}\}_m$ of prior distributions with exact posterior concentration $\mathcal{R}_\varepsilon^{m_\varepsilon} := \mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ)$ for $\theta^\circ \in \Theta$, i.e.,

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \mid Y}\Big(\mathcal{R}_\varepsilon^{m_\varepsilon} \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^{m_\varepsilon}\Big) = 1;$$

  - Consider for a given $\theta^\circ \in \Theta$ the oracle rate

$$\mathcal{R}_\varepsilon^\circ := \mathcal{R}_\varepsilon^\circ(\theta^\circ) := \inf_m \mathcal{R}_\varepsilon^m(\theta^\circ).$$

  - Consider for a given $\Theta_{\mathfrak{a}} \subset \Theta$ the minimax rate

$$\mathcal{R}_\varepsilon^\star := \mathcal{R}_\varepsilon^\star(\mathfrak{a}) := \inf_m \sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathcal{R}_\varepsilon^m(\theta^\circ).$$

Objective: optimal prior choice

- A sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ with exact posterior concentration $\mathcal{R}_\varepsilon^\circ$, i.e.

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}|Y}\Big(\mathcal{R}_\varepsilon^\circ \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon^\circ} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ\Big) = 1,$$

is called oracle prior and adaptive if it does not depend on θ°.

Objective: optimal prior choice

- A sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ with exact posterior concentration $\mathcal{R}_\varepsilon^\circ$, i.e.

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}|Y}\Big(\mathcal{R}_\varepsilon^\circ \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon^\circ} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ\Big) = 1,$$

  is called oracle prior and adaptive if it does not depend on θ°.

- A sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ with exact posterior concentration $\mathcal{R}_\varepsilon^\star$, i.e.

$$\lim_{\varepsilon\to 0} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}|Y}\Big(\mathcal{R}_\varepsilon^\star \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star\Big) = 1.$$

  is called minimax prior and adaptive if it depends not on $\Theta_{\mathfrak{a}}$.

Outline

- Introduction
- Frequentist perspective reviewed
- Posterior concentration
- Adaptive posterior concentration
- Adaptive Bayes estimator

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator

---

- Risk: $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 = \sum_{j=1}^m \mathbb{V}\mathrm{ar}\, \widehat{\theta}_j^m + \sum_{j>m} (\theta_j^\circ)^2$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta^\circ_j + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}^m_j)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator

---

- Risk: $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 = \varepsilon \sum_{j=1}^m \lambda_j^{-2} + \sum_{j>m} (\theta^\circ_j)^2$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta^\circ_j + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}^m_j)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$,

- Risk: $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 = \varepsilon \sum_{j=1}^m \Lambda_j + \sum_{j>m} (\theta^\circ_j)^2$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}, \quad \overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$

- Risk: $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 = \varepsilon m \overline{\Lambda}_m + \sum_{j>m} (\theta_j^\circ)^2$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\quad \overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

- Risk: $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 = \varepsilon m \overline{\Lambda}_m + \mathfrak{b}_m^2$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j\geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\quad \overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\quad \mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Risk: $[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] \leqslant \mathbb{E}_{\theta^\circ}\|\widehat{\theta}^m - \theta^\circ\|^2 = \varepsilon m \overline{\Lambda}_m + \mathfrak{b}_m^2$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\quad \overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

- Risk: $[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] \leqslant \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] =: \mathcal{R}_\varepsilon^m(\theta^\circ)$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\quad \overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

- Risk: $[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] \leqslant \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] =: \mathcal{R}_\varepsilon^m(\theta^\circ)$

- oracle cut-off $m_\varepsilon^\circ := m_\varepsilon^\circ(\theta^\circ) := \arg\min_{m \geqslant 1} \{\mathcal{R}_\varepsilon^m(\theta^\circ)\}$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\quad \overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

- Risk: $[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] \leqslant \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] =: \mathcal{R}_\varepsilon^m(\theta^\circ)$

- oracle cut-off $m_\varepsilon^\circ := m_\varepsilon^\circ(\theta^\circ) := \arg\min_{m \geqslant 1} \{\mathcal{R}_\varepsilon^m(\theta^\circ)\}$
- oracle rate $\mathcal{R}_\varepsilon^\circ := \mathcal{R}_\varepsilon^\circ(\theta^\circ) := [\varepsilon m_\varepsilon^\circ \overline{\Lambda}_{m_\varepsilon^\circ} \vee \mathfrak{b}_{m_\varepsilon^\circ}^2] = \min_{m \geqslant 1} \mathcal{R}_\varepsilon^m(\theta^\circ)$

Projection estimator reviewed: oracle optimality

Observations $Y_j = \lambda_j \theta_j^\circ + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (\widehat{\theta}_j^m)_{j \geqslant 1} = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\quad \overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

- Risk: $[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] \leqslant \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] =: \mathcal{R}_\varepsilon^m(\theta^\circ)$

- oracle cut-off $m_\varepsilon^\circ := m_\varepsilon^\circ(\theta^\circ) := \arg\min_{m \geqslant 1} \{\mathcal{R}_\varepsilon^m(\theta^\circ)\}$
- oracle rate $\mathcal{R}_\varepsilon^\circ := \mathcal{R}_\varepsilon^\circ(\theta^\circ) := [\varepsilon m_\varepsilon^\circ \overline{\Lambda}_{m_\varepsilon^\circ} \vee \mathfrak{b}_{m_\varepsilon^\circ}^2] = \min_{m \geqslant 1} \mathcal{R}_\varepsilon^m(\theta^\circ)$
- oracle estimator $\widehat{\theta}^{m_\varepsilon^\circ}$, i.e., $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\circ} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ$

Illustration: typical smoothness condition

Consider a non-increasing weight sequence $\mathfrak{a} = (\mathfrak{a}_j)_{j \geqslant 1}$ and let

$$\Theta_{\mathfrak{a}} := \left\{ \theta \in \Theta : \sum_{j=1}^{\infty} \theta_j^2 / \mathfrak{a}_j =: \|\theta\|_{\mathfrak{a}}^2 \leqslant r \right\}$$

Illustration: typical smoothness condition

Consider a non-increasing weight sequence $\mathfrak{a} = (\mathfrak{a}_j)_{j \geqslant 1}$ and let

$$\Theta_{\mathfrak{a}} := \left\{ \theta \in \Theta : \sum_{j=1}^{\infty} \theta_j^2 / \mathfrak{a}_j =: \|\theta\|_{\mathfrak{a}}^2 \leqslant r \right\}$$

Illustration

- $(u_j)_j$ trigonometric basis
- $\mathfrak{a}_1 = 1, \mathfrak{a}_{2j} = (2j)^{-2p}, \mathfrak{a}_{2j+1} = (2j)^{-2p}$, $p > 0$

Illustration: typical smoothness condition

Consider a non-increasing weight sequence $\mathfrak{a} = (\mathfrak{a}_j)_{j\geqslant 1}$ and let

$$\Theta_{\mathfrak{a}} := \left\{ \theta \in \Theta : \sum_{j=1}^{\infty} \theta_j^2 / \mathfrak{a}_j =: \|\theta\|_{\mathfrak{a}}^2 \leqslant r \right\}$$

Illustration

- $(u_j)_j$ trigonometric basis
- $\mathfrak{a}_1 = 1, \mathfrak{a}_{2j} = (2j)^{-2p}, \mathfrak{a}_{2j+1} = (2j)^{-2p}$, $p > 0$
  - Sobolev ellipsoid of p-times differentiable, absolutely continuous, periodic functions

Illustration: typical smoothness condition

Consider a non-increasing weight sequence $\mathfrak{a} = (\mathfrak{a}_j)_{j \geqslant 1}$ and let

$$\Theta_{\mathfrak{a}} := \left\{ \theta \in \Theta : \sum_{j=1}^{\infty} \theta_j^2 / \mathfrak{a}_j =: \|\theta\|_{\mathfrak{a}}^2 \leqslant r \right\}$$

Illustration

- $(u_j)_j$ trigonometric basis
- $\mathfrak{a}_1 = 1, \mathfrak{a}_{2j} = (2j)^{-2p}, \mathfrak{a}_{2j+1} = (2j)^{-2p}$, $p > 0$
  - Sobolev ellipsoid of p-times differentiable, absolutely continuous, periodic functions
- $\mathfrak{a}_j = \exp(1 - |j|^{2p})$, $p > \frac{1}{2}$
  - analytical functions

Projection estimator reviewed: minimax optimality

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Risk: $\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2]$

Projection estimator reviewed: minimax optimality

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Risk: $\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{b}_m^2] \leqslant (1 \vee r)[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]$

Projection estimator reviewed: minimax optimality

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Maximal risk: $\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]$

Projection estimator reviewed: minimax optimality

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Maximal risk: $\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]$

- minimax cut-off $m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}) := \arg\min_{m \geqslant 1}\{[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]\}$

Projection estimator reviewed: minimax optimality

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

- Maximal risk: $\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]$

- minimax cut-off $m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}) := \arg\min_{m \geqslant 1} \{[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]\}$
- minimax rate $\mathcal{R}_\varepsilon^\star := \mathcal{R}_\varepsilon^\star(\mathfrak{a}) := [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$

Projection estimator reviewed: minimax optimality

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

▸ Maximal risk: $\sup_{\theta^\circ \in \Theta_\mathfrak{a}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]$

- minimax cut-off $m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}) := \arg\min_{m \geqslant 1} \{[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]\}$
- minimax rate $\mathcal{R}_\varepsilon^\star := \mathcal{R}_\varepsilon^\star(\mathfrak{a}) := [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$
- minimax estimator $\widehat{\theta}^{m_\varepsilon^\star}$, i.e., $\sup_{\theta^\circ \in \Theta_\mathfrak{a}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\star} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star$

Projection estimator reviewed: minimax optimality

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Maximal risk: $\sup_{\theta^\circ \in \Theta_\mathfrak{a}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \lesssim [\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]$

- minimax cut-off $m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}) := \arg\min_{m \geqslant 1}\{[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]\}$
- minimax rate $\mathcal{R}_\varepsilon^\star := \mathcal{R}_\varepsilon^\star(\mathfrak{a}) := [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$
- minimax estimator $\widehat{\theta}^{m_\varepsilon^\star}$, i.e., $\sup_{\theta^\circ \in \Theta_\mathfrak{a}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\star} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star$

(JJ & Schwarz, 2013)

- Lower bound: $\inf_{\widehat{\theta}} \sup_{\theta^\circ \in \Theta_\mathfrak{a}} \mathbb{E}\|\widehat{\theta} - \theta^\circ\|^2 \gtrsim \mathcal{R}_\varepsilon^\star$, if

$$\inf_\varepsilon [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \wedge \mathfrak{a}_{m_\varepsilon^\star}]/[\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}] > 0$$

Projection estimator reviewed: adaptation

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Select m by using a penalised minimum contrast criterion

Projection estimator reviewed: adaptation

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m} \sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m} (\theta_j^\circ)^2$

- Select $\widehat{m}$ by using a penalised minimum contrast criterion

$$\widehat{m} := \arg\min_{1 \leqslant m \leqslant M} \left\{ -\|\widehat{\theta}^m\| + 10 \varepsilon m \overline{\Lambda}_m \right\}$$

$$M := \max \left\{ 1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon m \overline{\Lambda}_m \leqslant \Lambda_1 \right\}$$

Projection estimator reviewed: adaptation

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

▸ Select $\widehat{m}$ by using a penalised minimum contrast criterion

$$\widehat{m} := \underset{1 \leqslant m \leqslant M}{\arg\min} \left\{ -\|\widehat{\theta}^m\| + 10\varepsilon m \overline{\Lambda}_m \right\}$$

$$M := \max\left\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon m \overline{\Lambda}_m \leqslant \Lambda_1 \right\}$$

- minimax optimal data-driven estimator, if $1 \leqslant \Lambda_{(m)}/\overline{\Lambda}_m \leqslant L_\lambda$.

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{\widehat{m}} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star(\mathfrak{a}) = [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$$

Projection estimator reviewed: adaptation

Observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon} Z_j$, $j \geqslant 1$

- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator
- $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^m \Lambda_j$ and $\mathfrak{b}_m^2 := \mathfrak{b}_m^2(\theta^\circ) := \sum_{j>m}(\theta_j^\circ)^2$

- Select $\widehat{m}$ by using a penalised minimum contrast criterion

$$\widehat{m} := \underset{1 \leqslant m \leqslant M}{\arg\min} \left\{ - \|\widehat{\theta}^m\| + 10 \varepsilon m \overline{\Lambda}_m \right\}$$

$$M := \max\left\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon m \overline{\Lambda}_m \leqslant \Lambda_1 \right\}$$

- minimax optimal data-driven estimator, if $1 \leqslant \Lambda_{(m)}/\overline{\Lambda}_m \leqslant L_\lambda$.

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{\widehat{m}} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star(\mathfrak{a}) = [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$$

Outline

- Introduction
- Frequentist perspective reviewed
- Posterior concentration
- Adaptive posterior concentration
- Adaptive Bayes estimator

Objective: optimal prior choice

- A sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ with exact posterior concentration $\mathcal{R}_\varepsilon^\circ$, i.e.

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}\,|Y}\Big(\mathcal{R}_\varepsilon^\circ \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon^\circ} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ\Big) = 1,$$

  is called oracle prior and adaptive if it does not depend on θ°.

- A sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ with exact posterior concentration $\mathcal{R}_\varepsilon^\star$, i.e.

$$\lim_{\varepsilon\to 0} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}\,|Y}\Big(\mathcal{R}_\varepsilon^\star \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star\Big) = 1.$$

  is called minimax prior and adaptive if it depends not on $\Theta_{\mathfrak{a}}$.

Sieve family of prior distributions

- Gaussian prior distribution for $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$
  - $\{\boldsymbol{\vartheta}_j\}_{j\geqslant 1}$ are independent, normally distributed
  - prior means 0 and prior variances $\varsigma = (\varsigma_j)_{j\geqslant 1}$:
    - $\boldsymbol{\vartheta}_j \sim \mathcal{N}(0, \varsigma_j)$, independent, $j \in \mathbb{N}$.

Sieve family of prior distributions

- Gaussian prior distribution for $\vartheta = (\vartheta_j)_{j \geqslant 1}$
  - $\{\vartheta_j\}_{j \geqslant 1}$ are independent, normally distributed
  - prior means 0 and prior variances $\varsigma = (\varsigma_j)_{j \geqslant 1}$:
    - $\vartheta_j \sim \mathcal{N}(0, \varsigma_j), \quad$ independent, $\quad j \in \mathbb{N}$.

- Sieve family of prior distributions $\{P_{\vartheta^m}\}_m$ depending on a hyper parameter m:
  - first m random parameters $\{\vartheta_j\}_{j=1}^m$ non-degenerate
  - independent random variables $\{\vartheta_j^m\}_{j \geqslant 1}$ with marginals:
    - $\vartheta_j^m \sim \mathcal{N}(0, \varsigma_j)$ for $1 \leqslant j \leqslant m$ and
    - $\vartheta_j^m \sim \delta_0$ for $j > m$.

Posterior distribution

- posterior distribution of $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j \geqslant 1}$ given $Y = (Y_j)_{j \geqslant 1}$
  - $\{\boldsymbol{\vartheta}_j\}_{j \geqslant 1}$ are conditionally independent, normally distributed
    - posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}[\boldsymbol{\vartheta}_j \,|\, Y] = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1}$
    - posterior mean $\theta_j^Y := \mathbb{E}[\boldsymbol{\vartheta}_j \,|\, Y] = \sigma_j \, (\lambda_j \varepsilon^{-1} Y_j)$

Posterior distribution

- posterior distribution of $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$ given $Y = (Y_j)_{j\geqslant 1}$
  - $\{\boldsymbol{\vartheta}_j\}_{j\geqslant 1}$ are conditionally independent, normally distributed
    - posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}[\boldsymbol{\vartheta}_j \,|\, Y] = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1}$
    - posterior mean $\theta_j^Y := \mathbb{E}[\boldsymbol{\vartheta}_j \,|\, Y] = \sigma_j \, (\lambda_j \varepsilon^{-1} Y_j)$

- posterior distribution of $\boldsymbol{\vartheta}^m = (\boldsymbol{\vartheta}_j^m)_{j\geqslant 1}$ given Y
  - $\{\boldsymbol{\vartheta}_j^m\}_{j\geqslant 1}$ are conditionally independent, normally distributed
    - posterior mean θ_j^Y and variance σ_j for $1 \leqslant j \leqslant m$
    - degenerate on 0 for $j > m$

Exact posterior concentration

- Let $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^{m}\Lambda_j$, $\Lambda_{(m)} := \max_{1\leqslant j\leqslant m}\Lambda_j$ and $\mathfrak{b}_m^2 := \sum_{j>m}(\theta_j^\circ)^2$.

Assumption A1. (Prior variances)

Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon\Lambda_{(m)} \leqslant \Lambda_1\}$.

There exists $d > 0$ such that $\varsigma_j \geqslant d[(\varepsilon\Lambda_j)^{1/2} \vee \varepsilon\Lambda_j]$ for all $1 \leqslant j \leqslant G_\varepsilon$.

Exact posterior concentration

- Let $\Lambda_j := \lambda_j^{-2}$, $\overline{\Lambda}_m := \frac{1}{m}\sum_{j=1}^{m}\Lambda_j$, $\Lambda_{(m)} := \max_{1\leqslant j\leqslant m}\Lambda_j$ and $\mathfrak{b}_m^2 := \sum_{j>m}(\theta_j^\circ)^2$.

Assumption A1. (Prior variances)

Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon\Lambda_{(m)} \leqslant \Lambda_1\}$.

There exists $d > 0$ such that $\varsigma_j \geqslant d[(\varepsilon\Lambda_j)^{1/2} \vee \varepsilon\Lambda_j]$ for all $1 \leqslant j \leqslant G_\varepsilon$.

Assumption A2. (Regular sub-family $\{P_{\vartheta^{m_\varepsilon}}\}_{m_\varepsilon}$)

$$\sup_\varepsilon(\varepsilon m_\varepsilon \Lambda_{(m_\varepsilon)})[\varepsilon m_\varepsilon \overline{\Lambda}_{m_\varepsilon} \vee \mathfrak{b}_{m_\varepsilon}^2]^{-1} < \infty$$

Exact posterior concentration

Proposition. Under Assumption A1. and A.2 holds

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon}|Y}\Big(\mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ) \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ)\Big) \geqslant 1 - ce^{-cm_\varepsilon}$$

Exact posterior concentration and oracle prior

Proposition. Under Assumption A1. and A.2 holds

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon}|Y}\Big(\mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ) \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ)\Big) \geqslant 1 - ce^{-cm_\varepsilon}$$

- oracle cut-off $m_\varepsilon^\circ := m_\varepsilon^\circ(\theta^\circ) := \arg\min_{m\geqslant 1}\{\mathcal{R}_\varepsilon^m(\theta^\circ)\}$
- oracle rate $\mathcal{R}_\varepsilon^\circ := \mathcal{R}_\varepsilon^\circ(\theta^\circ) := [\varepsilon m_\varepsilon^\circ \overline{\Lambda}_{m_\varepsilon^\circ} \vee \mathfrak{b}^2_{m_\varepsilon^\circ}] = \min_{m\geqslant 1}\mathcal{R}_\varepsilon^m(\theta^\circ)$

Exact posterior concentration and oracle prior

PROPOSITION. Under Assumption A1. and A.2 holds

$$\mathbb{E}_{\theta^\circ} P_{\vartheta^{m_\varepsilon}|Y}\Big(\mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ) \lesssim \|\vartheta^{m_\varepsilon} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^{m_\varepsilon}(\theta^\circ)\Big) \geqslant 1 - ce^{-cm_\varepsilon}$$

- oracle cut-off $m_\varepsilon^\circ := m_\varepsilon^\circ(\theta^\circ) := \arg\min_{m\geqslant 1}\{\mathcal{R}_\varepsilon^m(\theta^\circ)\}$
- oracle rate $\mathcal{R}_\varepsilon^\circ := \mathcal{R}_\varepsilon^\circ(\theta^\circ) := [\varepsilon m_\varepsilon^\circ \overline{\Lambda}_{m_\varepsilon^\circ} \vee \mathfrak{b}^2_{m_\varepsilon^\circ}] = \min_{m\geqslant 1} \mathcal{R}_\varepsilon^m(\theta^\circ)$

THEOREM. Under A1., A2. and in addition $m_\varepsilon^\circ \to \infty$ then

$$\lim_{\varepsilon\to\infty} \mathbb{E}_{\theta^\circ} P_{\vartheta^{m_\varepsilon^\circ}|Y}\Big(\mathcal{R}_\varepsilon^\circ \lesssim \|\vartheta^{m_\varepsilon^\circ} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ\Big) = 1$$

Exact posterior concentration and minimax prior

- minimax cut-off $m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}) := \arg\min_{m \geqslant 1}\{[\varepsilon m \overline{\Lambda}_m \vee \mathfrak{a}_m]\}$
- minimax rate $\mathcal{R}_\varepsilon^\star := \mathcal{R}_\varepsilon^\star(\mathfrak{a}) := [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$

Exact posterior concentration and minimax prior

- minimax cut-off $m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}) := \arg\min_{m\geqslant 1}\{[\varepsilon m\overline{\Lambda}_m \vee \mathfrak{a}_m]\}$
- minimax rate $\mathcal{R}_\varepsilon^\star := \mathcal{R}_\varepsilon^\star(\mathfrak{a}) := [\varepsilon m_\varepsilon^\star\overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$

Assumption A3. (Regular parameter space)

$$\inf_\varepsilon[\varepsilon m_\varepsilon^\star\overline{\Lambda}_{m_\varepsilon^\star} \wedge \mathfrak{a}_{m_\varepsilon^\star}][\varepsilon m_\varepsilon^\star\overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]^{-1} > 0$$

Exact posterior concentration and minimax prior

- minimax cut-off $m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}) := \arg\min_{m\geqslant 1}\{[\varepsilon m\overline{\Lambda}_m \vee \mathfrak{a}_m]\}$
- minimax rate $\mathcal{R}_\varepsilon^\star := \mathcal{R}_\varepsilon^\star(\mathfrak{a}) := [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]$

ASSUMPTION A3. (Regular parameter space)

$$\inf_\varepsilon [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \wedge \mathfrak{a}_{m_\varepsilon^\star}][\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]^{-1} > 0$$

THEOREM. Under A1, A2 and A3 holds

$$\lim_{\varepsilon\to\infty} \inf_{\theta^\circ\in\Theta_\mathfrak{a}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}|Y}\Big(\mathcal{R}_\varepsilon^\star \lesssim \|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star\Big) = 1$$

Remark:

- oracle prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ attains oracle rate $\mathcal{R}_\varepsilon^\circ$ as exact posterior concentration rate

Remark:

- oracle prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ attains oracle rate $\mathcal{R}_\varepsilon^\circ$ as exact posterior concentration rate
- minimax prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ attains minimax rate $\mathcal{R}_\varepsilon^\star$ as exact posterior concentration rate

Remark:

- oracle prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ attains oracle rate $\mathcal{R}_\varepsilon^\circ$ as exact posterior concentration rate
- minimax prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ attains minimax rate $\mathcal{R}_\varepsilon^\star$ as exact posterior concentration rate
- choice of m_ε° and $m_\varepsilon^\star$ depends, respectively, on θ° and $\Theta_{\mathfrak{a}}$

Remark:

- oracle prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ attains oracle rate $\mathcal{R}_\varepsilon^\circ$ as exact posterior concentration rate
- minimax prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ attains minimax rate $\mathcal{R}_\varepsilon^\star$ as exact posterior concentration rate
- choice of m_ε° and $m_\varepsilon^\star$ depends, respectively, on θ° and $\Theta_{\mathfrak{a}}$

Objective:

- put prior on hyperparameter m as outcome of a r.v. M

Outline

- Introduction
- Frequentist perspective reviewed
- Posterior concentration
- Adaptive posterior concentration
- Adaptive Bayes estimator

Hierarchical Prior

Suppose $\Lambda_{(k)} \leqslant C_\lambda \min_{j>k} \Lambda_j$ and let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant \Lambda_1\}$.

Hierarchical Prior

Suppose $\Lambda_{(k)} \leqslant C_\lambda \min_{j>k} \Lambda_j$ and let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant \Lambda_1\}$.

- random parameter M taking its values in $\{1, \ldots, G_\varepsilon\}$
  - prior distribution P_{M} given for $1 \leqslant m \leqslant G_\varepsilon$ by

$$p_{\mathsf{M}}(m) = P_{\mathsf{M}}(\mathsf{M} = m) \propto \exp(-3C_\lambda m/2) \prod_{j=1}^{m} (\varsigma_j/\sigma_j)^{1/2}$$

  - and $p_{\mathsf{M}}(m) = 0$ otherwise

Hierarchical Prior

Suppose $\Lambda_{(k)} \leqslant C_\lambda \min_{j>k} \Lambda_j$ and let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant \Lambda_1\}$.

- random parameter M taking its values in $\{1, \ldots, G_\varepsilon\}$
  - prior distribution P_{M} given for $1 \leqslant m \leqslant G_\varepsilon$ by

$$p_{\mathsf{M}}(m) = P_{\mathsf{M}}(\mathsf{M} = m) \propto \exp(-3C_\lambda m/2) \prod_{j=1}^{m} (\varsigma_j/\sigma_j)^{1/2}$$

  - and $p_{\mathsf{M}}(m) = 0$ otherwise

- random variables $Y = (Y_j)_{j \geqslant 1}$ and $\boldsymbol{\vartheta}^{\mathsf{M}} = (\boldsymbol{\vartheta}_j^{\mathsf{M}})_{j \geqslant 1}$ determined by
  - $Y_j = \boldsymbol{\vartheta}_j^{\mathsf{M}} + \sqrt{\varepsilon} Z_j$ and $\boldsymbol{\vartheta}_j^{\mathsf{M}} = \sqrt{\varsigma_j}\, V_j\, \mathbb{1}\{1 \leqslant j \leqslant \mathsf{M}\}$

  where $\{Z_j, V_j\}_{j \geqslant 1}$ are iid. standard normally distributed and independent of M.

Posterior distribution

- posterior distribution $P_{\mathsf{M}\,|\,Y}$ of M given Y
  - given for $1 \leqslant m \leqslant G_\varepsilon$ by

$$p_{\mathsf{M}\,|\,Y}(m) = P_{\mathsf{M}\,|\,Y}(\mathsf{M} = m) \propto \exp\left(\frac{1}{2}\Big\{\sum_{j=1}^{m} \sigma_j^{-1}(\theta_j^Y)^2 - 3C_\lambda m\Big\}\right)$$

  - and $p_{\mathsf{M}\,|\,Y}(m) = 0$ otherwise

Posterior distribution

- posterior distribution $P_{\mathrm{M}\,|\,Y}$ of M given Y
  - given for $1 \leqslant m \leqslant G_\varepsilon$ by

$$p_{\mathrm{M}\,|\,Y}(m) = P_{\mathrm{M}\,|\,Y}(\mathrm{M} = m) \propto \exp\left(\frac{1}{2}\Big\{\sum_{j=1}^{m} \sigma_j^{-1}(\theta_j^Y)^2 - 3C_\lambda m\Big\}\right)$$

  - and $p_{\mathrm{M}\,|\,Y}(m) = 0$ otherwise

- posterior distribution $P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,Y}$ of $\boldsymbol{\vartheta}^{\mathrm{M}} = (\boldsymbol{\vartheta}_j^{\mathrm{M}})_{j\geqslant 1}$ given Y
  - weighted mixture of posterior distributions of $\{\boldsymbol{\vartheta}^m\}_{m=1}^{G_\varepsilon}$, i.e.,

$$P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,Y} = \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M}\,|\,Y}(m)\, P_{\boldsymbol{\vartheta}^m\,|\,Y}$$

Adaptive oracle posterior concentration

Assumption A4.

There exists $L_\lambda \geqslant 1$ such that $1 \leqslant \Lambda_{(k)}/\overline{\Lambda}_k \leqslant L_\lambda$ and $\Lambda_{(kl)} \leqslant \Lambda_{(k)}\Lambda_{(l)}$.

Adaptive oracle posterior concentration

Assumption A4.

There exists $L_\lambda \geqslant 1$ such that $1 \leqslant \Lambda_{(k)}/\overline{\Lambda}_k \leqslant L_\lambda$ and $\Lambda_{(kl)} \leqslant \Lambda_{(k)}\Lambda_{(l)}$.

Lemma. Under A1 and A4 there exist $1 \leqslant G_\varepsilon^- \leqslant m_\varepsilon^\circ \leqslant G_\varepsilon^+ \leqslant G_\varepsilon$ s.t.

$$\mathbb{E}_{\theta^\circ} P_{\mathsf{M}\,|Y}(G_\varepsilon^- \leqslant \mathsf{M} \leqslant G_\varepsilon^+) \geqslant 1 - 4\exp(-C_\lambda m_\varepsilon^\circ/5 + \log G_\varepsilon).$$

Adaptive oracle posterior concentration

Assumption A4.

There exists $L_\lambda \geqslant 1$ such that $1 \leqslant \Lambda_{(k)}/\overline{\Lambda}_k \leqslant L_\lambda$ and $\Lambda_{(kl)} \leqslant \Lambda_{(k)}\Lambda_{(l)}$.

Lemma. Under A1 and A4 there exist $1 \leqslant G_\varepsilon^- \leqslant m_\varepsilon^\circ \leqslant G_\varepsilon^+ \leqslant G_\varepsilon$ s.t.

$$\mathbb{E}_{\theta^\circ} P_{\mathsf{M}\,|Y}(G_\varepsilon^- \leqslant \mathsf{M} \leqslant G_\varepsilon^+) \geqslant 1 - 4\exp(-C_\lambda m_\varepsilon^\circ/5 + \log G_\varepsilon).$$

Lemma. Under A1 and A4 there exist $1 \leqslant G_\varepsilon^{\star-} \leqslant m_\varepsilon^\star \leqslant G_\varepsilon^{\star+} \leqslant G_\varepsilon$ s.t.

$$\inf_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\mathsf{M}\,|Y}(G_\varepsilon^{\star-} \leqslant \mathsf{M} \leqslant G_\varepsilon^{\star+}) \geqslant 1 - 4\exp(-C_\lambda m_\varepsilon^\star/5 + \log G_\varepsilon).$$

Adaptive oracle posterior concentration

Assumption A5. (Regular parameter)

$$\inf_{\varepsilon}[\varepsilon m_{\varepsilon}^{\circ}\overline{\Lambda}_{m_{\varepsilon}^{\circ}} \wedge \mathfrak{b}_{m_{\varepsilon}^{\circ}}]/\mathcal{R}_{\varepsilon}^{\circ} > 0$$

Adaptive oracle posterior concentration

ASSUMPTION A5. (Regular parameter)

$$\inf_\varepsilon [\varepsilon m_\varepsilon^\circ \overline{\Lambda}_{m_\varepsilon^\circ} \wedge \mathfrak{b}_{m_\varepsilon^\circ}]/\mathcal{R}_\varepsilon^\circ > 0$$

LEMMA. Under A1, A4 and A5 holds

$$\sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m|Y}\left(\mathcal{R}_\varepsilon^\circ \lesssim \|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ\right) \leqslant 1 - ce^{-cG_\varepsilon^-}.$$

Adaptive oracle posterior concentration

Assumption A5. (Regular parameter)

$$\inf_{\varepsilon}[\varepsilon m^{\circ}_{\varepsilon}\overline{\Lambda}_{m^{\circ}_{\varepsilon}} \wedge \mathfrak{b}_{m^{\circ}_{\varepsilon}}]/\mathcal{R}^{\circ}_{\varepsilon} > 0$$

Lemma. Under A1, A4 and A5 holds

$$\sum_{m=G^{-}_{\varepsilon}}^{G^{+}_{\varepsilon}} \mathbb{E}_{\theta^{\circ}} P_{\boldsymbol{\vartheta}^{m}|Y}\Big(\mathcal{R}^{\circ}_{\varepsilon} \lesssim \|\boldsymbol{\vartheta}^{m} - \theta^{\circ}\|^{2} \lesssim \mathcal{R}^{\circ}_{\varepsilon}\Big) \leqslant 1 - ce^{-cG^{-}_{\varepsilon}}.$$

Theorem. Under A1, A4, A5 and $(\log G_{\varepsilon})/m^{\circ}_{\varepsilon} = o(1)$ holds

$$\lim_{\varepsilon\to\infty} \mathbb{E}_{\theta^{\circ}} P_{\boldsymbol{\vartheta}^{\mathsf{M}}|Y}\Big(\mathcal{R}^{\circ}_{\varepsilon} \lesssim \|\boldsymbol{\vartheta}^{\mathsf{M}} - \theta^{\circ}\|^{2} \lesssim \mathcal{R}^{\circ}_{\varepsilon}\Big) = 1$$

Adaptive minimax posterior concentration

Assumption A3. (Regular parameter space)

$$\inf_{\varepsilon}[\varepsilon m_{\varepsilon}^{\star}\overline{\Lambda}_{m_{\varepsilon}^{\star}} \wedge \mathfrak{a}_{m_{\varepsilon}^{\star}}][\varepsilon m_{\varepsilon}^{\star}\overline{\Lambda}_{m_{\varepsilon}^{\star}} \vee \mathfrak{a}_{m_{\varepsilon}^{\star}}]^{-1} > 0$$

Adaptive minimax posterior concentration

ASSUMPTION A3. (Regular parameter space)

$$\inf_{\varepsilon}[\varepsilon m_{\varepsilon}^{\star}\overline{\Lambda}_{m_{\varepsilon}^{\star}} \wedge \mathfrak{a}_{m_{\varepsilon}^{\star}}][\varepsilon m_{\varepsilon}^{\star}\overline{\Lambda}_{m_{\varepsilon}^{\star}} \vee \mathfrak{a}_{m_{\varepsilon}^{\star}}]^{-1} > 0$$

LEMMA. Under A1, A3 and A4 holds for all $\theta^{\circ} \in \Theta_{\mathfrak{a}}$

$$\sum_{m=G_{\varepsilon}^{\star-}}^{G_{\varepsilon}^{\star+}} \mathbb{E}_{\theta^{\circ}} P_{\boldsymbol{\vartheta}^m|Y}\Big(\|\boldsymbol{\vartheta}^m - \theta^{\circ}\|^2 \lesssim \mathcal{R}_{\varepsilon}^{\star}\Big) \geqslant 1 - ce^{-cG_{\varepsilon}^{\star-}}.$$

Adaptive minimax posterior concentration

ASSUMPTION A3. (Regular parameter space)

$$\inf_{\varepsilon}[\varepsilon m_{\varepsilon}^{\star}\overline{\Lambda}_{m_{\varepsilon}^{\star}} \wedge \mathfrak{a}_{m_{\varepsilon}^{\star}}][\varepsilon m_{\varepsilon}^{\star}\overline{\Lambda}_{m_{\varepsilon}^{\star}} \vee \mathfrak{a}_{m_{\varepsilon}^{\star}}]^{-1} > 0$$

LEMMA. Under A1, A3 and A4 holds for all $\theta^{\circ} \in \Theta_{\mathfrak{a}}$

$$\sum_{m=G_{\varepsilon}^{\star-}}^{G_{\varepsilon}^{\star+}} \mathbb{E}_{\theta^{\circ}} P_{\boldsymbol{\vartheta}^{m}|Y}\Big(\|\boldsymbol{\vartheta}^{m} - \theta^{\circ}\|^{2} \lesssim \mathcal{R}_{\varepsilon}^{\star}\Big) \geqslant 1 - ce^{-cG_{\varepsilon}^{\star-}}.$$

THEOREM. Under A1, A3, A4 and $(\log G_{\varepsilon})/m_{\varepsilon}^{\star} = o(1)$ holds

- $\lim_{\varepsilon\to\infty} \mathbb{E}_{\theta^{\circ}} P_{\boldsymbol{\vartheta}^{\mathsf{M}}|Y}\Big(\|\boldsymbol{\vartheta}^{\mathsf{M}} - \theta^{\circ}\|^{2} \lesssim \mathcal{R}_{\varepsilon}^{\star}\Big) = 1$ for all $\theta^{\circ} \in \Theta_{\mathfrak{a}}$;

Adaptive minimax posterior concentration

Assumption A3. (Regular parameter space)

$$\inf_\varepsilon [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \wedge \mathfrak{a}_{m_\varepsilon^\star}][\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star} \vee \mathfrak{a}_{m_\varepsilon^\star}]^{-1} > 0$$

Lemma. Under A1, A3 and A4 holds for all $\theta^\circ \in \Theta_{\mathfrak{a}}$

$$\sum_{m=G_\varepsilon^{\star-}}^{G_\varepsilon^{\star+}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m|Y}\left(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star\right) \geqslant 1 - c e^{-c G_\varepsilon^{\star-}}.$$

Theorem. Under A1, A3, A4 and $(\log G_\varepsilon)/m_\varepsilon^\star = o(1)$ holds

- $\lim_{\varepsilon\to\infty} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathsf{M}}|Y}\left(\|\boldsymbol{\vartheta}^{\mathsf{M}} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star\right) = 1$ for all $\theta^\circ \in \Theta_{\mathfrak{a}}$;
- $\lim_{\varepsilon\to\infty} \inf_{\theta^\circ\in\Theta_{\mathfrak{a}}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathsf{M}}|Y}\left(\|\boldsymbol{\vartheta}^{\mathsf{M}} - \theta^\circ\|^2 \lesssim K_\varepsilon \mathcal{R}_\varepsilon^\star\right) = 1$ for any $K_\varepsilon \to \infty$;

Outline

- Introduction
- Frequentist perspective reviewed
- Posterior concentration
- Adaptive posterior concentration
- Adaptive Bayes estimator

Adaptive data-driven Bayes estimator

Common Bayes estimator $\widehat{\theta} = (\widehat{\theta}_j)_{j\geqslant 1} = \mathbb{E}[\boldsymbol{\vartheta}^{\mathsf{M}} \,|\, Y]$ given by

$$\widehat{\theta}_j = \theta_j^Y P_{\mathsf{M}\,|\,Y}(j \leqslant \mathsf{M} \leqslant G_\varepsilon)\mathbb{1}_{\{1\leqslant j\leqslant G_\varepsilon\}}$$

Theorem. Under A1, A4, A5 and $\log(G_\varepsilon/\mathcal{R}_\varepsilon^\circ)/m_\varepsilon^\circ = o(1)$ holds

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ(\theta^\circ) = \min_{m\geqslant 1}\mathcal{R}_\varepsilon^m(\theta^\circ).$$

Adaptive data-driven Bayes estimator

Common Bayes estimator $\widehat{\theta} = (\widehat{\theta}_j)_{j\geqslant 1} = \mathbb{E}[\boldsymbol{\vartheta}^{\mathsf{M}} \,|\, Y]$ given by

$$\widehat{\theta}_j = \theta_j^Y P_{\mathsf{M}\,|\,Y}(j \leqslant \mathsf{M} \leqslant G_\varepsilon)\mathbb{1}_{\{1\leqslant j\leqslant G_\varepsilon\}}$$

THEOREM. Under A1, A4, A5 and $\log(G_\varepsilon/\mathcal{R}_\varepsilon^\circ)/m_\varepsilon^\circ = o(1)$ holds

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\circ(\theta^\circ) = \min_{m\geqslant 1}\mathcal{R}_\varepsilon^m(\theta^\circ).$$

THEOREM. Under A1, A3, A4 and $\log(G_\varepsilon/\mathcal{R}_\varepsilon^\star)/m_\varepsilon^\star = o(1)$ holds

$$\sup_{\theta^\circ\in\Theta_{\mathfrak{a}}}\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|^2 \lesssim \mathcal{R}_\varepsilon^\star = \min_{m\geqslant 1}\sup_{\theta^\circ\in\Theta_{\mathfrak{a}}}\mathcal{R}_\varepsilon^m(\theta^\circ)$$

Diffuse hierarchical prior

The posterior weights under a diffuse prior for $\boldsymbol{\vartheta}$ satisfy

$$p_{\mathrm{M}\,|\,Y}(m) = \frac{\exp\left(\frac{1}{2}\left\{\sum_{j=1}^{m}(\varepsilon\Lambda_j)^{-1}(Y_j/\lambda_j)^2 - 3C_\lambda m\right\}\right)}{\sum_{k=1}^{G_\varepsilon}\exp\left(\frac{1}{2}\left\{\sum_{j=1}^{k}(\varepsilon\Lambda_j)^{-1}(Y_j/\lambda_j)^2 - 3C_\lambda k\right\}\right)}$$

Diffuse hierarchical prior

The posterior weights under a diffuse prior for $\boldsymbol{\vartheta}$ satisfy

$$p_{\mathrm{M}\,|\,Y}(m) = \frac{\exp\left(-\frac{1}{2}\left\{-\|(Y/\lambda)^m\|_{\varepsilon\Lambda}^2 + 3C_\lambda m\right\}\right)}{\sum_{k=1}^{G_\varepsilon} \exp\left(-\frac{1}{2}\left\{-\|(Y/\lambda)^k\|_{\varepsilon\Lambda}^2 + 3C_\lambda k\right\}\right)}$$

where $\|\theta\|_{\varepsilon\Lambda}^2 := \sum_{j\geqslant 1}(\varepsilon\Lambda_j)^{-1}(\theta_j)^2$ and $(Y/\lambda)^k := \left(Y_j/\lambda_j \mathbb{1}\{1 \leqslant j \leqslant k\}\right)_{j\in\mathbb{N}}$.

Diffuse hierarchical prior

The posterior weights under a diffuse prior for $\boldsymbol{\vartheta}$ satisfy

$$p_{\mathsf{M}\,|\,Y}(m) = \frac{\exp\Big(-\frac{1}{2}\big\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\big\}\Big)}{\sum_{k=1}^{G_\varepsilon} \exp\Big(-\frac{1}{2}\big\{-\|(Y/\lambda)^k\|^2_{\varepsilon\Lambda} + 3C_\lambda k\big\}\Big)}$$

where $\|\theta\|^2_{\varepsilon\Lambda} := \sum_{j\geqslant 1}(\varepsilon\Lambda_j)^{-1}(\theta_j)^2$ and $(Y/\lambda)^k := \big(Y_j/\lambda_j \mathbb{1}_{\{1\leqslant j\leqslant k\}}\big)_{j\in\mathbb{N}}$.

The adaptive Bayes estimator $\widehat{\theta} = (\widehat{\theta}_j)_{j\geqslant 1} = \mathbb{E}[\boldsymbol{\vartheta}^{\mathsf{M}}\,|\,Y]$ is given by

$$\widehat{\theta}_j = \big(1 - \sum_{k=1}^{j-1} p_{\mathsf{M}\,|\,Y}(m)\big) \times \frac{Y_j}{\lambda_j} \times \mathbb{1}_{\{1\leqslant j\leqslant G_\varepsilon\}}.$$

Data-driven shrinked projection estimator

Consider weights

$$w_m^Y := \frac{\exp\Big(-\frac{1}{2\varepsilon}\big\{ -\|(Y/\lambda)^m\|^2 + 3C_\lambda \varepsilon m \overline{\Lambda}_m \big\}\Big)}{\sum_{k=1}^{G_\varepsilon} \exp\Big(-\frac{1}{2\varepsilon}\big\{ -\|(Y/\lambda)^k\|^2 + 3C_\lambda \varepsilon k \overline{\Lambda}_k \big\}\Big)}$$

Data-driven shrinked projection estimator

Consider weights

$$w_m^Y := \frac{\exp\Big(-\frac{1}{2\varepsilon}\big\{ -\|(Y/\lambda)^m\|^2 + 3C_\lambda \varepsilon m \overline{\Lambda}_m \big\}\Big)}{\sum_{k=1}^{G_\varepsilon} \exp\Big(-\frac{1}{2\varepsilon}\big\{ -\|(Y/\lambda)^k\|^2 + 3C_\lambda \varepsilon k \overline{\Lambda}_k \big\}\Big)}$$

and the data-driven shrinked projection estimator $\widehat{\theta} = (\widehat{\theta}_j)_{j \geqslant 1}$ given by

$$\widehat{\theta}_j = \big(1 - \sum_{k=1}^{j-1} w_m^Y\big) \times \frac{Y_j}{\lambda_j} \times \mathbb{1}\{1 \leqslant j \leqslant G_\varepsilon\}.$$

Data-driven shrinked projection estimator

Consider weights

$$w_m^Y := \frac{\exp\Big(-\frac{1}{2\varepsilon}\big\{ -\|(Y/\lambda)^m\|^2 + 3C_\lambda \varepsilon m \overline{\Lambda}_m \big\}\Big)}{\sum_{k=1}^{G_\varepsilon} \exp\Big(-\frac{1}{2\varepsilon}\big\{ -\|(Y/\lambda)^k\|^2 + 3C_\lambda \varepsilon k \overline{\Lambda}_k \big\}\Big)}$$

and the data-driven shrinked projection estimator $\widehat{\theta} = (\widehat{\theta}_j)_{j\geqslant 1}$ given by

$$\widehat{\theta}_j = \big(1 - \sum_{k=1}^{j-1} w_m^Y\big) \times \frac{Y_j}{\lambda_j} \times \mathbb{1}\{1 \leqslant j \leqslant G_\varepsilon\}.$$

Open question: oracle/minimax optimality up to a constant?

Conclusion

- Summary
  - oracle and minimax optimal concentration rates in an indirect sequence space model

Conclusion

- Summary
  - oracle and minimax optimal concentration rates in an indirect sequence space model
  - hierarchical prior leading to oracle and minimax optimal concentration rates and a fully-data driven Bayes estimate that is oracle and minimax optimal in an indirect sequence space model

Conclusion

- Summary
  - oracle and minimax optimal concentration rates in an indirect sequence space model
  - hierarchical prior leading to oracle and minimax optimal concentration rates and a fully-data driven Bayes estimate that is oracle and minimax optimal in an indirect sequence space model

- Perspectives
  - additional noise: $X_j = \lambda_j + \text{noise}$

Conclusion

- Summary
  - oracle and minimax optimal concentration rates in an indirect sequence space model
  - hierarchical prior leading to oracle and minimax optimal concentration rates and a fully-data driven Bayes estimate that is oracle and minimax optimal in an indirect sequence space model

- Perspectives
  - additional noise: $X_j = \lambda_j + \text{noise}$
  - non-diagonal inverse problems with noise in the operator

References

J.J., A. Simoni and R. Schenk (2015) *Adaptive Bayesian estimation in indirect Gaussian sequence space models.* Discussion paper (arXiv:1502.00184)

Thank you for your attention.

Adaptive Bayesian estimation in indirect Gaussian sequence space models

Jan Johannes

Ruprecht-Karls-Universität Heidelberg

"Mathematical statistics and inverse problems"

February 2016, CIRM, Marseille

joint work with Anna Simoni and Rudolf Schenk