Adaptive Bayesian estimation in indirect Gaussian sequence space models

Jan Johannes   Anna Simoni   Rudolf Schenk

Workshop on Inverse Problems, October 28, 2016

Outline

Motivation (I)

- Growing interest in oracle or minimax optimal nonparametric estimation and adaptation in the framework of statistical inverse problems.
- Choice of a tuning parameter. Oracle and minimax estimation is achieved, respectively, if the tuning parameter is set to an optimal value which relies
  - either on a knowledge of the unknown parameter of interest
  - or on certain characteristics of the unknown parameter of interest (such as smoothness).

Motivation (II)

- Both the parameter and its smoothness are unknown: then one wants to design a feasible and adaptive procedure to select the tuning parameter that achieves the oracle or minimax rate.

- We investigate a Bayesian procedure where the tuning parameter is endowed with a prior.

- Previous literature on Bayesian Statistical Inverse Problems: Knapik, Van der Vaart & Van Zanten (2011), Knapik, Szabo, Van der Vaart & Van Zanten (2014), Agapiou, Larsson & Stuart (2013), Ray (2013), Florens & Simoni (2012, 2016), ...

- We consider an indirect Gaussian sequence space model (iGSSM) (which is equivalent to an indirect Gaussian regression, *e.g.* Brown & Low (1996) and Meister (2011)).

Motivation (II)

- Both the parameter and its smoothness are unknown: then one wants to design a feasible and adaptive procedure to select the tuning parameter that achieves the oracle or minimax rate.
- We investigate a Bayesian procedure where the tuning parameter is endowed with a prior.
- Previous literature on Bayesian Statistical Inverse Problems: Knapik, Van der Vaart & Van Zanten (2011), Knapik, Szabo, Van der Vaart & Van Zanten (2014), Agapiou, Larsson & Stuart (2013), Ray (2013), Florens & Simoni (2012, 2016), ...
- We consider an indirect Gaussian sequence space model (iGSSM) (which is equivalent to an indirect Gaussian regression, *e.g.* Brown & Low (1996) and Meister (2011)).

Motivation (II)

- Both the parameter and its smoothness are unknown: then one wants to design a feasible and adaptive procedure to select the tuning parameter that achieves the oracle or minimax rate.
- We investigate a Bayesian procedure where the tuning parameter is endowed with a prior.
- Previous literature on Bayesian Statistical Inverse Problems: Knapik, Van der Vaart & Van Zanten (2011), Knapik, Szabo, Van der Vaart & Van Zanten (2014), Agapiou, Larsson & Stuart (2013), Ray (2013), Florens & Simoni (2012, 2016), ...
- We consider an indirect Gaussian sequence space model (iGSSM) (which is equivalent to an indirect Gaussian regression, *e.g.* Brown & Low (1996) and Meister (2011)).

Motivation (II)

- Both the parameter and its smoothness are unknown: then one wants to design a feasible and adaptive procedure to select the tuning parameter that achieves the oracle or minimax rate.
- We investigate a Bayesian procedure where the tuning parameter is endowed with a prior.
- Previous literature on Bayesian Statistical Inverse Problems: Knapik, Van der Vaart & Van Zanten (2011), Knapik, Szabo, Van der Vaart & Van Zanten (2014), Agapiou, Larsson & Stuart (2013), Ray (2013), Florens & Simoni (2012, 2016), . . .
- We consider an indirect Gaussian sequence space model (iGSSM) (which is equivalent to an indirect Gaussian regression, *e.g.* Brown & Low (1996) and Meister (2011)).

Indirect Gaussian sequence space model (iGSSM)

An observable sequence of random variables $Y = (Y)_{j \geqslant 1}$ obeys an iGSSM, if

$$Y_j = \lambda_j \theta_j + \sqrt{\varepsilon}\xi_j, \qquad j \in \mathbb{N}, \tag{1}$$

where:

- $\{\xi_j\}_{j \geqslant 1}$ *i.i.d.* $\mathcal{N}(0,1)$ are unobservable error terms,
- $0 < \varepsilon < 1$ is a known noise level (e.g. $\varepsilon = \frac{1}{\sqrt{n}}$)
- $\theta = (\theta_j)_{j \geqslant 1} \in \ell_2$ parameter sequence of interest.

Inverse Problem: Consider $\mathcal{F} = L^2[0,1]$ and transformation $T : \mathcal{F} \to \mathcal{F}$. So, g= Tf.

Representation:

- $f \in \mathcal{F} \leftrightarrow \theta \in \Theta := \ell^2$ via $\theta_j = \int_0^1 f(t)\psi_j(t)dt$
- Operator $T \leftrightarrow$ Eigenvalues λ

Indirect Gaussian sequence space model (iGSSM)

An observable sequence of random variables $Y = (Y)_{j \geqslant 1}$ obeys an iGSSM, if

$$Y_j = \lambda_j \theta_j + \sqrt{\varepsilon}\xi_j, \qquad j \in \mathbb{N}, \tag{1}$$

where:

- $\{\xi_j\}_{j \geqslant 1}$ *i.i.d.* $\mathcal{N}(0, 1)$ are unobservable error terms,
- $0 < \varepsilon < 1$ is a known noise level (e.g. $\varepsilon = \frac{1}{\sqrt{n}}$)
- $\theta = (\theta_j)_{j \geqslant 1} \in \ell_2$ parameter sequence of interest.

Inverse Problem: Consider $\mathcal{F} = L^2[0, 1]$ and transformation $T : \mathcal{F} \to \mathcal{F}$. So, g= Tf.
Representation:

- $f \in \mathcal{F} \leftrightarrow \theta \in \Theta := \ell^2$ via $\theta_j = \int_0^1 f(t)\psi_j(t)dt$
- Operator $T \leftrightarrow$ Eigenvalues λ

Indirect Gaussian sequence space model (iGSSM)

An observable sequence of random variables $Y = (Y)_{j\geqslant 1}$ obeys an iGSSM, if

$$Y_j = \lambda_j \theta_j + \sqrt{\varepsilon}\xi_j, \qquad j \in \mathbb{N}, \tag{1}$$

where:

- $\{\xi_j\}_{j\geqslant 1}$ *i.i.d.* $\mathcal{N}(0, 1)$ are unobservable error terms,
- $0 < \varepsilon < 1$ is a known noise level (e.g. $\varepsilon = \frac{1}{\sqrt{n}}$)
- $\theta = \left(\theta_j\right)_{j\geqslant 1} \in \ell_2$ parameter sequence of interest.

Inverse Problem: Consider $\mathcal{F} = L^2[0, 1]$ and transformation $T : \mathcal{F} \to \mathcal{F}$. So, g= Tf.

Representation:

- $f \in \mathcal{F} \leftrightarrow \theta \in \Theta := \ell^2$ via $\theta_j = \int_0^1 f(t)\psi_j(t)dt$
- Operator $T \leftrightarrow$ Eigenvalues λ

Indirect Gaussian sequence space model (iGSSM)

An observable sequence of random variables $Y = (Y)_{j \geqslant 1}$ obeys an iGSSM, if

$$Y_j = \lambda_j \theta_j + \sqrt{\varepsilon} \xi_j, \qquad j \in \mathbb{N}, \tag{1}$$

where:

- $\{\xi_j\}_{j \geqslant 1}$ *i.i.d.* $\mathcal{N}(0,1)$ are unobservable error terms,
- $0 < \varepsilon < 1$ is a known noise level (e.g. $\varepsilon = \frac{1}{\sqrt{n}}$)
- $\theta = \left(\theta_j\right)_{j \geqslant 1} \in \ell_2$ parameter sequence of interest.

Inverse Problem: Consider $\mathcal{F} = L^2[0,1]$ and transformation $T : \mathcal{F} \to \mathcal{F}$. So, g= Tf.

Representation:

- $f \in \mathcal{F} \leftrightarrow \theta \in \Theta := \ell^2$ via $\theta_j = \int_0^1 f(t)\psi_j(t)dt$
- Operator $T \leftrightarrow$ Eigenvalues λ

Bayesian perspective

- $\theta = (\theta_j)_{j\geqslant 1}$ outcome of ℓ_2-valued r.v. $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$
- $Y_j | \boldsymbol{\vartheta}_j = \theta_j \sim \mathcal{N}(\lambda_j \theta_j, \varepsilon)$, independent, $j \in \mathbb{N}$,
- 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:

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

- Typical priors in the GSSM literature: sieve, Gaussian series, mixture of normal, ...
- Typical priors in the iGSSM literature: Gaussian series, mixture of Gaussian series, sieve, Gaussian process, ...
- We: sieve priors.

- $\theta = (\theta_j)_{j \geqslant 1}$ outcome of ℓ_2-valued r.v. $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j \geqslant 1}$
- $Y_j | \boldsymbol{\vartheta}_j = \theta_j \sim \mathcal{N}(\lambda_j \theta_j, \varepsilon)$, independent, $j \in \mathbb{N}$,
- 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:

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

- Typical priors in the GSSM literature: sieve, Gaussian series, mixture of normal, ...
- Typical priors in the iGSSM literature: Gaussian series, mixture of Gaussian series, sieve, Gaussian process, ...
- We: sieve priors.

Bayesian perspective

- $\theta = (\theta_j)_{j \geqslant 1}$ outcome of ℓ_2-valued r.v. $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j \geqslant 1}$
- $Y_j | \boldsymbol{\vartheta}_j = \theta_j \sim \mathcal{N}(\lambda_j \theta_j, \varepsilon)$, independent, $j \in \mathbb{N}$,
- 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:

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

- Typical priors in the GSSM literature: sieve, Gaussian series, mixture of normal, ...
- Typical priors in the iGSSM literature: Gaussian series, mixture of Gaussian series, sieve, Gaussian process, ...
- We: sieve priors.

Bayesian perspective

- $\theta = (\theta_j)_{j\geqslant 1}$ outcome of ℓ_2-valued r.v. $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$
- $Y_j|\,\boldsymbol{\vartheta}_j = \theta_j \sim \mathcal{N}\big(\lambda_j\theta_j, \varepsilon\big)$, independent, $j \in \mathbb{N}$,
- 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:

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

- Typical priors in the GSSM literature: sieve, Gaussian series, mixture of normal, …
- Typical priors in the iGSSM literature: Gaussian series, mixture of Gaussian series, sieve, Gaussian process, …
- We: sieve priors.

Bayesian perspective

- $\theta = (\theta_j)_{j\geqslant 1}$ outcome of ℓ_2-valued r.v. $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$
- $Y_j | \boldsymbol{\vartheta}_j = \theta_j \sim \mathcal{N}(\lambda_j\theta_j, \varepsilon)$, independent, $j \in \mathbb{N}$,
- 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:

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

- Typical priors in the GSSM literature: sieve, Gaussian series, mixture of normal, ...
- Typical priors in the iGSSM literature: Gaussian series, mixture of Gaussian series, sieve, Gaussian process, ...
- We: sieve priors.

Bayesian perspective

- $\theta = (\theta_j)_{j\geqslant 1}$ outcome of ℓ_2-valued r.v. $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$
- $Y_j | \boldsymbol{\vartheta}_j = \theta_j \sim \mathcal{N}(\lambda_j \theta_j, \varepsilon)$, independent, $j \in \mathbb{N}$,
- 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:

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

- Typical priors in the GSSM literature: sieve, Gaussian series, mixture of normal, ...
- Typical priors in the iGSSM literature: Gaussian series, mixture of Gaussian series, sieve, Gaussian process, ...
- We: sieve priors.

Posterior concentration rate (I)

θ°= realization of the r.v. ϑ associated with the data-generating distribution.

Objective:

- For observations $Y_j | \vartheta_j = \theta_j^\circ \sim \mathcal{N}(\lambda_j {\theta^\circ}_j, \varepsilon)$
- Construct a prior P_ϑ (that depends on ε) and study frequentist properties of the associated posterior, *i.e.*

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\vartheta | Y}((K)^{-1} \Phi_\varepsilon \leqslant \|\vartheta - \theta^\circ\|^2 \leqslant K \Phi_\varepsilon) = 1$$

  with $1 \leqslant K < \infty$.

- Φ_ε is called exact posterior concentration rate.
- The rate Φ_ε depends on the prior P_ϑ, on θ° and $\lambda = (\lambda_j)_{j \geqslant 1}$.

θ°= realization of the r.v. ϑ associated with the data-generating distribution.

Objective:

- For observations $Y_j | \vartheta_j = \theta_j^\circ \sim \mathcal{N}(\lambda_j {\theta^\circ}_j, \varepsilon)$
- Construct a prior P_ϑ (that depends on ε) and study frequentist properties of the associated posterior, *i.e.*

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\vartheta \,|\, Y}((K)^{-1}\, \Phi_\varepsilon \leqslant \|\vartheta - \theta^\circ\|^2 \leqslant K\, \Phi_\varepsilon) = 1$$

  with $1 \leqslant K < \infty$.

- Φ_ε is called exact posterior concentration rate.
- The rate Φ_ε depends on the prior P_ϑ, on θ° and $\lambda = (\lambda_j)_{j \geqslant 1}$.

θ°= realization of the r.v. ϑ associated with the data-generating distribution.

Objective:

- For observations $Y_j | \vartheta_j = \theta_j^\circ \sim \mathcal{N}(\lambda_j {\theta^\circ}_j, \varepsilon)$
- Construct a prior P_{ϑ} (that depends on ε) and study frequentist properties of the associated posterior, *i.e.*

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\vartheta \,|\, Y}((K)^{-1} \Phi_\varepsilon \leqslant \|\vartheta - \theta^\circ\|^2 \leqslant K \, \Phi_\varepsilon) = 1$$

  with $1 \leqslant K < \infty$.

- Φ_ε is called exact posterior concentration rate.
- The rate Φ_ε depends on the prior P_{ϑ}, on θ° and $\lambda = (\lambda_j)_{j \geqslant 1}$.

Posterior concentration rate (II)

- **Oracle approach:** given a θ°, we derive a prior with smallest possible exact posterior concentration rate Φ°_ε (oracle prior and oracle posterior concentration rate).

- **Minimax approach:** given a class $\Theta_{\mathfrak{a}}$ of parameters, we construct a prior with exact posterior concentration rate $\Phi^\star_\varepsilon$ uniformly over $\Theta_{\mathfrak{a}}$, where $\Phi^\star_\varepsilon$ is the minimax rate.

- The oracle and minimax posterior concentration rates that we obtain do not involve a logarithmic term (which is usual in most of the nonparametric Bayesian literature).

- **Adaptation:** construction of a hierarchical prior $P_{\vartheta^{\mathrm{M}}}$ that is adaptive, *i.e.* given $\theta^\circ \in \ell_2$ or $\Theta_{\mathfrak{a}} \subset \ell_2$, the posterior distribution contracts, respectively, at the Φ°_ε rate or the $\Theta_{\mathfrak{a}}$ rate over $\Theta_{\mathfrak{a}}$ while $P_{\vartheta^{\mathrm{M}}}$ does not rely neither on the knowledge of θ° nor the class $\Theta_{\mathfrak{a}}$.

Posterior concentration rate (II)

- **Oracle approach:** given a θ°, we derive a prior with smallest possible exact posterior concentration rate Φ_ε° (oracle prior and oracle posterior concentration rate).
- **Minimax approach:** given a class $\Theta_{\mathfrak{a}}$ of parameters, we construct a prior with exact posterior concentration rate $\Phi_\varepsilon^\star$ uniformly over $\Theta_{\mathfrak{a}}$, where $\Phi_\varepsilon^\star$ is the minimax rate.
- The oracle and minimax posterior concentration rates that we obtain do not involve a logarithmic term (which is usual in most of the nonparametric Bayesian literature).
- **Adaptation:** construction of a hierarchical prior $P_{\vartheta^{\mathrm{M}}}$ that is adaptive, *i.e.* given $\theta^\circ \in \ell_2$ or $\Theta_{\mathfrak{a}} \subset \ell_2$, the posterior distribution contracts, respectively, at the Φ_ε° rate or the $\Theta_{\mathfrak{a}}$ rate over $\Theta_{\mathfrak{a}}$ while $P_{\vartheta^{\mathrm{M}}}$ does not rely neither on the knowledge of θ° nor the class $\Theta_{\mathfrak{a}}$.

Posterior concentration rate (II)

- **Oracle approach:** given a θ°, we derive a prior with smallest possible exact posterior concentration rate Φ°_ε (oracle prior and oracle posterior concentration rate).
- **Minimax approach:** given a class $\Theta_{\mathfrak{a}}$ of parameters, we construct a prior with exact posterior concentration rate $\Phi^\star_\varepsilon$ uniformly over $\Theta_{\mathfrak{a}}$, where $\Phi^\star_\varepsilon$ is the minimax rate.
- The oracle and minimax posterior concentration rates that we obtain do not involve a logarithmic term (which is usual in most of the nonparametric Bayesian literature).
- **Adaptation:** construction of a hierarchical prior $P_{\vartheta^{\mathrm{M}}}$ that is adaptive, *i.e.* given $\theta^\circ \in \ell_2$ or $\Theta_{\mathfrak{a}} \subset \ell_2$, the posterior distribution contracts, respectively, at the Φ°_ε rate or the $\Theta_{\mathfrak{a}}$ rate over $\Theta_{\mathfrak{a}}$ while $P_{\vartheta^{\mathrm{M}}}$ does not rely neither on the knowledge of θ° nor the class $\Theta_{\mathfrak{a}}$.

Posterior concentration rate (II)

- **Oracle approach:** given a θ°, we derive a prior with smallest possible exact posterior concentration rate Φ_ε° (oracle prior and oracle posterior concentration rate).
- **Minimax approach:** given a class $\Theta_{\mathfrak{a}}$ of parameters, we construct a prior with exact posterior concentration rate $\Phi_\varepsilon^\star$ uniformly over $\Theta_{\mathfrak{a}}$, where $\Phi_\varepsilon^\star$ is the minimax rate.
- The oracle and minimax posterior concentration rates that we obtain do not involve a logarithmic term (which is usual in most of the nonparametric Bayesian literature).
- **Adaptation:** construction of a hierarchical prior $P_{\boldsymbol{\vartheta}^{\mathrm{M}}}$ that is adaptive, *i.e.* given $\theta^\circ \in \ell_2$ or $\Theta_{\mathfrak{a}} \subset \ell_2$, the posterior distribution contracts, respectively, at the Φ_ε° rate or the $\Theta_{\mathfrak{a}}$ rate over $\Theta_{\mathfrak{a}}$ while $P_{\boldsymbol{\vartheta}^{\mathrm{M}}}$ does not rely neither on the knowledge of θ° nor the class $\Theta_{\mathfrak{a}}$.

Outline:

Projection estimator motivates prior:
observations $Y_j = \lambda_j \theta_{\circ j} + \sqrt{\varepsilon}\xi_j$, $j \geqslant 1$

- Sieve $\Theta_1 \subset \Theta_2 \subset \Theta_3 \subset \dots$
- $\widehat{\theta}^m = (Y_1/\lambda_1, \dots, Y_m/\lambda_m, 0, \dots)$ projection estimator

Prior conditional on $m \in \mathbb{N}$:

- Construct sequence of prior distributions $(P_{\vartheta^m})_{m \geqslant 1}$ depending on a hyper parameter m: given $m \in \mathbb{N}$
  - first m random parameters $\{\vartheta_j^m\}_{j=1}^m$ non-degenerated
  - $\{\vartheta_j^m\}_{j>m}$ degenerated
  - independent random variables $\{\vartheta_j^m\}_{j \geqslant 1}$ with marginals:
- Notation: $\vartheta^m = (\vartheta_j^m)_{j \geqslant 1}$.

Sieve prior (I)

Prior conditional on $m \in \mathbb{N}$:

- Construct sequence of prior distributions $(P_{\vartheta^m})_{m \geqslant 1}$ depending on a hyper parameter m: given $m \in \mathbb{N}$
    - first m random parameters $\{\vartheta_j^m\}_{j=1}^m$ non-degenerated
    - $\{\vartheta_j^m\}_{j>m}$ degenerated
    - independent random variables $\{\vartheta_j^m\}_{j \geqslant 1}$ with marginals:
        - $\vartheta_j^m \sim \mathcal{N}(\theta_j^\times, \varsigma_j), \qquad 1 \leqslant j \leqslant m$
        - and $\vartheta_j^m \sim \delta_{\theta_j^\times}, \qquad m < j.$
- Notation: $\vartheta^m = \left(\vartheta_j^m\right)_{j \geqslant 1}$.

Prior conditional on $m \in \mathbb{N}$:

- Construct sequence of prior distributions $(P_{\vartheta^m})_{m \geqslant 1}$ depending on a hyper parameter m: given $m \in \mathbb{N}$
  - first m random parameters $\{\vartheta_j^m\}_{j=1}^m$ non-degenerated
  - $\{\vartheta_j^m\}_{j>m}$ degenerated
  - independent random variables $\{\vartheta_j^m\}_{j \geqslant 1}$ with marginals:
    - $\vartheta_j^m \sim \mathcal{N}(\theta_j^\times, \varsigma_j), \qquad 1 \leqslant j \leqslant m$
    - and $\vartheta_j^m \sim \delta_{\theta_j^\times}, \qquad m < j.$
- Notation: $\vartheta^m = \left(\vartheta_j^m\right)_{j \geqslant 1}$.

Sieve prior (I)

Prior conditional on $m \in \mathbb{N}$:

- Construct sequence of prior distributions $(P_{\vartheta^m})_{m \geqslant 1}$ depending on a hyper parameter m: given $m \in \mathbb{N}$
  - first m random parameters $\{\vartheta_j^m\}_{j=1}^m$ non-degenerated
  - $\{\vartheta_j^m\}_{j>m}$ degenerated
  - independent random variables $\{\vartheta_j^m\}_{j \geqslant 1}$ with marginals:
    - $\vartheta_j^m \sim \mathcal{N}(\theta_j^\times, \varsigma_j), \qquad 1 \leqslant j \leqslant m$
    - and $\vartheta_j^m \sim \delta_{\theta_j^\times}, \qquad m < j.$

- Notation: $\vartheta^m = \left(\vartheta_j^m\right)_{j \geqslant 1}$.

Sieve prior (I)

Prior conditional on $m \in \mathbb{N}$:

- Construct sequence of prior distributions $(P_{\boldsymbol{\vartheta}^m})_{m \geqslant 1}$ depending on a hyper parameter m: given $m \in \mathbb{N}$
  - first m random parameters $\{\boldsymbol{\vartheta}_j^m\}_{j=1}^m$ non-degenerated
  - $\{\boldsymbol{\vartheta}_j^m\}_{j>m}$ degenerated
  - independent random variables $\{\boldsymbol{\vartheta}_j^m\}_{j \geqslant 1}$ with marginals:
    - $\boldsymbol{\vartheta}_j^m \sim \mathcal{N}(\theta_j^\times, \varsigma_j), \qquad 1 \leqslant j \leqslant m$
    - and $\boldsymbol{\vartheta}_j^m \sim \delta_{\theta_j^\times}, \qquad m < j.$
- Notation: $\boldsymbol{\vartheta}^m = \left(\boldsymbol{\vartheta}_j^m\right)_{j \geqslant 1}$.

Sieve prior (II)

Prior on M:

- The thresholding parameter m is a hyper-parameter: we introduce a prior on the r.v. M.

- Random thresholding parameter M taking values in $\{1, \dots, G_\varepsilon\}$ for some $G_\varepsilon \in \mathbb{N}$ with prior distribution P_M.

- Distribution of the r.v.s $\{\mathrm{Y}_j\}_{j \geqslant 1}$ and $\{\vartheta_j^\mathrm{M}\}_{j \geqslant 1}$, conditionally on M:

$$\mathrm{Y}_j = \lambda_j \vartheta^\mathrm{M} + \sqrt{\varepsilon}\xi_j \quad \text{and} \quad \vartheta_j^\mathrm{M} = \theta_j^\times + \sqrt{\varsigma_j}\eta_j \mathbb{1}_{\{1 \leqslant j \leqslant \mathrm{M}\}}$$

where $\{\xi_j, \eta_j\}_{j \geqslant 1}$ are iid. standard normal random variables independent of M.

Prior on M:

- The thresholding parameter m is a hyper-parameter: we introduce a prior on the r.v. M.
  - Random thresholding parameter M taking values in $\{1, \ldots, G_\varepsilon\}$ for some $G_\varepsilon \in \mathbb{N}$ with prior distribution P_{M}.
  - Distribution of the r.v.s $\{\mathrm{Y}_j\}_{j \geqslant 1}$ and $\{\vartheta_j^{\mathrm{M}}\}_{j \geqslant 1}$, conditionally on M:

$$\mathrm{Y}_j = \lambda_j \vartheta^{\mathrm{M}} + \sqrt{\varepsilon}\xi_j \quad \text{and} \quad \vartheta_j^{\mathrm{M}} = \theta_j^{\times} + \sqrt{\varsigma_j}\eta_j \mathbb{1}_{\{1 \leqslant j \leqslant \mathrm{M}\}}$$

where $\{\xi_j, \eta_j\}_{j \geqslant 1}$ are iid. standard normal random variables independent of M.

Prior on M:

- The thresholding parameter m is a hyper-parameter: we introduce a prior on the r.v. M.

  - Random thresholding parameter M taking values in $\{1, \dots, G_\varepsilon\}$ for some $G_\varepsilon \in \mathbb{N}$ with prior distribution P_{M}.

  - Distribution of the r.v.s $\{\mathrm{Y}_j\}_{j \geqslant 1}$ and $\{\boldsymbol{\vartheta}_j^{\mathrm{M}}\}_{j \geqslant 1}$, conditionally on M:

$$\mathrm{Y}_j = \lambda_j \boldsymbol{\vartheta}^{\mathrm{M}} + \sqrt{\varepsilon}\xi_j \quad \text{and} \quad \boldsymbol{\vartheta}_j^{\mathrm{M}} = \theta_j^\times + \sqrt{\varsigma_j}\eta_j \mathbb{1}_{\{1 \leqslant j \leqslant \mathrm{M}\}}$$

  where $\{\xi_j, \eta_j\}_{j \geqslant 1}$ are iid. standard normal random variables independent of M.

Posterior distribution (I)

- Posterior distribution $P_{\vartheta^m \,|\, \mathrm{Y}}$ of ϑ^m given Y:
  - $\{\vartheta_j^m\}_{j=1}^m$ are independent, normally distributed with $\forall j \in [1, m]$
    - posterior mean $\theta_j^{\mathrm{Y}} := \mathbb{E}[\vartheta_j^m \,|\, \mathrm{Y}] = \sigma_j(\varsigma_j^{-1}\theta_j^{\times} + \lambda_j \varepsilon^{-1}\, \mathrm{Y}_j)$,
    - posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}(\vartheta_j \,|\, \mathrm{Y}) = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1}$.
  - $\{\vartheta_j^m\}_{j>m}$ degenerate on $\theta_j^{\times}$ for $j > m$.

- Posterior mean estimator of θ: $\widehat{\theta}^m = \left(\widehat{\theta}_j^m\right)_{j\geqslant 1} := \mathbb{E}[\vartheta^m \,|\, \mathrm{Y}]$ given for $j \geqslant 1$ by

$$\widehat{\theta}_j^m := \theta_j^{\mathrm{Y}}\,\mathbb{1}_{\{j \leqslant m\}} + \theta_j^{\times}\,\mathbb{1}_{\{j > m\}}.$$

Posterior distribution (I)

- Posterior distribution $P_{\vartheta^m \mid Y}$ of ϑ^m given Y:
  - $\{\vartheta_j^m\}_{j=1}^m$ are independent, normally distributed with $\forall j \in [1, m]$
    - posterior mean $\theta_j^Y := \mathbb{E}[\vartheta_j^m \mid Y] = \sigma_j(\varsigma_j^{-1}\theta_j^\times + \lambda_j \varepsilon^{-1} Y_j)$,
    - posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}(\vartheta_j \mid Y) = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1}$.
  - $\{\vartheta_j^m\}_{j>m}$ degenerate on $\theta_j^\times$ for $j > m$.
- Posterior mean estimator of θ: $\widehat{\theta}^m = \left(\widehat{\theta}_j^m\right)_{j \geqslant 1} := \mathbb{E}[\vartheta^m \mid Y]$ given for $j \geqslant 1$ by

$$\widehat{\theta}_j^m := \theta_j^Y \mathbb{1}_{\{j \leqslant m\}} + \theta_j^\times \mathbb{1}_{\{j > m\}}.$$

Posterior distribution (I)

- Posterior distribution $P_{\vartheta^m \,|\, \mathrm{Y}}$ of ϑ^m given Y:
  - $\{\vartheta_j^m\}_{j=1}^m$ are independent, normally distributed with $\forall j \in [1, m]$
    - posterior mean $\theta_j^{\mathrm{Y}} := \mathbb{E}[\vartheta_j^m \,|\, \mathrm{Y}] = \sigma_j(\varsigma_j^{-1}\theta_j^{\times} + \lambda_j \varepsilon^{-1}\, \mathrm{Y}_j)$,
    - posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}(\vartheta_j \,|\, \mathrm{Y}) = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1}$.
  - $\{\vartheta_j^m\}_{j>m}$ degenerate on $\theta_j^{\times}$ for $j > m$.
- Posterior mean estimator of θ: $\widehat{\theta}^m = \left(\widehat{\theta}_j^m\right)_{j\geqslant 1} := \mathbb{E}[\vartheta^m \,|\, \mathrm{Y}]$ given for $j \geqslant 1$ by

$$\widehat{\theta}_j^m := \theta_j^{\mathrm{Y}} \mathbb{1}_{\{j \leqslant m\}} + \theta_j^{\times} \mathbb{1}_{\{j > m\}}.$$

- Posterior distribution $P_{\vartheta^m \mid Y}$ of ϑ^m given Y:
  - $\{\vartheta_j^m\}_{j=1}^m$ are independent, normally distributed with $\forall j \in [1, m]$
    - posterior mean $\theta_j^Y := \mathbb{E}[\vartheta_j^m \mid Y] = \sigma_j(\varsigma_j^{-1}\theta_j^\times + \lambda_j \varepsilon^{-1} Y_j)$,
    - posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}(\vartheta_j \mid Y) = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1}$.
  - $\{\vartheta_j^m\}_{j>m}$ degenerate on $\theta_j^\times$ for $j > m$.
- Posterior mean estimator of θ: $\widehat{\theta}^m = \left(\widehat{\theta}_j^m\right)_{j \geqslant 1} := \mathbb{E}[\vartheta^m \mid Y]$ given for $j \geqslant 1$ by

$$\widehat{\theta}_j^m := \theta_j^Y \mathbb{1}_{\{j \leqslant m\}} + \theta_j^\times \mathbb{1}_{\{j > m\}}.$$

- Posterior distribution $P_{\vartheta^m \,|\, Y}$ of ϑ^m given Y:
  - $\{\vartheta_j^m\}_{j=1}^m$ are independent, normally distributed with $\forall j \in [1, m]$
    - posterior mean $\theta_j^{Y} := \mathbb{E}[\vartheta_j^m \,|\, Y] = \sigma_j(\varsigma_j^{-1}\theta_j^{\times} + \lambda_j \varepsilon^{-1}\, Y_j)$,
    - posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}(\vartheta_j \,|\, Y) = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1}$.
  - $\{\vartheta_j^m\}_{j>m}$ degenerate on $\theta_j^{\times}$ for $j > m$.

- Posterior mean estimator of θ: $\widehat{\theta}^m = \big(\widehat{\theta}_j^m\big)_{j\geqslant 1} := \mathbb{E}[\vartheta^m \,|\, Y]$ given for $j \geqslant 1$ by

$$\widehat{\theta}_j^m := \theta_j^{Y}\, \mathbb{1}_{\{j \,\leqslant\, m\}} + \theta_j^{\times}\, \mathbb{1}_{\{j \,>\, m\}}.$$

Remark. Improper prior:

$$\theta^\times = \left(\theta_j^\times\right)_{j\geqslant 1} \equiv 0 \quad \text{and} \quad \varsigma = \left(\varsigma_j\right)_{j\geqslant 1} \equiv \infty.$$

Posterior mean and variance:

$$\widehat{\theta}^m = Y_j/\lambda_j \mathbb{1}_{\{1\leqslant j\leqslant m\}} \quad \text{and} \quad \sigma = \varepsilon/\lambda^2 = (\varepsilon/\lambda_j^2)\mathbb{1}_{\{1\leqslant j\leqslant m\}}$$

$\forall m \in \mathbb{N}$, $\widehat{\theta}^m$ corresponds to an orthogonal projection estimator.

Posterior distribution (III)

- Posterior mean under the hierarchical prior: $\widehat{\theta} := \mathbb{E}[\vartheta^{\mathsf{M}} \mid \mathsf{Y}]$ satisfies
  - for $j > G_\varepsilon$: $\widehat{\theta}_j = \theta_j^\times$ and
  - for all $1 \leqslant j \leqslant G_\varepsilon$:

$$\widehat{\theta}_j = \theta_j^\times \, P(1 \leqslant \mathsf{M} < j | \mathsf{Y}) + \theta_j^{\mathsf{Y}} \, P(j \leqslant \mathsf{M} \leqslant G_\varepsilon | \mathsf{Y}).$$

- With the improper prior: the posterior mean is

$$\widehat{\theta}_j = P(j \leqslant \mathsf{M} \leqslant G_\varepsilon | \mathsf{Y}) \times Y_j / \lambda_j \mathbb{1}_{\{1 \leqslant j \leqslant G_\varepsilon\}}.$$

Posterior distribution (III)

- Posterior mean under the hierarchical prior: $\widehat{\theta} := \mathbb{E}[\vartheta^{\mathsf{M}} \,|\, \mathsf{Y}]$ satisfies
  - for $j > G_\varepsilon$: $\widehat{\theta}_j = \theta_j^{\times}$ and
  - for all $1 \leqslant j \leqslant G_\varepsilon$:

$$\widehat{\theta}_j = \theta_j^{\times}\, P(1 \leqslant \mathsf{M} < j | \,\mathsf{Y}) + \theta_j^{\mathsf{Y}}\, P(j \leqslant \mathsf{M} \leqslant G_\varepsilon | \,\mathsf{Y}).$$

- With the improper prior: the posterior mean is

$$\widehat{\theta}_j = P(j \leqslant \mathsf{M} \leqslant G_\varepsilon | \,\mathsf{Y}) \times Y_j/\lambda_j \mathbb{1}_{\{1 \leqslant j \leqslant G_\varepsilon\}}.$$

- Posterior mean under the hierarchical prior: $\widehat{\theta} := \mathbb{E}[\vartheta^{\mathsf{M}} \,|\, \mathsf{Y}]$ satisfies
  - for $j > G_\varepsilon$: $\widehat{\theta}_j = \theta_j^\times$ and
  - for all $1 \leqslant j \leqslant G_\varepsilon$:

$$\widehat{\theta}_j = \theta_j^\times \, P(1 \leqslant \mathsf{M} < j | \, \mathsf{Y}) + \theta_j^{\mathsf{Y}} \, P(j \leqslant \mathsf{M} \leqslant G_\varepsilon | \, \mathsf{Y}).$$

- With the improper prior: the posterior mean is

$$\widehat{\theta}_j = P(j \leqslant \mathsf{M} \leqslant G_\varepsilon | \, \mathsf{Y}) \times \, Y_j / \lambda_j \mathbb{1}_{\{1 \leqslant j \leqslant G_\varepsilon\}}.$$

Define for $j, m \in \mathbb{N}$:

$$\Lambda_j := \lambda_j^{-2}, \qquad \text{and} \qquad \Lambda_{(m)} := \max_{1 \leqslant j \leqslant m} \Lambda_j.$$

Set $\Lambda_1 = 1$ w.l.g.

Assumption A.1

Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant 1\}$. There exists a finite constant $d > 0$ such that

$$\varsigma_j \geqslant d \varepsilon^{1/2} \Lambda_j^{1/2}$$

for all $1 \leqslant j \leqslant G_\varepsilon$ and for all $\varepsilon \in (0, 1)$.

Define for $j, m \in \mathbb{N}$:

$$\Lambda_j := \lambda_j^{-2}, \qquad \text{and} \qquad \Lambda_{(m)} := \max_{1 \leqslant j \leqslant m} \Lambda_j.$$

Set $\Lambda_1 = 1$ w.l.g.

Assumption A.1

Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant 1\}$*. There exists a finite constant* $d > 0$ *such that*

$$\varsigma_j \geqslant d \varepsilon^{1/2} \Lambda_j^{1/2}$$

for all $1 \leqslant j \leqslant G_\varepsilon$ *and for all* $\varepsilon \in (0, 1)$*.*

Define for $j, m \in \mathbb{N}$:

$$\Lambda_j := \lambda_j^{-2}, \qquad \text{and} \qquad \Lambda_{(m)} := \max_{1 \leqslant j \leqslant m} \Lambda_j.$$

Set $\Lambda_1 = 1$ w.l.g.

Assumption A.1

Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant 1\}$*. There exists a finite constant* $d > 0$ *such that*

$$\varsigma_j \geqslant d \varepsilon^{1/2} \Lambda_j^{1/2}$$

for all $1 \leqslant j \leqslant G_\varepsilon$ *and for all* $\varepsilon \in (0, 1)$*.*

Prior Variances that satisfy Assumption A.1:

- a constant $\zeta_j \geqslant 1$;
- $\zeta_j = \infty$;
- $\zeta_j \geqslant d \left(\frac{\Lambda_{(j)}}{\Lambda_{(G_\varepsilon)}} \right)^{1/2}$.

$[P]$ case where $\lambda_j^2 \asymp j^{-2a}$ with $a > 0$:

- In this case $\varepsilon^{1/2} \Lambda_j^{1/2} \leqslant 1$ for all $1 \leqslant j \leqslant G_\varepsilon$.
- $\zeta_j \geqslant d\, j^a \sqrt{\varepsilon}$.

Prior Variances that satisfy Assumption A.1:

- a constant $\zeta_j \geqslant 1$;
- $\zeta_j = \infty$;
- $\zeta_j \geqslant d \left(\frac{\Lambda_{(j)}}{\Lambda_{(G_\varepsilon)}}\right)^{1/2}$.

$[P]$ case where $\lambda_j^2 \asymp j^{-2a}$ with $a > 0$:

- In this case $\varepsilon^{1/2}\Lambda_j^{1/2} \leqslant 1$ for all $1 \leqslant j \leqslant G_\varepsilon$.
- $\zeta_j \geqslant d\, j^a \sqrt{\varepsilon}$.

Outline:

Non-asymptotic posterior tails bounds (I)

Define:

$$\mathfrak{b}_m := \sum_{j>m} (\theta_j^\circ - \theta_j^\times)^2, \quad m\overline{\sigma}_m := \sum_{j=1}^{m} \sigma_j \quad \text{with } \sigma_j = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1};$$

$$\sigma_{(m)} := \max_{1 \leqslant j \leqslant m} \sigma_j \quad \text{and} \quad \mathfrak{r}_m := \sum_{j=1}^{m} (\mathbb{E}_{\theta^\circ}[\theta_j^Y] - \theta_j^\circ)^2 = \sum_{j=1}^{m} \frac{\sigma_j^2}{\varsigma_j^2} (\theta_j^\times - \theta_j^\circ)^2$$

where

- $\mathfrak{b}_m$ and $\mathfrak{r}_m$ characterize the squared bias of the Bayes estimator of θ°:

$$\|\mathbb{E}_{\theta^\circ}[\widehat{\theta}^m] - \theta^\circ\|^2 = \mathfrak{b}_m + \mathfrak{r}_m,$$

- $m\overline{\sigma}_m$ is the expectation, taken w.r.t. the posterior distribution, of $\|\vartheta^m - \widehat{\theta}^m\|^2$,
- $\sigma_{(m)}$ is the maximum posterior variance.

Non-asymptotic posterior tails bounds (I)

Define:

$$\mathfrak{b}_m := \sum_{j>m}(\theta^\circ_j - \theta^\times_j)^2, \quad m\overline{\sigma}_m := \sum_{j=1}^{m} \sigma_j \quad \text{with } \sigma_j = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1};$$

$$\sigma_{(m)} := \max_{1\leqslant j\leqslant m} \sigma_j \quad \text{and} \quad \mathfrak{r}_m := \sum_{j=1}^{m}(\mathbb{E}_{\theta^\circ}[\theta^Y_j] - \theta^\circ_j)^2 = \sum_{j=1}^{m} \frac{\sigma_j^2}{\varsigma_j^2}(\theta^\times_j - \theta^\circ_j)^2$$

where

- $\mathfrak{b}_m$ and $\mathfrak{r}_m$ characterize the squared bias of the Bayes estimator of θ°:

$$\|\mathbb{E}_{\theta^\circ}[\widehat{\theta}^m] - \theta^\circ\|^2 = \mathfrak{b}_m + \mathfrak{r}_m,$$

- $m\overline{\sigma}_m$ is the expectation, taken w.r.t. the posterior distribution, of $\|\vartheta^m - \widehat{\theta}^m\|^2$,
- $\sigma_{(m)}$ is the maximum posterior variance.

Proposition 3.1

For all $m \in \mathbb{N}$, *for all* $\varepsilon > 0$ *and for all* $0 < c < 1/5$ *we have*

$$\mathbb{E}_{\theta^\circ} P_{\vartheta^m \mid Y}\left(\|\vartheta^m - \theta^\circ\|^2 > \mathfrak{b}_m + 3m\overline{\sigma}_m + 3\, m\, \sigma_{(m)}/2 + 4\mathfrak{r}_m \right) \leqslant 2 \exp\left(-\frac{m}{36}\right); \tag{2}$$

$$\mathbb{E}_{\theta^\circ} P_{\vartheta^m \mid Y}\left(\|\vartheta^m - \theta^\circ\|^2 < \mathfrak{b}_m + m\overline{\sigma}_m - 4\, c\, (m\, \sigma_{(m)} + \mathfrak{r}_m) \right)\big) \leqslant 2 \exp\left(-\frac{c^2 m}{2}\right). \tag{3}$$

▸ Proof

This is a non asymptotic result.

Consider sub-family $\{P_{\vartheta^{m_\varepsilon}}\}_{m_\varepsilon}$ in dependence of ε.

Assumption A.2

There exist constants $0 < \varepsilon_\circ := \varepsilon_\circ(\theta^\circ, \lambda, \theta^\times, \varsigma) < 1$ *and* $1 \leqslant K := K(\theta^\circ, \lambda, \theta^\times, \varsigma) < \infty$ *such that* $\{P_{\vartheta^{m_\varepsilon}}\}_{m_\varepsilon}$ *satisfies*

$$\sup_{0<\varepsilon<\varepsilon_\circ} (\mathfrak{r}_{m_\varepsilon} \vee m_\varepsilon \sigma_{(m_\varepsilon)})/(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}) \leqslant K.$$

Proposition 3.2 (Posterior consistency)

Let Assumption A.2 be satisfied. If $m_\varepsilon \to \infty$, *then*

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\vartheta^{m_\varepsilon} \mid Y}\big((10K)^{-1}[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}] \leqslant \|\vartheta^{m_\varepsilon} - \theta^\circ\|^2 \leqslant 10K[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}]\big) = 1.$$

Moreover, if $m_\varepsilon \overline{\sigma}_{m_\varepsilon} = o(1)$ *as* $\varepsilon \to 0$, *then* $[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}] = o(1)$.

Consider sub-family $\{P_{\vartheta^{m_\varepsilon}}\}_{m_\varepsilon}$ in dependence of ε.

Assumption A.2

There exist constants $0 < \varepsilon_\circ := \varepsilon_\circ(\theta^\circ, \lambda, \theta^\times, \varsigma) < 1$ *and* $1 \leqslant K := K(\theta^\circ, \lambda, \theta^\times, \varsigma) < \infty$ *such that* $\{P_{\vartheta^{m_\varepsilon}}\}_{m_\varepsilon}$ *satisfies*

$$\sup_{0<\varepsilon<\varepsilon_\circ} (\mathfrak{r}_{m_\varepsilon} \vee m_\varepsilon \sigma_{(m_\varepsilon)})/(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}) \leqslant K.$$

Proposition 3.2 (Posterior consistency)

Let Assumption A.2 be satisfied. If $m_\varepsilon \to \infty$, *then*

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\vartheta^{m_\varepsilon} \mid Y}\big((10K)^{-1}[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}] \leqslant \|\vartheta^{m_\varepsilon} - \theta^\circ\|^2 \leqslant 10K[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}]\big) = 1.$$

Moreover, if $m_\varepsilon \overline{\sigma}_{m_\varepsilon} = o(1)$ *as* $\varepsilon \to 0$, *then* $[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}] = o(1)$.

Consider sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ in dependence of ε.

Assumption A.2

There exist constants $0 < \varepsilon_\circ := \varepsilon_\circ(\theta^\circ, \lambda, \theta^\times, \varsigma) < 1$ *and* $1 \leqslant K := K(\theta^\circ, \lambda, \theta^\times, \varsigma) < \infty$ *such that* $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ *satisfies*

$$\sup_{0<\varepsilon<\varepsilon_\circ} (\mathfrak{r}_{m_\varepsilon} \vee m_\varepsilon \sigma_{(m_\varepsilon)})/(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}) \leqslant K.$$

Proposition 3.2 (Posterior consistency)

Let Assumption A.2 be satisfied. If $m_\varepsilon \to \infty$, *then*

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \mid Y}\Big((10K)^{-1}[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}] \leqslant \|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2 \leqslant 10K[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}]\Big) = 1.$$

Moreover, if $m_\varepsilon \overline{\sigma}_{m_\varepsilon} = o(1)$ *as* $\varepsilon \to 0$, *then* $[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}] = o(1)$.

Posterior consistency (II)

- Proposition 3.2 establishes that $(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$ is up to a constant a lower and upper bound of the posterior concentration.
- Since $(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1} \to 0$, it is a posterior concentration rate.

We can consider two cases:

- $\lambda_j \to \infty$. Then $m_\varepsilon \overline{\sigma}_{m_\varepsilon} \asymp \varepsilon c, \forall m_\varepsilon$;
- $\lambda_j = O(1)$ or $\lambda_j = o(1)$, then we have to chose m_ε such that $m_\varepsilon \to \infty$ and $m_\varepsilon \overline{\sigma}_{m_\varepsilon} = o(1)$ and Proposition 3.2 gives consistency but the convergence can be arbitrarily slow.

- Proposition 3.2 establishes that $(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$ is up to a constant a lower and upper bound of the posterior concentration.
- Since $(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1} \to 0$, it is a posterior concentration rate.

We can consider two cases:

- $\lambda_j \to \infty$. Then $m_\varepsilon \overline{\sigma}_{m_\varepsilon} \asymp \varepsilon c, \forall m_\varepsilon$;
- $\lambda_j = O(1)$ or $\lambda_j = o(1)$, then we have to chose m_ε such that $m_\varepsilon \to \infty$ and $m_\varepsilon \overline{\sigma}_{m_\varepsilon} = o(1)$ and Proposition 3.2 gives consistency but the convergence can be arbitrarily slow.

Consistency of the Bayes estimator

$(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$ is also an upper bound of the frequentist risk of $\widehat{\theta}^{m_\varepsilon}$:

Proposition 3.3 (Bayes estimator consistency)

Let Assumption A.2 be satisfied. Consider the Bayes estimator $\widehat{\theta}^{m_\varepsilon} := \mathbb{E}[\vartheta^{m_\varepsilon} \mid Y]$ then

$$\mathbb{E}_{\theta^\circ} \| \widehat{\theta}^{m_\varepsilon} - \theta^\circ \|^2 \leqslant (2 + K)[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}]$$

and consequently $\mathbb{E}_{\theta^\circ} \| \widehat{\theta}^{m_\varepsilon} - \theta^\circ \|^2 = o(1)$ if $m_\varepsilon \to \infty$ and $m_\varepsilon \overline{\sigma}_{m_\varepsilon} = o(1)$ as $\varepsilon \to 0$. Proof

Consistency of the Bayes estimator

$(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$ is also an upper bound of the frequentist risk of $\widehat{\theta}^{m_\varepsilon}$:

Proposition 3.3 (Bayes estimator consistency)

Let Assumption A.2 be satisfied. Consider the Bayes estimator $\widehat{\theta}^{m_\varepsilon} := \mathbb{E}[\boldsymbol{\vartheta}^{m_\varepsilon} \mid Y]$ then

$$\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|^2 \leqslant (2+K)[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon \overline{\sigma}_{m_\varepsilon}]$$

and consequently $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|^2 = o(1)$ if $m_\varepsilon \to \infty$ and $m_\varepsilon \overline{\sigma}_{m_\varepsilon} = o(1)$ as $\varepsilon \to 0$. ▸ Proof

Posterior consistency (III)

Recall: $\Lambda_j := \lambda_j^{-2}$ and $\Lambda_{(m)} := \max_{1 \leqslant j \leqslant m} \Lambda_j$ for $j, m \in \mathbb{N}$ and define

$$\overline{\Lambda}_m := m^{-1} \sum_{j=1}^{m} \Lambda_j \quad \text{and} \quad \Phi_\varepsilon^m := [\mathfrak{b}_m \vee \varepsilon\, m \overline{\Lambda}_m], \qquad \text{for } m \in \mathbb{N}.$$

Assume Assumption A.1 holds.

If, in addition, $\exists$ constant $1 \leqslant L := L(\theta^\circ, \lambda, \theta^\times) < \infty$ such that

$$\sup_{0<\varepsilon<1} \varepsilon\, m_\varepsilon \Lambda_{(m_\varepsilon)} (\Phi_\varepsilon^{m_\varepsilon})^{-1} \leqslant L \tag{4}$$

and $m_\varepsilon \leqslant G_\varepsilon$ $\forall \varepsilon \to 0$, then $\{P_{\vartheta^{m_\varepsilon}}\}_{m_\varepsilon}$ satisfies Assumption A.2 with

$$K := ((1 + d^{-1}) \vee d^{-2} \|\theta^\circ - \theta^\times\|^2) L.$$

In the polynomial case, (4) is satisfied.

Posterior consistency (III)

Recall: $\Lambda_j := \lambda_j^{-2}$ and $\Lambda_{(m)} := \max_{1\leqslant j\leqslant m} \Lambda_j$ for $j, m \in \mathbb{N}$ and define

$$\overline{\Lambda}_m := m^{-1}\sum_{j=1}^{m} \Lambda_j \quad \text{and} \quad \Phi_\varepsilon^m := [\mathfrak{b}_m \vee \varepsilon\, m\,\overline{\Lambda}_m], \qquad \text{for } m \in \mathbb{N}.$$

Assume Assumption A.1 holds.

If, in addition, $\exists$ constant $1 \leqslant L := L(\theta^\circ, \lambda, \theta^\times) < \infty$ such that

$$\sup_{0<\varepsilon<1} \varepsilon\, m_\varepsilon\, \Lambda_{(m_\varepsilon)} (\Phi_\varepsilon^{m_\varepsilon})^{-1} \leqslant L \tag{4}$$

and $m_\varepsilon \leqslant G_\varepsilon$ $\forall \varepsilon \to 0$, then $\{P_{\vartheta^{m_\varepsilon}}\}_{m_\varepsilon}$ satisfies Assumption A.2 with

$$K := ((1 + d^{-1}) \vee d^{-2}\|\theta^\circ - \theta^\times\|^2)L.$$

In the polynomial case, (4) is satisfied.

Corollary 3.4

Under Assumption A.1 consider a sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ *such that (4) is satisfied and* $m_\varepsilon \leqslant G_\varepsilon$ $\forall \varepsilon \to 0$*, then* $\forall \varepsilon > 0$ *and* $0 < c < 1/(8K)$ *with* $K = ((1+d^{-1}) \vee d^{-2}\|\theta^\circ - \theta^\times\|^2)L$ *it holds:*

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \mid Y}\big(\|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2 > (4 + (11/2)K)\Phi_\varepsilon^{m_\varepsilon}\big) \leqslant 2\exp(-\frac{m_\varepsilon}{36}); \tag{5}$$

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \mid Y}\big(\|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2 < (1 - 8\,c\,K)(1 + d^{-1})^{-1}\Phi_\varepsilon^{m_\varepsilon}\big) \leqslant 2\exp(-c^2 m_\varepsilon/2). \tag{6}$$

Moreover, $\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|^2 \leqslant (2 + K)\Phi_\varepsilon^{m_\varepsilon}$*.*

Minimise the rate $\Phi_\varepsilon^{m_\varepsilon}$ for each θ° separately. Define $\forall \varepsilon > 0$

$$m_\varepsilon^\circ := \arg\min_{m \geqslant 1} \{\Phi_\varepsilon^m\} \text{ and}$$

$$\Phi_\varepsilon^\circ := \Phi_\varepsilon^{m_\varepsilon^\circ} = \min_{m \geqslant 1} \Phi_\varepsilon^m = \min_{m \geqslant 1} [\mathfrak{b}_m \vee \varepsilon\, m \overline{\Lambda}_m].$$

Theorem 3.5 (Oracle Bayes estimator)

Consider the family $\{\widehat{\theta}^m\}_m$ of Bayes estimators. Under Assumption A.1 we have

(i) $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\circ} - \theta^\circ\|^2 \leqslant (2 + d^{-2}\|\theta^\circ - \theta^\times\|^2)\Phi_\varepsilon^\circ$ *and*

(ii) $\inf_{m \geqslant 1} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \geqslant (1 + 1/d)^{-2}\Phi_\varepsilon^\circ$ *for all* $\varepsilon \in (0, \varepsilon_o)$.

Minimise the rate $\Phi_\varepsilon^{m_\varepsilon}$ for each θ° separately. Define $\forall \varepsilon > 0$

$$m_\varepsilon^\circ := \arg\min_{m \geqslant 1} \{\Phi_\varepsilon^m\} \text{ and}$$

$$\Phi_\varepsilon^\circ := \Phi_\varepsilon^{m_\varepsilon^\circ} = \min_{m \geqslant 1} \Phi_\varepsilon^m = \min_{m \geqslant 1} [\mathfrak{b}_m \vee \varepsilon\, m\, \overline{\Lambda}_m].$$

Theorem 3.5 (Oracle Bayes estimator)

Consider the family $\{\widehat{\theta}^m\}_m$ of Bayes estimators. Under Assumption A.1 we have

(i) $\mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\circ} - \theta^\circ\|^2 \leqslant (2 + d^{-2}\|\theta^\circ - \theta^\times\|^2)\Phi_\varepsilon^\circ$ *and*

(ii) $\inf_{m \geqslant 1} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^m - \theta^\circ\|^2 \geqslant (1 + 1/d)^{-2}\Phi_\varepsilon^\circ$ *for all* $\varepsilon \in (0, \varepsilon_o)$.

Theorem 3.6 (Oracle posterior concentration rate)

Suppose that Assumption A.1 holds true and that there exists a constant $1 \leqslant L^\circ := L^\circ(\theta^\circ, \lambda, \theta^\times) < \infty$ *such that*

$$\sup_{0<\varepsilon<1} \varepsilon\, m_\varepsilon^\circ\, \Lambda_{(m_\varepsilon^\circ)} (\Phi_\varepsilon^\circ)^{-1} \leqslant L^\circ. \tag{7}$$

If in addition $\mathfrak{b}_m > 0$ *for all* $m \geqslant 1$ *and* $K^\circ := 10((1 + d^{-1}) \vee d^{-2}\|\theta^\circ - \theta^\times\|^2)L^\circ$*, then*

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}\,|\,Y}((K^\circ)^{-1}\Phi_\varepsilon^\circ \leqslant \|\boldsymbol{\vartheta}^{m_\varepsilon^\circ} - \theta^\circ\|^2 \leqslant K^\circ \Phi_\varepsilon^\circ) = 1.$$

Oracle concentration rate (III)

If $\mathfrak{b}_m = 0$ for some m:

- the Bayes estimator attains the parametric rate.
- Corollary 3.4 implies that the near-oracle prior family $\{P_{\boldsymbol{\vartheta}^{\tilde{m}_\varepsilon}}\}_{\tilde{m}_\varepsilon}$ is such that

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\vartheta^{\tilde{m}_\varepsilon} \mid Y}\big((K^\circ)^{-1} \Gamma_\varepsilon \varepsilon \leqslant \|\vartheta^{\tilde{m}_\varepsilon} - \theta^\circ\|^2 \leqslant K^\circ \Gamma_\varepsilon \varepsilon\big) = 1$$

where $\Gamma_\varepsilon = \tilde{m}_\varepsilon \Lambda_{\tilde{m}_\varepsilon}$ is a slowly $\uparrow$ sequence depending on $\tilde{m}_\varepsilon$.

Find a uniform rate over a class of parameters

$$\Theta_{\mathfrak{a}}^{r} := \left\{\theta \in \ell_2^{\mathfrak{a}} : \|\theta - \theta^{\times}\|_{\mathfrak{a}}^2 \leqslant r\right\}$$

where:

- $\mathfrak{a} = (\mathfrak{a}_j)_{j\geqslant 1}$ is a strictly positive and non-increasing sequence with $\mathfrak{a}_1 = 1$ and $\lim_{j\to\infty} \mathfrak{a}_j = 0$;
- for $\theta \in \ell_2$, $\|\theta\|_{\mathfrak{a}}^2 := \sum_{j\geqslant 1} \theta_j^2/\mathfrak{a}_j$ and $\ell_2^{\mathfrak{a}}$ is the completion of ℓ_2 with respect to $\|\cdot\|_{\mathfrak{a}}$.
- Assume $\theta^{\circ} \in \Theta_{\mathfrak{a}}^{r}$ and therefore, $\mathfrak{b}_m(\theta^{\circ}) \leqslant \mathfrak{a}_m r$.

Define

$$m_{\varepsilon}^{\star} := \arg\min_{m\geqslant 1} \left\{\mathfrak{a}_m \vee \varepsilon\, m\, \overline{\Lambda}_m\right\} \text{ and}$$

$$\Phi_{\varepsilon}^{\star} := \left[\mathfrak{a}_{m_{\varepsilon}^{\star}} \vee \varepsilon\, m_{\varepsilon}^{\star}\, \overline{\Lambda}_{m_{\varepsilon}^{\star}}\right] \quad \text{for all } \varepsilon > 0. \qquad (8)$$

Note that

$$\Phi_{\varepsilon}^{\circ} = \min_{m\geqslant 1}[\mathfrak{b}_m \vee \varepsilon\, m\, \overline{\Lambda}_m] \leqslant (1 \vee r) \min_{m\geqslant 1}[\mathfrak{a}_m \vee \varepsilon\, m\, \overline{\Lambda}_m] = (1 \vee r)\Phi_{\varepsilon}^{\star}.$$

Find a uniform rate over a class of parameters

$$\Theta_{\mathfrak{a}}^{r} := \left\{\theta \in \ell_2^{\mathfrak{a}} : \|\theta - \theta^{\times}\|_{\mathfrak{a}}^2 \leqslant r\right\}$$

where:

- $\mathfrak{a} = (\mathfrak{a}_j)_{j\geqslant 1}$ is a strictly positive and non-increasing sequence with $\mathfrak{a}_1 = 1$ and $\lim_{j\to\infty} \mathfrak{a}_j = 0$;
- for $\theta \in \ell_2$, $\|\theta\|_{\mathfrak{a}}^2 := \sum_{j\geqslant 1} \theta_j^2/\mathfrak{a}_j$ and $\ell_2^{\mathfrak{a}}$ is the completion of ℓ_2 with respect to $\|\cdot\|_{\mathfrak{a}}$.
- Assume $\theta^{\circ} \in \Theta_{\mathfrak{a}}^{r}$ and therefore, $\mathfrak{b}_m(\theta^{\circ}) \leqslant \mathfrak{a}_m r$.

Define

$$m_{\varepsilon}^{\star} := \operatorname*{arg\,min}_{m\geqslant 1} \left\{\mathfrak{a}_m \vee \varepsilon\, m \overline{\Lambda}_m\right\} \text{ and}$$

$$\Phi_{\varepsilon}^{\star} := [\mathfrak{a}_{m_{\varepsilon}^{\star}} \vee \varepsilon\, m_{\varepsilon}^{\star}\, \overline{\Lambda}_{m_{\varepsilon}^{\star}}] \quad \text{for all } \varepsilon > 0. \quad (8)$$

Note that

$$\Phi_{\varepsilon}^{\circ} = \min_{m\geqslant 1}[\mathfrak{b}_m \vee \varepsilon\, m \overline{\Lambda}_m] \leqslant (1 \vee r) \min_{m\geqslant 1}[\mathfrak{a}_m \vee \varepsilon\, m \overline{\Lambda}_m] = (1 \vee r)\Phi_{\varepsilon}^{\star}.$$

Find a uniform rate over a class of parameters

$$\Theta_{\mathfrak{a}}^{r} := \left\{\theta \in \ell_2^{\mathfrak{a}} : \|\theta - \theta^{\times}\|_{\mathfrak{a}}^2 \leqslant r\right\}$$

where:

- $\mathfrak{a} = (\mathfrak{a}_j)_{j\geqslant 1}$ is a strictly positive and non-increasing sequence with $\mathfrak{a}_1 = 1$ and $\lim_{j\to\infty} \mathfrak{a}_j = 0$;
- for $\theta \in \ell_2$, $\|\theta\|_{\mathfrak{a}}^2 := \sum_{j\geqslant 1} \theta_j^2/\mathfrak{a}_j$ and $\ell_2^{\mathfrak{a}}$ is the completion of ℓ_2 with respect to $\|\cdot\|_{\mathfrak{a}}$.
- Assume $\theta^{\circ} \in \Theta_{\mathfrak{a}}^{r}$ and therefore, $\mathfrak{b}_m(\theta^{\circ}) \leqslant \mathfrak{a}_m r$.

Define

$$m_{\varepsilon}^{\star} := \arg\min_{m\geqslant 1} \left\{\mathfrak{a}_m \vee \varepsilon\, m\, \overline{\Lambda}_m\right\} \text{ and}$$

$$\Phi_{\varepsilon}^{\star} := [\mathfrak{a}_{m_{\varepsilon}^{\star}} \vee \varepsilon\, m_{\varepsilon}^{\star}\, \overline{\Lambda}_{m_{\varepsilon}^{\star}}] \quad \text{for all } \varepsilon > 0. \quad (8)$$

Note that

$$\Phi_{\varepsilon}^{\circ} = \min_{m\geqslant 1}[\mathfrak{b}_m \vee \varepsilon\, m\, \overline{\Lambda}_m] \leqslant (1 \vee r) \min_{m\geqslant 1}[\mathfrak{a}_m \vee \varepsilon\, m\, \overline{\Lambda}_m] = (1 \vee r)\Phi_{\varepsilon}^{\star}.$$

We now consider $\widehat{\theta}^{m_\varepsilon^\star}$ and $\{P_{\vartheta^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ which do not depend on θ° but only on $\Theta_{\mathfrak{a}}^r$.

Theorem 3.7 (Minimax optimal Bayes estimator)

Let Assumption A.1 be satisfied. Considering the Bayes estimator $\widehat{\theta}^{m_\varepsilon^\star} := \mathbb{E}[\vartheta^{m_\varepsilon^\star} \mid Y]$ *we have*

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\star} - \theta^\circ\|^2 \leqslant (2 + r/d^2)(1 \vee r)\Phi_\varepsilon^\star \quad \textit{for all } \varepsilon \in (0, \varepsilon_o).$$

The rate $\Phi_\varepsilon^\star$ provides up to a constant a lower bound for $\sup_{\theta \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_\theta \|\widehat{\theta} - \theta\|^2$ over $\Theta_{\mathfrak{a}}^r$ (see *e.g.* Johannes and Schwarz 2013) if the next assumption is satisfied.

We now consider $\widehat{\theta}^{m_\varepsilon^\star}$ and $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ which do not depend on θ° but only on $\Theta_{\mathfrak{a}}^r$.

Theorem 3.7 (Minimax optimal Bayes estimator)

Let Assumption A.1 be satisfied. Considering the Bayes estimator $\widehat{\theta}^{m_\varepsilon^\star} := \mathbb{E}[\boldsymbol{\vartheta}^{m_\varepsilon^\star} \,|\, Y]$ *we have*

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\star} - \theta^\circ\|^2 \leqslant (2 + r/d^2)(1 \vee r)\Phi_\varepsilon^\star \quad \textit{for all } \varepsilon \in (0, \varepsilon_o).$$

The rate $\Phi_\varepsilon^\star$ provides up to a constant a lower bound for $\sup_{\theta \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_\theta \|\widehat{\theta} - \theta\|^2$ over $\Theta_{\mathfrak{a}}^r$ (see *e.g.* Johannes and Schwarz 2013) if the next assumption is satisfied.

Assumption A.3

Let $\mathfrak{a}$ and λ be sequences such that

$$0 < \kappa^\star := \inf_{0<\varepsilon<\varepsilon_o} \left\{ \frac{[\mathfrak{a}_{m_\varepsilon^\star} \wedge \varepsilon\, m_\varepsilon^\star\, \overline{\Lambda}_{m_\varepsilon^\star}]}{\Phi_\varepsilon^\star} \right\} \leqslant 1.$$

Theorem 3.8 (Minimax optimal posterior conc. rate)

Let Assumptions A.1 and A.3 hold true. If $\exists$ *a constant* $1 \leqslant L^\star < \infty$ *such that*

$$\sup_{0<\varepsilon<\varepsilon_o} \frac{\varepsilon\, m_\varepsilon^\star\, \Lambda_{(m_\varepsilon^\star)}}{\Phi_\varepsilon^\star} \leqslant L^\star \tag{9}$$

and $K^\star := 10((1+1/d) \vee r/d^2)(1 \vee r)(L^\star/\kappa^\star)$*, then*

$$\lim_{\varepsilon\to 0} \inf_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\vartheta^{m_\varepsilon^\star} \mid Y}((K^\star)^{-1}\Phi_\varepsilon^\star \leqslant \|\vartheta^{m_\varepsilon^\star} - \theta^\circ\|^2 \leqslant K^\star \Phi_\varepsilon^\star) = 1.$$

The rate $\Phi_\varepsilon^\star$ provides up to a constant a lower and an upper bound for the posterior concentration rate associated with $\{P_{\vartheta^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$.

Typical choices of the sequences $\mathfrak{a}$ and λ.

[P-P] Consider $\mathfrak{a}_j \asymp j^{-2p}$ and $\lambda_j^2 \asymp j^{-2a}$ with $p > 0$ and $a > 0$ then $m_\varepsilon^\star \asymp \varepsilon^{-1/(2p+2a+1)}$ and $\Phi_\varepsilon^\star \asymp \varepsilon^{2p/(2a+2p+1)}$.

[E-P] Consider $\mathfrak{a}_j \asymp \exp(-j^{2p}+1)$ and $\lambda_j^2 \asymp j^{-2a}$ with $p > 0$ and $a > 0$ then $m_\varepsilon^\star \asymp |\log\varepsilon - \frac{2a+1}{2p}(\log|\log\varepsilon|)|^{1/(2p)}$ and $\Phi_\varepsilon^\star \asymp \varepsilon|\log\varepsilon|^{(2a+1)/(2p)}$.

[P-E] Consider $\mathfrak{a}_j \asymp j^{-2p}$ and $\lambda_j^2 \asymp \exp(-j^{2a}+1)$, with $p > 0$ and $a > 0$ then $m_\varepsilon^\star \asymp |\log\varepsilon - \frac{2p+(2a-1)_+}{2a}(\log|\log\varepsilon|)|^{1/(2a)}$ and $\Phi_\varepsilon^\star \asymp |\log\varepsilon|^{-p/a}$.

Outline:

1. Introduction
2. Bayesian perspective
3. Posterior consistency
4. Adaptive Bayesian approach

Consider the $[P]$ case $\lambda_j^2 \asymp j^{-2a}$.

Let $C_\lambda \geqslant 1$ and $L_\lambda \geqslant 1$ be finite constants such that for all $k, l \in \mathbb{N}$:

(i) $(k+1)^{-2a} \leqslant C_\lambda k^{-2a}$;

(ii) $1 \leqslant \Lambda_{(k)}/\overline{\Lambda}_k \leqslant L_\lambda$.

- Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant 1\}$.
- Random thresholding parameter M taking values in $\{1, \dots, G_\varepsilon\}$ with prior distribution P_{M} defined as for $1 \leqslant m \leqslant G_\varepsilon$:

$$p_{\mathrm{M}}(m) := P_{\mathrm{M}}(\mathrm{M} = m) = \frac{\exp(-3C_\lambda m/2) \prod_{j=1}^{m} (\varsigma_j/\sigma_j)^{1/2}}{\sum_{k=1}^{G_\varepsilon} \exp(-3C_\lambda k/2) \prod_{j=1}^{k} (\varsigma_j/\sigma_j)^{1/2}}.$$

Remember: $\varsigma_j/\sigma_j = (\lambda_j^2 \varepsilon^{-1} \zeta_j + 1)$.

Consider the $[P]$ case $\lambda_j^2 \asymp j^{-2a}$.

Let $C_\lambda \geqslant 1$ and $L_\lambda \geqslant 1$ be finite constants such that for all $k, l \in \mathbb{N}$:
(i) $(k+1)^{-2a} \leqslant C_\lambda k^{-2a}$;
(ii) $1 \leqslant \Lambda_{(k)}/\overline{\Lambda}_k \leqslant L_\lambda$.

- Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant 1\}$.
- Random thresholding parameter M taking values in $\{1, \dots, G_\varepsilon\}$ with prior distribution P_{M} defined as for $1 \leqslant m \leqslant G_\varepsilon$:

$$p_{\mathrm{M}}(m) := P_{\mathrm{M}}(\mathrm{M} = m) = \frac{\exp(-3C_\lambda m/2) \prod_{j=1}^{m} (\varsigma_j/\sigma_j)^{1/2}}{\sum_{k=1}^{G_\varepsilon} \exp(-3C_\lambda k/2) \prod_{j=1}^{k} (\varsigma_j/\sigma_j)^{1/2}}.$$

Remember: $\varsigma_j/\sigma_j = (\lambda_j^2 \varepsilon^{-1} \zeta_j + 1)$.

Consider the $[P]$ case $\lambda_j^2 \asymp j^{-2a}$.

Let $C_\lambda \geqslant 1$ and $L_\lambda \geqslant 1$ be finite constants such that for all $k, l \in \mathbb{N}$:
(i) $(k+1)^{-2a} \leqslant C_\lambda k^{-2a}$;
(ii) $1 \leqslant \Lambda_{(k)}/\overline{\Lambda}_k \leqslant L_\lambda$.

- Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant 1\}$.
- Random thresholding parameter M taking values in $\{1, \ldots, G_\varepsilon\}$ with prior distribution P_{M} defined as for $1 \leqslant m \leqslant G_\varepsilon$:

$$p_{\mathrm{M}}(m) := P_{\mathrm{M}}(\mathrm{M} = m) = \frac{\exp(-3C_\lambda m/2) \prod_{j=1}^{m} (\varsigma_j/\sigma_j)^{1/2}}{\sum_{k=1}^{G_\varepsilon} \exp(-3C_\lambda k/2) \prod_{j=1}^{k} (\varsigma_j/\sigma_j)^{1/2}}.$$

Remember: $\varsigma_j/\sigma_j = (\lambda_j^2 \varepsilon^{-1} \zeta_j + 1)$.

The posterior distribution $P_{\mathrm{M}\,|\,\mathrm{Y}}$ of M is given by

$$\begin{aligned} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) &= P_{\mathrm{M}\,|\,\mathrm{Y}}(\mathrm{M}=m) \\ &= \frac{\exp(-\frac{1}{2}\{-\|\widehat{\theta}^m - \theta^\times\|_\sigma^2 + 3C_\lambda m\})}{\sum_{k=1}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|\widehat{\theta}^k - \theta^\times\|_\sigma^2 + 3C_\lambda k\})} \end{aligned}$$

where $\|\theta\|_\sigma^2 := \sum_{j\geqslant 1} \theta_j^2/\sigma_j$ for $\theta \in \ell_2$ and
$\widehat{\theta}_j^m := \theta_j^{\mathrm{Y}} \mathbb{1}_{\{j\leqslant m\}} + \theta_j^\times \mathbb{1}_{\{j>m\}}$.

$P_{\mathrm{M}\,|\,\mathrm{Y}}$ is concentrating in a neighborhood of $m_\varepsilon^\circ := \arg\min_{m\geqslant 1} \{\Phi_\varepsilon^m\}$ (given by $[G_\varepsilon^-, G_\varepsilon^+]$) as ε tends to zero (if $m_\varepsilon^\circ/(\log G_\varepsilon) \to \infty$), where $\forall \varepsilon \in (0, \varepsilon_\circ)$

$$G_\varepsilon^- := \min\{m \in \{1,\dots,m_\varepsilon^\circ\} : \mathfrak{b}_m \leqslant 8L_\lambda C_\lambda(1+1/d)\Phi_\varepsilon^\circ\} \quad \text{and}$$

$$G_\varepsilon^+ := \max\left\{m \in \{m_\varepsilon^\circ,\dots,G_\varepsilon\} : m \leqslant 5L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ\right\}.$$

Lemma

The posterior distribution $P_{\mathrm{M}\,|\,\mathrm{Y}}$ of M is given by

$$\begin{aligned} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) &= P_{\mathrm{M}\,|\,\mathrm{Y}}(\mathrm{M}=m) \\ &= \frac{\exp(-\frac{1}{2}\{-\|\widehat{\theta}^m - \theta^\times\|_\sigma^2 + 3C_\lambda m\})}{\sum_{k=1}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|\widehat{\theta}^k - \theta^\times\|_\sigma^2 + 3C_\lambda k\})} \end{aligned}$$

where $\|\theta\|_\sigma^2 := \sum_{j\geqslant 1} \theta_j^2/\sigma_j$ for $\theta \in \ell_2$ and
$\widehat{\theta}_j^m := \theta_j^{\mathrm{Y}} \mathbb{1}_{\{j\,\leqslant\, m\}} + \theta_j^\times \mathbb{1}_{\{j\,>\,m\}}$.

$P_{\mathrm{M}\,|\,\mathrm{Y}}$ is concentrating in a neighborhood of $m_\varepsilon^\circ := \underset{m\geqslant 1}{\arg\min}\,\{\Phi_\varepsilon^m\}$
(given by $[G_\varepsilon^-, G_\varepsilon^+]$) as ε tends to zero (if $m_\varepsilon^\circ/(\log G_\varepsilon) \to \infty$),
where $\forall \varepsilon \in (0, \varepsilon_\circ)$

$$G_\varepsilon^- := \min\{m \in \{1,\dots,m_\varepsilon^\circ\} : \mathfrak{b}_m \leqslant 8L_\lambda C_\lambda(1+1/d)\Phi_\varepsilon^\circ\} \quad \text{and}$$

$$G_\varepsilon^+ := \max\left\{m \in \{m_\varepsilon^\circ,\dots,G_\varepsilon\} : m \leqslant 5L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ\right\}.$$

▸ Lemma

Posterior of M (II)

Assumption A.4

Let $\theta^{\times}$, θ° and λ be sequences such that

$$0 < \kappa^{\circ} := \inf_{0<\varepsilon<\varepsilon_o} \left\{ \frac{[\mathfrak{b}_{m_\varepsilon^\circ} \wedge \varepsilon\, m_\varepsilon^\circ\, \overline{\Lambda}_{m_\varepsilon^\circ}]}{\Phi_\varepsilon^\circ} \right\} \leqslant 1.$$

The posterior $P_{\vartheta^{\mathrm{M}}\,|\,\mathrm{Y}}$ of $\vartheta^{\mathrm{M}} = \left(\vartheta_j^{\mathrm{M}}\right)_{j\geqslant 1}$ associated with the hierarchical prior is a weighted mixture of the posterior $\{P_{\vartheta^m\,|\,\mathrm{Y}}\}_{m=1}^{G_\varepsilon}$: $P_{\vartheta^{\mathrm{M}}\,|\,\mathrm{Y}} = \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\vartheta^m\,|\,\mathrm{Y}}$.

Assumption A.4

Let $\theta^\times$, θ° and λ be sequences such that

$$0 < \kappa^\circ := \inf_{0<\varepsilon<\varepsilon_o} \left\{ \frac{[\mathfrak{b}_{m_\varepsilon^\circ} \wedge \varepsilon\, m_\varepsilon^\circ\, \overline{\Lambda}_{m_\varepsilon^\circ}]}{\Phi_\varepsilon^\circ} \right\} \leqslant 1.$$

The posterior $P_{\boldsymbol{\vartheta}^{\mathsf{M}}\,|\,\mathsf{Y}}$ of $\boldsymbol{\vartheta}^{\mathsf{M}} = \left(\boldsymbol{\vartheta}_j^{\mathsf{M}}\right)_{j\geqslant 1}$ associated with the hierarchical prior is a weighted mixture of the posterior $\{P_{\boldsymbol{\vartheta}^m\,|\,\mathsf{Y}}\}_{m=1}^{G_\varepsilon}$: $P_{\boldsymbol{\vartheta}^{\mathsf{M}}\,|\,\mathsf{Y}} = \sum_{m=1}^{G_\varepsilon} p_{\mathsf{M}\,|\,\mathsf{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathsf{Y}}$.

Lemma 4.1

If Assumptions A.1 and A.4 hold true then for all $\varepsilon \in (0, \varepsilon_\circ)$:

(i) $\sum_{G_\varepsilon^- \leqslant m \leqslant G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \mid Y}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^o \Phi_\varepsilon^\circ\big) \leqslant 74 \exp(-G_\varepsilon^-/36)$;

(ii) $\sum_{G_\varepsilon^- \leqslant m \leqslant G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \mid Y}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 < (K^o)^{-1} \Phi_\varepsilon^\circ\big) \leqslant 4(K^\circ)^2 \exp\left(-\frac{G_\varepsilon^-}{(K^\circ)^2}\right)$,

where
$K^\circ := 10((1 + 1/d) \vee \|\theta^\circ - \theta^\times\|^2/d^2) L_\lambda^2 (8 C_\lambda (1 + 1/d) \vee D^\circ \Lambda_{(D^\circ)})$
with $D^\circ := D^\circ(\theta^\times, \theta^\circ, \lambda) := \lceil 5 L_\lambda / \kappa^\circ \rceil$.

Theorem 4.2 (Oracle posterior concentration rate)

Let Assumptions A.1 and A.4 hold true. If in addition $(\log G_\varepsilon)/m_\varepsilon^\circ \to 0$ as $\varepsilon \to 0$, then

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathsf{M}}\,|\,\mathsf{Y}}\big((K^\circ)^{-1}\Phi_\varepsilon^\circ \leqslant \|\boldsymbol{\vartheta}^{\mathsf{M}} - \theta^\circ\|^2 \leqslant K^\circ \Phi_\varepsilon^\circ\big) = 1$$

where K° is given in Lemma 4.1.

Proof

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

$$\begin{aligned}
\widehat{\theta}_j &= \theta_j^{\times}, \quad \text{for } j > G_\varepsilon \qquad \text{and} \\
\widehat{\theta}_j &= \theta_j^{\times}\, P(1 \leqslant \mathsf{M} < j | \, Y) + \theta_j^{Y}\, P(j \leqslant \mathsf{M} \leqslant G_\varepsilon \,|\, Y), \quad \text{for } 1 \leqslant j \leqslant G_\varepsilon.
\end{aligned}$$

Theorem 4.3 (Oracle optimal Bayes estimator)

If Assumptions A.1 and A.4 hold and $\log(G_\varepsilon/\Phi_\varepsilon^\circ)/m_\varepsilon^\circ \to 0$ *as* $\varepsilon \to 0$*, then there exists a constant* $K^\circ := K^\circ(\theta^\circ, \theta^\times, \lambda, d, L) < \infty$ *such that*

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|^2 \leqslant K^\circ \Phi_\varepsilon^\circ$$

for all $\varepsilon \in (0, \varepsilon_\circ)$.

Data-driven Bayesian estimation (III)

Theorem 4.4 (Minimax optimal posterior conc. rate)

Let Assumptions A.1 and A.3 hold true and $(\log G_\varepsilon)/m_\varepsilon^\star \to 0$ *as* $\varepsilon \to 0$, *then*

(i) for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ *we have*

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}} \mid \mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2 \leqslant K^\star \Phi_\varepsilon^\star) = 1$$

where
$K^\star := 16((1+1/d) \vee r/d^2) L_\lambda^2 (8C_\lambda(1+1/d) \vee D^\star \Lambda_{(D^\star)})(1 \vee r)$
with $D^\star := D^\star(\mathfrak{a}, \lambda) := \lceil 5L_\lambda/\kappa^\star \rceil$*;*

(ii) for any monotonically $\nearrow$ *and unbounded sequence* $(K_\varepsilon)_\varepsilon$*:*

$$\lim_{\varepsilon\to 0} \inf_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}} \mid \mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2 \leqslant K_\varepsilon \Phi_\varepsilon^\star) = 1.$$

Theorem 4.5 (Minimax optimal Bayes estimate)

Under Assumptions A.1 and A.3 consider the Bayes estimator $\widehat{\theta} := \mathbb{E}[\boldsymbol{\vartheta}^{\mathsf{M}} \,|\, \mathsf{Y}]$. *If in addition* $\log(G_\varepsilon / \Phi_\varepsilon^\star)/m_\varepsilon^\star \to 0$ *as* $\varepsilon \to 0$, *then there exists* $K^\star := K^\star(\Theta_{\mathfrak{a}}^r, \lambda, d) < \infty$ *such that*

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} \|\widehat{\theta} - \theta^\circ\|^2 \leqslant K^\star \Phi_\varepsilon^\star$$

for all $\varepsilon \in (0, \varepsilon_\star)$.

Remark. Recall the improper prior family $\{P_{\vartheta^m}\}_m$ with $\theta^\times = \big(\theta_j^\times\big)_{j\geqslant 1} \equiv 0 \quad \text{and} \quad \varsigma = \big(\varsigma_j\big)_{j\geqslant 1} \equiv \infty.$

The Bayes estimator is $\widehat{\theta}^m = \mathbb{E}[\vartheta^m \,|\, Y] = (Y/\lambda)^m$.

The posterior probability of M is :

$$P_{M\,|\,Y}(M = m) \propto \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\}).$$

Hence, $\widehat{\theta} = \big(\widehat{\theta}_j\big)_{j\geqslant 1} = \mathbb{E}[\vartheta^M \,|\, Y]$ equals the shrunk orthogonal projection estimator given by

$$\widehat{\theta}_j = \frac{\sum_{m=j}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})}{\sum_{m=1}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})} \times \frac{Y_j}{\lambda_j} \mathbb{1}_{\{1 \leqslant j \leqslant G_\varepsilon\}}.$$

Remark. Recall the improper prior family $\{P_{\vartheta^m}\}_m$ with $\theta^\times = (\theta^\times_j)_{j\geqslant 1} \equiv 0$ and $\varsigma = (\varsigma_j)_{j\geqslant 1} \equiv \infty.$

The Bayes estimator is $\widehat{\theta}^m = \mathbb{E}[\vartheta^m \,|\, Y] = (Y/\lambda)^m$.

The posterior probability of M is :

$$P_{M\,|\,Y}(M = m) \propto \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\}).$$

Hence, $\widehat{\theta} = (\widehat{\theta}_j)_{j\geqslant 1} = \mathbb{E}[\vartheta^M \,|\, Y]$ equals the shrunk orthogonal projection estimator given by

$$\widehat{\theta}_j = \frac{\sum_{m=j}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})}{\sum_{m=1}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})} \times \frac{Y_j}{\lambda_j} \mathbb{1}_{\{1 \leqslant j \leqslant G_\varepsilon\}}.$$

Data-driven Bayesian estimation (V)

Remark. Recall the improper prior family $\{P_{\vartheta^m}\}_m$ with
$\theta^\times = (\theta_j^\times)_{j\geqslant 1} \equiv 0$ and $\varsigma = (\varsigma_j)_{j\geqslant 1} \equiv \infty$.

The Bayes estimator is $\widehat{\theta}^m = \mathbb{E}[\vartheta^m \,|\, Y] = (Y/\lambda)^m$.

The posterior probability of M is :

$$P_{M\,|\,Y}(M = m) \propto \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\}).$$

Hence, $\widehat{\theta} = (\widehat{\theta}_j)_{j\geqslant 1} = \mathbb{E}[\vartheta^M \,|\, Y]$ equals the shrunk orthogonal projection estimator given by

$$\widehat{\theta}_j = \frac{\sum_{m=j}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})}{\sum_{m=1}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})} \times \frac{Y_j}{\lambda_j} \mathbb{1}_{\{1 \leqslant j \leqslant G_\varepsilon\}}.$$

Data-driven Bayesian estimation (V)

Remark. Recall the improper prior family $\{P_{\vartheta^m}\}_m$ with

$\theta^\times = (\theta^\times_j)_{j\geqslant 1} \equiv 0 \quad$ and $\quad \varsigma = (\varsigma_j)_{j\geqslant 1} \equiv \infty.$

The Bayes estimator is $\widehat{\theta}^m = \mathbb{E}[\vartheta^m \,|\, Y] = (Y/\lambda)^m$.

The posterior probability of M is :

$$P_{M\,|\,Y}(M = m) \propto \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\}).$$

Hence, $\widehat{\theta} = (\widehat{\theta}_j)_{j\geqslant 1} = \mathbb{E}[\vartheta^M \,|\, Y]$ equals the shrunk orthogonal projection estimator given by

$$\widehat{\theta}_j = \frac{\sum_{m=j}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})}{\sum_{m=1}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda} + 3C_\lambda m\})} \times \frac{Y_j}{\lambda_j}\mathbb{1}_{\{1\leqslant j\leqslant G_\varepsilon\}}.$$

Data-driven Bayesian estimation (VI)

- Denote $\Upsilon(\widehat{\theta}^m) := -(1/2)\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda}$ (contrast) and $\mathrm{pen}_m := 3/2 C_\lambda m$ (penalty term).
- Hence, the j-th shrinkage weight is proportional to $\sum_{m=j}^{G_\varepsilon} \exp(-\{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\})$.
- In comparison to a classical model selection approach where a data-driven estimator $\widehat{\theta}^{\widehat{m}} = (Y/\lambda)^{\widehat{m}}$ is obtained by selecting the dimension parameter $\widehat{m}$ as $\widehat{m} = \arg\min_{1 \leqslant m \leqslant G_\varepsilon} \{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\}$,
- following the Bayesian approach each of the G_ε components of the data-driven Bayes estimator is shrunk proportional to the associated values of the penalized contrast criterion.

Data-driven Bayesian estimation (VI)

- Denote $\Upsilon(\widehat{\theta}^m) := -(1/2)\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda}$ (contrast) and $\mathrm{pen}_m := 3/2 C_\lambda m$ (penalty term).
- Hence, the j-th shrinkage weight is proportional to $\sum_{m=j}^{G_\varepsilon} \exp(-\{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\})$.
- In comparison to a classical model selection approach where a data-driven estimator $\widehat{\theta}^{\widehat{m}} = (Y/\lambda)^{\widehat{m}}$ is obtained by selecting the dimension parameter $\widehat{m}$ as $\widehat{m} = \arg\min_{1 \leqslant m \leqslant G_\varepsilon} \{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\}$,
- following the Bayesian approach each of the G_ε components of the data-driven Bayes estimator is shrunk proportional to the associated values of the penalized contrast criterion.

Data-driven Bayesian estimation (VI)

- Denote $\Upsilon(\widehat{\theta}^m) := -(1/2)\|(Y/\lambda)^m\|^2_{\varepsilon\Lambda}$ (contrast) and $\mathrm{pen}_m := 3/2 C_\lambda m$ (penalty term).
- Hence, the j-th shrinkage weight is proportional to $\sum_{m=j}^{G_\varepsilon} \exp(-\{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\})$.
- In comparison to a classical model selection approach where a data-driven estimator $\widehat{\theta}^{\widehat{m}} = (Y/\lambda)^{\widehat{m}}$ is obtained by selecting the dimension parameter $\widehat{m}$ as $\widehat{m} = \arg\min_{1 \leqslant m \leqslant G_\varepsilon} \{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\}$,
- following the Bayesian approach each of the G_ε components of the data-driven Bayes estimator is shrunk proportional to the associated values of the penalized contrast criterion.

Adaptive Bayesian estimation in indirect Gaussian sequence space models

Jan Johannes   Anna Simoni   Rudolf Schenk

Workshop on Inverse Problems, October 28, 2016

Sketch of the proof

This can be proved since $\{\vartheta_j^m - \theta_j^\circ\}_{j=1}^m | Y \sim ind\ \mathcal{N}(\theta_j^Y - \theta_j^\circ, \sigma_j)$ by using:

Lemma 4.6 (Birgé 2001 and Laurent & Massart 2000)

Let $\{X_j\}_{j\geqslant 1}$ ind $\mathcal{N}(\alpha_j, \beta_j^2)$, $\alpha_j \in \mathbb{R}$ and standard deviation $\beta_j \geqslant 0$, $j \in \mathbb{N}$. For $m \in \mathbb{N}$ set $S_m := \sum_{j=1}^m X_j^2$ and consider $v_m \geqslant \sum_{j=1}^m \beta_j^2$, $t_m \geqslant \max_{1\leqslant j\leqslant m} \beta_j^2$ and $r_m \geqslant \sum_{j=1}^m \alpha_j^2$. Then for all $c \geqslant 0$:

$$\sup_{m\geqslant 1} \exp\Big(\frac{c(c \wedge 1)(v_m + 2r_m)}{4t_m}\Big) P\big(S_m - \mathbb{E}S_m \leqslant -c(v_m + 2r_m)\big) \leqslant 1; \tag{10}$$

$$\sup_{m\geqslant 1} \exp\Big(\frac{c(c \wedge 1)(v_m + 2r_m)}{4t_m}\Big) P\big(S_m - \mathbb{E}S_m \geqslant \frac{3c}{2}(v_m + 2r_m)\big) \leqslant 1. \tag{11}$$

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|^2 = \mathbb{E}_{\theta^\circ}\sum_{j=1}^{m_\varepsilon}(\theta_j^{\curlyvee} - \theta_j^\circ)^2 + \sum_{j>m_\varepsilon}(\theta_j^{\times} - \theta_j^\circ)^2$$
$$= \sum_{j=1}^{m_\varepsilon}\sigma_j(\sigma_j\lambda_j^2\varepsilon^{-1}) + \mathfrak{r}_{m_\varepsilon} + \mathfrak{b}_{m_\varepsilon}, \quad (12)$$

which together with $\sigma_j\lambda_j^2\varepsilon^{-1} \leqslant 1$ implies

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|^2 \leqslant \mathfrak{b}_{m_\varepsilon} + m_\varepsilon\overline{\sigma}_{m_\varepsilon} + \mathfrak{r}_{m_\varepsilon}.$$

By Assumption A.2: $\mathfrak{r}_{m_\varepsilon} \leqslant K[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]$.

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|^2 = \mathbb{E}_{\theta^\circ}\sum_{j=1}^{m_\varepsilon}(\theta_j^{Y} - \theta_j^\circ)^2 + \sum_{j>m_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2$$
$$= \sum_{j=1}^{m_\varepsilon}\sigma_j(\sigma_j\lambda_j^2\varepsilon^{-1}) + \mathfrak{r}_{m_\varepsilon} + \mathfrak{b}_{m_\varepsilon}, \quad (12)$$

which together with $\sigma_j\lambda_j^2\varepsilon^{-1} \leqslant 1$ implies

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|^2 \leqslant \mathfrak{b}_{m_\varepsilon} + m_\varepsilon\overline{\sigma}_{m_\varepsilon} + \mathfrak{r}_{m_\varepsilon}.$$

By Assumption A.2: $\mathfrak{r}_{m_\varepsilon} \leqslant K[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]$.

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon}-\theta^\circ\|^2 = \mathbb{E}_{\theta^\circ}\sum_{j=1}^{m_\varepsilon}(\theta_j^Y-\theta_j^\circ)^2 + \sum_{j>m_\varepsilon}(\theta_j^\times-\theta_j^\circ)^2$$

$$= \sum_{j=1}^{m_\varepsilon}\sigma_j(\sigma_j\lambda_j^2\varepsilon^{-1}) + \mathfrak{r}_{m_\varepsilon} + \mathfrak{b}_{m_\varepsilon}, \quad (12)$$

which together with $\sigma_j\lambda_j^2\varepsilon^{-1}\leqslant 1$ implies

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon}-\theta^\circ\|^2 \leqslant \mathfrak{b}_{m_\varepsilon} + m_\varepsilon\overline{\sigma}_{m_\varepsilon} + \mathfrak{r}_{m_\varepsilon}.$$

By Assumption A.2: $\mathfrak{r}_{m_\varepsilon}\leqslant K[\mathfrak{b}_{m_\varepsilon}\vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]$.

Lemma 4.7

If Assumptions A.1 holds true then for all $\varepsilon \in (0, \varepsilon_\circ)$

(i) $\mathbb{E}_{\theta^\circ} P_{\mathsf{M}\,|\,\mathsf{Y}}(1 \leqslant \mathsf{M} < G_\varepsilon^-) \leqslant 2\exp\big(-\frac{C_\lambda}{5} m_\varepsilon^\circ + \log G_\varepsilon\big)$;

(ii) $\mathbb{E}_{\theta^\circ} P_{\mathsf{M}\,|\,\mathsf{Y}}(G_\varepsilon^+ < \mathsf{M} \leqslant G_\varepsilon) \leqslant 2\exp\big(-\frac{C_\lambda}{5} m_\varepsilon^\circ + \log G_\varepsilon\big)$.

Observe that $\mathfrak{b}_{m_\varepsilon^\circ} \geqslant \kappa^\circ \Phi_\varepsilon^\circ > 0$ due to Assumption A.4 which implies $\mathfrak{b}_k > 0\ \forall k \in \mathbb{N}$ and, hence $m_\varepsilon^\circ \to \infty$ as $\varepsilon \to 0$.

Sketch of the proof of Theorem 4.2 (I)

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}} \mid \mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ) = \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M} \mid \mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m \mid \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$= \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon^- - 1} p_{\mathrm{M} \mid \mathrm{Y}}(m) P_{\vartheta^m \mid \mathrm{Y}}(\|\vartheta^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$+ \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} p_{\mathrm{M} \mid \mathrm{Y}}(m) P_{\vartheta^m \mid \mathrm{Y}}(\|\vartheta^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$+ \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^+ + 1}^{G_\varepsilon} p_{\mathrm{M} \mid \mathrm{Y}}(m) P_{\vartheta^m \mid \mathrm{Y}}(\|\vartheta^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$\leqslant \mathbb{E}_{\theta^\circ} P_{\mathrm{M} \mid \mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^-) + \mathbb{E}_{\theta^\circ} P_{\mathrm{M} \mid \mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon)$$

$$+ \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\vartheta^m \mid \mathrm{Y}}(\|\vartheta^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$\leqslant 4 \exp\big(-m_\varepsilon^\circ \{C_\lambda/5 - \log G_\varepsilon/m_\varepsilon^\circ\}\big) + 74 \exp(-G_\varepsilon^-/36) \qquad (13)$$

Sketch of the proof of Theorem 4.2 (I)

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}} \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ) = \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M} \,|\, \mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$= \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon^- - 1} p_{\mathrm{M} \,|\, \mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$+ \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} p_{\mathrm{M} \,|\, \mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$+ \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^+ + 1}^{G_\varepsilon} p_{\mathrm{M} \,|\, \mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$\leqslant \mathbb{E}_{\theta^\circ} P_{\mathrm{M} \,|\, \mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^-) + \mathbb{E}_{\theta^\circ} P_{\mathrm{M} \,|\, \mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon)$$

$$+ \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$\leqslant 4 \exp\big(-m_\varepsilon^\circ \{C_\lambda/5 - \log G_\varepsilon/m_\varepsilon^\circ\}\big) + 74 \exp(-G_\varepsilon^-/36) \qquad (13)$$

Sketch of the proof of Theorem 4.2 (I)

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ) = \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$= \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon^- - 1} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$+ \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$+ \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^+ + 1}^{G_\varepsilon} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$\leqslant \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^-) + \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon)$$

$$+ \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi_\varepsilon^\circ)$$

$$\leqslant 4 \exp\left(- m_\varepsilon^\circ \{ C_\lambda / 5 - \log G_\varepsilon / m_\varepsilon^\circ \} \right) + 74 \exp(-G_\varepsilon^- / 36) \quad (13)$$

Sketch of the proof of Theorem 4.2 (I)

$$
\begin{aligned}
\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2 > K^\circ \Phi^\circ_\varepsilon) &= \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi^\circ_\varepsilon) \\
&= \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon^- -1} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi^\circ_\varepsilon) \\
&\quad + \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi^\circ_\varepsilon) \\
&\quad + \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^+ +1}^{G_\varepsilon} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi^\circ_\varepsilon) \\
&\leqslant \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^-) + \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon) \\
&\qquad + \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 > K^\circ \Phi^\circ_\varepsilon) \\
&\leqslant 4 \exp\big(- m^\circ_\varepsilon \{ C_\lambda/5 - \log G_\varepsilon / m^\circ_\varepsilon \} \big) + 74 \exp(-G_\varepsilon^-/36) \qquad (13)
\end{aligned}
$$

Sketch of the proof of Theorem 4.2 (II)

On the other side

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathsf{M}}\,|\,\mathsf{Y}}(\|\boldsymbol{\vartheta}^{\mathsf{M}} - \theta^\circ\|^2 < (K^\circ)^{-1}\Phi_\varepsilon^\circ) \leqslant \mathbb{E}_{\theta^\circ} P_{\mathsf{M}\,|\,\mathsf{Y}}(1 \leqslant \mathsf{M} < G_\varepsilon^-)$$

$$+\mathbb{E}_{\theta^\circ} P_{\mathsf{M}\,|\,\mathsf{Y}}(G_\varepsilon^+ < \mathsf{M} \leqslant G_\varepsilon) + \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} p_{\mathsf{M}\,|\,\mathsf{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathsf{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2 < (K^\circ)^{-1}\Phi_\varepsilon^\circ)$$

$$\leqslant 4\exp\big(-m_\varepsilon^\circ\{C_\lambda/5 - \log G_\varepsilon/m_\varepsilon^\circ\}\big) + 4(K^\circ)^2\exp(-G_\varepsilon^-/(K^\circ)^2) \qquad (14)$$

By combining (13) and (14) we obtain the assertion of the theorem since $G_\varepsilon^-, m_\varepsilon^\circ \to \infty$ and $\log G_\varepsilon/m_\varepsilon^\circ = o(1)$ as $\varepsilon \to 0$ which completes the proof.