Statistical Modeling of Complex Systems and Processes

Frequentist analysis of Bayesian methods for statistical ill-posed inverse problems

Xavier Loizeau

Supervisors: Prof. Dr. Jan Johannes, Prof. Dr. Claudia Schillings

The inverse Gaussian sequence space model

Consider an indirect Gaussian sequence space model consisting of:

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

The goal is to recover θ° and derive an upper bound.

Bayesian paradigm

We adopt a Bayesian point of view:

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

Within this framework we define the estimator: $\widehat{\theta} := \mathbb{E}^n_{\boldsymbol{\theta}|Y}[\boldsymbol{\theta}]$.

We are interested in the behavior of $\mathbb{P}^n_{\boldsymbol{\theta}|Y}$ as n tends to infinite.

In particular, the question of oracle and minimax concentration (resp. convergence) is answered for the posterior distribution (resp. posterior mean).

Hierarchical prior

- Consider a random hyper-parameter M, with values in a subset of $\mathbb{N}$, acting like a threshold:

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

- If we denote $\mathbb{P}_M$ the distribution of M (to be specified later), then

$$\mathbb{P}^n_{\boldsymbol{\theta}|Y} = \sum_{m\in\mathbb{N}} \mathbb{P}^n_{\boldsymbol{\theta}|M=m,Y} \cdot \mathbb{P}^n_{M=m|Y}.$$

- Hence, given M, the posterior is

$$\forall j > m, \quad \boldsymbol{\theta}_j | M = m, Y \sim \delta_0,$$
$$\forall j \leq m, \quad \boldsymbol{\theta}_j | M = m, Y \sim \mathcal{N}\left(\frac{Y_j \cdot n \cdot \lambda_j}{1 + n \cdot \lambda_j^2}, \frac{1}{1 + n \cdot \lambda_j^2}\right).$$

Remark: the family of hierarchical priors with deterministic threshold M is called family of sieve priors.

Review : optimal posterior concentration

In [4], under a pragmatic Bayesian point of view; that is, the existence of a true parameter θ° is accepted; it is shown that, by choosing $\mathbb{P}_M$ suitably:

- the Bayes estimator $\widehat{\theta}$ converges with,
  - oracle optimal rate for the quadratic risk which means, $\forall \theta^\circ \in \Theta^\circ, \exists C^\circ \in [1,\infty[: \forall n \in \mathbb{N}, \exists \Phi_n^\circ \in \mathbb{R}$:

$$\inf_{m\in\mathbb{N}} \mathbb{E}^n_{\theta^\circ}\left[\left\|\tilde{\theta}^m - \theta^\circ\right\|^2\right] \geq \Phi_n^\circ(\theta^\circ), \qquad \mathbb{E}^n_{\theta^\circ}\left[\left\|\widehat{\theta} - \theta^\circ\right\|^2\right] \leq C^\circ\Phi_n^\circ(\theta^\circ);$$

  - minimax optimal rate for the maximal risk over some ellipsoid Θ_a°, that is to say, $\exists C^\star \in [1,\infty[: \forall n \in \mathbb{N}, \exists \Phi_n^\star(a) \in \mathbb{R}$:

$$\inf_{\tilde{\theta}} \sup_{\theta^\circ\in\Theta_a^\circ} \mathbb{E}^n_{\theta^\circ}\left[\left\|\tilde{\theta} - \theta^\circ\right\|^2\right] \geq \Phi_n^\star(a), \qquad \sup_{\theta^\circ\in\Theta_a^\circ} \mathbb{E}^n_{\theta^\circ}\left[\left\|\widehat{\theta} - \theta^\circ\right\|^2\right] \leq C^\star\Phi_n^\star(a),$$

  where $\inf_{\tilde{\theta}}$ is taken over all possible estimators of θ°;
- the posterior distribution concentrates with,
  - oracle optimal rate for the quadratic loss which means, $\forall \theta^\circ \in \Theta^\circ, \exists K^\circ \in [1,\infty[$:

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

  - minimax optimal rate over Θ_a°, that is to say, for any unbounded sequence $K_n \in \mathbb{R}^{\mathbb{N}}$:

$$\lim_{n\to\infty} \sup_{\theta^\circ\in\Theta_a^\circ} \mathbb{E}^n_{\theta^\circ}\left[\mathbb{P}^n_{\boldsymbol{\theta}|Y}\left(\|\boldsymbol{\theta} - \theta^\circ\|^2 \leq K_n\Phi_n^\star(a)\right)\right] = 1.$$

Bayesian formulation of optimality

Theorem 1. For all θ° in Θ°,

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

where $\inf_{\mathbb{Q}_{\boldsymbol{\theta}}}$ is taken over all sieve priors; establishing a Bayesian formulation of oracle optimality.

Theorem 2. We also have, for some Sobolev's ellipsoids Θ_a°,

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

where $\inf_{\mathbb{Q}_{\boldsymbol{\theta}}}$ is taken over all possible sieve priors.

Iterated posterior

In the spirit of [1], we then generate a posterior family by introducing an iteration parameter η:

- for $\eta = 1$, the prior distribution is $\mathbb{P}_{\boldsymbol{\theta}^1} = \mathbb{P}_{\boldsymbol{\theta}}$, the likelihood $\mathbb{P}^n_{Y^1|\boldsymbol{\theta}^1} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and the posterior distribution is $\mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1} = \mathbb{P}^n_{\boldsymbol{\theta}|Y}$,
- for $\eta = 2$, we take the posterior for $\eta = 1$ as prior, hence, the prior distribution is $\mathbb{P}^n_{\boldsymbol{\theta}^2} = \mathbb{P}^n_{\boldsymbol{\theta}^1|Y^1}$, the likelihood is kept the same $\mathbb{P}^n_{Y^2|\boldsymbol{\theta}^2} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and we compute the posterior distribution with the same observations Y, which we note $\mathbb{P}^n_{\boldsymbol{\theta}^2|Y^2}$,
- ...
- for any value of $\eta > 1$, the prior is $\mathbb{P}^n_{\boldsymbol{\theta}^\eta} = \mathbb{P}^n_{\boldsymbol{\theta}^{\eta-1}|Y^{\eta-1}}$ and we compute the posterior with the same likelihood $\mathbb{P}_{Y^\eta|\boldsymbol{\theta}^\eta} = \mathbb{P}^n_{Y|\boldsymbol{\theta}}$ and same observation Y which gives $\mathbb{P}^n_{\boldsymbol{\theta}^\eta|Y^\eta}$.

This iteration procedure corresponds to giving more and more weight to the observations.

Theorem 3. We have shown that any element of the family defined this way also has the properties given previously, including in the limit case, called self informative Bayes carrier and its posterior mean the self informative limit.

Additional results

- The self informative Bayes carrier is a point mass on the frequentist model selection estimate as in [5].
- In circular deconvolution model in presence of beta mixing data, methodology leads to oracle and minimax optimal fully data driven estimator.
- We are currently working on an extension of this method to the real line deconvolution model.

References

[1] O. Bunke and J. Johannes. Selfinformative limits of Bayes estimates and generalized maximum likelihood. *Statistics*, 39(6):483–502, 2005.

[2] J. Johannes, F. Comtes, and X. Loizeau. Data-driven shrinked deconvolution estimator (in prep.). 2017.

[3] J. Johannes and X. Loizeau. Adaptive Bayesian estimation and self informative limit in indirect Gaussian sequence space models (in preparation). 2017.

[4] J. Johannes, A. Simoni, and R. Schenk. Adaptive Bayesian estimation in indirect Gaussian sequence space model. 2016.

[5] P. Massart. Concentration inequalities and model selection. In Springer, editor, *Ecole d'Été de Probabilités de Saint-Flour XXXIII*, volume 1869, 2003.

Personal information

- 2012-2013: BSc in Mathematics, University of Rennes 1, France
- 2012-2015: Master of Science (MSc) in Statistics (Diplôme d'ingénieur) ENSAI - National School for Statistics and Information Analysis, Rennes, France
- 2013-2015: MSc in Mathematical Statistics, University of Rennes 1
- 2015-2018: Research assistant, Heidelberg University
- 2015-2018: PhD student in the RTG 1953, Heidelberg University

Supported by DFG

RTG 1953