Minimax Goodness-of-Fit Testing in Ill-Posed Inverse Problems with Partially Unknown Operators

C. Marteau[1] and T. Sapatinas[2]

Mathematical Statistics and Inverse Problems

1. Université Lyon I - Institut Camille Jordan
2. University of Cyprus

Outline

Outline

Inverse problem model

Let H, K be Hilbert spaces.

Goal : inference on $f \in H$ from noisy and indirect observations

$$Y = Af + \epsilon\xi,$$

where $A : H \to K$ denotes a (compact) operator, ξ a Gaussian white noise and $\epsilon > 0$.

Inverse problem model

Let H, K be Hilbert spaces.

Goal : inference on $f \in H$ from noisy and indirect observations

$$Y = Af + \epsilon\xi,$$

where $A : H \to K$ denotes a (compact) operator, ξ a Gaussian white noise and $\epsilon > 0$.

For all $g \in K$, we can observe

$$\langle Y, g\rangle = \langle Af, g\rangle + \epsilon\langle \xi, g\rangle,$$

where $\langle \xi, g\rangle \sim \mathcal{N}(0, \|g\|^2)$.

In the case where the operator A is compact, the problem is **ill-posed**.

Non-parametric goodness-of-fit tests

Given a (known) benchmark function f_0, the goal is to test

$$H_0 : f = f_0,$$

against

$$H_1 : f \in \mathcal{F},$$

where $\mathcal{F}$ denotes a function set (made precise next slide).

Non-parametric goodness-of-fit tests

Given a (known) benchmark function f_0, the goal is to test

$$H_0 : f = f_0,$$

against

$$H_1 : f \in \mathcal{F},$$

where $\mathcal{F}$ denotes a function set (made precise next slide).

Several contributions in this setting and related domains :

- Ermakov (2006) - Ingster, Sapatinas and Suslina (2012) - Laurent, Loubes and M. (2012)
- Butucea (2007)
- Bissantz, Claeskens, Holzmann and Munk (2009) - Ingster, Sapatinas and Suslina (2011)

Non-parametric goodness-of-fit tests

In our context, the alternative can be considered as follows

$$H_1 : f - f_0 \in \mathcal{E}_a, \ \|f - f_0\|^2 \geq r_\epsilon^2,$$

where

- $f - f_0 \in \mathcal{E}_a$ denotes some constraints on the smoothness of $f - f_0$,
- $\|f - f_0\|^2 \geq r_\epsilon^2$ denotes some 'energy' condition between f and f_0. The term r_ϵ is called separation radius.

For a given constraint $f \in \mathcal{E}_a$, the task is to determine the minimal (achievable) radius r_ϵ for which H_0 and H_1 can be separated with prescribed errors (Type I and Type II).

Non-parametric goodness-of-fit tests

Question : What happens when we have some uncertainty on the operator A at hand ?

Several contributions in an estimation context (quantitative) : Cavalier and Hengartner (2005) - Hoffmann and Reiss (2008) - Johannes and Schwarz (2013) - Delattre, Hoffmann, Picard and Vareschi (2012), ...

In this talk, we propose an attempt in a qualitative context (investigation on the separation rates). This may provide outcome for related models (IV regression, density model,...).

Singular value decomposition

Call $(b_k^2)_{k\geq 1}$ the eigenvalues of $A^\star A$ and $(\phi_k)_{k\geq 1}$ the associated eigenvectors. Let $(\psi_k)_{k\geq 1}$ the basis verifying, for all $k \in \mathbb{N}$:

$$\begin{cases} A\phi_k = b_k\psi_k, \\ A^\star\psi_k = b_k\phi_k. \end{cases}$$

Singular value decomposition

Call $(b_k^2)_{k\geq 1}$ the eigenvalues of $A^\star A$ and $(\phi_k)_{k\geq 1}$ the associated eigenvectors. Let $(\psi_k)_{k\geq 1}$ the basis verifying, for all $k \in \mathbb{N}$:

$$\begin{cases} A\phi_k = b_k \psi_k, \\ A^\star \psi_k = b_k \phi_k. \end{cases}$$

This leads to the sequence space model :

$$\langle Y, \psi_k \rangle = Y_k = b_k \theta_k + \epsilon \xi_k, \ \ k \in \mathbb{N},$$

with $\theta_k = \langle f, \phi_k \rangle$. The ξ_k are i.i.d. standard Gaussian random variables.

Singular value decomposition

Call $(b_k^2)_{k\geq 1}$ the eigenvalues of $A^\star A$ and $(\phi_k)_{k\geq 1}$ the associated eigenvectors. Let $(\psi_k)_{k\geq 1}$ the basis verifying, for all $k \in \mathbb{N}$:

$$\begin{cases} A\phi_k = b_k \psi_k, \\ A^\star \psi_k = b_k \phi_k. \end{cases}$$

This leads to the sequence space model :

$$\langle Y, \psi_k \rangle = Y_k = b_k \theta_k + \epsilon \xi_k, \; k \in \mathbb{N},$$

with $\theta_k = \langle f, \phi_k \rangle$. The ξ_k are i.i.d. standard Gaussian random variables.

When the operator A is compact, $b_k \to 0$ as $k \to +\infty$. Classical settings

$$b_k \sim k^{-\beta} \quad \text{or} \quad b_k \sim e^{-\beta k} \; \forall k \in \mathbb{N}.$$

The model

The following observations are available

$$\begin{cases} Y_j = b_j\theta_j + \varepsilon\,\xi_j, & j \in \mathbb{N}, \\ X_j = b_j + \sigma\,\eta_j, & j \in \mathbb{N}, \end{cases}$$

where ϵ, σ are known noise levels, ξ_j and η_j denotes i.i.d. standard Gaussian random variables (independent of each other).

Given a fixed $\theta_0 \neq 0$ and $\theta_0 \in \mathcal{E}_a$, we want to test

$$H_0 : \theta = \theta_0,$$

against

$$H_1 : \theta - \theta_0 \in \mathcal{E}_a,\ \|\theta - \theta_0\|^2 \geq r_{\epsilon,\sigma}^2.$$

The model

We want to test

$$H_0 : \theta = \theta_0,$$

against

$$H_1 : \theta - \theta_0 \in \mathcal{E}_a, \ \|\theta - \theta_0\|^2 \geq r_{\epsilon,\sigma}^2,$$

where

$$\mathcal{E}_a = \left\{ \nu \in l^2(\mathbb{N}), \ \sum_{j \in \mathbb{N}} a_j^2 \nu_j^2 \leq 1 \right\}.$$

In the sequel, define

$$\Theta_a(r_{\epsilon,\sigma}) = \left\{ \nu = (\nu_k)_k \ s.t. \ \sum_{k \in \mathbb{N}^\star} a_k^2 \nu_k^2 \leq 1 \quad \text{and} \quad \|\nu\|^2 \geq r_{\epsilon,\sigma}^2 \right\},$$

Notations

Let $\Psi_\alpha = \Psi_\alpha(X, Y) \in \{0, 1\}$ a given level-α test.

Notations

Let $\Psi_\alpha = \Psi_\alpha(X, Y) \in \{0, 1\}$ a given level-α test. Define

- Maximal Type II error probability

$$\boldsymbol{\beta}_{\varepsilon,\sigma}(\Theta_a(r_{\varepsilon,\sigma}), \Psi_\alpha) := \sup_{\theta:\, \theta - \theta_0 \in \Theta_a(r_{\varepsilon,\sigma})} \mathbb{P}_{\theta,b}(\Psi_\alpha = 0).$$

Notations

Let $\Psi_\alpha = \Psi_\alpha(X, Y) \in \{0, 1\}$ a given level-α test. Define

- Maximal Type II error probability

$$\boldsymbol{\beta}_{\varepsilon,\sigma}(\Theta_a(r_{\varepsilon,\sigma}), \Psi_\alpha) := \sup_{\theta:\, \theta - \theta_0 \in \Theta_a(r_{\varepsilon,\sigma})} \mathbb{P}_{\theta,b}(\Psi_\alpha = 0).$$

- Separation radius of Ψ_α

$$r_{\varepsilon,\sigma}(\mathcal{E}_a, \Psi_\alpha, \beta) := \inf\left\{ r_{\varepsilon,\sigma} > 0 : \ \boldsymbol{\beta}_{\varepsilon,\sigma}(\Theta_a(r_{\varepsilon,\sigma}), \Psi_\alpha) \le \beta \right\},$$

Notations

Let $\Psi_\alpha = \Psi_\alpha(X, Y) \in \{0, 1\}$ a given level-α test. Define

- Maximal Type II error probability

$$\boldsymbol{\beta}_{\varepsilon,\sigma}(\Theta_a(r_{\varepsilon,\sigma}), \Psi_\alpha) := \sup_{\theta:\, \theta-\theta_0 \in \Theta_a(r_{\varepsilon,\sigma})} \mathbb{P}_{\theta,b}(\Psi_\alpha = 0).$$

- Separation radius of Ψ_α

$$r_{\varepsilon,\sigma}(\mathcal{E}_a, \Psi_\alpha, \beta) := \inf\left\{ r_{\varepsilon,\sigma} > 0 : \; \boldsymbol{\beta}_{\varepsilon,\sigma}(\Theta_a(r_{\varepsilon,\sigma}), \Psi_\alpha) \leq \beta \right\},$$

- Minimax separation radius over $\mathcal{E}_a$

$$\tilde{r}_{\varepsilon,\sigma} := \inf_{\tilde{\Psi}_\alpha:\, \boldsymbol{\alpha}_{\varepsilon,\sigma}(\tilde{\Psi}_\alpha) \leq \alpha} r_{\varepsilon,\sigma}(\mathcal{E}_a, \tilde{\Psi}_\alpha, \beta).$$

Program

All along this talk, we will use the following guideline

- Propose a testing procedure.
- Control its Type I error.
- Compute the associated separation radius (non-asymptotic point of view).
- Check the optimality of the procedure (lower bounds).

To conclude the talk, we present asymptotic separation rates in this setting, for both mildly and severely ill-posed problems, and ordinary and super smooth functions.

Outline

A testing procedure

We use the following testing procedure

$$\Psi_{D,M} := \mathbf{1}_{\{T_{D,M} > t_{1-\alpha,D}\}},$$

where

$$T_{D,M} := \sum_{j=1}^{D} \left(\frac{Y_j}{b_j} - \theta_{j,0} \right)^2,$$

$$t_{1-\alpha,D} = \varepsilon^2 \sum_{j=1}^{D} b_j^{-2} + C_\alpha \varepsilon^2 \sqrt{\sum_{j=1}^{D} b_j^{-4}}.$$

A testing procedure

We use the following testing procedure

$$\Psi_{D,M} := \mathbf{1}_{\{T_{D,M} > t_{1-\alpha,D}(X)\}},$$

where

$$T_{D,M} := \sum_{j=1}^{D\wedge M} \left(\frac{Y_j}{X_j} - \theta_{j,0} \right)^2,$$

M denotes a random bandwidth (made precise later) and

$$t_{1-\alpha,D}(X) = \varepsilon^2 \sum_{j=1}^{D\wedge M} X_j^{-2} + C_\alpha \varepsilon^2 \sqrt{\sum_{j=1}^{D\wedge M} X_j^{-4}} + C_\alpha \left[\sigma^2 \ln^{3/2}(1/\sigma) \vee a_{D\wedge M}^{-2} \right]$$

Remarks I

Some comments

- Many alternative ways to perform a test, e.g. based on the direct data (Y_k is compared to $X_k\theta_{k,0}$) : see for instance Laurent, Loubes, M. (2011).
- The constant C_α involved in the procedure is explicit.
- The bandwidth M is defined as

$$M := \inf\{j \in \mathbb{N} : |X_j| \leq \sigma h_j\} - 1,$$

where $h_j \sim \sqrt{\ln(j/\alpha)}$. For all $j \leq M$, X_j is a 'good' approximation of b_j with high probability.

Remarks II (Heuristic)

Under H_0,

$$\begin{aligned} T_{D,M} &= \sum_{j=1}^{D\wedge M} \left(\frac{Y_j}{X_j} - \theta_{j,0} \right)^2, \\ &= \sum_{j=1}^{D\wedge M} \left[\left(\frac{b_j}{X_j} - 1 \right) \theta_{j,0} + \varepsilon X_j^{-1} \xi_j \right]^2, \\ &\sim \sigma^2 \sum_{j=1}^{D\wedge M} b_j^{-2} \theta_{j,0}^2 \eta_j^2 + \epsilon^2 \sum_{j=1}^{D\wedge M} X_j^{-2} \xi_j^2, \end{aligned}$$

Remarks II (Heuristic)

Under H_0,

$$T_{D,M} \sim \sigma^2 \sum_{j=1}^{D \wedge M} b_j^{-2} \theta_{j,0}^2 \eta_j^2 + \epsilon^2 \sum_{j=1}^{D \wedge M} X_j^{-2} \xi_j^2,$$

Two different situations may occur

- If $\sup_j b_j^{-2} a_j^{-2}$ is bounded, we obtain a bound of order σ^2 for the first term (up to a log term).
- In the other case, we can prove that $\sup_{j \leq M} \sigma^2 b_j^{-2} \eta_j^2$ is bounded with high probability, which leads to the additional term $a_{D \wedge M}^{-2}$ in the threshold.

This explain in some sense the presence of the term $\left[\sigma^2 \ln^{3/2}(1/\sigma) \vee a_{D \wedge M}^{-2}\right]$.

A non-asymptotic upper bound

Let

$$D^\dagger \sim \arg\min_{D\in\mathbb{N}} \left\{ \varepsilon^2 \sqrt{\sum_{j=1}^{D\wedge M} X_j^{-4}} + \left[\sigma^2 \ln^{3/2}(1/\sigma) \vee a_{D\wedge M}^{-2}\right] \right\}.$$

Proposition

There exists $\sigma_0 \in]0,1[$ such that, for all $0 < \sigma \leq \sigma_0$ and for each $\varepsilon > 0$,

$$\begin{aligned} \tilde{r}^2_{\epsilon,\sigma} &\leq r^2_{\varepsilon,\sigma}(\mathcal{E}_a, \Psi_{D,M}, \beta) \\ &\leq C_{\alpha,\beta} \inf_{D\in\mathbb{N}} \left[\epsilon^2 \sqrt{\sum_{j=1}^{D\wedge M_1} b_j^{-4}} + \left[\sigma^2 \ln^{3/2}(1/\sigma) \vee a_{D\wedge M_0}^{-2}\right] \right], \end{aligned}$$

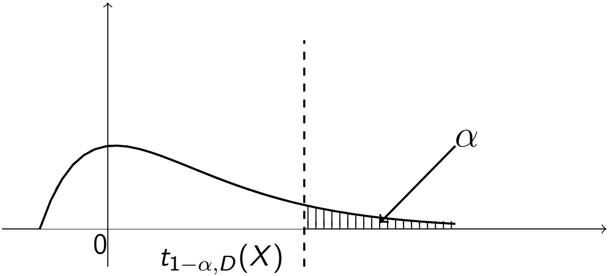
This bound is non-asymptotic. No condition required on both sequences $(b_j)_j$ and $(a_j)_j$.

Sketch of the proof

Remark that, for any fixed D, conditionally to X

$$\begin{aligned}\mathbb{P}_\theta(\Psi_{D,M} = 0/X) &= \mathbb{P}(T_{D,M} \leq t_{1-\alpha,D}(X)/X), \\ &\leq \frac{\beta}{2} + \mathbf{1}_{t_{1-\alpha,D}(X) > t_{\beta/2,D}(\theta,X)},\end{aligned}$$

where $t_{\beta/2,D}(\theta, X)$ denotes the $\beta/2$ quantile of $T_{D,M}$ when $\theta \neq \theta_0$.

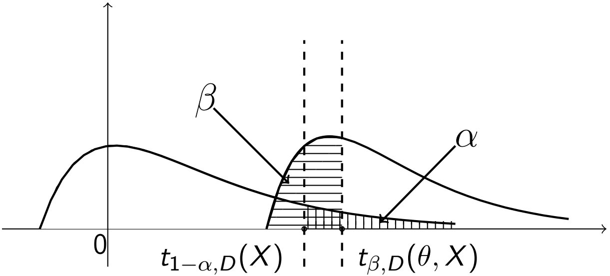


Sketch of the proof

Remark that, for any fixed D, conditionally to X

$$\begin{aligned} \mathbb{P}_\theta(\Psi_{D,M} = 0/X) &= \mathbb{P}(T_{D,M} \leq t_{1-\alpha,D}(X)/X), \\ &\leq \frac{\beta}{2} + \mathbf{1}_{t_{1-\alpha,D}(X) > t_{\beta/2,D}(\theta,X)}, \end{aligned}$$

where $t_{\beta/2,D}(\theta, X)$ denotes the $\beta/2$ quantile of $T_{D,M}$ when $\theta \neq \theta_0$.



Sketch of the proof

In order to conclude the proof, we have to check that

$$t_{1-\alpha,D}(X) > t_{\beta/2,D}(\theta, X),$$

occurs with probability bounded by $\beta/2$. This occurs as soon as

$$\|\theta - \theta_0\|^2 \gtrsim \epsilon^2 \sqrt{\sum_{j=1}^{D\wedge M_1} b_j^{-4}} + \left[\sigma^2 \ln^{3/2}(1/\sigma) \vee a_{D\wedge M_0}^{-2}\right].$$

Then, prove that similar properties hold as soon as D is allowed to depend on X.

Outline

A non-asymptotic lower bound

Proposition

For every $\varepsilon > 0$ and $\sigma > 0$,

$$\tilde{r}_{\varepsilon,\sigma} \geq \max(\tilde{r}_{0,\sigma}, \tilde{r}_{\varepsilon,0}). \qquad (1)$$

In particular, there exists $C_{\alpha,\beta}$ and a bandwidth M_2 (made precise latter on) such that

$$\tilde{r}^2_{\varepsilon,\sigma} \geq C_{\alpha,\beta} \left\{ \sigma^2 \max_{1 \leq D \leq M_2} [b_D^{-2} a_D^{-2}] \right\} \vee \left\{ \sup_{D \in \mathbb{N}} \left[\varepsilon^2 \sqrt{\sum_{j=1}^{D} b_j^{-4}} \wedge a_D^{-2} \right] \right\}.$$

Similar arguments are proposed in, e.g., Delattre, Hoffmann, Picard and Vareschi (2012).

A non-asymptotic lower bound

$$\begin{cases} Y_j = b_j\theta_j + \varepsilon\, \xi_j, & j \in \mathbb{N}, \\ X_j = b_j + \sigma\, \eta_j, & j \in \mathbb{N}, \end{cases}$$

In some sense, we address separately the testing problem in the 'extreme' cases $\epsilon = 0, \sigma \neq 0$ and $\epsilon \neq 0, \sigma = 0$. Two different questions are at hand :

- What can be done when the operator is completely known ? (already addressed in the literature)
- Can we trust the observations on the operator ?

Lower bound when $\sigma = 0$

We deal with

$$Y_j = b_j\theta_j + \varepsilon\,\xi_j, \quad j \in \mathbb{N}.$$

Theorem (Laurent, Loubes, M. (2012)

There exists a constant $C_{\alpha,\beta}$ such that

$$\tilde{r}^2_{\varepsilon,0} \geq C_{\alpha,\beta} \sup_{D\in\mathbb{N}} \left[\varepsilon^2 \sqrt{\sum_{j=1}^{D} b_j^{-4}} \wedge a_D^{-2} \right].$$

The proof uses an (appropriate) uniform a priori on θ.

Lower bound when $\epsilon = 0$

We deal with

$$\begin{cases} Y_j = b_j \theta_j, & j \in \mathbb{N}, \\ X_j = b_j + \sigma\, \eta_j, & j \in \mathbb{N}, \end{cases}$$

Theorem (M. and Sapatinas (2015))

There exists a constant $C_{\alpha,\beta}$ such that

$$\tilde{r}^2_{0,\sigma} \geq C_{\alpha,\beta} \left\{ \sigma^2 \max_{1 \leq D \leq M_2} [b_D^{-2} a_D^{-2}] \right\},$$

where

$$M_2 = \sup \left\{ D \in \mathbb{N} : b_D^2 \geq \mathcal{C} \sigma^2 \right\},$$

for some (explicit) constant $\mathcal{C}$.

Remarks

Some comments

- The proof is uses a Bayesian a priori on the sequence of eigenvalues, which makes the vecteur Y and X dependent.
- This lower bound is of the same order of the one proposed for an estimation task.
- The term M_2 is the deterministic analogue of the bandwidth M introduced for the upper bound.

Outline

Considered setting

We are interested in the behavior of the rates in terms of the values of ϵ, σ.

Concerning the operator, we alternatively consider mildly and severely ill-posed problems, namely

$$b_k \sim k^{-t} \quad \text{or} \quad b_k \sim e^{-kt} \quad k \in \mathbb{N},$$

Different kind of smoothness are displayed

$$a_k \sim k^{s} \quad \text{or} \quad a_k \sim e^{ks} \quad k \in \mathbb{N}.$$

We compute the separation rates in terms of the noise levels ϵ and σ for each sub-case.

Separation rates : ordinary-smooth functions

Upper bounds

Goodness-of-Fit Testing Problem	*ordinary-smooth* $a_j \sim j^s$
mildly ill-posed $b_j \sim j^{-t}$	$\varepsilon^{4s/(2s+2t+1/2)} \vee [\sigma \ln^{3/4}(1/\sigma)]^{2[(s/t)\wedge 1]}$
severely ill-posed $b_j \sim \exp\{-jt\}$	$(\ln(1/\varepsilon))^{-2s} \vee [\ln(1/\sigma \ln^{-1/2}(1/\sigma))]^{-2s}$

Lower bounds

Goodness-of-Fit Testing Problem	*ordinary-smooth* $a_j \sim j^s$
mildly ill-posed $b_j \sim j^{-t}$	$\varepsilon^{4s/(2s+2t+1/2)} \vee \sigma^{2[(s/t)\wedge 1]}$
severely ill-posed $b_j \sim \exp\{-jt\}$	$(\ln(1/\varepsilon))^{-2s} \vee [\ln(1/\sigma)]^{-2s}$

Separation rates : super-smooth functions

Upper bounds

Goodness-of-Fit Testing Problem	super-smooth $a_j \sim e^{js}$
mildly ill-posed $b_j \sim j^{-t}$	$\varepsilon^2(\ln(1/\varepsilon))^{2t+1/2} \vee \sigma^2 \ln^{3/2}(1/\sigma)$
severely ill-posed $b_j \sim \exp\{-jt\}$	$(\ln(1/\varepsilon))^{-2s} \vee [\ln(1/\sigma \ln^{-1/2}(1/\sigma))]^{-2s}$

Lower bounds

Goodness-of-Fit Testing Problem	super-smooth $a_j \sim e^{js}$
mildly ill-posed $b_j \sim j^{-t}$	$\varepsilon^2(\ln(1/\varepsilon))^{2t+1/2} \vee \sigma^2$
severely ill-posed $b_j \sim \exp\{-jt\}$	$\varepsilon^{2s/(s+t)} \vee \sigma^{2[(s/t)\wedge 1]}$

Conclusion

Perspectives and related topics :

- Deal with the case $\theta_0 = 0$ (or more generally $\|\theta_0\| \leq \rho$).
- Adadaption to the smoothness,
- Find the optimal constants.
- Related models (density, IV regression, ...)

References

[1] N. Bissantz, G. Claeskens, H. Holzmann, and A. Munk. Testing for lack of fit in inverse regression - with applications to biophotonic imaging. J. R. Stat. Soc. Ser. B Stat. Methodol., 71(1) :25-48, 2009.

[2] L. Cavalier and N.W. Hengartner. Adaptive estimation for inverse problems with noisy operators. Inverse Problems, 21(4) :1345-1361, 2005.

[3] S. Delattre, M. Hoffmann, D. Picard, and T. Vareschi. Blockwise SVD with error in the operator and application to blind deconvolution. Electronic Journal of Statistics, 6 :2274- 2308, 2012.

[4]M. Hoffmann and M. Reiss. Nonlinear estimation for linear inverse problems with error in the operator. Annals of Statistics, 36(1) :310-336, 2008.

[5] Yu.I. Ingster, T. Sapatinas, and I.A. Suslina. Minimax signal detection in ill-posed inverse problems. Annals of Statistics, 40 :1524-1549, 2012.

[6] J. Johannes and M. Schwarz. Adaptive Gaussian inverse regression with partially unknown operator. Communications in Statistics - Theory and Methods, 42(7) :1343-1362, 2013.