Minimax testing of a composite null hypothesis defined via a quadratic functional

Joint work with L. Comminges

Asymptotic Statistics and Related Topics

Tokyo, Japan

Arnak S. Dalalyan
ENSAE / CREST / GENES

Motivation 1

Testing the relevance of a group of variables

- We observe a sampled signal

$$f : \mathbb{R}^d \to \mathbb{R} \qquad \mathbf{t} = (t^1, \ldots, t^d)^\top \mapsto f(\mathbf{t})$$

  in a noisy environment.
- The dimension d is large.
- Based on a training sample, some variable selection procedure suggests the irrelevance of the subset of variables $\mathbf{t}^{J^c} := \{t^j : j \in J^c\}$.
- Based on a testing sample we would like to check the irrelevance of J^c.

This amounts to testing the hypothesis $\mathbf{E}[\mathbf{Var}(f(\mathbf{t})|\mathbf{t}^J)] = 0$.

Motivation 2
Testing the validity of a partial linear model

- We observe a sampled signal obeying the partial linear model :

$$f(\mathbf{t}) = g(\mathbf{t}^J) + \beta^\top \mathbf{t}^{J^c}$$

in a noisy environment.
- g, J and β are unknown.
- The dimension d is large, but the cardinal of J is small.
- For a given set J_0, we would like to test the hypothesis $J = J_0$.

This amounts to testing the hypothesis $\mathbf{Var}[\nabla_{J_0^c} f(\mathbf{t})] = 0$.

Motivation 3

Testing the equality of two norms

- Two noisy (sub)images g_1 and g_2 are observed.
- The goal is to check whether they coincide up to a rotation and illumination change : $g_1(\boldsymbol{z}) = g_2(\mathtt{R}\boldsymbol{z}) + a$, $\forall \boldsymbol{z} \in D \subset \mathbb{R}^2$, for some orthogonal matrix $\mathtt{R}$ and some $a \in \mathbb{R}$.
- This requires testing the hypothesis

$$H_0 : \quad \exists(\mathtt{R}, a) \quad \text{s.t.} \quad g_1(\boldsymbol{z}) = g_2(\mathtt{R}\boldsymbol{z}) + a, \quad \forall \boldsymbol{z} \in D \qquad (1)$$

which is usually very time-consuming (involves a nonlinear and nonconvex minimization step).

A simpler strategy is to start with testing $H_0' :$ $\mathbf{Var}[g_1(\mathbf{Z})] = \mathbf{Var}[g_2(\mathbf{Z})]$, and to reject the hypothesis H_0 if H_0' is rejected.

Unifying framework

Testing the nullspace of a quadratic functional in regression

Model: $x_i = f(t_i)+\xi_i$; $i=1,\ldots,n$; t_i in $\mathbf{R}^d$

Hypothesis: $Q[f] = 0$

↓

SVD of Q: $\{q_l\}$, $\{\psi_l(.)\}$

Smoothness: f in Σ, $C[f]<1$

↓

SVD of C: $\{c_l\}$, $\{\varphi_l(.)\}$

Assumption: the two bases $\{\psi_l(.)\}$ and $\{\varphi_l(.)\}$ coincide

Relation to previous work

	Non Gaussian	Sampled	Multi-variate	Beyond $Q = I$	Beyond $Q \succeq 0$
Ingster & Stepanova 2011	x	x	✓	x	x
Ingster & Sapatinas 2009	x	✓	✓	x	x
Ingster, Sapatinas & Suslina 2012	x	x	x	✓	x
Laurent, Loubes & Marteau 2011	x	x	x	✓	x
Comminges & D. 2012	✓	✓	✓	✓	✓

Remark The approach adopted in the first three references is purely asymptotic, whereas Laurent et al. (2011) obtained nonasymptotic rates of separation.

Overview of our results
Testing procedure

- We observe $\{(x_i, \mathbf{t}_i)\}_{i=1,\ldots,n} \subset \mathbb{R} \times [0,1]^d$ such that

$$x_i = f(\mathbf{t}_i) + \xi_i, \quad f(\mathbf{t}) = \sum_{\ell \in \mathscr{L}} \theta_\ell[f] \varphi_\ell(\mathbf{t}),$$

  where ξ_i iid with $\mathbf{E}[\xi_1] = 0$ and $\mathbf{t}_i \overset{\text{iid}}{\sim} \mathcal{U}[0,1]^d$.
- We wish to test the hypothesis

$$H_0 : \; Q[f] = \sum_{\ell \in \mathscr{L}} q_\ell \theta_\ell[f]^2 = 0 \qquad H_1 : \; |Q[f]| > \rho^2.$$

- Each $\theta_\ell[f]^2$ is unbiasedly estimated by

$$\widehat{\theta_\ell^2} = \frac{1}{n(n-1)} \sum_{i \neq i'} x_i x_{i'} \varphi_\ell(\mathbf{t}_i) \varphi_\ell(\mathbf{t}_{i'}).$$

- Given a sequence of weights $\boldsymbol{w} = \{w_\ell\}$, we estimate $Q[f]$ by

$$\widehat{Q}_n^{\boldsymbol{w}} = \sum_{\ell \in \mathscr{L}} w_\ell q_\ell \widehat{\theta_\ell^2}.$$

- **Test :** we fix a threshold $u > 0$ and reject H_0 if $|\widehat{Q}_n^{\boldsymbol{w}}| > u$.

For any estimator $\widehat{Q}_n$, we can write $\widehat{Q}_n = Q[f] + \epsilon_n[f]$.

- Under H_0 : $|\widehat{Q}_n| \le \sup_{f\in\mathcal{F}_0} |\epsilon_n[f]|$.
- Under H_1 : $|\widehat{Q}_n| \ge \rho^2 - \sup_{f\in\mathcal{F}_1(\rho)} |\epsilon_n[f]|$.
- The testing statistic $\widehat{Q}_n$ leads to a consistent test if

$$\sup_{f\in\mathcal{F}_0} |\epsilon_n[f]| < \rho^2 - \sup_{f\in\mathcal{F}_1(\rho)} |\epsilon_n[f]| \qquad \text{(with prob. } 1-\gamma).$$

- Let $\rho_n(\widehat{Q})$ be the smallest possible $\rho > 0$ satisfying

$$\sup_{f\in\mathcal{F}_0} |\epsilon_n[f]| + \sup_{f\in\mathcal{F}_1(\rho)} |\epsilon_n[f]| < \rho^2, \qquad \text{(with prob. } 1-\gamma).$$

- **Minimax rate of separation :** $\rho_n^* \asymp \inf_{\widehat{Q}_n} \rho_n(\widehat{Q})$.

Where the difference with the minimax rate of estimation comes from : replacing $\sup_{f\in\mathcal{F}_1(\rho)}$ with $\sup_{\rho>0}\sup_{f\in\mathcal{F}_1(\rho)}$ leads to the minimax rate of estimation, but this is sub-optimal !

Overview of our results
Minimax rates of separation

- Let us call the ratio $|q_\ell|/c_\ell$ the importance of the axis φ_ℓ.
- Let $\mathcal{N}(T)$ be the set of indices with importance $\geq T > 0$.
- Let $M(T) = \sum_{\ell\in\mathcal{N}(T)} q_\ell^2$.
- In the general case, the minimax rate of separation is given by

$$(\rho^*_{n,\gamma})^2 = \inf_{T>0}\Big(\frac{4\big(B_1 M(T) + B_2 n\big)^{1/2}}{n\gamma^{1/2}} + 2\sqrt{2}T\Big)$$
$$\asymp \inf_{T>0}\Big(\frac{M(T)^{1/2}}{n} + T\Big)\bigvee n^{-1/2}.$$

- Interestingly, in the case of positive $Q \succeq 0$,

$$(\rho^*_{n,\gamma})^2 \asymp \inf_{T>0}\Big(\frac{M(T)^{1/2}}{n} + T\Big).$$

- In both cases, the test defined using the statistic $\widehat{Q}_n^w$ with the weights $w_\ell = \mathbb{1}(|q_\ell|/c_\ell \geq T)$ achieves the optimal rate.

Relation to the norm estimation

Phase transition/ “Elbow” effect

Let us assume the simple case $q_\ell^2 = 1$ and $c_\ell = \sum_{j=1}^d \ell_j^{2\sigma_j}$, $\ell \in \mathbb{Z}^d$.
One can check that $M(T) \asymp T^{-d/(2\bar\sigma)}$ where $\bar\sigma^{-1} = \frac{1}{d}\sum \sigma_j^{-1}$.

In hypotheses testing :

- If Q is positive, the mmx rate of separation is
$$(\rho_n^*)^2 \asymp n^{-4\bar\sigma/(4\bar\sigma+d)}.$$
- If Q is neither positive nor negative, the mmx rate of separation is
$$(\rho_n^*)^2 \asymp n^{-(4\bar\sigma/(4\bar\sigma+d)\,\wedge\,1/2)}.$$

Relation to the norm estimation

Phase transition/ “Elbow” effect

Let us assume the simple case $q_\ell^2 = 1$ and $c_\ell = \sum_{j=1}^d \ell_j^{2\sigma_j}$, $\ell \in \mathbb{Z}^d$.
One can check that $M(T) \asymp T^{-d/(2\bar\sigma)}$ where $\bar\sigma^{-1} = \frac{1}{d}\sum \sigma_j^{-1}$.

In hypotheses testing :

- If Q is positive, the mmx rate of separation is
$$(\rho_n^*)^2 \asymp n^{-4\bar\sigma/(4\bar\sigma+d)}.$$
- If Q is neither positive nor negative, the mmx rate of separation is
$$(\rho_n^*)^2 \asymp n^{-(4\bar\sigma/(4\bar\sigma+d)\,\bigwedge\,1/2)}.$$

In functional estimation :

- If $Q[f] = \|f\|_2$, the mmx rate of estimation is (Lepski et al. ’99)
$$r_n^* \asymp n^{-2\bar\sigma/(4\bar\sigma+d)}.$$
- If $Q[f] = \|f\|_2^2$, the mmx rate of estimation is (Donoho and Nussbaum ’90)
$$r_n^* \asymp n^{-(4\bar\sigma/(4\bar\sigma+d)\,\bigwedge\,1/2)}.$$

Main result I
Positive functionals

Theorem 1. Assume that $E[\xi_1^4] < \infty$ and for every $T > 0$, the set $\mathcal{N}(T) = \{\ell : q_\ell \geq Tc_\ell\}$ is finite. For a $\gamma \in (0,1)$, let $T_{n,\gamma}$ be such that :

$$\Big(\tfrac{n(n-1)}{2}\sum_\ell (q_\ell - Tc_\ell)_+^2\Big)^{1/2} = \Big(\sum_\ell c_\ell(q_\ell - Tc_\ell)_+\Big)(2z_{1-\gamma/2} + o(1)).$$

Let us define

$$\rho^*_{n,\gamma} = \left\{\frac{\sum_{l\in\mathscr{L}} q_\ell(q_\ell - T_{n,\gamma}c_\ell)_+}{\sum_{l\in\mathscr{L}} c_\ell(q_\ell - T_{n,\gamma}c_\ell)_+}\right\}^{1/2}.$$

If several conditions are fulfilled, then the test based on the array

$$\widehat{w}^*_{l,n} = \left(1 - \frac{T_{n,\gamma}c_\ell}{q_\ell}\right)_+$$

satisfies $\gamma_n(\mathcal{F}_0, \mathcal{F}_1(\rho^*_{n,\gamma}), \widehat{\phi}^*_n) \leq \gamma + o(1)$, as $n \to \infty$.

Testing partial derivatives

- Let $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ and $\boldsymbol{\sigma} \in \mathbb{R}_+^d$ be two given vectors.
- Let $Q[f] = \|\partial^{\sum_j \alpha_j} f / \partial t_1^{\alpha_1} \ldots \partial t_d^{\alpha_d}\|_2^2, \quad C[f] = \sum_{j=1}^d \|\partial^{\sigma_j} f / \partial t_j^{\sigma_j}\|_2^2.$
- Let us define δ, $\bar\sigma$, (κ_j) and κ by

$$\delta = \textstyle\sum_{j=1}^d \alpha_j/\sigma_j, \qquad \frac{1}{\bar\sigma} = \frac{1}{d}\sum_{j=1}^d \frac{1}{\sigma_j}.$$

- If $\qquad \delta < 1 \quad$ and $\quad \bar\sigma > d/4,$

  then the exact mmx rate $\rho_{n,\gamma}^*$ is given by $\rho_{n,\gamma}^* = C_\gamma^* \rho_n^*(1 + o(1)),$
- where the minimax rate ρ_n^* and the exact separation constant are

$$\rho_n^* = n^{-\frac{2\bar\sigma(1-\delta)}{4\bar\sigma+d}},$$

and $C_\gamma^* = \big(4z_{1-\gamma/2}^2 \kappa C(d, \boldsymbol{\sigma}, \boldsymbol{\alpha})\big)^{\frac{\bar\sigma(1-\delta)}{4\bar\sigma+d}} (1 + 2\kappa^{-1})^{\frac{2(1+\delta)\bar\sigma+d}{2(4\bar\sigma+d)}}$ with $\kappa_j = \frac{1}{2\sigma_j} + \frac{\alpha_j}{\sigma_j}\frac{4\bar\sigma+d}{2\bar\sigma(1-\delta)}$ and $\kappa = \sum_{j=1}^d \kappa_j$ and $C(d, \boldsymbol{\sigma}, \boldsymbol{\alpha}) = \pi^{-d}\frac{\prod_{i=1}^d \Gamma(\kappa_i)}{\big(\prod_{i=1}^d \sigma_i\big)(1-\delta)\Gamma(\kappa+2)}.$

Conclusion

- We established minimax rates of separation in the model of regression with random design for null hypotheses corresponding to the nullspace of a general quadratic functionals.
- In the case of positive functionals, we also proved sharp-minimax optimality of the proposed procedure.
- When comparing two norms, the minimax rate of separation is : $\rho_n^* = n^{-\frac{2\bar{\sigma}}{4\bar{\sigma}+d} \wedge \frac{1}{4}}$. This rate shows that the watershed between the two regimes corresponds to the condition $\bar{\sigma} = d/4$. In other terms, we are in the regular regime when $\bar{\sigma} > d/4$. It is interesting to note, even if we are unable to establish a direct connection, that this is also the regime under which the Sobolev embedding $W_2^{\sigma} \subset L_4([0, 1]^d)$ holds true.
- Open questions : adaptation to the unknown smoothness, unknown noise level, the case of (sparse) Besov bodies,...

Thank You