Parametric, mismatched and semiparametric estimation: how to reconcile robustness and efficiency

Stefano Fortunati

Laboratoire des signaux et systèmes (L2S) & DR2I-IPSA,
Paris, France.

Seminar at ENSAI

Rennes

January 24, 2024

Outline of the presentation

Parametric models: a recall

Misspecified models

The MCRB and the Mismatched ML estimator

Example: Misspecified estimation in ES distributions

Semiparametric models

The SCRB and the RAL estimators

Efficient semiparametric estimators

Introduction

- Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the set of n observations collected from a random experiment, such that $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbb{R}^m$, $\forall i$.
- We suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are *iid* with "true" distribution P_0 and related density (pdf) $p_0 = dP_0 d\mu$, [1] i.e. $\mathbf{x}_i \sim p_0$, $\forall i$.
- In point estimation, we are interested in evaluating some functional of p_0, say $\nu(p_0)$.
- However, p_0 is generally unknown, at least to some extent.
- The lack of *a priori* knowledge on the data generating density p_0 can be formalized in the concept of *statistical models*.

[1] Here μ is the Lebesgue measure over $\mathbb{R}^m$.

Parametric models

- The most widely used statistical models are the *parametric* ones.
- A parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ is defined as a set of pdfs that are parametrized by a finite-dimensional parameter vector $\boldsymbol{\theta}$:

$$\mathcal{P}_{\boldsymbol{\theta}} \triangleq \{p_X(\mathbf{x}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}.$$

- The underlying parametric assumption is that there exists $\boldsymbol{\theta}_0 \in \Theta \subseteq \mathbb{R}^q$, such that:

$$\mathcal{P}_{\boldsymbol{\theta}} \ni p_X(\mathbf{x}|\boldsymbol{\theta}_0) = p_0(\mathbf{x}), \ \forall \mathbf{x} \in \mathcal{X}. \qquad \text{(A1)}$$

- The (lack of) knowledge about the data generating process is summarized in $\boldsymbol{\theta}$ that needs to be estimated.

The score vector and Fisher Information

- A key element in parametric inference is the **score vector**:

$$\mathbf{s}_{\boldsymbol{\theta}} \triangleq \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}).$$

- "Under some regularity conditions" [2], and under Assumption A1, the **Fisher Information Matrix** (FIM) is defined as:

$$\mathbf{I}(\boldsymbol{\theta}_0) \triangleq E_0\left\{\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x};\boldsymbol{\theta}_0)\mathbf{s}_{\boldsymbol{\theta}}^T(\mathbf{x};\boldsymbol{\theta}_0)\right\}, \quad \mathbf{x} \sim p_0,$$

where $E_0\{g\} \triangleq \int g dP_0$.

- The FIM plays a central role in defining the concept of *efficiency* of an estimator of the parameter $\boldsymbol{\theta}_0 \in \Theta$.

[2] Due to the limited time of the talk, we will not discuss them here. Moreover, we will omit to repeat this "magic" sentence in the following derivations.

The Cramér-Rao Inequality

- Let $\hat{\boldsymbol{\theta}}_n \triangleq \hat{\boldsymbol{\theta}}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be an estimator of $\boldsymbol{\theta}_0 \in \Theta$ based on the set of n iid observations.

- The estimator $\hat{\boldsymbol{\theta}}_n$ is said to be **unbiased** if:

$$E_0\{\hat{\boldsymbol{\theta}}_n\} \triangleq \int \hat{\boldsymbol{\theta}}(\mathbf{x}_1, \ldots, \mathbf{x}_n) dP_0(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \boldsymbol{\theta}_0.$$

- Under Assumption A1, the following inequality on the Mean Squared Error (MSE) of $\hat{\boldsymbol{\theta}}_n$ holds true:

Cramér-Rao Inequality: Any *unbiased* estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$, derived in $\mathcal{P}_{\boldsymbol{\theta}}$ from $\{\mathbf{x}_i \sim p_0\}_{i=1}^n$ iid observations, satisfies:

$$n \cdot E\left\{(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T\right\} \geq \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \triangleq \mathrm{CRB}(\boldsymbol{\theta}_0)$$

(Asymptotic) Efficient estimators

- An estimator $\hat{\boldsymbol{\theta}}_n$ is said to be (asymptotically) **efficient** if it satisfies the following two properties:

  P1 $\sqrt{n}$*-consistent*:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = O_P(1).^3$$

  P2 *Asymptotically Gaussian and efficient*:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \underset{n\to\infty}{\overset{d}{\sim}} \mathcal{N}(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}) = \mathcal{N}(\mathbf{0}, \mathrm{CRB}(\boldsymbol{\theta}_0)),$$

  where $\underset{n\to\infty}{\overset{d}{\sim}}$ indicates the convergence in distribution.

- There are at least two estimators satisfying P1 and P2: the Maximum Likelihood (ML) and the One-Step (OS) estimators.

[3] Let x_i be a sequence of random variables. Then $x_i = O_P(1)$ if for any $\epsilon > 0$, there exists a finite $N > 0$ and a finite $L > 0$, s.t. $\Pr\{|x_i| > N\} < \epsilon, \forall i > L$ (*stochastic boundedness*).

The Maximum Likelihood estimator (MLE)

- By indicating as:

$$L_n(\boldsymbol{\theta}) \triangleq \sum_{i=1}^{n} \ln p_X(\mathbf{x}_i|\boldsymbol{\theta})$$

the *likelihood function*, the MLE $\hat{\boldsymbol{\theta}}_{ML,n}$ can be expressed as:

$$L_n(\hat{\boldsymbol{\theta}}_{ML,n}) = \max\left\{L_n(\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\right\}.$$

- Let us now introduce the **central sequence** as:

$$\Delta_n(\boldsymbol{\theta}) \triangleq n^{-1/2} \sum_{i=1}^{n} \mathbf{s}(\mathbf{x}_i; \boldsymbol{\theta}),$$

- Then, en equivalent version of the MLE, can be obtained from the set of *estimating equations*:

$$\Delta_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML,n}} = \mathbf{0}.$$

One-Step Le Cam's estimators

- Let us introduce a **preliminary estimator** $\boldsymbol{\theta}^{\star}$ as a $\sqrt{n}-$consistent, but not necessarily efficient, estimator of $\boldsymbol{\theta}_0$.
- Le Cam showed that the following "one-step" estimator satisfies the properties P1 and P2:

$$\hat{\boldsymbol{\theta}}_{OS,n} = \boldsymbol{\theta}_n^{\star} + n^{-1/2}[\mathbf{I}(\boldsymbol{\theta}_n^{\star})]^{-1}\Delta_n(\boldsymbol{\theta}_n^{\star}).$$

- The one-step estimator $\hat{\boldsymbol{\theta}}_{OS,n}$ is asymptotically equivalent to the MLE $\hat{\boldsymbol{\theta}}_{ML,n}$.
- **!Note!**: the existence of the one-step estimator requires much weaker regularity conditions with respect to the MLE.

Outline of the presentation

Parametric models: a recall

Misspecified models
- The MCRB and the Mismatched ML estimator
- Example: Misspecified estimation in ES distributions

Semiparametric models
- The SCRB and the RAL estimators
- Efficient semiparametric estimators

Model misspecification

- **Classical "matched" assumption A1**: the model, assumed to derive the estimation algorithm, contains the true pdf p_0.
- All the results on the ML and OP estimators and the CRB rely on Assumption A1.
- However, much evidence from everyday practice shows that this assumption is often violated.
- **Model misspecification**: the assumed parametric model may not contain the true pdf p_0.

Model misspecification

- There are two main reasons for model misspecification:
  1. An **imperfect knowledge** of the true data generating process that leads to a wrong specification of the data pdf.
  2. The true parametric model is known but it is **too involved** to pursue the optimal "matched" estimator.

- One may be forced (1) or may prefer (2) to derive an estimator by assuming a *simpler* but *misspecified* data model.

- This sub-optimal procedure may lead to some degradation in the estimation performance.

Formal description of the misspecification

- Our observations $\{\mathbf{x}_i \sim p_0\}_{i=1}^n$ are *iid* with pdf p_0 belonging to a possibly non-parametric model $\mathcal{P}$.

- To characterize the statistical behavior of $\mathbf{x}_i, \forall i$, we adopt a different parametric pdf, say $f_X(\cdot|\boldsymbol{\gamma})$, with $\boldsymbol{\gamma} \in \Gamma \subseteq \mathbb{R}^p$.

- The adopted pdf $f_X(\cdot|\boldsymbol{\gamma})$ is assumed to belong to a *possibly misspecified* parametric model :

$$\mathcal{F}_\gamma \triangleq \{f_X(\mathbf{x}|\boldsymbol{\gamma}), \boldsymbol{\gamma} \in \Gamma\}.$$

- The classical "matched" assumption A1 requires:

$$\exists \bar{\boldsymbol{\gamma}} \in \Gamma,\ f_X(\mathbf{x}|\bar{\boldsymbol{\gamma}}) = p_0(\mathbf{x}),\ \forall \mathbf{x} \in \mathcal{X},$$

or, equivalently, that $p_0 \in \mathcal{F}_\gamma$.

Formal description of the misspecification

- If the previous assumption is violated, the model $\mathcal{F}_\gamma$ is *misspecified*. Formally:

$$f_X(\mathbf{x}|\gamma) \neq p_0(\mathbf{x}), \; \forall \gamma \in \Gamma,$$

  or, equivalently, that $p_0(\mathbf{x}) \notin \mathcal{F}_\gamma$.

- This misspecified scenario raises three main questions: [4]
  1. Which parameter vector $\gamma \in \Gamma$ make sense to estimate?
  2. How do the properties of an estimator, e.g. *unbiasedness* and *efficiency*, change in this misspecified framework?
  3. Is it still possible to derive lower bounds on the error covariance of any mismatched estimator?

[4] S. Fortunati, F. Gini, M. S. Greco and C. D. Richmond, "Performance Bounds for Parameter Estimation under Misspecified Models: Fundamental Findings and Applications", *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 142-157, Nov. 2017.

The pseudo-true parameter vector

- Suppose that there exist a unique interior point $\gamma_0 \in \Gamma$, s.t.:

$$\gamma_0 \triangleq \underset{\gamma \in \Gamma}{\operatorname{argmin}} \{ D(p_0 \parallel \mathcal{F}_\gamma) \}$$

  where

$$D(p_0 \parallel \mathcal{F}_\gamma) \triangleq \int \ln \left(\frac{p_0(\mathbf{x})}{f_X(\mathbf{x}|\gamma)} \right) dP_0(\mathbf{x}),$$

  is the **Kullback-Leibler divergence** (KLD) between the true p_0 and the assumed model $\mathcal{F}_\gamma$.

- The **pseudo-true parameter vector** γ_0 is the vector in Γ the minimizes the KLD between p_0 and $\mathcal{F}_\gamma$.

- It may make sense to estimate it.

Information matrices under misspecification

- Let $\mathbf{A}_{\gamma_0}$ be the matrix defined as:

$$\mathbf{A}_{\gamma_0} \triangleq E_0 \left\{ \nabla_\gamma \nabla_\gamma^T \ln f_X(\mathbf{x}|\gamma_0) \right\}, \quad \mathbf{x} \sim p_0.$$

- Let $\mathbf{B}_{\gamma_0}$ be the matrix defined as:

$$\mathbf{B}_{\gamma_0} \triangleq E_0 \left\{ \nabla_\gamma \ln f_X(\mathbf{x}|\gamma_0) \nabla_\gamma^T \ln f_X(\mathbf{x}|\gamma_0) \right\}, \quad \mathbf{x} \sim p_0.$$

- If the model is correctly specified, i.e. if $\exists \bar{\gamma} \in \Gamma$ such that $f_X(\mathbf{x}|\bar{\gamma}) = p_0(\mathbf{x})$, then:
  1. $\gamma_0 = \bar{\gamma}$, i.e. the pseudo-true parameter is equal to the true one (in the classical "matched" sense),
  2. $\mathbf{B}_{\gamma_0} = -\mathbf{A}_{\gamma_0} = \mathbf{I}(\bar{\gamma})$, where $\mathbf{I}(\bar{\gamma})$ is the Fisher Information Matrix for the ("matched" in this case) model $\mathcal{F}_\gamma$.

The Misspecified Cramér-Rao Bound

- Given our $\{\mathbf{x}_i \sim p_0\}_{i=1}^n$ *iid* observations, let's consider an estimator $\hat{\gamma}_n$ derived in the misspecified model $\mathcal{F}_\gamma$.

- **Misspecified (MS)-unbiasedness**: the estimator $\hat{\gamma}_n$ is said to be MS-unbiased iff:

$$E_0\{\hat{\gamma}_n\} \triangleq \int \hat{\gamma}(\mathbf{x}_1, \ldots, \mathbf{x}_n) dP_0(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \gamma_0.$$

Misspecified CRB: Any *MS-unbiased* estimator $\hat{\gamma}_n$ of γ_0, derived in $\mathcal{F}_\gamma$ from $\{\mathbf{x}_i \sim p_0\}_{i=1}^n$ iid observations, satisfies: [5,6,7]

$$n \cdot E_0\left\{(\hat{\gamma}_n - \gamma_0)(\hat{\gamma}_n - \gamma_0)^T\right\} \geq \mathbf{A}_{\gamma_0}^{-1}\mathbf{B}_{\gamma_0}\mathbf{A}_{\gamma_0}^{-1} \triangleq \mathrm{MCRB}(\gamma_0).$$

[5] Q. H. Vuong, "Cramér-Rao bounds for misspecified models", *Working paper 652, Division of the Humanities and Social Sciences*, Caltech, October 1986.

[6] S. Fortunati, F. Gini, M. S. Greco, "The Constrained Misspecified Cramér-Rao Bound", *IEEE Signal Process. Letters*, vol. 23, No. 5, pp. 718-721, May 2016.

[7] S. Fortunati, "Misspecified Cramér-Rao Bounds for Complex Unconstrained and Constrained Parameters," EUSIPCO 2017, Kos, Greece, 28 Aug. 2017–2 Sept. 2017

The Misspecified ML estimator (MML)

- The MML estimator, defined on the possibly misspecified parametric model $\mathcal{F}_\gamma$, is given by:

$$\widehat{\gamma}_{MML,n} \triangleq \underset{\gamma \in \Gamma}{\mathrm{argmax}} \sum_{i=1}^{n} \ln f_X(\mathbf{x}_i|\gamma), \quad \mathbf{x}_i \sim p_0.$$

Properties: the MML estimator $\widehat{\gamma}_{MML,n}$ is: [8,9]

1. $\sqrt{n}$-*MS-consistent*:

$$\sqrt{n}\left(\hat{\gamma}_{MML,n} - \gamma_0\right) = O_P(1).$$

2. *Asymptotically Gaussian and MS-efficient*:

$$\sqrt{n}\left(\hat{\gamma}_{MML,n} - \gamma_0\right) \underset{n \to \infty}{\overset{d}{\sim}} \mathcal{N}(\mathbf{0}, \mathbf{A}_{\gamma_0}^{-1}\mathbf{B}_{\gamma_0}\mathbf{A}_{\gamma_0}^{-1}) = \mathcal{N}(\mathbf{0}, \mathrm{MCRB}(\gamma_0)),$$

[8] P. J. Huber, "The behavior of Maximum Likelihood Estimates under Nonstandard Conditions," *Proc. of the Fifth Berkeley Symposium in Mathematical Statistics and Probability.* Berkley: University of California Press, 1967

[9] H. White, "Maximum likelihood estimation of misspecified models", *Econometrica* vol.50,pp.1-25, Jan. 1982.

To fix the ideas...

- Estimating the correlation structure, i.e. the covariance matrix, of a dataset is a central problem in many applications:

  1. Dimensionality reduction and Principal Component Analysis,
  2. Signal/Image Denoising,
  3. Adaptive detection in radar/sonar systems,
  4. Graph signal processing,
  5. ...

- A general working assumption (motivated by the CLT) consists of assuming the data as Gaussian-distributed.

- However, this assumption is generally violated in practical applications where the data may be better characterized by heavy-tailed distributions.

A set of heavy-tailed distributions

- A family of non-Gaussian/heavy-tailed distribution is the class of (real) **Elliptically Symmetric (ES)** distributions.
- Thanks to their flexibility, ES distributions represent a reliable data model in many applications. [10]
- The Gaussian, Generalized Gaussian, K-distribution, t-distribution and all the compound-Gaussian distributions belong to the ES class.
- The ES model is particularly useful in applications with *impulsive noise* and/or *spiky data*.

[10] J.-P. Delmas, M. N. El Korso, S. Fortunati, F. Pascal, *Elliptically Symmetric Distributions in Signal Processing and Machine Learning*, Springer book, in press.

Real ES distributions

- A *zero-mean* ES distributed random vector $\mathbf{x} \in \mathbb{R}^m$ has pdf:

$$p_0(\mathbf{x}) = 2^{-m/2}|\boldsymbol{\Sigma}|^{-1/2} h_{\boldsymbol{\eta}}(\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}) \triangleq ES_N(\mathbf{0}, \boldsymbol{\Sigma}, h_{\boldsymbol{\eta}}).$$

  - $h_{\boldsymbol{\eta}} \in \mathcal{H}$, $h_{\boldsymbol{\eta}} : \mathbb{R}_0^+ \to \mathbb{R}^+$ is the *density generator* possibly parameterized by a set of parameters $\boldsymbol{\eta}$.
  - $\boldsymbol{\Sigma} \in \mathcal{M}_l$ is the (full rank) scatter matrix.
  - To avoid the scale ambiguity between $\boldsymbol{\Sigma}$ and $h_{\boldsymbol{\eta}}$, the shape matrix is generally introduced as:

$$\mathbf{V} = \boldsymbol{\Sigma}/s(\boldsymbol{\Sigma}), \; s : \mathcal{M}_l \to \mathbb{R}^+.$$

- **Stochastic representation**: If $\mathbf{x} \sim ES_m(\mathbf{0}, \boldsymbol{\Sigma}, h_{\boldsymbol{\eta}})$, then

$$\mathbf{x} =_d \sqrt{\mathcal{Q}} \boldsymbol{\Sigma}^{1/2} \mathbf{u}$$

  - $\mathbf{u} \sim \mathcal{U}(S^{m-1})$ is uniformly distributed on the unit $m-1$ sphere,
  - $\mathcal{Q} \sim p_{\mathcal{Q}}(q) = c_m q^{m/2-1} h_{\boldsymbol{\eta}}(q)$ is called *2nd-order modular variate*.

An example: the t-distribution

- The pdf of t-distributed data can be obtained from the ES family by specifying the density generator:

$$h_\lambda(t) = \frac{\Gamma(m+\lambda/2)}{(\lambda\pi)^{m/2}\Gamma(\lambda/2)}\left(1+\frac{t}{\lambda}\right)^{-\frac{m+\lambda}{2}},$$

  where λ controls the data non-Gaussianity.

- To guarantee the finitness of the second order moments (i.e. the existence of the covariance matrix), we need $\lambda > 2$.

- Note that for values of $\lambda \to 2$ the data are heavy-tailed, while for $\lambda \to \infty$ the data tends to be Gaussian.

A common misspecified scenario in ES data

- Our *iid* observations $\{\mathbf{x}_i \sim p_0\}_{i=1}^n$ are ES-distributed, that is $p_0 \sim ES_m(\mathbf{0}, \mathbf{\Sigma}_0, h_0)$.
- The "true" density generator is the t- one: $h_0(t) = h_{\lambda_0}(t)$.
- The practitioner decides to built a ML estimator for $\mathbf{V}_0$ on the misspecified Gaussian model $\mathcal{F}_\gamma$, such that:

$$\mathcal{F}_\gamma = \left\{ f_X(\mathbf{x}|\gamma) = 2^{-m/2}|\mathbf{\Sigma}|^{-1/2} g_{\sigma^2}(\mathbf{x}^T \mathbf{\Sigma}^{-1} \mathbf{x}), \gamma \in \Gamma \right\},$$

  - $g_{\sigma^2}(t) = (\pi\sigma^2)^{-N} \exp\left(-t/\sigma^2\right)$,
  - $\gamma \in \Gamma \triangleq \{\gamma = (\mathrm{vec}(\mathbf{V})^T, \sigma^2)^T | \mathbf{V} = \mathbf{\Sigma}/s(\mathbf{\Sigma})\}$,

- Clearly, $f_X(\mathbf{x}|\gamma) \neq p_0(\mathbf{x})$, $\forall \gamma \in \Gamma$: model mismatch!

Misspecified scatter matrix estimation

- We need to evaluate the pseudo-true parameter vector

$$\gamma_0 \triangleq \underset{\gamma \in \Gamma}{\operatorname{argmin}} \left\{ D(p_0 \parallel \mathcal{F}_\gamma) \right\},$$

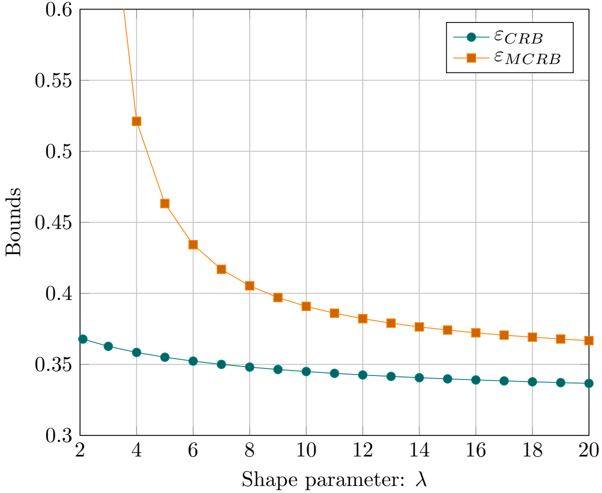
  where p_0 is the t-distribution and $\mathcal{F}_\gamma$ is the Gaussian model.

- It can be shown [11] that $\gamma_0 = (\mathrm{vec}(\mathbf{V}_0)^T, \sigma_0^2)^T$, where $\mathbf{V}_0$ is the true shape matrix.

- We can now compare:
  - $\varepsilon_{CRB} \triangleq ||\mathrm{CRB}(\mathbf{V}_0)||_F$ evaluated for the true t-distribution,
  - $\varepsilon_{MCRB} \triangleq ||\mathrm{MCRB}(\mathbf{V}_0)||_F$ evaluated for the misspecified Gaussian distribution.

[11] S. Fortunati, F. Gini, M. S. Greco, "The Misspecified Cramér-Rao Bound and its Application to the Scatter Matrix estimation in Complex Elliptically Symmetric distributions," IEEE Trans. Signal Processing, vol. 64, no. 9, pp. 2387-2399, 2016.

Misspecified estimation performance: Bounds



- For small values of λ (highly non-Gaussian data), the estimation losses due to model mismatching rapidly increase!

Our contributions

- Constrained MCRB for real and complex parameters, [12,13]
- Application of the misspecification theory to the covariance estimation of elliptically distributed data, [14,15]
- Misspecification theory in the presence of depended data: application to GNSS system and massive MIMO radar. [16,17]

[12] S. Fortunati, F. Gini, M. S. Greco, "The Constrained Misspecified Cramér-Rao Bound", *IEEE Signal Process. Letters*, vol. 23, No. 5, pp. 718-721, May 2016.

[13] S. Fortunati, "Misspecified Cramér-Rao Bounds for Complex Unconstrained and Constrained Parameters," EUSIPCO 2017, Kos, Greece, 28 Aug. 2017–2 Sept. 2017

[14] S. Fortunati, F. Gini, M. S. Greco, "The Misspecified Cramér-Rao Bound and its Application to the Scatter Matrix estimation in Complex Elliptically Symmetric distributions", *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2387-2399, May 2016.

[15] A. Mennad, S. Fortunati, M. N. El Korso, A. Younsi, A. M. Zoubir, A. Renaux, "Slepian-Bangs-type formulas and the related Misspecified Cramér-Rao Bounds for Complex Elliptically Symmetric Distributions", *Signal Processing*, 142C (2018) pp. 320-329.

[16] S. Fortunati, L. Sanguinetti, F. Gini, M. S. Greco and B. Himed, "Massive MIMO Radar for Target Detection", *IEEE Transactions on Signal Processing*, vol. 68, pp. 859-871, 2020.

[17] S. Fortunati, L. Ortega, "On the efficiency of misspecified Gaussian inference in nonlinear regression: application to time-delay and Doppler estimation" (preprint), submitted to *IEEE Transactions on Signal Processing*, October 2023.

Outline of the presentation

The need of robust procedures

- As expected, a wrong specification of the model leads to performance degradation.
- **Robustness**: we need some estimation procedures that:
  1. are efficient under nominal assumptions,
  2. continue to be "reliable" under (strong) departure of the true data model from the nominal one.
- There are many different definitions of robustness in statistic.
- Here, we focus on the *robustness with respect to the lack of a (partial or full) knowledge of the true data model*.
- We consider this missing knowledge as an additional **nuisance (functional) parameter** in the data model.

Semiparametric models

- A semiparametric model $\mathcal{P}_{\boldsymbol{\theta},h}$ is a set of pdfs characterized by a finite-dimensional parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ along with a *function*, i.e. an infinite-dimensional parameter, $h \in \mathcal{H}$:

$$\mathcal{P}_{\boldsymbol{\theta},h} \triangleq \{p_X(\mathbf{x}|\boldsymbol{\theta}, h), \boldsymbol{\theta} \in \Theta, h \in \mathcal{H}\}.$$

- Usually, $\boldsymbol{\theta}$ is the (finite-dimensional) parameter of interest while h can be considered as a nuisance parameter.

- Many inference problems can be cast in the semiparametric framework:

  1. Inference in ES distributions,
  2. Estimation with missing data,
  3. Non-linear regression and inverse problems,
  4. Time series analysis, ...

ES distributions as semiparametric model

- The family of ES distributions is a perfect example of *semiparametric model*.

- The (zero-mean) ES semiparametric model is obtained from the parametric one by relaxing the *unrealistic assumption* on the a-priori knowledge of the density generator:

$$\mathcal{P}_{\boldsymbol{\theta},h} \triangleq \left\{ p_X(\mathbf{x}|\boldsymbol{\theta}) = 2^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}h(\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}), \boldsymbol{\theta} \in \Theta, h \in \mathcal{H} \right\},$$

where the parameter of interest is

$$\boldsymbol{\theta} \in \Theta \triangleq \{\boldsymbol{\theta} = \mathrm{vec}(\mathbf{V}) | \mathbf{V} = \boldsymbol{\Sigma}/s(\boldsymbol{\Sigma})\},$$

while $h \in \mathcal{H}$ is a nuisance function.

Inference in semiparametric model

- To derive inference procedures in semiparametric models, the estimation problem can be framed in the contest of Hilbert spaces. [18,19]

- Describing the problem through the Hilber space geometry allows us to "override" the difference between finite and infinite-dimensional parameters.

- Any semiparametric infernce scheme is based on the following key ingredients:

  1. The score vector of the parameter of interest $\mathbf{s}(\mathbf{x};\boldsymbol{\theta}_0, h_0)$,
  2. The *nuisance tangent space* $\mathcal{T}_{h_0}$,
  3. The *efficient score vector* $\bar{\mathbf{s}}(\mathbf{x};\boldsymbol{\theta}_0, h_0)$.

[18] P.J. Bickel, C.A.J Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, 1993.

[19] S. Fortunati, "Semiparametric estimation in elliptical distributions", in *Elliptically Symmetric Distributions in Signal Processing and Machine Learning*, Springer book, in press.

The basic ingredients

- The score vector of the parameter of interest is defined as in the parametric case as:

$$\mathbf{s}(\mathbf{x};\boldsymbol{\theta}_0,h_0) \triangleq \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0,h_0).$$

- To define the *nuisance tangent space* $\mathcal{T}_{h_0}$ and the associated projection operator $\Pi(\cdot|\mathcal{T}_{h_0})$ we need the notion of *regular parametric sub-models*.

- The **efficient score vector** is defined as the residual of $\mathbf{s}(\mathbf{x};\boldsymbol{\theta}_0,h_0)$ after projecting it onto $\mathcal{T}_{h_0}$:

$$\bar{\mathbf{s}}(\mathbf{x};\boldsymbol{\theta}_0,h_0) \triangleq \mathbf{s}(\mathbf{x};\boldsymbol{\theta}_0,h_0) - \Pi(\mathbf{s}(\mathbf{x};\boldsymbol{\theta}_0,h_0)|\mathcal{T}_{h_0}),$$

- Let us finally introduce the **efficient information matrix** as:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0) \triangleq E_0\{\bar{\mathbf{s}}(\mathbf{x};\boldsymbol{\theta}_0,h_0)\bar{\mathbf{s}}(\mathbf{x};\boldsymbol{\theta}_0,h_0)^T\}.$$

A lower bound in semiparametric estimation

- Let $\{\mathbf{x}_i\}_{i=1}^n$ be a set of iid observations, such that $\mathbf{x}_i \sim p_0(\mathbf{x}; \boldsymbol{\theta}_0, h_0) \in \mathcal{P}_{\boldsymbol{\theta},h}\ \forall i$.

- The class of *Regular and Asymptotically Linear (RAL) estimators* is defined as:

  1. $\sqrt{n}$-consistent: $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_P(1)$,
  2. Asymptotically normal: $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \underset{n\to\infty}{\overset{d}{\sim}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi}(\boldsymbol{\theta}_0, h_0))$.

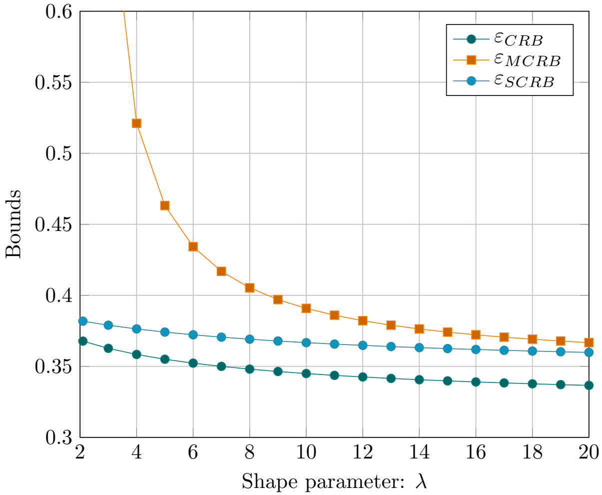
- E.g.: the robust M-estimators belong to this class.

Semiparametric CRB (SCRB): Any *RAL* estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$, derived in $\mathcal{P}_{\boldsymbol{\theta},h}$ from $\{\mathbf{x}_i \sim p_0\}_{i=1}^n$ iid observations, satisfies: [20]

$$\boldsymbol{\Xi}(\boldsymbol{\theta}_0, h_0) \geq \bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)^{-1} \triangleq \mathrm{SCRB}(\boldsymbol{\theta}_0|h_0).$$

[20] P.J. Bickel, C.A.J Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, 1993.

Ex: Bounds for the ES estimation problem



- The SCRB is in between the “unrealistic” CRB and the MCRB where a wrong model is assumed.

Efficient semiparametric estimators

- Let us focus on the efficient estimation of $\boldsymbol{\theta}_0 \in \Theta$ in the presence of the unknown function $h_0 \in \mathcal{H}$.
- Clearly, Maximum Likelihood estimation is not an option.
- What about the One-Step (OS) estimator? Could we generalize it to semiparametric model?
- We have two possible way to do this:
  1. Use a non-parametric estimation of the nuisance function,
  2. Rank-based approach.

Construction of semiparametric OS estimators

- If the nuisance function $h_0 \in \mathcal{H}$ where known, we could define the **efficient central sequence** as:

$$\overline{\Delta}_n(\boldsymbol{\theta}, h_0) \triangleq n^{-1/2} \sum_{i=1}^{n} \bar{\mathbf{s}}(\mathbf{x}_i; \boldsymbol{\theta}, h_0).$$

- By assuming to have a preliminary estimator $\boldsymbol{\theta}^\star$, [21] an OS estimator of $\boldsymbol{\theta}_0$ can be build as:

$$\hat{\boldsymbol{\theta}}_{OS,n} = \boldsymbol{\theta}_n^\star + n^{-1/2}[\bar{\mathbf{I}}(\boldsymbol{\theta}_n^\star|h_0)]^{-1}\overline{\Delta}_n(\boldsymbol{\theta}_n^\star, h_0). \qquad \text{(OS1)}$$

- Clearly, this estimator cannot be implemented since $h_0 \in \mathcal{H}$ is unknown.

- What if we had a preliminary estimator $h^\star$ for h_0 as well?

[21] Recall: $\boldsymbol{\theta}^\star$ is a $\sqrt{n}-$consistent but not necessarily efficient estimator of $\boldsymbol{\theta}_0$

Construction of semiparametric OS estimators

- Suppose to have two preliminary estimators $\boldsymbol{\theta}^\star$ and $h^\star$ for $\boldsymbol{\theta}_0$ and h_0, respectively.

- Then, the Theorem 1 in Sec. 7.8 of [22] tells us that:

$$\hat{\boldsymbol{\theta}}_{OS,n} = \boldsymbol{\theta}_n^\star + n^{-1/2}[\bar{\mathbf{I}}(\boldsymbol{\theta}_n^\star | h^\star)]^{-1}\overline{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_n^\star, h^\star),$$

  is an **efficient RAL estimator**, i.e.:

  1. $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_P(1)$,
  2. $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \underset{n\to\infty}{\overset{d}{\sim}} \mathcal{N}(\mathbf{0}, [\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)]^{-1}) = \mathcal{N}(\mathbf{0}, \mathrm{SCRB}(\boldsymbol{\theta}_0|h_0))$.

- **!Problem!**: Finding $h^\star$, i.e. a non-parametric estimator for h_0 having a rate of convergence of $\sqrt{n}$, may be difficult.

[22] P.J. Bickel, C.A.J Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, 1993.

A rank-based alternative

- As shown in [23], *in the case of ES-distributed data*, efficient semiparametric estimators may be built by using the **ranks**.

- *A recall*: let $\{x_i\}_{i=1}^n$ be a set of n continuous iid **univariate** random variables with distribution P_X.

- Let us introduce the vector of *order statistics* as:

$$\mathbf{v}_X \triangleq [x_{n(1)}, x_{n(2)}, \ldots, x_{n(n)}]^T,$$

  whose entries $x_{n(1)} < x_{n(2)} < \cdots < x_{n(n)}$ are the values of $\{x_i\}_{i=1}^n$ ordered in an ascending way.

- The rank $r_i \in \mathbb{N}$ of x_i is the position index of x_i in $\mathbf{v}_X$.

[23] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. optimal R-estimation of shape,", *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.

The two key properties of the ranks

- Let $\mathbf{r}_X \triangleq [r_1, \ldots, r_n]^T \in \mathbb{N}^n$ be the vector collecting the ranks of $\{x_i \sim P_X\}_{i=1}^n$. Then:

R1 The rank vector $\mathbf{r}_X$ is uniformly distributed on the set of all $n!$ permutations on $\{1, 2, \ldots, n\}$, regardless the actual distribution P_X.

R2 Let $P_X^{(n)}$ the *empirical distribution* of x_i defined as, for $z \in \mathbb{R}$:

$$P_X^{(n)}(z) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{x_i \leqslant z\}} = \begin{cases} 0 & z < \min_i\{x_i\} \\ \frac{r_i}{n+1} & x_i = \max_j\{x_j | x_j \leqslant z\} \end{cases} .$$

Then, we have that (Glivenko-Cantelli Theorem):

$$\max_{1 \leqslant i \leqslant n} \left| \frac{r_i}{n+1} - P_X(x_i) \right| \to 0 \; a.s. \; n \to \infty,$$

where $P_X^{(n)}(x_i) = \frac{r_i}{n+1}$.

Semiparametric rank-based R-estimators

- Let us now go back to the semiparametric OS estimator introduced in (OS1) as:

$$\hat{\boldsymbol{\theta}}_{OS,n} = \boldsymbol{\theta}_n^\star + n^{-1/2}[\bar{\mathbf{I}}(\boldsymbol{\theta}_n^\star|h_0)]^{-1}\overline{\Delta}_n(\boldsymbol{\theta}_n^\star, h_0).$$

- From the two properties of the ranks, we may derive:

  1. A $\sqrt{n}$-consistent estimator of the *efficient information matrix*:

  $$\widehat{\boldsymbol{\Upsilon}}(\boldsymbol{\theta}_0, \mathbf{r}) = \bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0) + o_P(1) = \bar{\mathbf{I}}(P_0) + o_P(1)$$

  2. A $\sqrt{n}$-consistent estimator of the *efficient central sequence*:

  $$\widetilde{\boldsymbol{\Delta}}(\boldsymbol{\theta}_0, \mathbf{r}) = \overline{\Delta}_n(\boldsymbol{\theta}_0, h_0) + o_P(1) = \overline{\Delta}_n(P_0) + o_P(1)$$

- Thanks to R1, $\widehat{\boldsymbol{\Upsilon}}(\boldsymbol{\theta}_0, \mathbf{r})$ and $\widetilde{\boldsymbol{\Delta}}(\boldsymbol{\theta}_0, \mathbf{r})$ are *"distribution-free"*, **robust** estimates of $\bar{\mathbf{I}}(P_0)$ and $\overline{\Delta}_n(P_0)$!

Semiparametric rank-based R-estimators

- A *robust and efficient* semiparametric estimator is given by:

$$\hat{\boldsymbol{\theta}}_{R,n} = \boldsymbol{\theta}_n^\star + n^{-1/2}[\widehat{\boldsymbol{\Upsilon}}(\boldsymbol{\theta}_n^\star, \mathbf{r})]^{-1}\widetilde{\boldsymbol{\Delta}}(\boldsymbol{\theta}_n^\star, \mathbf{r}). \qquad \text{(OS2)}$$

- **Crucial point**: the ranks $\mathbf{r}$ of what?

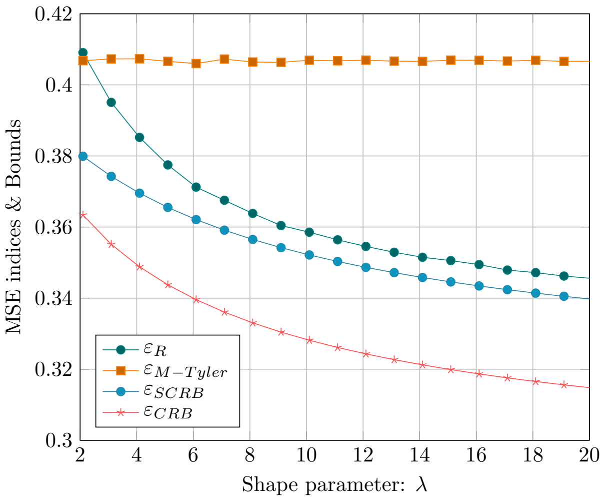
- As shown in [24], for the case of ES-distributed data, (OS2) can be obtained thanks to the stochastic representation: If $\mathbf{x} \sim ES_m(\mathbf{0}, \boldsymbol{\Sigma}, h_{\boldsymbol{\eta}})$, then

$$\mathbf{x} =_d \sqrt{\mathcal{Q}}\boldsymbol{\Sigma}^{1/2}\mathbf{u}$$

  - $\mathbf{u} \sim \mathcal{U}(S^{m-1})$ is uniformly distributed on the unit $m-1$ sphere,
  - $\mathcal{Q} \sim p_{\mathcal{Q}}(q) = c_m q^{m/2-1} h_{\boldsymbol{\eta}}(q)$ is *2nd-order modular variate*.
  - All the "information" on the true distribution P_0 of $\mathbf{x}$ is contained in the **univariate** $\mathcal{Q}$.

[24] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. optimal R-estimation of shape,", *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.

Ex: Semiparametric estimation performance



- The Tyler *M*-estimator fails to be efficient while the *R*-estimator approaches the SCRB.

Our contributions

- *R*-estimators for complex elliptical data [25,26]
  - Computational efficiency,
  - Robustness to outliers,
  - Extensive comparison with other robust estimators.

- Constrained real and complex SCRB for scatter matrix estimation and semiparametric Slepian-Bangs formula. [27,28]

- Application to array processing.

---

[25] S. Fortunati, A. Renaux, F. Pascal, "Robust semiparametric efficient estimators in complex elliptically symmetric distributions", *IEEE Transactions on Signal Processing*, vol. 68, pp. 5003-5015, 2020.

[26] S. Fortunati, A. Renaux, F. Pascal, "Joint Estimation of Location and Scatter in Complex Elliptical Distributions: A robust semiparametric and computationally efficient *R*-estimator of the shape matrix", *Journal of Signal Processing Systems*, July 2021.

[27] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir and and M. Rangaswamy, "Semiparametric Inference and Lower Bounds for Real Elliptically Symmetric Distributions", *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 164-177, 1 Jan.1, 2019.

[28] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir and and M. Rangaswamy, "Semiparametric CRB and Slepian-Bangs Formulas for Complex Elliptically Symmetric Distributions", *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5352-5364, 15 Oct.15, 2019.

Semiparametric rank-based R-estimators

- What happens for all the other cases in which the observation vector $\mathbf{x}$ does not admit a stochastic representation as the one of the ES-data?

- How could we define *multidimensional ranks* ? No ordering can be established in multivariate case...

- More specifically, how could we define a multivariate version of the ranks, say $\mathbf{R}_i$, able to keep the two crucial properties:

  1. *Distribution-freeness* : the distribution of $\mathbf{R}_i$ does not depend of the true distribution P_0,
  2. *"Glivenko-Cantelli" behavior* : there exists a sort of empirical distribution, based on the multivariate ranks $\mathbf{R}_i$, able to converge to (some functional of) P_0 as $n \to \infty$.

A possible way out...

- In a recent series of works [29,30,31], the concept of **center-outward ranks** has been proposed.
- Center-outward ranks satisfy the two required properties of distribution-freeness and "Glivenko-Cantelli" behavior.
- Some preliminary applications have been proposed in the context of time series analysis.
- However, the general construction of general "center-outward" R-estimators is still an **open question**.

[29] M. Hallin, E. del Barrio, J. Cuesta-Albertos, C. Matrán. "Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach." *Annals of Statistics* 49 (2) 1139 - 1165, April 2021

[30] M. Hallin, D. Hlubinka, Š. Hudecová (2023) "Efficient Fully Distribution-Free Center-Outward Rank Tests for Multiple-Output Regression and MANOVA", *Journal of the American Statistical Association*, 118:543, 1923-1939

[31] M. Hallin, D. La Vecchia, H. Liu (2022) "Center-Outward R-Estimation for Semiparametric VARMA Models", *Journal of the American Statistical Association*, 117:538, 925-938

Concluding remarks

- Parametric estimation has been discussed in three different frameworks:
  1. *Parametric*: perfect knowledge of the data model,
  2. *Misspecified*: wrong assumption on the data model.
  3. *Semiparaemtric*: no or partial assumption on the data model, functional nuisance parameters are allowed.

- For the tree cases, we investigate the efficiency:
  1. *Parametric*: ML or One Step (OS) estimators and CRB,
  2. *Misspecified*: MML estimator and MCRB.
  3. *Semiparametric*: OS semiparametric estimator and SCRB.

One Step, rank-based, semiparametric estimators are able to reconcile robustness (distribution-freeness) and (semiparametric) efficiency!

Future works

- The construction of One Step semiparametric estimators is a challenging and largely unexplored problem.

- We saw that two approaches are possible:
  1. Use a non-parametric estimation of the nuisance function,
  2. Rank-based approach.

- *Problem with 1)*: Derivation of $\sqrt{n}$-consistent preliminary estimators of the nuisance function.

- *Problem with 2)*:
  1. A rank-based semiparametric estimator has been constructed only in the case of the ES data (thanks to their stochastic representation).
  2. The use of the center-outward ranks remains largely unexplored.

Thanks for your attention!

Laboratoire Signaux & Systèmes

Backup slides

Semiparametric models: Missing data

- Let $\mathbf{z} \triangleq (\mathbf{x}^T, \mathbf{y}^T)^T$ be a *complete* dataset, where:
  - $\mathbf{x}$ is the *observed* (available) dataset.
  - $\mathbf{y}$ is the *unobservable* (missing) dataset.

- **Problem**: Estimate $\boldsymbol{\theta} \in \Theta$ from the observed dataset $\mathbf{x}$ when the pdf p_Y of the missing data $\mathbf{y}$ is unknown.

- The pdf p_X of the observed dataset can be expressed as:

$$p_X(\mathbf{x}|\boldsymbol{\theta}) = \int_{\mathcal{Y}} p_{X,Y}(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) d\mathbf{y} = \int_{\mathcal{Y}} p_{X|Y}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) p_Y(\mathbf{y}) d\mathbf{y}.$$

- The set of all the pdfs of the observed dataset $\mathbf{x}$ is a *semiparametric mixture model* of the form :

$$\mathcal{P}_{\boldsymbol{\theta}, p_Z} \triangleq \{p_X | p_X(\mathbf{x}|\boldsymbol{\theta}, p_Y), \boldsymbol{\theta} \in \Theta, p_Y \in \mathcal{K}\}.$$

Semiparametric models: Non-linear regression

- Let us consider the general non-linear regression model:

$$\mathbf{x} = f(\mathbf{z}, \boldsymbol{\theta}) + \boldsymbol{\epsilon},$$

  - $\boldsymbol{\theta} \in \Theta$: parameter vector to be estimated,
  - $f \in \mathcal{F}$: possibly unknown non-linear function,
  - $\mathbf{z}$: random vector with possibly unknown pdf $p_Z \in \mathcal{K}$,
  - $\boldsymbol{\epsilon}$: random noise with possibly unknown pdf $p_\epsilon \in \mathcal{E}$

- The set of all pdfs for $\mathbf{x}$ is a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\theta}, f, p_Z, p_\epsilon} \triangleq \{p_X(\mathbf{x}|\boldsymbol{\theta}, f, p_Z, p_\epsilon), \boldsymbol{\theta} \in \Theta, f \in \mathcal{F}, p_Z \in \mathcal{K}, p_\epsilon \in \mathcal{E}\}.$$

- This model is a general form of a *semiparametric regression model*.

Semiparametric models: Autoregressive processes

- Consider the $\mathrm{AR}(p)$ process:

$$x_n = \sum_{i=1}^{p} \theta_i x_{n-i} + w_n, \quad n \in (-\infty, \infty)$$

  - $\boldsymbol{\theta} \triangleq [\theta_1, \ldots, \theta_p]$: parameter vector to be estimated.
  - w_n: i.i.d. innovations with unknown pdf $p_w \in \mathcal{W}$,

- Let $\mathbf{x} \in \mathbb{R}^N$ a vector of N observations from an $\mathrm{AR}(p)$.

- The set of all possible pdfs for $\mathbf{x} \in \mathbb{R}^N$ is a semiparametric model:

$$\mathcal{P}_{\boldsymbol{\theta}, p_w} \triangleq \{p_X | p_X(\mathbf{x}|\boldsymbol{\theta}, p_w), \boldsymbol{\theta} \in \Theta, p_w \in \mathcal{W}\}.$$

Hellinger differentiability

- Let $p_X(\mathbf{x}|\boldsymbol{\theta})$ be a parametric pdf with $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$.

- We indicate with $u_{\boldsymbol{\theta}}(\mathbf{x})$ the following parametric map:

$$\begin{aligned} u_{\boldsymbol{\theta}} : \Theta &\to L_2(\mu) \\ \boldsymbol{\theta} &\mapsto u_{\boldsymbol{\theta}}(\mathbf{x}) \triangleq \sqrt{p_X(\mathbf{x}|\boldsymbol{\theta})}, \end{aligned}$$

- Then $u_{\boldsymbol{\theta}}$ is Hellinger (Fréchet) differentiable in $\boldsymbol{\theta}_0$ if there exists a vector $\dot{\mathbf{u}}_{\boldsymbol{\theta}_0} \equiv \dot{\mathbf{u}}_{\boldsymbol{\theta}_0}(\mathbf{x})$ such that:

$$||u_{\boldsymbol{\theta}_0+\mathbf{h}} - u_{\boldsymbol{\theta}_0} - \dot{\mathbf{u}}_{\boldsymbol{\theta}_0}^T \mathbf{h}|| = o(\textstyle\sum_i h_i^2), \quad \mathbf{h} \to 0,$$

  where $||u_{\boldsymbol{\theta}}||^2 = \langle u_{\boldsymbol{\theta}}, u_{\boldsymbol{\theta}} \rangle = \int u_{\boldsymbol{\theta}}^2 d\mu$.

- $\dot{\mathbf{u}}_{\boldsymbol{\theta}_0} \equiv \dot{\mathbf{u}}_{\boldsymbol{\theta}_0}(\mathbf{x})$ is the Hellinger derivative of $u_{\boldsymbol{\theta}}$ in $\boldsymbol{\theta}_0$.

Hellinger derivative and score vector

- Recall that the score vector of $p_X(\mathbf{x}|\boldsymbol{\theta})$ in $\boldsymbol{\theta}$ is defined as:

$$\mathbf{s}_{\boldsymbol{\theta}} \triangleq \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}).$$

- If for all $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ [BKRW, Sec. 2.1, Prep. 1]:
  - $p_X(\mathbf{x}|\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$ for almost all $\mathbf{x}$,
  - $\left(\sum_i [\mathbf{s}_{\boldsymbol{\theta}}]_i^2\right)^{1/2} \in L_2(P_{\boldsymbol{\theta}})$,
  - The FIM $\mathbf{I}(\boldsymbol{\theta}) \triangleq \int \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})\mathbf{s}_{\boldsymbol{\theta}}^T(\mathbf{x}) p_X(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}$ is non-singular and continuous in $\boldsymbol{\theta}$,

  then [BKRW, Sec. 2.1], we have that:

$$\dot{\mathbf{u}}_{\boldsymbol{\theta}} = \frac{1}{2} u_{\boldsymbol{\theta}} \mathbf{s}_{\boldsymbol{\theta}}.$$

Parametric submodels (1/3)

- Let us recall the semiparametric model:

$$\mathcal{P}_{\boldsymbol{\theta},h} \triangleq \{p_X(\mathbf{x}|\boldsymbol{\theta},h), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q, h \in \mathcal{H}\}.$$

- The *i-th parametric submodel* of $\mathcal{P}_{\boldsymbol{\theta},h}$ is defined as:

$$\mathcal{P}_{\boldsymbol{\theta},\nu_i} = \{p_X(\mathbf{x}|\boldsymbol{\theta},\nu_i(\cdot,\boldsymbol{\eta})), \boldsymbol{\theta} \in \Theta, \boldsymbol{\eta} \in \Gamma_i\},$$

where:

$$\begin{aligned} \nu_i : \Gamma_i &\to \mathcal{H} \\ \boldsymbol{\eta} &\mapsto \nu_i(\cdot,\boldsymbol{\eta}), \end{aligned}$$

- The function $\nu_i \in \mathcal{H}$ is a *known* function parametrized by a vector of *unknown* parameters.

Parametric submodels (2/3)

- Denote the "true semiparametric vector" and the related true pdf as $(\boldsymbol{\theta}_0^T, h_0)^T$ and $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\theta}_0, h_0)$, respectively.

- For every $i \in \mathcal{I}$, the *i-th parametric submodel*:

$$\mathcal{P}_{\boldsymbol{\theta},\nu_i} = \left\{ p_X(\mathbf{x}|\boldsymbol{\theta}, \nu_i(\cdot, \boldsymbol{\eta}), \boldsymbol{\theta} \in \Theta, \boldsymbol{\eta} \in \Gamma_i \right\},$$

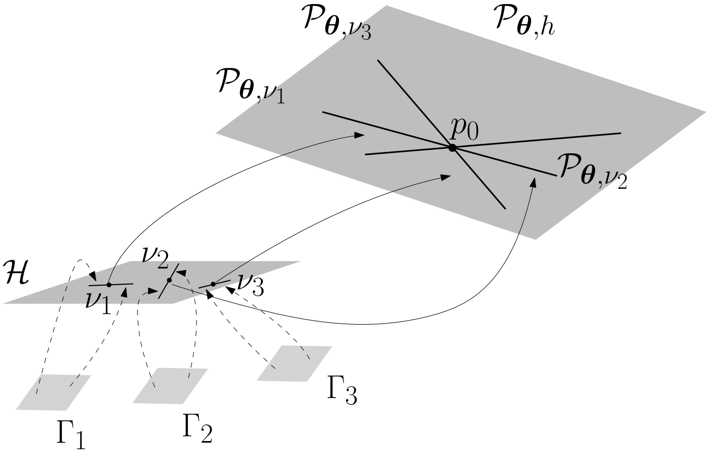
has to satisfy the following three conditions :

C0) $\nu_i : \Gamma_i \to \mathcal{H}$ is a smooth parametric map,

C1) $\mathcal{P}_{\boldsymbol{\theta},\nu_i} \subseteq \mathcal{P}_{\boldsymbol{\theta},g}$,

C2) $p_0(\mathbf{x}) \in \mathcal{P}_{\boldsymbol{\theta},\nu_i}$, i.e. there exists a vector $(\boldsymbol{\theta}_0^T, \boldsymbol{\eta}_0^T)^T$ such that $p_X(\mathbf{x}|\boldsymbol{\theta}_0, \nu_i(\cdot, \boldsymbol{\eta}_0)) = p_X(\mathbf{x}|\boldsymbol{\theta}_0, h_0) \triangleq p_0(\mathbf{x})$.

Parametric submodels (3/3)



Semiparametric nuisance tangent space

- For every parametric submodel:

$$\mathcal{P}_{\boldsymbol{\theta},\nu_i} = \{p_X(\mathbf{x}|\boldsymbol{\theta},\nu_i(\cdot,\boldsymbol{\eta})), \boldsymbol{\theta} \in \Theta, \boldsymbol{\eta} \in \Gamma_i\},$$

we have a relevant nuisance tangent space:

$$\mathcal{T}_{\boldsymbol{\eta}_{0,i}} \triangleq \{\mathbf{t}_i | \mathbf{t}_i = \mathbf{A}_i \mathbf{s}_{\boldsymbol{\eta}_{0,i}} : \mathbf{A}_i \text{ is any matrix in } \mathbb{R}^{q \times d_i}\},$$

where $\mathbf{s}_{\boldsymbol{\eta}_{0,i}} \triangleq \nabla_{\boldsymbol{\eta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, \nu_i(\cdot,\boldsymbol{\eta}_0))$.

- The **semiparametric nuisance tangent space** is defined as:[32]

$$\mathcal{T}_{h_0} \triangleq \overline{\bigcup_{\{\mathcal{P}_{\boldsymbol{\theta},\nu_i}\}_{i\in\mathcal{I}}} \mathcal{T}_{\boldsymbol{\eta}_{0,i}}}$$

[32] The closure $\overline{\mathcal{A}}$ of a set $\mathcal{A}$ is defined as the smallest closed set that contains $\mathcal{A}$, or equivalently, as the set of all elements in $\mathcal{A}$ together with all the limit points of $\mathcal{A}$.

Center-outward distribution function

- Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the set of n i.i.d. observations, such that $\mathbb{R}^m \supseteq \mathcal{X} \ni \mathbf{x}_i \sim P_0, \ \forall i$, where P_0 is the distribution function.

- Let us express n as:

$$n = n_R n_S + n_0, \ n_R, n_S, n_0 \in \mathbb{N}, \ 0 \leq n_0 < \min(n_R, n_S).$$

- Consider a regular grid $\mathcal{G}$ of $n_R n_S$ points in the unit ball $\mathcal{S}^m$ obtained as the intersection between:
  - a "regular" n_S-tuple $(\mathbf{u}_1, \ldots, \mathbf{u}_{n_S})$ of unit vectors,
  - the n_R hyperspheres centered at the origin with radii $\frac{1}{n_R+1}, \frac{2}{n_R+1}, \ldots, \frac{n_R}{n_R+1}$,

  along with n_0 copies of the origin.

Center-outward distribution function

- We define $\mathbf{F}^{(n)}(\mathbf{x}_i),\ i = 1, \ldots, n$ as the solution of an optimal coupling problem between the observations and $\mathcal{G}$.

- Let $\mathcal{T}$ denote the set of all possible bijective mapping between $\{\mathbf{x}_i\}_{i=1}^n$ and the n points in $\mathcal{G}$.

- **Definition**: The empirical *center-outward* distribution is the random mapping

$$\mathbf{F}^{(n)} : (\mathbf{x}_1, \ldots, \mathbf{x}_n) \mapsto (\mathbf{F}^{(n)}(\mathbf{x}_1), \ldots, \mathbf{F}^{(n)}(\mathbf{x}_n))$$

satisfying

$$\sum\nolimits_{i=1}^{n} ||\mathbf{x}_i - \mathbf{F}^{(n)}(\mathbf{x}_i)||^2 = \min\nolimits_{T \in \mathcal{T}} \sum_{i=1}^{n} ||\mathbf{x}_i - T(\mathbf{x}_i)||^2.$$

where the set $\{\mathbf{F}^{(n)}(\mathbf{x}_i)\}_{i=1}^n$ coincides with the n points in $\mathcal{G}$.

Center-outward ranks

- From the definition of center-outward empirical distribution $\mathbf{F}^{(n)}$, we can introduce the *center-outward ranks* as:

$$R_i \triangleq (n_R + 1)||\mathbf{F}^{(n)}(\mathbf{x}_i)||.$$

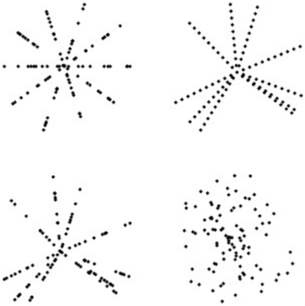


Figure 1: Four examples of possible grids. the image is taken from the master thesis of Veronika Roubinova *Center-outward ranks and signs and their applications in statistical tests*, Charles University, Prague,2023.