Laboratoire Signaux & Systèmes

Matched, mismatched and semiparametric inference in elliptical distributions

Stefano Fortunati

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

TéSA Scientific Seminar

Toulouse

Thursday, November 17

Outline of the presentation

Parametric models
- The CRB and the Maximum Likelihood (ML) estimator
- Parametric estimation in CES distributions

Misspecified models
- The MCRB and the Mismatched ML estimator
- Misspecified estimation in CES distributions

Semiparametric models
- The SCRB and the R-estimators
- Semiparametric inference in CES distributions

Introduction

- Let $\mathbf{x}_1, \ldots, \mathbf{x}_L$ be the set of L observations collected from a random experiment.

- We indicate with $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_L)$ their joint "true" probability density function (pdf).

- In point estimation, we are interested in evaluating some functional of $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_L)$, say $\nu(p_0)$.

- However, $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_L)$ is generally unknown, at least to some extent.

- The lack of *a priori* knowledge on $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_L)$ can be formalized in the concept of *statistical models*.

Parametric models

- In signal processing (SP) application, 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 $\boldsymbol{\theta}$:

$$\mathcal{P}_{\boldsymbol{\theta}} \triangleq \{p_X(\mathbf{x}_1, \ldots, \mathbf{x}_L | \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}.$$

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

$$\mathcal{P}_{\boldsymbol{\theta}} \ni p_X(\mathbf{x}_1, \ldots, \mathbf{x}_L | \boldsymbol{\theta}_0) = p_0(\mathbf{x}_1, \ldots, \mathbf{x}_L). \quad \text{(A1)}$$

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

The Maximum Likelihood (ML) estimator

- We suppose that our observations $\mathbf{x}_1, \dots, \mathbf{x}_L$ are *iid* with "true" distribution $p_0(\mathbf{x})$, i.e. $\mathbf{x}_l \sim p_0, \forall l$.

- The *Maximum Likelihood* (ML) estimator, defined on the parametric model $\mathcal{P}_{\boldsymbol{\theta}}$, is given by:

$$\widehat{\boldsymbol{\theta}}_{L,ML} \triangleq \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \prod_{l=1}^{L} p_X(\mathbf{x}_l|\boldsymbol{\theta}), \quad \mathbf{x}_l \sim p_0.$$

- The ML estimator is a cornerstone of the parametric estimation due to the following two optimality properties:
  1. *Consistency*,
  2. *Asymptotic Gaussianity and efficiency*.

- To well understand them, first we need to introduce the *Fisher Information Matrix* (FIM) $\mathbf{I}(\boldsymbol{\theta})$.

Fisher Information and Cramér-Rao Bound

- "Under some regularity conditions" [1], and under Assumption (A1), the FIM is defined as:

$$\mathbf{I}(\boldsymbol{\theta}) \triangleq E\left\{\nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}}^T \ln p_X(\mathbf{x}|\boldsymbol{\theta})\right\}$$
$$\triangleq -E\left\{\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \ln p_X(\mathbf{x}|\boldsymbol{\theta})\right\}, \quad \mathbf{x} \sim p_0.$$

Cramér-Rao Bound: Any *unbiased* estimator $\hat{\boldsymbol{\theta}}_L$ of $\boldsymbol{\theta}_0$, derived in $\mathcal{P}_{\boldsymbol{\theta}}$ from $\{\mathbf{x}_l \sim p_0\}_{l=1}^{L}$ iid observations, satisfies:

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

where the unbiasedness condition must hold, i.e. $\forall L \in \mathcal{N}$:

$$E_0\{\hat{\boldsymbol{\theta}}_L\} \triangleq \int \hat{\boldsymbol{\theta}}_L(\mathbf{x}_1, \ldots, \mathbf{x}_L) p_0(\mathbf{x}_1, \ldots, \mathbf{x}_L) d\mathbf{x}_1, \ldots, d\mathbf{x}_L = \boldsymbol{\theta}_0.$$

[1] 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 optimality of the ML estimator

- Why is the ML estimator so popular in applications?

Under Assumption (A1), the ML estimator $\widehat{\boldsymbol{\theta}}_{L,ML}$ is:

1. $\sqrt{L}$-*consistent*:

$$\sqrt{L}\left(\hat{\boldsymbol{\theta}}_{L,ML} - \boldsymbol{\theta}_0\right) = O_P(1).^2$$

2. *Asymptotically Gaussian and efficient*:

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

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

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

Covariance/scatter matrix estimation

- 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 **Complex Elliptically Symmetric (CES)** distributions.
- Thanks to their flexibility, CES distributions represent a reliable data model in many applications. [3]
- The complex Gaussian, Generalized Gaussian, K-distribution, complex t-distribution and all the compound-Gaussian distributions belong to the CES class.
- The CES model is particularly useful in applications with *impulsive noise* and/or *spiky data*.

[3] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex Elliptically Symmetric Distributions: Survey, New Results and Applications", *IEEE Trans. on Signal Processing*, vol. 60, no. 11, pp. 5597-5625, Nov. 2012.

CES distributions (1/2)

- A CES distributed random vector $\mathbf{x} \in \mathbb{C}^N$ admits a pdf:

$$p_X(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1} h((\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})) \triangleq CES_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h).$$

  - $h \in \mathcal{H}$, $h : \mathbb{R}_0^+ \to \mathbb{R}^+$ is the *density generator*,
  - $\boldsymbol{\mu} \in \mathbb{C}^N$ is the location vector,
  - $\boldsymbol{\Sigma} \in \mathcal{M}_N$ is the (full rank) scatter matrix.

- Note that $\boldsymbol{\Sigma}$ and h are not jointly identifiable:

$$CES_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h(t)) \equiv CES_N(\boldsymbol{\mu}, c\boldsymbol{\Sigma}, h(ct)), \ \forall c > 0.$$

- To avoid this identifiability problem, we introduce the *shape matrix* as a normalized version of $\boldsymbol{\Sigma}$:

$$\mathbf{V} \triangleq \boldsymbol{\Sigma}/s(\boldsymbol{\Sigma}).$$

CES distributions (2/2)

- Typical examples of *scale function* $s(\cdot)$ are:

$$s(\mathbf{\Sigma}) = [\mathbf{\Sigma}]_{11}, \quad s(\mathbf{\Sigma}) = \mathrm{tr}(\mathbf{\Sigma})/N \quad s(\mathbf{\Sigma}) = |\mathbf{\Sigma}|^{1/N}.$$

- Not that, *under finite second order moments*, if the scale $s(\mathbf{\Sigma}) = \mathrm{tr}(\mathbf{\Sigma})/N$ is adopted, we have that:

$$\mathbf{C} \triangleq E\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^H\} \equiv \mathbf{\Sigma} = \sigma^2 \mathbf{V},$$

  where:

  1. $\mathbf{C}$ is the covariance matrix of the CES-distributed vector $\mathbf{x} \sim CES_N(\boldsymbol{\mu}, \mathbf{\Sigma}, h)$,
  2. $\sigma^2 = \mathrm{tr}(\mathbf{\Sigma})/N = E\{\mathbf{x}^H \mathbf{x}\}/N$ is the statistical power of $\mathbf{x}$.

- Unless otherwise stated, in the following we always implicitly adopt the scale $s(\mathbf{\Sigma}) = \mathrm{tr}(\mathbf{\Sigma})/N$.

The complex t-distribution

- We suppose to collect $\{\mathbf{x}_l\}_{l=1}^{L}$, *zero mean*, iid observations that we assume to be t-distributed.

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

$$h_0(t) = \frac{1}{\pi^N} \frac{\Gamma(N+\lambda)}{\Gamma(\lambda)} \left(\frac{\lambda}{\eta}\right)^{\lambda} \left(\frac{\lambda}{\eta} + t\right),$$

  1. λ: *shape parameter* controlling the data non-Gaussianity,
  2. η: *scale parameter* controlling the data power $\sigma^2 = \frac{\lambda}{\eta(\lambda-1)}$.

- To guarantee the finitness of the second order moments (i.e. the existence of the covariance matrix), we need $\lambda > 1$.

- Note that for values of $\lambda \to 1$ the data are heavy-tailed, while for $\lambda \to \infty$ the data tends to be Gaussian.

The parametric t-model

- Under this t-assumption, the parametric model characterizing the random experiment is:

$$\mathcal{P}_{\boldsymbol{\theta}} \triangleq \left\{ p_X(\mathbf{x}|\boldsymbol{\theta}) = |\boldsymbol{\Sigma}|^{-1} h_{0,\lambda,\eta}\left(\mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x}\right), \boldsymbol{\theta} \in \Theta \right\}.$$

- The parameter space is defined as:

$$\boldsymbol{\theta} \in \Theta \triangleq \left\{ \boldsymbol{\theta} = (\mathrm{vec}(\mathbf{V})^T, \lambda, \eta)^T | \mathbf{V} = N\boldsymbol{\Sigma}/\mathrm{tr}(\boldsymbol{\Sigma}) \right\},$$

  where λ and η can be considered as nuisance parameter while the shape matrix $\mathbf{V}$ is the parameter of interest.

- Optimal inference in this t-model will require the derivation of the Joint ML estimator for the three parameter in $\boldsymbol{\theta} \in \Theta$.

The parametric t-model: ML estimator for $\mathbf{V}$

- Deriving the joint ML estimator for $\boldsymbol{\theta} \in \Theta$ is a prohibitive task! No closed form exists.

- Given λ and η, the ML estimator of $\mathbf{V}$ is given by the convergence point of the iterative procedure: [4]

$$\begin{cases} \widehat{\boldsymbol{\Sigma}}^{(k+1)} = \frac{N+\lambda}{L} \sum_{l=1}^{L} \frac{\mathbf{x}_l^H \mathbf{x}_l}{\mathbf{x}_l^H \left[\widehat{\boldsymbol{\Sigma}}^{(k)}\right]^{-1} \mathbf{x}_l + \lambda/\eta} \\ \widehat{\mathbf{V}}_{JML}^{(k+1)} \triangleq N\widehat{\boldsymbol{\Sigma}}^{(k+1)} / \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}^{(k+1)}) \end{cases},$$

where, as starting point, we use $\boldsymbol{\Sigma}^{(0)} = \mathbf{I}_N$.

- We substitute λ and η with two sub-optimal but consistent estimators derived using the Method of Moments. [5]

[4] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex Elliptically Symmetric Distributions: Survey, New Results and Applications", *IEEE Trans. on Signal Processing*, vol. 60, no. 11, pp. 5597-5625, Nov. 2012.

[5] S. Fortunati, F. Gini and M. Greco, "Matched, mismatched, and robust scatter matrix estimation and hypothesis testing in complex t-distributed data," *EURASIP J. Adv. Signal Process.* 2016, 123 (2016).

The parametric t-model: performance

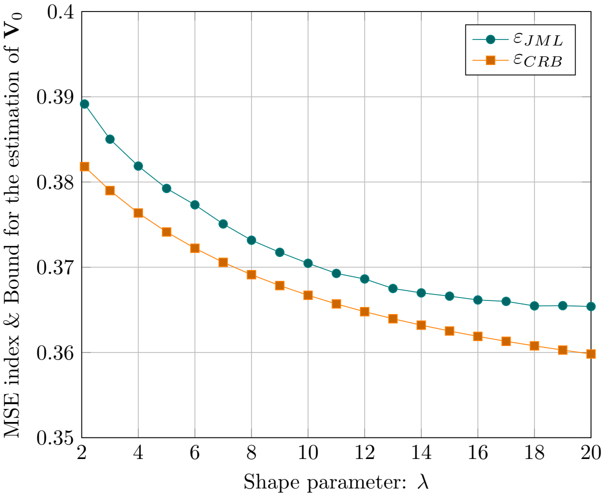
- Let $\boldsymbol{\theta}_0 = (\mathrm{vec}(\mathbf{V}_0), \lambda_0, \eta_0)^T$ be the true parameter vector.
- The constrained joint Cramér-Rao Bound $\mathrm{CRB}(\boldsymbol{\theta}_0)$ for $\boldsymbol{\theta} \in \Theta$ has been derived in [6] and it is not reported here.
- We compare the performance of the joint ML algorithm for the estimation of $\mathbf{V}_0$ in terms of *Mean Squared Error (MSE)*

$$\varepsilon_{JML} = ||E\{(\mathrm{vec}(\widehat{\mathbf{V}}_{JML}) - \mathrm{vec}(\mathbf{V}_0))(\mathrm{vec}(\widehat{\mathbf{V}}_{JML}) - \mathrm{vec}(\mathbf{V}_0))^H\}||_F.$$

- As performance bound, we plot: $\varepsilon_{CRB} = ||\mathrm{CRB}(\boldsymbol{\theta}_0)||_F$.
- Number of observations: **finite sample regime** $L = 5N$.

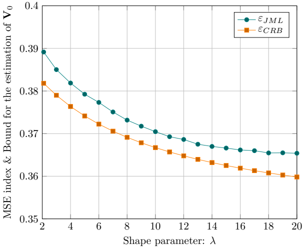
[6] S. Fortunati,F. Gini and M. Greco, "Matched, mismatched, and robust scatter matrix estimation and hypothesis testing in complex t-distributed data," *EURASIP J. Adv. Signal Process.* 2016, 123 (2016).

The parametric t-model: performance



- Under the "matched model" Assumption (A1), the MSE of $\widehat{\mathbf{V}}_{JML}$ estimator is close to the CRB.

The parametric t-model: performance



- However, the JML is not fully efficient, i.e. its MSE is not exactly equal to the CRB, for three reasons:
  1. Suboptimal MoM-estimators of λ and η,
  2. The constraint on $\boldsymbol{\Sigma}$ has been imposed in a suboptimal way, without relying on a constrained optimization procedure.
  3. We are not in the asymptotic regime. In fact, we assumed $L = 5N$, i.e. L does not tend to infinity for a given N.

- **What if the model assumption is wrong?**

Outline of the presentation

Model misspecification

- **Classical "matched" assumption**: the true data model and the model assumed to derive the estimation algorithm are the same, i.e. the model is correctly specified.
- All the results on the ML estimator and the CRB rely on this implicit assumption.
- However, much evidence from everyday practice shows that this assumption is often violated.
- **Model misspecification**: the assumed data model (i.e. the data pdf) differs from the true model.

Model misspecification

- There are two main reasons for model misspecification:
  1. An **imperfect knowledge** of the true data model that leads to a wrong specification of the data pdf.
  2. The true data 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 suboptimal procedure may lead to some degradation in the overall system performance.

Formal description of the misspecification

- Our observations $\{\mathbf{x}_l\}_{l=1}^{L}$ are *iid* with "true" distribution $p_0(\mathbf{x})$ belonging to a possibly non-parametric model $\mathcal{P}$.

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

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

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

- The classical "matched" assumption (A1) requires:

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

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

Formal description of the misspecification

- If the previous assumption is violated, the model $\mathcal{F}_\gamma$ is *misspecified*. Formally: [7]

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

  or, equivalently, that $p_0(\mathbf{x}) \in \mathcal{P} \not\subseteq \mathcal{F}_\gamma$.

- This misspecified scenario raises two main questions:
  1. How will the classical statistical properties of an estimator, e.g. *unbiasedness*, *consistency* and *efficiency*, change in this misspecified model framework?
  2. Is it still possible to derive lower bounds on the error covariance of any mismatched estimator?

[7] 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

- "Under some regularity conditions", there exist a unique interior point γ_0 of Γ, such that:

$$\gamma_0 \triangleq \underset{\gamma\in\Gamma}{\mathrm{argmin}}\left\{-E_0\{\ln f_X(\mathbf{x}|\gamma)\}\right\} = \underset{\gamma\in\Gamma}{\mathrm{argmin}}\left\{D_\gamma(p_0 \parallel f_X)\right\},$$

  where $E_0\{g(\mathbf{x})\} \triangleq \int g(\mathbf{x})p_0(\mathbf{x})d\mathbf{x}$ and

$$D_\gamma(p_0 \parallel f_X) \triangleq \int \ln\left(\frac{p_0(\mathbf{x})}{f_X(\mathbf{x}|\gamma)}\right)p_0(\mathbf{x})d\mathbf{x}$$

  is the **Kullback-Leibler divergence** (KLD) between the true pdf and the assumed pdf.

- The *pseudo-true parameter vector* γ_0 is the point the minimizes the KLD between the true and the assumed pdfs.

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 $\{\mathbf{x}_l \sim p_0\}_{l=1}^{L}$ *iid* observations, let's build an estimator $\hat{\gamma}_L$ assuming the possibly misspecified model $\mathcal{F}_\gamma$.

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

$$E_0\{\hat{\gamma}_L\} \triangleq \int \hat{\gamma}_L(\mathbf{x}_1, \ldots, \mathbf{x}_L) p_0(\mathbf{x}_1, \ldots, \mathbf{x}_L) d\mathbf{x}_1, \ldots, d\mathbf{x}_L = \gamma_0.$$

Misspecified CRB: Any *MS-unbiased* estimator $\hat{\gamma}_L$ of γ_0, derived in $\mathcal{F}_\gamma$ from $\{\mathbf{x}_l \sim p_0\}_{l=1}^{L}$ iid observations, satisfies: [8,9,10]

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

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

[9] 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.

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

The Mismatched ML estimator (MML)

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

$$\widehat{\gamma}_{L,MML} \triangleq \underset{\gamma \in \Gamma}{\text{argmax}} \prod_{l=1}^{L} f_X(\mathbf{x}_l|\gamma), \quad \mathbf{x}_l \sim p_0.$$

Properties: the MML estimator $\widehat{\gamma}_{L,MML}$ is: [11,12]

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

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

2. *Asymptotically Gaussian and MS-efficient*:

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

[11] 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

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

A common misspecified scenario in CES data

- Our *iid* observations $\{\mathbf{x}_l \sim p_0\}_{l=1}^{L}$ are CES-distributed, that is $p_0 \sim CES_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0, h_0)$.

- The "true" density generator is supposed to be the one of the t-distribution: $h_{0,\lambda_0,\eta_0}(t) = \frac{1}{\pi^N} \frac{\Gamma(N+\lambda_0)}{\Gamma(\lambda_0)} \left(\frac{\lambda_0}{\eta_0}\right)^{\lambda_0} \left(\frac{\lambda_0}{\eta_0} + t\right)$.

- The practitioner decides to built a ML estimator for $\boldsymbol{\Sigma}_0$ on the misspecified Gaussian model $\mathcal{F}_\gamma$, such that:

$$\mathcal{F}_\gamma = \left\{ f_X(\mathbf{x}|\gamma) = |\boldsymbol{\Sigma}|^{-1} g_{0,\sigma_X^2}\left(\mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x}\right), \gamma \in \Gamma \right\}.$$

where $g_{0,\sigma_X^2}(t) = (\pi \sigma_X^2)^{-N} \exp\left(-t/\sigma_X^2\right)$ and

$$\gamma \in \Gamma \triangleq \left\{ \gamma = (\mathrm{vec}(\mathbf{V})^T, \sigma_X^2)^T | \mathbf{V} = N\boldsymbol{\Sigma}/\mathrm{tr}(\boldsymbol{\Sigma}) \right\},$$

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

Misspecified scatter matrix estimation

- Let's recall the Sample Covariance Matrix, i.e. the ML estimator under Gaussian assumption as:

$$\mathrm{SCM} \triangleq \frac{1}{L}\sum_{l=1}^{L} \mathbf{x}_l \mathbf{x}_l^H .$$

- The Mismatched ML (MML) estimator can be derived as:

$$\begin{cases} \widehat{\mathbf{V}}_{MML} = \frac{N}{\mathrm{tr}(\mathrm{SCM})}\mathrm{SCM} \\ \hat{\sigma}_X^2 = \frac{1}{N \cdot L}\sum_{l=1}^{L} \mathbf{x}_l^H \widehat{\mathbf{V}}_{MML}^{-1} \mathbf{x}_l \end{cases} .$$

- The practitioner should now answer the following questions:
  1. Is $\widehat{\mathbf{V}}_{MML}$ a MS-consistent estimator for $\mathbf{V}_0 = N\boldsymbol{\Sigma}_0/\mathrm{tr}(\boldsymbol{\Sigma}_0)$?
  2. Is it efficient wrt the MCRB?
  3. Is its performance loss wrt the matched case acceptable?

Misspecified scatter matrix estimation

- To answer the first question, we need to evaluate the pseudo-true parameter vector

$$\gamma_0 \triangleq \underset{\gamma \in \Gamma}{\operatorname{argmin}} \{ D_\gamma(p_0 \parallel f_X) \},$$

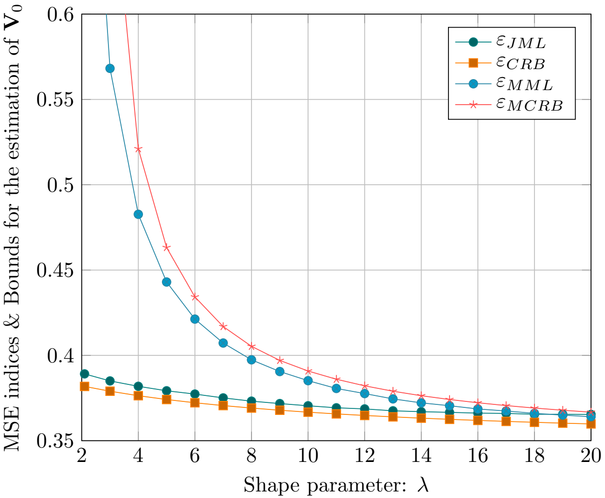
where p_0 is the t-distribution and f_X is the Gaussian one.

- It can be shown [13] that $\gamma_0 = (\mathrm{vec}(\mathbf{V}_0)^T, \sigma_0^2)^T$ then:
  1. $\sqrt{L}\left(\widehat{\mathbf{V}}_{MML} - \mathbf{V}_0\right) = O_P(1)$,
  2. $\sqrt{L}\left(\widehat{\sigma}_X^2 - \sigma_0^2\right) = O_P(1)$, where $\sigma_0^2 = \frac{\lambda_0}{\eta_0(\lambda_0 - 1)}$.

- The practitioner can use $\widehat{\mathbf{V}}_{MML}$ since it converge to the true shape matrix!

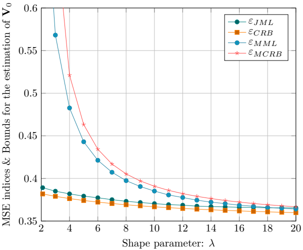
[13] S. Fortunati, F. Gini, M. S. Greco, "The Misspecified Cramér-Rao Bound and its Application to the Scatter Matrix estimation in Complex Elliptically Symmetric distributions," IEEE Trans. Signal Processing, vol. 64, no. 9, pp. 2387-2399, 2016.

Misspecified estimation performance: MSE



- Comparison among the MSE of the "matched" $\widehat{\mathbf{V}}_{JML}$ and $\widehat{\mathbf{V}}_{MML}$, and the related performance bounds, CRB and MCRB.

Misspecified estimation performance: MSE



- For small values of λ (highly non-Gaussian data), the estimation losses due to model mismatching rapidly increase!
    1. When $\lambda \to \infty$ (the data tend to be Gaussian distributed), the MSE of $\widehat{\mathbf{V}}_{MML}$ and of $\widehat{\mathbf{V}}_{JML}$ coincides: *no misspecification*.
    2. The MSE of $\widehat{\mathbf{V}}_{MML}$ is slightly below the MCRB because of the residual bias.

- **How can we overcome the misspecification?**

Outline of the presentation

Parametric models

- The CRB and the Maximum Likelihood (ML) estimator
- Parametric estimation in CES distributions

Misspecified models

- The MCRB and the Mismatched ML estimator
- Misspecified estimation in CES distributions

Semiparametric models

- The SCRB and the R-estimators
- Semiparametric inference in CES distributions

Semiparametric models

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

$$\mathcal{P}_{\boldsymbol{\theta},h} \triangleq \left\{ p_X(\mathbf{x}_1, \ldots, \mathbf{x}_L | \boldsymbol{\theta}, h), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q, h \in \mathcal{H} \right\}.$$

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

- Most of the SP inference problems can be cast in the semiparametric framework:

    1. Inference in CES distributions (as we will see here),
    2. Estimation with missing data,
    3. Non-linear regression and inverse problems,
    4. Time series analysis, ...

CES distributions as semiparametric model

- The CES distributions family is a perfect example of *semiparametric model*.
- The (zero-mean and *iid*) CES semiparametric model can be obtained from the parametric one by relaxing the *unrealistic assumption* on the a-priori knowledge of the density generator:

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

where the parameter of interest is

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

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

Inference in semiparametric model

- Semiparametric inference requires sophisticated tools in functional analysis and asymptotic statistics (e.g. the *Hájek–Le Cam convolution theorem*).
- Here, only a very short and non-exhaustive introduction will be provided aiming at highlighting some **crucial results**.
- As rigorously discussed in [14], any semiparametric infernce scheme is based on the following key ingredients:
  1. The score vector of the parameter of interest $\mathbf{s}(\mathbf{x};\boldsymbol{\theta},h)$,
  2. The *nuisance tangent space* $\mathcal{T}_h$,
  3. The *efficient score vector* $\bar{\mathbf{s}}(\mathbf{x};\boldsymbol{\theta},h)$.

[14] 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.

The basic ingredients

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

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

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

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

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

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

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

A lower bound in semiparametric estimation

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

- The class of *Regular and Asymptotically Linear (RAL) estimators* is defined as:
  1. $\sqrt{L}$-consistent: $\sqrt{L}\left(\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}_0\right) = O_P(1)$,
  2. Asymptotically normal: $\sqrt{L}(\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}_0) \underset{L\to\infty}{\overset{d}{\sim}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi}(\boldsymbol{\theta}_0, h_0))$.

- The ML and all the robust estimators belong to this class.

Semiparametric CRB (SCRB): Any *RAL* estimator $\hat{\boldsymbol{\theta}}_L$ of $\boldsymbol{\theta}_0$, derived in $\mathcal{P}_{\boldsymbol{\theta},h}$ from $\{\mathbf{x}_l \sim\}_{l=1}^L$ iid observations, satisfies: [15]

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

[15] 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.

Efficient semiparametric estimators

- Let us focus on the efficient estimation of the parameter of interest $\boldsymbol{\theta} \in \Theta$ in the presence of the unknown function $h \in \mathcal{H}$.
- Clearly, Maximum Likelihood estimation is not an option.
- Is there any other "optimal" procedure for deriving asymptotically efficient estimates other then the ML one?
- The answer is positive and it is given by the semiparametric *rank-based (R-) Le Cam's "one step" estimators.* [16,17,18]

[16] L. Le Cam, "Locally asymptotically normal families of distributions," *University of California Publications Statist.*, vol. 3, 1960, pp. 37-98.

[17] 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.

[18] M. Hallin, B. J. M. Werker, "Semi-parametric efficiency, distribution-freeness and invariance," *Bernoulli*, vol. 9, no. 1, pp. 137-165, 2003.

Semiparametric efficient R-estimators

- Let $\{\mathbf{x}_l \sim p_0(\mathbf{x}; \boldsymbol{\theta}_0, h_0)\}_{l=1}^{L}$ be a set of iid observations.

- In the seminal paper [19], a *rank-based* (R-) class of one step estimators has been proposed:

$$\hat{\boldsymbol{\theta}}_{L,R} = \hat{\boldsymbol{\theta}}_L^\star + L^{-1/2} \widehat{\boldsymbol{\Upsilon}}_{\hat{\boldsymbol{\theta}}_L^\star}^{-1} \widetilde{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\theta}}_L^\star}.$$

  - $\boldsymbol{\theta}_L^\star$ is a sub-optimal (consistent, not efficient) estimator of $\boldsymbol{\theta}$,
  - $\widehat{\boldsymbol{\Upsilon}}_{\hat{\boldsymbol{\theta}}_L^\star}$ is a *rank-based*, $\sqrt{L}$-consistent estimator of the *efficient information matric* $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)$,
  - $\widetilde{\boldsymbol{\Delta}}_{\boldsymbol{\theta}_L^\star} \triangleq \sum_{l=1}^{L} \widetilde{\varphi}(\mathbf{x}_l, \boldsymbol{\theta}_L^\star)$, where $\widetilde{\varphi}$ is a distributionally-free, *rank-based* approximation of the *efficient score vector*.

- No non-parametric estimator $\hat{h}_L$ of $h \in \mathcal{H}$ is required!

[19] M. Hallin, B. J. M. Werker, "Semi-parametric efficiency, distribution-freeness and invariance," *Bernoulli*, vol. 9, no. 1, pp. 137-165, 2003.

Semiparametric efficient R-estimators

- An R-estimator built on $\mathcal{P}_{\boldsymbol{\theta},h}$ satisfies the same optimality properties of the ML estimator built on $\mathcal{P}_{\boldsymbol{\theta}}$! [20]

Under any possible $h \in \mathcal{H}$, the R-estimator $\hat{\boldsymbol{\theta}}_{L,R}$ is:

1. $\sqrt{L}$-*consistent*:

$$\sqrt{L}\left(\hat{\boldsymbol{\theta}}_{L,R} - \boldsymbol{\theta}_0\right) = O_P(1).$$

2. *Asymptotically Gaussian and "efficient"*:

$$\sqrt{L}\left(\hat{\boldsymbol{\theta}}_{L,R} - \boldsymbol{\theta}_0\right) \underset{L\to\infty}{\overset{d}{\sim}} \mathcal{N}(\mathbf{0}, \bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)^{-1}) = \mathcal{N}(\mathbf{0}, \mathrm{SCRB}(\boldsymbol{\theta}_0|h_0)).$$

- **Note:** A classical alternative to R-estimators are the robust M-estimators. However, **the M-estimators are not efficient!**

[20] M. Hallin, B. J. M. Werker, "Semi-parametric efficiency, distribution-freeness and invariance," *Bernoulli*, vol. 9, no. 1, pp. 137-165, 2003.

R-estimators and SCRB for the CES model

- The SCRB for the estimation of the scatter matrix $\overline{\mathbf{V}}$ in CES-distributed data can be found in [1], [2].

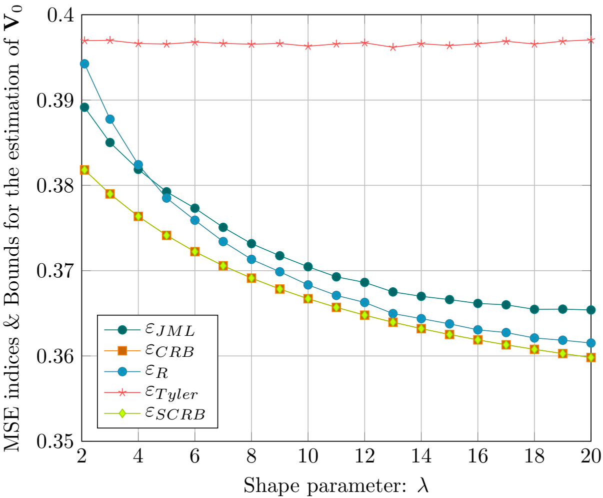
- The R-estimator in real elliptical data has been proposed by Hallin, Oja and Paindaveine in [21].

- In [3], [4], its generalization to the complex field with additional discussion about:
  - Discussion about optimal setting,
  - Robustness to outliers,
  - Extensive comparison with other robust estimators.

$$\widehat{\mathbf{V}}_R = \widehat{\mathbf{V}}^\star + \frac{1}{\hat{\alpha}}\left(\mathbf{W} - [\mathbf{W}]_{1,1}\,\widehat{\mathbf{V}}^\star\right).$$

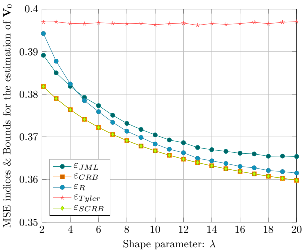
[21] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. Optimal R-estimation of shape," *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.

Semiparametric estimation performance: MSE



- Comparison among the MSE of $\widehat{\mathbf{V}}_R$ (R-estimator), $\widehat{\mathbf{V}}_{Ty}$ (Tyler M-estimator), $\widehat{\mathbf{V}}_{JML}$, the CRB and the SCRB.

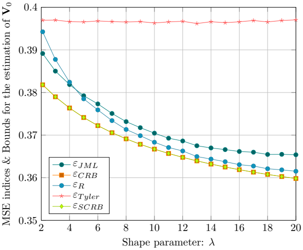
Semiparametric estimation performance: MSE



- Regarding the bounds:
  1. The CRB on $\mathbf{V}_0$ obtained for joint estimation of $(\mathrm{vec}(\mathbf{V}_0), \lambda_0, \eta_0)$ is equal to the SCRB on $\mathbf{V}_0$!
  2. Not knowing the parameters of the density generator is equivalent to not knowing its full functional form!
  3. This is true for all the CES distributions, not only for the *t*-distribution. [22]

[22] M. Hallin, D. Paindaveine, "Parametric and semiparametric inference for shape: the role of the scale functional," *Statistics & Decisions*, vol. 24, no. 3, 2006, pp. 327-350.

Semiparametric estimation performance: MSE



- Regarding the estimators:
  1. The popular Tyler's covariance estimator is not efficient.
  2. The *R*-estimator is (almost) semiparametrically efficient!
  3. The performance of the sub-optimal JML estimator of $(\mathrm{vec}(\mathbf{V}_0), \lambda_0, \eta_0)$ is even lower than the one of the *R*-estimator.
  4. The *R*-estimator can *almost* achieve the *parametric* CRB!

Concluding remarks

- In this talk, the estimation of the (normalized) covariance matrix has been addressed under three different frameworks:
  1. *Parametric*: perfect knowledge of the functional form of the density generator up to its parameters,
  2. *Misspecified*: wrong assumption on the density generator.
  3. *Semiparaemtric*: no assumption on the density generator.

- For the tree cases, we provided an efficient estimator and the related MSE bound:
  1. *Parametric*: JML estimator and CRB,
  2. *Misspecified*: MML estimator and MCRB.
  3. *Semiparaemtric*: R-estimator and SCRB.

An R-estimator is able to reach almost the same performance of the JML estimator without the need of any assumption of the density generator!

Future works

- Short-term perspectives:
  1. Rigorous analysis of the *almost efficiency* of the R-estimator (*Chernoff-Savage result*),
  2. Derivation of an R-estimator of the eigenspace of the scatter matrix.
- Medium-term perspectives:
  1. Semiparametric Mahalanobis distance,
  2. Applications to clustering and distance learning.
- Long-term perspectives:
  1. Semiparametric statistics and missing data,
  2. Applications to array processing, image reconstruction, ecc...

R-estimators and SCRB for the CES model

1. 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.

2. 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.

3. 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.

4. 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.

5. S. Fortunati, A. Renaux, F. Pascal, "Properties of a new R-estimator of shape matrices", *EUSIPCO 2020*, Amsterdam, the Netherlands, August 24-28, 2020.

6. S. Fortunati, A. Renaux, F. Pascal, "Robust Semiparametric DOA Estimation in non-Gaussian Environment", *2020 IEEE Radar Conference*, Florence, Italy, September 21-25, 2020

7. S. Fortunati, A. Renaux, F. Pascal, "Robust Semiparametric Joint Estimators of Location and Scatter in Elliptical Distributions", *IEEE International Workshop on Machine Learning for Signal Processing*, Aalto University, Espoo, Finland, September 21-24, 2020.

8. All the code about real and complex R-estimator is provided in my GitHub page.

Merci de votre 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:

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

Semiparametric "one step" estimators

- Let $\{\mathbf{x}_l\}_{l=1}^{L}$ be a set of iid observations, such that $\mathbf{x}_l \sim p_0(\mathbf{x}; \boldsymbol{\theta}_0, h_0) \in \mathcal{P}_{\boldsymbol{\theta},h}\ \forall l$.
- Let $\boldsymbol{\theta}_L^\star$ and $h_L^\star$ two sub-optimal (consistent but not efficient) estimators of $\boldsymbol{\theta}$ and of the nuisance function h.
- An asymptotically efficient, one step estimator $\hat{\boldsymbol{\theta}}_{L,OS}$ of $\boldsymbol{\theta}$ can then be obtained as:

$$\hat{\boldsymbol{\theta}}_{L,OS} = \boldsymbol{\theta}_L^\star + \frac{1}{L}\left[\mathbf{C}(\boldsymbol{\theta}_L^\star, h_L^\star)\right]^{-1} \sum\nolimits_{l=1}^{L} \bar{\mathbf{s}}(\mathbf{x}_l, \boldsymbol{\theta}_L^\star, h_L^\star),$$

$$\mathbf{C}(\boldsymbol{\theta}_L^\star, h_L^\star) \triangleq \frac{1}{L}\left[\sum\nolimits_{l=1}^{L} \bar{\mathbf{s}}(\mathbf{x}_l, \boldsymbol{\theta}_L^\star, h_L^\star)\bar{\mathbf{s}}(\mathbf{x}_l, \boldsymbol{\theta}_L^\star, h_L^\star)^T\right].$$

- The **critical requirement** is the non-parametric estimator $h_L^\star$! Can we avoid it?

Ranks (1/2)

- Let $\{x_l\}_{l=1}^{L}$ be a set of L continuous i.i.d. random variables with pdf p_X.

- Define the vector of the *order statistics* as

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

  whose entries

  $$x_{L(1)} < x_{L(2)} < \cdots < x_{L(L)}$$

  are the values of $\{x_l\}_{l=1}^{L}$ ordered in an ascending way.[23]

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

[23] Note that, since $x_l, \forall l$ are continuous random variable the equality occurs with probability 0.

Ranks (2/2)

- Let $\mathbf{r}_X \triangleq [r_1, \dots, r_L]^T \in \mathbb{N}^L$ be the vector collecting the ranks.
- Let $\mathcal{K}$ be the family of score functions $K : (0,1) \to \mathbb{R}$ that are continuous, square integrable and that can be expressed as the difference of two monotone increasing functions.

Let $\{x_l\}_{l=1}^{L}$ be a set of i.i.d. random variables s.t. $x_l \sim p_X, \forall l$. Then, we have:

1. The vectors $\mathbf{v}_X$ and $\mathbf{r}_X$ are independent,
2. Regardless the actual pdf p_X, the rank vector $\mathbf{r}_X$ is uniformly distributed on the set of all $L!$ permutations on $\{1, 2, \dots, L\}$,
3. For each $l = 1, \dots, L$, $K\left(\frac{r_l}{L+1}\right) = K(u_l) + o_P(1)$, where $K \in \mathcal{K}$ and $u_l \sim \mathcal{U}[0,1]$ is a random variable uniformly distributed in $(0,1)$.

A semiparametric efficient R-estimator (1/2)

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^\star) + \frac{1}{L\hat{\alpha}}\left[\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^H\right]^{-1} \times \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right)\mathrm{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H),$$

- $\{r_l^\star\}_{l=1}^L$ are the ranks of the r. v. $\hat{Q}_l^\star \triangleq \mathbf{x}_l^H[\widehat{\mathbf{V}}_1^\star]^{-1}\mathbf{x}_l$,
- $\hat{\mathbf{u}}_l^\star \triangleq \frac{[\widehat{\mathbf{V}}_1^\star]^{-1/2}\mathbf{x}_l}{\sqrt{\hat{Q}_l^\star}}$,
- $K_h(\cdot)$ is a *score* function based on $h \in \mathcal{H}$,
- $\hat{\alpha}$ is a data-dependent "cross-information" term,
- $\widehat{\mathbf{V}}_1^\star$ is a preliminary $\sqrt{L}$-consistent estimator of $\mathbf{V}_{1,0}$.

A semiparametric efficient R-estimator (2/2)

- The previous "vectorized" version of the R-estimator requires the unnecessary calculation of $N^2 \times N^2$ matrices with a consequent waste of computational resources.

- To overcome this problem, we have recently proposed the following "matrix" version of the same R-estimator: [24]

$$\widehat{\mathbf{V}}_{1,R} = \widehat{\mathbf{V}}_1^\star + \frac{1}{\hat{\alpha}}\left(\mathbf{W} - [\mathbf{W}]_{1,1}\,\widehat{\mathbf{V}}_1^\star\right)$$

where:

$$\mathbf{W} \triangleq L^{-1/2}(\widehat{\mathbf{V}}_1^\star)^{1/2}\mathbf{R}(\widehat{\mathbf{V}}_1^\star)^{1/2}.$$

$$\mathbf{R} \triangleq \frac{1}{\sqrt{L}}\sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right)\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H.$$

[24] 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," *MLSP Special Issue of the Journal of Signal Processing Systems*, 2021.

Tyler's M-estimator

- Tyler's M-estimator $\widehat{\mathbf{V}}_{Ty}$) is given by the convergence point ($k \to \infty$) of:

$$\left\{ \begin{array}{l} \widehat{\boldsymbol{\Sigma}}_{Ty}^{(k+1)} = \frac{N}{L} \sum_{l=1}^{L} [\hat{Q}_l^{(k)}]^{-1} \mathbf{x}_l \mathbf{x}_l^H \\ \widehat{\mathbf{V}}_{Ty}^{(k+1)} \triangleq N \widehat{\boldsymbol{\Sigma}}_{Ty}^{(k+1)} / \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{Ty}^{(k+1)}). \end{array} \right.$$

  where:

$$\hat{Q}_l^{(k)} = \mathbf{x}_l^H [\widehat{\mathbf{V}}_1^{(k)}]^{-1} \mathbf{x}_l,$$

- The estimator $\widehat{\mathbf{V}}_{Ty}$ is $\sqrt{L}$-consistent under any (unknown) density generator $h \in \mathcal{H}$.

- **It is not semiparametric efficient**.

Two possible score functions

- *van der Waerden* (Gaussian-based) score function:

$$K_{\mathbb{C}vdW}(u) = \Phi_G^{-1}(u), \quad u \in (0,1),$$

where Φ_G indicates the cdf of a Gamma-distributed random variable with parameters $(N, 1)$.

- t_ν-Student-based score function:

$$K_{\mathbb{C}t_\nu}(u) = \frac{N(2N+\nu)F_{2N,\nu}^{-1}(u)}{\nu + 2NF_{2N,\nu}^{-1}(u)}, \quad u \in (0,1),$$

where $F_{2N,\nu}(u)$ stands for the Fisher cdf with $2N$ and $\nu \in (0,\infty)$ degrees of freedom.

- We refer to [25] for a discussion on how to build score functions.

[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.

Finite-sample breakdown point (BP)

- Let $Z = \{\mathbf{z}_l\}_{l=1}^{L} \sim CES(\mathbf{0}, \mathbf{V}_1, h_0)$ be the "pure" GG data set.
- Let $Z_\varepsilon = \{\mathbf{z}_l\}_{l=1}^{L} \sim f_{Z_\varepsilon}$ be the ε-*contaminated* data set s.t.:

$$f_{Z_\varepsilon}(\mathbf{z}|\mathbf{V}_1, h_0, \varrho) = (1-\varepsilon) CES(\mathbf{0}, \mathbf{V}_1, h_0) + \varepsilon q_Z(\varrho),$$

  where $\varepsilon \in [0, 1/2]$ is a contamination parameter and the outlier $\tilde{\mathbf{z}} = \tau^{-1}\mathbf{u} \sim q_Z(\varrho)$ where $\tau \sim \mathrm{Gam}(\varrho, 1/\varrho)$.

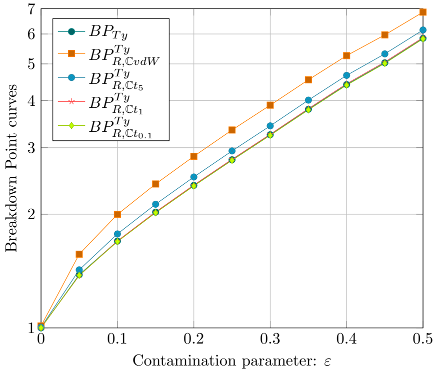
- Then the finite-sample BP curves can be evaluated as: [26]

$$BP(\varepsilon) \triangleq \max\left\{\lambda_1(\varepsilon), 1/\lambda_N(\varepsilon)\right\},$$

  where $\lambda_i(\varepsilon)$ is the i-th ordered eigenvalue of the matrix $[\widehat{\mathbf{V}}(Z)]^{-1}\widehat{\mathbf{V}}(Z_\varepsilon)$, s.t. $\lambda_1(\varepsilon) \geq \cdots \geq \lambda_N(\varepsilon)$.

[26] L. D¨umberg and D. E. Tyler, "On the breakdown properties of some multivariate M-functionals," *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 247–264, 2005.

Finite-sample breakdown point (BP)



- All the BP curves, related to the resulting *R*-estimator, remain close to the Tyler's one for every value of ε.
- t_ν-scores with a small value of ν lead to more robust estimators.

Empirical Influence Function (EIF)

- Let $Z = \{\mathbf{z}_l\}_{l=1}^{L} \sim CES(\mathbf{0}, \mathbf{V}_1, h_0)$ be the "pure" GG data set.
- Let $\tilde{\mathbf{z}} = \tau^{-1}\mathbf{u}$ be the outlier where $\tau \sim \mathrm{Gam}(\varrho, 1/\varrho)$. [27]
- Then the empirical influence function (EIF) can be evaluated as: [28]

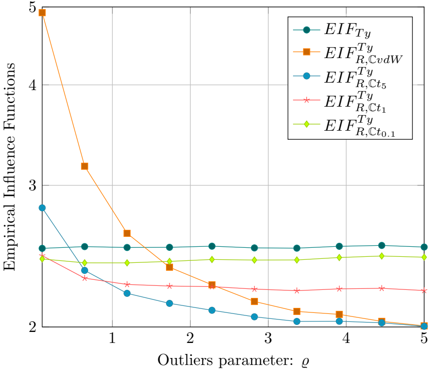
$$EIF \triangleq (L+1)||\widehat{\mathbf{V}}(Z) - \widehat{\mathbf{V}}(Z, \tilde{\mathbf{z}})||_F.$$

- the EIF gives us a measure of the impact that a single outlier $\tilde{\mathbf{z}}$ has on the shape matrix estimator $\widehat{\mathbf{V}}$ when it is added to the "pure" data set Z.

[27] Note that we can obtain "arbitrarily large" outlier by generating arbitrarily small values of $\tau \sim \mathrm{Gam}(\varrho, 1/\varrho)$ by letting $\varrho \to 0$.

[28] C. Croux, "Limit behavior of the empirical influence function of the median," *Statistics & Probability Letters*, vol. 37, no. 4, pp. 331 – 340, 1998.

Empirical Influence Function (EIF)



- The EIFs of the proposed R-estimator remain bounded and close to the one of the Tyler's estimator for arbitrarily large values of $||\tilde{\mathbf{z}}||$ ($\varrho \to 0$).