Semiparametric inference in elliptical distributions: lower bounds and efficient estimators.

Stefano Fortunati

L2S, CentraleSupélec
(Post-doc with the ANR MARGARITA project)

S^3 seminar
Friday, January 31th, 2020

Outline of the talk

Why semiparametric models?

CRB in parametric models with finite-dimensional nuisance parameters: classical approach

CRB in parametric models with finite-dimensional nuisance parameters: "Hilbert-space-based" approach

Extension to semiparametric models

SCRB for elliptical distributions

Efficient estimators of the shape matrix in elliptical distributions

Parametric models

- A parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ is defined as a set of pdfs that are parametrized by a finite-dimensional parameter vector $\boldsymbol{\theta}$:

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

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

- **Pros**: Parametric inference procedures are generally "simple" due to the finite dimensionality of $\boldsymbol{\theta}$.

- **Cons**: A parametric model could be too restrictive and a *misspecification problem*[1] may occur.

[1] 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.

Non-parametric models

- A non-parametric model $\mathcal{P}_p$ is a collection of pdfs possibly satisfying some functional constraints (i.e. *symmetry*):

$$\mathcal{P}_p \triangleq \{p_X(\mathbf{x}_1, \ldots, \mathbf{x}_M) \in \mathcal{K}\},$$

  where $\mathcal{K}$ is some constrained set of pdfs.

- **Pros**: The risk of model misspecification is minimized.

- **Cons**: In non-parametric inference we have to face with infinite-dimensional estimation problem.

- **Cons**: Non-parametric inference may be a prohibitive task due to the large amount of required data.

Semiparametric models

- A semiparametric model[2] $\mathcal{P}_{\boldsymbol{\theta},g}$ 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, $g \in \mathcal{L}$:

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

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

- **Pros**: All parametric signal models involving an unknown noise distribution are semiparametric models.

- **Cons**: Tools from functional analysis are needed.

[2] 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.

Examples: CES distributions

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

$$p_X(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c_{N,g} |\boldsymbol{\Sigma}|^{-1} g((\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})),$$

  - $c_{N,g}$ is a normalizing constant,
  - $g \in \mathcal{G}$, $g : \mathbb{R}_0^+ \to \mathbb{R}^+$ is the *density generator*,
  - $\boldsymbol{\mu} \in \mathbb{C}^N$ is the mean value,
  - $\boldsymbol{\Sigma} \in \mathcal{M}_N$ is the (full rank) scatter matrix.

- The set of all CES pdfs is a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\mu},\boldsymbol{\Sigma},g} \triangleq \left\{ p_X | p_X(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, g), \boldsymbol{\mu} \in \mathbb{C}^N, \boldsymbol{\Sigma} \in \mathcal{M}_N, g \in \mathcal{G} \right\}.$$

- This semiparametric model is a particular instance of the more general set of *semiparametric group models*.

Examples: 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 :

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

Examples: 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:

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

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

Examples: 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:

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

Outline of the talk

Why semiparametric models?

CRB in parametric models with finite-dimensional nuisance parameters: classical approach

CRB in parametric models with finite-dimensional nuisance parameters: "Hilbert-space-based" approach

Extension to semiparametric models

SCRB for elliptical distributions

Efficient estimators of the shape matrix in elliptical distributions

Score vectors in parametric models

- Let us consider the following *parametric model* involving a finite-dimensional vector of nuisance parameters:

$$\mathcal{P}_{\boldsymbol{\theta},\boldsymbol{\eta}} \triangleq \left\{ p_X(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{\eta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q, \boldsymbol{\eta} \in \Gamma \subseteq \mathbb{R}^d \right\},$$

  - $\boldsymbol{\theta} \in \Theta$: vector of the parameters of interest to be estimated,
  - $\boldsymbol{\eta} \in \Gamma$: vector of the (unknown) nuisance parameters.

- Denote with $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ the true value of $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\eta} \in \Gamma$, respectively. Then $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)$.

- **Score vectors** of the parametric model $\mathcal{P}_{\boldsymbol{\theta},\boldsymbol{\eta}}$ in $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$:

$$\mathbf{s}_{\boldsymbol{\theta}_0} \triangleq \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0,\boldsymbol{\eta}_0), \quad \mathbf{s}_{\boldsymbol{\eta}_0} \triangleq \nabla_{\boldsymbol{\eta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0,\boldsymbol{\eta}_0).$$

The Fisher Information Matrix (FIM)

- The FIM for the parametric model $\mathcal{P}_{\boldsymbol{\theta},\boldsymbol{\eta}}$ is given by:

$$\mathbf{I}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \triangleq \begin{pmatrix} E_0\left\{\mathbf{s}_{\boldsymbol{\theta}_0}\mathbf{s}_{\boldsymbol{\theta}_0}^T\right\} & E_0\left\{\mathbf{s}_{\boldsymbol{\theta}_0}\mathbf{s}_{\boldsymbol{\eta}_0}^T\right\} \\ E_0\left\{\mathbf{s}_{\boldsymbol{\eta}_0}\mathbf{s}_{\boldsymbol{\theta}_0}^T\right\} & E_0\left\{\mathbf{s}_{\boldsymbol{\eta}_0}\mathbf{s}_{\boldsymbol{\eta}_0}^T\right\} \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\theta}_0} & \mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\eta}_0} \\ \mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\eta}_0}^T & \mathbf{I}_{\boldsymbol{\eta}_0\boldsymbol{\eta}_0} \end{pmatrix},$$

  where $E_0\{h\} \triangleq \int h(\mathbf{x})p_0(\mathbf{x})d\mathbf{x}$.

- Let $\hat{\boldsymbol{\theta}}(\mathbf{x})$ be an *unbiased* estimator of $\boldsymbol{\theta}_0$: $E_0\{\hat{\boldsymbol{\theta}}(\mathbf{x})\} = \boldsymbol{\theta}_0$.

- How can we derive the CRB on the estimation of $\boldsymbol{\theta}_0$ in the presence of the unknown nuisance parameter vector $\boldsymbol{\eta}_0$?

Parametric CRB: classical approach

- The Cramér-Rao inequality provides us with a lower bound on the error covariance matrix of $\hat{\boldsymbol{\theta}}(\mathbf{x})$ when $\boldsymbol{\eta}_0$ is unknown:

$$E_0 \left\{ (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}_0)^T \right\} \geq \mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0).$$

- *Classical approach*: $\mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)$ can be obtained from the FIM using the Matrix Inversion Lemma:

$$\mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0) \triangleq \left(\mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\theta}_0} - \mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\eta}_0} \mathbf{I}_{\boldsymbol{\eta}_0\boldsymbol{\eta}_0}^{-1} \mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\eta}_0}^T \right)^{-1}.$$

- It is possible to obtain this same result by using a geometrical, "**Hilbert-space-based**" approach.

Outline of the talk

Why semiparametric models?

CRB in parametric models with finite-dimensional nuisance parameters: classical approach

CRB in parametric models with finite-dimensional nuisance parameters: "Hilbert-space-based" approach

Extension to semiparametric models

SCRB for elliptical distributions

Efficient estimators of the shape matrix in elliptical distributions

Three basic ingredients

- Let $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^q$ be a vector-valued function of the (random) observation vector $\mathbf{x}$.

- **Hilbert space** of all the zero-mean random functions:

$$\mathcal{H}^q = \{\mathbf{h} | \mathbf{h}(\mathbf{x}) \in \mathbb{R}^q, E\{\mathbf{h}\} = 0, E\{\mathbf{h}^T\mathbf{h}\} < \infty\},$$

- **Nuisance tangent space** of the parametric model $\mathcal{P}_{\boldsymbol{\theta},\boldsymbol{\eta}}$

$$\mathcal{T}_{\boldsymbol{\eta}_0} \triangleq \{\mathbf{C}\mathbf{s}_{\boldsymbol{\eta}_0} : \mathbf{C} \text{ is any matrix in } \mathbb{R}^{q\times d}\} \subset \mathcal{H}^q,$$

where $q = \dim(\boldsymbol{\theta}_0)$ and $r = \dim(\boldsymbol{\eta}_0)$.

- **Projection operator** of $\mathbf{h} \in \mathcal{H}^q$ onto $\mathcal{T}_{\boldsymbol{\eta}_0}$:

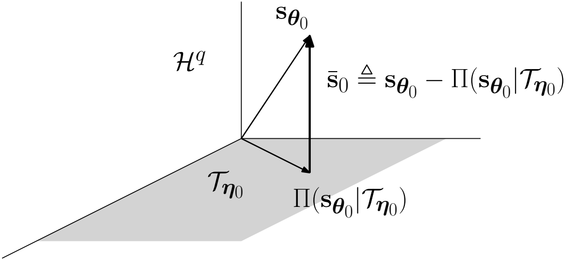
$$\Pi(\mathbf{h}|\mathcal{T}_{\boldsymbol{\eta}_0}) \triangleq E\{\mathbf{h}\mathbf{s}_{\boldsymbol{\eta}_0}^T\}\mathbf{I}_{\boldsymbol{\eta}_0\boldsymbol{\eta}_0}^{-1}\mathbf{s}_{\boldsymbol{\eta}_0}.$$

The efficient score vector

- Let us define the **efficient score vector** as:

$$\bar{\mathbf{s}}_0 \triangleq \mathbf{s}_{\boldsymbol{\theta}_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{\boldsymbol{\eta}_0})$$
$$= \mathbf{s}_{\boldsymbol{\theta}_0} - E\{\mathbf{s}_{\boldsymbol{\theta}_0}\mathbf{s}_{\boldsymbol{\eta}_0}^T\}\mathbf{I}_{\boldsymbol{\eta}_0\boldsymbol{\eta}_0}^{-1}\mathbf{s}_{\boldsymbol{\eta}_0}.$$

- $\bar{\mathbf{s}}_0$ is the residual of $\mathbf{s}_{\boldsymbol{\theta}_0}$ after projecting it onto the nuisance tangent space $\mathcal{T}_{\boldsymbol{\eta}_0}$.



Evaluation of the CRB using $\bar{\mathbf{s}}_0$

- Let us define the efficient FIM as:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0) \triangleq E_0\left\{\bar{\mathbf{s}}_0\bar{\mathbf{s}}_0^T\right\}.$$

- Through direct calculation, we get:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0) = \mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\theta}_0} - \mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\eta}_0}\mathbf{I}_{\boldsymbol{\eta}_0\boldsymbol{\eta}_0}^{-1}\mathbf{I}_{\boldsymbol{\theta}_0\boldsymbol{\eta}_0}^T.$$

- The inverse of $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)$ is exactly the $\mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)$ previously derived by means of the Matrix Inversion Lemma:

$$\left[E\left\{\bar{\mathbf{s}}_0\bar{\mathbf{s}}_0^T\right\}\right]^{-1} \triangleq \left[\bar{\mathbf{I}}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)\right]^{-1} = \mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0).$$

Outline of the talk

Why semiparametric models?

CRB in parametric models with finite-dimensional nuisance parameters: classical approach

CRB in parametric models with finite-dimensional nuisance parameters: "Hilbert-space-based" approach

Extension to semiparametric models

SCRB for elliptical distributions

Efficient estimators of the shape matrix in elliptical distributions

The three basic ingredients

- In summary, to derive the $\mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)$, we only need:

    1. The Hilbert space $\mathcal{H}^q$,
    2. The nuisance tangent space $\mathcal{T}_{\boldsymbol{\eta}_0} \subset \mathcal{H}^q$ of the parametric model $\mathcal{P}_{\boldsymbol{\theta},\boldsymbol{\eta}}$ at $\boldsymbol{\eta}_0$,
    3. The projection operator onto $\mathcal{T}_{\boldsymbol{\eta}_0}$: $\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{\boldsymbol{\eta}_0})$.

- Important fact: None of them require the finite dimensionality of the nuisance parameters.

- This alternative way to calculate the CRB can be extended to semiparametric models.

- To make this extension possible, we have to introduce the concept of *parametric submodel*.[3]

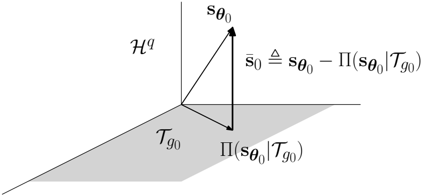
[3] Some additional details in the backup slides.

The projection operator $\Pi(\cdot|\mathcal{T}_{g_0})$

- Let $\mathcal{T}_{g_0} \subset \mathcal{H}^q$ the *semiparametric nuisance tangent space*.

- The *existence* and the *uniqueness* of the projection operator $\Pi(\cdot|\mathcal{T}_{g_0})$ is guaranteed by the Projection Theorem.

- The **semiparametric efficient score vector** for the estimation of $\boldsymbol{\theta}_0 \in \Theta$ in the presence of the nuisance function $g_0 \in \mathcal{L}$ is:
$$\bar{\mathbf{s}}_0 \triangleq \mathbf{s}_{\boldsymbol{\theta}_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{g_0}).$$



The Semiparametric CRB (SCRB)

Theorem 1. A lower bound on the *asymptotic* error covariance matrix of "any" robust estimator of $\boldsymbol{\theta}_0$ in the presence of the nuisance function $g_0 \in \mathcal{L}$ is given by: [4,5,6]

$$\mathrm{SCRB}(\boldsymbol{\theta}_0|g_0) = [\bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0)]^{-1},$$

where $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0) \triangleq E_0\{\bar{\mathbf{s}}_0\bar{\mathbf{s}}_0^T\}$ is the *semiparametric FIM* (SFIM) and:

$$\bar{\mathbf{s}}_0 \triangleq \mathbf{s}_{\boldsymbol{\theta}_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{g_0}).$$

[4] J. M. Begun, W. J. Hall, W.-M. Huang, and J. A. Wellner, "Information and asymptotic efficiency in parametric-nonparametric models", *The Annals of Statistics*, vol. 11, no. 2, pp. 432-452, 1983.

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

[6] W. K. Newey, "Semiparametric efficiency bounds", Journal of Applied Econometrics, vol. 5, no. 2, pp. 99-135, 1990.

A bound for any robust estimator

- The SCRB is a lower bound for the MSE of any *Regular and Asymptotically Linear (RAL) estimator*.

- All the robust M-, R-, L- estimators belong to this class.

- Every RAL estimator $\hat{\boldsymbol{\theta}}_M(\mathbf{x}_1, \ldots, \mathbf{x}_M) \equiv \hat{\boldsymbol{\theta}}_M$ is:
  1. $\sqrt{M}$-consistent: $\sqrt{M}\left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0\right) = O_P(1)$, [7]
  2. Asymptotically normal: $\sqrt{M}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0) \underset{M \to \infty}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi}(\boldsymbol{\theta}_0, g_0))$.

- Consequently, the following inequality holds:

$$\boldsymbol{\Xi}(\boldsymbol{\theta}_0, g_0) \geq \mathrm{SCRB}(\boldsymbol{\theta}_0 | g_0).$$

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

Evaluation of $\Pi(\cdot|\mathcal{T}_{g_0})$

- The crucial step to evaluate $\mathrm{SCRB}(\boldsymbol{\theta}_0|g_0)$ is in determining the semiparametric efficient score vector:

$$\bar{\mathbf{s}}_0 \triangleq \mathbf{s}_{\boldsymbol{\theta}_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{g_0}).$$

- To this end, we need to:
  1. Calculate $\mathbf{s}_{\boldsymbol{\theta}_0} = \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, g_0)$ (easy task),
  2. Evaluate the projection $\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{g_0})$ (difficult task).

- Two possible approaches:
  1. Least Favorable Submodel,
  2. Projection as a conditional expectation.

Outline of the talk

Why semiparametric models?

CRB in parametric models with finite-dimensional nuisance parameters: classical approach

CRB in parametric models with finite-dimensional nuisance parameters: “Hilbert-space-based” approach

Extension to semiparametric models

SCRB for elliptical distributions

Efficient estimators of the shape matrix in elliptical distributions

Evaluation of the SCRB for the RES class

$$p(\mathbf{x}|\boldsymbol{\theta}_0, g_0) = 2^{-N/2}|\boldsymbol{\Sigma}_0|^{-1/2} g_0(\mathbf{x}^T \boldsymbol{\Sigma}_0^{-1}\mathbf{x}),$$
$$\boldsymbol{\theta}_0 = \mathrm{vecs}(\boldsymbol{\Sigma}_0).$$

- **Problem**: Find the (constrained) SCRB on the estimation of the scatter matrix $\boldsymbol{\Sigma}_0$ when the density generator g_0 is unknown.[8,9]
- To avoid the ambiguity between $\boldsymbol{\Sigma}_0$ and g_0, we put a constraint on the scatter matrix:

$$\mathbf{c}(\boldsymbol{\Sigma}_0) = \mathbf{0}.$$

[8] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, 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, Jan 2019.

[9] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, 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, Oct 2019.

Evaluation of the SFIM $\bar{\mathbf{I}}(\Sigma_0|g_0)$

- The SFIM $\bar{\mathbf{I}}(\boldsymbol{\Sigma}_0|g_0)$ can be obtained as:

$$\bar{\mathbf{I}}(\boldsymbol{\Sigma}_0|g_0) = \frac{2E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}}{N(N+2)} \times \\ \times \mathbf{D}_N^T \left(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1} - \frac{1}{N}\mathrm{vec}(\boldsymbol{\Sigma}_0^{-1})\mathrm{vec}(\boldsymbol{\Sigma}_0^{-1})^T \right) \mathbf{D}_N.$$

  - N: dimension of the data vector $\mathbf{x} \in \mathbb{R}^N$,
  - $\mathcal{Q} =_d \mathbf{x}^H \boldsymbol{\Sigma}_0^{-1}\mathbf{x}$,
  - $\psi_0(q) \triangleq d \ln g_0(q)/dq$,
  - *Duplication matrix*: $\mathbf{D}_N \mathrm{vecs}(\mathbf{A}) = \mathrm{vec}(\mathbf{A})$, $\forall \mathbf{A}$ symmetric.

- Due to the scale-ambiguity problem, $\bar{\mathbf{I}}(\boldsymbol{\Sigma}_0|g_0)$ is singular.

Evaluation of the constrained SCRB (CSCRB)

- To avoid the scale-ambiguity problem, we need to put a constraint on $\boldsymbol{\Sigma}_0$, i.e. $\mathbf{c}(\boldsymbol{\Sigma}_0) = \mathbf{0}$.
- Let $\mathbf{J}_{\mathbf{c}}(\boldsymbol{\Sigma}_0)$ be the Jacobian matrix of the constraint, then there exists a matrix $\mathbf{U}$ s.t.:

$$\mathbf{J}_{\mathbf{c}}(\boldsymbol{\Sigma}_0)\mathbf{U} = \mathbf{0}, \qquad \mathbf{U}^T\mathbf{U} = \mathbf{I}.$$

- The constrained $\mathrm{SCRB}(\boldsymbol{\Sigma}_0|g_0)$ can be expressed as:[10,11]

$$\mathrm{CSCRB}(\boldsymbol{\Sigma}_0|g_0) = \mathbf{U}\left(\mathbf{U}^T\bar{\mathbf{I}}(\boldsymbol{\Sigma}_0|g_0)\mathbf{U}\right)^{-1}\mathbf{U}^T.$$

[10] T. J. Moore, R. J. Kozick and B. M. Sadler, "The Constrained Cramér–Rao Bound from the perspective of fitting a model", *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 564-567, Aug. 2007

[11] S. Fortunati, F. Gini and M. S. Greco, "The Constrained Misspecified Cramér–Rao Bound," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 718-721, May 2016.

Two "semiparametric" estimators (1/2)

- The efficiency w.r.t. the CSCRB of two estimators is investigated:
  - the constrained Sample Covariance matrix (CSCM),
  - the constrained Tyler's estimator (C-Tyler),

- We impose a constraint on the trace: $\mathrm{tr}(\boldsymbol{\Sigma}_0) = N$.

- The CSCM is given by:

$$\begin{cases} \hat{\boldsymbol{\Sigma}}_{SCM} \triangleq \dfrac{1}{M} \sum_{m=1}^{M} \bar{\mathbf{x}}_m \bar{\mathbf{x}}_m^T \\ \hat{\boldsymbol{\Sigma}}_{CSCM} \triangleq \dfrac{N}{\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{SCM})} \hat{\boldsymbol{\Sigma}}_{SCM} \end{cases},$$

Two "semiparametric" estimators (2/2)

- The C-Tyler estimator is given by the convergence point of the following recursion:

$$\begin{cases} \mathbf{S}_T^{(k+1)} = \dfrac{1}{M}\displaystyle\sum_{m=1}^{M} \varphi(t^{(k)})\bar{\mathbf{x}}_m\bar{\mathbf{x}}_m^T \\ \hat{\boldsymbol{\Sigma}}_T^{(k+1)} = N\mathbf{S}_T^{(k+1)}/\mathrm{tr}(\mathbf{S}_T^{(k+1)}) \end{cases},$$

where $t^{(k)} = \bar{\mathbf{x}}_m^T(\hat{\boldsymbol{\Sigma}}_T^{(k)})^{-1}\bar{\mathbf{x}}_m$ and the starting point is $\hat{\boldsymbol{\Sigma}}_T^{(0)} = \mathbf{I}$.

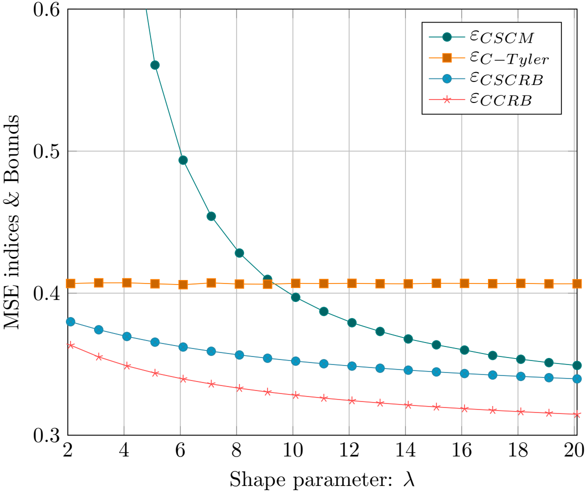
- The weight function $\varphi(t)$ for Tyler's estimator is:

$$\varphi_{Tyler}(t) = N/t.$$

Simulation setup

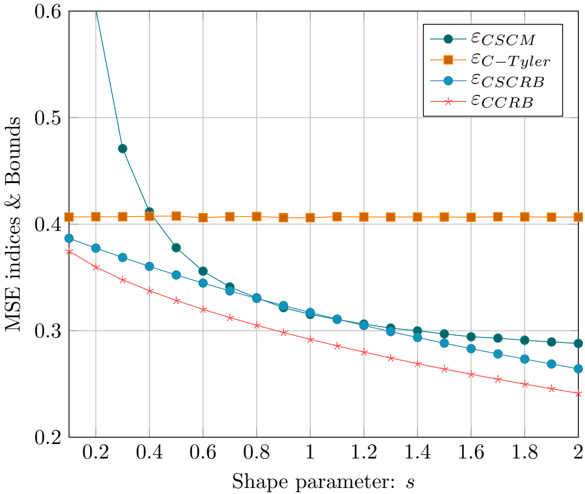
- Two different "true" distributions are considered:
  1. The t-distribution,
  2. The Generalized Gaussian (GG) distribution.
- Simulation parameters
  - $[\boldsymbol{\Sigma}_0]_{i,j} = \rho^{|i-j|}$, $i,j = 1,\dots,N$. Moreover $\rho = 0.8$ and $N = 8$,
  - The data power is chosen to be $\sigma_X^2 = E_{\mathcal{Q}}\{\mathcal{Q}\}/N = 4$,
  - The number of the available i.i.d. data vectors is $M = 5N$.
- Performance indices:
  - MSE index: $\varepsilon_\alpha \triangleq ||E\{\mathrm{vecs}(\hat{\boldsymbol{\Sigma}}_\alpha - \boldsymbol{\Sigma}_0)\mathrm{vecs}(\hat{\boldsymbol{\Sigma}}_\alpha - \boldsymbol{\Sigma}_0)^T\}||_F$,
  - CSCRB index: $\varepsilon_{CSCRB,\boldsymbol{\Sigma}_0} \triangleq ||[\mathrm{CSCRB}(\boldsymbol{\theta}_0|g_0)]_{\boldsymbol{\Sigma}_0}||_F$,
  - CCRB index: $\varepsilon_{CCRB,\boldsymbol{\Sigma}_0} \triangleq ||[\mathrm{CCRB}(\boldsymbol{\theta}_0)]_{\boldsymbol{\Sigma}_0}||_F$.

t-distribution



- When $\lambda \to \infty$, the t-distribution tends the Gaussian one.

GG distribution



- When $s = 1$, the data are Gaussian-distributed.

Main facts

- The *constrained* SCM reaches the SCRB when the data tends to be Gaussian distributed.
- However, its performance degrades significantly in heavy-tailed data.
- The performance of the *constrained* Tyler estimator is invariant w.r.t. the data non-Gaussianity (*robustness*).
- However, it is not efficient w.r.t. the SCRB!
- Is it possible to derive **robust** and **semiparametric efficient** estimator of the scatter matrix in elliptical distributions?

Outline of the talk

Why semiparametric models?

CRB in parametric models with finite-dimensional nuisance parameters: classical approach

CRB in parametric models with finite-dimensional nuisance parameters: "Hilbert-space-based" approach

Extension to semiparametric models

SCRB for elliptical distributions

Efficient estimators of the shape matrix in elliptical distributions

The semiparametric shape matrix estimation

- Let us introduce the *shape* matrix $\mathbf{V}$ as:

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

where $s : \mathcal{M}_N^{\mathbb{R}} \to \mathbb{R}_0^+$ is a scale functional.

- We choose the simplest one, i.e. $s(\boldsymbol{\Sigma}) = [\boldsymbol{\Sigma}]_{1,1}$, that leads to

$$[\mathbf{V}_1]_{1,1} = 1 \Rightarrow \phi \triangleq \underline{\mathrm{vecs}}(\mathbf{V}_1),$$

where $\mathrm{vecs}(\mathbf{V}_1) = [1, \underline{\mathrm{vecs}}(\mathbf{V}_1)^T]^T$.

- The following pdf is uniquely defined: $\forall\, \mathbf{V}_1 \in \mathcal{M}_N^{\mathbb{R}}$, $\forall\, g \in \mathcal{G}$

$$p_X(\mathbf{x}|\phi, g) = 2^{-N/2}|\mathbf{V}_1|^{-1/2} g\left(\mathbf{x}^T\mathbf{V}_1^{-1}\mathbf{x}\right).$$

Le Cam's "one-step" estimators (1/4)

- Let us consider the *parametric* model (g_0 is known):

$$\mathcal{P}_\phi = \{ p_X | p_X(\mathbf{x}|\phi, g_0) \\ = 2^{-N/2} |\mathbf{V}_1|^{-1/2} g_0 \left(\mathbf{x}^T \mathbf{V}_1^{-1} \mathbf{x} \right), \mathbf{V}_1 \in \mathcal{M}_N^{\mathbb{R}} \}.$$

- The Maximum Likelihood estimator for ϕ is:

$$\hat{\phi}_{ML} \triangleq \underset{\phi \in \Omega}{\operatorname{argmax}} \sum_{m=1}^{M} \ln p_X(\mathbf{x}_m | \phi, g_0).$$

- Solving the optimization problem may result to be a prohibitive task.

- In some cases, $\hat{\phi}_{ML}$ may not even exist.

Le Cam's "one-step" estimators (2/4)

- Recall the definition of score vector:

$$\mathbf{s}_\phi(\mathbf{x}_m) \triangleq \nabla_\phi \ln p_X(\mathbf{x}_m|\phi, g_0).$$

- Let us define the *central sequence* as:

$$\boldsymbol{\Delta}_\phi(\mathbf{x}_1, \ldots, \mathbf{x}_M) \equiv \boldsymbol{\Delta}_\phi \triangleq M^{-1/2} \sum_{m=1}^{M} \mathbf{s}_\phi(\mathbf{x}_m).$$

- Under Cramér-type regularity conditions, if $\hat{\phi}_{ML}$ exists, then it satisfies:

$$\boldsymbol{\Delta}_\phi(\mathbf{x}_1, \ldots, \mathbf{x}_M)\big|_{\phi=\hat{\phi}_{ML}} = \mathbf{0},$$

Le Cam's "one-step" estimators (3/4)

- A new estimator $\hat{\phi}$ can be obtained by the one-step Newton-Raphson iteration:

$$\hat{\phi} = \tilde{\phi} - \left[\nabla_{\phi}^{T} \mathbf{\Delta}_{\tilde{\phi}}\right]^{-1} \mathbf{\Delta}_{\tilde{\phi}},$$

where $\tilde{\phi}$ is a "good" starting point.

- $\nabla_{\phi}^{T} \mathbf{\Delta}_{\tilde{\phi}}$ indicates the Jacobian matrix of $\mathbf{\Delta}_{\phi}$ evaluated at $\tilde{\phi}$.

Key point. It can be shown that:

$$\nabla_{\phi}^{T} \mathbf{\Delta}_{\phi} \equiv -M^{1/2}\mathbf{I}(\phi) + o_P(1), \quad \forall \phi \in \mathcal{M}_N^{\mathbb{R}},$$

where $\mathbf{I}(\phi)$ is the Fisher Information Matrix (FIM):

$$\mathbf{I}(\phi) \triangleq E_{\phi,g_0}\left\{\mathbf{s}_{\phi}(\mathbf{x})\mathbf{s}_{\phi}^{T}(\mathbf{x})\right\}.$$

Le Cam's "one-step" estimators (4/4)

Theorem 2. A "one-step" estimator of ϕ_0 is defined as:[12]

$$\hat{\phi} = \hat{\phi}^\star + M^{-1/2}\mathbf{I}(\hat{\phi}^\star)^{-1}\boldsymbol{\Delta}_{\hat{\phi}^\star},$$

where $\hat{\phi}^\star$ is any preliminary $\sqrt{M}$-consistent estimator of ϕ_0.

Properties:

P1 $\sqrt{M}$-consistency:

$$\sqrt{M}\left(\hat{\phi} - \phi_0\right) = O_P(1),$$

P2 Asymptotic normality and efficiency:

$$\sqrt{M}\left(\hat{\phi} - \phi_0\right) \underset{M\to\infty}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}(\phi_0)^{-1}),$$

where $\mathbf{I}(\phi_0)^{-1} \equiv \mathrm{CCRB}(\phi_0)$.

[12] L. Le Cam and G. L. Yang, *Asymptotics in Statistics: Some Basic Concepts (second edition)*. Springer series in statistics, 2000.

Extension to semiparametric models (1/2)

- Theorem 2 is valid in parametric models.
- *Semiparametric extension*: $\phi_0 = \underline{\mathrm{vecs}}(\mathbf{V}_1)$ has to be estimated in the presence of the unknown density generator g_0.
- Let us introduce the *efficient central sequence* as:

$$\overline{\boldsymbol{\Delta}}_\phi(\mathbf{x}_1, \ldots, \mathbf{x}_L) \equiv \overline{\boldsymbol{\Delta}}_\phi \triangleq M^{-1/2} \sum_{m=1}^{M} \bar{\mathbf{s}}_\phi(\mathbf{x}_m),$$

where $\bar{\mathbf{s}}_\phi(\mathbf{x}) \triangleq \mathbf{s}_\phi(\mathbf{x}) - \Pi(\mathbf{s}_\phi|\mathcal{T}_{g_0})$ is the efficient score vector.

- Let us also recall the SFIM:

$$\bar{\mathbf{I}}(\phi|g_0) \triangleq E_{\phi,g_0}\{\bar{\mathbf{s}}_\phi(\mathbf{x})\bar{\mathbf{s}}_\phi(\mathbf{x})^T\}.$$

Extension to semiparametric models (2/2)

Theorem 3. A semiparametric "one-step" estimator of ϕ_0 is defined as:[13]

$$\hat{\phi}_s = \hat{\phi}^\star + M^{-1/2}\bar{\mathbf{I}}(\hat{\phi}^\star|g_0)^{-1}\overline{\boldsymbol{\Delta}}_{\hat{\phi}^\star},$$

where $\hat{\phi}^\star$ is any preliminary $\sqrt{M}$-consistent estimator of ϕ_0.

Properties:

P1 $\sqrt{M}$-consistency: $\sqrt{M}\left(\hat{\phi} - \phi_0\right) = O_P(1)$,

P2 Asymptotic normality and efficiency:

$$\sqrt{M}\left(\hat{\phi} - \phi_0\right) \underset{M\to\infty}{\sim} \mathcal{N}(\mathbf{0}, \bar{\mathbf{I}}(\phi_0|g_0)^{-1}),$$

and $\bar{\mathbf{I}}(\phi_0|g_0)^{-1} \equiv \mathrm{CSCRB}(\phi_0|g_0)$.

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

Have we finished?

- To derive a semiparametric efficient estimator of the shape matrix $\phi = \underline{\mathrm{vecs}}(\mathbf{V}_1)$, we need:
  1. The efficient central sequence:

$$\overline{\mathbf{\Delta}}_{\mathbf{V}_1} = -L^{-1/2}\mathbf{K}_{\mathbf{V}_1} \sum_{m=1}^{M} Q_m \psi_0(Q_m)\mathrm{vec}(\mathbf{u}_m\mathbf{u}_m^T).$$

  2. The efficient Semiparametric FIM

$$\bar{\mathbf{I}}(\underline{\mathrm{vecs}}(\mathbf{V}_1)|g_0) = \frac{2E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}}{N(N+2)}\mathbf{K}_{\mathbf{V}_1}\mathbf{K}_{\mathbf{V}_1}^T.$$

  3. A preliminary $\sqrt{M}$-consistent estimator, i.e. $\widehat{\mathbf{V}}_1^\star$.

- $Q_m \triangleq \mathbf{x}^H \mathbf{\Sigma}_0^{-1}\mathbf{x} =_d \mathcal{Q} \sim P_{\mathcal{Q}}(g_0)$,
- $\psi_0(q) \triangleq d \ln g_0(q)/dq$,
- $\mathbf{u}_m \sim \mathcal{U}(\mathbb{R}S^N)$.
- $\mathbf{K}_{\mathbf{V}_1} \triangleq \mathbf{D}_N^1(\mathbf{V}_1^{-1/2} \otimes \mathbf{V}_1^{-1/2})\Pi^{\perp}_{\mathrm{vec}(\mathbf{I}_N)}$.

Have we finished? Not yet...

- **Bad news**: The efficient central sequence $\overline{\mathbf{\Delta}}_{\mathbf{V}_1}$ and the SFIM $\bar{\mathbf{I}}(\underline{\mathrm{vecs}}(\mathbf{V}_1)|g_0)$ depends of the *unknown* density generator g_0.

- Theorem 3 provides us with a *pseudo*-estimator that cannot be implemented in practice.

- Is there any way out? **Rank-based statistics!** [14]

- In their seminal paper,[15] Hallin and Werker proposed an *invariance-based* approach to solve semiparametric estimation problems.

- **Main idea**: Find a distribution-free approximation of the efficient central sequence $\overline{\mathbf{\Delta}}_{\mathbf{V}_1}$!

[14] The definition of rank is given in the backup slides.

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

A semiparametric efficient R-estimator (1/2)

- Building upon this fundamental results on rank-based invariance, Hallin, Oja and Paindaveine[16] obtained a
  1. *distributionally-robust* and
  2. (almost) *semiparametric efficient*,

  R-estimator of the shape matrix!

$$\underline{\mathrm{vecs}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vecs}}(\widehat{\mathbf{V}}_1^\star) + M^{-1/2}\widehat{\boldsymbol{\Upsilon}}^{-1}\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}.$$

- $\widehat{\boldsymbol{\Upsilon}}$ represents an estimate of $\bar{\mathbf{I}}(\underline{\mathrm{vecs}}(\mathbf{V}_1)|g_0)$.
- $\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}$ is a distributionally-free approximation of the efficient central sequence $\overline{\boldsymbol{\Delta}}_{\mathbf{V}_1}$.

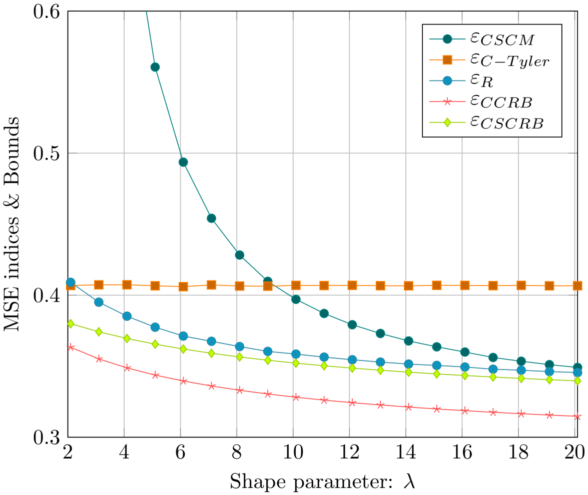
[16] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. optimal R-estimation of shape," *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.

A semiparametric efficient R-estimator (2/2)

$$\underline{\mathrm{vecs}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vecs}}(\widehat{\mathbf{V}}_1^\star) - \frac{1}{M\hat{\alpha}} \left[\mathbf{K}_{\widehat{\mathbf{V}}_1^\star} \mathbf{K}_{\widehat{\mathbf{V}}_1^\star}^T\right]^{-1}$$
$$\times \mathbf{K}_{\widehat{\mathbf{V}}_1^\star} \sum\nolimits_{m=1}^{M} K_{\mathrm{vdW}} \left(\frac{r_l^\star}{M+1}\right) \mathrm{vec}(\hat{\mathbf{u}}_m^\star (\hat{\mathbf{u}}_m^\star)^T),$$

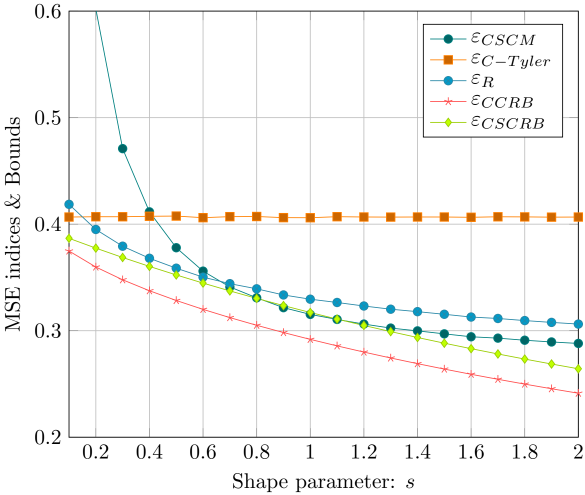
- $\{r_m^\star\}_{m=1}^M$ are the ranks of the r. v. $\hat{Q}_m^\star \triangleq \mathbf{x}_m^T [\widehat{\mathbf{V}}_1^\star]^{-1} \mathbf{x}_m$,
- $\hat{\mathbf{u}}_m^\star \triangleq \frac{[\widehat{\mathbf{V}}_1^\star]^{-1/2} \mathbf{x}_m}{\sqrt{\hat{Q}_m^\star}}$,
- $K_{\mathrm{vdW}}(\cdot)$ is the *van der Waerden* score function,
- $\hat{\alpha}$ is a data-dependent "cross-information" term,
- $\widehat{\mathbf{V}}_1^\star$ is a preliminary $\sqrt{M}$-consistent estimator of $\mathbf{V}_1$.

t-distribution



- When $\lambda \to \infty$, the t-distribution tends the Gaussian one.

GG distribution



- When $s = 1$, the data are Gaussian-distributed.

Conclusions

- The wide applicability of the semiparametric framework has been discussed.
- The theory underlying the Semiparametric CRB (SCRB) has been introduced.
- We derived a **closed form expression** of the SCRB for the scatter matrix estimation problem in elliptical distributions.
- Building upon the Le Cam's "one-step" estimators, a general procedure to obtain semiparametric efficient estimators has been provided.
- The derivation of a distributionally-robust and semiparametric efficient estimator for the shape matrix has been addressed.

Our current work

- We derived a **complex extension** of the R-estimator that is of great importance in Signal Processing applications.
- We are working on the derivation of an **efficient estimator** of the "cross-information" term $\hat{\alpha}$.
- What about the **asymptotic distribution** of the derived the R-estimator?
- Which is the behavior of the R-estimator as the **data dimension** N goes to infinity?
- **Applications**: Radar/Sonar processing, beamforming, image processing,...

Many thanks for your attention!

Any question?

Backup slides

Hilbert spaces

Definition

A Hilbert space $\mathcal{F}$ is a *normed vector space*

1. equipped with an *inner product* $\langle \cdot, \cdot \rangle$ and,
2. *complete* with respect to the norm $|| \cdot || = \sqrt{\langle \cdot, \cdot \rangle}$.

- A normed (metric) space is complete when every Cauchy sequences in $\mathcal{F}$ converges to an element of $\mathcal{F}$.

- $f_1, f_2, \cdots$ is a Cauchy sequence if, for every $\varepsilon > 0$ there is a positive integer N such that for all $i, j > N$, we have that:

$$||f_i - f_j|| < \varepsilon.$$

The square-integrable functions

- Let $(\mathcal{X}, \mathfrak{F}, \mu)$ be a measure space where $\mathcal{X} \subseteq \mathbb{R}^N$, $\mathfrak{F}$ is the Borel σ-algebra on $\mathcal{X}$ and μ is a measure on $\mathfrak{F}$.

- Then, $L_2(\mu)$ is the space of all the measurable functions s. t.

$$L_2(\mu) = \left\{ f : \mathcal{X} \to \mathbb{R} \,\middle|\, \int_{\mathcal{X}} |f(\mathbf{x})|^2 d\mu(\mathbf{x}) < \infty \right\}.$$

- The $L_2(\mu)$ space is an Hilbert space with the following inner product:

$$\langle f_1, f_2 \rangle \triangleq \int_{\mathcal{X}} f_1(\mathbf{x}) f_2(\mathbf{x}) d\mu(\mathbf{x}).$$

- For the standard Lebesgue measure: $d\mu(\mathbf{x}) = d\mathbf{x}$.

The space of scalar zero-mean functions

- Let $(\mathcal{X}, \mathfrak{F}, P_X)$ be a probability space where $\mathcal{X} \subseteq \mathbb{R}^N$ is the sample space, $\mathfrak{F}$ is the Borel σ-algebra on $\mathcal{X}$ and P_X is a probability measure. [17]

- Let $\mathcal{H}$ be the Hilbert space defined as:

$$\mathcal{H} = \left\{ h : \mathcal{X} \to \mathbb{R} \,\middle|\, E_X\{h\} = 0, E_X\{|h|^2\} < \infty \right\}.$$

- The expectation operator $E_X\{\cdot\}$ is

$$E_X\{h\} \triangleq \int_{\mathcal{X}} h(\mathbf{x}) dP_X(\mathbf{x}) = \int_{\mathcal{X}} h(\mathbf{x}) p_X(\mathbf{x}) d\mathbf{x},$$

where p_X is the probability density function (pdf).

- The inner product in $\mathcal{H}$ is: $\langle h_1, h_2 \rangle \triangleq E_X\{h_1 h_2\}$.

[17] Some additional definitions are given in the backup slides.

The projection theorem (1/2)

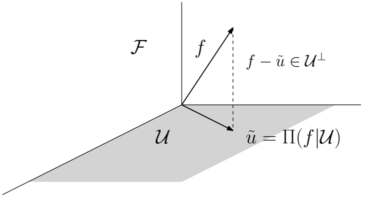
Theorem

Let $\mathcal{U}$ be a closed subspace of an Hilbert space $\mathcal{F}$ and take $f \in \mathcal{F}$. We call

$$d(f,\mathcal{U}) \triangleq \inf_{u\in\mathcal{U}} ||f-u||, \quad f \in \mathcal{F},$$

the distance of f to $\mathcal{U}$. Then there exists a unique element $\tilde{u} \in \mathcal{U}$ for which

$$||f-\tilde{u}|| = d(f,\mathcal{U}).$$



The projection theorem (2/2)

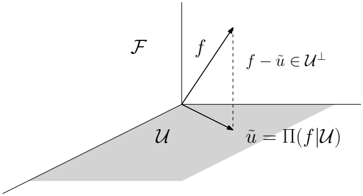
- f can be uniquely written as:

$$f = \tilde{u} + (f - \tilde{u}),$$

where $\tilde{u} \triangleq \Pi(f|\mathcal{U}) \in \mathcal{U}$ and $f - \tilde{u} \in \mathcal{U}^{\perp}$.

- $\tilde{u}$ is uniquely determined by the orthogonality constraint:

$$\langle f - \tilde{u}, u \rangle = \langle f - \Pi(f|\mathcal{U}), u \rangle = 0, \quad \forall u \in \mathcal{U}.$$



The linear span

- A *q-replicating* Hilbert space $\mathcal{F}^q$ is obtained by the Cartesian product of the q copies of $\mathcal{F}$ as $\mathcal{F}^q \triangleq \mathcal{F} \times \cdots \times \mathcal{F}$, then:

$$\mathcal{F}^q \ni \mathbf{f} = (f_1, f_2, \cdots, f_q)^T, \quad f_i \in \mathcal{F}.$$

- The inner product of $\mathcal{F}^q$ is induced by the one in $\mathcal{F}$:

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{i=1}^{q} \langle f_i, g_i \rangle .$$

- **Linear span**: Let $\mathbf{u} = (u_1, \cdots, u_k)^T$ be a column vector of k elements of $\mathcal{F}$. The *linear span* of the vector $\mathbf{u}$, defined as:

$$\mathcal{V} \triangleq \{\mathbf{v} | \mathbf{v} = \mathbf{A}\mathbf{u}, \mathbf{A} \text{ is any matrix in } \mathbb{R}^{q \times k}\},$$

is a *finite-dimensional* subspace of $\mathcal{F}^q$.

Projection onto a finite-dimensional subspace

$$\mathcal{V} \triangleq \{\mathbf{v} | \mathbf{v} = \mathbf{A}\mathbf{u}, \mathbf{A} \text{ is any matrix in } \mathbb{R}^{q \times k}\}.$$

- If $u_1, \ldots, u_k$ are linearly independent in $\mathcal{F}$, $\dim(\mathcal{V}) = kq$.
- The projection of a generic element $\mathbf{f} \in \mathcal{F}^q$ onto the subspace $\mathcal{V}$ is given by:

$$\Pi(\mathbf{f}|\mathcal{V}) = \left\langle \mathbf{f}, \mathbf{u}^T \right\rangle \left\langle \mathbf{u}, \mathbf{u}^T \right\rangle^{-1} \mathbf{u},$$

where

$$\left[\left\langle \mathbf{f}, \mathbf{u}^T \right\rangle\right]_{i,j} \triangleq \langle f_i, u_j \rangle, \quad \begin{array}{l} i = 1, \ldots, q, \\ j = 1, \ldots, k, \end{array}$$

$$\left[\left\langle \mathbf{u}, \mathbf{u}^T \right\rangle\right]_{i,j} \triangleq \langle u_i, u_j \rangle, \ i, j = 1, \ldots, k.$$

The vector-valued zero-mean functions

- Let $(\mathcal{X}, \mathfrak{F}, P_X)$ be a probability space.
- Let $\mathcal{H}^q$ be the q-replicating Hilbert space:

$$\begin{aligned} \mathcal{H}^q &= \mathcal{H} \times \cdots \times \mathcal{H} \\ &= \left\{ \mathbf{h} : \mathcal{X} \to \mathbb{R}^q \middle| E_X\{\mathbf{h}\} = \mathbf{0}, E_X\{\mathbf{h}^T \mathbf{h}\} < \infty \right\}, \end{aligned}$$

- The induced inner product is:

$$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle \triangleq E_X\{\mathbf{h}_1^T \mathbf{h}_2\}.$$

- The *covariance matrix* of $\mathbf{h} \in \mathcal{H}^q$ is:

$$\mathbf{C}_X(\mathbf{h}) \triangleq E_X\{\mathbf{h}\mathbf{h}^T\}.$$

Projection onto finite-dimensional subspaces

- Let $\mathbf{u} = (u_1, \cdots, u_k)^T$ be a column vector of k arbitrary elements of $\mathcal{H}$ and let $\mathcal{V}$ be its linear span.

- The orthogonal projection of an arbitrary element $\mathbf{h} \in \mathcal{H}^q$ onto $\mathcal{V}$ is unique and it is given by:

$$\begin{aligned} \Pi(\mathbf{h}|\mathcal{V}) &= E_X\{\mathbf{h}\mathbf{u}^T\}E_X\{\mathbf{u}\mathbf{u}^T\}^{-1}\mathbf{u} \\ &= E_X\{\mathbf{h}\mathbf{u}^T\}\mathbf{C}_X(\mathbf{u})^{-1}\mathbf{u}. \end{aligned}$$

- Linear Minimum Mean Square Error (LMMSE) estimator:

  1. $\mathrm{MSE} \triangleq ||\mathbf{h} - \mathbf{A}\mathbf{u}||^2$ is minimized by $\Pi(\mathbf{h}|\mathcal{V})$, then $\hat{\mathbf{h}}_{LMMSE} = E_X\{\mathbf{h}\mathbf{u}^T\}\mathbf{C}_X(\mathbf{u})^{-1}\mathbf{u}$.
  2. The "orthogonality principle" is nothing but the Projection Theorem.

Score vectors as elements of $\mathcal{H}^r$ (1/2)

- Let us go back to the *parametric model*:

$$\mathcal{P}_{\boldsymbol{\theta},\boldsymbol{\eta}} \triangleq \left\{ p_X(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{\eta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q, \boldsymbol{\eta} \in \Gamma \subseteq \mathbb{R}^d \right\},$$

  - $\boldsymbol{\theta} \in \Theta$ is the vector of the parameters of interest,
  - $\boldsymbol{\eta} \in \Gamma$ is the vector of the (unknown) nuisance parameters,
  - $\gamma \triangleq \left(\boldsymbol{\theta}^T, \boldsymbol{\eta}^T\right)^T \in \mathbb{R}^r,\ r = q + d.$
  - $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ is the "true" pdf.

- The **score vector** for the true parameter vector γ_0 is:

$$\mathbf{s}_{\gamma_0} \triangleq \nabla_\gamma \ln p_X(\mathbf{x}|\gamma_0) = \begin{pmatrix} \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \\ \nabla_{\boldsymbol{\eta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{s}_{\boldsymbol{\theta}_0} \\ \mathbf{s}_{\boldsymbol{\eta}_0} \end{pmatrix}$$

  - $\mathbf{s}_{\boldsymbol{\theta}_0}$ is $q \times 1$ the score vector of the parameters of interest,
  - $\mathbf{s}_{\boldsymbol{\eta}_0}$ is $d \times 1$ the nuisance score vector.

Score vectors as elements of $\mathcal{H}^r$ (2/2)

- Under standard regularity conditions:

$$
\begin{aligned}
E_0\left\{\mathbf{s}_{\gamma_0}\right\} &= \int_{\mathcal{X}} \nabla_\gamma \ln p_X(\mathbf{x}|\gamma_0) dP_0(\mathbf{x}) \\
&= \int_{\mathcal{X}} \frac{\nabla_\gamma p_X(\mathbf{x}|\gamma_0)}{p_0(\mathbf{x})} p_0(\mathbf{x}) d\mathbf{x} = \nabla_\gamma \int_{\mathcal{X}} p_X(\mathbf{x}|\gamma_0) d\mathbf{x} = 0,
\end{aligned}
$$

and $E_0\left\{\mathbf{s}_{\gamma_0}^T \mathbf{s}_{\gamma_0}\right\} < \infty$.

- Then, by definition[18] of $\mathcal{H}^r$:

$$
\mathcal{H}^r \ni \mathbf{s}_{\gamma_0} = \begin{pmatrix} \mathbf{s}_{\boldsymbol{\theta}_0} \\ \mathbf{s}_{\boldsymbol{\eta}_0} \end{pmatrix} \quad \Rightarrow \quad \mathbf{s}_{\boldsymbol{\theta}_0} \in \mathcal{H}^q, \quad \mathbf{s}_{\boldsymbol{\eta}_0} \in \mathcal{H}^d.
$$

[18] $\mathcal{H}^r = \left\{\mathbf{h} : \mathcal{X} \to \mathbb{R}^r \,\middle|\, E_0\{\mathbf{h}\} = \mathbf{0},\, E_0\{\mathbf{h}^T\mathbf{h}\} < \infty\right\}$.

Parametric submodels (1/3)

- Let us recall the semiparametric model:

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

- The **i-th parametric submodel** of $\mathcal{P}_{\boldsymbol{\theta},g}$ is defined as:[19]

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

  where:

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

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

[19] An example on how tu build parametric submodels of the CES distributions is given in the backup slides.

Parametric submodels (2/3)

- Denote the "true semiparametric vector" and the related true pdf as $(\boldsymbol{\theta}_0^T, g_0)^T$ and $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\theta}_0, g_0)$, respectively.
- For every $i \in \mathcal{I}$, the *i-th parametric submodel*:

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

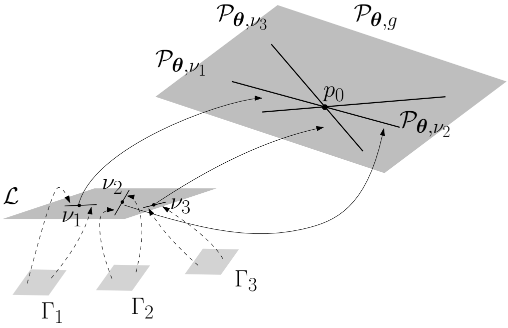
has to satisfy the following three conditions:

C0) $\nu_i : \Gamma_i \to \mathcal{L}$ 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 $p_X(\mathbf{x}|\boldsymbol{\theta}_0, \nu_i(\mathbf{x}, \boldsymbol{\eta}_0)) = p_X(\mathbf{x}|\boldsymbol{\theta}_0, g_0) \triangleq p_0(\mathbf{x})$.

Parametric submodels (3/3)



- The generalization to the semiparametric framework can be done in two steps:

  1. Exploit the obtained results in the set of (artificial) parametric submodels $\{\mathcal{P}_{\boldsymbol{\theta},\nu_i}\}_{i\in\mathcal{I}}$,
  2. "Take the limit" to generalize them in the infinite-dimensional semiparametric framework.

Semiparametric nuisance tangent space

- For every parametric submodel:

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

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

- The **semiparametric nuisance tangent space** is defined as:[20]

$$\mathcal{T}_{g_0} \triangleq \overline{\bigcup_{\{\mathcal{P}_{\boldsymbol{\theta},\nu_i}\}_{i\in\mathcal{I}}} \mathcal{T}_{\boldsymbol{\eta}_{0,i}}} \subseteq \mathcal{H}^q.$$

[20] 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 $\mathcal{A}$.

Parametric submodels of the CES model (1/3)

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

$$p_X(\mathbf{x}; \boldsymbol{\Sigma}) = c_{N,g} |\boldsymbol{\Sigma}|^{-1} g(\mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x}) \triangleq CES_N(\mathbf{x}; \boldsymbol{\Sigma}, g),$$

- $\mathcal{G} \ni g : \mathbb{R}_0^+ \to \mathbb{R}^+$ is the *density generator* and

$$\mathcal{G} \triangleq \left\{ g : \mathbb{R}_0^+ \to \mathbb{R}^+ \middle| \int_0^\infty t^{N-1} g(t) dt < \infty \right\}$$

- The set of all CES pdfs is a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\Sigma},g} \triangleq \{ p_X | p_X(\mathbf{x}|\boldsymbol{\Sigma}, g), \boldsymbol{\Sigma} \in \mathcal{M}_N, g \in \mathcal{G} \}.$$

- How can we build a parametric submodel of $\mathcal{P}_{\boldsymbol{\Sigma},g}$?

Parametric submodels of the CES model (2/3)

- The set of all the density generator $\mathcal{G}$ is a convex set!

Proof

For every $g_0, g_1 \in \mathcal{G}$ and for every $\eta \in [0,1]$, we have that:

1. $\eta g_1(t) + (1-\eta) g_0(t)$ is a function of $t \triangleq \mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x}$,
2. By linearity, $\int_0^\infty t^{N-1} [\eta g_1(t) + (1-\eta) g_0(t)] dt < \infty$,

then $\eta g_1 + (1-\eta) g_0 \in \mathcal{G}$ and consequently $\mathcal{G}$ is a convex set.

- Then it is immediate to verify that:

$$\begin{aligned} CES_N(\mathbf{x}; \mathbf{\Sigma}, g_0) &= CES_N(\mathbf{x}; \mathbf{\Sigma}, \eta g_1 + (1-\eta) g_0) \\ &= \eta CES_N(\mathbf{x}; \mathbf{\Sigma}, g_1) + (1-\eta) CES_N(\mathbf{x}; \mathbf{\Sigma}, g_0). \end{aligned}$$

- $\mathcal{P}_{\mathbf{\Sigma}, g}$ is a convex set as well!

Parametric submodels of the CES model (3/3)

- Let us define a smooth parametric map as:

$$\begin{aligned} \nu_i &: [0,1] \to \mathcal{G} \\ &\quad \eta \mapsto \nu_i(t,\eta) \triangleq \eta g_i(t) + (1-\eta) g_0(t), \end{aligned}$$

  where g_i is a generic density generator while g_0 is the true one.

- The relevant *i-th* parametric submodel is then given by:

$$\mathcal{P}_{\boldsymbol{\Sigma},\nu_{\eta_i}} = \{ p_X | p_X(\mathbf{x}|\boldsymbol{\Sigma}, \eta g_i + (1-\eta) g_0), \boldsymbol{\Sigma} \in \mathcal{M}_N, \eta \in [0,1] \}.$$

- It is immediate to verify that this submodel satisfies the conditions C0, C1 and C2 given in slide 64.

- In particular, Condition C2 is verified by choosing $\eta = 0$.

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.[21]

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

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

Ranks (2/2)

- Let $\mathbf{r}_X \triangleq [r_1, \ldots, r_L]^T \in \mathbb{N}^L$ 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 $\{1, 2, \ldots, L\}$,
3. For each $l = 1, \ldots, L$, $K\left(\frac{r_l}{L+1}\right) = K(u_l) + o_P(1)$, where $K \in \mathcal{K}$ and $u_l \sim \mathcal{U}[0,1]$ is a random variable uniformly distributed in $(0,1)$.