The Misspecified and Semiparametric lower bounds and their application to inference problems with Complex Elliptically Symmetric (CES) distributed data

Stefano Fortunati and Fulvio Gini

Dip. Ingegneria dell'Informazione, University of Pisa,

EUSIPCO 2019 - A Coruña, Spain
Monday, September 2, 2019

Part II - 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.
- Semiparametric interpretation of Real and Complex ES distributions.
- Examples.

Part II - 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

Semiparametric interpretation of Real and Complex ES distributions

Examples

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 $\boldsymbol{\theta}$:

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

- 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 [1,2,3,4,5,6].

[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*):

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

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, $g \in \mathcal{L}$ [7]:

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

- 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 [8]:

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

  - $c_{N,g}$ 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* [9, Sec. 4.2].

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 [9, Sec. 4.5], [10]:

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

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:

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

- This model is a general form of a *semiparametric regression model* [9, Sec. 4.3].

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 [11,12]:

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

Part II - 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

Semiparametric interpretation of Real and Complex ES distributions

Examples

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 (see e.g. [13, Sec. 10.7]):

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

- *Classical approach*: $\mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)$ can be obtained from the FIM using the Matrix Inversion Lemma [14]:

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

- It is possible to obtain this same result by using a geometrical, "**Hilbert-space-based**" approach [7].

Part II - 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

Semiparametric interpretation of Real and Complex ES distributions

Examples

Hilbert spaces

Definition ([9, A.1, A.2],[15])

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 $\mathfrak{F}$. [3]

- 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: $d\mu(\mathbf{x}) = d\mathbf{x}$.

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

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. [4]

- Let $\mathcal{H}$ be the Hilbert space defined as [10, Ch. 2]:

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

- 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: $\langle h_1, h_2 \rangle \triangleq E_X\{h_1 h_2\}$.

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

The projection theorem (1/2)

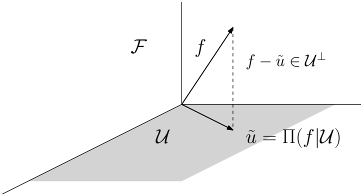
Theorem

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)

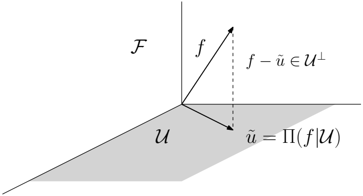
- f can be uniquely written as:

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

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 $\mathcal{F}$, $\dim(\mathcal{V}) = kq$. [5]
- The projection of a generic element $\mathbf{f} \in \mathcal{F}^q$ onto the subspace $\mathcal{V}$ is given by [9, A.2], [10, Sec. 2.4]:

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

where

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

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

[5] The proof of this result is in the backup slides (see also [10, Sec. 2.4]).

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 [10, Ch. 2]:

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

- 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 [9, A.2], [10, Sec. 2.4]:

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

- 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,
  - $\boldsymbol{\gamma} \triangleq \left(\boldsymbol{\theta}^T, \boldsymbol{\eta}^T\right)^T \in \mathbb{R}^r, \; r = q + d$.
  - $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ 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 [16]:

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

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

- Then, by definition[6] of $\mathcal{H}^r$:

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

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

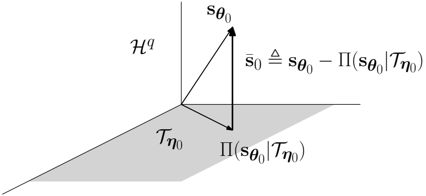
The efficient score vector

- The *nuisance tangent space*[7] $\mathcal{T}_{\boldsymbol{\eta}_0}$ is defined as the linear span of $\mathbf{s}_{\boldsymbol{\eta}_0}$ in $\mathcal{H}^q$ [10, Ch. 3]:

$$\mathcal{T}_{\boldsymbol{\eta}_0} \triangleq \{\mathbf{t} | \mathbf{t} = \mathbf{A}\mathbf{s}_{\boldsymbol{\eta}_0}, \mathbf{A} \text{ is any matrix in } \mathbb{R}^{q \times d}\} \subset \mathcal{H}^q.$$

- Let us define the **efficient score vector** as [9, Ch. 2]:

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



[7] The geometrical intuition behind this terminology is given in the backup slides.

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

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

- 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:

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

Part II - 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

Semiparametric interpretation of Real and Complex ES distributions

Examples

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 [7].

- 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*.

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**[8] of $\mathcal{P}_{\boldsymbol{\theta},g}$ is defined as [10, Sec. 4.2], [9, Sec. 3.1], [17,18,11], :

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

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.

[8] An explicit example of parametric submodel 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*:

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

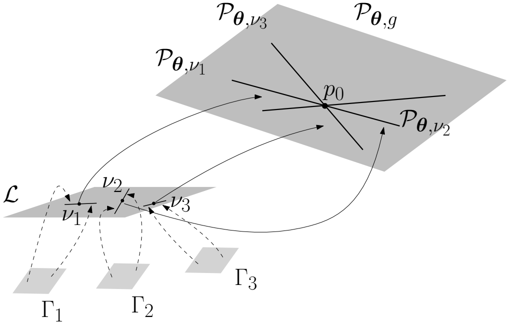
has to satisfy the following three conditions [10, Sec. 4.2]:

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 (1/2)

- 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:[9]

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

[9] 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 $\mathcal{A}$.

Semiparametric nuisance tangent space (2/2)

- Recall that the Hilbert space $\mathcal{H}^q$ is a complete normed space with norm:

$$||\mathbf{h}_1 - \mathbf{h}_2|| = \sqrt{E_0\{(\mathbf{h}_1 - \mathbf{h}_2)^T(\mathbf{h}_1 - \mathbf{h}_2)\}}, \quad \forall \mathbf{h}_1, \mathbf{h}_2 \in \mathcal{H}^q.$$

- The semiparametric nuisance tangent space $\mathcal{T}_{g_0} \subseteq \mathcal{H}^q$ can be expressed as [10, Sec. 4.4],[19],[18]: [10]

$$\mathcal{T}_{g_0} \triangleq \left\{\mathbf{h} \in \mathcal{H}^q | \forall \varepsilon > 0, \exists i \in \mathcal{I} : ||\mathbf{h} - \mathbf{A}_i \mathbf{s}_{\boldsymbol{\eta}_{0,i}}|| < \varepsilon\right\}$$

- Unlike $\mathcal{T}_{\boldsymbol{\eta}_{0,i}}$ that has finite dimension, $\mathcal{T}_{g_0}$ is in general an infinite-dimensional subspace of $\mathcal{H}^q$.

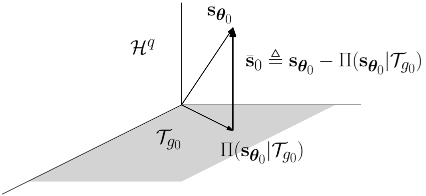
[10] A more explicit definition of the nuisance tangent space requires the notion of *Hellinger differentiability* [19],[9, Sec. 3.2]. See also the backup slides.

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

- 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 [9, Sec. 3.3]:

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



The Semiparametric CRB (SCRB) (1/2)

Theorem ([9, Sec. 3.4], [19], [10, Theo. 4.2], [18]):
A lower bound on the MSE of "any" [11] robust estimator of $\boldsymbol{\theta}_0$ in the presence of the nuisance function $g_0 \in \mathcal{L}$ is given by:

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

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:

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

[10] 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.

[9, Sec. 3.4] P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner, *Effient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, 1993.

[11] The class of estimators to which the SCRB applies is discussed ahead.

The Semiparametric CRB (SCRB) (2/2)

- The expression of $\mathrm{SCRB}(\boldsymbol{\theta}_0|g_0)$ is formally equivalent to $\mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)$ derived for finite-dimensional nuisance vectors.
- The Hilbert-space-based approach allows to handle both finite and infinite-dimensional nuisance parameters.
- The $\mathrm{SCRB}(\boldsymbol{\theta}_0|g_0)$ is higher than any $\mathrm{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_{0,i})$ derived in the *i-th* parametric submodel.
- A semiparametric model contains less information on $\boldsymbol{\theta}_0$ than any of its possible parametric submodel.

A bound for any robust estimator

- The SCRB is a lower bound for the MSE of any *Regular and Asymptotically Linear (RAL) estimator* [9, Sec. 2.2 and Ch. 7], [10, Ch.3], [20, Ch. 4] [21,18,22,23].

- All the robust M-, S-, L- estimators belong to this class [24]:

- It can be shown that every RAL estimator is:
  1. Consistent: $\hat{\boldsymbol{\theta}}(\mathbf{x}_1, \ldots, \mathbf{x}_M) \triangleq \hat{\boldsymbol{\theta}}_M \underset{M \to \infty}{\longrightarrow} \boldsymbol{\theta}_0$,
  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 [9, Ch. 2 and 3]:

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

- Note that efficient estimators may not exist [25].

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 Favourable Submodel (if it exists) [12],
  2. **Projection as a conditional expectation**.

[12] Some additional details are given in the backup slides.

Projection and conditional expectation (1/3)

- We defined $\mathcal{H}^q$ as the Hilbert space of the q-dimensional zero-mean function on the probability space $(\mathcal{X}, \mathfrak{F}, P_X)$:

$$\mathbf{h} \equiv \mathbf{h}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^N.$$

- Let $f : \mathbb{R}^N \to \mathbb{R}$ be a measurable function. We define a *statistic* V of the random vector $\mathbf{x}$ as:

$$V =_d f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X}.$$

- Let $\mathfrak{G}(V) \subseteq \mathfrak{F}$ be the sub-σ algebra generated by V. [13]

- The set of the q-dim zero-mean functions on $(\mathcal{X}, \mathfrak{G}(V), P_X)$ is a closed linear subspace, say $\mathcal{V}$, of $\mathcal{H}^q$ [26, Theo. 23.2].

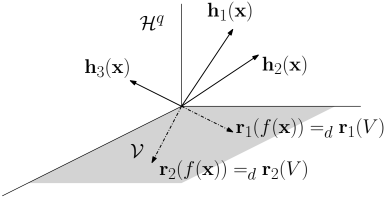
[13] Additional details are given in the backup slides.

Projection and conditional expectation (2/3)

- Let $\mathbf{r} \in \mathcal{H}^q$ be a zero-mean function of $\mathbf{x} \in \mathcal{X}$ through the function f, i.e.: [14]

$$\mathbf{r} \equiv \mathbf{r}(f(\mathbf{x})) =_d \mathbf{r}(V) \in \mathcal{V} \subseteq \mathcal{H}^q.$$

- Consequently, $\mathbf{r} \in \mathcal{H}^q$ can be considered as a q-dimensional function defined on $(\mathcal{X}, \mathfrak{G}(V), P_X)$ with $\mathfrak{G}(V) \subseteq \mathfrak{F}$.



[14] The symbol "$=_d$" means "has the same distribution as".

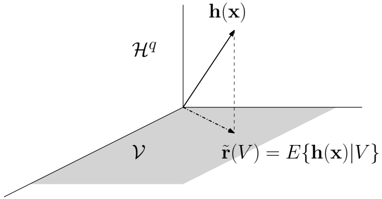
Projection and conditional expectation (3/3)

- The conditional expectation $E\{\mathbf{h}|V\}$ is the unique element in $\mathcal{V}$, such that [26, Def. 23.3, Theo. 23.3][15]:

$$\langle \mathbf{h} - E\{\mathbf{h}|V\}, \mathbf{r}\rangle \triangleq E\left\{\left(\mathbf{h} - E\{\mathbf{h}|V\}\right)^T \mathbf{r}\right\} = 0, \quad \forall \mathbf{r} \in \mathcal{V}.$$

Given the Projection Theorem, the previous definition implies:

$$\Pi(\cdot|\mathcal{V}) = E\{\cdot|V\}.$$



[15] This definition is consistent with the classical one [26, Ch. 32]. See the proof in the backup slides.

Part II - 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

Semiparametric interpretation of Real and Complex ES distributions

Examples

Spherically Symmetric (SS) distributions

- Let $\mathbf{z} \in \mathbb{R}^N$ be a real-valued random vector.
- Let $\mathcal{O}$ be the set of all unitary transformations:

$$\begin{aligned} \mathcal{O} \ni O : \mathbb{R}^N &\to \mathbb{R}^N \\ \mathbf{z} &\mapsto O(\mathbf{z}) = \mathbf{O}\mathbf{z}, \end{aligned}$$

  for any unitary matrix $\mathbf{O}$, i.e $\mathbf{O}^T\mathbf{O} = \mathbf{O}\mathbf{O}^T = \mathbf{I}$.

- Then, $\mathbf{z}$ is said to be SS-distributed if its distribution is invariant to any unitary transformations $\mathbf{O} \in \mathcal{O}$, i.e.

$$\mathbf{z} =_d \mathbf{O}\mathbf{z}.$$

- We indicate with $\mathcal{S}$ the class of all SS-distributions.

Properties of the (SS) distributions (1/4)

Property P1 [16]

- The SS-distributed random vector $\mathbf{z} \sim SS(g)$ has a pdf:

$$p_Z(\mathbf{z}) = 2^{-N/2} g\left(||\mathbf{z}||^2\right),$$

where $\mathcal{G} \ni g$, is a function, called *density generator* and

$$\mathcal{G} = \left\{ g : \mathbb{R}_0^+ \to \mathbb{R}^+ \,\middle|\, \int_0^\infty t^{N/2-1} g(t) dt < \infty \right\}.$$

- The set of all SS pdfs can be described as:

$$\mathcal{S} = \left\{ p_Z | p_Z(\mathbf{z}) = 2^{-N/2} g\left(||\mathbf{z}||^2\right), \forall g \in \mathcal{G} \right\}.$$

[16] See [27] or [28, Ch. 3] for the proofs of these properties. A comprehensive list is also summarized in [29].

Property P2

- Let $s_N \triangleq 2\pi^{N/2}/\Gamma(N/2)$ be the surface area of the unit sphere $\mathbb{R}S^N$ in $\mathbb{R}^N$.

- The pdf of $\mathcal{Q} =_d ||\mathbf{z}||^2$, called *2nd-order modular variate*, is:

$$p_{\mathcal{Q}}(q) = s_N 2^{-N/2-1} q^{N/2-1} g\left(q\right).$$

- The pdf of $\mathcal{R} \triangleq \sqrt{\mathcal{Q}} =_d ||\mathbf{z}||$, called *modular variate*, is:

$$p_{\mathcal{R}}(r) = s_N 2^{-N/2} r^{N-1} g\left(r^2\right).$$

Properties of the (SS) distributions (3/4)

Property P3: *Stochastic Representation Theorem*

- Let $\mathbf{u} \sim \mathcal{U}(\mathbb{R}S^N)$ be a random vector uniformly distributed on $\mathbb{R}S^N$, i.e. $||\mathbf{u}|| = 1$.

- If $\mathbf{z} \in \mathbb{R}^N$ is SS-distributed, i.e. $\mathbf{z} \sim SS(g)$, then:

$$\mathbf{z} =_d \sqrt{\mathcal{Q}}\mathbf{u} =_d \mathcal{R}\mathbf{u},$$

- Moreover, $\mathcal{Q}$ and $\mathbf{u}$ (or $\mathcal{R}$ and $\mathbf{u}$) are independent.

- P2 and P3 imply that, not knowing the density generator g has an impact only on the pdf of the r.v. $\mathcal{R}$ (or $\mathcal{Q}$).

Properties of the (SS) distributions (4/4)

Property P4: *Invariant statistic*

- By definition of SS distributions, $||\cdot||$ is an *invariant statistic* since [30, Ch. 6]
$$||\mathbf{z}|| =_d ||\mathbf{O}\mathbf{z}||,$$
for every unitary matrix $\mathbf{O} \in \mathcal{O}$.

- Moreover, given two SS-distributed r.v. $\mathbf{z}_1$ and $\mathbf{z}_2$, we have:
$$||\mathbf{z}_1|| =_d ||\mathbf{z}_2|| \;\Rightarrow\; \mathbf{z}_1 =_d \mathbf{O}\mathbf{z}_2, \quad \forall \mathbf{O} \in \mathcal{O}.$$

- Then, the modular variate $\mathcal{R} =_d ||\mathbf{z}||$ is a *maximal invariant statistic* for the set of the SS-distributed random vectors.

Tangent space and invariance

- Let $\mathcal{A}$ be a group of transformations from $\mathbb{R}^N$ into itself:

$$\begin{aligned} \mathcal{A} \ni \alpha : \mathbb{R}^N &\to \mathbb{R}^N \\ \mathbf{z} &\mapsto \alpha(\mathbf{z}), \end{aligned}$$

- Suppose that $\mathcal{P}$ is a set of pdfs which are invariant with respect to $\mathcal{A}$, i.e.:

$$\mathcal{P} = \left\{ p_Z | p_Z(\alpha(\mathbf{z})) = p_Z(\mathbf{z}); \forall \alpha \in \mathcal{A}, \forall \mathbf{z} \in \mathbb{R}^N \right\}.$$

- Then, the tangent space $\mathcal{T}$ of $\mathcal{P}$ is given by [9, App. 3]: [17]

$$\mathcal{T} = \left\{ h \in \mathcal{H} | h(\alpha(\mathbf{z})) = h(\mathbf{z}), \forall \alpha \in \mathcal{A}, \forall \mathbf{z} \in \mathbb{R}^N \right\}$$

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

Projection and invariance

If there exists an invariant statistic D for $\mathbf{z} \sim p_Z$ s.t.:

$$D =_d D(\alpha(\mathbf{z})), \quad \forall \alpha \in \mathcal{A},$$

then the projection operator on $\mathcal{T}$ can be calculated as [9, App. 3]:

$$\Pi(\cdot|\mathcal{T}) = E\{\cdot|D\}.$$

Example: SS distributions

- The tangent space $\mathcal{T}_{\mathcal{S}}$ is given by:

$$\mathcal{T}_{\mathcal{S}} = \left\{ h \in \mathcal{H} | h(||\mathbf{z}||) = h(\mathbf{z}), \forall \mathbf{z} \in \mathbb{R}^N \right\},$$

- $\Pi(\cdot|\mathcal{T}_{\mathcal{S}}) = E\{\cdot|\mathcal{R}\}$ where $\mathcal{R} =_d ||\mathbf{z}||$ is the modular variate.

Parametric group models (1/2)

- Let $\mathcal{A}$ be a group of *parametric* transformations from $\mathbb{R}^N$ into itself:

$$\mathcal{A} = \{\alpha | \alpha(\cdot; \boldsymbol{\theta}) \triangleq \alpha_{\boldsymbol{\theta}}(\cdot); \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}.$$

  - $\alpha_{\boldsymbol{\theta}}^{-1}(\cdot)$ defines the inverse of $\alpha_{\boldsymbol{\theta}}(\cdot)$,
  - $(\alpha_{\boldsymbol{\theta}_2} \circ \alpha_{\boldsymbol{\theta}_1})(\cdot) \triangleq \alpha_{\boldsymbol{\theta}_2}(\alpha_{\boldsymbol{\theta}_1}(\cdot))$ denotes the composition,
  - $\boldsymbol{\theta}_e$ indicates the parameter vector that characterizes the identity transformation $\alpha_{\boldsymbol{\theta}_e}$, s.t. $\alpha_{\boldsymbol{\theta}_e}(\cdot) = \cdot$.

Example: Let us define $\boldsymbol{\theta} \triangleq [\mu, \sigma]^T$, then:

$$\alpha_{\boldsymbol{\theta}}(z) \triangleq \mu + \sigma z,$$

$$\alpha_{\boldsymbol{\theta}}^{-1}(z) = (z - \mu)/\sigma, \qquad \boldsymbol{\theta}_e \triangleq [0, 1]^T.$$

Parametric group models (2/2)

- Let $\mathbf{z} \in \mathbb{R}^N$ be a random vector s.t. $\mathbf{z} \sim p_Z(\mathbf{z})$.
- The *parametric group model*, generated by the action of $\mathcal{A}$ on $\mathbf{z}$ can be expressed as:

$$\mathcal{P}_{\boldsymbol{\theta}} = \left\{ p_X | p_X(\mathbf{x}|\boldsymbol{\theta}) = |\mathbf{J}(\alpha_{\boldsymbol{\theta}}^{-1})(\mathbf{x})| p_Z(\alpha_{\boldsymbol{\theta}}^{-1}(\mathbf{x})); \boldsymbol{\theta} \in \Theta \right\},$$

where:

  - $[\mathbf{J}(\alpha_{\boldsymbol{\theta}}^{-1})(\mathbf{x})]_{i,j} \triangleq \partial[\alpha^{-1}(\mathbf{x}; \boldsymbol{\theta})]_i / \partial\theta_j$ is the Jacobian matrix of the inverse transformation $\alpha_{\boldsymbol{\theta}}^{-1}$,
  - $|\cdot|$ defines the (absolute value of the) determinant of the Jacobian matrix.

Semiparametric group models (1/2)

- If p_Z is allowed to vary in a function set $\mathcal{L}$, we get a *semiparametric group model*:

$$\mathcal{P}_{\boldsymbol{\theta},p_Z} = \{p_X | \, p_X(\mathbf{x}|\boldsymbol{\theta}, p_Z) = |\mathbf{J}(\alpha_{\boldsymbol{\theta}}^{-1})(\mathbf{x})| p_Z(\alpha_{\boldsymbol{\theta}}^{-1}(\mathbf{x})); \\ \boldsymbol{\theta} \in \Theta, p_Z \in \mathcal{L}\} \, .$$

- The calculation of the projection operator can be greatly simplified!
  1. Evaluate the projection on the semiparametric nuisance tangent space at the identity $\alpha_{\boldsymbol{\theta}_e}$.
  2. "Translate" the projection in any other $\boldsymbol{\theta}$ of the parameter space Θ.

Semiparametric group models (2/2)

- $\mathcal{T}_{p_{Z,0}}(\boldsymbol{\theta}_e) \subseteq \mathcal{H}^q$: Semiparametric nuisance tangent space at the identity $\boldsymbol{\theta}_e$.
- $\mathcal{T}_{p_{Z,0}}(\boldsymbol{\theta}) \subseteq \mathcal{H}^q$: Semiparametric nuisance tangent space at a generic $\boldsymbol{\theta} \in \Theta$.

The projection operator on $\mathcal{T}_{p_{Z,0}}(\boldsymbol{\theta})$ can be obtained as [9, Sec. 4.2, Lemma 3]:

$$\Pi(\cdot|\mathcal{T}_{p_{Z,0}}(\boldsymbol{\theta})) = \Pi(\cdot \circ \alpha_{\boldsymbol{\theta}}|\mathcal{T}_{p_{Z,0}}(\boldsymbol{\theta}_e)) \circ \alpha_{\boldsymbol{\theta}}^{-1}, \quad \forall \boldsymbol{\theta} \in \Theta.$$

From SS to RES distributions (1/2)

- Let us define the parameter space $\Theta \subseteq \mathbb{R}^q$ as:

$$\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^q | \boldsymbol{\theta} = [\boldsymbol{\mu}^T, \mathrm{vecs}(\boldsymbol{\Sigma})^T]^T; \boldsymbol{\mu} \in \mathbb{R}^N, \boldsymbol{\Sigma} \in \mathcal{M}_N\}.$$

- We can define the group of parametric transformations $\mathcal{A}$ as:

$$\begin{aligned} \mathcal{A} \ni \alpha_{\boldsymbol{\theta}} : \mathbb{R}^N &\to \mathbb{R}^N, \ \forall \boldsymbol{\theta} \in \Theta \\ \mathbf{z} &\mapsto \alpha_{\boldsymbol{\theta}}(\mathbf{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}. \end{aligned}$$

  - The identity $\alpha_{\boldsymbol{\theta}_e}$ is parametrized by $\boldsymbol{\theta}_e = [\mathbf{0}^T, \mathrm{vecs}(\mathbf{I})^T]^T$,
  - The inverse is simply given by:

$$\alpha_{\boldsymbol{\theta}}^{-1}(\cdot) = \boldsymbol{\Sigma}^{-1/2}(\cdot - \boldsymbol{\mu}).$$

From SS to RES distributions (2/2)

- A random vector $\mathbf{x} \in \mathbb{R}^N$ is said to be RES-distributed if it can be expressed as:

$$\mathbf{x} = \alpha_{\boldsymbol{\theta}}(\mathbf{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z} =_d \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Sigma}^{1/2}\mathbf{u},$$

  - $\mathbf{z} \sim SS(g)$ is an SS-distributed random vector,
  - $\mathbf{u} \sim \mathcal{U}(\mathbb{R}S^N)$ and $\mathcal{R} = \sqrt{\mathcal{Q}}$ is the modular variate, s.t.:

$$\mathcal{Q} =_d ||\mathbf{z}||^2 = ||\alpha_{\boldsymbol{\theta}}^{-1}(\mathbf{x})||^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}).$$

- RES distributions represent a semiparametric group model:

$$\mathcal{P}_{\boldsymbol{\theta},g} = \left\{ p_X | p_X(\mathbf{x}|\boldsymbol{\theta}, g) = 2^{-N/2}|\boldsymbol{\Sigma}|^{-1/2} g(||\alpha_{\boldsymbol{\theta}}^{-1}(\mathbf{x})||^2); \right. \\ \left. \boldsymbol{\theta} \in \Theta, g \in \mathcal{G} \right\},$$

Part II - 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

Semiparametric interpretation of Real and Complex ES distributions

Examples

Evaluation of the SCRB for the RES class

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

$$\boldsymbol{\theta}_0 = [\boldsymbol{\mu}_0^T, \mathrm{vecs}(\boldsymbol{\Sigma}_0)^T]^T.$$

- **Problem**: Find the (Constrained) SCRB on the estimation of the mean vector $\boldsymbol{\mu}_0$ and of the scatter matrix $\boldsymbol{\Sigma}_0$ when the density generator g_0 is unknown.

- To avoid the ambiguity between $\boldsymbol{\Sigma}_0$ and g_0, we put a constraint on the scatter matrix:

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

- All the details can be found in [29].

Evaluation of the SCRB for the RES class

Step A: Evaluation of the score vector $\mathbf{s}_{\boldsymbol{\theta}_0}$

- By definition:

$$\mathbf{s}_{\boldsymbol{\theta}_0} = \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, g_0) = \begin{pmatrix} \mathbf{s}_{\boldsymbol{\mu}_0} \\ \mathbf{s}_{\mathrm{vecs}(\boldsymbol{\Sigma}_0)} \end{pmatrix}$$

- Through direct calculation, we get:

$$\mathbf{s}_{\boldsymbol{\mu}_0} =_d -2\sqrt{\mathcal{Q}}\psi_0(\mathcal{Q})\boldsymbol{\Sigma}_0^{-1/2}\mathbf{u},$$

$$\mathbf{s}_{\mathrm{vecs}(\boldsymbol{\Sigma}_0)} =_d -\mathbf{D}_N^T \Big(2^{-1}\mathrm{vec}(\boldsymbol{\Sigma}_0^{-1})+ \\ +\mathcal{Q}\psi_0(\mathcal{Q})\boldsymbol{\Sigma}_0^{-1/2} \otimes \boldsymbol{\Sigma}_0^{-1/2}\mathrm{vec}(\mathbf{u}\mathbf{u}^T)\Big),$$

  - $\psi_0(t) \triangleq d \ln g_0(t)/dt,$
  - *Duplication matrix*: $\mathbf{D}_N \mathrm{vecs}(\mathbf{A}) = \mathrm{vec}(\mathbf{A})$, $\forall \mathbf{A}$ symmetric.

Evaluation of the SCRB for the RES class

Step B: Evaluation of the projection operator $\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{g0})$

- Due to the group structure underlying the RES class, $\mathcal{T}_{g0}$ evaluated at the group identity $\boldsymbol{\theta}_e$ is given by:

$$\mathcal{T}_{g0}(\boldsymbol{\theta}_e) = \{\mathbf{l}|\mathbf{l} = h\mathbf{a}; h \in \mathcal{T}_{\mathcal{S}}, \mathbf{a} \in \mathbb{R}^q\};$$

where $\mathcal{T}_{\mathcal{S}}$ is the tangent space of the SS distributions:

$$\mathcal{T}_{\mathcal{S}} = \left\{h \in \mathcal{H}|h(||\mathbf{x}||) = h(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^N\right\},$$

- Using the property of the semiparametric group model:

$$\begin{aligned}\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{g0}(\boldsymbol{\theta}_0)) &= \Pi\left(\mathbf{s}_{\boldsymbol{\theta}_0} \circ \alpha_{\boldsymbol{\theta}_0}|\mathcal{T}_{g0}(\boldsymbol{\theta}_e)\right) \circ \alpha_{\boldsymbol{\theta}_0}^{-1} \\ &= E\left\{\mathbf{s}_{\boldsymbol{\theta}_0} \circ \alpha_{\boldsymbol{\theta}_0}|\mathcal{R}\right\} \circ \alpha_{\boldsymbol{\theta}_0}^{-1}.\end{aligned}$$

Evaluation of the SCRB for the RES class

- Through direct calculation (see [29] for the details):

$$\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{g_0}) = \begin{pmatrix} \Pi(\mathbf{s}_{\boldsymbol{\mu}_0}|\mathcal{T}_{g_0}) \\ \Pi(\mathbf{s}_{\mathrm{vecs}(\boldsymbol{\Sigma}_0)}|\mathcal{T}_{g_0}) \end{pmatrix}$$

$$=_d \begin{pmatrix} \mathbf{0} \\ -\mathbf{D}_N^T \left(\frac{1}{2} + \frac{1}{N}\mathcal{Q}\psi_0(\mathcal{Q})\right) \mathrm{vec}(\boldsymbol{\Sigma}_0^{-1}) \end{pmatrix}.$$

- The score function $\mathbf{s}_{\boldsymbol{\mu}_0}$ of the mean value is orthogonal to the nuisance tangent space $\mathcal{T}_{g_0}$,

- Not knowing the true g_0 does not have any impact in the (asymptotic) estimation performance of $\boldsymbol{\mu}_0$ [21].

Evaluation of the SCRB for the RES class

Step C: Evaluation of the semiparametric FIM $\bar{\mathbf{I}}(\boldsymbol{\theta}_0, g_0)$

- The efficient score vector $\bar{\mathbf{s}}_0$ can then be expressed as:

$$\bar{\mathbf{s}}_0 = \mathbf{s}_{\boldsymbol{\theta}_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0}(\mathbf{x})|\mathcal{T}_{g_0})$$
$$=_d \begin{pmatrix} -2\sqrt{\mathcal{Q}}\psi_0(\mathcal{Q})\boldsymbol{\Sigma}_0^{-1/2}\mathbf{u} \\ -\mathbf{D}_N^T\mathcal{Q}\psi_0(\mathcal{Q})\left(\boldsymbol{\Sigma}_0^{-1/2}\otimes\boldsymbol{\Sigma}_0^{-1/2}\mathrm{vec}(\mathbf{u}\mathbf{u}^T) - \frac{\mathrm{vec}(\boldsymbol{\Sigma}_0^{-1})}{N}\right) \end{pmatrix}.$$

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

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0) = E_0\{\bar{\mathbf{s}}_0\bar{\mathbf{s}}_0^T\}$$
$$= \begin{pmatrix} \mathbf{C}_0(\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}) & \mathbf{0} \\ \mathbf{0}^T & \mathbf{C}_0(\bar{\mathbf{s}}_{\mathrm{vecs}(\boldsymbol{\Sigma}_0)}) \end{pmatrix},$$

where $\mathbf{C}_0(\mathbf{h}) \triangleq E_0\{\mathbf{h}\mathbf{h}^T\}$, $\forall \mathbf{h} \in \mathcal{H}^q$.

Evaluation of the SCRB for the RES class

- Through direct calculation of the expectation, we get:

$$\mathbf{C}_0(\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}) = \frac{4E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\}}{N}\boldsymbol{\Sigma}_0^{-1},$$

and

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

- The block-diagonal structure of $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0)$ implies that the estimates of vector $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are asymptotically decoupled.

- $\boldsymbol{\mu}_0$ can be substituted with any consistent estimator without affecting the asymptotic performance of the scatter matrix estimator.

Evaluation of the SCRB for the RES class

Step D: Evaluation of the constrained $\mathrm{SCRB}(\boldsymbol{\theta}_0|g_0)$

- 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 $\mathbf{J_c}(\boldsymbol{\Sigma}_0)$ be the Jacobian matrix of the constraint, then there exists a matrix $\mathbf{U}$ s.t. [31,32]:

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

- The constrained $\mathrm{SCRB}(\boldsymbol{\theta}_0|g_0)$ can be expressed as:

$$\mathrm{CSCRB}(\boldsymbol{\theta}_0|g_0) = \begin{pmatrix} \frac{N}{4E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\}}\boldsymbol{\Sigma}_0 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{U}\left(\mathbf{U}^T\mathbf{C}_0(\bar{\mathbf{s}}_{\mathrm{vecs}(\boldsymbol{\Sigma}_0)})\mathbf{U}\right)^{-1}\mathbf{U}^T \end{pmatrix}.$$

Numerical results

- Let $\{\mathbf{x}_m\}_{m=1}^{M}$ be a set of M i.i.d. RES-distributed data, s.t.:

$$\mathbf{x}_m \sim RES_N(\mathbf{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0), \quad m = 1, \dots, M.$$

- Let us define $\{\bar{\mathbf{x}}_m\}_{m=1}^{M}$ as the set of M vectors such that:

$$\bar{\mathbf{x}}_m = \mathbf{x}_m - \hat{\boldsymbol{\mu}}, \quad m = 1, \dots, M,$$

and $\hat{\boldsymbol{\mu}}$ is the sample mean estimator, i.e.

$$\hat{\boldsymbol{\mu}} \triangleq M^{-1} \sum_{m=1}^{M} \mathbf{x}_m.$$

- $\hat{\boldsymbol{\mu}}$ is a consistent and unbiased estimator.

Three "semiparametric" estimators (1/3)

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

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

- The CSCM is given by:

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

Three “semiparametric” estimators (2/3)

- The C-Tyler and the C-Hub are given by the convergence point of the following recursion:

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

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

- The weight function $\varphi(t)$ for Tyler’s estimator is [33,8]:

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

Three "semiparametric" estimators (3/3)

- The weight function for Huber's estimator is given by [24,34]

$$\varphi_{Hub}(t) = \begin{cases} 1/b & t \leqslant \delta^2 \\ \delta^2/(tb) & t > \delta^2 \end{cases},$$

  and
  - $\delta = F_{\chi^2_N}(u)$, [18]
  - $b = F_{\chi^2_{N+2}}(\delta^2) + \delta^2(1 - F_{\chi^2_N}(\delta^2))/N$ [8], [34].

- u is a tuning parameter that controls the trade-off between robustness and efficiency.

- For $u \to 1$ Huber's estimator is equal to the SCM, while for $u \to 0$ Huber's estimator tends to Tyler's estimator.

[18] $F_{\chi^2_N}(\cdot)$ indicates the distribution of a chi-squared random variable with N degrees of freedom.

Simulation setup

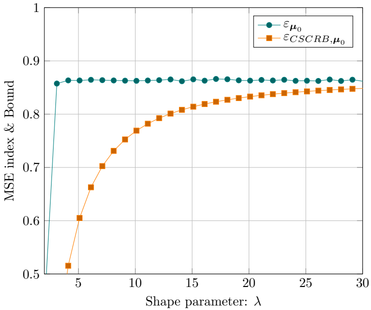
- Two different "true" distributions are considered:
  1. The t-distribution,
  2. The Generalized Gaussian (GG) distribution.

- Simulation parameters
  - $[\boldsymbol{\Sigma}_0]_{i,j} = \rho^{|i-j|}$, $\rho = 0.8$ $i, j = 1, \ldots, N$. Moreover $N = 8$,
  - The data power is chosen to be $\sigma_X^2 = E_{\mathcal{Q}}\{\mathcal{Q}\}/N = 4$,
  - The data mean value is chosen to be $[\boldsymbol{\mu}_0]_i = 1$, $i = 1, \ldots, N$,
  - The number of the available i.i.d. data vectors is $M = 3N = 24$,
  - The tuning parameter u of Huber's estimator $u = 0.5$.

- The MSE of the scatter matrix estimators is compared with:
  1. The $\mathrm{CSCRB}(\boldsymbol{\theta}_0|g_0)$ previously derived,
  2. The classical constrained CRB, i.e. $\mathrm{CCRB}(\boldsymbol{\theta}_0)$, evaluated under perfect knowledge of the density generator [35,36].

t-distribution - Mean vector

$$\varepsilon_{\boldsymbol{\mu}_0} \triangleq ||E\{(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T\}||_F, \quad \varepsilon_{CSCRB,\boldsymbol{\mu}_0} \triangleq ||[\mathrm{CSCRB}(\boldsymbol{\theta}_0|g_0)]_{\boldsymbol{\mu}_0}||_F.$$

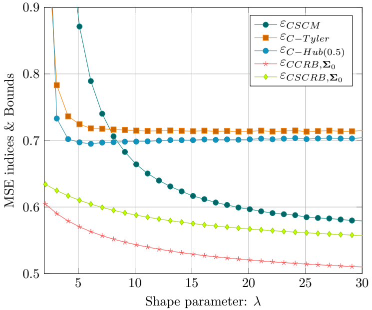


- For the estimation of $\boldsymbol{\mu}_0$, CSCRB coincides with CCRB.
- When the shape parameter λ goes to infinity, the t-distribution tends to a Gaussian one.
- Then, for $\lambda \to \infty$, the sample mean tends to be efficient.

t-distribution - Scatter matrix

$$\varepsilon_\alpha \triangleq ||E\{(\mathrm{vecs}(\hat{\mathbf{\Sigma}}_\alpha) - \mathrm{vecs}(\mathbf{\Sigma}_0))(\mathrm{vecs}(\hat{\mathbf{\Sigma}}_\alpha) - \mathrm{vecs}(\mathbf{\Sigma}_0))^T\}||_F,$$

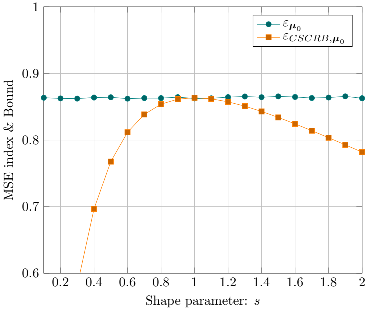
$$\varepsilon_{CSCRB,\mathbf{\Sigma}_0} \triangleq ||[\mathrm{CSCRB}(\boldsymbol{\theta}_0|g_0)]_{\mathbf{\Sigma}_0}||_F, \quad \varepsilon_{CCRB,\mathbf{\Sigma}_0} \triangleq ||[\mathrm{CCRB}(\boldsymbol{\theta}_0)]_{\mathbf{\Sigma}_0}||_F.$$



- The CSCM tends to be efficient w.r.t. the CSCRB as $\lambda \to \infty$.
- Both C-Tyler's and C-Huber's estimators are not efficient with respect to the CSCRB.

GG distribution - Mean vector

$$\varepsilon_{\boldsymbol{\mu}_0} \triangleq ||E\{(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T\}||_F, \quad \varepsilon_{CSCRB,\boldsymbol{\mu}_0} \triangleq ||[\mathrm{CSCRB}(\boldsymbol{\theta}_0|g_0)]_{\boldsymbol{\mu}_0}||_F.$$

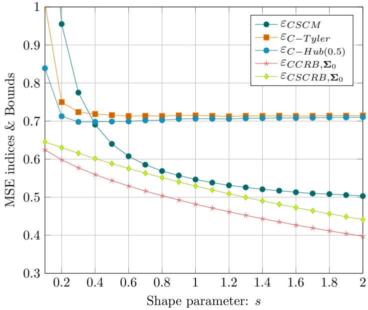


- When $s = 1$, the GG distribution is exactly Gaussian one.
- Hence, for $s = 1$, the sample mean is an efficient estimator.

GG distribution - Scatter matrix

$$\varepsilon_\alpha \triangleq ||E\{(\text{vecs}(\hat{\boldsymbol{\Sigma}}_\alpha) - \text{vecs}(\boldsymbol{\Sigma}_0))(\text{vecs}(\hat{\boldsymbol{\Sigma}}_\alpha) - \text{vecs}(\boldsymbol{\Sigma}_0))^T\}||_F,$$

$$\varepsilon_{CSCRB,\boldsymbol{\Sigma}_0} \triangleq ||[\text{CSCRB}(\boldsymbol{\theta}_0|g_0)]_{\boldsymbol{\Sigma}_0}||_F, \quad \varepsilon_{CCRB,\boldsymbol{\Sigma}_0} \triangleq ||[\text{CCRB}(\boldsymbol{\theta}_0)]_{\boldsymbol{\Sigma}_0}||_F.$$



- The lack of knowledge of the particular density generator has an higher impact when the tails of the true distribution become lighter [37].

The SCRB for the CES class

- The derivation of:[19]
  - SCRB for the estimation of the mean vector and of the scatter matrix in CES distributed random vectors,
  - The Semiparametric Slepian-Bangs formula,
  - The Semiparametric Stochastic CRB (SSCRB),

  can be found in [38]:

  S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "Semiparametric CRB and Slepian-Bangs formulas for Complex Elliptically Symmetric distributions," *accepted in IEEE Transactions on Signal Processing*, 2019. [Online]. Available: http://arxiv.org/abs/1902.09541.

- The application of these theoretical results to Direction of Arrival (DOA) estimation problems is discussed in [39]:

  S. Fortunati, F. Gini, M. S. Greco, "Semiparametric stochastic CRB for DOA estimation in elliptical data model," in 2019 27th European Signal Processing Conference, *EUSIPCO*, Sep. 2019.

[19] Additional details are given in the backup slides.

Conclusions

- We provided a fresh look to the Semiparametric Cramér-Rao Bound (SCRB) by showing its relations with the classical (parametric) CRB [7].
- The link between parametric and semiparametric framework is given by the Hilbert-space geometry underling any inference problem.
- The application of the SCRB to the scatter matrix estimation in RES and CES distributed data has been discussed.
- Future works will explore possible applications of the semiparametric inference to well-known signal processing problems, in particular the *semiparametric detection*.

Acknowledgement

- Prof. Maria Sabrina Greco,
  University of Pisa, Italy.
- Prf. Abdelhak M. Zoubir,
  Technische Universität Darmstadt, Germany.
- Prof. Christ D. Richmond,
  Arizona State University, USA.
- Prof. Muralidhar Rangaswamy,
  U.S. AFRL, Sensors Directorate, Wright-Patterson AFB, USA.

References

[1] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2263–2278, 2015.

[2] S. Fortunati, F. Gini, and M. Greco, "The Misspecified Cramér-Rao bound and its application to scatter matrix estimation in complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2387 – 2399, 2016.

[3] S. Fortunati, F. Gini, and M. S. Greco, "Chapter 4 - parameter bounds under misspecified models for adaptive radar detection," in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2018, pp. 197 – 252. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128118870000043

[4] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 142–157, Nov 2017.

References

[5] S. Fortunati, F. Gini, and M. S. Greco, "Matched, mismatched, and robust scatter matrix estimation and hypothesis testing in complex t-distributed data," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 123, 2016. [Online]. Available: https://doi.org/10.1186/s13634-016-0417-0

[6] S. Fortunati, "Misspecified Cramér-Rao bounds for complex unconstrained and constrained parameters," in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 1644–1648.

[7] S. Fortunati, F. Gini, M. Greco, A. M. Zoubir, and M. Rangaswamy, "A fresh look at the semiparametric Cramér-Rao bound," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 261–265.

[8] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

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

References

[10] A. Tsiatis, *Semiparametric Theory and Missing Data*. Springer series in statistics, 2006.

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

[12] M. Hallin and B. J. Werker, "Optimal testing for semi-parametric AR models from Gaussian Lagrange multipliers to autoregression rank scores and adaptive tests," in *Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life*, N. Balakrishnan, M. Nikulin, M. Mesbah, and N. Limnios, Eds. Birkhäuser, Boston, MA: Statistics for Industry and Technology, 2004, pp. 295–350.

[13] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.

[14] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," nov 2012, version 20121115. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274

References

[15] L. Debnath and P. Mikusiński, *Introduction to Hilbert Spaces with Applications (Second Edition)*. Academic Press, 1999.

[16] C. R. Blyth, "Necessary and sufficient conditions for inequalities of cramér-rao type," *The Annals of Statistics*, vol. 2, no. 3, pp. 464–473, 1974.

[17] Y. Koshevnik and B. Levit, "On a non-parametric analogue of the information matrix," *Theory of Probability & Its Applications*, vol. 21, no. 4, pp. 738–753, 1977.

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

[19] 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.

[20] H. Rieder, *Robust Asymptotic Statistics*. Springer series in statistics, 1994.

[21] P. J. Bickel, "On adaptive estimation," *The Annals of Statistics*, vol. 10, no. 3, pp. 647–671, 1982.

References

[22] W. K. Newey, “The asymptotic variance of semiparametric estimators,” *Econometrica*, vol. 62, no. 6, pp. 1349–1382, 1994.

[23] C. A. J. Klaassen, “Consistent estimation of the influence function of locally asymptotically linear estimators,” *Ann. Statist.*, vol. 15, no. 4, pp. 1548–1562, 12 1987.

[24] P. J. Huber and E. M. Ronchetti, *Robust Statistics (Second Edition)*. Wiley Series in Probability and Statistics, 2009.

[25] Y. Ritov and P. J. Bickel, “Achieving information bounds in non and semiparametric models,” *Ann. Statist.*, vol. 18, no. 2, pp. 925–938, 06 1990. [Online]. Available: https://doi.org/10.1214/aos/1176347633

[26] J. Jacod and P. Protter, *Probability Essentials*. Springer series in statistics, 2004.

[27] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*. Monographs on Statistics and Applied Probability, Springer US, 1990.

References

[28] C. D. Richmond, "Adaptive array signal processing and performance analysis in non-Gaussian environments," Ph.D. dissertation, Massachusetts Institute of Technology, 1996. [Online]. Available: https://dspace.mit.edu/handle/1721.1/11005

[29] 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.

[30] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer Texts in Statistics, 2004.

[31] 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.

[32] P. Stoica and B. C. Ng, "On the Cramér-Rao Bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, July 1998.

References

[33] D. E. Tyler, "A distribution-free *m*-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 03 1987. [Online]. Available: https://doi.org/10.1214/aos/1176350263

[34] E. Ollila, I. Soloveychik, D. E. Tyler, and A. Wiesel, "Simultaneous penalized M-estimation of covariance matrices using geodesically convex optimization," *Submitted to Journal of Multivariate Analysis*. [Online]. Available: https://arxiv.org/abs/1608.08126

[35] M. Greco and F. Gini, "Cramér-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing,*, vol. 61, no. 24, pp. 6401–6409, 2013.

[36] O. Besson and Y. I. Abramovich, "On the Fisher Information Matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1130–1133, Nov 2013.

[37] M. Hallin and D. Paindaveine, "Parametric and semiparametric inference for shape: the role of the scale functional," *Statistics & Decisions*, vol. 24, no. 3, pp. 327–350, 2009.

References

[38] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, “Semiparametric CRB and Slepian-Bangs formulas for Complex Elliptically Symmetric distributions,” *Accepted in IEEE Transactions on Signal Processing*, 2019. [Online]. Available: http://arxiv.org/abs/1902.09541

[39] S. Fortunati, F. Gini, and M. Greco, “Semiparametric stochastic CRB for DOA estimation in elliptical data model,” in *2019 27th European Signal Processing Conference (EUSIPCO)*, Sep. 2019. [Online]. Available: https://arxiv.org/abs/1903.00403

[40] P. Krishnaiah and J. Lin, “Complex elliptically symmetric distributions,” *Communications in Statistics - Theory and Methods*, vol. 15, no. 12, pp. 3693–3718, 1986.

[41] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[42] A. van den Bos, “Complex gradient and hessian,” *IEE Proceedings - Vision, Image and Signal Processing*, vol. 141, no. 6, pp. 380–383, Dec 1994.

[43] R. Remmert, *Theory of Complex Functions*. New York: Springer, 1991.

References

[44] P. J. Schreirer and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: the Theory of Improper and Noncircular Signals*. Cambridge UK: Cambridge Univ. Press, 2010.

[45] J. Eriksson, E. Ollila, and V. Koivunen, "Essential statistics and tools for complex random variables," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5400–5408, Oct 2010.

[46] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5101–5125, Nov 2011.

[47] H. Li and T. Adalı, "Complex-valued adaptive signal processing using nonlinear functions," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 765615, Feb 2008.

[48] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus." [Online]. Available: https://arxiv.org/abs/0906.4835

[49] A. Hjørungnes, *Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications*. Cambridge University Press, 2011.

References

[50] D. Slepian, "Estimation of signal parameters in the presence of noise," *Transactions of the IRE Professional Group on Information Theory*, vol. 3, no. 3, pp. 68–89, 1954.

[51] W. J. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Yale University, New Haven, CT, USA, 1971.

[52] J. Delmas and H. Abeida, "Stochastic Cramér-Rao bound for noncircular signals with application to DOA estimation," *IEEE Transactions on Signal Processing*, vol. 52, no. 11, pp. 3192–3199, Nov 2004.

[53] H. Abeida, "Fisher Information Matrix for non-circular Complex Elliptical Symmetric distributions with application to DOA estimation," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017.

[54] A. Mennad, S. Fortunati, M. N. E. Korso, A. Younsi, A. M. Zoubir, and A. Renaux, "Slepian-bangs-type formulas and the related misspecified Cramér-Rao bounds for complex elliptically symmetric distributions," *Signal Processing*, vol. 142, pp. 320 – 329, 2018.

References

[55] E. Ollila and V. Koivunen, “Influence function and asymptotic efficiency of scatter matrix based array processors: Case MVDR beamformer,” *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 247–259, Jan 2009.

[56] P. Stoica, E. G. Larsson, and A. B. Gershman, “The stochastic CRB for array processing: a textbook derivation,” *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, May 2001.

[57] P. Stoica and A. Nehorai, “Performance study of conditional and unconditional direction-of-arrival estimation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, Oct 1990.

Backup slides

σ-algebras and measures

- Let $\mathcal{X}$ be some set and let $2^{\mathcal{X}}$ represent its power set. Then a subset $\mathfrak{F} \subseteq 2^{\mathcal{X}}$ is called a σ-algebra if (see e.g. [26, Ch. 2]):
  1. $\mathcal{X} \in \mathfrak{F}$,
  2. If $A \in \mathcal{X}$ is in $\mathfrak{F}$, then so is its complement, $\mathcal{X} \setminus A$,
  3. If $\{A_i\}_{i \in \mathbb{N}} \in \mathfrak{F}$, then so $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.
- A function $\mu : \mathfrak{F} \to [0, \infty)$ is called a measure if:
  1. $\mu(\emptyset) = 0$ (*Null empty set*),
  2. For all countable collections $\{A_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\mathfrak{F}$, $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ (*Countable additivity*).
- The couple $(\mathcal{X}, \mathfrak{F})$ is a *measurable space*, while the triplet $(\mathcal{X}, \mathfrak{F}, \mu)$ is a *measure space*.

Probability spaces and random variables

- A probability space is a measure space $(\Omega, \mathfrak{D}, P)$ where:
    1. Ω is the *sample space* that represents the set of all possible outcomes of a random experiment,
    2. $\mathfrak{D}$ is the σ-algebra on Ω,
    3. P is a probability measure, that is a measure $P : \mathfrak{D} \to [0, 1]$ satisfying $P(\Omega) = 1$.

- Let $(\Omega, \mathfrak{D}, P)$ be a probability space and $(\mathcal{X}, \mathfrak{F})$ a measurable space.

A *random variable* (r.v.) X is a measurable function $X : \Omega \to \mathcal{X}$, that is for every subset $A \in \mathfrak{F}$, its preimage

$$X^{-1}(A) \triangleq \{\omega \in \Omega | X(\omega) \in A\},$$

is an element of the σ-algebra $\mathfrak{D}$, i.e. $X^{-1}(B) \in \mathfrak{D}$.

Distribution and density functions

- A r.v. allows us to "transport" the probability structure, defined in the abstract space $(\Omega, \mathfrak{D}, P)$, in $(\mathcal{X}, \mathfrak{F})$.

- Specifically, a new probability measure can be defined on $(\mathcal{X}, \mathfrak{F})$ as follows:

$$P_X(A) \triangleq P\left(\{\omega \in \Omega | X(\omega) \in A\}\right) = P(X^{-1}(A)), \quad A \in \mathfrak{F}.$$

- Consequently, the triplet $(\mathcal{X}, \mathfrak{F}, P_X)$ is a probability space.

- **Example:** If $\mathcal{X} \equiv \mathbb{R}$ and $\mathfrak{F}$ is the Borel σ-algebra on $\mathbb{R}$, then P_X is the *distribution* of X [26, Ch. 11].

- The *density* p_X of X is a measurable function satisfying:

$$P_X\left((-\infty, x]\right) = \int_{-\infty}^{x} p_X(a)da, \quad \forall x \in \mathbb{R}.$$

Sub-σ-algebra generated by a transformation

- Let $(\mathcal{X}, \mathfrak{F}, P_X)$ be a probability space as previously defined.
- Let $T : (\mathcal{X}, \mathfrak{F}) \to (\mathcal{Y}, \mathfrak{L})$ a measurable transformation on $\mathcal{X}$.
- The preimage of T, i.e.:

$$\mathfrak{G}(T) \triangleq \left\{G \in \mathfrak{F} | G = T^{-1}(A),\ A \in \mathfrak{L}\right\}$$

  may be a coarser subset of $\mathfrak{F}$!
- It can be shown that $\mathfrak{G}(T)$ is a σ-algebra [26, Theo. 8.1] and, clearly, $\mathfrak{G}(T) \subseteq \mathfrak{F}$.
- $\mathfrak{G}(T)$ is then indicated as the sub-σ-algebra generated by the transformation T [26, Def. 23.3].

Proof: Finite-dimensionality of the linear span

Theorem

Let $\mathbf{u} = (u_1, \cdots, u_k)^T$ be a column vector of k arbitrary elements of an infinite-dimensional Hilbert space $\mathcal{F}$. The linear span of $\mathbf{u}$, 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 $\mathcal{F}^q$. Moreover, if $u_1, \cdots, u_k$ are linearly independent in $\mathcal{F}$, then $\dim(\mathcal{V}) = kq$.

Proof

- Assume that the entries of $\mathbf{u}$ are linearly independent.
- The dimension of a (finite-dimensional) space is equal to the minimum number of linearly independent vectors required to span it.

Proof: Finite-dimensionality of the linear span

- Then if $\mathcal{V}$ has dimension qk, there must exist qk linearly independent q-dimensional vectors such that $\mathcal{V} = \mathrm{span}\{\mathbf{v}_{11}, \ldots, \mathbf{v}_{1k}, \mathbf{v}_{q1}, \ldots, \mathbf{v}_{q \cdot k}\}$.

- Each vector $\mathbf{v}_{ij}$, $i = 1, \ldots, q; j = 1, \ldots, k$ can be constructed by putting all except the i-th entry equal to 0 and the i-th entry equal to $u_j \in \mathcal{F}$ for $j = 1, \ldots, k$, i.e:

$$\begin{array}{cccccccc} \mathbf{v}_{11} & \cdots & \mathbf{v}_{1k} & \mathbf{v}_{21} & \cdots & \mathbf{v}_{2k} & \\ \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \\ \begin{pmatrix} v_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \cdots & \begin{pmatrix} v_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ v_1 \\ \vdots \\ 0 \end{pmatrix} & \cdots & \begin{pmatrix} 0 \\ v_k \\ \vdots \\ 0 \end{pmatrix} & \cdots \end{array}$$

- By visual inspection, it is immediate to verify that they are linearly independent and this conclude the proof.

Parametric submodels of the CES model (1/3)

- A CES (zero-mean) random vector $\mathbf{x} \in \mathbb{C}^N$ admits a pdf [8]:

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

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

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

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

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

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

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:

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

  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:

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

- It is immediate to verify that this submodel satisfies the conditions C0, C1 and C2 given in slide 32.

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

Hellinger differentiability

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

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

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

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

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

where $||u_{\boldsymbol{\theta}}||^2 = \langle u_{\boldsymbol{\theta}}, u_{\boldsymbol{\theta}} \rangle = \int u_{\boldsymbol{\theta}}^2(\mathbf{x}) d\mathbf{x}$.

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

A geometrical intuition (1/4)

- Since $u_{\boldsymbol{\theta}}(\mathbf{x}) \triangleq \sqrt{p_X(\mathbf{x}|\boldsymbol{\theta})}$, we have that:

$$||u_{\boldsymbol{\theta}}||^2 = \langle u_{\boldsymbol{\theta}}, u_{\boldsymbol{\theta}} \rangle = \int p_X(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} = 1, \quad \forall \boldsymbol{\theta} \in \Theta.$$

- $u_{\boldsymbol{\theta}}$ can be interpreted as a differentiable map between Θ and (a subset of) the *surface* $S(L_2)$ of the unit sphere in L_2.

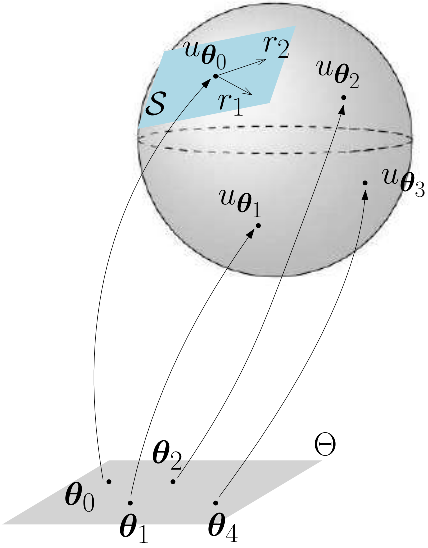
- Given a point on $S(L_2)$, say $u_{\boldsymbol{\theta}_0}$, the tangent space $\mathcal{S} \subseteq L_2$ of $\mathcal{S}_0$ at $u_{\boldsymbol{\theta}_0}$ is defined by the orthogonality condition:

$$\langle r, u_{\boldsymbol{\theta}_0} \rangle = 0 \quad \Leftrightarrow \quad r \in \mathcal{S}.$$

- Note that the tangent space $\mathcal{S}_0$ is a subset of L_2, while previously we defined it as a subset of $\mathcal{H}$.[20]

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

A geometrical intuition (2/4)



A geometrical intuition (3/4)

- Are the two definition consistent?

- Let us define the (locally) one-to-one transformation:

$$\begin{aligned} H_0 : \mathcal{S} &\to \mathcal{H} \\ r &\mapsto H_0(r) \triangleq \frac{2r}{u_{\boldsymbol{\theta}_0}} = h. \end{aligned}$$

- Then, we have:

$$\begin{aligned} r \in \mathcal{S} \Rightarrow \langle r, u_{\boldsymbol{\theta}_0} \rangle &= \int r(\mathbf{x}) u_{\boldsymbol{\theta}_0}(\mathbf{x}) d\mathbf{x} = 0 \\ &\Rightarrow 2^{-1} \int h(\mathbf{x}) u^2_{\boldsymbol{\theta}_0}(\mathbf{x}) = 2^{-1} \int h(\mathbf{x}) p(\mathbf{x}|\boldsymbol{\theta}_0) d\mathbf{x} = 0 \\ &\Rightarrow E_X\{h\} = 0 \Rightarrow h \in \mathcal{H}. \end{aligned}$$

A geometrical intuition (4/4)

- The vice-versa is as follows:

$$h \in \mathcal{H} \Rightarrow E_X\{h\} = \int h(\mathbf{x})p(\mathbf{x}|\boldsymbol{\theta}_0)d\mathbf{x} = 0$$
$$\Rightarrow 2\int r(\mathbf{x})u_{\boldsymbol{\theta}_0}^{-1}(\mathbf{x})p(\mathbf{x}|\boldsymbol{\theta}_0)d\mathbf{x} = 2\int r(\mathbf{x})u_{\boldsymbol{\theta}_0}(\mathbf{x})d\mathbf{x} = 0$$
$$\Rightarrow \langle r, u_{\boldsymbol{\theta}_0}\rangle = 0 \Rightarrow r \in \mathcal{S}.$$

Then the two definition are consistent [9, Sec. 3.1, Prep. 3]:

$$\langle r, u_{\boldsymbol{\theta}_0}\rangle = 0,\ \forall r \in \mathcal{S} \quad \Leftrightarrow \quad E_X\{h\} = 0,\ \forall h \in \mathcal{H}.$$

Hellinger derivative and score vector

- Recall that the score vector of $p_X(\mathbf{x}|\boldsymbol{\theta})$ in $\boldsymbol{\theta}_0$ is defined as:

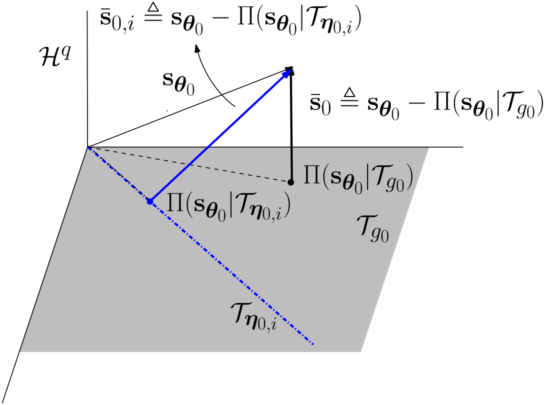
$$\mathbf{s}_{\boldsymbol{\theta}_0} \triangleq \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0).$$

- If for all $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ [9, Sec. 2.1, Prep. 1]:
  - $p_X(\mathbf{x}|\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$ for almost all $\mathbf{x}$,
  - $\left(\sum_i [\mathbf{s}_{\boldsymbol{\theta}_0}]_i^2\right)^{1/2} \in L_2(P_0)$,
  - The FIM $\mathbf{I}(\boldsymbol{\theta}) \triangleq \int \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x})\mathbf{s}_{\boldsymbol{\theta}}^T(\mathbf{x})p_X(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}$ is non-singular and continuous in $\boldsymbol{\theta}$,

  then [9, Sec. 2.1], we have that:

$$\dot{\mathbf{u}}_{\boldsymbol{\theta}_0} = \frac{1}{2} u_{\boldsymbol{\theta}_0} \mathbf{s}_{\boldsymbol{\theta}_0}, \quad \dot{\mathbf{u}}_{\boldsymbol{\theta}_0} \in \mathcal{S}^q, \ \mathbf{s}_{\boldsymbol{\theta}_0} \in \mathcal{H}^q.$$

The Semiparametric CRB (SCRB)



$$\mathcal{T}_{\boldsymbol{\eta}_{0,i}} \subseteq \mathcal{T}_{g_0}, \forall i \in \mathcal{I} \quad \Rightarrow \quad ||\bar{\mathbf{s}}_{0,i}|| \geq ||\bar{\mathbf{s}}_0||, \forall i \in \mathcal{I}$$

$$\Rightarrow \quad E_0\{\bar{\mathbf{s}}_{0,i}\bar{\mathbf{s}}_{0,i}^T\} \geq E_0\{\bar{\mathbf{s}}_0\bar{\mathbf{s}}_0^T\} \triangleq \bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0)$$

The Least Favourable Submodel (1/2)

- The Least Favourable Submodel (LFS) (if it exists) is the $\bar{i}$*-th* parametric submodel of $\mathcal{P}_{\boldsymbol{\theta},g}$ s.t.:

$$\sup_{\{\mathcal{P}_{\boldsymbol{\theta},\nu_i}\}} \left[E_0\{\bar{\mathbf{s}}_{0,i}\bar{\mathbf{s}}_{0,i}^T\}\right]^{-1} = \max_{\{\mathcal{P}_{\boldsymbol{\theta},\nu_i}\}} \left[E_0\{\bar{\mathbf{s}}_{0,i}\bar{\mathbf{s}}_{0,i}^T\}\right]^{-1}$$
$$= \bar{\mathbf{I}}(\boldsymbol{\theta}_0|\nu_{\bar{i}})^{-1},$$

- Let us define as Least Favourable Direction (LFD) the score vector [9, Sec. 3.1], [11, Sec. 2.2]:

$$\mathbf{s}_{\boldsymbol{\eta}_{0,\bar{i}}}(\mathbf{x}) = \nabla_{\boldsymbol{\eta}} \ln p_X(\mathbf{x}|\gamma_0, \nu_{\bar{i}}(\mathbf{x}, \boldsymbol{\eta}_0)),$$

- Then, as shown previously, for the parametric case:

$$\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{\boldsymbol{\eta}_{0,\bar{i}}}) = E_0\{\mathbf{s}_{\boldsymbol{\theta}_0}\mathbf{s}_{\boldsymbol{\eta}_{0,\bar{i}}}^T\}\mathbf{C}_0(\mathbf{s}_{\boldsymbol{\eta}_{0,\bar{i}}})^{-1}\mathbf{s}_{\boldsymbol{\eta}_{0,\bar{i}}}.$$

The Least Favourable Submodel (2/2)

- The existence of a LFS depends on the "level of richness" of the set of the parametric submodels $\{\mathcal{P}_{\boldsymbol{\theta},\nu_i}\}_{i \in \mathcal{I}}$.
- Unfortunately, the existence of a LFS needs to be verified on a case-by-case basis.
- Moreover, if it exists, figuring out which such LFS is, is not an easy task (see [11] for some hints on this).
- We refer to [9] for an exhaustive list of semiparametric models that admits a LFS expressible in "closed-form".

Conditional expectation: a remark (1/2)

- Let $h \equiv h(X)$ be a function of the random variable (r.v.) X.
- We defined the conditional expectation as $E\{h(X)|Y\}$ as the unique function of the r.v. Y such that:

$$E\{[h(X) - E\{h(X)|Y\}]Y\} = 0.$$

- The explicit "operative definition" of $E\{h(X)|Y\}$ is:

$$\begin{aligned} E\{h(X)|Y\} &\triangleq \int_{\mathcal{X}} h(x) p_{X|Y}(x|y) dx \\ &= \int_{\mathcal{X}} h(x) \frac{p_{X,Y}(x,y)}{p_Y(y)} dx, \end{aligned}$$

where $p_{X,Y}$ is the joint pdf of X and Y, $p_{X|Y}$ is the conditional pdf of X given Y and p_Y is the pdf of Y.

Conditional expectation: a remark (2/2)

- Are the two definitions consistent?

$$E\{[h(X) - E\{h(X)|Y\}]Y\} = 0 \Rightarrow$$
$$\int_{\mathcal{X},\mathcal{Y}} [h(X) - E\{h(X)|Y = y\}]p_{X,Y}(x,y)dxdy = 0$$

$$\begin{aligned}
\int_{\mathcal{X},\mathcal{Y}} & h(x)p_{X,Y}(x,y)dxdy \\
&= \int_{\mathcal{X},\mathcal{Y}} E\{h(X)|Y = y\}p_{X,Y}(x,y)dxdy \\
&= \int_{\mathcal{Y}} E\{h(X)|Y = y\}p_Y(y)dy \\
&= \int_{\mathcal{Y}} \left[\int_{\mathcal{X}} h(x)\frac{p_{X,Y}(x,y)}{p_Y(y)}dx\right] p_Y(y)dy \\
&= \int_{\mathcal{X},\mathcal{Y}} h(x)p_{X,Y}(x,y)dxdy. \qquad \blacksquare
\end{aligned}$$

From RES to CES distributions (1/3)

Definition ([40], [28], [8] and [41, Ch. 4])

- Let $\mathbf{x}_R \in \mathbb{R}^N$ and $\mathbf{x}_I \in \mathbb{R}^N$ be two real random vectors.

- $\mathbf{z} \triangleq \mathbf{x}_R + j\mathbf{x}_I \in \mathbb{C}^N$ is said to be CES-distributed with mean vector $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\mu} = \boldsymbol{\mu}_R + j\boldsymbol{\mu}_I \in \mathbb{C}^N \quad \boldsymbol{\Sigma} = \mathbf{C}_1 + j\mathbf{C}_2 \in \mathbb{C}^{N\times N},$$

  iff $\tilde{\mathbf{x}} \triangleq (\mathbf{x}_R^T, \mathbf{x}_I^T)^T \in \mathbb{R}^{2N}$ is RES-distributed with mean vector $\tilde{\boldsymbol{\mu}} = (\boldsymbol{\mu}_R^T, \boldsymbol{\mu}_I^T)^T$ and scatter matrix $\tilde{\boldsymbol{\Sigma}}$ satisfying:

$$\tilde{\boldsymbol{\Sigma}} = \frac{1}{2}\begin{pmatrix} \mathbf{C}_1 & -\mathbf{C}_2 \\ \mathbf{C}_2 & \mathbf{C}_1 \end{pmatrix},$$

  where $\mathbf{C}_1$ is symmetric and $\mathbf{C}_2$ is skew-symmetric.

From RES to CES distributions (2/3)

- Let $\tilde{\mathbf{x}} \sim RES_{2N}(\tilde{\mathbf{x}}; \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, g)$ be a RES-distributed random vector.

- When the scatter matrix $\tilde{\boldsymbol{\Sigma}}$ has full rank, we have that:

$$\begin{aligned} RES_{2N}(\tilde{\mathbf{x}}; \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, g) &\triangleq p_{\tilde{X}}(\tilde{\mathbf{x}}; \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, g) \\ &= 2^{-(2N)/2} |\tilde{\boldsymbol{\Sigma}}|^{-1/2} g\left((\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}})^T \right) \\ &= |\boldsymbol{\Sigma}|^{-1} g\left(2(\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right) \\ &= p_Z(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h) \triangleq CES_N(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, h), \end{aligned}$$

where $h(t) \triangleq g(2t)$.

- The functional form of the density generator remains unchanged except for the scaling factor 2 of its argument.

From RES to CES distributions (2/3)

- There exists a one-to-one mapping between a subset of the RES distributions and the (circular) CES distributions.

- The semiparametric theory already developed for the RES class holds true for the CES class as well.

- In particular, CES distributions are a *semiparametric group model* generated by the set of Complex Spherically Symmetric (CSS) distributions [28, Sec. 3.5] through the action of:

$$\alpha_{(\boldsymbol{\mu},\boldsymbol{\Sigma})} : \mathbb{C}^N \to \mathbb{C}^N, \ \forall \boldsymbol{\mu}, \boldsymbol{\Sigma}$$
$$CSS(g) \sim \mathbf{z} \mapsto \alpha_{(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\mathbf{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}.$$

The SCRB for the CES class

- The steps to derive the SCRB for the CES class follow exactly the ones already discussed for the RES one.
- Difference: the mean vector $\boldsymbol{\mu}$ and the scatter matrix $\boldsymbol{\Sigma}$ are complex quantities!
- The *Wirtinger* or $\mathbb{CR}$ -calculus has to be used to evaluate the derivatives [42,43,44,45,46,47,48,49].
- All the details can be found in [38].

Slepian-Bangs (SB) formula

- Introduced by Slepian and Bangs in [50] and [51], the SB formula has been extensively used for many years in array processing.

- The "classic" SB formula is a compact expression of the Fisher Information Matrix (FIM) for parameter estimation under a Gaussian data model [13, Appendix 3C].

- Specifically:
  - $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$: deterministic parameter vector,
  - $\mathbf{z} \sim CN(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$: complex Gaussian random vector.

- Then the SB formula provides us with a closed-form expression of the FIM for the estimation of $\boldsymbol{\theta} \in \Theta$.

Semiparametric Slepian-Bangs (SSB) formula

- Generalizations to:
  1. Non-circular complex Gaussian distributions [52],
  2. CES distributions [36],
  3. Non-circular CES distributions [53],
  4. Model misspecification under Gaussianity assumption [1],
  5. Model misspecification under CES assumption [54],
  6. Semiparametric model under CES assumption [38].

- Let $\mathbb{C}^N \ni \mathbf{z} \sim CES_N(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}), h)$ be a CES-distributed random vector parameterized by $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$.

- The semiparametric SB (SSB) formula in [38] provides the efficient FIM for the estimation of $\boldsymbol{\theta}$ in the presence of an *unknown*, nuisance density generator $h \in \mathcal{G}$.

Semiparametric Stochastic CRB (SSCRB)

- Assume to have an array of N sensors and K narrowband sources impinging on the array from $\{\nu_1, \dots, \nu_K\}$ directions.
- Data snapshots $\mathbf{z}_m \sim CES_N(\mathbf{z}; \mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\nu}, \boldsymbol{\Gamma}, \sigma^2), h_0)$, $\forall m$ whose density generator $h_0 \in \bar{\mathcal{G}}$ is unknown and [55]:

$$\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}(\boldsymbol{\nu}, \boldsymbol{\Gamma}, \sigma^2) = \mathbf{A}(\boldsymbol{\nu})\boldsymbol{\Gamma}\mathbf{A}(\boldsymbol{\nu})^H + \sigma^2 \mathbf{I}_N.$$

- The $\mathrm{SSCRB}(\boldsymbol{\nu}_0|\boldsymbol{\zeta}_0, \sigma_0^2, h_0)$ [38,39] generalizes the classical, Gaussian-based, SCRB [56,57] since:
    1. The Gaussianity assumption is replaced by the more general CES assumption,
    2. The additional infinite-dimensional *nuisace* parameter h_0 is taken into account.