



HAL
open science

Semiparametric estimation in elliptical distributions

Stefano Fortunati

► **To cite this version:**

| Stefano Fortunati. Semiparametric estimation in elliptical distributions. 2013. hal-04141679v2

HAL Id: hal-04141679

<https://hal.science/hal-04141679v2>

Preprint submitted on 1 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semiparametric estimation in elliptical distributions

Stefano Fortunati

Abstract This chapter has the twofold aim of introducing in an intuitive and accessible manner the general framework of semiparametric inference and then of showing how it can be fruitfully applied to the joint estimation of the *location vector* and the *covariance* (or *scatter*) matrix of a set of elliptically distributed observations in the presence of an unknown density generator. A semiparametric model is a set of probability density functions (pdfs) parameterized by a finite-dimensional parameter vector of interest and by an infinite-dimensional nuisance parameter, i.e. a function, whose estimation is not strictly required. The presence of this additional functional unknown will clearly lead to some performance losses in the estimation of the finite-dimensional parameter vector of interest. The first goal of this chapter is then to show how the classical estimation theory can be generalized in order to take into account an infinite-dimensional nuisance term. In particular, the three building blocks of the semiparametric theory, that are *the Hilbert space of score vectors*, *the nuisance tangent space* and *the related projection operator*, will be introduced. By means of these abstract concepts, we define the semiparametric counterpart of the Fisher Information Matrix (FIM) and the related Semiparametric Efficiency Bound. After having prepared the theoretical ground, the focus of the second part of the chapter is on the application of the general semiparametric inference framework to the joint estimation of the location vector and of the scatter matrix in Real Elliptically Symmetric (RES) distributed random vectors. A closed form expression for the semiparametric FIM and the related bound will be provided. We conclude the chapter by presenting the class of the *R*-estimators of the scatter matrix.

Stefano Fortunati

Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette & DR2I-IPSA, 94200, Ivry sur Seine e-mail: stefano.fortunati@centralesupelec.fr

1 Introduction

Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be n observation vectors collected from a random experiment and let $P_0(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be their (joint) “true” distribution admitting a (joint) “true” pdf $p_0(\mathbf{x}_1, \dots, \mathbf{x}_n)$. In this chapter we always assume that the observations are *independent* and *identically distributed* (i.i.d.) random vectors, consequently the joint pdf can be expressed as $p_0(\mathbf{x}_1, \dots, \mathbf{x}_n) = \prod_{i=1}^n p_0(\mathbf{x}_i)$. In the following, we indicate simply as $\mathbf{x} \sim p_0$ a generic random vector distributed according the (marginal) “true” pdf p_0 . Note that, in practice, the true pdf p_0 is not fully a-priori known, then the main task of any statistical procedure is to infer the missing knowledge on p_0 from the collected data. Inference methods rely then on the definition of a *model*, i.e. a set of pdfs that are able to statistically characterize the data-generating process. The nature of the model to be used depends on the available amount of a-priori information: the more precise our knowledge of the real-world phenomenon is, the narrower the model can be. The most commonly adopted models are the *parametric* ones.

As the name suggests, a parametric model \mathcal{P}_θ is a set of pdfs parameterized only by a finite-dimensional parameter vector $\theta \in \Theta \subseteq \mathbb{R}^q$:

$$\mathcal{P}_\theta \stackrel{\text{def}}{=} \{p_X(\mathbf{x}|\theta), \theta \in \Theta \subseteq \mathbb{R}^q\}.$$

If a parametric model is adopted, the information to be inferred from the observations is summarized in the (finite-dimensional) parameter vector θ to be estimated. More formally, if the true data pdf $p_0(\mathbf{x})$ belongs to \mathcal{P}_θ , this implies that there exists $\theta_0 \in \Theta$ such that $p_0(\mathbf{x}) = p_X(\mathbf{x}|\theta_0)$. However, in many practical applications, the available a-priori knowledge about the data-generating process may result to be insufficient to validate the chosen parametric model and we could run the risk that the true data pdf p_0 does not belong to the assumed model \mathcal{P}_θ . This *model misspecification* [6] will lead to some performance degradation in the estimation of θ that may result to be unacceptable. In this case, one may decide to rely on a more general *non-parametric model*. A non-parametric model is a collection of pdfs that possibly satisfy some functional constraints, i.e. the symmetry or the finiteness of their moments up to a given order:

$$\mathcal{P}_p \stackrel{\text{def}}{=} \{p_X(\mathbf{x}) \in \mathcal{K}\},$$

where \mathcal{K} is some constrained set of pdfs. If, on one hand, the generality of a non-parametric model minimizes the risk of model misspecification, non-parametric inference requires a large amount of homogeneous data that may be difficult (or even impossible) to collect in some applications.

The *semiparametric models* have been then proposed in order to keep the inherent finite-dimensionality of the parameters of interest without sacrificing the generality of the statistical description of a data-generating process [39]. Specifically, a semi-parametric model is a set of pdfs characterized by a finite-dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^q$ along with some infinite-dimensional parameter, i.e. a function, $g \in \mathcal{G}$:

$$\mathcal{P}_{\theta,g} \stackrel{\text{def}}{=} \{p_X(\mathbf{x}|\theta, g), \theta \in \Theta \subseteq \mathbb{R}^q, g \in \mathcal{G}\}.$$

A huge amount of practical inference problems can be described using this semi-parametric formalism (see [39], [2], [29] and the reference therein). Here some examples of remarkable importance for the signal processing community [5]:

1. Inference in Real Elliptically Symmetric (RES) Distributions

As discussed in the Chapter 1, a random vector $\mathbf{x} \in \mathbb{R}^m$ is said to be RES-distributed if it has a relevant pdf of the form:

$$p_X(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})), \quad (1)$$

where $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is called *density generator*. In almost all practical applications involving the elliptical model, we are generally interested in the estimation of the *location vector* $\boldsymbol{\mu} \in \mathbb{R}^m$ and/or of the *covariance/scatter matrix* $\boldsymbol{\Sigma} \in \mathcal{M}$ where \mathcal{M} indicates the set of all the positive definite, Hermitian $m \times m$ matrices. Therefore, the density generator g can be considered as a nuisance function whose estimation is not required. Note that, as discussed in Sec. 2.1 of the Chapter 1, $\boldsymbol{\Sigma} \in \mathcal{M}$ and $g \in \mathcal{G}$ are not jointly indefinable. To avoid the scale ambiguity, we introduce the symmetric and positive definite *shape matrix* as

$$\mathbf{V} = \boldsymbol{\Sigma}/s(\boldsymbol{\Sigma}) \in \mathcal{M}_s \stackrel{\text{def}}{=} \{\mathbf{V} \in \mathcal{M} | s(\mathbf{V}) = 1\} \subset \mathcal{M} \quad (2)$$

where $s : \mathcal{M} \rightarrow \mathbb{R}^+$ is a scalar homogeneous functional (see [15, 31], [10, Sec. II]) and \mathcal{M}_s is a constrained smooth sub-manifold of \mathcal{M} whose dimension is $\dim(\mathcal{M}_s) = \dim(\mathcal{M}) - 1$. It is immediate to verify that the RES family can be interpreted as a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\mu}, \mathbf{V}, g} \stackrel{\text{def}}{=} \left\{ p_X(\mathbf{x}) = |\mathbf{V}|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})), (\boldsymbol{\mu}, \mathbf{V}) \in \Theta, g \in \mathcal{G} \right\}, \quad (3)$$

where $\Theta \stackrel{\text{def}}{=} \mathbb{R}^m \times \mathcal{M}_s$. We will investigate the semiparametric elliptical model extensively in the second part of the chapter.

2. Estimation with 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.

Suppose now that we want to estimate the value of a parameter vector $\boldsymbol{\theta} \in \Theta$ from the observed dataset \mathbf{x} when the pdf p_Y of the missing data \mathbf{y} is *unknown*. Direct application of the basic rules of conditional distributions allows one to express the pdf p_X of the observed dataset as:

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

Consequently, the set of all the pdfs of the observed dataset \mathbf{x} can be framed in a *semiparametric mixture model* of the form [2, Sec. 4.5], [35]:

$$\mathcal{P}_{\boldsymbol{\theta}, p_Y} \stackrel{\text{def}}{=} \{p_X(\mathbf{x}|\boldsymbol{\theta}, p_Y), \boldsymbol{\theta} \in \Theta, p_Y \in \mathcal{K}\}.$$

3. **Non-linear regression** According to the amount of a-priori knowledge that we have on the observation model, the non-linear regression can be conveniently described using the semiparametric formalism:

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

- $\boldsymbol{\theta} \in \Theta$: 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 [2, Sec. 4.3]:

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

4. **Autoregressive processes (AR)** Let us consider the AR(p) process:

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

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

The aim is to estimate $\boldsymbol{\theta}$ from a vector \mathbf{x} of autoregressive observations. Usually, a Gaussianity assumption is adopted for the innovation process $\{w_n\}$. This assumption however is generally hardly justifiable. One again, the semiparametric formalism provide us with a suitable framework to statistically describe this problem. In fact, let $\mathbf{x} \in \mathbb{R}^m$ a vector of m observations from the above-mentioned AR(p). The set of all possible pdfs for $\mathbf{x} \in \mathbb{R}^m$ is a semiparametric model [18]:

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

Unlike the classical estimation theory, in semiparametric inference we need to handle the presence of a nuisance function. The bridge allowing us to go from the classical, “finite-dimensional”, estimation theory to the semiparametric one is the general theory of Hilbert spaces.

2 Estimation and Hilbert spaces

The aim of this section is then to introduce the mathematical framework that will allow us to handle the infinite-dimensional (functional) nuisance involved in a semiparametric model. We will proceed in two steps: firstly, we frame the classical finite-dimensional estimation theory in the geometrical context of Hilbert spaces. Secondly, we will extend these outcomes to the semiparametric framework. Our

discussion will be tutorial in nature, but the reader can find the whole formal mathematical derivation in [2, Sec. 2 and 3] and in [35].

2.1 Background on Hilbert spaces, linear spans and projections

Defining the geometry of a problem means to specify (at least) two elements: the linear vector space in which our problem is embedded and the related inner product. This two concept can be formalized by the definition of Hilbert space [34, Ch. 4]:

Definition 1 A Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a real (or complex) inner product space that is also a complete metric space with respect to the distance function $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ induced by the inner product $\langle \cdot, \cdot \rangle$.

The inner product allows us to define the concept of orthogonality. In particular, two elements $u, h \in \mathcal{H}$ are said to be *orthogonal*, denoted as $u \perp h$, iff $\langle u, h \rangle = 0$. They are called *orthonormal* if, in addition, $\|u\| = \|h\| = 1$. The next theorem is a key result for Hilbert spaces since it define the projection operator.

Theorem 1 (The Projection Theorem)

Let \mathcal{U} be a closed subspace of an Hilbert space \mathcal{H} . For any $h \in \mathcal{H}$, there exist a unique $u_0 \in \mathcal{U}$ such that (s.t.)

$$u_0 = \underset{u \in \mathcal{U}}{\operatorname{argmin}} \|h - u\|, \quad h \in \mathcal{H}. \quad (4)$$

Moreover, $u_0 \in \mathcal{U}$ is uniquely determined by the orthogonality constraint

$$\langle h - u_0, u \rangle = 0, \quad \forall u \in \mathcal{U}. \quad (5)$$

We indicate $u_0 = \Pi(h|\mathcal{U}) \in \mathcal{U}$ as the projection of $h \in \mathcal{H}$ onto the subspace $\mathcal{U} \subset \mathcal{H}$ and $\Pi(\cdot|\mathcal{U})$ the related projection operator.

Finding the projection of an element $h \in \mathcal{H}$ onto a generic subspace $\mathcal{U} \subset \mathcal{H}$ is, in general, a non trivial task. However, there is a case of fundamental importance for the subsequent development in which this projection can be explicitly derived. In particular, let us build a vector $\mathbf{v} = (v_1, \dots, v_r)$ of r arbitrary elements of \mathcal{H} . Let us define the *linear span* of \mathbf{v} in \mathcal{H} as:

$$\mathcal{U} \stackrel{\text{def}}{=} \{u \in \mathcal{H} | u = \mathbf{a}^T \mathbf{v} : \mathbf{a} \text{ is any vector in } \mathbb{R}^r\} \subset \mathcal{H}. \quad (6)$$

If $\{v_1, \dots, v_r\}$ are linearly independent, the linear span \mathcal{U} is a *finite-dimensional* linear subspace of \mathcal{H} of dimension r . Moreover, it can be shown that the unique orthogonal projection of an arbitrary element $h \in \mathcal{H}$ onto \mathcal{U} in (6) can be expressed as [35, Sec. 2]:

$$\Pi(h|\mathcal{U}) = \langle h, \mathbf{v}^T \rangle \langle \mathbf{v}, \mathbf{v}^T \rangle^{-1} \mathbf{v}, \quad (7)$$

where

$$\langle \mathbf{h}, \mathbf{v}^T \rangle \stackrel{\text{def}}{=} (\langle h, v_1 \rangle \cdots \langle h, v_r \rangle) \in \mathbb{R}^{1 \times r}, \quad (8)$$

$$\langle \mathbf{v}, \mathbf{v}^T \rangle \stackrel{\text{def}}{=} \begin{pmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_r \rangle \\ \vdots & \ddots & \vdots \\ \langle v_r, v_1 \rangle & \cdots & \langle v_r, v_r \rangle \end{pmatrix} \in \mathbb{R}^{r \times r}, \quad (9)$$

that is assumed to be invertible.

Let us now take it a step further and define the space $\mathcal{H}_q \stackrel{\text{def}}{=} \mathcal{H} \times \cdots \times \mathcal{H}$ as an Hilbert space obtained by the Cartesian product of q copies of \mathcal{H} whose inner product is induced by the one of \mathcal{H} . Specifically, let $\mathbf{h} = (h_1, \cdots, h_q)^T$ and $\mathbf{w} = (w_1, \cdots, w_q)^T$ be two vectors in \mathcal{H}_q . Then, the inner product of \mathcal{H}_q is given by $\langle \mathbf{h}, \mathbf{w} \rangle = \sum_{i=1}^q \langle h_i, w_i \rangle$. This type of space is generally called q -replicating Hilbert space. What about the linear span of \mathcal{H}_q ? Let $\mathbf{v} = (v_1, \cdots, v_r)^T$ be a column vector of r arbitrary elements of \mathcal{H} . The linear span of \mathbf{v} in \mathcal{H}_q is given by [35, Sec. 2]:

$$\mathcal{U}_q \stackrel{\text{def}}{=} \{ \mathbf{u} \in \mathcal{H}_q \mid \mathbf{u} = \mathbf{A}\mathbf{v} : \mathbf{A} \text{ is any matrix in } \mathbb{R}^{q \times r} \} \subset \mathcal{H}_q. \quad (10)$$

If the entries of \mathbf{v} are linearly independent, the dimension of \mathcal{U}_q is $q \cdot r$. Moreover, the projection of an arbitrary $\mathbf{h} \in \mathcal{H}_q$ onto the linear span \mathcal{U}_q in (10) is given by:

$$\Pi(\mathbf{h} | \mathcal{U}_q) = \langle \mathbf{h}, \mathbf{v}^T \rangle \langle \mathbf{v}, \mathbf{v}^T \rangle^{-1} \mathbf{v}, \quad (11)$$

$$\langle \mathbf{h}, \mathbf{v}^T \rangle \stackrel{\text{def}}{=} \begin{pmatrix} \langle h_1, v_1 \rangle & \cdots & \langle h_1, v_r \rangle \\ \vdots & \ddots & \vdots \\ \langle h_q, v_1 \rangle & \cdots & \langle h_q, v_r \rangle \end{pmatrix} \in \mathbb{R}^{q \times r}, \quad (12)$$

and the matrix $\langle \mathbf{v}, \mathbf{v}^T \rangle$ is defined as in (9) and it is assumed to be invertible.

2.2 The Hilbert space \mathcal{H}_q of the q -variate random functions

In this subsection we introduce a replicating Hilbert space that plays a fundamental role in the geometrical interpretation of the estimation theory. Let $(\mathcal{X}, \mathfrak{F}, P_X)$ be a probability space where the sample space \mathcal{X} is a subset of \mathbb{R}^m , \mathfrak{F} is the Borel σ -algebra on \mathcal{X} and P_X is a probability measure. We assume that the related cdf $P_X(\mathbf{x})$ admits a pdf $p_X(\mathbf{x})$ with respect to Lebesgue measure, s.t. $dP_X(\mathbf{x}) = p_X(\mathbf{x})d\mathbf{x}$.

Let \mathcal{H} be the set of scalar random functions on the sample space \mathcal{X} , i.e. $h : \mathcal{X} \rightarrow \mathbb{R}$ satisfying the following two properties:

1. *Zero-mean*: $\mathbb{E}(h) = \int_{\mathcal{X}} h(\mathbf{x})dP_X(\mathbf{x}) = 0$
2. *Finite variance*: $\mathbb{E}(h^2) = \int_{\mathcal{X}} h^2(\mathbf{x})dP_X(\mathbf{x}) = \sigma_h^2 < +\infty$

The (infinite-dimensional) linear space \mathcal{H} can be endowed with the inner product:

$$\langle h_1, h_2 \rangle \stackrel{\text{def}}{=} E(h_1 h_2) = \int_{\mathcal{X}} h_1(\mathbf{x}) h_2(\mathbf{x}) dP_X(\mathbf{x}), \quad \forall h_1, h_2 \in \mathcal{H}. \quad (13)$$

Then, $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is an Hilbert space and the squared norm associated to its inner product in (13) is $\|h\|^2 = E(h^2) = \sigma_h^2$ that is the variance of $h \in \mathcal{H}$.

We can now define the q -replicating Hilbert space $\mathcal{H}_q = \mathcal{H} \times \cdots \times \mathcal{H}$ as the linear space of the multivariate random functions $\mathbf{h} : \mathcal{X} \rightarrow \mathbb{R}^q$ endowed by the following inner product induced by the one of \mathcal{H} :

$$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle \stackrel{\text{def}}{=} E(\mathbf{h}_1^T \mathbf{h}_2) = \sum_{i=1}^q E(h_{1,i} h_{2,i}). \quad (14)$$

From the replicating nature of \mathcal{H}_q , it is immediate to verify that every $\mathbf{h} \in \mathcal{H}_q$ satisfies the two properties:

1. *Zero-mean*: $E(\mathbf{h}) = \int_{\mathcal{X}} \mathbf{h}(\mathbf{x}) dP_X(\mathbf{x}) = \mathbf{0}$
2. *Finite variance*: $E(\mathbf{h}^T \mathbf{h}) = \langle \mathbf{h}, \mathbf{h} \rangle = \|\mathbf{h}\|^2 < +\infty$.

Furthermore, since \mathbf{h} is a q -variate random function, we can define its $q \times q$ covariance matrix $\mathbf{C}(\mathbf{h})$ in the usual way as:

$$\mathbf{C}(\mathbf{h}) \stackrel{\text{def}}{=} E(\mathbf{h} \mathbf{h}^T) = \begin{pmatrix} E(h_1 h_1) & \cdots & E(h_1 h_q) \\ \vdots & \ddots & \vdots \\ E(h_q h_1) & \cdots & E(h_q h_q) \end{pmatrix} \in \mathbb{R}^{q \times q}. \quad (15)$$

Let us investigate the geometrical structure of \mathcal{H}_q . In particular, we want to derive an explicit expression of the orthogonal projection of a generic element \mathbf{h} into a linear span of \mathcal{H}_q . Following the discussion given in subsec. 2.1, let us define $\mathbf{v} = (v_1, \cdots, v_r)^T$ as a column vector of r arbitrary element of \mathcal{H} . Then, its linear span \mathcal{U}_q is defined as in (10). Let us consider now the problem of finding the orthogonal projection of an arbitrary element $\mathbf{h} \in \mathcal{H}_q$ onto \mathcal{U}_q , i.e. $\Pi(\mathbf{h}|\mathcal{U}_q)$. From the Projection Theorem (Theo. 1), we know that this projection is unique and it can be explicitly written using (11) as:

$$\Pi(\mathbf{h}|\mathcal{U}_q) = E(\mathbf{h} \mathbf{v}^T) E(\mathbf{v} \mathbf{v}^T)^{-1} \mathbf{v} = E(\mathbf{h} \mathbf{v}^T) [\mathbf{C}(\mathbf{v})]^{-1} \mathbf{v}. \quad (16)$$

Let us now see how this general concepts on Hilbert space may provide an original look on the classical (finite-dimensional) estimation theory.

2.3 Estimation in parametric models: a geometric interpretation

Let $(\mathcal{X}, \mathfrak{F}, P_X)$ be the same probability space introduced in subsec. 2.2, and let P_X and p_X the related cdf and pdf, s.t. $dP_X(\mathbf{x}) = p_X(\mathbf{x}) d\mathbf{x}$. Let us consider the following *parametric model*:

$$\mathcal{P}_{\theta, \eta} \stackrel{\text{def}}{=} \{p_X(\mathbf{x}|\theta, \eta), \theta \in \Theta \subseteq \mathbb{R}^q, \eta \in \Gamma \subseteq \mathbb{R}^r\}, \quad (17)$$

where the q -dimensional vector $\boldsymbol{\theta}$ is the vector of the *parameters of interest* while the r -dimensional vector $\boldsymbol{\eta}$ is the vector of the *nuisance parameters*. In the rest of this section, we always indicate the *true* parameter vectors as $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$, and the related *true* pdf and distribution as $p_0(\mathbf{x}) = p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ and $P_0(\mathbf{x})$, respectively. Moreover, the expectation operator and the covariance matrix evaluated w.r.t. the true pdf will be indicated as $E_0(\cdot)$ and $\mathbf{C}_0(\cdot)$, respectively.

Let us start by defining a key ingredient in estimation theory: the *score vector*.

Definition 2 Let $\mathcal{P}_{\boldsymbol{\theta}, \boldsymbol{\eta}}$ be a parametric model as in (17). The *score vector of the parameters of interest* $\mathbf{s}_{\boldsymbol{\theta}_0}$ and the *nuisance score vector* $\mathbf{s}_{\boldsymbol{\eta}_0}$ are defined, respectively, as the q - and r -variate random functions s.t.:

$$[\mathbf{s}_{\boldsymbol{\theta}_0}]_i \stackrel{\text{def}}{=} [\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0)]_i = [\nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)]_i = \left. \frac{\partial \ln p_X(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\eta}_0)}{\partial \theta_i} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad (18)$$

$$[\mathbf{s}_{\boldsymbol{\eta}_0}]_j \stackrel{\text{def}}{=} [\mathbf{s}_{\boldsymbol{\eta}}(\mathbf{x}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0)]_j = [\nabla_{\boldsymbol{\eta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)]_j = \left. \frac{\partial \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, \boldsymbol{\eta})}{\partial \eta_j} \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0}, \quad (19)$$

where $i = 1, \dots, q$ and $j = 1, \dots, r$. Under the usual regularity conditions (collected in e.g. [23, sec. 6.2 and 6.3]) that allow for the order inversion between integral and derivative operators, the score vectors are zero-mean random vectors:

$$E_0(\mathbf{s}_{\boldsymbol{\theta}_0}) = \mathbf{0}, \quad E_0(\mathbf{s}_{\boldsymbol{\eta}_0}) = \mathbf{0}. \quad (20)$$

Moreover, if $\|\mathbf{s}_{\boldsymbol{\theta}_0}\| < \infty$ and $\|\mathbf{s}_{\boldsymbol{\eta}_0}\| < \infty$, we can immediately deduce that the score vectors belongs to two replicating Hilbert spaces of the type defined in subsec. 2.2:

$$\mathbf{s}_{\boldsymbol{\theta}_0} \in \mathcal{H}_q, \quad \mathbf{s}_{\boldsymbol{\eta}_0} \in \mathcal{H}_r. \quad (21)$$

This result allow us to use the geometric structure of \mathcal{H}_q and \mathcal{H}_r to our estimation-oriented goal. First of all, we can note that the covariance matrix in (15) of $\mathbf{s}_{\boldsymbol{\theta}_0}$ and $\mathbf{s}_{\boldsymbol{\eta}_0}$ is nothing but the Fisher Information Matrices (FIMs) of the parameters of interest and of the nuisance ones:

$$\mathbf{C}_0(\mathbf{s}_{\boldsymbol{\theta}_0}) = E_0(\mathbf{s}_{\boldsymbol{\theta}_0} \mathbf{s}_{\boldsymbol{\theta}_0}^T) \stackrel{\text{def}}{=} \mathbf{I}_{\boldsymbol{\theta}_0} \quad \mathbf{C}_0(\mathbf{s}_{\boldsymbol{\eta}_0}) = E_0(\mathbf{s}_{\boldsymbol{\eta}_0} \mathbf{s}_{\boldsymbol{\eta}_0}^T) \stackrel{\text{def}}{=} \mathbf{I}_{\boldsymbol{\eta}_0}. \quad (22)$$

For further reference, let us also introduce the cross-information matrices as:

$$\mathbf{I}_{\boldsymbol{\theta}_0 \boldsymbol{\eta}_0} \stackrel{\text{def}}{=} E_0(\mathbf{s}_{\boldsymbol{\theta}_0} \mathbf{s}_{\boldsymbol{\eta}_0}^T) = \left[E_0(\mathbf{s}_{\boldsymbol{\eta}_0} \mathbf{s}_{\boldsymbol{\theta}_0}^T) \right]^T \stackrel{\text{def}}{=} \mathbf{I}_{\boldsymbol{\eta}_0 \boldsymbol{\theta}_0}^T. \quad (23)$$

Specifically, using the findings related to the projection operator onto the finite-dimensional linear span of \mathcal{H}_q presented in subsec. (2.2), we can introduce two fundamental concepts that will be extensively used in the rest of the chapter:

Definition 3 The nuisance tangent space. The linear span of \mathcal{H}_q spanned by the entries of the nuisance score vector $\mathbf{s}_{\boldsymbol{\eta}_0}$, i.e.

$$\mathcal{T}_{\boldsymbol{\eta}_0} \stackrel{\text{def}}{=} \{ \mathbf{u} \in \mathcal{H}_q | \mathbf{u} = \mathbf{A} \mathbf{s}_{\boldsymbol{\eta}_0} : \mathbf{A} \text{ is any matrix in } \mathbb{R}^{q \times r} \} \subset \mathcal{H}_q, \quad (24)$$

is called *nuisance tangent space of the parametric model* $\mathcal{P}_{\theta,\eta}$ at the true nuisance parameter vector $\boldsymbol{\eta}_0$.

Definition 4 The **efficient score vector** $\bar{\mathbf{s}}_0$ is defined as the residual of the score vector of the parameters of interest \mathbf{s}_{θ_0} after projecting it onto the nuisance tangent space \mathcal{T}_{η_0} defined in (24) (see e.g. [2, Ch. 2]):

$$\bar{\mathbf{s}}(\mathbf{x}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = \bar{\mathbf{s}}_0 \stackrel{\text{def}}{=} \mathbf{s}_{\theta_0} - \Pi(\mathbf{s}_{\theta_0} | \mathcal{T}_{\eta_0}) = \mathbf{s}_{\theta_0} - \mathbf{I}_{\theta_0 \eta_0} \mathbf{I}_{\eta_0}^{-1} \mathbf{s}_{\eta_0}. \quad (25)$$

Note that, in the last equality, we used the explicit expression for the projection operator given in (16) and the definition of the FIM and of the cross-information matrix in (23). Consequently, the **efficient FIM** can be defined as:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0 | \boldsymbol{\eta}_0) \stackrel{\text{def}}{=} \text{E}_0(\bar{\mathbf{s}}_0 \bar{\mathbf{s}}_0^T) = \mathbf{I}_{\theta_0} - \mathbf{I}_{\theta_0 \eta_0} \mathbf{I}_{\eta_0}^{-1} \mathbf{I}_{\theta_0 \eta_0}^T. \quad (26)$$

2.3.1 Efficiency and Cramér-Rao inequality in parametric models

Score vectors and FIMs are the two building blocs of a well-known result in parametric models: the *Cramér-Rao inequality*. We show that this classical result can be obtained in a different, yet equivalent way, by using the geometrical approach previously introduced. Let $\mathcal{P}_{\theta,\eta}$ be a parametric model as in (17) and let $\mathbf{I}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ be the joint FIM for the parameter vector of interest and for the nuisance one:

$$\mathbf{I}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = \begin{pmatrix} \mathbf{I}_{\theta_0} & \mathbf{I}_{\theta_0 \eta_0} \\ \mathbf{I}_{\theta_0 \eta_0}^T & \mathbf{I}_{\eta_0} \end{pmatrix}. \quad (27)$$

Suppose now to have a set of n i.i.d. observations s.t. $\mathbf{x}_i \sim p_0, \forall i = 1, \dots, n$. We define as $\hat{\boldsymbol{\theta}}_n \stackrel{\text{def}}{=} \hat{\boldsymbol{\theta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ an estimator of $\boldsymbol{\theta}_0 \in \Theta$ in the presence of the nuisance vector $\boldsymbol{\eta}_0 \in \Gamma$. Then, if $\hat{\boldsymbol{\theta}}_n$ is unbiased, i.e. $\text{E}_0(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\theta}_0, \forall n$, the following inequality holds (see e.g. [23, Secs. 2.7 and 6.4]):

$$n \text{E}_0 \left((\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T \right) \geq \text{CRB}(\boldsymbol{\theta}_0 | \boldsymbol{\eta}_0), \quad (28)$$

where $\text{CRB}(\boldsymbol{\theta}_0 | \boldsymbol{\eta}_0)$ can be obtained from the joint FIM using the Matrix Inversion Lemma [32]:

$$\text{CRB}(\boldsymbol{\theta}_0 | \boldsymbol{\eta}_0) \stackrel{\text{def}}{=} \left[\mathbf{I}_{\theta_0} - \mathbf{I}_{\theta_0 \eta_0} \mathbf{I}_{\eta_0}^{-1} \mathbf{I}_{\theta_0 \eta_0}^T \right]^{-1}. \quad (29)$$

It is immediate to verify that this classical result can be obtained by using the efficient FIM defined in (26), i.e.

$$\text{CRB}(\boldsymbol{\theta}_0 | \boldsymbol{\eta}_0) = \bar{\mathbf{I}}(\boldsymbol{\theta}_0 | \boldsymbol{\eta}_0)^{-1}. \quad (30)$$

This result allow us to provide an intuitive explanation of the efficient score vector in (25): $\bar{\mathbf{s}}_0$ is able to characterize the information that the parametric model $\mathcal{P}_{\theta,\eta}$ has on the parameter vector of interest $\boldsymbol{\theta}_0 \in \Theta$ when a additional nuisance term

is present. This fact is of crucial importance since, unlike the classical, “matrix-oriented”, derivation of the Cramér-Rao inequality, the one based on the efficient score vector and the related efficient FIM does not require the nuisance term to be finite-dimensional. Then, it may be extended to the semiparametric context, as we will see in the next section. Before moving on, let us answer to the following question: is there any estimator of $\theta_0 \in \Theta$ able to achieve the $\text{CRB}(\theta_0|\eta_0)$, at least asymptotically in n ? Under some regularity condition on the parametric model (see e.g. [23, sec. 6.1 and 6.2]), the *Maximum Likelihood* (ML) estimator is a good candidate. By indicating with

$$L_n(\theta, \eta) \stackrel{\text{def}}{=} \sum_{i=1}^n \ln p_X(\mathbf{x}_i|\theta, \eta) \quad (31)$$

the *joint likelihood function* of a set of n i.i.d. observations, the joint ML estimator $(\hat{\theta}_{ML,n}, \hat{\eta}_{ML,n})$ can be expressed as [23, sec. 6.2]:

$$L_n(\hat{\theta}_{ML,n}, \hat{\eta}_{ML,n}) = \max \{L_n(\theta, \eta); \theta \in \Theta, \eta \in \Gamma\} \quad (32)$$

Moreover, under the regularity conditions listed in e.g. [23, Secs. 6.2 and 6.3], the ML estimator $\hat{\theta}_{ML,n}$ of the parameter vector of interest θ_0 satisfies the two following, well-known, properties:

P1 \sqrt{n} -consistency: $\sqrt{n}(\hat{\theta}_{ML,n} - \theta_0) = O_{P_X}(1)$.¹

P2 Asymptotic normality and efficiency: $\sqrt{n}(\hat{\theta}_{ML,n} - \theta_0) \underset{n \rightarrow \infty}{\sim} \mathcal{N}(\mathbf{0}, \text{CRB}(\theta_0|\eta_0))$.

We note in passing that, when the ML estimator in (32) exists and it is unique, it can be expressed as the solution of the following non-linear system of *score-based estimating equations*:

$$\begin{cases} n^{-1/2} \sum_{i=1}^n \mathbf{s}_\theta(\mathbf{x}_i; \theta, \eta) \Big|_{\theta=\hat{\theta}_{S,n}, \eta=\hat{\eta}_{S,n}} = \mathbf{0}, \\ n^{-1/2} \sum_{i=1}^n \mathbf{s}_\eta(\mathbf{x}_i; \theta, \eta) \Big|_{\theta=\hat{\theta}_{S,n}, \eta=\hat{\eta}_{S,n}} = \mathbf{0}, \end{cases} \quad (33)$$

where \mathbf{s}_θ and \mathbf{s}_η are the score vectors defined in (18) and (19), respectively

Even if it is the most popular, the ML estimator in (32) is not the only one having the desired properties P1 and P2. Without rigorously listing the required regularity conditions, in the following provide two other \sqrt{n} -consistent (P1) and efficient (P2) estimators:

- *Efficient score-based estimating equations*: Let $\eta_n^\star = \eta_n^\star(\theta_0)$ be a sub-optimal, \sqrt{n} -consistent but not necessarily efficient, estimator of the nuisance parameter vector η_0 . Note that, η_n^\star may depend on the true parameter vector θ_0 . For further reference, we indicate η_n^\star as a *preliminary estimator*. Then, we can define $\hat{\theta}_{ES,n}$ as the estimator of the parameter vector of interest θ_0 obtained as the solution of the following estimating equations:

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

$$n^{-1/2} \sum_{i=1}^n \bar{\mathbf{s}}(\mathbf{x}_i; \boldsymbol{\theta}, \boldsymbol{\eta}_n^*(\boldsymbol{\theta})) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ES,n}} = \mathbf{0}, \quad (34)$$

where $\bar{\mathbf{s}}(\mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\eta})$ is the efficient score vector given in (25). Under the ‘‘necessary’’ regularity conditions [35, Ex. 3.2], it can be proved that $\hat{\boldsymbol{\theta}}_{ES,n}$ satisfies the properties P1 and P2.

- *One-step estimators*: Suppose that, to find an approximate solution of the estimating equations in (34), we use the first (one-step) iteration of the Newton-Raphson method as:

$$\tilde{\boldsymbol{\theta}}_{ES,n} = \tilde{\boldsymbol{\theta}} - [\mathbf{J}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\eta}_n^*(\tilde{\boldsymbol{\theta}}))]^{-1} \bar{\Delta}_n(\tilde{\boldsymbol{\theta}}, \boldsymbol{\eta}_n^*(\tilde{\boldsymbol{\theta}})), \quad (35)$$

where $\tilde{\boldsymbol{\theta}}$ is the starting point of the Newton-Raphson iteration applied to the function given in (34) and called *efficient central sequence*:

$$\bar{\Delta}(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\theta}, \boldsymbol{\eta}) = \bar{\Delta}_n(\boldsymbol{\theta}, \boldsymbol{\eta}) \stackrel{\text{def}}{=} n^{-1/2} \sum_{i=1}^n \bar{\mathbf{s}}(\mathbf{x}_i; \boldsymbol{\theta}, \boldsymbol{\eta}), \quad (36)$$

whose Jacobian matrix is given by:

$$\mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\eta}_n^*(\boldsymbol{\theta})) \stackrel{\text{def}}{=} \nabla_{\boldsymbol{\theta}}^T [\bar{\Delta}_n(\boldsymbol{\theta}, \boldsymbol{\eta}_n^*(\boldsymbol{\theta}))] = n^{-1/2} \sum_{i=1}^n \nabla_{\boldsymbol{\theta}}^T \bar{\mathbf{s}}(\mathbf{x}_i; \boldsymbol{\theta}, \boldsymbol{\eta}_n^*(\boldsymbol{\theta})). \quad (37)$$

In his seminal and fundamental works [21],[22, Ch. 6] Le Cam showed that it is possible to construct an efficient one-step estimator $\hat{\boldsymbol{\theta}}_{OP,n}$, inspired by the Newton-Raphson iteration in (35), as:

$$\hat{\boldsymbol{\theta}}_{OP,n} = \boldsymbol{\theta}_n^* + n^{-1/2} [\mathbf{C}(\boldsymbol{\theta}_n^*, \boldsymbol{\eta}_n^*)]^{-1} \bar{\Delta}_n(\boldsymbol{\theta}_n^*, \boldsymbol{\eta}_n^*), \quad (38)$$

where $\boldsymbol{\theta}_n^*$ is a preliminary, \sqrt{n} -consistent but possibly not efficient, estimator of the parameter vector of interest $\boldsymbol{\theta}_0$ and:

$$\mathbf{C}(\boldsymbol{\theta}_n^*, \boldsymbol{\eta}_n^*) = \frac{1}{n} \left[\sum_{i=1}^n \bar{\mathbf{s}}(\mathbf{x}_i, \boldsymbol{\theta}_n^*, \boldsymbol{\eta}_n^*) \bar{\mathbf{s}}(\mathbf{x}_i, \boldsymbol{\theta}_n^*, \boldsymbol{\eta}_n^*)^T \right], \quad (39)$$

is the sample estimate of the efficient FIM $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0)$ in (26).² Remarkably, Le Cam showed that the estimator $\hat{\boldsymbol{\theta}}_{OP,n}$ in (38) satisfies the properties P1 and P2 (see e.g. [2, Theo. 2 in Sec. 2.5 and Theo. 1 in Sec. 7.8]).

We are now ready to generalize these results to the semiparametric framework.

2.4 Extension to Semiparametric Models

Let $(\mathcal{X}, \mathfrak{F}, P_X)$ be the same probability space defined in sec. 2.3. We have already defined in sec. 1 a *semiparametric model* as:

² To link this result with the Jacobian matrix of the efficient central sequence in (37), we can note that, under the necessary regularity conditions, $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_0) = \text{E}_0(\bar{\mathbf{s}}_0 \bar{\mathbf{s}}_0^T) = -\text{E}_0(\nabla_{\boldsymbol{\theta}_0}^T \bar{\mathbf{s}}_0)$.

$$\mathcal{P}_{\theta, g} \stackrel{\text{def}}{=} \{p_X(\mathbf{x}|\theta, g), \theta \in \Theta \subseteq \mathbb{R}^q, g \in \mathcal{G}\}. \quad (40)$$

where $\theta \in \Theta$ is q -dimensional vector of the parameters of interest, while $g \in \mathcal{G}$ is a nuisance function. Following the notation introduced before, we denote the true “semiparametric vector” as $(\theta_0, g_0) \in \Theta \times \mathcal{G}$, and consequently the true pdf is denoted as $p_0(\mathbf{x}) = p_X(\mathbf{x}|\theta_0, g_0)$. From an abstract, “Hilbert space-based” standpoint, there are no conceptual differences between the definition of a parametric model in (17) and the one of a semiparametric model in (40). In particular, both the parametric and the semiparametric models can be considered as a family of pdfs indexed by a vector belonging to some parameter space. In the case of parametric models, such a parameter space is a finite-dimensional space obtained as the Cartesian product of two finite-dimensional spaces, $(\theta^T, \eta^T)^T \in \Theta \times \Gamma \subseteq \mathbb{R}^q \times \mathbb{R}^r$, i.e. the space of the parameters of interest Θ and the nuisance parameter space Γ . In the semiparametric case, the parameter space is the Cartesian product of the finite-dimensional space of the parameters of interest Θ and an infinite-dimensional set \mathcal{G} of nuisance functions, i.e. $(\theta, g) \in \Theta \times \mathcal{G}$. Even if, from a conceptual standpoint, there is no difference between parametric and semiparametric models, we cannot directly apply the tools developed in the parametric framework to the semiparametric one because of the infinite-dimensional nature of the nuisance parameter $g \in \mathcal{G}$. To develop an inference theory for semiparametric models, we need to generalize the concepts of efficient score vector in the presence of infinite dimensional nuisance parameters.

The link between parametric and semiparametric inference is the concept of *parametric sub-model* of a semiparametric model. Formally, the i -th parametric sub-model of $\mathcal{P}_{\theta, g}$ is defined as [35, Sec. 4.2], [2, Sec. 3.1],[18, Sec. 2.2]:³

$$\mathcal{P}_{\theta, v_i} \stackrel{\text{def}}{=} \{p_X(\mathbf{x}|\theta, v_i(\cdot, \cdot, \eta)), \theta \in \Theta \subseteq \mathbb{R}^q, \eta \in \Gamma_i \subseteq \mathbb{R}^{r_i}\}, \text{ where:} \quad (41)$$

$$\begin{aligned} v_i : \mathcal{X} \times \Theta \times \Gamma_i &\rightarrow \mathcal{G} \\ \eta &\mapsto v_i(\mathbf{x}, \theta, \eta), \end{aligned} \quad (42)$$

is a *known* function parametrized by a vector of *unknown* parameters. In particular, for every $i \in \mathcal{I}$, $\mathcal{P}_{\theta, v_i}$ in (41) is a parametric model of the form already defined in (17) satisfying the following three conditions [35, Sec. 4.2]:

- C0) $v_i : \mathcal{X} \times \Theta \times \Gamma_i \rightarrow \mathcal{G}$ is a smooth parametric map $\forall i \in \mathcal{I}$,
- C1) $\mathcal{P}_{\theta, v_i} \subseteq \mathcal{P}_{\theta, g}$, $\forall i \in \mathcal{I}$,
- C2) $p_0(\mathbf{x}) \in \mathcal{P}_{\theta, v_i}$, i.e. $\forall i \in \mathcal{I}$ there exists a vector $(\theta_0^T, \eta_0^T)^T \in \Theta \times \Gamma_i \subseteq \mathbb{R}^q \times \mathbb{R}^{r_i}$ such that $p_X(\mathbf{x}|\theta_0, v_i(\cdot, \cdot, \eta_0)) = p_X(\mathbf{x}|\theta_0, g_0)$.

Condition C1 tells us that all the pdfs that compose each possible parametric submodel $\mathcal{P}_{\theta, v_i}$ must belong to the semiparametric model $\mathcal{P}_{\theta, g}$ as well. Moreover, Condition C2 highlights the fact that each parametric submodel $\mathcal{P}_{\theta, v_i}$ must contain the “true” data pdf, i.e. there must exist a parameter vector $\eta_0 \in \Gamma_i$ such that $p_X(\mathbf{x}|\theta_0, v_i(\cdot, \cdot, \eta_0)) = p_0(\mathbf{x})$ for every $i \in \mathcal{I}$. It must be stressed that, unlike a

³ Note that the notation $p_X(\mathbf{x}|\theta, v_i(\cdot, \cdot, \eta))$ indicates that the pdf p_X is a composition of two functions, i.e. w and v_i , such that $p_X(\mathbf{x}|\theta, v_i(\cdot, \cdot, \eta)) = (w \circ v_i)(\mathbf{x}, \theta, \eta) = w(\mathbf{x}, \theta, v_i(\mathbf{x}, \theta, \eta))$.

classical parametric model, a parametric sub-model is only an artificial tool that is exploited to develop the semiparametric theory but that cannot be directly used for data analysis. In fact, since we do not know the true data pdf p_0 , we cannot explicitly build a set of parametric sub-models satisfying the condition C2.

Roughly speaking, using a parametric sub-model $\mathcal{P}_{\theta, \nu_i}$ in place of $\mathcal{P}_{\theta, g}$ we are actually approximating the infinite-dimensional parameter $g \in \mathcal{G}$ with the function $\nu_i(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\eta})$, parameterized by the finite-dimensional nuisance parameter vector $\boldsymbol{\eta} \in \Gamma_i \subseteq \mathbb{R}^{r_i}$ whose dimension r_i depends on the particular choice of ν_i . This approximation may be made finer and finer by choosing a sequence of function $\nu_i(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\eta}) \in \mathcal{G}$ that converges to $g \in \mathcal{G}$. The way to generalize the classical theory developed for parametric models to semiparametric framework should now be more clear: we could exploit the outcomes obtained in a sequence of (artificial) parametric sub-models $\{\mathcal{P}_{\theta, \nu_i}\}_{i \in I}$ to obtain a generalization of the concept of nuisance tangent space and efficient score function.

Let us start by providing a semiparametric generalization of nuisance tangent space. In this chapter, we define the *semiparametric nuisance tangent space* according to the definition given in [29] and [35, Sec. 4.4], while a more general (but more abstract) definition is given in [2, Sec. 3.2], [1] and [37, Ch. 25]. At first, let us recall that the Hilbert space \mathcal{H}_q of the q -variate, zero-mean, random functions is a metric space with (squared) distance given by $\|\mathbf{h}_1 - \mathbf{h}_2\|^2 = \mathbb{E}_0((\mathbf{h}_1 - \mathbf{h}_2)^T(\mathbf{h}_1 - \mathbf{h}_2))$. The notion of convergence adopted in the following is based on this definition of distance.

Definition 5 The **semiparametric nuisance tangent space** \mathcal{T}_{g_0} of $\mathcal{P}_{\theta, g}$ at g_0 is defined as the closure ⁴ of the union of all the nuisance tangent spaces

$$\mathcal{T}_{\eta_{0,i}} \stackrel{\text{def}}{=} \{\mathbf{u} \in \mathcal{H}_q \mid \mathbf{u} = \mathbf{C}_i \mathbf{s}_{\eta_{0,i}} : \mathbf{C}_i \text{ is any matrix in } \mathbb{R}^{q \times r_i}\} \subset \mathcal{H}_q \quad (43)$$

of the parametric submodels $\{\mathcal{P}_{\theta, \nu_i}\}_{i \in I} \subseteq \mathcal{P}_{\theta, g}$. Specifically, $\mathcal{T}_{g_0} \subseteq \mathcal{H}_q$ is the space of all q -variate, zero-mean, random functions $\mathbf{u} \in \mathcal{H}_q$ for which there exists a sequence $\{\mathbf{C}_i \mathbf{s}_{\eta_{0,i}}\}_{i \in I}$ such that $\|\mathbf{u} - \mathbf{C}_i \mathbf{s}_{\eta_{0,i}}\| \rightarrow 0$, where, according to (19), $\mathbf{s}_{\eta_{0,i}} = \nabla_{\boldsymbol{\eta}} \ln p_X(\mathbf{x} \mid \boldsymbol{\theta}_0, \nu_i(\cdot, \cdot, \boldsymbol{\eta}_0)) \in \mathcal{H}_{r_i}$ is the nuisance score vector of the parametric sub-model $\mathcal{P}_{\theta, \nu_i}$ [35, Sec. 4.4],[1],[29]. Using this definition, the semiparametric nuisance tangent space can be simply indicated as:

$$\mathcal{T}_{g_0} = \overline{\bigcup_{\{\mathcal{P}_{\theta, \nu_i}\}_{i \in I}} \mathcal{T}_{\eta_{0,i}}} \subseteq \mathcal{H}_q. \quad (44)$$

Note that, the closure of a union of linear spaces doesn't need to be linear, in general. However, as discussed in [1, Assumption S] and [35, Sec. 4.4, Remark 5]), \mathcal{T}_{g_0} is a q -replicating Hilbert space in the vast majority of the non-pathological semiparametric models. The closure and the linearity of \mathcal{T}_{g_0} guarantee the existence and the uniqueness of the projection operator $\Pi(\cdot \mid \mathcal{T}_{g_0})$ through Theorem 1. However, unlike its parametric counterpart in (24), the semiparametric nuisance tangent space

⁴ 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} .

is generally an *infinite-dimensional* linear space. Consequently, a general, closed form expression for $\Pi(\cdot|\mathcal{T}_{g_0})$ is no longer available.

In the wake of what we have done in sec. 2.3 for a parametric model involving a finite-dimensional nuisance vector, let us try to define the semiparametric counterpart of the efficient score vector in (25). To this end, we need three ingredients: the score vector of the parameters of interest \mathbf{s}_{θ_0} , the semiparametric nuisance tangent space \mathcal{T}_{g_0} and the related projection operator $\Pi(\cdot|\mathcal{T}_{g_0})$. We have already defined \mathcal{T}_{g_0} and $\Pi(\cdot|\mathcal{T}_{g_0})$ in Definition 5, while \mathbf{s}_{θ_0} is defined exactly as in (18) for the parametric case where we need to substitute the finite-dimensional nuisance vector $\boldsymbol{\eta}_0$ with the nuisance function g_0 , i.e. $\mathbf{s}_{\theta_0} = \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, g_0)$. Consequently, **the semiparametric efficient score vector** $\bar{\mathbf{s}}_0$ is defined as the residual of the score vector of the parameters of interest \mathbf{s}_{θ_0} after projecting it onto the semiparametric nuisance tangent space \mathcal{T}_{g_0} :

$$\bar{\mathbf{s}}(\mathbf{x}; \boldsymbol{\theta}_0, g_0) = \bar{\mathbf{s}}_0 \stackrel{\text{def}}{=} \mathbf{s}_{\theta_0} - \Pi(\mathbf{s}_{\theta_0}|\mathcal{T}_{g_0}). \quad (45)$$

Moreover, the related **efficient semiparametric FIM** (SFIM) can be defined as:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0) \stackrel{\text{def}}{=} E_0(\bar{\mathbf{s}}_0 \bar{\mathbf{s}}_0^T). \quad (46)$$

It is immediate to verify that the previous two definitions are perfectly in line with the one given in Definition 4 for the parametric case. However, it is important to stress again here that, unlike the parametric case, a general explicit expression for $\Pi(\cdot|\mathcal{T}_{g_0})$ is not available when \mathcal{T}_{g_0} is infinite-dimensional and it has to be calculated on a case-by-case basis. We will come back on this central point ahead in the section.

2.4.1 Semiparametric efficiency and related bound

The definition of the semiparametric efficient score vector and the related efficient SFIM provided in (45) and (46) lead us directly to the generalization of the Cramér-Rao inequality to the semiparametric framework. The following theorem stating it can be found in [35, Theo. 4.1] and [38, Theo. 4.2]. A more abstract and general formulation can be found in [1], [2, Sec. 3.4] and [37, Ch. 25].

Theorem 2 *The Semiparametric Cramér-Rao Bound (SCRB) for the estimation of the finite-dimensional vector $\boldsymbol{\theta}_0 \in \Theta$ in the semiparametric model $\mathcal{P}_{\boldsymbol{\theta},g}$ is given by:*

$$\text{SCRB}(\boldsymbol{\theta}_0|g_0) \stackrel{\text{def}}{=} \sup_{\{\mathcal{P}_{\boldsymbol{\theta},v_i}\}} [E_0(\bar{\mathbf{s}}_{0,i} \bar{\mathbf{s}}_{0,i}^T)]^{-1} = [E_0(\bar{\mathbf{s}}_0 \bar{\mathbf{s}}_0^T)]^{-1} = [\bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0)]^{-1}, \quad (47)$$

where $\{\mathcal{P}_{\boldsymbol{\theta},v_i}\}_{i \in I}$ is the set of all possible parametric sub-models of $\mathcal{P}_{\boldsymbol{\theta},g}$, as defined in (41). Moreover, $\bar{\mathbf{s}}_{0,i}$ indicates the efficient score vector of the i -th parametric sub-model that, according to (25), is given by $\bar{\mathbf{s}}_0 = \mathbf{s}_{\theta_0} - \Pi(\mathbf{s}_{\theta_0}|\mathcal{T}_{\eta_{0,i}})$ where $\mathcal{T}_{\eta_{0,i}}$ is defined in (43). Finally, $\bar{\mathbf{s}}_0$ and $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|g_0)$ are the semiparametric efficient score vector and the efficient SFIM defined in (45) and (46), respectively.

The proof is long and non-trivial, and we left the interesting reader to check it in e.g. [35, Theo. 4.1] and [38, Theo. 4.2]. Here, we limit ourselves to two observations:

- The definition of the semiparametric efficient score vector, of the efficient SFIM and of the related SCRB provided in (45), (46) and (47) are formally equivalent to their parametric counterparts given in (25), (26) and (30). Only the definition of the nuisance tangent space and of the related projection operator changes. This is due to the fact that the abstract Hilbert space-based framework introduced in sec. 2.2 is able to handle in a unified way both finite- and infinite-dimensional parameter vectors.
- From (47), the $\text{SCRB}(\boldsymbol{\theta}_0|g_0)$ in $\mathcal{P}_{\boldsymbol{\theta},g}$ is defined as the supremum of all the $\text{CRB}(\boldsymbol{\theta}_0|\boldsymbol{\eta}_{0,i})$ in each parametric sub-models $\{\mathcal{P}_{\boldsymbol{\theta},v_i}\}_{i \in I}$. This confirm the intuition that a semiparametric model contains less information on the parameter vector of interest $\boldsymbol{\theta}_0 \in \Theta$ than any of its possible parametric sub-models.

The following two points, closely linked to the semiparametric efficiency bound, must now be discussed:

- Q1 Which is the class of estimators to which the $\text{SCRB}(\boldsymbol{\theta}_0|g_0)$ applies?
 Q2 Is there any estimator able to achieve it?

Even if of crucial importance, answering to this two question is not a trivial task and they are the subject of intensive studies in statistics. Here, without any claim of completeness, we will try to provide some general hints leaving to the interested reader the task of further investigating the topic.

Regarding the first question, we can start by observing that we are looking for a class of estimators able to estimate the finite-dimensional parameter vector $\boldsymbol{\theta}_0 \in \Theta$ without assuming any knowledge on the nuisance function $g_0 \in \mathcal{G}$. In other words, a semiparametric estimator should be *robust* with respect to the lack of a complete specification of the functional form of the pdf of the observed dataset $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. As shown in [2, Sec. 2.2 and Ch. 7], [35, Ch.3], [33, Ch. 4] and the references therein, this concept can formalized in the class of *regular and asymptotically linear (RAL) estimators* [2, Sec. 2.2 and Ch. 7]. It is worth mentioning that all the robust M -, S -, L -, R - estimators belong to this class [33]. Specifically, a semiparametric RAL estimator $\hat{\boldsymbol{\theta}}_n \stackrel{\text{def}}{=} \hat{\boldsymbol{\theta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ of the finite-dimensional parameter vector $\boldsymbol{\theta}_0 \in \Theta$ from a set of n i.i.d. data $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ satisfy the following two properties:

1. \sqrt{n} -consistency: $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{P_X}(1)$,
2. *Asymptotically normality*: $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \underset{n \rightarrow \infty}{\sim} \mathcal{N}(\mathbf{0}, \Xi(\boldsymbol{\theta}_0, g_0))$.

Consequently, the following inequality holds [2, Ch. 2 and 3]:

$$\Xi(\boldsymbol{\theta}_0, g_0) \geq \text{SCRB}(\boldsymbol{\theta}_0|g_0). \quad (48)$$

This answer to the first point Q1.

Let us now move to the question Q2: is it possible to characterize the RAL estimator with the lowest error covariance matrix, i.e. the one that achieve the semi-parametric lower bound (this means to have an equality in (48))? Roughly speaking,

here we are looking for the semiparametric counterpart of the ML estimator discussed in sec. 2.3.1. This “one million dollars question” is still largely open and many fundamental aspect remain unresolved. Here, we limit ourselves to provide a very preliminary and informal sketch of answer. Basically, an efficient semiparametric estimator can be built by using the *one-step* procedure that we introduced for parametric models in sec. 2.3.1. We can note in fact that the *one-step* estimator in (38) is the only one among the four efficient estimators presented in sec. 2.3.1 that does not require the finite-dimensionality of the nuisance parameter space and then it may be generalized to semiparametric models [2, Ch. 7], [18], [37, Ch. 25].

Suppose to have at our disposal two preliminary, sub-optimal (\sqrt{n} -consistent but not necessarily efficient) estimators, say θ_n^* and g_n^* , of the parameter vector of interest $\theta_0 \in \Theta$ and of the nuisance function $g_0 \in \mathcal{G}$. An asymptotically efficient, one-step estimator $\hat{\theta}_{OP,n}$ of $\theta_0 \in \Theta$ can then be obtained as [2, Th. 1, Sec. 7.8]:

$$\hat{\theta}_{OP,n} = \theta_n^* + n^{-1/2} [\mathbf{C}(\theta_n^*, g_n^*)]^{-1} \bar{\Delta}_n(\theta_n^*, g_n^*), \quad (49)$$

where, the *efficient central sequence* is given by:

$$\bar{\Delta}(\mathbf{x}_1, \dots, \mathbf{x}_n; \theta_n^*, g_n^*) = \bar{\Delta}_n(\theta_n^*, g_n^*) \stackrel{\text{def}}{=} n^{-1/2} \sum_{i=1}^n \bar{\mathbf{s}}(\mathbf{x}_i; \theta_n^*, g_n^*). \quad (50)$$

Since $\bar{\mathbf{s}}_0$ depends on the true value of the parameter vector of interest θ_0 and on the true nuisance function g_0 , in the definition of the central sequence we exploit the term $\bar{\mathbf{s}}(\cdot, \theta_n^*, g_n^*)$ where θ_0 and g_0 are substituted with their two \sqrt{n} -consistent preliminary estimates θ_n^* and g_n^* . Moreover, the matrix $\mathbf{C}(\theta_n^*, g_n^*)$ represents the sample estimate of the efficient SFIM $\bar{\mathbf{I}}(\theta_0|g_0) = E_0(\bar{\mathbf{s}}_0 \bar{\mathbf{s}}_0^T)$ in (46) and can be expressed as:

$$\mathbf{C}(\theta_n^*, g_n^*) = \frac{1}{n} \left[\sum_{i=1}^n \bar{\mathbf{s}}(\mathbf{x}_i, \theta_n^*, g_n^*) \bar{\mathbf{s}}(\mathbf{x}_i, \theta_n^*, g_n^*)^T \right]. \quad (51)$$

According to [2, Theo.1 and Cor. 1], it can be show that, “under suitable regularity conditions”, $\hat{\theta}_{OP,n}$ in (49) satisfy the following properties:

1. \sqrt{n} -consistency: $\sqrt{n}(\hat{\theta}_{OP,n} - \theta_0) = O_{P_X}(1)$,
2. *Asymptotic normality and efficiency*: $\sqrt{n}(\hat{\theta}_{OP,n} - \theta_0) \underset{n \rightarrow \infty}{\sim} \mathcal{N}(\mathbf{0}, \text{SCR}(\theta_0|g_0))$.

It is worth stressing that the two preliminary estimators, even if sub-optimal, needs to be \sqrt{n} -consistent. While it is not difficult to derive \sqrt{n} -consistent estimators for the finite-dimensional parameter θ_0 , this may be a critical point for the functional estimator g^* whose rate of convergence is generally slower than \sqrt{n} [37, Sec. 25.8]. This may strongly limit the exploitability of semiparametric one-step estimator of the form in (49). As recently pointed out in [18], there may be some *rank-based* alternative procedure allowing us to avoid the need of the non-parametric \sqrt{n} -consistent estimator of the nuisance function $g \in \mathcal{G}$. We will further discuss this point in sec. 3.3 for the case of elliptical distributions.

2.4.2 Calculation of the projection operator $\Pi(\cdot|\mathcal{T}_{g_0})$

Before moving to the second part of the chapter, let us discuss one last fundamental point. All the previous results on semiparametric estimation rely on the definition of the semiparametric efficient score vector \bar{s}_0 in (45), which, in turn, is based on the calculation of the projection operator $\Pi(\cdot|\mathcal{T}_{g_0})$. Even if $\Pi(\cdot|\mathcal{T}_{g_0})$ is formally well-defined, i.e. it always exists and it is unique, its direct calculation could be involved and sometime impossible to obtain in closed form. Fortunately, many semiparametric models of interest for practical estimation problems are characterized by some additional regularity or/and invariance properties that could be exploited to derive $\Pi(\cdot|\mathcal{T}_{g_0})$. Specifically, in this subsection, we analyze the case in which the projection operator can be evaluated as a conditional expectation. As we will see in the second part of the chapter, this methodology will be readily applied to obtain the projection operator in the semiparametric model of the elliptical distributions.

In section 2.2, we defined \mathcal{H}_q as the Hilbert space of the q -variate, zero-mean, random functions on the probability space $(\mathcal{X}, \mathfrak{F}, P_X)$ with inner product given in (14) as $\langle \mathbf{h}_1, \mathbf{h}_2 \rangle = E(\mathbf{h}_1^T \mathbf{h}_2)$. Let us now introduce $\mathfrak{D}(\mathbf{x}) \subseteq \mathfrak{F}$ as the sub-sigma algebra in \mathfrak{F} generated by some (possibly multivariate) transformation d of the random vector \mathbf{x} , such that $D = d(\mathbf{x})$. As shown in e.g. [20, Ch. 23] and [2, App. 3], the set of all the q -variate, zero-mean, random functions on the probability space $(\mathcal{X}, \mathfrak{D}(\mathbf{x}), P_X)$ is a closed linear subspace, say \mathcal{D} , of the Hilbert space $\mathcal{H}_q \supseteq \mathcal{D}$. Consequently, according to Theorem 1, the projection of a generic element $\mathbf{h} \in \mathcal{H}_q$ onto the linear subspace \mathcal{D} , i.e. $\Pi(\mathbf{h}|\mathcal{D})$ exists and it is unique.

The question that we aim at answering here is: is there any relation between the projection operator $\Pi(\cdot|\mathcal{D})$ and the *conditional expectation* $E(\cdot|\mathfrak{D}(\mathbf{x})) = E(\cdot|D)$? To this end, let us provide a geometrical definition of the conditional expectation (see [20, Ch. 23] and [2, App. 3] for a more formal discussion).

Definition 6 Let $\mathbf{h} \in \mathcal{H}_q$ and $\mathbf{u} \in \mathcal{D} \subseteq \mathcal{H}_q$ be two q -variate, zero-mean, random functions on the probability spaces $(\mathcal{X}, \mathfrak{F}, P_X)$ and $(\mathcal{X}, \mathfrak{D}(\mathbf{x}), P_X)$ with $\mathfrak{D}(\mathbf{x}) \subseteq \mathfrak{F}$. Moreover, let D be a random vector defined on the sub-sigma algebra $\mathfrak{D}(\mathbf{x})$. Then the conditional expectation $E(\mathbf{h}|D)$ is the unique element in \mathcal{D} , s.t.:

$$\langle \mathbf{h} - E(\mathbf{h}|\mathfrak{D}(\mathbf{x})), \mathbf{u} \rangle = E((\mathbf{h} - E(\mathbf{h}|D))^T \mathbf{u}) = 0, \quad \forall \mathbf{u} \in \mathcal{D} \subseteq \mathcal{H}_q \quad (52)$$

It is immediate to verify that the definition of conditional expectation given in (52) is exactly equal to the one of the orthogonal projection given in (5) of Theorem 1 that define the projection operator. As a consequence, we have the desired relation:

$$\Pi(\cdot|\mathcal{D}) = E(\cdot|D). \quad (53)$$

Why should this relation help us? The answer is that, in some semiparametric model of interest, the semiparametric nuisance tangent space *i*) has some invariance structure with respect to a given group of transformations and *ii*) it admits a characterization through the sub-sigma algebras generated by the maximal invariant statistic of the relevant group of transformation [24]. As we will see in the next

section, the semiparametric model of the elliptical distributions belongs to the class of semiparametric models satisfying these two requirements.

3 Semiparametric estimation in Real Elliptically Symmetric (RES) Distributions

The second part of the chapter is dedicated to the application of the general semiparametric theory previously introduced to the problem of the joint estimation of the location vector and of the scatter matrix of a set of elliptically distributed real observations in the presence of an unknown density generator. An exhaustive presentation of the class of RES distributions can be found in the Chapter 1 where a comprehensive list of definitions, properties and examples (together with the relevant notation and nomenclature) is provided. Here we limit ourselves to recall only some fundamental points that will be extensively used in the rest of this section.

We start by recalling here the definition of *spherically distributed* random vector (see the Definition 2 in the Chapter 1). Let $\mathbf{x}_s \in \mathbb{R}^m$ be a real-valued random vector and let \mathcal{O} be the set of all the orthogonal transformations such that:

$$\begin{aligned} \mathcal{O} \ni O : \mathbb{R}^m &\rightarrow \mathbb{R}^m \\ \mathbf{x}_s &\mapsto O(\mathbf{x}_s) = \mathbf{O}\mathbf{x}_s, \end{aligned} \quad (54)$$

for any orthogonal matrix \mathbf{O} , i.e for any $\mathbf{O} \in \mathbb{R}^{m \times m}$ such that $\mathbf{O}^T \mathbf{O} = \mathbf{O} \mathbf{O}^T = \mathbf{I}_m$. Then, \mathbf{x}_s is said to be spherically distributed if its distribution is invariant to any orthogonal transformations $\mathbf{O} \in \mathcal{O}$:

$$\mathbf{x}_s =_d \mathbf{O}\mathbf{x}_s \quad (55)$$

As a consequence of the properties already given in the Definition 2 of the Chapter 1, the set of spherically distributed random vectors admits a *maximal invariant statistic* w.r.t. \mathcal{O} . For completeness, we recall here the definition of maximal invariant statistic [24, Ch. 6]. Let $\mathcal{D} = \{d\}$ be a group of one-to-one transformations on a sample space \mathcal{X} and let T an invariant statistic such that $T(\mathbf{x}) =_d T(d(\mathbf{x}))$, $\forall \mathbf{x} \in \mathcal{X}$ and $\forall d \in \mathcal{D}$. Then, T is a maximal invariant statistic on \mathcal{X} w.r.t. \mathcal{D} if $T(\mathbf{x}_1) =_d T(\mathbf{x}_2)$ implies that $\mathbf{x}_1 =_d d(\mathbf{x}_2)$, $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ and $\forall d \in \mathcal{D}$. Let us now go back to the spherically distributed random vectors. From the *stochastic representation* given in sec. 2.2 of the Chapter 1, we have:

$$\mathbf{x}_s =_d \sqrt{Q} \mathbf{u} =_d \mathcal{R} \mathbf{u}, \quad \mathcal{R} = \|\mathbf{x}_s\| \quad (56)$$

where $Q \sim p_Q(q)$ (the *second-order modular variate*) and $\mathcal{R} \sim p_{\mathcal{R}}(r)$ (the *modular variate*), have pdfs given in eq. (13) of the Chapter 1. The random vector $\mathbf{u} \sim \mathcal{U}(\mathbb{R}S^m)$ is uniformly distributed on the real unit sphere of dimension $m-1$ and, consequently, $\|\mathbf{u}\| = 1$, $E(\mathbf{u}) = \mathbf{0}$ and $E(\mathbf{u}\mathbf{u}^T) = m^{-1}\mathbf{I}_m$. Moreover, Q and \mathbf{u} (or \mathcal{R} and \mathbf{u}) are

independent. As a direct consequence, we have that $\|\mathbf{x}_s\|^2$ and $\mathbf{x}_s/\|\mathbf{x}_s\| (= {}_d \mathbf{u})$ are independent as well.

The stochastic representation in (56) provides us with a one-to-one mapping between a spherically distributed vector \mathbf{x}_s and a couple $(\mathcal{R}, \mathbf{u})$. Moreover, for any couple of spherically distributed random vectors \mathbf{x}_s and \mathbf{y}_s we have that:

$$\|\mathbf{x}_s\| =_d \|\mathbf{y}_s\| \Rightarrow \mathbf{x}_s =_d \mathbf{O} \mathbf{y}_s, \quad \forall \mathbf{O} \in \mathcal{O}, \quad (57)$$

where \mathcal{O} is the group of the orthogonal transformations defined in (54). Consequently, according to the definition previously provided, $R = \|\mathbf{x}\|$ is a maximal invariant statistic for the set of spherically distributed random vectors.

3.1 The RES distributions as a semiparametric group model

A semiparametric group model is a statistical model generated by the *action* of a group of transformations, say \mathcal{A} , on a random vector \mathbf{x} that is distributed according to a given class of distributions [2, Sec. 4.2]. As we will show in the following, this abstract definition fits with the RES distributions class. Note that, all the subsequent results can be obtained for the class of the circular Complex Elliptically Symmetric (C-CES) distribution as well. The interested reader can found the details in [7].

Definition 7 Let $\boldsymbol{\mu} \in \mathbb{R}^m$ be a real vector and let $\boldsymbol{\Sigma} \in \mathcal{M}$ be a positive-definite, symmetric matrix such that $\boldsymbol{\Sigma} = \mathbf{T}\mathbf{T}^T$. Among all the possible \mathbf{T} we always choose $\mathbf{T} = \boldsymbol{\Sigma}^{1/2}$. Let us define the group \mathcal{A} of the affine transformations $\alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\cdot)$ s.t.:

$$\begin{aligned} \mathcal{A} \ni \alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\cdot) : \mathbb{R}^m &\rightarrow \mathbb{R}^m, \quad \forall \boldsymbol{\mu} \in \mathbb{R}^m, \forall \boldsymbol{\Sigma} \in \mathcal{M} \\ \mathbb{R}^m \ni \mathbf{y} &\mapsto \alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{y}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{y}. \end{aligned} \quad (58)$$

Note that the neutral element of the group \mathcal{A} is given by $\alpha_e(\cdot) = \alpha_{\mathbf{0}, \mathbf{I}_m}(\cdot)$ while the inversion operator is $\alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^{-1}(\cdot) = \boldsymbol{\Sigma}^{-1/2}(\cdot - \boldsymbol{\mu})$.

From the stochastic representation provided in eq. (7) of sec. 2.2 of the Chapter 1, we can readily deduce that the set of RES distributions is closed under the action of the group \mathcal{A} on the set of the spherically distributed random vectors. In particular, we have that:

$$\mathbf{x} =_d \alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{x}_s) =_d \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{x}_s =_d \boldsymbol{\mu} + \sqrt{Q} \boldsymbol{\Sigma}^{1/2} \mathbf{u} =_d \boldsymbol{\mu} + \mathcal{R} \boldsymbol{\Sigma}^{1/2} \mathbf{u}. \quad (59)$$

Moreover, by noticing that the Jacobian of the inverse operator $\alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^{-1}(\cdot)$ is $|\mathbf{J}(\alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^{-1})| = |\boldsymbol{\Sigma}^{-1/2}| = |\boldsymbol{\Sigma}|^{-1/2}$, in the absolutely continuous case, we can deduce the pdf p_X of any RES-distributed random vector \mathbf{x} from the one of the relevant spherically distributed random vector \mathbf{x}_s given in eq. (4) of sec. 2.2 of the Chapter 1 as $p(\mathbf{x}_s) = g(\|\mathbf{x}_s\|^2)$:

$$\text{RES}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g) = p_X(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, g) = |\boldsymbol{\Sigma}|^{-1/2} g(\|\alpha_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^{-1}(\mathbf{x})\|^2), \quad (60)$$

that is in line with the expression given in (1) in sec. 1. Finally, in order to avoid the scale ambiguity between the scatter matrix Σ and the density generator g , we rewrite $p(\mathbf{x}; \boldsymbol{\mu}, \Sigma, g)$ as function of the *shape matrix* $\mathbf{V} = \Sigma/s(\Sigma) \in \mathcal{M}_s$ already defined in (2). Finally, the set of RES distribution can be cast as a *semiparametric group model* where the finite-dimensional parameter space is given by:⁵

$$\Theta \stackrel{\text{def}}{=} \{\boldsymbol{\theta} \in \mathbb{R}^{q-1} | \boldsymbol{\theta} = (\boldsymbol{\mu}^T, v(\mathbf{V})^T)^T; \boldsymbol{\mu} \in \mathbb{R}^m, \mathbf{V} \in \mathcal{M}_s\}, \quad (61)$$

where $q = m(m+3)/2$ and the “-1” is due to the fact that $\dim(\mathcal{M}_s) = \dim(\mathcal{M}) - 1$ as discussed in sec. 1. For further reference, we introduce the “unconstrained” finite-dimensional parameter space as:

$$\Theta_u \stackrel{\text{def}}{=} \{\boldsymbol{\theta}_u \in \mathbb{R}^q | \boldsymbol{\theta}_u = (\boldsymbol{\mu}^T, v(\mathbf{V})^T)^T; \boldsymbol{\mu} \in \mathbb{R}^m, \mathbf{V} \in \mathcal{M}\}, \quad (62)$$

The infinite-dimensional parameter space is given by the set $\mathcal{G} \ni g$ of all the possible density generators. Note that, for technical reasons, we always assume that the set \mathcal{G} is made up of all the density generators satisfying the following two additional properties: 1) the limit $\lim_{t \rightarrow +\infty} t^{m/2} g(t)$ exists and 2) $\int_0^\infty t^{m/2} g(t) dt < \infty$ that corresponds to impose $E\{Q\} < \infty$.

The semiparametric model that we are going to consider in the following is:

$$\mathcal{P}_{\boldsymbol{\mu}, \mathbf{V}, g} \stackrel{\text{def}}{=} \left\{ p_X(\mathbf{x}) = |\mathbf{V}|^{-1/2} g(\|\alpha_{\boldsymbol{\mu}, \mathbf{V}}^{-1}(\mathbf{x})\|^2), (\boldsymbol{\mu}^T, v(\mathbf{V})^T) \in \Theta, g \in \mathcal{G} \right\}. \quad (63)$$

In the following, we will indicate as $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}_0^T, v(\mathbf{V}_0)^T)^T \in \Theta$ the true finite-dimensional vector of interest that we need to estimate and as $g_0 \in \mathcal{G}$ the true nuisance function. Consequently, the true but unknown, data pdf is $p_0(\mathbf{x}) = |\mathbf{V}_0|^{-1/2} g_0(\|\alpha_{\boldsymbol{\mu}_0, \mathbf{V}_0}^{-1}(\mathbf{x})\|^2)$. After having formally placed the set of the RES distributions in the semiparametric framework, in the next two sections we will first evaluate the semiparametric Cramér-Rao Bound (SCRb), defined in Theorem 2, for the joint estimation of the location vector $\boldsymbol{\mu}_0$ and the shape matrix \mathbf{V}_0 . Secondly, a class of one-step estimators of \mathbf{V}_0 of the type given in (49) will be presented.

3.2 The Semiparametric Cramér-Rao Bound for RES distributions

The derivation of a SCRb for estimation problems involving elliptical data has been firstly investigated by Bickel in [3], where a bound for the estimation of the inverse of the scatter matrix has been derived. Further discussion and analysis related to the joint estimation of $\boldsymbol{\mu}$ and \mathbf{V} have been also presented in [2, Sec. 4.2 and sec. 6.3]. More recently, in a series of papers [14, 13, 15, 16, 31, 30], Hallin, Paindaveine

⁵ The operator $v(\mathbf{A})$ indicates the $m(m+1)/2$ -dimensional vector of the entries of the lower (or upper) sub-matrix of a symmetric \mathbf{A} . This notation has been adopted in order to keep the consistency with the Chapter 1. Note however that in our previous publications (see e.g. [7, 8, 10]) it was expressed as $\text{vecs}(\cdot)$.

and their coauthors exploited an original semiparametric generalization of the Le Cam's theory on Local Asymptotical Normal (LAN) families of distributions [18] to derive optimal inference algorithms on elliptically distributed data. However, even if profound and of crucial importance, the work of Hallin and Paindavaine requires a strong knowledge of the Le Cam's theory, and this falls outside the scope of this chapter. Our aim here is to show how the geometrical concepts previously introduced can be applied in a concrete problem as the joint estimation of $\boldsymbol{\mu}$ and \mathbf{V} in elliptical data. To this end, we will basically follow the line of reasoning proposed in [2, Secs. 4.2 and 6.3], along with the results in [8, 7].

As stated in Theorem 2, the basic ingredient that we need to derive the SCRB is the efficient score vector $\bar{\mathbf{s}}_{u,0} = \mathbf{s}_{\theta_{u,0}} - \Pi(\mathbf{s}_{\theta_{u,0}}|\mathcal{T}_{g_0})$ that, in turn, depends on the score vector of the parameters of interest $\mathbf{s}_{\theta_{u,0}}$, the semiparametric nuisance tangent space \mathcal{T}_{g_0} and the related projection operator $\Pi(\cdot|\mathcal{T}_{g_0})$.

3.2.1 Step 1: Calculation of $\mathbf{s}_{\theta_{u,0}}$

According to Definition 2, the score vector $\mathbf{s}_{\theta_{u,0}}$ of the “unconstrained” parameters of interest $\boldsymbol{\theta}_{u,0} = (\boldsymbol{\mu}_0^T, v(\mathbf{V}_0)^T)^T \in \Theta_u$ in (62) can be expressed as:

$$\mathbf{s}_{\theta_{u,0}} = \nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0, g_0) = \begin{pmatrix} \mathbf{s}_{\boldsymbol{\mu}_0}(\mathbf{x}) \\ \mathbf{s}_{v(\mathbf{V}_0)}(\mathbf{x}) \end{pmatrix}, \quad \boldsymbol{\theta}_0 \in \Theta_u \quad (64)$$

where:

$$\mathbf{s}_{\boldsymbol{\mu}_0} = \nabla_{\boldsymbol{\mu}} \ln p_X(\mathbf{x}|\boldsymbol{\mu}_0, \mathbf{V}_0, g_0), \quad (65)$$

$$\mathbf{s}_{v(\mathbf{V}_0)} = \nabla_{v(\mathbf{V})} \ln p_X(\mathbf{x}|\boldsymbol{\mu}_0, \mathbf{V}_0, g_0). \quad (66)$$

By using the equality:

$$\mathcal{Q}_0 = \|\alpha_{\boldsymbol{\mu}_0, \mathbf{V}_0}^{-1}(\mathbf{x})\|^2 = (\mathbf{x} - \boldsymbol{\mu}_0)^T \mathbf{V}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) =_d \mathcal{Q}, \quad (67)$$

through direct calculation [27, Ch. 8], [8], we have

$$\mathbf{s}_{\boldsymbol{\mu}_0} = -2\psi_0(\mathcal{Q}_0)\mathbf{V}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0) =_d -2\sqrt{\mathcal{Q}}\psi_0(\mathcal{Q})\mathbf{V}_0^{-1/2}\mathbf{u}, \quad (68)$$

where the last two equalities follow directly from the Stochastic Representation of a RES vector given in (59) and

$$\psi_0(t) \stackrel{\text{def}}{=} \frac{1}{g_0(t)} \frac{dg_0(t)}{dt}. \quad (69)$$

Moreover, after some matrix calculation, we get [8]:

$$\begin{aligned} \mathbf{s}_{v(\mathbf{V}_0)} &= -\mathbf{D}_m^T \left(\frac{1}{2} \text{vec}(\mathbf{V}_0^{-1}) + \psi_0(Q_0) \left(\mathbf{V}_0^{-1} \otimes \mathbf{V}_0^{-1} \right) \text{vec}((\mathbf{x} - \boldsymbol{\mu}_0)(\mathbf{x} - \boldsymbol{\mu}_0)^T) \right) \\ &= {}_d -\mathbf{D}_m^T \left(\frac{1}{2} \text{vec}(\mathbf{V}_0^{-1}) + Q\psi_0(Q) \left(\mathbf{V}_0^{-1/2} \otimes \mathbf{V}_0^{-1/2} \right) \text{vec}(\mathbf{u}\mathbf{u}^T) \right), \end{aligned} \quad (70)$$

where, as before, the last two equalities follow from (59) while \mathbf{D}_m is the $m^2 \times m(m+1)/2$ duplication matrix that is implicitly defined by the equality $\mathbf{D}_m v(\mathbf{A}) = \text{vec}(\mathbf{A})$ for any $m \times m$ symmetric matrix \mathbf{A} [25, 26].

Parametric Cramér-Rao bound If the density generator g_0 is known, the FIM for the estimation of $\boldsymbol{\theta}_{u,0} = (\boldsymbol{\mu}_0^T, v(\mathbf{V}_0)^T)^T \in \Theta_u$, can be obtained as a block matrix of the form:

$$\mathbf{I}(\boldsymbol{\theta}_{u,0}) = \mathbb{E}_0(\mathbf{s}_{\boldsymbol{\theta}_{u,0}} \mathbf{s}_{\boldsymbol{\theta}_{u,0}}^T) = \begin{pmatrix} \mathbf{C}_0(\mathbf{s}_{\boldsymbol{\mu}_0}) & \mathbf{0} \\ \mathbf{0}^T & \mathbf{C}_0(\mathbf{s}_{v(\mathbf{V}_0)}) \end{pmatrix}, \quad (71)$$

that is in line with the one provided in eqs. (88) and (89) of sec. 6.2 of the Chapter 1. Through direct calculation, it is easy to verify that:

$$\mathbf{C}_0(\mathbf{s}_{\boldsymbol{\mu}_0}) = 4m^{-1} \mathbb{E}(Q\psi_0^2(Q)) \mathbf{V}_0^{-1}, \quad (72)$$

$$\mathbf{C}_0(\mathbf{s}_{v(\mathbf{V}_0)}) = \mathbf{D}_m^T \left(a_1 \text{vec}(\mathbf{V}_0^{-1}) \text{vec}(\mathbf{V}_0^{-1})^T + a_2 \mathbf{V}_0^{-1} \otimes \mathbf{V}_0^{-1} \right) \mathbf{D}_m \quad (73)$$

$$a_1 = \frac{a_2}{2} - \frac{1}{4}, \quad (74)$$

$$a_2 = \frac{2\mathbb{E}(Q^2\psi_0^2(Q))}{m(m+2)}. \quad (75)$$

Finally, the CRB on the estimation of $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}_0^T, v(\mathbf{V}_0)^T)^T \in \Theta$ in (61) with the constraint on the shape $s(\mathbf{V}_0) = 1$ can be obtained by using the same procedure discussed in [28]. Specifically, let us define the Jacobian of the constraint as $\mathbf{J}_s(\mathbf{V}) = \nabla_{v(\mathbf{V}_0)}(s(\mathbf{V}_0) - 1)$ and $\mathbf{U}(\mathbf{V}_0) \equiv \mathbf{U}_{\mathbf{V}_0}$ as the matrix whose columns form a basis for an orthonormal null space of $\mathbf{J}_s(\mathbf{V}_0)$. Then, the Constrained CRB (CCRB) for the estimation of $\boldsymbol{\theta}_0 \in \Theta$ when the density generator is known is given by:

$$\text{CCRB}(\boldsymbol{\theta}_0) = \begin{pmatrix} \frac{m}{4\mathbb{E}(Q\psi_0^2(Q))} \mathbf{V}_0 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{U}_{\mathbf{V}_0} \left(\mathbf{U}_{\mathbf{V}_0}^T \mathbf{C}_0(\mathbf{s}_{v(\mathbf{V}_0)}) \mathbf{U}_{\mathbf{V}_0} \right)^{-1} \mathbf{U}_{\mathbf{V}_0}^T \end{pmatrix}. \quad (76)$$

Note that the block-diagonal structure of $\mathbf{I}(\boldsymbol{\theta}_{u,0})$ in (71) implies that not knowing the mean value $\boldsymbol{\mu}_0$ does not have any impact on the asymptotic performance in the estimation of the shape matrix \mathbf{V}_0 and vice versa.

3.2.2 Step 2: Calculation of the projection $\Pi(\mathbf{s}_{\boldsymbol{\theta}_{u,0}} | \mathcal{T}_{g_0})$

Let us now move to the crucial step of the evaluation of the projection $\Pi(\mathbf{s}_{\boldsymbol{\theta}_{u,0}} | \mathcal{T}_{g_0})$ of $\mathbf{s}_{\boldsymbol{\theta}_{u,0}}$ in (64) onto the semiparametric nuisance tangent space \mathcal{T}_{g_0} at the true, but

unknown, density generator g_0 . To this end, we apply a fundamental property of the semiparametric group model discussed in [2, Sec. 4.2, Lemma 3]:

Lemma 1 *Let \mathcal{A} be the group defined in (58). Let \mathcal{T}_{g_0} and \mathcal{S}_{g_0} be the semiparametric nuisance tangent spaces of the RES and of the spherically symmetric distributions at the true density generator g_0 , respectively. Then, \mathcal{T}_{g_0} can be obtained from \mathcal{S}_{g_0} as:*

$$\mathcal{T}_{g_0} = \left\{ \mathbf{u} \in \mathcal{H}_q \mid \mathbf{u} = \mathbf{v} \circ \alpha_{\mu_0, \mathbf{v}_0}^{-1} : \mathbf{v} \in \mathcal{S}_{g_0} \subseteq \mathcal{H}_q \right\}, \quad (77)$$

where \circ indicates the composition of functions. Moreover, the projection of the score vector of the parameters of interest $\mathbf{s}_{\theta_{u,0}} \in \mathcal{H}_q$ on \mathcal{T}_{g_0} can be expressed as:

$$\Pi(\mathbf{s}_{\theta_{u,0}} | \mathcal{T}_{g_0}) = \Pi(\mathbf{s}_{\theta_{u,0}} \circ \alpha_{\mu_0, \mathbf{v}_0} | \mathcal{S}_{g_0}) \circ \alpha_{\mu_0, \mathbf{v}_0}^{-1}. \quad (78)$$

Intuitively, Lemma 1 can be explained as follows. The set of spherically symmetric distributions is characterized by the neutral element $\alpha_e(\cdot) = \alpha_{\mathbf{0}, \mathbf{I}_m}(\cdot)$ of the group \mathcal{A} in (58). We can derive the semiparametric nuisance tangent space \mathcal{S}_{g_0} and the related projection operator $\Pi(\cdot | \mathcal{S}_{g_0})$ at α_e , then we can translate them in the true RES distribution p_0 characterized by an affine transformation $\alpha_{\mu_0, \mathbf{v}_0}$ by means of the relations (77) and (78).

The following theorem, whose proof is reported in Appendix, provide us with an explicit expression of the semiparametric nuisance tangent space \mathcal{S}_{g_0} of the spherically symmetric distribution at g_0 .

Theorem 3 *Let \mathcal{P}_g be the non-parametric model of the spherically symmetric pdfs:*

$$\mathcal{P}_g \stackrel{\text{def}}{=} \{p_X(\mathbf{x}) = g(\|\mathbf{x}\|^2), g \in \mathcal{G}\} \quad (79)$$

Its semiparametric nuisance tangent space $\mathcal{S}_{g_0} = \mathcal{L} \times \dots \times \mathcal{L}$ is a q -replicating Hilbert space such that:

$$\mathcal{S}_{g_0} = \{\mathbf{u} \in \mathcal{H}_q \mid \mathbf{u} = (l_1, l_2, \dots, l_q)^T, l_j \in \mathcal{L}, j = 1, \dots, q\}, \quad (80)$$

$$\mathcal{L} = \{l \in \mathcal{H} \mid l(\mathbf{x}) = l(\mathbf{O}\mathbf{x}), \forall \mathbf{O} \in \mathcal{O}\} = \{l \in \mathcal{H} \mid l(\mathbf{x}) = l(\|\mathbf{x}\|)\}, \quad (81)$$

where \mathcal{O} is the group of orthogonal transformations defined in (54).

As shown in (57), $\|\mathbf{x}\| = R$ is a maximal invariant statistic for the set of spherically distributed random vectors. Then, following the discussion previously given in sec. 2.4.2, the set \mathcal{L} in (81) can be equivalently characterized as the set of the univariate random functions defined on the probability space $(\mathcal{X}, \mathfrak{D}(\mathbf{x}), P_X)$, where $\mathfrak{D}(\mathbf{x}) \subseteq \mathfrak{F}$ is the sub-sigma algebra generated by the modular variate $\|\mathbf{x}\| = R$. Then, as a consequence of the Definition 6 (see also [2, Sec. 6.3]), we have that the projection onto \mathcal{S}_{g_0} can be expressed as a conditional expectation w.r.t. the modular variate R :

$$\Pi(\mathbf{h} | \mathcal{S}_{g_0}) = \mathbb{E}(\mathbf{h} | R), \quad \forall \mathbf{h} \in \mathcal{H}_q. \quad (82)$$

Finally, the desired $\Pi(\mathbf{s}_{\theta_0} | \mathcal{T}_{g_0})$ can be obtained by combining the results in (78) and in (82). Specifically, after involved but standard calculation (see [8] for all the

details), we get:

$$\Pi(\mathbf{s}_{\theta_{u,0}}|\mathcal{T}_{g_0}) = \begin{pmatrix} \Pi(\mathbf{s}_{\mu_0}|\mathcal{T}_{g_0}) \\ \Pi(\mathbf{s}_{v(\mathbf{V}_0)}|\mathcal{T}_{g_0}) \end{pmatrix}, \quad (83)$$

where:

$$\Pi(\mathbf{s}_{\mu_0}|\mathcal{T}_{g_0}) =_d -2\sqrt{Q}\psi_0(Q)\mathbf{V}_0^{-1/2}\mathbf{E}(\mathbf{u}) = \mathbf{0}, \quad (84)$$

$$\Pi(\mathbf{s}_{v(\mathbf{V}_0)}|\mathcal{T}_{g_0}) =_d -\mathbf{D}_m^T \left(\frac{1}{2} + \frac{1}{m}Q\psi_0(Q) \right) \text{vec}(\mathbf{V}_0^{-1}). \quad (85)$$

Note that, eq. (84) tells us that the score vector of the location vector \mathbf{s}_{μ_0} is orthogonal to the nuisance tangent space \mathcal{T}_{g_0} :

$$\Pi(\mathbf{s}_{\mu_0}|\mathcal{T}_{g_0}) = \mathbf{0} \Rightarrow \mathbf{s}_{\mu_0} \perp \mathcal{T}_{g_0}. \quad (86)$$

This implies that the missing a-priori knowledge of the true density generator g_0 does not have any (asymptotic) impact on the performance of an estimator of the location vector μ_0 [11].

3.2.3 Step 3: Calculation of the efficient SFIM $\bar{\mathbf{I}}(\theta_{u,0}|g_0)$ and of SCRB($\theta_0|g_0$)

By subtracting (83) from (64), the semiparametric efficient score vector in (45) can be evaluated as:

$$\begin{aligned} \bar{\mathbf{s}}_{u,0} &= \mathbf{s}_{\theta_{u,0}} - \Pi(\mathbf{s}_{\theta_{u,0}}|\mathcal{T}_{g_0}) = \begin{pmatrix} \bar{\mathbf{s}}_{\mu_0} \\ \bar{\mathbf{s}}_{v(\mathbf{V}_0)} \end{pmatrix} \\ &= \begin{pmatrix} -2\sqrt{Q}\psi_0(Q)\mathbf{V}_0^{-1/2}\mathbf{u} \\ -\mathbf{D}_m^T Q\psi_0(Q) \left(\mathbf{V}_0^{-1/2} \otimes \mathbf{V}_0^{-1/2} \text{vec}(\mathbf{u}\mathbf{u}^T) - \frac{1}{m} \text{vec}(\mathbf{V}_0^{-1}) \right) \end{pmatrix}. \end{aligned} \quad (87)$$

Let us now introduce the matrix

$$\Pi_{\text{vec}(\mathbf{I}_m)}^\perp = \mathbf{I}_{m^2} - m^{-1} \text{vec}(\mathbf{I}_m) \text{vec}(\mathbf{I}_m)^T. \quad (88)$$

as the orthogonal projection matrix on the orthogonal complement of $\text{span}(\text{vec}(\mathbf{I}_m))$, and the matrix $\mathbf{N}_{\mathbf{V}_0}$ as

$$\mathbf{N}_{\mathbf{V}_0} \stackrel{\text{def}}{=} \left(\mathbf{V}_0^{-1/2} \otimes \mathbf{V}_0^{-1/2} \right) \Pi_{\text{vec}(\mathbf{I}_m)}^\perp. \quad (89)$$

Then, the semiparametric efficient score vector $\bar{\mathbf{s}}_{v(\mathbf{V}_0)}$ can be expressed as:

$$\bar{\mathbf{s}}_{v(\mathbf{V}_0)} =_d -\mathbf{D}_m^T Q\psi_0(Q) \mathbf{N}_{\mathbf{V}_0} \text{vec}(\mathbf{u}\mathbf{u}^T). \quad (90)$$

We can now explicitly evaluate the efficient SFIM $\bar{\mathbf{I}}(\theta_0|g_0)$ as [8]:

$$\bar{\mathbf{I}}(\theta_{u,0}|g_0) = \mathbf{E}_0(\bar{\mathbf{s}}_{u,0} \bar{\mathbf{s}}_{u,0}^T) = \begin{pmatrix} \mathbf{C}_0(\bar{\mathbf{s}}_{\mu_0}) & \mathbf{0} \\ \mathbf{0}^T & \mathbf{C}_0(\bar{\mathbf{s}}_{v(\mathbf{V}_0)}) \end{pmatrix}, \quad (91)$$

where, as for the FIM in (71), the semiparametric efficient cross information matrix $E_0(\bar{\mathbf{s}}_{\mu_0} \bar{\mathbf{s}}_{v(\mathbf{V}_0)}^T)$ is nil. From the previous results and through some algebra, we get:

$$\mathbf{C}_0(\bar{\mathbf{s}}_{\mu_0}) = \mathbf{C}_0(\mathbf{s}_{\mu_0}) = 4m^{-1}E(Q\psi_0^2(Q))\mathbf{V}_0^{-1}, \quad (92)$$

$$\begin{aligned} \mathbf{C}_0(\bar{\mathbf{s}}_{v(\mathbf{V}_0)}) &= a_2 \mathbf{D}_m^T \left(\mathbf{V}_0^{-1} \otimes \mathbf{V}_0^{-1} - \frac{1}{m} \text{vec}(\mathbf{V}_0^{-1}) \text{vec}(\mathbf{V}_0^{-1})^T \right) \mathbf{D}_m \\ &= a_2 \mathbf{D}_m^T \left[\left(\mathbf{V}_0^{-1/2} \otimes \mathbf{V}_0^{-1/2} \right) \Pi_{\text{vec}(\mathbf{I}_m)}^\perp \left(\mathbf{V}_0^{-1/2} \otimes \mathbf{V}_0^{-1/2} \right) \right] \mathbf{D}_m \\ &= a_2 \mathbf{D}_m^T \mathbf{N}_{\mathbf{V}_0} \mathbf{N}_{\mathbf{V}_0}^T \mathbf{D}_m. \end{aligned} \quad (93)$$

where a_2 is given in (75). Note also that in the second inequality we use the fact that the projection matrix $\Pi_{\text{vec}(\mathbf{I}_m)}^\perp$ is idempotent and that \mathbf{V}_0 is symmetric. It is important to highlight that the block-diagonal structure of the efficient SFIM $\bar{\mathbf{I}}(\theta_{u,0}|g_0)$ implies that the estimate of the location vector μ_0 and of the shape matrix \mathbf{V}_0 are asymptotically decorrelated. Consequently, for the estimation of \mathbf{V}_0 we may center the data by means of any \sqrt{n} -consistent estimator of μ_0 without any (asymptotic) impact on the estimation performance of \mathbf{V}_0 [11]. Note that this result can also be obtained by using the semiparametric Le Cam theory as shown in [14, 13].

We are finally able to state the desired result: an explicit expression of the Semiparametric Cramér-Rao Bound for the estimation of $\theta_0 = (\mu_0^T, v(\mathbf{V}_0)^T)^T \in \Theta$ in RES-distributed data with the constraint on the shape $s(\mathbf{V}_0) = 1$. By introducing the matrix $\mathbf{U}(\mathbf{V}_0) \equiv \mathbf{U}_{\mathbf{V}_0}$ as we did in (76), the *constrained* SCRB($\theta_0|g_0$) can be expressed as:

$$\text{CSCRB}(\theta_0|g_0) = \begin{pmatrix} \frac{m}{4E(Q\psi_0^2(Q))} \mathbf{V}_0 & \mathbf{0} \\ \mathbf{0}^T & \frac{m(m+2)}{2E(Q^2\psi_0^2(Q))} \mathbf{U}_{\mathbf{V}_0} \left(\mathbf{U}_{\mathbf{V}_0}^T \mathbf{D}_m^T \mathbf{N}_{\mathbf{V}_0} \mathbf{N}_{\mathbf{V}_0}^T \mathbf{D}_m \mathbf{U}_{\mathbf{V}_0} \right)^{-1} \mathbf{U}_{\mathbf{V}_0}^T \end{pmatrix}. \quad (94)$$

We note, in passing, that the corresponding bounds for C-CES distributed observation can be found in [7] along with the semiparametric generalization of the celebrated Slepian-Bangs formula.

Let us close the chapter by answering to this fundamental question: *Is it possible to derive a RAL estimator of the shape matrix \mathbf{V}_0 able to achieve the CSCRB($\theta_0|g_0$)?*

3.3 The rank-based R -estimators of the shape matrix

Semiparametric efficient estimators of the finite-dimensional parameter vector $\theta_0 \in \Theta$ in the presence of a nuisance function $g \in \mathcal{G}$ has been already introduced in sec. 2.4.1. In particular, we saw that the *semiparametric one-step estimators* in (49) have the remarkable property to be asymptotically semiparametric efficient RAL estimators. In this section, we will specialize the general theory presented in sec.

2.4.1 to the estimation of the shape matrix $\mathbf{V}_0 \in \mathcal{M}_s$ from a set of n i.i.d. observations $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, in the presence of an unknown density generator $g_0 \in \mathcal{G}$. In order to keep the discussion as clear as possible, in the following we assume the location vector is nil, i.e. $\boldsymbol{\mu} = \mathbf{0}$. Then, the finite-dimensional parameter to be estimated is simply given by $\mathbf{V}_0 \in \mathcal{M}_s$. This reduction of the parameter space can be done without loss of generality, since as already discussed in sec. 3.2.2, the score vector of $\boldsymbol{\mu}$ is orthogonal to the nuisance tangent space, i.e. $\Pi(\mathbf{s}_{\boldsymbol{\mu}_0} | \mathcal{T}_{g_0}) = \mathbf{0}$. Consequently, when $\boldsymbol{\mu} \neq \mathbf{0}$ we could always center the data by means of any \sqrt{n} -consistent estimator $\hat{\boldsymbol{\mu}}_n$ without any (asymptotic) impact on the estimation of $\mathbf{V}_0 \in \mathcal{M}_s$ [11]. However, in order to have a good finite-sample performance, we suggest to use a \sqrt{n} -consistent and *robust* location estimator as the one proposed in [19]. A further reduction of the parameter space can be done by choosing, among all the possible constraints of the shape matrix, the one that forces its first top-left element to be equal to 1:

$$\mathbf{V}_{0,1} = \boldsymbol{\Sigma}_0 / [\boldsymbol{\Sigma}_0]_{11} \Rightarrow s(\mathbf{V}_{0,1}) = [\mathbf{V}_{0,1}]_{11} = 1. \quad (95)$$

Note that, in the following we will use the subscript 1, i.e. \mathbf{A}_1 to indicate that the generic matrix \mathbf{A} satisfies the constraint in (95), that is $[\mathbf{A}]_{11} = 1$. To notice the impact that this choice has on the definition of the parameter space, let us start by introducing some useful notation. Let \mathbf{A} be an $m \times m$ symmetric matrix, then $\underline{v}(\mathbf{A})$ is implicitly defined as $\underline{v}(\mathbf{A}) = [[\mathbf{A}]_{11}, \underline{v}(\mathbf{A})^\top]^\top$.⁶ Moreover, if $[\mathbf{A}]_{11} = 0$, then \mathbf{M}_m is the $(m(m+1)/2 - 1) \times m^2$ matrix s.t. $\mathbf{M}_m^\top \underline{v}(\mathbf{A}) = \text{vec}(\mathbf{A})$. Note that \mathbf{M}_m^\top can be obtained from the *duplication matrix* \mathbf{D}_m by removing its first column. Consequently, the finite-dimensional parameter vector to estimate is given by:

$$\boldsymbol{\theta}_0 = \underline{v}(\mathbf{V}_{1,0}) \in \Theta \subseteq \mathbb{R}^{m(m+1)/2-1}, \quad [\mathbf{V}_{0,1}]_{11} = 1. \quad (96)$$

For ease of readability, let us recall here the expression of a semiparametric one-step estimator given in (49) as:

$$\hat{\boldsymbol{\theta}}_{op,n} = \boldsymbol{\theta}_n^* + n^{-1/2} [\mathbf{C}(\boldsymbol{\theta}_n^*, g_n^*)]^{-1} \bar{\Delta}_n(\boldsymbol{\theta}_n^*, g_n^*),$$

and try to list the ingredient that we need to implement it for our estimation problem:

1. A \sqrt{n} -consistent preliminary estimator $\boldsymbol{\theta}_n^*$ of the constrained shape matrix. A good candidate for $\boldsymbol{\theta}_n^*$ is the celebrated Tyler M -estimator [36] given in sec. 6.4 of the Chapter 1, since it is proven to be \sqrt{n} -consistent under any (unknown) density generator. We indicate this estimator as $\mathbf{V}_{1,n}^*$, where the subscript “1” remember us that the constraint in (95) needs to be imposed.
2. The efficient central sequence $\bar{\Delta}_n(\mathbf{V}_{1,n}^*, g_n^*)$. By specializing the definition of semiparametric efficient score vector for \mathbf{V}_0 given in (90) to the parameter space (96), we have that [10]:

$$\bar{\mathbf{s}}_{\underline{v}(\mathbf{V}_{1,0})} =_d -\mathbf{Q}\psi_0(\mathbf{Q})\mathbf{K}_{\mathbf{V}_{1,0}}\text{vec}(\mathbf{u}\mathbf{u}^T). \quad (97)$$

⁶ Note that, in our previous publications (see e.g. [7, 8, 10]), $\underline{v}(\cdot)$ was expressed as $\text{vecs}(\cdot)$

where, for notation simplicity, we define the matrix

$$\mathbf{K}_{\mathbf{V}_{1,0}} \stackrel{\text{def}}{=} \mathbf{M}_m \left(\mathbf{V}_{1,0}^{-1/2} \otimes \mathbf{V}_{1,0}^{-1/2} \right) \Pi_{\text{vec}(\mathbf{I}_m)}^\perp, \quad (98)$$

Consequently, the efficient central sequence can be expressed as:

$$\bar{\Delta}_n(\mathbf{V}_{1,n}^\star, g_n^\star) = -n^{-1/2} \mathbf{K}_{\mathbf{V}_{1,n}^\star} \sum_{i=1}^n Q_i^\star \psi_n^\star(Q_i^\star) \text{vec}(\mathbf{u}_i^\star (\mathbf{u}_i^\star)^T), \quad (99)$$

where, from the stochastic representation in (56) and (59), we have:

$$Q_i^\star = \mathbf{x}_i^T [\mathbf{V}_{1,n}^\star]^{-1} \mathbf{x}_i, \quad \mathbf{u}_i^\star = [Q_i^\star \mathbf{V}_{1,n}^\star]^{-1/2} \mathbf{x}_i, \quad (100)$$

and, from (69), $\psi_n^\star(t) = \frac{1}{g_n^\star(t)} \frac{d g_n^\star(t)}{dt}$, where $g_n^\star(t)$ is a \sqrt{n} -consistent non-parametric estimator of the density generator $g_0 \in \mathcal{G}$.

3. An estimate $\mathbf{C}(\mathbf{V}_{1,n}^\star, g_n^\star)$ of the efficient SFIM $\bar{\mathbf{I}}(\theta_0|g_0)$ evaluated in (93). From its definition provided in (51) and from the previous results, we have that:

$$\mathbf{C}(\mathbf{V}_{1,n}^\star, g_n^\star) = a_n^\star \mathbf{K}_{\mathbf{V}_{1,n}^\star} \mathbf{K}_{\mathbf{V}_{1,n}^\star}^T, \quad (101)$$

where a_n^\star is a \sqrt{n} -consistent estimator of the term $a_2 = \frac{2\text{E}(Q^2 \psi_0^2(Q))}{m(m+2)}$ in (75).

By taking a careful look at the previous list of ingredients, it is immediate to realize that the critical point for the implementation of this semiparametric efficient estimator is the derivation of \sqrt{n} -consistent estimators for the density generator $g_n^\star(t)$ and for the two related quantities $\psi_n^\star(t)$ and a_n^\star . As discussed in sec. 2.4.1, finding a non-parametric estimator able to converge at the \sqrt{n} -rate is a difficult, or even impossible, task. What can we do then? Is there any other way out? Fortunately, a positive answer to this question has been recently found by Hallin, Oja and Paindaveine in their seminal work [13]. Even if of great importance, their results strongly rely on an invariance-based extension of the semiparametric Le Cam theory discussed in [18], whose explanation falls outside the scope of this paper. Here, we limit ourselves to a provides only some hints, without any claim of completeness nor of mathematical rigor. The groundbreaking idea proposed in [13] is the use of a *rank-based approximation* $\tilde{\Delta}_n(\mathbf{V}_{1,n}^\star)$ [12], [37, Ch. 13] (hence the name of *R-estimator*) of the efficient central sequence $\bar{\Delta}_n(\mathbf{V}_{1,n}^\star, g_n^\star)$ that does not require any non-parametric estimator g_n^\star of the density generator $g_0 \in \mathcal{G}$.

By letting the interested reader to discover the technical details and the proofs in [13, 10], in the following we will provide the necessary tools to implement an *R-estimator* of the shape matrix. At first, we need to introduce the concept of *ranks*. To this end, let us order the n ‘‘preliminary estimated’’ modular variates $\{Q_1^\star, Q_2^\star, \dots, Q_n^\star\}$ in (100) in an ascending way as $Q_{n(1)}^\star < Q_{n(2)}^\star < \dots < Q_{n(n)}^\star$. Then, the rank r_i^\star of Q_i^\star is its position index in the ordered sequence [37, Ch. 13]. Let us now introduce the following two rank-based functionals:

1. *The g_n^\star -free approximation of the efficient central sequence.* By using, as before, the Tyler M -estimator as the \sqrt{n} -consistent preliminary estimator of the constrained shape matrix in (95), i.e. $\mathbf{V}_{1,n}^\star$, the semiparametric efficient central sequence $\bar{\Delta}_n(\boldsymbol{\theta}_n^\star, g_n^\star)$ in (99) can be approximated as:

$$\tilde{\Delta}(\mathbf{V}_1^\star) = n^{-1/2} \mathbf{K}_{\mathbf{V}_1^\star} \sum_{i=1}^n K\left(\frac{r_i^\star}{n+1}\right) \text{vec}(\mathbf{u}_i^\star (\mathbf{u}_i^\star)^T), \quad (102)$$

where r_i^\star is the rank of Q_i^\star , already defined in (100) along with \mathbf{u}_i^\star . The function $K : (0, 1) \rightarrow \mathbb{R}^+$ belongs to the set \mathcal{K} of continuous, square integrable functions that can be expressed as the difference of two monotone increasing functions [13]. Here, due to its ‘‘optimality property’’ (see [30] for some additional discussion) we use the *van der Waerden* function:

$$K_{vdW}(t) \stackrel{\text{def}}{=} \Phi_G^{-1}(t), \quad (103)$$

where Φ_G^{-1} indicates the inverse cumulative distribution function of a χ^2 -distributed random variable with m degrees of freedom.

2. *The g_n^\star -free approximation of a_2 in (75).* As proved in [13], a \sqrt{n} -consistent, g_n^\star -free, estimator of $a_2 = \frac{2\mathbb{E}(Q^2 \psi_0^2(Q))}{m(m+2)}$ is given by:

$$\hat{\alpha}_n = \frac{\|\Delta(\mathbf{V}_1^\star + n^{-1/2} \mathbf{H}^0) - \tilde{\Delta}(\mathbf{V}_1^\star)\|}{\|\mathbf{K}_{\mathbf{V}_1^\star} \mathbf{K}_{\mathbf{V}_1^\star}^T \underline{v}(\mathbf{H}^0)\|}, \quad (104)$$

where \mathbf{H}^0 is a ‘‘small perturbation’’, symmetric, matrix such that $[\mathbf{H}^0]_{11} = 0$. Following [10], we set $\mathbf{H}^0 = (\mathbf{G} + \mathbf{G}^H)/2$ where $[\mathbf{G}]_{ij} \sim \mathcal{CN}(0, \nu^2)$, $[\mathbf{G}]_{11} = 0$ and ν is an hyper-parameter that has to be small enough to guarantee that $\hat{\mathbf{V}}_1^\star + n^{-1/2} \mathbf{H}^0$ remains a positive-definite matrix. Consequently, we have that a g_n^\star -free approximation of the efficient SFIM can be obtained as:

$$\tilde{\mathbf{C}}(\hat{\mathbf{V}}_{1,n}^\star) = \hat{\alpha}_n \mathbf{K}_{\mathbf{V}_{1,n}^\star} \mathbf{K}_{\mathbf{V}_{1,n}^\star}^T. \quad (105)$$

Finally, by putting all the previous results together, we have that a (almost) semiparametric efficient one-step R -estimator of the shape matrix is given by:

$$\begin{aligned} \underline{v}(\hat{\mathbf{V}}_{R,1,n}) &= \underline{v}(\mathbf{V}_{1,n}^\star) + n^{-1/2} [\tilde{\mathbf{C}}(\mathbf{V}_{1,n}^\star)]^{-1} \tilde{\Delta}(\mathbf{V}_1^\star) \\ &= \underline{v}(\mathbf{V}_{1,n}^\star) + \frac{1}{n \hat{\alpha}_n} \left[\mathbf{K}_{\mathbf{V}_{1,n}^\star} \mathbf{K}_{\mathbf{V}_{1,n}^\star}^T \right]^{-1} \mathbf{K}_{\mathbf{V}_1^\star} \sum_{i=1}^n K\left(\frac{r_i^\star}{n+1}\right) \text{vec}(\mathbf{u}_i^\star (\mathbf{u}_i^\star)^T). \end{aligned} \quad (106)$$

As shown in [13], $\underline{v}(\hat{\mathbf{V}}_{R,1,n})$ is a RAL estimator of $\mathbf{V}_{1,0}$ satisfying the following properties:

1. *\sqrt{n} -consistent:* $\sqrt{n}(\underline{v}(\hat{\mathbf{V}}_{R,1,n}) - \mathbf{V}_0) = O_{P_X}(1)$,

2. *Asymptotically normal*: $\sqrt{n}(\underline{v}(\widehat{\mathbf{V}}_{R,1,n} - \mathbf{V}_0)) \underset{n \rightarrow \infty}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{\Phi}(\mathbf{V}_0, g_0, K))$.

It is worth highlighting that its asymptotic error covariance matrix $\mathbf{\Phi}(\theta_0, g_0, K)$ depends on the choice of the rank-based function K along with the true shape matrix $\mathbf{V}_{1,0}$ and the true density generator $g_0 \in \mathcal{G}$. Moreover, due to the constrained parameter space in (96), induced by the specific choice of the constraint in (95), the efficiency inequality in (48) can be explicitly expressed as:

$$\mathbf{\Phi}(\mathbf{V}_0, g_0, K) \geq \text{CSCRB}(\mathbf{V}_0 | g_0) = \frac{m(m+2)}{2\mathbb{E}(\mathcal{Q}^2 \psi_0^2(\mathcal{Q}))} \left[\mathbf{K}_{\mathbf{V}_{1,0}} \mathbf{K}_{\mathbf{V}_{1,0}}^T \right]^{-1}. \quad (107)$$

Before moving to the simulation results, some comments are in order:

- The R -estimator in (106) has been generalized to the case of Complex Elliptically Symmetric distributed data in [10, 11, 9] where an investigation of its robustness properties has been also performed.
- Building upon the seminal results in [13, Prop. 3.1], in [11] the complex-valued R -estimator has been recast in a matrix form able to be more efficient in terms of numerical calculation.
- The Matlab and Python codes related to the implementation of the different versions of the R -estimator can be found in [4].
- If we want to constraint the estimated shape matrix with a criterion different from the one in (95), we can just re-normalize the estimated shape matrix as:

$$\widehat{\mathbf{V}}_{R,s,n} = \widehat{\mathbf{V}}_{R,1,n} / s(\widehat{\mathbf{V}}_{R,1,n}), \quad (108)$$

without any impact on the *asymptotic* properties of [15]. However, it must be noted that generally the non-asymptotic, *finite-sample* estimation performance may be impacted by this re-scaling.

3.3.1 Finite-sample performance of R -estimators

In this subsection, we compare, through numerical simulation, the estimation performance of the Tyler M -estimator with the one of the R -estimator in (106). The main aim is to quantify the efficiency losses described by the inequality (107) in a finite-sample regime, i.e. when the number of observations is finite. To this end, we generate a set of *zero-mean* i.i.d. t -distributed data whose distribution is described in sec. 5.2 of the Chapter 1. The scatter matrix is assumed to have a Toeplitz structure $[\mathbf{\Sigma}_0]_{ij} = \rho^{|i-j|}$, $i, j = 1, \dots, m$ where $\rho = 0.8$ and $m = 8$. The performance index, used to compare the Tyler and the R estimators is the following:

$$\varepsilon_\varphi \stackrel{\text{def}}{=} \left\| \mathbb{E}_0 \left(\underline{v}(\widehat{\mathbf{V}}_{\varphi,1,n} - \mathbf{V}_{1,0}) \underline{v}(\widehat{\mathbf{V}}_{\varphi,1,n} - \mathbf{V}_{1,0})^T \right) \right\|_F, \quad (109)$$

where $\varphi = \{\text{Tyler}, R\}$ indicates the particular estimator under test, $\mathbf{V}_{1,0} = \mathbf{\Sigma}_0 / [\mathbf{\Sigma}_0]_{11}$ and $\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}^T \mathbf{A})$ is the Frobenius norm of matrix \mathbf{A} . This Mean Square Error

(MSE) index will be compared with the CCRB in (76) and with the CSCRb in (107). Note that the scalars a_1 and a_2 (in (74) and (75) respectively), needed to define the bounds, can be evaluated, for the t -distribution, as [8]:

$$a_1 = -\frac{1}{2(m+2+\nu)}, \quad a_2 = \frac{2E(Q^2\psi_0^2(Q))}{m(m+2)} = \frac{\nu+m}{2(m+2+\nu)}. \quad (110)$$

Finally, for the CCRB in (76) and the (107), the following indices will be reported:

$$\varepsilon_{CCRB} = \|\text{CCRB}(\mathbf{V}_0)\|_F, \quad \varepsilon_{CSCRb} = \|\text{CCRB}(\mathbf{V}_0|g_0)\|_F. \quad (111)$$

We present the behavior of the MSE indices and of the bounds as function of the number of i.i.d. observations n in Fig. 1 and as function of the degrees of freedom ν of the t -distribution in Fig. 2. We recall, in passing, that when $\nu \rightarrow 0^+$, the data tend to be extremely heavy-tailed, while when $\nu \rightarrow +\infty$, the t -distribution collapse into the Gaussian one.

The first thing to note is that the distance between ε_{CCRB} and ε_{CSCRb} quantifies the efficiency loss due to the missing knowledge of the density generator (parametric vs. semiparametric estimation). This distance does not goes to zero as n goes to infinity and it increases as $\nu \rightarrow +\infty$, i.e. when the observation are less heavy-tailed. This result remains true for any RES distributions (not only for the t -distribution) [13, 15]. Let us now focus on the performance of the Tyler and R -estimators. Both of them are semiparametric in nature since they does not require the a priori knowledge of the density generator characterizing the data generating process. If we compare their MSE indices, i.e. ε_{Tyler} and ε_R , as function of n as in Fig. 1, we can notice a small performance improvement of the R -estimator with respect to the Tyler's one. The benefit of the R -estimator are more evident if we evaluate ε_{Tyler} and ε_R as function of the degrees of freedom ν (the non-Gaussianity parameter) as in Fig. 2. As expected, the constrained Tyler M -estimator has a MSE index that is constant with respect to ν . On the other hand, the R -estimator tends to be adaptive with respect to the changing value of ν . Unfortunately, it fails to be semiparametric efficient, i.e. it is not able to achieve the CSCRb in (107). There are three main reasons for this:

- As discussed in sec. 3.3, the R -estimator relies on the estimation of the term a_2 in (75). The estimator \hat{a}_n given in (104) is \sqrt{n} -consistent but not optimal in the MSE sense. Then, to improve the global performance of an R -estimator, we should look for a “better” estimator of a_2 in (75).
- As shown in (107), the asymptotic error covariance matrix $\Phi(\mathbf{V}_0, g_0, K)$ depends of the choice of the rank function K . A natural question is then: *would it be possible to find the optimal function K that minimize the error covariance under any density generator?* Even if a preliminary answer to this question has been provided in [30], this problem is still open.
- Regarding the Fig. 2, it must be noted that ε_R is evaluated in a finite-sample regime ($n = 3m$) and not in asymptotic conditions needed to guarantee the optimality of the semiparametric estimator in (106). Moreover, it is worth underling that in this

non-asymptotic regime, the constraint in (108) that we choose to put of the scatter matrix may have an impact on the MSE of the related shape matrix estimator.

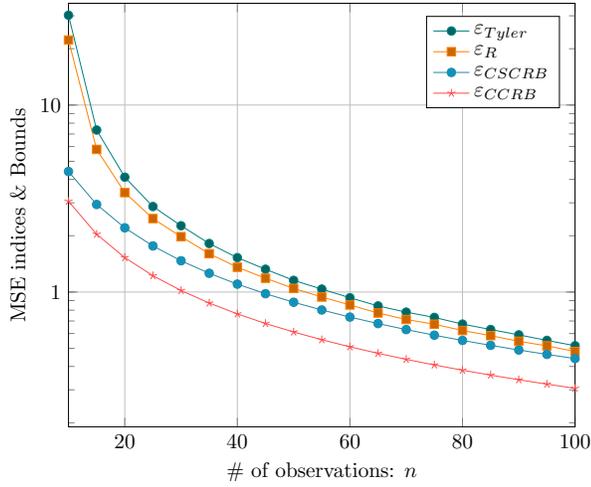


Fig. 1 MSE indices and bounds as function of the number of observations n ($\nu = 5, m = 8$).

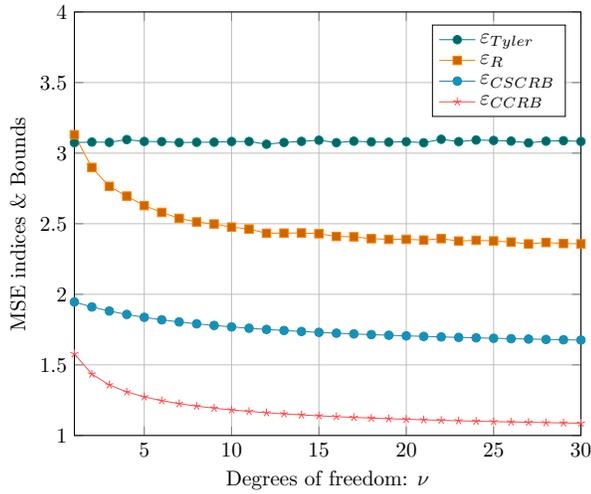


Fig. 2 MSE indices and bounds as function of the degrees of freedom ν of the t -distribution ($n = 3m, m = 8$).

4 Conclusions

The aim of this chapter was to firstly provide the reader with a general and tutorial in nature introduction about the estimation problem in semiparametric models. All the necessary mathematical tools have been presented along with some fundamental results as the *Semiparametric Efficiency Bound*. In the second part, we showed how to apply the general framework to the semiparametric estimation of the shape matrix of RES-distributed data in the presence of an unknown density generator. Specifically, the Constrained Semiparametric Cramér-Rao Bound (CSCRB) for the joint estimation of the location vector and of the shape matrix has been proposed together with a rank-based R -estimator of the shape matrix. Here we focused only on the RES case, but the interested reader can find the generalization of all the above mentioned results to CES-distributed data in [7, 10, 11, 9, 4]. We note in passing the semiparametric version of the celebrated Slepian-Bangs formula has also been derived in [7]. Despite the large amount of works already done in the field of the semiparametric statistics, some crucial problem still merit our effort to reach a fully satisfactory solution. Here we limit ourselves to cite the ones related to the RES distributions:

- As previously discussed, the derivation of a (at least asymptotically) semiparametric efficient estimator is still an open problem. Even if the R -estimator proposed by Hallin, Oja and Paidaveine in [13] has represented a breakthrough, we still have room for improvement as shown in Figs. 1 and 2.
- In some application, we are more interested in a direct estimation of the eigenspace of the shape matrix instead of the shape matrix itself. Would it be possible then to derive some semiparametric efficient estimators of the eigenvectors and eigenvalues of the shape matrix for elliptically distributed data? This problem has been recently addressed in [16, 17] but further works need to be done in order to have an exploitable practical estimator.

Acknowledgements I would like to warmly thank all those who have collaborated with me on this subject. Specifically, in chronological order, Prof. Fulvio Gini, Prof. Maria Sabrina Greco (University of Pisa), Prof. Abdelhak Zoubir (Technische Universität Darmstadt), Prof. Muralidhar Rangaswamy (U.S. AFRL, Sensors Directorate), Prof. Alexandre Renaux and Prof. Frédéric Pascal (Université Paris-Saclay, CNRS, CentraleSupélec, L2S).

Appendix: proof of Theorem 3

In this Appendix, we provide the proof of Theorem 3. Before starting the proof, it is worth recalling here that the Hilbert space \mathcal{H} is the set of scalar random functions on the sample space \mathcal{X} , i.e. $h : \mathcal{X} \rightarrow \mathbb{R}$ satisfying the following two properties:

1. *Zero-mean*: $E(h) = \int_{\mathcal{X}} h(\mathbf{x}) p_X(\mathbf{x}) d\mathbf{x} = 0$
2. *Finite variance*: $E(h^2) = \int_{\mathcal{X}} h^2(\mathbf{x}) p_X(\mathbf{x}) d\mathbf{x} = \sigma_h^2 < +\infty$

Proof: The proof follows the one discussed in [35, Theo. 4.4]. Let us start by defining the generic i -th parametric sub-model of the non-parametric model \mathcal{P}_g in (79) as

$$\mathcal{P}_{v_i} \stackrel{\text{def}}{=} \{p_X(\mathbf{x}|\boldsymbol{\eta}) = v_i(\|\mathbf{x}\|, \boldsymbol{\eta}), \boldsymbol{\eta} \in \Gamma_i \subseteq \mathbb{R}^{r_i}\}, \text{ where:} \quad (112)$$

where, according to the definition of parametric sub-model in sec. 2.4:

- $v_i : \mathcal{X} \times \Gamma_i \rightarrow \mathcal{G}$ is a smooth parametric map $\forall i \in \mathcal{I}$,
- Let g_0 be the true density generator, then $\exists \boldsymbol{\eta}_0 \in \Gamma_i$, s.t. $v_i(\|\mathbf{x}\|, \boldsymbol{\eta}_0) = g_0(\|\mathbf{x}\|^2)$.

According to the definition in (43), the i -th nuisance tangent space of \mathcal{P}_{v_i} in (112) is given by:

$$\mathcal{T}_{\boldsymbol{\eta}_{0,i}} \stackrel{\text{def}}{=} \{\mathbf{u} \in \mathcal{H}_q \mid \mathbf{u} = \mathbf{C}_i \mathbf{s}_{\boldsymbol{\eta}_{0,i}} : \mathbf{C}_i \text{ is any matrix in } \mathbb{R}^{q \times r_i}\} \subset \mathcal{H}_q \quad (113)$$

where the nuisance score vector is expressed as:

$$\mathbf{s}_{\boldsymbol{\eta}_{0,i}} \equiv \mathbf{s}_{\boldsymbol{\eta}_{0,i}}(\|\mathbf{x}\|) = \nabla_{\boldsymbol{\eta}} v_i(\|\mathbf{x}\|, \boldsymbol{\eta}_0). \quad (114)$$

From the standard properties of the score vectors, we have that :

$$\mathbb{E}_0([\mathbf{s}_{\boldsymbol{\eta}_{0,i}}]_j) = 0 \quad \mathbb{E}_0([\mathbf{s}_{\boldsymbol{\eta}_{0,i}}]_j^2) < +\infty \Rightarrow [\mathbf{s}_{\boldsymbol{\eta}_{0,i}}]_j \in \mathcal{L} \quad (115)$$

then $\mathbf{s}_{\boldsymbol{\eta}_{0,i}} \in \mathcal{S}_{g_0}$. Consequently:

$$\mathcal{T}_{\boldsymbol{\eta}_{0,i}} \subseteq \mathcal{S}_{g_0}, \quad \forall i \in \mathcal{I}, \quad (116)$$

that is, all the nuisance tangent spaces $\mathcal{T}_{\boldsymbol{\eta}_{0,i}}$ of any possible parametric sub-model \mathcal{P}_{v_i} are contained in \mathcal{S}_{g_0} .

To complete the proof, we now need to prove that any element in \mathcal{S}_{g_0} can be expressed as an element of some parametric submodel $\mathcal{T}_{\boldsymbol{\eta}_{0,i}}$. Let us start by choosing an arbitrary element $\mathbf{h}(\|\mathbf{x}\|)$ of \mathcal{S}_{g_0} , that is a zero-mean q -variate random function with finite variance. As parametric sub-model, we choose the following one:

$$\mathcal{P}_v = \{p_X(\mathbf{x}|\boldsymbol{\eta}) = v(\|\mathbf{x}\|, \boldsymbol{\eta}), \boldsymbol{\eta} \in \Gamma \subseteq \mathbb{R}^q\}, \text{ where:} \quad (117)$$

$$v(\|\mathbf{x}\|, \boldsymbol{\eta}) = g_0(\|\mathbf{x}\|^2) \left(1 + \boldsymbol{\eta}^T \mathbf{h}(\|\mathbf{x}\|)\right), \quad (118)$$

and $\boldsymbol{\eta} \in \Gamma \subseteq \mathbb{R}^q$ is sufficiently small to guarantee that $(1 + \boldsymbol{\eta}^T \mathbf{h}(\|\mathbf{x}\|)) \geq 0 \quad \forall \mathbf{x}$, and then $p_X(\mathbf{x}|\boldsymbol{\eta}) = v(\|\mathbf{x}\|, \boldsymbol{\eta}) \geq 0$. Moreover, from the definition of $\mathbf{h}(\|\mathbf{x}\|)$ as an element of \mathcal{S}_{g_0} , we have that:

$$\begin{aligned} \int p_X(\mathbf{x}|\boldsymbol{\eta}) d\mathbf{x} &= \int g_0(\|\mathbf{x}\|^2) d\mathbf{x} + \boldsymbol{\eta}^T \int \mathbf{h}(\|\mathbf{x}\|) g_0(\|\mathbf{x}\|^2) d\mathbf{x} \\ &= 1 + \boldsymbol{\eta}^T \mathbb{E}_0(\mathbf{h}(\|\mathbf{x}\|)) = 1 + 0 = 1, \end{aligned} \quad (119)$$

then $p_X(\mathbf{x}|\boldsymbol{\eta}) = v(\|\mathbf{x}\|, \boldsymbol{\eta})$ represents a proper pdf. For this specific parametric sub-model, we have that the nuisance score vector is given by:

$$\mathbf{s}_{\eta_0} = \nabla_{\eta} \nu(\|\mathbf{x}\|, \eta_0) = \nabla_{\eta} \left[g_0(\|\mathbf{x}\|^2) \left(1 + \boldsymbol{\eta}^T \mathbf{h}(\|\mathbf{x}\|) \right) \right] \Big|_{\boldsymbol{\eta}=\eta_0=\mathbf{0}} = \mathbf{h}(\|\mathbf{x}\|). \quad (120)$$

Finally, if we pose \mathbf{C}_i equal to the identity matrix \mathbf{I} in the nuisance tangent space (113) (also given in (43)), from (120), we immediately have that any arbitrary element $\mathbf{h}(\|\mathbf{x}\|)$ of \mathcal{S}_{g_0} can be expressed as an element of the nuisance tangent space of the parametric model in (117). This concludes the proof, i.e. \mathcal{S}_{g_0} is the semiparametric nuisance tangent space of the set of the spherically symmetric distributions.

References

1. Begun, J.M., Hall, W.J., Huang, W.M., Wellner, J.A.: Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics* **11**(2), 432–452 (1983)
2. Bickel, P., Klaassen, C., Ritov, Y., Wellner, J.: *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press (1993)
3. Bickel, P.J.: On adaptive estimation. *The Annals of Statistics* **10**(3), 647–671 (1982)
4. Fortunati, S.: <https://github.com/StefanoFor?tab=repositories> (2020)
5. Fortunati, S., Gini, F., Greco, M., Zoubir, A.M., Rangaswamy, M.: A fresh look at the semiparametric Cramér-Rao bound. In: 2018 26th European Signal Processing Conference (EUSIPCO), pp. 261–265 (2018)
6. Fortunati, S., Gini, F., Greco, M.S., Richmond, C.D.: Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications. *IEEE Signal Processing Magazine* **34**(6), 142–157 (2017)
7. Fortunati, S., Gini, F., Greco, M.S., Zoubir, A.M., Rangaswamy, M.: Semiparametric CRB and Slepian-Bangs formulas for complex elliptically symmetric distributions. *IEEE Transactions on Signal Processing* **67**(20), 5352–5364 (2019)
8. Fortunati, S., Gini, F., Greco, M.S., Zoubir, A.M., Rangaswamy, M.: Semiparametric inference and lower bounds for real elliptically symmetric distributions. *IEEE Transactions on Signal Processing* **67**(1), 164–177 (2019)
9. Fortunati, S., Renaux, A., Pascal, F.: Properties of a new R -estimator of shape matrices. *EUSIPCO 2020* (2020)
10. Fortunati, S., Renaux, A., Pascal, F.: Robust semiparametric efficient estimators in complex elliptically symmetric distributions. *IEEE Transactions on Signal Processing* **68**, 5003–5015 (2020)
11. Fortunati, S., Renaux, A., Pascal, F.: Robust semiparametric joint estimators of location and scatter in elliptical distributions. *Journal of Signal Processing Systems* **94**, 133–146 (2022)
12. Hájek, J.: Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39**(2), 325–346 (1968)
13. Hallin, M., Oja, H., Paindaveine, D.: Semiparametrically efficient rank-based inference for shape II. Optimal R -estimation of shape. *The Annals of Statistics* **34**(6), 2757–2789 (2006)
14. Hallin, M., Paindaveine, D.: Semiparametrically efficient rank-based inference for shape I. Optimal rank-based tests for sphericity. *The Annals of Statistics* **34**(6), 2707–2756 (2006)
15. Hallin, M., Paindaveine, D.: Parametric and semiparametric inference for shape: the role of the scale functional. *Statistics & Decisions* **24**(3), 327–350 (2009)
16. Hallin, M., Paindaveine, D., Verdebout, T.: Optimal rank-based testing for principal components. *The Annals of Statistics* **38**(6), 3245–3299 (2010)
17. Hallin, M., Paindaveine, D., Verdebout, T.: Efficient r -estimation of principal and common principal components. *Journal of the American Statistical Association* **109**(507), 1071–1083 (2014)
18. Hallin, M., Werker, B.J.M.: Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9**(1), 137–165 (2003)

19. Hettmansperger, T.P., Randles, R.H.: A practical affine equivariant multivariate median. *Biometrika* **89**(4), 851–860 (2002)
20. Jacod, J., Protter, P.: *Probability Essentials*. Springer series in statistics (2004)
21. Le Cam, L.: Locally asymptotically normal families of distributions. In: *Univ. California Publ. Statist.*, vol. 3, pp. 37–98 (1960)
22. LeCam, L., L.Yang, G.: *Asymptotics in Statistics: Some Basic Concepts*, 2 edn. Berlin, Germany: Springer series in statistics (2000)
23. Lehmann, E.L.: *Theory of Point Estimation*. John Wiley and Sons Inc., New York. (1983)
24. Lehmann, E.L., Romano, J.P.: *Testing Statistical Hypotheses*. Springer Texts in Statistics (2004)
25. Magnus, J.R., Neudecker, H.: The commutation matrix: Some properties and applications. *The Annals of Statistics* **7**(2), 381–394 (1979)
26. Magnus, J.R., Neudecker, H.: The elimination matrix: Some lemmas and applications. *SIAM Journal on Algebraic Discrete Methods* **1**(4), 422–449 (1980)
27. Magnus, J.R., Neudecker, H.: *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3 edn. (1999)
28. Moore, T.J., Kozick, R.J., Sadler, B.M.: The constrained Cramér–Rao bound from the perspective of fitting a model. *IEEE Signal Processing Letters* **14**(8), 564–567 (2007). DOI 10.1109/LSP.2006.891316
29. Newey, W.K.: Semiparametric efficiency bounds. *Journal of Applied Econometrics* **5**(2), 99–135 (1990)
30. Paindaveine, D.: A Chernoff-Savage result for shape: on the non-admissibility of pseudo-Gaussian methods. *Journal of Multivariate Analysis* **97**(10), 2206 – 2220 (2006)
31. Paindaveine, D.: A canonical definition of shape. *Statistics & Probability Letters* **78**(14), 2240 – 2247 (2008)
32. Petersen, K.B., Pedersen, M.S.: *The matrix cookbook* (2012). URL <http://www2.imm.dtu.dk/pubdb/p.php?3274>. Version 20121115
33. Rieder, H.: *Robust Asymptotic Statistics*. Springer series in statistics (1994)
34. Rudin, W.: *Real and Complex Analysis*, 1 edn. McGraw-Hill (1987)
35. Tsiatis, A.: *Semiparametric Theory and Missing Data*. Springer series in statistics (2006)
36. Tyler, D.E.: A distribution-free M -estimator of multivariate scatter. *The Annals of Statistics* **15**(1), 234–251 (1987)
37. van der Vaart, A.W.: *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press (1998)
38. Vermeulen, K.: *Semiparametric efficiency*. Ph.D. thesis, Universiteit Gent (2011)
39. Wellner, J.A.: Semiparametric models: Progress and problems. In: ISI (ed.) *Bulletin of the International Statistical Institute*, **4**, vol. 51 (1985)