A fresh look at the Semiparametric Cramér-Rao Bound

Stefano Fortunati[1,2], Fulvio Gini[1], Maria Greco[1]   Abdelhak M. Zoubir[2]   Muralidhar Rangaswamy[3]

[1]Dept. of Information Engineering   [2]Signal Processing Group   [3]U.S. AFRL, Sensors Directorate

University of Pisa   Technische Universität Darmstadt   Wright-Patterson AFB, OH, USA

{s.fortunati,f.gini,m.greco}@iet.unipi.it   zoubir@spg.tu-darmstadt.de   Muralidhar.Rangaswamy@us.af.mil

Abstract—This paper aims at providing a fresh look at semiparametric estimation theory and, in particular, at the Semiparametric Cramér-Rao Bound (SCRB). Semiparametric models are characterized by a finite-dimensional parameter vector of interest and by an infinite-dimensional nuisance function that is often related to an unspecified functional form of the density of the noise underlying the observations. We summarize the main motivations and the intuitive concepts about semiparametric models. Then we provide a new look at the classical estimation theory based on a geometrical Hilbert space-based approach. Finally, the semiparametric version of the Cramér-Rao Bound for the estimation of the finite-dimensional vector of the parameters of interest is provided.

I. Introduction

Any scientific experiment which aims to gain some knowledge about a real-word phenomenon starts with the data collection that will be later used to infer information. In statistical Signal Processing (SP) applications, all the available knowledge about the observed phenomenon is summarized in the probability distribution of the collected data. More formally, let $\mathbf{x}_1, \ldots, \mathbf{x}_M$ be M measurements collected from a random experiment, then $P_0(\mathbf{x}_1, \ldots, \mathbf{x}_M)$ is their (joint) "true" distribution that will be the basic ingredient of any inference method and, in particular of point estimation. In an estimation problem, in fact, we are interested in the evaluation of some functional of $P_0(\mathbf{x}_1, \ldots, \mathbf{x}_M)$, say $\nu(P_0)$. The mean value, the median, the covariance of the data are only some simple examples of $\nu(P_0)$. Clearly, any inference problem implies, as a first step, the estimation or, at least a reasonable guess, of the true data distribution $P_0(\mathbf{x}_1, \ldots, \mathbf{x}_M)$ or of the relevant probability density function (pdf) $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_M)$. To this end, we have to define a family of distributions (or pdfs) that are able to statistically characterize the collected observations. The set of possible distributions for a given random experiment is called *model*. According to the available amount of a-priori knowledge, two different classes of models can be exploited for data analysis and statistical inference: the *parametric* and the *non-parametric models*.

A parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ is defined as a set of pdfs for the acquired dataset $\mathbf{x}_1, \ldots, \mathbf{x}_M$ that are parametrized by a finite-dimensional parameter vector $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$:

$$\mathcal{P}_{\boldsymbol{\theta}} \triangleq \{p_X(\mathbf{x}_1, \ldots, \mathbf{x}_M | \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}. \qquad (1)$$

Clearly, if a parametric model is adopted, the knowledge about the real-word phenomenon of interest is summarized in the (finite-dimensional) parameter vector $\boldsymbol{\theta}$. In particular, if the true data pdf $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_M)$ belongs to $\mathcal{P}_{\boldsymbol{\theta}}$, this implies that there exists $\boldsymbol{\theta}_0 \in \Theta$ such that $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_M) = p_X(\mathbf{x}_1, \ldots, \mathbf{x}_M | \boldsymbol{\theta}_0)$ and then the model is said to be correctly specified. In a parametric model, every pdf is completely characterized by a finite number of parameters, i.e. the d entries of the vector $\boldsymbol{\theta}$, and this makes the subsequent inference procedure relatively simple even when only a small number of measurements is available. However, in some practical applications, and in particular when our a-priori knowledge about the experiment at hand is limited, a parametric model could result to be too restrictive and we could run the risk that the true data pdf $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_M)$ falls outside the assumed model $\mathcal{P}_{\boldsymbol{\theta}}$. If this happens, the parametric model is said to be *misspecified* ([1], [2]). In order to avoid the model misspecification (that, of course will lead to some performance degradation of the inference procedure), one can decide to characterize the statistical behavior of the collected data using a more general non-parametric model.

A non-parametric model is a collection of pdfs that possibly satisfy some functional constraints, i.e. symmetry around their mean value, and can be indicated as:

$$\mathcal{P}_p \triangleq \{p_X(\mathbf{x}_1, \ldots, \mathbf{x}_M) \in \mathcal{K}\}, \qquad (2)$$

where $\mathcal{K}$ is some constrained set of pdfs. Of course, using a non-parametric model, the risk of model misspecification is minimized since $\mathcal{P}_p$ is able to embrace a wider range of pdfs. On the other hand, the "expanse" of $\mathcal{P}_p$ could represent a big issue. In fact, to obtain an estimate of the true data pdf, we have to face an infinite-dimensional estimation problem. This could represent a prohibitive task in practical applications where the number of available data is limited and it usually results to be too small to estimate the full "shape" of $p_0(\mathbf{x}_1, \ldots, \mathbf{x}_M)$.

The concept of *semiparametric models* has been introduced to be a compromise between the "parsimony" of parametric models and the "realism" of the non-parametric ones [3]. A semiparametric model is a set of pdfs characterized by a finite-dimensional parameter $\boldsymbol{\theta} \in \Theta$ along with some infinite-dimensional parameter $l \in \mathcal{L}$, where $\mathcal{L}$ is some set of

functions:

$$\mathcal{P}_{\boldsymbol{\theta},l} \triangleq \{p_X(\mathbf{x}_1,\ldots,\mathbf{x}_M|\boldsymbol{\theta},l), \boldsymbol{\theta} \in \Theta, l \in \mathcal{L}\}. \quad (3)$$

In many practical applications involving semiparametric models, the parameters of interest are the ones collected in (a subvector of) the finite-dimensional vector $\boldsymbol{\theta}$, while the infinite-dimensional parameter l can be considered as a *nuisance* parameter, i.e. a parameter that interferes with the inference process but whose estimation is not required. Classical estimation theory has been fully developed in the case in which both the parameters of interest and the nuisance parameters can be collected in a finite dimensional vector (see e.g. [4], Sect. 10.7). In particular, we can easily derive a Cramér-Rao inequality for the covariance matrix of any unbiased estimator in the presence of finite-dimensional nuisance parameters. The issue we address in this paper is the following: is it possible to generalize the classical theories in order to take an infinite-dimensional nuisance function into account?

A huge amount of practical inference problems can be described using a semiparametric formalism (see [3], [5], [6] and references therein), but its potential has not been fully exploited by the SP community as yet. To the best of our knowledge, there are only few attempts to use a semiparametric approach in SP applications. Two examples are the works [7] and [8], where a semiparametric approach has been applied to blind source separation and to nonlinear regression, respectively. Our aim here, is to provide a fresh look at semiparametric estimation and at the SCRB, that can be exploited by a wide audience of SP practitioners.

The rest of the paper is organized as follows. In Sect. II, we show that one of the most widely used non-Gaussian model, i.e. the class of Real Elliptically Symmetric (RES) distributions ([9], [10]), is semiparametric in nature. In Sect. III, we discuss a geometrical reinterpretation of classical estimation theory and, in particular, of the CRB in the presence of a finite-dimensional nuisance vector. In Sect. IV, previously introduced geometrical tools will be used to generalize the parametric CRB to the semiparametric framework. Finally, a discussion on some open problems and on possible future research directions is provided in Sect. V.

II. A SEMIPARAMETRIC MODEL: THE RES DISTRIBUTIONS

Before providing some hints about the theory of semiparametric estimation and about the SCRB, let us discuss an example of a semiparametric model that will help us to clarify the basic ideas. We focus our attention on the class of RES distributions [9], [10]. This family of distributions has been recognized to be one of the more suitable and general model to statistically characterize the non-Gaussian, heavy-tailed, nature of the disturbance in many practical applications [11], [12].

A random vector $\mathbf{x} \in \mathbb{R}^N$ is said to be RES-distributed if it has a relevant pdf of the form:

$$p_X(\mathbf{x}) = 2^{-N/2}|\boldsymbol{\Sigma}|^{-1/2}g((\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})), \quad (4)$$

where $g : \mathbb{R}_0^+ \to \mathbb{R}^+$ is called *density generator*. The RES class encompasses the whole family of the (real) Compound

Gaussian (CG) distributions as a special case. Moreover, it contains non-Gaussian and non-CG distributions as well, as for example, the Generalized Gaussian (GG) distribution. Even if the RES class and its complex counterpart, the CES class, are two celebrated disturbance models, their intrinsic semiparametric structure has not been investigated as yet.

In almost all practical applications involving the RES model, we are generally interested in the estimation of the mean value $\boldsymbol{\mu}$ and/or of the scatter matrix $\boldsymbol{\Sigma}$ of the RES distributions, irrespective of the particular density generator g. Therefore, g can be considered as a nuisance function. It is immediate to verify that the RES family can be interpreted as a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\mu},\boldsymbol{\Sigma},g} \triangleq \{p_X(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},g), (\boldsymbol{\mu},\boldsymbol{\Sigma}) \in \mathbb{R}^N \times \mathcal{M}, g \in \mathcal{G}\}, \quad (5)$$

where $\mathcal{M}$ indicates the set of all the positive definite, symmetric $N \times N$ matrices and $\mathcal{G}$ indicates the set of the density generators. Due to the elliptical symmetry that characterizes the RES class, the relevant semiparametric model is a particular instance of a *semiparametric group model* (see [5] (Sect. 4.2)). Before providing a discussion on the mathematical tools needed to handle estimation problems in the presence of an (infinite-dimensional) nuisance function, we first sketch, in the next section, a new geometrical reading of the classical estimation theory in the presence of finite-dimensional nuisance parameters that will be the basis for its generalization to the semiparametric framework.

III. ESTIMATION IN PARAMETRIC MODELS: A GEOMETRICAL INTERPRETATION

We now introduce a geometrical interpretation of the finite-dimensional estimation theory that can be extended to semiparametric estimation. The classical parametric estimation theory can be reformulated using three main ingredients:

- the Hilbert space of all the zero-mean, vector-valued, random functions of the observation vectors,
- a notion of *tangent space* for a statistical model;
- an orthogonal projection operator.

Before introducing these three geometrical objects, it is worth pointing out that here we assume to deal with real random vectors and real parameters. The extension to the complex field falls outside the scope of this paper. Moreover, unless otherwise stated, we assume to have a single random observation vector $\mathbf{x}$ at our disposal. This is not a big limitation, since the extension to the case of a set of M *independent* and *identically distributed* (i.i.d.) observation vectors is straightforward. The generalization to the non i.i.d. case is discussed in [13]. Lastly, due to the lack of space, we will not provide any details about the formal definition of a Hilbert space, a linear subspace of a Hilbert space and of the projection operator. We refer the reader to the book [14] or to [5] (Sections A.1 and A.2).

Let us start from the parametric model defined in (1) where $\boldsymbol{\theta}$ is a d-dimensional parameter vector and $\mathbf{x}$ is an observation vector. We could be interested in estimating a subvector of $\boldsymbol{\theta}$. Then $\boldsymbol{\theta}$ can be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\eta}^T)^T \in \Gamma \times \Omega \subseteq$

$\mathbb{R}^q \times \mathbb{R}^r$ where the q-dimensional vector $\boldsymbol{\gamma}$ is the vector of the *parameters of interest* while the r-dimensional vector $\boldsymbol{\eta}$ is the vector of the *nuisance parameters*. Note that $q + r = d$. In the reminder of this section, we always indicate the *true* parameter vector as $\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0^T, \boldsymbol{\eta}_0^T)^T$ and the related *true* pdf as $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\theta}_0)$. Consequently, we have that $\mathbf{x} \sim p_0(\mathbf{x})$ where $\sim$ stands for "is distributed according to". Finally, $E_0\{\cdot\}$ indicates the expectation operator taken with respect to the true pdf $p_0(\mathbf{x})$.

A. The Hilbert space of the q-dimensional random functions

We introduce here a Hilbert space that plays a fundamental role in the following development. Consider the set $\mathcal{H}^q$ of all the q-dimensional vector-valued functions of the random vector $\mathbf{x} \sim p_0(\mathbf{x})$ such that:

$$\mathcal{H}^q = \{\mathbf{h}|\mathbf{h}(\mathbf{x}) \in \mathbb{R}^q, E_0\{\mathbf{h}\} = 0, E_0\{\mathbf{h}^T\mathbf{h}\} < \infty\}, \quad (6)$$

where, in the expectation operator, we dropped the dependence of $\mathbf{h}$ on $\mathbf{x}$ for notation simplicity. This convention will be adopted from here onwards. It is immediate to verify that $\mathcal{H}^q$ is a Hilbert space whose inner product is defined through the expectation operator, i.e. $\langle \mathbf{h}_1, \mathbf{h}_2 \rangle = E_0\{\mathbf{h}_1^T\mathbf{h}_2\}, \forall \mathbf{h}_1, \mathbf{h}_2 \in \mathcal{H}^q$ [15] (Sect. 2). Since $\mathbf{h}$ is a q-dimensional function of a random vector $\mathbf{x}$, we can define its $q \times q$ covariance matrix in the usual way as:

$$\mathbf{C}_0(\mathbf{h}) \triangleq E_0\{\mathbf{h}\mathbf{h}^T\} \in \mathbb{R}^{q \times q}. \quad (7)$$

Let us investigate the geometrical structure of $\mathcal{H}^q$. Specifically, we focus on the derivation of an explicit expression of the orthogonal projection of a generic element $\mathbf{h}$ into a finite-dimensional subspace of $\mathcal{H}^q$. Following [15] (Sect. 2.4), let us define $\mathbf{v} = [v_1, \cdots, v_k]^T$ as a column vector of k arbitrary elements of $\mathcal{H}^1$ that is the Hilbert space of functions obtained from (6) by choosing $q = 1$. As a finite-dimensional subspace of $\mathcal{H}^q$, we consider the linear span of vector $\mathbf{v}$ defined as:

$$\mathcal{U} \triangleq \{\mathbf{A}\mathbf{v} : \mathbf{A} \text{ is any matrix in } \mathbb{R}^{q \times k}\}. \quad (8)$$

We want to find the orthogonal projection of an arbitrary element $\mathbf{h} \in \mathcal{H}^q$ onto $\mathcal{U}$, i.e. $\Pi(\mathbf{h}|\mathcal{U})$. From the Projection Theorem (see e.g. [5] (Sect. A.2)), we know that this projection exists, it is unique and it can be explicitly written as:

$$\Pi(\mathbf{h}|\mathcal{U}) = E_0\{\mathbf{h}\mathbf{v}^T\}\mathbf{C}_0^{-1}(\mathbf{v})\mathbf{v}. \quad (9)$$

For the sake of clarity, it is worth recalling that $\Pi(\mathbf{h}|\mathcal{U})$, $\mathbf{h}$ and $\mathbf{v}$ are all vector-valued functions of the random observation vector $\mathbf{x}$. As we will show in Subsect. III-C, this explicit expression of the projection operator is the key for a geometrical description of the Cramér-Rao inequality in the presence of a finite-dimensional nuisance parameter vector.

B. The parametric nuisance tangent space

In order to define the parametric nuisance tangent space, we first need to recall the notion of *score vector*. Let $\mathcal{P}_{\boldsymbol{\theta}}$ be a parametric model as in (1) and let $\mathbf{x} \sim p_0(\mathbf{x})$. Then, the score vector for $\mathbf{x}$ in $\boldsymbol{\theta}_0$, indicated as $\mathbf{s}_{\boldsymbol{\theta}_0} \equiv \mathbf{s}_{\boldsymbol{\theta}_0}(\mathbf{x})$, is the

d-dimensional vector-valued function whose entries, for $i = 1, \ldots, d$, are defined as:

$$[\mathbf{s}_{\boldsymbol{\theta}_0}]_i \triangleq [\nabla_{\boldsymbol{\theta}} \ln p_X(\mathbf{x}|\boldsymbol{\theta}_0)]_i = \left.\frac{\partial \ln p_X(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \quad (10)$$

Note that, if $\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0^T, \boldsymbol{\eta}_0^T)^T$, the score vector in (10) can be partitioned as $\mathbf{s}_{\boldsymbol{\theta}_0} = (\mathbf{s}_{\boldsymbol{\gamma}_0}^T, \mathbf{s}_{\boldsymbol{\eta}_0}^T)^T$. Under the usual regularity conditions that allow for the order inversion between integral and derivative operators, the score vector is a zero-mean random vector, i.e. $E_0\{\mathbf{s}_{\boldsymbol{\theta}_0}\} = \mathbf{0}$ and each of its entries has finite variance, i.e. $E_0\{[\mathbf{s}_{\boldsymbol{\theta}_0}]_i^2\} < \infty$. Consequently, each entry of $\mathbf{s}_{\boldsymbol{\theta}_0}$ belongs to $\mathcal{H}^1$ obtained from (6) with $q = 1$. Then, using the procedure discussed in Subsection III-A, we can define a finite-dimensional subspace of $\mathcal{H}^q$ (see (8)) as the linear span generated by the entries of the nuisance score vector $\mathbf{s}_{\boldsymbol{\eta}_0}$. In particular, we define the *nuisance tangent space* as the finite-dimensional subspace of $\mathcal{H}^q$ spanned by the entries of $\mathbf{s}_{\boldsymbol{\eta}_0}$, i.e.

$$\mathcal{T}_{\boldsymbol{\eta}_0} \triangleq \{\mathbf{C}\mathbf{s}_{\boldsymbol{\eta}_0} : \mathbf{C} \text{ is any matrix in } \mathbb{R}^{q \times r}\}, \quad (11)$$

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

C. The efficient score vector and the Cramér-Rao Bound

As before, let $\mathcal{P}_{\boldsymbol{\theta}}$ be a parametric model as in (1) and let $\mathbf{x} \sim p_0(\mathbf{x})$ be the random observation vector, then the Fisher Information Matrix (FIM) for $\boldsymbol{\theta}_0$ is defined as follows:

$$\mathbf{I}(\boldsymbol{\theta}_0) = E_0\{\mathbf{s}_{\boldsymbol{\theta}_0}\mathbf{s}_{\boldsymbol{\theta}_0}^T\} = \begin{pmatrix} \mathbf{C}_0(\mathbf{s}_{\boldsymbol{\gamma}_0}) & \mathbf{I}_{\boldsymbol{\gamma}_0\boldsymbol{\eta}_0} \\ \mathbf{I}_{\boldsymbol{\gamma}_0\boldsymbol{\eta}_0}^T & \mathbf{C}_0(\mathbf{s}_{\boldsymbol{\eta}_0}) \end{pmatrix} \quad (12)$$

where $\mathbf{I}_{\boldsymbol{\gamma}_0\boldsymbol{\eta}_0} \triangleq E_{p_0}\{\mathbf{s}_{\boldsymbol{\gamma}_0}\mathbf{s}_{\boldsymbol{\eta}_0}^T\}$ (see e.g. [4]). Moreover, let $\hat{\boldsymbol{\gamma}}(\mathbf{x})$ be an *unbiased* estimator of the vector of the parameter of interest $\boldsymbol{\gamma}_0$ in the presence of the nuisance vector $\boldsymbol{\eta}_0$. Then, the Cramér-Rao inequality on the error covariance matrix of $\hat{\boldsymbol{\gamma}}(\mathbf{x})$ can be easily established as:

$$\begin{aligned} E_0\{(\hat{\boldsymbol{\gamma}}(\mathbf{x}) - \boldsymbol{\gamma}_0)(\hat{\boldsymbol{\gamma}}(\mathbf{x}) - \boldsymbol{\gamma}_0)^T\} &\geq \text{CRB}(\boldsymbol{\gamma}_0|\boldsymbol{\eta}_0) \\ &\triangleq \left(\mathbf{C}_0(\mathbf{s}_{\boldsymbol{\gamma}_0}) - \mathbf{I}_{\boldsymbol{\gamma}_0\boldsymbol{\eta}_0}\mathbf{C}_0^{-1}(\mathbf{s}_{\boldsymbol{\eta}_0})\mathbf{I}_{\boldsymbol{\gamma}_0\boldsymbol{\eta}_0}^T\right)^{-1}. \end{aligned} \quad (13)$$

In particular, $\text{CRB}(\boldsymbol{\gamma}_0|\boldsymbol{\eta}_0)$ can be obtained as the top-left sub-matrix of the inverse of the FIM $\mathbf{I}(\boldsymbol{\theta}_0)$ in (12) whose explicit expression, given in (13), follows directly from the application of the Matrix Inversion Lemma. Interestingly enough, this result can also be obtained using the geometrical approach discussed in Subsects. III-A and III-B. To prove this, we firstly introduce the notion of *efficient score vector* ([5] (Sect. 2) and [15] (Sect. 3.4)). The efficient score vector $\mathbf{s}_0^\star \equiv \mathbf{s}_0^\star(\mathbf{x})$ is defined as the residual of the score vector of the parameters of interest $\mathbf{s}_{\boldsymbol{\gamma}_0}$ after projecting it onto the nuisance tangent space $\mathcal{T}_{\boldsymbol{\eta}_0}$ defined in (11):

$$\begin{aligned} \mathbf{s}_0^\star &\triangleq \mathbf{s}_{\boldsymbol{\gamma}_0} - \Pi(\mathbf{s}_{\boldsymbol{\gamma}_0}|\mathcal{T}_{\boldsymbol{\eta}_0}) \\ &= \mathbf{s}_{\boldsymbol{\gamma}_0} - E_0\{\mathbf{s}_{\boldsymbol{\gamma}_0}\mathbf{s}_{\boldsymbol{\eta}_0}^T\}\mathbf{C}_0^{-1}(\boldsymbol{\eta}_0)\mathbf{s}_{\boldsymbol{\eta}_0}, \end{aligned} \quad (14)$$

where, for the last equality, we used the explicit projection formula derived in (9). Roughly speaking, the efficient score vector can be used to quantify the amount of information carried by the true data pdf, $p_X(\mathbf{x}|\boldsymbol{\theta}_0)$, about the vector of

parameters of interest $\boldsymbol{\gamma}_0$ when the nuisance vector $\boldsymbol{\eta}_0$ is unknown. This information can be summarized in the *efficient FIM* defined as:

$$\begin{aligned} \mathbf{I}^\star(\boldsymbol{\theta}_0) &= \mathbf{C}_0(\mathbf{s}_0^\star) = E_0\{\mathbf{s}_0^\star(\mathbf{s}_0^\star)^T\} \\ &= \mathbf{C}_0(\mathbf{s}_{\boldsymbol{\gamma}_0}) - \mathbf{I}_{\boldsymbol{\gamma}_0\boldsymbol{\eta}_0}\mathbf{C}_0^{-1}(\mathbf{s}_{\boldsymbol{\eta}_0})\mathbf{I}_{\boldsymbol{\gamma}_0\boldsymbol{\eta}_0}^T, \end{aligned} \quad (15)$$

where the last equality follows by substituting the expression of the efficient score vector, given in (14), in the expectation operator in (15). It is immediate to verify that the inverse of the efficient FIM is equal to the CRB on the estimation of $\boldsymbol{\gamma}_0$ in the presence of the nuisance vector $\boldsymbol{\eta}_0$ given in (13), i.e.

$$\mathrm{CRB}(\boldsymbol{\gamma}_0|\boldsymbol{\eta}_0) = (\mathbf{I}^\star(\boldsymbol{\theta}_0))^{-1}. \quad (16)$$

Equality in (16) provides the link between the classical approach to estimation theory and the geometrical one, introduced in Subsects. III-A and III-B. More important, it shows us that, in order to derive a CRB for estimation problems in the presence of nuisance parameters, we only need two geometrical objects: the nuisance tangent space $\mathcal{T}_{\boldsymbol{\eta}_0}$ and a projection operator on $\mathcal{T}_{\boldsymbol{\eta}_0}$, i.e. $\Pi(\cdot|\mathcal{T}_{\boldsymbol{\eta}_0})$. Remarkably, none of the previous geometrical objects requires the finite-dimensionality of the nuisance parameters, hence they can be readily applied to the semiparametric framework.

IV. EXTENSION TO SEMIPARAMETRIC MODELS

The geometrical framework previously introduced can be extended to the semiparametric case. In order to maintain the notation as consistent as possible with the one used in Sect. III, we define as *semiparametric model* the set of densities:

$$\mathcal{P}_{\boldsymbol{\gamma},l} \triangleq \{p_X(\mathbf{x}|\boldsymbol{\gamma},l), \boldsymbol{\gamma} \in \Gamma \subseteq \mathbb{R}^q, l \in \mathcal{L}\}, \quad (17)$$

where, as before, $\boldsymbol{\gamma}$ is a q-dimensional vector of the parameters of interest, while l is a nuisance function belonging to some set $\mathcal{L}$. We denote the true "semiparametric vector" as $(\boldsymbol{\gamma}_0^T, l_0)^T \in \Gamma \times \mathcal{L}$, and consequently the true pdf is $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\gamma}_0, l_0)$. Clearly, if we set the unknown nuisance function $l \in \mathcal{L}$ to be the true one, i.e. l_0, the model $\mathcal{P}_{\boldsymbol{\gamma},l_0} \equiv \mathcal{P}_{\boldsymbol{\gamma}}$ can be considered as a parametric model, where each pdf is indexed by the finite-dimensional parameter vector $\boldsymbol{\gamma} \in \Gamma$. As discussed in Sect. I, we cannot directly apply inference methods developed in the parametric framework to the semiparametric one because of the non-parametric, infinite-dimensional nature of the nuisance function $l \in \mathcal{L}$. This dimensionality problem can be overcome by introducing the concept of *parametric submodel* of a semiparametric model ([5] (Sect. 3.1, Def. 1) and [15] (Sect. 4.2)). Specifically, the *i-th* parametric submodel of the semiparametric model $\mathcal{P}_{\boldsymbol{\gamma},l}$, denoted as:

$$\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})} = \{p_X(\mathbf{x}|\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})), \boldsymbol{\gamma} \in \Gamma, \boldsymbol{\eta} \in \Omega_i\}, \quad (18)$$

is defined as a class of parametric pdfs indexed by a finite-dimensional parameter vector $(\boldsymbol{\gamma}^T, \boldsymbol{\eta}^T)^T \in \Gamma \times \Omega_i \subseteq \mathbb{R}^q \times \mathbb{R}^{r_i}$, such that, for every $i \in I$:

C0) $\nu_i(\boldsymbol{\eta}) : \Omega_i \to \mathcal{L}$ is a smooth parametric map,
C1) $\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})} \subseteq \mathcal{P}_{\boldsymbol{\gamma},l}$,
C2) $p_0(\mathbf{x}) \in \mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}$, i.e. there exists a vector $(\boldsymbol{\gamma}_0^T, \boldsymbol{\eta}_0^T)^T$ such that $p_X(\mathbf{x}|\boldsymbol{\gamma}_0,\nu_i(\boldsymbol{\eta}_0)) = p_X(\mathbf{x}|\boldsymbol{\gamma}_0,l_0) \triangleq p_0(\mathbf{x})$.

Condition C1 tells us that all the pdfs that compose each possible parametric submodel $\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}$ have to belong to the semiparametric model $\mathcal{P}_{\boldsymbol{\gamma},l}$ as well. Moreover, Condition C2 highlights the fact that each parametric submodel $\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}$ must contain the true pdf $p_0(\mathbf{x})$. Roughly speaking, by using a parametric submodel $\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}$ in place of $\mathcal{P}_{\boldsymbol{\gamma},l}$ we are actually identifying the infinite-dimensional parameter $l \in \mathcal{L}$ with the finite-dimensional nuisance parameter vector $\boldsymbol{\eta} \in \Omega_i \subseteq \mathbb{R}^{r_i}$ whose dimension r_i depends on the choice of the particular parametric submodel. The way to generalize the theory developed for parametric models to semiparametric models should now be clear. The idea is to exploit the finite-dimensional statistical results in the (artificial) set of parametric submodels $\{\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}\}_{i \in I}$ and then "take the limit" to generalize them to the infinite-dimensional semiparametric framework.

A. The semiparametric nuisance tangent space

Let us now define a key element of the semiparametric theory, i.e. the semiparametric nuisance tangent space $\mathcal{T}_{l_0}$, according to the definition given in [6] and [15] (Sect. 4.4). Note that a more general (but more abstract) definition is given in [5] (Sect. 3.2) and in [16]. At first, let us recall that the Hilbert space $\mathcal{H}^q$ in (6) is a metric space with squared distance given by $||\mathbf{h}_1 - \mathbf{h}_2||^2 = E_0\{(\mathbf{h}_1 - \mathbf{h}_2)^T(\mathbf{h}_1 - \mathbf{h}_2)\}$. The semiparametric nuisance tangent space $\mathcal{T}_{l_0}$ of the semiparametric model $\mathcal{P}_{\boldsymbol{\gamma},l}$ is defined as the closure of the union of all nuisance tangent spaces

$$\mathcal{T}_{\boldsymbol{\eta}_{0,i}} \triangleq \{\mathbf{C}_i\mathbf{s}_{\boldsymbol{\eta}_{0,i}} : \mathbf{C}_i \text{ is any matrix in } \mathbb{R}^{q \times r_i}\} \quad (19)$$

of the parametric submodels $\{\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}\}_{i \in I} \subseteq \mathcal{P}_{\boldsymbol{\gamma},l}$. Specifically, $\mathcal{T}_{l_0} \subseteq \mathcal{H}^q$ is the subspace of all q-dimensional, zero-mean, vector-valued, random functions $\mathbf{t} \in \mathcal{H}^q$ for which there exists a sequence $\{\mathbf{C}_i\mathbf{s}_{\boldsymbol{\eta}_{0,i}}\}_{i \in I}$ such that:

$$||\mathbf{t} - \mathbf{C}_i\mathbf{s}_{\boldsymbol{\eta}_{0,i}}|| \underset{i \in I}{\longrightarrow} 0, \quad (20)$$

where $\mathbf{s}_{\boldsymbol{\eta}_{0,i}}$ is the nuisance score vector of the parametric submodel $\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}$ and the matrices $\mathbf{C}_i$ have appropriate dimensions, i.e. $\mathbf{C}_i \in \mathbb{R}^{q \times r_i}$ when the nuisance parameter vector $\boldsymbol{\eta}$ belongs to $\mathbb{R}^{r_i}$. Using this definition, the semiparametric nuisance tangent space can be simply set up as:

$$\mathcal{T}_{l_0} = \overline{\bigcup_{\{\mathcal{P}_{\boldsymbol{\gamma},\nu_i(\boldsymbol{\eta})}\}_{i \in I}} \mathcal{T}_{\boldsymbol{\eta}_{0,i}}}. \quad (21)$$

Note that 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 $\mathcal{A}$. In our case, the limit points are those defined by the convergence points of all the sequences in (20). The semiparametric nuisance tangent space $\mathcal{T}_{l_0}$ is assumed to be a closed and linear subspace of the Hilbert space $\mathcal{H}^q$ (see [16] (Assumption S), [15] (Sect. 4.4, Remark 5)). Then, the existence and the uniqueness of the orthogonal projection operator onto $\mathcal{T}_{l_0}$, i.e. $\Pi(\cdot|\mathcal{T}_{l_0})$, is guaranteed by the Projection Theorem.

B. The Semiparametric Cramér-Rao Bound

The projection operator $\Pi(\cdot|\mathcal{T}_{l_0})$ leads us to the definition of the semiparametric counterpart of the CRB, reported in Theorem 1.

Theorem 1: The Semiparametric Cramér-Rao Bound (SCRB) for the estimation of the finite-dimensional vector $\boldsymbol{\gamma}_0$ in the presence of the nuisance function $l_0 \in \mathcal{L}$ is given by:

$$\mathrm{SCRB}(\boldsymbol{\gamma}_0|l_0) \triangleq \sup_{\{\mathcal{P}_{\boldsymbol{\gamma},\boldsymbol{\eta}_i}\}_{i \in I}} \mathbf{C}_0^{-1}(\mathbf{s}_{0,i}^{\star}) = [\bar{\mathbf{I}}(\boldsymbol{\gamma}_0|l_0)]^{-1}, \quad (22)$$

where $\mathbf{s}_{0,i}^{\star} = \mathbf{s}_{\boldsymbol{\gamma}_0} - \Pi(\mathbf{s}_{\boldsymbol{\gamma}_0}|\mathcal{T}_{\boldsymbol{\eta}_{0,i}})$ (see (14)) is the efficient score vector for the *i-th* parametric submodel $\mathcal{P}_{\boldsymbol{\gamma},\boldsymbol{\eta}_i}$ while

$$\bar{\mathbf{I}}(\boldsymbol{\gamma}_0|l_0) \triangleq E_0\{\bar{\mathbf{s}}_0(\bar{\mathbf{s}}_0)^T\}, \quad (23)$$

is the *semiparametric efficient FIM* and $\bar{\mathbf{s}}_0 \equiv \bar{\mathbf{s}}_0(\mathbf{x})$ is the *semiparametric efficient score vector* defined as:

$$\bar{\mathbf{s}}_0 = \mathbf{s}_{\boldsymbol{\gamma}_0} - \Pi(\mathbf{s}_{\boldsymbol{\gamma}_0}|\mathcal{T}_{l_0}), \quad (24)$$

where $\Pi(\mathbf{s}_{\boldsymbol{\gamma}_0}|\mathcal{T}_{l_0})$ is the orthogonal projection of $\mathbf{s}_{\boldsymbol{\gamma}_0}$ on the semiparametric nuisance tangent space. Note that $\bar{\mathbf{s}}_0$, $\mathbf{s}_{\boldsymbol{\gamma}_0}$ and $\Pi(\mathbf{s}_{\boldsymbol{\gamma}_0}|\mathcal{T}_{l_0})$ are, in general, q-dimensional vector-valued random functions of the observation vector $\mathbf{x}$.

This theorem can be found in [15] (Theorem 4.1) and in [6]. A more abstract and general formulation can be found in [16] and in [5] (Section 3.4).

Two comments are in order. First of all, it is immediate to verify that the expression of the SCRB in (22) and of the semiparametric efficient score function in (24) are formally equivalent to the ones introduced in Sect. III-C in the case of finite-dimensional nuisance parameters. The only difference is in the definition of the nuisance tangent space. This confirms the intuition that, from and abstract and geometrical stand-point, the parametric and the semiparametric frameworks are equivalent. Secondly, from (22), it is clear that $\mathrm{SCRB}(\boldsymbol{\gamma}_0|l_0)$ is higher that any $\mathrm{CRB}(\boldsymbol{\gamma}_0|\boldsymbol{\eta}_{0,i}) = \mathbf{C}_0^{-1}(\mathbf{s}_{0,i}^{\star})$ derived for the *i-th* parametric submodel. In words, this means that a semiparametric model contains less information on the parameter vector of interest $\boldsymbol{\gamma}_0$ than any of its possible parametric submodel.

The SCRB is of great practical usefulness since it provides a lower bound to the error covariance of any *robust* estimator of the finite-dimensional parameter vector $\boldsymbol{\gamma}_0$, i.e. of any estimator of $\boldsymbol{\gamma}_0$ that does not rely on the a-priori knowledge of the nuisance function l_0. For example, the SCRB in (22) can be used to obtain a lower bound for *any* robust estimator of the scatter matrix $\boldsymbol{\Sigma}$ of a set of RES-distributed observation vectors ([17], [18]).

V. CONCLUSION

The aim of this paper is to provide a fresh look at semiparametric estimation, and in particular at the Semiparametric Cramér-Rao Bound (SCRB), that can be usable by a wide audience of SP practitioners. In particular, the possible applications and the potential advantages, that we may have in addressing classical problems by using this approach, have been discussed. Of course, a huge amount of work still remain to be done. Among the numerous open issues, the biggest challenge

for the application of the semiparametric theory to practical inference problems is the calculation of the projection operator $\Pi(\cdot|\mathcal{T}_{l_0})$. The monograph [5] provided us with many examples of this calculation for different semiparametric models. A different and more general (since it can be applied also in non i.i.d. cases) approach is discussed in [13]. Our current effort is devoted to the investigation of semiparametric inference methods in the context of RES and CES distributions. In particular, we are looking for a closed form expression of the SCRB for the estimation of the mean vector and the scatter matrix to assess the performance of various robust M-estimators.

REFERENCES

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

[2] S. Fortunati, M. S. Greco, and F. Gini, "Asymptotic robustness of Kelly's GLRT and Adaptive Matched Filter detector under model misspecification," in *ISI World Statistics Congress 2017 (ISI2017)*, 2017.

[3] J. A. Wellner, "Semiparametric models: Progress and problems," in *Bulletin of the International Statistical Institute*, ser. 4, vol. 51, 1985.

[4] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Elsevier, 1993.

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

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

[7] S.-I. Amari and J. F. Cardoso, "Blind source separation-semiparametric statistical approach," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2692–2700, Nov 1997.

[8] U. Hammes, E. Wolsztynski, and A. M. Zoubir, "Transformation-based robust semiparametric estimation," *IEEE Signal Processing Letters*, vol. 15, pp. 845–848, 2008.

[9] C. D. Richmond, "Adaptive array signal processing and performance analysis in non-Gaussian environments," Ph.D. dissertation, Massachusetts Institute of Technology, 1996.

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

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

[12] S. Fortunati, F. Gini, and M. S. Greco, "Matched, mismatched, and robust scatter matrix estimation and hypothesis testing in complex t-distributed data," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 123, 2016.

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

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

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

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

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

[18] D. Paindaveine, "A canonical definition of shape," *Statistics & Probability Letters*, vol. 78, no. 14, pp. 2240 – 2247, 2008.