The Cramér-Rao Inequality on Singular Statistical Models

Hông Vân Lê[1(✉)], Jürgen Jost[2], and Lorenz Schwachhöfer[3]

[1] Mathematical Institute of ASCR, Zitna 25, 11567 Praha, Czech Republic
hvle@math.cas.cz
[2] Max-Planck-Institut Für Mathematik in den Naturwissenschaften, Inselstrasse 22, 04103 Leipzig, Germany
jost@mis.mpg.de
[3] Technische Universität Dortmund, Vogelpothsweg 87, 44221 Dortmund, Germany
Lorenz.Schwachhoefer@math.uni-dortmund.de

Abstract. We introduce the notions of essential tangent space and reduced Fisher metric and extend the classical Cramér-Rao inequality to 2-integrable (possibly singular) statistical models for general φ-estimators, where φ is a V-valued feature function and V is a topological vector space. We show the existence of a φ-efficient estimator on strictly singular statistical models associated with a finite sample space and on a class of infinite dimensional exponential models that have been discovered by Fukumizu. We conclude that our general Cramér-Rao inequality is optimal.

1 *k*-integrable Parametrized Measure Models and the Reduced Fisher Metric

In this section we recall the notion of a *k-integrable parametrized measure model* (Definitions 1, 3). Then we give a characterization of k-integrability (Theorem 1), which is important for later deriving the classical Cramér-Rao inequalities from our general Cramér-Rao inequality. Finally we introduce the notion of *essential tangent space* of a 2-integrable parametrized measure model (Definition 4) and the related notion of *reduced Fisher metric*.

Notations. For a measurable space Ω and a finite measure μ_0 on Ω we denote

$$\begin{aligned}
\mathcal{P}(\Omega) &:= \{\mu \ : \ \mu \text{ a probability measure on } \Omega\},\\
\mathcal{M}(\Omega) &:= \{\mu \ : \ \mu \text{ a finite measure on } \Omega\},\\
\mathcal{S}(\Omega) &:= \{\mu \ : \ \mu \text{ a signed finite measure on } \Omega\},\\
\mathcal{S}(\Omega, \mu_0) &= \{\mu = \phi\, \mu_0 \ : \ \phi \in L^1(\Omega, \mu_0)\}.
\end{aligned}$$

Hông Vân Lê—Speaker, partially supported by RVO: 6798584.

F. Nielsen and F. Barbaresco (Eds.): GSI 2017, LNCS 10589, pp. 552–560, 2017.
https://doi.org/10.1007/978-3-319-68445-1_64

Definition 1 ([AJLS2016b, Definition 4.1]). *Let Ω be a measurable space.*

1. *A* parametrized measure model *is a triple* $(M, \Omega, \mathbf{p})$ *where* M *is a (finite or infinite dimensional) Banach manifold and* $\mathbf{p} : M \to \mathcal{M}(\Omega) \subset \mathcal{S}(\Omega)$ *is a Frechét-*C^1*-map, which we shall call simply a* C^1*-map.*
2. *The triple* $(M, \Omega, \mathbf{p})$ *is called a* statistical model *if it consists only of probability measures, i.e., such that the image of* $\mathbf{p}$ *is contained in* $\mathcal{P}(\Omega)$*.*
3. *We call such a model* dominated by μ_0 *if the image of* $\mathbf{p}$ *is contained in* $\mathcal{S}(\Omega, \mu_0)$*. In this case, we use the notation* $(M, \Omega, \mu_0, \mathbf{p})$ *for this model.*

Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model. It follows from [AJLS2016b, Proposition 2.1] that for all $\xi \in M$ the differential $d_\xi \mathbf{p}(V)$ is dominated by $\mathbf{p}(\xi)$. Hence *the logarithmic derivative of* $\mathbf{p}$ *at* ξ *in direction* V [AJLS2016b, (4.2)]

$$\partial_V \log \mathbf{p}(\xi) := \frac{d\{d_\xi \mathbf{p}(V)\}}{d\mathbf{p}(\xi)} \tag{1}$$

is an element in $L^1(\Omega, \mathbf{p}(\xi))$. If measures $\mathbf{p}(\xi)$, $\xi \in M$, are dominated by μ_0, we also write

$$\mathbf{p}(\xi) = p(\xi) \cdot \mu_0 \text{ for some } p(\xi) \in L^1(\Omega, \mathbf{p}_0). \tag{2}$$

Definition 2 ([AJLS2016b, Definition 4.2]). *We say that a parametrized model* $(M, \Omega, \mu_0, \mathbf{p})$ *has a* regular density function *if the density function* $p : \Omega \times M \to \mathbb{R}$ *satisfying (2) can be chosen such that for all* $V \in T_\xi M$ *the partial derivative* $\partial_V p(.;\xi)$ *exists and lies in* $L^1(\Omega, \mu_0)$ *for some fixed* μ_0*.*

If the model has a positive regular density function, we have

$$\partial_V \log \mathbf{p}(\xi) = \partial_V \log p. \tag{3}$$

Next we recall the notion of k-integrability. On the set $\mathcal{M}(\Omega)$ we define the preordering $\mu_1 \le \mu_2$ if μ_2 dominates μ_1. Then $(\mathcal{M}(\Omega), \le)$ is a directed set, meaning that for any pair $\mu_1, \mu_2 \in \mathcal{M}(\Omega)$ there is a $\mu_0 \in \mathcal{M}(\Omega)$ dominating both of them (e.g. $\mu_0 := \mu_1 + \mu_2$).

For fixed $r \in (0, 1]$ and measures $\mu_1 \le \mu_2$ on Ω we define the linear embedding

$$\imath^{\mu_1}_{\mu_2} : L^{1/r}(\Omega, \mu_1) \longrightarrow L^{1/r}(\Omega, \mu_2), \qquad \phi \longmapsto \phi \left(\frac{d\mu_1}{d\mu_2}\right)^r.$$

Observe that

$$\begin{aligned} \|\imath^{\mu_1}_{\mu_2}(\phi)\|_{1/r} &= \left|\int_\Omega |\imath^{\mu_1}_{\mu_2}(\phi)|^{1/r}\, d\mu_2\right|^r = \left|\int_\Omega |\phi|^{1/r} \frac{d\mu_1}{d\mu_2}\, d\mu_2\right|^r \\ &= \left|\int_\Omega |\phi|^{1/r}\, d\mu_1\right|^r = \|\phi\|_{1/r}. \end{aligned} \tag{4}$$

It has been proved that $\imath_{\mu_2}^{\mu_1}$ is an isometry [AJLS2016b, (2.6)]. Moreover, $\imath_{\mu_2}^{\mu_1}\imath_{\mu_3}^{\mu_2} = \imath_{\mu_3}^{\mu_1}$ whenever $\mu_1 \le \mu_2 \le \mu_3$. Then we define the *space of r-th roots of measures on Ω* to be the directed limit over the directed set $(\mathcal{M}(\Omega), \le)$

$$\mathcal{S}^r(\Omega) := \varinjlim L^{1/r}(\Omega, \mu). \tag{5}$$

By [AJLS2016b, (2.9)] the space $\mathcal{S}^r(\Omega)$ is a Banach space provided with the norm $||\phi||_{1/r}$ defined in (4).

Denote the equivalence class of $\phi \in L^{1/r}(\Omega, \mu)$ by $\phi\mu^r$, so that $\mu^r \in \mathcal{S}^r(\Omega)$ is the equivalence class represented by $1 \in L^{1/r}(\Omega, \mu)$.

In [AJLS2016b, Proposition 2.2], for $r \in (0, 1]$ and $0 < k \le 1/r$ we defined a map

$$\tilde{\pi}^k : \mathcal{S}^r(\Omega) \to \mathcal{S}^{rk}(\Omega),\ \phi \cdot \mu^r \mapsto \mathrm{sign}(\phi)|\phi|^{\mathrm{k}}\mu^{\mathrm{rk}}.$$

For $1 \le k \le 1/r$ the map $\tilde{\pi}^k$ is a C^1-map between Banach spaces [AJLS2016b, (2.13)]. Using the same analogy, we set [AJLS2016b, (4.3)]

$$\mathbf{p}^{1/k} := \tilde{\pi}^{1/k} \circ \mathbf{p} : M \to \mathcal{S}^{1/k}(\Omega) \tag{6}$$

and

$$d_\xi \mathbf{p}^{1/k}(V) := \frac{1}{k}\partial_V \log \mathbf{p}(\xi)\ \mathbf{p}^{1/k}(\xi) \in \mathcal{S}^{1/k}(\Omega, \mathbf{p}(\xi)). \tag{7}$$

Definition 3 ([JLS2017a, Definition 2.6]). *A parametrized measure model $(M, \Omega, \mathbf{p})$ is called* k-integrable*, if the map $\mathbf{p}^{1/k}$ from (6) is a Fréchet-C^1-map.*

The k-integrability of parametrized measure models can be characterized in different ways.

Theorem 1 ([JLS2017a, Theorem 2.7]). *Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model. Then the model is k-integrable if and only if the map*

$$V \longmapsto \|d\mathbf{p}^{1/k}(V)\|_k < \infty \tag{8}$$

defined on TM is continuous.

Thus, $(M, \Omega, \mathbf{p})$ is k-integrable if and only if the map $d\mathbf{p}^{1/k} : M \to \mathcal{S}^{1/k}(\Omega)$ from (7) is well defined (i.e., $\partial_V \log \mathbf{p}(\xi) \in L^k(\Omega, \mathbf{p}(\xi))$) and continuous. In particular, the definition of k-integrability in Definition 3 above is equivalent to that in [AJLS2016b, Definition 4.4] and [AJLS2015, Definition 2.4].

Remark 1. 1. The Fisher metric $\mathfrak{g}$ on a 2-integrable parametrized measure model $(M, \Omega, \mathbf{p})$ is defined as follows for $v, w \in T_\xi M$

$$\mathfrak{g}_\xi(v, w) := \langle \partial_v \log \mathbf{p}; \partial_w \log \mathbf{p}\rangle_{L^2(\Omega, \mathbf{p}(\xi))} = \langle d\mathbf{p}^{1/2}(v); d\mathbf{p}^{1/2}(w)\rangle_{\mathcal{S}^{1/2}(\Omega)} \tag{9}$$

2. The standard notion of a statistical model always assumes that it is dominated by some measure and has a positive regular density function (e.g. [Borovkov1998, p. 140, 147], [BKRW1998, p. 23], [AN2000, Sect. 2.1], [AJLS2015, Definition 2.4]). In fact, the definition of a parametrized measure model or statistical model in [AJLS2015, Definition 2.4] is equivalent to a parametrized measure model or statistical model with a positive regular density function in the sense of Definition 2, see also [AJLS2016] for detailed discussion.

Let $(M, \Omega, \mathbf{p})$ be a 2-integrable parametrized measure model. Formula (9) shows that the kernel of the Fisher metric $\mathfrak{g}$ at $\xi \in M$ coincides with the kernel of the map $\Lambda_\xi : T_\xi M \to L^2(\Omega, \mathbf{p}(\xi))$, $V \mapsto \partial_V(\log \mathbf{p})$. In other words, the degeneracy of the Fisher metric $\mathfrak{g}$ is caused by the non-effectiveness of the parametrisation of the family $\mathbf{p}(\xi)$ by the map $\mathbf{p}$. The tangent cone $T_{\mathbf{p}(\xi)}\mathbf{p}(M)$ of the image $\mathbf{p}(M) \subset \mathcal{S}(\Omega)$ is isomorphic to the quotient $T_\xi M / \ker \Lambda_x$. This motivates the following

Definition 4. ([JLS2017a, Definition 2.9]). *The quotient $\hat{T}_\xi M := T_\xi M / \ker \Lambda_\xi$ will be called the essential tangent space of M at ξ.*

Clearly the Fisher metric $\mathfrak{g}$ descends to a non-degenerated metric $\hat{\mathfrak{g}}$ on $\hat{T}M$, which we shall call the *reduced Fisher metric*.

Denote by $\hat{T}^{\hat{\mathfrak{g}}}M$ the fiberwise completion of $\hat{T}M$ wrt the reduced Fisher metric $\hat{\mathfrak{g}}$. Its inverse $\hat{\mathfrak{g}}^{-1}$ is a well-defined quadratic form on the fibers of the dual bundle $\hat{T}^{*,\hat{\mathfrak{g}}^{-1}}M$, which we can therefore identify with $\hat{T}^{\hat{\mathfrak{g}}}M$.

2 The General Cramér-Rao Inequality

In this section we assume that $(M, \Omega, \mathbf{p})$ is a 2-integrable measure model. We introduce the notion of a regular function on a measure space Ω (Definition 5), state a rule of differentiation under integral sign (Proposition 1) and derive a general Cramér-Rao inequality (Theorem 2).

We set for $k \in \mathbb{N}^+$

$$L^k_M(\Omega) := \{f \in L^k(\Omega, \mathbf{p}(\xi)) \text{ for all } \xi \in M\}.$$

Definition 5. *Let $(M, \Omega, \mathbf{p})$ be a parametrized measure model. We call an element $f \in L^k_M(\Omega)$* regular *if the function $\xi \mapsto \|f\|_{L^k(\Omega, \mathbf{p}(\xi))}$ is locally bounded, i.e. if for all $\xi_0 \in M$*

$$\limsup_{\xi \to \xi_0} \|f\|_{L^k(\Omega, \mathbf{p}(\xi))} < \infty.$$

The regularity of a function f is important for the validity of differentiation under the integral sign.

Proposition 1. *Let $k, k' > 1$ be dual indices, i.e. $k^{-1} + k'^{-1} = 1$, and let $(M, \Omega, \mathbf{p})$ be a k'-integrable parametrized measure model. If $f \in L^k_M(\Omega)$ is regular, then the map*

$$M \longrightarrow \mathbb{R}, \qquad \xi \longmapsto \mathbb{E}_{\mathbf{p}(\xi)}(f) = \int_\Omega f \, d\mathbf{p}(\xi) \tag{10}$$

is Gatêaux-differentiable, and for $X \in TM$ the Gâteaux-derivative is

$$\partial_X \mathbb{E}_{\mathbf{p}(\xi)}(f) = \mathbb{E}_{\mathbf{p}(\xi)}(f \, \partial_X \log \mathbf{p}(\xi)) = \int_\Omega f \, \partial_X \log \mathbf{p}(\xi) \, d\mathbf{p}(\xi). \tag{11}$$

Let V be a topological vector space over the real field $\mathbb{R}$, possibly infinite dimensional. We denote by V^M the vector space of all V-valued functions on M. A V-valued function φ will stand for the coordinate functions on M, or in general, a feature of M (cf. [BKRW1998]). Let V^* denote the dual space of V. Later, for $l \in V^*$ we denote the composition $l \circ \varphi$ by φ^l. This should be considered as the l-th coordinate of φ.

Assume that $(M, \Omega, \mathbf{p})$ is a 2-integrable parametrized measure model. A Gateaux-differentiable function f on M whose differential df vanishes on $\ker d\mathbf{p} \subset TP$ will be called a *visible function*.

Recall that an *estimator* is a map $\hat\sigma : \Omega \to M$. If $k, k' > 1$ are dual indices, i.e., $k^{-1} + k'^{-1} = 1$, and given a k'-integrable parametrized measure model $(M, \Omega, \mathbf{p})$ and a function $\varphi \in V^M$, we define

$$L^k_\varphi(M, \Omega) := \{\hat\sigma : \Omega \to M \mid \varphi^l \circ \hat\sigma \in L^k_M(\Omega) \text{ for all } l \in V^*\}.$$

We call an estimator $\hat\sigma \in L^k_\varphi(M, \Omega)$ *φ-regular* if $\varphi^l \circ \hat\sigma \in L^k_M(\Omega)$ is regular for all $l \in V^*$.

Any $\hat\sigma \in L^k_\varphi(M, \Omega)$ induces a V^{**}-valued function $\varphi_{\hat\sigma}$ on M by computing the expectation of the composition $\varphi \circ \hat\sigma$ as follows

$$\langle \varphi_{\hat\sigma}(\xi), l \rangle := \mathbb{E}_{\mathbf{p}(\xi)}(\varphi^l \circ \hat\sigma) = \int_\Omega \varphi^l \circ \hat\sigma \, d\mathbf{p}(\xi) \tag{12}$$

for any $l \in V^*$. If $\hat\sigma \in L^k_\varphi(M, \Omega)$ is φ-regular, then Proposition 1 immediately implies that $\varphi_{\hat\sigma} : M \to V^{**}$ is visible with Gâteaux-derivative

$$\langle \partial_X \varphi_{\hat\sigma}(\xi), l \rangle = \int_\Omega \varphi^l \circ \hat\sigma \cdot \partial_X \log \mathbf{p}(\xi) \, \mathbf{p}(\xi). \tag{13}$$

Let $pr : TM \to \hat{T}M$ denote the natural projection.

Definition 6. ([JLS2017a, Definition 3.8]). *A section $\xi \mapsto \nabla_{\hat{\mathfrak{g}}} f(\xi) \in \hat{T}^{\hat{\mathfrak{g}}}_\xi M$ will be called the generalized Fisher gradient of a visible function f, if for all $X \in T_\xi M$ we have*

$$df(X) = \hat{\mathfrak{g}}(pr(X), \nabla_{\hat{\mathfrak{g}}} f).$$

If the generalized gradient belongs to $\hat{T}M$ we will call it the Fisher gradient.

We set (cf. [Le2016])

$$\mathcal{L}_1^k(\Omega) := \{(f, \mu) |\, \mu \in \mathcal{M}(\Omega) \text{ and } f \in L^k(\Omega, \mu)\}.$$

For a map $\mathbf{p} : P \to \mathcal{M}(\Omega)$ we denote by $\mathbf{p}^*(\mathcal{L}_1^k(\Omega))$ the pull-back "fibration" (also called the fiber product) $P \times_{\mathcal{M}(\Omega)} \mathcal{L}_1^k(\Omega)$.

Definition 7. ([JLS2017a, Definition 3.10]). *Let h be a visible function on M. A section*

$$M \to \mathbf{p}^*(\mathcal{L}_1^2(\Omega)),\ \xi \mapsto \nabla h_\xi \in L^2(\Omega, \mathbf{p}(\xi)),$$

is called a pre-gradient of h, if for all $\xi \in M$ and $X \in T_\xi M$ we have

$$dh(X) = \mathbb{E}_{\mathbf{p}(\xi)}((\partial_X \log \mathbf{p}) \cdot \nabla h_\xi).$$

Proposition 2. ([JLS2017a, Proposition 3.12]).

1. *Let $(M, \Omega, \mathbf{p})$ be a 2-integrable measure model and $f \in L^2_M(\Omega, V)$ is a regular function. Then the section of the pullback fibration $\mathbf{p}^*(\mathcal{L}_1^2(\Omega))$ defined by $\xi \mapsto f \in L^2(\Omega, \mathbf{p}(\xi))$ is a pre-gradient of the visible function $E_{\mathbf{p}(\xi)}(f)$.*
2. *Let $(P, \Omega, \mathbf{p})$ be a 2-integrable statistical model and $f \in L^2_P(\Omega, V)$. Then the section of the pullback fibration $\mathbf{p}^*(\mathcal{L}_1^2(\Omega))$ defined by $\xi \mapsto f - \mathbb{E}_{\mathbf{p}(\xi)}(f) \in L^2(\Omega, \mathbf{p}(\xi))$ is a pre-gradient of the visible function $E_{\mathbf{p}(\xi)}(f)$.*

For an estimator $\hat\sigma \in L^2_\varphi(P, \Omega)$ we define the *variance of $\hat\sigma$ w.r.t. φ* to be the quadratic form $V^\varphi_{\mathbf{p}(\xi)}[\hat\sigma]$ on V^* such that for all $l, k \in V^*$ we have [JLS2017a, (4.3)]

$$V^\varphi_{\mathbf{p}(\xi)}[\hat\sigma](l, k) := E_{\mathbf{p}(\xi)}[(\varphi^l \circ \hat\sigma - E_{\mathbf{p}(\xi)}(\varphi^l \circ \hat\sigma)) \cdot (\varphi^k \circ \hat\sigma - E_{\mathbf{p}(\xi)}(\varphi^k \circ \hat\sigma))]. \quad (14)$$

We regard $||d\varphi^l_{\hat\sigma}||^2_{\hat{\mathfrak{g}}^{-1}}(\xi)$ as a quadratic form on V^* and denote the latter one by $(\hat{\mathfrak{g}}^\varphi_{\hat\sigma})^{-1}(\xi)$, i.e.

$$(\hat{\mathfrak{g}}^\varphi_{\hat\sigma})^{-1}(\xi)(l, k) := \langle d\varphi^l_{\hat\sigma}, d\varphi^k_{\hat\sigma}\rangle_{\hat{\mathfrak{g}}^{-1}}(\xi).$$

Theorem 2 (General Cramér-Rao inequality) ([JLS2017a, Theorem 4.4]). *Let $(P, \Omega, \mathbf{p})$ be a 2-integrable statistical model, φ a V-valued function on P and $\hat\sigma \in L^2_\varphi(P, \Omega)$ a φ-regular estimator. Then the difference $V^\varphi_{\mathbf{p}(\xi)}[\hat\sigma] - (\hat{\mathfrak{g}}^\varphi_{\hat\sigma})^{-1}(\xi)$ is a positive semi-definite quadratic form on V^* for any $\xi \in P$.*

Remark 2. Assume that V is finite dimensional and φ is a coordinate mapping. Then $\mathfrak{g} = \hat{\mathfrak{g}}$, $d\varphi^l = d\xi^l$, and abbreviating $b^\varphi_{\hat\sigma}$ as b, we write

$$(\mathfrak{g}^\varphi_{\hat\sigma})^{-1}(\xi)(l, k) = \langle \sum_i (\frac{\partial \xi^l}{\partial \xi^i} + \frac{\partial b^l}{\partial \xi^i}) d\xi^i, \sum_j (\frac{\partial \xi^k}{\partial \xi^j} + \frac{\partial b^k}{\partial \xi^j}) d\xi^j \rangle_{\mathfrak{g}^{-1}}(\xi). \quad (15)$$

Let $D(\xi)$ be the linear transformation of V whose matrix coordinates are

$$D(\xi)^l_k := \frac{\partial b^l}{\partial \xi^k}.$$

Using (15) we rewrite the Cramér-Rao inequality in Theorem 2 as follows

$$V_\xi[\hat{\sigma}] \geq (E + D(\xi))\mathfrak{g}^{-1}(\xi)(E + D(\xi))^T. \tag{16}$$

The inequality (16) coincides with the Cramér-Rao inequality in [Borovkov1998, Theorem 1.A, p. 147]. By Theorem 1, the condition (R) in [Borovkov1998, p. 140, 147] for the validity of the Cramér-Rao inequality is essentially equivalent to the 2-integrability of the (finite dimensional) statistical model with positive density function under consideration, more precisely Borokov ignores/excludes the points $x \in \Omega$ where the density function vanishes for computing the Fisher metric. Borovkov also uses the φ-regularity assumption, written as $\mathbb{E}_\theta((\theta^*)^2) < c < \infty$ for $\theta \in \Theta$, see also [Borovkov1998, Lemma 1, p. 141] for a more precise formulation. Classical versions of Cramér-Rao inequalities, as in e.g. [CT2006], [AN2000], are special cases of the Cramér-Rao inequality in [Borovkov1998]. We refer the reader to [JLS2017a] for comparison of our Cramér-Rao inequality with more recent Cramér-Rao inequalities in parametric statistics.

3 Optimality of the General Cramér-Rao Inequality

To investigate the optimality of our general Cramér-Rao inequality we introduce the following

Definition 8 ([JLS2017b]). *Assume that φ is a V-valued function on P, where $(P, \Omega, \mathbf{p})$ is a 2-integrable statistical model. A φ-regular estimator $\hat{\sigma} \in L^2_\varphi(P, \Omega)$ will be called φ-efficient, if $V^\varphi_{\mathbf{p}(\xi)} = (\hat{\mathfrak{g}}^\varphi_{\hat{\sigma}})^{-1}(\xi)$ for all $\xi \in P$.*

If a statistical model $(P, \Omega, \mathbf{p})$ admits a φ-efficient estimator, the Cramér-Rao inequality is optimal on $(P, \Omega, \mathbf{p})$.

Example 1. Assume that $(P \subset \mathbb{R}^n, \Omega \subset \mathbb{R}^n, \mathbf{p})$ is a minimal full regular exponential family, $\varphi : P \to \mathbb{R}^n$ - the canonical embedding $P \to \mathbb{R}^n$, and $\hat{\sigma} : \Omega \to P$ - the mean value parametrization. Then it is well known that $\hat{\sigma}$ is an unbiased φ-efficient estimator, see e.g. [Brown1986, Theorem 3.6, p. 74]. Let S be a submanifold in P and $f : P' \to P$ is a blowing-up of P along S, i.e. f is a smooth surjective map such that $\ker df$ is non -trivial exactly at $f^{-1}(S)$. Then $(P', \Omega, \mathbf{p} \circ f)$ is a strictly singular statistical model which admits an unbiased φ-efficient estimator, since $(P, \Omega, \mathbf{p})$ admits unbiased φ-efficient estimator.

Example 2. Let Ω_n be a finite set of n elements. Let $A : \Omega_n \to \mathbb{R}^d_+$ be a map, where $d \leq m - 1$. We define an exponential family $P^A(\cdot|\theta) \subset \mathcal{M}(\Omega_m)$ with parameter θ in $\mathbb{R}^d$ as follows.

$$P^A(x|\theta) = Z_A(\theta) \cdot \exp\langle\theta, A(x)\rangle, \text{ for } \theta \in \mathbb{R}^d, \text{ and } x \in \Omega_m. \tag{17}$$

Here $Z_A(\theta)$ is the normalizing factor such that $P^A(\cdot|\theta) \cdot \mu_0$ is a probability measure, where μ_0 is the counting measure on Ω_m: $\mu_0(x_i) = 1$ for $x_i \in \Omega_m$.

Denote $A^l(x) := \langle l, A(x)\rangle$ for $l \in (\mathbb{R}^d)^*$. We set

$$\hat{\sigma} : \Omega_n \to \mathbb{R}^d,\ x \mapsto \log A(x) := (\log A^1(x), \cdots, \log A^d(x)),$$

$$\varphi : \mathbb{R}^d \to \mathbb{R}^d_+ \subset \mathbb{R}^d,\ \theta \mapsto \exp \theta.$$

Then $\hat{\sigma}$ is a (possibly biased) φ-efficient estimator [JLS2017b]. Using blowing-up, we obtain strictly singular statistical models admitting (possibly biased) φ-efficient estimators.

In [Fukumizu2009] Fukumizu constructed a large class of infinite dimensional exponential families using reproducing kernel Hilbert spaces (RKHS). Assume that Ω is a topological space and μ is a Borel probability measure such that $sppt(\mu) = \Omega$. Let $k : \Omega \times \Omega \to \mathbb{R}$ be a continuous positive definite kernel on Ω. It is known that for a positive definite kernel k on Ω there exists a unique RKHS $\mathcal{H}_k$ such that

1. $\mathcal{H}_k$ consists of functions on Ω,
2. Functions of the form $\sum_{i=1}^m a_i k(\cdot, x_i)$ are dense in $\mathcal{H}_k$,
3. For all $f \in \mathcal{H}_k$ we have $\langle f, k(\cdot, x)\rangle = f(x)$ for all $x \in \Omega$,
4. $\mathcal{H}_k$ contains the constant functions $c_{|\Omega}$, $c \in \mathbb{R}$.

For a given positive definite kernel k on Ω we set

$$\hat{k} : \Omega \to \mathcal{H}_k,\ \hat{k}(x) := k(\cdot, x).$$

Theorem 3 ([JLS2017b]). *Assume that Ω is a complete topological space and μ is a Borel probability measure with $sppt(\mu) = \Omega$. Suppose that a kernel k on Ω is bounded and satisfies the following relation whenever $x, y \in \Omega$*

$$\hat{k}(x) - \hat{k}(y) = c_{|\Omega} \in \mathcal{H}_k \Longrightarrow c_{|\Omega} = 0 \in \mathcal{H}_k. \tag{18}$$

Let

$$\mathcal{P}_\mu := \{f \in L^1(\Omega, \mu) \cap C^0(\Omega) |\ f > 0 \text{ and } \int_\Omega f d\mu = 1\}.$$

Set

$$\mathbf{p} : \mathcal{P}_\mu \to \mathcal{M}(\Omega),\ f \mapsto f \cdot \mu_0.$$

Then there exists a map $\varphi : \mathcal{P}_\mu \to \mathcal{H}_k$ such that $(\mathcal{P}_\mu, \Omega, \mathbf{p})$ admits a φ-efficient estimator.

References

[AJLS2015] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. Probab. Theory Relat. Fields **162**, 327–364 (2015)

[AJLS2016] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry. Springer (2017). Ergebnisse der Mathematik und ihrer Grenzgebiete

[AJLS2016b] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Parametrized measure models. (accepted for Bernoulli Journal). arXiv:1510.07305

[AN2000] Amari, S., Nagaoka, H.: Methods of Information Geometry. Translations of Mathematical Monographs, vol. 191. American Mathematical Society (2000)

[BKRW1998] Bickel, P., Klaassen, C.A.J., Ritov, Y., Wellner, J.A.: Efficient and Adaptive Estimation for Semiparametric Models. Springer, New York (1998)

[Borovkov1998] Borovkov, A.A.: Mathematical Statistics. Gordon and Breach Science Publishers (1998)

[Brown1986] Brown, L.D.: Fundamentals of Statistical Families with Applications in Statistical Decision Theory, Lecture Notes-Monograph Series, vol. 9. IMS (1986)

[CT2006] Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley, Hoboken (2006)

[Fukumizu2009] Fukumizu, K.: Exponential manifold by reproducing kernel Hilbert spaces. In: Gibilisco, P., Riccomagno, E., Rogantin, M.-P., Winn, H. (eds.) Algebraic and Geometric methods in Statistics, pp. 291–306. Cambridge University Press (2009)

[JLS2017a] Jost, J., Lê, H.V., Schwachhöfer, L.: The Cramér-Rao inequality on singular statistical models I. arXiv:1703.09403

[JLS2017b] Jost, J., Lê, H.V., Schwachhöfer, L.: The Cramér-Rao inequality on singular statistical models II (2017, preprint)

[Le2016] Lê, H.V.: The uniqueness of the Fisher metric as information metric. AISM **69**, 879–896 (2017)