ISSN 2279-9362

Collegio Carlo Alberto

A Class of Non-Parametric Deformed Exponential Statistical Models

Luigi Montrucchio and Giovanni Pistone

No. 562
December 2018

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

A CLASS OF NON-PARAMETRIC DEFORMED EXPONENTIAL STATISTICAL MODELS

MONTRUCCHIO, LUIGI AND GIOVANNI PISTONE

Abstract. We study the class on non-parametric deformed statistical models where the deformed exponential has linear growth at infinity and is sub-exponential at zero. We discuss the convexity and regularity of the normalization operator, the form of the deformed statistical divergences and their convex duality, the properties of the escort densities, and the affine manifold structure of the statistical bundle.

1. Introduction

Let $\mathcal{M}$ be a family of (strictly) positive probability densities on the probability space $(\mathbb{X}, \mathcal{X}, \mu)$. At each $p \in \mathcal{M}$, the Hilbert space of square-integrable random variables $L^2(p \cdot \mu)$ provides a fiber that sits at each $p \in \mathcal{P}$, so we can define the *Hilbert bundle* with base $\mathcal{M}$. Such a bundle is a convenient framework for Information Geometry, cf. [1] and the non-parametric version in [16, 14].

If $\mathcal{M}$ is an exponential manifold in the sense of [16], there exists a splitting of each fiber $L(p \cdot \mu) = H_p \oplus H_p^\perp$, such that each H_p contains a dense vector sub-space which is an expression of the tangent space $T_p\mathcal{M}$ of the manifold. Moreover, the geometry on $\mathcal{M}$ is affine and Hessian.

When the sample space is finite and $\mathcal{M}$ is the full set $\mathcal{P}$ of positive probability densities, then $H_p = L_0^2(p)$ and each $H_p \simeq T_p\mathcal{M}$. A similar situation occurs when $\mathcal{M}$ is a finite-dimensional exponential family. It is difficult to devise set-ups other than those mentioned above, where the identification of the Hilbert fiber with the tangent space holds true. In fact, a necessary condition would be the topological linear isomorphism between the fibers.

There have been many alternative proposals on how to define a manifold $\mathcal{M}$ of positive probability densities modeled on an Hilbert space. A successful one has been introduced by N.J. Newton [13] using what he calls the "balanced chart" $p \mapsto \log p + p - 1 \in L_0^2(\mu)$. This chart is a "deformation" of the usual logarithmic representation and it is an instance of "deformed logarithm" as defined by J. Naudts [12]. In this approach the Hilbert bundle is trivial as all the fibers coincide with $L_0^2(\mu)$.

In this paper, we take out this approach showing how to define the affine structure of the relevant Hilbert bundle by the use of deformed exponential families as defined [12] but allowing for a general reference measure as done by R.F. Vigelis and C.C. Cavalcante [18]. We use the representation $p = \exp_A(v)$, where $\exp_A$ is an exponential-like function which has a linear growth at $+\infty$ and is dominate by an exponential at $-\infty$.

The formalism of deformed exponentials is reviewed in a special case on Sec. 2. The following Sec. 3 is devoted to the adaptation of deformed exponential families to the non-parametric case. In Sec. 4 we discuss the form of the divergence which is natural in our case. Sec. 5 discusses the construction of the Hilbert statistical bundle.

A partial version of this piece of research has been presented at the GSI 2017 Conference [11] and we refer to that paper for some of the proofs.

2. Deformed exponential

We review first a special case of the deformed exponential formalism of [12].

We assume to be given a function A from $]0, +\infty[$ onto $]0, a[$, strictly increasing, continuously differentiable, such that $\|A'\|_\infty < \infty$. It follows $a = \|A\|_\infty$ and $A(x) \leq \|A'\|_\infty x$, so that $\int_0^1 d\xi/A(\xi)\ d\xi = +\infty$.

The A-logarithm is the function

$$\log_A(x) = \int_1^x \frac{1}{A(\xi)}\, d\xi, \quad x \in]0, +\infty[\ .$$

The A-logarithm is strictly increasing from $-\infty$ to $+\infty$, its derivative $\log_A'(x) = 1/A(x)$ is positive and strictly decreasing for all $x > 0$, hence it is strictly concave.

By inverting the A-logarithm, one obtains the A-exponential, $\exp_A = \log_A^{-1}$. Hence, the function $\exp_A \colon]-\infty, +\infty[\to]0, +\infty[$ is strictly increasing, strictly convex, and is the solution of the Cauchy problem

$$\exp_A'(y) = A(\exp_A(y)), \quad \exp_A(0) = 1 \ . \tag{1}$$

As a consequence, we have the linear bound

$$|\exp_A(y_1) - \exp_A(y_2)| \le \|A\|_\infty |y_1 - y_2| \ . \tag{2}$$

The behavior of the A-logarithm is linear for large arguments and super-logarithmic for small arguments. To derive explicit bounds, define

$$\alpha_1 = \min_{x \le 1} \frac{A(x)}{x} \ , \quad \alpha_2 = \max_{x \le 1} \frac{A(x)}{x} \ ,$$

namely the best constants such that $\alpha_1 x \le A(x) \le \alpha_2 x$ if $x \le 1$. Note that $\alpha_1 \ge 0$ and $\alpha_2 > 0$. If $\alpha_1 > 0$, it follows that

$$\frac{1}{\alpha_2} \log x \le \log_A x \le \frac{1}{\alpha_1} \log x \ , \quad x \le 1 \ . \tag{3}$$

If $\alpha_1 = 0$, the left inequality only holds.

For $x \ge 1$ we have $A(1) \le A(x) < 1$, hence

$$x - 1 < \log_A x \le \frac{1}{A(1)}(x-1) \ , \quad x \ge 1 \ . \tag{4}$$

2.1. **Examples.** The main example of A-logarithm is the the N.J. Newton A-logarithm [13], with

$$A(x) = 1 - \frac{1}{1+x} = \frac{x}{1+x} \ ,$$

so that

$$\log_A(y) = \log x + x - 1 \ .$$

There is a simple algebraic expression for the product,

$$\log_A(x_1 x_2) = \log_A(x_1) + \log_A(x_2) + (x_1 - 1)(x_2 - 1) \ .$$

Other similar examples are available in the literature. One is a special case of the G. Kaniadakis' exponential of [8] i.e.,

$$\exp_A(y) = y + \sqrt{1+y^2} \ ,$$

whose inverse is easily derived from the relation

$$y + \sqrt{1+y^2} - \frac{1}{y + \sqrt{1+y^2}} = 2y.$$

The inverse is

$$\log_A x = \frac{x - x^{-1}}{2} \ ,$$

which in turn provides

$$A(\xi) = \frac{2\xi^2}{1+\xi^2} \ .$$

A remarkable feature of the G. Kaniadakis' exponential is

$$\exp_A(y)\exp_A(-y) = \left(y + \sqrt{1+y^2}\right)\left(-y + \sqrt{1+y^2}\right) = 1$$

Notice that the A function on the N.J. Newton exponential is concave, while the A function of the G. Kaniadakis exponential is not.

Another example is $A(\xi) = 1 - 2^{-\xi}$, which gives $\log_A(x) = \log_2(1 - 2^{-x})$ and $\exp_A(y) = \log_2(1 + 2^y)$.

A notable example of deformed exponential that does not fit into our set of assumptions is the Tsallis logarithm with parameter 1/2 of [17],

$$\log_{1/2} x = 2\left(\sqrt{x} - 1\right) = \int_1^x \frac{1}{\sqrt{\xi}}\, d\xi \ .$$

In this case, $\log_{1/2}(0+) = -\int_0^1 d\xi/\sqrt{\xi} = -2$, so that the inverse is not defined for all real numbers. The Tsallis logarithm provides models with heavy tails, which is not the case in our setting.

2.2. **Superposition operator.** The deformed exponential is used to represent positive probability densities in the form $p(x) = \exp_A[u(x)]$, where u is a random variable on the probability space $(\mathbb{X}, \mathcal{X}, \mu)$. Because of that, we are interested in the properties of the *superposition operator*

$$S_A \colon u \mapsto \exp_A \circ u \tag{5}$$

in some convenient functional setting. See e.g. [2, Ch. 1] and [3, Ch. 3] about superposition operators.

It is clear from Eq. (2) that $\exp_A(u) \le 1 + \|A\|_\infty |u|$, which in turn implies that the superposition operator S_A maps $L^\alpha(\mu)$ to itself for all $\alpha \in [1, +\infty]$ and the mapping is uniformly Lipschitz with constant $\|A\|_\infty$. Notice that we are assuming that μ is a finite measure. The superposition operator $S_A \colon L^\alpha(\mu) \to L^\alpha(\mu)$ is 1-to-1 and its image consists of all positive random variables f such that $\log_A f \in L^\alpha(\mu)$.

Proposition 1. (1) *For all $\alpha \in [1, \infty]$, the superposition operator S_A of Eq. (5) is Gateaux-differentiable with derivative*

$$dS_A(u)[h] = A(\exp_A(u))h \ . \tag{6}$$

(2) *For all $\alpha > \beta \ge 1$, the superposition operator S_A of Eq. (5) is Fréchet-differentiable from $L^\alpha(\mu)$ to $L^\beta(\mu)$.*

Proof. (1) Eq. (1) implies that for each couple of random variables $u, h \in L^\alpha(\mu)$ we have

$$\lim_{t \to 0} t^{-1} \left(\exp_A(u + th) - \exp_A(u)\right) - A(\exp_A(u))h = 0$$

point-wise. Moreover, for each $\alpha \in [1, \infty[$ we derive, by Jensen inequality, that for $t > 0$ it holds

$$\left|t^{-1}\left(\exp_A(u + th) - \exp_A(u)\right) - A(\exp_A(u))h\right|^\alpha \le \\ t^{-1} |h|^\alpha \int_0^t \left|A(\exp_A(u + rh)) - A(\exp_A(u))\right|^\alpha \, dr \le \left(2\|A\|_\infty\right)^\alpha |h|^\alpha \ .$$

Now, bounded converge forces the limit to hold in $L^\alpha(\mu)$. For $t < 0$, change h to $-h$.

If $\alpha = \infty$, we can use the second order bound

$$\left|t^{-1}\left(\exp_A(u + th) - \exp_A(u)\right) - A(\exp_A(u))h\right| = \\ |t|^{-1} h^2 \left|\int_0^t (t - r)\frac{d}{dr} A(\exp_A(u + rh))\, dr\right| \le \frac{t}{2} \|h\|_\infty^2 \left\|A'\right\|_\infty \|A\|_\infty \ .$$

As $\|A' \cdot A\|_\infty < \infty$, then the RHS goes to 0 as $t \to 0$ uniformly for each $h \in L^\infty(\mu)$.

(2) Given $u, h \in L^\alpha(\mu)$, let us use again the Taylor formula to get

$$\int \left|\exp_A(u+h) - \exp_A(u) - A(\exp_A(u))h\right|^\beta \ d\mu \leq \int |h|^\beta \int_0^1 \left|A(\exp_A(u+rh)) - A(\exp_A(u))\right|^\beta \ dr \ d\mu \ .$$

By using Hölder inequality with conjugate exponents α/β and $\alpha/(\alpha-\beta)$ the RHS is bounded by

$$\left(\int |h|^\alpha \ d\mu\right)^{\frac{\beta}{\alpha}} \left(\iint \left|A(\exp_A(u+rh)) - A(\exp_A(u))\right|^{\frac{\alpha\beta}{\alpha-\beta}} \ dr \ d\mu\right)^{\frac{\alpha-\beta}{\alpha}} ,$$

hence,

$$\|h\|_{L^\alpha(\mu)}^{-1} \left\|\exp_A(u+h) - \exp_A(u) - A(\exp_A(u))h\right\|_{L^\beta(\mu)} \leq \left(\iint \left|A(\exp_A(u+rh)) - A(\exp_A(u))\right|^{\frac{\alpha\beta}{\alpha-\beta}} \ dr \ d\mu\right)^{\frac{\alpha-\beta}{\alpha\beta}} .$$

In order to show that the RHS tend to zero as $\|h\|_{L^\alpha(\mu)} \to 0$, observe that for all $\delta > 0$ we have

$$\left|A(\exp_A(u+rh)) - A(\exp_A(u))\right| \leq \begin{cases} 2\,\|A\|_\infty & \text{always,} \\ \|A'\|_\infty \|A\|_\infty \delta & \text{if } |h| \leq \delta, \end{cases}$$

so that, decomposing the double integral as $\iint = \iint_{|h|\leq\delta} + \iint_{|h|>\delta}$, we obtain

$$\iint \left|A(\exp_A(u+rh)) - A(\exp_A(u))\right|^\gamma \ dr \ d\mu \leq (2\,\|A\|_\infty)^\gamma \mu\{|h| > \delta\} + \left(\left\|A'\right\|_\infty \|A\|_\infty \delta\right)^\gamma \leq (2\,\|A\|_\infty)^\gamma \delta^{-\alpha} \int |h|^\alpha \ d\mu + \left(\left\|A'\right\|_\infty \|A\|_\infty \delta\right)^\gamma ,$$

where $\gamma = \alpha\beta/(\alpha-\beta)$ and we have used Cebičev inequality. Now it is clear that the last bound implies the conclusion for each $\alpha < \infty$. The case $\alpha = \infty$ follows *a fortiori.*

□

It is not generally true for $\alpha = \beta$ that the superposition operator S_A is Fréchet differentiable, cf. [2, §1.2]. Here is a well known counter-example. Assume μ is a non-atomic probability measure. For each $\lambda \in \mathbb{R}$ and $\delta > 0$ define the simple function

$$h_{\lambda,\delta}(x) = \begin{cases} \lambda & \text{if } |x| \leq \delta, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that for each $\alpha \in [1, +\infty[$ we have

$$\lim_{\delta\to 0} \|h_{\lambda,\delta}\|_{L^\alpha(\mu)} = \lim_{\delta\to 0} |\lambda|\, \mu\{|x| \leq \delta\}^{1/\alpha} = 0 \ .$$

Differentiability at 0 in $L^\alpha(\mu)$ would imply for all λ

$$0 = \lim_{\delta\to 0} \frac{\|\exp_A(h_{\lambda,\delta}) - 1 - A(1)h_{\lambda,\delta}\|_{L^\alpha(\mu)}}{\|h_{\lambda,\delta}\|_{L^\alpha(\mu)}} = \lim_{\delta\to 0} \frac{|\exp_A(\lambda) - 1 - A(1)\lambda|\, \mu\{x|\,|x| \leq \delta\}^{1/\alpha}}{|\lambda|\, \mu\{x|\,|x| \leq \delta\}^{1/\alpha}} = \left|\frac{\exp_A(\lambda) - 1}{\lambda} - A(1)\right| ,$$

hence a contradiction.

We conclude this section by observing that it is also interesting to study the action of the superposition operator on spaces of differentiable functions, for example Gauss-Sobolev spaces

[10]. Assume that μ is the standard Gaussian measure on $\mathbb{R}^n$, and u is a differentiable function such that u, $\frac{\partial}{\partial x_i} u \in L^2(\mu)$, $i = 1, \dots, n$. It follows $\exp_A(u) \in L^2(\mu)$ and also $\frac{\partial}{\partial x_i} \exp_A(u) \in L^2(\mu)$ because

$$\frac{\partial}{\partial x_i} \exp_A(u(x)) = A(\exp_A(u(x))\frac{\partial}{\partial x_i} u(x) \ .$$

We do not pursue this line of investigation here.

3. Deformed exponential family based on $\exp_A$

In the spirit of [18, 4], we consider the deformed exponential curve in the space of positive measures on $(\mathbb{X}, \mathcal{X})$ given by

$$t \mapsto \mu_t = \exp_A(tu + \log_A p) \cdot \mu \ , \quad u \in L^1(\mu) \ .$$

We have $\exp_A(x + y) \leq \|A\|_\infty x^+ + \exp_A(y)$, because the inequality holds for $x \leq 0$ as $\exp_A$ is increasing and for $x = x^+ > 0$ the inequality follows from Eq. (2). As a consequence, each μ_t is a finite measure, $\mu_t(\mathbb{X}) \leq \|A\|_\infty \int (tu)^+ \ d\mu + 1$, with $\mu_0 = p \cdot \mu$. The curve is actually continuous and differentiable in $L^1(\mu)$ because the point-wise derivative of the density $p_t = \exp_A(tu + \log_A(p))$ is $\dot{p}_t = A(p_t)u$ so that $|\dot{p}_t| \leq \|A\|_\infty |u|$. In conclusion $\mu_0 = p$ and $\dot{\mu}_0 = A\,(p)\,u$.

Notice that there are two ways to normalize to total mass 1 the density p_t, either dividing by a normalizing constant $Z(t)$ to get the statistical model $t \mapsto \exp_A(tu - \log_A p)/Z(t)$ or, subtracting a constant $\psi(t)$ from the argument to get the model $t \mapsto \exp_A(tu - \psi(t) + \log_A(p))$. In the standard exponential case the two methods lead to the same result, which is not the case for deformed exponentials where $\exp_A(\alpha + \beta) \neq \exp_A(\alpha) \exp_A(\beta)$. We choose in the present paper the latter option.

Here we use the ideas of [12, 18, 4] to construct deformed non-parametric exponential families. Recall that we are given: the probability space $(\mathbb{X}, \mathcal{X}, \mu)$; the set $\mathcal{P}$ of positive probability densities and the function A satisfies the conditions listed in Section 2. Throughout this section, the density $p \in \mathcal{P}$ will be fixed.

The following proposition is taken from [11] where a detailed proof is given.

Proposition 2.

(1) *The mapping* $L^1(\mu) \ni u \mapsto \exp_A(u + \log_A p) \in L^1(\mu)$ *has full domain and is* $\|A\|_\infty$ *-Lipschitz. Consequently, the mapping*

$$u \mapsto \int g \exp_A(u + \log_A p) \ d\mu$$

is $\|g\|_\infty \cdot \|A\|_\infty$*-Lipschitz for each bounded function* g.

(2) *For each* $u \in L^1(\mu)$ *there exists a unique constant* $K_p(u) \in \mathbb{R}$ *such that* $\exp_A(u - K_p(u) + \log_A p) \cdot \mu$ *is a probability.*

(3) *It holds* $K_p(u) = u$ *if, and only if,* u *is constant. In such a case,*

$$\exp_A(u - K_p(u) + \log_A p) \cdot \mu = p \cdot \mu \ .$$

Otherwise, $\exp_A(u - K_p(u) + \log_A p) \cdot \mu \neq p \cdot \mu$.

(4) *A density* q *is of the form* $q = \exp_A(u - K_p(u) + \log_A p)$*, with* $u \in L^1(\mu)$ *if, and only if,* $\log_A q - \log_A p \in L^1(\mu)$.

(5) *If* $u, v \in L^1(\mu)$ *and*

$$\exp_A(u - K_p(u) + \log_A p) = \exp_A(v - K_p(v) + \log_A p) \ ,$$

then $u - v$ *is constant.*

(6) *The functional* $K_p \colon L^1(\mu) \to \mathbb{R}$ *is translation invariant. More specifically,* $K_p(u + c) = K_p(u) + cK_p(1)$ *holds for all* $c \in \mathbb{R}$.

(7) $K_p : L^1(\mu) \to \mathbb{R}$ *is continuous and convex.*

We now discuss the form of the sub-gradient of the convex continuous function K_p. We refer to [6, Part I] for the general theory of convex functions in infinite dimension.

3.1. **Escort density.** For each positive density $q \in \mathcal{P}$, its *escort density* is

$$\operatorname{escort}(q) = \frac{A(q)}{\int A(q)\, d\mu} \ ,$$

see [12]. Notice that $0 \leq A(q) \leq A(\|q\|_\infty) \leq \|A\|_\infty$. In particular $\widetilde{q} = \operatorname{escort}(q)$ is a bounded positive density.

Assume $\operatorname{escort}(q_1) = \operatorname{escort}(q_2)$ for μ-almost all x. Say, $\int A \circ q_1\, d\mu \geq \int A \circ q_2\, d\mu$. Then $A(q_1(x)) \leq A(q_2(x))$, for μ-almost all x. Since A is strictly increasing, it follows $q_1(x) \leq q_2(x)$ for μ-almost all x, which, in turn, implies $q_1 = q_2$ μ-a.s. because both μ-integrals are equal to 1. In conclusion, the escort mapping is a.s. injective.

We want to discuss the image of the escort mapping.

Proposition 3. (1) *A bounded positive density $\widetilde{q}$ is an escort density if, and only if,*

$$\lim_{\alpha \uparrow \|A\|_\infty} \int A^{-1}\left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty}\right)\, d\mu \geq 1 \ . \tag{7}$$

(2) *The condition* (7) *holds if $\mu\left\{\widetilde{q} = \|\widetilde{q}\|_\infty\right\} > 0$. In particular, every simple density is an escort density.*

(3) *If $\widetilde{q}_1 = \operatorname{escort}(q_1)$ is an escort density, and q_2 is a bounded positive density such that*

$$\mu\left\{\widetilde{q}_1 > t\,\|\widetilde{q}_1\|_\infty\right\} \leq \mu\left\{q_2 > t\,\|q_2\|_\infty\right\}, \quad t > 0 \ ,$$

then q_2 is an escort density.

Proof. (1) Let be given a $\widetilde{q} \in \mathcal{P} \cap L^\infty(\mu)$, and consider the mapping

$$f(\alpha) = \int A^{-1}\left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty}\right)\, d\mu, \quad \alpha \in [0,1[\ .$$

We have $f(0) = 0$ and the mapping is finite, increasing, continuous. It is clear that the range condition in Eq. (7) is necessary because $\widetilde{q} = \operatorname{escort}(q)$ implies $q = A^{-1}\left(\left(\int A(q)\, d\mu\right)\widetilde{q}\right)$ and, in turn, $1 = \int A^{-1}\left(\left(\int A(q)\, d\mu\right)\widetilde{q}\right)\, d\mu$ because q is a probability density. We can take $\alpha = \int A(q)\, d\mu\ \|\widetilde{q}\|_\infty \leq \|A\|_\infty$ to satisfy the range condition. Conversely, if the rank condition is satisfied, there exists $\alpha \leq \|A\|_\infty$ such that $q = A^{-1}\left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty}\right)$ is positive probability density whose escort is $\widetilde{q}$.

(2) The special case of Item 2. follows from the inequality

$$\int A^{-1}\left(\alpha \frac{\widetilde{q}}{\|\widetilde{q}\|_\infty}\right)\, d\mu \geq A^{-1}(\alpha)\mu\left\{\widetilde{q} = \|\widetilde{q}\|_\infty\right\} \ .$$

(3) For each bounded positive density q we have

$$\int A^{-1}\left(\frac{q}{\|q\|_\infty}\right)\, d\mu = \int_0^{+\infty} \mu\left\{\frac{q}{\|q\|_\infty} > A(t)\right\}\, dt =$$

$$\int_0^{\|A\|_\infty} \mu\left\{\frac{q}{\|q\|_\infty} > s\right\} \frac{1}{A'\left(A^{-1}(s)\right)}\, ds \ .$$

Now the necessary condition of Item 3. follows from Item 1. and our assumptions. $\square$

The previous proposition shows that the range of the escort mapping is uniformly dense as it contains all simple densities. Moreover, in the partial order induced by the rearrangement of the normalized density (that is for each q the mapping $t \mapsto \mu\left\{\frac{q}{\|q\|_\infty} > t\right\}$), it contains the full right interval of each element. But the range of the escort mapping is not the full set of bounded positive densities, unless the σ-algebra $\mathcal{X}$ is a finite partition. To provide an example, consider on the Lebesgue unit interval the densities $q_\delta(x) \propto (1 - x^{1/\delta})$, $\delta > 0$, and $A(x) = x/(1+x)$. It turns out that q_δ is an escort density if, and only if, $\delta \leq 1$.

3.2. **Gradient of K_p.** Prop. 2 shows that the functional K_p is a global solution of a functional equation. We now give local properties of K_p by the implicit function theorem.

For each $u \in L^1(\mu)$, we write

$$q(u) = \exp_A(u - K_p(u) + \log_A p)$$

and $\widetilde{q}(u) = \text{escort}\,(q(u))$ denotes its escort density.

Proposition 4.

(1) *The functional $K_p \colon L^1(\mu) \to \mathbb{R}$ is Gateaux-differentiable with derivative*

$$\frac{d}{dt} K_p(u + tv)\bigg|_{t=0} = \int v\widetilde{q}(u)\ d\mu\ .$$

It follows that $K_p \colon L^1(\mu) \to \mathbb{R}$ is monotone and globally Lipschitz.

(2) *For every $u, v \in L^1(\mu)$, the inequality*

$$K_p(u + v) - K_p(u) \geq \int v\widetilde{q}(u)\ d\mu$$

holds i.e., the density $\widetilde{q}(u) \in L^\infty(\mu)$ is the unique sub-gradient of K_p at u.

Proof. (1) Consider the equation

$$F(t, \kappa) = \int \exp_A(u + tv - \kappa + \log_A p)\ d\mu - 1, \quad t, \kappa \in \mathbb{R}\ .$$

so that $\kappa = K_p(u + tv)$. The implicit function theorem applies by derivation under the integral because of the bounds

$$\left|\frac{\partial}{\partial t} \exp_A(u + tv - \kappa + \log_A p)\right| = \\ |A(\exp_A(u + tv - \kappa + \log_A p))v| \leq \|A\|_\infty |v|$$

and

$$\left|\frac{\partial}{\partial \kappa} \exp_A(u + tv - \kappa + \log_A p)\right| = |A(\exp_A(u + tv - \kappa + \log_A p))| \leq \|A\|_\infty\ .$$

Moreover the partial derivative with respect to κ is never zero. Therefore there exists the derivative $(d\kappa/dt)_{t=0}$ which is the desired Gateaux derivative. As $\widetilde{q}(u)$ is positive and bounded, then K_p is monotone and globally Lipschitz.

(2) Thanks to convexity of $\exp_A$ and the derivation formula, we have

$$\exp_A(u + v - K_p(u + v) + \log_A p) \geq q + A(q)(v - (K_p(u + v) - K_p(v)))\ ,$$

where $q = \exp_A(u - K_p(u) + \log_A p)$ If we take μ-integral of both sides,

$$0 \geq \int vA(q)\ d\mu - (K_p(u + v) - K_p(v)) \int A(q)\ d\mu\ .$$

Isolating the increment $K_p(u + v) - K_p(v)$, the desired inequality obtains. Therefore, $\widetilde{q}(u)$ is a sub-gradient of K_p at u. From Item 1. we deduce that $\widetilde{q}(u)$ is the unique sub-gradient and further $\widetilde{q}(u)$ is the Gateaux differential of K_p at u.

□

We can also prove a special Fréchet-differentiability as follows.

Proposition 5. *Let $\alpha \geq 2$.*

(1) *The superposition operator*

$$L^\alpha(\mu) \ni v \mapsto \exp_A(v + \log_A p) \in L^1(\mu)$$

is continuously Fréchet differentiable with derivative

$$d \exp_A(v) = (h \mapsto A(\exp_A(v + \log_A p))h) \in \mathcal{L}(L^\alpha(\mu), L^1(\mu))\ .$$

(2) *The functional $K_p : L^\alpha(\mu) \to \mathbb{R}$, implicitly defined by the equation*

$$\int \exp_A(v - K_p(v) + \log_A p)\ d\mu = 1, \quad v \in L^\alpha(\mu)$$

is continuously Fréchet differentiable with derivative

$$dK_p(v) = (h \mapsto \int h\widetilde{q}(v)\ d\mu)\ ,$$

where $\widetilde{q}(u) = \operatorname{escort}(q(u))$.

Proof. (1) Setting $\beta = 1$ in Prop. 1.2, we get easily the assertion. It remains just to check that the Fréchet derivative is continuous i.e., that the Fréchet derivative is a continuous map $L^\alpha(\mu) \to \mathcal{L}(L^\alpha(\mu), L^1(\mu))$. If $\|h\|_{L^\alpha(\mu)} \leq 1$ and $v, w \in L^\alpha(\mu)$ we have

$$\int |(A[\exp_A(v + \log_A p)] - A[\exp_A(w + \log_A p)])h|\ d\mu$$
$$\leq \|A[\exp_A(v + \log_A p) - A[\exp_A(w + \log_A p)]\|_{L^\sigma(\mu)}\ ,$$

where $\sigma = \alpha / (\alpha - 1)$ is the conjugate exponent of α. On the other hand,

$$\begin{aligned} & \|A[\exp_A(v + \log_A p) - A[\exp_A(w + \log_A p)]\|_{L^\sigma(\mu)} \\ \leq\ & \|A'\|_\infty \|A\|_\infty \|v - w\|_{L^\sigma(\mu)} \end{aligned}$$

and so the map $L^\alpha(\mu) \to \mathcal{L}(L^\alpha(\mu), L^1(\mu))$ is continuous whenever $\alpha \geq \sigma$, i.e., $\alpha \geq 2$.

(2) Fréchet differentiability of K_p is a consequence of the Implicit Function Theorem in Banach spaces, see [5], applied to the C^1-mapping

$$L^\alpha(\mu) \times \mathbb{R} \ni (v, \kappa) \mapsto \int \exp_A(v - \kappa + \log_A p)\ d\mu\ .$$

The value of the derivative is given by Proposition 4.

□

4. A-Divergence

In analogy with the standard exponential case, define the A-divergence between probability densities as

$$D_A(q\|p) = \int (\log_A q - \log_A p) \operatorname{escort}(q)\ d\mu, \ \text{ for } q, p \in \mathcal{P}\ .$$

Let us check that D_A is well defined that is, $(\log_A q - \log_A p)$ is quasi-integrable. As $\log_A$ is strictly concave with derivative $1/A$ we have

$$\log_A(x) \leq \log_A(y) + \frac{1}{A(y)}(x - y)$$

for all $x, y > 0$ and with equality if, and only if, $x = y$. Hence

$$A(y)(\log_A(y) - \log_A(x)) \geq y - x\ . \tag{8}$$

It follows in particular that

$$A(y)(\log_A y - \log_A x) \geq -|y - x|$$

hence the quasi-integrability is proved and $D_A(\cdot\|\cdot)$ is a well defined, possibly extended valued, function.

Observe further that by Prop. 2.4, if $q = \exp_A(u - K_p(u) + \log_A p)$, then $\log_A q - \log_A p \in L^1(\mu)$, and so $D_A(q\|p) < \infty$.

The binary relation D_A satisfies Gibbs' inequality hence it is a faithful divergence.

Proposition 6. *We have $D_A(q\|p) \geq 0$ and $D_A(q\|p) = 0$ if and only if $p = q$.*

Proof. From inequality (8) it follows

$$\begin{aligned} D_A(q\|p) &= \frac{1}{\int A(q)\,d\mu} \int (\log_A q - \log_A p)\, A(q)\; d\mu \\ &\geq \frac{1}{\int A(q)\,d\mu} \int (q-p) \quad d\mu = 0. \end{aligned}$$

Moreover, equality holds if and only if $p = q$ μ-a.e. ☐

Now we give a variational formula in the spirit of the classical one by Donsker-Varadhan. In equation

(9) $$q = \exp_A(u - K_p(u) + \log_A p), \quad u \in L^1(\mu)\ ,\ q \in \mathcal{P}\ ,$$

the random variable u is identified up to a constant for any given q. There are at least two options for selecting an interesting representative in the equivalence class.

One option is to assume $\int u\widetilde{p}\,d\mu = 0$ with $\widetilde{p} = \operatorname{escort}(p)$, the integral being well defined as the escort density is bounded. Such a choice is that used in the construction of the non-parametric exponential manifold, see [16, 15]. In this case we can solve Eq. (9) for $u - K(u)$ to get

(10) $$K_p(u) = \mathrm{E}_{\widetilde{p}}[\log_A p - \log_A q] = D_A(p\|q),$$

with $\mathrm{E}_{\widetilde{p}}[u] = 0$ and $q = \exp_A(u - K_p(u) + \log_A p)$.

A second option is to assume in Eq. (9) the random variable u to be centered with respect to $\widetilde{q} = \operatorname{escort}(q)$, i.e., $\mathrm{E}_{\widetilde{q}}[u] = 0$. This representation is of special interest in Statistical Physics, see for example [9].

To avoid confusion we rewrite Eq. (9) as

(11) $$q = \exp_A(v + H_p(v) + \log_A p), \quad v \in L^1(\mu), \quad \mathrm{E}_{\widetilde{q}}[v] = 0,$$

so that

$$D_A(q\|p) = \mathrm{E}_{\widetilde{q}}[\log_A q - \log_A p] = H_p(v),$$

where $\mathrm{E}_{\widetilde{q}}[v] = 0$.

In conclusion, we have two notable representation of the same probability density q, namely

$$\exp_A(u - K_p(u) + \log_A p) = \exp_A(v + H_p(v) + \log_A p)$$

which implies $u - v = K_p(u) + H_p(v)$. This, in turn, implies

(12) $$-\mathrm{E}_{\widetilde{p}}[v] = \mathrm{E}_{\widetilde{q}}[u] = K_p(u) + H_p(v).$$

The previous discussion is actually related to the computation of the convex conjugate of K_p in the duality $L^\infty(\mu) \times L^1(\mu)$ as we see now. Let us denote by $\overline{\mathcal{P}}$ the set of all probability densities that is, the closure in $L^1(\mu)$ of $\mathcal{P}$. The operator $\eta \mapsto \hat{\eta}$ denotes the inverse of the escort operator that is, $\eta = \operatorname{escort}(\hat{\eta})$, see Sec. 3.1.

Proposition 7. (1) *The convex conjugate mapping of K_p,*

(13) $$K_p^*(w) = \sup_{u \in L^1(\mu)} \left(\int wu\; d\mu - K_p(u) \right), \quad w \in L^\infty(\mu)$$

has domain contained into $\overline{\mathcal{P}} \cap L^\infty(\mu)$.

(2) *At each η in the image of the escort mapping, that is $\eta = \operatorname{escort}(\hat{\eta}) = dK_p(v)$, with $\hat{\eta} = q(v) = \exp_A(v - K_p(v) + \log_A p)$, the conjugate $K_p^*(\eta)$ is given by the* Legendre transform*,*

$$K_p^*(\eta) = \int v \ \operatorname{escort}(q(v))\; d\mu - K_p(v)\ ,$$

so that $K_p^*(\eta) = H_p(v) = D_A(q(u)\|p)$. In particular, K_p^* is finite on the image of the escort mapping.

Proof. (1) It follows from the fact that K_p is monotone and translation invariant. Actually, from the definition in Eq. (13) it follows

$$K_p^*(w) \geq \sup_{u \in L^1(\mu), u \leq 0} \left(\int wu \, d\mu - K_p(u) \right) \geq \sup_{u \in L^1(\mu), u \leq 0} \int wu \, d\mu$$

since $K_p(u) \leq 0$ if $u \leq 0$. If w is not positive, then there exists an element $u_0 \leq 0$ such that $\int wu_0 \, d\mu > 0$. Hence $K_p^*(w) = +\infty$. Now consider the case $w \geq 0$ and $u = \lambda \in \mathbb{R}$, $\lambda > 0$. We have $K_p(\lambda) = \lambda$ and

$$K_p^*(w) \geq \sup_{\lambda > 0} \left(\lambda \int w \, d\mu - \lambda \right) ,$$

which is $+\infty$ unless $\int w \, d\mu = 1$. Summarizing, $K_p^*(w) < \infty$ implies $w \in \overline{\mathcal{P}}$ i.e., the domain of K_p^* is contained in $\mathcal{P} \cap L^\infty(\mu)$.

(2) The concave and Gateaux differentiable function $u \mapsto \int \eta u \, d\mu - K_p(u)$ has derivative at u is $\eta - dK_p(u) = \eta - \operatorname{escort}(q(u))$ with $q(u) = \exp_A(u - K_p(u) + \log_p)$. As $\eta = \operatorname{escort}(q(v))$ by assumption, the derivative is zero at v and the sup in the definition of K_p^* is attained at that point. The value is $K_p^*(\eta) = \int \eta v \, d\mu - K_p(v)$.

□

Notice that, given any $\eta \in \overline{\mathcal{P}} \cap L^\infty(\mu)$ and $\epsilon > 0$, there exist a simple $\eta_\epsilon \in \mathcal{P} \cap L^\infty(\mu)$ such that $\|\eta - \eta_\epsilon\|_\infty < \epsilon$. Now, η_ϵ belongs to the image of the escort mapping because of Prop. 3.2, hence $K_p^*(\eta_\epsilon) < \infty$ so that the uniform closure of the image of the escort mapping is the full $\overline{\mathcal{P}} \cap L^\infty(\mu)$.

5. Hilbert bundle based on $\exp_A$

We discuss now the Hilbert manifold of probability densities as defined in [13]. With respect to that reference, we consider a slightly more general set-up. We use a general A function, provide an atlas of charts, and define a linear bundle as an expression of the tangent space.

Let $\mathcal{P}(\mu)$ denote the set of all μ-densities on the probability space $(\mathbb{X}, \mathcal{X}, \mu)$ of the kind

$$q = \exp_A(u - K_1(u)), \quad u \in L^2(\mu), \quad \mathrm{E}_\mu[u] = 0 \; . \tag{14}$$

Notice that $1 \in \mathcal{P}(\mu)$ because we can take $u = 0$.

Proposition 8.

(1) *$\mathcal{P}(\mu)$ is the set of all densities q such that $\log_A q \in L^2(\mu)$, in which case $u = \log_A q - \mathrm{E}_\mu[\log_A q]$.*

(2) *If $A'(0) > 0$, then $\mathcal{P}(\mu)$ is the set of all densities q such that both q and $\log q$ are in $L^2(\mu)$.*

(3) *Assume $A'(0) > 0$. On a product space with reference probability measures μ_1 and μ, and densities respectively q_1 and q_2, it holds $(q_1 \cdot \mu_1) \otimes (q_2 \cdot \mu_2) = (q_1 \otimes q_2) \cdot (\mu_1 \otimes \mu_2)$. Moreover, $q_1 \in \mathcal{P}(\mu_1)$ and $q_2 \in \mathcal{P}(\mu_2)$ if, and only if, $(q_1 \otimes q_2) \in \mathcal{P}(\mu_1 \otimes \mu_2)$.*

Proof. (1) If Eq. (14) holds, then $\log_A q = u - K_1(q) \in L^2(\mu)$. Conversely, if $v = \log_A q \in L^2(\mu)$, then Prop. (2) implies

$$q = \exp_A(v) = \exp_A(v - c - K_1(v - c)), \quad c \in \mathbb{R}$$

and we can take $c = \mathrm{E}_\mu[v]$ to satisfy Eq. (14) with $u = \log_A q - \mathrm{E}_\mu[\log_A q]$.

(2) Write $|\log_A q|^2 = |\log_A q|^2 (q < 1) + |\log_A q|^2 (q \geq 1)$, and use the bounds in Eq. (3) and Eq. (4) to get

$$\mathrm{E}_\mu\left[|\log_A q|^2\right] \leq \frac{1}{\alpha_1^2} \mathrm{E}_\mu\left[|\log q|^2 (q < 1)\right] + \mathrm{E}_\mu\left[|q - 1|^2 (q \geq 1)\right] \leq \frac{1}{\alpha_1^2} \mathrm{E}_\mu\left[|\log q|^2\right] + \mathrm{E}_\mu\left[q^2\right] - 1 \; .$$

By using the other two bounds, we get

$$\mathrm{E}_\mu\left[|\log_A q|^2\right] \geq \frac{1}{\alpha_2^2}\mathrm{E}_\mu\left[|\log q|^2\right] + \alpha_1(\mathrm{E}_\mu\left[q^2\right] - 1) \ .$$

(3) We use the previous item. $q_1 \otimes q_2 \in \mathcal{P}(\mu_1 \otimes \mu_2)$ if and only if both $q_1 \otimes q_2$ and $\log(q_1 \otimes q_2)$ are in $L^2(\mu_1 \otimes \mu_2)$. The first condition is equivalent to both $q_1 \in L^2(\mu_1)$ and $q_2 \in L^2(\mu_2)$. The second condition is $\log q_1 + \log q_2 \in L^2(\mu_1 \otimes \mu_2)$. We have

$$\mathrm{E}_{\mu_1\otimes\mu_2}\left[(\log q_1 + \log q_2)^2\right] = \mathrm{E}_{\mu_1}\left[(\log q_1)^2\right] + \mathrm{E}_{\mu_2}\left[(\log q_2)^2\right] + 2\left|\mathrm{E}_{\mu_1}\left[\log q_1\right]\right|\left|\mathrm{E}_{\mu_2}\left[\log q_2\right]\right|$$

because $\mathrm{E}_{\mu_i}\left[\log q_i\right] \leq \mathrm{E}_{\mu_1}\left[q_i - 1\right] = 0$. It follows that the second condition is equivalent to $\log q_1 \in L^2(\mu_1)$ and $\log q_2 \in L^2(\mu_2)$.

□

We proceed now to define an Hilbert bundle with base $\mathcal{P}(\mu)$. The notion of Hilbert bundle has been introduced in Information Geometry by [1]. We use here an adaptation to the A-exponential of arguments elaborated by [7, 14]. Notice that the construction depends in a essential way on the special conditions we are assuming for the present class of deformed exponential.

At each $q \in \mathcal{P}(\mu)$ the escort density $\widetilde{q}$ is bounded, so that we can define the fiber given by the Hilbert spaces

$$H_q = \left\{u \in L^2(\mu) | \mathrm{E}_{\widetilde{q}}\left[u\right] = 0\right\}$$

with scalar product $\langle u, v\rangle_q = \int uv \ d\mu$. The Hilbert bundle is

$$H\mathcal{P}(\mu) = \left\{(q, u) | q \in \mathcal{P}(\mu), u \in H_q\right\} \ .$$

For each $p, q \in \mathcal{P}(\mu)$ the mapping $\mathbb{U}_p^q u = u - \mathrm{E}_{\widetilde{q}}\left[u\right]$ is a continuous linear mapping from H_p to H_q. We have $\mathbb{U}_q^r \mathbb{U}_p^q = \mathbb{U}_p^r$. In particular, $\mathbb{U}_q^p \mathbb{U}_p^q$ is the identity on H_p, hence $\mathbb{U}_p^q$ is an isomorphism of H_p onto H_q.

In the following proposition we introduce an affine atlas of charts and use it to define our Hilbert bundle which is an expression of the tangent bundle. The velocity of a curve $t \mapsto p(t) \in \mathcal{P}(\mu)$ is expressed in the Hilbert bundle by the so called A-score that, in our case, takes the form $A(p(t))^{-1}\dot{p}(t)$, with $\dot{p}(t)$ computed in $L^1(\mu)$.

The following proposition is taken from [11] where a detailed proof is given.

Proposition 9. (1) *Fix $p \in \mathcal{P}(\mu)$. A positive density q can be written as*

$$q = \exp_A(u - K_p(u) + \log_A p), \ \text{with } u \in L^2(\mu) \text{ and } \mathrm{E}_{\widetilde{p}}\left[u\right] = 0,$$

if, and only if, $q \in \mathcal{P}(\mu)$.

(2) *For each $p \in \mathcal{P}(\mu)$ the mapping*

$$s_p \colon \mathcal{P}(\mu) \ni q \mapsto \log_A q - \log_A p + D_A(p\|q) \in H_p$$

is injective and surjective, with inverse $e_p(u) = \exp_A(u - K_p(u) + \log_A p)$.

(3) *The atlas $\{s_p | p \in \mathcal{P}(\mu)\}$ is affine with transitions*

$$s_q \circ e_p(u) = \mathbb{U}_p^q u + s_p(q) \ .$$

(4) *The expression of the velocity of the differentiable curve $t \mapsto p(t) \in \mathcal{P}(\mu)$ in the chart s_p is $ds_p(p(t))/dt \in H_p$. Conversely, given any $u \in H_p$, the curve*

$$p \colon t \mapsto \exp_A(tu - K_p(tu) + \log_A p)$$

has $p(0) = p$ and has velocity at $t = 0$ expressed in the chart s_p by u. If the velocity of a curve is expressed in the chart s_p by $t \mapsto \dot{u}(t)$, then its expression in the chart s_q is $\mathbb{U}_p^q \dot{u}(t)$.

(5) *If $t \mapsto p(t) \in \mathcal{P}(\mu)$ is differentiable with respect to the atlas then it is differentiable as a mapping in $L^1(\mu)$. It follows that the A-score is well-defined and is the expression of the velocity of the curve $t \mapsto p(t)$ in the moving chart $t \mapsto s_{p(t)}$.*

6. Final remarks

A non-parametric Hilbert manifold based on a deformed exponential representation of positive densities has been firstly introduced by N. J. Newton [13]. We have derived regularity properties of the normalizing functional K_p and discussed the relevant Fenchel conjugation. With respect to the original version, we allow for an atlas containing charts centered at each density in the model. Moreover, we discuss explicitly the Hilbert bundle on the Hilbert manifold. Though K_p is a convex function, it should be remarked we do not follow the standard development that uses it as a potential function to derive a Fisher metric from its Hessian.

Acknowledgement. L. Montrucchio is Honorary Fellow of the Collegio Carlo Alberto Foundation. G. Pistone is a member of GNAMPA-INdAM and acknowledges the support of de Castro Statistics and Collegio Carlo Alberto.

References

[1] Amari, S.: Dual connections on the Hilbert bundles of statistical models. In: Geometrization of statistical theory (Lancaster, 1987), pp. 123–151. ULDM Publ. (1987)

[2] Ambrosetti, A., Prodi, G.: A primer of nonlinear analysis, *Cambridge Studies in Advanced Mathematics*, vol. 34. Cambridge University Press (1993)

[3] Appell, J., Zabrejko, P.P.: Nonlinear superposition operators, *Cambridge Tracts in Mathematics*, vol. 95. Cambridge University Press (1990). DOI 10.1017/CBO9780511897450. URL `http://dx.doi.org/10.1017/CBO9780511897450`

[4] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry, *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*, vol. 64. Springer, Cham (2017)

[5] Dieudonné, J.: Foundations of Modern Analysis. Academic press (1960)

[6] Ekeland, I., Témam, R.: Convex analysis and variational problems, *Classics in Applied Mathematics*, vol. 28, english edn. Society for Industrial and Applied Mathematics (SIAM) (1999). DOI 10.1137/1.9781611971088. URL `http://dx.doi.org/10.1137/1.9781611971088`. Translated from the French

[7] Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. IDAQP **1**(2), 325–347 (1998)

[8] Kaniadakis, G.: Non-linear kinetics underlying generalized statistics. Physica A **296**(3-4), 405–425 (2001)

[9] Landau, L.D., Lifshitz, E.M.: Course of theoretical physics. Vol. 5: Statistical physics. Translated from the Russian by J. B. Sykes and M. J. Kearsley. Second revised and enlarged edition. Pergamon Press, Oxford-Edinburgh-New York (1968)

[10] Malliavin, P.: Integration and probability, *Graduate Texts in Mathematics*, vol. 157. Springer-Verlag (1995). With the collaboration of Hlne Airault, Leslie Kay and Grard Letac, Edited and translated from the French by Kay, With a foreword by Mark Pinsky

[11] Montrucchio, L., Pistone, G.: Deformed exponential bundle: the linear growth case. In: F. Nielsen, F. Barbaresco (eds.) Geometric Science of Information, pp. 239–246. Springer (2017). Third International Conference, GSI 2017, Paris, France, November 7-9, 2017, Proceedings

[12] Naudts, J.: Generalised thermostatistics. Springer-Verlag London Ltd. (2011). DOI 10.1007/978-0-85729-355-8. URL `http://dx.doi.org/10.1007/978-0-85729-355-8`

[13] Newton, N.J.: An infinite-dimensional statistical manifold modelled on Hilbert space. J. Funct. Anal. **263**(6), 1661–1681 (2012). DOI 10.1016/j.jfa.2012.06.007. URL `http://dx.doi.org/10.1016/j.jfa.2012.06.007`

[14] Pistone, G.: Nonparametric information geometry. In: F. Nielsen, F. Barbaresco (eds.) Geometric science of information, *Lecture Notes in Comput. Sci.*, vol. 8085, pp. 5–36. Springer, Heidelberg (2013). First International Conference, GSI 2013 Paris, France, August 28-30, 2013 Proceedings

[15] Pistone, G., Rogantin, M.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernoulli **5**(4), 721–760 (1999)

[16] Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. Ann. Statist. **23**(5), 1543–1561 (1995)

[17] Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. J. Statist. Phys. **52**(1-2), 479–487 (1988)

[18] Vigelis, R.F., Cavalcante, C.C.: On ϕ-families of probability distributions. Journal of Theoretical Probability **26**, 870–884 (2013)

L. Montrucchio: Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy

G. Pistone: de Castro Statistics, Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy

E-mail address: `luigi.montrucchio@unito.it, giovanni.pistone@carloalberto.org`