AISM (2007) 59:27–56
DOI 10.1007/s10463-006-0096-y

Exponential statistical manifold

Alberto Cena · Giovanni Pistone

Received: 2 May 2006 / Revised: 25 September 2006 /
Published online: 16 December 2006

Abstract We consider the non-parametric statistical model $\mathcal{E}(p)$ of all positive densities q that are connected to a given positive density p by an open exponential arc, i.e. a one-parameter exponential model $p(t)$, $t \in I$, where I is an open interval. On this model there exists a manifold structure modeled on Orlicz spaces, originally introduced in 1995 by Pistone and Sempi. Analytic properties of such a manifold are discussed. Especially, we discuss the regularity of mixture models under this geometry, as such models are related with the notion of e- and m-connections as discussed by Amari and Nagaoka.

Keywords Information geometry · Statistical manifold · Orlicz space · Moment generating functional · Cumulant generating functional · Kullback–Leibler divergence

1 Introduction

1.1 Content

In the present paper we follow closely the discussion of Information Geometry developed by Amari and coworkers, see e.g. in Amari (1982), Amari (1985), Amari and Nagaoka (2000), with the specification that we want to construct a Banach manifold structure in the classical sense, see e.g. Bourbaki (1971) or Lang (1995), without any restriction to parametric models.

A. Cena · G. Pistone (✉)
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
e-mail: giovanni.pistone@polito.it

A. Cena
e-mail: alberto.cena@gmail.com

We build on the previous work on the theory of Statistical Manifolds modeled on Orlicz spaces as defined in Pistone and Sempi (1995), Pistone and Rogantin (1999), Gibilisco and Pistone (1998) and the unpublished PhD thesis Cena (2002). We give a number of new results and improvements of the basic theory, together with results from Cena (2002). Concerning more advanced topics, a related paper, currently in progress, will discuss fiber bundles and the theory of Amari's α-connections in the same framework. Some of the material presented here in detail was presented by the second author, without proofs, in the IGAIA 2005 Tokyo Conference, see Pistone (2005).

The rest of this introductory section contains a review of relevant facts related with the topology of Orlicz spaces.

Section 2 is devoted to the properties of the moment generating functional and the cumulant generating functional as defined and discussed in Pistone and Sempi (1995), Pistone and Rogantin (1999). The statement and proof of the analyticity of both functionals is taken from Cena (2002) and published here for the first time. Most of this section is devoted to a detailed discussion of the exponential function as an analytic mapping. The main point is to explain why in this case the radius of convergence of the exponential Taylor series is finite. Finally, a new proof of Fréchet-differentiability follows from analyticity.

Section 3 contains a discussion of densities which are connected by open mixture and exponential arcs (one dimensional models); here 'open' means that the parameter takes values in an open real interval. Moreover, we define and discuss the properties of the maximal exponential model. In particular, the Kullback–Leibler divergence $D(q \,\|\, p) = \mathrm{E}_p\left[\frac{q}{p}\log\left(\frac{q}{p}\right)\right]$ takes a special form on the maximal exponential model. Apart from Theorem 19, which is taken from Pistone and Rogantin (1999) and it is reproduced here for the sake of clarity, the rest is new, at least in the present form.

The definition of the Exponential Statistical Manifold in Pistone and Sempi (1995) was based on a system of charts whose image was the open unit ball of an Orlicz space. In Sect. 4 we introduce a new equivalent definition, based on an extension of such charts. In fact, we show that the general exponential model $u \mapsto \mathrm{e}^{u-K_p(u)}p$, on the maximal open set where it is defined, is the inverse of a chart. However, the old definition still remains of interest, because some properties can be proved only on the smaller domains. The final part of the section is devoted to the topology of the exponential statistical manifold. It contains the adaptation to the new framework of what was discussed in the previous papers, in particular e-convergence and the disconnection of the exponential statistical manifold into maximal exponential models. Especially, Proposition 29 considers a simple case of an interesting question raised by Streater and concerning the compatibility of mixture models with the exponential structure.

Section 5 discusses the construction of a manifold supporting the mixture geometry. As we are unable to do that on positive densities, our manifold is supported by measurable functions with integral 1. The charts we use are a non-parametric version of the expectation parameterization for parametric exponential models. The relation with the divergence functional is discussed. Most of the results are published here for the first time.

In Sect. 7 we prove the C^∞-regularity of the divergence in both variables with the consequence that the Fisher information has the same regularity. The inclusion of the exponential structure into the L^2 structure was presented in Gibilisco and Pistone (1998). In Sect. 7 it is proved that the inclusion of the strictly positive densities $\mathcal{M}$ into the set $\mathcal{P}$ of random variables with unit integral is C^∞ for the relevant structures. The results in this section are new.

1.2 Notations

For all of this paper, we are given a probability space $(X, \mathcal{X}, \mu)$ and we will denote by $\mathcal{M}$ the set of all densities which are positive μ-a.s. The set $\mathcal{M}$ is thought to be the maximal regular statistical model. Here, 'regular' means that the support of each density in the model is the same X. We want to endow this maximal model with a manifold structure in such a way that each specific statistical model could be considered as a sub-manifold of $\mathcal{M}$. This ultimate goal is not fully achieved in the present paper, because it requires an unconventional definition of sub-manifold, cf. the discussion in Sect. 7.

Locally, at each $p \in \mathcal{M}$, the model space for the manifold is an Orlicz space of centered random variables. We refer to Lang (1995) for the theory of manifolds modeled on Banach spaces and to Krasnosel'skii and Rutickii (1961) and Rao and Ren (2002) for the theory of Orlicz spaces. The rest of this sub-section contains a compact review of these matters.

The Young function $\Phi_1(x) = \cosh x - 1$ is used here instead of the equivalent and more commonly used Young function $\Phi_2(x) = \mathrm{e}^{|x|} - |x| - 1$. Its conjugate Young function is denoted by $\Psi_1(y) = \int_0^y \sinh^{-1}(s)\, \mathrm{d}s$. It is the smallest function such that $\Psi_1(y) \geq xy - \Phi_1(x)$. The function $y \mapsto (1+y)\log(1+y) - y$, $y \geq 0$, is equivalent to $\Psi_1(y)$.

A real random variable u belongs to the vector space $L^{\Phi_1}(p)$ if $\mathrm{E}_p[\Phi_1(\alpha u)] < +\infty$ for some $\alpha > 0$. This space is a Banach space when endowed with the Luxemburg norm $\|\cdot\|_{\Phi_1,p}$; this norm is defined by the assumption that its closed unit ball consists of all u's such that $\mathrm{E}_p[\Phi_1(u)] \leq 1$. Then, the open unit ball $B(0,1)$ consists of those u's such that αu is in the closed unit ball for some $\alpha > 1$. A sequence u_n, $n = 1, 2, \ldots$ is convergent to 0 for such a norm if and only if for all $\epsilon > 0$ there exists a $n(\epsilon)$ such that $n > n(\epsilon)$ implies $\mathrm{E}_p\left[\Phi_1(\frac{u_n}{\epsilon})\right] \leq 1$. Note that $|u| \leq |v|$ implies

$$\mathrm{E}_p\left[\Phi_1\left(\frac{u}{\|v\|_{\Phi_1,p}}\right)\right] \leq \mathrm{E}_p\left[\Phi_1\left(\frac{v}{\|v\|_{\Phi_1,p}}\right)\right] \leq 1$$

so that $\|u\|_{\Phi_1,p} \leq \|v\|_{\Phi_1,p}$.

As we systematically connect Orlicz spaces defined at different points of statistical models, we will use frequently the following lemma. It is a slight but important improvement upon the similar statements in Pistone and Rogantin (1999) and Cena (2002).

Lemma 1 *Let $p, q \in \mathcal{M}$ and let Φ_0 be a Young function. If the Orlicz spaces $L^{\Phi_0}(p)$ and $L^{\Phi_0}(q)$ are equal as sets, then their norms are equivalent.*

Proof Note that p and q have to be equivalent, otherwise the two Orlicz spaces cannot be equal as sets. We prove the equivalence of norms by a standard argument for function spaces. Let $\{u_n\} \subset L^{\Phi_0}(p)$ be a sequence converging in norm to 0. Let us assume it does not converge in the norm of $L^{\Phi_0}(q)$. By possibly considering a subsequence, we suppose $\|u_n\|_{\Phi_0,q} > \epsilon$ for some $\epsilon > 0$. By the convergence in norm $\|\cdot\|_{\Phi_0,p}$, there is a sub-sequence $\{u_{n_k}\}$ such that

$$\|u_{n_k}\|_{\Phi_0,p} < \frac{1}{2^k}.$$

Since

$$\sum_{k=1}^{\infty} \|k u_{n_k}\|_{\Phi_0,p} \leq \sum_{k=1}^{\infty} \frac{k}{2^k} < \infty$$

and $\left(L^{\Phi_0}(p), \|\cdot\|_{\Phi_0,p}\right)$ is complete, the series $\sum k|u_{n_k}|$ converges in norm to $r \in L^{\Phi_0}(p)$. Moreover, the partial sums $r_m = \sum_{k=1}^{m} k \left|u_{n_k}\right|$ are increasing and convergent to r a.s., then $r_m \leq r$, or $|u_{n_k}| \leq \frac{r}{k}$. As r is in $L^{\Phi_0}(q)$ by the equality of the vector spaces, then $\frac{r}{k} \to 0$ in $L^{\Phi_0}(q)$ and $\lim_{k\to\infty} u_{n_k} = 0$ in $L^{\Phi_0}(q)$, contradicting the assumption. Hence the identity map is continuous and, by symmetry, it is an homeomorphism, i.e. the norms are equivalent. □

The condition $u \in L^{\Phi_1}(p)$ is equivalent to the existence of the moment generating function $g(t) = \mathrm{E}_p\left[\mathrm{e}^{tu}\right]$ on a neighborhood of 0. The case when such a moment generating function is defined on all of the real line is special, as the following lemma shows. In fact, the space of bounded random variables is not dense in the space $L^{\Phi_1}(p)$, cf. Rao and Ren (2002). Especially, the sequence of truncations of a generic element of such a space does not converge in general to the original variable. On the other side, the lemma below characterizes in terms of the moment generating function those random variables for which the truncations converge. The restriction of our construction to such a class of random variables in order to save the density of the class of bounded random variables in the model space would considerably restrict the scope of our definition of manifold. The opposite option has been discussed and developed in detail in Grasselli (2001).

Lemma 2 *Let $u \in L^{\Phi_1}(p)$ and let $u_n = (|u| \leq n)u$ be its sequence of truncations at n. Then $\lim_{n\to\infty} \|u - u_n\|_{\Phi_1,p} = 0$ if and only if the moment generating function g of u is defined on $\mathbb{R}$.*

Proof If the moment generating function of u is defined on $\mathbb{R}$, then for $\epsilon > 0$

$$\Phi_1\left(\frac{u-u_n}{\epsilon}\right) = \Phi_1\left(\frac{(|u|>n)u}{\epsilon}\right) \leq \Phi_1\left(\frac{u}{\epsilon}\right),$$

where the left-end side goes a.s. to 0, while the right-end side is integrable. Vice-versa,

$$\begin{aligned} \mathrm{E}_p\left[\Phi_1\left(\frac{u}{\epsilon}\right)\right] &= \mathrm{E}_p\left[\Phi_1\left(\frac{u}{\epsilon}\right)(|u|\leq n)\right] + \mathrm{E}_p\left[\Phi_1\left(\frac{u}{\epsilon}\right)(|u|>n)\right] \\ &\leq \Phi_1\left(\frac{n}{\epsilon}\right) + \mathrm{E}_p\left[\Phi_1\left(\frac{u(|u|>n)}{\epsilon}\right)\right], \end{aligned}$$

where the last term is lesser or equal to 1 for some n. □

In fact, the Banach space $L^{\Phi_1}(p)$ is not separable, unless the basic space has a finite number of atoms. In this sense it is an un-natural choice from the point of view of functional analysis and manifold's theory. However, $L^{\Phi_1}(p)$ is natural for statistics because for each $u \in L^{\Phi_1}(p)$ the Laplace transform of u is well defined at 0, then the one-dimensional exponential model $p(\theta) \propto \mathrm{e}^{\theta u} \cdot p$ is well defined.

However, the space $L^{\Psi_1}(p)$ is separable and its dual space is $L^{\Phi_1}(p)$, the duality pairing being $(u, v) \mapsto \mathrm{E}_p[uv]$. This duality extends to a continuous chain of spaces:

$$L^{\Phi_1}(p) \sqsubset L^a(p) \sqsubset L^b(p) \sqsubset L^{\Psi_1}(p), \quad 1 < b \leq 2, \ \frac{1}{a} + \frac{1}{b} = 1,$$

where $\sqsubset$ denotes continuous injection.

From the duality pairing of conjugate Orlicz spaces and the characterization of the closed unit ball it follows a definition of a dual norm on $L^{\Psi_1}(p)$:

$$N_p(v) = \sup\left\{\mathrm{E}_p[uv] \mid \mathrm{E}_p[\Phi_1(u)] \leq 1\right\}.$$

This norm is equivalent to the Ψ_1-Luxemburg norm.

2 Moment generating functional and cumulant generating functional

Let $p \in \mathcal{M}$ be given. We shall define an analytic function between the open unit ball $B(0,1)$ of the Orlicz space $L^{\Phi_1}(p)$ and the Lebesgue space $L^a(p)$, $a \geq 1$. We refer to Upmeier (1985) for the relevant theory of Banach valued analytic functions.

Lemma 3 *For each $a \geq 1$, $n \in \mathbb{N}^*$ and $u \in B(0,1)$, let $\lambda_{a,n}(u)$ be defined by*:

$$\lambda_{a,n}(u) : \begin{cases} L^{\Phi_1}(p) \times \cdots \times L^{\Phi_1}(p) \to L^a(p) \\ (w_1, \ldots, w_n) \mapsto \dfrac{w_1}{a} \cdots \dfrac{w_n}{a} \, \mathrm{e}^{\frac{u}{a}} \end{cases}$$

Then each $\lambda_{a,n}(u)$ is a continuous, symmetric, n-multi-linear map from $\left(L^{\Phi_1}(p)\right)^n$ to $L^a(p)$.

Proof Consider $r = (1 - \|u\|_{\Phi_1,p})/n$. Let $v_1, \ldots, v_n \in L^{\Phi_1}(p)$ with $\|v_i\|_{\Phi_1,p} = 1$. Then

$$\left\| u + r\sum_{i=1}^{n} |v_i| \right\|_{\Phi_1,p} \leq \|u\|_{\Phi_1,p} + rn = 1$$

that is $u + r\sum_{i=1}^{n}|v_i| \in \overline{B}(0,1)$, hence $\mathrm{E}_p\left[\cosh\left(u + r\sum_{i=1}^{n}|v_i|\right)\right] \leq 2$. Since $|rv_i|^a/a^a < \mathrm{e}^{|rv_i|}$ for $i = 1, \ldots, n$,

$$\begin{aligned}\mathrm{E}_p\left[\left|r\frac{v_1}{a}\right|^a \cdots \left|r\frac{v_n}{a}\right|^a \mathrm{e}^u\right] &\leq \mathrm{E}_p\left[\mathrm{e}^{u + r\sum_{i=1}^{n}|v_i|}\right] \\ &\leq 2\mathrm{E}_p\left[\cosh\left(u + r\sum_{i=1}^{n}|v_i|\right)\right] \leq 4,\end{aligned}$$

hence

$$\mathrm{E}_p\left[\left|\frac{v_1}{a}\cdots\frac{v_n}{a}\mathrm{e}^{\frac{u}{a}}\right|^a\right] \leq \frac{4}{r^{na}} = \frac{4n^{na}}{\left(1 - \|u\|_{\Phi_1,p}\right)^{na}} = C_n(u) < \infty.$$

For each $w_1, \ldots, w_n \in L^{\Phi_1}(p)\setminus\{0\}$, if define $v_i = w_i/\|w_i\|_{\Phi_1,p}$,

$$\mathrm{E}_p\left[\left|\frac{w_1}{a}\cdots\frac{w_n}{a}\mathrm{e}^{\frac{u}{a}}\right|^a\right] = \mathrm{E}_p\left[\left|\frac{v_1}{a}\cdots\frac{v_n}{a}\mathrm{e}^{\frac{u}{a}}\right|^a\right]\|w_1\|^a_{\Phi_1,p}\cdots\|w_n\|^a_{\Phi_1,p}.$$

Then $\lambda_{a,n}(u)$ is a continuous multi-linear symmetric function, since

$$\begin{aligned}\left\|\lambda_{a,n}(u)\cdot(w_1,\ldots,w_n)\right\|_a &= \left(\mathrm{E}_p\left[\left|\frac{w_1}{a}\cdots\frac{w_n}{a}\mathrm{e}^{\frac{u}{a}}\right|^a\right]\right)^{\frac{1}{a}} \\ &\leq (C_n(u))^{\frac{1}{a}}\|w_1\|_{\Phi_1,p}\cdots\|w_n\|_{\Phi_1,p}.\end{aligned} \tag{1}$$

□

If $u \in B(0,1)$, then $\mathrm{e}^{u/a} \in L^a(p)$. In fact, since $\|u\|_{\Phi_1,p} < 1$,

$$\mathrm{E}_p\left[\left|\mathrm{e}^{\frac{u}{a}}\right|^a\right] \leq \mathrm{E}_p\left[\mathrm{e}^{|u|}\right] \leq 2\mathrm{E}_p[\cosh|u|] \leq 2\mathrm{E}_p\left[\cosh\frac{|u|}{\|u\|_{\Phi_1,p}}\right] \leq 4.$$

Definition 4 *For each $a \geq 1$, $n \in \mathbb{N}$ and $u \in B(0,1)$, we can define the* continuous n-homogeneous polynomial $\hat{\lambda}_{a,n}(u)$ of degree *n from $L^{\Phi_1}(p)$ to $L^a(p)$ as follows*: *if $n = 0$*

$$\hat{\lambda}_{a,0}(u) := \mathrm{e}^{\frac{u}{a}}$$

otherwise $\hat{\lambda}_{a,n}(u)$ *is determined by its polar form* $\lambda_{a,n}(u)$*,*

$$\hat{\lambda}_{a,n}(u)\cdot w = \lambda_{a,n}(u)\cdot(w,\ldots,w) \quad \forall\, w \in L^{\Phi_1}(p)\,.$$

The radius of convergence of the Taylor series of the real function $\mathrm{e}^{u/a}$ is $+\infty$. However, the composition operator $u \mapsto \mathrm{e}^{u/a}$ could have a restricted domain and a finite radius of convergence as a map between Banach spaces.

Lemma 5 *Let* $a \geq 1$*, then*

$$A(v):\begin{cases} L^{\Phi_1}(p) \to L^{a}(p) \\ v \mapsto \displaystyle\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{v}{a}\right)^{n} \end{cases}$$

is a power series from $L^{\Phi_1}(p)$ *to* $L^{a}(p)$ *with radius of convergence* $\hat{\rho} \geq 1$*.*

Proof $A = \sum_{n=0}^{\infty}\frac{1}{n!}\hat{\lambda}_{a,n}(0)$ is the power series from $L^{\Phi_1}(p)$ to $L^{a}(p)$ defined by the n-homogeneous polynomials $\hat{\lambda}_{a,n}(0)$.

For each degree n, in the space $\mathcal{P}^{n}\left(L^{\Phi_1}(p), L^{a}(p)\right)$ of all the n-homogeneous polynomials from $L^{\Phi_1}(p)$ to $L^{a}(p)$ the norms of $\frac{1}{n!}\hat{\lambda}_{a,n}(0)$ are uniformly bounded. In fact, for each $v \in L^{\Phi_1}(p)$ with $\|v\|_{\Phi_1,p} = 1$, from $\frac{1}{n!}\left(\frac{|v|}{a}\right)^{n} < \mathrm{e}^{|v|/a}$ it follows that

$$\mathrm{E}_p\left[\left|\frac{1}{n!}\left(\frac{v}{a}\right)^{n}\right|^{a}\right] \leq \mathrm{E}_p\left[\mathrm{e}^{|v|}\right] \leq 2\mathrm{E}_p\left[\cosh|v|\right] \leq 4,$$

hence

$$\left\|\frac{1}{n!}\hat{\lambda}_{a,n}(0)\right\|_{\mathcal{P}^{n}} = \sup_{\|v\|_{\Phi_1,p}=1}\left\|\frac{1}{n!}\hat{\lambda}_{a,n}(0)\cdot v\right\|_{a} \leq 4^{\frac{1}{a}}.$$

By Cauchy–Hadamard Formula $1/\hat{\rho} = \limsup_{n\to\infty}\left\|\frac{1}{n!}\hat{\lambda}_{a,n}(0)\right\|_{\mathcal{P}^{n}}^{1/n}$, one can conclude that $1/\hat{\rho} \leq \lim_{n\to\infty} 4^{1/(na)} = 1$ and $\hat{\rho} \geq 1$. □

It was correctly pointed out by one of the anonymous Referee of this paper that this lemma does not tell if and in which cases the radius of convergence is actually $+\infty$ as in the ordinary exponential Taylor series. The following example shows a case with finite radius of convergence. Let $p(x) = \mathrm{e}^{-x}$ $(0 < x < \infty)$ be the exponential distribution and let us consider $A : L^{\Phi_1}(p) \to L^{1}(p)$. For each $\alpha \in \mathbb{R}$, the random variable $\alpha x \in L^{\Phi_1}(p)$ and $\|\alpha x\|_{\Phi_1,p} = |\alpha|\sqrt{2}$. As $\left\|\frac{(\alpha x)^{n}}{n!}\right\|_{L^{1}(p)} = |\alpha|^{n}$, if $|\alpha| \geq 1$, then the general term of the series $A(\alpha x)$ does not converge to 0 in $L^{1}(p)$. From this follows that $\hat{\rho} < \sqrt{2}$.

Definition 6 *For each $a \geq 1$ we call* exponential function *the mapping* $\exp_{p,a}$ *between the open unit ball of the Orlicz space $L^{\Phi_1}(p)$ and the Lebesgue space $L^a(p)$*

$$\exp_{p,a} : \begin{cases} B(0,1) \to L^a(p) \\ v \mapsto \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{v}{a}\right)^n \end{cases}$$

We remark that $\exp_{p,a}(v) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{v}{a}\right)^k$ converges absolutely in $L^a(p)$ for each $v \in B(0,1)$ and it converges uniformly for each v in the closed ball $\bar{B}(0,r)$ with $r < 1$. Moreover, the sequence of partial sums $s_n = \sum_{k=0}^{n} \frac{1}{k!} \left(\frac{v}{a}\right)^k$ converges to $\exp_{p,a}(v)$ also in probability and s_n converges a.s. and in probability to $\mathrm{e}^{v/a}$, so $\exp_{p,a}(v) = \mathrm{e}^{v/a}$ a.s..

Proposition 7 *The exponential function* $\exp_{p,a}$ *satisfies the following properties*:

(1) $\exp_{p,a}(0) = 1$;
(2) *for each $u, v \in B(0,1)$ such that $u + v \in B(0,1)$*

$$\exp_{p,a}(u+v) = \exp_{p,a}(u)\exp_{p,a}(v)\,;$$

(3) *for each $u \in B(0,1)$,* $\exp_{p,a}(u)$ *has an inverse $\left(\exp_{p,a}(u)\right)^{-1}$ in $L^a(p)$ and*

$$\left(\exp_{p,a}(u)\right)^{-1} = \exp_{p,a}(-u)\,.$$

Proof (1) $\exp_{p,a}(0) = \hat{\lambda}_{a,0}(0) = \mathrm{e}^0 = 1$.

(2) Since in a Banach space absolute convergence implies unconditional convergence, with a rearrangement of terms we have:

$$\begin{aligned} \exp_{p,a}(u+v) &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{m=0}^{n} \binom{n}{m} \left(\frac{u}{a}\right)^{n-m} \left(\frac{v}{a}\right)^m \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{u}{a}\right)^k \sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{v}{a}\right)^m = \exp_{p,a}(u)\exp_{p,a}(v)\,. \end{aligned}$$

(3) It follows from (1) and (2) if $v = -u$. □

Theorem 8 *Let $a \geq 1$. Mapping* $\exp_{p,a}$ *is an analytic function. In a neighborhood of each $u_0 \in B(0,1)$ it can be expanded in the Taylor series*

$$\exp_{p,a}(u) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{u-u_0}{a}\right)^n \mathrm{e}^{\frac{u_0}{a}}\,. \tag{2}$$

Proof Let $u_0 \in B(0,1)$ and consider the power series $\sum_{n=0}^{\infty} \frac{1}{n!} \hat{\lambda}_{a,n}(u_0)$. Power series are analytic in a ball $B(0,\rho)$ where ρ is the radius of restricted convergence. Using Inequality (1) of Lemma 3, one can determine the following bound for the norm of the continuous n-multi-linear mapping $\frac{1}{n!}\lambda_{a,n}(u_0)$:

$$\left\| \frac{1}{n!}\lambda_{a,n}(u_0) \right\|_{\mathcal{L}^n} \leq \frac{(C_n(u_0))^{1/a}}{n!}$$

where C_n is defined in the proof of Lemma 3. Then, by Cauchy–Hadamard Formula $1/\rho = \limsup_{n\to\infty} \left\| \frac{1}{n!}\lambda_{a,n}(u_0) \right\|_{\mathcal{L}_n}^{1/n}$, one can check $\rho \geq (1 - \|u_0\|_{\Phi_1,p})/\mathrm{e} = r$, that is the series $\sum_{n=0}^{\infty} \frac{1}{n!}\hat{\lambda}_{a,n}(u_0)$ is analytic in a ball with radius greater than r.
Also, for u such that $\|u - u_0\|_{\Phi_1,p} < r$

$$\exp_{p,a}(u) = \sum_{n=0}^{\infty} \frac{1}{n!}\hat{\lambda}_{a,n}(u_0) \cdot (u - u_0).$$

□

We are now able to improve the results of Pistone and Sempi (1995) about the regularity of the moment generating functional.

Definition 9 *The* moment generating functional *is*

$$M_p : \begin{cases} L^{\Phi_1}(p) \to \mathbb{R} \cup \{+\infty\} \\ u \mapsto \mathrm{E}_p\left[\mathrm{e}^u\right] \end{cases}$$

Note that M_p, if restricted to the open unit ball $B(0,1)$, coincides with the expected value at p of $\exp_{p,1}$.

Theorem 10 *The moment generating functional M_p satisfies the following properties*:

(1) $M_p(0) = 1$; *otherwise, for each centered random variable $u \neq 0$, $M_p(u) > 1$.*
(2) *M_p is convex and lower semi-continuous, and its proper domain*

$$\operatorname{dom} M_p = \left\{u \in L^{\Phi_1}(p) \mid M_p(u) < \infty\right\}$$

is a convex set which contains the open unit ball $B(0,1) \subset L^{\Phi_1}(p)$; in particular the interior of such a domain is a non empty convex set.
(3) *M_p is infinitely Gâteaux-differentiable in the interior of its proper domain, the nth-derivative at $u \in \overset{\circ}{\operatorname{dom}} M_p$ in the direction $v \in L^{\Phi_1}(p)$ being*

$$\left.\frac{\mathrm{d}^n}{\mathrm{d}t^n} M_p(u + tv)\right|_{t=0} = \mathrm{E}_p\left[v^n \mathrm{e}^u\right];$$

(4) *M_p is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of $L^{\Phi_1}(p)$, the nth-derivative at $u \in B(0,1)$ evaluated in $(v_1,\dots,v_n) \in L^{\Phi_1}(p) \times \dots \times L^{\Phi_1}(p)$ is*

$$\mathbf{D}^n M_p(u) \cdot (v_1,\dots,v_n) = \mathrm{E}_p\left[v_1 \cdots v_n \mathrm{e}^u\right].$$

In particular, $\mathbf{D} M_p(0) = \mathrm{E}_p$.

Proof (1), (2) and (3) are proved in Pistone and Sempi (1995).

(4) For each $u \in B(0,1)$ and $n \in \mathbb{N}$, we have $\mathrm{E}_p\left[\hat{\lambda}_{1,n}(u)\right] \in \mathcal{P}^n\left(L^{\Phi_1}(p);\mathbb{R}\right)$ and $\sum_{n=0}^{\infty} \frac{1}{n!}\mathrm{E}_p\left[\hat{\lambda}_{1,n}(u)\right]$ is a power series from $L^{\Phi_1}(p)$ to $\mathbb{R}$ with positive radius of convergence. In a neighborhood of each $u_0 \in B(0,1)$, by (2) we have

$$M_p(u) = \mathrm{E}_p\left[\exp_{p,1}(u)\right] = \int \sum_{n=0}^{\infty} \frac{1}{n!}\lambda_{1,n}(u_0)(u-u_0)^n\, p\mathrm{d}\mu.$$

Integrating term by term we obtain the following expansion in power series about u_0:

$$\begin{aligned} M_p(u) &= \sum_{n=0}^{\infty} \frac{1}{n!}\mathrm{E}_p\left[\hat{\lambda}_{1,n}(u_0)(u-u_0)\right] \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}\mathrm{E}_p\left[(u-u_0)^n \mathrm{e}^{u_0}\right]. \end{aligned}$$

Hence M_p is an analytic function. Its nth-derivative at u in the directions $(v_1,\dots,v_n) \in L^{\Phi_1}(p) \times \dots \times L^{\Phi_1}(p)$ is

$$\mathbf{D}^n M_p(u) \cdot (v_1,\dots,v_n) = \mathrm{E}_p\left[\lambda_{1,n}(u)\cdot(v_1,\dots,v_n)\right] = \mathrm{E}_p\left[v_1 \cdots v_n\, \mathrm{e}^u\right].$$

□

We introduce here the notations $B_p = \left\{u \in L^{\Phi_1}(p) \mid \mathrm{E}_p[u] = 0\right\}$ and $\mathcal{V}_p = B_p \cap B(0,1)$ that is

$$\mathcal{V}_p = \left\{u \in B_p \mid \|u\|_{\Phi_1,p} < 1\right\}.$$

Definition 11 *The* cumulant generating functional *is*

$$K_p : \begin{cases} B_p \to \mathbb{R} \cup \{+\infty\} \\ u \mapsto \log\left(M_p(u)\right) \end{cases}$$

For convenience, the moment generating functional is defined on the space L^{Φ_1}, while the cumulant generating functional is defined on the corresponding space of centered random variables.

Theorem 12 *The cumulant generating functional K_p satisfies the following properties*:

(1) $K_p(0) = 0$; *otherwise, for each $u \neq 0$, $K_p(u) > 0$.*
(2) K_p *is convex and lower semi-continuous, and its proper domain*

$$\operatorname{dom} K_p = \left\{u \in B_p \mid K_p(u) < \infty\right\}$$

is a convex set which contains the open unit ball $\mathcal{V}_p$; in particular the interior of such a domain is a non empty convex set.
(3) K_p *is infinitely Gâteaux-differentiable in the interior of its proper domain.*
(4) K_p *is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of $\mathcal{V}_p$.*

Proof All immediate from the definition of K_p. □

3 Families of Orlicz spaces

In the theory of statistical models, it is usual to associate to each density p in the model a space of p-centered random variables to represent scores or estimating functions. In fact, if the one-parameter statistical model $p(t)$, $t \in I$, I open interval, is regular enough, then $u(t) = \frac{\mathrm{d}}{\mathrm{d}t} \log p(t)$ satisfies $\mathrm{E}_{p(t)}[u(t)] = 0$ for all $t \in I$. A general estimating function is such a $u(t)$, other than the score. It is crucial to discuss how the relevant spaces of p-centered random variables depend on the variation of the density p. In particular, we are interested in the variation of the spaces $B_p = L_0^{\Phi_1}(p)$ and ${}^*B_p = L_0^{\Psi_1}(p)$ along a one-dimensional statistical model $p(t)$, $t \in I$. In Information Geometry, those spaces are models for the tangent and cotangent spaces of the statistical models. On two different points of a regular model, they must be isomorphic, or, in particular, equal.

In order to introduce our discussion, we are going to present a peculiar notion of connection by arcs, which is different from what is usually meant with this name. Note that, given $p, q \in \mathcal{M}$, the exponential model $p(\theta) \propto p^{1-\theta} q^{\theta}$, $0 \leq \theta \leq 1$ connects the two given densities as end points of a curve, sometimes called Hellinger arc. In fact, such a density in exponential form is

$$\exp\left(\theta \log \frac{q}{p}\right) \cdot p,$$

where $\log \frac{q}{p}$ is not in the exponential Orlicz space at p unless θ can be extended to assume negative values. As the Hellinger arc need not be continuous for the topology we are going to put on the set of all positive densities $\mathcal{M}$, we insist our one-parameter exponential models to be defined on *open* intervals.

In order to introduce our discussion, we start with an elementary example. The family of beta densities $p(t) \propto x^t$ is an exponential model with $t \in I = \left]-1, +\infty\right[$. We are going to consider general cases of the following facts:

(1) Given $t_1, t_2 \in I$, $t_2 > t_1$, the densities $p = p(t_1)$ and $q = p(t_2)$ are connected by an open exponential model proportional to $x^{(1-\theta)t_1} x^{\theta t_2}$ with $\theta \in \left]-\frac{t_1+1}{t_2-t_1}, +\infty\right[\supset [0,1]$.
(2) Given $t_1, t_2 \in I$, the densities $p = p(t_1)$ and $q = p(t_2)$ are connected by the mixture model $p(\lambda) = (1-\lambda)p + \lambda q$, $\lambda \in [0,1]$, then by a closed mixture model, but it is impossible to extend λ to take value either smaller than zero or bigger than 1 without violating the non-negativity. Then, p and q are not connected by an open mixture model.
(3) At all points $t \in I$, the exponential Orlicz spaces $L^{\Phi_1}(p(t))$ coincide. This is shown by considering the proper domain of the convex function $(\theta, t) \mapsto \int_0^1 e^{\theta x} x^t dx$.
(4) Given $t_1, t_2 \in I$, $t_2 > t_1$, we have $L^2(p(t_1)) \neq L^2(p(t_2))$ and $L^{\Psi_1}(p(t_1)) \neq L^{\Psi_1}(p(t_2))$. This can be checked by considering test random variables, e.g. x^α: if $\alpha \in \left]-\frac{1+t_2}{2}, -\frac{1+t_1}{2}\right]$, then $x^\alpha \in L^2(p(t_2))$ and $x^\alpha \notin L^2(p(t_1))$.

We consider first the case of mixture models.

Definition 13 *We say that $p, q \in \mathcal{M}$ are* connected by an open mixture arc *if there exist an open interval I and a mixture model $p(t)$, $t \in I$, containing both p and q at $t_0, t_1 \in I$, respectively.*

Proposition 14 *The relation in Definition 13 is an equivalence relation.*

Proof Reflexivity and symmetry follow from the definition, as a single point p is part of the mixture model $t \mapsto (1-t)p + tp$. For transitivity, consider two open mixture models $p(t)$, $t \in I$ and $q(s)$, $s \in J$, with $p = p(t_0)$, $q = p(t_1) = q(s_0)$, $r = q(s_1)$ (see Fig. 1):

$$p(t) = \frac{t_1 - t}{t_1 - t_0} p + \frac{t - t_0}{t_1 - t_0} q \quad q(s) = \frac{s_1 - s}{s_1 - s_0} q + \frac{s - s_0}{s_1 - s_0} r$$

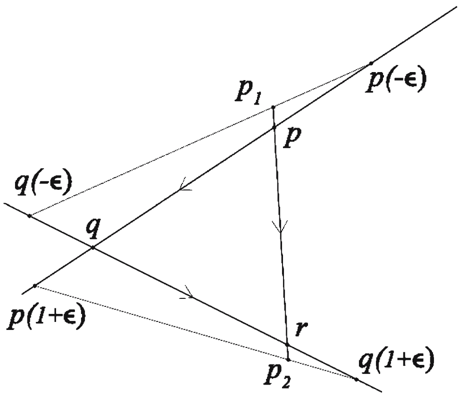


Fig. 1 Proof of Proposition 14

or, under a change of parameters,

$$p(\lambda) = (1-\lambda)p + \lambda q \quad q(\theta) = (1-\theta)q + \theta r$$

with λ and θ defined in a pair of open intervals containing $[0,1]$. Any convex combination of a $p(\lambda)$ with a $q(\theta)$ is positive, then it is a density. For some $\epsilon > 0$, consider the convex combination p_1 of $p(-\epsilon)$ with $q(-\epsilon)$

$$p_1 = \left(1 - \frac{\epsilon}{1+2\epsilon}\right) p(-\epsilon) + \frac{\epsilon}{1+2\epsilon} q(-\epsilon) = \frac{(1+\epsilon)^2}{1+2\epsilon} p - \frac{\epsilon^2}{1+2\epsilon} r$$

and the convex combination p_2 of $p(1+\epsilon)$ with $q(1+\epsilon)$

$$p_2 = \left(1 - \frac{1+\epsilon}{1+2\epsilon}\right) p(1+\epsilon) + \frac{1+\epsilon}{1+2\epsilon} q(1+\epsilon) = -\frac{\epsilon^2}{1+2\epsilon} p + \frac{(1+\epsilon)^2}{1+2\epsilon} r.$$

The mixture model $r(k) = (1-k)p_1 + kp_2$, $k \in (0,1)$, contains $p = r(\epsilon^2/(1+2\epsilon+2\epsilon^2))$ and $r = r((1+\epsilon)^2/(1+2\epsilon+2\epsilon^2))$. □

Proposition 15 *Let Φ_0 be any Young function. If p and q are connected by an open mixture arc, then*:

(1) *The ratio q/p is bounded above and below by positive constants.*
(2) $L^{\Phi_0}(p) = L^{\Phi_0}(q)$.

Proof (1) If two densities p and q are connected by an open mixture model $p(\lambda) = (1-\lambda)p + \lambda q$ with $\lambda \in]\alpha, 1+\beta[\supset [0,1]$, as $p(-\alpha)$ and $p(1+\beta)$ are non negative

$$\frac{\beta}{1+\beta} \leq \frac{q}{p} \leq \frac{1+\alpha}{\alpha}. \tag{3}$$

(2) Let $u \in L^{\Phi_0}(p)$, that is $\mathrm{E}_p[\Phi_0(\alpha u)] < +\infty$ for some $\alpha > 0$. By assumption, there exists a point r on the mixture model such that $p = (1-\theta)r + \theta q$, $\theta \in]0,1[$. From

$$\int \Phi_0(\alpha u)\left[(1-\theta)r + \theta q\right] \mathrm{d}\mu < +\infty \Rightarrow \int \Phi_0(\alpha u) q \,\mathrm{d}\mu < +\infty$$

we get $u \in L^{\Phi_0}(q)$. □

Second, we consider exponential models.

Definition 16 *We say that $p, q \in \mathcal{M}$ are* connected by an open exponential arc *if there exist $r \in \mathcal{M}$, a random variable u and an open interval I, such that $p(t) \propto \mathrm{e}^{tu} r$, $t \in I$, is an exponential model containing both p and q at t_0, t_1 respectively. By the change of parameter $s = t - t_0$, we can always reduce to the case where $r = p$ and $u \in L^{\Phi_1}(p)$.*

Proposition 17 *The relation in Definition 16 is an equivalence relation.*

Proof Reflexivity and symmetry follow from the definition. For transitivity, consider the open exponential models

$$p(t) \propto \mathrm{e}^{tu} q, \quad r(t) \propto \mathrm{e}^{tv} q, \;\; t \in \left]-\epsilon, 1+\epsilon\right[$$

with $p(1) = p$, $r(1) = r$, $u, v \in L^{\Phi_1}(q)$. The exponential model

$$q(\theta) \propto \mathrm{e}^{(1-\theta)u+\theta v} q \propto \mathrm{e}^{\theta(v-u)} p$$

is defined, by convexity, for θ in an open neighborhood of $[0,1]$ and it is such that $q(0) = p$ and $q(1) = r$. □

In order to prove our key results, e.g. Theorem 19 below, we need the following lemma.

Lemma 18 *Let Φ be a strictly convex symmetric function and $\Phi_0 = \Phi - \Phi(0)$ the related Young function. Let $p(t) \propto f(t)$, $t \in I$, I open interval, be such that for each random variable w the function*

$$g(\theta, t) = \mathrm{E}(\Phi(\theta w) f(t)), \quad \theta \in \mathbb{R}, \;\; t \in I$$

is convex. Then, for all $t \in I$ the spaces $L^{\Phi_0}(p(t))$ are equal.

Proof We prove equality of the Orlicz spaces by a convexity argument. Let $t_0, t_1 \in I$ and assume $w \in L^{\Phi_0}(p(t_0))$. Then, the function $\theta \mapsto g(\theta, t_0)$ is finite on a symmetric interval J. As I is open, there exist a $\bar{t} \in I$ such that t_1 is between t_0 and $\bar{t}$. The function $t \mapsto g(0,t)$ is finite on the interval whose end points are $t_0, \bar{t}$. Because of the convexity, the function g is finite on the convex hull of $J \times \{t_0\} \cup \{0\} \times [t_0, \bar{t}]$. Especially, the function $\theta \mapsto g(\theta, t_1)$ is finite on a symmetric interval, then it follows $w \in L^{\Phi_0}(p(t_1))$. □

Theorem 19 *Let p and q be densities connected by an open exponential arc. Then the Banach spaces $L^{\Phi_1}(p)$ and $L^{\Phi_1}(q)$ are equal as vector spaces and their norms are equivalent.*

Proof Equality as vector spaces follows from Lemma 18 with $\Phi(x) = \cosh(x)$ and $f(t) \propto \mathrm{e}^{tu} p$. Equivalence of the norms follows from Lemma 1. □

The following definition was introduced in Pistone and Sempi (1995).

Definition 20 *Let us denote by $\mathcal{S}_p$ the interior of the proper domain of the cumulant generating functional* $\overset{\circ}{\mathrm{dom}}\, K_p$. *For every density $p \in \mathcal{M}$, the* maximal exponential model at *p is defined to be the family of densities*

$$\mathcal{E}(p) := \left\{ \mathrm{e}^{u - K_p(u)} p \mid u \in \mathcal{S}_p \right\} \subseteq \mathcal{M}.$$

In the original paper, such a definition was introduced to discuss the connected components of the exponential manifold. Here, we will show in Theorem 25 below that there is a unique chart of the manifold whose domain is $\mathcal{S}_p$ and the image is $\mathcal{E}(p)$. We give now equivalent conditions to check if a given density q belongs to the maximal exponential model at p. The main point is that being in the same maximal exponential model and being connected by an open exponential arc is the same property, cf. 17.

Theorem 21 *The following statements are equivalent*:

(1) $q \in \mathcal{E}(p)$;
(2) $q \in \mathcal{M}$ *is connected to* p *by an open exponential arc*;
(3) $\mathcal{E}(p) = \mathcal{E}(q)$;
(4) $\log \frac{q}{p}$ *belongs to both* $L^{\Phi_1}(p)$ *and* $L^{\Phi_1}(q)$.

Proof If $q \in \mathcal{E}(p)$, then $q \propto e^u p$ for some $u \in \mathcal{S}_p$. As $tu \in \mathcal{S}_p$ for t in an open interval containing both 0 and 1 because $\{tu\} \cap \mathcal{S}_p$ is a convex open set containing both 0 and u in the line generated by u, then $p(t) \propto e^{tu} p$ is an open exponential arc containing q. This shows that (1) implies (2).

Let us assume (2). Then, the exponential arc $p(t) \propto e^{tu} p$ contains q at $t = 1$ and is defined at some $\bar{t} > 1$. Let v have a distance from u smaller than $(\bar{t} - 1)/\bar{t}$. Define $\bar{u} = \bar{t}u$ and $\bar{v} = \bar{t}(v - u)/(\bar{t} - 1)$. The norm of $\bar{v}$ is smaller than 1, so that $\bar{v}$ belong to $\mathcal{S}_p$ and v is a convex combination of $\bar{v}$ and $\bar{u}$. This shows that u is in $\mathcal{S}_p$ because of the convexity, then (1) is true.

Now we show that (1) and (2) imply (4). If $q \in \mathcal{E}(p)$, then $\log \frac{q}{p}$ differs by a constant from an element of $\mathcal{S}_p$, in particular it belongs to $L^{\Phi_1}(p)$. Also, as $L^{\Phi_1}(p) = L^{\Phi_1}(q)$ because of the property (2) and Theorem 19, then $\log \frac{q}{p} \in L^{\Phi_1}(q)$.

If (3) holds, then (2) is true. Vice-versa, by Proposition 17, the two equivalence classes $\mathcal{E}(p)$ and $\mathcal{E}(q)$ are the same equivalence class.

Finally, $\log \frac{q}{p}$ belongs to the Orlicz space $L^{\Phi_1}(p)$ if and only if its moment generating functional is defined in an open interval containing 0 which implies that there is an exponential arc containing p in its interior. The second part of the assumption rules the case of q in a similar way. □

In the following proposition we have collected a number of properties of the maximal exponential model $\mathcal{E}(p)$ which are relevant for its manifold structure.

Proposition 22 *Assume* $q \in \mathcal{E}(p)$. *Then the following statements are true.*

(1) *The mapping*

$$U_{pq} : \begin{cases} L^{\Psi_1}(p) \to L^{\Psi_1}(q) \\ v \mapsto v\dfrac{p}{q} \end{cases}$$

is an isomorphism of Banach spaces.

(2) $q/p \in L^{\Psi_1}(p)$.

(3) $D\left(q \,\|\, p\right) = \mathbf{D}K_p(u)\cdot u - K_p(u)$ *with* $q = \mathrm{e}^{u-K_p(u)}p$, *in particular* $D(q \,\|\, p) < +\infty$.

(4)

$$B_q = L_0^{\Phi_1}(q) = \left\{ u \in L^{\Phi_1}(p) \,|\, \mathrm{E}_p\left[u\frac{q}{p}\right] = 0 \right\}.$$

(5) $u \mapsto u - \mathrm{E}_q\left[u\right]$ *is an isomorphism of* B_p *onto* B_q.

Proof (1) Let $v \in L^{\Psi_1}(p)$ and consider, if it exists, the norm N_q of vp/q in $L^{\Psi_1}(q)$

$$\begin{aligned} N_q\left(v\frac{p}{q}\right) &= \sup\left\{\mathrm{E}_q\left[uv\frac{p}{q}\right] \,|\, \mathrm{E}_q\left[\Phi_1(u)\right] \le 1\right\} \\ &= \sup\left\{\mathrm{E}_p\left[uv\right] \,|\, \mathrm{E}_q\left[\Phi_1(u)\right] \le 1\right\} \\ &\le \sup\left\{\alpha\mathrm{E}_p\left[\frac{u}{\alpha}v\right] \,|\, \mathrm{E}_p\left[\Phi_1\left(\frac{u}{\alpha}\right)\right] \le 1\right\} \\ &\le \alpha N_p(v) < \infty, \end{aligned}$$

where α is such that for all u $\|u\|_{\Phi_1,p} \le \alpha \|u\|_{\Phi_1,q}$. Since $N_q\left(v\,p/q\right) < \alpha N_p(v)$ is bounded, $v\,p/q$ is an element of $L^{\Psi_1}(q)$ and the linear mapping U_{pq} is continuous. The same is true for the inverse map $w \mapsto U_{qp}(w) = wq/p \in L^{\Psi_1}(p)$.

(2) q/p is the image of 1 under the map U_{pq}.

(3) As K_p is Gâteaux-differentiable in the interior of its proper domain,

$$\begin{aligned} D\left(q \,\|\, p\right) &= \mathrm{E}_q\left[u\right] - K_p\left(u\right) \\ &= \frac{\mathrm{E}_p\left[u\mathrm{e}^u\right]}{\mathrm{E}_p\left[\mathrm{e}^u\right]} - K_p\left(u\right) \\ &= \frac{\mathbf{D}M_p\left(u\right)\cdot u}{M_p\left(u\right)} - K_p\left(u\right) \\ &= \mathbf{D}K_p\left(u\right)\cdot u - K_p\left(u\right). \end{aligned}$$

(4) As the Orlicz space $L^{\Phi_1}(q)$ is equal to $L^{\Phi_1}(p)$, then $u \in B_q$ means $u \in L^{\Phi_1}(q) = L^{\Phi_1}(p)$ and $\mathrm{E}_q\left[u\right] = \mathrm{E}_p\left[u\frac{q}{p}\right] = 0$.

(5) The map is linear with inverse the mapping $v \mapsto v - \mathrm{E}_p\left[v\right]$ of B_q onto B_p. To prove continuity consider, for any $u \in B_p$, the following bound

$$\left\|u - \mathrm{E}_q\left[u\right]\right\|_{\Phi_1,q} \le \|u\|_{\Phi_1,q} + \left|\mathrm{E}_q\left[u\right]\right| \|1\|_{\Phi_1,q} \le C\,\|u\|_{\Phi_1,p},$$

where equivalence of $L^{\Phi_1}(p)$ and $L^{\Phi_1}(q)$ and continuity of $\mathrm{E}_q\left[\,\cdot\,\right]$ are exploited. □

Two remarks are of order. First, the necessary condition $q/p \in L^{\Psi_1}(p) = L^{\Psi_1}(q)$ is a statement of finite relative entropy. We do not know if it is actually sufficient.

Second, the characterization of all B_q's as subspaces of codimension 1 of the same model space is crucial for the construction of the exponential manifold, its tangent bundle and both its e- and m-connections, as in Amari and Nagaoka (2000).

4 Exponential manifold

If $p, q \in \mathcal{M}$ are connected by an open exponential arc, then the random variable $u \in \mathcal{S}_p$ such that $q \propto \mathrm{e}^u p$ is unique and it is equal to $\log \frac{q}{p} - \mathrm{E}_p\left[\log \frac{q}{p}\right]$. In fact, $q \propto \mathrm{e}^u p$ for some $u \in L^{\Phi_1}(p)$ if and only if $u - \log \frac{q}{p}$ is a constant. If $u \in \mathcal{S}_p \subset B_p$, then $u - \log \frac{q}{p} = K_p(u)$ and, as u is centered, it follows that $-\mathrm{E}_p\left[\log \frac{q}{p}\right] = K_p(u)$ and $u = \log \frac{q}{p} - \mathrm{E}_p\left[\log \frac{q}{p}\right]$. Indeed, u is the projection of $\log \frac{q}{p}$ onto B_p in the split $L^{\Phi_1}(p) = B_p \oplus \langle 1 \rangle$.

Definition 23 *We define two one-to-one mappings: the parameterization $(\mathcal{S}_p, e_p)$ and the chart $(\mathcal{E}(p), s_p)$, respectively, as*

$$e_p : \begin{cases} \mathcal{S}_p \to \mathcal{E}(p) \subseteq \mathcal{M} \\ u \mapsto \mathrm{e}^{u - K_p(u)} \cdot p \end{cases}$$

$$s_p : \begin{cases} \mathcal{E}(p) \to \mathcal{S}_p \\ q \mapsto \log \dfrac{q}{p} - \mathrm{E}_p\left[\log \dfrac{q}{p}\right] \end{cases}$$

As the interior of the proper domain of K_p contains the unit ball $\mathcal{V}_p \subset B_p$, the current definition is an extension of the charts $s_p : \mathcal{U}_p \to \mathcal{V}_p$ introduced in previous papers. In statistical terms, the coordinate $s_p(q)$ represents the density $q \in \mathcal{E}(p)$ with its centered log-likelihood s_p. The charts $(\mathcal{E}(p), s_p)$ are defined on domains which are either equal or disjoint. This fact simplifies the proof of the properties of the corresponding atlas.

Theorem 24 *If $p_1, p_2 \in \mathcal{E}(p)$ for a pair of densities $p_1, p_2 \in \mathcal{M}$, then the transition mapping $s_{p_2} \circ e_{p_1}$ is the restriction of an affine function*

$$s_{p_2} \circ e_{p_1} : \begin{cases} \mathcal{S}_{p_1} \to \mathcal{S}_{p_2} \\ u \mapsto u + \log \dfrac{p_1}{p_2} - \mathrm{E}_{p_2}\left[u + \log \dfrac{p_1}{p_2}\right] \end{cases}$$

Proof Let $p_1 = \mathrm{e}^{w - K_{p_2}(w)} p_2$ with $w = \log \frac{p_1}{p_2} - \mathrm{E}_{p_2}\left[\log \frac{p_1}{p_2}\right] \in \mathcal{S}_{p_2}$. For any $q \in \mathcal{E}(p)$ there exist $u_1 = s_{p_1}(q) \in \mathcal{S}_{p_1}$ and $u_2 = s_{p_2}(q) \in \mathcal{S}_{p_2}$. With a direct calculation one can check that

$$u_2 = u_1 - \mathrm{E}_{p_2}[u_1] + \log \frac{p_1}{p_2} - \mathrm{E}_{p_2}\left[\log \frac{p_1}{p_2}\right] = u_1 - \mathrm{E}_{p_2}[u_1] + w.$$

□

With the notations of the previous proof, observe that the image under s_{p_1} of the open exponential arc $\mathrm{e}^{tu_1 - K_{p_1}(tu_1)} p_1$ connecting p_1 to q lies on the line of B_{p_1} generated by u_1 and that it is mapped by the overlap map $s_{p_2} \circ e_{p_1}$ into the affine line generated by the projection of u_1 onto B_{p_2} with a translation of $w = s_{p_2}(p_1)$.

The derivative of the transition map $s_{p_2} \circ e_{p_1}$ is the isomorphism of B_{p_1} onto B_{p_2}

$$B_{p_1} \ni u \mapsto u - \mathrm{E}_{p_2}[u] \in B_{p_2}.$$

The new manifold structure defined above is equivalent to the old one in Pistone and Sempi (1995).

Theorem 25 *The atlases*

$$\mathcal{A} = \left\{\left(\mathcal{E}(p), s_p\right) \mid p \in \mathcal{M}\right\} \quad \text{and} \quad \mathcal{B} = \left\{\left(\mathcal{U}_p, s_p\right) \mid p \in \mathcal{M}\right\}$$

are equivalent.

Proof We must prove that all overlap maps mixing charts of the two atlases have open domains and are of class C^∞. In this case the intersection of the two domains is always of the form $\mathcal{U}_p \cap \mathcal{E}(q)$ which is either empty or equal to $\mathcal{U}_p$. The overlap map is a restriction of the map computed in the previous theorem.

□

Definition 26 *The* exponential statistical manifold *is the manifold defined by either one of the atlases in Theorem 25.*

Notice that for every density $p \in \mathcal{M}$, the maximal exponential model $\mathcal{E}(p)$ is the connected component of the exponential statistical manifold $\mathcal{M}$ containing p. In fact, all points of $\mathcal{E}(p)$ are connected by continuous exponential arcs and $\mathcal{E}(p)$ is both open, being the image of a chart, and closed, being the complement of all the others.

We do not require $\mathcal{M}$ to be a topological space as in Pistone and Sempi (1995). However, a metric topology is induced by the equivalent atlases $\mathcal{A}$ and $\mathcal{B}$. As we are going to use a specific characterization of such topology to prove the continuity of open mixture models, we recall such a characterization.

Definition 27 *The sequence* $\{p_n\}$, $n \in \mathbb{N}$, *in* $\mathcal{M}$ *is* e-convergent (*exponentially convergent*) *to* p *if* $\{p_n\}$ *tends to* p *in* μ*-measure as* $n \to \infty$ *and moreover sequences* $\left\{\frac{p_n}{p}\right\}$ *and* $\left\{\frac{p}{p_n}\right\}$ *are eventually bounded in each* $L^\alpha(p)$, $\alpha > 1$, *that is*

$$\forall \alpha > 1, \ \limsup_{n\to\infty} \mathrm{E}_p\left[\left(\frac{p_n}{p}\right)^\alpha\right] < \infty \quad \text{and} \quad \limsup_{n\to\infty} \mathrm{E}_p\left[\left(\frac{p}{p_n}\right)^\alpha\right] < \infty.$$

Here is an equivalent statement.

Proposition 28 *A sequence* $\{p_n\}$*,* $n \in \mathbb{N}$*, is e-convergent to* p *if and only if sequences* $\{p_n/p\}$ *and* $\{p/p_n\}$ *are convergent to* 1 *in each* $L^\alpha(p)$*,* $\alpha > 1$*.*

Proof Let us assume

$$\mathrm{E}_p\left[\left|\frac{p_n}{p} - 1\right|^\alpha\right] \to 0 \quad \text{and} \quad \mathrm{E}_p\left[\left|\frac{p}{p_n} - 1\right|^\alpha\right] \to 0 \tag{4}$$

for each $\alpha > 1$. Since $L^\alpha(p) \sqsubset L^1(p)$, the convergences (4) hold also for $\alpha = 1$ and, in particular,

$$\int |p_n - p| \,\mathrm{d}\mu = 2\left(1 - \int \min(p_n, p)\mathrm{d}\mu\right) \to 0$$

which is equivalent to $p_n \xrightarrow{\mu} p$.

In order to prove the boundedness condition of the sequence $\{p_n/p\}$ it suffices to use Minkowsky's Inequality and the convergence of $\{p_n/p\}$ to 1 in $L^\alpha(p)$,

$$\mathrm{E}_p\left[\left|\frac{p_n}{p}\right|^\alpha\right] = \mathrm{E}_p\left[\left|\frac{p_n}{p} - 1 + 1\right|^\alpha\right] \le \left[\underbrace{\left(\mathrm{E}_p\left[\left|\frac{p_n}{p} - 1\right|^\alpha\right]\right)^{1/\alpha}}_{\longrightarrow 0} + 1\right]^\alpha$$

and $\limsup_{n\to\infty} \mathrm{E}_p\left[|p_n/p|^\alpha\right] \le 1$. The same argument applies to $\{p/p_n\}$.

Conversely, let the sequence $\{p_n\}$ be e-convergent to p. We remark that, for each $\alpha > 1$, Minkowsky's Inequality

$$\left(\mathrm{E}_p\left[\left|\frac{p_n}{p} - 1\right|^\alpha\right]\right)^{1/\alpha} \le \left(\mathrm{E}_p\left[\left|\frac{p_n}{p}\right|^\alpha\right]\right)^{1/\alpha} + 1$$

and the boundedness condition of e-convergence definition show that the sequence $\left\{\mathrm{E}_p\left[|p_n/p - 1|^\alpha\right]\right\}$ is eventually bounded. As p_n converges to p in $L^1(\mu)$, then p_n/p converges to 1 in $L^1(p)$.

Let $a \in \left]\alpha - 1, \alpha\right[$ be given. Hölder's Inequality and the convergence $p_n/p \to 1$ in $L^1(p)$ show that $p_n/p \to 1$ in $L^\alpha(p)$:

$$\mathrm{E}_p\left[\left|\frac{p_n}{p} - 1\right|^\alpha\right] \le \left(\mathrm{E}_p\left[\left|\frac{p_n}{p} - 1\right|\right]\right)^{\alpha - a} \left(\mathrm{E}_p\left[\left|\frac{p_n}{p} - 1\right|^{\frac{a}{1-\alpha+a}}\right]\right)^{1-\alpha+a} \to 0.$$

The convergence $p/p_n \to 1$ in $L^\alpha(p)$ follows from

$$\begin{aligned}\mathrm{E}_p\left[\left|\frac{p}{p_n}-1\right|^\alpha\right] &= \mathrm{E}_p\left[\left(\frac{p}{p_n}\right)^\alpha\left|1-\frac{p_n}{p}\right|^\alpha\right] \\ &\le \left(\mathrm{E}_p\left[\left(\frac{p}{p_n}\right)^{2\alpha}\right]\right)^{1/2}\left(\mathrm{E}_p\left[\left|1-\frac{p_n}{p}\right|^{2\alpha}\right]\right)^{1/2} \to 0.\end{aligned}$$

□

This characterization of e-convergence could be used to prove directly the continuity of open exponential arcs. However, this continuity follows also from the C^∞-regularity of open exponential arcs, as images of lines.

We consider the *closed* mixture model connecting two points on a maximal exponential model.

Proposition 29 *If $q \in \mathcal{E}(p)$, then the mixture model $p(\lambda) = (1-\lambda)p + \lambda q \in \mathcal{E}(p)$ for $\lambda \in [0,1]$.*

Proof If $q \in \mathcal{E}(p)$, then the exponential model $p^{1-\theta}q^\theta$ is defined for $\theta \in]-\alpha, 1+\alpha[$. As x^θ is a convex function for each $\theta \in]-\alpha, 0[\cup]1, 1+\alpha[$,

$$\mathrm{E}_p\left[\left(\frac{p(\lambda)}{p}\right)^\theta\right] \le (1-\lambda) + \lambda \mathrm{E}_p\left[\left(\frac{q}{p}\right)^\theta\right] < +\infty \quad \lambda \in [0,1].$$

For each $\theta \in [0,1]$ there is nothing to prove, because we have an Hellinger arc. This shows that $p^{1-\theta}p(\lambda)^\theta$ is an open exponential model connecting p and $p(\lambda)$.
□

We use now e-convergence to prove continuity of an open mixture arc.

Proposition 30 *An open mixture arc*

$$p(\cdot) : \begin{cases}]-\alpha, 1+\beta[\to \mathcal{M} \\ t \mapsto (1-t)p + tq \end{cases}$$

is a continuous path.

Proof Let $\{t_n\}$ be a sequence converging to $\bar{t} \in]-\alpha, 1+\beta[$. We shall prove that $\{p(t_n)\}$ is e-convergent to $p(\bar{t})$.

It follows from $p(t_n) - p(\bar{t}) = (t_n - \bar{t})(q-p)$ that $\{p(t_n)\}$ converges to $p(\bar{t})$ in $L^1(\mu)$. Then, $p(t_n) \to p(\bar{t})$ in μ-measure, an also $p(t_n)/p(\bar{t}) \to 1$, $p(\bar{t})/p(t_n) - 1$ in μ-measure. For each n, the open mixture arc $p(t)$ can be written as

$$p(t) = \frac{t_n - t}{t_n - \bar{t}} p(\bar{t}) + \frac{t - \bar{t}}{t_n - \bar{t}} p(t_n).$$

If $t_n > \bar{t}$ then, under the change of parameter $t \mapsto \lambda = (t - \bar{t})/(t_n - \bar{t})$, the open mixture arc is

$$p(\lambda) = (1-\lambda)\, p(\bar{t}) + \lambda p(t) \quad \text{with } \lambda \in \left]-\frac{\alpha + \bar{t}}{t_n - \bar{t}},\ 1 + \frac{1+\beta - t_n}{t_n - \bar{t}}\right[.$$

By (3) of Proposition 15, the ratio $\frac{p(t_n)}{p(\bar{t})}$ is bounded above and below,

$$\frac{1+\beta-t_n}{1+\beta-\bar{t}} \leq \frac{p(t_n)}{p(\bar{t})} \leq \frac{\alpha+t_n}{\alpha+\bar{t}}.$$

A similar argument shows that if $t_n < \bar{t}$ then

$$\frac{\alpha+t_n}{\alpha+\bar{t}} \leq \frac{p(t_n)}{p(\bar{t})} \leq \frac{1+\beta-t_n}{1+\beta-\bar{t}}$$

and e-convergence follows from

$$\forall\, a > 1 \quad \lim_{n\to\infty} \mathrm{E}_{p(\bar{t})}\left[\left(\frac{p(t_n)}{p(\bar{t})}\right)^a\right] = \lim_{n\to\infty} \mathrm{E}_{p(\bar{t})}\left[\left(\frac{p(\bar{t})}{p(t_n)}\right)^a\right] = 1.$$

□

The two previous results show that the closed mixture arc between two densities connected by an open exponential arc is contained in their maximal exponential model and it is e-continuous in its interior. In general, it is not continuous the end-points. Regularity and smoothness mixture models can be obtained only in weaker topologies and manifold structures, as it is done in the following section.

5 Mixture manifold

Let $p \in \mathcal{M}$ be a probability density. For each $u \in \mathcal{V}_p$ and $q = \mathrm{e}^{u-K_p(u)}p$, the derivative of K_p at u, $\mathbf{D}K_p(u) \in B_p^*$, is the linear mapping

$$\mathbf{D}K_p(u) \cdot v = \mathrm{E}_p\left[\left(\frac{q}{p} - 1\right) v\right], \quad v \in B_p$$

and $\mathbf{D}K_p(u)$ is identified to its gradient $q/p - 1 \in {}^*B_p$. Mapping $\mathcal{U}_p \ni q \mapsto q/p - 1 \in {}^*B_p = \left\{v \in L^{\Psi_1}(p) \,|\, \mathrm{E}_p[v] = 0\right\}$ cannot be a chart because its values are bounded below by -1. We consider the set of all densities $p \geq 0$

$$\mathcal{P}_{\geq} := \left\{p \in L^1(\mu) \mid p \geq 0,\ \int p\, \mathrm{d}\mu = 1\right\}$$

and the set

$$\mathcal{P} := \left\{ p \in L^1(\mu) \mid \int p \mathrm{d}\mu = 1 \right\}.$$

Observe that $\mathcal{M} \subseteq \mathcal{P}_\geq \subseteq \mathcal{P}$. For each $q \in \mathcal{P}$ there exists an element $\tilde{q} \in \mathcal{P}_\geq$ defined by

$$\tilde{q} := \frac{|q|}{\int |q| \mathrm{d}\mu}.$$

For each probability density $p \in \mathcal{P}_\geq$, let us introduce the subset ${}^*\mathcal{U}_p$ of $\mathcal{P}$ defined by

$${}^*\mathcal{U}_p := \left\{ q \in \mathcal{P} \mid \frac{q}{p} \in L^{\Psi_1}(p) \right\}.$$

Then consider the map η_p defined on ${}^*\mathcal{U}_p$ by

$$\eta_p : \begin{cases} {}^*\mathcal{U}_p \to {}^* B_p \\ q \mapsto \dfrac{q}{p} - 1 \end{cases}$$

Since $\eta_p(q)$, $q \in \mathcal{U}_p \subset \mathcal{E}(p)$, can be identified with $\mathrm{E}_q[v]$, it could be called *expectation parameter*. This mapping is bijective and its inverse is

$$\eta_p^{-1} : {}^*B_p \ni u \mapsto (u+1)p \in {}^*\mathcal{U}_p.$$

The collection of sets $\left\{{}^*\mathcal{U}_p\right\}_{p \in \mathcal{P}_\geq}$ is a covering of $\mathcal{P}$. In fact, for each $q \in \mathcal{P}$ we have $q \in {}^*\mathcal{U}_{\tilde{q}}$ and $\tilde{q} \in \mathcal{P}_\geq$.

Let us characterize the elements of ${}^*\mathcal{U}_p \bigcap \mathcal{P}_\geq$: they are all the probability densities with finite divergence with respect to p.

Proposition 31 *Let $p \in \mathcal{M}$ be given. For each $q \in \mathcal{P}$, the divergence $D(\tilde{q} \,\|\, p)$ of the probability density $\tilde{q}$ with respect to p is finite if and only if $q \in {}^*\mathcal{U}_p$:*

$$D(\tilde{q} \,\|\, p) = \mathrm{E}_p\left[\frac{\tilde{q}}{p} \log\left(\frac{\tilde{q}}{p}\right)\right] < \infty \quad \Leftrightarrow \quad q \in {}^*\mathcal{U}_p.$$

Proof Let $q \in \mathcal{P}$. We assume that $D\left(\tilde{q} \parallel p\right) < \infty$. By the inequality $\Psi_1(x) \leq 1 + x\log(x)$ for $x > 0$ we have

$$\mathrm{E}_p\left[\Psi_1\left(\left(\int |q|\,\mathrm{d}\mu\right)^{-1}\frac{q}{p}\right)\right] = \mathrm{E}_p\left[\Psi_1\left(\frac{\tilde{q}}{p}\right)\right]$$
$$\leq 1 + \mathrm{E}_p\left[\frac{\tilde{q}}{p}\log\left(\frac{\tilde{q}}{p}\right)\right]$$
$$= 1 + D\left(\tilde{q} \parallel p\right) < \infty.$$

Hence $\frac{q}{p} \in L^{\Psi_1}(p)$ and, by definition, $q \in {}^*\mathcal{U}_p$.

Conversely, let $q \in {}^*\mathcal{U}_p$, that is $\frac{q}{p} \in L^{\Psi_1}(p)$. Since $L^{\Psi_1}(p)$ is linear, $\frac{\tilde{q}}{p} = \left(\int |q|\,\mathrm{d}\mu\right)^{-1}\frac{q}{p} \in L^{\Psi_1}(p)$ and since $\left(1+\frac{\tilde{q}}{p}\right)\log\left(1+\frac{\tilde{q}}{p}\right)$ is p integrable, from inequality

$$x\log^+(x) \leq (1+x)\log(1+x), \quad x > 0$$

where $\log^+(x) := \max\{0, \log(x)\}$, we have

$$\mathrm{E}_p\left[\frac{\tilde{q}}{p}\log\left(\frac{\tilde{q}}{p}\right)\right] \leq \mathrm{E}_p\left[\frac{\tilde{q}}{p}\log^+\left(\frac{\tilde{q}}{p}\right)\right]$$
$$\leq \mathrm{E}_p\left[\left(1+\frac{\tilde{q}}{p}\right)\log\left(1+\frac{\tilde{q}}{p}\right)\right] < \infty$$

hence $D\left(\tilde{q} \parallel p\right)$ is finite. □

Recall that the divergence $D(q \parallel p)$ of a probability density $q \in \mathcal{P}_\geq$ with respect to $p \in \mathcal{M}$ is always well defined and it is finite if and only if $\log\left(\frac{q}{p}\right) \in L^1(q)$.

Proposition 32 *Let $p \in \mathcal{M}$ be given, then $\mathcal{U}_p \subset {}^*\mathcal{U}_p$.*

Proof If $q \in \mathcal{U}_p$, there exists $u \in \mathcal{V}_p \subset B_p$ such that $q = \mathrm{e}^{u-K_p(u)}p$. Since $L^{\Phi_1}(q) = L^{\Phi_1}(p)$ the random variable u is q-integrable and we have

$$\mathrm{E}_q\left[\log\left(\frac{q}{p}\right)\right] = \mathrm{E}_q[u] - K_p(u) < \infty.$$

□

Proposition 33 *Let p_1 and p_2 be two positive densities in the same connected component $\mathcal{E}(p)$. Then the linear map*

$$P^m_{p_1p_2} : \begin{cases} {}^*B_{p_1} \to {}^*B_{p_2} \\ u \mapsto u\,\frac{p_1}{p_2} \end{cases}$$

is an isomorphism of Banach spaces.

Proof $P^m_{p_1 p_2}$ is the restriction of the $U_{p_1 p_2}$ of Theorem 22. If $u \in {}^*B_{p_1}$ then $\mathrm{E}_{p_2}\left[u p_1/p_2\right] = \mathrm{E}_{p_1}[u] = 0$ and $up_1/p_2 \in {}^*B_{p_2}$. □

Proposition 34 *Let p_1 and p_2 be a pair of positive densities in the same connected component $\mathcal{E}(p) \subset \mathcal{M}$ for some density $p \in \mathcal{M}$, then ${}^*\mathcal{U}_{p_1} = {}^*\mathcal{U}_{p_2}$.*

Proof Let $p_1, p_2 \in \mathcal{E}(p)$. First we note that $p_1, p_2 \in {}^*\mathcal{U}_{p_1} \bigcap {}^*\mathcal{U}_{p_2}$. In fact, by Proposition 22(1) we have $\frac{p_1}{p_2} \in L^{\Psi_1}(p_2)$ and we conclude that $p_1 \in {}^*\mathcal{U}_{p_2}$. Similarly, we see that $p_2 \in {}^*\mathcal{U}_{p_1}$.

Every $q \in {}^*\mathcal{U}_{p_1}$ can be written as $(u+1)\,p_1$ where $u = \eta_{p_1}(q) \in {}^*B_{p_1}$ and we have

$$\frac{q}{p_2} = u\frac{p_1}{p_2} + \frac{p_1}{p_2} = P^m_{p_1,p_2}(u) + \frac{p_1}{p_2} \in L^{\Psi_1}(p_2)\,.$$

Hence $q \in {}^*\mathcal{U}_{p_2}$ and ${}^*\mathcal{U}_{p_1} \subseteq {}^*\mathcal{U}_{p_2}$. In the same way we prove the opposite inclusion. □

For each pair $p_1, p_2 \in \mathcal{E}(p)$ we can define the overlap map

$$\eta_{p_2} \circ \eta_{p_1}^{-1} : \begin{cases} {}^*B_{p_1} \to {}^*B_{p_2} \\ u \mapsto u\,\dfrac{p_1}{p_2} + \dfrac{p_1}{p_2} - 1 \end{cases}$$

The function $\eta_{p_2} \circ \eta_{p_1}^{-1}$ is a $\mathcal{C}^\infty$-affine map since it can be written as the sum of the continuous linear map $P^m_{p_1,p_2}$ and the constant $\frac{p_1}{p_2} - 1 \in {}^*B_{p_2}$ so it a $\mathcal{C}^\infty$-affine map.

Definition 35 *Let $p \in \mathcal{M}$ be fixed. ${}^*\mathcal{E}(p)$ is the subset of $\mathcal{P}$ defined by*

$${}^*\mathcal{E}(p) = \left\{ q \in \mathcal{P} \mid \frac{q}{p} \in L^{\Psi_1}(p) \right\}$$

Note that ${}^*\mathcal{E}(p) = {}^*\mathcal{U}_{p_\alpha}$ for each $p_\alpha \in \mathcal{E}(p)$.

${}^*\mathcal{E}(p)$ has the structure of a manifold modeled on the Banach space ${}^*B_p$.

Theorem 36 *Let $p \in \mathcal{M}$ be given. The collection of charts*

$$\left\{ \left({}^*\mathcal{U}_{p_\alpha}, \eta_{p_\alpha}\right) \mid p_\alpha \in \mathcal{E}(p) \right\}$$

is an affine $\mathcal{C}^\infty$-atlas on ${}^*\mathcal{E}(p)$.

Proof The collection of sets $\left\{{}^*\mathcal{U}_{p_\alpha} \mid p_\alpha \in \mathcal{E}(p)\right\}$ covers ${}^*\mathcal{E}(p)$. For each pair $p_1, p_2 \in \mathcal{E}(p)$ the set $\eta_{p_1}\left({}^*\mathcal{U}_{p_1} \bigcap {}^*\mathcal{U}_{p_2} = {}^*\mathcal{E}(p)\right) = {}^*B_{p_1}$ is clearly open in ${}^*B_{p_1}$ and we have just observed that the transition mapping $\eta_{p_2} \circ \eta_{p_1}^{-1}$ is an $\mathcal{C}^\infty$-affine function. □

We conclude this section with the derivation of a local Pythagorean-type relation in our framework.

Let $p \in \mathcal{M}$ be given and let $s_p : \mathcal{U}_p \to \mathcal{V}_p$ and $\eta_p : {}^*\mathcal{U}_p \to {}^*B_p$ be charts respectively in $\mathcal{E}(p)$ and ${}^*\mathcal{E}(p)$. Let $q \in \mathcal{U}_p$, $u = s_p(q)$ and $0 \le r \in {}^*\mathcal{U}_p$ be given. Consider the duality

$$\begin{aligned}\left\langle \eta_p(r), s_p(q) \right\rangle &= \mathrm{E}_p\left[\eta_p(r)\, s_p(q)\right] = \mathrm{E}_p\left[\left(\frac{r}{p} - 1\right) u\right] \\ &= \mathrm{E}_r[u].\end{aligned}$$

As

$$u = \log\left(\frac{q}{p}\right) - \mathrm{E}_p\left[\log\left(\frac{q}{p}\right)\right] = \log\left(\frac{q}{p}\right) + D(p \,\|\, q),$$

we have

$$\begin{aligned}\mathrm{E}_r[u] &= \mathrm{E}_r\left[\log\left(\frac{q}{p}\right)\right] + D(p \,\|\, q) \\ &= \mathrm{E}_r\left[\log\left(\frac{q}{r}\right) + \log\left(\frac{r}{p}\right)\right] + D(p \,\|\, q) \\ &= -D(r \,\|\, q) + D(r \,\|\, p) + D(p \,\|\, q).\end{aligned}$$

In particular, $\left\langle \eta_p(r), s_p(q) \right\rangle = 0$ implies the relation

$$D(r \,\|\, q) = D(r \,\|\, p) + D(p \,\|\, q).$$

6 Regularity of some maps

The regularity of the divergence function and of the Fisher information follows naturally from our framework and does not require any ad-hoc assumption.

Proposition 37 *For each density $p \in \mathcal{M}$, the divergence*

$$D(\cdot \,\|\, \cdot) : \mathcal{E}(p) \times \mathcal{E}(p) \to \mathbb{R}$$

is of class $\mathcal{C}^\infty$.

Proof The proof uses the local atlas $\mathcal{B}$. Given a pair of charts $\left(\mathcal{U}_{p_1}, s_{p_1}\right)$ and $\left(\mathcal{U}_{p_2}, s_{p_2}\right)$ of $\mathcal{E}(p)$, we consider the local representative $D_{p_1,p_2} = D \circ \left(e_{p_1}, e_{p_2}\right) : \mathcal{V}_{p_1} \times \mathcal{V}_{p_2} \to \mathbb{R}$.

The cumulant generating functionals $K_{p_i} \in \mathcal{C}^\infty\left(\mathcal{V}_{p_i}\right)$, $i = 1, 2$, and, for each $u_i \in \mathcal{V}_{p_i}$ and $v_i \in B_{p_i}$,

$$\mathbf{D}K_{p_i}(u_i) \cdot v_i = \mathrm{E}_{p_i}\left[v_i \mathrm{e}^{u_i - K_{p_i}(u_i)}\right].$$

We shall write D_{p_1,p_2} as a linear combination of smooth functions. For all $(q_1, q_2) \in \mathcal{U}_{p_1} \times \mathcal{U}_{p_2}$, if $q_i = e_{p_i}(u_i)$, then

$$\begin{aligned}
D(q_1 \| q_2) &= D_{p_1,p_2}(u_1, u_2) \\
&= \mathrm{E}_{p_1}\left[\log\left(\frac{\mathrm{e}^{u_1 - K_{p_1}(u_1)} p_1}{\mathrm{e}^{u_2 - K_{p_2}(u_2)} p_2}\right) \mathrm{e}^{u_1 - K_{p_1}(u_1)}\right] \\
&= \mathrm{E}_{p_1}\left[\left(u_1 - u_2 - \log\frac{p_2}{p_1} + \mathrm{E}_{p_1}\left[u_2 + \log\frac{p_2}{p_1}\right]\right) \mathrm{e}^{u_1 - K_{p_1}(u_1)}\right] \\
&\quad - K_{p_1}(u_1) + K_{p_2}(u_2) - \mathrm{E}_{p_1}\left[u_2 + \log\frac{p_2}{p_1}\right] \\
&= \mathrm{E}_{p_1}\left[\left(u_1 - s_{p_1} \circ e_{p_2}(u_2)\right) \mathrm{e}^{u_1 - K_{p_1}(u_1)}\right] \\
&\quad - K_{p_1}(u_1) + K_{p_2}(u_2) - \mathrm{E}_{p_1}\left[u_2 + \log\frac{p_2}{p_1}\right] \\
&= \mathbf{D}K_{p_1}(u_1) \cdot \left(u_1 - s_{p_1} \circ e_{p_2}(u_2)\right) \\
&\quad - K_{p_1}(u_1) + K_{p_2}(u_2) - \mathrm{E}_{p_1}\left[u_2 + \log\frac{p_2}{p_1}\right].
\end{aligned}$$

□

The partial derivatives $\mathbf{D}_i D_{p_1,p_2} : \mathcal{V}_{p_1} \times \mathcal{V}_{p_2} \to \left(B_{p_i}\right)^*$ at $(u_1, u_2) \in \mathcal{V}_{p_1} \times \mathcal{V}_{p_2}$ applied, respectively, to $w_1 \in B_{p_1}$ and $w_2 \in B_{p_2}$ are

$$\mathbf{D}_1 D_{p_1,p_2}(u_1, u_2) \cdot w_1 = \mathbf{D}^2 K_{p_1}(u_1) \cdot \left(u_1 - s_{p_1} \circ e_{p_2}(u_2), w_1\right) \tag{5}$$

and

$$\begin{aligned}
&\mathbf{D}_2 D_{p_1,p_2}(u_1, u_2) \cdot w_2 \\
&\quad = -\mathbf{D}K_{p_1}(u_1) \cdot \left(w_2 - \mathrm{E}_{p_1}[w_2]\right) + \mathbf{D}K_{p_2}(u_2) \cdot w_2 - \mathrm{E}_{p_1}[w_2].
\end{aligned}$$

If we assume $q_1, q_2 \in \mathcal{U}_{p_\alpha}$ with, as usual, $u_i = s_{p_\alpha}(q_i)$ for $i = 1, 2$, then

$$D(q_1 \| q_2) = \mathbf{D}K_{p_\alpha}(u_1) \cdot (u_1 - u_2) - K_{p_\alpha}(u_1) + K_{p_\alpha}(u_2) \tag{6}$$

When $q_2 = p_\alpha$, which is equivalent to $u_2 = 0$, Eq. (6) reduces to

$$D(q_1 \| p_\alpha) = \mathbf{D}K_{p_\alpha}(u_1) \cdot u_1 - K_{p_\alpha}(u_1)$$

and when $q_1 = p_\alpha$, which is equivalent to $u_1 = 0$, it reduces to

$$D(p_\alpha \| q_2) = K_{p_\alpha}(u_2).$$

Since $K_{p_\alpha} \in \mathcal{C}^\omega\left(\mathcal{V}_{p_\alpha}\right)$, Eq. (6) allows us to write $D\left(q_1 \,\|\, q_2\right)$ as

$$\begin{aligned} D\left(q_1 \,\|\, q_2\right) &= \mathbf{D}K_{p_\alpha}\left(u_1\right)\cdot\left(u_1-u_2\right) - K_{p_\alpha}\left(u_1\right) \\ &\quad + \sum_{n\geq 0}\frac{1}{n!}\mathbf{D}^n K_{p_\alpha}\left(u_1\right)\cdot\left(u_2-u_1\right)^n \\ &= \sum_{n\geq 2}\frac{1}{n!}\mathbf{D}^n K_{p_\alpha}\left(u_1\right)\cdot\left(u_2-u_1\right)^n . \end{aligned} \tag{7}$$

From (7), it is easy to see that for all $u \in \mathcal{V}_{p_1}$ and $\left(w_1, w_2\right) \in B_{p_\alpha} \times B_{p_\alpha}$

$$\mathbf{D}^2_{22} D_{p_\alpha p_\alpha}\left(u,u\right)\cdot\left(w_1,w_2\right) = \mathbf{D}^2 K_{p_\alpha}\left(u\right)\cdot\left(w_1,w_2\right).$$

Taking the derivative of (5) with respect to the first variable, we obtain also

$$\mathbf{D}^2_{11} D_{p_\alpha p_\alpha}\left(u,u\right)\cdot\left(w_1,w_2\right) = \mathbf{D}^2 K_{p_\alpha}\left(u\right)\cdot\left(w_1,w_2\right). \tag{8}$$

In particular, these derivatives are the covariance

$$\mathbf{D}^2_{ii} D_{p_\alpha, p_\alpha}\left(u,u\right)\cdot\left(w_1,w_2\right) = \mathrm{E}_q\left[\left(w_1 - \mathrm{E}_q\left[w_1\right]\right)\left(w_2 - \mathrm{E}_q\left[w_2\right]\right)\right]$$

with $q = e_{p_\alpha}\left(u\right)$.

We shall show that we can take the bilinear form $\mathbf{D}^2 K_{p_\alpha}\left(u\right) \in \mathcal{L}^2\left(B_q\right)$ as the extension to the non-parametric case of the Fisher metric.

Let $S = \left\{p_\theta = p\left(x,\theta\right) : \theta = \left(\theta^1,\ldots,\theta^n\right) \in \Theta \subseteq \mathbb{R}^n\right\}$ be a parametric n-dimensional statistical model on $\left(\mathcal{X}, \mathcal{X}, \mu\right)$. Given a point p_θ, the Fisher information matrix $G\left(\theta\right) = \left[g_{ij}\left(\theta\right)\right]$ of S at p_θ is defined by

$$g_{ij}\left(\theta\right) = \mathrm{E}_{p_\theta}\left[\frac{\partial \log p\left(x,\theta\right)}{\partial\theta^i}\frac{\partial \log p\left(x,\theta\right)}{\partial\theta^j}\right].$$

We assume that $g_{ij}\left(\theta\right)$ is finite for all i,j and θ and that G is positive definite so it determines a Riemannian metric, the so called Fisher metric. The Fisher matrix gives the second order approximation of the divergence:

$$D\left(p_{\theta_1} \,\|\, p_{\theta_0}\right) = \frac{1}{2} g_{ij}\left(\theta_0\right)\left(\theta^i_1 - \theta^i_0\right)\left(\theta^j_1 - \theta^j_0\right) + o\left(\left\|\theta_1 - \theta_0\right\|^2\right). \tag{9}$$

If we fix θ_0 and we consider the function $D\left(\cdot \,\|\, p_{\theta_0}\right) : S \to \mathbb{R}$, Eq. (9) means

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j} D\left(p_\theta \,\|\, p_{\theta_0}\right)\Big|_{\theta=\theta_0} = g_{ij}\left(\theta_0\right)$$

which can be compared with (8).

The next proposition is a technical result on Orlicz spaces to be applied to the proof of the regularity on the inclusion of the exponential manifold into the mixture manifold.

Proposition 38 *For each density $p \in \mathcal{M}$, the injection α of ${}^*B_p$ into $\left(B_p\right)^*$*

$$\alpha : {}^*B_p \ni {}^*u \mapsto \mathrm{E}_p\left[{}^*u \,\cdot\right] \in \left(B_p\right)^*$$

has closed range $\operatorname{im}\alpha$*, therefore* $\operatorname{im}\alpha$ *is a subspace of $\left(B_p\right)^*$ and α is a isomorphism of ${}^*B_p$ into its range in $\left(B_p\right)^*$.*

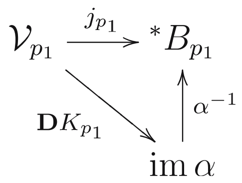
Proof From the general theory of Orlicz spaces, see Rao and Ren (2002) the dual space of $L^{\Psi_1}(p)$ is $L^{\Phi_1}(p)$ and there is a continuous injection of $L^{\Psi_1}(p)$ into the dual of $L^{\Phi_1}(p)$. The proposition follows by checking the effect of considering centered random variables. □

Proposition 39 *For each density $p \in \mathcal{M}$, the inclusion $j : \mathcal{E}(p) \hookrightarrow {}^*\mathcal{E}(p)$ is of class $\mathcal{C}^\infty$.*

Proof For each $p_1 \in \mathcal{E}(p)$ the local representative of j relative to the pair of charts $\left(\mathcal{U}_{p_1}, s_{p_1}\right)$ of $\mathcal{E}(p)$ and $\left({}^*\mathcal{U}_{p_1}, \eta_{p_1}\right)$ of ${}^*\mathcal{E}(p)$ is

$$j_{p_1} = \eta_{p_1} \circ j \circ e_{p_1} : \mathcal{V}_{p_1} \ni u \mapsto \mathrm{e}^{u - K_{p_1}(u)} - 1 \in {}^*B_p.$$

Observe that, for all $u \in \mathcal{V}_{p_1}$, $j_{p_1}(u) = \nabla K_{p_1}(u) = \alpha^{-1}\left(\mathbf{D}K_{p_1}(u)\right)$, that is the following diagram



is commutative. This shows that the map $j_{p_1} = \alpha^{-1} \circ \mathbf{D}K_{p_1}$ belongs to $\mathcal{C}^\infty\left(\mathcal{V}_{p_1}\right)$.

We shall show that the derivative of j_{p_1} is the mapping

$$\mathbf{D}j_{p_1} = \mathbf{D}\nabla K_{p_1} : \mathcal{V}_{p_1} \ni u \to \frac{q}{p}\left(\cdot - \mathrm{E}_q\left[\,\cdot\,\right]\right) \in \mathcal{L}\left(B_{p_1}, \left({}^*B_{p_1}\right)\right),$$

where $q = e_{p_1}(u)$. In fact, employing the chain rule and the linearity of α^{-1}, we evaluate the derivative at u applied to $w \in B_{p_1}$:

$$\begin{aligned}
\mathbf{D}j_{p_1}(u) \cdot w &= \alpha^{-1}\left(\mathbf{D}^2 K_{p_1}(u) \cdot (w, \cdot)\right) \\
&= \alpha^{-1}\left(\mathrm{E}_p\left[\frac{q}{p}\left(w - \mathrm{E}_q[w]\right)\,\cdot\right]\right) \\
&= \frac{q}{p}\left(w - \mathrm{E}_q[w]\right)
\end{aligned}$$

$((w - \mathrm{E}_q[w])\, q/p \in {}^*B_{p_1}$, see Pistone and Rogantin (1999, Proposition. 16-f)).
□

7 Discussion

In this paper we are following a specific track to the development of Information Geometry, i.e. the construction of a classical manifold structure. This is done by developing in the natural way the original suggestion of Efron to look at the exponential structure. Other options are present in the literature.

The most classical and most successful non-parametric construction of Information Geometry is based on the embedding $p \mapsto 2\sqrt{p}$ from the probability density simplex into the L^2 sphere or radius 2, followed by the pull-back of the geometry of the sphere to the probability density simplex. The L^2-sphere is a Riemannian manifold, but the embedding cannot define an atlas because the co-domain of the embedding has empty interior. Variants of this basic Hilbert embedding were used, see e.g. Burdet, Combe and Nencka (2001). In the same vein, Eguchi (2005) has a different L_0^2 representation based on the mapping $u \mapsto \frac{1}{2} - \frac{1}{2}\sigma^2(u) + \frac{1}{2}(1-u)^2 = g$ which is defined on the unit L_0^2 open ball and takes its values in the set of densities which are bounded below by a positive constant. If u is bounded, then g is bounded above and away from zero. The choice of the mentioned Authors to look for an hilbertian structure is actually the best from the point of view of Statistics, because it is grounded on the original idea due to Rao of looking to Fisher information as a metric tensor.

The discovery of the duality between the exponential and the manifold structure by Nagaoka together with our approach to the m-manifold , could lead to an other intermediate option, i.e. to define a manifold were the regularity of the maps is defined in a weak sense. See also the discussion in Zhang and Hästö (2006).

The real need for Statistics of the manifold structure we insist on is of course questionable. Our construction will eventually be fully justified only if all the basic structures of interest in Statistics will be embedded in the framework we propose, which is not fully done at this stage.

Acknowledgements We gratefully acknowledge the contribution of suggestions, comments and critics made by the two anonymous Referees of this Journal. We thank D. Imparato and B. Trivellato for the careful reading of the first version of this paper.

References

Amari, S. (1982). Differential geometry of curved exponential families. Curvature and information loss. *The Annals of Statistics*, *10*, 357–387.

Amari, S. (1985). *Differential-geometrical methods in statistics*. In *Lecture Notes in Statistics, Vol. 28*. New York Berlin Heidelberg: Springer.

Amari, S., Nagaoka, H. (2000). *Methods of information geometry (translated from the 1993 Japanese original by Daishi Harada)*. Providence, RI: American Mathematical Society.

Bourbaki, N. (1971). Variétés differentielles et analytiques. Fascicule de résultats / Paragraphes 1 à 7. No. XXXIII in Éléments de mathématiques, Hermann, Paris.

Burdet, G., Combe, P., Nencka, H. (2001). On real Hilbertian info-manifolds. In P. Sollich et al. (Eds.), *Disordered and complex systems*. Proceedings of the international conference, London, UK, July 10–14, 2000. Melville, NY: American Institute of Physics. AIP Conf. Proc. *553*, 153–158 (2001).

Cena, A. (2002). Geometric structures on the non-parametric statistical manifold. PhD Thesis. Dottorato in Matematica, Università di Milano.

Eguchi, S. (2005). Tubular modelling approach to statistical method for observational studies. In 2nd International symposium on information geometry and its applications, Tokyo.

Gibilisco, P., Pistone, G. (1998). Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, *1*(2), 325–347.

Grasselli, M. R. (2001). Dual connections in nonparametric classical information geometry. Tech. Rep. math-ph/0104031 v1, arXiv.

Krasnosel'skii, M. A., Rutickii, Y. B. (1961). Convex functions and Orlicz spaces. Noordhoff, Groningen, russian original: (1958) Fizmatgiz, Moskva.

Lang, S. (1995). Differential and Riemannian manifolds. Graduate texts in mathematics, Vol. 160, 3rd edn. New York Berlin Heidelberg: Springer.

Pistone, G. (2005). Advances in the geometry of non-parametric exponential models. In Proceedings of the 2nd international symposium on information geometry and its applications, December 12–16, 2005, University of Tokyo (pp. 72–81).

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

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

Rao, M. M., Ren, Z. D. (2002). Applications of Orlicz spaces. *Monographs and textbooks in pure and applied mathematics*, Vol. 250. New York: Marcel Dekker Inc.

Upmeier, H. (1985). *Symmetric Banach manifolds and Jordan C^*-algebras*. North-Holland Mathematics Studies, Vol. 104. Notas de Matemática [Mathematical Notes], 96. Amsterdam: North-Holland Publishing Co.

Zhang, J., Hästö, P. (2006). Statistical manifold as an affine space: a functional equation approach. *Journal of Mathematical Psychology*, *50*(1), 60–65.