J Theor Probab (2013) 26:870–884
DOI 10.1007/s10959-011-0400-5

On φ-Families of Probability Distributions

Rui F. Vigelis · Charles C. Cavalcante

Received: 1 June 2011 / Revised: 27 September 2011 / Published online: 21 December 2011

Abstract We generalize the exponential family of probability distributions. In our approach, the exponential function is replaced by a φ-function, resulting in a φ-family of probability distributions. We show how φ-families are constructed. In a φ-family, the analogue of the cumulant-generating function is a normalizing function. We define the φ-divergence as the Bregman divergence associated to the normalizing function, providing a generalization of the Kullback–Leibler divergence. A formula for the φ-divergence where the φ-function is the Kaniadakis κ-exponential function is derived.

Keywords Exponential family of probability distributions · Musielak–Orlicz spaces · Bregman divergence

Mathematics Subject Classification (2000) 60E05

1 Introduction

Let (T, Σ, μ) be a σ-finite, non-atomic measure space. We denote by $\mathcal{P}_\mu = \mathcal{P}(T, \Sigma, \mu)$ the family of all probability measures on T that are equivalent to the measure μ. The probability family $\mathcal{P}_\mu$ can be represented as (we adopt the same

R.F. Vigelis (✉)
Computer Engineering, Campus Sobral, Federal University of Ceará, Fortaleza, CE, Brazil
e-mail: rfvigelis@gmail.com

C.C. Cavalcante
Wireless Telecommunication Research Group, Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza, CE, Brazil
e-mail: charles@gtel.ufc.br

symbol $\mathcal{P}_\mu$ for this representation)

$$\mathcal{P}_\mu = \{ p \in L^0 : p > 0 \text{ and } \mathbb{E}[p] = 1 \},$$

where L^0 is the linear space of all real-valued, measurable functions on T, with equality μ-a.e., and $\mathbb{E}[\cdot]$ denotes the expectation with respect to the measure μ.

The family $\mathcal{P}_\mu$ can be equipped with a structure of C^∞-Banach manifold, using the Orlicz space $L^{\Phi_1}(p) = L^{\Phi_1}(T, \Sigma, p \cdot \mu)$ associated to the Orlicz function $\Phi_1(u) = \exp(u) - 1$, for $u \geq 0$. With this structure, $\mathcal{P}_\mu$ is called the *exponential statistical manifold*, whose construction was proposed in [15] and developed in [3, 5, 14]. Each connected component of the exponential statistical manifold gives rise to an *exponential family of probability distributions* $\mathcal{E}_p$ (for each $p \in \mathcal{P}_\mu$). Each element of $\mathcal{E}_p$ can be expressed as

$$\boldsymbol{e}_p(u) = e^{u - K_p(u)} p, \quad \text{for } u \in \mathcal{B}_p, \tag{1}$$

for a subset $\mathcal{B}_p$ of the Orlicz space $L^{\Phi_1}(p)$. K_p is the cumulant-generating functional $K_p(u) = \log \mathbb{E}_p[e^u]$, where $\mathbb{E}_p[\cdot]$ is the expectation with respect to $p \cdot \mu$. If c is a measurable function such that $p = e^c$, then (1) can be rewritten as

$$\boldsymbol{e}_p(u) = e^{c + u - K_p(u) \cdot \mathbf{1}_T}, \quad \text{for } u \in \mathcal{B}_p, \tag{2}$$

where $\mathbf{1}_A$ is the indicator function of a subset $A \subseteq T$. A generalization of expression (1) was given in [13], where the exponential function is replaced by a κ-exponential function. In our generalization, we make use of expression (2).

In the *φ-family of probability distributions* $\mathcal{F}_c^\varphi$, which we propose, the exponential function is replaced by the so called *φ-function* $\varphi: T \times \overline{\mathbb{R}} \to [0, \infty]$. The function $\varphi(t, \cdot)$ has a “shape” which is similar to that of an exponential function, with an arbitrary rate of increasing. For example, we found that the κ-exponential function satisfies the definition of φ-functions. As in the exponential family, the φ-families are the connected component of $\mathcal{P}_\mu$, which is endowed with a structure of C^∞-Banach manifold, using φ in the place of an exponential function. Let c be any measurable function such that $\varphi(t, c(t))$ belongs to $\mathcal{P}_\mu$. The elements of the φ-family of probability distributions $\mathcal{F}_c^\varphi$ are given by

$$\boldsymbol{\varphi}_c(u)(t) = \varphi\big(t, c(t) + u(t) - \psi(u) u_0(t)\big), \quad \text{for } u \in \mathcal{B}_c^\varphi, \tag{3}$$

for a subset $\mathcal{B}_c^\varphi$ of a Musielak–Orlicz space L_c^φ. The *normalizing function* $\psi: \mathcal{B}_c^\varphi \to [0, \infty)$ and the measurable function $u_0: T \to [0, \infty)$ in (3) replaces K_p and $\mathbf{1}_T$ in (2), receptively. The function u_0 is not arbitrary. In the text, we will show how u_0 can be chosen.

We define the *φ-divergence* as the Bregman divergence associated to the normalizing function ψ, providing a generalization of the Kullback–Leibler divergence. Then geometrical aspects related to the φ-family can be developed, since the Fisher information (on which the Information Geometry [1, 9] is based) is derived from the divergence. A formula for the φ-divergence where the φ-function is the Kaniadakis' κ-exponential function [6, 11] is derived, which we called the *κ-divergence*.

We expect that an extension of our work will provide advances in other areas, like in Information Geometry or in the non-parametric, non-commutative setting [4, 12]. The rest of this paper is organized as follows. Section 2 deals with the topics of

Musielak–Orlicz spaces we will use in the construction of the φ-family of probability distributions. In Sect. 3, the exponential statistical manifold is reviewed. The construction of the φ-family of probability distributions is given in Sect. 4. Finally, the φ-divergence is derived in Sect. 5.

2 Musielak–Orlicz Spaces

In this section we provide a brief introduction to Musielak–Orlicz (function) spaces, which are used in the construction of the exponential and φ-families. A more detailed exposition about these spaces can be found in [7, 10, 16].

We say that $\Phi: T \times [0,\infty] \to [0,\infty]$ is a *Musielak–Orlicz function* when, for μ-a.e. $t \in T$,

(i) $\Phi(t,\cdot)$ is convex and lower semi-continuous,
(ii) $\Phi(t,0) = \lim_{u\downarrow 0} \Phi(t,u) = 0$ and $\Phi(t,\infty) = \infty$,
(iii) $\Phi(\cdot,u)$ is measurable for all $u \geq 0$.

Items (i)–(ii) guarantee that $\Phi(t,\cdot)$ is not equal to 0 or ∞ on the interval $(0,\infty)$. A Musielak–Orlicz function Φ is said to be an *Orlicz function* if the functions $\Phi(t,\cdot)$ are identical for μ-a.e. $t \in T$.

Define the functional $I_\Phi(u) = \int_T \Phi(t,|u(t)|)\,d\mu$, for any $u \in L^0$. The *Musielak–Orlicz space*, *Musielak–Orlicz class*, and *Morse–Transue space*, are given by

$$L^\Phi = \{u \in L^0 : I_\Phi(\lambda u) < \infty \text{ for some } \lambda > 0\},$$
$$\tilde{L}^\Phi = \{u \in L^0 : I_\Phi(u) < \infty\},$$

and

$$E^\Phi = \{u \in L^0 : I_\Phi(\lambda u) < \infty \text{ for all } \lambda > 0\},$$

respectively. If the underlying measure space (T,Σ,μ) have to be specified, we write $L^\Phi(T,\Sigma,\mu)$, $\tilde{L}^\Phi(T,\Sigma,\mu)$ and $E^\Phi(T,\Sigma,\mu)$ in the place of L^Φ, $\tilde{L}^\Phi$ and E^Φ, respectively. Clearly, $E^\Phi \subseteq \tilde{L}^\Phi \subseteq L^\Phi$. The Musielak–Orlicz space L^Φ can be interpreted as the smallest vector subspace of L^0 that contains $\tilde{L}^\Phi$, and E^Φ is the largest vector subspace of L^0 that is contained in $\tilde{L}^\Phi$.

The Musielak–Orlicz space L^Φ is a Banach space when it is endowed with the *Luxemburg norm*

$$\|u\|_\Phi = \inf\left\{\lambda > 0 : I_\Phi\left(\frac{u}{\lambda}\right) \leq 1\right\},$$

or the *Orlicz norm*

$$\|u\|_{\Phi,0} = \sup\left\{\left|\int_T uv\,d\mu\right| : v \in \tilde{L}^{\Phi^*} \text{ and } I_{\Phi^*}(v) \leq 1\right\},$$

where $\Phi^*(t,v) = \sup_{u\geq 0}(uv - \Phi(t,u))$ is the *Fenchel conjugate* of $\Phi(t,\cdot)$. These norms are equivalent and the inequalities $\|u\|_\Phi \leq \|u\|_{\Phi,0} \leq 2\|u\|_\Phi$ hold for all $u \in L^\Phi$.

If we can find a non-negative function $f \in \tilde{L}^\Phi$ and a constant $K > 0$ such that

$$\Phi(t, 2u) \le K\Phi(t, u), \quad \text{for all } u \ge f(t),$$

then we say that Φ satisfies the Δ_2-*condition*, or belong to the Δ_2-*class* (denoted by $\Phi \in \Delta_2$). When the Musielak–Orlicz function Φ satisfies the Δ_2-*condition*, E^Φ coincides with L^Φ. On the other hand, if Φ is finite-valued and does not satisfy the Δ_2-*condition*, then the Musielak–Orlicz class $\tilde{L}^\Phi$ is not open and its interior coincides with

$$B_0\big(E^\Phi, 1\big) = \Big\{u \in L^\Phi : \inf_{v \in E^\Phi} \|u - v\|_{\Phi,0} < 1\Big\},$$

or, equivalently, $B_0(E^\Phi, 1) \subsetneq \tilde{L}^\Phi \subsetneq \overline{B}_0(E^\Phi, 1)$.

3 The Exponential Statistical Manifold

This section starts with the definition of a C^k-Banach manifold [8]. A C^k-*Banach manifold* is a set M and a collection of pairs $(U_\alpha, \boldsymbol{x}_\alpha)$ (α belonging to some indexing set), composed by open subsets U_α of some Banach space X_α, and injective mappings $\boldsymbol{x}_\alpha : U_\alpha \to M$, satisfying the following conditions:

(bm1) the sets $\boldsymbol{x}_\alpha(U_\alpha)$ cover M, i.e., $\bigcup_\alpha \boldsymbol{x}_\alpha(U_\alpha) = M$;
(bm2) for any pair of indices α, β such that $\boldsymbol{x}_\alpha(U_\alpha) \cap \boldsymbol{x}_\beta(U_\beta) = W \neq \emptyset$, the sets $\boldsymbol{x}_\alpha^{-1}(W)$ and $\boldsymbol{x}_\beta^{-1}(W)$ are open in X_α and X_β, respectively; and
(bm3) the *transition map* $\boldsymbol{x}_\beta^{-1} \circ \boldsymbol{x}_\alpha : \boldsymbol{x}_\alpha^{-1}(W) \to \boldsymbol{x}_\beta^{-1}(W)$ is a C^k-isomorphism.

The pair $(U_\alpha, \boldsymbol{x}_\alpha)$ with $p \in \boldsymbol{x}_\alpha(U_\alpha)$ is called a *parametrization* (or *system of coordinates*) of M at p; and $\boldsymbol{x}_\alpha(U_\alpha)$ is said to be a *coordinate neighborhood* at p.

The set M can be endowed with a topology in a unique way such that each $\boldsymbol{x}_\alpha(U_\alpha)$ is open, and the $\boldsymbol{x}_\alpha$'s are topological isomorphisms. We note that if $k \ge 1$ and two parametrizations $(U_\alpha, \boldsymbol{x}_\alpha)$ and $(U_\beta, \boldsymbol{x}_\beta)$ are such that $\boldsymbol{x}_\alpha(U_\alpha)$ and $\boldsymbol{x}_\beta(U_\beta)$ have a non-empty intersection, then from the derivative of $\boldsymbol{x}_\beta^{-1} \circ \boldsymbol{x}_\alpha$ we see that X_α and X_β are isomorphic.

Two collections $\{(U_\alpha, \boldsymbol{x}_\alpha)\}$ and $\{(V_\beta, \boldsymbol{x}_\beta)\}$ satisfying (bm1)–(bm3) are said to be C^k-*compatible* if their union also satisfies (bm1)–(bm3). It can be verified that the relation of C^k-compatibility is an equivalence relation. An equivalence class of C^k-compatible collections $\{(U_\alpha, \boldsymbol{x}_\alpha)\}$ on M is said to define a C^k-*differentiable structure* on X.

Now we review the construction of the exponential statistical manifold. We consider the Musielak–Orlicz space $L^{\Phi_1}(p) = L^{\Phi_1}(T, \Sigma, p \cdot \mu)$, where the Orlicz function $\Phi_1 : [0, \infty) \to [0, \infty)$ is given by $\Phi_1(u) = e^u - 1$, and p is a probability density in $\mathcal{P}_\mu$. The space $L^{\Phi_1}(p)$ corresponds to the set of all functions $u \in L^0$ whose *moment-generating function* $\widehat{u}_p(\lambda) = \mathbb{E}_p[e^{\lambda u}]$ is finite in a neighborhood of 0.

For every function $u \in L^0$ we define the *moment-generating functional*

$$M_p(u) = \mathbb{E}_p\big[e^u\big],$$

and the *cumulant-generating functional*

$$K_p(u) = \log M_p(u).$$

Clearly, these functionals are not expected to be finite for every $u \in L^0$. Denote by $\mathcal{K}_p$ the interior of the set of all functions $u \in L^{\Phi_1}(p)$ whose moment-generating functional $M_p(u)$ is finite. Equivalently, a function $u \in L^{\Phi_1}(p)$ belongs to $\mathcal{K}_p$ if and only if $M_p(\lambda u)$ is finite for every λ in some neighborhood of [0, 1]. The closed subspace of *p-centered* random variables

$$B_p = \left\{u \in L^{\Phi_1}(p) : \mathbb{E}_p[u] = 0\right\}$$

is taken to be the coordinate Banach space. The *exponential parametrization* $\boldsymbol{e}_p: \mathcal{B}_p \to \mathcal{E}_p$ maps $\mathcal{B}_p = B_p \cap \mathcal{K}_p$ to the *exponential family* $\mathcal{E}_p = \boldsymbol{e}_p(\mathcal{B}_p) \subseteq \mathcal{P}_\mu$, according to

$$\boldsymbol{e}_p(u) = e^{u - K_p(u)} p, \quad \text{for all } u \in \mathcal{B}_p.$$

$\boldsymbol{e}_p$ is a bijection from $\mathcal{B}_p$ to its image $\mathcal{E}_p = \boldsymbol{e}_p(\mathcal{B}_p)$, whose inverse $\boldsymbol{e}_p^{-1}: \mathcal{E}_p \to \mathcal{B}_p$ can be expressed as

$$\boldsymbol{e}_p^{-1}(q) = \log\left(\frac{q}{p}\right) - \mathbb{E}_p\left[\log\left(\frac{q}{p}\right)\right], \quad \text{for } q \in \mathcal{E}_p.$$

Since $K_p(u) < \infty$ for every $u \in \mathcal{K}_p$, we find that $\boldsymbol{e}_p$ can be extended to $\mathcal{K}_p$. The restriction of $\boldsymbol{e}_p$ to $\mathcal{B}_p$ guarantees that $\boldsymbol{e}_p$ is bijective.

Given two probability densities p and q in the same connected component of $\mathcal{P}_\mu$, the exponential probability families $\mathcal{E}_p$ and $\mathcal{E}_q$ coincide, and the exponential spaces $L^{\Phi_1}(p)$ and $L^{\Phi_1}(q)$ are isomorphic (see [14, Proposition 5]). Hence, $\mathcal{B}_p = \boldsymbol{e}_p^{-1}(\mathcal{E}_p \cap \mathcal{E}_q)$ and $\mathcal{B}_q = \boldsymbol{e}_q^{-1}(\mathcal{E}_p \cap \mathcal{E}_q)$. The transition map $\boldsymbol{e}_q^{-1} \circ \boldsymbol{e}_p : \mathcal{B}_p \to \mathcal{B}_q$, which can be written as

$$\boldsymbol{e}_q^{-1} \circ \boldsymbol{e}_p(u) = u + \log\left(\frac{p}{q}\right) - \mathbb{E}_q\left[u + \log\left(\frac{p}{q}\right)\right], \quad \text{for all } u \in \mathcal{B}_p,$$

is a C^∞-function. Clearly, $\bigcup_{p \in \mathcal{P}_\mu} e_p(\mathcal{B}_p) = \mathcal{P}_\mu$. Thus the collection $\{(\mathcal{B}_p, \boldsymbol{e}_p)\}_{p \in \mathcal{P}_\mu}$ satisfies (bm1)–(bm2). Hence $\mathcal{P}_\mu$ is a C^∞-Banach manifold, which is called the *exponential statistical manifold.*

4 Construction of the φ-Family of Probability Distributions

The generalization of the exponential family is based on the replacement of the exponential function by a *φ-function* $\varphi: T \times \overline{\mathbb{R}} \to [0, \infty]$ that satisfies the following properties, for μ-a.e. $t \in T$:

(a1) $\varphi(t, \cdot)$ is convex and injective,
(a2) $\varphi(t, -\infty) = 0$ and $\varphi(t, \infty) = \infty$,
(a3) $\varphi(\cdot, u)$ is measurable for all $u \in \mathbb{R}$.

In addition, we assume a positive, measurable function $u_0: T \to (0, \infty)$ can be found such that, for every measurable function $c: T \to \mathbb{R}$ for which $\varphi(t, c(t))$ is in $\mathcal{P}_\mu$, we have

(a4) $\varphi(t, c(t) + \lambda u_0(t))$ is μ-integrable for all $\lambda > 0$.

The choice for $\varphi(t,\cdot)$ injective with image $[0,\infty]$ is justified by the fact that a parametrization of $\mathcal{P}_\mu$ maps real-valued functions to positive functions. Moreover, by (a1), $\varphi(t,\cdot)$ is continuous and strictly increasing. From (a3), the function $\varphi(t,u(t))$ is measurable if and only if $u\colon T\to\mathbb{R}$ is measurable. Replacing $\varphi(t,u)$ by $\varphi(t,u_0(t)u)$, a "new" function $u_0=1$ is obtained, satisfying (a4).

Example 1 The *Kaniadakis' κ-exponential* $\exp_\kappa\colon\mathbb{R}\to(0,\infty)$ for $\kappa\in[-1,1]$ is defined as

$$\exp_\kappa(u)=\begin{cases}\left(\kappa u+\sqrt{1+\kappa^2u^2}\right)^{1/\kappa}, & \text{if } \kappa\neq 0,\\ \exp(u), & \text{if } \kappa=0.\end{cases}$$

The inverse of $\exp_\kappa$ is the *Kaniadakis' κ-logarithm*

$$\ln_\kappa(u)=\begin{cases}\frac{u^\kappa-u^{-\kappa}}{2\kappa}, & \text{if } \kappa\neq 0,\\ \ln(u), & \text{if } \kappa=0.\end{cases}$$

Some algebraic properties of the ordinary exponential and logarithm functions are preserved:

$$\exp_\kappa(u)\exp_\kappa(-u)=1,\qquad \ln_\kappa(u)+\ln_\kappa\left(u^{-1}\right)=0.$$

For a measurable function $\kappa\colon T\to[-1,1]$, we define the *variable κ-exponential* $\exp_\kappa\colon T\times\mathbb{R}\to(0,\infty)$ as

$$\exp_\kappa(t,u)=\exp_{\kappa(t)}(u),$$

whose inverse is called the *variable κ-logarithm*:

$$\ln_\kappa(t,u)=\ln_{\kappa(t)}(u).$$

Assuming that $\kappa_-=\operatorname{ess\,inf}|\kappa(t)|>0$, the variable κ-exponential $\exp_\kappa$ satisfies (a1)–(a4). The verification of (a1)–(a3) is easy. Moreover, we notice that $\exp_\kappa(t,\cdot)$ is strictly convex. We can write for $\alpha\geq 1$

$$\begin{aligned}\exp_\kappa(t,\alpha u)&=\left(\kappa(t)\alpha u+\alpha\sqrt{1/\alpha^2+\kappa(t)^2u^2}\right)^{1/\kappa(t)}\\&\leq\alpha^{1/|\kappa(t)|}\left(\kappa(t)u+\sqrt{1+\kappa(t)^2u^2}\right)^{1/\kappa(t)}\\&\leq\alpha^{1/\kappa_-}\exp_\kappa(t,u).\end{aligned}$$

By the convexity of $\exp_\kappa(t,\cdot)$, we obtain for any $\lambda\in(0,1)$

$$\begin{aligned}\exp_\kappa(t,c+u)&\leq\lambda\exp_\kappa\left(t,\lambda^{-1}c\right)+(1-\lambda)\exp_\kappa\left(t,(1-\lambda)^{-1}u\right)\\&\leq\lambda^{1-1/\kappa_-}\exp_\kappa(t,c)+(1-\lambda)^{1-1/\kappa_-}\exp_\kappa(t,u).\end{aligned}$$

Thus any positive function u_0 such that $\mathbb{E}[\exp_\kappa(u_0)]<\infty$ satisfies (a4).

Let $c\colon T\to\mathbb{R}$ be a measurable function such that $\varphi(t,c(t))$ is μ-integrable. We define the Musielak–Orlicz function

$$\Phi(t,u)=\varphi\big(t,c(t)+u\big)-\varphi\big(t,c(t)\big),$$

and denote L^Φ, $\tilde{L}^\Phi$ and E^Φ by L_c^φ, $\tilde{L}_c^\varphi$ and E_c^φ, respectively. Since $\varphi(t, c(t))$ is μ-integrable, the Musielak–Orlicz space L_c^φ corresponds to the set of all functions $u \in L^0$ for which $\varphi(t, c(t) + \lambda u(t))$ is μ-integrable for every λ contained in some neighborhood of 0.

Let $\mathcal{K}_c^\varphi$ be the set of all functions $u \in L_c^\varphi$ such that $\varphi(t, c(t) + \lambda u(t))$ is μ-integrable for every λ in a neighborhood of $[0, 1]$. Denote by $\boldsymbol{\varphi}$ the operator acting on the set of real-valued functions $u: T \to \mathbb{R}$ given by $\boldsymbol{\varphi}(u)(t) = \varphi(t, u(t))$. For each probability density $p \in \mathcal{P}_\mu$, we can take a measurable function $c: T \to \mathbb{R}$ such that $p = \boldsymbol{\varphi}(c)$. The first import result in the construction of the φ-family is given below.

Lemma 2 *The set $\mathcal{K}_c^\varphi$ is open in L_c^φ.*

Proof Take any $u \in \mathcal{K}_c^\varphi$. We can find $\varepsilon \in (0, 1)$ such that $\mathbb{E}[\boldsymbol{\varphi}(c + \alpha u)] < \infty$ for every $\alpha \in [-\varepsilon, 1 + \varepsilon]$. Let $\delta = [\frac{2}{\varepsilon}(1 + \varepsilon)(1 + \frac{\varepsilon}{2})]^{-1}$. For any function $v \in L_c^\varphi$ in the open ball $B_\delta = \{w \in L_c^\varphi : \|w\|_\Phi < \delta\}$, we have $I_\Phi(\frac{v}{\delta}) \le 1$. Thus $\mathbb{E}[\boldsymbol{\varphi}(c + \frac{1}{\delta}|v|)] \le 2$. Taking any $\alpha \in (0, 1 + \frac{\varepsilon}{2})$, we denote $\lambda = \frac{\alpha}{1+\varepsilon}$. In virtue of

$$\frac{\alpha}{1-\lambda} = \frac{\alpha}{1 - \frac{\alpha}{1+\varepsilon}} \le \frac{1 + \frac{\varepsilon}{2}}{1 - \frac{1+\frac{\varepsilon}{2}}{1+\varepsilon}} = \frac{2}{\varepsilon}(1 + \varepsilon)\left(1 + \frac{\varepsilon}{2}\right) = \frac{1}{\delta},$$

it follows that

$$\begin{aligned}
\boldsymbol{\varphi}\big(c + \alpha(u + v)\big) &= \boldsymbol{\varphi}\left(\lambda\left(c + \frac{\alpha}{\lambda}u\right) + (1 - \lambda)\left(c + \frac{\alpha}{1-\lambda}v\right)\right) \\
&\le \lambda\boldsymbol{\varphi}\left(c + \frac{\alpha}{\lambda}u\right) + (1 - \lambda)\boldsymbol{\varphi}\left(c + \frac{\alpha}{1-\lambda}v\right) \\
&\le \lambda\boldsymbol{\varphi}\big(c + (1 + \varepsilon)u\big) + (1 - \lambda)\boldsymbol{\varphi}\left(c + \frac{1}{\delta}|v|\right). \qquad (4)
\end{aligned}$$

For $\alpha \in (-\frac{\varepsilon}{2}, 0)$, we can write

$$\begin{aligned}
\boldsymbol{\varphi}\big(c + \alpha(u + v)\big) &\le \frac{1}{2}\boldsymbol{\varphi}(c + 2\alpha u) + \frac{1}{2}\boldsymbol{\varphi}(c + 2\alpha v) \\
&\le \frac{1}{2}\boldsymbol{\varphi}(c + 2\alpha u) + \frac{1}{2}\boldsymbol{\varphi}\big(c + |v|\big). \qquad (5)
\end{aligned}$$

By (4) and (5), we get $\mathbb{E}[\boldsymbol{\varphi}(c + \alpha(u + v))] < \infty$, for any $\alpha \in (-\frac{\varepsilon}{2}, 1 + \frac{\varepsilon}{2})$. Hence the ball of radius δ centered at u is contained in $\mathcal{K}_c^\varphi$. Therefore, the set $\mathcal{K}_c^\varphi$ is open. □

Clearly, for $u \in \mathcal{K}_c^\varphi$ the function $\boldsymbol{\varphi}(c+u)$ is not necessarily in $\mathcal{P}_\mu$. The *normalizing function* $\psi: \mathcal{K}_c^\varphi \to \mathbb{R}$ is introduced in order to make the density

$$\boldsymbol{\varphi}\big(c + u - \psi(u)u_0\big)$$

contained in $\mathcal{P}_\mu$, for any $u \in \mathcal{K}_c^\varphi$. We have to find the functions for which the normalizing function exists. For a function $u \in L_c^\varphi$, suppose that $\boldsymbol{\varphi}(c + u - \alpha u_0)$ is μ-integrable for some $\alpha \in \mathbb{R}$. Then u is in the closure of the set $\mathcal{K}_c^\varphi$. Indeed, for any $\lambda \in (0, 1)$,

$$\begin{aligned}\boldsymbol{\varphi}(c+\lambda u) &= \boldsymbol{\varphi}\left(\lambda(c+u-\alpha u_0)+(1-\lambda)\left(c+\frac{\lambda}{1-\lambda}\alpha u_0\right)\right)\\ &\le \lambda\boldsymbol{\varphi}(c+u-\alpha u_0)+(1-\lambda)\boldsymbol{\varphi}\left(c+\frac{\lambda}{1-\lambda}\alpha u_0\right).\end{aligned}$$

Since the function u_0 satisfies (a4), we see that $\boldsymbol{\varphi}(c+\lambda u)$ is μ-integrable. Hence the maximal, open domain of ψ is contained in $\mathcal{K}_c^{\varphi}$.

Proposition 3 *If the function u is in $\mathcal{K}_c^{\varphi}$, then there exists a unique $\psi(u) \in \mathbb{R}$ for which $\boldsymbol{\varphi}(c+u-\psi(u)u_0)$ is a probability density in $\mathcal{P}_\mu$.*

Proof We will show that if the function u is in $\mathcal{K}_c^{\varphi}$, then $\boldsymbol{\varphi}(c+u+\alpha u_0)$ is μ-integrable for every $\alpha \in \mathbb{R}$. Since u is in $\mathcal{K}_c^{\varphi}$, we can find $\varepsilon > 0$ such that $\boldsymbol{\varphi}(c+(1+\varepsilon)u)$ is μ-integrable. Taking $\lambda = \frac{1}{1+\varepsilon}$, we can write

$$\begin{aligned}\boldsymbol{\varphi}(c+u+\alpha u_0) &= \boldsymbol{\varphi}\left(\lambda\left(c+\frac{1}{\lambda}u\right)+(1-\lambda)\left(c+\frac{1}{1-\lambda}\alpha u_0\right)\right)\\ &\le \lambda\boldsymbol{\varphi}\left(c+\frac{1}{\lambda}u\right)+(1-\lambda)\boldsymbol{\varphi}\left(c+\frac{1}{1-\lambda}\alpha u_0\right).\end{aligned}$$

Thus $\boldsymbol{\varphi}(c+u+\alpha u_0)$ is μ-integrable. By the Dominated Convergence Theorem, the map $\alpha \mapsto J(\alpha) = \mathbb{E}[\boldsymbol{\varphi}(c+u+\alpha u_0)]$ is continuous, tends to 0 as $\alpha \to -\infty$, and goes to infinity as $\alpha \to \infty$. Since $\varphi(t,\cdot)$ is strictly increasing, it follows that $J(\alpha)$ is also strictly increasing. Therefore, there exists a unique $\psi(u) \in \mathbb{R}$ for which $\boldsymbol{\varphi}(c+u-\psi(u)u_0)$ is a probability density in $\mathcal{P}_\mu$. □

The function $\psi: \mathcal{K}_c^{\varphi} \to \mathbb{R}$ can take both positive and negative values. However, if the domain of ψ is restricted to a subspace of L_c^{φ}, its image will be contained in $[0,\infty)$. We denote by $\boldsymbol{\varphi}'_+$ the operator acting on the set of real-valued functions $u: T \to \mathbb{R}$ given by $\boldsymbol{\varphi}'_+(u)(t) = \varphi'_+(t,u(t))$, where $\varphi'_+(t,\cdot)$ is the right-derivative of $\varphi(t,\cdot)$. Define the closed subspace

$$B_c^{\varphi} = \left\{u \in L_c^{\varphi} : \mathbb{E}\left[u\boldsymbol{\varphi}'_+(c)\right] = 0\right\},$$

and let $\mathcal{B}_c^{\varphi} = B_c^{\varphi} \cap \mathcal{K}_c^{\varphi}$. By the convexity of $\varphi(t,\cdot)$, we have

$$u\varphi'_+\big(t,c(t)\big) \le \varphi\big(t,c(t)+u\big) - \varphi\big(t,c(t)\big), \quad \text{for all } u \in \mathbb{R}.$$

Hence, for any $u \in \mathcal{B}_c^{\varphi}$, we get

$$1 = \mathbb{E}\left[u\boldsymbol{\varphi}'_+(c)\right] + \mathbb{E}\left[\boldsymbol{\varphi}(c)\right] \le \mathbb{E}\left[\boldsymbol{\varphi}(c+u)\right] < \infty.$$

Thus it follows that $\psi(u) \ge 0$ in order to find that $\boldsymbol{\varphi}(c+u-\psi(u)u_0)$ is in $\mathcal{P}_\mu$.

For each measurable function $c: T \to \mathbb{R}$ such that $p = \boldsymbol{\varphi}(c)$ is the probability density in $\mathcal{P}_\mu$, we associate a parametrization $\boldsymbol{\varphi}_c: \mathcal{B}_c^{\varphi} \to \mathcal{F}_c^{\varphi}$ that maps any function u in $\mathcal{B}_c^{\varphi}$ to a probability density in $\mathcal{F}_c^{\varphi} = \varphi_c(\mathcal{B}_c^{\varphi}) \subseteq \mathcal{P}_\mu$ according to

$$\boldsymbol{\varphi}_c(u) = \boldsymbol{\varphi}\big(c+u-\psi(u)u_0\big).$$

Clearly, we have $\mathcal{P}_\mu = \bigcup\{\mathcal{F}_c^\varphi : \boldsymbol{\varphi}(c) \in \mathcal{P}_\mu\}$. Moreover, the map $\boldsymbol{\varphi}_c$ is a bijection from $\mathcal{B}_c^\varphi$ to $\mathcal{F}_c^\varphi$. If the functions $u, v \in \mathcal{B}_c^\varphi$ are such that $\boldsymbol{\varphi}_c(u) = \boldsymbol{\varphi}_c(v)$, then the difference $u - v = (\psi(u) - \psi(v))u_0$ is in B_c^φ. Consequently, $\psi(u) = \psi(v)$ and then $u = v$.

Suppose that the measurable functions $c_1, c_2 : T \to \mathbb{R}$ are such that $p_1 = \boldsymbol{\varphi}(c_1)$ and $p_2 = \boldsymbol{\varphi}(c_2)$ belong to $\mathcal{P}_\mu$. The parametrizations $\boldsymbol{\varphi}_{c_1} : \mathcal{B}_{c_1}^\varphi \to \mathcal{F}_{c_1}^\varphi$ and $\boldsymbol{\varphi}_{c_2} : \mathcal{B}_{c_2}^\varphi \to \mathcal{F}_{c_2}^\varphi$ related to these functions have transition map

$$\boldsymbol{\varphi}_{c_2}^{-1} \circ \boldsymbol{\varphi}_{c_1} : \boldsymbol{\varphi}_{c_1}^{-1}\big(\mathcal{F}_{c_1}^\varphi \cap \mathcal{F}_{c_2}^\varphi\big) \to \boldsymbol{\varphi}_{c_2}^{-1}\big(\mathcal{F}_{c_1}^\varphi \cap \mathcal{F}_{c_2}^\varphi\big).$$

Let $\psi_1 : \mathcal{B}_{c_1}^\varphi \to [0, \infty)$ and $\psi_2 : \mathcal{B}_{c_2}^\varphi \to [0, \infty)$ be the normalizing functions associated to c_1 and c_2, respectively. Assume that the functions $u \in \mathcal{B}_{c_1}^\varphi$ and $v \in \mathcal{B}_{c_2}^\varphi$ are such that $\boldsymbol{\varphi}_{c_1}(u) = \boldsymbol{\varphi}_{c_2}(v) \in \mathcal{F}_{c_1}^\varphi \cap \mathcal{F}_{c_2}^\varphi$. Then we can write

$$v = c_1 - c_2 + u - \big(\psi_1(u) - \psi_2(v)\big)u_0.$$

Since the function v is in $B_{c_2}^\varphi$, if we multiply this equation by $\boldsymbol{\varphi}'_+(c_2)$ and integrate with respect to the measure μ, we obtain

$$0 = \mathbb{E}\big[(c_1 - c_2 + u)\boldsymbol{\varphi}'_+(c_2)\big] - \big(\psi_1(u) - \psi_2(v)\big)\mathbb{E}\big[u_0\boldsymbol{\varphi}'_+(c_2)\big].$$

Thus the transition map $\boldsymbol{\varphi}_{c_2}^{-1} \circ \boldsymbol{\varphi}_{c_1}$ can be expressed as

$$\boldsymbol{\varphi}_{c_2}^{-1} \circ \boldsymbol{\varphi}_{c_1}(w) = c_1 - c_2 + w - \frac{\mathbb{E}[(c_1 - c_2 + w)\boldsymbol{\varphi}'_+(c_2)]}{\mathbb{E}[u_0\boldsymbol{\varphi}'_+(c_2)]}u_0, \tag{6}$$

for every $w \in \boldsymbol{\varphi}_{c_1}^{-1}(\mathcal{F}_{c_1}^\varphi \cap \mathcal{F}_{c_2}^\varphi)$. Clearly, this transition map will be of class C^∞ if we show that the functions w and $c_1 - c_2$ are in $L_{c_2}^\varphi$, and the spaces $L_{c_1}^\varphi$ and $L_{c_2}^\varphi$ have equivalent norms. It is not hard to verify that if two Musielak–Orlicz spaces are equal as sets, then their norms are equivalent (see [10, Theorem 8.5]). We make use of the following:

Proposition 4 *Assume that the measurable functions $\widetilde{c}, c : T \to \mathbb{R}$ satisfy $\mathbb{E}[\varphi(t, \widetilde{c}(t))] < \infty$ and $\mathbb{E}[\varphi(t, c(t))] < \infty$. Then $L_{\widetilde{c}}^\varphi \subseteq L_c^\varphi$ if and only if $\widetilde{c} - c \in L_c^\varphi$.*

Proof Suppose that $\widetilde{c} - c$ is not in L_c^φ. Let $A = \{t \in T : \widetilde{c}(t) < c(t)\}$. For $\lambda \in [0, 1]$, we have

$$\begin{aligned}\mathbb{E}\big[\boldsymbol{\varphi}\big(c + \lambda(\widetilde{c} - c)\big)\big] &= \mathbb{E}\big[\boldsymbol{\varphi}\big(c + \lambda(\widetilde{c} - c)\big)\mathbf{1}_{T\setminus A}\big] + \mathbb{E}\big[\boldsymbol{\varphi}\big(c + \lambda(\widetilde{c} - c)\big)\mathbf{1}_A\big] \\ &\le \mathbb{E}\big[\boldsymbol{\varphi}\big(c + (\widetilde{c} - c)\big)\mathbf{1}_{T\setminus A}\big] + \mathbb{E}\big[\boldsymbol{\varphi}(c)\mathbf{1}_A\big] \\ &\le \mathbb{E}\big[\boldsymbol{\varphi}(\widetilde{c})\big] + \mathbb{E}\big[\boldsymbol{\varphi}(c)\big] < \infty.\end{aligned}$$

Since $\widetilde{c} - c \notin L_c^\varphi$, for any $\lambda > 0$, there holds $\mathbb{E}[\boldsymbol{\varphi}(c - \lambda(\widetilde{c} - c))] = \infty$. From

$$\begin{aligned}\mathbb{E}\big[\boldsymbol{\varphi}\big(c - \lambda(\widetilde{c} - c)\big)\big] &= \mathbb{E}\big[\boldsymbol{\varphi}\big(c - \lambda(\widetilde{c} - c)\big)\mathbf{1}_{T\setminus A}\big] + \mathbb{E}\big[\boldsymbol{\varphi}\big(c - \lambda(\widetilde{c} - c)\big)\mathbf{1}_A\big] \\ &\le \mathbb{E}\big[\boldsymbol{\varphi}\big(c + \lambda(c - \widetilde{c})\big)\mathbf{1}_A\big],\end{aligned}$$

we see that $(c - \widetilde{c})\mathbf{1}_A$ does not belong to L_c^φ. Clearly, $(c - \widetilde{c})\mathbf{1}_A \in L_{\widetilde{c}}^\varphi$. Consequently, $L_{\widetilde{c}}^\varphi$ is not contained in L_c^φ.

Conversely, assume $\widetilde{c} - c \in L_c^\varphi$. Let w be any function in $L_{\widetilde{c}}^\varphi$. We can find $\varepsilon > 0$ such that $\mathbb{E}[\boldsymbol{\varphi}(\widetilde{c} + \lambda w)] < \infty$, for every $\lambda \in (-\varepsilon, \varepsilon)$. Consider the convex function

$$g(\alpha,\lambda)=\mathbb{E}\big[\boldsymbol{\varphi}\big(c+\alpha(\widetilde{c}-c)+\lambda w\big)\big].$$

This function is finite for $\lambda=0$ and α in the interval $(-\eta,1]$, for some $\eta>0$. Moreover, $g(1,\lambda)$ is finite for every $\lambda\in(-\varepsilon,\varepsilon)$. By the convexity of g, we see that g is finite in the convex hull of the set $1\times(-\varepsilon,\varepsilon)\cup(-\eta,1]\times 0$. We find that $g(0,\lambda)$ is finite for every λ in some neighborhood of 0. Consequently, $w\in L_c^{\varphi}$. Since $w\in L_{\widetilde{c}}^{\varphi}$ is arbitrary, the inclusion $L_{\widetilde{c}}^{\varphi}\subseteq L_c^{\varphi}$ follows. □

Lemma 5 *If the function u is in $\mathcal{K}_c^{\varphi}$ and we denote $\widetilde{c}=c+u-\psi(u)u_0$, then the spaces L_c^{φ} and $L_{\widetilde{c}}^{\varphi}$ are equal as sets.*

Proof The inclusion $L_{\widetilde{c}}^{\varphi}\subseteq L_c^{\varphi}$ follows from Proposition 4. Since $u\in\mathcal{K}_c^{\varphi}$, we have

$$\mathbb{E}\big[\boldsymbol{\varphi}(\widetilde{c}+\lambda u)\big]\le\mathbb{E}\big[\boldsymbol{\varphi}\big(c+(1+\lambda)u\big)\big]<\infty,$$

for every λ in a neighborhood of 0. Thus $c-\widetilde{c}=-u+\psi(u)u_0$ belongs to $L_{\widetilde{c}}^{\varphi}$. From Proposition 4, we obtain $L_{\widetilde{c}}^{\varphi}\subseteq L_c^{\varphi}$. □

By Lemma 5, if we denote $c_1+u-\psi_1(u)u_0=\widetilde{c}=c_2+v-\psi_2(v)u_0$, we find that the spaces $L_{c_1}^{\varphi}$, $L_{\widetilde{c}}^{\varphi}$ and $L_{c_2}^{\varphi}$ are equal as sets. In (6), the function w is in $L_{c_2}^{\varphi}$ and consequently c_1-c_2 is in $L_{c_2}^{\varphi}$. Therefore, the transition map $\boldsymbol{\varphi}_{c_2}^{-1}\circ\boldsymbol{\varphi}_{c_1}$ is of class C^{∞}.

Since $\boldsymbol{\varphi}_{c_2}^{-1}\circ\boldsymbol{\varphi}_{c_1}$ is of class C^{∞}, the set $\boldsymbol{\varphi}_{c_1}^{-1}(\mathcal{F}_{c_1}^{\varphi}\cap\mathcal{F}_{c_2}^{\varphi})$ is open $B_{c_1}^{\varphi}$. The φ-families $\mathcal{F}_c^{\varphi}$ are maximal in the sense that if two φ-families $\mathcal{F}_{c_1}^{\varphi}$ and $\mathcal{F}_{c_2}^{\varphi}$ have non-empty intersection, then they coincide.

Lemma 6 *For a function u in $\mathcal{B}_c^{\varphi}$, denote $\widetilde{c}=c+u-\psi(u)u_0$. Then $\mathcal{F}_c^{\varphi}=\mathcal{F}_{\widetilde{c}}^{\varphi}$.*

Proof Let v be a function in $\mathcal{B}_c^{\varphi}$. Then there exists $\varepsilon>0$ such that, for every $\lambda\in(-\varepsilon,1+\varepsilon)$, the function $\boldsymbol{\varphi}(c+\lambda v+(1-\lambda)u)$ is μ-integrable. Consequently, $\varphi(\widetilde{c}+\lambda(v-u))$ is μ-integrable for all $\lambda\in(-\varepsilon,1+\varepsilon)$. Thus the difference $v-u$ is in $\mathcal{K}_{\widetilde{c}}^{\varphi}$ and

$$w=v-u-\frac{\mathbb{E}[(v-u)\boldsymbol{\varphi}'_+(\widetilde{c})]}{\mathbb{E}[u_0\boldsymbol{\varphi}'_+(\widetilde{c})]}u_0 \tag{7}$$

belongs to $\mathcal{B}_{\widetilde{c}}^{\varphi}$. Let $\widetilde{\psi}:\mathcal{B}_{\widetilde{c}}^{\varphi}\to[0,\infty)$ be the normalizing function associated to $\widetilde{c}$. Then the probability density $\boldsymbol{\varphi}(\widetilde{c}+w-\widetilde{\psi}(w)u_0)$ is in $\mathcal{F}_{\widetilde{c}}^{\varphi}$. This probability density can be expressed as $\boldsymbol{\varphi}(c+v-ku_0)$ for a constant k. According to Proposition 3, there exists a unique $\psi(u)\in\mathbb{R}$ such that the probability density $\boldsymbol{\varphi}(c+v-\psi(v)u_0)$ is in $\mathcal{F}_c^{\varphi}$. Therefore, $\mathcal{F}_c^{\varphi}\subseteq\mathcal{F}_{\widetilde{c}}^{\varphi}$.

Using the same arguments as in the previous paragraph, we obtain $c=\widetilde{c}+w-\widetilde{\psi}(w)u_0$, where the function $w\in\mathcal{B}_{\widetilde{c}}^{\varphi}$ is given in (7) with $v=0$. Thus $\mathcal{F}_{\widetilde{c}}^{\varphi}\subseteq\mathcal{F}_c^{\varphi}$. □

By Lemma 6, if we denote $c_1+u-\psi_1(u)u_0=\widetilde{c}=c_2+v-\psi_2(v)u_0$, then we have the equality $\mathcal{F}_{c_1}^{\varphi}=\mathcal{F}_{\widetilde{c}}^{\varphi}=\mathcal{F}_{c_2}^{\varphi}$.

The results obtained in these lemmas are summarized in the next proposition.

Proposition 7 *Let $c_1, c_2 : T \to \mathbb{R}$ be measurable functions such that the probability densities $p_1 = \boldsymbol{\varphi}(c_1)$ and $p_2 = \boldsymbol{\varphi}(c_2)$ are in $\mathcal{P}_\mu$. Suppose $\mathcal{F}^\varphi_{c_1} \cap \mathcal{F}^\varphi_{c_2} \neq \emptyset$. Then the Musielak–Orlicz spaces $L^\varphi_{c_1}$ and $L^\varphi_{c_2}$ are equal as sets, and have equivalent norms. Moreover, $\mathcal{F}^\varphi_{c_1} = \mathcal{F}^\varphi_{c_2}$.*

Thus we can state:

Proposition 8 T*he collection $\{(\mathcal{B}^\varphi_c, \boldsymbol{\varphi}_c)\}_{\boldsymbol{\varphi}(c)\in\mathcal{P}_\mu}$ satisfies (bm1)–(bm2), equipping $\mathcal{P}_\mu$ with a C^∞-differentiable structure.*

5 Divergence

In this section we define the divergence between two probability distributions. The entities found in Information Geometry [1, 9], like the Fisher information, connections, geodesics, etc., are all derived from the divergence taken in the considered family. The divergence we will found is the Bregman divergence [2] associated to the normalizing function $\psi : \mathcal{K}^\varphi_c \to [0, \infty)$. We show that our divergence does not depend on the parametrization of the φ-family $\mathcal{F}^\varphi_c$.

Let S be a convex subset of a Banach space X. Given a convex function $f : S \to \mathbb{R}$, the *Bregman divergence* $B_f : S \times S \to [0, \infty)$ is defined as

$$B_f(y, x) = f(y) - f(x) - \partial_+ f(x)(y - x),$$

for all $x, y \in S$, where $\partial_+ f(x)(h) = \lim_{t\downarrow 0}(f(x + th) - f(x))/t$ denotes the *right-directional derivative* of f at x in the direction of h. The right-directional derivative $\partial_+ f(x)(h)$ exists and defines a sublinear functional. If the function f is strictly convex, the divergence satisfies $B_f(y, x) = 0$ if and only if $x = y$.

Let X and Y be Banach spaces, and $U \subseteq X$ be an open set. A function $f : U \to Y$ is said to be *Gâteaux-differentiable* at $x_0 \in U$ if there exists a bounded linear map $A : X \to Y$ such that

$$\lim_{t\to 0} \frac{1}{t} \big\| f(x_0 + th) - f(x_0) - Ah \big\| = 0,$$

for every $h \in X$. The *Gâteaux derivative* of f at x_0 is denoted by $A = \partial f(x_0)$. If the limit above can be taken uniformly for every $h \in X$ such that $\|h\| \le 1$, then the function f is said to be *Fréchet-differentiable* at x_0. The *Fréchet derivative* of f at x_0 is denoted by $A = Df(x_0)$.

Now we verify that $\psi : \mathcal{K}^\varphi_c \to \mathbb{R}$ is a convex function. Take any $u, v \in \mathcal{K}^\varphi_c$ such that $u \neq v$. Clearly, the function $\lambda u + (1 - \lambda)v$ is in $\mathcal{K}^\varphi_c$, for any $\lambda \in (0, 1)$. By the convexity of $\varphi(t, \cdot)$, we can write

$$\begin{aligned}
&\mathbb{E}\big[\boldsymbol{\varphi}\big(c + \lambda u + (1 - \lambda)v - \lambda\psi(u)u_0 - (1 - \lambda)\psi(v)u_0\big)\big] \\
&\quad \le \lambda\mathbb{E}\big[\boldsymbol{\varphi}\big(c + u - \psi(u)u_0\big)\big] + (1 - \lambda)\mathbb{E}\big[\boldsymbol{\varphi}\big(c + v - \psi(v)u_0\big)\big] = 1.
\end{aligned}$$

Since $\boldsymbol{\varphi}(c + \lambda u + (1 - \lambda)v - \psi(\lambda u + (1 - \lambda)v)u_0)$ has μ-integral equal to 1, we can conclude that the following inequality holds:

$$\psi\big(\lambda u + (1 - \lambda)v\big) \le \lambda\psi(u) + (1 - \lambda)\psi(v).$$

So we can define the Bregman divergence B_ψ from to the normalizing function ψ.

The Bregman divergence $B_\psi: \mathcal{B}_c^\varphi \times \mathcal{B}_c^\varphi \to [0,\infty)$ associated to the normalizing function $\psi: \mathcal{B}_c^\varphi \to [0,\infty)$ is given by

$$B_\psi(v,u) = \psi(v) - \psi(u) - \partial_+\psi(u)(v-u).$$

Then we define the divergence $D_\psi: \mathcal{B}_c^\varphi \times \mathcal{B}_c^\varphi \to [0,\infty)$ related to the φ-family $\mathcal{F}_c^\varphi$ as

$$D_\psi(u,v) = B_\psi(v,u).$$

The entries of B_ψ are inverted in order that D_ψ corresponds in some way to the *Kullback–Leibler divergence* $D_{\mathrm{KL}}(p,q) = \mathbb{E}[p\log(\frac{p}{q})]$. Assuming that $\varphi(t,\cdot)$ is continuously differentiable, we will find an expression for $\partial\psi(u)$.

Lemma 9 *Assume that $\varphi(t,\cdot)$ is continuously differentiable. For any $u \in \mathcal{K}_c^\varphi$, the linear functional $f_u: L_c^\varphi \to \mathbb{R}$ given by $f_u(v) = \mathbb{E}[v\boldsymbol{\varphi}'(c+u)]$ is bounded.*

Proof Every function $v \in L_c^\varphi$ with norm $\|v\|_{\Phi,0} \le 1$ satisfies $I_\Phi(v) \le \|v\|_{\Phi,0}$. Then we obtain

$$\mathbb{E}\big[\boldsymbol{\varphi}\big(c+|v|\big)\big] = I_\Phi(v) + \mathbb{E}\big[\boldsymbol{\varphi}(c)\big] \le 2.$$

Since $u \in \mathcal{K}_c^\varphi$, we can find $\lambda \in (0,1)$ such that $\mathbb{E}[\boldsymbol{\varphi}(c+\frac{1}{\lambda}u)] < \infty$. We can write

$$\begin{aligned}(1-\lambda)\mathbb{E}\big[|v|\boldsymbol{\varphi}'(c+u)\big] &\le \mathbb{E}\big[\boldsymbol{\varphi}\big(c+u+(1-\lambda)|v|\big)\big] - \mathbb{E}\big[\boldsymbol{\varphi}(c+u)\big] \\ &= \mathbb{E}\left[\boldsymbol{\varphi}\left(\lambda\left(c+\frac{1}{\lambda}u\right) + (1-\lambda)\big(c+|v|\big)\right)\right] - \mathbb{E}\big[\boldsymbol{\varphi}(c+u)\big] \\ &\le \lambda\mathbb{E}\left[\boldsymbol{\varphi}\left(c+\frac{1}{\lambda}u\right)\right] + (1-\lambda)\mathbb{E}\big[\boldsymbol{\varphi}\big(c+|v|\big)\big] \\ &\quad - \mathbb{E}\big[\boldsymbol{\varphi}(c+u)\big].\end{aligned}$$

Thus the absolute value of $f_u(v) = \mathbb{E}[v\boldsymbol{\varphi}'(c+u)]$ is bounded by some constant for $\|v\|_{\Phi,0} \le 1$. □

Lemma 10 *Assume that $\varphi(t,\cdot)$ is continuously differentiable. Then the normalizing function $\psi: \mathcal{K}_c^\varphi \to \mathbb{R}$ is Gâteaux-differentiable and*

$$\partial\psi(u)v = \frac{\mathbb{E}[v\boldsymbol{\varphi}'(c+u-\psi(u)u_0)]}{\mathbb{E}[u_0\boldsymbol{\varphi}'(c+u-\psi(u)u_0)]}. \tag{8}$$

Proof According to Lemma 9, the expression in (8) defines a bounded linear functional. Fix functions $u \in \mathcal{K}_c^\varphi$ and $v \in L_c^\varphi$. In virtue of Proposition 4, we can find $\varepsilon > 0$ such that $\mathbb{E}[\boldsymbol{\varphi}(c+u+\lambda|v|)] < \infty$, for every $\lambda \in [-\varepsilon,\varepsilon]$. Define

$$g(\lambda,k) = \mathbb{E}\big[\boldsymbol{\varphi}(c+u+\lambda v - ku_0)\big],$$

for any $\lambda \in (-\varepsilon,\varepsilon)$ and $k \ge 0$. Since $\mathcal{K}_c^\varphi$ is open, there exist a sufficiently small $\alpha_0 > 0$ such that $u + \lambda v + \alpha|v|$ is in $\mathcal{K}_c^\varphi$ for all $\alpha \in [-\alpha_0,\alpha_0]$. We can write

$$\frac{g(\lambda+\alpha,k)-g(\lambda,k)}{\alpha}=\mathbb{E}\left[\frac{1}{\alpha}\left\{\boldsymbol{\varphi}\left(c+u+(\lambda+\alpha)v-ku_0\right)\right.\right.$$
$$\left.\left.-\boldsymbol{\varphi}(c+u+\lambda v-ku_0)\right\}\right].$$

The function in the expectation above is dominated by the μ-integrable function $\frac{1}{\alpha_0}\{\boldsymbol{\varphi}(c+u+\lambda v+\alpha_0|v|-ku_0)-\boldsymbol{\varphi}(c+u+\lambda v-ku_0)\}$. By the Dominated Convergence Theorem,

$$\mathbb{E}\left[\frac{1}{\alpha}\left\{\boldsymbol{\varphi}\left(c+u+(\lambda+\alpha)v-ku_0\right)-\boldsymbol{\varphi}(c+u+\lambda v-ku_0)\right\}\right]$$
$$\to\mathbb{E}\left[v\boldsymbol{\varphi}'(c+u+\lambda v-ku_0)\right],\quad \text{as } \alpha\to 0,$$

and, consequently,

$$\frac{\partial g}{\partial\lambda}(\lambda,k)=\mathbb{E}\left[v\boldsymbol{\varphi}'(c+u+\lambda v-ku_0)\right].$$

Since $v\boldsymbol{\varphi}'(c+u+\lambda v-ku_0)$ is dominated by the μ-integrable function $|v|\boldsymbol{\varphi}'(c+u+\varepsilon|v|-ku_0)$, we obtain for any sequence $\lambda_n\to\lambda$,

$$\mathbb{E}\left[v\boldsymbol{\varphi}'(c+u+\lambda_n v-ku_0)\right]\to\mathbb{E}\left[v\boldsymbol{\varphi}'(c+u+\lambda v-ku_0)\right],\quad \text{as } n\to\infty.$$

Thus $\frac{\partial g}{\partial\lambda}(\lambda,k)$ is continuous with respect to λ. Analogously, it can be shown that

$$\frac{\partial g}{\partial k}(\lambda,k)=-\mathbb{E}\left[u_0\boldsymbol{\varphi}'(c+u+\lambda v-ku_0)\right],$$

and $\frac{\partial g}{\partial k}(\lambda,k)$ is continuous with respect to k. The equality $g(\lambda,k(\lambda))=\mathbb{E}[\boldsymbol{\varphi}(c+u+\lambda v-k(\lambda)u_0)]=1$ defines $k(\lambda)=\psi(u+\lambda v)$ as an implicit function of λ. Notice that $\frac{\partial g(0,k)}{\partial k}<0$. By the Implicit Function Theorem, the function $k(\lambda)=\psi(u+\lambda v)$ is continuously differentiable in a neighborhood of 0, and has derivative

$$\frac{\partial k}{\partial\lambda}(0)=-\frac{(\partial g/\partial\lambda)(0,k(0))}{(\partial g/\partial k)(0,k(0))}.$$

Consequently,

$$\partial\psi(u)(v)=\frac{\partial\psi(u+\lambda v)}{\partial\lambda}(0)=\frac{\mathbb{E}[v\boldsymbol{\varphi}'(c+u-\psi(u)u_0)]}{\mathbb{E}[u_0\boldsymbol{\varphi}'(c+u-\psi(u)u_0)]}.$$

Thus the expression in (8) is the Gâteaux-derivative of ψ. □

Lemma 11 *Assume that $\varphi(t,\cdot)$ is continuously differentiable. Then the divergence D_ψ does not depend on the parametrization of $\mathcal{F}_c^\varphi$.*

Proof For any $w\in\mathcal{B}_c^\varphi$, we denote $\widetilde{c}=c+w-\psi(w)u_0$. Given $u,v\in\mathcal{B}_c^\varphi$, select $\widetilde{u},\widetilde{v}\in\mathcal{B}_{\widetilde{c}}^\varphi$ such that $\boldsymbol{\varphi}_{\widetilde{c}}(\widetilde{u})=\boldsymbol{\varphi}_c(u)$ and $\boldsymbol{\varphi}_{\widetilde{c}}(\widetilde{v})=\boldsymbol{\varphi}_c(v)$. Let $\widetilde{\psi}\colon\mathcal{B}_{\widetilde{c}}^\varphi\to[0,\infty)$ be the normalizing function associated to $\widetilde{c}$. These definitions provide

$$\widetilde{c}+\widetilde{u}-\widetilde{\psi}(\widetilde{u})u_0=c+u-\psi(u)u_0,$$

and

$$\widetilde{c} + \widetilde{v} - \widetilde{\psi}(\widetilde{v})u_0 = c + v - \psi(v)u_0.$$

Subtracting these equations, we obtain

$$\bigl[-\widetilde{\psi}(\widetilde{v}) + \widetilde{\psi}(\widetilde{u})\bigr]u_0 + (\widetilde{v} - \widetilde{u}) = \bigl[-\psi(v) + \psi(u)\bigr]u_0 + (v - u)$$

and, consequently,

$$\widetilde{\psi}(\widetilde{v}) - \widetilde{\psi}(\widetilde{u}) - \frac{\mathbb{E}[(\widetilde{v} - \widetilde{u})\boldsymbol{\varphi}'(\widetilde{c} + \widetilde{u} - \widetilde{\psi}(\widetilde{u})u_0)]}{\mathbb{E}[u_0\boldsymbol{\varphi}'(\widetilde{c} + \widetilde{u} - \widetilde{\psi}(\widetilde{u})u_0)]}$$

$$= \psi(v) - \psi(u) - \frac{\mathbb{E}[(v - u)\boldsymbol{\varphi}'(c + u - \psi(u)u_0)]}{\mathbb{E}[u_0\boldsymbol{\varphi}'(c + u - \psi(u)u_0)]}.$$

Therefore, $D_{\widetilde{\psi}}(\widetilde{u}, \widetilde{v}) = D_\psi(u, v)$. □

Let $p = \boldsymbol{\varphi}_c(u)$ and $q = \boldsymbol{\varphi}_c(v)$, for $u, v \in \mathcal{B}_c^\varphi$. We denote the divergence between the probability densities p and q by

$$D(p \parallel q) = D_\psi(u, v).$$

According to Lemma 11, $D(p \parallel q)$ is well-defined if p and q are in the same φ-family. We will find an expression for $D(p \parallel q)$ where p and q are given explicitly. For $u = 0$, we have $D(p \parallel q) = D_\psi(0, v) = \psi(v)$, and then

$$D(p \parallel q) = \frac{\mathbb{E}[(-v + \psi(v)u_0)\boldsymbol{\varphi}'(c)]}{\mathbb{E}[u_0\boldsymbol{\varphi}'(c)]}.$$

Therefore, the divergence between probability densities p and q in the same φ-family can be expressed as

$$D(p \parallel q) = \frac{\mathbb{E}[\frac{\boldsymbol{\varphi}^{-1}(p) - \boldsymbol{\varphi}^{-1}(q)}{(\boldsymbol{\varphi}^{-1})'(p)}]}{\mathbb{E}[\frac{u_0}{(\boldsymbol{\varphi}^{-1})'(p)}]}. \tag{9}$$

Clearly, the expectation in (9) may not be defined if p and q are not in the same φ-family. We extend the divergence in (9) by setting $D(p \parallel q) = \infty$ if p and q are not in the same φ-family. With this extension, the divergence is denoted by D_φ and is called the *φ-divergence*. By the strict convexity of $\varphi(t, \cdot)$, we have the inequality $\varphi^{-1}(t, u) - \varphi^{-1}(t, v) \geq (\varphi^{-1})'(t, u)(u - v)$ for any $u, v > 0$, with equality if and only if $u = v$. Hence D_φ is always non-negative, and $D_\varphi(p \parallel q)$ is equal to zero if and only if $p = q$.

Example 12 With the variable κ-exponential $\exp_\kappa(t, u) = \exp_{\kappa(t)}(u)$ in the place of $\varphi(t, u)$, whose inverse $\varphi^{-1}(t, u)$ is the variable κ-logarithm $\ln_\kappa(t, u) = \ln_{\kappa(t)}(u)$, we rewrite (9) as

$$D(p \parallel q) = \frac{\mathbb{E}[\frac{\mathbf{ln}_\kappa(p) - \mathbf{ln}_\kappa(q)}{\mathbf{ln}'_\kappa(p)}]}{\mathbb{E}[\frac{u_0}{\mathbf{ln}'_\kappa(p)}]}, \tag{10}$$

where $\mathbf{ln}_\kappa(p)$ denotes $\ln_{\kappa(t)}(p(t))$. Since the κ-logarithm $\ln_\kappa(u) = \frac{u^\kappa - u^{-\kappa}}{2\kappa}$ has derivative $\ln'_\kappa(u) = \frac{1}{u}\frac{u^\kappa + u^{-\kappa}}{2}$, the numerator and denominator in (10) result in

$$\mathbb{E}\left[\frac{\mathbf{ln}_\kappa(p) - \mathbf{ln}_\kappa(q)}{\mathbf{ln}'_\kappa(p)}\right] = \mathbb{E}\left[\frac{\frac{p^\kappa - p^{-\kappa}}{2\kappa} - \frac{q^\kappa - q^{-\kappa}}{2\kappa}}{\frac{1}{p}\frac{p^\kappa + p^{-\kappa}}{2}}\right] = \frac{1}{\kappa}\mathbb{E}_p\left[\frac{p^\kappa - p^{-\kappa}}{p^\kappa + p^{-\kappa}} - \frac{q^\kappa - q^{-\kappa}}{p^\kappa + p^{-\kappa}}\right]$$

and

$$\mathbb{E}\left[\frac{u_0}{\mathbf{ln}'_\kappa(p)}\right] = \mathbb{E}_p\left[\frac{2u_0}{p^\kappa + p^{-\kappa}}\right],$$

respectively. Thus (10) can be rewritten as

$$D_\kappa(p \parallel q) = \frac{1}{\kappa}\frac{\mathbb{E}_p[\frac{p^\kappa - p^{-\kappa}}{p^\kappa + p^{-\kappa}} - \frac{q^\kappa - q^{-\kappa}}{p^\kappa + p^{-\kappa}}]}{\mathbb{E}_p[\frac{2u_0}{p^\kappa + p^{-\kappa}}]},$$

which we called the *κ-divergence*.

Acknowledgements We are indebted to the anonymous referees for significant comments and suggestions leading to the current version. This work received financial support from CAPES—Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

References

1. Amari, S., Nagaoka, H.: Methods of Information Geometry. Translations of Mathematical Monographs, vol. 191. Am. Math. Soc., Providence (2000). Translated from the 1993 Japanese original by Daishi Harada
2. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. U.S.S.R. Comput. Math. Math. Phys. **7**(3), 200–217 (1967)
3. Cena, A., Pistone, G.: Exponential statistical manifold. Ann. Inst. Stat. Math. **59**(1), 27–56 (2007)
4. Gibilisco, P., Riccomagno, E., Rogantin, M.P., Wynn, H.P. (eds.): Algebraic and Geometric Methods in Statistics. Cambridge University Press, Cambridge (2010)
5. Grasselli, M.R.: Dual connections in nonparametric classical information geometry. Ann. Inst. Stat. Math. **62**(5), 873–896 (2010)
6. Kaniadakis, G.: Statistical mechanics in the context of special relativity. Phys. Rev. E **66**(5), 056125 (2002)
7. Krasnosel'skiĭ, M.A., Rutickiĭ, J.B.: Convex Functions and Orlicz Spaces. Noordhoff, Groningen (1961). Translated from the first Russian edition by Leo F. Boron
8. Lang, S.: Differential and Riemannian Manifolds. Graduate Texts in Mathematics, vol. 160, 3rd edn. Springer, New York (1995)
9. Murray, M.K., Rice, J.W.: Differential Geometry and Statistics. Monographs on Statistics and Applied Probability, vol. 48. Chapman & Hall, London (1993)
10. Musielak, J.: Orlicz Spaces and Modular Spaces. Lecture Notes in Mathematics, vol. 1034. Springer, Berlin (1983)
11. Naudts, J.: Generalised Thermostatistics. Springer, London (2011)
12. Petz, D.: Quantum Information Theory and Quantum Statistics. Theoretical and Mathematical Physics. Springer, Berlin (2008)
13. Pistone, G.: κ-exponential models from the geometrical viewpoint. Eur. Phys. J. B **70**(1), 29–37 (2009)
14. Pistone, G., Rogantin, M.P.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernoulli **5**(4), 721–760 (1999)
15. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. Ann. Stat. **23**(5), 1543–1561 (1995)
16. Rao, M.M., Ren, Z.D.: Theory of Orlicz Spaces. Monographs and Textbooks in Pure and Applied Mathematics, vol. 146. Dekker, New York (1991)