On the ϕ-family of probability distributions✩

Rui F. Vigelis*, Charles C. Cavalcante

Wireless Telecommunication Research Group, Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza-CE, Brazil

Abstract

We generalize the exponential family of probability distributions $\mathcal{E}_\mu$. In our approach, the exponential function is replaced by the $\phi$-function, resulting in the $\phi$-family of probability distributions $\mathcal{F}_\phi$. We provide how $\phi$-families are constructed. In the $\phi$-family, the analogous of the cumulant-generating functional is a normalizing function. We define the $\phi$-divergence as the Bregman divergence associated to the normalizing function, providing a generalization of the Kullback–Leibler divergence. We found that the Kaniadakis’ $\kappa$-exponential function satisfies the definition of $\phi$-functions. A formula for the $\phi$-divergence where the $\phi$-function is the $\kappa$-exponential function is derived.

Keywords: Exponential family of probability distributions, Musielak-Orlicz spaces, Bregman divergence

1. Introduction

Let $(T, \Sigma, \mu)$ be a $\sigma$-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 $\mu$. The probability family $\mathcal{P}_\mu$ can be represented as (we adopt the same 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 $\mu$-a.e., and $\mathbb{E}[\cdot]$ denotes the expectation with respect to the measure $\mu$.

The family $\mathcal{P}_\mu$ can be equipped with a structure of $C^\infty$-Banach manifold, using the Orlicz space $L^{\Phi_1}(\mu) = L^{\Phi_1}(T, \Sigma, p \cdot \mu)$ 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

✩This work received financial support from CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.
*Corresponding author

Email addresses: rfvigelis@gtel.ufc.br (Rui F. Vigelis), charles@gtel.ufc.br (Charles C. Cavalcante)

Preprint submitted to Journal of Mathematical Analysis and Applications January 20, 2013
statistical manifold., whose construction was proposed in [1] and developed in [2, 3, 4]. 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

\[ e_p(u) = e^{u - K_p(u)} p, \quad \text{for } u \in B_p, \quad (1) \]

for a subset $B_p$ 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

\[ e_p(u) = e^{c + u - K_p(u) \cdot 1_T}, \quad \text{for } u \in B_p, \quad (2) \]

where $1_A$ is the indicator function of a subset $A \subseteq T$.

In the $\varphi$-family of probability distributions $\mathcal{F}_\varphi$, which we propose, the exponential function is replaced by the so called $\varphi$-function $\varphi: T \times \mathbb{R} \rightarrow [0, \infty]$. 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 $\kappa$-exponential function satisfies the definition of $\varphi$-functions. As in the exponential family, the $\varphi$-families are the connected component of $\mathcal{P}_\mu$, which is endowed with a structure of $C^\infty$-Banach manifold, using $\varphi$ 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 $\varphi$-family of probability distributions $\mathcal{F}_\varphi$ are given by

\[ \varphi_c(u)(t) = \varphi(t, c(t) + u(t) - \psi(u) u_0(t)), \quad \text{for } u \in B_c^\varphi, \quad (3) \]

for a subset $B_c^\varphi$ of a Musielak–Orlicz space $L_{\varphi}^\varphi$. The normalizing function $\psi: B_c^\varphi \rightarrow [0, \infty)$ and the measurable function $u_0: T \rightarrow [0, \infty)$ in (3) replaces $K_p$ and $1_T$ in (2), respectively. The function $u_0$ is not arbitrary. In the text, we will show how $u_0$ can be chosen.

We define the $\varphi$-divergence as the a Bregman divergence associated to the normalizing function $\psi$, providing a generalization of the Kullback–Leibler divergence. Then geometrical aspects related to the $\varphi$-family can be developed, since the Fisher information (from which the Information Geometry [5, 6] is based) is derived from the divergence. A formula for the $\varphi$-divergence where the $\varphi$-function is the $\kappa$-exponential function is derived, which we called the $\kappa$-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 [7, 8]. The rest of this paper is organized as follows. Section 2 deals with the topics of Musielak–Orlicz spaces we will use in the the construction of the $\varphi$-family of probability distributions. In Section 3 the exponential statistical manifold is reviewed. The construction of the $\varphi$-family of probability distributions is given in Section 4. Finally, the $\varphi$-divergence is derived in Section 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 $\varphi$-families. A more detailed exposition about these spaces can be found in [9, 10, 11].

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

(i) $\Phi(t, \cdot)$ is convex and lower semi-continuous,

(ii) $\Phi(t, 0) = \lim_{u \to 0} \Phi(t, u) = 0$ 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 $\infty$ on the interval $(0, \infty)$.

A Musielak–Orlicz function $\Phi$ is said to be an Orlicz function if the functions $\Phi(t, \cdot)$ are identical for $\mu$-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

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

$$\hat{L}^\Phi = \{ u \in L^0 : I_\Phi(u) < \infty \},$$

and

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

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

The Musielak–Orlicz space $L^\Phi$ 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

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

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 $f \in \hat{L}^\Phi$ and a constant $K > 0$ such that

$$\Phi(t, 2u) \leq K \Phi(t, u), \text{ for all } u \geq f(t),$$

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

\[
B_0(E^\Phi, 1) = \{ u \in L^\Phi : \inf_{v \in E^\Phi} \| u - v \|_{\Phi, 0} < 1 \},
\]
or, equivalently, \( B_0(E^\Phi, 1) \subset \bar{L}^\Phi \subset \overline{\mathcal{B}_0}(E^\Phi, 1) \).

3. The exponential statistical manifold

This section starts with the definition of a \( C^k \)-Banach manifold. A \( C^k \)-Banach manifold is a set \( M \) and a collection of pairs \((U_\alpha, x_\alpha)\) (\( \alpha \) belonging to some indexing set), composed by open subsets \( U_\alpha \) of some Banach space \( X_\alpha \), and injective mappings \( x_\alpha : U_\alpha \rightarrow M \), satisfying the following conditions:

(bm1) the sets \( x_\alpha(U_\alpha) \) cover \( M \), i.e., \( \bigcup_\alpha x_\alpha(U_\alpha) = M \);

(bm2) for any pair of indices \( \alpha, \beta \) such that \( x_\alpha(U_\alpha) \cap x_\beta(U_\beta) = W \neq \emptyset \), the sets \( x_\alpha^{-1}(W) \) and \( x_\beta^{-1}(W) \) are open in \( X_\alpha \) and \( X_\beta \), respectively; and

(bm3) the transition map \( x_\beta^{-1} \circ x_\alpha : x_\alpha^{-1}(W) \rightarrow x_\beta^{-1}(W) \) is a \( C^k \)-isomorphism.

The pair \((U_\alpha, x_\alpha)\) with \( p \in x_\alpha(U_\alpha) \) is called a parametrization (or system of coordinates) of \( M \) at \( p \); and \( x_\alpha(U_\alpha) \) 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 \( x_\alpha(U_\alpha) \) is open, and the \( x_\alpha \)'s are topological isomorphisms. We note that if \( k \geq 1 \) and two parametrizations \((U_\alpha, x_\alpha)\) and \((U_\beta, x_\beta)\) are such that \( x_\alpha(U_\alpha) \) and \( x_\beta(U_\beta) \) have a non-empty intersection, then from the derivative of \( x_\beta^{-1} \circ x_\alpha \) we have that \( X_\alpha \) and \( X_\beta \) are isomorphic.

Two collections \( \{(U_\alpha, x_\alpha)\} \) and \( \{(U_\beta, x_\beta)\} \) 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 \( \{(U_\alpha, x_\alpha)\} \) 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 \( \Phi_1 : [0, \infty) \rightarrow [0, \infty) \) 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 \( \hat{u}_p(t) = \mathbb{E}[e^{tu}] \) is finite in a neighborhood of 0.

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

\[
M_p(u) = \mathbb{E}[e^u],
\]
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 \( K_p \) 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 \( K_p \) if and only if \( M_p(\lambda u) \) is finite for every \( \lambda \) in some neighborhood of \([0, 1]\). The closed subspace of \( p \)-centered random variables

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

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

\[ e_p(u) = e^{u - K_p(u)} p, \quad \text{for all } u \in B_p. \]

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

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

Since \( K_p(u) < \infty \) for every \( u \in K_p \), we have that \( e_p \) can be extended to \( K_p \). The restriction of \( e_p \) to \( B_p \) guarantees that \( e_p \) 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 [2, Proposition 5]). Hence, \( B_p = e_q^{-1}(\mathcal{E}_p \cap \mathcal{E}_q) \) and \( B_q = e_q^{-1}(\mathcal{E}_p \cap \mathcal{E}_q) \). The transition map \( e_q^{-1} \circ e_p : B_p \to B_q \), which can be written as

\[ e_q^{-1} \circ e_p(u) = u + \log \left( \frac{q}{p} \right) - \mathbb{E}_q \left[ u + \log \left( \frac{q}{p} \right) \right], \quad \text{for all } u \in B_p, \]

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

4. Construction of the \( \varphi \)-family of probability distributions

The generalization of the exponential family is based on the replacement of the exponential function by a \( \varphi \)-function \( \varphi : T \times \mathbb{R} \to [0, \infty] \) that satisfies the following properties, for \( \mu \)-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 that

(a4) $\varphi(t, c(t) + \lambda u_0(t))$ is $\mu$-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 $u : T \to \mathbb{R}$ 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** [12, 13, 14]. The Kaniadakis’ $\kappa$-exponential $\exp_\kappa : \mathbb{R} \to (0, \infty)$ for $\kappa \in [-1, 1]$ is defined as

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

The inverse of $\exp_\kappa$ is the Kaniadakis’ $\kappa$-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:

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

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

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

whose inverse is called the variable $\kappa$-logarithm:

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

Assuming that $\kappa_- = \operatorname{ess inf}|\kappa(t)| > 0$, the variable $\kappa$-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 $\alpha \geq 1$

$$
\exp_\kappa(t, \alpha u) = (\kappa(t)\alpha u + \alpha \sqrt{1/\alpha^2 + \kappa(t)^2 u^2})^{1/\kappa}
\leq \alpha^{1/\kappa} \kappa(t) u + \sqrt{1 + \kappa(t)^2 u^2})^{1/\kappa}
\leq \alpha^{1/\kappa} \exp_\kappa(t, u).
$$

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

$$
\exp_\kappa(t, c + u) \leq \lambda \exp_\kappa(t, \lambda^{-1}c) + (1 - \lambda) \exp_\kappa(t, (1 - \lambda)^{-1}u)
\leq \lambda^{1-1/\kappa} \exp_\kappa(t, c) + (1 - \lambda)^{1-1/\kappa} \exp_\kappa(t, u).
$$

Thus any positive function $u_0$ such that $\mathbb{E}[\exp_\kappa(u_0)] < \infty$ satisfies (a4).
Let \( c: T \to \mathbb{R} \) be a measurable function such that \( \varphi(t, c(t)) \) is \( \mu \)-integrable. We define the Musielak–Orlicz function
\[
\Phi(t, u) = \varphi(t, c(t) + u) - \varphi(t, c(t)).
\]
and denote \( L^\Phi, \hat{L}^\Phi \) and \( E^\Phi \) by \( L^\Phi_x, \hat{L}^\Phi_x \) and \( E^\Phi_x \), respectively. Since \( \varphi(t, c(t)) \) is \( \mu \)-integrable, the Musielak–Orlicz space \( L^\Phi_x \) corresponds to the set of all functions \( u \in L^0 \) for which \( \varphi(t, c(t) + \lambda u(t)) \) is \( \mu \)-integrable for every \( \lambda \) contained in some neighborhood of \( 0 \). By the convexity of \( \varphi(t, \cdot) \), we have
\[
u \varphi'(t, c(t)) \leq \varphi(t, c(t) + u) - \varphi(t, c(t)), \quad \text{for all } u \in \mathbb{R}. \tag{4}
\]
Hence every function \( u \) in \( L^\Phi_x \) belongs to the weighted Lebesgue space \( L^\Phi_{1, \mu} \) where \( w(t) = \varphi'(t, c(t)) \).

Let \( K^\varphi_x \) be the set of all functions \( u \in L^\Phi_x \) such that \( \varphi(t, c(t) + \lambda u(t)) \) is \( \mu \)-integrable for every \( \lambda \) in a neighborhood of \( [0, 1] \). Denote by \( \varphi \) the operator acting on the set of real-valued functions \( u: T \to \mathbb{R} \) given by \( \varphi(u)(t) = \varphi(t, u(t)) \). For each probability density \( p \in \mathcal{P}_1 \), we can take a measurable function \( c: T \to \mathbb{R} \) such that \( p = \varphi(c) \). The first import result in the construction of the \( \varphi \)-family is given below.

**Lemma 2.** The set \( K^\varphi_x \) is open in \( L^\Phi_x \).

**Proof.** Take any \( u \in K^\varphi_x \). We can find \( \varepsilon \in (0, 1) \) such that \( \mathbb{E}[\varphi(c + \alpha u)] < \infty \) for every \( \alpha \in [-\varepsilon, 1 + \varepsilon] \). Let \( \delta = \left( \varepsilon(1 + \varepsilon)(1 + \frac{\alpha}{2}) \right)^{-1} \). For any function \( v \in L^\Phi_x \) in the open ball \( B_\delta = \{ w \in L^\Phi_x : \| w \|_\Phi < \delta \} \), we have \( I_\Phi(\frac{\alpha}{2}) \leq 1 \). Thus \( \mathbb{E}[\varphi(c + \frac{\alpha}{2} |v|)] \leq 2 \). Taking any \( \alpha \in (0, 1 + \frac{\varepsilon}{2}) \), we denote \( \lambda = \frac{\alpha}{1 + \varepsilon} \). In virtue of
\[
\frac{\alpha}{1 - \lambda} = \frac{\alpha}{1 - \frac{\alpha}{1 + \varepsilon}} \leq \frac{1 + \varepsilon}{1 - \frac{\alpha}{1 + \varepsilon}} = \frac{2}{\varepsilon}(1 + \varepsilon)(1 + \frac{\alpha}{2}) = \frac{1}{\delta},
\]
it follows that
\[
\varphi(c + \alpha(u + v)) = \varphi(\lambda(c + \frac{\alpha}{\lambda} u) + (1 - \lambda)(c + \frac{\alpha}{1 - \lambda} v))
\leq \lambda \varphi(c + \frac{\alpha}{\lambda} u) + (1 - \lambda) \varphi(c + \frac{\alpha}{1 - \lambda} v)
\leq \lambda \varphi(c + (1 + \varepsilon) u) + (1 - \lambda) \varphi(c + \frac{1}{\varepsilon} |v|). \tag{5}
\]
For \( \alpha \in (-\frac{\varepsilon}{2}, 0) \), we can write
\[
\varphi(c + \alpha(u + v)) \leq \frac{1}{2} \varphi(c + 2 \alpha u) + \frac{1}{2} \varphi(c + 2 \alpha v)
\leq \frac{1}{2} \varphi(c + 2 \alpha u) + \frac{1}{2} \varphi(c + |v|). \tag{6}
\]
By (5) and (6), we get \( \mathbb{E}[\varphi(c + \alpha(u + v))] < \infty \), for any \( \alpha \in (-\frac{\varepsilon}{2}, 1 + \frac{\varepsilon}{2}) \). Hence the ball of radius \( \delta \) centered at \( u \) is contained in \( K^\varphi_x \). Therefore, the set \( K^\varphi_x \) is open. \( \square \)
Clearly, for \( u \in \mathcal{K}_c^\phi \) the function \( \varphi(c + u) \) is not necessarily in \( \mathcal{P}_\mu \). The normalizing function \( \psi: \mathcal{K}_c^\phi \to \mathbb{R} \) is introduced in order to make the density

\[
\varphi(c + u - \psi(u)u_0)
\]

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

\[
\varphi(c + \lambda u) = \varphi(\lambda(c + u - \alpha u_0) + (1 - \lambda)(c + \frac{1}{1-\lambda} \alpha u_0)) \\
\leq \lambda \varphi(c + u - \alpha u_0) + (1 - \lambda) \varphi(c + \frac{1}{1-\lambda} \alpha u_0).
\]

Since the function \( u_0 \) satisfies (a4), we obtain that \( \varphi(c + \lambda u) \) is \( \mu \)-integrable. Hence the maximal, open domain of \( \psi \) is contained in \( \mathcal{K}_c^\phi \).

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

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

\[
\varphi(c + u + \alpha u_0) = \varphi(\lambda(c + \frac{1}{\lambda} u) + (1 - \lambda)(c + \frac{1}{1-\lambda} \alpha u_0)) \\
\leq \lambda \varphi(c + \frac{1}{\lambda} u) + (1 - \lambda) \varphi(c + \frac{1}{1-\lambda} \alpha u_0).
\]

Thus \( \varphi(c + u + \alpha u_0) \) is \( \mu \)-integrable. By the Dominated Convergence Theorem, the map \( \alpha \mapsto J(\alpha) = \mathbb{E}[\varphi(c + u + \alpha u_0)] \) 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 \( \varphi(c + u - \psi(u)u_0) \) is a probability density in \( \mathcal{P}_\mu \).

The function \( \psi: \mathcal{K}_c^\phi \to \mathbb{R} \) can take both positive and negative values. However, if the domain of \( \psi \) is restricted to a subspace of \( L_c^\phi \), its image will be contained in \( [0, \infty) \). Denote the closed subspace

\[
B_c^\phi = \{ u \in L_c^\phi : \mathbb{E}[\varphi'(c)] = 0 \},
\]

and let \( B_c^\phi \) be \( B_c^\phi \cap \mathcal{K}_c^\phi \). Supposing that \( u \in B_c^\phi \), it follows that \( \mathbb{E}[\varphi'(c)] = 0 \) and \( \mathbb{E}[\varphi(c + u)] < \infty \); and, according to inequality (4), we have

\[
1 = \mathbb{E}[\varphi'(c)] + \mathbb{E}[\varphi(c)] \leq \mathbb{E}[\varphi(c + u)] < \infty.
\]

If \( u \in \mathcal{K}_c^\phi \) belongs to the subspace \( B_c^\phi \), the integral of \( \varphi(c + u) \) is greater than or equal to 1. Subtracting \( \psi(u)u_0 \), the integral decreases to 1, and we obtain that \( \varphi(c + u - \psi(u)u_0) \) is in \( \mathcal{P}_\mu \).

For each measurable function \( c: T \to \mathbb{R} \) such that the probability density \( p = \varphi(c) \) belongs to \( \mathcal{P}_\mu \), we associate a parametrization \( \varphi_c: B_c^\phi \to \mathcal{F}_c^\phi \) that
maps each function $u$ in $\mathcal{B}_c^\mu$ to a probability density in $\mathcal{F}_c^\mu = \varphi_c(\mathcal{B}_c^\mu) \subseteq \mathcal{P}_\mu$ according to
\[
\varphi_c(u) = \varphi(c + u - \psi(u)u_0).
\]
Clearly, we have $\mathcal{P}_\mu = \bigcup \{ \mathcal{F}_c^\mu : \varphi_c \in \mathcal{P}_\mu \}$. Moreover, the map $\varphi_c$ is a bijection from $\mathcal{B}_c^\mu$ to $\mathcal{F}_c^\mu$. If the functions $u, v \in \mathcal{B}_c^\mu$ are such that $\varphi_c(u) = \varphi_c(v)$, then the difference $u - v = (\psi(u) - \psi(v))u_0$ is in $\mathcal{B}_c^\mu$. 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 $p_1 = \varphi(c_1)$ and $p_2 = \varphi(c_2)$ belong to $\mathcal{P}_\mu$. The parametrizations $\varphi_{c_1} : \mathcal{B}_{c_1}^\mu \to \mathcal{F}_{c_1}^\mu$ and $\varphi_{c_2} : \mathcal{B}_{c_2}^\mu \to \mathcal{F}_{c_2}^\mu$ related to these functions have transition map
\[
\varphi_{c_2}^{-1} \circ \varphi_{c_1} : \varphi_{c_1}^{-1}(\mathcal{F}_{c_1}^\mu \cap \mathcal{F}_{c_2}^\mu) \to \varphi_{c_2}^{-1}(\mathcal{F}_{c_1}^\mu \cap \mathcal{F}_{c_2}^\mu).
\]
Let $\psi_1 : \mathcal{B}_{c_1}^\mu \to \mathbb{R}$ and $\psi_2 : \mathcal{B}_{c_2}^\mu \to \mathbb{R}$ be the normalizing functions associated to $c_1$ and $c_2$, respectively. Assume that the functions $u \in \mathcal{B}_{c_1}^\mu$ and $v \in \mathcal{B}_{c_2}^\mu$ are such that $\varphi_{c_1}(u) = \varphi_{c_2}(v) \in \mathcal{F}_{c_1}^\mu \cap \mathcal{F}_{c_2}^\mu$. Then we can write
\[
v = c_1 - c_2 + u - (\psi_1(u) - \psi_2(v))u_0.
\]
Since the function $v$ is in $\mathcal{B}_{c_2}^\mu$, if we multiply this equation by $\varphi'(c_2)$ and integrate with respect to the measure $\mu$, we obtain
\[
0 = \mathbb{E}[(c_1 - c_2 + u)\varphi'(c_2)] - (\psi_1(u) - \psi_2(v)) \mathbb{E}[u_0\varphi'(c_2)].
\]
Thus the transition map $\varphi_{c_2}^{-1} \circ \varphi_{c_1}$ can be expressed as
\[
\varphi_{c_2}^{-1} \circ \varphi_{c_1}(w) = c_1 - c_2 + w - \frac{\mathbb{E}[(c_1 - c_2 + w)\varphi'(c_2)]}{\mathbb{E}[u_0\varphi'(c_2)]}u_0, \quad (7)
\]
for every $w \in \varphi_{c_1}^{-1}(\mathcal{F}_{c_1}^\mu \cap \mathcal{F}_{c_2}^\mu)$. Clearly, this transition map will be of class $C^\infty$ if we show that the functions $w$ and $c_1 - c_2$ are in $L_c^\mu$, and the spaces $L_c^\mu$ and $L_c^\nu$ 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 [9, Theorem 8.5]). We make use of the following:

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

**Proof.** Suppose that $\bar{c} - c$ is not in $L_c^\mu$. Let $A = \{ t \in T : \bar{c}(t) < c(t) \}$. For $\lambda \in [0, 1]$, we have
\[
\mathbb{E}[\varphi(c + \lambda(\bar{c} - c))] = \mathbb{E}[\varphi(c + \lambda(\bar{c} - c))1_{T \setminus A}] + \mathbb{E}[\varphi(c + \lambda(\bar{c} - c))1_A] \\
\leq \mathbb{E}[\varphi(c + (\bar{c} - c))1_{T \setminus A}] + \mathbb{E}[\varphi(c)1_A] \\
\leq \mathbb{E}[\varphi(\bar{c})] + \mathbb{E}[\varphi(c)] < \infty.
\]
Since $\bar{c} - c \notin L_c^\mu$, for any $\lambda > 0$, there holds $\mathbb{E}[\varphi(c - \lambda(\bar{c} - c))] = \infty$. From
\[
\mathbb{E}[\varphi(c - \lambda(\bar{c} - c))] = \mathbb{E}[\varphi(c - \lambda(\bar{c} - c))1_{T \setminus A}] + \mathbb{E}[\varphi(c - \lambda(\bar{c} - c))1_A] \\
\leq \mathbb{E}[\varphi(c + \lambda(c - \bar{c}))1_A],
\]

we obtain that \((c - \tilde{c})1_A\) does not belong to \(L^\infty_c\). Clearly, \((c - \tilde{c})1_A \in L^\infty_c\). Consequently, \(L^\infty_c\) is not contained in \(L^\infty_c\).

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

\[
g(\alpha, \lambda) = E[\varphi(c + \alpha(\tilde{c} - c) + \lambda w)].
\]

This function is finite for \(\lambda = 0\) and \(\alpha\) 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 have that \(g\) is finite in the convex hull of the set \(1 \times (-\varepsilon, \varepsilon) \cup (-\eta, 1] \times 0\). We obtain that \(g(0, \lambda)\) is finite for every \(\lambda\) in some neighborhood of \(0\). Consequently, \(w \in L^\infty_c\). Since \(w \in L^\infty_c\) is arbitrary, the inclusion \(L^\infty_c \subseteq L^\infty_c\) follows. \(\blacksquare\)

**Lemma 5.** If the function \(u\) is in \(K^\infty_c\) and we denote \(\tilde{c} = c + u - \varphi(u)u_0\), then the spaces \(L^\infty_c\) and \(L^\infty_c\) are equal as sets.

**Proof.** The inclusion \(L^\infty_c \subseteq L^\infty_c\) follows from Proposition 5. Since \(u \in K^\infty_c\), we have

\[
E[\varphi(\tilde{c} + \lambda u)] \leq E[\varphi(c + (1 + \lambda)u)] < \infty,
\]

for every \(\lambda\) in a neighborhood of \(0\). Thus \(c - \tilde{c} = -u + \varphi(u)u_0\) belongs to \(L^\infty_c\). From Proposition 5, we obtain \(L^\infty_c \subseteq L^\infty_c\). \(\blacksquare\)

By Lemma 6, if we denote \(c_1 + u + \varphi(v)u_0 = \tilde{c} = c_2 + v - \varphi_2(v)u_0\), we have that the spaces \(L^\infty_{c_1}, L^\infty_{c_2}\) and \(L^\infty_{c_1}\) are equal as sets. In \((7)\), the function \(w\) is in \(L^\infty_c\) and consequently \(c_1 - c_2\) is in \(L^\infty_c\). Therefore, the transition map \(\varphi^{-1}_{c_2} \circ \varphi_{c_1}\) is of class \(C^\infty\).

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

**Lemma 6.** For a function \(v\) in \(B^c_{c_2}\), denote \(\tilde{c} = c + u - \varphi(u)u_0\). Then \(\mathcal{F}^c_{c_2} = \mathcal{F}^c_{\tilde{c}}\).

**Proof.** Let \(v\) be a function in \(B^c_{c_2}\). Then there exists \(\varepsilon > 0\) such that, for every \(\lambda \in (-\varepsilon, 1 + \varepsilon)\), the function \(\varphi(c + \lambda v + (1 - \lambda)u)\) is \(\mu\)-integrable. Consequently, \(\varphi(c + \lambda v - u)\) is \(\mu\)-integrable for all \(\lambda \in (-\varepsilon, 1 + \varepsilon)\). Thus the difference \(v - u\) is in \(K^\infty_c\) and

\[
w = v - u - \frac{E[(v - u)\varphi'(c)]}{E[u_0\varphi'(c)]} u_0
\]

belongs to \(B^c_{c_2}\). Let \(\tilde{\varphi} : B^c_{c_2} \rightarrow [0, \infty)\) be the normalizing function associated to \(\tilde{c}\). Then the probability density \(\varphi(c + w - \tilde{\varphi}(w)u_0)\) is in \(\mathcal{F}^c_{\tilde{c}}\). This probability density can be expressed as \(\varphi(c + v - ku_0)\) for a constant \(k\). According to Proposition 8, there exists a unique \(\varphi(u) \in \mathbb{R}\) such that the probability density \(\varphi(c + v - \varphi(v)u_0)\) is in \(\mathcal{F}^c_{\tilde{c}}\). Therefore, \(\mathcal{F}^c_{c_2} \subseteq \mathcal{F}^c_{\tilde{c}}\).

Using the same arguments as in the previous paragraph, we obtain that \(c = \tilde{c} + w - \tilde{\varphi}(w)u_0\), where the function \(w \in B^c_{c_2}\) is given in \((8)\) with \(v = 0\). Thus \(\mathcal{F}^c_{c_2} \subseteq \mathcal{F}^c_{\tilde{c}}\). \(\blacksquare\)
By Lemma 6, if we denote $c_1 + u - \psi_1(u)u_0 = \bar{c} = c_2 + v - \psi_2(v)u_0$, then we have the equality $F_{c_1}^\varphi = F_{c_2}^\varphi = F_{c_2}^\varphi$.

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 $p_1 = \varphi(c_1)$ and $p_2 = \varphi(c_2)$ are in $\mathcal{P}_\mu$. Suppose $F_{c_1}^\varphi \cap F_{c_2}^\varphi \neq \emptyset$. Then the Musielak–Orlicz spaces $h$ the limit above can be taken uniformly for every $\varphi$.

By the convexity of $h$ for every $\varphi$, the function $f$ is strictly convex, the divergence satisfies $\partial h$ derivative $\partial f(x)(h)$ exists and defines a sublinear functional. If the function $f$ is strictly convex, the divergence satisfies $B_f(y, x)$ if and only if $x = y$.

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

$$B_f(y, x) = f(y) - f(x) - \varphi_1 f(x)(y - x),$$

for all $x, y \in S$, where $\varphi_1 f(x)(h) = \lim_{t \to 0} (f(x + th) - f(x))/t$ denotes the right-directional derivative of $f$ at $x$ in the direction of $h$. The right-directional derivative $\varphi_1 f(x)(h)$ 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

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

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

$$\mathbb{E}[\varphi(c + \lambda u + (1 - \lambda)v - \psi(u)u_0 - (1 - \lambda)\psi(v)u_0)]$$

$$\leq \lambda \mathbb{E}[\varphi(c + u - \psi(u)u_0)] + (1 - \lambda)\mathbb{E}[\varphi(c + v - \psi(v)u_0)] = 1.$$
Since $\varphi(c + \lambda u + (1 - \lambda)v - \psi(\lambda u + (1 - \lambda)v)u_0)$ has $\mu$-integral equal to 1, we can conclude that the following inequality holds:

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

So we can define the Bregman divergence $B_\psi$ from to the normalizing function $\psi$.

The Bregman divergence $B_\psi: \mathcal{B}_E^2 \times \mathcal{B}_E^2 \to [0, \infty)$ associated to the normalizing function $\psi: \mathcal{B}_E^2 \to [0, \infty)$ is given by

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

Then we define the divergence $D_\psi: \mathcal{B}_E^2 \times \mathcal{B}_E^2 \to [0, \infty)$ related to the $\varphi$-family $\mathcal{F}_\mathcal{E}$ as

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

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

**Lemma 8.** Assume that $\varphi(t, \cdot)$ is continuously differentiable. For any $u \in \mathcal{K}_E^2$, the linear functional $f_u: L_E^\infty \to \mathbb{R}$ given by $f_u(v) = \mathbb{E}[v \varphi'(c + u)]$ is bounded.

**Proof.** Every function $v \in L_E^\infty$ with norm $\|v\|_{\Phi,0} \leq 1$ satisfies $I_\Phi(v) \leq \|u\|_{\Phi,0}$. Then we obtain

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

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

$$(1 - \lambda) \mathbb{E}[v|\varphi'(c + u)] \leq \mathbb{E}[\varphi(c + u + (1 - \lambda)v)] - \mathbb{E}[\varphi(c + u)]$$

$$= \mathbb{E}[\varphi(\lambda(c + \frac{1}{\lambda}u) + (1 - \lambda)(c + |v|))] - \mathbb{E}[\varphi(c + u)]$$

$$\leq \lambda \mathbb{E}[\varphi(c + \frac{1}{\lambda}u)] + (1 - \lambda) \mathbb{E}[\varphi(c + |v|)] - \mathbb{E}[\varphi(c + u)].$$

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

**Lemma 9.** Assume that $\varphi(t, \cdot)$ is continuously differentiable. Then the normalizing function $\psi: \mathcal{K}_E^2 \to \mathbb{R}$ is Gâteaux-differentiable and

$$\partial \psi(u)v = \frac{\mathbb{E}[v \varphi'(c + u - \psi(u))u_0]}{\mathbb{E}[u_0 \varphi'(c + u - \psi(u))]}, \quad (9)$$

**Proof.** According to Lemma 8, the expression in (9) defines a bounded linear functional. Fix functions $u \in \mathcal{K}_E^2$ and $v \in L_E^\infty$. In virtue of Proposition 4 we can find $\varepsilon > 0$ such that $\mathbb{E}[\varphi(c + u + \lambda|v|)] < \infty$, for every $\lambda \in [-\varepsilon, \varepsilon]$. Define

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

$$R_{\psi}(\lambda, k) = \mathbb{E}[\varphi'(c + u + \lambda v - ku_0)].$$
for any \( \lambda \in (-\varepsilon, \varepsilon) \) and \( k \geq 0 \). Since \( \mathcal{K}_c^\varepsilon \) is open, there exist a sufficiently small \( \alpha_0 > 0 \) such that \( u + \lambda v + \alpha|v| \) is in \( \mathcal{K}_c^\varepsilon \) for all \( \alpha \in [-\alpha_0, \alpha_0] \). We can write
\[
g(\lambda + \alpha, k) - g(\lambda, k) = \frac{1}{\alpha} \mathbb{E} \left[ \varphi(c + u + (\lambda + \alpha)v - ku_0) - \varphi(c + u + \lambda v - ku_0) \right].
\]
The function in the expectation above is dominated by the \( \mu \)-integrable function
\[
\frac{1}{\alpha} \{ \varphi(c + u + \lambda v + \alpha_0|v| - ku_0) - \varphi(c + u + \lambda v - ku_0) \}.
\]
By the Dominated Convergence Theorem,
\[
\mathbb{E} \left[ \frac{1}{\alpha} \{ \varphi(c + u + (\lambda + \alpha)v - ku_0) - \varphi(c + u + \lambda v - ku_0) \} \right]
\]
\[
\to \mathbb{E}[v \varphi'(c + u + \lambda v - ku_0)], \quad \text{as } \alpha \to 0,
\]
and, consequently,
\[
\frac{\partial g}{\partial \lambda}(\lambda, k) = \mathbb{E}[v \varphi'(c + u + \lambda v - ku_0)].
\]
Since \( v \varphi'(c + u + \lambda v - ku_0) \) is dominated by the \( \mu \)-integrable function \( |v| \varphi'(c + u + \varepsilon|v| - ku_0) \), we obtain for any sequence \( \lambda_n \to \lambda \),
\[
\mathbb{E}[v \varphi'(c + u + \lambda_n v - ku_0)] \to \mathbb{E}[v \varphi'(c + u + \lambda v - ku_0)], \quad \text{as } n \to \infty.
\]
Thus \( \frac{\partial g}{\partial \lambda}(\lambda, k) \) is continuous with respect to \( \lambda \). Analogously, it can be shown that
\[
\frac{\partial g}{\partial k}(\lambda, k) = -\mathbb{E}[u_0 \varphi'(c + u + \lambda v - ku_0)],
\]
and \( \frac{\partial g}{\partial k}(\lambda, k) \) is continuous with respect to \( k \). The equality \( g(\lambda, k(\lambda)) = \mathbb{E}[\varphi(c + u + \lambda v - k(\lambda)u_0)] = 1 \) defines \( k(\lambda) = \psi(u + \lambda v) \) as an implicit function of \( \lambda \). 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,
\[
\partial \psi(u)(v) = \frac{\partial \psi(u + \lambda v)}{\partial \lambda}(0) = \frac{\mathbb{E}[v \varphi'(c + u - \psi(u)u_0)]}{\mathbb{E}[u_0 \varphi'(c + u - \psi(u)u_0)]},
\]
Thus the expression in (\ref{eq:gauteaux_derivative}) is the Gâteaux-derivative of \( \psi \).

**Lemma 10.** Assume that \( \varphi(t, \cdot) \) is continuously differentiable. Then the divergence \( D_\psi \) does not depend on the parametrization of \( \mathcal{F}_c^\varepsilon \).

**Proof.** For any \( w \in \mathcal{B}_c^\varepsilon \), we denote \( \tilde{c} = c + w - \psi(w)u_0 \). Given \( u, v \in \mathcal{B}_c^\varepsilon \), select \( \tilde{u}, \tilde{v} \in \mathcal{B}_c^\varepsilon \) such that \( \varphi_c(\tilde{u}) = \varphi_c(u) \) and \( \varphi_c(\tilde{v}) = \varphi_c(v) \). Let \( \tilde{\psi} : \mathcal{B}_c^\varepsilon \to [0, \infty) \) be the normalizing function associated to \( \tilde{c} \). These definitions provide
\[
\tilde{c} + \tilde{u} - \tilde{\psi}(\tilde{u})u_0 = c + u - \psi(u)u_0,
\]

and
\[ \tilde{c} + \tilde{v} - \tilde{\psi}(\tilde{u})u_0 = c + v - \psi(u_0). \]
Subtracting these equations, we obtain
\[ [ -\tilde{\psi}(\tilde{v}) + \tilde{\psi}(\tilde{u})]u_0 + (\tilde{v} - \tilde{u}) = [-\psi(v) + \psi(u)]u_0 + (v - u) \]
and, consequently,
\[
\tilde{\psi}(\tilde{v}) - \tilde{\psi}(\tilde{u}) - \frac{E[(\tilde{v} - \tilde{u})\phi'(\tilde{c} + \tilde{u} - \tilde{\psi}(\tilde{u})u_0)]}{E[u_0\phi'(\tilde{c} + \tilde{u} - \tilde{\psi}(\tilde{u})u_0)]} = \psi(v) - \psi(u) - \frac{E[(v - u)\phi'(c + u - \psi(u)u_0)]}{E[u_0\phi'(c + u - \psi(u)u_0)].}
\]
Therefore, \( D_{\psi}(\tilde{u}, \tilde{v}) = D_{\phi}(u, v). \)

Let \( p = \varphi_c(u) \) and \( q = \varphi_c(v) \), for \( u, v \in B_\varphi^c \). We denote the divergence between the probability densities \( p \) and \( q \) by
\[
D(p \parallel q) = D_{\phi}(u, v).
\]
According to Lemma 10, \( D(p \parallel q) \) is well-defined if \( p \) and \( q \) are in the same \( \varphi \)-family. We will find an expression for \( D(p \parallel q) \) where \( p \) and \( q \) are given explicitly. For \( u = 0 \), we have \( D(p \parallel q) = D_{\phi}(0, v) = \psi(v) \), and then
\[
D(p \parallel q) = \frac{E[(-v + \psi(v)u_0)\phi'(c)]}{E[u_0\phi'(c)]}.
\]
Therefore, the divergence between probability densities \( p \) and \( q \) in the same \( \varphi \)-family can be expressed as
\[
D(p \parallel q) = \frac{\mathbb{E}\left[\frac{\varphi^{-1}(p) - \varphi^{-1}(q)}{\varphi^{-1}'(p)}\right]}{\mathbb{E}\left[\frac{u_0}{\varphi^{-1}'(p)}\right]}.
\]
Clearly, the expectation in (10) may not be defined if \( p \) and \( q \) are not in the same \( \varphi \)-family. We extend the divergence in (10) by setting \( D(p \parallel q) = \infty \) if \( p \) and \( q \) are not in the same \( \varphi \)-family. With this extension, the divergence is denoted by \( D_{\varphi} \) and is called the \( \varphi \)-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 \( D_{\varphi} \) is always non-negative, and \( D_{\varphi}(p \parallel q) \) is equal to zero if and only if \( p = q \).

**Example 11.** With the variable \( \kappa \)-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 \( \kappa \)-logarithm \( \ln_{\kappa}(t, u) = \ln_{\kappa(t)}(u) \), we rewrite (10) as
\[
D(p \parallel q) = \frac{\mathbb{E}\left[\frac{\ln_{\kappa}(p) - \ln_{\kappa}(q)}{\ln_{\kappa}'(p)}\right]}{\mathbb{E}\left[\frac{u_0}{\ln_{\kappa}'(p)}\right]},
\]
where \( \ln_{\kappa}(p) \) denotes \( \ln_{\kappa}(t)(p(t)) \). Since the \( \kappa \)-logarithm \( \ln_{\kappa}(u) = \frac{u^{\kappa} - u^{-\kappa}}{2\kappa} \) has derivative \( \ln'_{\kappa}(u) = \frac{1}{u}u^{\kappa} + u^{-\kappa} \), the numerator and denominator in (11) result in

\[
E\left[ \frac{\ln_{\kappa}(p) - \ln_{\kappa}(q)}{\ln'_{\kappa}(p)} \right] = E\left[ \frac{\frac{p^{\kappa} - p^{-\kappa}}{2\kappa} - \frac{q^{\kappa} - q^{-\kappa}}{2\kappa}}{\frac{1}{p^{\kappa} + p^{-\kappa}} - \frac{q^{\kappa} - q^{-\kappa}}{p^{\kappa} + p^{-\kappa}}} \right] = \frac{1}{\kappa} E_p\left[ \frac{p^{\kappa} - p^{-\kappa}}{p^{\kappa} + p^{-\kappa}} - \frac{q^{\kappa} - q^{-\kappa}}{p^{\kappa} + p^{-\kappa}} \right]
\]

and

\[
E\left[ \frac{u_0}{\ln_{\kappa}(p)} \right] = E_p\left[ \frac{2u_0}{p^{\kappa} + p^{-\kappa}} \right]
\]

respectively. Thus (11) can be rewritten as

\[
D_{\kappa}(p \parallel q) = \frac{1}{\kappa} \frac{E_p\left[ \frac{p^{\kappa} - p^{-\kappa}}{p^{\kappa} + p^{-\kappa}} - \frac{q^{\kappa} - q^{-\kappa}}{p^{\kappa} + p^{-\kappa}} \right]}{E_p\left[ \frac{2u_0}{p^{\kappa} + p^{-\kappa}} \right]},
\]

which we called the \( \kappa \)-divergence.

References

[1] G. Pistone, C. Sempi, An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one, Ann. Statist. 23 (5) (1995) 1543–1561. doi:10.1214/aos/1176324311.

[2] G. Pistone, M. P. Rogantin, The exponential statistical manifold: mean parameters, orthogonality and space transformations, Bernoulli 5 (4) (1999) 721–760. doi:10.2307/3318699.

[3] A. Cena, G. Pistone, Exponential statistical manifold, Ann. Inst. Statist. Math. 59 (1) (2007) 27–56. doi:10.1007/s10463-006-0096-y.

[4] M. R. Grasselli, Dual connections in nonparametric classical information geometry, Ann. Inst. Statist. Math. 62 (5) (2010) 873–896. doi:10.1007/s10463-008-0191-3.

[5] S.-i. Amari, H. Nagaoka, Methods of information geometry, Vol. 191 of Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 2000, translated from the 1993 Japanese original by Daishi Harada.

[6] M. K. Murray, J. W. Rice, Differential geometry and statistics, Vol. 48 of Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1993.
[7] P. Gibilisco, E. Riccomagno, M. P. Rogantin, H. P. Wynn (Eds.), Algebraic and geometric methods in statistics, Cambridge University Press, Cambridge, 2010.

[8] D. Petz, Quantum information theory and quantum statistics, Theoretical and Mathematical Physics, Springer-Verlag, Berlin, 2008.

[9] J. Musielak, Orlicz spaces and modular spaces, Vol. 1034 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1983.

[10] M. A. Krasnosel’$'$skiš, J. B. Rutickiš, Convex functions and Orlicz spaces, Translated from the first Russian edition by Leo F. Boron, P. Noordhoff Ltd., Groningen, 1961.

[11] M. M. Rao, Z. D. Ren, Theory of Orlicz spaces, Vol. 146 of Monographs and Textbooks in Pure and Applied Mathematics, Marcel Dekker Inc., New York, 1991.

[12] G. Kaniadakis, M. Lissia, A. M. Scarfone, Deformed logarithms and entropies, Phys. A 340 (1-3) (2004) 41–49, news and expectations in thermo-statistics. \( \text{doi:10.1016/j.physa.2004.03.075} \)

[13] J. Naudts, Estimators, escort probabilities, and φ-exponential families in statistical physics, JIPAM. J. Inequal. Pure Appl. Math. 5 (4) (2004) Article 102, 15 pp. (electronic).

[14] J. Naudts, Generalised exponential families and associated entropy functions, Entropy 10 (3) (2008) 131–149. \( \text{doi:10.3390/entropy-e10030131} \)

[15] L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, USSR Computational Mathematics and Mathematical Physics 7 (3) (1967) 200–217. \( \text{doi:10.1016/0041-5553(67)90040-7} \)