Sup-Sums Principles for F -Divergence and a New Definition for t -Entropy

V. I. Bakhtin^{1,2} · A. V. Lebedev^{2,3}

Received: 29 February 2020 / Revised: 14 August 2020 / Published online: 4 November 2020
© The Author(s) 2020

Abstract

The article presents new sup-sums principles for integral F -divergence for arbitrary convex functions F on the whole real axis and arbitrary (not necessarily positive and normalized) measures. Among applications of these results, we work out a new ‘integral’ definition for t -entropy explicitly establishing its relation to Kullback–Leibler divergence.

Keywords F -divergence · Kullback–Leibler divergence · Sup-sums principle · Partition of unity · t -entropy

Mathematics Subject Classification (2020) 26D15 · 37A35 · 47B37 · 62H20 · 94A17

1 Introduction

The notion of F -divergence was introduced and originally studied in analysis of probability distributions by [2, 14, 19]. It is defined in the following way. Let Q and P be two probability distributions over a space Ω such that Q is absolutely continuous with respect to P . Then, for a convex function $F : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that $F(1) = 0$, the F -divergence $D_F(Q \| P)$ of Q from P is defined as

✉ V. I. Bakhtin
bakhtin@tut.by

A. V. Lebedev
lebedev@bsu.by

¹ John Paul II Catholic University of Lublin, Lublin, Poland

² Belarusian State University, Minsk, Belarus

³ University of Białystok, Białystok, Poland

$$D_F(Q\|P) := \int_{\Omega} F\left(\frac{dQ}{dP}\right) dP, \quad (1)$$

where dQ/dP is the Radon–Nikodym derivative of Q with respect to P .

Since its introduction, the F -divergence has been intensively exploited and analysed due to the fact that by taking appropriate functions F one arrives here at numerous important divergences such as Kullback–Leibler divergence, Hellinger distance, Pearson χ^2 -divergence, etc.

A comprehensive analysis of F -divergence was worked out by Liese and Vajda in [16] where a sup-sums principle for space partitions was established as well [16, Theorem 16].

In fact, formula (1) can be extended to arbitrary real-valued measures Q . Moreover, for such measures, Q , the value $D_F(Q\|P)$ possesses a substantial statistical meaning. In [13,20,23], it was shown that the value $e^{-nD_F(Q\|P)}$ determines the asymptotics for conditional probabilities of large deviations for a certain family of weighted empirical measures that are close to Q , where F is the rate function for large deviations for the sequence of random weights. In [12], the F -divergences for real-valued measures Q were applied for parametric estimation and testing.

The object of the present article is a general F -divergence associated with an arbitrary convex function F that is defined on the whole real axis and can take infinite values and arbitrary real-valued (not necessarily positive, normalized, and absolutely continuous) measures. For this F -divergence, we derive a number of new sup-sums principles exploiting as measurable so also continuous partitions of unity (Theorems 10, 12, and 14 of the article). In particular, they disclose the passage procedure from the F -divergence on a finite phase space to the F -divergence on an arbitrary measurable space. This passage involves additional components $F'(\pm\infty)$.

On the base of sup-sums principles obtained, we derive the corresponding sup-sums principle for Kullback–Leibler divergence (Theorem 15) leading also naturally to its new definition for measures that are not probability ones. The initial variant of the sup-sums principle for a particular case of F -divergence (mutual information) for mutually absolutely continuous probability measures was established by Gelfand, Kolmogorov, and Yaglom in [15].

As one more substantial application of integral sup-sums principles, we obtain a new formula for t -entropy. The t -entropy plays a fundamental role in the spectral analysis of operators associated with dynamical systems (cf. Theorem 20) and is a key ingredient in ‘entropy statistic theorem’. The latter statement, in the spectral theory of weighted shift and transfer operators, plays the role analogous to Shannon–McMillan–Breiman theorem in information theory [1,18] and its important corollary known as ‘asymptotic equipartition property’ [10, p. 135]. Up to now, the definition of t -entropy has been formulated in a rather sophisticated manner in terms of actions of transfer operators on continuous partitions of unity (for more details, see Sect. 5). In Theorem 21, we give a fundamentally new ‘integral’ definition of t -entropy explicitly establishing its relation to Kullback–Leibler divergence.

2 Sup-Sums F -Divergence

Consider an arbitrary convex function $F: \mathbb{R} \rightarrow (-\infty, +\infty]$. Let

$$F'(+\infty) := \lim_{t \rightarrow +\infty} \frac{F(t)}{t}, \quad F'(-\infty) := \lim_{t \rightarrow -\infty} \frac{F(t)}{t}. \quad (2)$$

Obviously, both limits do exist, and the value of $F'(+\infty)$ may be finite or equal to $+\infty$ while $F'(-\infty)$ may be finite or equal to $-\infty$.

We adopt the following agreement. The product $0F(x/0)$ for $x \neq 0$ will be defined as the limit $\lim_{t \rightarrow +0} tF(x/t)$, and for $x = 0$, it will be assumed to be zero. In other words,

$$0F\left(\frac{x}{0}\right) = \begin{cases} xF'(+\infty), & \text{if } x > 0, \\ xF'(-\infty), & \text{if } x < 0, \\ 0, & \text{if } x = 0. \end{cases} \quad (3)$$

Let μ be a finite nonnegative measure and let ν be a finite real-valued measure, both defined on the same measurable space $(X, \mathfrak{A})$. For measurable functions g on $(X, \mathfrak{A})$, we write

$$\mu[g] := \int_X g \, d\mu, \quad \nu[g] := \int_X g \, d\nu$$

(provided the integrals do exist).

By a *measurable partition of unity*, we will understand a finite set $G = \{g_1, \dots, g_k\}$ of nonnegative measurable functions on $(X, \mathfrak{A})$ such that $\sum_i g_i \equiv 1$.

Now we introduce the main object of the paper. For any convex function $F: \mathbb{R} \rightarrow (-\infty, +\infty]$, the *sup-sums F -divergence* $\rho_F(\mu, \nu)$ is defined as

$$\rho_F(\mu, \nu) := \sup_G \sum_{g \in G} \mu[g] F\left(\frac{\nu[g]}{\mu[g]}\right), \quad (4)$$

where the supremum is taken over the set of all measurable partitions of unity G , and we assume that if $\mu[g] = 0$, then the corresponding summand in the right-hand part is defined according to convention (3).

The relation of the sup-sums F -divergence to the usual (integral) F -divergence will be uncovered in the next section.

The following two lemmas present important properties of the function $sF(x/s)$ used in the definition (4).

Lemma 1 *For any convex function F and all $s, t \geq 0$ and $x, y \in \mathbb{R}$,*

$$(s+t)F\left(\frac{x+y}{s+t}\right) \leq sF\left(\frac{x}{s}\right) + tF\left(\frac{y}{t}\right). \quad (5)$$

Each convex function F on the real axis is superlinear, i.e.,

$$F(t) \geq At + B \quad (6)$$

for some constants $A, B \in \mathbb{R}$ and all $t \in \mathbb{R}$.

Lemma 2 *If a convex function F satisfies condition (6), then for all $s \geq 0$ and $x \in \mathbb{R}$,*

$$sF\left(\frac{x}{s}\right) \geq Ax + Bs. \quad (7)$$

Now we proceed to description of the technical properties of $\rho_F(\mu, \nu)$.

Lemma 3 *The value of $\rho_F(\mu, \nu)$ does not change if we use countable partitions of unity G in (4) instead of finite ones.*

Proposition 4 *The function $\rho_F(\mu, \nu)$ is subadditive with respect to the pair (μ, ν) . That is, for any finite nonnegative measures μ_1, μ_2 and any finite real-valued measures ν_1, ν_2 ,*

$$\rho_F(\mu_1 + \mu_2, \nu_1 + \nu_2) \leq \rho_F(\mu_1, \nu_1) + \rho_F(\mu_2, \nu_2). \quad (8)$$

For any measure ν and a bounded measurable function f on a measurable space $(X, \mathfrak{A})$, we define a real-valued measure $f\nu$ by the formula

$$f\nu[g] := \nu[fg] = \int_X gf \, d\nu, \quad g \in L^1(X, \nu).$$

Proposition 5 *Let μ, ν be finite measures, where μ is nonnegative and ν is real-valued, and let f_1, f_2 be nonnegative bounded measurable functions on $(X, \mathfrak{A})$. Then,*

$$\rho_F((f_1 + f_2)\mu, (f_1 + f_2)\nu) = \rho_F(f_1\mu, f_1\nu) + \rho_F(f_2\mu, f_2\nu).$$

This means that the function $\rho_F(f\mu, f\nu)$ is additive with respect to f .

Generally, a real-valued measure ν is decomposed into three components

$$\nu = \nu_a + \nu_s^+ + \nu_s^-, \quad (9)$$

where ν_a is absolutely continuous, ν_s^+ is positive and singular, and ν_s^- is negative and singular (with respect to μ).

The next result describes the corresponding decomposition of $\rho_F(\mu, \nu)$.

Theorem 6 *Let μ, ν be finite measures on a measurable space $(X, \mathfrak{A})$, where μ is nonnegative and ν is real-valued. Then,*

$$\rho_F(\mu, \nu) = \rho_F(\mu, \nu_a) + \rho_F(0, \nu_s^+) + \rho_F(0, \nu_s^-), \quad (10)$$

where each term may be finite or equal to $+\infty$, and

$$\rho_F(0, v_s^+) = v_s^+(X)F'(+\infty), \quad (11)$$

$$\rho_F(0, v_s^-) = v_s^-(X)F'(-\infty). \quad (12)$$

Here, we assume that if $v_s^+ = 0$ or $v_s^- = 0$ then the corresponding product in the right-hand part of (11) or (12) is zero regardless of the value of $F'(\pm\infty)$.

There is quite a number of objects in analysis where one has to exploit not measurable partitions of unity but continuous ones (one of them will be considered in Sect. 5). To discuss this setting in our context, we need the next definition.

Let X be a topological space and let μ, ν be finite Borel measures, where μ is nonnegative and ν is real-valued. For any convex function $F: \mathbb{R} \rightarrow (-\infty, +\infty]$ set

$$\rho_{F,c}(\mu, \nu) := \sup_G \sum_{g \in G} \mu[g] F\left(\frac{\nu[g]}{\mu[g]}\right), \quad (13)$$

where the supremum is taken over the set of all (finite) continuous partitions of unity G , and we assume that if $\mu[g] = 0$ then the corresponding summand in the right-hand part is defined according to convention (3).

Theorem 7 *Let μ, ν be finite Borel measures on a metric space X , where μ is non-negative and ν is real-valued. Then, for any convex lower semicontinuous function F ,*

$$\rho_{F,c}(\mu, \nu) = \rho_F(\mu, \nu).$$

Remark 8 In fact instead of metrizable of X in Theorem 7, it suffices to require the density of the set of continuous functions $C(X)$ in the space $L^1(X, \mu + |\nu|)$ (which is always true for metrizable space X or, as a variant, for *regular* measures μ, ν).

Now, let us prove the above formulated results.

Proof of Lemma 2 If $s > 0$, then (7) follows immediately from (6). Note that (6) and (2) imply inequalities

$$F'(+\infty) \geq A, \quad F'(-\infty) \leq A. \quad (14)$$

In turn along with (3), they imply (7), provided $s = 0$ and $x \neq 0$. Finally, in case $s = 0$ and $x = 0$, both sides in (7) become zero. $\square$

Proof of Lemma 1 If $s, t > 0$, then by convexity of F ,

$$sF\left(\frac{x}{s}\right) + tF\left(\frac{y}{t}\right) = (s+t)\left(\frac{s}{s+t}F\left(\frac{x}{s}\right) + \frac{t}{s+t}F\left(\frac{y}{t}\right)\right) \geq (s+t)F\left(\frac{x+y}{s+t}\right).$$

Consider the case when $s > 0$ and $t = 0$.

If $y = 0$, then (5) turns into the equality $sF(x/s) = sF(x/s)$.

Suppose now that $y > 0$. If at least one summand in the right-hand part of (5) is infinite, then (5) holds true. If both summands $sF(x/s)$ and $0F(y/0) = yF'(+\infty)$ are finite then the function F must be finite and continuous on the whole interval $(x/s, +\infty)$. Hence, in (5), one can pass to a limit as $t \rightarrow +0$ and obtain the desired inequality

$$sF\left(\frac{x+y}{s}\right) \leq sF\left(\frac{x}{s}\right) + 0F\left(\frac{y}{0}\right).$$

The case $y < 0$ is treated similarly.

It remains to analyse the case $s, t = 0$ and $x, y \neq 0$. If x and y have the same sign (say $x, y > 0$), then (5) turns into equality:

$$(x+y)F'(+\infty) = xF'(+\infty) + yF'(+\infty).$$

Suppose x, y have different signs (say $x < 0$ and $y > 0$). Recall that $F'(+\infty) \geq F'(-\infty)$ (see (14)). Therefore, in any case,

$$xF'(-\infty) + yF'(+\infty) \geq \begin{cases} (x+y)F'(-\infty), & \text{if } x+y < 0, \\ (x+y)F'(+\infty), & \text{if } x+y > 0, \\ 0, & \text{if } x+y = 0, \end{cases}$$

which means that

$$0F\left(\frac{x}{0}\right) + 0F\left(\frac{y}{0}\right) \geq 0F\left(\frac{x+y}{0}\right).$$

Thus, Lemma 1 is proved in all cases. $\square$

Proof of Lemma 3 Consider a countable partition of unity $G = \{g_1, g_2, \dots\}$. First, we prove that in this case, the sum in (4) is well-defined, i.e., that the limit

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mu[g_i] F\left(\frac{v[g_i]}{\mu[g_i]}\right) \quad (15)$$

does exist, being either finite or equal to $+\infty$.

Set $h_n = \sum_{i \geq n} g_i$. Then, by Levi's monotone convergence theorem,

$$\lim_{n \rightarrow \infty} \mu[h_n] = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} |v|[h_n] = 0, \quad (16)$$

where $|v|$ denotes the total variation of v . Lemma 2 implies that

$$\sum_{i=n}^m \mu[g_i] F\left(\frac{v[g_i]}{\mu[g_i]}\right) \geq \sum_{i=n}^m (Av[g_i] + B\mu[g_i]) \geq -|A||v|[h_n] - |B|\mu[h_n]. \quad (17)$$

It follows from (16) and (17) that for any $\varepsilon > 0$, there exists N such that for all $n > N$ and $m \geq n$,

$$\sum_{i=n}^m \mu[g_i] F\left(\frac{\nu[g_i]}{\mu[g_i]}\right) > -\varepsilon. \quad (18)$$

Now, we have two possibilities: if for any $\varepsilon > 0$ there exists N such that for all $n > N$ and $m \geq n$,

$$\sum_{i=n}^m \mu[g_i] F\left(\frac{\nu[g_i]}{\mu[g_i]}\right) < \varepsilon, \quad (19)$$

then limit (15) does exist (being finite when all the summands in (15) are finite and equal to $+\infty$ when there is at least one infinite summand); otherwise, if assumption (19) fails, using its negation and (18) one can easily show that the limit still exists and equals $+\infty$.

Now, let us check equivalence of finite and countable partitions for use in (4).

Each finite partition of unity G in (4) may be transformed into a countable one by adding countably many zero elements, so transition from finite to countable partitions cannot decrease the value of $\rho_F(\mu, \nu)$. Thus, it suffices to prove that it cannot increase as well.

Let $\rho_F(\mu, \nu)$ be defined by (4) using countable partitions G . Then, for any $c < \rho_F(\mu, \nu)$, there exists a countable partition of unity $G = \{g_1, g_2, \dots\}$ such that

$$\sum_{i=1}^{\infty} \mu[g_i] F\left(\frac{\nu[g_i]}{\mu[g_i]}\right) > c. \quad (20)$$

Set $h_n = \sum_{i \geq n} g_i$. Combining Lemma 2 and (16), we obtain

$$\liminf_{n \rightarrow \infty} \mu[h_n] F\left(\frac{\nu[h_n]}{\mu[h_n]}\right) \geq \liminf_{n \rightarrow \infty} (A\nu[h_n] + B\mu[h_n]) = 0. \quad (21)$$

Consider a finite partition of unity $G_n = \{g_1, \dots, g_{n-1}, h_n\}$. Now, (20) and (21) imply

$$\sup_n \left(\sum_{i=1}^{n-1} \mu[g_i] F\left(\frac{\nu[g_i]}{\mu[g_i]}\right) + \mu[h_n] F\left(\frac{\nu[h_n]}{\mu[h_n]}\right) \right) \geq \sum_{i=1}^{\infty} \mu[g_i] F\left(\frac{\nu[g_i]}{\mu[g_i]}\right) > c.$$

Then, (20) is valid for some G_n instead of G , which along with arbitrariness of the constant $c < \rho_F(\mu, \nu)$ implies the statement of Lemma 3. $\square$

Proof of Theorem 4 If g is an element of a measurable partition of unity G , then by Lemma 1,

$$(\mu_1[g] + \mu_2[g])F\left(\frac{\nu_1[g] + \nu_2[g]}{\mu_1[g] + \mu_2[g]}\right) \leq \mu_1[g]F\left(\frac{\nu_1[g]}{\mu_1[g]}\right) + \mu_2[g]F\left(\frac{\nu_2[g]}{\mu_2[g]}\right).$$

Summing this up over $g \in G$ and passing to suprema gives (8). $\square$

Proof of Theorem 5 From Theorem 4, it follows that

$$\rho_F((f_1 + f_2)\mu, (f_1 + f_2)v) \leq \rho_F(f_1\mu, f_1v) + \rho_F(f_2\mu, f_2v).$$

So, it suffices to prove the inverse inequality.

By definition, for any $c_i < \rho_F(f_i\mu, f_iv)$, $i = 1, 2$, there exist measurable partitions of unity G_i , $i = 1, 2$, such that

$$\sum_{g \in G_i} \mu[f_i g]F\left(\frac{\nu[f_i g]}{\mu[f_i g]}\right) > c_i, \quad i = 1, 2. \quad (22)$$

For each $g \in G_i$ let us define the function

$$h_g = \begin{cases} f_i g / (f_1 + f_2), & \text{if } f_1 + f_2 > 0, \\ g/2, & \text{if } f_1 + f_2 = 0. \end{cases}$$

Evidently, the collection $H = \{h_g \mid g \in G_1 \cup G_2\}$ forms a measurable partition of unity. Note that for each $g \in G_i$, we have the equality $(f_1 + f_2)h_g = f_i g$. Therefore,

$$\sum_{h_g \in H} \mu[(f_1 + f_2)h_g]F\left(\frac{\nu[(f_1 + f_2)h_g]}{\mu[(f_1 + f_2)h_g]}\right) = \sum_{i=1}^2 \sum_{g \in G_i} \mu[f_i g]F\left(\frac{\nu[f_i g]}{\mu[f_i g]}\right). \quad (23)$$

From (22), (23) it follows that

$$\rho_F((f_1 + f_2)\mu, (f_1 + f_2)v) > c_1 + c_2$$

and, by arbitrariness of $c_i < \rho_F(f_i\mu, f_iv)$,

$$\rho_F((f_1 + f_2)\mu, (f_1 + f_2)v) \geq \rho_F(f_1\mu, f_1v) + \rho_F(f_2\mu, f_2v).$$

$\square$

Proof of Theorem 6 The space X can be decomposed into three disjoint measurable parts, say $X = X_a \sqcup X_s^+ \sqcup X_s^-$, such that the measures μ and ν_a are supported on X_a while ν_s^+ , ν_s^- are, respectively, supported on X_s^+ , X_s^- . Denote by f_a , f_s^+ , f_s^- characteristic functions of these disjoint parts. Then,

$$f_a\mu = \mu, \quad f_av = \nu_a, \quad f_s^+\mu = 0, \quad f_s^+v = \nu_s^+, \quad f_s^-\mu = 0, \quad f_s^-v = \nu_s^-,$$

and hence (10) follows from Theorem 5.

Proofs of equalities (11) and (12) are similar. For example,

$$\rho_F(0, v_s^+) = \sup_G \sum_{g \in G} 0F\left(\frac{v_s^+[g]}{0}\right) = \sup_G \sum_{g \in G} v_s^+[g]F'(+\infty) = v_s^+(X)F'(+\infty).$$

□

To prove Theorem 7, we need the next

Lemma 9 *Let μ be a positive finite Borel measure on a topological space X such that $C(X)$ is dense in $L^1(X, \mu)$. Then, for any measurable partition of unity $G = \{g_1, \dots, g_n\}$ on X and any $\varepsilon > 0$, there exists a continuous partition of unity $H = \{h_1, \dots, h_n\}$ on X such that $\|h_i - g_i\| < \varepsilon$ in $L^1(X, \mu)$ for all $i \in \overline{1, n}$.*

Proof Choose a small $\delta > 0$ and approximate each g_i by a continuous function f_i satisfying $\|f_i - g_i\| < \delta$ in the space $L^1(X, \mu)$. Without loss of generality, we can assume that the functions f_i are strictly positive (which can always be guaranteed by replacing each f_i by $\max\{f_i, 0\} + \gamma$ with a small $\gamma > 0$). Now define a continuous partition of unity with elements

$$h_i := \frac{f_i}{\sum_{j=1}^n f_j}, \quad i = 1, \dots, n.$$

Clearly,

$$|h_i - f_i| = \left| \frac{1 - \sum_{j=1}^n f_j}{\sum_{j=1}^n f_j} f_i \right| \leq \left| 1 - \sum_{j=1}^n f_j \right| \leq \sum_{j=1}^n |g_j - f_j|,$$

which implies the estimate

$$\|h_i - g_i\| \leq \|h_i - f_i\| + \|f_i - g_i\| \leq n\delta + \delta.$$

Since δ is arbitrary, this finishes the proof. □

Proof of Theorem 7 Since any continuous partition of unity is measurable it follows that $\rho_{F,C}(\mu, \nu) \leq \rho_F(\mu, \nu)$, and it is enough to prove the opposite inequality.

As in the proof of Theorem 6, the space X can be decomposed into three disjoint parts, $X = X_a \sqcup X_s^+ \sqcup X_s^-$, such that the measures μ and ν_a are supported on X_a while ν_s^+ , ν_s^- are, respectively, supported on X_s^+ , X_s^- . Denote by f_a , f_s^+ , f_s^- the characteristic functions of these disjoint parts.

Theorem 6 gives the following representation of $\rho_F(\mu, \nu)$:

$$\rho_F(\mu, \nu) = \sup_G \sum_{g \in G} \mu[g]F\left(\frac{\nu_a[g]}{\mu[g]}\right) + 0F\left(\frac{\nu_s^+[f_s^+]}{0}\right) + 0F\left(\frac{\nu_s^-[f_s^-]}{0}\right). \quad (24)$$

Suppose that $\nu_s^+[f_s^+] > 0$ and $\nu_s^-[f_s^-] < 0$ (otherwise the corresponding summands in the right-hand side of (24) may be omitted).

Note that in (24), one can assume that $\mu[g] > 0$ for all g since on the one hand the summands with $\mu[g] = 0$ are equal to 0 according to definition and on the other hand once $\mu[g'] = 0$ and $\mu[g''] > 0$ the pair g', g'' can be replaced by one element $g = g' + g''$ in the partition G that does not change the sum in (24) due to absolute continuity of ν_a with respect to μ .

Now recalling lower semicontinuity of F and definition of $F'(\pm\infty)$, the proof of theorem completes by applying Lemma 9 to partitions of unity $G' = \{f_ag \mid g \in G\} \cup \{f_s^+, f_s^-\}$ in the space $L^1(X, \mu + |\nu|)$. $\square$

3 Sup-Sums Principles for Integral F -Divergence

In this section, we present a number of the principal results of the article uncovering interrelation between sup-sums F -divergences and integral F -divergence.

Theorem 10 (sup-sums principle for partitions of unity) *Let μ and ν be two finite measures on a measurable space $(X, \mathfrak{A})$, where μ is nonnegative and ν is real-valued, and let $\nu = \nu_a + \nu_s^+ + \nu_s^-$ be the decomposition (9). Then,*

$$\rho_F(\mu, \nu_a) = \int_X F\left(\frac{d\nu_a}{d\mu}\right) d\mu, \quad (25)$$

and

$$\rho_F(\mu, \nu) = \int_X F\left(\frac{d\nu_a}{d\mu}\right) d\mu + \nu_s^+(X)F'(+\infty) + \nu_s^-(X)F'(-\infty). \quad (26)$$

Here, $d\nu_a/d\mu$ denotes the Radon–Nikodym derivative, and we assume that if $\nu_s^+ = 0$ or $\nu_s^- = 0$, then the corresponding product in the right-hand part of (26) is zero regardless of the value of $F'(\pm\infty)$.

Corollary 11 *For any $f \in L^1(X, \mu)$ and any convex function $F: \mathbb{R} \rightarrow (-\infty, +\infty]$,*

$$\int_X F(f) d\mu = \sup_G \sum_{g \in G} \mu[g] F\left(\frac{\mu[fg]}{\mu[g]}\right),$$

where the supremum is taken over all measurable partitions of unity G .

Along with partitions of unity one can also use space partitions. Namely, by a measurable partition of space X , we mean a finite family $\Gamma = \{\Delta_1, \dots, \Delta_k\}$ of sets $\Delta_i \in \mathfrak{A}$ such that $\Delta_1 \sqcup \dots \sqcup \Delta_k = X$.

For any convex function $F: \mathbb{R} \rightarrow (-\infty, +\infty]$ put

$$\rho_{F,X}(\mu, \nu) := \sup_{\Gamma} \sum_{\Delta \in \Gamma} \mu(\Delta) F\left(\frac{\nu(\Delta)}{\mu(\Delta)}\right), \quad (27)$$

where the supremum is taken over the set of all measurable partitions Γ of space X , and we assume that if $\mu(\Delta) = 0$, then the corresponding summand in the right-hand part is defined according to convention (3).

The argument of the proof of Lemma 3 shows that expression (27) preserves its value whether we use finite or countable measurable partitions of the space X .

The next statement is a ‘space’ variant of Theorem 10.

Theorem 12 (sup-sums principle for space partitions) *Under the assumptions of Theorem 10, we have $\rho_{F,X}(\mu, \nu) = \rho_F(\mu, \nu)$. Thus, the equalities (25) and (26) are valid with $\rho_{F,X}(\mu, \nu)$ used instead of $\rho_F(\mu, \nu)$.*

Remark 13 In the classical situation, i.e., for a convex function $F: (0, +\infty) \rightarrow \mathbb{R}$ and probability measures μ and ν the sup-sums principle (25), (26) with $\rho_{F,X}(\mu, \nu)$ was established by Vajda [24]. A different proof, based on generalized Taylor expansion of a convex function, is given in [16, Theorem 16]. The paper [16] is a good source of information on the classical F -divergence.

In many fields of analysis, one has to use continuous partitions of unity (see, in particular, Sect. 5). The next theorem presents the corresponding refinement of Theorem 10.

Theorem 14 (sup-sums principle for continuous partitions) *Let X be a topological space and μ and ν be two Borel finite measures, where μ is nonnegative and ν is real-valued. If the set $C(X)$ of continuous functions is dense in $L^1(X, \mu + |\nu|)$ (which is always true for a metrizable space X or, as a variant, for regular measures μ and ν) and F is a convex lower semicontinuous function then (25) and (26) are valid with $\rho_{F,C}(\cdot, \cdot)$ (13) substituted for $\rho_F(\cdot, \cdot)$.*

It is worth mentioning that there are at least two different ways to define the value of $\rho_F(\mu, \nu)$ for a measure ν that is not absolutely continuous with respect to μ . Let us explain them in the case of finite set $X = \{1, \dots, K\}$. In this case, the measures μ, ν have the form $\mu = (\mu_1, \dots, \mu_K)$, $\nu = (\nu_1, \dots, \nu_K)$ and then

$$\rho_F(\mu, \nu) = D_F(\nu \parallel \mu) = \sum_{i=1}^K \mu_i F\left(\frac{\nu_i}{\mu_i}\right).$$

The question is how to define the product $\mu_i F(\nu_i/\mu_i)$ when $\mu_i = 0$ and $\nu_i \neq 0$.

The first way (adopted in the present paper as well as in [16, 24]) is to put

$$0F\left(\frac{\nu_i}{0}\right) = \lim_{\mu_i \rightarrow +0} \mu_i F\left(\frac{\nu_i}{\mu_i}\right).$$

Under this approach, the function $\rho_F(\mu, \nu)$ depends continuously on μ . Namely this property enables us to establish in Theorem 14 a link between sup-sums principles for measurable and continuous partitions of unity, that is inevitable for applications to the spectral objects in Sect. 5.

Alternatively, $-\rho_F(\mu, \nu)$ may be treated as the exponential rate for conditional probabilities of large deviations for a certain family of weighted empirical measures. Then, we have to put $0F(v_i/0) = +\infty$ and, respectively, $\rho_F(\mu, \nu) = +\infty$ whenever ν is not absolutely continuous with respect to μ (for details see [11, 13, 20, 23]).

Under the second approach to the definition of F -divergence, the analogues of all the above-stated results (Propositions 4, 5 and Theorems 7, 10, 12) can be formulated and proved. However, in this setting, they will be meaningful only for absolutely continuous measures ν , while the singular case becomes trivial.

Proof of Theorem 10 Note that (26) follows from (25) along with Theorem 6.

Let us check that for each (no matter finite or countable) measurable partition of unity G , we have

$$\sum_{g \in G} \mu[g] F\left(\frac{\nu_a[g]}{\mu[g]}\right) \leq \int_X F\left(\frac{d\nu_a}{d\mu}\right) d\mu. \quad (28)$$

Indeed,

$$\begin{aligned} \sum_{g \in G} \mu[g] F\left(\frac{\nu_a[g]}{\mu[g]}\right) &= \sum_{\mu[g] > 0} \mu[g] F\left(\int_X \frac{g}{\mu[g]} d\nu_a\right) \\ &= \sum_{\mu[g] > 0} \mu[g] F\left(\int_X \frac{g}{\mu[g]} \frac{d\nu_a}{d\mu} d\mu\right) \\ &\leq \sum_{\mu[g] > 0} \mu[g] \int_X \frac{g}{\mu[g]} F\left(\frac{d\nu_a}{d\mu}\right) d\mu \\ &= \int_X \sum_{g \in G} g F\left(\frac{d\nu_a}{d\mu}\right) d\mu = \int_X F\left(\frac{d\nu_a}{d\mu}\right) d\mu, \end{aligned}$$

where we used Jensen's inequality for the probability measure $(g/\mu[g])d\mu$ and that by convention (3) and absolute continuity of ν_a all the summands with $\mu[g] = 0$ are zero.

From (28), it follows that the left-hand part in (25) does not exceed the right-hand one, and to finish the proof of Theorem 10 we have to verify the inequality

$$\sup_G \sum_{g \in G} \mu[g] F\left(\frac{\nu_a[g]}{\mu[g]}\right) \geq \int_X F\left(\frac{d\nu_a}{d\mu}\right) d\mu. \quad (29)$$

For the convex function F under consideration, there exists a partition of real axis by three points $-\infty \leq a \leq b \leq c \leq +\infty$ such that

- (i) $F(y) = +\infty$ for $y < a$ and $y > c$;
- (ii) $F(y)$ is nonincreasing, finite and continuous on (a, b) ;
- (iii) $F(y)$ is nondecreasing, finite and continuous on (b, c) .

Let us decompose X into seven subsets

$$X_{<a}, \quad X_a, \quad X_{(a,b)}, \quad X_b, \quad X_{(b,c)}, \quad X_c, \quad X_{>c}, \quad (30)$$

defined, respectively, by the conditions

$$\frac{dv_a}{d\mu}(x) < a, \quad \frac{dv_a}{d\mu}(x) = a, \quad a < \frac{dv_a}{d\mu}(x) < b, \quad \dots, \quad \frac{dv_a}{d\mu}(x) = c, \quad \frac{dv_a}{d\mu}(x) > c.$$

Some of these sets may be empty; for example, if the function F decreases on (a, c) , then $b = c$ and $X_{(b,c)} = \emptyset$, and if F is finite everywhere, then the sets $X_{<a}$, X_a , X_c , $X_{>c}$ will be empty.

Evidently, it is enough to prove inequality (29) for each of the sets (30) separately and then sum the components. In doing so, partitions of unity G on these sets should also be defined separately.

For the sets $X_{<a}$, X_a , X_b , X_c , $X_{>c}$ (some of them may be empty) inequality (29) is verified easily: it is sufficient to take a trivial partition G consisting of a single unit function on the set considered.

Now consider the set $X_{(a,b)}$. Let us take an arbitrary number $\varepsilon > 0$ and set

$$Y_i = \{y \in (a, b) \mid i\varepsilon \leq F(y) < i\varepsilon + \varepsilon\}, \quad i \in \mathbb{Z}, \quad (31)$$

$$X_i = \left\{x \in X \mid \frac{dv_a}{d\mu}(x) \in Y_i\right\}, \quad i \in \mathbb{Z}. \quad (32)$$

Clearly, the sets X_i form a partition of $X_{(a,b)}$ and their characteristic functions (that we denote by g_i) form a measurable partition of unity on $X_{(a,b)}$.

Note that by monotonicity of F on (a, b) , the sets Y_i are convex. Therefore, if $\mu(X_i) > 0$, then

$$\frac{v_a(X_i)}{\mu(X_i)} = \frac{1}{\mu(X_i)} \int_{X_i} \frac{dv_a}{d\mu} d\mu \in Y_i,$$

and by definition of Y_i , we have

$$i\varepsilon \leq F\left(\frac{v_a(X_i)}{\mu(X_i)}\right) < i\varepsilon + \varepsilon. \quad (33)$$

Now, (31), (32), (33) imply that

$$\begin{aligned} \int_{X_{(a,b)}} F\left(\frac{dv_a}{d\mu}\right) d\mu &= \sum_{i \in \mathbb{Z}} \int_{X_i} F\left(\frac{dv_a}{d\mu}\right) d\mu \leq \sum_{i \in \mathbb{Z}} (i\varepsilon + \varepsilon) \mu(X_i) \\ &\leq \sum_{i \in \mathbb{Z}} \mu(X_i) \left(F\left(\frac{v_a(X_i)}{\mu(X_i)}\right) + \varepsilon \right) \\ &= \sum_{i \in \mathbb{Z}} \mu[g_i] F\left(\frac{v_a[g_i]}{\mu[g_i]}\right) + \varepsilon \mu(X_{(a,b)}). \end{aligned}$$

By arbitrariness of ε , this implies inequality (29) for the set $X_{(a,b)}$.

For the set $X_{(b,c)}$, it is verified in the same way. Thus, Theorem 10 is proved. $\square$

Proof of Corollary 11 Take ν such that $d\nu/d\mu = f$ and apply (25). $\square$

Proof of Theorem 12 Every space partition $\Delta_1, \dots, \Delta_k$ is defined by the partition of unity consisting of the corresponding characteristic functions. Thus, $\rho_{F,X}(\mu, \nu) \leq \rho_F(\mu, \nu)$. The rest of the proof coincides with the ending part of the proof of Theorem 10 (starting from formula (29)), where only the space partitions are used. $\square$

Proof of Theorem 14 Apply Theorem 10 along with Theorem 7 bearing in mind Remark 8. $\square$

4 Sup-Sums Principle for Kullback–Leibler Divergence, etc

If μ and ν are probability measures on $(X, \mathfrak{A})$ and μ is absolutely continuous with respect to ν , then *Kullback–Leibler divergence* D_{KL} is defined as

$$D_{\text{KL}}(\mu \| \nu) := \int_X \ln \left(\frac{d\mu}{d\nu} \right) d\mu. \quad (34)$$

The principal philosophy behind the results we are going to discuss is not new. Namely, an analogue of Theorem 15 for *space partitions* of X (cf. (27)) goes back to Gelfand, Kolmogorov, and Yaglom [15]. However, the results obtained in the foregoing section give this field a new flavour and among the basic novelties here is the use of *continuous partitions of unity* (see, in particular, Remarks 17 and 16), which serves as an inevitable apparatus in the analysis of the objects in Sect. 5.

The results of the foregoing section lead to the next

Theorem 15 *Under the above conditions on μ and ν ,*

$$D_{\text{KL}}(\mu \| \nu) = D_{\text{KL}}(\mu \| \nu_a) = \rho_{-\ln}(\mu, \nu) = \rho_{-\ln}(\mu, \nu_a) \quad (35)$$

$$= \sup_G \sum_{g \in G} \mu[g] \ln \left(\frac{\mu[g]}{\nu[g]} \right) = \sup_G \sum_{g \in G} \mu[g] \ln \left(\frac{\mu[g]}{\nu_a[g]} \right), \quad (36)$$

where ν_a is the absolutely continuous component of ν with respect to μ and the supremum is taken over all (finite or countable) measurable partitions of unity G on X , and we assume that if $\mu[g] = 0$, then the corresponding summand in the sums vanishes regardless of the second multiplier $\ln(\mu[g]/\nu[g])$ or $\ln(\mu[g]/\nu_a[g])$.

Proof According to (2), we have $-\ln'(+\infty) = 0$. Hence, by Theorem 10,

$$\rho_{-\ln}(\mu, \nu) = \rho_{-\ln}(\mu, \nu_a) = \int_X -\ln \left(\frac{d\nu_a}{d\mu} \right) d\mu. \quad (37)$$

It is easily seen that outside a set of zero measure μ ,

$$0 < \frac{d\mu}{dv_a} = \frac{d\mu}{dv} < +\infty.$$

Therefore,

$$\int_X \ln\left(\frac{d\mu}{dv}\right) d\mu = \int_X \ln\left(\frac{d\mu}{dv_a}\right) d\mu = \int_X -\ln\left(\frac{dv_a}{d\mu}\right) d\mu. \quad (38)$$

From (37), (38), we obtain equalities (35).

Recall that, μ is absolutely continuous with respect to ν and hence with respect to ν_a as well. So if $\mu[g] \neq 0$, then $\nu[g] \neq 0$ and $\nu_a[g] \neq 0$ for any element g of a measurable partition of unity on X . From this and definition (4) of $\rho_{-\ln}(\mu, \nu)$, it follows that

$$\rho_{-\ln}(\mu, \nu) = \sup_G \sum_{g \in G} \mu[g] \left(-\ln\left(\frac{\nu[g]}{\mu[g]}\right) \right) = \sup_G \sum_{g \in G} \mu[g] \ln\left(\frac{\mu[g]}{\nu[g]}\right),$$

where all summands with $\mu[g] = 0$ are supposed to be zero. The analogous equality for $\rho_{-\ln}(\mu, \nu_a)$ may be got in the same way. Thus, Theorem 15 is proved. $\square$

Remark 16 The theorem just proved along with formula (38) naturally suggests an extension of the definition of Kullback–Leibler divergence onto measures that are neither necessarily probability ones, nor mutually absolutely continuous. Namely, for any finite positive measures μ, ν on a measurable space $(X, \mathfrak{A})$ let us define the *generalized Kullback–Leibler divergence* $D_{\text{KL}}(\mu \| \nu)$ by the formula

$$D_{\text{KL}}(\mu \| \nu) := \int_X -\ln\left(\frac{dv_a}{d\mu}\right) d\mu. \quad (39)$$

The reasoning from the proof of Theorem 15 shows that $D_{\text{KL}}(\mu \| \nu)$ defined in this way satisfies equalities (35) and (36) as well.

Remark 17 If X is a topological space and μ and ν are Borel measures such that the set $C(X)$ of continuous functions is dense in $L^1(X, \mu)$ and $L^1(X, \nu)$ (which is always true for a metrizable space X or, as a variant, for regular measures μ, ν) then recalling Theorems 12 and 14 one concludes that when applying (35) and (36) to definition (39), we can equally use continuous (finite or countable) partitions of unity.

Remark 18 As is known apart from Kullback–Leibler divergence many common divergences are special cases of F -divergence, corresponding to a suitable choice of F . For example, Hellinger distance corresponds to the function $F(t) = 1 - \sqrt{t}$, total variation distance corresponds to $F(t) = |t - 1|$, Pearson χ^2 -divergence corresponds to $F(t) = (t - 1)^2$, and for the function $F(t) = (t^\alpha - t)/(\alpha^2 - \alpha)$, we obtain the so-called α -divergence.

Thus, by choosing the corresponding convex functions F , one can write out the ‘sup-sums principles’ of Theorem 15 type for them where again one can exploit not only measurable but also continuous partitions of unity. Moreover, for example, for total variation distance, Pearson χ^2 -divergence and α -divergence one naturally arrives at consideration of real-valued (not necessarily nonnegative) measures.

Remark 19 In the paper [22], the result of Theorem 15 type was established for a sigma-finite measure ν and a measure μ which is absolutely continuous with respect to ν .

5 New Definition for t -Entropy

In this section, as an application of Theorems 14 and 15, we obtain a new formula for t -entropy that clarifies its relationship with Kullback–Leibler divergence.

The t -entropy (we recall its definition below) is a principal object of spectral analysis of operators associated with dynamical systems. In particular, in the series of articles [3–8], a relation between t -entropy and spectral radii of the corresponding operators has been established. Namely, it was shown that t -entropy is the Fenchel–Legendre dual to the spectral exponent of operators in question.

For transparency of presentation, let us recall the mentioned objects and results.

Hereafter, X is a Hausdorff compact space, $C(X)$ is the algebra of continuous functions on X taking real values and equipped with the max-norm, and $\alpha : X \rightarrow X$ is an arbitrary continuous mapping. The corresponding dynamical system will be denoted by (X, α) .

Recall that, a *transfer operator* $A : C(X) \rightarrow C(X)$, associated with a given dynamical system, is defined in the following way:

- (a) A is a positive linear operator (i.e., it maps nonnegative functions to nonnegative ones); and
- (b) the following *homological identity* for A is valid:

$$A(g \circ \alpha \cdot f) = gAf, \quad g, f \in C(X).$$

As an important and popular example of transfer operators one can take say the classical Perron–Frobenius operator, that is, the operator having the form

$$Af(x) := \sum_{y \in \alpha^{-1}(x)} a(y)f(y),$$

where $a \in C(X)$ is fixed. This operator is well defined when α is a local homeomorphism.

Let A be a certain transfer operator in $C(X)$. In what follows, we denote by A_φ the family of transfer operators in $C(X)$ given by the formula

$$A_\varphi f := A(e^\varphi f), \quad \varphi \in C(X).$$

Let us denote by $\lambda(\varphi)$ the *spectral potential* of A_φ , defined by the formula

$$\lambda(\varphi) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|A_\varphi^n\| = \ln(r(A_\varphi)),$$

where $r(A_\varphi)$ is the spectral radius of operator A_φ .

We denote by $M(X)$ the set of all probability Borel measures on X . Recall that, a measure $\mu \in M(X)$ is called α -invariant iff $\mu(g) = \mu(g \circ \alpha)$ for all $g \in C(X)$. The family of α -invariant probability measures on X is denoted by $M_\alpha(X)$.

The t -entropy is a certain functional on $M(X)$ denoted by $\tau(\mu)$ (its detailed definition will be given below).

The substantial importance of t -entropy is demonstrated by the following variational principle.

Theorem 20 ([6], Theorem 5.6) *Let $A: C(X) \rightarrow C(X)$ be a transfer operator for a continuous mapping $\alpha: X \rightarrow X$ of a compact Hausdorff space X . Then,*

$$\lambda(\varphi) = \max_{\mu \in M_\alpha(X)} (\mu[\varphi] + \tau(\mu)), \quad \varphi \in C(X).$$

One vividly notes the resemblance of this theorem to the Ruelle–Walters variational principle for the topological pressure [21,25,26] uncovering its relation with Kolmogorov–Sinai entropy.

Among the principal ingredients in the proofs of the results leading to Theorem 20 is the so-called ‘entropy statistic theorem’. This theorem plays in the spectral theory of weighted shift and transfer operators the role analogous to Shannon–McMillan–Breiman theorem in information theory [1,18] and its important corollary known as ‘asymptotic equipartition property’ [10, p. 135]. The variational principles containing t -entropy and the objects therein serve as key ingredients of the thermodynamical formalism (see [4,7,17] and the sources quoted there).

Being so important, t -entropy at the same time is rather sophisticated object to calculate. The description of t -entropy not leaning on Fenchel–Legendre duality is not elementary, and it took a substantial time and effort to obtain its ‘accessible’ definition.

Namely, originally t -entropy $\tau(\mu)$ was defined in three steps (see, for example, [6]).

Definition 1 Firstly, for a given $\mu \in M(X)$, any $n \in \mathbb{N}$, and any continuous partition of unity $G = \{g_1, \dots, g_k\}$ we set

$$\tau_n(\mu, G) := \sup_{m \in M(X)} \sum_{g_i \in G} \mu[g_i] \ln \frac{m[A^n g_i]}{\mu[g_i]}. \quad (40)$$

Here, if $\mu[g_i] = 0$ for some $g_i \in G$ then the corresponding summand in (40) is assumed to be zero regardless the value $m[A^n g_i]$; if $m[A^n g_i] = 0$ for some $g_i \in G$ and at the same time $\mu[g_i] > 0$, then $\tau_n(\mu, G) = -\infty$.

Secondly, we put

$$\tau_n(\mu) := \inf_G \tau_n(\mu, G); \quad (41)$$

here the infimum is taken over all continuous partitions of unity G in $C(X)$.

And finally, the t -entropy $\tau(\mu)$ is defined as

$$\tau(\mu) := \inf_{n \in \mathbb{N}} \frac{\tau_n(\mu)}{n}.$$

Recently, in [9], the authors proved that for $\mu \in M_\alpha(X)$ (note that only such measures are essential for Theorem 20), the t -entropy could be defined in two steps.

Definition 2 First, we set

$$\tau_n(\mu) := \inf_G \sum_{g \in G} \mu[g] \ln \frac{\mu[A^n g]}{\mu[g]}, \quad (42)$$

where the infimum is taken over the set of all continuous partitions of unity G , and we assume that if $\mu[g] = 0$, then the corresponding summand in the right-hand part of the equality is equal to 0 independently of the value of $\mu[A^n g]$.

Now, $\tau(\mu)$ is defined as

$$\tau(\mu) := \inf_{n \in \mathbb{N}} \frac{\tau_n(\mu)}{n}.$$

In other words, in the original definition of t -entropy, one does not need to calculate the supremum in (40) but can simply put $m = \mu$ there. In [9], it was proved that this leads to the same value of $\tau_n(\mu)$ in (42) as in (41).

Of course, two steps are better than three, but even this two-steps definition of t -entropy is also rather sophisticated.

Note parenthetically that, if one identifies a Borel measure μ on X with a linear functional $\mu: C(X) \rightarrow \mathbb{R}$ given by

$$\mu[f] := \int_X f \, d\mu$$

then, by Riesz's theorem, there exists the only one regular Borel measure on X defining the same functional. Thus, since the definition of t -entropy leans only on integration of continuous functions (forming partitions of unity), it suffices to determine the t -entropy only for regular measures μ (which are the measures considered, for instance, in Theorem 14).

The next theorem in essence gives a new definition of t -entropy and transparently establishes its relation to Kullback–Leibler divergence.

Theorem 21 (*t*-entropy via Kullback–Leibler divergence) *Let A be a transfer operator for a dynamical system (X, α) then for any regular measure $\mu \in M_\alpha(X)$, we have*

$$\tau_n(\mu) = \int_X \ln \frac{d(A^{*n}\mu)_a}{d\mu} d\mu = -D_{\text{KL}}(\mu \| A^{*n}\mu)$$

and

$$\tau(\mu) = \inf_{n \in \mathbb{N}} \frac{1}{n} \int_X \ln \frac{d(A^{*n}\mu)_a}{d\mu} d\mu = - \sup_{n \in \mathbb{N}} \frac{D_{\text{KL}}(\mu \| A^{*n}\mu)}{n},$$

where $A^*: C(X)^* \rightarrow C(X)^*$ is the operator adjoint to A .

Proof Apply Theorem 15 and Remark 17 to (42). Namely, set $\nu = A^{*n}\mu$ in this equality (so that $\mu[A^n g] = \nu[g]$) and apply formulae (34)–(36). $\square$

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Algoet, P.H., Cover, T.M.: A sandwich proof of the Shannon–McMillan–Breiman theorem. *Ann. Probab.* **16**(2), 899–909 (1988)
2. Ali, S.M., Silvey, S.D.: A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc. Ser. B* **28**(1), 131–142 (1966)
3. Antonevich, A.B., Bakhtin, V.I., Lebedev, A.V.: Thermodynamics and spectral radius. *Nonlinear Phenom. Complex Syst.* **4**(4), 318–321 (2001)
4. Antonevich, A.B., Bakhtin, V.I., Lebedev, A.V., Sarzhinsky, D.S.: Legendre analysis, thermodynamic formalism and spectra of Perron–Frobenius operators. *Dokl. Math.* **67**(3), 343–345 (2003)
5. Antonevich, A.B., Bakhtin, V.I., Lebedev, A.V.: Spectra of operators associated with dynamical systems: from ergodicity to the duality principle. In: *Twenty Years of Białowieża: A Mathematical Anthology*. World Scientific Monograph Series in Mathematics, Chapter 7, vol. 8, pp. 129–161 (2005)
6. Antonevich, A.B., Bakhtin, V.I., Lebedev, A.V.: On *t*-entropy and variational principle for the spectral radii of transfer and weighted shift operators. *Ergod. Theory Dyn. Syst.* **31**, 995–1045 (2011)
7. Antonevich, A.B., Bakhtin, V.I., Lebedev, A.V.: A road to the spectral radius of transfer operators. *Contemp. Math.* **567**, 17–51 (2012)
8. Bakhtin, V.I.: On *t*-entropy and variational principle for the spectral radius of weighted shift operators. *Ergod. Theory Dyn. Syst.* **30**, 1331–1342 (2010)
9. Bakhtin, V.I., Lebedev, A.V.: A new definition of *t*-entropy for transfer operators. *Entropy* **19**(573), 1–6 (2017)
10. Billingsley, P.: *Ergodic Theory and Information*. Wiley, New York (1965)
11. Broniatowski, M., Keziou, A.: Minimization of divergences on sets of signed measures. *Stud. Sci. Math. Hung.* **43**(4), 403–442 (2006)
12. Broniatowski, M., Keziou, A.: Parametric estimation and tests through divergences and duality technique. *J. Multivar. Anal.* **100**(1), 16–36 (2009)

13. Broniatowski, M.: A weighted bootstrap procedure for divergence minimization problems. In: Analytical Methods in Statistics. Springer Proceedings in Mathematics & Statistics, vol. 193, pp. 1–22 (2015)
14. Csiszár, I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Magyar. Tud. Akad. Mat. Kutató Int. Közl. **8**, 85–108 (1963)
15. Gelfand, I.M., Kolmogorov, A.N., Yaglom, A.M.: On the general definition of the amount of information. Dokl. Akad. Nauk SSSR **11**, 745–748 (1956)
16. Liese, F., Vajda, I.: On divergences and informations in statistics and information theory. IEEE Trans. Inf. Theory **52**(10), 4394–4412 (2006)
17. Lopes, A., Mengue, J., Mohr, J., Souza, R.: Entropy and variational principle for one-dimensional lattice systems with a general a priori probability: positive and zero temperature. Ergod. Theory Dyn. Syst. **35**, 1925–1961 (2015)
18. McMillan, B.: The basic theorems of information theory. Ann. Math. Stat. **24**, 196–219 (1953)
19. Morimoto, T.: Markov processes and the H-theorem. J. Phys. Soc. Jpn. **18**(3), 328–331 (1963)
20. Najim, J.: A Cramér type theorem for weighted random variables. Electron. J. Probab. **7**(4), 1–32 (2002)
21. Ruelle, D.: Statistical mechanics on a compact set with Z^V action satisfying expansiveness and specification. Trans. Am. Math. Soc. **185**, 237–252 (1973)
22. Sokol, E.E.: Introduction of the Kullback–Leibler information function by means of partitions of the probability space. J. Belarus. State Univ. Math. Inform. **1**, 59–67 (2018)
23. Thashorras, J., Wintenberger, O.: Large deviations for bootstrapped empirical measures. Bernoulli **20**(4), 1845–1878 (2014)
24. Vajda, I.: On the f -divergence and singularity of probability measures. Period. Math. Hung. **1**, 223–234 (1972)
25. Walters, P.: A variational principle for the pressure of continuous transformations. Am. J. Math. **97**(4), 937–971 (1975)
26. Walters, P.: An Introduction to Ergodic Theory. Springer, Berlin (1982)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.