KYBERNETIKA — VOLUME *61* (2025), NUMBER 4, PAGES 481–491

METRIZATION OF POWERS OF THE JENSEN–SHANNON DIVERGENCE

Kazuki Okamura

Metrization of statistical divergences is valuable in both theoretical and practical aspects. One approach to obtaining metrics associated with divergences is to consider their fractional powers. Motivated by this idea, Osán, Bussandri, and Lamberti (2018) studied the metrization of fractional powers of the Jensen–Shannon divergence between multinomial distributions and posed an open problem. In this short note, we provide an affirmative answer to their conjecture. Moreover, our method is also applicable to fractional powers of f-divergences between Cauchy distributions.

Keywords: Jensen–Shannon divergence, metric, multinomial distribution, Cauchy distribution

Classification: 94A17, 60A10, 60E05, 28A33

1. INTRODUCTION

Dissimilarity between probability distributions is a fundamental topic in probability, statistics, and related fields such as machine learning, and has been extensively studied ([18]). Statistical divergences serve as canonical measures of dissimilarity. One of the most widely used divergences is the Kullback-Leibler divergence (KLD), also known as relative entropy. It has numerous theoretical and practical applications. In particular, it naturally appears as the rate function in Sanov's theorem in large deviation theory, describing the exponential decay rate of rare events. In the framework of information geometry, the KLD generalizes the squared Euclidean distance, and for exponential families, it satisfies a Pythagorean theorem. However, the square root of the KLD is generally not a metric: it can be asymmetric and violate the triangle inequality. Another commonly used divergence is the total variation distance (TVD). Unlike the KLD, the TVD is a bounded metric. However, the TVD between two distributions which are singular to each other always equals 2. Furthermore, closed-form expressions are often difficult to obtain, and one typically must rely on numerical approximations.

The Jensen–Shannon divergence (JSD), defined via the KLD, is also referred to as the information radius or total divergence from the average. The JSD is always well-defined, symmetric, and bounded ([11]). It has found applications across numerous research disciplines and admits both statistical and information-theoretic interpretations. In

DOI: 10.14736/kyb-2025-4-0481

statistical inference, the JSD provides both lower and upper bounds on the Bayes error probability, while in information theory, it is related to mutual information ([8]). Various generalizations and related notions of the JSD have been proposed ([13, 16]).

From a theoretical standpoint, metric spaces are one of the most foundational frameworks in mathematics. Metrizing divergences is also significant in practical applications, especially in the design of efficient algorithms in computational geometry ([3]). For instance, the triangle inequality can accelerate proximity queries ([23]) and k-means clustering ([5]). In general, symmetric divergences are not metrics, so it is natural to consider fractional powers (moments) of these divergences to obtain associated metric structures. Sufficient conditions for fractional powers of Csiszár's f-divergences to form metrics are given in [9, 17]. It is well-known that the square root of the JSD satisfies the triangle inequality ([1, 6, 21]). This, along with the TVD, constitutes a canonical statistical metric distance.

Osán, Bussandri, and Lamberti [15] considered the JSD as a special case of a Csiszár divergence and provided a sufficient condition for the power of the JSD between multinomial distributions to define a metric. In [15, Conjecture 1], they conjecture that the pth power of the JSD between multinomial distributions is *not* a metric for $p > 1/2$. The square root of the JSD admits an isometric embedding into Hilbert spaces ([7]). Since in general pth powers of distances on Hilbert spaces are *not* distances when $p > 1$, this indirectly supports the conjecture. However, the embedding is far from being surjective, so this fact cannot be directly used in the proof. To the best of our knowledge, this problem remains open.

One aim of this paper is to prove [15, Conjecture 1]. Our approach is elementary and self-contained, differing significantly from [15]. While our method is somewhat similar to the proof of [14, Theorem 28], we do not utilize the metric transformation introduced there. Furthermore, we present an alternative proof of [15, Proposition 1], which is considerably simpler than the original.

Our elementary method also applies to the Cauchy distribution, a canonical example of a heavy-tailed distribution. For Cauchy distributions, f-divergences are always symmetric ([14, 22]), which motivates the question of whether powers of f-divergences form metrics for general convex functions f. We prove that the pth power of the f-divergence between Cauchy distributions fails to be a metric for $p > 1/2$, for a broad class of differentiable convex functions f on $(0, \infty)$, including those corresponding to the KLD and the JSD, but excluding the TVD. Our proof relies on an expression of f-divergences given by Verdú [22].

We include a short note in Appendix A on the fact that the square root of the JSD satisfies the triangle inequality. This implies that the JSD defines a regular semi-metric, meaning that its local properties are similar to those of a metric. See [2] for further details.

2. FRAMEWORK

Let X be a set with a sigma-algebra and μ be a positive measure on X. For the discrete and continuous distributions, μ is usually taken as the counting measure and the Lebesgue measure, respectively. Let P and Q be two probability measures on X with density functions p and q with respect to μ, respectively. The Kullback-Leibler

divergence between P and Q is defined by

$$D_{KL}(P:Q) := \int_X \log\left(\frac{p(x)}{q(x)}\right) p(x)\mu(\mathrm{d}x).$$

The Jensen–Shannon divergence between P and Q is defined by

$$D_{JS}(P:Q) := \frac{1}{2}\left(D_{KL}\left(P:\frac{P+Q}{2}\right) + D_{KL}\left(Q:\frac{P+Q}{2}\right)\right).$$

There are generalizations of this divergence. For example, [13] replaces $(P+Q)/2$ with a quasi-arithmetic mean.

We can define them by using the Radon-Nikodym derivative. The Kullback-Leibler divergence is asymmetric in general, but the Jensen–Shannon divergence is always symmetric. We also remark that P and Q are both absolutely continuous with respect to $(P+Q)/2$, so the Jensen–Shannon divergence is always defined. If P is not absolutely continuous with respect to Q, then, $D_{KL}(P:Q) = +\infty$. These are canonical examples of f-divergences.

We let the entropy be

$$H(P) := \int_X -p(x)\log p(x)\mu(\mathrm{d}x).$$

Then,

$$D_{JS}(P:Q) = H\left(\frac{P+Q}{2}\right) - \frac{H(P)+H(Q)}{2}.$$

We now recall the definition of a metric. Let S be a non-empty set. We call a function $d : S \times S \to [0,\infty)$ a distance function if it satisfies the following three conditions:

1. $d(x,y) = 0$ if and only if $x = y$.
2. (symmetry) $d(x,y) = d(y,x)$ for $x, y \in S$.
3. (triangle inequality) $d(x,z) \leq d(x,y) + d(y,z)$ for $x, y, z \in S$.

For such d, we call a pair (S,d) a metric space. This is a fundamental notion in geometry.

3. METRIZATION OF JENSEN–SHANNON DIVERGENCES BETWEEN THE MULTINOMIAL DISTRIBUTIONS

Throughout this section, we set $\log = \log_2$, so that $0 \leq D_{JS}(P:Q) \leq 1$, and, $D_{JS}(P:Q)$ is smaller than the total variation distance. The natural-log version differs only by a constant factor of $\ln 2$.

We let $X = \{1, 2, \cdots, n\}$ and μ be the counting measure on X. For $n \geq 2$, let $\mathcal{P}_n := \{(p_i)_i : \sum_i p_i = 1, p_i > 0\}$ and $\overline{\mathcal{P}_n} := \{(p_i)_i : \sum_i p_i = 1, p_i \geq 0\}$. For $P = (p_i)_{i=1}^n$ and $Q = (q_i)_{i=1}^n$ in $\mathcal{P}_n$,

$$D_{KL}(P:Q) = \sum_{i=1}^n p_i \log_2\left(\frac{p_i}{q_i}\right),$$

and

$$D_{JS}(P:Q)=\sum_{i=1}^{n}-\frac{p_i}{2}\log_2\left(\frac{p_i+q_i}{2p_i}\right)-\frac{q_i}{2}\log_2\left(\frac{p_i+q_i}{2q_i}\right).$$

Our main result is:

Theorem 3.1. Let $\alpha>1/2$. Then, $D_{JS}(P:Q)^\alpha$ is not a metric on $\mathcal{P}_n$.

Proof. We first deal with the case that $n=2$. Let $P_t:=(t,1-t)$, $0\le t\le 1$. For $t\in[0,1/2)$, let $f(t):=D_{JS}(P_{1/2-t}:P_{1/2+t})$ and $g(t):=D_{JS}(P_{1/2-t}:P_{1/2})=D_{JS}(P_{1/2}:P_{1/2+t})$. Let $F(t):=f(t)^\alpha-2g(t)^\alpha$. It suffices to show that $F(t)>0$ for some t. Since $F(0)=0$, by the mean-value theorem, it suffices to show that $F'(t)>0$ for every t sufficiently close to 0. Since $F'(t)=\alpha(f'(t)f(t)^{\alpha-1}-2g'(t)g(t)^{\alpha-1})$, it suffices to show that

$$\left(\frac{g(t)}{f(t)}\right)^{1-\alpha}>2\frac{g'(t)}{f'(t)}, \tag{1}$$

for every t sufficiently close to 0.

We see that $f(t)=1-H(P_{1/2+t})$ and $g(t)=H(P_{(1+t)/2})-\dfrac{H(P_{1/2+t})+1}{2}$. Hence,

$$f'(t)=-\frac{d}{dt}H(P_{1/2+t}) \text{ and } g'(t)=\frac{d}{dt}H(P_{(1+t)/2})-\frac{1}{2}\frac{d}{dt}H(P_{1/2+t}).$$

Since $H(P_s)=-s\log_2 s-(1-s)\log_2(1-s)$, $0\le s\le 1$, we see that $\dfrac{d}{dt}H(P_{(1+t)/2})=-\dfrac{1}{2}\log\left(\dfrac{1+t}{1-t}\right)$ and $\dfrac{d}{dt}H(P_{1/2+t})=-\log\left(\dfrac{1+2t}{1-2t}\right)$. Hence,

$$2\frac{g'(t)}{f'(t)}=1-2\frac{\frac{d}{dt}H(P_{(1+t)/2})}{\frac{d}{dt}H(P_{1/2+t})}=1-\frac{\log\frac{1+t}{1-t}}{\log\frac{1+2t}{1-2t}}.$$

Since $\lim\limits_{s\to 0}\dfrac{\log\left(\frac{1+s}{1-s}\right)}{2s}=1$, we see that $\lim\limits_{t\to 0}2\dfrac{g'(t)}{f'(t)}=\dfrac{1}{2}$.

We recall that $f(0)=g(0)=0$. Then, by l'Hospital's theorem,

$$\lim_{t\to 0}\frac{g(t)}{f(t)}=\lim_{t\to 0}\frac{g'(t)}{f'(t)}=\frac{1}{4}.$$

Hence, $\lim\limits_{t\to 0}\left(\dfrac{g(t)}{f(t)}\right)^{1-\alpha}=\left(\dfrac{1}{4}\right)^{1-\alpha}$. Since $\alpha>1/2$, $\left(\dfrac{1}{4}\right)^{1-\alpha}>\dfrac{1}{2}$. Thus we have Eq. (1). The proof of Theorem 1 is completed for $n=2$.

We now deal with the case of $n\ge 3$. We can naturally embed $\overline{\mathcal{P}_2}$ into $\overline{\mathcal{P}_n}$ by a map $(p_1,p_2)\mapsto(p_1,p_2,0,\cdots,0)$. Since $P\mapsto H(P)$ is continuous with respect to P on $\overline{\mathcal{P}_n}$, we can find P_1,P_2 and P_3 in $\mathcal{P}_n$ such that $D_{JS}(P_1:P_3)^\alpha>D_{JS}(P_1:P_2)^\alpha+D_{JS}(P_2:P_3)^\alpha$. The proof of Theorem 3.1 is completed for $n\ge 3$. □

Remark 3.2. In general, $x^\beta + y^\beta \le (x+y)^\beta$ for $x, y \ge 0$ and $\beta \ge 1$, and, if $x^\beta + y^\beta = (x+y)^\beta$, then, $\beta = 1$ or $xy = 0$. Hence, if a function $d : S \times S \to [0, \infty)$ is *not* a metric on a set S, then, $d^\beta(x, y)$ is *not* a metric S. Since it is known that $D_{JS}(P : Q)^{1/2}$ is a metric, this gives an alternative proof of [15, Proposition 1]. which is much easier than the proof given in it.

4. METRIZATION OF F-DIVERGENCES BETWEEN THE CAUCHY DISTRIBUTIONS

For $\mu \in \mathbb{R}$ and $\sigma > 0$, the density function of the univariate Cauchy distribution is given by $p_{\mu,\sigma}(x) := \dfrac{\sigma}{\pi} \dfrac{1}{(x-\mu)^2 + \sigma^2}$, $x \in \mathbb{R}$. For a continuous function f on $(0, \infty)$, the f-divergence is defined by

$$D_f(p_{\mu_1,\sigma_1} : p_{\mu_2,\sigma_2}) := \int_{\mathbb{R}} f\left(\frac{p_{\mu_2,\sigma_2}(x)}{p_{\mu_1,\sigma_1}(x)}\right) p_{\mu_1,\sigma_1}(x)\,\mathrm{d}x.$$

The following result is crucial in our proof.

Theorem 4.1. (Verdú [22], Eq. (189) in Theorem 10) Let f be a continuous function on $(0, \infty)$. Then,

$$D_f(p_{\mu_1,\sigma_1} : p_{\mu_2,\sigma_2}) = \int_0^\pi f\left(\frac{1}{\zeta + \sqrt{\zeta^2 - 1}\cos\theta}\right) \frac{\mathrm{d}\theta}{\pi},$$

where $\zeta := 1 + \dfrac{(\mu_2 - \mu_1)^2 + (\sigma_2 - \sigma_1)^2}{2\sigma_1\sigma_2}$.

In particular, every f-divergence is a function of ζ. This quantity is also known as maximal invariant with respect to an action of the special linear group $SL(2, \mathbb{R})$ to the upper-half plane $\mathbb{H} := \{\mu + \sigma i : \mu \in \mathbb{R}, \sigma > 0\}$ with complex parameter, considered by McCullagh [12]. For example, we obtain the JSD if we let

$$f(u) = f_{JS}(u) := \frac{1}{2}\left(u \log \frac{2u}{1+u} - \log \frac{1+u}{2}\right).$$

Theorem 4.2. Let f be a convex function on $(0, \infty)$ such that $f(1) = 0$, f is in C^2 class on an open neighborhood of 1, and $f''(1) > 0$. Let $\alpha > 1/2$. Then, $D_f(p_{0,\sigma_1} : p_{0,\sigma_2})^\alpha$ is not a metric on $(0, \infty)$.

This result is applicable to a large class of f-divergences including the KLD and the JSD. However, the regularity assumption for f is crucial. Obviously, the conclusion fails for the TVD, which is obtained by $f(u) = f_{TV}(u) := |u - 1|/2$.

P r o o f . We will show that

$$D_f(p_{0,\sigma_1} : p_{0,\sigma_2})^\alpha + D_f(p_{0,\sigma_2} : p_{0,\sigma_3})^\alpha < D_f(p_{0,\sigma_1} : p_{0,\sigma_3})^\alpha$$

where $(\sigma_1, \sigma_2, \sigma_3) = (e^{-t}, 1, e^t)$ for sufficiently small $t > 0$. For $t > 0$, let

$$h(t) := \int_0^\pi f\left(\frac{1}{\cosh(t) + \sinh(t)\cos\theta}\right)\frac{\mathrm{d}\theta}{\pi}.$$

Then, by Theorem 4.1, $h(t) = D_f(p_{0,\sigma_1} : p_{0,\sigma_2}) = D_f(p_{0,\sigma_2} : p_{0,\sigma_3})$ and $h(2t) = D_f(p_{0,\sigma_1} : p_{0,\sigma_3})$. Hence, it suffices to show that $2h(t)^\alpha < h(2t)^\alpha$ for some $t > 0$. We remark that

$$\lim_{t\to+0} \cosh(t) + \sinh(t)\cos\theta = 1 \tag{2}$$

and

$$\lim_{t\to+0} \sinh(t) + \cosh(t)\cos\theta = \cos\theta \in [-1, 1]. \tag{3}$$

Under the assumption of f, we can exchange the derivative wit respect to t and the integral wit respect to θ, so we obtain that there exists a sufficiently small $\delta_0 > 0$ such that for every $0 < t < \delta_0$,

$$h'(t) = \int_0^\pi -\frac{\sinh(t) + \cosh(t)\cos\theta}{(\cosh(t) + \sinh(t)\cos\theta)^2} f'\left(\frac{1}{\cosh(t) + \sinh(t)\cos\theta}\right)\frac{\mathrm{d}\theta}{\pi},$$

and,

$$h''(t) = \int_0^\pi \frac{(\sinh(t) + \cosh(t)\cos\theta)^2}{(\cosh(t) + \sinh(t)\cos\theta)^4} f''\left(\frac{1}{\cosh(t) + \sinh(t)\cos\theta}\right)\frac{\mathrm{d}\theta}{\pi}$$

$$+ \int_0^\pi \frac{2(\sinh(t) + \cosh(t)\cos\theta)^2 - (\cosh(t) + \sinh(t)\cos\theta)^2}{(\cosh(t) + \sinh(t)\cos\theta)^3} f'\left(\frac{1}{\cosh(t) + \sinh(t)\cos\theta}\right)\frac{\mathrm{d}\theta}{\pi}.$$

We recall that $\displaystyle\int_0^\pi \cos\theta\,\mathrm{d}\theta = \int_0^\pi \cos(2\theta)\,\mathrm{d}\theta = 0$ and $\displaystyle\int_0^\pi \cos^2\theta\,\mathrm{d}\theta = \frac{\pi}{2}$. By this, (2), and (3), we see that

$$\lim_{t\to+0} h(t) = \lim_{t\to+0} h'(t) = 0$$

and

$$\lim_{t\to+0} h''(t) = \frac{f''(1)}{2} > 0.$$

By l'Hospital's theorem,

$$\lim_{t\to+0} \frac{h(2t)}{h(t)} = \lim_{t\to+0} \frac{2h'(2t)}{h'(t)} = \lim_{t\to+0} \frac{4h''(2t)}{h''(t)} = 4.$$

Since $\alpha > 1/2$, we see that $2h(t)^\alpha < h(2t)^\alpha$ for sufficiently small $t > 0$. This completes the proof. □

Remark 4.3. (i) In the case of the TVD, $\lim_{t\to+0} h'(t) = \dfrac{1}{\pi} > 0$, and hence, by l'Hospital's theorem, we have that $\lim_{t\to+0} \dfrac{h(2t)}{h(t)} = 2$.
(ii) In [22, Theorem 10], it is assumed that f is convex and right-continuous at 0. However, for every (μ_1, σ_1) and (μ_2, σ_2),

$$0 < \inf_{x\in\mathbb{R}} \frac{p_{\mu_2,\sigma_2}(x)}{p_{\mu_1,\sigma_1}(x)} \le \sup_{x\in\mathbb{R}} \frac{p_{\mu_2,\sigma_2}(x)}{p_{\mu_1,\sigma_1}(x)} < +\infty,$$

so we do not need to assume that f is defined at 0. This property does not hold for normal distributions.

A. ON THE SQUARE ROOT OF JENSEN–SHANNON DIVERGENCE

Fuglede and Topsøe [7] stated that the square root of the JSD is a metric on the space of probability measures over a given measure space. Acharyya, Banerjee, and Boley [1] provided a proof of this result. However, some parts of the arguments of [1, 7] are sketchy, and we offer more details here. While we follow the overall strategy used in [1, 7], we believe that several components of our approach are more elementary, transparent, and simpler than those in [1]. Our arguments make use of the *Lambert W function*.

Let P, Q be two probability measures on a measurable space X. Let $M := (P+Q)/2$. Let the Jensen–Shannon divergence between P and Q be

$$D_{JS}(P:Q) := \frac{1}{2}\left(\int_X \log\frac{\mathrm{d}P}{\mathrm{d}M}\,\mathrm{d}P + \int_X \log\frac{\mathrm{d}Q}{\mathrm{d}M}\,\mathrm{d}Q\right).$$

Let $\phi(z) := z\log z, z \ge 0$. Then, this is convex. Let

$$\psi(x,y) := \sqrt{\frac{\phi(x)+\phi(y)}{2} - \phi\left(\frac{x+y}{2}\right)},\ x, y \ge 0.$$

Let λ be a probability measure on X such that $P \ll \lambda$ and $Q \ll \lambda$. For ease of notation, we let $f := dP/d\lambda$ and $g := dQ/d\lambda$. Then,

$$D_{JS}(P:Q) = \int_X \psi(f,g)\,\mathrm{d}\lambda.$$

Let P, Q, R be three probability measures on a measure space X. Let λ be a probability measure on X such that $P \ll \lambda$, $Q \ll \lambda$ and $R \ll \lambda$. Let $f := dP/d\lambda$, $g := dQ/d\lambda$ and $h = dR/d\lambda$. By the Minkowski inequality, in order to show that

$$\sqrt{D_{JS}(P:R)} \le \sqrt{D_{JS}(P:Q)} + \sqrt{D_{JS}(Q:R)},$$

which is equivalent to

$$\int_X \psi(f,g)\,\mathrm{d}\lambda \le \int_X \psi(f,g)\,\mathrm{d}\lambda + \int_X \psi(f,h)\,\mathrm{d}\lambda,$$

it suffices to show that:

Proposition A.1.

$$\sqrt{\psi(x,z)} \le \sqrt{\psi(x,y)} + \sqrt{\psi(y,z)},\ x,y,z \ge 0.$$

By Schoenberg's theorem [20], in order to show Proposition A.1, it suffices to show that:

Proposition A.2. (Acharyya et al. [1], Lemma 4) If $k(x,y) := \phi(x+y) = (x+y)\log(x+y)$, then, $(x,y) \mapsto \exp(\beta k(x,y))$ is a positive-definite kernel for every $\beta > 0$.

Let $W(x)$ be the inverse function of a C^∞ function $z \mapsto z\exp(z)$ on $(-1,\infty)$. This is called the Lambert W function, and $W \in C^\infty((-1/e,\infty))$. By [10], $W(\cdot)$ is a Bernstein function. Since a map $x \mapsto 1/(1+x)$ is a completely monotone function, by [19, Theorem 3.7 (ii)], a map $x \mapsto 1/(1+W(x))$ is also a completely monotone function.

Hence, by Bernstein's theorem (cf. [19, Theorem 1.4]), there exists a unique probability measure μ on $(0,\infty)$ such that

$$\int_0^\infty \exp(-tx)\mu(\mathrm{d}x) = \frac{1}{1+W(t)},\ t > 0.$$

Let $0 < s < 1$. Then, by a disintegration formula (cf. [24], p. 63),

$$\int_0^\infty x^s \mu(\mathrm{d}x) = \frac{s}{\Gamma(1-s)} \int_0^\infty t^{-s-1} \left(1 - \int_0^\infty \exp(-tx)\mu(\mathrm{d}x)\right) \mathrm{d}t.$$

We see that

$$\int \left(1 - \int_0^\infty \exp(-tx)\mu(\mathrm{d}x)\right) \mathrm{d}t = \int t^{-s-1} \frac{W(t)}{1+W(t)}\,\mathrm{d}t = -s^{s-1}\Gamma(1-s, sW(t)) + C,$$

where $\Gamma(,)$ is the incomplete Gamma function and C is the integral constant.

Since $\lim_{t\to+0} W(t) = 0$ and $\lim_{t\to+\infty} W(t) = +\infty$,

$$\int_0^\infty x^s \mu(\mathrm{d}x) = s^s = \exp(\phi(s)),\ 0 < s < 1.$$

We remark that for every $n \ge 1$ and $t > 0$,

$$\sup_{x>0} x^n \exp(-tx) < +\infty.$$

Hence, for each $n \ge 1$,

$$\frac{\partial^n}{\partial t^n} \int_0^\infty \exp(-tx)\mu(\mathrm{d}x) = \int_0^\infty (-x)^n \exp(-tx)\mu(\mathrm{d}x),\ t > 0.$$

By the monotone convergence theorem, for each $n \ge 1$,

$$\int_0^\infty x^{2n}\mu(\mathrm{d}x) = \lim_{t\to+0} \frac{\partial^{2n}}{\partial t^{2n}} \int_0^\infty \exp(-tx)\mu(\mathrm{d}x) = \lim_{t\to+0} \frac{\partial^{2n}}{\partial t^{2n}} \frac{1}{1+W(t)}.$$

This limit is finite since $W \in C^\infty((-1/e, \infty))$. Hence,

$$\int_0^\infty x^s \mu(\mathrm{d}x) < +\infty,\ s > 0.$$

Hence,

$$F(z) := \int_0^\infty x^z \mu(\mathrm{d}x),\ z \in \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\},$$

is well-defined and holomorphic. By the identity theorem for holomorphic functions,

$$\int_0^\infty x^s \mu(\mathrm{d}x) = s^s = \exp(\phi(s)),\ s > 0.$$

Let $\nu := \mu \circ (\log)^{-1} = \mu \circ \exp$ be a probability measure. Then,

$$\int_{-\infty}^\infty \exp(sy)\nu(\mathrm{d}y) = s^s = \exp(\phi(s)),\ s > 0.$$

Let $\beta > 0$. Then,

$$\int_{-\infty}^\infty \exp(s(\beta y - \log\beta))\nu(\mathrm{d}y) = s^s = \exp(\beta\phi(s)),\ s > 0.$$

Let $b_1, \cdots, b_n > 0$ and $c_1, \cdots, c_n \in \mathbb{R}$. Then,

$$\sum_{i,j=1}^n c_i c_j \exp(\beta\phi(b_i + b_j)) = \int_{-\infty}^\infty \left(\sum_{i=1}^n c_i \exp(b_i(\beta y - \log\beta))\right)^2 \nu(\mathrm{d}y) \geq 0.$$

This completes the proof of Proposition A.2.

Remark A.3. We see that

$$\int_{-\infty}^\infty \exp(ity)\nu(\mathrm{d}y) = \exp\left(-\frac{\pi}{2}|t| + it\log|t|\right),\ t \in \mathbb{R}. \tag{4}$$

Hence, ν is an asymmetric stable distribution with $\alpha = 1$. The function $\phi(t) := \int_{-\infty}^\infty \exp(ity)\nu(\mathrm{d}y)$ is not an entire function, so we cannot apply [4, Theorem 1].

The arguments in the proof of [1, Lemma 4] implicitly assumes that if (4) holds, then, $\int_{-\infty}^\infty \exp(sy)\nu(\mathrm{d}y) < +\infty$ for every $s > 0$. However, the proof is not written in it. One easy way to resolve this is to use an integral expression of the density function g of ν given in [24, Theorem 2.2.3] as follows:

$$g(x) = \frac{1}{2}\int_{-1}^1 U(t)\exp(x - \exp(x)U(t))\,\mathrm{d}t,\ x \in \mathbb{R},$$

where

$$U(t) := \frac{\pi}{2}\frac{1-t}{\cos(\pi t/2)}\exp\left(-\frac{\pi}{2}(1-t)\tan(\frac{\pi}{2}t)\right).$$

We see that

$$\lim_{t\to -1+0} U(t) = \frac{1}{e}, \quad \lim_{t\to 1-0} U(t) = +\infty,$$

and,

$$\int_{-1}^{1} U(t)\exp(-\exp(x)U(t))\,\mathrm{d}t \le \exp(-(n+2)x)\int_{-1}^{1}\frac{(n+2)!}{U(t)^{n+1}}\,\mathrm{d}t,\ n\ge 1.$$

Then, we see that for every $n \ge 1$,

$$\int_{\mathbb{R}} \exp(nx)g(x)\,\mathrm{d}x < +\infty.$$

Now we can use the identity theorem as above, and obtain that

$$\int_{\mathbb{R}} \exp(sx)g(x)\,\mathrm{d}x = \exp(\phi(s)),\ s>0.$$

ACKNOWLEDGEMENT

The author wishes to give his thanks to the referee for his or her comments, and to Prof. Frank Nielsen for notifying me of the conjecture by Osán, Bussandri, and Lamberti. The author was supported by JSPS KAKENHI 22K13928.

(Received April 2, 2025)

REFERENCES

[1] S. Acharyya, A. Banerjee, and D. Boley: Bregman divergences and triangle inequality. In: Proc. 2013 SIAM International Conference on Data Mining, SIAM 2013, pp. 476–484. SIAM.

[2] K. Chrząszcz, J. Jachymski, and F. Turoboś: On characterizations and topology of regular semimetric spaces. Publ. Math. Debr. *93* (2018), 87–105. DOI:10.5486/PMD.2018.8049

[3] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf: Computational Geometry. Algorithms and Applications. (Second rev. edition.) Springer, Berlin 2000.

[4] W. Ehm, M. G. Genton, and T. Gneiting: Stationary covariances associated with exponentially convex functions. Bernoulli *9* (2004), 607–615. DOI:10.3150/bj/1066223271

[5] C. Elkan: Using the triangle inequality to accelerate k-means. In: Proc. 20th International Conference on Machine Learning (ICML-03), 2003, pp. 147–153.

[6] D. M. Endres and J. E. Schindelin: A new metric for probability distributions. IEEE Trans. Inform. Theory *49* (2003), 1858–1860. DOI:10.1109/TIT.2003.813506

[7] B. Fuglede and F. Topsøe: Jensen–Shannon divergence and Hilbert space embedding. In: Proc. International Symposium on Information Theory, 2004. ISIT, IEEE 2004, p. 31.

[8] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H. E. Stanley: Analysis of symbolic sequences using the Jensen–Shannon divergence. Phys. Rev. E (3), *65* (2002). DOI:10.1103/physreve.65.041905

[9] P. Kafka, F. Österreicher, and I. Vincze: On powers of f-divergences defining a distance. Stud. Sci. Math. Hung. *26* (1991), 415–422. DOI:10.3109/10826089109058894

[10] G. A. Kalugin and D. J. Jeffrey: Unimodal sequences show that Lambert W is Bernstein. C. R. Math. Acad. Sci. Soc. R. Can. *33* (2011), 50–56.

[11] J. Lin: Divergence measures based on the Shannon entropy. IEEE Trans. Inform. Theory *37* (1991), 145–151. DOI:10.1109/18.61115

[12] P. McCullagh: Möbius transformation and Cauchy parameter estimation. Ann. Statist. *24* (1996), 787–808.

[13] F. Nielsen: On the Jensen-Shannon symmetrization of distances relying on abstract means. Entropy *21* (2019), 485. DOI:10.3390/e21050485

[14] F. Nielsen and K. Okamura: On f-divergences between Cauchy distributions. IEEE Trans. Inform. Theory *69* (2023), 3150–3171. DOI:10.1109/TIT.2022.3231645

[15] T. M. Osán, D. G. Bussandri, and P. W. Lamberti: Monoparametric family of metrics derived from classical Jensen–Shannon divergence. Physica A, *495* (2018), 336–344. DOI:10.1016/j.physa.2017.12.073

[16] T. M. Osán, D. G. Bussandri, and P. W. Lamberti: Quantum metrics based upon classical Jensen–Shannon divergence. Physica A *594* (2022). DOI:10.1016/j.physa.2022.127001

[17] F. Österreicher and I. Vajda: A new class of metric divergences on probability spaces and its applicability in statistics. Ann. Inst. Stat. Math. *55* (2003), 639–653. DOI:10.1007/BF02517812

[18] S. T. Rachev, L. B. Klebanov, S. V. Stoyanov, and F. J. Fabozzi: The Methods of Distances in the Theory of Probability and Statistics. Springer, New York 2013.

[19] R. L. Schilling, R. Song, and Z. Vondraček: Bernstein functions, volume 37 of De Gruyter Studies in Mathematics. Walter de Gruyter Co., Berlin 2010.

[20] I. J. Schoenberg: Metric spaces and positive definite functions. Trans. Amer. Math. Soc. *44* (1938), 522–536. DOI:10.1090/S0002-9947-1938-1501980-0

[21] I. Vajda: On metric divergences of probability measures. Kybernetika *45* (2009), 885–900. DOI:10.1145/1932682.1869533

[22] S. Verdú: The Cauchy distribution in information theory. Entropy *25* (2023), 346. DOI:10.3390/e25020346

[23] P. N. Yianilos: Data structures and algorithms for nearest neighbor. In: Proce. Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM *66* (1993), p. 311.

[24] V. M. Zolotarev: One-Dimensional Stable Distributions. Americal Mathematical Society, 1986.

Kazuki Okamura, Department of Mathematics, Faculty of Science, Shizuoka University, 836, Ohya, Suruga-ward, Shizuoka, 422-8529. Japan.
e-mail: okamura.kazuki@shizuoka.ac.jp