PHYSICAL REVIEW E **69**, 026128 (2004)

Observability of Rényi's entropy

Petr Jizba* and Toshihico Arimitsu†
Institute of Physics, University of Tsukuba, Ibaraki 305-8571, Japan
(Received 29 July 2003; published 27 February 2004)

Despite recent claims we argue that Rényi's entropy is an observable quantity. It is shown that, contrary to popular belief, the reported domain of instability for Rényi entropies has zero measure (Bhattacharyya measure). In addition, we show that the instabilities can be easily emended by introducing a coarse graining into an actual measurement. We also clear up any doubts regarding the observability of Rényi's entropy in (multi)fractal systems and in systems with absolutely continuous probability density functions.

DOI: 10.1103/PhysRevE.69.026128 PACS number(s): 05.90.+m, 65.40.Gr, 47.53.+n

I. INTRODUCTION

Thermodynamical or statistical concept of entropy, though deeply rooted in physics, is rigorously defined only for equilibrium systems or, at best, for adiabatically evolving systems. In fact, the very existence of the entropy in thermodynamics is attributed to Carathéodory's inaccessibility theorem [1] and the statistical interpretation behind the thermodynamical entropy is then usually provided via the ergodic hypothesis [2,3]. It is, however, a highly nontrivial matter to find a proper conceptual ground for entropy of systems away from equilibrium, nonergodic systems, or equilibrium systems with "exotic" non-Gibbsian statistics (multifractals, percolation, polymers, or protein folding provide examples). It is frequently said that entropy is a measure of disorder, and while this needs many qualifications and clarifications it is generally believed that this does represent something essential about it. Information theory might be then viewed as a pertinent mathematical framework capable of quantifying the "measure of disorder." It is an undoubted advantage of information theoretic approaches that whenever one can measure (or control) information one can also measure (or control) the associated entropy, as the latter is essentially an average information about a system in question [4,5].

In recent years there have been many attempts to extend the equilibrium concept of entropy to more generic situations by applying various generalizations of the information theory. Systems with (multi)fractal structure, long-range interactions, and long-time memories might serve as examples. Among a multitude of information entropies Shannon's entropy, Rényi entropies, and Tsallis-Havrda-Charvat (THC) nonextensive entropies [6] have found utility in a wide range of physical problems. Shannon's entropy is known to reproduce the usual Gibbsian thermodynamics and is frequently used in such areas as astronomy, geophysics, biology, medical diagnosis, and economics (for the latest developments in Shannon's entropy applications the interested reader may consult Ref. [7] and citations therein). Rényi entropies were conveniently applied, for instance, in multiparticle hadronic systems [8], fractional diffusion processes [9], or in multifractal systems [10]. THC entropy was recently used in a study of systems with strong long-range correlations and in systems with long-time memories [11].

Despite the information theoretic origin there has been raised some doubt regarding the observability of Rényi entropies [12]. Some authors went even as far as to claim that instabilities in systems with large number of microstates completely invalidate the use of Rényi entropies in all physical problems [13]. This is rather surprising since Rényi's entropy is routinely measured in numerous situations ranging from coding theory and cryptography [14] (where it regulates the optimality of coding), through chaotic dynamical systems [10] (where it determines the generalized dimensions for strange attractors) and earthquake analysis [15] (where it is used to evaluate the distribution of earthquake epicenters and lacunarity) to nonparametric mathematical statistics (where it prescribes the price of constituent information). Besides, Rényi entropies directly provide measurable bounds in quantum-information uncertainty relations [16].

In the present paper we aim to revise Lesche's condition of observability. We illustrate this in various contexts: systems with a finite number of microstates, systems with an infinite (but countable) number of microstates, systems with absolutely continuous probability density functions (PDF's), and multifractals. We show that it is not quite as simple to define the ubiquitous concept of observability. We propose a less restrictive observability condition and demonstrate that Rényi entropies are observable in this new framework. In what follows we will give some considerations in favor of the above statement.

The paper is organized in the following way. In Sec. II we discuss Lesche's criterion of observability which frequently forms a core argument against observability of Rényi entropies. We argue that the criterion is unnecessarily restrictive and, in fact, many standard physical phenomena which are observed and measured in the real world do not obey Lesche's condition. In Sec. III we present some essentials of Rényi entropies required in the main body of the paper. In Sec. IV A we argue that for the finite number of microstates Rényi entropies easily conform with Lesche's criterion, i.e., they are observable. In Sec. IV B we extend our analysis to a countably infinite number of microstates. Here appearance of instabilities may be observed. The latter can be traced to a large sensitivity of Rényi entropies to (ultra)rare-event sys-

*Email address: petr@cm.ph.tsukuba.ac.jp
†Email address: arimitsu@cm.ph.tsukuba.ac.jp

tems. We demonstrate that when the coarse graining is included into realistic measurements, the instabilities get "diluted" and Rényi entropies once again obey Lesche's condition. In Sec. IV C we propose a more realistic criterion of observability where we allow for a certain amount of instability points, provided the latter ones have measure zero. To this extent we employ Bhattacharyya statistical measure—i.e., natural measure on the space of nonparametric statistics. We prove that the Bhattacharyya measure of the above "critical" distributions is, in fact, zero. Finally, we analyze in Sec. V systems with continuous probability distributions and multifractal systems. We find that the very nature of the absolute continuity of PDF's and the multifractality prohibits *per se* an appearance of instability points.

II. LESCHE'S CRITERION OF OBSERVABILITY

In order to explain fully the apparent inconsistencies in the recent claims concerning nonobservability of Rényi entropies we feel it is necessary to briefly review the main points of Lesche's observability criterion. While we hope to discuss all the salient points, a full discussion can be found in Ref. [12]. Our discussion will be in terms of a scalar quantity $G(\boldsymbol{x})$. Following Ref. [12], a necessary condition for $G(\boldsymbol{x})$ with the state[1] variable $\boldsymbol{x}\in X\subset\mathbb{R}^n$ to be observable is the following. Let

$$\|\boldsymbol{x}-\boldsymbol{x}'\|_1=\sum_k^n |x_k-x_k'|$$

be the Hölder l_1 metric on $\mathbb{R}^n$, then $\forall\,\varepsilon>0$ there exists ($\boldsymbol{x}$ independent) $\delta_\varepsilon>0$ such that for any pair $\boldsymbol{x},\boldsymbol{x}'$ one has

$$\|\boldsymbol{x}-\boldsymbol{x}'\|_1\leqslant\delta_\varepsilon\Rightarrow\frac{|G(\boldsymbol{x})-G(\boldsymbol{x}')|}{G_{max}}<\varepsilon. \tag{1}$$

From a strict mathematical standpoint Eq. (1) is, in fact, the definition of the uniform metric continuity of $G(\boldsymbol{x})$ on the state space X. Informally Eq. (1) states that points from X which are close in sense of $\|\cdots\|_1$ are mapped via G to points which are close in $|\cdots|$ metric. Lesche's criterion is thus nothing but the condition of stability of $G(\boldsymbol{x})$ under a measurement. In fact, the continuity criterion ensures that a small error in a state variable $\boldsymbol{x}$ will not bring in repeated experiments violent fluctuations in measured data. The *uniform* continuity in Eq. (1) is then a key ingredient to secure that the size of the changes in $G(\boldsymbol{x})$ depends only on the size of the changes in $\boldsymbol{x}$ but not on $\boldsymbol{x}$ itself. This condition excludes, for example, systems whose statistical fluctuations in $G(\boldsymbol{x})$ would change too dramatically with a small change in the state variable $\boldsymbol{x}$.

When $G(\boldsymbol{x})$ is bounded we can recast Lesche's condition of observability into an equivalent but more expedient form; namely (inverse) Lipschitz continuity condition [17]. In this case, a quantity G: $X\subset\mathbb{R}^n\mapsto\mathbb{R}$ is observable in Lesche's sense if and only if for every $\varepsilon>0$ there exists ($\boldsymbol{x}$ independent and finite) K_ε such that for any pair $\boldsymbol{x},\boldsymbol{x}'\in X$ one has

$$|G(\boldsymbol{x})-G(\boldsymbol{x}')|\leqslant K_\varepsilon\|\boldsymbol{x}-\boldsymbol{x}'\|_1+\varepsilon. \tag{2}$$

We will practically employ the condition (2) in Sec. IV A.

Criteria (1) and (2) get generalized in the case when $n\to\infty$. This should be expected as the uniform continuity may not survive in the large n limit. To avoid such situations Lesche postulated that the mapping

$$G:\ \bigcup_{n=1}^{\infty} X_n\mapsto\mathbb{R} \tag{3}$$

with $X_n\subset\mathbb{R}^n$, taken as a function of n, converges to a uniformly continuous function in a uniform manner, i.e., $\forall\,\varepsilon>0$ there exists $\delta_\varepsilon>0$ such that $\forall\ \boldsymbol{x},\boldsymbol{x}'\in\mathbb{R}^n$ and $\forall\ n\in\mathbb{Z}^+$

$$\|\boldsymbol{x}-\boldsymbol{x}'\|_1\leqslant\delta_\varepsilon\Rightarrow\frac{|G(\boldsymbol{x})-G(\boldsymbol{x}')|}{G_{max}}<\varepsilon. \tag{4}$$

The uniform convergence is then reflected in the fact that δ_ε is both $\boldsymbol{x}$ and n independent.

Let us add a couple of remarks concerning the aforementioned observability conditions. Lesche's condition, as illustrated above, is based on the notion of measurability. This is, however, not the only possible way how to define observability. It is well known that various alternative concepts exist in literature. For instance, one may use the approach based on distinguishability [18] or detectability [19]. In fact, the condition based on measurability, and namely the condition of uniform continuity, might be often too tight. Indeed, there are clearly many quantities which are not uniformly continuous in their state variables (e.g., they are discontinuous in a finite number of points in the state space) and which are, nevertheless, perfectly detectable and well defined away from the singularity domain. Note, for instance, that although pressure and latent heat in first order phase transitions are discontinuous in temperature, and similarly susceptibility in second order phase transitions is nonanalytic in temperature, there is still no reason to dismiss pressure, latent heat, and susceptibility as observables. Discontinuous or nonanalytic *state* functions are not exclusive to phase transitions only. Actually, such a type of behavior is common to many different situations—formation of shocks in nonlinear wave propagation, mechanical systems involving small masses and large damping, electric-circuit systems with large resistance and small inductance, catastrophe and bifurcation theories, to name a few.

[1]Here and throughout, the state space X represents the sample space of mathematical statistics, i.e., the space over which the probability distributions operate. In simple situations this coincides with the set of all possible outcomes in some experiment. Generally, the elements of X can represent probability distributions themselves, provided a suitable measure is defined. This fact will be used in Sec. IV.

III. RÉNYI ENTROPIES

Rényi entropies constitute a one-parametric family of information entropies labeled by Rényi's parameter $\alpha \in \mathbb{R}^+$ and fulfill the additivity with respect to the composition of statistically independent systems. The special case with $\alpha = 1$ corresponds to ordinary Shannon's entropy. It might be shown that Rényi entropies belong to the class of mixing homomorphic functions [12] and that they are analytic for α's which lie in $I \cup IV$ quadrants of the complex plane [20]. In order to address the observability issue it is important to distinguish three situations.

A. Discrete probability distribution case

Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be a random variable admitting n different events (be it outcomes of some experiment or microstates of a given macrosystem), and let $\mathcal{P} = \{p_1, \ldots, p_n\}$ be the corresponding probability distribution. Information theory then ensures that the most general information measures (i.e., entropy) compatible with the additivity of independent events are those of Rényi [4]:

$$\mathcal{I}_\alpha(\mathcal{P}) = \frac{1}{(1-\alpha)} \log_2 \left(\sum_{k=1}^{n} p_k^\alpha \right). \tag{5}$$

Form (5) is valid even in the limiting case when $n \to \infty$. If, however, n is finite then Rényi entropies are bounded both from below and from above: $\log_2 (p_k)_{\max} \leq \mathcal{I}_\alpha \leq \log_2 n$. In addition, Rényi entropies are monotonically decreasing functions in α, so namely $\mathcal{I}_{\alpha_1} < \mathcal{I}_{\alpha_2}$ if and only if $\alpha_1 > \alpha_2$. One can reconstruct the entire underlying probability distribution knowing all Rényi distributions via Widder-Stiltjes inverse formula [20]. In the latter case the leading order contribution comes from $\mathcal{I}_1(\mathcal{P})$, i.e., from Shannon's entropy. Typical playground of Eq. (5) is in a coding theory [21], cryptography [14], and in the theory of statistical inference [4]. The parameter α might be then related with the price of constituent information. It should be admitted that in discrete cases the conceptual connection of $\mathcal{I}_\alpha(\mathcal{P})$ with actual physical problems is still an open issue. The interested reader can find some further practical applications of discrete Rényi entropies, for instance, in Refs. [20,22]

B. Continuous probability distribution case

Let M be a support on which is defined a continuous PDF $\mathcal{F}(\boldsymbol{x})$. We will assume that the support (or outcome space) can be generally a fractal set. By covering the support with the mesh $M^{(l)}$ of d–dimensional (disjoint) cubes $M_k^{(l)}$ ($k = 1, \ldots, n$) of size l^d we may define the integrated probability in kth cube as

$$p_{nk} = \mathcal{F}(\boldsymbol{x}_i) l^d, \quad \boldsymbol{x}_i \in M_k^{(l)}. \tag{6}$$

The latter specifies the mesh probability distribution $\mathcal{P}_n = \{p_{n1}, \ldots, p_{nn}\}$. Infinite precision of measurements (i.e., with $l \to 0$) often brings infinite information. In fact, it is more sensible to consider the relative information entropy rather than absolute one as the most "junk" information comes from the uniform distribution $\mathcal{E}_n$. It was shown in Refs. [4,20] that in the $n \to \infty$ (i.e., $l \to 0$) limit it is possible to define finite information measure compatible with information theory axioms. This *renormalized* Rényi's entropy—negentropy (or information gain), reads

$$\tilde{\mathcal{I}}_\alpha(\mathcal{F}) \equiv \lim_{n\to\infty} [\mathcal{I}_\alpha(\mathcal{P}_n) - \mathcal{I}_\alpha(\mathcal{E}_n)] = \frac{1}{(1-\alpha)} \log_2 \left(\frac{\int_M d\mu \mathcal{F}^\alpha(\boldsymbol{x})}{\int_M d\mu 1/V^\alpha} \right). \tag{7}$$

Here V is the corresponding volume. Equation (7) can be viewed as a generalization of the Kullback-Leibler relative entropy [24]. It is possible to introduce a simpler alternative prescription as

$$\begin{aligned} \mathcal{I}_\alpha(\mathcal{F}) &\equiv \lim_{n\to\infty} [\mathcal{I}_\alpha(\mathcal{P}_n) - \mathcal{I}_\alpha(\mathcal{E}_n)|_{V=1}] \\ &= \lim_{n\to\infty} [\mathcal{I}_\alpha(\mathcal{P}_n) + D \log_2 l] \\ &= \frac{1}{(1-\alpha)} \log_2 \left(\int_M d\mu \mathcal{F}^\alpha(\boldsymbol{x}) \right). \end{aligned} \tag{8}$$

In both previous cases the measure μ is the Hausdorff measure [23]:

$$\mu(d;l) = \sum_{k\text{th box}} l^d \xrightarrow{l\to 0} \begin{cases} 0 & \text{if } d < D, \\ \infty & \text{if } d > D, \end{cases}$$

with D being the Hausdorff dimension of the support. Rényi entropies (7) and (8) are defined if and only if the corresponding integral $\int_M d\mu \mathcal{F}^\alpha(\boldsymbol{x})$ exists. Equations (7) and (8) indicate that asymptotic expansion for $\mathcal{I}_\alpha(\mathcal{P}_n)$ has the form

$$\begin{aligned} \mathcal{I}_\alpha(\mathcal{P}_n) &= -D \log_2 l + \mathcal{I}_\alpha(\mathcal{F}) + o(1) \\ &= -D \log_2 l + \tilde{\mathcal{I}}_\alpha(\mathcal{F}) + \log_2 V_n + o(1). \end{aligned} \tag{9}$$

Here V_n is the prefractal volume and the symbol $o(1)$ is the residual error which tends to 0 for $l \to 0$. In contrast to the discrete case, Rényi entropies $\mathcal{I}_\alpha(\mathcal{F})$ are not positive here.

Information measures $\tilde{\mathcal{I}}_\alpha(\mathcal{F})$ and $\mathcal{I}_\alpha(\mathcal{F})$ have been so far mostly applied in the theory of statistical inference [25] and in chaotic dynamical systems [10]. Let us note finally that one may view the discrete distributions as a special case of the continuous PDF's, provided the outcome space (or sample space) is discrete. In such a situation the Hausdorff dimension D is zero and Eq. (8) reduces directly to Eq. (5).

C. Multifractal systems

Multifractals can be viewed as statistical systems where both cells in the covering mesh and integrated probabilities scale as some power of l. Grouping all the integrated probabilities according to their scaling exponents (Lipshitz-

Hölder exponents), say a, we effectively divide the support into the ensemble of intertwined unifractals, each with its own fractal dimension $f(a)$. Exponents $f(a)$ are called the singularity spectrum. In multifractal analysis it is customary to introduce yet another pair of scaling exponents, namely, the correlation exponent $\tau(\alpha)$ which prescribes scaling of the partition function and "inverse temperature" α. These two descriptions are related via Legendre transformation

$$\tau(\alpha)=\min_a(\alpha a-f(a)). \tag{10}$$

As in the case of continuous PDF's, the renormalization of Rényi entropies is required to extract relevant finite information—negentropy. It is possible to show that the following renormalized Rényi's entropy complies with the axiomatics of the information theory [20]:

$$\begin{aligned}\mathcal{I}_\alpha(\mu_{\mathcal{P}})&\equiv\lim_{l\to 0}[\mathcal{I}_\alpha(\mathcal{P}_n)-\mathcal{I}_\alpha(\mathcal{E}_n)|_{V=1}]\\ &=\lim_{l\to 0}\left(\mathcal{I}_\alpha(\mathcal{P}_n)+\frac{\tau(\alpha)}{(\alpha-1)}\log_2 l\right)\\ &=\frac{1}{(1-\alpha)}\log_2\left(\int_a d\mu_{\mathcal{P}}^{(\alpha)}(a)\right).\end{aligned} \tag{11}$$

Here the multifractal measure is defined as [23]

$$\mu_{\mathcal{P}}^{(\alpha)}(d;l)=\sum_{k\text{th box}}\frac{p_{nk}^{\alpha}}{l^d}\overset{l\to 0}{\to}\begin{cases}0 & \text{if } d<\tau(\alpha),\\ \infty & \text{if } d>\tau(\alpha).\end{cases}$$

Rényi entropies $\mathcal{I}_\alpha(\mu_{\mathcal{P}})$ are defined if and only if the corresponding integrals $\int_a d\mu_{\mathcal{P}}^{(\alpha)}(a)$ exist. Equation (11) implies the following asymptotic expansion for $\mathcal{I}_\alpha(\mathcal{P}_n)$:

$$\mathcal{I}_\alpha(\mathcal{P}_n)=-D(\alpha)\log_2 l+\mathcal{I}_\alpha(\mu_{\mathcal{P}})+o(1). \tag{12}$$

Here

$$D(\alpha)\equiv\frac{\tau(\alpha)}{(\alpha-1)}=\lim_{l\to 0}\frac{\mathcal{I}_\alpha(\mathcal{P}_n)}{\log_2(1/l)}, \tag{13}$$

is the, so called, generalized dimension [23]. Note also that for systems of Sec. III B $D(\alpha)$ is α independent.

Let us stress that Rényi's entropy of multifractal systems is a more convenient tool than the ordinary Shannon's entropy. It is possible to show that one can obtain Shannon's entropy for any unifractal by merely changing the Rényi parameter. In fact, Rényi's parameter coincides in this case with the singularity spectrum [20].

IV. OBSERVABILITY OF RÉNYI ENTROPIES: DISCRETE PROBABILITY DISTRIBUTION

A. Finite case

It is quite simple to see that for systems with a finite number of outcomes (e.g., systems with a finite number of microstates) Lesche's criterion of observability is fulfilled. The proof goes as follows.[2] We first use the inequality $\ln x \leqslant x-1$ and assume that $\Sigma_k p_k^\alpha \geqslant \Sigma_k q_k^\alpha$, then

$$\begin{aligned}|\mathcal{I}_\alpha(\mathcal{P})-\mathcal{I}_\alpha(\mathcal{Q})|&\leqslant\frac{1}{|1-\alpha|}\left(\frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n q_i^\alpha}-1\right)\\ &=\frac{1}{|1-\alpha|\sum_{i=1}^n q_k^\alpha}\sum_{i=1}^n (p_i^\alpha-q_i^\alpha).\end{aligned}$$

This might be written in the invariant form as

$$\begin{aligned}|\mathcal{I}_\alpha(\mathcal{P})-\mathcal{I}_\alpha(\mathcal{Q})|&\leqslant\frac{1}{|1-\alpha|c(\alpha,\mathcal{P},\mathcal{Q})}\left|\sum_{i=1}^n (p_i^\alpha-q_i^\alpha)\right|\\ &\leqslant\frac{1}{|1-\alpha|d(\alpha,n)}\left|\sum_{i=1}^n (p_i^\alpha-q_i^\alpha)\right|.\end{aligned} \tag{14}$$

Here $c(\alpha,\mathcal{P},\mathcal{Q})=\min(\Sigma_i p_i^\alpha,\Sigma_i q_i^\alpha)$ and

$$d(\alpha,n)=\begin{cases}1 & \text{if } 0<\alpha\leqslant 1,\\ n^{1-\alpha} & \text{if } \alpha\geqslant 1.\end{cases}$$

To find the efficient estimate for $|\Sigma_k(p_k^\alpha-q_k^\alpha)|$ in terms of $\|\mathcal{P}-\mathcal{Q}\|_1$ we utilize the following trick. Let us define the function

$$\mathcal{A}(s,\mathcal{P})=\sum_{k=1}^n [p_k-f(s)]\theta(p_k-f(s)). \tag{15}$$

Here $\theta(\cdots)$ is the Heaviside step function and f: $[a,b]\mapsto[0,1]$ is some invertible function. Both $f(s)$, a, and b will be chosen at the latter stage so as to facilitate our computations. Note also that

$$\max\{0;[1-nf(s)]\}\leqslant\mathcal{A}(s,\mathcal{P})\leqslant 1. \tag{16}$$

An important property of $\mathcal{A}(s,\mathcal{P})$ is the following straightforward inequality:

$$\begin{aligned}|\mathcal{A}(s,\mathcal{P})-\mathcal{A}(s,\mathcal{Q})|&\leqslant\sum_{k=1}^n |[p_k-f(s)]\theta(p_k-f(s))\\ &\quad-[q_k-f(s)]\theta(q_k-f(s))|\\ &\leqslant\sum_{k=1}^n |p_k-q_k|=\|\mathcal{P}-\mathcal{Q}\|_1,\end{aligned} \tag{17}$$

which is valid for any $s\in[a,b]$. Note further that

[2]For simplicity's sake we use in this subsection a natural logarithm instead of $\log_2$.

$$\int_a^b \mathcal{A}(s,\mathcal{P})ds = \sum_{k=1}^n \int_{f(a)}^{f(b)} (p_k - x)\,\theta(p_k - x)[f^{-1}(x)]'dx$$
$$= \sum_{k=1}^n \left\{ \theta(p_k - f(a))\left([f(a) - p_k]a + \int_{f(a)}^{p_k} f^{-1}(x)dx \right) + \theta(p_k - f(b)) \times \left([p_k - f(b)]b + \int_{p_k}^{f(b)} f^{-1}(x)dx \right) \right\}. \tag{18}$$

Here we have used the fact that p_k's must lie somewhere between $f(a)$ and $f(b)$. If we now chose $f(x) = (x/\alpha)^{1/(\alpha-1)}$ with

$$a = \begin{cases} \infty & \text{if } 0 < \alpha < 1 \\ 0 & \text{if } \alpha > 1 \end{cases} \quad \text{and} \quad b = \alpha$$

[so $f(a) = 0$, $f(b) = 1$] we obtain

$$\left| \int_a^b [\mathcal{A}(s,\mathcal{P}) - \mathcal{A}(s,\mathcal{Q})]ds \right| = \left| \sum_{k=1}^n (p_k^\alpha - q_k^\alpha) \right|. \tag{19}$$

Applying Eqs. (17) and (19) we may write for $\alpha > 1$,

$$\left| \sum_{k=1}^n (p_k^\alpha - q_k^\alpha) \right| \leqslant \left\{ \int_0^c n(s/\alpha)^{1/(\alpha-1)}ds + \int_c^\alpha \left| \mathcal{A}(s,\mathcal{P}) - \mathcal{A}(s,\mathcal{Q}) \right| ds \right\}$$
$$\leqslant n(\alpha-1)(c/\alpha)^{\alpha/\alpha-1} + (\alpha - c)\|\mathcal{P} - \mathcal{Q}\|_1. \tag{20}$$

So if we take $c = \alpha(\varepsilon/n^\alpha)^{(\alpha-1)/\alpha}$ (this assures that $\mathcal{A}(s,\mathcal{P}) \geqslant [1 - nf(s)] > 0$ for $s \in (0,c]$), then

$$|\mathcal{I}_\alpha(\mathcal{P}) - \mathcal{I}_\alpha(\mathcal{Q})| \leqslant K_\varepsilon^{(1)} \|\mathcal{P} - \mathcal{Q}\|_1 + \varepsilon, \tag{21}$$

with $K_\varepsilon^{(1)} = [\alpha/(\alpha-1)][(n^\alpha/\varepsilon)^{(\alpha-1)/\alpha} - 1]\varepsilon^{(\alpha-1)/\alpha}$.

In case when $0 < \alpha < 1$ we may utilize Eqs. (16), (17), and (19) to obtain

$$\left| \sum_{k=1}^n (p_k^\alpha - q_k^\alpha) \right| \leqslant \left\{ \int_\alpha^{\tilde{c}} \left| \mathcal{A}(s,\mathcal{P}) - \mathcal{A}(s,\mathcal{Q}) \right| ds + \int_{\tilde{c}}^\infty n(s/\alpha)^{1/(\alpha-1)}ds \right\}$$
$$\leqslant (\tilde{c} - \alpha)\|\mathcal{P} - \mathcal{Q}\|_1 + n(1-\alpha)(\tilde{c}/\alpha)^{\alpha/(\alpha-1)}. \tag{22}$$

By setting $\tilde{c} = \alpha(\varepsilon/n)^{(\alpha-1)/\alpha}$ (this assures that $\mathcal{A}(s,\mathcal{P}) \geqslant [1 - nf(s)] > 0$ for $s \in [\tilde{c},\infty)$) we have

$$|\mathcal{I}_\alpha(\mathcal{P}) - \mathcal{I}_\alpha(\mathcal{Q})| \leqslant K_\varepsilon^{(2)} \|\mathcal{P} - \mathcal{Q}\|_1 + \varepsilon, \tag{23}$$

with $K_\varepsilon^{(2)} = [\alpha/(1-\alpha)][(\varepsilon/n)^{(\alpha-1)/\alpha} - 1]$. Note particularly that $\lim_{\alpha \to 1_+} K_\varepsilon^{(1)} = \ln(n/\varepsilon)$ and $\lim_{\alpha \to 1_-} K_\varepsilon^{(2)} = \ln(n/\varepsilon)$. This indicates that the Lipschitz conditions (21) and (23) can be analytically continued to $\alpha = 1$. This reconfirms the well known result that Shannon's entropy is Lipschitz.

Finally note that Eqs. (21) and (23) represent the Lesche criterion (2). Hence, in cases when the state space corresponds to the space of all possible probability distributions assigned to a definite (finite) number of outcomes (microstates), Rényi entropies are measurable in Lesche's sense.

B. Infinite limit case

As was already mentioned in Sec. II, the situation becomes be more delicate in the large n limit. This is because for the sake of uniform metric continuity at any n one might require that also the limiting case should obey the uniform continuity. To tackle statistical systems with a countable infinity of microstates[3] we will illustrate first that by introducing a coarse graining into a realistic measurement, alleged Lesche's counterexamples do not apply.

In his paper [12] Lesche proposed the following examples to demonstrate the nonobservability of Rényi entropies. In $\alpha > 1$ he picked up two distributions, namely ($i = 1, \ldots, n$),

$$\mathcal{P} = \left\{ p_i = \frac{1}{n-1}(1 - \delta_{1i}) \right\},$$
$$\mathcal{P}' = \left\{ p_i' = \frac{\delta}{2}\delta_{1i} + \left(1 - \frac{\delta}{2}\right)\left(\frac{1 - \delta_{1i}}{n-1}\right) \right\},$$
$$\|\mathcal{P} - \mathcal{P}'\|_1 = \delta. \tag{24}$$

Lesche then went on to show that these two distributions do not fulfill the uniform continuity in the large n limiting case. Let us now show that the coarse graining (which is naturally present in any realistic measurement) will restore the uniform continuity for the large n limit case.

We will assume, for simplicity's sake, that the discrete probability distributions (24) are living on the unit lattice with equidistantly distributed lattice (i.e., support) points. In the spirit of Lesche's paper we assume that the true probability distribution on the interval [0,1] is obtained in $n \to \infty$ limit (i.e., when the lattice spacing tends to zero). As usually, we will keep $n \gg 1$ finite during calculations and set to infinity only at the very last stage. Because every actual measurement has a certain resolution capacity we will further assume that a realistic measurement can sample the unit interval through a window of width $1/k$ ($k \ll n$) (so k windows will cover the support space). In this case one can know only

[3] Such systems often appear in various physical situations. (Countable) Markov chains, Fermi-Pasta-Ulam lattice models, or symbolic dynamical models being examples.

integrated probabilities, hence $\mathcal{P}\to\mathcal{P}_{(k)}$ and $\mathcal{P}'\to\mathcal{P}'_{(k)}$. As in every window there are n/k underlying p_i's we have ($i=1,\dots,k$)

$$\mathcal{P}_{(k)}=\left\{p_i^{(k)}=\frac{1}{n-1}\left(\frac{n}{k}-\delta_{1i}\right)\right\}, \tag{25}$$

$$\mathcal{P}'_{(k)}=\left\{p_i'^{(k)}=\frac{\delta}{2}\delta_{1i}+\frac{(1-\delta/2)}{n-1}\left(\frac{n}{k}-\delta_{1i}\right)\right\},$$

$$\|\mathcal{P}_{(k)}-\mathcal{P}'_{(k)}\|_1=\delta.$$

Using the fact that $\mathcal{I}_{\alpha\,\max}=\log_2 k$ we have

$$\begin{aligned}\frac{|\mathcal{I}_\alpha(\mathcal{P}_{(k)})-\mathcal{I}_\alpha(\mathcal{P}'_{(k)})|}{\mathcal{I}_{\alpha\,\max}}&=\left|\frac{1}{(1-\alpha)}\log_2\left[\frac{\left(\frac{1}{n-1}\right)^\alpha\left(\frac{n}{k}-1\right)^\alpha+(k-1)\left(\frac{1}{n-1}\right)^\alpha\left(\frac{n}{k}\right)^\alpha}{\left[\frac{\delta}{2}+\frac{(1-\delta/2)}{n-1}\left(\frac{n}{k}-1\right)\right]^\alpha+(k-1)\left(\frac{1-\delta/2}{n-1}\right)^\alpha\left(\frac{n}{k}\right)^\alpha}\right]\right|\times(\log_2 k)^{-1}\\ &\overset{n\to\infty}{\to}\left|\frac{1}{(1-\alpha)}\log_2\left[\frac{\left(\frac{1}{k}\right)^\alpha+(k-1)\left(\frac{1}{k}\right)^\alpha}{\left(\frac{\delta}{2}+\frac{(1-\delta/2)}{k}\right)^\alpha+(k-1)\left(\frac{1-\delta/2}{k}\right)^\alpha}\right]\Big/\log_2 k\right|\\ &=\left|\frac{1}{1-\alpha}-\frac{1}{1-\alpha}\log_2\left[\left(1+\frac{\delta}{2}(k-1)\right)^\alpha+(k-1)\left(1-\frac{\delta}{2}\right)^\alpha\right]\Big/\log_2 k\right|\\ &=\left(\frac{\delta}{2}\right)^2\frac{\alpha}{2}\,\frac{(k-1)}{\ln k}+O(\delta^3).\end{aligned} \tag{26}$$

It is now simple to see that Lesche's condition is easily fulfilled, as for arbitrarily small ε there exist δ_ε, namely,

$$\delta_\varepsilon\leqslant 2\sqrt{\frac{\varepsilon}{k-1}\ln(k)^{2/\alpha}}, \tag{27}$$

for which the metric proximity $\|\mathcal{P}_{(k)}-\mathcal{P}'_{(k)}\|_1\leqslant\delta_\varepsilon$ implies the proximity of outcomes, i.e., $|\mathcal{I}_\alpha(\mathcal{P}_{(k)})-\mathcal{I}_\alpha(\mathcal{P}'_{(k)})|/\log_2 k\leqslant\varepsilon$. This result is clearly independent of n because whenever n is finite the outcome of the preceding section applies and for $n\to\infty$ the validity has been just proven.

We proceed analogously for $\alpha<1$. In this case Lesche's counterexamples were provided by two distributions ($i=1,\dots,n$)

$$\mathcal{P}=\{p_i=\delta_{1i}\},$$

$$\mathcal{P}'=\left\{p_i'=\left(1-\frac{\delta}{2}\right)\delta_{1i}+\frac{1}{n-1}\,\frac{\delta}{2}(1-\delta_{1i})\right\},$$

$$\|\mathcal{P}-\mathcal{P}'\|_1=\delta. \tag{28}$$

As before, we can obtain integrated probability distributions which read ($i=1,\dots,k$)

$$\mathcal{P}_{(k)}=\{p_i^{(k)}=\delta_{1i}\},$$

$$\mathcal{P}'_{(k)}=\left\{p'^{(k)}_i=\left(1-\frac{\delta}{2}\right)\delta_{1i}+\frac{1}{n-1}\,\frac{\delta}{2}\left(\frac{n}{k}-\delta_{1i}\right)\right\},$$

$$\|\mathcal{P}_{(k)}-\mathcal{P}'_{(k)}\|_1=\delta, \tag{29}$$

and so

$$\begin{aligned}\frac{|\mathcal{I}_\alpha(\mathcal{P}_{(k)})-\mathcal{I}_\alpha(\mathcal{P}'_{(k)})|}{\mathcal{I}_{\alpha\,\max}}&=\left|\frac{1}{(1-\alpha)}\log_2\left[\left(1-\frac{\delta}{2}\,\frac{n(k-1)}{k(n-1)}\right)^\alpha\right.\right.\\ &\quad\left.\left.+(k-1)\left(\frac{\delta}{2}\,\frac{n}{k(n-1)}\right)^\alpha\right]\right|\times(\log_2 k)^{-1}\\ &\overset{n\to\infty}{\to}\frac{1}{(1-\alpha)}\left|\log_2\left[\left(1-\frac{\delta}{2}\,\frac{k-1}{k}\right)^\alpha\right.\right.\\ &\quad\left.\left.+(k-1)\left(\frac{\delta}{2k}\right)^\alpha\right]\right|\Big/\log_2 k\\ &\leqslant\left(\frac{\delta}{2k}\right)^2\frac{\alpha}{2}\,\frac{(k-1)^2}{\ln k}+O(\delta^3).\end{aligned} \tag{30}$$

Here the inequality

$$x^\alpha-\alpha x\geqslant 0 \quad \text{for} \quad x\in[0,1],\alpha\in[0,1]$$

was used on the last line. Consequently we again see that for sufficiently small ε there exist δ_ε, namely,

$$\delta_\varepsilon\leqslant\frac{2k}{(k-1)}\sqrt{\varepsilon\,\ln(k)^{2/\alpha}}, \tag{31}$$

which satisfies Lesche's condition. Note, that from Eqs. (27) and (31) it follows that our argument naturally includes also the case $\alpha = 1$ (i.e., Shannon's entropy) as in all steps leading to Eqs. (27) and (31) we have well defined limits $\alpha \rightarrow 1_+$ and $\alpha \rightarrow 1_-$, respectively.

C. Region of instability

In the preceding section we have found that Lesche's counterexamples can be bypassed by introducing a coarsened resolution into a measurement process. Let us now show that even when the coarsening is not employed the Lesche instability points have zero measure in the space of all discrete infinite distributions—Bhattacharyya's measure [26]—and hence they do not affect a measurement in most practical situations.

The key observation is that Lesche's counterexamples single out a very narrow class of probability distributions. In particular, they imply that when $\alpha > 1$, only distributions with high peak probabilities create problems. Similarly, in cases where $\alpha < 1$ only distributions with an infinite number of microstates having a negligible overall probability exhibit a critical type of behavior. We now demonstrate that the above probability distributions have a very small relevance in the actual measurement. For this purpose we remind the reader the concept of Bhattacharyya measure [26].

Suppose that $\mathcal{X}$ is a discrete random variable with n different values, $\mathbb{P}_n$ is the probability space affiliated with $\mathcal{X}$, and $\mathcal{P} = \{p_1, \ldots, p_n\}$ is a sample probability distribution from $\mathbb{P}_n$. Because $\mathcal{P}$ is non-negative and summable to unity, it follows that the square-root likelihood $\xi_i = \sqrt{p_i}$ exists for all $i = 1, \ldots, n$, and it satisfies the normalization condition

$$\sum_{i=1}^{n} (\xi_i)^2 = 1. \tag{32}$$

We see that $\boldsymbol{\xi}$ can be regarded as a unit vector in the Hilbert space $\mathcal{H} = \mathbb{R}^n$. Now, let $\mathcal{P}^{(1)}$ and $\mathcal{P}^{(2)}$ denote a pair of probability distributions and $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$ the corresponding elements in Hilbert space. Then the inner product

$$\cos\phi = \sum_{i=1}^{n} \xi_i^{(1)} \xi_i^{(2)} = 1 - \frac{1}{2} \sum_{i=1}^{n} (\xi_i^{(1)} - \xi_i^{(2)})^2 \tag{33}$$

defines the angle ϕ that can be interpreted as a distance between two probability distributions. More precisely, if $\mathcal{S}^{n-1}$ is the unit sphere in the n-dimensional Hilbert space, then ϕ is the spherical (or geodesic) distance between the points on $\mathcal{S}^{n-1}$ determined by $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$. Clearly, the maximal possible distance, corresponding to orthogonal distributions, is given by $\phi = \pi/2$. This follows from the fact that $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$ are non-negative, and hence they are located only on the positive orthant of $\mathcal{S}^{n-1}$. Spherical geometry on $\mathcal{S}^{n-1}$ then naturally induces the measure—Bhattacharyya measure. The corresponding geodesic distance ϕ is the, so called, Bhattacharyya distance. We remark that the surface "area" of the orthant $(\mathcal{S}^{n-1})^+$, i.e., the volume of the probability space $\mathbb{P}_n$, is

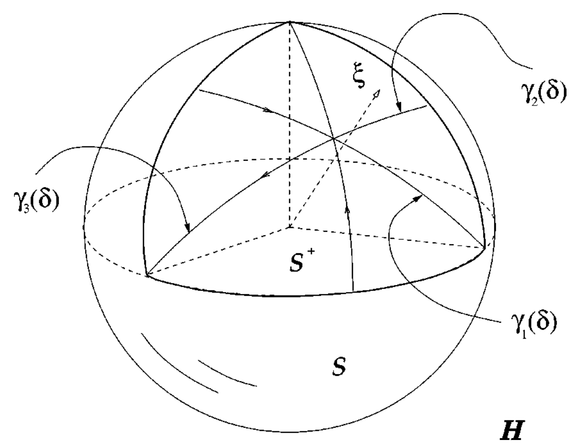

FIG. 1. The family of Lesche's critical distributions ($\alpha > 1$). A statistical system can be represented by points $\boldsymbol{\xi}$ on a positive orthant $\mathcal{S}^+$ of the unit sphere $\mathcal{S}$ in a real Hilbert space $\mathcal{H}$. 1-parametric families of Leshe's critical distributions are then represented by arcs $\gamma_i(\delta) = \{\xi_k(\delta) = \sqrt{\delta\delta_{ik}/2 + (1-\delta/2)[(1-\delta_{ik})/(n-1)]}; k \in 1, \ldots, n; \delta \in [0,2]\}$. Depicted example corresponds to $\mathcal{S} = \mathcal{S}^2$.

$$V_{n-1}(\mathbb{P}_n) \equiv V_{n-1}((\mathcal{S}^{n-1})^+) = \frac{1}{2^n} \int d\Omega^n = \frac{\pi^{n/2}}{2^{n-1}\Gamma\left(\frac{n}{2}\right)}. \tag{34}$$

The Bhattacharyya measure of any set $\mathcal{A} \subseteq (\mathcal{S}^{n-1})^+$ is then

$$\mu_B(\mathcal{A}) = \frac{V_{n-1}(\mathcal{A})}{V_{n-1}(\mathbb{P}_n)}, \tag{35}$$

and so particularly the normalization $\mu_B(\mathbb{P}_n) = 1$ holds. The reader may see that the Bhattacharyya measure is indeed a very natural concept. In fact, Eq. (35) implies that the latter is just the Haar measure on $\mathcal{S}^{n-1}$. One could possibly adopt some other (not spherical) metric on the the probability space $(\mathcal{S}^{n-1})^+$, but because all nonsingular metric measures are on compact manifolds equivalent (i.e., they differ only by finite multiplicative functions—Jacobians) the Bhattacharyya measure will be fully satisfactory for our purpose. Actually the exclusiveness of Bhattacharyya measure in nonparametric statistics was already emphasized, for instance, in Ref. [27]. The naturalness and simplicity of Bhattacharyya's measure have been also appreciated in various areas of physics and engineering ranging from quantum mechanics [28] to statistical pattern recognition and signal processing [29].

1. *α>1 case*

Let us now look at the Bhattacharyya measure of the family of Lesche's critical distributions corresponding to $\alpha > 1$. In this case the relation (24) suggests that the critical distributions form the 1-parametric family of distributions parametrized by δ. Figure 1 indicates that there are clearly n such families. In contrast to the orthant surface which has dimension $D = n - 1$, the countable set of linelike 1-parametric families has the topological dimension $D=1$ and hence the Bhattacharyya measure of Lesche's critical distributions is plainly zero.

We wish to ask whether some extension of Eq. (24) might have the nonzero measure. We will illustrate now that the answer is negative. In fact, we will show that with Bhattacharyya measure approaching 1 (in the limit of large n) all distributions $\mathcal{P}\in\mathbb{P}_n$ inevitably fulfil Lesche's condition (4). Inasmuch, all distributions which exhibit the critical behavior encountered in Ref. [12] have $\mu_B\to 0$ as $n\to\infty$. To prove this we employ the following isoperimetric inequality (also known as Levy's lemma) [30]. Let f: $\mathcal{S}^{n-1}\mapsto\mathbb{R}$ be a K-Lipshitz function, i.e., for any pair $\boldsymbol{\xi}^{(1)},\boldsymbol{\xi}^{(2)}\in\mathcal{S}^{n-1}$,

$$\|f(\boldsymbol{\xi}^{(1)})-f(\boldsymbol{\xi}^{(2)})\|\leqslant K\|\boldsymbol{\xi}^{(1)}-\boldsymbol{\xi}^{(2)}\|_2. \tag{36}$$

Then

$$\frac{V_{n-1}\left(\boldsymbol{\xi}\in\mathcal{S}^{n-1};\left|f(\boldsymbol{\xi})-\int_{\mathcal{S}^{n-1}}f\,d\mu\right|>C\right)}{V_{n-1}(\mathcal{S}^{n-1})}\leqslant 4e^{-\vartheta C^2 n/K}, \tag{37}$$

where μ is the Haar measure on $\mathcal{S}^{n-1}$ and ϑ is an absolute (i.e., n-independent) constant whose precise form is not important here.[4]

Let us choose $f(\boldsymbol{\xi})=\|\boldsymbol{\xi}\|_{2\alpha}$. Using the triangle inequality we have

$$|\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}-\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}|\leqslant\|\boldsymbol{\xi}^{(1)}-\boldsymbol{\xi}^{(2)}\|_{2\alpha}\leqslant\|\boldsymbol{\xi}^{(1)}-\boldsymbol{\xi}^{(2)}\|_2, \tag{38}$$

so $\|\boldsymbol{\xi}\|_{2\alpha}$ is 1-Lipshitz function. In addition,

$$\begin{aligned}\|\mathcal{P}-\mathcal{Q}\|_1&=\sum_i |(\xi_i^{(1)})^2-(\xi_i^{(2)})^2|\\&=\sum_i |\xi_i^{(1)}-\xi_i^{(2)}|(\xi_i^{(1)}+\xi_i^{(2)})\geqslant\|\boldsymbol{\xi}^{(1)}-\boldsymbol{\xi}^{(2)}\|_2^2.\end{aligned} \tag{39}$$

So particularly when two distributions are δ close then their representative points on the sphere fulfil the inequality

$$|\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}-\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}|\leqslant\sqrt{\delta}. \tag{40}$$

The next step is to calculate the mean $\int_{\mathcal{S}^{n-1}}f(\boldsymbol{\xi})d\mu$. As it stands, this is a quite difficult task but fortunately we may take advantage of the fact that

$$\begin{aligned}\int_{\mathcal{S}^{n-1}}\sum_i^n |\xi_i|^{2\alpha}d\mu(\boldsymbol{\xi})&=n\int_{\mathcal{S}^{n-1}}|\xi_1|^{2\alpha}d\mu(\boldsymbol{\xi})\\&=\frac{n\int_0^\pi|\cos(\theta)|^{2\alpha}[\sin(\theta)]^{n-2}d\theta}{\int_{\mathcal{S}^{n-1}}[\sin(\theta)]^{n-2}d\theta}\\&=\frac{n\Gamma(n/2)\Gamma(\alpha+1/2)}{\sqrt{\pi}\Gamma(n/2+\alpha)}\\&\sim\frac{\Gamma(\alpha+1/2)2^\alpha}{\sqrt{\pi}}n^{1-\alpha}.\end{aligned} \tag{41}$$

[Note that Eq. (41) is true for all $\alpha>0$.] Using Jensen's inequality we then have

$$\begin{aligned}E(\|\boldsymbol{\xi}\|_{2\alpha})&\equiv\int_{\mathcal{S}^{n-1}}\|\boldsymbol{\xi}\|_{2\alpha}d\mu\leqslant\sqrt[2\alpha]{\int_{\mathcal{S}^{n-1}}(\|\boldsymbol{\xi}\|_{2\alpha})^{2\alpha}d\mu}\\&=\sqrt[2\alpha]{\frac{\Gamma(\alpha+1/2)2^\alpha}{\sqrt{\pi}}}n^{1/2\alpha-1/2}.\end{aligned} \tag{42}$$

On the other hand, because all distributions from $\mathbb{P}_n$ fulfill the condition

$$n^{1-\alpha}\leqslant\sum_{i=1}^n p_i^\alpha\leqslant 1,\quad \alpha\geqslant 1, \tag{43}$$

we have that $E(\|\boldsymbol{\xi}\|_{2\alpha})\geqslant n^{1/2\alpha-1/2}$. Thus the mean value of $\|\boldsymbol{\xi}\|_{2\alpha}$ goes to zero as $b(n^{1/2\alpha-1/2})$ where $b=b(n,\alpha)$ is some bounded function of n. Collecting results (41) and (42) together we can recast Levy's lemma into form

$$\begin{aligned}&\mu_B(|\|\boldsymbol{\xi}\|_{2\alpha}-E(\|\boldsymbol{\xi}\|_{2\alpha})|\leqslant C)\\&\quad\geqslant 1-4e^{-\vartheta C^2 n}\\&\quad\Rightarrow\mu_B(|\|\boldsymbol{\xi}\|_{2\alpha}-E(\|\boldsymbol{\xi}\|_{2\alpha})|\leqslant\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^p)\\&\quad\geqslant 1-4\exp(-\vartheta\epsilon^2b^2n^{\{1-p[(\alpha-1)/\alpha]\}}),\end{aligned} \tag{44}$$

for some $\epsilon>0$. Note that due to symmetry of $f(\boldsymbol{\xi})$ we were allowed to exchange in Eq. (37) the averaging over the surface of $\mathcal{S}^{n-1}$ for the averaging over the positive octant $(\mathcal{S}^{n-1})^+$. Result (44) implies that for any $\epsilon>0$ and any $1<p<\alpha/(\alpha-1)$ the inequalities

$$\begin{aligned}\|\boldsymbol{\xi}\|_{2\alpha}&\geqslant E(\|\boldsymbol{\xi}\|_{2\alpha})\{1-\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^{p-1}\}\\&\geqslant E(\|\boldsymbol{\xi}\|_{2\alpha})e^{-2\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^{p-1}},\\\|\boldsymbol{\xi}\|_{2\alpha}&\leqslant E(\|\boldsymbol{\xi}\|_{2\alpha})\{1+\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^{p-1}\}\\&\leqslant E(\|\boldsymbol{\xi}\|_{2\alpha})e^{[\epsilon E(\|\boldsymbol{\xi}\|_{2\alpha})]^{p-1}},\end{aligned} \tag{45}$$

hold for almost all $\boldsymbol{\xi}\in\mathbb{P}_n$ (their Bhattacharyya measure is arbitrarily close to 1 as n increases). The fact that "well behaved" functions are at large n practically constant on al-

[4]The metric $\|\cdots\|_2$ appearing in the lemma represents the Euclidean distance inherited from $\mathbb{R}^n$ (this is also called the chordal metric). Note that $\|\boldsymbol{\xi}^{(1)}-\boldsymbol{\xi}^{(2)}\|_2=2\sin(\phi/2)\leqslant\phi$, with ϕ representing the Bhattacharyya distance.

most entire sphere is known as the *concentration measure phenomenon* [30–32]. In passing, the reader may notice that the relation (44) is a variant of Bernstein-Hoeffding's large deviation inequality [31,33].

Using now Minkowski's triangle inequality

$$\begin{aligned} |\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}-\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}| &\leqslant |\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}-E(\|\boldsymbol{\xi}\|_{2\alpha})| \\ &\quad + |\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}-E(\|\boldsymbol{\xi}\|_{2\alpha})| \\ &\leqslant 2\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^p, \end{aligned}$$

and bearing in mind Eq. (40) we can choose $\sqrt{\delta} \geqslant 2\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^p$. Consequently (for $n\geqslant 3$)

$$\begin{aligned} \frac{|\mathcal{I}_\alpha(\mathcal{P})-\mathcal{I}_\alpha(\mathcal{Q})|}{\mathcal{I}_{\alpha\,\max}} &= \frac{2\alpha}{(\alpha-1)\log_2 n}\left|\log_2\left(\frac{\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}}{\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}}\right)\right| \\ &\leqslant \frac{2\alpha}{(\alpha-1)}\left|\ln\left(\frac{e^{\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^{p-1}}}{e^{-2\epsilon[E(\|\boldsymbol{\xi}\|_{2\alpha})]^{p-1}}}\right)\right| \\ &= \frac{6\alpha\epsilon}{(\alpha-1)}[E(\|\boldsymbol{\xi}\|_{2\alpha})]^{p-1} \\ &\leqslant \frac{6\alpha}{(\alpha-1)}\left(\frac{\delta}{4}\right)^{(p-1)/2p}. \end{aligned} \tag{46}$$

Thus we see that one can always find an appropriate δ_ε for every ε, namely,

$$\delta_\varepsilon \leqslant 4\left(\frac{\varepsilon(\alpha-1)}{6\alpha}\right)^{2p/(p-1)}, \tag{47}$$

and so the observability condition (4) is satisfied in all cases for which inequalities (45) hold.

2. $0<\alpha<1$ case

A similar analysis can be performed for critical distributions in the $\alpha<1$ case. The corresponding 1-parametric families of Lesche's critical distributions are represented by arcs

$$\mathfrak{s}_i(\delta)=\left\{\xi_k(\delta)=\sqrt{\left(1-\frac{\delta}{2}\right)\delta_{ik}+\frac{\delta}{2}\left(\frac{1-\delta_{ik}}{n-1}\right)};\; k\in\hat{n};\delta\in[0,2]\right\}.$$

These arcs are identical to arcs $\gamma_i(\delta)$ depicted in Fig. 1, only the orientation is reversed. Consequently the Bhattacharyya measure is again zero in this case.

We may now ask whether there exists some generalization of Eq. (25) such that the corresponding measure μ_B is nonzero. The answer is again negative. We show now that this is a consequence of the fact that almost all distributions $\mathcal{P} \in \mathbb{P}_n$ fulfill Lesche's observability condition (4), while Bhattacharyya's measure of those distributions which do not comply with the condition (4) tends to 0 at large n.

To prove this we utilize once again Levy's lemma. In this case we make identification $f(\boldsymbol{\xi})=\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}/E(\|\boldsymbol{\xi}^{(2)}\|_{2\alpha})$. Similarly as in the previous case we must determine first the asymptotic behavior of the mean $E(\|\boldsymbol{\xi}\|_{2\alpha})$. This can be achieved by employing Jensen's inequality

$$\sqrt[2\alpha]{\frac{\Gamma(\alpha+1/2)2^\alpha}{\sqrt{\pi}}}\, n^{\frac{1}{2\alpha}-\frac{1}{2}} = \sqrt[2\alpha]{\int_{\mathcal{S}^{n-1}}(\|\boldsymbol{\xi}\|_{2\alpha})^{2\alpha}d\mu} \;\leq\; \int_{\mathcal{S}^{n-1}}\|\boldsymbol{\xi}\|_{2\alpha}d\mu, \tag{48}$$

together with the inequality

$$1\leqslant \sum_{i=1}^n p_i^\alpha \leqslant n^{1-\alpha}, \quad 0<\alpha<1. \tag{49}$$

Therefore $E(\|\boldsymbol{\xi}\|_{2\alpha})$ is unbounded at large n and it approaches infinity as $a(n^{1/2\alpha-1/2})$ [$a=a(n,\alpha)$ is some function with lower and upper bounds in n]. Employing now the estimate

$$\begin{aligned} |\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}-\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}| &\leqslant \|\boldsymbol{\xi}^{(1)}-\boldsymbol{\xi}^{(2)}\|_{2\alpha} \\ &\leqslant \|\boldsymbol{\xi}^{(1)}-\boldsymbol{\xi}^{(2)}\|_2 n^{1/2\alpha-1/2} \leqslant \sqrt{\delta}n^{1/2\alpha-1/2} \end{aligned} \tag{50}$$

(where the triangle and Hölder inequalities were successively applied) we obtain that $f(\boldsymbol{\xi})$ is $1/\underline{a}$-Lipshitz. Here $\underline{a}$ is the lower bound[5] of a. Levy's lemma then implies that

$$\mu_B\left(\left|\frac{\|\boldsymbol{\xi}\|_{2\alpha}}{E(\|\boldsymbol{\xi}\|_{2\alpha})}-1\right|\leqslant\epsilon\right)\geqslant 1-4e^{-\vartheta\underline{a}\epsilon^2 n} \tag{51}$$

for any $\epsilon>0$. Result (51) suggests that for a sufficiently small ϵ ($\epsilon\leqslant 1{,}59\ldots$) the inequality

$$e^{-2\epsilon}\leqslant 1-\epsilon\leqslant\frac{\|\boldsymbol{\xi}\|_{2\alpha}}{E(\|\boldsymbol{\xi}\|_{2\alpha})}\leqslant 1+\epsilon\leqslant e^{\epsilon} \tag{52}$$

holds for almost all $\boldsymbol{\xi}\in\mathbb{P}_n$ ($\mu_B\to 1$ as $n\to\infty$). So we again encounter the concentration of measure phenomenon—at large n almost all Bhattacharyya measure is concentrated on $\boldsymbol{\xi}$'s fulfilling the condition $\|\boldsymbol{\xi}\|_{2\alpha}\approx E(\|\boldsymbol{\xi}\|_{2\alpha})$. Using now

$$\begin{aligned} \left|\frac{\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}}{E(\|\boldsymbol{\xi}\|_{2\alpha})}-\frac{\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}}{E(\|\boldsymbol{\xi}\|_{2\alpha})}\right| &\leqslant \left|\frac{\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}}{E(\|\boldsymbol{\xi}\|_{2\alpha})}-1\right|+\left|\frac{\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}}{E(\|\boldsymbol{\xi}\|_{2\alpha})}-1\right| \\ &\leqslant 2\epsilon, \end{aligned} \tag{53}$$

and bearing in mind Eq. (50) we can set $\delta=4\epsilon^2\underline{a}^2$. Consequently (for $n\geqslant 3$)

---

[5] Clearly $\underline{a} \;\geq\; \sqrt[2\alpha]{\frac{\Gamma(\alpha+1/2)2^\alpha}{\sqrt{\pi}}} \;\geq\; 0.529\ldots$

$$\frac{|\mathcal{I}_\alpha(\mathcal{P})-\mathcal{I}_\alpha(\mathcal{Q})|}{\mathcal{I}_{\alpha\,\max}} = \frac{2\alpha}{(1-\alpha)\log_2 n}\left|\log_2\left(\frac{\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}}{\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}}\right)\right| \leqslant \frac{2\alpha}{(1-\alpha)\ln n}\left|\ln\left(\frac{e^{\epsilon}}{e^{-2\epsilon}}\right)\right| = \frac{6\epsilon\alpha}{(1-\alpha)\ln n} \leqslant \frac{3\sqrt{\delta}\alpha}{(1-\alpha)\underline{a}}. \tag{54}$$

As in the previous case we can conclude that it is always possible to find an appropriate δ_ε for every ε, namely,

$$\delta_\varepsilon \leqslant \left(\frac{\underline{a}(1-\alpha)\varepsilon}{3\alpha}\right)^2. \tag{55}$$

So the observability condition (4) is satisfied in all cases for which Eq. (52) holds. In passing, we should mention that the underlying reason behind the relations (44) and (51) lies in the fact that n-spheres $\mathcal{S}^n$ equipped with the Bhattacharyya distance ϕ_n and Haar measure μ_n form the so called *normal Levy family* [30,34]. It can be shown [30] that the concentration measure phenomenon is an inherent property of any Levy family.

The moral of this section can be summarized in the following way. Whenever one selects as the state space for Rényi entropies the space of all discrete statistics then a nonuniform continuity behavior [i.e., violation of Lesche's condition (4)] can be observed for a certain set of distribution functions in the limit of large n. We demonstrated that the cardinality of such critical distributions is of zero Bhattacharyya measure in the space of all $n\rightarrow\infty$ probability distributions. One may relate those zero measure distributions to the so called l_α-bounded distributions (i.e., distributions whose l_α norm has a nonzero lower bound for $\alpha>1$ and a finite upper bound for $\alpha<1$). This can be plainly seen from the fact that for l_α-bounded distributions the critical conditions (41) and (52) cannot be satisfied.

V. OBSERVABILITY OF RÉNYI ENTROPIES: CONTINUOUS PROBABILITY DISTRIBUTIONS AND MULTIFRACTALS

Let us briefly illustrate here that the conditions of absolutely continuous PDF's or multifractality are themselves sufficiently restrictive to ensure that the instabilities discussed in the preceding section do not occur. To see this let us consider Eqs. (9) and (12). The latter imply that for any $\mathcal{P}_n$ and $\mathcal{P}'_n$ for which the renormalized Rényi entropy exists the following identity holds:

$$\frac{|\mathcal{I}_\alpha(\mathcal{P}_n)-\mathcal{I}_\alpha(\mathcal{P}'_n)|}{\mathcal{I}_{\alpha\,\max}} = \frac{|-D(\alpha)\log_2 l+\mathcal{I}^r_\alpha+D(\alpha)\log_2 l-\mathcal{I}^{r\,\prime}_\alpha+o(1)|}{D(\alpha)\log_2(1/l)} = \frac{|\mathcal{I}^r_\alpha-\mathcal{I}^{r\,\prime}_\alpha|}{D(\alpha)\log_2(1/l)}+o(1). \tag{56}$$

Superscript r denotes renormalized quantities. Note particularly that $\mathcal{I}^r_\alpha$ are by construction finite and n (i.e., l) independent. Using the fact that $\ln x\leqslant(x-1)$ together with Hölder inequality and Eq. (39) we have for two δ-close distributions

$$|\mathcal{I}_\alpha(\mathcal{P}_n)-\mathcal{I}_\alpha(\mathcal{P}'_n)| \leqslant \frac{2\alpha k}{|1-\alpha|}\,\frac{|\,\|\boldsymbol{\xi}^{(1)}\|_{2\alpha}-\|\boldsymbol{\xi}^{(2)}\|_{2\alpha}|}{\min(\|\boldsymbol{\xi}\|_{2\alpha})} \leqslant \frac{2\alpha k}{|1-\alpha|}\,\frac{\max(\|\boldsymbol{\xi}\|_{2\alpha})}{\min(\|\boldsymbol{\xi}\|_{2\alpha})}\sqrt{\delta} \tag{57}$$

with $k=1/\ln 2$. Realizing that Eqs. (9) and (12) imply

$$\|\boldsymbol{\xi}\|_{2\alpha} = e^{[(1-\alpha)/2\alpha][-D(\alpha)\log_2 l+\mathcal{I}^r_\alpha+o(1)]}, \tag{58}$$

we can straightforwardly write that

$$|\mathcal{I}^r_\alpha-\mathcal{I}^{r\,\prime}_\alpha|+o(1) \leqslant \frac{2\alpha k}{|1-\alpha|}\sqrt{\delta}e^{[(1-\alpha)/2\alpha]|(\mathcal{I}^r_\alpha)_{\max}-(\mathcal{I}^r_\alpha)_{\min}|+o(1)} \equiv \frac{2\alpha k}{|1-\alpha|}\mathcal{B}\sqrt{\delta}. \tag{59}$$

Here $\mathcal{B}$ is an absolute constant representing the upper bound for the exponential. Gathering results (56) and (59) together we can finally write (for $n\geqslant 2$)

$$\frac{|\mathcal{I}_\alpha(\mathcal{P}_n)-\mathcal{I}_\alpha(\mathcal{P}'_n)|}{\mathcal{I}_{\alpha\,\max}} \leqslant |\mathcal{I}^r_\alpha-\mathcal{I}^{r\,\prime}_\alpha|+o(1) \leqslant \frac{2\alpha k}{|1-\alpha|}\mathcal{B}\sqrt{\delta}. \tag{60}$$

It is then clear in this case that one can easily find an appropriate δ_ε for every ε, namely,

$$\delta_\varepsilon \leqslant \left(\frac{\varepsilon(1-\alpha)}{2\alpha k\mathcal{B}}\right)^2, \tag{61}$$

represents a correct choice. So for all pairs $\mathcal{P}_n$ and $\mathcal{P}'_n$ which lead in $n\rightarrow\infty$ limit to continuous PDF's (or multifractals) the Leshe condition (4) applies. It is therefore the very definition of systems with absolutely continuous PDF's/ multifractals [incorporated in Eqs. (8) and (11)] that naturally avoids the situations with instability points confronted in the preceding section.

VI. CONCLUSIONS

In this paper we have attempted to make sense of the recent claims concerning a *total* nonobservability of Rényi's entropy. We have found that problems have arisen from uncritical use of Lesche's observability criterion. We have proved that the latter criterion, as it stands, does not rule out observability of Rényi entropies in a large class of systems, systems with a finite number of microstates or multifractals being examples. This is so because the structure of the space of distribution functions (or PDF's) over which such systems operate essentially prohibits the existence of "critical" situations considered by Lesche. In cases where such situations are encountered, namely, in systems with (countable) infinity of microstates, we argue that Lesche's uniform continuity condition is too tight to serve as a decisive criterion for the observability.

In previous works the uniform continuity condition was used to force observability upon state functions. As we have shown, it is not just unnecessary to do this but it also causes the Lesche criterion to produce incorrect results in certain cases. By identifying the probability distribution with a state variable, this has led to confusion about the observability of Rényi's entropy. Once the uniform continuity condition is dropped, we can clear up these confusing points. For this purpose we present a more intuitive concept of observability by allowing the quantity in question to have a certain amount of "critical" points provided that the cardinality of the critical points in the state space is of zero measure.

It is definitely interesting to know what the "critical regions" correspond to. In case of Rényi entropies we offer a partial reply to this question. Namely, for systems with (countable) infinity of microstates we show that the critical regions correspond to the δ-vicinity of l_α-bounded distributions. Basically such distributions correspond to (ultra) rare events which are frequently encountered, e.g., in particle detection (double β or tritium decays being examples). We have proved that the Bhattacharyya measure of these distributions must be zero. As l_α-bounded distributions are not existent in (coarse-grained) multifractals or in systems with continuous PDF's, neither in systems at thermal equilibrium, there is no *a priori* reason to disregard Rényi entropies as observable in the aforementioned instances. On the other hand, it is known that many systems undergo "statistics transitions" (stock market bidding and continuous phase transitions with their exponential-law–power-law distribution "transitions" may serve as examples). It might be also expected that in dynamical systems away from equilibrium, transitions to l_α-bounded statistics may play a relevant role. In any case, one can turn the sensitivity of Rényi entropies to a virtue as it could be used as a diagnostic instrument for an analysis of (ultra)rare-event systems, similarly as, for instance, temperature sensitivity of the susceptibility is used as a diagnostic tool in continuous phase transitions. We believe that further investigation in this direction would be of a great value.

Let us finally stress that there is also a conceptual reason why the observability in the manner of Lesche should be viewed with some hint of scepticism. This is because the observability treated in such a framework is not a unique concept. Indeed, Lesche's condition can brand a quantity as observable under one choice of state variables and as nonobservable under a different choice, even if two such choices overlap in the scope of physical situations they describe. A typical example is the Gibbs-Shannon entropy. Here, according to the above criterion, the entropy is observable if the probability distribution is chosen as the state variable [13,12]. On the other hand, if temperature and pressure are state variables then entropy develops discontinuity in any system which undergoes first order phase transition (Clausius-Clapeyron equation) and hence it is not for such systems a uniformly continuous function of state variables, and according to Eq. (1) [or Eq. (2)] it is doomed to be nonobservable. In this connection it is interesting to notice that because the parameter α plays formally the role of inverse temperature [22,35] one may expect that various limits may not commute similarly as in Gibbsian statistical physics. Namely, we may anticipate that $\lim_{\alpha\to 1}\lim_{n\to\infty} \neq \lim_{n\to\infty}\lim_{\alpha\to 1}$. In fact, Lesche [12] and other authors [13] applied the sequence of limits $\lim_{n\to\infty}\lim_{\alpha\to 1}$. In such a case they concluded that Rényi entropy of order 1 (Shannon's entropy) is observable while the rest of Rényi entropies is not (despite the fact that Rényi entropies are analytic in $\alpha \in \mathbb{R}^+$, see Ref. [20]). On the other hand, when one utilizes the "thermodynamical" order, i.e., $\lim_{\alpha\to 1}\lim_{n\to\infty}$, then also Rényi's entropy of order 1 develops instability points (this may be easily checked by noticing that unobservability argument presented in Ref. [12] is continuous in $\alpha=1$). The latter seems to support our previous comment that Shannon's entropy should not be uniformly continuous in the space of discrete distribution functions in order to account, for instance, for the first-order phase transitions.

ACKNOWLEDGMENTS

P.J. would like to gratefully acknowledge discussions with C. Isham and D. S. Brody. P.J. would also like to thank the Japanese Society for the Promotion of Science for financial support.

[1] C. Caratheodory, Math. Ann. **67**, 355 (1909); H.A. Buchdahl, Am. J. Phys. **17**, 212 (1949).
[2] R. Balian, *From Microphysics to Macrophysics, Methods and Applications of Statistical Physics* (Springer-Verlag, Heidelberg, 1991), Vol. 1.
[3] A.I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949).
[4] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970); *Selected Papers of Alfred Rényi* (Akadémia Kiado, Budapest, 1976), Vol. 2.
[5] P. Jizba, in *Decoherence and Entropy in Complex Systems*, edited by H.-T. Eltze, Lecture Notes in Physics Vol. 633 (Springer-Verlag, Berlin, 2003), p. 362.
[6] C. Tsallis, J. Stat. Phys. **52**, 479 (1988); J.H. Havrda and F. Charvat, Kybernetika **3**, 30 (1967).
[7] http://astrosun.tn.cornell.edu/staff/loredo/bayes/
[8] A. Bialas and W. Czyz, Acta Phys. Pol. B **31**, 2803 (2000); Phys. Rev. D **61**, 074021 (2000); M.K. Suleymanov, M. Sumbera, and I. Zborovsky, e-print hep-ph/0304206.
[9] C. Essex, C. Schultzky, A. Franz, and K.H. Hoffmann, Physica A **284**, 299 (2000).
[10] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman, Phys. Rev. A **33**, 1141 (1986); M.H. Jensen, L.P. Kadanoff, A. Libchaber, I. Procaccia, and J. Stavans, Phys. Rev. Lett. **55**, 2798 (1985); K. Tomita, H. Hata, T. Horita, H. Mori, and T. Morita, Prog. Theor. Phys. **80**, 963 (1988); H.G.E.

Hentschel and I. Procaccia, Physica D **8**, 435 (1983).
[11] S. Abe and Y. Okamoto, *Nonextensive Statistical Mechanics and Its Applications* (Springer-Verlag, New York, 2001).
[12] B. Lesche, J. Stat. Phys. **27**, 419 (1982).
[13] S. Abe, Phys. Rev. E **66**, 046134 (2002).
[14] C. Cachin, Ph.D. thesis, ETH Zurich, 1997, ftp://ftp.inf.ethz.ch/pub/publications/dissertations/th12187.ps.gz
[15] D. Harte, *Multifractals Theory and Applications* (Chapman and Hall/CRC, New York, 2001); M.B. Geilikman, T.V. Golubeva, and V.F. Pisarenko, Earth Planet. Sci. Lett. **99**, 127 (1990).
[16] H. Maassen and J.B.M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).
[17] See e.g., W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1991).
[18] See, e.g., F. Berthier, J.-P. Diard, and L. Pronzato, Automatica **35**, 1605 (1999).
[19] See, e.g., E.D. Sontag and Y. Wang, Syst. Control Lett. **29**, 279 (1997).
[20] P. Jizba and T. Arimitsu, Ann. Phys. (N.Y.) (to be published), e-print cond-mat/0207707.
[21] L.L. Campbell, Inf. Control. **8**, 423 (1965).
[22] H. Sakaguchi, Prog. Theor. Phys. **81**, 732 (1989).
[23] J. Feder, *Fractals* (Plenum Press, New York, 1988).
[24] S. Kullback and R. Leibler, Ann. Math. Stat. **22**, 79 (1951).
[25] T. Arimitsu and N. Arimitsu, Physica A **295**, 177 (2001); J. Phys. A **33**, L235 (2000); **34**, 673(E) (2001); Physica A **305**, 218 (2002); J. Phys.: Condens. Matter **14**, 2237 (2002); , e-print cond-mat/0306042.
[26] A. Bhattacharyya, Bull. Calcutta Math. Soc. **35**, 99 (1943); D.C. Brody and L.P. Hughston, Proc. R. Soc. London, Ser. A **457**, 1343 (2001).
[27] N. Thacker, F.J. Aherne, and P.I. Rockett, Kybernetika **34**, 363 (1997); D.C. Brody and L.P. Hughston, in *Disordered and Complex Systems*, edited by Peter Sollich, A.C.C. Coolen, L.P. Hughston, and R.F. Streater, AIP Conf. Proc. No. 553 (AIP, Melville, NY, 2001), pp. 281–288; S.S. Dragomir, e-print math.PR/0304240.
[28] W.K. Wootters, Phys. Rev. D **23**, 357 (1981); H. Araki and G. Raggio, Lett. Math. Phys. **6**, 237 (1982); D.C. Brody and L.P. Hughston, Phys. Rev. Lett. **77**, 2851 (1996); Proc. R. Soc. London, Ser. A **454**, 2445 (1998).
[29] K. Kukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press, New York, 1990); R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification* (Wiley, London, 2000); T. Kailath, IEEE Trans. Commun. **15**, 52 (1967).
[30] V.D. Milman and G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces* (Springer-Verlag, New York, 1980).
[31] K.M. Ball, *An Elementary Introduction to Modern Convex Geometry*, edited by Silvio Levy (Cambridge University Press, New York, 1997).
[32] R.J. Gardner, Bull. Am. Math. Soc. **39**, 355 (2002).
[33] D. Williams, *Probability with Martingales* (Cambridge University Press, Cambridge, 1991).
[34] M. Ledoux, *Concentration of Measure and Logarithmic Sobolev Inequalities*, edited by J. Azéma, M. Émery, M. Ledoux, and M. Yor, Lecture Notes in Mathematics Vol. 1709 (Springer, Berlin, 1999).
[35] P. Jizba and T. Arimitsu, in *Nonequilibrium and Nonlinear Dynamics in Nuclear and Other Finite Systems*, edited by Zhuxia Li, Ke Wu, Xizhen Wu, Enguang Zhao, and Fumihiko Sakata, AIP Conf. Proc. No. 597 (AIP, Melville, NY, 2001), pp. 341.