On the stability of generalized entropies



2009 J. Phys. A: Math. Theor. 42 075205

(http://iopscience.iop.org/1751-8121/42/7/075205)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **42** (2009) 075205 (9pp) doi:10.1088/1751-8113/42/7/075205

# On the stability of generalized entropies

**Xuelian Cao**[1] **and Shunlong Luo**[2]

[1] School of Mathematics and Statistics, Huazhong University of Science and Technology, 430074 Wuhan, People's Republic of China

[2] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100080 Beijing, People's Republic of China

E-mail: hustcxl@qq.com and luosl@amt.ac.cn

Received 28 October 2008, in final form 12 December 2008
Published 21 January 2009
Online at stacks.iop.org/JPhysA/42/075205

**Abstract**

By the use of a probabilistic coupling technique and generalizing a recent result of Zhang (2007 *Lett. Math. Phys.* **80** 171), we derive several sharp inequalities concerning the uniform continuity of some generalized entropies. These inequalities capture certain aspects of the interplay between the entropies and the uniform distance (variational distance), and establish the stability of the involved entropies.

PACS numbers: 02.50.Cw, 89.70.+c
Mathematics Subject Classification: 94A17, 62B10

## 1. Introduction

In information theory and statistics, various generalized entropic quantities play a fundamental role. Some prototypical examples are the Shannon entropy [9], the Rényi entropy [8] and the Tsallis entropy [11]. A basic and important issue for these entropic quantities is their stability (or robustness): if a slight change of the underlying state (probability distribution) only causes a uniformly small deviation of the corresponding entropies, then the entropies are stable. To appreciate the subtlety and to gain an intuitive insight into this problem, let us first review briefly some related results.

Following Lesche [5], we will discuss the stability of a general state functional which is often a kind of entropic quantity. The framework is as follows: for any positive integer $n$, let $\mathcal{P}_n$ be the set of $n$-dimensional probability distributions (which is here interpreted as a state space), i.e.,

$$\mathcal{P}_n = \left\{ p = (p_1, p_2, \ldots, p_n), \sum_{i=1}^{n} p_i = 1, p_i \geqslant 0, \forall\, i = 1, 2, \ldots, n \right\}.$$

Let $J : \mathcal{P}_n \to R$ be a general state functional, and put

$$J^{\max}(n) = \max_{p \in \mathcal{P}_n} J(p).$$

If, for any $\varepsilon > 0$, there exists $\delta > 0$, such that

$$|J(p) - J(q)| \leqslant \varepsilon J^{\max}(n)$$

holds for all $p, q \in \mathcal{P}_n$ satisfying $\sum_{i=1}^{n} |p_i - q_i| \leqslant \delta$, then we say that $J$ has the Lesche stability. Note that the stability is with respect to the variation of probability distributions, not with respect to $n$ (which is any but a fixed natural number). Lesche [5] showed that, for probability distributions supported on any finite set, the Shannon entropy has the desired stability, and argued that in sharp contrast, the Rényi $\alpha$-entropy for $\alpha \neq 1$ is not stable. In mathematical terms, the latter is not *uniformly* continuous with respect to the uniform distance.

Due to its importance, the stability issue has been further investigated by many authors. For example, Abe [1] showed that the Tsallis entropies [11], though trivially related to the Rényi entropies from the mathematical viewpoint, are stable for probability distributions supported on any finite set. Moreover, Abe *et al* [2], Curado and Nobre [4], and Naudts [7] discussed the stabilities for certain generalized entropies. By the use of a probabilistic coupling technique, Zhang established a sharp inequality which controls the changes of the Tsallis entropies in terms of the uniform distance, and thus characterizes the stability of the Tsallis entropies from a more precise and quantitative perspective [12]. All the above results are considered for probability distributions supported on finite sets, and for those supported on infinite sets, even the Shannon entropy is not stable [10]. This can also be heuristically seen from inequality (8) and example 1 in section 3, since when $n \to \infty$, the right-hand side of inequality (8) tends to infinite.

In this paper, we will prove several sharp inequalities which immediately provide some quantitative characterizations of the stability of generalized entropies. Our results, being *quantitative* rather than qualitative, are stronger than the above ones. The probabilistic coupling technique introduced by Zhang plays a crucial role [12]. Given two probability distributions supported on the same set, this technique couples the two probability distributions (marginals) into a joint bivariate probability distribution with a further property relating their non-coincidence probability and their uniform distance (see lemma 2).

We will only consider probability distributions with finite support and assume that the number of supporting points equals $n$. Let $X$ be a finite random variable supported on $n$ points, with the probability distribution $\{P_X(i), i = 1, 2, \ldots, n\}$. The generalized entropy we considered is

$$S_f(P_X) := \sum_i P_X(i) f(P_X(i)), \tag{1}$$

where $f : [0, 1] \to [0, \infty)$ is a nonnegative continuous function with $f(1) = 0$. It is clear that if no further conditions are imposed for the function $f$, not much can be said about the general properties of the entropy $S_f$ and its stability. We will provide some rather general and easily verified sufficient conditions for the stability of the above quantities. As simple particular consequences, we recapitulate the stabilities for the Shannon entropy and the Tsallis entropies.

The remainder of this paper is organized as follows. In section 2, we discuss some general properties of the generalized entropy $S_f$ and prepare some useful lemmas. The main results, which characterize the variation of $S_f$ in terms of the uniform distance of the underlying states, are presented in section 3. Finally, section 4 concludes with some general discussion.

## 2. Preliminary lemmas

Based on physical motivations and mathematical considerations, we will assume that the function $f : [0, 1] \to [0, \infty)$ satisfies the following conditions:

(a) $f(1) = 0$;
(b) $\widetilde{f}(x) := xf(x)$ is concave;
(c) $f(xy) \leqslant f(x) + f(y)$.

The family of such functions is rather rich and contains many examples which are of statistical and physical relevance. A lot of examples will be given in the following section after we prove theorem 1.

Some simple but important properties of the generalized entropy $S_f$ are summarized as follows.

**Lemma 1.** *Under the above conditions (a)–(c), the generalized entropy $S_f$ defined by equation (1) has the following properties:*

*(1) (Maximum entropy). It holds that*

$$S_f(P_X) \leqslant f\left(\frac{1}{n}\right),$$

*for any $P_X \in \mathcal{P}_n$. The equality holds if $P_X$ is the uniform distribution: $P_X(i) = \frac{1}{n}, i = 1, 2, \ldots, n$.*

*(2) (Concavity). Let $P_X$ and $P_Y$ be two probability distributions, then for all $\lambda \in [0, 1]$,*

$$S_f(\lambda P_X + (1-\lambda)P_Y) \geqslant \lambda S_f(P_X) + (1-\lambda)S_f(P_Y). \tag{2}$$

*(3) (Monotonicity). For any bivariate probability distribution $P_{XY}$ with marginals $P_X$ and $P_Y$, we have*

$$S_f(P_X) \leqslant S_f(P_{XY}). \tag{3}$$

*(4) (Subadditivity). For any bivariate probability distribution $P_{XY}$, it holds that*

$$S_f(P_{XY}) \leqslant S_f(P_X) + S_f(P_Y). \tag{4}$$

**Proof.** (1) By the concavity of $\widetilde{f}(x) = xf(x)$, we readily have

$$\begin{aligned} S_f(P_X) &= \sum_i \widetilde{f}(P_X(i)) \\ &= n \sum_i \frac{1}{n} \widetilde{f}(P_X(i)) \\ &\leqslant n\widetilde{f}\left(\sum_i \frac{1}{n} P_X(i)\right) \\ &= n\widetilde{f}\left(\frac{1}{n}\right) = f\left(\frac{1}{n}\right). \end{aligned}$$

(2) This is also a direct consequence of the concavity of $\widetilde{f}(x)$:

$$\begin{aligned} S_f(\lambda P_X(i) + (1-\lambda)P_Y(i)) &= \sum_i \widetilde{f}(\lambda P_X(i) + (1-\lambda)P_Y(i)) \\ &\geqslant \sum_i (\lambda \widetilde{f}(P_X(i)) + (1-\lambda)\widetilde{f}(P_Y(i))) \\ &= \lambda S_f(P_X) + (1-\lambda)S_f(P_Y). \end{aligned}$$

(3) From the concavity of $\widetilde{f}$ we obtain the subadditivity of $\widetilde{f}$, that is,

$$\widetilde{f}(x+y) \leqslant \widetilde{f}(x) + \widetilde{f}(y).$$

To see this, note that

$$\widetilde{f}(\lambda x + (1-\lambda)y) \geqslant \lambda \widetilde{f}(x) + (1-\lambda)\widetilde{f}(y), \qquad \forall\, \lambda \in [0,1].$$

Let $y = 0$ and note that $\widetilde{f}(0) = 0$, we obtain

$$\widetilde{f}(\lambda x) \geqslant \lambda \widetilde{f}(x).$$

In particular, by putting $\lambda = \frac{x}{x+y}, \frac{y}{x+y}$, we have

$$\widetilde{f}(x+y) = \frac{x}{x+y}\widetilde{f}(x+y) + \frac{y}{x+y}\widetilde{f}(x+y) \leqslant \widetilde{f}(x) + \widetilde{f}(y).$$

Consequently,

$$\begin{aligned} S_f(P_X) &= \sum_i \widetilde{f}(P_X(i)) \\ &= \sum_i \widetilde{f}\left(\sum_j P_{XY}(i,j)\right) \\ &\leqslant \sum_i \sum_j \widetilde{f}(P_{XY}(i,j)) \\ &= S_f(P_{XY}). \end{aligned}$$

(4) From condition (c), we have

$$\widetilde{f}(xy) = xyf(xy) \leqslant xy(f(x)+f(y)) = y\widetilde{f}(x) + x\widetilde{f}(y). \tag{5}$$

Let

$$P_{Y|X}(j|i) = P(Y=j|X=i) = \frac{P_{XY}(i,j)}{P_X(i)}$$

be the conditional probability. By inequality (5) and the concavity of $\widetilde{f}$, we have

$$\begin{aligned} &S_f(P_{XY}) - S_f(P_X) - S_f(P_Y) \\ &\quad= \sum_{i,j} \widetilde{f}(P_{X,Y}(i,j)) - \sum_i \widetilde{f}(P_X(i)) - \sum_j \widetilde{f}(P_Y(j)) \\ &\quad= \sum_{i,j} \widetilde{f}(P_X(i)P_{Y|X}(j|i)) - \sum_i \widetilde{f}(P_X(i)) - \sum_j \widetilde{f}\left(\sum_i P_X(i)P_{Y|X}(j|i)\right) \\ &\quad\leqslant \sum_{i,j}(P_X(i)\widetilde{f}(P_{Y|X}(j|i)) + P_{Y|X}(j|i)\widetilde{f}(P_X(i))) \\ &\qquad - \sum_i \widetilde{f}(P_X(i)) - \sum_j \sum_i P_X(i)\widetilde{f}(P_{Y|X}(j|i)) \\ &\quad= \sum_{i,j} P_{Y|X}(j|i)\widetilde{f}(P_X(i)) - \sum_i \widetilde{f}(P_X(i)) \\ &\quad= \sum_i \widetilde{f}(P_X(i)) \sum_j P_{Y|X}(j|i) - \sum_i \widetilde{f}(P_X(i)) \\ &\quad= 0. \end{aligned}$$

The proof is complete. □

For two probability distributions $P_X$ and $P_Y$, the uniform distance (also called $L^1$-distance, variational distance or the Kolmogorov distance) is defined as

$$\|P_X - P_Y\| := \frac{1}{2}\sum_i |P_X(i) - P_Y(i)|. \tag{6}$$

The fact $1/2$ is incorporated not only to ensure the mathematical convenience such that the distance lies in [0, 1], but also for some intrinsic significance because

$$\|P_X - P_Y\| = \sup_A |P_X(A) - P_Y(A)|,$$

which plays a crucial role in distinguishing probability distributions.

We will need the following key lemma which is extracted from the proof of theorem 1 in [12], and may be of independent interest.

**Lemma 2.** *For any two probability distributions $P_X$ and $P_Y$ with the same support, there exists a bivariate probability distribution $P_{XY}$ with marginals $P_X(i) = \sum_j P_{XY}(i, j)$ and $P_Y(j) = \sum_i P_{XY}(i, j)$, such that*

$$P_{XY}(i, i) = \min\{P_X(i), P_Y(i)\}, \qquad \forall\, i. \tag{7}$$

*Moreover, let X and Y be the corresponding random variables, then it holds that*

$$P(X \neq Y) = \|P_X - P_Y\|,$$

*and for $i \neq j$,*

$$P_{XY}(i, j) \leqslant \|P_X - P_Y\|.$$

## 3. Uniform variations of generalized entropies and stability

With the above preparations, we now proceed to present our main results.

Consider the generalized entropy defined by equation (1). Let $f$ satisfy the conditions (a)–(c) stated at the beginning of the previous section. For a binary probability distribution $(\varepsilon, 1 - \varepsilon)$ denote its generalized entropy by

$$S_f(\varepsilon) := \varepsilon f(\varepsilon) + (1 - \varepsilon) f(1 - \varepsilon).$$

**Theorem 1.** *Let $P_X$ and $P_Y$ be two probability distributions with the uniform distance*

$$\varepsilon := \|P_X - P_Y\| \equiv \frac{1}{2}\sum_i |P_X(i) - P_Y(i)|.$$

*Then*

$$|S_f(P_X) - S_f(P_Y)| \leqslant \varepsilon f\left(\frac{1}{n-1}\right) + S_f(\varepsilon). \tag{8}$$

**Proof.** Without loss of generality, we may assume that

$$S_f(P_X) \geqslant S_f(P_Y).$$

Let $P_{XY}$ be the joint probability distribution guaranteed by lemma 2, then by the monotonicity (3) in lemma 1, we have

$$|S_f(P_X) - S_f(P_Y)| = S_f(P_X) - S_f(P_Y) \leqslant S_f(P_{XY}) - S_f(P_Y). \tag{9}$$

Consider the random variable

$$\Theta := \begin{cases} 0, & \text{if} \quad X = Y, \\ 1, & \text{if} \quad X \neq Y, \end{cases}$$

and let $P_{\Theta Y}(\theta, j)$ be the joint probability distribution of $(\Theta, Y)$, then

$$P_{\Theta Y}(0, j) = P_{XY}(j, j), \qquad P_{\Theta Y}(1, j) = \sum_{i:i\neq j} P_{XY}(i, j). \tag{10}$$

If we put

$$q_j := \sum_{i:i\neq j} P_{XY}(i, j),$$

then

$$\sum_j q_j = \sum_{i\neq j} P_{XY}(i, j) = P(X \neq Y) = \varepsilon,$$

and

$$S_f(P_\Theta) = S_f(\varepsilon). \tag{11}$$

Now we write the right-hand side of inequality (9) as

$$S_f(P_{XY}) - S_f(P_Y) = S_f(P_{XY}) - S_f(P_{\Theta Y}) + S_f(P_{\Theta Y}) - S_f(P_Y) \tag{12}$$

and want to estimate the above two differences separately.

First, by equation (10) and noting that $\widetilde{f}(x) = xf(x)$, we have

$$\begin{aligned} S_f(P_{XY}) - S_f(P_{\Theta Y}) &= \sum_{i,j} \widetilde{f}(P_{X,Y}(i, j)) - \sum_{\theta,j} \widetilde{f}(P_{\Theta,Y}(\theta, j)) \\ &= \sum_{i,j} \widetilde{f}(P_{X,Y}(i, j)) - \sum_{\theta=0}^{1} \sum_j \widetilde{f}(P_{\Theta,Y}(\theta, j)) \\ &= \sum_{i,j} \widetilde{f}(P_{X,Y}(i, j)) - \sum_j \widetilde{f}(P_{XY}(j, j)) - \sum_j \widetilde{f}(q_j) \\ &= \sum_j \sum_{i:i\neq j} \widetilde{f}(P_{XY}(i, j)) - \sum_j \widetilde{f}(q_j) \\ &= \sum_j q_j \left( \sum_{i:i\neq j} \frac{P_{XY}(i, j)}{q_j} f(P_{XY}(i, j)) - f(q_j) \right) \\ &= \sum_j q_j \sum_{i:i\neq j} \widetilde{f}\left( \frac{P_{XY}(i, j)}{q_j} \right) \frac{f(P_{XY}(i, j)) - f(q_j)}{f\left(\frac{P_{XY}(i,j)}{q_j}\right)}. \end{aligned}$$

From condition (c), for $i \neq j$, we have

$$f(P_{XY}(i, j)) = f\left( \frac{P_{XY}(i, j)}{q_j} q_j \right) \leqslant f\left( \frac{P_{XY}(i, j)}{q_j} \right) + f(q_j),$$

which implies that (noting that $f$ is nonnegative)

$$\frac{f(P_{XY}(i, j)) - f(q_j)}{f\left(\frac{P_{XY}(i,j)}{q_j}\right)} \leqslant 1. \tag{13}$$

Consequently,

$$S_f(P_{XY}) - S_f(P_{\Theta Y}) \leqslant \sum_j q_j \sum_{i:i\neq j} \widetilde{f}\left( \frac{P_{XY}(i, j)}{q_j} \right).$$

Since each term $\sum_{i:i\neq j}\widetilde{f}\big(\frac{P_{XY}(i,j)}{q_j}\big)$ is the generalized entropy of the probability distribution $\big\{\frac{P_{XY}(i,j)}{q_j} : i = 1, 2, \ldots, n, i \neq j\big\}$ supported on $n-1$ points, by lemma 1 (1), it is bounded by $f\big(\frac{1}{n-1}\big)$. Therefore, we have

$$S_f(P_{XY}) - S_f(P_{\Theta Y}) \leqslant \sum_j q_j f\left(\frac{1}{n-1}\right) = \varepsilon f\left(\frac{1}{n-1}\right). \tag{14}$$

Second, from lemma 1 (4) (subadditivity), with $\Theta$ playing the role of $X$ there, we have

$$S_f(P_{\Theta Y}) - S_f(P_Y) \leqslant S_f(P_\Theta). \tag{15}$$

The desired inequality (8) now readily follows from the combination of inequalities (9), (12), (14) and (15). □

**Remark.** If condition (c) is replaced by the stronger inequality

$$f(xy) \leqslant y^\beta f(x) + f(y), \tag{16}$$

for some $\beta \geqslant 0$, then in the proof of theorem 1, inequality (13) can be modified as

$$\frac{f(P_{XY}(i,j)) - f(q_j)}{f\big(\frac{P_{XY}(i,j)}{q_j}\big)} \leqslant q_j^\beta,$$

and proceeding similarly as the proof of theorem 1, we readily have the stronger result

$$|S_f(P_X) - S_f(P_Y)| \leqslant \varepsilon^{\beta+1} f\left(\frac{1}{n-1}\right) + S_f(\varepsilon). \tag{17}$$

Let us consider some concrete examples.

**Example 1.** Let $f(x) = -\log x$, then the corresponding generalized entropy $S_f$ reduces to the Shannon entropy. Since the conditions (a)–(c) are met for such a $f$, we recover the stability of the Shannon entropy.

**Example 2.** Let $\phi(q)$ be continuous and satisfy $\phi(q)(1-q) > 0$, for $q \neq 1$. Put $f(x) = \frac{x^{q-1}-1}{\phi(q)}$, then $S_f$ is the nonadditive entropy. When $q > 1$, the above $f$ satisfies conditions (a)–(c), and moreover, $f(xy) - f(x) = x^{q-1} f(y)$, consequently, by the above remark, we have

$$|S_f(P_X) - S_f(P_Y)| \leqslant \varepsilon^q f\left(\frac{1}{n-1}\right) + S_f(\varepsilon) \tag{18}$$

which implies the stability of nonadditive entropies. The case $\phi(q) = 1-q$ corresponding to the case of the Tsallis entropy.

**Example 3.** More generally, for any concave function $g : [0,\infty) \to [0,\infty)$ satisfying $g(0) = 0$, let $f(x) = g(-\log x)$, and suppose that $\widetilde{f}(x) = xf(x)$ is concave, then $f$ satisfies conditions (a)–(c), and thus the corresponding generalized entropies $S_f$ satisfy inequality (8) and are stable.

**Example 4.** Let $\mu$ be any probability measure on $[0,\infty)$ and put

$$f(x) = 1 - \int_0^\infty x^t \, \mathrm{d}\mu(t).$$

Then $f$ satisfies conditions (a)–(c), and thus the corresponding entropy is stable.

If we relax condition (c), then we still have the following estimate of continuity.

**Theorem 2.** *Suppose that $f$ is nonnegative, $f(1) = 0$, and $\widetilde{f}(x) := xf(x)$ is concave. Further assume that $S_f$ is subadditive. Then*

$$|S_f(P_X) - S_f(P_Y)| \leqslant n(n-1)\varepsilon f(\varepsilon) + S_f(\varepsilon),$$

*for sufficiently small $\varepsilon := \frac{1}{2}\sum_i |P_X(i) - P_Y(i)|$.*

**Proof.** This is similar to theorem 1. First note that

$$\begin{aligned} S_f(P_{XY}) - S_f(P_{\Theta Y}) &= \sum_j \sum_{i:i\neq j} \widetilde{f}(P_{XY}(i,j)) - \sum_j \widetilde{f}(q_j) \\ &\leqslant \sum_j \sum_{i:i\neq j} \widetilde{f}(P_{XY}(i,j)). \end{aligned}$$

From the concavity of $\widetilde{f}$ and the fact that $\widetilde{f}(0) = \widetilde{f}(1) = 0$, we know $\widetilde{f}$ is increasing on $[0, \varepsilon]$ for sufficiently small $\varepsilon$. From the construction of $P_{XY}$, we know that for $i \neq j$ (lemma 2), it holds that

$$P_{XY}(i,j) \leqslant \varepsilon.$$

Therefore,

$$S_f(P_{XY}) - S_f(P_{\Theta Y}) \leqslant \sum_j \sum_{i:i\neq j} \widetilde{f}(\varepsilon) = n(n-1)\varepsilon f(\varepsilon). \tag{19}$$

Second, from the subadditivity of the generalized entropy, inequality (15) also holds. The desired result follows from the combination of inequalities (9), (12), (15), (19). □

## 4. Discussion

We have investigated some characteristic properties of generalized entropies. In particular, we have proved three inequalities which bound the variations of the entropies in terms of the uniform distance of the underlying states, and thus have demonstrated the stability of such entropies. The results are established for finite systems, and cannot be extended to the case when the support of the underlying probability distributions is infinite. On the other hand, the results can be extended to the finite-dimensional quantum case. Apart from their mathematical significance, the results may be of interest in the investigation of non-extensive thermodynamics and anomalous physical systems.

Generalized entropies considered in this paper can be reinterpreted as the $f$-divergence introduced by Csiszár [3]. Due to the importance and wide applications of this measure [6], it will also be desirable to investigate the interplay between the $f$-divergence and the uniform distance by means of the method of probabilistic coupling. This is left for further study.

## Acknowledgments

The authors are grateful to the referees for helpful suggestions. This work was supported by NSFC, Grant. No. 10771208, and by the science Fund for Creative Research Groups (10721101).

## References

[1] Abe S 2002 Stability of Tsallis entropy and instabilities of Rényi and normalized Tsallis entropies: a basis of $q$-exponential distributions *Phys. Rev.* E **66** 046134

[2] Abe S, Kaniadakis G and Scarfone A M 2004 Stabilities of generalized entropies *J. Phys. A: Math. Gen.* **37** 10513–9

[3] Csiszár I and Körner J 1981 *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Akademiai: Budapest)

[4] Curado E M F and Nobre F D 2004 On the stability of analytic entropic forms *Physica* A **335** 94–106

[5] Lesche B 1982 Instability of Rényi entropies *J. Stat. Phys.* **27** 419–22

[6] Liese F and Vaida I 2006 On divergences and information in statistics and information theory *IEEE Trans. Inf. Theory* **52** 4394–412

[7] Naudts J 2004 Continuity of a class of entropies and relative entropies *Rev. Math. Phys.* **16** 809–22

[8] Rényi A 1970 *Foundations of Probability* (Holden-Day: San Francisco)

[9] Shannon C E and Weaver W 1949 *The Mathematical Theory of Communication* (Champaign, IL: University of Illinois Press)

[10] Topsoe F 2001 Basic concepts, identities and inequalities—the toolkit of information theory *Entropy* **3** 162–90

[11] Tsallis C 1998 Possible generalization of Bolzmann–Gibbs statistics *J. Stat. Phys.* **52** 479–87

[12] Zhang Z M 2007 Uniform estimates on the Tsallis entropies *Lett. Math. Phys.* **80** 171–81