South African Statistical Journal
Vol. 54, No. 2, 163–175
https://doi.org/10.37920/sasj.2020.54.2.3

TSALLIS' ENTROPIES — AXIOMATICS, ASSOCIATED *f*-DIVERGENCES AND FISHER'S INFORMATION

Tertius de Wet[1]
Department of Statistics and Actuarial Science, University of Stellenbosch, South Africa
e-mail: *tdewet@sun.ac.za*

Ferdinand Österreicher
Department of Mathematics, University of Salzburg, Austria
e-mail: *Ferdinand.Oesterreicher@sbg.ac.at*

Maximilian Thaler
Department of Mathematics, University of Salzburg, Austria
e-mail: *Maximilian.Thaler@sbg.ac.at*

In a previous paper, de Wet and Österreicher (2016) showed how *Arimoto's extended class of entropies* generates a family of *f*-divergences leading to approximation results and finally to Fisher's information in a limiting way. In the current paper, the so-called *Tsallis class of entropies* is used in a similar fashion to generate a new family of *f*-divergences with analogous properties. The approximation properties are proved in a form which significantly generalizes the corresponding results in the above mentioned paper.

Key words: Dissimilarity measure (*f*-divergence), Fisher's information, Measure of information (entropy).

1. Introduction

In a seminal paper Havrda and Charvát (1967) introduced a new class of entropies depending on a parameter $\alpha > 0$, which has since become known as *Tsallis' class of entropies*. For a probabilty distribution $P = (t, 1-t)$, $t \in [0,1]$, with two states, these entropies are given by

$$h_\alpha(t) = \begin{cases} \frac{1}{\alpha-1} \cdot (1 - (t^\alpha + (1-t)^\alpha) & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\}, \\ -(t \ln t + (1-t)\ln(1-t)) & \text{for} \quad \alpha = 1\ . \end{cases}$$

In Section 2 we review the axiomatics of Tsallis' entropies and their basic properties. Section 3 points out an interesting result for the special case $\alpha = 2$ regarding Ahlswede's so-called "*Identification Entropy*".

Csiszár (1963) and Ali and Silvey (1966) introduced the concept $I_f(Q,P)$ of *f*-divergence which is a measure of deviation of two probability distributions Q and P, given in terms of a convex function f defined on $[0,\infty)$. A well-known example is Pearson's χ^2-deviation, where $f(u) = (u-1)^2$. Vajda (1972 and 1973) extended Pearson's χ^2-divergence to the family of the χ^α-divergences $\chi^\alpha(Q,P)$, $\alpha \in$

[1]Corresponding author.
MSC2010 subject classifications. 94A17, 62B10, 82B30.

$[1,\infty)$, associated with the functions $f(u) = |u-1|^\alpha$. In Section 4 we present the class of f-divergences associated with Tsallis' entropies, which is given in terms of the convex functions

$$\varphi_\alpha(u) = \begin{cases} \frac{1+u}{\alpha-1} \cdot \left(\frac{1+u^\alpha}{(1+u)^\alpha} - \frac{1}{2^{\alpha-1}} \right) & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\}, \\ (1+u)\ln \frac{2}{1+u} + u \ln u & \text{for} \quad \alpha = 1 \; . \end{cases}$$

Section 6 contains a general version of Theorem 3 of de Wet and Österreicher (2016), relating f-divergences to Fisher's information for a large class of functions f including these examples. The proof relies on an inequality of the form

$$\left| I_f(Q,P) - \frac{f''(1)}{2} \cdot \chi^2(Q,P) \right| \leq c(f) \cdot \chi^3(Q,P), \;\; c(f) \in (0,\infty),$$

which is derived in Section 5.

2. Axiomatics and properties of Tsallis' entropies

We introduce Tsallis' entropies quoting the axiomatic approach developed in Havrda and Charvát (1967) and Daróczy (1970). The key background reference is Csiszár's (2008) paper entitled "*Axiomatic Characterizations of Information Measures*". In the first part of the section our notation follows Daróczy's (1970) paper except for the normalization, which is taken over from Tsallis (1988).

Theorem 1 (Havrda and Charvát, 1967; Daróczy, 1970)**.** *Let α be a fixed parameter in $(0,\infty)\backslash\{1\}$ and let $\mathcal{P}_k$ be the set of all probability distributions $P_k = (p_1,\ldots,p_k)$ on a set $\Omega_k = \{\omega_1,\ldots,\omega_k\}$ with k elements, $k \in \mathbb{N}$. Furthermore, let the functions $(p_1,\ldots,p_k) \mapsto H_k^\alpha(p_1,\ldots,p_k)$ satisfy the following conditions.*

(1) Symmetry: $H_k^\alpha(p_1,\ldots,p_k)$ is invariant under the permutations of $p_1,\ldots,p_k$ for $k = 3$.

(2) Specification:

$$H_2^\alpha\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1-(\frac{1}{2})^{\alpha-1}}{\alpha-1} \; .$$

(3) α-Recursivity: For all $k \in \mathbb{N}\backslash\{1\}$ the functions H_k^α satisfy

$$\begin{aligned} &H_k^\alpha(p_1,\ldots,p_{k-1},p_k) \\ &= H_{k-1}^\alpha(p_1,\ldots,p_{k-1}+p_k) + (p_{k-1}+p_k)^\alpha \cdot H_2^\alpha\left(\frac{p_{k-1}}{p_{k-1}+p_k}, \frac{p_k}{p_{k-1}+p_k}\right). \end{aligned}$$

Then for all $k \in \mathbb{N}$ the functions $(p_1,\ldots,p_k) \mapsto H_k^\alpha(p_1,\ldots,p_k)$ have the form

$$H_k^\alpha(p_1,\ldots,p_k) = -\sum_{j=1}^k p_j \cdot \frac{p_j^{\alpha-1}-1}{\alpha-1} = \frac{1-\sum_{j=1}^k p_j^\alpha}{\alpha-1} \; .$$

Remark 1. The limiting case $\alpha = 1$ is Shannon's (1948) resp. von Neumann's (1927) entropy:

$$H_k^1(p_1,\ldots,p_k) := \lim_{\alpha\to 1} H_k^\alpha(p_1,\ldots,p_k) = -\sum_{j=1}^k p_j \cdot \ln p_j \, .$$

Looking at the first representation of H_k^α in Theorem 1 we get this as a consequence of the formula

$$\lim_{\varepsilon \to 0} \frac{x^\varepsilon - 1}{\varepsilon} = \ln x, \;\; x > 0,$$

which is well-known from elementary real analysis (see Euler, 1748, page 100). From a purely analytical point of view it is therefore quite natural to choose the approximating functions as generalizations of the logarithm and thus the functions H_k^α as generalizations of Shannon's entropy. Generalizing the logarithm in this way has also proved fruitful elsewhere in Statistical Physics as in constructing simple models for intermittency with explicitely given invariant measures (see Thaler, 2000; Venegeroles, 2015; Thaler, 1980).

The by now common notation, fixed in the following definition, refers to Tsallis (1988).

Definition 1. For $\alpha > 0$, $k \in \mathbb{N}$ and a probability distribution $(p_1, \ldots, p_k)$ the quantity $H_k^\alpha(p_1, ..., p_k)$ is called the *Tsallis entropy* of $(p_1, \ldots, p_k)$ with parameter α.

Remark 2. For $\alpha = 2$ and $k \in \mathbb{N}$

$$H_k^2(p_1, ..., p_k) = 1 - \sum_{j=1}^{k} p_j^2 = 1 - \frac{1}{k} - \sum_{j=1}^{k} \left(p_j - \frac{1}{k}\right)^2 \;,$$

which is known as the *quadratic entropy* or the *Gini-Simpson index*. The corresponding so-called *measure of concentration* is

$$H_k^2\left(\frac{1}{k}, ..., \frac{1}{k}\right) - H_k^2\,(p_1, ..., p_k) = \sum_{i=1}^{k} \left(p_i - \frac{1}{k}\right)^2 \;.$$

Brukner and Zeilinger (1999) introduced this quantity as an appropriate measure of information for the specific use in quantum measurement. For a detailed history of the case $\alpha = 2$ we refer to the paper of Österreicher and Casquilho (2018).

Remark 3. In Havrda and Charvát (1967) in addition to the assumptions given in Theorem 1 the following conditions are assumed for each $\alpha \in (0, \infty) \setminus \{1\}$:

- $H_1^\alpha(1) = 0$,
- *Symmetry* for all $k \in \mathbb{N} \setminus \{1\}$,
- *Expansibiliy*: i.e.

$$H_k^\alpha(p_1, ..., p_{k-1}, 0) = H_{k-1}^\alpha(p_1, ..., p_{k-1}) \;\; \forall\, k \in \mathbb{N} \setminus \{1\},$$

- *Continuity* of all functions $(p_1, ..., p_k) \mapsto H_k^\alpha(p_1, ..., p_k), \;\; k \in \mathbb{N} \setminus \{1\}$,

which – according to Theorem 5 of Daróczy's paper (1970) – turn out to be unnecessary.

Remark 4. In his axiomatic approach to Shannon's entropy Faddeev (1956) assumes α-recursivity for $\alpha = 1$, symmetry of the functions $H_k^1, k \in \mathbb{N} \setminus \{1\}$, continuity of H_2^1 and positivity of H_2^1 for at least one pair of argument values. In Theorem 1, the last two assumptions are compensated by the specification property.

In the sequel we state basic properties of Tsallis' entropies, switching to the notation in de Wet and Österreicher (2017):

$$S_\alpha(P_k) := H_k^\alpha(p_1, \ldots, p_k), \text{ where } P_k = (p_1, \ldots, p_k)\ (\alpha > 0,\ k \in \mathbb{N}).$$

(a) *Positivity:* $S_\alpha(P_k) \geq 0$, and $S_\alpha(P_k) = 0$ iff P_k is degenerate.

(b) *Concavity:* Let $P_{1k} = (p_{11}, ..., p_{1k}),\ P_{2k} = (p_{21}, ..., p_{2k}) \in \mathcal{P}_k$ and $\lambda \in (0, 1)$. Then

$$\lambda \cdot S_\alpha(P_{1k}) + (1 - \lambda) \cdot S_\alpha(P_{2k}) \leq S_\alpha(\lambda \cdot P_{1k} + (1 - \lambda) \cdot P_{2k}),$$

where equality holds if and only if $P_{1k} = P_{2k}$. For $\alpha \neq 1$ this is immediate from

$$\frac{d^2}{dx^2}\left(\frac{x^\alpha}{\alpha - 1}\right) = \alpha \cdot x^{\alpha-2} > 0,\ x > 0.$$

(c) *Maximum Property:* If $\frac{1}{2} \cdot P_{1k} + \frac{1}{2} \cdot P_{2k} = (\frac{1}{k}, ..., \frac{1}{k})$, then the concavity of S_α implies

$$\frac{1}{2} \cdot S_\alpha(P_{1k}) + \frac{1}{2} \cdot S_\alpha(P_{2k}) \leq S_\alpha\left(\left(\frac{1}{k}, ..., \frac{1}{k}\right)\right) = \frac{1 - (\frac{1}{k})^{\alpha-1}}{\alpha - 1},$$

where equality holds if and only if $P_{1k} = P_{2k} = (\frac{1}{k}, ..., \frac{1}{k})$.

(d) *α-Additivity:* Let $P_k = (p_1, ..., p_k) \in \mathcal{P}_k,\ Q_m = (q_1, ..., q_m) \in \mathcal{P}_m$. Then a straightforward calculation shows that

$$S_\alpha(P \times Q) = S_\alpha(P) + S_\alpha(Q) - (\alpha - 1) \cdot S_\alpha(P) \cdot S_\alpha(Q).$$

In particular,

$$S_1(P \times Q) = S_1(P) + S_1(Q)$$

and

$$S_2(P \times Q) = S_2(P) + S_2(Q) - S_2(P) \cdot S_2(Q).$$

(e) *Subadditivity*: The α-additivity obviously implies

$$S_\alpha(P \times Q) \leq S_\alpha(P) + S_\alpha(Q) \quad \forall\, \alpha \in [1, \infty).$$

Remark 5. According to Proposition 4 in de Wet and Österreicher (2017) the function $\alpha \mapsto S_\alpha(P)$ is strictly decreasing on $(0, \infty)$ for each $k \geq 2$ and each non-degenerate probability distribution $P \in \mathcal{P}_k$. Concerning the so-called *functions of uncertainty*, in terms of which Tsallis' class of entropies is defined, we refer to the remark following this Proposition.

3. Ahlswede's identfication entropy

In this section we recall a result concerning an interpretation of the quadratic entropy within the framework of optimal source coding.

For this purpose let $P = (p_1, ..., p_m)$ be a probabilty distribution on $\Omega = \{\omega_1, ..., \omega_m\}$, satisfying $p_1 \geq ... \geq p_m > 0$ and let X be a discrete random variable with values in Ω and distribution P. We consider (Ω, P, X) as the source of a q-ary prefix code

$$C_q : \omega \to \cup_{j\geq 1}\{0, \dots, q-1\}^j : \omega \mapsto c(\omega),$$

where $q \in \mathbb{N}\backslash\{1\}$. Denoting by $\|c(\omega)\|$ the length of the code word $c(\omega)$ the quantity

$$\bar{L}_{C_q}(P) = \sum_{k=1}^{m} p_k \cdot \|c(\omega_k)\|$$

is the expected value of the code word length. Furthermore, let

$$H_q(P) = -\sum_{k=1}^{m} p_k \cdot \log_q p_k$$

denote Shannon's q-ary entropy. Then, provided that C_q is an optimal code, the inequalities

$$H_q(P) \leq \bar{L}_{C_q}(P) < H_q(P) + 1$$

hold. On the other hand, let the random variable C be defined by $C = C_q(X)$. We want to use our code C_q for noiseless identification, that is, being interested in a fixed $\omega \in \Omega$ we want to know whether the actual value of C equals $c(\omega)$. To this end we successively compare the components and stop when the first different letter or identity of the words occur. Let $L(P, c(\omega))$ denote the expected number of checkings and let, finally,

$$L_{C_q}(P, P) = \sum_{k=1}^{m} p_k \cdot L(P, c(\omega_k)) .$$

Ahlswede and Cai (2006) presented in Theorem 2 of their paper the following surprising result linking $\bar{L}_{C_q}(P)$ and Shannon's q-ary entropy $H_q(P)$ with the quadratic entropy $S_2(P)$ for an optimal code C_q.

Theorem 2 (Ahlswede and Cai, 2006)**.** *Let* C_q *be a q-ary prefix code. Then the following statements are equivalent.*

(i) $\bar{L}_{\mathrm{C}_q}(P) = H_q(P)$,

(iii) $L_{C_q}(P, P) = \frac{1-\sum_{k=1}^{m} p_k^2}{1-1/q} =: H_{I,q}$.

The authors call $H_{I,q}$ the *identification entropy*. (For the missing statement (ii) we refer to the original article.)

4. Associated *f*-divergences

The aim of this section is to introduce the *f*-divergences associated with Tsallis' entropies. For convenience we include the relevant basic definitions, following de Wet and Österreicher (2016).

Let $(\Omega, \mathcal{A})$ be a measurable space, μ a σ-finite measure on $(\Omega, \mathcal{A})$ and $\mathcal{P}(\Omega, \mathcal{A})$ the set of probability distributions on $(\Omega, \mathcal{A})$ dominated by μ. Furthermore, let $\mathcal{F}$ be the set of convex functions $f : [0, \infty) \mapsto (-\infty, \infty]$ which are finite on $(0, \infty)$ and continuous on $[0, \infty)$. For $f \in \mathcal{F}$ let the function $f^* \in \mathcal{F}$ be defined by

$$f^*(u) = u \cdot f(1/u) \quad \text{for} \quad u \in (0, \infty). \tag{1}$$

Additionally we define for any $f \in \mathcal{F}$

$$0 \cdot f(v/0) := v \cdot f^*(0), \; v \geq 0, \tag{2}$$

with the usual conventions for multiplication by ∞. Then $x \cdot f^*(y/x) = y \cdot f(x/y)$ holds for all $x \in [0, \infty)$ and all $y \in [0, \infty)$.

Definition 2. The function $f^* \in \mathcal{F}$ defined by (1) is called the *$*$-conjugate function* of f. A function $f \in \mathcal{F}$ which satisfies $f^* \equiv f$ is called *$*$-self conjugate*.

Definition 3 (Csiszár, 1963; Ali and Silvey, 1966)**.** Let $P, Q \in \mathcal{P}(\Omega, \mathcal{A})$ and let $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ be the Radon-Nikodym-derivatives of P and Q with respect to μ. Then

$$I_f(Q, P) := \int f(\frac{q}{p}) \cdot p \, d\mu = \int_{\{p>0\}} f(\frac{q}{p}) \cdot p \, d\mu + f^*(0) \cdot Q(\{p = 0\})$$

is called the *f-divergence of* Q and P.

Prominent examples are the *f*-divergences $\chi^\alpha(P, Q)$ associated with the functions $f_\alpha(u) = |u - 1|^\alpha$, $\alpha \in [1, \infty)$, which generalize Pearson's χ^2-deviation $\chi^2(P, Q)$. The generalizing class was introduced and investigated by Vajda (1972 and 1973).

Remark 6. (a) $I_f(Q, P) \geq f(1)$ for all $f \in \mathcal{F}$ and all $P, Q \in \mathcal{P}(\Omega, \mathcal{A})$.
(b) If the function $f \in \mathcal{F}$ is $*$-self conjugate, the corresponding *f*-divergence is symmetric, i.e. $I_f(Q, P) = I_f(P, Q) \;\; \forall P, Q \in \mathcal{P}(\Omega, \mathcal{A})$.

In the sequel we restrict ourselves to the subset $\tilde{\mathcal{F}}_1$ of $*$-self conjugate functions $f \in \mathcal{F}$ which satisfy $f(0) < \infty$, $f(1) = 0$, and are strictly convex in $u = 1$. Functions $f \in \tilde{\mathcal{F}}_1$ turn out to be positive on $[0, \infty) \setminus \{1\}$ and the property $f(u) = u\, f(1/u)$, $u \in (0, \infty)$, entails

$$f(u) \sim f(0) \cdot u \quad \text{for} \quad u \to \infty \, . \tag{3}$$

Moreover, the following estimate holds for $f \in \tilde{\mathcal{F}}_1$:

$$f(u) \; \leq \; f(0) \cdot |u - 1|, \; u \in [0, \infty). \tag{4}$$

In fact, for $u \in [0, 1]$ convexity of f implies $f(u) \leq f(0)(1 - u)$. For $u \geq 1$ we then get

$$f(u) = u\, f(1/u) \leq u\, f(0)(1 - 1/u) = f(0)(u - 1).$$

Owing to (4) and (2),

$$I_f(Q,P) \le f(0) \cdot \int |q-p|\, d\mu = f(0) \cdot \chi^1(Q,P) \ \text{ for all } \ P,Q \in \mathcal{P}(\Omega,\mathcal{A}).$$

In particular, $I_f(Q,P)$ is finite for all $f \in \tilde{\mathcal{F}}_1$ and all $P,Q \in \mathcal{P}(\Omega,\mathcal{A})$. Note also that $f'(1) = 0$ if $f \in \tilde{\mathcal{F}}_1$ is differentiable at $u = 1$.

The functions in $\tilde{\mathcal{F}}_1$ are in one-to-one correspondence with the entropy functions in the sense of the following definition (cf. Österreicher and Vajda, 2003).

Definition 4. A continuous concave function $h : [0,1] \to \mathbb{R}$ is called an *entropy* if $h(0) = h(1) = 0$, h is symmetric with respect to $1/2$ and strictly concave in $t = 1/2$. If $P = (t, 1-t)$, $t \in [0,1]$, is a discrete probability distribution with two states, then $h(t)$ is the *entropy of P* with respect to h.

The proof of the following statement, which relies on the observation that $\varphi(0) = h(1/2)$, is straightforward.

Proposition. *If h is an entropy, then the function* $\varphi : [0,\infty) \mapsto \mathbb{R}$ *defined in terms of*

$$\varphi(u) = (1+u) \cdot \left(h\left(\frac{1}{2}\right) - h\left(\frac{u}{1+u}\right)\right), \quad u \in [0,\infty),$$

is an element of $\tilde{\mathcal{F}}_1$. *On the other hand, if* $\varphi \in \tilde{\mathcal{F}}_1$, *then the function* h *defined by*

$$h(t) = \varphi(0) - (1-t)\,\varphi\left(\frac{t}{1-t}\right), \ t \in [0,1), \ \ h(1) = 0,$$

is an entropy.

The Tsallis entropy of a probability distribution $P = (t, 1-t)$, $t \in [0,1]$, is given by $h_\alpha(t)$ where

$$h_\alpha(t) = \begin{cases} \frac{1}{\alpha-1} \cdot (1 - (t^\alpha + (1-t)^\alpha)) & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\}, \\ -(t \cdot \ln t + (1-t) \cdot \ln(1-t)) & \text{for} \quad \alpha = 1, \end{cases}$$

with

$$h_\alpha(1/2) = \begin{cases} \frac{1}{\alpha-1} \cdot \left(1 - \frac{1}{2^{\alpha-1}}\right) & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\}, \\ \ln 2 & \text{for} \quad \alpha = 1\ . \end{cases}$$

The corresponding functions in the sense of the Proposition are

$$\varphi_\alpha(u) = \begin{cases} \frac{1+u}{\alpha-1} \cdot \left(\frac{1+u^\alpha}{(1+u)^\alpha} - \frac{1}{2^{\alpha-1}}\right) & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\}, \\ (1+u)\ln\frac{2}{1+u} + u \ln u & \text{for} \quad \alpha = 1\ . \end{cases} \tag{5}$$

It is easily verified that the functions h_α are entropies in the sense of Definition 4, or, equivalently, that φ_α belongs to $\tilde{\mathcal{F}}_1$. A crucial observation hereby is that

$$\varphi''_\alpha(u) = \alpha\, \frac{1+u^{\alpha-2}}{(1+u)^{\alpha+1}} > 0 \ \ \forall\, u \in (0,\infty), \ \forall\, \alpha \in (0,\infty). \tag{6}$$

Definition 5. The *f*-divergences given in terms of the functions φ_α as defined in (5) are called *f-divergences associated with Tsallis' class of entropies.*

Remark 7. For $\alpha \in (0,\infty)\backslash\{1\}$ positivity of φ_α on $[0,\infty) \setminus \{1\}$ is obvious from the representation

$$\varphi_\alpha(u) = \frac{2}{(\alpha-1)(1+u)^{\alpha-1}}\left(\frac{1+u^\alpha}{2} - \left(\frac{1+u}{2}\right)^\alpha\right).$$

For all $\alpha \in (0,\infty)$ we have $\varphi_\alpha(1) = \varphi'_\alpha(1) = 0$ and $\varphi''_\alpha(1) = \alpha/2^\alpha$. Therefore

$$\varphi_\alpha(u) \sim \frac{\alpha}{2^{\alpha+1}} \cdot (u-1)^2 \ \text{ for } \ u \to 1$$

and thus

$$\varphi_\alpha(u) \sim \frac{\alpha}{2^{\alpha-1}} \cdot \varphi_2(u) \ \text{ for } \ u \to 1.$$

The defining formula for φ_α immediately confirms the asymptotic relation stated in (3) above:

$$\varphi_\alpha(u) \sim \varphi_\alpha(0) \cdot u \ \text{ for } \ u \to \infty\ .$$

The behaviour of $\varphi_\alpha(u)$ as $u \to \infty$ is described more precisely by observing that for all $u \in [0,\infty)$

$$\varphi_\alpha(u) > \begin{cases} \varphi_\alpha(0)(1+u) + \frac{(1+u)^{1-\alpha}}{\alpha-1} - \frac{\alpha}{\alpha-1}, & \text{if} \quad \alpha \in (0,\infty) \setminus \{1\}, \\ \varphi_1(0)(1+u) - \ln(1+u) - 1, & \text{if} \quad \alpha = 1, \end{cases}$$

where in each case the right hand side is asymptotic to φ_α at infinity in the sense that the difference between $\varphi_\alpha(u)$ and the right hand side tends to 0 as u tends to infinity.

To see this for $\alpha \in (0,\infty) \setminus \{1\}$ consider the function

$$\begin{aligned} \psi_\alpha(u) \quad &:= \quad \varphi_\alpha(u) - \left(\varphi_\alpha(0)(1+u) + \frac{(1+u)^{1-\alpha}}{\alpha-1} - \frac{\alpha}{\alpha-1}\right) \\ &= \quad \frac{1}{\alpha-1}\left(\alpha - \frac{(1+u)^\alpha - u^\alpha}{(1+u)^{\alpha-1}}\right), \ u \in [0,\infty). \end{aligned}$$

From the second term it is clear that $\lim_{u\to\infty} \psi_\alpha(u) = 0$. The first term and formula (6) immediately yield

$$\psi''_\alpha(u) = \alpha\, \frac{u^{\alpha-2}}{(1+u)^{\alpha+1}} \ > \ 0, \ u \in (0,\infty),$$

so that ψ_α is strictly convex on $[0,\infty)$. Together these properties imply that ψ_α is positive on $[0,\infty)$, completing the argument for $\alpha \in (0,\infty) \setminus \{1\}$.

For $\alpha = 1$ the assertion is easily verified by calculating the corresponding difference. Note that the shape of the asymptote for this case is in accordance with

$$-\frac{(1+u)^{1-\alpha}}{\alpha-1} + \frac{\alpha}{\alpha-1} = \frac{(1+u)^{1-\alpha}-1}{1-\alpha} + 1 \to \ln(1+u) + 1 \ \text{ for } \ \alpha \to 1.$$

Evidently, for $\alpha > 1$ the line

$$u \mapsto \varphi_\alpha(0)(1+u) - \frac{\alpha}{\alpha-1}, \ u \in [0,\infty),$$

is a lower asymptote as well.

Remark 8. The f-divergence corresponding to

$$\varphi_2(u) = \frac{1}{2}\frac{(u-1)^2}{1+u}\,, u \in [0,\infty),$$

is the symmetrized χ^2-divergence

$$I_{\varphi_2}(Q,P) = \frac{1}{2}\int_{\{p+q>0\}} \frac{(q-p)^2}{p+q}\, d\mu\,,$$

which was investigated by Puri and Vincze (1988) within the frame of f-divergences given in terms of the family of functions $\Phi_\beta \in \tilde{\mathcal{F}}_1$ with

$$\Phi_\beta(u) = \frac{1}{2}\frac{|u-1|^\beta}{(1+u)^{\beta-1}},\ \ \beta \in [1,\infty).$$

We also note that φ_3 and φ_2 only differ by a constant factor:

$$\varphi_3(u) = \frac{3}{4}\cdot \varphi_2(u),\ u \in [0,\infty).$$

The f-divergence corresponding to $\alpha = 1$ is the *symmetrized Kullback-Leibler-divergence*. According to Österreicher (2013), Section 4, the cases $\alpha = 1$ and $\alpha \in \{2,3\}$ are also elements of the *"Class of Perimeter-Type Divergences"*.

Remark 9. It is well-known that the square root of $I_{\varphi_1}(Q,P)$ and $I_{\varphi_\alpha}(Q,P), \alpha \in \{2,3\}$, are metrics in the topological sense. Recently F. Österreicher (unpublished notes) proved that this result holds true for all parameters $\alpha \in [2,3] \cup \{4\}$ and conjectures that it extends at least to all $\alpha \in [1,4]$. To prove or disprove this conjecture and, in addition, to find out which of the remaining elements of the class $I_{\varphi_\alpha}(Q,P)$, $\alpha \in (0,\infty)$, may be converted into a metric by applying a suitable scaling function is a challenging task. The interested reader may find helpful hints in the paper by Kafka, Österreicher and Vincze (1991).

Remark 10. In Definition 5, the parameter region can – in principle – be extended from $\alpha \in (0,\infty)$ to $\alpha \in \mathbb{R}\backslash\{0\}$ in terms of

$$\tilde{\varphi}_\alpha(u) = \operatorname{sgn}(\alpha)\cdot \varphi_\alpha(u),\ \ u \in [0,\infty)\ .$$

The functions $\tilde{\varphi}_\alpha$, $\alpha \in (-\infty,0)$, fulfill all properties of the elements in $\tilde{\mathcal{F}}_1$ except that $\tilde{\varphi}_\alpha(0) = \infty$. Therefore the inequality in the lemma stated below obviously does not hold in this case. For this reason we have restricted the parameter to $\alpha \in (0,\infty)$. Although he mainly stressed the case $\alpha > 0$, Tsallis had also considered the parameter space $\mathbb{R}\backslash\{0\}$ in his original paper (1988).

5. A basic inequality

Our next goal is to derive the results of Sections 3 and 4 in de Wet and Österreicher (2016) for the class of f-divergences associated with Tsallis' class of entropies. We shall in fact prove the results for a general subclass of $\tilde{\mathcal{F}}_1$ which includes these examples as well as the f-divergences considered in the above reference with the sole exception of the case $\alpha = 0$. The core analytic estimate is provided by the following elementary lemma.

Lemma. *Let $f : [0,\infty) \to \mathbb{R}$ be a continuous function which is twice differentiable in a neighborhood of $u = 1$ and has finite third derivative at $u = 1$. Then the following assertions are equivalent.*

(a) $f(1) = f'(1) = 0$ *and* $f(u) = O(u^3)$ *for* $u \to \infty$.

(b) There exists a real constant $c(f)$ *such that*

$$\left| f(u) - \frac{f''(1)}{2}\,(u-1)^2 \right| \leq c(f)\,|u-1|^3 \ \ \textit{for all } u \in [0,\infty).$$

Proof. The inequality in (b) obviously implies the conditions in (a). Assuming these define the function r by

$$r(u) := \left(f(u) - \frac{f''(1)}{2} \cdot (u-1)^2 \right) \Big/ (u-1)^3, \ \ u \in [0,\infty)\backslash\{1\},$$

which is continuous on its domain. Using de l'Hospital's rule we see that $\lim_{u\to 1} r(u) = f^{(3)}(1)/6$. Therefore r extends continously to $[0,\infty)$. The O-condition guarantees that r is bounded on $[A,\infty)$ for some $A > 0$. On $[0, A]$ the function r is bounded by continuity, and we conclude

$$c(f) := \sup_{u\in[0,\infty)} |r(u)| < \infty. \qquad \blacksquare$$

Remark 11. If "O" is replaced by "o" the supremum in the definition of $c(f)$ is in fact a maximum.

As an immediate consequence we get the following approximation result.

Theorem 3. *If $f \in \tilde{\mathcal{F}}_1$ is twice differentiable in a neighbourhood of $u = 1$ and has finite third derivative at $u = 1$, then there exists a real constant $c(f)$ such that*

$$\left| I_f(Q,P) - \frac{f''(1)}{2} \cdot \chi^2(Q,P) \right| \leq c(f) \cdot \chi^3(Q,P) \quad \forall\, P,Q \in \mathcal{P}(\Omega,\mathcal{A})\,.$$

Proof. A function f in $\tilde{\mathcal{F}}_1$ is continuous on $[0,\infty)$ and satisfies $f(1) = 0$. Differentiability at $u = 1$ implies $f'(1) = 0$. As $f(u) \sim f(0)\cdot u$ for $u \to \infty$, the condition $f(u) = o(u^3)$ for $u \to \infty$ is satisfied. Therefore the inequality in part (b) of the Lemma holds true and integration yields the result provided that $Q(\{p = 0\}) = 0$. If $Q(\{p = 0\}) > 0$, both $\chi^2(Q,P)$ and $\chi^3(Q,P)$ are infinite and thus the assertion is obviously true. $\blacksquare$

Example 1. Our functions φ_α, $\alpha \in (0,\infty)$, clearly satisfy the conditions of Theorem 3 and thus admit the asserted inequality.

Concerning the constant $c(\varphi_\alpha)$, $\alpha \in (0,\infty)$, we have a rich list of indications that $c(\varphi_\alpha) = |r_\alpha(0)|$ where r_α is the function r in the proof of the Lemma associated with φ_α. To prove or disprove this conjecture seems to be a demanding problem.

Example 2. The f-divergences based on the extended Arimoto entropies, defined by

$$\phi_\alpha(u) = \begin{cases} \frac{sgn(\alpha)\cdot(1+u)}{1-\alpha} \cdot \left(\frac{(1+u^{1/\alpha})^\alpha}{1+u} - 2^{\alpha-1} \right) & \text{for} \quad \alpha \in \mathbb{R}\backslash\{0,1\}, \\ (1+u)\ln\frac{2}{1+u} + u\ln u & \text{for} \quad \alpha = 1, \\ |u-1|/2 & \text{for} \quad \alpha = 0, \end{cases}$$

which were studied in the paper by de Wet and Österreicher (2016), also satisfy the conditions of our Theorem 3 except for $\alpha = 0$. Hence the approximation result holds for all $\alpha \in \mathbb{R} \setminus \{0\}$.

6. Fisher's information

In Section 4 of de Wet and Österreicher (2016) it is shown how the f-divergences in Example 2 are connected with Fisher's information. Theorem 3 allows us to generalize the corresponding formula. For hints to the historical background of the formula we refer to de Wet and Österreicher (2016). Interesting insights into various areas of science – especially of statistical physics – from the point of view of Fisher's information are provided in Frieden (2004).

Let $(\Omega, \mathcal{A})$ be a measurable space, μ a σ-finite measure on $(\Omega, \mathcal{A})$ and P_θ, $\theta \in \Theta$, a family of probability distributions given in terms of densities $p_\theta = \frac{dP_\theta}{d\mu}$, where Θ is an open subinterval of the real line. Furthermore, let the support

$$\Omega_+ = \{x \in \Omega : p_\theta(x) > 0\}$$

be independent of the parameter $\theta \in \Theta$, and assume that the derivatives $\frac{\partial}{\partial\theta} p_\theta(x)$ exist for $x \in \Omega_+$. Then

$$I(\theta) = \int_{\Omega_+} \left(\frac{\partial}{\partial\theta} \ln p_\theta(x) \right)^2 dP_\theta(x),\ \theta \in \Theta,$$

is called *Fisher's information*. For $\alpha \in [1, \infty)$ and $\theta, \theta_0 \in \Theta, \theta \neq \theta_0$,

$$\frac{\chi^\alpha(P_\theta, P_{\theta_0})}{|\theta - \theta_0|^\alpha} = \int_{\Omega_+} \frac{1}{p_{\theta_0}(x)^{\alpha-1}} \left| \frac{p_\theta(x) - p_{\theta_0}(x)}{\theta - \theta_0} \right|^\alpha d\mu(x).$$

If the pointwise convergence of the integrand for $\theta \to \theta_0$ is dominated by a μ-integrable function we have

$$\lim_{\theta \to \theta_0} \frac{\chi^\alpha(P_\theta, P_{\theta_0})}{|\theta - \theta_0|^\alpha} = \int_{\Omega_+} \left| \frac{\partial}{\partial\theta} \ln p_\theta(x)_{|\theta=\theta_0} \right|^\alpha dP_{\theta_0}(x). \tag{7}$$

Leaving aside the problem of formulating conditions on the family P_θ, $\theta \in \Theta$, guaranteeing (7) for $\alpha \in \{2, 3\}$ we can establish the following result, taking over the proof from de Wet and Österreicher (2016).

Theorem 4. *Let $f \in \tilde{\mathcal{F}}_1$ satisfy the conditions in Theorem 3 and let $\theta_0 \in \Theta$ be fixed. If (7) holds for $\alpha \in \{2, 3\}$ with finite limits and the function $\theta \mapsto I_f(P_\theta, P_{\theta_0})$, $\theta \in \Theta$, is differentiable in a neighbourhood of θ_0 and twice differentiable at θ_0, then*

$$\frac{d^2}{d\theta^2} I_f(P_\theta, P_{\theta_0}) \,|_{\theta=\theta_0} = f''(1) \cdot I(\theta_0).$$

In particular, the Theorem applies to the f-divergences associated with Tsallis' class of entropies, the prime subject of the present article.

7. Concluding remarks

In this paper the Tsallis class of entropies is used as the basis to generate a new class of f-divergences. It is shown that members of this class can be approximated by a χ^2-divergence and its relationship to Fisher's information is established.

Such a new class of divergences creates many possible areas for application, especially in statistical inference. Measures of divergence have long been applied to statistical inference, see e.g. Pardo

(2006). Areas of application cover inter alia estimation, including robust estimation via minimum distance estimation (see e.g. Maji et al., 2019), model selection (e.g. Cavanaugh ,2004) and goodness-of-fit tests (e.g. Bitaraf, Rezaei and Yousefzadeh, 2017). Recently divergence measures have also found application in deriving robust estimators in the area of extreme value analyses (e.g. Dierckx, Goegebeur and Guillou, 2013; Ghosh, 2017). Our interest is particularly in applying the proposed divergences to constructing goodness-of-fit tests and in deriving robust estimators in extreme value analyses. In both these application areas, the free parameter α in the Tsallis entropy-based divergence allows for flexibility to attain potentially good efficiency and robustness properties. Current and future research will focus and report on work in these areas.

References

Ahlswede, R. and Cai, A. (2006). An interpretation of identification entropy. *IEEE Transactions on Information Theory*, **52**, 4198–4207.

Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B*, **28**, 131–142.

Bitaraf, M., Rezaei, M., and Yousefzadeh, F. (2017). Goodness-of-fit tests based on Verma Kullback–Leibler information. *Communications in Statistics – Theory and Methods*, **46**, 12152–12164.

Brukner, Č. and Zeilinger, A. (1999). Operationally invariant information in quantum measurements. *Physical Review Letters*, **83**, 3354.

Cavanaugh, J. E. (2004). Criteria for linear model selection based on Kullback's symmetric divergence. *Australian & New Zealand Journal of Statistics*, **46**, 257–274.

Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **8**, 85–107.

Csiszár, I. (2008). Axiomatic characterizations of information measures. *Entropy*, **10**, 261–273.

Daróczy, Z. (1970). Generalized information functions. *Information and Control*, **16**, 36–51.

de Wet, T. and Österreicher, F. (2016). Fisher's information and the class of *f*-divergences based on extended Arimoto's entropies. *South African Statistical Journal*, **50**, 43–64.

de Wet, T. and Österreicher, F. (2017). A note on extended Arimoto's entropies. *South African Statistical Journal*, **51**, 285–294.

Dierckx, G., Goegebeur, Y., and Guillou, A. (2013). An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *Journal of Multivariate Analysis*, **121**, 70–86.

Euler, L. (1748). *Introduction to Analysis of the Infinite. Book 1*. Springer, New York. Translated by J. D. Blanton (1988).

Faddeev, D. K. (1956). On the concept of entropy of a finite probabilistic scheme (in Russian). *Russian Mathematical Surveys*, **11**, 227–231.

Frieden, B. R. (2004). *Science from Fisher Information: A Unification*. Cambridge University Press, Cambridge.

GHOSH, A. (2017). Divergence based robust estimation of the tail index through an exponential regression model. *Statistical Methods & Applications*, **26**, 181–213.

HAVRDA, J. AND CHARVÁT, F. (1967). Quantification method of classification processes. Concept of structural α-entropy. *Kybernetika*, **3**, 30–35.

KAFKA, P., ÖSTERREICHER, F., AND VINCZE, I. (1991). On powers of *f*-divergences defining a distance. *Studia Scientiarum Mathematicarum Hungarica*, **26**, 415–422.

MAJI, A., GHOSH, A., BASU, A., AND PARDO, L. (2019). Robust statistical inference based on the C-divergence family. *Annals of the Institute of Statistical Mathematics*, **71**, 1289–1322.

ÖSTERREICHER, F. (2013). Distances based on the perimeter of the risk set of a testing problem. *Austrian Journal of Statistics*, **42**, 3–19.

ÖSTERREICHER, F. AND CASQUILHO, J. A. P. (2018). On the Gini–Simpson index and its generalisation – a historic note. *South African Statistical Journal*, **52**, 129–137.

ÖSTERREICHER, F. AND VAJDA, I. (2003). A new class of metric divergences on probability spaces and its applicability in statistics. *Annals of the Institute of Statistical Mathematics*, **55**, 639–653.

PARDO, L. (2006). *Statistical Inference Based on Divergence Measures*. Chapman & Hall, Boca Raton, Florida.

PURI, M. L. AND VINCZE, I. (1988). Information and mathematical statistics. *In Proceedings of the 4th Conference on Asymptotic Statistics*. Charles University, Prague, 447–456.

SHANNON, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379–423 and 623–656.

THALER, M. (1980). Estimates of the invariant densities of endomorphisms with indifferent fixed points. *Israel Journal of Mathematics*, **37**, 303–314.

THALER, M. (2000). The asymptotics of the Perron-Frobenius operator of a class of interval maps preserving infinite measures. *Studia Mathematica*, **143**, 103–119.

TSALLIS, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, **52**, 479–487.

VAJDA, I. (1972). On the *f*-divergence and singularity of probability measures. *Periodica Mathematica Hungarica*, **2**, 223–234.

VAJDA, I. (1973). χ^α-divergence and generalized Fisher's information. *In Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*. Prague, 873–886.

VENEGEROLES, R. (2015). Exact invariant measures: How the strength of measure settles the intensity of chaos. *Physical Review E*, **91**, 062914.

VON NEUMANN, J. (1927). Thermodynamik quantenmechanischer Gesamtheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 273–291.

Manuscript received 2020-01-21, revised 2020-06-14, accepted 2020-06-20.