A NOTE ON EXTENDED ARIMOTO'S ENTROPIES

T. de Wet [1]
University of Stellenbosch, South Africa
e-mail: *tdewet@sun.ac.za*

F. Österreicher
University of Salzburg, Austria

Key words: Dissimilarity measure (f-divergence), Measure of information (Entropy).

Abstract: Arimoto (1971) introduced, among other things, a class of entropies for probability distributions on any finite set of elements, which includes Shannon's entropy (1948) as a special case. Restricted to the set $\mathscr{P}_2$ of probability distributions $P = (t, 1-t)$ on a set with only two elements his class of entropies is given in terms of

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

As Vajda (2009) extended a certain class of Csiszár's f-divergences, which are closely related to Arimoto's entropies to all parameters $\alpha \in \mathbb{R}$, the authors of this note generalised the special case of Arimoto's entropies for probability distributions $P \in \mathscr{P}_2$ to all $\alpha \in \mathbb{R}$ (De Wet and Österreicher, 2016). It turns out that these entropies are given for negative $\alpha = -k,\ k \in (0,\infty)$, by

$$h_{-k}(t) = \frac{1}{1+k}\frac{t(1-t)}{[t^{1/k}+(1-t)^{1/k}]^k}\ ,\ \ t \in [0,1].$$

In the present note their extension to probability distributions $P \in \mathscr{P}_n$ for $n \geq 2$ is investigated. In addition, a comparison of Arimoto's extended class of entropies with Rényi's and Tsallis' classes, is given. For the axiomatic characterization of the latter two classes of entropies we refer to the survey paper by Csiszár (2008).

1. Main Part

Let $\mathscr{P}_n$ be the set of all probability distributions $P = (p_1, ..., p_n)$ on a set $\Omega = \{\omega_1, ..., \omega_n\}$ with n elements, $n \in \mathbb{N}\backslash\{1\}$. All entropies H considered in this paper (a) are symmetric, i.e. they depend on the probabilities $p_i,\ i \in \{1, ..., n\}$, however, not on their order; more precisely, it holds

$$H((p_{\pi_1}, ..., p_{\pi_n})) = H((p_1, ..., p_n))$$

[1] Corresponding author.

AMS: 94A17, 62B10

for all permutations $(\pi_1, ..., \pi_n)$ of $\{1, ..., n\}$ and all $n \geq 2$, and (b) attain their maximal values $H(P)$ if and only if P is a uniform distribution $P_n = (\frac{1}{n}, ..., \frac{1}{n}) \in \mathscr{P}_n,\ n \geq 2$.

Arimoto's class of entropies is defined in terms of the so-called *functions of uncertainty*

$$\psi_\alpha(s) = \begin{cases} \frac{1-s^{1-\alpha}}{1-\alpha} & \text{for} \quad \alpha \in [0,\infty)\backslash\{1\} \\ -\ln s & \text{for} \quad \alpha = 1 \end{cases} \quad ,\ s \in [0,1].$$

In view of Arimoto (1971, p. 183), Vajda (1989, p. 322), and Csiszár (1974, Section 4), the following corresponding definition seems appropriate.

Definition 1 The function $\psi : [0,1] \mapsto \bar{\mathbb{R}}$ is called a *function of uncertainty* if it is strictly monotone decreasing and satisfies $\psi(1) = 0$ and, consequently, $\psi(s) > 0\ \forall\ s \in [0,1)$. [2]

Furthermore, let $P = (p_1, ..., p_n),\ Q = (q_1, ..., q_n) \in \mathscr{P}_n$ and

$$H_{\psi_\alpha}(Q,P) = \sum_{j=1}^{n} p_j \cdot \psi_\alpha(q_j)\ .$$

Then

$$H_\alpha(P) := H_{\psi_\alpha}(Q^*_\alpha, P) = \min(H_{\psi_\alpha}(Q,P),\ Q \in \mathscr{P}_n)$$

and the optimal distribution Q^*_α is given by

$$Q^*_\alpha = (s_\alpha(1), ..., s_\alpha(n)) \quad \text{and} \quad s_\alpha(j) = \frac{p_j^{1/\alpha}}{\sum_{i=1}^{n} p_i^{1/\alpha}},\ \ j \in \{1, ..., n\},\ \alpha \in (0,\infty).$$

Arimoto's class of entropies is consequently given by

$$H_\alpha(P) = \begin{cases} \frac{1}{1-\alpha}[1 - (\sum_{j=1}^{n} p_j^{1/\alpha})^\alpha] & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\} \\ -\sum_{j=1}^{n} p_j \ln p_j & \text{for} \quad \alpha = 1 \\ 1 - p_{\max} & \text{for} \quad \alpha = 0\ , \end{cases}$$

where $p_{\max} = \max(p_j,\ j \in \{1, ..., n\})$.

Special cases:

- For $\alpha = 1$ the optimal distribution is $Q^*_1 = P$ and $H_1(P)$ is Shannon's entropy (1948).
- For $\alpha = 0$ let $\Omega_{\max} = \{\omega_j : p_j = p_{\max},\ j \in \{1, ..., n\}\}$. Then every probability distribution Q^*_0 with support $\Omega_{\max}$ is optimal.

Both special cases are limiting cases of Arimoto's class of entropies.

[2] Note that the functions of uncertainty for Arimoto's extended class of entropies are also convex; that the convexity of the uncertainty functions for Rényi's class of entropies, however, is only given for the parameters $\alpha \in [1, \alpha_R],\ \alpha_R \simeq 1.53$, and that the convexity of the functions of uncertainty for Tsallis' class of entropies is only given for the parameters $\alpha \in [1,2] \cup [3,5]$.

Remark 1 As already mentioned in the abstract, the authors' extension of Arimoto's class of entropies (1971) for probability distributions $P \in \mathscr{P}_2$ to all $\alpha \in \mathbb{R}$ (De Wet and Österreicher, 2016) was stimulated by Vajda's (2009) extension of a certain class $\varphi_\alpha,\ \alpha \in [0,\infty)$, of Csiszár's f-divergences to all $\alpha \in \mathbb{R}$. The latter was studied in a series of papers, including those of 1993 and 2003, by Vajda and Österreicher. For Csiszár's seminal paper (1963) on f-divergences and further results on this topic we also refer to the pertinent literature.

Theorem 1 Let $k \in (0,\infty)$ and $P = (p_1,...,p_n) \in \mathscr{P}_n$ satisfying $p_j > 0\ \forall\ j \in \{1,...,n\},\ n \geq 2$. Furthermore, let

$$\psi_{-k}(s) = \frac{(1-s)^{1+k}}{1+k},\ s \in [0,1],$$

be the functions of uncertainty and

$$H_{-k}(P) := H_{\psi_{-k}}(Q^*_{-k},P) = \min(H_{\psi_{-k}}(Q,P),\ Q \in \mathscr{P}_n).$$

Then the optimal distribution

$$Q^*_{-k} = (s_{-k}(1),...,s_{-k}(n))$$

is given by

$$s_{-k}(j) = 1-(n-1)\cdot\frac{p_j^{-1/k}}{\sum_{i=1}^n p_i^{-1/k}},\quad j \in \{1,...,n\},$$

and Arimoto's extended entropy is for $\alpha = -k,\ k \in (0,\infty)$,

$$H_{-k}(P) = \frac{(n-1)^{1+k}}{1+k}\cdot\left[\sum_{j=1}^n p_j^{-1/k}\right]^{-k} = \frac{\prod_{i=1}^n p_i}{[\sum_{j=1}^n[\prod_{i=1,i\neq j}^n p_i]^{1/k}]^k}\ .$$

Example For the special case $k = 1$, which is related to the squared Puri-Vincze distance ($4\cdot\varphi_{-1}(u) = \frac{1}{2}\frac{(u-1)^2}{1+u}$, (Puri and Vincze, 1988)), it holds

$$H_{-1}(P) = \frac{(n-1)^2}{2}\cdot\left[\sum_{j=1}^n p_j^{-1}\right]^{-1} = \frac{(n-1)^2}{2}\cdot\frac{\prod_{i=1}^n p_i}{\sum_{j=1}^n\prod_{i=1,i\neq j}^n p_i}\ .$$

Proof. Applying the method of Lagrange multipliers yields

$$\frac{\partial}{\partial\tilde{p}_j}[\sum_{i=1}^n p_i\cdot\frac{(1-\tilde{p}_i)^{1+k}}{1+k}+\lambda\cdot\sum_{i=1}^n(\tilde{p}_i-1)] = -p_j\cdot(1-\tilde{p}_j)^k+\lambda = 0$$

and hence

$$\tilde{p}_j = 1-\lambda^{1/k}\cdot p_j^{-1/k}\ .$$

Summation of the latter gives

$$n-\lambda^{1/k}\cdot\sum_{j=1}^n p_j^{-1/k} = 1 \quad\text{and thus}\quad \lambda^{1/k} = \frac{n-1}{\sum_{j=1}^n p_j^{-1/k}}\ .$$

Hence

$$s_{-k}(j) = \tilde{p}_j = 1 - (n-1) \cdot \frac{p_j^{-1/k}}{\sum_{i=1}^n p_i^{-1/k}}, \quad j \in \{1, ..., n\}.$$

Therefore, for Arimoto's extended entropy it holds that

$$\begin{aligned} \frac{1+k}{(n-1)^{1+k}} \cdot H_{-k}(P) &= \frac{1}{(n-1)^{1+k}} \sum_{j=1}^n p_j \cdot \left[1 - (1 - (n-1) \frac{p_j^{-1/k}}{\sum_{i=1}^n p_i^{-1/k}}) \right]^{1+k} \\ &= \sum_{j=1}^n p_j \cdot \left[\frac{p_j^{-1/k}}{\sum_{i=1}^n p_i^{-1/k}} \right]^{1+k} = \frac{\sum_{j=1}^n p_j \cdot p_j^{-\frac{1+k}{k}}}{[\sum_{i=1}^n p_i^{-1/k}]^{1+k}} \\ &= \frac{\sum_{j=1}^n p_j^{-1/k}}{[\sum_{i=1}^n p_i^{-1/k}]^{1+k}} = [\sum_{i=1}^n p_i^{-1/k}]^{-k} . \end{aligned}$$

Finally, because of

$$\sum_{j=1}^n \frac{1}{p_j^{1/k}} = \frac{\sum_{j=1}^n [\prod_{i=1, i \neq j}^n p_i]^{1/k}}{[\prod_{i=1}^n p_i]^{1/k}}$$

and thus

$$\left[\sum_{j=1}^n p_j^{-1/k} \right]^{-k} = \frac{\prod_{i=1}^n p_i}{[\sum_{j=1}^n [\prod_{i=1, i \neq j}^n p_i]^{1/k}]^k} ,$$

the former yields the alternative representation

$$H_{-k}(P) = \frac{(n-1)^{1+k}}{1+k} \cdot \frac{\prod_{i=1}^n p_i}{[\sum_{j=1}^n [\prod_{i=1, i \neq j}^n p_i]^{1/k}]^k} .$$

■

Special Case For $n = 2$ and $P = (t, 1-t) \in \mathscr{P}_2$ this, indeed, yields

$$h_{-k}(t) = \frac{1}{1+k} \cdot \frac{t \cdot (1-t)}{[t^{1/k} + (1-t)^{1/k}]^k}$$

with the limiting case $h_0(t) = \min(t, 1-t) = 1 - \max(t, 1-t)$.

The following general property is easily verified.

General Property: Let $P_n = (\frac{1}{n}, ..., \frac{1}{n}) \in \mathscr{P}_n$ be the uniform distribution on a set $\Omega = \{\omega_1, ..., \omega_n\}$ with $n \geq 2$ elements. Then for every $\alpha \in \mathbb{R}$ the corresponding optimal distribution $Q_\alpha^* = Q_\alpha^*(P_n)$ equals P_n and its entropy is

$$H_\alpha(P_n) \equiv \psi_\alpha \left(\frac{1}{n} \right).$$

Proposition 1 It is easily seen that the class of functions

$$s \mapsto \psi_\alpha(s), \; s \in [0,1], \quad \alpha \in \mathbb{R},$$

of uncertainty is for all $s \in (0,1)$ strictly monotone increasing on $\mathbb{R}$. Therefore the class $H_\alpha(P),\ \alpha \in \mathbb{R}$, of entropies is also strictly monotone increasing on $[0,\infty)$ for all non-degenerate probability distributions $P \in \mathscr{P}_n$ and all $n \geq 2$ and on $(-\infty,0)$ for all probability distributions $P \in \mathscr{P}_n$ satisfying $p_{\min} > 0$ and all $n \geq 2$.

Remark 2 It is also easily seen and a well-known fact that Arimoto's entropy for $\alpha = 0$ is also the limiting case

$$\lim_{\alpha \searrow 0} H_\alpha(P) = H_0(P)$$

for all distributions $P \in \mathscr{P}_n$ and all $n \geq 2$. Because of this fact and the validity of $\lim_{k \searrow 0} H_{-k}(P) = H_0(P)$ for distributions $P = (t, 1-t) \in \mathscr{P}_2$ one might be tempted to assume that this holds true for all distributions $P \in \mathscr{P}_n$ and $n \geq 2$. However, this is not necessarily the case, as can be seen from the following Proposition.

Proposition 2 Let, for $n > 2$,

$$p_{\min} = \min(p_j,\ j \in \{1,...,n\}) > 0 \ .$$

Then

$$\lim_{k \searrow 0} H_{-k}(P) = (n-1) \cdot p_{\min} \leq 1 - p_{\max} \ ,$$

where equality holds true if and only if P is a permutation of a distribution $\tilde{P}_n = (\frac{1-t}{n-1}, ..., \frac{1-t}{n-1}, t),\ t \in \left[\frac{1}{n}, 1\right)$.

Proof. Let $\beta = \frac{1}{k}$ and $x_j = \frac{1}{p_j},\ j \in \{1,...,n\}$. Then, since

$$\lim_{\beta \nearrow \infty} (\sum_{j=1}^{n} x_j^\beta)^{1/\beta} = \max(x_j : j \in \{1,...,n\}),$$

it holds that

$$\begin{aligned} H_{-k}(P) &= \frac{(n-1)^{1+k}}{1+k} \cdot \frac{1}{(\sum_{j=1}^{n} (\frac{1}{p_j})^{1/k})^k} = \frac{(n-1)^{1+1/\beta}}{1+1/\beta} \cdot \frac{1}{(\sum_{j=1}^{n} (\frac{1}{p_j})^\beta)^{1/\beta}} \\ &\nearrow (n-1) \cdot \frac{1}{\max(\frac{1}{p_j})} = (n-1) \cdot \frac{1}{1/p_{\min}} = (n-1) \cdot p_{\min} \ . \end{aligned}$$

Let $P = (p_1, ..., p_{n-1}, p_n)$ with $p_n = p_{\max}$. Then

$$(n-1) \cdot p_{\min} \leq \sum_{j=1}^{n-1} p_j = 1 - p_n = 1 - p_{\max} \ ,$$

where equality holds if and only if $P = \tilde{P}_n = (\frac{1-t}{n-1}, ..., \frac{1-t}{n-1}, t),\ t = p_{\max} \in \left[\frac{1}{n}, 1\right)$. ■

Note that, obviously, Arimoto's extended entropies $H_{-k}(P)$ for $k \in (0,\infty)$ and probability distributions $P = (p_1, ..., p_n) \in \mathscr{P}_n$ and $n > 2$ can only be defined if $p_{\min} > 0$ [3]. Hence the property

$$H((p_1, ..., p_{n-1}, 0)) = H((p_1, ..., p_{n-1})),$$

[3] Note that the limit of $H_{-k}(P)$ for $p_{\min} \searrow 0$ vanishes.

the so-called *expansibility*, which is valid for nearly every measure of entropy – including Arimoto's entropies for all $\alpha \in [0,\infty)$ – cannot be fulfilled for the class of entropies $H_{-k}(P)$ for $k \in (0,\infty)$.

2. Comparison of Arimoto's Class of Entropies with Two Further Classes

As for Arimoto's class of entropies the special case of Shannon's entropy is also included in the following two classes of entropies for the parameter $\alpha = 1$ in a limiting way. To begin with, we state it therefore in the terminology of these classes

$$R_1(P) = S_1(P) = H_1(P) = -\sum_{j=1}^{n} p_j \ln p_j \ ,$$

in order to avoid repeating its definition within that of the classes $R_\alpha(P),\ \alpha \in [0,\infty]$, and $S_\alpha(P),\ \alpha \in [0,\infty)$.

Rényi's Class of Entropies of Order $\alpha \in [0,\infty]$, is given by (Schützenberger, 1954; Rényi, 1961):

$$R_\alpha(P) = \begin{cases} \frac{1}{1-\alpha}\ln(\sum_{j=1}^{n} p_j^\alpha) & \text{for} \quad \alpha \in [0,\infty)\backslash\{1\} \\ -\ln(p_{\max}) & \text{for} \quad \alpha = \infty \end{cases}$$

and for the uniform distribution $P_n = (\frac{1}{n},\ldots,\frac{1}{n})$ therefore $R_\alpha(P_n) \equiv -\ln(\frac{1}{n})$.

Proposition 3 $R_\alpha(P)$ is strictly monotone decreasing on $[0,\infty]$ for all non-degenerate probability distributions $P \in \mathscr{P}_n,\ P \neq P_n = (\frac{1}{n},\ldots,\frac{1}{n})$, and $n \geq 2$.

Proof. Let $\alpha \in (0,\infty)\backslash\{1\}$ and

$$\rho_\alpha(P) = (1-\alpha)^2 \cdot \frac{\partial}{\partial\alpha}\frac{\ln(\sum_{j=1}^{n} p_j^\alpha)}{1-\alpha}.$$

Let $Q = (q_1,\ldots,q_n)$ and $R = (r_1,\ldots,r_n) \in \mathscr{P}_n$ be two probability distributions, furthermore let

$$I(Q \parallel R) = \sum_{j=1}^{n} q_j \ln(\frac{q_j}{r_j})$$

be the I-divergence of Q and R and, finally, let

$$Q_\alpha = (q_\alpha(1),\ldots,q_\alpha(n)) \quad \text{with} \quad q_\alpha(j) = \frac{p_j^\alpha}{\sum_{i=1}^{n} p_i^\alpha},\ j \in \{1,\ldots,n\}.$$

Then it turns out that

$$\rho_\alpha(P) = -I(Q_\alpha \parallel P) \begin{cases} < 0 & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\} \\ = 0 & \text{for} \quad \alpha = 1\ , \end{cases}$$

the latter since $I(Q \parallel R) > 0$ if and only if $R \neq Q$. ∎

Remark 3 Let $\alpha \in [0,\infty)\backslash\{1\}$, $P=(p_1,...,p_n) \in \mathscr{P}_n$ and $Q_\alpha \in \mathscr{P}_n$ as above. (A) Then, as can be easily seen, the difference $R_\alpha(P)-H_1(P)$ equals the $(1-\alpha)$-th part of the I-divergence of the distributions P and Q_α, i.e. it holds

$$R_\alpha(P)=H_1(P)+\frac{1}{1-\alpha}I(P\parallel Q_\alpha)\ .$$

(B) Now let, in addition, $R=(r_1,...,r_n) \in \mathscr{P}_n$, whereby we assume for simplicity $r_{\min}>0$. Then the term

$$D_\alpha(P\parallel R)=\frac{1}{\alpha-1}\ln(\sum_{j=1}^{n} p_j^\alpha \cdot r_j^{1-\alpha})\ ,\ \ \alpha \in (0,\infty)\backslash\{1\},$$

called *Rényi's divergence of order* α *of P and R*, is an extension of its limiting case $\lim_{\alpha\to 1} D_\alpha(P\parallel R)=I(P\parallel R)$. Rényi's entropy and Rényi's divergence of order α obey the following relationship

$$R_\alpha(P)=H_1(P_n)-D_\alpha(P\parallel P_n)\ .$$

Remark 3 is the result of a very valuable suggestion by the Referee whom we also owe the reference to the paper by Van Erven and Harremoës (2014).

Already Arimoto (1971) stated that in view of $\ln u<u-1,\ u\in(0,\infty)\backslash\{1\}$, the following sequences of inequalities hold for all non-degenerate probability distributions $P\in\mathscr{P}_n,\ P\neq P_n$, and all $n\geq 2$

$$\begin{array}{llll} H_1(P)<R_\alpha(P)<H_{1/\alpha}(P) & \text{for} & \alpha\in(0,1) & \text{and} \\ H_1(P)>R_\alpha(P)>H_{1/\alpha}(P) & \text{for} & \alpha\in(1,\infty]. & \end{array}$$

Remark 4 For the special case $n=2$ Rényi's class of entropies of a distribution $P_t=(t,1-t)\in\mathscr{P}_2$ is given by

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

Furthermore let, (see Österreicher and Vajda, 1993, section II),

$$\begin{aligned} \psi_\alpha^R(s) :=& R_\alpha(s)+(1-s)\cdot R'_\alpha(s) \\ =& \begin{cases} \frac{1}{1-\alpha}(\alpha(\frac{s^{\alpha-1}}{s^\alpha+(1-s)^\alpha}-1)+\ln(s^\alpha+(1-s)^\alpha) & \text{for} \quad \alpha\in[0,\infty)\backslash\{1\} \\ -\ln s & \text{for} \quad \alpha=1 \\ -\ln(\max(s,1-s))+sgn(\frac{1}{2}-s)\cdot\frac{1-s}{\max(s,1-s)} & \text{for} \quad \alpha=\infty\ . \end{cases} \end{aligned}$$

Rényi's entropy $R_\alpha(P_t)$ is, as a function of $t\in[0,1]$, concave on the whole interval $[0,1]$ if and only if $\alpha\in(0,2]$. Therefore, for these and only these α the functions $s\mapsto\psi_\alpha^R(s)$ are strictly monotone decreasing on the whole interval $[0,1]$ and are, because of $\psi_\alpha^R(1)=1$, functions of uncertainty in the sense of Definition 1. As a matter of fact, the functions ψ_α^R are the corresponding functions of uncertainty for all $\alpha\in(0,2]$ and the optimal distribution for P_t equals $Q_\alpha^*(P_t)\equiv P_t$. Nonetheless, the family of linear functions

$$t\mapsto\bar{h}_\alpha(t,s)=t\cdot\psi_\alpha^R(s)+(1-t)\cdot\psi_\alpha^R(1-s),\ \ s\in[0,1],$$

is the family of supporting resp. generating lines of the entropy function $t\mapsto\bar{h}_\alpha(t,t)=R_\alpha(P_t),\ t\in[0,1]$ also for the parameters $\alpha\in\{0\}\cup(2,\infty]$.

Tsallis' Class of Entropies of Order $\alpha \in [0,\infty)$ is given by (Havrda and Charvát, 1967; Daróczy, 1970; Tsallis, 1988):

$$S_\alpha(P) = \frac{1}{\alpha - 1}\left[1 - \sum_{j=1}^{n} p_j^\alpha\right] \text{ for } \alpha \in [0,\infty)\backslash\{1\}$$

and for the uniform distribution $P_n = (\frac{1}{n},...,\frac{1}{n})$, therefore

$$S_\alpha(P_n) = \begin{cases} \frac{1}{\alpha-1}\cdot\left[1-(\frac{1}{n})^{\alpha-1}\right] & \text{for} \quad \alpha \in [0,\infty)\backslash\{1\} \\ -\ln(\frac{1}{n}) & \text{for} \quad \alpha = 1\ . \end{cases}$$

Proposition 4 $S_\alpha(P)$ is strictly monotone decreasing on $[0,\infty)$ for all non-degenerate probability distributions $P \in \mathscr{P}_n$ and all $n \geq 2$.

Proof. Let $\alpha \in (0,\infty)\backslash\{1\}$ and

$$\sigma_\alpha(P) = (\alpha-1)^2 \cdot \frac{\partial}{\partial\alpha}\frac{1-\sum_{j=1}^n p_j^\alpha}{\alpha-1}\ .$$

Then applying $\ln u < u-1,\ u \in (0,\infty)\backslash\{1\}$, yields

$$\begin{aligned} \sigma_\alpha(P) &= \sum_{j=1}^{n} p_j^\alpha \cdot (1+\ln p_j^{1-\alpha}) - 1 \\ &< \sum_{j=1}^{n} p_j^\alpha \cdot p_j^{1-\alpha} - 1 = \sum_{j=1}^{n} p_j - 1 = 0. \end{aligned}$$

■

Since the functions

$$\delta_\alpha(u) = \frac{1-u^{1/\alpha}}{1-1/\alpha} - \frac{1-u}{\alpha-1},\ u \in (0,\infty), \quad \alpha \in (0,\infty)\backslash\{1\},$$

satisfy $\delta_\alpha(1) = \delta'_\alpha(1) = 0$ and $\delta''_\alpha(u) = \frac{u^{1/\alpha-2}}{\alpha} > 0$ it holds $\delta_\alpha(u) > 0\ \forall\ u \in (0,\infty)\backslash\{1\}$ and, by inserting for $u = \sum_{j=1}^n p_j^\alpha$, therefore

$$S_\alpha(P) < H_{1/\alpha}(P) \quad \text{for} \quad \alpha \in (0,\infty)\backslash\{1\}.$$

In addition, for all non-degenerate distributions $P \in \mathscr{P}_n,\ P \neq P_n$, and all $n \geq 2$ it holds

$$\begin{aligned} &H_1(P) < R_\alpha(P) < S_\alpha(P) \quad \text{for} \quad \alpha \in [0,1) \qquad \text{and} \\ &H_1(P) > R_\alpha(P) > S_\alpha(P) \quad \text{for} \quad \alpha \in (1,\infty)\ . \end{aligned}$$

Remark 5 Again, let $n = 2$. Then for every $\alpha \in (0,\infty)$ the function

$$\psi_\alpha^S(s) = \begin{cases} \frac{1-\alpha\cdot s^{\alpha-1}}{\alpha-1} + s^\alpha + (1-s)^\alpha & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\} \\ -\ln s & \text{for} \quad \alpha = 1 \end{cases}, \ s \in [0,1],$$

which is similarly defined as $\psi_\alpha^R(s)$ in Remark 4, is monotone decreasing and satisfies $\psi_\alpha(1) = 0$ and is, therefore, a function of uncertainty in the sense of Definition 1.

(A) For the special case $n = 2$ the class of entropies corresponding to this class of functions of uncertainty is Tsallis' class of entropies

$$S_\alpha(P_t) = \begin{cases} \frac{1-(t^\alpha+(1-t)^\alpha)}{\alpha-1} & \text{for} \quad \alpha \in (0,\infty)\setminus\{1\} \\ -(t\ln t + (1-t)\ln(1-t)) & \text{for} \quad \alpha = 1 \;, \end{cases}$$

where the optimal distribution for $P_t = (t, 1-t) \in \mathscr{P}_2$ equals $Q^*_\alpha(P_t) \equiv P_t$.

(B) The above functions of uncertainty for the special cases $\alpha = 2$ and $\alpha = 3$ are

$$\psi^S_2(s) = 2(1-s)^2 \quad \text{and} \quad \psi^S_3(s) = \frac{3}{2}(1-s)^2$$

and, therefore, are the quadruple resp. triple of the function of uncertainty of the special case $k = 1$ of Arimoto's extended class of entropies. Consequently, for the special case $n = 2$ the corresponding entropies are the quadruple resp. triple of that of Arimoto's extended class for $k = 1$. Since $H_{-k}(P)$, $k \in (0,\infty)$, is limited to distributions $P \in \mathscr{P}_n$, which satisfy $p_{\min} > 0$, except for $n = 2$, the latter cannot apply for $n > 2$.

(C) Note that for the special case $\alpha = 2$

$$\kappa(P) - \frac{1}{n} = I(P) := \sum_{j=1}^{n} (p_j - \frac{1}{n})^2 = 1 - \frac{1}{n} - S_2(P)$$

holds, where $\kappa(P) = \sum_{j=1}^n p_j^2$ is Friedman's time-honoured *Index of Coincidence* (1920) and $I(P)$ the measure of information proposed by Brukner and Zeilinger (2001) in their paper. For further applications of $\kappa(P)$ see, e.g., Österreicher (2008, Section 2.4).

Remark 6 Let $P \in \mathscr{P}_n$ and $n \geq 2$. Then owing to

$$\kappa(P) = \sum_{j=1}^{n} p_j^2 \leq p_{\max} \cdot \sum_{j=1}^{n} p_j = p_{\max}$$

it holds

$$H_0(P) = 1 - p_{\max} \leq 1 - \sum_{j=1}^{n} p_j^2 = S_2(P) \;,$$

with equality iff $P = P_n = (\frac{1}{n}, ..., \frac{1}{n})$. For the final statement let $P \in \mathscr{P}_n$ satisfy $p_{\min} > 0$ for all $n \geq 2$. Then it holds that

$$H_{2-\alpha}(P) < S_\alpha(P) \;\; \forall\, \alpha \in (2,\infty) \quad \text{or} \quad H_{-k}(P) < S_{2+k}(P) \;\; \forall\, k \in (0,\infty).$$

In order to shorten our paper we skip the proof of this result. Of course, it is available from the authors upon request. Here we only mention that the key step of our proof is to verify that the lower bound

$$\lambda(t,k,n) = 1 - t^{1+k} - \frac{(n-1)^{1+k} \cdot t \cdot (1-t)}{((n-1)^{1+1/k} \cdot t^{1/k} + (1-t)^{1/k})^k}$$

of the difference

$$(1+k) \cdot (S_{2+k}(P_n(t)) - H_{-k}(P_n(t))),$$

$P_n(t) = (p_1, ..., p_{n-1}, t) \in \mathscr{P}_n$, is positive for all $t = p_{\max} \in \left[\frac{1}{n}, 1\right)$, all $k \in (0,\infty)$ and $n \geq 2$.

References

ARIMOTO, S. (1971). Information-theoretical considerations on estimation problems. *Information and Control*, **19**, 181–194.

BRUKNER, Č. AND ZEILINGER, A. (2001). Conceptual inadequacy of the Shannon information in quantum measurements. *Physical Review A*, **63** (022113), 1–11.

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. (1974). Information measures: A critical survey. *In Transactions of the Seventh Prague Conference on Information Theory*, volume A. Academia Prague, pp. 73–86.

CSISZÁR, I. (2008). Axiomatic characterizations of information measures. *Entropy*, **10** (3), 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.

FRIEDMAN, W. F. (1920). The index of coincidence and its application in cryptography. Technical Report 22, Riverbank Publication.

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

ÖSTERREICHER, F. (2008). Informationstheorie. Lecture notes, University of Salzburg.

ÖSTERREICHER, F. AND VAJDA, I. (1993). Statistical information and discrimination. *IEEE Transactions on Information Theory*, **39** (3), 1036–1039.

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** (3), 639–653.

PURI, M. L. AND VINCZE, I. (1988). Information and mathematical statistics. *In Proceedings of the IVth Conference on Asymptotic Statistics*. Prague.

RÉNYI, A. (1961). On measures of Entropy and information. *In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1. Berkeley, pp. 547–561.

SCHÜTZENBERGER, M. P. (1954). Contribution aux applications statistic de la théorie de l'information. *Publications de l'Institut de statistique de l'Université de Paris*, **3**, 3–117.

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

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

VAJDA, I. (1989). *Theory of Statistical Inference and Information*. Kluwer Academic Publishers: Dordrecht-Boston-London.

VAJDA, I. (2009). On metric divergences of probability distributions. *Kybernetika*, **45** (6), 885–900.

VAN ERVEN, T. AND HARREMOËS, P. (2014). Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, **60** (7), 3797–3820.

Manuscript received, 2016-12-22, revised, 2017-04-05, accepted, 2017-04-11.