The role of boundedness and nonnegativity in characterizing entropies of degree α

By GYULA MAKSA (Debrecen)

Dedicated to Professor Zoltán Daróczy on his 50th birthday

1. Introduction. We denote the set of all complete n-ary probability distributions by Γ_n $(n=2, 3, \dots)$, that is

$$\Gamma_n = \Big\{(p_1, \dots, p_n)\colon p_i \geqq 0,\ i = 1, \dots, n;\ \sum_{i=1}^{n} p_i = 1\Big\}.$$

For a fixed real number α, the entropy of degree α is a sequence $\{H_n^\alpha\}_{n=2}^\infty$ of functions where $H_n^\alpha\colon \Gamma_n \to \mathbf{R}$ ($\mathbf{R}$ denotes the set of all real numbers) is defined by

$$H_n^\alpha(p_1, \dots, p_n) = \begin{cases} (2^{1-\alpha}-1)^{-1}\Big(\sum_{i=1}^{n} p_i^\alpha - 1\Big) & \text{if} \quad \alpha \neq 1 \\ -\sum_{i=1}^{n} p_i \log p_i & \text{if} \quad \alpha = 1 \end{cases}$$

with the conventions $0^\alpha=0$ for all $\alpha\in\mathbf{R}$, $\log=\log_2$ and $0\log 0=0$. The entropy of degree 1 is the Shannon entropy introduced by SHANNON [15]. For $\alpha>0$, the entropy of degree α was introduced by DARÓCZY [6]. Because of the convention $0^0=0$ none of the functions H_n^α is constant. For example, $H_n^0(p_1, \dots, p_n)+1$ gives how many probabilities in $(p_1, \dots, p_n)\in\Gamma_n$ are different from zero.

The problem of characterization of the entropies of degree α is the following: What properties have to be imposed upon a sequence of functions $I_n\colon \Gamma_n\to\mathbf{R}$ $(n=2, 3, \dots)$ in order that the equality

$$I_n(p_1, \dots, p_n) = H_n^\alpha(p_1, \dots, p_n)$$

should hold for some $\alpha\in\mathbf{R}$ and for all $(p_1, \dots, p_n)\in\Gamma_n$, $n=2, 3, \dots$. This problem is raised and extensively discussed in the book of ACZÉL and DARÓCZY [2].

We begin with three usual *axioms for the sequence* $\{I_n\}_{n=2}^\infty$ $(I_n\colon \Gamma_n\to\mathbf{R}, n=2, 3, \dots)$

$(\mathrm{A})_\alpha$ *α-additivity:*

$$I_{nm}(p_1 q_1, \dots, p_1 q_m, \dots, p_n q_1, \dots, p_n q_m) =$$
$$= I_n(p_1, \dots, p_n)+I_m(q_1, \dots, q_m)+(2^{1-\alpha}-1)\, I_n(p_1, \dots, p_n)\, I_m(q_1, \dots, q_m)$$

for some $\alpha\in\mathbf{R}$ and for all $(p_1, \dots, p_n)\in\Gamma_n$, $(q_1, \dots, q_m)\in\Gamma_m$; $n, m=2, 3, \dots$.

(B) *Sum property:* There exists a function $f\colon [0, 1]\to\mathbf{R}$ such that

$$I_n(p_1, \dots, p_n) = \sum_{i=1}^{n} f(p_i)$$

for all $(p_1, \dots, p_n)\in\Gamma_n$, $n=2, 3, \dots$. (The function f is called a generating function of $\{I_n\}_{n=2}^{\infty}$.)

(C) *Normalization:*

$$I_2\left(\frac{1}{2}, \frac{1}{2}\right) = 1.$$

These properties do not characterize $\{H_n^{\alpha}\}_{n=2}^{\infty}$ therefore the authors investigating this problem required a further property. Supposing that a sequence of functions $I_n\colon \Gamma_n\to\mathbf{R}$ $(n=2, 3, \dots)$ satisfies $(\mathrm{A})_\alpha$ with some $\alpha\in\mathbf{R}$ and (B) we have the following system of functional equations for the generating function f of $\{I_n\}_{n=2}^{\infty}$:

$$(1)_\alpha \qquad \sum_{i=1}^{n}\sum_{j=1}^{m} f(p_i q_j) = \sum_{i=1}^{n} f(p_i)+\sum_{j=1}^{m} f(q_j)+(2^{1-\alpha}-1)\sum_{i=1}^{n} f(p_i)\sum_{j=1}^{m} f(q_j),$$

$$(p_1, \dots, p_n)\in\Gamma_n, \quad (q_1, \dots, q_m)\in\Gamma_m; \; n, m = 2, 3, \dots .$$

In the case $\alpha=1$ this system of functional equations was first studied by CHAUNDY and MCLEOD [5]. They proved that the continuous solutions are of the form

$$(2) \qquad f(x) = cx\log x, \quad x\in[0, 1]$$

with some $c\in\mathbf{R}$ and with the convention $0\log 0=0$. ACZÉL and DARÓCZY [1] proved the same supposing only that f is continuous and the equation $(1)_1$ holds for all $n=m\geqq 2$. DARÓCZY [7] determined the measurable solutions of $(1)_1$ in the case $n=3$, $m=2$, $f(1)=0$. The author [13] proved the following result: If $n\geqq 3$ and $m\geqq 2$ are fixed in $(1)_1$ and f is a solution of $(1)_1$ which is bounded on a set of positive Lebesgue measure then

$$f(x) = cx\log x+bx+a, \quad x\in[0, 1]$$

with some $a, b, c\in\mathbf{R}$. In the case $\alpha\neq 1$ supposing that $(1)_\alpha$ holds for all $n\geqq 2$, $m\geqq 2$ the continuous solutions were determined by BEHARA and NATH [3], KANNAPPAN [10] and MITTAL [14]. For fixed $n\geqq 3$, $m\geqq 2$ LOSONCZI [11] found the measurable solutions of $(1)_\alpha$.

From the results mentioned above such a type of characterization theorems can be obtained for $\{H_n^{\alpha}\}_{n=2}^{\infty}$ in which $(\mathrm{A})_\alpha$, (B), (C) and a regularity property of the generating function are supposed. Our purpose in this paper is to give characterization theorems for $\{H_n^{\alpha}\}_{n=2}^{\infty}$ in which all conditions concern the sequence $\{I_n\}_{n=2}^{\infty}$ itself and we suppose nothing on the generating function. (We suppose its existence only.)

Therefore our *further axioms for* $\{I_n\}_{n=2}^{\infty}$ will be the following:

$(\mathrm{D})_1$ *Boundedness:* There exists $K\in\mathbf{R}$ such that

$$|I_3(p_1, p_2, p_3)| \leqq K, \quad (p_1, p_2, p_3)\in\Gamma_3.$$

(D) *Boundedness:* The function $t \to I_2(t, 1-t)$, $t \in [0, 1]$ is bounded on a subset of positive Lebesgue measure of $[0, 1]$.

(F) *Nonnegativity:* The function $t \to I_2(t, 1-t)$, $t \in [0, 1]$ is nonnegative on a subset of positive Lebesgue measure of $[0, 1]$.

2. The case $\alpha = 1$. First, we prove a characterization theorem for the Shannon entropy.

Theorem 1. *A sequence $\{I_n\}_{n=2}^{\infty}$ of functions I_n: $\Gamma_n \to \mathbf{R}$ $(n=2, 3, \ldots)$ is the Shannon entropy $\{H_n^1\}_{n=2}^{\infty}$ if and only if $\{I_n\}_{n=2}^{\infty}$ has the properties* $(A)_1$, (B), (C) *and* $(D)_1$.

Proof. It can easily be verified that the Shannon entropy satisfies the axioms $(A)_1$, (B), (C) and $(D)_1$ thus we only deal with the converse. Suppose that $\{I_n\}_{n=2}^{\infty}$ satisfies $(A)_1$, (B), (C) and $(D)_1$ and let f: $[0, 1] \to \mathbf{R}$ be a generating function of $\{I_n\}_{n=2}^{\infty}$. Then f satisfies $(1)_1$ for all $n \geqq 2$, $m \geqq 2$. With the substitution $(1, 0, \ldots, 0) \in \Gamma_n$, $(q_1, \ldots, q_m) \in \Gamma_m$, $(1)_1$ implies that $(n-1)(m-1)f(0) = f(1)$. Since this is valid for all $n \geqq 2$, $m \geqq 2$ we have that $f(0) = f(1) = 0$.

Let F be the periodic extension of f to $\mathbf{R}$ with the period 1 and define the function g on $\mathbf{R}^2$ by

$$g(x, y) = F(x+y) - F(x) - F(y). \tag{3}$$

We shall show that g is bounded on $\mathbf{R}^2$. Since g is periodic in both variables with the period 1, it is enough to show that g is bounded on $[0, 1] \times [0, 1]$. Moreover, because of the identity

$$g(x, y) = g(x, 1-x) + g(y, 1-y) - g(1-x, 1-y) - g(2-x-y, x+y-1)$$

it is enough to show that g is bounded on the set

$$\triangle = \{x, y): x, y, x+y \in [0, 1]\}.$$

Let $(x, y) \in \Delta$. Then, by (B) and (3), we get

$$I_3(1-x-y, x+y, 0) - I_3(x, y, 1-x-y) = f(x+y) - f(x) - f(y) = g(x, y),$$

therefore $(D)_1$ implies that g is bounded on Δ. Thus g is bounded on $\mathbf{R}^2$, too.

Applying Theorem 1.2 of de Bruijn [4] (or the stability theorem of Hyers [9]) we obtain that F is the sum of a function bounded on $\mathbf{R}$ and a function a: $\mathbf{R} \to \mathbf{R}$ satisfying the Cauchy functional equation $a(x+y) = a(x) + a(y)$ for all $x, y \in \mathbf{R}$. It follows that

$$f(x) = f^*(x) + a(x), \quad x \in [0, 1]$$

where f^*: $[0, 1] \to \mathbf{R}$ is bounded and we may suppose that $a(1) = 0$. Therefore f^* is a bounded solution of $(1)_1$, thus, by [13], f^* has the form (2) with some $c \in \mathbf{R}$. Finally, for $(p_1, \ldots, p_n) \in \Gamma_n$ we have

$$I_n(p_1, \ldots, p_n) = \sum_{i=1}^{n} f(p_i) = \sum_{i=1}^{n} f^*(p_i) + a(1) = c \sum_{i=1}^{n} p_i \log p_i = -cH_n^1(p_1, \ldots, p_n)$$

where, by (C), $c = -1$ and the *proof is complete.*

In this theorem the boundedness $(D)_1$ cannot be replaced by boundedness from one side or, in particular, by nonnegativity. In connection with this fact we present an *example*.

A function $d\colon \mathbf{R}\to\mathbf{R}$ is called a real derivation if

$$d(x+y) = d(x)+d(y) \quad \text{and} \quad d(xy) = x\,d(y)+y\,d(x)$$

hold for all $x, y\in\mathbf{R}$. It is known that there exist nonidentically zero real derivations (see ZARISKI and SAMUEL [16]). Define the function f on $[0, 1]$ by

$$f(x) = \begin{cases} \dfrac{d(x)^2}{x} - x\log x & \text{if} \quad x\in]0, 1] \\ 0 & \text{if} \quad x = 0 \end{cases}$$

where d is a nonidentically zero real derivation, and define I_n on Γ_n by

$$I_n(p_1, \ldots, p_n) = \sum_{i=1}^{n} f(p_i) \quad (n = 2, 3, \ldots). \tag{4}$$

Using the properties of d we find that

$$f(xy) = xf(y)+yf(x)+2\,d(x)\,d(y); \quad x, y\in[0, 1]. \tag{5}$$

Since $d(1)=0$, (5) and (4) imply that $\{I_n\}_{n=2}^{\infty}$ defined by (4) has the property (A_1). (B) and (C) are obviously satisfied by $\{I_n\}_{n=2}^{\infty}$. It has been proved by DARÓCZY and MAKSA [8] that

$$\frac{d(u+v)^2}{u+v} \leqq \frac{d(u)^2}{u} + \frac{d(v)^2}{v}$$

for all positive numbers u and v. Since

$$-(u+v)\log(u+v) \leqq -u\log u - v\log v \quad (u > 0,\ v > 0)$$

is also true, we have that $f(x+y)\leqq f(x)+f(y)$ for all $(x, y)\in\Delta$. Using this inequality several times, (4) implies that

$$I_n(p_1, \ldots, p_n) \geqq f\Big(\sum_{i=1}^{n} p_i\Big) = f(1) = 0$$

for all $(p_1, \ldots, p_n)\in\Gamma_n$, $n=2, 3, \ldots$. Finally, we show that $\{I_n\}_{n=2}^{\infty}$ is different from $\{H_n^1\}_{n=2}^{\infty}$. Indeed, suppose that $I_2(x, 1-x)=H_2^1(x, 1-x)$ for all $x\in[0, 1]$. Then

$$0 = \frac{d(x)^2}{x} + \frac{d(1-x)^2}{1-x} = \frac{d(x)^2}{x(1-x)}$$

for all $x\in]0, 1[$, which implies the contradiction that d is identically zero.

3. The case $\alpha\neq 1$. In this case the generating functions of $\{I_n\}_{n=2}^{\infty}$ ($I_n\colon \Gamma_n\to\mathbf{R}$, $n=2, 3, \ldots$) satisfying $(A)_\alpha$ and (B) have been determined by LOSONCZI and MAKSA [12]. The following theorem is a simple consequence of this result so its proof is omitted.

Theorem 2. *Let* $\alpha\neq 1$. *The sequence* $\{I_n\}_{n=2}^{\infty}$ *of functions* $I_n\colon \Gamma_n\to\mathbf{R}$ $(n=2, 3, \ldots)$ *satisfies* $(A)_\alpha$, (B) *and* (C) *if and only if there exists a function* $h\colon [0, 1]\to\mathbf{R}$ *such that*

$$h(xy)=h(x)h(y);\quad x, y\in[0, 1] \tag{6}$$

$$h(0)=0,\quad h\left(\frac{1}{2}\right)=2^{-\alpha} \tag{7}$$

and

$$I_n(p_1, \ldots, p_n)=(2^{1-\alpha}-1)^{-1}\left(\sum_{i=1}^{n}h(p_i)-1\right) \tag{8}$$

for all $(p_1, \ldots, p_n)\in\Gamma_n$ *and* $n=2, 3, \ldots$.

In connection with the solutions $h\colon [0, 1]\to\mathbf{R}$ of (6) we mention the well-known facts that $h(x)\geqq 0$ for all $x\in[0, 1]$, and if h is bounded on a subset of $[0, 1]$ of positive Lebesgue measure and (7) is also satisfied then

$$h(x)=x^\alpha,\quad x\in[0, 1] \tag{9}$$

with the convention $0^\alpha=0$.

Thus we can easily prove the following two theorems.

Theorem 3. *Let* $\alpha\neq 1$. *A sequence* $\{I_n\}_{n=2}^{\infty}$ *of functions* $I_n\colon \Gamma_n\to\mathbf{R}$ $(n=2, 3, \ldots)$ *is the entropy of degree* α $\{H_n^\alpha\}_{n=2}^{\infty}$ *if and only if* $\{I_n\}_{n=2}^{\infty}$ *has the properties* $(A)_\alpha$, (B), (C) *and* (D).

Proof. By Theorem 2, $\{I_n\}_{n=2}^{\infty}$ satisfies $(A)_\alpha$, (B) and (C) if and only if there exists $h\colon [0, 1]\to\mathbf{R}$ with the properties (6), (7) and (8). If (D) holds too then

$$0\leqq h(t)=(2^{1-\alpha}-1)I_2(t, 1-t)-h(1-t)+1\leqq(2^{1-\alpha}-1)I_2(t, 1-t)+1,\quad t\in[0, 1]$$

shows that h is also bounded on a subset of positive Lebesgue measure of $[0, 1]$. Thus (9) and (8) imply that $\{I_n\}_{n=2}^{\infty}=\{H_n^\alpha\}_{n=2}^{\infty}$. Since $h(x)=x^\alpha$, $x\in[0, 1]$ satisfies (6) and (7) and Theorem 2 yields the converse.

Theorem 4. *Let* $\alpha>1$. *A sequence* $\{I_n\}_{n=2}^{\infty}$ *of functions* $I_n\colon \Gamma_n\to\mathbf{R}$ $(n=2, 3, \ldots)$ *is the entropy of degree* α $\{H_n^\alpha\}_{n=2}^{\infty}$ *if and only if* $\{I_n\}_{n=2}^{\infty}$ *has the properties* $(A)_\alpha$, (B), (C) *and* (E).

Proof. Applying again Theorem 2, $\{I_n\}_{n=2}^{\infty}$ satisfies $(A)_\alpha$, (B) and (C) if and only if there exists $h\colon [0, 1]\to\mathbf{R}$ with the properties (6), (7) and (8). If (E) holds too then

$$0\leqq h(t)=(2^{1-\alpha}-1)I_2(t, 1-t)-h(1-t)+1\leqq 1$$

is valid on a subset of positive Lebesgue measure of $[0, 1]$. Thus we have again (9) and Theorem 2 implies Theorem 4.

In the case $\alpha<1$ Theorem 4 does not work. In what follows we present an *example.* First, we verify an inequality.

Lemma. *Let* $a, b, \alpha\in\mathbf{R}$ *such that* $a>0, b>0$ *and* $\alpha<1$. *Then*

$$b(a+1)^\alpha<b^{a+1}+a^\alpha. \tag{10}$$

PROOF. Let $a>0$ be fixed and

$$\varphi(b) = b(a+1)^\alpha - b^{a+1} - a^\alpha, \quad b > 0.$$

Then $\varphi'(b)=(a+1)^\alpha-(a+1)b^a=(a+1)[(a+1)^{\alpha-1}-b^a]$. This implies that $\varphi'(b)\geqq 0$ if and only if $b\in]0, b_0]$ where $b_0=(a+1)^{\alpha-1/a}$. Thus we have

$$\varphi(b) \leqq \varphi(b_0) = (a+1)^{((\alpha-1)/a)+\alpha} - (a+1)^{((\alpha-1)/a)(a+1)} - a^\alpha =$$
$$= a[(a+1)^{(\alpha-1)(1+(1/a))} - a^{\alpha-1}] < a[(a+1)^{\alpha-1} - a^{\alpha-1}] < 0$$

which proves (10).

Let $\alpha<1$ be fixed and define the function H on $[0, +\infty[$ by

$$H(x) = \begin{cases} x^\alpha 2^{d(x)/x} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \end{cases}$$

where d is a nonidentically zero real derivation. An easy calculation shows that

(11) $$H(xy) = H(x)H(y); \quad x, y\in[0, +\infty[$$

and $H\left(\frac{1}{2}\right)=2^{-\alpha}$, $H(0)=0$. Let $x>0$ and $a=x$, $b=2^{-(d(x)/x(x+1))}$. Then, applying our lemma and using that $d(x)=d(x+1)$, (10) implies that

$$(x+1)^\alpha 2^{-(d(x)/x(x+1))} < 2^{-(d(x)/x)} + x^\alpha$$

and

$$(x+1)^\alpha 2^{d(x+1)/(x+1)} < x^\alpha 2^{d(x)/x} + 1,$$

that is $H(x+1)<H(x)+1$.

This inequality and (11) give that

(12) $$H(x+y) \leqq H(x)+H(y); \quad x, y\in[0, +\infty[.$$

Let now $h(x)=H(x)$ if $x\in[0, 1]$ and

$$I_n(p_1, \ldots, p_n) = (2^{1-\alpha}-1)^{-1}\left(\sum_{i=1}^n h(p_i)-1\right); \quad (p_1, \ldots, p_n)\in\Gamma_n, \ n = 2, 3, \ldots.$$

Then it follows from Theorem 2 that $\{I_n\}_{n=2}^\infty$ satisfies $(A)_\alpha$, (B) and (C). Using (12) several times, we obtain

$$I_n(p_1, \ldots, p_n) = (2^{1-\alpha}-1)^{-1}\left(\sum_{i=1}^n H(p_i)-1\right) \geqq (2^{1-\alpha}-1)^{-1}\left(H\left(\sum_{i=1}^n p_i\right)-1\right) =$$
$$= (2^{1-\alpha}-1)^{-1}\left(H(1)-1\right) = 0$$

for all $(p_1, \ldots, p_n)\in\Gamma_n$, $n=2, 3, \ldots$. Finally, we show that $\{I_n\}_{n=2}^\infty$ is different from $\{H_n^\alpha\}_{n=2}^\infty$. Indeed, suppose that $I_2(x, 1-x)=H_2(x, 1-x)$ for all $x\in[0, 1]$. Then $h(x)+h(1-x)=x^\alpha+(1-x)^\alpha$, whence

$$0 \leqq h(x) \leqq h(x)+h(1-x) = x^\alpha+(1-x)^\alpha$$

which shows that h is bounded on a subset of positive Lebesgue measure of [0, 1] and thus we have (9). This implies that d is identically zero which is a contradiction.

References

[1] J. ACZÉL and Z. DARÓCZY, Characterisierung der Entropien positiver Ordnung und der Shannonschen Entropie, *Acta Math. Acad. Sci. Hungar.*, **14** (1963), 95—121.

[2] J. ACZÉL and Z. DARÓCZY, On measures of information and their characterizations, *Academic Press* (New York—San Francisco—London, 1975).

[3] M. BEHARA and P. NATH, Additive and non-additive entropies of finite measurable partitions, Probability and Information Theory II, *Lecture Notes in Math.*, **296** (1973), 102—138.

[4] N. G. DE BRUIJN, Functions whose differences belong to a given class, *Nieuw Arch. Wisk.*, **23** (1951), 194—218.

[5] T. W. CHAUNDY and J. B. MCLEOD, On a functional equation, *Edinburgh Math. Notes*, **43** (1960), 7—8.

[6] Z. DARÓCZY, Generalized information functions, *Information and Control*, **16** (1970), 36—51.

[7] Z. DARÓCZY, On the measurable solutions of a functional equation, *Acta Math. Acad. Sci. Hungar.*, **22** (1971), 11—14.

[8] Z. DARÓCZY and GY. MAKSA, Nonnegative information functions, In analytic Function Methods in Probability and Statistics, *Colloquia Math. Soc. J. Bolyai*, **21** (1979), 65—76.

[9] D. H. HYERS, On the stability of the linear functional equation, *Proc. Math. Acad. Sci. U.S.A.*, **27** (1941), 222—224.

[10] PL. KANNAPPAN, On a generalization of some measures in information theory, *Glasnik Mat.*, **9** (29), 1974), 81—93.

[11] L. LOSONCZI, A characterization of entropies of degree α, *Metrika*, **28** (1981), 237—244.

[12] L. LOSONCZI and GY. MAKSA, On some functional equations of the information theory, *Acta Math. Acad. Sci. Hungar.*, **39** (1982), 73—82.

[13] GY. MAKSA, On the bounded solutions of a functional equation, *Acta Math. Acad. Sci. Hungar.*, **37** (1981), 445—450.

[14] D. P. MITTAL, On continuous solutions of a functional equation, *Metrika*, **22** (1970), 31—40.

[15] C. E. SHANNON, A mathematical theory of communication, *Bell System Tech. J.*, **27** (1948), 379—423.

[16] O. ZARISKI and P. SAMUEL, Commutative algebra, *D. Van Nostrand*, (Princeton, 1958).

DEPARTMENT OF MATHEMATICS
LAJOS KOSSUTH UNIVERSITY
H—4010 DEBRECEN, PF. 12
HUNGARY

(Received December 30, 1986)