

Corina Reischer, Dan Simovici, and Ivan Stojmenovic

## AN ALGEBRAIC APPROACH TO ENTROPY

ABSTRACT. We introduce and characterize the algebraic notion of entropy for functions between finite sets using common operations applied to functions. Unlike the notion of entropy of a probability distribution, the entropy of a function has an algebraic rather than a probabilistic character, though the two notions are clearly related. We point to applications of this notion to the study of finite functions. Then, we extend this notion to equivalence relations and we point to some other areas of interest.

### 1. Introduction

The purpose of this Note is to introduce an algebraic counterpart of the notion of entropy which is applicable to the study of functions between finite sets. Such functions play an essential role in modeling multi-valued logic circuitry and we believe that functional entropy can be applied to the study of iteration and decomposition properties.

After we introduce the notion of functional entropy we characterize this notion using a set of axioms that makes exclusive use of operations on functions. This follows a long standing interest in axiomatizations of similar concepts (see [1, 2, 3]).

A function is defined as a set  $f$  of ordered pairs such that  $(a, b), (a, b') \in f$  imply  $b = b'$ . Following standard terminology, the domain of  $f$  and the range of  $f$  are defined by

$$\begin{aligned}\text{Dom}(f) &= \{a \mid (a, b) \in f\}, \\ \text{Ran}(f) &= \{b \mid (a, b) \in f\},\end{aligned}$$

respectively.

If  $\text{Dom}(f) = A$  and  $\text{Ran}(f) \subseteq B$ , we use the notation  $f : A \rightarrow B$ .

Let  $f : A \rightarrow B$  be a function between two finite sets  $A$  and  $B$ . For every  $b \in B$  define the number  $\alpha_b = |\{a \in A \mid f(a) = b\}|$

**1.1. Definition.** The *entropy* of the function  $f$  is the number  $\mathcal{H}(f)$  defined by

$$\mathcal{H}(f) = \sum_{b \in B} \alpha_b \log \alpha_b.$$

Unless otherwise specified we assume that all logarithms are in base 2.

We need to consider several operations on functions between finite sets.

Let  $A, C$  be two disjoint sets. If  $f : A \rightarrow B$  and  $g : C \rightarrow D$  are two functions and  $B, D$  are disjoint sets, define the function  $f||g : A \cup C \rightarrow B \cup D$  as

$$f||g(z) = \begin{cases} f(z) & \text{if } z \in A, \\ g(z) & \text{if } z \in C. \end{cases}$$

One could easily see that this partial operation on functions is commutative and associative.

If  $A, B, C, D$  are four arbitrary sets and  $f : A \rightarrow B$ ,  $g : C \rightarrow D$ , define the function  $f \times g : A \times C \rightarrow B \times D$  as  $f \times g(a, c) = (f(a), g(c))$  for every  $a \in A$  and  $c \in C$ .

Finally, if  $f : A \rightarrow B$  and  $g : B \rightarrow C$ , the composition of  $f$  and  $g$  is the function  $gf : A \rightarrow C$ , defined by  $gf(a) = g(f(a))$ .

The proof of the next statement is immediate.

**1.2. Lemma.** *If  $f : A \rightarrow B$ ,  $g : C \rightarrow D$  and  $A \cap C = B \cap D = \emptyset$ , then  $\mathcal{H}(f|g) = \mathcal{H}(f) + \mathcal{H}(g)$ .*

**1.3. Theorem.** *For every function  $f : A \rightarrow B$  we have*

$$|A| \log \frac{|A|}{|f(A)|} \leq \mathcal{H}(f) \leq |A| \log |A|.$$

*Proof.* The notion of function entropy is related to the notion of entropy of a random variable. If  $f : A \rightarrow B$  and  $B = \{b_0, \dots, b_{n-1}\}$ , consider the random variable

$$X_f : \begin{pmatrix} b_0 & \cdots & b_{n-1} \\ p_0 & \cdots & p_{n-1} \end{pmatrix},$$

where  $p_i = |f^{-1}(b_i)|/|A|$  for  $0 \leq i \leq n-1$ . The entropy of  $X$  is  $\mathcal{H}(X) = \sum_{0 \leq i \leq n-1} p_i \log \frac{1}{p_i}$ . This allows us to write

$$\begin{aligned} \mathcal{H}(X_f) &= \sum_{0 \leq i \leq n-1} |f^{-1}(b_i)|/|A| \log \frac{|A|}{|f^{-1}(b_i)|} \\ &= \log |A| - \frac{1}{|A|} \mathcal{H}(f). \end{aligned}$$

It is well known fact that  $0 \leq \mathcal{H}(X_f) \leq \log |f(A)|$ ; therefore, we obtain  $|A| \log \frac{|A|}{|f(A)|} \leq \mathcal{H}(f) \leq |A| \log |A|$ .  $\square$

**1.4. Corollary.** *For every constant function  $c$  defined on the finite set  $A$  we have*

$$\mathcal{H}(c) = \max\{\mathcal{H}(f) | \text{Dom}(f) = A\}.$$

*If  $f : A \rightarrow A$ , then  $\mathcal{H}(f) = 0$  if and only if  $f$  is a bijection.*

Denote by  $H(A)$  the number

$$H(A) = \sup\{\mathcal{H}(f) | f : A \rightarrow B \text{ for some } B\}.$$

**1.5. Lemma.** *Let  $A_0, \dots, A_{n-1}$  be  $n$  finite sets and  $n \geq 1$ . We have*

$$H(A_0 \times \cdots \times A_n) = \sum_{i=0}^{n-1} |A_0| \cdots |A_{i-1}| H(A_i) |A_{i+1}| \cdots |A_{n-1}|.$$

*Proof.* The argument is straightforward and it is left to the reader.  $\square$

**1.6. Lemma.** *Let  $A, B$  be two finite sets and let  $\pi_A, \pi_B$  be the projection mappings  $\pi_A : A \times B \rightarrow A$ ,  $\pi_B : A \times B \rightarrow B$ . If  $|A| \leq |B|$ , then  $\mathcal{H}(\pi_A) \geq \mathcal{H}(\pi_B)$ .*

*Proof.* Let  $A = \{a_0, \dots, a_{m-1}\}$  and  $B = \{b_0, \dots, b_{n-1}\}$  be two finite sets, where  $m \leq n$ . We have

$$\mathcal{H}(\pi_A) = \sum_{0 \leq i \leq m-1} |\pi_A^{-1}(a_i)| \log |\pi_A^{-1}(a_i)| = mn \log n.$$

Similarly,  $\mathcal{H}(\pi_B) = nm \log m$  and, since  $m \leq n$ , we obtain  $\mathcal{H}(\pi_B) \leq \mathcal{H}(\pi_A)$ .  $\square$

## 2. An Axiomatization of Functional Entropy

We are introducing the functional entropy  $\mathcal{H}(f)$  as a numerical characteristic of functions satisfying the following axioms:

(ENT1)  $\mathcal{H}(f \times g) = |A|\mathcal{H}(g) + |C|\mathcal{H}(f)$ , if  $f : A \rightarrow B$  and  $g : C \rightarrow D$ .

(ENT2)  $\mathcal{H}(f|g) = \mathcal{H}(f) + \mathcal{H}(g)$ , if  $f : A \rightarrow B$ ,  $g : C \rightarrow D$  and  $A \cap C = B \cap D = \emptyset$ .

(ENT3) If  $\alpha$  and  $\beta$  are bijections, then  $\mathcal{H}(\beta f \alpha) = \mathcal{H}(f)$ .

(ENT4) If  $f : A \rightarrow B$  and  $g : B \rightarrow C$ , then  $\mathcal{H}(gf) \geq \mathcal{H}(f)$ .

(ENT5) For every finite set  $A$  there is a function  $f$  defined on  $A$  such that  $\mathcal{H}(f) = \sup\{\mathcal{H}(g) | g : A \rightarrow B \text{ for some set } B\}$ . Moreover, if  $A, B$  are two finite sets such that  $|A| \leq |B|$ , then the largest entropy of a function defined on  $A$  is less or equal than the largest entropy of a function defined on  $B$ .

Note that all these axioms are formulated in terms of functions and functional operations. This suggests that the algebraic notion of functional entropy is a natural one.

In order to prove Theorem 2.4 we need to show several preliminary results.

**2.1. Lemma.** *If  $f : A \rightarrow B$  is a function, and  $f' : A \rightarrow f(A)$  is the corestriction of  $f$  to  $f(A)$ , then  $\mathcal{H}(f) = \mathcal{H}(f')$ .*

*Proof.* Consider the mapping  $g : B \rightarrow f(A)$  defined by  $g(b) = b$  if  $b \in f(A)$  and  $g(b) = b_0$  otherwise, where  $b_0$  is a fixed element of  $B$ . Since  $gf = f'$ , by (ENT4), we have  $\mathcal{H}(f') \geq \mathcal{H}(f)$ . On the other hand, if  $i : f(A) \rightarrow B$  is the identity mapping given by  $i(b) = b$  for all  $b \in B$ , then  $if' = f$ , which gives  $\mathcal{H}(f) \geq \mathcal{H}(f')$ . Therefore, we conclude that  $\mathcal{H}(f) = \mathcal{H}(f')$ .  $\square$

Lemma 2.1 shows that, as far as the entropy is concerned, we can assume that the functions we are dealing with are surjective.

**2.2. Lemma.** *Let  $c$  be a constant function defined on the set  $A$ . For every function  $f$  defined on  $A$  we have  $\mathcal{H}(c) \geq \mathcal{H}(f)$ .*

*Furthermore, two constant functions defined on sets with the same cardinality have the same entropy.*

*Proof.* By Lemma 2.1, we may assume that  $c$  is surjective. Suppose that  $c : A \rightarrow B$  and  $f : A \rightarrow D$ , where  $B = \{b\}$ . If  $l : D \rightarrow B$  is the mapping given by  $l(d) = b$  for every  $d \in D$ , we have  $c = lf$ , so by axiom (ENT4), we have  $\mathcal{H}(c) = \mathcal{H}(lf) \geq \mathcal{H}(f)$ .

Let now  $c : A \rightarrow B$  and  $c' : A' \rightarrow B'$  be two constant function. In view of Lemma 2.1 we may assume that  $B = \{b\}$  and  $B' = \{b'\}$ . By hypothesis, there exists a bijection  $\alpha : A \rightarrow A'$ . If  $\beta : B' \rightarrow B$  is the bijection given by  $\beta(b') = b$ , then  $c = \beta c \alpha$ , so  $\mathcal{H}(c) = \mathcal{H}(c')$  by ENT3.  $\square$

Let  $A$  be a finite set and let  $h(A)$  be the largest entropy of a function defined on  $A$ . Lemma 2.2 shows that for any constant function  $c$  defined on  $A$  we have  $\mathcal{H}(c) = h(A)$ .

**2.3. Corollary.** *There exists a mapping  $h : \mathbb{N} \rightarrow \mathbb{R}$  such that  $h(n)$  is the largest entropy of a function defined on a finite set having  $n$  elements for  $n \geq 2$ . Furthermore, we if  $h(2) = 1$ , then  $h(n) = n \log n$  for some  $k \in \mathbb{R}$ .*

*Proof.* The existence of the mapping  $h$  is an immediate consequence of Lemma 2.2. Suppose now that  $A, B$  are two finite sets and let  $c_A, c_B$  be two surjective constant functions defined on  $A$  and  $B$ , respectively. Observe that the function  $c_A \times c_B$  is a constant function defined on  $A \times B$ . Therefore,  $\mathcal{H}(c_A \times c_B) = |A|\mathcal{H}(c_A) + |B|\mathcal{H}(c_B)$  because of (ENT1). If  $|A| = m$  and  $|B| = n$  this amounts to  $h(mn) = mh(n) + nh(m)$  for every  $m, n \in \mathbb{N}$ .

If  $m \leq n$ , then by (ENT5), we have  $h(m) \geq h(n)$ . Since  $h(2) = 2$ , Theorem A.6 from Appendix A shows that  $h(n) = n \log n$ .  $\square$

**2.4. Theorem.** *If  $\mathcal{H}(f)$  satisfies the axioms (ENT1)-(ENT5) and*

$$\max\{\mathcal{H}(f) \mid f \text{ is defined over a set } A \text{ with } |A| = 2\} = 2,$$

*then, for every finite function  $f$  we have:*

$$\mathcal{H}(f) = \sum \{|f^{-1}(b)| \log |f^{-1}(b)| \mid b \in \text{Ran}(f)\}$$

*for some  $k \in \mathbf{R}$ .*

*Proof.* Let  $f : A \rightarrow B$  be a finite function, where  $B = \{b_0, \dots, b_{n-1}\}$ . Define  $A_i$  by  $A_i = \{a \in A \mid f(a) = b_i\} = f^{-1}(b_i)$  for  $0 \leq i \leq n-1$  and consider the constant functions  $c_i : A_i \rightarrow \{b_i\}$  for  $0 \leq i \leq n-1$ .

It is easy to see that  $f = c_0 \parallel \dots \parallel c_{n-1}$ , so

$$\mathcal{H}(f) = \sum_{0 \leq i \leq n-1} \mathcal{H}(c_i) = \sum_{0 \leq i \leq n-1} m_i \log m_i,$$

where  $m_i = |f^{-1}(b_i)|$  for  $0 \leq i \leq n-1$ .  $\square$

### 3. The Entropy of Equivalence Relations

A partition of a set  $A$  is defined as a collection of subsets of  $A$ ,  $\Pi = \{K_i \mid i \in I\}$  such that  $i \neq j$  implies  $K_i \cap K_j = \emptyset$  and  $\bigcup \{K_i \mid i \in I\} = A$ . The sets  $K_i$  are referred as the *blocks* of the partition  $\Pi$ .

A partition  $\Pi$  is finer than a partition  $\Pi'$  of a set  $A$  if for every block  $K_i$  of  $\Pi$  there is a block  $H_j$  of  $\Pi'$  such that  $K_i \subseteq H_j$ . In this case, we write  $\Pi \leq \Pi'$ . The set of partitions of a set  $A$  becomes a partial ordered set  $(\mathbf{Part}(A), \leq)$  which is, in fact, a lattice.

The set of equivalences of a set  $A$  is denoted by  $\mathbf{Equiv}(A)$ . It is a well known that  $(\mathbf{Equiv}(A), \subseteq)$  is also a lattice and is isomorphic to  $(\mathbf{Part}(A), \leq)$ .

Let  $\rho \in \mathbf{Equiv}(A)$  and  $\sigma \in \mathbf{Equiv}(C)$ . The *product*  $\rho \times \sigma$  is the equivalence on  $A \times C$  defined by  $((a, c), (a', c')) \in \rho \times \sigma$  if  $(a, a') \in \rho$  and  $(c, c') \in \sigma$  for every  $a, a' \in A$  and  $c, c' \in C$ .

If  $A \cap C = \emptyset$  and  $\rho \in \mathbf{Equiv}(A)$ ,  $\sigma \in \mathbf{Equiv}(C)$  the supremum of  $\{\rho, \sigma\}$  in the lattice  $(\mathbf{Equiv}(A \cup C), \subseteq)$  is the equivalence that corresponds to the partition that consists of the equivalence classes of  $\rho$  and the equivalence classes of  $\sigma$ . We use the notation  $\rho + \sigma$  for  $\sup\{\rho, \sigma\}$  to emphasize that  $A$  and  $C$  are supposed to be disjoint. If  $\Pi = \{H_i \mid i \in I\}$  is a partition of the set  $A$ , define the function  $\beta_\Pi : A \rightarrow \Pi$  by  $\beta(a) = H_i$  if  $a \in H_i$ . The entropy of the partition  $\Pi$  is defined as

$$\mathcal{H}(\Pi) = \mathcal{H}(\beta_\Pi).$$

In other words, we have  $\mathcal{H}(\Pi) = \sum_{i \in I} |H_i| \log |H_i|$ .

If  $\rho \in \mathbf{Equiv}(A)$  we define  $\mathcal{H}(\rho)$ , the entropy of  $\rho$  as being the entropy of the partition  $\Pi_\rho$  that consists of the equivalence classes of  $\rho$ .

We leave to the reader the verification of the next lemma.

**3.1. Lemma.** *Let  $h : \{x \in \mathbf{R} \mid x > 1\} \rightarrow \mathbf{R}$  be a function defined by*

$$h(x) = (x+a+1) \ln(x+a+1) + (x-1) \ln(x-1) - (x+a) \ln(x+a) - x \ln x.$$

*If  $a \geq 0$  we have  $h(x) > 0$  for every  $x \in \mathbf{R}$ ,  $x > 1$ .*

Lemma 3.1 shows that if an element of a set  $A$  “migrates” from a smaller block of a partition to a larger one, the entropy of the new partition is larger than the one of the previous partition. Therefore, the largest partition (that consists of a single block) is also the partition with the largest entropy.

A function  $f : A \rightarrow B$  is defined, up to a permutation of its range, by its kernel partition  $\Pi_f = \{f^{-1}(b) \mid f^{-1}(b) \neq \emptyset, b \in B\}$  or, by its kernel equivalence  $\ker_f = \{(a, a') \in A \times A \mid f(a) = f(a')\}$ . Observe that we have  $\mathcal{H}(f) = \mathcal{H}(\Pi_f)$ .

Let  $\rho$  be an equivalence,  $\rho \subseteq A \times A$ . If  $\alpha : A \rightarrow B$  is a bijection, then the  $\alpha$ -conjugate of  $\rho$  is the equivalence

$$\rho_\alpha = \{(b_1, b_2) \in B \times B \mid b_1 = \alpha(a_1), \\ b_2 = \alpha(a_2), \text{ and } (a_1, a_2) \in \rho\},$$

or, equivalently, the relation  $\alpha^{-1}\rho\alpha$ .

**3.2. Lemma.** For any functions  $f : A \rightarrow B$  and  $g : C \rightarrow D$  we have  $\ker_{f \times g} = \ker_f \times \ker_g$ . If  $A \cap C = B \cap D = \emptyset$ , then  $\ker_{f||g} = \ker_f + \ker_g$ .

For every function  $f : A \rightarrow B$  and bijections  $\alpha : D \rightarrow A, \beta : B \rightarrow E$  we have  $\ker_{\beta f \alpha} = (\ker_f)_\alpha$ .

*Proof.* We give the proof only for the last part of the Lemma. We have  $(d, d') \in \ker_{\beta f \alpha}$  if and only if  $\beta f \alpha(d) = \beta f \alpha(d')$  which happens if and only if  $f \alpha(d) = f \alpha(d')$  because  $\beta$  is a bijection. Therefore,  $\ker_{\beta f \alpha} = (\ker_f)_\alpha$ .  $\square$

**3.3. Theorem.** Let  $\mathcal{H}$  be a function defined on the class of all equivalences which satisfies the following axioms:

(EQENT1) If  $\rho \in \mathbf{Equiv}(A)$  and  $\sigma \in \mathbf{Equiv}(C)$ , then

$$\mathcal{H}(\rho \times \sigma) = |A|\mathcal{H}(\rho) + |C|\mathcal{H}(\sigma).$$

(EQENT2) If  $A \cap C = \emptyset$  and  $\rho \in \mathbf{Equiv}(A)$  and  $\sigma \in \mathbf{Equiv}(C)$ , then

$$\mathcal{H}(\rho + \sigma) = \mathcal{H}(\rho) + \mathcal{H}(\sigma).$$

(EQENT3) If  $\rho, \rho' \in \mathbf{Equiv}(A)$  and  $\rho \subseteq \rho'$ , then  $\mathcal{H}(\rho) \leq \mathcal{H}(\rho')$ .

(EQENT4) If  $\rho \in \mathbf{Equiv}(A)$  and  $\alpha : D \rightarrow A$ , then  $\mathcal{H}(\rho) = \mathcal{H}(\rho_\alpha)$ .

Then, for every equivalence  $\rho$  on  $A$  we have  $\mathcal{H}(\rho) = \sum_{i \in I} |H_i| \log |H_i|$ , where  $\{H_i \mid i \in I\}$  is the set of equivalence classes of  $\rho$ .

*Proof.* In order to show that the conditions are sufficient, consider the function  $\mathcal{K}$  defined on the class of all functions by  $\mathcal{K}(f) = \mathcal{H}(\ker_f)$ . We claim that  $\mathcal{K}$  satisfies the conditions ENT1-ENT5. Lemma 3.2 allows us to write:

$$\begin{aligned} \mathcal{K}(f \times g) &= \mathcal{H}(\ker_{f \times g}) = \mathcal{H}(\ker_f \times \ker_g) \\ &= |A|\mathcal{H}(\ker_f) + |C|\mathcal{H}(\ker_g) = |A|\mathcal{K}(f) + |C|\mathcal{K}(g), \end{aligned}$$

by (EQENT1).

Let  $A, B, C, D$  be four sets such that  $A \cap C = B \cap D = \emptyset$  and let  $f : A \rightarrow B, g : C \rightarrow D$  be two functions. Lemma 3.2 and axiom (EQENT2) allow us to write:

$$\begin{aligned} \mathcal{K}(f||g) &= \mathcal{H}(\ker_{f||g}) = \mathcal{H}(\ker_f + \ker_g) \\ &= \mathcal{H}(\ker_f) + \mathcal{H}(\ker_g) = \mathcal{K}(f) + \mathcal{K}(g). \end{aligned}$$

Let now  $f : A \rightarrow B$  be a function and let  $\alpha$  and  $\beta$  be bijections. By Lemma 3.2 and (EQENT4) we have

$$\begin{aligned} \mathcal{K}(\beta f \alpha) &= \mathcal{H}(\ker_{\beta f \alpha}) = \mathcal{H}((\ker_f)_\alpha) \\ &= \mathcal{H}(\ker_f) = \mathcal{K}(f). \end{aligned}$$

Assume now that  $f : A \rightarrow B$  and  $g : B \rightarrow C$ . If  $(a, a') \in \ker_f$  we have  $f(a) = f(a')$ , so  $g(f(a)) = g(f(a'))$  which gives  $(a, a') \in \ker_{gf}$ . Therefore,  $\ker_f \subseteq \ker_{gf}$ . By axiom (EQENT3) we have  $\mathcal{H}(\ker_f) \leq \mathcal{H}(\ker_{gf})$  and this gives  $\mathcal{K}(f) \leq \mathcal{K}(gf)$  thus showing that  $\mathcal{K}$  satisfies axiom (ENT4).

Since every equivalence  $\rho$  on a finite set  $A$  is such that  $\rho \subseteq \omega_A$ , where  $\omega_A = A \times A$ , it follows that  $\mathcal{H}(\ker_f) \leq \mathcal{H}(\ker_{\omega_A})$ . Thus, we obtain  $\mathcal{K}(f) \leq \mathcal{K}(k)$ , where  $k$  is a constant function on  $A$  (since the kernel equivalence of any constant function equals  $\omega_A$ ).

Let now  $A, A'$  be two finite sets such that  $|A| \leq |A'|$ . There exists an injection  $f : A \rightarrow A'$  and we can write  $A' = B \cup (A' - B)$ , where  $B = f(A)$ . If  $\rho$  is an equivalence on  $A$ , define the equivalence  $\rho^f$  on  $A'$  by  $\rho^f = \rho_f + \omega_{A'-B}$ . By (EQENT2) we have

$$\begin{aligned} \mathcal{H}(\rho^f) &= \mathcal{H}(\rho_f) + \mathcal{H}(\omega_{A'-B}) \\ &= \mathcal{H}(\rho) + \mathcal{H}(\omega_{A'-B}) \geq \mathcal{H}(\rho). \end{aligned}$$

Then, if  $\rho = \omega_A$ , it follows immediately that  $\mathcal{H}(\omega_{A'}) \geq \mathcal{H}(\omega_A^f) \geq \mathcal{H}(\omega_A)$ . This, in turn, implies that  $\mathcal{K}(k') \geq \mathcal{K}(k)$ , where  $k, k'$  are constant functions defined on  $A$  and  $A'$  respectively, so  $\mathcal{K}(\cdot)$  satisfies axiom (ENT5).

Let  $f : A \rightarrow B$  a function between the finite sets  $A$  and  $B$ . Since  $\mathcal{K}(\cdot)$  satisfies all axioms of the entropy of functions we have  $\mathcal{K}(f) = \sum_{b \in B} |f^{-1}(b)| \log |f^{-1}(b)|$ .

Let  $\rho$  be an arbitrary equivalence on a set  $A$ . Since  $\rho = \ker_{f_\rho}$ , where  $f_\rho(a) = [a]_\rho$ , it follows that  $\mathcal{H}(\rho) = \mathcal{K}(f_\rho) = \sum_{i \in I} |H_i| \log |H_i|$ , where  $\{H_i | i \in I\}$  is the set of equivalence classes of  $\rho$ . □

The necessity of axioms (EQENT1)-(EQENT4) is obvious and Theorem 3.3 gives, therefore, a uniqueness result for entropies of equivalences.

#### 4. Entropies of Classes of Functions

The closer a function  $f$  is to an injection the smaller its entropy is; in particular, the entropy of an injection is equal to zero. It is interesting to note that iterating a function increases the entropy up to a certain amount.

Let  $f : \{0, \dots, n-1\} \rightarrow \{0, \dots, n-1\}$  be a function and let  $u, v \in \mathbb{N}$  be two natural numbers. If  $u > v$  axiom (ENT4) implies  $\mathcal{H}(f^u) \geq \mathcal{H}(f^v)$  because  $f^u = f^{u-v} f^v$ .

If  $\kappa(n)$  is the least common multiple of all numbers  $1, \dots, n$ ,  $p \geq n-1$ , an  $q \equiv p \pmod{\kappa(n)}$ , then  $f^p = f^q$  for every function  $f : \{0, \dots, n-1\} \rightarrow \{0, \dots, n-1\}$  (see [4]). Therefore, there exists a number  $h \leq n-1$  such that

$$\mathcal{H}(f) < \dots < \mathcal{H}(f^{h-1}) < \mathcal{H}(f^h) = \mathcal{H}(f^{h+1}) = \dots$$

for every  $n$ -valued function  $f : \{0, \dots, n-1\} \rightarrow \{0, \dots, n-1\}$ .

Determining the entropy of some function may point towards the impossibility of using certain circuit modules as component of other circuits. For instance, consider the functions  $f, g : \{0, 1, 2\}^2 \rightarrow \{0, 1, 2\}$  defined by

$$\begin{aligned} f(x_0, x_1) &= x_0 + x_1 \pmod{3} \\ g(x_0, x_1) &= \begin{cases} 0 & \text{if } x_0 + x_1 \leq 2 \\ 2 & \text{otherwise,} \end{cases} \end{aligned}$$

for every  $x_0, x_1 \in \{0, 1, 2\}$ . Clearly,  $f$  is the addition modulo 3 and  $g$  is a threshold function. It is easy to verify that  $\mathcal{H}(f) = 14.26$  and  $\mathcal{H}(g) = 20.26$ . Therefore, we may conclude that it is impossible to decompose  $f$  as  $f = hg$  for some function  $h : \{0, 1, 2\} \rightarrow \{0, 1, 2\}$ .

Let  $f : E_k^n \rightarrow E_k$  be an  $n$ -argument function over the set of logical values  $E_k = \{0, \dots, k-1\}$ . We saw that the maximum entropy of such a function is achieved for constants. Therefore, the maximum entropy of such a function is  $k^n \log k^n = nk^n \log k$ . The minimum is achieved when all values in the range have equal preimages, and so, the minimum is  $\sum_1^k k^{n-1} \log k^{n-1} = (n-1)k^n \log k$ .

**4.1. Definition.** The *entropic index* of a function  $f : E_k^n \rightarrow E_k$  is the quantity

$$\epsilon(f) = \frac{\mathcal{H}(f) - (n-1)k^n \log k}{k^n \log k}.$$

Since  $(n-1)k^n \log k \leq \mathcal{H}(f) \leq nk^n \log k$ , it follows that  $0 \leq \epsilon(f) \leq 1$ .

The upper (lower) entropic index of a class of functions  $\mathcal{F} \subseteq E_k^n \rightarrow E_k$  is  $\epsilon_1(\mathcal{F}) = \sup\{\epsilon(f) | f \in \mathcal{F}\}$ , and  $\epsilon_0(\mathcal{F}) = \inf\{\epsilon(f) | f \in \mathcal{F}\}$ , respectively. The *entropic range* of  $\mathcal{F}$  is the number  $\epsilon_1(\mathcal{F}) - \epsilon_0(\mathcal{F})$ .

**4.2. Lemma.** If  $\mathcal{F}$  contains any constant function, then  $\epsilon_1(\mathcal{F}) = 1$ ; if  $\mathcal{F}$  contains any projection, then  $\epsilon_0(\mathcal{F}) = 0$ .

Note that for every clone  $\mathcal{C}$ , we have  $\epsilon_0(\mathcal{C}) = 0$  because every clone contains the projections. Every maximal clone of operations on  $E_k$  (see [5]) contains constants with the exception of the maximal clone of self-dual functions. Therefore, the entropic range of such clones equals 1.

If  $p$  is a prime divisor of  $k$  and  $\pi : E_k \rightarrow E_k$  is a permutation of  $E_k$  that has  $k/p$  cycles of the same length  $p$ , we denote by  $\rho_{kp}(\pi)$  the binary relation:

$$\rho_{kp}(\pi) = \{(x, y) | x, y \in E_k, y = \pi(x)\}$$

The maximal clone of  $\pi$ -self-dual functions on  $E_k$  is defined as the set of all functions  $f : E_k^n \rightarrow E_k$  that preserve  $\rho_{kp}(\pi)$ . In other words, if  $(x_i, y_i) \in \rho_{kp}(\pi)$  for  $1 \leq i \leq n$ , then  $(f(x_1, \dots, x_n), f(y_1, \dots, y_n)) \in \rho_{kp}(\pi)$ . This is equivalent to saying

$$\pi(f(x_1, \dots, x_n)) = f(\pi(x_1), \dots, \pi(x_n)),$$

for every  $x_1, \dots, x_n \in E_k$ .

**4.3. Theorem.** Let  $p$  be a prime divisor of  $k$ .

If  $\mathcal{C}$  is the class of  $\pi$ -self-dual functions on  $E_k$ , where  $\pi$  is a permutation of  $E_k$  that has  $k/p$  cycles, then  $\epsilon_1(\mathcal{C}) = 1 - \log_k p$ .

*Proof.* For an element  $c$  of  $E_k$ , the elements  $c, \pi(c), \dots, \pi^{p-1}(c)$  are pairwise distinct. If  $(x_1, \dots, x_n) \in A_c^f$ , then  $(\pi^i(x_1), \dots, \pi^i(x_n)) \in A_{\pi^i(c)}^f$  for  $0 \leq i \leq p-1$ . The maximum entropy is achieved when there exists  $c$  such that every  $(x_1, \dots, x_n)$  is mapped by  $f$  in one of the elements  $\pi^i(c)$  for some  $i$ ,  $0 \leq i \leq p-1$ . Since, in this case, all sets  $A_{\pi^i(c)}^f$  will contain  $k^n/p$  elements the maximal entropy is

$$\sum_{1 \leq i \leq p} \frac{k^n}{p} \log \frac{k^n}{p} = k^n \log \frac{k^n}{p}.$$

Therefore, for such a function  $f$  we have:

$$\begin{aligned} \epsilon(f) &= \frac{k^n \log \frac{k^n}{p} - (n-1)k^n \log k}{k^n \log k} \\ &= \frac{\log k - \log p}{\log k} \\ &= 1 - \frac{\log p}{\log k} = 1 - \log_k p. \end{aligned}$$

Therefore  $\epsilon_1(\mathcal{F}) = 1 - \log_k p$ . □

There are many non-trivial classes of functions whose entropic range equals 1.

**4.4. Definition.** A function  $f : E_k^n \rightarrow E_k$  is *symmetric* if for every permutation  $\pi : E_k \rightarrow E_k$  we have  $f(x_1, \dots, x_n) = f(\pi(x_1), \dots, \pi(x_n))$ , for every  $x_1, \dots, x_n \in E_k$ .

A function  $f : E_k^n \rightarrow E_k$  is *linear* if there exist  $a_0, \dots, a_n \in E_k$  such that  $f(x_1, \dots, x_n) = a_0 + a_1 \cdot x_1 + \dots + a_n \cdot x_n$ , for every  $x_1, \dots, x_n$ . Here the multiplication and the addition are both modulo  $k$ .

A *threshold function* is a function  $f : E_k^n \rightarrow E_k$  if there exists a  $n$ -tuple  $(w_1, \dots, w_n) \in E_k^n$  and a  $(k - 1)$ -tuple  $(t_1, \dots, t_{k-1}) \in E_k^{k-1}$  such that

$$f(x_1, \dots, x_n) = \begin{cases} 0 & \text{if } \sum_{i=1}^n w_i x_i < t_1 \\ i & \text{if } t_i \leq \sum_{i=1}^n w_i x_i < t_{i+1}, \\ & \text{for } 1 \leq i \leq k - 2, \\ k - 1 & \text{if } \sum_{i=1}^n w_i x_i \geq t_{k-1}, \end{cases}$$

for every  $x_1, \dots, x_n$ . Here  $\sum_{i=1}^n w_i x_i$  is taken as a scalar product in  $\mathbf{R}^n$ .

We will denote the class of symmetric function by  $\mathcal{S}$  and the class of threshold functions by  $\mathcal{T}$ .

**4.5. Theorem.** For both  $\mathcal{S}$  and  $\mathcal{T}$  the entropic range equals 1.

*Proof.* Since constant functions are symmetric we have  $\epsilon_1(\mathcal{S}) = 1$ . On the other hand, the linear function  $f(x_1, \dots, x_n) = x_1 + \dots + x_n$  is also symmetric and we have  $|A_j^f| = k^{n-1}$  for every  $j$ ,  $0 \leq j \leq k - 1$  (as it can be verified easily by induction on  $n$ ). Therefore  $\epsilon(f) = 0$ , so  $\epsilon_0(\mathcal{S}) = 0$ .

Again, constant functions are threshold functions and thus  $\epsilon_1(\mathcal{T}) = 1$ . On the other hand, a function  $f : E_k^n \rightarrow E_k$  is threshold if the lattice of  $k^n$  points of the cube  $E_k^n$  can be separated by  $k - 1$  parallel hyperplanes into  $k$  layers so that  $f$  is constant on each layer and its values do not decrease as we move in a direction perpendicular on the hyperplanes. Therefore, if the vector  $w = (w_1, \dots, w_n)$  is chosen appropriately, it is possible to separate the  $k^n$  points into equal size layers and thus to achieve the minimum entropy. □

### 5. Further Problems

Other aspects of entropy, such as conditional entropy of a function, patterned after the standard notion of entropy of a probabilistic entropy seem to be worth looking into. Further study is needed in order to examine the behaviour of entropy with respect to standard operations on equivalences (intersection, union of permutable equivalences, etc.)

### 6. Acknowledgement

The authors would like to acknowledge helpful discussions with Herbert Kamowitz and Ethan Bolker and the contribution of Steven Scheiberg who supplied us with Theorem A.6.

### Appendix A. A Technical Result

The following technical Lemma is a special case of a result of Dirichlet (see[7], pp.235).

**A.1. Lemma.** Let  $\alpha$  be a real number and let  $Q$  be a positive integer. There exists  $m$  such that  $1 \leq m < Q$  and an integer  $n$  such that  $|m\alpha - n| < \frac{1}{Q} < \frac{1}{m}$ .

*Proof.* The argument is left to the reader. □

Another useful result is contained in the next lemma; we omit the proof.

**A.2. Lemma.** If  $n \leq m\epsilon < n + \epsilon$ , then there exist  $m', n'$  such that  $n' - \epsilon < m'\epsilon < n'$ .  
Similarly, if  $n - \epsilon < m\epsilon < n$ , there exist  $m'', m''$  such that  $n'' \leq m''\epsilon < n'' + \epsilon$ .

**A.3. Lemma.** If  $h : \mathbf{N} \rightarrow \mathbf{R}$  is a function such that

$$h(mn) = mh(n) + nh(m),$$

for every  $m, n \in \mathbf{N}$ , then  $h(p^k) = kp^{k-1}h(p)$  for every  $p, k \in \mathbf{N}$  and  $k \geq 1$ .

*Proof.* The argument is by induction on  $k$  and it is left to the reader. □

**A.4. Lemma.** If  $h : \mathbf{N} \rightarrow \mathbf{R}$  is a function such that

$$h(mn) = mh(n) + nh(m),$$

for every  $m, n \in \mathbf{N}$ , then

$$h(p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}) = p_1^{k_1} p_2^{k_2} \dots p_n^{k_n} \sum_{1 \leq i \leq n} \frac{k_i h(p_i)}{p_i}.$$

*Proof.* Let  $\ell : \mathbf{N} \rightarrow \mathbf{R}$  be the function given by

$$\ell(n) = \begin{cases} 0 & \text{if } n = 0, \\ \frac{h(n)}{n} & \text{if } n > 0. \end{cases}$$

Note that  $\ell(mn) = \ell(m) + \ell(n)$  and, therefore,

$$\ell(p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}) = \ell(p_1^{k_1}) + \ell(p_2^{k_2}) + \dots + \ell(p_n^{k_n}).$$

Since  $\ell(p^k) = \frac{k}{p} h(p)$  (because of Lemma A.3 we obtain:

$$\ell(p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}) = \sum_{1 \leq i \leq n} \frac{k_i h(p_i)}{p_i},$$

which gives immediately the equality of the Lemma. □

**A.5. Corollary.** Let  $h : \mathbf{N} \rightarrow \mathbf{R}$  be a function such that  $h(p) = p \log p$  if  $p = 1$  or if  $p$  is prime. If  $h(mn) = mh(n) + nh(m)$  for every  $m, n \in \mathbf{N}$ , then  $h(n) = n \log n$  for every  $n \in \mathbf{N}$ ,  $n \geq 1$ .

*Proof.* Since every positive integer  $n$  other than 1 can be written uniquely as a product of powers of primes  $n = p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}$ , we have

$$\begin{aligned} h(n) &= p_1^{k_1} p_2^{k_2} \dots p_n^{k_n} \sum_{1 \leq i \leq n} \frac{k_i h(p_i)}{p_i} \\ &= n \sum_{1 \leq i \leq n} k_i \log p_i \\ &= n \log n, \end{aligned}$$

for  $n \geq 2$ . □

The following result is due to S. Scheinberg [6]:

**A.6. Theorem.** Let  $h : \mathbf{N} \rightarrow \mathbf{R}$  be an increasing function such that

$$h(mn) = mh(n) + nh(m)$$

for every  $m, n \in \mathbf{N}$ . If  $h(2) = 2$ , then  $h(n) = n \log n$  for  $n \in \mathbf{N}$ .

*Proof.* Define the function  $b : \{n \in \mathbf{N} | n > 1\} \rightarrow \mathbf{R}$  by  $b(n) = h(n)/(n * \log n)$ .

We shall prove initially that if  $p > 2$  is a prime number, then  $b(p) \geq 1$ . Let  $\epsilon > 0$  be a real number. Taking  $Q < 1/\epsilon$  in Lemma A.1 we obtain the existence of  $m, n \in \mathbf{N}$  such that  $|m\alpha - n| < \epsilon$ . In other words, we have  $n - \epsilon < m\alpha < n + \epsilon$ . If  $n < m\alpha < n + \epsilon$ , then by Lemma A.2, there are  $m', n'$  such that  $n' - \epsilon < m'\epsilon < n'$ . If  $n - \epsilon < m\alpha < n$ , then the same lemma implies the existence of  $n'', m''$  such that  $n'' \leq m''\epsilon < n'' + \epsilon$ .

If we choose  $\alpha = \log p$ , then we may assume that there are  $m, n \in \mathbf{N}$ ,  $m, n \geq 1$  such that  $n \leq m \log p < n + \epsilon$ . Equivalently, we have  $2^n \leq p^m < 2^{n+\epsilon}$ . Since  $h$  is an increasing function, we obtain  $n2^n \leq h(p^m)$ , or  $n2^n \leq mp^{m-1}h(p)$ . Because of the definition of  $b$  we have  $n2^n \leq mp^m b(p) \log p$ , or  $n2^n \leq b(p)p^m \log p^m$ . In view of the previous inequality, this implies

$$n2^n \leq b(p)2^{n+\epsilon}(n + \epsilon),$$

or, equivalently,

$$b(p) \geq \frac{n}{2^\epsilon(n + \epsilon)}.$$

Taking  $\epsilon \rightarrow 0$  we obtain  $b(p) \geq 1$ .

Similarly, there exists a number  $m \in \mathbf{N}$  such that  $n - \epsilon < m \log p \leq n$ . A similar argument which makes use of Lemma A.2 shows that  $b(p) \leq 1$ , so  $b(p) = 1$ , which proves that  $h(p) = p \log p$  for every prime  $p$ .  $\square$

### References

- [1] Ingarden, R.S., Urbanik, K.: *Information without Probability*, Coll. Math., **9**, 281-304, 1962.
- [2] R enyi, A.: *On a Theorem of P. Erdős and its Applications in Information Theory*, *Mathematica*, **24**, pp. 341-344, 1959.
- [3] Lee, P. M.: *On the Axioms of Information Theory*, *Ann. Math. Statist.*, **35**, pp. 416-418, 1964.
- [4] Reischer, C., Simovici, D.A.: *Several Remarks on Iteration Properties of Switching Functions*, *Proceedings of the 12th International Symposium on Multiple-Valued Logic*, Paris, 1982, pp. 244-247.
- [5] Rosenberg, I.: *Completeness Properties of Multiple-Valued Logic Algebras*, ch. 6 in *Computer Science and Multiple-Valued Logic*, edited by D.C. Rine, North-Holland, 1984, pp. 150-192.
- [6] Scheinberg, S: *personal communication*.
- [7] Zygmund, A.: *Trigonometric Series*, Cambridge Mathematical Library, Cambridge University Press, 1990, 2nd edition.

### Authors' addresses

Corina Reischer  
University of Quebec  
at Trois-Rivieres,  
Dept. of Math. & Comp. Sci.  
Quebec, G9A 5H7, Canada

Dan Simovici  
Univ. of Massachusetts  
at Boston,  
Dept. of Math. & Comp. Sci.  
Boston, 02125 USA

Ivan Stojmenovic  
Univ. of Ottawa,  
Computer Science Dept.,  
Ottawa, Ontario K1N 9B4,  
Canada