RAIRO. INFORMATIQUE THÉORIQUE

J. ACZÉL
Z. DARÓCZY

A mixed theory of information. I : symmetric, recursive and measurable entropies of randomized systems of events

RAIRO. Informatique théorique, tome 12, n° 2 (1978), p. 149-155
<http://www.numdam.org/item?id=ITA_1978__12_2_149_0>

R.A.I.R.O. Informatique théorique/Theoretical Computer Science
(vol. 12, n° 2, 1978, p. 149 à 155)

A MIXED THEORY OF INFORMATION. I: SYMMETRIC, RECURSIVE AND MEASURABLE ENTROPIES OF RANDOMIZED SYSTEMS OF EVENTS (*) (1)

by J. ACZÉL (2) and Z. DARÓCZY (3)

Abstract. — *The paper contains the first result in a mixed theory of information, where measures of information may depend both upon the events and their probabilities. All such entropics that are* 3-*symmetric, recursive and measurable are determined.*

1. In the *probabilistic* theory of information (*see*, e. g., [3]) the entropies and other measures of information or uncertainty are supposed to depend *solely* upon the probabilities of the events (messages, outcomes of an experiment, weather, market situations, answers to a questionnaire, etc.). On the other hand, in the *nonprobabilistic* theory of information (*see*, e. g., [4, 7]) these measures do *not* depend upon the probabilities *at all*, only directly upon the events themselves.

After a result of B. Forte [5] in the similar case of random variables, one of us has proposed in [1, 2] a *mixed* theory of information, where measures of information may depend *both* upon the events *and* their probabilities. The present paper contains the first result in this direction. Generalizing an important theorem of Lee [9], we determine all 3-symmetric, recursive, and measurablee ntropies depending upon a system of events and their probabilities, which we will call a *randomized system of events*. We will also refer to entropies of randomized systems of events in short as "inset entropies" (inset: a map set within another map; but one may also consider it (1) as "in set"). Under the above conditions, they turn out to be essentially the sum of a Shannon entropy and of the expected value of a random variable.

2. Let B be a ring of sets (containing, with any two sets also their union and their difference, thus also their intersection and the empty set 0; *see* [6]). Denote

$$\Omega_n = \{(x_1, x_2, \ldots, x_n) \,|\, x_i \in B,\ x_i \cap x_j = 0 \text{ if } i \neq j;\ i, j = 1, 2, \ldots, n\}$$

(*) Reçu 14 juillet, 1977.

(1) This paper has been conceived at the meeting Colloque international du Centre national de Recherche scientifique, Les développements récents de la théorie d'information et leurs applications, organized by C.-F. Picard, July 4-8, 1977 in E.N.S.E.T. at Cachan, France.

(2) University of Waterloo, Ontario, Canada.
(3) L. Kossuth University, Debrecen, Hungary.

and

$$\Gamma_n = \left\{ (p_1, p_2, \ldots, p_n) \,\middle|\, \sum_{i=1}^{n} p_i = 1,\ p_i \geqq 0;\ i = 1, 2, \ldots, n \right\}$$

$(n = 2, 3, \ldots)$. We call

$$\begin{pmatrix} x_1, & x_2, & \ldots, & x_n \\ p_1, & p_2, & \ldots, & p_n \end{pmatrix} \in \Omega_n \times \Gamma_n,$$

a randomized system of events. We use *events* x_i as name for the elements of B, while the p_i are *probabilities*.

The sequence of mappings (inset entropy) $I_n : \Omega_n \times \Gamma_n \to R$ $(n = 2, 3, \ldots;$ R the set of reals) is *recursive* if, for all integers $n > 2$, and all

$$\begin{pmatrix} x_1, & x_2, & \ldots, & x_n \\ p_1, & p_2, & \ldots, & p_n \end{pmatrix} \in \Omega_n \times \Gamma_n,$$

$$I_n\begin{pmatrix} x_1, & x_2, & x_3, & \ldots, & x_n \\ p_1, & p_2, & p_3, & \ldots, & p_n \end{pmatrix} = I_{n-1}\begin{pmatrix} x_1 \cup x_2, & x_3, & \ldots, & x_n \\ p_1 + p_2, & p_3, & \ldots, & p_n \end{pmatrix} + (p_1+p_2) I_2 \begin{pmatrix} x_1, & x_2 \\ \dfrac{p_1}{p_1+p_2}, & \dfrac{p_2}{p_1+p_2} \end{pmatrix},$$

with the convention $0 . I_2 \begin{pmatrix} x_1, & x_2 \\ 0/0, & 0/0 \end{pmatrix} := 0$. This states how the uncertainty changes if an event is split into two; it is also connected to Huffman codes and algorithms. The sequence $\{ I_n \}$ is *k-symmetric* ($k \geqq 2$) if

$$I_k\begin{pmatrix} x_1, & \ldots, & x_k \\ p_1, & \ldots, & p_k \end{pmatrix} = I_k\begin{pmatrix} x_{r(1)}, & \ldots, & x_{r(k)} \\ p_{r(1)}, & \ldots, & p_{r(k)} \end{pmatrix},$$

for all $\begin{pmatrix} x_1, & \ldots, & x_k \\ p_1, & \ldots, & p_k \end{pmatrix} \in \Omega_k \times \Gamma_k$ and all permutations r on $\{ 1, 2, \ldots, k \}$ (meaning simply that the uncertainty does not depend upon the labelling of events). Finally, our inset entropy is *measurable* if the function

$$t \mapsto I_2\begin{pmatrix} x_1, & x_2 \\ 1-t, & t \end{pmatrix}, \tag{1}$$

is measurable on $]0, 1[$ for all fixed $(x_1, x_2) \in \Omega_2$.

THEOREM: *The sequence* $I_n : \Omega_n \times \Gamma_n \to R$ $(n = 2, 3, \ldots)$ *is recursive, 3-symmetric and measurable if, and only if, there exists a constant A and a* *function* $g : B \to R$ *such that*

$$I_n\begin{pmatrix} x_1, & \dots, & x_n \\ p_1, & \dots, & p_n \end{pmatrix} = g\left(\bigcup_{i=1}^{n} x_i\right) - \sum_{i=1}^{n} p_i g(x_i) - A \sum_{i=1}^{n} p_i \log p_i, \tag{2}$$

for all $\begin{pmatrix} x_1, & \dots, & x_n \\ p_1, & \dots, & p_n \end{pmatrix} \in \Omega_n \times \Gamma_n$ $(n = 2, 3, \dots)$ *with the convention*

$$0.\log 0 := 0. \tag{3}$$

3. *Proof:* It is obvious that any inset entropy given by (2) with arbitrary $A \in R$ and $g : B \to R$ is recursive, symmetric and measurable. Now we prove the converse.

Recursivity means, for $n = 3$,

$$I_3\begin{pmatrix} x_1, & x_2, & x_3 \\ p_1, & p_2, & p_3 \end{pmatrix} = I_2\begin{pmatrix} x_1 \cup x_2, & x_3 \\ p_1 + p_2, & p_3 \end{pmatrix} + (p_1+p_2) I_2\begin{pmatrix} x_1, & x_2 \\ \dfrac{p_1}{p_1+p_2}, & \dfrac{p_2}{p_1+p_2} \end{pmatrix}, \tag{4}$$

for all $\begin{pmatrix} x_1, & x_2, & x_3 \\ p_1, & p_2, & p_3 \end{pmatrix} \in \Omega_3 \times \Gamma_3$. We introduce a function $f : \Omega_2 \times [0, 1] \to R$ by

$$f(x_1, x_2; t) = I_2\begin{pmatrix} x_1, & x_2 \\ 1-t, & t \end{pmatrix}, \tag{5}$$

cf. (1).

Let $s \in [0, 1[$, $t \in [0, 1[$, $s+t \leqq 1$, but s and t else arbitrary. Then, from (4) and from the 3-symmetry, we have

$$\begin{aligned} & f(x_1 \cup x_2, x_3; t) + (1-t) f\left(x_1, x_2; \frac{s}{1-t}\right) \\ &= I_3\begin{pmatrix} x_1, & x_2, & x_3 \\ 1-s-t, & s, & t \end{pmatrix} = I_3\begin{pmatrix} x_1, & x_3, & x_2 \\ 1-s-t, & t, & s \end{pmatrix} \\ &= f(x_1 \cup x_3, x_2; s) + (1-s) f\left(x_1, x_3; \frac{t}{1-s}\right), \end{aligned} \tag{6}$$

for all $(x_1, x_2, x_3) \in \Omega_3$ and for all

$$(s, t) \in D := \{(s, t) \mid s \in [0, 1[, t \in [0, 1[, s+t \leqq 1\}.$$

For fixed $(x_1, x_2, x_3) \in \Omega_3$, we get from (6) with the notations

$$\left.\begin{aligned} f_1(s) &= f(x_1 \cup x_3, x_2; s), & f_2(u) &= f(x_1, x_3; u), \\ f_3(t) &= f(x_1 \cup x_2, x_3; t), & f_4(v) &= f(x_1, x_2, v), \end{aligned}\right\} \tag{7}$$

the equation

$$f_1(s)+(1-s)f_2\left(\frac{t}{1-s}\right)=f_3(t)+(1-t)f_4\left(\frac{t}{1-s}\right) \quad \text{for all } (s, t)\in D.$$

The general solutions, measurable on $]0, 1[$, have been determined for this equation in [8] (*cf.* [3]) as

$$\left.\begin{array}{l} f_j(t)=A[-t\log t-(1-t)\log(1-t)]+a_j t+b_j \\ \quad (t\in[0, 1[\text{ or } [0, 1];\, j=1, 2, 3, 4), \end{array}\right\} \tag{8}$$

with the convention (3). (There are certain relations among b_1, b_2, b_3 and b_4, which we will not need here. It is also unimportant how we fix the base of the logarithm.)

In the situation described by (7), when x_1, x_2 and x_3 are allowed to vary again, the coefficients A, a_j, b_j ($j = 1, 2, 3, 4$) in (8) may depend upon them. In particular, *see* (7),

$$\begin{aligned} &f(x_1, x_2;\, t) \\ &\quad = A(x_1, x_2)[-t\log t-(1-t)\log(1-t)]+a_4(x_1, x_2)t+b_4(x_1, x_2) \\ &\qquad f(x_1\cup x_2, x_3;\, t) \\ &\qquad\quad = A(x_1\cup x_2, x_3)[-t\log t-(1-t)\log(1-t)] \\ &\qquad\quad +a_3(x_1\cup x_2, x_3)t+b_3(x_1\cup x_2, x_3). \end{aligned} \tag{9}$$

But, as seen from (8), A has to be the same for f_3 and f_4, thus

$$A(x_1, x_2)=A(x_1\cup x_2, x_3) \quad \text{for all } (x_1, x_2, x_3)\in\Omega_3.$$

Substituting $x_1 = 0$, we get

$$A(x_2, x_3)=A(0, x_2). \tag{10}$$

So $A(x, y) = \alpha(x)$ is independent of y. Thus, combined with (10), we have that $\alpha(x_2) = \alpha(0) =$ constant, that is,

$$A \text{ is constant.} \tag{11}$$

If we substitute (9), with constant A, into (6) and compare the members linear in t on the left and right hand sides, we obtain, writing simply

$$a_4=a, \qquad b_4=b, \tag{12}$$

the equation

$$a(x_1\cup x_2, x_3)-b(x_1, x_2)=a(x_1, x_3) \quad \text{for all } (x_1, x_2, x_3)\in\Omega_3. \tag{13}$$

[We will *not* need the other equations obtainable by comparison of the two extremities of (6).] The substitution $x_3 = 0$ now gives, with the notation

$$g(x) = a(x, 0),$$

the equation

$$b(x_1, x_2) = g(x_1 \cup x_2) - g(x_1). \tag{14}$$

Resubstituting this into (13), we get

$$a(x_1 \cup x_2, x_3) - g(x_1 \cup x_2) = a(x_1, x_3) - g(x_1)$$

and, again with $x_1 = 0$,

$$a(x_2, x_3) = g(x_2) - G(x), \tag{15}$$

where we have written

$$G(x) = g(0) - a(0, x).$$

From (5), (9), (11), (12), (14) and (15) we have now

$$I_2\begin{pmatrix} x_1, & x_2 \\ 1-t, & t \end{pmatrix} = A\left[-t\log t - (1-t)\log(1-t)\right] + g(x_1 \cup x_2) - (1-t)\,g(x_1) - t\,G(x_2), \tag{16}$$

[with (3)]. But equation (4) and the 3-symmetry

$$I_3\begin{pmatrix} x_1, & x_2, & x_3 \\ p_1, & p_2, & p_3 \end{pmatrix} = I_3\begin{pmatrix} x_2, & x_1, & x_3 \\ p_2, & p_1, & p_3 \end{pmatrix},$$

show that I_2 is symmetric too (that is, our inset entropies are also 2-symmetric). Thus

$$I_2\begin{pmatrix} x_1, & x_2 \\ 1-t, & t \end{pmatrix} = I_2\begin{pmatrix} x_2, & x_1 \\ t, & 1-t \end{pmatrix}.$$

Comparison to (16) gives immediately

$$G(x) = g(x),$$

so that (16) goes over into

$$I_2\begin{pmatrix} x_1, & x_2 \\ 1-t, & t \end{pmatrix} = g(x_1 \cup x_2) - (1-t)\,g(x_1) - \mathrm{t}g(x_2) - A\left[(1-t)\log(1-t) + t\log t\right] \tag{17}$$

[with the convention (3)].

This shows that (2) holds for $n = 2$. Suppose it is true for $n-1$ then, by the recursivity and by (17),

$$
\begin{aligned}
I_n&\begin{pmatrix} x_1, & x_2, & \ldots, & x_n \\ p_1, & p_2, & \ldots, & p_n \end{pmatrix} \\
&= I_{n-1}\begin{pmatrix} x_1 \cup x_2, & x_3, & \ldots, & x_n \\ p_1 + p_2, & p_3, & \ldots, & p_n \end{pmatrix} + (p_1+p_2) I_2 \begin{pmatrix} x_1, & x_2 \\ \dfrac{p_1}{p_1+p_2}, & \dfrac{p_2}{p_1+p_2} \end{pmatrix} \\
&= g(x_1 \cup x_2 \cup \ldots \cup x_n) - (p_1+p_2) g(x_1 \cup x_2) \\
&\quad - \sum_{i=3}^{n} p_i g(x_i) - A\left[(p_1+p_2)\log(p_1+p_2) + \sum_{i=3}^{n} p_i \log p_i\right] \\
&\quad + (p_1+p_2)\left[g(x_1 \cup x_2) - \frac{p_1}{p_1+p_2} g(x_1) - \frac{p_2}{p_1+p_2} g(x_2)\right. \\
&\qquad\qquad \left. - A\frac{p_1}{p_1+p_2}\log\frac{p_1}{p_1+p_2} - A\frac{p_2}{p_1+p_2}\log\frac{p_2}{p_1+p_2}\right] \\
&= g\left(\bigcup_{i=1}^{n} x_i\right) - \sum_{i=1}^{n} p_i g(x_1) - A \sum_{i=1}^{n} p_i \log p_i
\end{aligned}
$$

(again with the convention (3), using the similar convention in the definition of recursivity), that is, (2) holds also for n. This concludes the proof.

4. REMARKS: The last member, $-\sum p_i \log p_i$ in (2) [with the convention (3)] is, of course, the *Shannon entropy* (*see*, e. g., [3]). If the system $x_1, x_2, \ldots, x_n$ of events is *complete*, that is, $\bigcup_{i=1}^{n} x_i$ is the whole space Ω (the certain event), then $g\left(\bigcup_{i=1}^{n} x_i\right) = C$ is a constant and, with the notation $h(x) = C - g(x)$, the first two members in (2) reduce to

$$\sum_{i=1}^{n} p_i h(x_i),$$

that is, to the *expected value of a random variable* [which the second member in (2) is also in the general case]. Thus, *in this case of complete systems of events, the general recursive,* 3-*symmetric and measurable inset entropies are sums of the expected value of an arbitrary random variable and of an arbitrary constant multiple of the Shannon entropy,*

$$I_n\begin{pmatrix} x_1, & \ldots, & x_n \\ p_1, & \ldots, & p_n \end{pmatrix} = \sum_{i=1}^{n} p_i h(x_i) - A \sum_{i=1}^{n} p_i \log p_i,$$

$$\left[(x_1, \ldots, x_n) \in \Omega_n, \bigcup_{i=1}^{n} x_i = \Omega;\ (p_1, \ldots, p_n) \in \Gamma_n, 0 \log 0 := 0\right].$$

There is a close resemblance between this representation and C. T. Ng's parallel composition law (5.8.C) in [10].

On the other hand, in the case of *incomplete* systems of events (when their union is a proper subset of the "whole space"), we may notice that the sum of probabilities is still 1. This means that we have *conditional* probabilities [observe, for instance, the probabilities assigned to x_1 and x_2 in the last member of the definition of recursivity, for instance in (4)] or measures geared to the union of the events (sets) in the inset entropy.

ACKNOWLEDGEMENTS

The authors are grateful for remarks made by C.T. Ng in connection with a previous version of this paper. The first auhor's research was supported in part by a National Research Council of Canada grant.

REFERENCES

1. J. Aczél, *Some Recent Results on Characterizations of Measures of Information* Colloque International du C.N.R.S., Les Développements Récents de la Théorie de l'Information et leurs Applications, E.N.S.E.T., Cachan, 4-8 juillet, 1977.
2. J. Aczél, *Some Recent Results on Characterizations of Measures of Information, Related to Coding*, I.E.E.E. Trans. Information Theory, Vol. IT-24, 1978.
3. J. Aczél and Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, New York, San Francisco, London, 1975.
4. B. Forte, *Measures of Information. The General Axiomatic Theory*, R.A.I.R.O., Vol. 3, série R-2, 1969, pp. 63-89.
5. B. Forte, *Subadditive Entropies for a Random Variable*, Boll. Un. Mat. Ital., (5), Vol. 14 B, 1977, pp. 118-133.
6. P. R. Halmos, *Measure Theory*, Van Nostrand, Princeton, N.J., Toronto, London, New York, 1950.
7. J. Kampé de Fériet, *La théorie généralisée de l'information et la mesure subjective d'information. Théories de l'Information*, Actes des Rencontres de Marseille-Luminy, 5-7 juin 1973, Springer, Berlin, Heidelberg, New York, 1974, pp. 1-15.
8. Pl. Kannappan and C. T. Ng, *Measurable Solutions of Functional Equations Related to Information Theory*, Proc. Amer. Math. Soc., Vol. 38, 1973, pp. 303-310.
9. P. M. Lee, *On the Axioms of Information Theory*, Ann. Math. Statist., Vol. 35, 1963, pp. 415-418.
10. C. T. Ng, *Universal Parallel Composition Laws and Their Representations*, Math. Scand., Vol. 40, 1977, pp. 25-45.