Taking into account the Remarks 2.1 and 2.2, we have to discuss the numerical stability of the methods $(\alpha A A^*)$ with respect to initial and rounding errors. An answer to this problem is given by the following.

Theorem 3.1. *Let* $\tilde{u}_k(\alpha)$, $\tilde{\dot{u}}_k(\alpha)$, $\tilde{u}_{k-\alpha}(\alpha)$, $\tilde{\dot{u}}_{k-\alpha}(\alpha)$ $k=\overline{1, N_1}$, *be the values given by the method* $(\alpha A A^*)$ *with the starting set* $\mathcal{M}^{(1)}_{AA^*}(\alpha)$. *If the hypotheses of the Theorem 3.1. hold and if there is a positive function* $R^*_p(\alpha)$ *defined on* R *such that,*

$$\max\left(\|\tilde{\rho}_{p,k}(\alpha)\|, \|\tilde{\tau}_{p,k}(\alpha)\|, \|\tilde{\rho}^*_{p,k}(\alpha)\|, \|\tilde{\tau}^*_{p,k}(\alpha)\| ;\ k=\overline{1, N_1}\right) \leqslant R^*_p(\alpha), \tag{3.19}$$

where $\tilde{\rho}_{p,k}(\alpha)$, $\tilde{\tau}_{p,k}(\alpha)$ *and* $\tilde{\rho}^*_{p,k}(\alpha)$, $\tilde{\tau}^*_{p,k}(\alpha)$ *are the terms appearing in (2.32) and (2.33) respectively, then for every* $\alpha \in R/Z$ *the method* $(\alpha A A^*)$ *is stable with respect to the initial and rounding errors.*

Proof. The proof, is similar to that of the Theorem 3.1. One shows that the sequences of vectors $(\tilde{\rho}_{p,k}(\alpha))$, $(\tilde{\tau}_{p,k}(\alpha))$, $(\tilde{\rho}^*_{p,k}(\alpha))$, $(\tilde{\tau}_{p,k}(\alpha))$, $(\dot{u}_k(\alpha) - \tilde{\dot{u}}_k(\alpha))$, $(u_k(\alpha) - \tilde{u}_k(\alpha))$, $(\dot{u}_{k-\alpha}(\alpha) - \tilde{\dot{u}}_{k-\alpha}(\alpha))$, $(u_{k-\alpha}(\alpha) - \tilde{u}_{k-\alpha}(\alpha))$ $k=\overline{1, N_1}$; $(\dot{u}_{j-\alpha}(\alpha) - \tilde{\dot{u}}_{j-\alpha}(\alpha))$, $(u_{j-\alpha}(\alpha) - \tilde{u}_{j-\alpha}(\alpha))$, $(f(t_{j-\alpha}, u_{j-\alpha}(\alpha), \dot{u}_{j-\alpha}(\alpha)) - f(t_{j-\alpha}, \tilde{u}_{j-\alpha}(\alpha), \tilde{\dot{u}}_{j-\alpha}(\alpha)))$ $j=\overline{-(p+1), N_1}$; $A_s(\alpha)$, $\overline{A}_s(\alpha)$, $A^*_s(\alpha)$, $\overline{A}^*_s(\alpha)$ $s=\overline{0, p}$, satisfy the conditions of the Lemma 2.1 with $R(\alpha) = R^*_p(\alpha)$, $E(\alpha) = \mathcal{J}(\alpha)$ and $L, K, G(\alpha)$ entering in the Theorem 3.1. In accordance with (2.11) we get

$$\mathcal{R}_k(\alpha) \leqslant (1 + hM(\alpha))^{k+1}\mathcal{J}(\alpha) + (1 + hM(\alpha))\frac{\overline{R}^*_p(\alpha)}{h},\ k=\overline{1, N_1}, \tag{3.20}$$

where

$$M(\alpha) = q(\alpha) L G_p(\alpha)(2p+1+T-t_0),\quad \overline{R}^{**}_p(\alpha) = 2q(\alpha)(T-t_0)(T-t_0+2)R^*_p(\alpha).$$

Remark 3.4. By Theorem 3.2 we obtain the next estimation for the total α-error up the step k of the family $(\alpha A A^*)$.

$$\mathcal{T}_k(\alpha) \leqslant (\mathcal{J}(\alpha) + h^{p+1}(\overline{R}_p(\alpha) + \overline{R}^*_p(\alpha) h^{-p-2}))\, e^{(T-t_0)M(\alpha)},\ k=\overline{1, N_1}. \tag{3.21}$$

We note that if $\mathcal{J}(\alpha)$ is of the order $O(h^{p+1})$ and $\overline{R}^*_p(\alpha)$ is of the order $O(h^{p+2})$, then $\mathcal{T}_k(\alpha)$ has the order $O(h^{p+1})$.

REFERENCES

1. Calistru N. — *A family of multistep methods for second-order differential equations*, (to appear).

Received 25. VIII. 1980

Department of Mathematics
Polytechnic Institute of Iaşi
Iaşi, R. S. România

Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi
Tomul XXVII, s. I a, 1981 f. 2

WHY SHANNON'S ENTROPY : A FURTHER REASON

BY

CARLO SEMPI

1. The definition of conditional entropy. Additive and subadditive entropies were characterized in [1] (here the notation of [2] will be adopted). For any finite discrete complete probability distribution $p = (p_1, p_2, \ldots, p_n) \in \Gamma_n := \{(q_1, q_2, \ldots, q_n) : q_i > 0, \sum_{i=1}^{n} q_i = 1\}$ $(n = 2, 3, \ldots)$ these entropies have the representation

$$\mathcal{H}(\mathcal{P}) := H_n(p_1, p_2, \ldots, p_n) = a\, H^s_n(p_1, p_2, \ldots, p_n) + b \log \mathcal{N}(\mathcal{P}), \tag{1}$$

where a and b are non-negative constants, $\mathcal{N}(\mathcal{P})$ denotes the number of non-zero componets in the distribution $\mathcal{P}$ and H^s_n is the Shannon's entropy

$$H^s_n(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i.$$

In order to define the conditional entropy for the additive and subadditive entropies (1), we shall consider the joint distribution $\mathcal{P} \times Q = \{\pi_{ij} ;\ i = 1, 2, \ldots, m ;\ j = 1, 2, \ldots, n\} \in \Gamma_{mn}$ together with its marginals $\pi = \left\{\sum_{j=1}^{n} \pi_{ij} p_i\right\} \in \Gamma_m$ and $Q = \left\{\sum_{i=1}^{m} \pi_{ij} = q_j\right\} \in \Gamma_n$. The conditional distributions $(q_{i1}, q_{i2}, \ldots, q_{in}) \in \Gamma_n$ $(i = 1, 2, \ldots, m)$, where $\pi_{ij} = p_i q_{ij}$ will also be useful.

As is well known, Shannon's entropy satisfies the relationship

$$\mathcal{H}(\mathcal{P} \times Q) = \mathcal{H}(\mathcal{P}) + \mathcal{H}(Q/\mathcal{P}),$$

which can also be used to define the conditional entropy (see [2]). This we shall do with the entropy (1).

A simple calculation shows that for (1), one has

$$\mathcal{H}(Q/\mathcal{P}) = \mathcal{H}(\mathcal{P} \times Q) - \mathcal{H}(\mathcal{P}) = -a \sum_{i=1}^{m} p_i \sum_{j=1}^{n} q_{ij} \log q_{ij} + b \log \mathcal{N}(\mathcal{P} \times Q)/\mathcal{P}. \tag{2}$$

This relations supports the following

Definition. *The conditional entropy* $\mathcal{H}(Q/\mathcal{P})$ *for the additive and subadditive entropies (1) is defined as being equal to the r. h. s. of (2) for every distribution* $\mathcal{P} \times Q \in \Gamma_{mn}$ $(m, n = 2, 3, \ldots)$.

2. A review of the properties of Shannon's conditional. entropy The first term in (2) is the Shannon's conditional entropy. Thus, for the sake of easy reference, it is useful to list its properties (see [2], [3], [4] e. g.). Let $\mathcal{P}$,$\mathcal{Q}$ and $\mathcal{R}$ be complete probability distributions. Then the following holds for Shannon's conditional entropy

(i) $\mathcal{H}(\mathcal{Q}/\mathcal{P}) \geqslant 0$,

(ii) $\mathcal{H}(\mathcal{Q}/\mathcal{P}) \leqslant \mathcal{H}(\mathcal{Q})$,

(iii) $\mathcal{H}(\mathcal{Q}/\mathcal{P}) = \mathcal{H}(\mathcal{Q})$ *if $\mathcal{P}$ and $\mathcal{Q}$ are independent*,

(iv) $\mathcal{H}(\mathcal{P}\times\mathcal{Q}\times\mathcal{R}) = \mathcal{H}(\mathcal{P}) + (\mathcal{H}(\mathcal{Q}/\mathcal{P}) + \mathcal{H}(\mathcal{R}/\mathcal{P}\times\mathcal{Q}) = \mathcal{H}(\mathcal{P}\times\mathcal{Q}) + \mathcal{H}(\mathcal{R}/\mathcal{P}\times\mathcal{Q})$

(v) $\mathcal{H}(\mathcal{Q}\times\mathcal{R}/\mathcal{P}) = \mathcal{H}(\mathcal{Q}/\mathcal{P}) + \mathcal{H}(\mathcal{R}/\mathcal{P}\times\mathcal{Q})$,

(vi) $\mathcal{H}(\mathcal{Q}\times\mathcal{R}/\mathcal{P}) \leqslant \mathcal{H}(\mathcal{Q}/\mathcal{P}) + \mathcal{H}(\mathcal{R}/\mathcal{P})$,

(vii) $\mathcal{H}(\mathcal{R}/\mathcal{P}\times\mathcal{Q}) \leqslant \mathcal{H}(\mathcal{R}/\mathcal{P})$.

3. The properties of (2). By virtue of the remark made at the beginning of section 2, it suffices to prove, or disprove, the above properties for the Hartley's conditional entropy

$$\mathcal{H}^{H}(\mathcal{Q}/\mathcal{P}) := b \log \mathcal{N}(\mathcal{P}\times\mathcal{Q})/\mathcal{N}(\mathcal{P}), \tag{3}$$

that is the last term in (2).

We shall start by proving the properties that (3) does satisfy

Theorem 3.1. *The Hartley's conditional entropy satisfies the properties (i) through (v).*

Corollary *the conditional entropy (2) satisfies the properties (i) through (v).*

The following simple lemmas will be needed in order to prove the theorem.

Lemma 1. $\mathcal{N}(\mathcal{P}) \leqslant \mathcal{N}(\mathcal{P}\times\mathcal{Q})$.

Proof. Since $\mathcal{P} = \{p_i = \sum_{j=1}^{n} \pi_{ij} : i = 1, 2, \ldots, m\}$ and $\pi_{ij} \geqslant 0$ $(i = 1, 2, \ldots, m\,;\ j = 1, 2, \ldots, n)$, to every non-zero element of $\mathcal{P}$ there corresponds at least a non-zero element of $\mathcal{P}\times\mathcal{Q}$; hence the assertion.

Lemma 2. $\mathcal{N}(\mathcal{P}\times\mathcal{Q}) \leqslant \mathcal{N}(\mathcal{P})\,\mathcal{N}(\mathcal{Q})$.

Proof. It is possible to regard the probabilities π_{ij} as the entries of a non-negative matrix π of order $m\times n$. Thus p_i is obtained by summing the entries in the i-th row and q_j by summing the entries in the j-th column. At least one of the entries of Π is positive since one has $\sum_{i=1}^{m}\sum_{j=1}^{n}\pi_{ij} = 1$. The result will be proved by finite induction.

If $\mathcal{N}(\mathcal{P}\times\mathcal{Q}) = 1$, then Π has exactly one positive entry (equal to 1) so that $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{Q}) = 1$ and the assertion is proved (equality holding).

Let the matix Π be modified in such a way that the number of non-zero entries is increased by one (and the other entries are modified in order to keep normalization). Thus $\mathcal{N}(\mathcal{P}'\times\mathcal{Q}') = 2$ and, at the same time either $\mathcal{N}(\mathcal{P}') = 2$ or $\mathcal{N}(\mathcal{Q}') = 2$ or both, so that in every case one has $2 = \mathcal{N}(\mathcal{P}'\times\mathcal{Q}') \leqslant \mathcal{N}(\mathcal{P}')\,\mathcal{N}(\mathcal{Q}')$.

Generally, assume $\mathcal{N}(\mathcal{P}\times\mathcal{Q}) = k$ with $k < mn$. Then modify the distribution $\mathcal{P}\times\mathcal{Q}$, and hence the matrix Π, by increasing the number of non-zero entries by one. Thus $\mathcal{N}(\mathcal{P}'\times\mathcal{Q}') = k+1$. In order to conclude the proof, we shall show that if $\mathcal{N}(\mathcal{P}\times\mathcal{Q}) \leqslant \mathcal{N}(\mathcal{P})\mathcal{N}(\mathcal{Q})$ then $\mathcal{N}(\mathcal{P}'\times\mathcal{Q}') \leqslant \mathcal{N}(\mathcal{P}')\mathcal{N}(\mathcal{Q}')$.

Now if $\mathcal{N}(\mathcal{P}\times\mathcal{Q}) < \mathcal{N}(\mathcal{P})\mathcal{N}(\mathcal{Q})$, the proof is immediate since $\mathcal{N}(\mathcal{P}) \leqslant \mathcal{N}(\mathcal{P}')$ and $\mathcal{N}(\mathcal{Q}) \leqslant \mathcal{N}(\mathcal{Q}')$ entail $\mathcal{N}(\mathcal{P}'\times\mathcal{Q}') = \mathcal{N}(\mathcal{P}\times\mathcal{Q}) + 1 \leqslant \mathcal{N}(\mathcal{P})\mathcal{N}(\mathcal{Q}) \leqslant \mathcal{N}(\mathcal{P}')\,\mathcal{N}(\mathcal{Q}')$.

Then assume $\mathcal{N}(\mathcal{P}\times\mathcal{Q}) = \mathcal{N}(\mathcal{P})\mathcal{N}(\mathcal{Q}) = k$. This means that two positive integers r ans s exist such that $k = rs$ and such that there r rows in π each with exactly s non-zero entries placed in such a way that there are exactly s columns with non-zero entries (these numbering r). Thus either $\mathcal{N}(\mathcal{P}') = \mathcal{N}(\mathcal{P}) + 1$ or $\mathcal{N}(\mathcal{P}') = \mathcal{N}(\mathcal{P}) + 1$ or both so that.

$$\mathcal{N}(\mathcal{P}'\times\mathcal{Q}') = k + 1 \leqslant \mathcal{N}(\mathcal{P})\,\mathcal{N}(\mathcal{Q}) + \min\{\mathcal{N}(\mathcal{P}), \mathcal{N}(\mathcal{Q})\} \leqslant \mathcal{N}(\mathcal{P}')\mathcal{N}(\mathcal{Q}'). \quad \text{Q.E.D.}$$

Remark. If $\mathcal{P}$ and $\mathcal{Q}$ are independent, then lemma 2 reads

$$\mathcal{N}(\mathcal{P}\times\mathcal{Q}) = \mathcal{N}(\mathcal{P})\,\mathcal{N}(\mathcal{Q}). \tag{4}$$

Proof of the theorem. (i) is an immediate consequence of (3) and of lemma. 1.

(ii) follows from lemma 2 ; one has in fact

$$\mathcal{H}^{H}(\mathcal{Q}/\mathcal{P}) = b \log \mathcal{N}(\mathcal{P}\times\mathcal{Q})/\mathcal{N}(\mathcal{P}) \leqslant b \log \mathcal{N}(\mathcal{P})\,\mathcal{N}(\mathcal{Q})/\mathcal{N}(\mathcal{Q}) = b \log \mathcal{N}(\mathcal{P}) = \mathcal{H}^{H}(\mathcal{Q}).$$

(iii) follows from (4) in the same way as (ii) ensued from lemma 2.

$$\mathcal{H}^{H}(\mathcal{P}\times\mathcal{Q}\times\mathcal{R}) = b \log \mathcal{N}(\mathcal{P}\times\mathcal{Q}\times\mathcal{R}) = b \log \frac{\mathcal{N}(\mathcal{P}\times\mathcal{Q}\times\mathcal{R})\mathcal{N}(\mathcal{P}\times\mathcal{Q})\mathcal{N}(\mathcal{P})}{\mathcal{N}(\mathcal{P}\times\mathcal{Q})\mathcal{N}(\mathcal{P})} =$$

(iv)
$$= b \log \frac{\mathcal{N}(\mathcal{P}\times\mathcal{Q}\times\mathcal{R})}{\mathcal{N}(\mathcal{P}\times\mathcal{Q})} + b \log \frac{\mathcal{N}(\mathcal{P}\times\mathcal{Q})}{\mathcal{N}(\mathcal{P})} + b \log \mathcal{N}(\mathcal{P}) =$$

$$= \mathcal{H}^{H}(\mathcal{R}/\mathcal{P}\times\mathcal{Q}) + \mathcal{H}^{H}(\mathcal{Q}/\mathcal{P}) + \mathcal{H}^{H}(\mathcal{P}).$$

$$\mathcal{H}^{H}(\mathcal{Q}\times\mathcal{R}/\mathcal{P}) = b \log \frac{\mathcal{N}(\mathcal{P}\times\mathcal{Q}\times\mathcal{P})}{\mathcal{N}(\mathcal{P})} = b \log \frac{\mathcal{N}(\mathcal{P}\times\mathcal{Q}\times\mathcal{R})\mathcal{N}(\mathcal{P}\times\mathcal{Q})}{\mathcal{N}(\mathcal{P}\times\mathcal{Q})\mathcal{N}(\mathcal{P})} =$$

(v)
$$= b \log \mathcal{N}(\mathcal{P}\times\mathcal{Q}\times\mathcal{R})/\mathcal{N}(\mathcal{P}\times\mathcal{Q}) + b \log \mathcal{N}(\mathcal{P}\times\mathcal{Q})/\mathcal{N}(\mathcal{P}) =$$

$$= \mathcal{H}^{H}(\mathcal{R}/\mathcal{P}\times\mathcal{Q}) + \mathcal{H}^{H}(\mathcal{Q}/\mathcal{P}). \quad \text{Q. E. D.}$$

It may be somewhat surprising that Hartley's entropy (3) does not satisfy properties (vi) and (vii) as will be shown by mens of an example.

One can easily check that (v) and (vi) imply (vii) and conversely that (vi) follows from (v) and (vii). Thus, since (v) has already been proved, it will suffice to show that one of (vi) and (vii) does not hold in order to show that both are generally false. We can therefore focus our attention on the property (vii). A trivial computation reveals that (vii) is equivalent, in the case of Hartleys conditional entropy (3), to the statement.

$$\mathcal{H}(\mathcal{P} \times \mathcal{Q} \times \mathcal{R})\,\mathcal{H}(\mathcal{P}) \leqslant \mathcal{H}(\mathcal{P} \times \mathcal{Q})\,\mathcal{H}(\mathcal{P} \times \mathcal{R}). \tag{5}$$

It would not be hard to list cases in which (5) holds. For instance, whenever $\mathcal{H}(\mathcal{P})=1$, the inequility (5) is true. However (5) is not true in general, as the following example shows.

Let $\mathcal{P} \times \mathcal{Q} \times \mathcal{R} \in \Gamma_{mnr}$ be such that the only positive entries are π_{111}, π_{121}, π_{112}, π_{122} and π_{211}. Thus $\mathcal{H}(\mathcal{P} \times \mathcal{Q} \times \mathcal{R})=5$, $\mathcal{H}(\mathcal{P} \times \mathcal{Q})=\mathcal{H}(\mathcal{P} \times \mathcal{R})=3$ and $\mathcal{H}(\mathcal{P})=2$; therefore

$$\mathcal{H}(\mathcal{P} \times \mathcal{Q} \times \mathcal{R})\,\mathcal{H}(\mathcal{P})=10>9=\mathcal{H}(\mathcal{P} \times \mathcal{Q})\,\mathcal{H}(\mathcal{P} \times \mathcal{R}).$$

4. Final remarks. The properties (vi) and (vii) of Shannon's conditional entropy are not translated into analogous properties of Hartley's conditional entropy, and, as a consequence, of the conditional entropy (2), with $b>0$ This has an important negative effect on the application of the additive and subadditive entropies (1), at least as long as one adopt the definition of conditional entropy used in this note. In fact the property (vii) is essential (see [3] or [4]) in order to define the entropy of a sequence of discrete random variables. This is no longer possible in view of our result. More important, it is even impossible to speak about an analogue of Mac Millan's theorem in connexion with the additive and subadditive entropies (1), if $b>0$. Thus one is inevitably lead to the choice $b=0$, thus justifying what has always been done in the approach to Statistical Mechanics through Information Theory. There it is entirely natural to use the additive and subadditive entropies (see [5] and [6]), but nevertheless one has always chosen $b=0$.

It is an open question whether a different definition of conditional entropy would yield also the crucial properties (vi) and (vii).

REFERENCES

1. Aczél, J., B. Forte and C. T. Ng. — *Why the Shannon and Hartley Entropies are 'Natural*, Advances Appl. Probability 6, 131—146 (1974).
2. Aczél J. and Z. Daróczy — *On Measures of Information and Their Characterization*, Academic Press, New York-San Fran isco-London, 1975.
3. Ash R. B. — *Information Theory*, Interscience, New York, 1965.
4. Ash R. B. and M. F. Gardner — *Topics in Stochastic Processes*, Academic Press, New York-San Francisco-London, 1975.
5. Sempi C. — *Entropies in Information Theory and Statistical Mechanics. Related Methods*, Ph. D. Thesis, University of Waterloo, Ontario, Canada, 1974.
6. Forte B. and C. Sempi — *Maximizing Conditional Entropies : a Derivation of Quantal Statistics*, Rend. Mat. (6), 9, 551—566 (1976).

Received 20. VI. 1980

Istituto di Matematica
dell'Università,
Lecce
73100 Italy

Analele ştiinţifice ale Universităţii „Al.I.Cuza" din Iaşi
Tomul XXVII, s.I a. 1981, f.2

REMARKS ON DISTANCES IN GRAPHS

By

ROBERT A. MELTER and IOAN TOMESCU

It is standard practice to define the distance $d(x, y)$ between two points in a connected graph as the length of the shortest path between x and y. In [2] the Boolean distance $b(x, y)$ is defined to be the set of all points on all paths from x to y, a path from x to y containing, by definition [1], pairwise defferent nodes only.

This distance satisfies the following properties :

1) $b(x, y)=\varnothing$ if and only if $x=y$;

2) $b(x, y)=b(y, x)$ for any points x, y ;

3) $b(x, y)\subset b(x, z)\cup b(z, y)$ for any three points x, y, z of G.

In this paper we first investigate certain connections between the numerical distance and the Boolean distance.

For a connected graph G having point set X with $|X|=n$ let $\rho(x, y)=|b(x, y)|$ for any $x, y\in X$. This function is a numerical metric on X. Indeed. $\rho(x, y)=0$, if and only if $d(x, y)=\varnothing$ or $x=y$, $\rho(x, y)=\rho(y, x)$ and for any three points $x, y, z \in X$:

$$\rho(x, y)=|b(x, y)|\leqslant |b(x, z)\cup b(z, y)|\leqslant |b(x, z)|+|b(z, y)|=$$
$$=\rho(x, z)+\rho(z, y).$$

For any point $x\in X$. let $e(x)=\max\limits_{y\in X} \rho(x, y)$.

Definition 1. *A point $x_0\in X$ will be called an s-center of G if $e(x_0)=\min\limits_{y\in X} e(x)$.*

If T is a tree then it is clear that $\rho(x, y)$ $(x\neq y)$ is greater by one than the numerical distance between x and y. Hence centers and s-centers as defined above coincide for trees. In Figure 1 the number next to each point x is equal to $e(x)$.

We derive $\min e(x)=8$ and the s-centers are two points which lie in an acyclic block ; they are marked with small circles. The graph has a unique center ; it is adjacent to an s-center and is marked with a small square. From the definition it follows that s-centers are „near" the big blocks of the graph. However, this notion is not useful for 2-connected graphs, since if G is 2-connected. $b(x, y)=X$ for any $x\neq y$ and the s-radius of G equals $n=|X|$ and any point is an s-center. It is well known [1] that the centers