RAIRO. INFORMATIQUE THÉORIQUE

SILVIU GUIAŞU
CORINA REISCHER

Some remarks on entropic distance, entropic measure of connexion and Hamming distance

RAIRO. Informatique théorique, tome 13, nº 4 (1979), p. 395-407
<http://www.numdam.org/item?id=ITA_1979__13_4_395_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 13, n° 4, 1979, p. 395 à 407)

SOME REMARKS ON ENTROPIC DISTANCE, ENTROPIC MEASURE OF CONNEXION AND HAMMING DISTANCE (*)

by Silviu GUIAŞU (1) and Corina REISCHER (2)

Abstract. — *The relationship between entropic distance, entropic measure of connexion or interdependence and Hamming distance is investigated. Some applications to the classification of the families of curves are also given.*

Résumé. — *On étudie la relation entre la distance entropique, la mesure entropique de connexion ou interdépendance et la distance de Hamming, avec quelques applications à la classification de familles de courbes.*

1. INTRODUCTION

Using Shannon's conditional entropy, it is possible to define an interesting global entropic distance on the set of finite probability spaces (*see* Horibe [3]). On the other hand Watanabe [7] has introduced an entropic measure of connexion or interdependence between finite probability spaces, extending Shannon's information rate from communication theory. Finally, Hamming distance between vectors is an important tool in algebraic coding theory. In the present paper we intend to establish some connexions between these three concepts. In the second paragraph the relationship between entropic distance and entropic measure of interdependence is analysed. In the third paragraph both entropic measure of interdependence and entropic distance are used for classifying the families of curves. Any such family may be classified according either to the monotony of the curves or to the values taken on by these curves. In the fourth paragraph the relationship between entropic distance and Hamming distance is investigated. Some advantages of entropic distance for coping with insertions and deletions of symbols during transmission or for synchronization are underlined.

(*) Reçu juin 1978.
(1) Faculty of Mathematics, University of Bucharest, Romania.
(2) Dept. of Mathematics, University of Quebec at Trois-Rivières, Canada.

R.A.I.R.O. Informatique théorique/Theoretical Informatics, 0399-0540/1979/395/$ 4.00

2. ENTROPIC DISTANCE AND ENTROPIC MEASURE OF CONNEXION

By X, Y, Z we denote finite probability spaces and $H(X)$ is the entropy of the finite probability space X. The entropic distance (Shannon's metric) between X and Y is (according to Horibe [3]):

$$\rho(X, Y)=H(X \mid Y)+H(Y \mid X). \tag{1}$$

According to Watanabe [7], the entropic measure of connexion or interdependence between X and Y is

$$W(X \otimes Y; X, Y)=H(X)+H(Y)-H(X \otimes Y), \tag{2}$$

where $X \otimes Y$ denotes the product probability space. For brevity, we shall write $H(X, Y)$ and $W(X, Y)$ instead of $H(X \otimes Y)$ and $W(X \otimes Y; X, Y)$ respectively.

The connexion between the entropic distance and the entropic measure of interdependence is given by the following proposition:

PROPOSITION 1: *We have*

$$\rho(X, Y)=H(X, Y)-W(X, Y). \tag{3}$$

Proof: From the well-known property of the entropy of product probability spaces we have

$$H(X \mid Y)=H(X, Y)-H(Y), \tag{4}$$

$$H(Y \mid X)=H(X, Y)-H(X), \tag{5}$$

where $H(X \mid Y)$ is the entropy of X conditioned by Y. Introducing these equalities into (1) and taking into account (2) we get (3).

Q.E.D.

REMARK: According to proposition 1, the entropic distance between X and Y shows us how many uncertainty we still have on the product probability space $X \otimes Y$ (or on the product probabilistic experiment $X \otimes Y$) if we remove the interdependence (i. e. the connexion) between X and Y.

PROPOSITION 2: *We have*

$$0 \leqq \rho(X, Y) \leqq H(X)+H(Y). \tag{6}$$

Proof: From (4) and (5) we have

$$H(X \mid Y) \leqq H(X, Y), \qquad H(Y \mid X) \leqq H(X, Y),$$

and then

$$0 \leqq \max\{H(X \mid Y), H(Y \mid X)\} \leqq H(X, Y) \leqq H(X)+H(Y). \tag{7}$$

Also, because

$$W(X,\ Y)=H(X)-H(X \mid Y)=H(Y)-H(Y \mid X),$$

we get

$$0 \leqq W(X,\ Y) \leqq \max\{H(X),\ H(Y)\}. \qquad (8)$$

Therefore (6) holds, where the second inequality becomes equality if X and Y are independent.

Q.E.D.

REMARKS: 1) if for any elementary event x of X there is an elementary event y_x of Y such that $p(y_x \mid x)=1$, then $H(Y \mid X)=0$,

and

$$\rho(X,\ Y)=H(X,\ Y)-W(X,\ Y)=H(X)+H(Y \mid X)-H(X)+H(Y \mid X)=0;$$

2) if X and Y are independent then

$$H(X,\ Y)=H(X)+H(Y), \qquad W(X,\ Y)=0,$$

which implies

$$\rho(X,\ Y)=H(X)+H(Y).$$

3. CLASSIFICATION OF FAMILIES OF CURVES

We may apply the entropic measure of connexion for classifying the families of curves with respect to their interaction. If a family of curves is given we may study their interactions and classify them according to: (*a*) *monotony*; (*b*) *vertical connexion* (or Riemann interaction); (*c*) *horizontal connexion* (or Lebesgue interaction).

For giving an example let us take a family of four curves, defined by the following functions: for the curve e_k the corresponding function is

$$y=(-1)^r\{T^2\,2^{-2(k+1)}-[x-(2r+1)\,T2^{-(k+1)}]^2\}^{1/2}, \qquad (9)$$

if

$$2r\,T2^{-(k+1)} \leqq x \leqq 2(r+1)\,T2^{-(k+1)}, \qquad (10)$$

where $r=0,\ 1,\ 2,\ \ldots,\ 2^k-1$; $k=1,\ 2,\ 3,\ 4$.

Of course, each function has a period equal to the half of the period of the previous function (see *fig*. 1). Thus, the curve e_k has the period $T2^{-(k-1)}$.

(*a*) *Classification according to the monotony.*

We shall take into account the partition of the interval $[0,\ T]$ generated by the centers and the extremities of the 16 semicircles of the curve e_4. This partition contains 32 intervals of length $T2^{-5}$. Taking into account the monotony of the

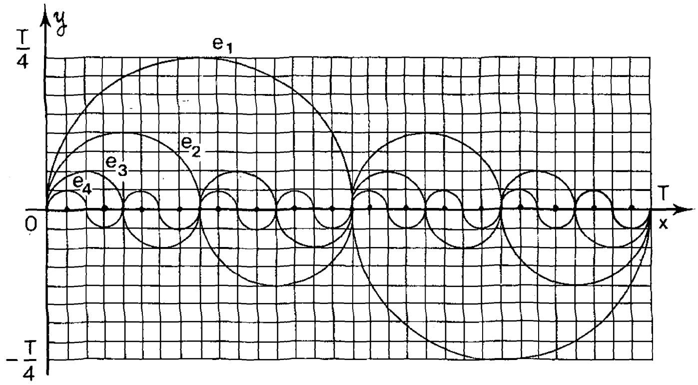

Figure 1

functions e_k $(k=1, 2, 3, 4)$, on these intervals we obtain the table I, where we put 0 if the function e_k increase and 1 if the function e_k decreases on the respective interval of the partition. Now, each curve $\{e_k\}$ may be considered as a finite probability space with two elementary events (0 and 1) whose probabilities are the corresponding relative frequencies which may be computed without any

TABLE I

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
e_1...	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
e_2...	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
e_3...	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
e_4...	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
e_1...	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
e_2...	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
e_3...	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
e_4...	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

difficulty from the table I. The product probability space $\{e_i, e_j\}$ has four elementary events, namely the four columns

$$\begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{matrix}$$

whose relative frequencies may be computed from the rows e_i and e_j of the table I. Similarly for the product probability spaces $\{e_i, e_j, e_k\}$ and $\{e_1, e_2, e_3, e_4\}$. We get

$$H(\{e_i\})=\log_2 2=1 \qquad (i=1, 2, 3, 4);$$
$$H(\{e_i, e_j\})=\log_2 4=2 \qquad (i, j=1, 2, 3, 4;\ i\neq j);$$
$$H(\{e_i, e_j, e_k\})=\log_2 8=3 \qquad (i, j, k=1, 2, 3, 4;\ i\neq j\neq k),$$
$$H(\{e_1, e_2, e_3, e_4\})=\log_2 16=4;$$
$$W(\{e_i\}, \{e_j\})=H(\{e_i\})+H(\{e_j\})-H(\{e_i, e_j\})=0 \qquad (i\neq j);$$
$$W(\{e_i, e_j\}, \{e_k\})=0; \qquad W(\{e_i, e_j\}, \{e_k, e_r\})=0 \qquad (i\neq j\neq k\neq r);$$
$$W(\{e_i\}, \{e_j\}, \{e_k\})=H(\{e_i\})+H(\{e_j\})+H(\{e_k\})-H(\{e_i, e_j, e_k\})=0;$$
$$W(\{e_i, e_j\}, \{e_k\}, \{e_r\})=0; \qquad W(\{e_1\}, \{e_2\}, \{e_3\}, \{e_4\})=0.$$

We obtain here a family of curves which is totally independent. Accordingly, the same analysis may be maid taking $k>4$ in (9). With respect to the monotony we have here no interdependence at all between any disjoint sets of curves. This computation gives a justification to the conjecture formulated by Greek mathematician and philosopher Pythagoras according to which the perfect harmony may be obtained by a set of oscillating strings for which each oscillating string has its wave-length equal to the half of the wave-length of other oscillating string, like in our example. In such a case we have no interference between oscillating strings. The classification is given in figure 2.

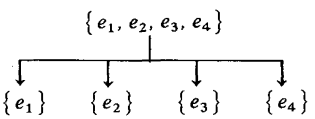

Figure 2

(*b*) *Classification according to the vertical connexion.*

Returning to our example, we may classify the family of four curves according to the values taken on by the respective curves in the points determined by the ends of the intervals (10) of the partition considered above. We have 32 such points. Let us denote

$$a=T2^{-5}\sqrt{15}; \qquad b=T2^{-4}\sqrt{7}; \qquad c=T2^{-5}\sqrt{39};$$
$$d=T2^{-3}\sqrt{3}; \qquad e=T2^{-5}\sqrt{55}; \qquad f=T2^{-4}\sqrt{15};$$
$$g=T2^{-5}\sqrt{63}; \qquad h=T2^{-2}; \qquad i=T2^{-5}\sqrt{7};$$
$$j=T2^{-4}\sqrt{3}; \qquad k=T2^{-3}; \qquad l=T2^{-5}\sqrt{3};$$
$$m=T2^{-4}; \qquad n=T2^{-5}.$$

Taking into account the equalities (9), the values taken on by these four functions in the ends of the intervals (10) are given in the table II. The values of the entropies and of interdependences between different subsets of curves are given in the column R of the table III. Now, for classifying our family of curves we adopt the following natural strategy underlined by Watanabe [7] (*see* also

TABLE II

	0	1	2	3	4	5	6	7	8	9	10
e_1	0	a	b	c	d	e	f	g	h	g	f
e_2	0	i	j	a	k	a	j	i	0	$-i$	$-j$
e_3	0	l	m	l	0	$-l$	$-m$	$-l$	0	l	m
e_4	0	n	0	$-n$	0	n	0	$-n$	0	n	0
	11	12	13	14	15	16	17	18	19	20	21
e_1	e	d	c	b	a	0	$-a$	$-b$	$-c$	$-d$	$-e$
e_2	$-a$	$-k$	$-a$	$-j$	$-i$	0	i	j	a	k	a
e_3	l	0	$-l$	$-m$	$-l$	0	l	m	l	0	$-l$
e_4	$-n$	0	n	0	$-n$	0	n	0	$-n$	0	n
	22	23	24	25	26	27	28	29	30	31	32
e_1	$-f$	$-g$	$-h$	$-g$	$-f$	$-e$	$-d$	$-c$	$-b$	$-a$	0
e_2	j	i	0	$-i$	$-j$	$-a$	$-k$	$-a$	$-j$	$-i$	0
e_3	$-m$	$-l$	0	l	m	l	0	$-l$	$-m$	$-l$	0
e_4	0	$-n$	0	n	0	$-n$	0	n	0	$-n$	0

Guiaşu [1]): At each step of the decomposition of a set of curves into disjoint subsets we pick up the decomposition characterized by the smallest amount of interconnexion between selected subsets. In this way the connexion between the curves belonging to the same selected subset is the largest one. In our case, the classification is given in the figure 3. The total connexion between all curves of the family is given by the amount

$$W(\{e_1\}, \{e_2\}, \{e_3\}, \{e_4\}) = W(\{e_1, e_2, e_3, e_4\}; \{e_1\}, \{e_2\}, \{e_3\}, \{e_4\})$$
$$= \sum_{i=1}^{4} H(\{e_i\}) - H(\{e_1, e_2, e_3, e_4\}) = 6.0345.$$

The three branching points of the decomposition given in figure 3 are characterized by the smallest amounts of interconnexion between selected subsets. Of course, in this case only the first branching point is "economic", being characterized by a small quantity of interaction between selected subsets. Thus, a natural classification has to stop at this step of the decomposition. Of course, this classification depends on the given partition of the interval $[0, T]$ on

TABLE III

	R	L		R	L
$H(\{e_1\})$	4.0637	0.8343	$W(\{e_2\}, \{e_3, e_4\})$	1.9825	0.3804
$H(\{e_2\})$	3.1162	1.6457	$W(\{e_3\}, \{e_1, e_2\})$	2.2395	0.7622
$H(\{e_3\})$	2.2395	1.3216	$W(\{e_3\}, \{e_1, e_4\})$	2.0095	0.8628
$H(\{e_4\})$	1.4838	0.8343	$W(\{e_3\}, \{e_2, e_4\})$	2.2395	0.7045
$H(\{e_1, e_2\})$	4.8687	2.0275	$W(\{e_4\}, \{e_1, e_2\})$	1.4838	0.5102
$H(\{e_1, e_3\})$	4.7537	1.7804	$W(\{e_4\}, \{e_1, e_3\})$	1.4838	0.8343
$H(\{e_1, e_4\})$	4.5237	1.3216	$W(\{e_4\}, \{e_2, e_3\})$	1.4838	0.4525
$H(\{e_2, e_3\})$	3.8572	2.2051	$W(\{e_1\}, \{e_2, e_3, e_4\})$	3.0522	0.4525
$H(\{e_2, e_4\})$	3.8572	1.9698	$W(\{e_2\}, \{e_1, e_3, e_4\})$	3.0012	0.8392
$H(\{e_3, e_4\})$	2.7235	1.3216	$W(\{e_3\}, \{e_1, e_2, e_4\})$	2.2395	1.0863
$H(\{e_1, e_2, e_3\})$	4.8687	2.5869	$W(\{e_4\}, \{e_1, e_2, e_3\})$	1.4838	0.8343
$H(\{e_1, e_2, e_4\})$	4.8687	2.3516	$W(\{e_1, e_2\}, \{e_3, e_4\})$	2.7235	0.7622
$H(\{e_1, e_3, e_4\})$	4.7537	1.7804	$W(\{e_1, e_3\}, \{e_2, e_4\})$	3.7322	1.1637
$H(\{e_2, e_3, e_4\})$	3.8572	2.2051	$W(\{e_1, e_4\}, \{e_2, e_3\})$	3.5122	0.9398
$H(\{e_1, e_2, e_3, e_4\})$	4.8687	2.5869	$W(\{e_1\}, \{e_2\}, \{e_3\})$	4.5507	1.2147
$W(\{e_1\}, \{e_2\})$	2.3112	0.4525	$W(\{e_1\}, \{e_2\}, \{e_4\})$	3.7950	0.9627
$W(\{e_1\}, \{e_3\})$	1.5495	0.3755	$W(\{e_1\}, \{e_3\}, \{e_4\})$	3.0333	1.2098
$W(\{e_1\}, \{e_4\})$	1.0238	0.3470	$W(\{e_2\}, \{e_3\}, \{e_4\})$	2.9823	1.5965
$W(\{e_2\}, \{e_3\})$	1.4985	0.7622	$W(\{e_1\}, \{e_2\}, \{e_3, e_4\})$	5.0347	1.2147
$W(\{e_2\}, \{e_4\})$	0.7428	0.5102	$W(\{e_1\}, \{e_3\}, \{e_2, e_4\})$	5.2917	1.6735
$W(\{e_3\}, \{e_4\})$	0.9998	0.8343	$W(\{e_1\}, \{e_4\}, \{e_2, e_3\})$	4.5360	1.2868
$W(\{e_1\}, \{e_2, e_3\})$	3.0522	0.4525	$W(\{e_2\}, \{e_3\}, \{e_1, e_4\})$	5.0107	1.7020
$W(\{e_1\}, \{e_2, e_4\})$	3.0522	0.4525	$W(\{e_2\}, \{e_4\}, \{e_1, e_3\})$	4.4850	1.6735
$W(\{e_1\}, \{e_3, e_4\})$	2.0335	0.3755	$W(\{e_3\}, \{e_4\}, \{e_1, e_2\})$	3.7233	1.5965
$W(\{e_2\}, \{e_1, e_3\})$	3.0012	1.0745	$W(\{e_1\}, \{e_2\}, \{e_3\}, \{e_4\})$	6.0345	2.0490
$W(\{e_2\}, \{e_1, e_4\})$	2.7712	0.6157			

the axis Ox. This is the reason for which we call it the classification according to Riemann interaction between curves.

(c) *Classification according to the horizontal connexion.*

Let us consider a partition of the interval $[-T/4, T/4]$ on the axis Oy, containing the disjoint intervals

$$\left[\frac{(k-1)T}{32}, \frac{kT}{32}\right), \quad (k=-7, -6, \ldots, 0, 1, \ldots, 6, 7), \quad \left[\frac{7T}{32}, \frac{8T}{32}\right].$$

Through the ends of these intervals we draw the 17 parallels to the axis Ox. The number of intersection points between these parallels and the curves of our family are introduced in the table IV. The values of corresponding entropies and entropic interdependences are given in the column L of the table III. The classification in this case is given in the figure 4. Of course, this classification

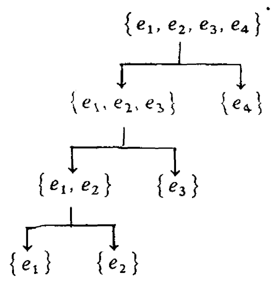

Figure 3

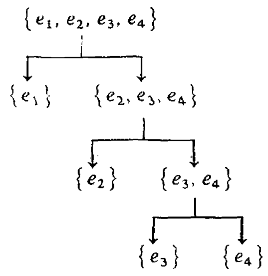

Figure 4

TABLE IV

	$\frac{T}{4}$	$\frac{7T}{32}$	$\frac{3T}{16}$	$\frac{5T}{32}$	$\frac{T}{8}$	$\frac{3T}{32}$	$\frac{T}{16}$	$\frac{T}{32}$	0
e_1	1	2	2	2	2	2	2	2	3
e_2	0	0	0	0	2	4	4	4	5
e_3	0	0	0	0	0	0	4	8	9
e_4	0	0	0	0	0	0	0	8	17
	$\frac{-T}{32}$	$\frac{-T}{16}$	$\frac{-3T}{32}$	$\frac{-T}{8}$	$\frac{-5T}{32}$	$\frac{-3T}{32}$	$\frac{-7T}{32}$	$\frac{-T}{4}$	
e_1	2	2	2	2	2	2	2	1	
e_2	4	4	4	2	0	0	0	0	
e_3	8	4	0	0	0	0	0	0	
e_4	8	0	0	0	0	0	0	0	

depends on the partition of the interval $[-T/4, T/4]$ on the axis Oy. It is a classification according to Lebesgue interactions between the curves.

4. ENTROPIC DISTANCE AND HAMMING DISTANCE

Let E_n be the linear space of all vectors having n components belonging to Galois Field GF (2). For any vector $s \in E_n$ let $w(s)$ be its weight, i. e. the number of non-zero components of s. For any pair of vectors $s, s^* \in E_n$ we denote by $w(s^*, s)$ the number of positions where we have the component 1 in the vector s^* and the component 0 in the vector s. Let also $d(s, s^*)$ be Hamming distance between the vectors s and s^*. Then

$$d(s, s^*) = w(s \mid s^*) + w(s^* \mid s). \tag{11}$$

To each vector $s \in E_n$ we attach the entropy

$$H(\{s\}) = H_2\left(\frac{w(s)}{n}, \frac{n-w(s)}{n}\right)$$
$$= -\frac{w(s)}{n}\log_2\frac{w(s)}{n} - \frac{n-w(s)}{n}\log_2\frac{n-w(s)}{n}. \quad (12)$$

For any subset

$$E^* = \{s_1, s_2, \ldots, s_r\} \subset E_n$$

we define the inner connexion, or the total connexion, between the vectors belonging to E^*, as being

$$W^{\text{tot}}(E^*) = W(E^*; \{s_1\}, \ldots, \{s_r\}) = \sum_{i=1}^{r} H(\{s_i\}) - H(E^*), \quad (13)$$

where $H(E^*)$ is the entropy of the random distribution given by the relative frequencies of the columns in the matrix whose rows are the vectors $s_1, \ldots, s_r$.

The interdependence between two vectors $s, s^* \in E_n$ is given by

$$W(\{s\}, \{s^*\}) = H(\{s\}) + H(\{s^*\}) - H(\{s, s^*\}) \quad (14)$$

where

$$H(\{s, s^*\})$$
$$= H_4\left(\frac{w(s, s^*)}{n}, \frac{w(s \mid s^*)}{n}, \frac{w(s^* \mid s)}{n}, \frac{n - w(s, s^*) - w(s \mid s^*) - w(s^* \mid s)}{n}\right), \quad (15)$$

where $w(s, s^*)$ is the number of components where both s and s^* have the letter 1. The entropic distance between s and s^* is

$$\rho(s, s^*) = H(\{s, s^*\}) - W(\{s\}, \{s^*\})$$
$$= 2H(\{s, s^*\}) - H(\{s\}) - H(\{s^*\}) \quad (16)$$

while the Hamming distance is given by

$$d(s, s^*) = w(s - s^*). \quad (17)$$

PROPOSITION 3: *For any pair $s, s^* \in E_n$ we have*

$$\rho(s, s^*) = 2\lambda\left(\frac{d(s, s^*)}{n}\right) + 2\frac{d(s, s^*)}{n}\lambda\left(\frac{w(s \mid s^*)}{d(s, s^*)}\right)$$
$$+ 2\left(1 - \frac{d(s, s^*)}{n}\right)\lambda\left(\frac{w(s, s^*)}{n - d(s, s^*)}\right) - \lambda\left(\frac{w(s)}{n}\right) - \lambda\left(\frac{w(s^*)}{n}\right), \quad (18)$$

where

$$\lambda(t)=H_2(t,\,1-t)=-t\log_2 t-(1-t)\log_2(1-t),\qquad t\in[0,\,1].$$

Proof: For any random distribution

$$p_i>0,\quad \sum_{i=1}^{n} p_i=1,$$

the corresponding entropy is equal to

$$H_n(p_1,\,\ldots,\,p_n)=-\sum_{i=1}^{n} p_i\log_2 p_i$$

According to the well-known properties of discrete entropy we obtain

$$\begin{aligned}
&H(\{s,\,s^*\})\\
&\quad=H_4\left(\frac{w(s,\,s^*)}{n},\,\frac{w(s\,|\,s^*)}{n},\,\frac{w(s^*\,|\,s)}{n},\,1-\frac{w(s,\,s^*)+w(s\,|\,s^*)+w(s^*\,|\,s)}{n}\right)\\
&\quad=H_3\left(\frac{w(s,\,s^*)}{n},\,1-\frac{w(s,\,s^*)+d(s,\,s^*)}{n},\,\frac{d(s,\,s^*)}{n}\right)\\
&\qquad+\frac{d(s,\,s^*)}{n}H_2\left(\frac{w(s\,|\,s^*)}{d(s,\,s^*)},\,\frac{w(s^*\,|\,s)}{d(s,\,s^*)}\right)\\
&\quad=H_2\left(\frac{d(s,\,s^*)}{n},\,1-\frac{d(s,\,s^*)}{n}\right)+\frac{d(s,\,s^*)}{n}H_2\left(\frac{w(s\,|\,s^*)}{d(s,\,s^*)},\,\frac{w(s^*\,|\,s)}{d(s,\,s^*)}\right)\\
&\qquad+\left(1-\frac{d(s,\,s^*)}{n}\right)H_2\left(\frac{w(s,\,s^*)}{n-d(s,\,s^*)},\,1-\frac{w(s,\,s^*)}{n-d(s,\,s^*)}\right).
\end{aligned}$$

According to (16) we get

$$\begin{aligned}
\rho(s,\,s^*)=&\,2H_2\left(\frac{d(s,\,s^*)}{n},\,1-\frac{d(s,\,s^*)}{n}\right)-2\frac{d(s,\,s^*)}{n}H_2\left(\frac{w(s\,|\,s^*)}{d(s,\,s^*)},\,\frac{w(s^*\,|\,s)}{d(s,\,s^*)}\right)\\
&+2\left(1-\frac{d(s,\,s^*)}{n}\right)H_2\left(\frac{w(s,\,s^*)}{n-d(s,\,s^*)},\,1-\frac{w(s,\,s^*)}{n-d(s,\,s^*)}\right)\\
&-H_2\left(\frac{w(s)}{n},\,1-\frac{w(s)}{n}\right)-H_2\left(\frac{w(s^*)}{n},\,1-\frac{w(s^*)}{n}\right),
\end{aligned}$$

i. e. just (18).

Q.E.D.

REMARK: Obviously, if

$$d(s, s^*)=0,$$

then

$$H(\{s\})=H(\{s^*\})=H(\{s, s^*\})=W(\{s\}, \{s^*\}),$$

which implies

$$\rho(s, s^*)=0.$$

The converse assertion is not true.

PROPOSITION 4: *If for $s, s^* \in E_n$ we have $d(s, s^*)=n$ then $\rho(s, s^*)=0$.*

Proof: If

$$d(s, s^*)=n,$$

we have

$$w(s)=n-w(s^*),$$

and according to (18) we can write

$$\rho(s, s^*)=2H_2\left(\frac{w(s\,|\,s^*)}{n}, \frac{w(s^*\,|\,s)}{n}\right)-2H_2\left(\frac{w(s^*)}{n}, \frac{n-w(s^*)}{n}\right)=0$$

because

$$w(s\,|\,s^*)=w(s)=n-w(s^*),$$
$$w(s^*\,|\,s)=w(s^*).$$

Q.E.D.

It is well-known the importance of Hamming distance in algebraic coding theory. The entropic distance can also be used with this respect. How the independent errors modify the entropic distance can be seen in the following proposition:

PROPOSITION 5: *If the vector $s \in E_n$ is transmitted and t independent errors occur such that u components 1 are changed into 0 and v components 0 are changed into 1 $(u+v=t)$, we obtain at the receiver the vector $s^* \in E_n$ for which*

$$\rho(s, s^*)=\lambda\left(\frac{w(s)}{n}\right)+2\left(1-\frac{w(s)}{n}\right)\lambda\left(\frac{v}{n-w(s)}\right)+2\frac{w(s)}{n}\lambda\left(\frac{u}{w(s)}\right)-\lambda\left(\frac{w(s)-u+v}{n}\right). \quad (19)$$

Proof: If the vector s is transmitted and u components 1 are changed into 0 and v components 0 are changed into 1, according to (16) we get

$$\begin{aligned}\rho(s, s^*)&=2H(\{s, s^*\})-H(\{s\})-H(\{s^*\})\\ &=-H_2\left(\frac{w(s)}{n}, \frac{n-w(s)}{n}\right)-H_2\left(\frac{w(s)-u+v}{n}, \frac{n-w(s)+u-v}{n}\right)\\ &\quad+2H_4\left(\frac{w(s)-u}{n}, \frac{u}{n}, \frac{v}{n}, \frac{n-w(s)-v}{n}\right). \quad (20)\end{aligned}$$

But

$$H_4\left(\frac{w(s)-u}{n}, \frac{u}{n}, \frac{v}{n}, \frac{n-w(s)-v}{n}\right)=H_3\left(\frac{w(s)}{n}, \frac{v}{n}, \frac{n-w(s)-v}{n}\right)$$

$$+\frac{w(s)}{n}H_2\left(\frac{w(s)-u}{w(s)}, \frac{u}{w(s)}\right)=H_2\left(\frac{w(s)}{n}, \frac{n-w(s)}{n}\right)$$

$$+\frac{n-w(s)}{n}H_2\left(\frac{v}{n-w(s)}, \frac{n-w(s)-v}{n-w(s)}\right)+\frac{w(s)}{n}H_2\left(\frac{w(s)-u}{w(s)}, \frac{u}{w(s)}\right). \quad (21)$$

From (20) and (21) we obtain (19).

Q.E.D.

The advantage of using the entropic distance instead of Hamming distance can be seen in such common situations when some letters of the transmitted message are deleted or when some letters are inserted into the message during the transmission. With respect to these errors Hamming distance is very vulnerable while the entropic distance works quite well. For example let us take the binary vector $s=01010101$ and let us suppose that an initial symbol 1 is inserted and all the other symbols are shifted, receiving the vector $s^*=10101010$ of the same length 8. The effect of this usual kind of error on Hamming distance is very severe, namely $d(s, s^*)=8$ while $\rho(s, s^*)=0$. If the initial symbol 0 is inserted into the message s and all the other symbols are shifted, receiving the vector $s^*=00101010$ of length 8, we get $d(s, s^*)=7$ (i. e. 87.5% from the maximum Hamming distance) while $\rho(s, s^*)=0.8568$ (i. e. 42.84% from the maximum entropic distance). The entropic distance gives good results when we want to recognize the structure, or the pattern, of the word. Thus, from the viewpoint of the entropic distance there is no difference between the vector $s=01010101$ and the vectors $s_1^*=10101010$ and $s_2^*=02020202$ because $\rho(s, s_1^*)=\rho(s, s_2^*)=0$ because s, s_1^* and s_2^* have the same pattern (i. e. the alternation of two symbols) while $d(s, s_1^*)=8$ and $d(s, s_2^*)=4$.

On the other hand, the entropic distance is more flexible than Hamming distance. Thus let us consider the table IV. We can see that Hamming distance makes no difference between the pairs of vectors $(\{e_1\}, \{e_3\})$ and $(\{e_1\}, \{e_4\})$ because

$$d(\{e_1\}, \{e_3\})=d(\{e_1\}, \{e_4\})=17$$

while

$$\rho(\{e_1\}, \{e_3\})=1.4049, \qquad \rho(\{e_1\}, \{e_4\})=0.9746$$

Also, we get

$$d(\{e_1, e_3\}, \{e_2, e_4\})=d(\{e_1, e_4\}, \{e_2, e_3\})=15,$$

while

$$\rho(\{e_1, e_3\}, \{e_2, e_4\})=1.4236, \qquad \rho(\{e_1, e_4\}, \{e_2, e_3\})=1.6471.$$