Note di Matematica Vol.III, 15-27 (1983)

A CHARACTERIZATION OF A NEW SOURCE ENTROPY

B. FORTE (*) and C.C.A. SASTRI (**)

Abstract. *A new source entropy was introduced recently*[2] *and was shown, for discrete memoryless channels, to be an improvement over the classical one. It is shown here that this new entropy can be characterized by certain simple and natural properties.*

Classically, in communication theory, source entropy (for any channel) is defined as follows: if H_n denotes the uncertainty associated with a message of length n, theii the source entropy is $H_\infty \equiv \lim_{n\to\infty} \frac{H_n}{n}$. A new source entropy was introduced recently in [2]: Let p_n (n=0,1,2,...) be the probability that a message is of length n and li the uncertainty associated iiith a message given

---

(*) Department of Applied Mathematice, Univercity of Waterloo, Waterloo, Ontario, Canada, M2L 3G1, recearch cupported by NSERC grant # A7677.

(**) Department of Mathematics, Statistics and Computing Science, **Dalhoucie** Univercity, Halifax, Nova Scotia, Canada, B3H 4H8, recearch cupported by NSERC grant # A4825.

that it ic of length n. Then the cource entropy is

$$I \equiv \frac{\sum_n H_n p_n}{\sum_n n p_n}$$

We characterize thic entropy, in effect, by characterizing the numerator, namely $\sum_n H_n p_n$. (Alternatively, one can regard $\sum_n H_n p_n$ asthe cource entropy and I ac the source entropy per unit length of the mescage).

In practice all mecsagec, however long, are finite. The limit of $\frac{H_n}{n}$ ac $n \to \infty$ hac been used because, a priori, the length of a mes-cage ic unknown but could be large. For practical purpoces, H_n is often replaced by $\frac{H_n}{n}$ for some maximal n. In that case, I gives the came result if $p_n = 1$ for the maximal n and 0 otherwise.

The principal advantage in adopting I for the source entropy inctead of H_∞ is the following[2]. Using the notation of [1], consider a message of length n consicting of lettcrs from an alphabet of size M. Define $\langle P_e(n) \rangle$, the average probability of error per unit length in cuch a mecsage by

$$\langle P_e(n) \rangle = \frac{1}{n} P_e(n) \equiv \frac{1}{n} \sum_{i=1}^{n} P_e(i,n) \quad ,$$

where $P_e(i,n)$ denotec the probability of an error in the i^{th} position. Then define the average probability $\langle P \rangle$ of error per unit length (in an arbitrary meccage) by $<P_e> = \lim_{n\to\infty} <P_e(n)>$. An example wac given in [2] to chow that this limit doec not always exist. If, however, the average probability of error per unit length is defined by $\frac{E(P_e(n))}{E(n)}$, where E denotec expectation, then it exictc whenever $E(n) < \infty$. (The statisticc collected from past experience can be used to ectimate the probabilities used in computing these expectationc.) For a discrete memorylece channel, replacement of $<P_e>$ bv $\frac{E(P_e(n))}{E(n)}$ and of H_∞ by I yields a better lower bound in the classical Fano inequality[1]

$$<P_e> \log (M-1) + H(P_e) \geq H_\infty - \frac{\tau_s}{\tau_c} C,$$

where C is the capacity of the channel, H the two-event shannon entropy, τ_s the time interval between sourcc letters, and τ_c the intervai between channel lettere. For details, the render is referred to [2].

The aim of thic notc ic to show that the new definition ic notad hoc and in fact that it has certain "natural" properties which, conversely, characterize this new entropy. This is achieved by appealing twicc to a theorem, proved in [3], charactcrizing the entropy associated with a random vector. A statement and explanation of thc theorem follows.

Let $X = (X_1, X_2, \ldots, X_n)$ be a random vcctor with rcal components.

Suppose that $X_1, X_2, \ldots, X_n$ are discrete random variables s.t. the range of X_i is $(a_{i1}, a_{i2}, \ldots, a_{im_i})$. Denote the collection of ranges by $X^{(n)} = ((a_{i}, a_{i}, \ldots, a_{i} \quad ,2,\ldots,n)$. Let the joint probability dictribution be $\Pi_{X^{(n)}} \in \Gamma_{m_1 m_2 \ldots m_n}$, where

$$\Gamma_{m_1 m_2 \ldots m_n} \equiv \{\pi_{j_1 j_2 \ldots j_n}, \; j_i = 1,2,\ldots,m_i \;, \quad i=1,2,\ldots,n \;|$$

$$\pi_{j_1 j_2 \ldots j_n} \geq 0; \; \sum_{j_1 j_2 \ldots j_n} \pi_{j_1 j_2 \ldots j_n} = 1\} \;.$$

Let $I^{(n)}_{m_1 m_2 \ldots m_n}(X^{(n)}; \Pi_{X^{(n)}})$ denote the entropy associated with X, i.e., the uncertainty about which values its components take. Observe that this uncertainty can be regarded as the uncertainty associated with a message of length n, the ith character of which can be any one of an alphabet of m_i characters. To establish the correspondence, we need anly set $a_{i,j} = j_i$ iff the ith character of the message is the j_ith character of the ith alphabet. Observe also that thic entropy ic quite general in the sense that it could depend not only on the probability distribution $\Pi_{X^{(n)}}$ but alco on the range $X^{(n)}$ (thc actual content for messages) and the sizes of the ranges of the componentc of X (the lengts of the alphabets for messages). Consider now the following propertiec for the entropy:

1. SUB-ADDITIVITY

$$I^{(n)}_{m_1 m_2 \ldots m_n}(X^{(n)};\Pi_{X^{(n)}}) \leq I^{(k)}_{m_1 m_2 \ldots m_k}(X^{(k)};\Pi_{X^{(k)}}) +$$

$$+ I^{(n-k)}_{m_{k+1} \ldots m_n}(X^{(n-k)};\Pi_{X^{(n-k)}})$$

where $\Pi_{X^{(k)}}$ and $\Pi_{X^{(n-k)}}$ are marginal distributions of $\Pi_{X^{(n)}}$ defined in the usual manner,

$$X^{(k)} = \{(a_{i1}, a_{i2}, \ldots, a_{im_i}) \quad , \quad i = 1,2,\ldots,k\}$$

$$X^{(n-k)} = \{(a_{i1}, a_{i2}, \ldots, a_{im_i}) \quad , \quad i = k+1,\ldots,n\}$$

This property means that if a message of length n is divided into two blocks, one of length k and the other of length n-k, then the uncertainty about the full message cannot be bigger than the sum of the uncertainties about the two blocks. If, however, the two blocks are independent, the uncertainties should add up, i.e. the entropy should have the following property.

2. ADDITIVITY

$$I^{(n)}_{m_1 m_2 \ldots m_n}(X^{(n)};\Pi_{X^{(n)}}) = I^{(k)}_{m_1 m_2 \ldots m_k}(X^{(k)};\Pi_{X^{(k)}}) +$$

$$(X^{(n-k)};\Pi$$

whenever the distribution $\Pi_{X^{(n)}}$ is the product of its marginals

$\Pi_{X^{(k)}}$ and $\Pi_{X^{(n-k)}}$

3. LOCAL SYMMETRY

For any positive integer s, let $X^{(2s)} = \{(a_{i1},\dots,a_{im_i}),\ i = 1,2,\dots,2s\}$, $X^{(s)} = \{(a_{i1},\dots,a_{im_i}),\ i=1,2,\dots,s$, $X'^{(s)} = \{(a_{i1},a_{i2},\dots,a_{im_i}),\ i=s+1,\dots,2s\}$. Let $a_{i+s,k} = a_{ik}$ $(i=1,2..s,\ k = 1,2,\dots,m_i)$ and assume that $X^{(s)}$ and $X'^{(s)}$ are independent.

Then

$$I^{(2s)}_{m_1m_2\dots m_sm_1m_2\dots m_s}(X^{(2s)};\Pi_{X^{(2s)}}) = I^{(2s)}_{m_1m_2\dots m_sm_1m_2\dots m_s}(X^{(2s)};\Pi^{T(\vec{h},\vec{k})}_{X^{(2s)}}),$$

where $\vec{h} = (h_1,h_2,\dots,h_s)$, $\vec{k} = (k_1,k_2,\dots,k_s)$ $(1 \le h_i\ , k_i \le m_i)$

and $\Pi^{T(\vec{h},\vec{k})}_{X^{(2s)}}$ is obtained from $\Pi_{X^{(2s)}}$ by interchanging the entries

$\pi_{h_1h_2\dots h_sk_1k_2\dots k_s}$ and $\pi_{k_1k_2\dots k_sh_1h_2\dots h_s}$. This property is a technical necessity, but it can be motivated in commutation theory as follows. Suppose that the sender of the message is a spy who has infiltrated enemy territory and that the receiver is his employer. Suppose there is an understanding between them that whenever he sends a specific message it is meant to be another specific message (of the same length). Clearly then the interchange of these message should not affect the uncertainty.

4. BOUNDEDNESS.

For each $q \in [0,1]$, let

$$\Pi'_{X^{(n)}} = \{\pi_{j_1 j_2 \ldots j_n} : \pi_{h_1 h_2 \ldots h_n} = q, \pi_{k_1 k_2 \ldots k_n} = 1-q \text{ and } \pi_{j_1 j_2 \ldots j_n} = 0 \text{ otherwice}\}.$$

For fixed $X^{(n)}$, let $f^{(n)}_{h_1 h_2 \ldots j_n k_1 k_2 \ldots k_n}(q)$ denote the function that takes q to $I^{n}_{m_1 m_2 \ldots m_n}(X^{(n)};\Pi'_{X^{(n)}})$. Intuitively, we expect that $f^{(n)}_{h_1 h_2 \ldots h_n k_1 k_2 \ldots k_n}(0) = f^{(n)}_{h_1 h_2 \ldots j_n k_1 k_2 \ldots k_n}(1) = 0$. We shall require that there exict real numbers $M^{(n)}_{h_1 \ldots h_n k_1 \ldots k_n}$

$$f^{(n)}_{h_1 h_2 \ldots h_n k_1 k_2 \ldots k_n}(q) \geq M^{(n)}_{h_1 h_2 \ldots h_n k_1 k_2 \ldots k_n} \quad \forall q \in [0,1] .$$

Thic regularity property is a technical necessity

THEOREM 3. If and only if $I^{(n)}_{m_1 m_2 \ldots m_n}(X^{(n)};\Pi_{X^{(n)}})$ has Properties 1-4, it has the following form:

$$I^{(n)}_{m_1 m_2 \ldots m_n}(X^{(n)};\Pi_{X^{(n)}}) = -A \sum_{j_1 j_2 \ldots j_n} \pi_{j_1 j_2 \ldots j_n} \log \pi_{j_1 j_2 \ldots j_n}$$

$$+ \sum_{i=1} \sum_{j_i=1} B_{j_i}(a_{i1}, a_{i2}, \ldots, a_{i}$$

$$^{(n)}(X^{(n)} : S$$

where A is a non-negative constant, the B'_is are arbitrary real-va lued functions of their arguments and

$$\text{Supp}\,(\Pi_{X^{(n)}}) = \{(j_1 j_2 \ldots j_n) \mid \pi_{j_1 j_2 \ldots j_n} > 0,$$

$$p^{(i)}_{j_i} = \sum_{j_1} \sum_{j_2} \ldots \sum_{j_{i-1}} \sum_{j_{i+1}} \ldots \sum_{j_n} \pi_{j_1 j_2 \ldots j_n}$$

$G^{(n)}$ being a function of its arguments that satisfies 1-3.

Now consider a source which sends messages of arbitrary but finite length.

Let P_i denote the probability that the length of a message will be i and let H_i denote the uncertainty associated with a message given that it is of length i. Let

$$H^{(1)} = \{H_1, H_2, \ldots\}$$

$$P^{(1)} = \{P_1, P_2, \ldots\}$$

It is natural to assume that the source entropy $I^{(1)}$ for such a source depends on $H^{(1)}$ and $P^{(1)}$ — indeed, it can be regarded as the uncertainty associated with the random variable which takes the values $H_1, H_2, \ldots$ with probabilities $P_1, P_2, \ldots$. Thus the cource entropy $I^{(1)}(P^{(1)}; H^{(1)})$ is an example of the kind of entropy characterized by the above theorem. Moreover, properties 1-4 become meaningful for such an entropy when suitably interpreted. For example, sub-additivity means that if there are two sources sending messages — say, two spies — and one regards them together as a single source, then the source entropy of the composite source should not exceed the sum of the source entropies of the constituent sources, with equality holding when the two sources are independent (ad ditivity). Similarly, local symmetry means invarinncc under the interchnngc of sourccs in specific instancces. Finally, boundednecss on onc side is a weak regularity property — in fact it would be reasonable to assume that the source entropy is non-ncgntivc. All these propcrties could be written down explicitly, but we shall not do so for thc sake of brevity.

It follows thcn fi-om the characterization thcoreni — with obvious changes in notntion — that

$$I^{(1)}(P^{(1)};\ H^{(1)}) = -A\sum_i P_i \log P_i + \sum_i B_i(H_1,H_2,\ldots)P_i$$

$$+ G(H_1,H_2,\ldots;\ \mathrm{Supp}\,(P^{(1)})) .$$

Suppose that in nddition to propcrties 1-4, $I^{(1)}(P^{(1)};\ H^{(1)})$ has the following two propcrties:

5. If $P_j = 0\ \forall\ j \neq i$ and $P_i = 1$, $I^{(1)}(P^{(1)}, H^{(1)}) = H_i$

6. If $P_1 = \frac{1}{2}$, $P_2 = \frac{1}{2}$, $H_1 = H_2$, $P_i = 0\ \forall i \neq 1,2$ then

$$I^{(1)}(P^{(1)},\ H^{(1)}) = H_1 .$$

Property 5 implies that $H_i = B_i(H_1,H_2,\ldots) + G$.

Hence

$$\sum_i B_i P_i + G = \sum_i (B_i + G)P_i = \sum_i H_i P_i .$$

so

$$I^{(1)}(P^{(1)};\ H^{(1)}) = -A\sum_i P_i \log P_i + \sum_i H_i P_i .$$

Property 6 gives

$$H_1 = -A \log 2 + H_1$$

which implies that A=0. Thus we have

$$I^{(1)}(P^{(1)}; H^{(1)}) = \sum_n H_n P_n$$

The conditional entropies H_n can themselves be characterized by properties 1-4 plus a few additional ones. For instance, in order to recover the Shannon entropy, we shall assume the following in addition to properties 1-4:

7. H_n depends only on the distribution $\Pi_{X^{(n)}}$ and not on the content $X^{(n)}$.

8. For $1 \leq h_i, k_i \leq m_i$, $1 \leq i \leq n$,

$$\lim_{q\to 0^+} f^{(n)}_{h_1h_2\ldots h_nk_1k_2\ldots k_n}(q) = 0 = f^{(n)}_{h_1h_2\ldots h_nk_1k_2\ldots k_n}(0)$$

Observe that property 8 is a continuity property traditionally used to eliminate the Hartley entropy.

Now properties 1-4 and 7 imply that

$$H = -A \sum_{j_1j_2\ldots j_n} \pi_{j_1j_2\ldots j_n} \log \pi_{j_1j_2\ldots j_n} + \sum_{i=1}^{n} \sum_{j_i=1}^{m_i} B_{j_i} P^{(i)}_{j_i} + G^{(n)}(\mathrm{Supp}(\Pi_{X^{(n)}})),$$

where the B_i are arbitrary constants and $G^{(n)}$ is a function of the specified argument that satisfies 1-3. In particular we have

$$f_{h_1h_2\ldots h_nk_1k_2\ldots k_n}(q) = -A(q\log q+(1-q)\log(1-q))$$

$$+\sum_{r=1}^{n}(B_{h_r}-B_{k_r})q+\sum_{s=1}^{n}B_{k_s}+G^{(n)}((h_1,h_2,\ldots,h_n),(k_1,k_2,\ldots,k_n))$$

Property 8 then givec

$$0 = \sum_{s=1}^{n}B_{k_s} + G^{(n)}((h_1,h_2,\ldots,h_n),(k_1,k_2,\ldots,k_n)).$$

Since this must be true for every choice of $(h_1,h_2,\ldots,h_n)$ and $(k_1,k_2,\ldots,k_n)$, $G^{(n)}$ must be a constant for each n, and

$$B_{k_s} = \text{conct.} =: B, \quad s = 1,2,\ldots,n \quad , k_s = 1,2,\ldots,m_s$$

Hence $G^{(n)}(\mathrm{Supp}(\Pi_{X^{(n)}}) = -nB$, whenever the cardinality of $\mathrm{Supp}(\Pi_{X^{(n)}})$ is 1 or 2.

Let $\Pi'_{X^{(n)}}$ and $\Pi''_{X^{(n)}}$ be any two probability distributionc on the same $X^{(n)}$

Let

$$C := \mathrm{Sup}(\Pi'_{X^{(n)}})\cap\mathrm{Supp}(\Pi''_{X^{(n)}})$$

$$D' := \mathrm{sup}(\Pi'_{X^{(n)}})/C$$

$$D'' := \mathrm{supp}(\Pi''_{X^{(n)}})/C .$$

By impocing propertiec 2,3 and 1 in that order on $G^{(n)}(\mathrm{Supp}(\Pi_{X^{(n)}})$

with a standard technique[3] we obtain

$$G^{(n)}(D')-G^{(n)}(D'\cup D'') \leq G^{(n)}(D\cup D')-G^{(n)}(C\cup D'') \leq G^{(n)}(D'\cup D'')-G^{(n)}(D'')$$

By choosing $\Pi'_{X^{(n)}}$ and $\Pi''_{X^{(n)}}$ so that **D'** and **D"** have cardinality 1,

$$G^{(n)}(C\cup D') - G^{(n)}(C\cup D'') = 0.$$

Hence $G^{(n)}(\mathrm{Supp}(\Pi_{X^{(n)}}))$ depends only on the cardinaiity of $\mathrm{Supp}(\Pi_{X^{(n)}})$. By choosing $\Pi''_{(n)}$ so that $D'' = \emptyset$ one can recognize that $G^{(n)}(\mathrm{Supp}(\Pi_{X^{(n)}}))$ is a non-decreasing function of the cardinality of $\mathrm{Supp}(\Pi_{X^{(n)}}$. A last recourse to additivity yields

$$G^{(n)}(\mathrm{Supp}(\Pi_{X^n})) = -nB$$

for all possible v'alues of the cardinality of Supp (Π_{X^n}).

Hence

$$H = -A\sum_{j_1j_2\ldots j_n} \pi_{j_1j_2\ldots j_n}\log \pi_{j_1j_2\ldots j_n} + nB + G^{(n)}(\mathrm{Supp}(\Pi_{X^{(n)}}))$$

$$= -A\sum_{j_1j_2\ldots j_n} \pi_{j_1j_2\ldots j_n}\log \pi_{j_1j_2\ldots j_n}$$

Acknowledgement. We wish to thank an anonymous referee for making certain remarks which led to a clarification of the origina1 paper

REFERENCES

[1] R.G.GALLAGER: *"Information Theory and Reliable Communication"*, John Wiley and Sons, New York (1968).

[2] B.FORTE and A.CIAMPI: *"Source Entropy for Messages of Random Length"*, to appear in Rendiconti di Matematica in 1982.

[3] B.FORTE and M.LO SCHIAVO: *"Non-Expansible, Additive and Subadditive Entropies for a Random Vector"*, submitted to Utilitas Mathematica.

Lavoro pervenuto alla Redazione il 30 Luglio 1982
ed accettato per la pubblicazione il 18 Maggio 1983
su parere favorevole di G. Andreassi e P. Benvenuti