

ON THE AMOUNT OF INFORMATION

HU KUO TING

(Translated by D. Lieberman)

Introduction

1°. In the present article we shall consider random variables $X_i, i = 1, 2, \dots$, which take on a finite number of values. We let $P(X_{i_1}, X_{i_2}, \dots, X_{i_n})$ denote the joint probability for values of the variables $X_{i_1}, X_{i_2}, \dots, X_{i_n}$, and $P(X_{i_1}, X_{i_2}, \dots, X_{i_n} | X_{j_1}, X_{j_2}, \dots, X_{j_m})$ the conditional probability for values of the variables $X_{i_1}, X_{i_2}, \dots, X_{i_n}$ if the values of the variables $X_{j_1}, X_{j_2}, \dots, X_{j_m}$ are given.

It is well known that the amount of information¹ of the variables X_1, X_2 can be written as

$$\begin{aligned} H(X_1) &= - \sum_{x_1} p(x_1) \log p(x_1), \quad H(X_2) = - \sum_{x_2} p(x_2) \log p(x_2), \\ H(X_1, X_2) &= - \sum_{x_1, x_2} p(x_1, x_2) \log p(x_1, x_2), \\ I(X_1, X_2) &= \sum_{x_1, x_2} p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1) \cdot p(x_2)}, \\ H_{X_2}(X_1) &= \sum_{x_2} p(x_2) \left[- \sum_{x_1} p(x_1 | x_2) \log p(x_1 | x_2) \right]. \end{aligned}$$

Moreover, the following conditions are satisfied (see [1] and [2]):

$$\begin{aligned} (1) \quad & I(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2), \\ (2) \quad & I(X_1, X_2) = H(X_1) - H_{X_2}(X_1), \\ (3) \quad & H_{X_2}(X_1) \leq H(X_1). \end{aligned}$$

Similarly, for three variables X_1, X_2, X_3 , the amount of information has the form

$$\begin{aligned} H(X_1, X_2, X_3) &= - \sum_{x_1, x_2, x_3} p(x_1, x_2, x_3) \log p(x_1, x_2, x_3), \\ I(X_1, X_2, X_3) &= \sum_{x_1, x_2, x_3} p(x_1, x_2, x_3) \log \frac{p(x_1, x_2, x_3)}{p(x_1, x_2) \cdot p(x_3)}, \\ H_{X_3}(X_1, X_2) &= \sum_{x_3} p(x_3) \left[- \sum_{x_1, x_2} p(x_1, x_2 | x_3) \log p(x_1, x_2 | x_3) \right], \\ H_{(X_2, X_3)}(X_1) &= \sum_{x_2, x_3} p(x_2, x_3) \left[- \sum_{x_1} p(x_1 | x_2, x_3) \log p(x_1 | x_2, x_3) \right], \\ I_{X_2}(X_1, X_2) &= \sum_{x_2} p(x_2) \left[\sum_{x_1, x_3} p(x_1, x_3 | x_2) \log \frac{p(x_1, x_3 | x_2)}{p(x_1, x_2) p(x_3, x_2)} \right], \end{aligned}$$

and the conditions

$$\begin{aligned} (4) \quad & H_{(X_2, X_3)}(X_1) \leq H_{X_2}(X_1), \\ (5) \quad & I(X_1, X_2, X_3) = I(X_2, X_3) + I_{X_2}(X_1, X_3), \\ (6) \quad & I(X_1, X_2, X_3) + I(X_1, X_2) = I(X_1, (X_2, X_3)) + I(X_2, X_3) \end{aligned}$$

are satisfied (see [1], [2] and [3]).

2°. It is interesting to note that if sets A_1, A_2, A_3 and an additive function φ on them are put into correspondence with the variables X_1, X_2, X_3 and their distribution P in the following manner ($i, j = 1, 2, 3$):

¹ We use the term amount of information to denote any information functional; for example, $H(X_1), H(X_1, X_2), I(X_1, X_2), H_{X_2}(X_1), H((X_1, X_2), X_3), \dots$ etc.

$$\begin{aligned}
H(X_i) &\rightarrow \varphi(A_i), \\
H(X_i, X_j) &\rightarrow \varphi(A_i \cup A_j), \\
I(X_i, X_j) &\rightarrow \varphi(A_i \cap A_j), \\
H_{X_j}(X_i) &\rightarrow \varphi(A_i - A_j), \\
H((X_1, X_2), X_3) &\rightarrow \varphi((A_1 \cup A_2) \cup A_3), \\
I((X_1, X_2), X_3) &\rightarrow \varphi((A_1 \cup A_2) \cap A_3), \\
H_{X_3}(X_1, X_2) &\rightarrow \varphi((A_1 \cup A_2) - A_3), \\
H_{(X_2, X_3)}(X_1) &\rightarrow \varphi(A_1 - (A_2 \cup A_3)), \\
I_{X_2}(X_1, X_3) &\rightarrow \varphi((A_1 \cap A_3) - A_2),
\end{aligned}$$

then the following relations between values of the additive function φ will correspond to equations (1)–(6):

$$\begin{aligned}
(1') \quad & \varphi(A_1 \cap A_2) = \varphi(A_1) + \varphi(A_2) - \varphi(A_1 \cup A_2), \\
(2') \quad & \varphi(A_1 \cap A_2) = \varphi(A_1) - \varphi(A_1 - A_2), \\
(3') \quad & \varphi(A_1 - A_2) \leq \varphi(A_1), \\
(4') \quad & \varphi(A_1 - (A_2 \cup A_3)) \leq \varphi(A_1 - A_2), \\
(5') \quad & \varphi((A_1 \cup A_2) \cap A_3) = \varphi(A_2 \cap A_3) + \varphi((A_1 \cap A_3) - A_2), \\
(6') \quad & \varphi((A_1 \cup A_2) \cap A_3) + \varphi(A_1 \cap A_2) = \varphi(A_1 \cap (A_2 \cup A_3)) + \varphi(A_2 \cap A_3).
\end{aligned}$$

In the present work we shall establish, in a mathematically rigorous form, such dual relations generated by the amount of information and an additive set function.

I would like to thank Yu. V. Prokhorov warmly for his interest in the present work and useful advice.

Fundamental Theorems

3°. Let us take a sequence of sets A_i , $i = 1, 2, \dots$, and consider a function φ on the ring \mathfrak{U} generated by the sequence A_i , $i = 1, 2, \dots$, such that

- 1) for any $E \in \mathfrak{U}$, $-\infty < \varphi(E) < +\infty$,
- 2) for any nonintersecting sets E_1, E_2, \dots, E_n of \mathfrak{U} , the following relation holds ²:

$$\varphi(E_1) + \varphi(E_2) + \dots + \varphi(E_n) = \varphi(E_1 + E_2 + \dots + E_n).$$

Let us consider the following three collections of values of the function φ :

$$\begin{aligned}
(7) \quad & \varphi(A_1) < \infty, \\
& \varphi(A_1 \cap A_2) < \infty, \quad \varphi(A_1 \cap \bar{A}_2) < \infty, \quad \varphi(\bar{A}_1 \cap \bar{A}_2) < \infty, \\
& \varphi(A_1 \cap A_2 \cap A_3) < \infty, \quad \varphi(A_1 \cap A_2 \cap \bar{A}_3) < \infty, \quad \dots, \quad \varphi(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) < \infty, \\
& \dots \dots \dots
\end{aligned}$$

(where $A_1 \cap \bar{A}_2 = A_1 - A_2$, $\bar{A}_1 \cap A_2 = A_2 - A_1$, $A_1 \cap A_2 \cap \bar{A}_3 = A_1 \cap A_2 - A_3$, \dots).

$$\begin{aligned}
(8) \quad & \varphi(A_i) < \infty, & i = 1, 2, \dots, \\
& \varphi(A_i \cap A_j) < \infty, & i < j = 1, 2, \dots, \\
& \varphi(A_i \cap A_j \cap A_k) < \infty, & i < j < k = 1, 2, \dots, \\
& \dots \dots \dots
\end{aligned}$$

and finally

² Here and in what follows the summation sign "+", with regard to sets, denotes unions of nonintersecting sets.

(9)

It is easily seen respectively): the several of the values

4°. Let us introduce. We start with the

$$H\left(\bigcup_{v=1}^n X_{i_v}\right)$$

Then we define

$$H\left(\bigcap_{v=1}^n X_{i_v}\right)$$

Similarly, we define set function. For

$$H(X_i \sim ($$

Thus, we have $X_{i_1}, X_{i_2}, \dots, X_{i_n}$ operations \cup, \cap . Many of the

(10)

are always non-negative take on negative $H(\bigcap_{i=1}^n X_i)$, $n >$

random variable

$$X_1(w) =$$

Then

Amounts of information, and others

5°. Theorem

³ We use the fuse $X_i \sim X_j$ with

$$(9) \quad \begin{aligned} \varphi(A_i) &< \infty, & i = 1, 2, \dots, \\ \varphi(A_i \cup A_j) &< \infty, & i < j = 1, 2, \dots, \\ \varphi(A_i \cup A_j \cup A_k) &< \infty, & i < j < k = 1, 2, \dots, \\ &\dots\dots\dots \end{aligned}$$

It is easily seen that the function φ on \mathfrak{A} is defined uniquely by its value (7) (or (8) or (9), respectively): the value of φ on any $E \in \mathfrak{A}$ can be represented in the form of an algebraic sum of several of the values from the collection (7) (or (8) or (9), respectively).

4°. Let us introduce the following amounts of information of the variables $X_i, i = 1, 2, \dots$. We start with the notation

$$H(X) = H\left(\bigcup_{v=1}^n X_{i_v}\right) = H(X_{i_1}, X_{i_2}, \dots, X_{i_n}) \\ = - \sum_{x_{i_1}, x_{i_2}, \dots, x_{i_n}} p(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \log p(x_{i_1}, x_{i_2}, \dots, x_{i_n}), \quad n = 1, 2, \dots$$

Then we define

$$pb \quad H\left(\bigcap_{v=1}^n X_{i_v}\right) = \sum_{v \leq n} H(X_{i_v}) - \sum_{v_1 < v_2 \leq n} H(X_{i_{v_1}} \cup X_{i_{v_2}}) + \dots + (-1)^{n+1} H\left(\bigcup_{v=1}^n X_{i_v}\right).$$

Similarly, we define other amounts of information in accordance with identities for an additive set function. For example,³

$$\begin{aligned} H(X_i \sim X_j) &= H(X_i) - H(X_i \cap X_j), \\ H((X_i \cap X_j) \cup X_k) &= H(X_k) - H(X_i \cap X_j) + H(X_i \cap X_j \cap X_k), \\ H(X_i \sim (X_j \cup X_k)) &= H(X_i) - H(X_i \cap X_j) - H(X_i \cap X_k) + H(X_i \cap X_j \cap X_k), \\ H(X_i \sim (X_j \cap X_k)) &= H(X_i) - H(X_i \cap X_j \cap X_k), \\ &\dots\dots\dots \end{aligned}$$

Thus, we have defined all amounts of information $H(Q(X_{i_1}, X_{i_2}, \dots, X_{i_n}))$ for the variables $X_{i_1}, X_{i_2}, \dots, X_{i_n}$, where $Q(X_{i_1}, X_{i_2}, \dots, X_{i_n})$ denotes a symbol generated by a finite number of operations \cup, \cap or \sim on $X_{i_1}, X_{i_2}, \dots, X_{i_n}$.

Many of these amounts of information, as for example

$$(10) \quad H(X_i), \quad H(X_i \sim X_j), \quad H(X_i \cap X_j), \quad H((X_i \cap X_j) \sim X_k)$$

are always non-negative. But it should be noted that, in general, amounts of information can take on negative values. It is not hard to see, for example, that the amount of information $H(\bigcap_{i=1}^n X_i)$, $n > 2$, can be negative. In fact, let, in the probability field

$$\left(\begin{matrix} \omega_1, & \omega_2, & \omega_3 \\ \frac{1}{2}, & \frac{1}{4}, & \frac{1}{4} \end{matrix} \right),$$

random variables X_1, X_2, X_3 be given such that

$$X_1(\omega) = \begin{cases} 1, & \omega = \omega_1, \\ 0, & \omega = \omega_2, \omega_3, \end{cases} \quad X_2(\omega) = \begin{cases} 1, & \omega = \omega_2, \\ 0, & \omega = \omega_1, \omega_3, \end{cases} \quad X_3(\omega) = \begin{cases} 1, & \omega = \omega_3, \\ 0, & \omega = \omega_1, \omega_2. \end{cases}$$

Then

$$H(X_1 \cap X_2 \cap X_3) = -0.3774 < 0.$$

Amounts of information which are always non-negative will be called *proper amounts of information*, and others will be called *improper amounts of information*.

5°. **Theorem 1.** For a given sequence of variables X_1, X_2, \dots and their distribution P there

³ We use the symbol " \sim " rather than the symbol " $-$ " in $H(X_i \sim X_j)$ in order not to confuse $X_i \sim X_j$ with the difference of the variables $X_i - X_j$.

exists a corresponding sequence of sets A_1, A_2, \dots and an additive function φ on the ring \mathfrak{A} generated by the sequence $A_i, i = 1, 2, \dots$, such that

$$H(Q(X_{i_1}, X_{i_2}, \dots, X_{i_n})) = \varphi(Q(A_{i_1}, A_{i_2}, \dots, A_{i_n}))$$

for all collections of variables $X_{i_1}, X_{i_2}, \dots, X_{i_n}$ and all operations Q .

NOTE. Theorem 1 permits us to assert the validity of identities concerning amounts of information by using the corresponding identities for additive set functions. For example, the validity of (1), (2), (5) and (6) follows immediately from the well known identities (1'), (2'), (5') and (6') in § 2°. As regards the inequalities between amounts of information, further considerations are necessary. For example, since the amounts of information $H(X_1 \cap X_2)$ and $H((X_1 \cap X_2) \sim X_2)$ are proper, the following inequalities hold:

$$\begin{aligned} H(X_1 \sim X_2) &= H(X_1) - H(X_1 \cap X_2) \leq H(X_1), \\ H(X_1 \sim (X_2 \cup X_3)) &= H(X_1 \sim X_2) - H((X_1 \cap X_3) \sim X_2) \leq H(X_1 \sim X_2). \end{aligned}$$

Before proving the theorem, we prove the following lemma.

Lemma 1.1. Let $\varphi^{(i)}, \varphi^{(ij)}, \varphi^{(ijk)}, \dots$ ($1 \leq i < j < k$ are integers) be an arbitrarily prescribed collection of real numbers. Then there exists a sequence of sets A_1, A_2, \dots and an additive function φ on the ring \mathfrak{A} generated by the sequence $A_i, i = 1, 2, \dots$, such that

$$\varphi^{(i)} = \varphi(A_i), \quad \varphi^{(ij)} = \varphi(A_i \cup A_j), \quad \varphi^{(ijk)} = \varphi(A_i \cup A_j \cup A_k) \dots$$

PROOF. In view of the existence of a one-to-one correspondence between the collections (7) and (9), and noting that the values of (7) can be calculated from the values of (9), by addition and subtraction, we can proceed by analogy, and obtain from the numbers $\varphi^{(i)}, \varphi^{(ij)}, \varphi^{(ijk)}, \dots$, $i < j < k \dots = 1, 2, \dots$, the numbers $\varphi_1, \varphi_{12}, \varphi_{1\bar{2}}, \varphi_{\bar{1}2}, \varphi_{123}, \varphi_{12\bar{3}}, \varphi_{\bar{1}2\bar{3}}, \varphi_{1\bar{2}3}, \varphi_{\bar{1}23}, \varphi_{1\bar{2}\bar{3}}, \varphi_{\bar{1}2\bar{3}}, \dots$, such that

$$\begin{aligned} \varphi_1 &= \varphi_{12} + \varphi_{1\bar{2}}, \\ \varphi_{12} &= \varphi_{123} + \varphi_{12\bar{3}}, \quad \varphi_{1\bar{2}} = \varphi_{1\bar{2}3} + \varphi_{1\bar{2}\bar{3}}, \quad \varphi_{\bar{1}2} = \varphi_{\bar{1}23} + \varphi_{\bar{1}2\bar{3}}, \quad \text{etc.} \end{aligned}$$

Without difficulty, we can find a sequence of nonempty sets $A_0, A_1, A_{\bar{1}}, A_{12}, A_{1\bar{2}}, A_{\bar{1}2}, A_{1\bar{2}\bar{3}}, A_{12\bar{3}}, A_{1\bar{2}3}, \dots, A_{1\bar{2}\bar{3}}, \dots$, such that

$$\begin{aligned} A_0 &= A_1 + A_{\bar{1}}, \\ A_1 &= A_{12} + A_{1\bar{2}}, \quad A_{\bar{1}} = A_{\bar{1}2} + A_{\bar{1}\bar{2}}, \\ A_{12} &= A_{123} + A_{12\bar{3}}, \dots, A_{1\bar{2}} = A_{1\bar{2}3} + A_{1\bar{2}\bar{3}}, \quad \text{etc.} \end{aligned}$$

We construct a finitely-additive set function φ with values on $A_1, A_{12}, A_{1\bar{2}}, A_{\bar{1}2}, A_{123}, A_{12\bar{3}}, \dots, A_{1\bar{2}\bar{3}}, \dots$, defined as follows:

$$\begin{aligned} \varphi(A_1) &= \varphi_1, \\ \varphi(A_{12}) &= \varphi_{12}, \quad \varphi(A_{1\bar{2}}) = \varphi_{1\bar{2}}, \quad \varphi(A_{\bar{1}2}) = \varphi_{\bar{1}2}, \\ \varphi(A_{123}) &= \varphi_{123}, \quad \varphi(A_{12\bar{3}}) = \varphi_{12\bar{3}}, \dots, \varphi(A_{1\bar{2}3}) = \varphi_{1\bar{2}3}, \\ &\dots \end{aligned}$$

Then the sequence of sets

$$A_1 = A_1, \quad A_2 = A_{12} + A_{1\bar{2}}, \quad A_3 = A_{123} + A_{12\bar{3}} + A_{1\bar{2}3} + A_{1\bar{2}\bar{3}}, \dots$$

and the additive function φ on the corresponding ring \mathfrak{A} is, in fact, just what we require.

PROOF OF THEOREM 1. Taking §§ 3° and 4° into account, we need only prove the existence of a sequence of sets A_1, A_2, \dots and an additive function φ on the corresponding ring \mathfrak{A} such that for $1 \leq i < j < k$

$$\begin{aligned} H(X_i) &= \varphi(A_i), \\ H(X_i \cup X_j) &= \varphi(A_i \cup A_j), \\ H(X_i \cup X_j \cup X_k) &= \varphi(A_i \cup A_j \cup A_k), \\ &\dots \end{aligned}$$

Let $\varphi^{(i)} = H(X_i)$, $\varphi^{(i)}$ from Lemma 1.1 that such that for $1 \leq i$

Thus, Theorem 1 is

Corollary 1.1.

$$H(Q(X_1^{(1)}))$$

PROOF. Taking §

$$H((X_1^{(1)} \cup$$

which is easily deriv
Let

We let $H_{X_{n+1}}(C$
 $f(p(x_1, x_2, \dots, x_n|x_n$

$$H_{X_{n+1}}$$

For example,

Corollary 1.2.

$$H(Q$$

PROOF. Taking §

$$H((X_1 \cup$$

To do this, it is suffi

$$H(X$$

Thus, Corollary 1.2
It is easily seen

$$y_1 = (X$$

the amounts of info

(11)

are proper.

Lemma 2.1. 2

⁴ Here and in v

Let $\varphi^{(i)} = H(X_i)$, $\varphi^{(ij)} = H(X_i \cup X_j)$, $\varphi^{(ijk)} = H(X_i \cup X_j \cup X_k)$, \dots . It follows immediately from Lemma 1.1 that there exists a sequence of sets A_1, A_2, \dots and an additive function φ on \mathfrak{A} such that for $1 \leq i < j < k$

$$\begin{aligned} H(X_i) &= \varphi^{(i)} = \varphi(A_i), \\ H(X_i \cup X_j) &= \varphi^{(ij)} = \varphi(A_i \cup A_j), \\ H(X_i \cup X_j \cup X_k) &= \varphi^{(ijk)} = \varphi(A_i \cup A_j \cup A_k), \\ &\dots \end{aligned}$$

Thus, Theorem 1 is proved.

Corollary 1.1.

$$\begin{aligned} H(Q((X_1^{(1)} \cup X_1^{(2)} \cup \dots \cup X_1^{(m)}), X_2, X_3, \dots, X_n)) \\ = H(Q((X_1^{(1)}, X_1^{(2)}, \dots, X_1^{(m)}), X_2, X_3, \dots, X_n)). \end{aligned}$$

PROOF. Taking §§ 3° and 4° into account, we need only prove the validity of the equality

$$\begin{aligned} H((X_1^{(1)} \cup X_1^{(2)} \cup \dots \cup X_1^{(m)}) \cup X_2 \cup X_3 \cup \dots \cup X_n) \\ = H((X_1^{(1)}, X_1^{(2)}, \dots, X_1^{(m)}), X_2, X_3, \dots, X_n), \end{aligned}$$

which is easily derived directly from the definitions. Thus, Corollary 1.1 is proved.

Let

$$f(p(x_1, x_2, \dots, x_n)) \equiv H(Q(X_1, X_2, \dots, X_n)).$$

We let $H_{X_{n+1}}(Q(X_1, X_2, \dots, X_n))$ denote the mathematical expectation of the function $f(p(x_1, x_2, \dots, x_n|x_{n+1}))$ of X_{n+1} with respect to the distribution $P(X_{n+1})$, i. e.

$$H_{X_{n+1}}(Q(X_1, X_2, \dots, X_n)) = \sum_{x_{n+1}} P(X_{n+1}) f(p(x_1, x_2, \dots, x_n|x_{n+1})).$$

For example,

$$H_{X_2}(X_1) = \sum_{x_2} P(X_1) \left[- \sum_{x_1} p(x_1|x_2) \log p(x_1|x_2) \right].$$

Corollary 1.2.

$$H(Q(X_1, X_2, \dots, X_n) \sim X_{n+1}) = H_{X_{n+1}}(Q(X_1, X_2, \dots, X_n)).$$

PROOF. Taking §§ 3° and 4° into account, we need only prove the validity of the equality

$$H((X_1 \cup X_2 \cup \dots \cup X_n) \sim X_{n+1}) = H_{X_{n+1}}(X_1 \cup X_2 \cup \dots \cup X_n).$$

To do this, it is sufficient to prove that $H(X_1 \sim X_2) = H_{X_2}(X_1)$. But

$$\begin{aligned} H(X_1 \sim X_2) &= H(X_1) - H(X_1 \cap X_2) \\ &= \sum_{x_2} p(x_2) \left[- \sum_{x_1} p(x_1|x_2) \log p(x_1|x_2) \right] = H_{X_2}(X_1). \end{aligned}$$

Thus, Corollary 1.2 is proved.

It is easily seen from Corollaries 1.1 and 1.2 and equation (10) that for any symbols

$$\begin{aligned} y_1 &= (X_1^{(1)} \cup X_1^{(2)} \cup \dots \cup X_1^{(n_1)}), \quad y_2 = (X_2^{(1)} \cup X_2^{(2)} \cup \dots \cup X_2^{(n_2)}), \\ y_3 &= (X_3^{(1)} \cup X_3^{(2)} \cup \dots \cup X_3^{(n_3)}) \end{aligned}$$

the amounts of information

$$(11) \quad H(y_1), \quad H(y_1 \sim y_2), \quad H(y_1 \cap y_2), \quad H((y_1 \cap y_2) \sim y_3)$$

are proper.

Lemma 2.1. The random variables X_1 and X_2 are independent \Leftrightarrow $H(X_1|X_2) = 0$.

⁴ Here and in what follows, the symbol " \Leftrightarrow " denotes equivalence of propositions.

$$H(X_1|X_2) \neq H_{X_2}(X_1)$$

Grove
Problems
notation???

Lemma 2.2. X_1 with probability $p(x_1, x_2) = 1$ is a function of $X_2 \Leftrightarrow H(X_1 \sim X_2) = 0$.

Theorem 2. Let $(X_1, X_2, \dots; p)$ correspond to $(A_1, A_2, \dots; \varphi)$ in the sense of Theorem 1, and let the variation of the additive function φ be non-zero everywhere except for an empty set. Then the following correspondences hold:

1. X_i and X_j are independent $\Leftrightarrow A_i \cap A_j = \emptyset$.
2. $X_i = f(X_j)$ with $p(x_i, x_j) = 1 \Leftrightarrow A_i \subset A_j$ for all $i, j = 1, 2, \dots$.

The proof of the Theorem follows directly from Lemmas 2.1, 2.2 and Theorem 1.

Corollary 2.1. Let there be given the variables X_1, X_2, \dots with distribution P , and the functions $f_1(X_1), f_2(X_2), \dots$. Then there exist corresponding sets A_1, A_2, \dots ; and $A'_i, i = 1, 2, \dots$, $A'_1 \subset A_1, A'_2 \subset A_2, \dots$ and an additive function φ on the ring \mathfrak{A} generated by the sequence $A_i, i = 1, 2, \dots$, such that for all collections of variables $X_{i_1}, X_{i_2}, \dots, X_{i_m}$ ($i_v, v = 1, 2, \dots, m$, are positive or negative integers) and all operations Q the following equality is satisfied:

$$H(Q(X_{i_1}, X_{i_2}, \dots, X_{i_m})) = \varphi(Q(A_{i_1}, A_{i_2}, \dots, A_{i_m})),$$

and

$$X_{-i} = f_i(X_i), \quad A_{-i} = A'_i \quad i = 1, 2, \dots$$

In view of Corollary 2.1, it is easily seen that

- 1) $H(f(X_1) \cap X_2) \leq H(X_1 \cap X_2)$,
- 2) $H(f(X_1) \cup X_2) \leq H(X_1 \cup X_2)$,
- 3) $H(f(X_1) \cap X_2) = H(X_1 \cap X_2) \Leftrightarrow H(X_2 \sim f(X_1)) = H(X_2 \sim X_1)$ (see [4]).

Theorem 3. In order that the variables X_1, X_2, \dots form a Markov chain it is necessary and sufficient that for any collection $X_{i_1}, X_{i_2}, \dots, X_{i_n}, i_1 < i_2 < \dots < i_n$, the equality

$$H(X_{i_1} \cap X_{i_2} \cap \dots \cap X_{i_n}) = H(X_{i_1} \cap X_{i_n})$$

holds.

PROOF. Let us take the variables $X_{i_1}, X_{i_2}, X_{i_3}, i_1 < i_2 < i_3$. We shall prove that they form a Markov chain $\Leftrightarrow H(X_{i_1} \cap X_{i_2} \cap X_{i_3}) = H(X_{i_1} \cap X_{i_3})$. This is evident from the following equivalence relations:

$$\begin{aligned} H(X_{i_1} \cap X_{i_2} \cap X_{i_3}) &= H(X_{i_1} \cap X_{i_3}) \\ &\Leftrightarrow H((X_{i_1} \cap X_{i_3}) \sim X_{i_2}) = H_{X_{i_2}}(X_{i_1} \cap X_{i_3}) \\ &= \sum_{x_{i_2}} p(x_{i_2}) \left[\sum_{x_{i_1}, x_{i_3}} p(x_{i_1}, x_{i_3} | x_{i_2}) \log \frac{p(x_{i_1}, x_{i_3} | x_{i_2})}{p(x_{i_1} | x_{i_2}) p(x_{i_3} | x_{i_2})} \right] = 0 \end{aligned}$$

\Leftrightarrow for any $p(x_{i_2}) > 0$,

$$\sum_{x_{i_1}, x_{i_3}} p(x_{i_1}, x_{i_3} | x_{i_2}) \log \frac{p(x_{i_1}, x_{i_3} | x_{i_2})}{p(x_{i_1} | x_{i_2}) p(x_{i_3} | x_{i_2})} = 0$$

\Leftrightarrow for any $p(x_{i_2}) > 0$,

$$p(x_{i_1}, x_{i_3} | x_{i_2}) = p(x_{i_1} | x_{i_2}) p(x_{i_3} | x_{i_2})$$

$\Leftrightarrow X_{i_1}, X_{i_2}, X_{i_3}$ form a Markov chain.

Now let us take four variables $X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}, i_1 < i_2 < i_3 < i_4$. Assuming that our proposition is true for $n \leq 3$, we prove that it is also true for $n \leq 4$. Let us suppose that $X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}$ form a Markov chain. Then for fixed X_{i_2} the pair X_{i_1}, X_{i_3} is independent of X_{i_4} :

$$H_{X_{i_2}}(X_{i_1} \cap X_{i_4}) = 0, \quad H_{X_{i_2}}(X_{i_3} \cap X_{i_4}) = 0, \quad H_{X_{i_2}}((X_{i_1}, X_{i_3}) \cap X_{i_4}) = 0.$$

Therefore,

$$H_{X_{i_2}}(X_{i_1} \cap X_{i_2} \cap X_{i_4}) = H_{X_{i_2}}(X_{i_1} \cap X_{i_4}) + H_{X_{i_2}}(X_{i_3} \cap X_{i_4}) - H_{X_{i_2}}((X_{i_1}, X_{i_3}) \cap X_{i_4}) = 0.$$

Consequently,

Conversely, let u

$$H(X_{i_1} \cap X_{i_2})$$

and

Then

$$H_{X_{i_3}}(X_{i_1} \cap$$

and consequently,

$$H_{X_{i_3}}$$

Thus, for fixed X_{i_2} the variable X_{i_1} is chain.

Proceeding in thi

Lemma 4.1. In form a Markov chain The proof follow

Theorem 4. In a) for all collect

or, equivalently, b) for all collect

H

PROOF. Proposit b) follows immediate

$$H\left(\bigcup_{v=1}^n X_{i_v}\right) =$$

Theorem 5. If of sets A_1, A_2, \dots, A_n , $A_i, i = 1, 2, \dots, n$, suc

PROOF. It follow A_1, A_2, \dots and an ε $X_{i_1}, X_{i_2}, \dots, X_{i_n}, i$

where

$$(12) \quad H(\dots)$$

$$H(X_{i_1} \cap X_{i_2} \cap X_{i_3} \cap X_{i_4}) = H(X_{i_1} \cap X_{i_2} \cap X_{i_4}) \\ - H_{X_{i_3}}(X_{i_1} \cap X_{i_2} \cap X_{i_4}) = H(X_{i_1} \cap X_{i_4}).$$

Conversely, let us suppose that

$$H(X_{i_1} \cap X_{i_2} \cap X_{i_3}) = H(X_{i_1} \cap X_{i_3}), \\ H(X_{i_1} \cap X_{i_2} \cap X_{i_4}) = H(X_{i_1} \cap X_{i_4}), \quad H(X_{i_2} \cap X_{i_3} \cap X_{i_4}) = H(X_{i_2} \cap X_{i_4}), \\ H(X_{i_1} \cap X_{i_3} \cap X_{i_4}) = H(X_{i_1} \cap X_{i_4})$$

and

$$H(X_{i_1} \cap X_{i_2} \cap X_{i_3} \cap X_{i_4}) = H(X_{i_1} \cap X_{i_4}).$$

Then

$$H_{X_{i_3}}(X_{i_1} \cap X_{i_2} \cap X_{i_4}) = H(X_{i_1} \cap X_{i_2} \cap X_{i_4}) - H(X_{i_1} \cap X_{i_2} \cap X_{i_3} \cap X_{i_4}) \\ = H(X_{i_1} \cap X_{i_4}) - H(X_{i_1} \cap X_{i_4}) = 0,$$

and consequently,

$$H_{X_{i_3}}((X_{i_1}, X_{i_2}) \cap X_{i_4}) = H_{X_{i_3}}(X_{i_1} \cap X_{i_4}) + H_{X_{i_3}}(X_{i_2} \cap X_{i_4}) \\ - H_{X_{i_3}}(X_{i_1} \cap X_{i_2} \cap X_{i_4}) = 0.$$

Thus, for fixed X_{i_3} the pair X_{i_1}, X_{i_2} is independent of X_{i_4} . We can show, similarly, that for fixed X_{i_2} the variable X_{i_1} is independent of the pair X_{i_3}, X_{i_4} , i. e. $X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}$ form a Markov chain.

Proceeding in this way, we see by induction that the theorem holds.

Lemma 4.1. *In order that X_1, X_2, \dots be independent, it is necessary and sufficient that they form a Markov chain and are pairwise independent.*

The proof follows directly from the definition of a Markov chain.

Theorem 4. *In order that X_1, X_2, \dots be independent, it is necessary and sufficient that*

a) *for all collections $X_{i_1}, X_{i_2}, \dots, X_{i_n}$, $i_1 < i_2 < \dots < i_n$,*

$$H(X_{i_1} \cap X_{i_2} \cap \dots \cap X_{i_n}) = 0,$$

or, equivalently,

b) *for all collections $X_{i_1}, X_{i_2}, \dots, X_{i_n}$, $i_1 < i_2 < \dots < i_n$,*

$$H(X_{i_1}, X_{i_2}, \dots, X_{i_n}) = H(X_{i_1}) + H(X_{i_2}) + \dots + H(X_{i_n}).$$

PROOF. Proposition a) follows directly from Theorem 3 and Lemma 4.1, and proposition b) follows immediately from proposition a) and the obvious identity

$$H\left(\bigcup_{\nu=1}^n X_{i_\nu}\right) = \sum_{\nu=1}^n H(X_{i_\nu}) - \sum_{\nu_1 < \nu_2 \leq n} H(X_{i_{\nu_1}} \cap X_{i_{\nu_2}}) + \dots + (-1)^{n+1} H\left(\bigcap_{\nu=1}^n X_{i_\nu}\right).$$

Theorem 5. *If X_1, X_2, \dots form a Markov chain, then there exists a corresponding sequence of sets A_1, A_2, \dots and a non-negative additive function φ on the ring \mathfrak{A} generated by the sequence $A_i, i = 1, 2, \dots$, such that for all collections $X_{i_1}, X_{i_2}, \dots, X_{i_n}$ and all operations Q*

$$H(Q(X_{i_1}, X_{i_2}, \dots, X_{i_n})) = \varphi(Q(A_{i_1}, A_{i_2}, \dots, A_{i_n})).$$

PROOF. It follows immediately from Theorems 1 and 3 that there exists a sequence of sets A_1, A_2, \dots and an additive function φ on the corresponding ring \mathfrak{A} such that for all collections $X_{i_1}, X_{i_2}, \dots, X_{i_n}$, $i_1 < i_2 < \dots < i_n$, and all operations Q

$$H(Q(X_{i_1}, X_{i_2}, \dots, X_{i_n})) = \varphi(Q(A_{i_1}, A_{i_2}, \dots, A_{i_n})),$$

where

$$(12) \quad H(X_{i_1} \cap X_{i_2} \cap \dots \cap X_{i_n}) = H(X_{i_1} \cap X_{i_n}) = \varphi(A_{i_1} \cap A_{i_n}) \\ = \varphi(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}).$$

In order to prove that the function φ is non-negative, we need only prove, in view of § 3°, that all values of the function φ in the collection (7) are non-negative. We know from (11) that for any $i < j < k < l$

$$\begin{aligned}\varphi(A_i \cap \bar{A}_j) &= \varphi(A_i - A_j) = H(X_i \sim X_j) \geq 0, \\ \varphi(A_i \cap A_j \cap \bar{A}_k) &= \varphi((A_i \cap A_j) - A_k) = H((X_i \cap X_j) \sim X_k) \geq 0, \\ \varphi(A_i \cap A_j) &= \varphi(A_i - A_j) = H(X_j \sim X_i) \geq 0, \\ \varphi(\bar{A}_i \cap A_j \cap \bar{A}_k) &= \varphi(A_j - (A_i \cup A_k)) = H(X_j \sim (X_i \cup X_k)) \geq 0, \\ \varphi(\bar{A}_i \cap A_j \cap A_k \cap \bar{A}_l) &= \varphi((A_j \cap A_k) - (A_i \cup A_l)) = H((X_j \cap X_k) \sim (X_i \cup X_l)) \geq 0.\end{aligned}$$

In view of condition (11), it is not hard to verify that for a portion of the values of collection (7) we have, for any $1 < m_1 < m_2 < m$,

$$\begin{aligned}\varphi(A_1)(\bigcap_{1 < i \leq m} \bar{A}_i) &= \varphi(A_1 \cap \bar{A}_2), \\ \varphi((\bigcap_{1 \leq i \leq m} A_i)(\bigcap_{m_1 < i \leq m} \bar{A}_i)) &= \varphi(A_1 \cap A_{m_1} \cap \bar{A}_{m_1+1}), \\ \varphi((\bigcap_{1 \leq i \leq m_1} \bar{A}_i)(\bigcap_{m_1 < i \leq m} A_i)) &= \varphi(\bar{A}_{m_1} \cap A_{m_1+1} \cap A_m), \\ \varphi((\bigcap_{1 \leq i \leq m_1} \bar{A}_i)(A_{m_1+1})(\bigcap_{m_1+1 < i \leq m} \bar{A}_i)) &= \varphi(\bar{A}_{m_1} \cap A_{m_1+1} \cap \bar{A}_{m_1+2}), \\ \varphi((\bigcap_{1 \leq i \leq m_1} \bar{A}_i)(\bigcap_{m_1 < i \leq m_2} A_i)(\bigcap_{m_2 < i \leq m} \bar{A}_i)) &= \varphi(\bar{A}_{m_1} \cap A_{m_1+1} \cap A_{m_2} \cap \bar{A}_{m_2+1}),\end{aligned}$$

and all the remaining values of collection (7) are equal to zero. For example:

$$\varphi(A_1 \cap \bar{A}_2 \cap A_3) = \varphi(A_1 \cap A_3) - \varphi(A_1 \cap A_2 \cap A_3) = \varphi(A_1 \cap A_3) - \varphi(A_1 \cap A_3) = 0.$$

Thus, the function φ is always non-negative.

The following proposition follows immediately from Theorem 5: if X_1, X_2, X_3, X_4 form a Markov chain, then

1. $H(X_1 \cap X_4) = H(X_1 \cap X_2 \cap X_3 \cap X_4) \leq H(X_2 \cap X_3)$ (the converse of Shannon's proposition);
2. $H((X_1 \cap X_2) \sim X_3) = H(X_1 \cap X_2) - H(X_1 \cap X_2 \cap X_3) \leq H(X_1 \cap X_2)$ (see [5]).

6°. So far, we have been concerned only with random variables which take on a finite number of values. Now, we generalize the definition of the amount of information for variables $X_i, i = 1, 2, \dots$, taking on an arbitrary number of values as follows:

$$H\left(\bigcup_{i=1}^n X_i\right) = \sup_{(f_1, f_2, \dots, f_n)} H\left(\bigcup_{i=1}^n f_i(X_i)\right), \quad n = 1, 2, \dots,$$

with the condition

$$\sup_{i=1}^n H\left(\bigcup_{i=1}^n f_i(X_i)\right) < \infty,$$

where the upper bound is taken over all possible collections (f_1, f_2, \dots, f_n) of measurable functions which take a finite number of values. In accordance with the identities of the additive set function we define, analogously,

$$H\left(\bigcap_{i=1}^n X_i\right) = \sum_{i \leq n} H(X_i) - \sum_{i < j \leq n} H(X_i \cap X_j) + \dots + (-1)^{n+1} H\left(\bigcup_{i=1}^n X_i\right), \quad n = 1, 2, \dots,$$

$$H(X_i \sim X_j) = H(X_i) - H(X_i \cap X_j),$$

$$H((X_i \cap X_j) \cup X_k) = H(X_k) - H(X_i \cap X_j) + H(X_i \cap X_j \cap X_k).$$

Then, we can easily prove that all the above results are valid not only for variables taking on a finite number of values, but also for variables taking on an arbitrary number of values.

Received by the editors
April 28, 1960

- [1] A. YA. KHINCI VIII, 3
- [2] A. YA. KHINC 1(67), 1
- [3] R. L. DOBRUSI tions, 3.
- [4] I. M. GEL'FAN random pp. 3-5.
- [5] I. P. TSAREGR Prob. A

In this paper function is constructed between entropies

STATISTICAL

Let p and q respectively, by the $2n$ -dimensional

It follows that function F_{pq} is

(1.1)

where $w = (w_1,$

For our proposition in E^n : probability of probability of intersection with the set of Appendix, Definition triangle inequality distance between q and p in E^n , it is certain. Accordingly we

¹ Through of w , which d

REFERENCES

- [1] A. YA. KHINCHIN, *The concept of entropy in the theory of probability*, Uspekhi Matem. Nauk, VIII, 3 (55), 1953, pp. 3-20. (In Russian.)
- [2] A. YA. KHINCHIN, *Fundamental theorems of information theory*, Uspekhi Matem. Nauk XI, 1(67), 1956, pp. 17-25. (In Russian.)
- [3] R. L. DOBRUSHIN, *Information transmission in a channel with feedback*, Theory Prob. Applications, 3, 1958, pp. 367-383. (English translation.)
- [4] I. M. GEL'FAND and A. M. YAGLOM, *On the calculation of the amount of information regarding a random function contained in another such function*, Uspekhi Matem. Nauk, XII, 3, 1957, pp. 3-52 (In Russian.)
- [5] I. P. TSAREGRADSKII, *A note on the capacity of a stationary channel with finite memory*, Theory Prob. Applications, 3, 1958, pp. 79-91. (English translation.)

ON THE AMOUNT OF INFORMATION

HU KUO TING (MOSCOW-TIENTSIN)

(Summary)

In this paper a connection between the amount of information and a certain additive set function is constructed. This connection enables us not only to obtain all known algebraic relations between entropies and various information quantities, but also all new algebraic relations.

STATISTICAL METRIC SPACES ARISING FROM SETS OF RANDOM VARIABLES
IN EUCLIDEAN n -SPACE

B. SCHWEIZER AND A. SKLAR

Introduction

Let p and q be random vectors in n -dimensional Euclidean space E^n with distributions given, respectively, by the n -dimensional distribution functions G_p , G_q , and joint distribution given by the $2n$ -dimensional distribution function H_{pq} , so that

$$\begin{aligned} H_{pq}(w_1, \dots, w_n, +\infty, \dots, +\infty) &= G_p(w_1, \dots, w_n), \\ H_{pq}(+\infty, \dots, +\infty, w_{n+1}, \dots, w_{2n}) &= G_q(w_{n+1}, \dots, w_{2n}). \end{aligned}$$

It follows that $d(p, q)$, the distance between p and q , is a random variable whose distribution function F_{pq} is completely determined by H_{pq} and is given by

$$(1.1) \quad F_{pq}(x) = \begin{cases} 0, & x \leq 0, \\ \int_{|u-v| < x} \dots \int dH_{pq}(w), & x > 0, \end{cases}$$

where $w = (w_1, w_2, \dots, w_{2n})$, $u = (w_1, \dots, w_n)$, and $v = (w_{n+1}, \dots, w_{2n})$ ¹.

For our purposes it is convenient to think of the random variable p as a «particle» whose position in E^n is determined probabilistically by its distribution function G_p . Then $G_p(u)$ is the probability of finding the particle « p » in the region $\{v; v_1 < u_1, \dots, v_n < u_n\}$; and $F_{pq}(x)$ is the probability of finding the particles « p » and « q » removed from each other by a distance less than x . Using this interpretation, it is readily seen that such a set of random variables $\{p, q, \dots\}$ together with the set of associated distance distribution functions $\{F_{pq}\}$ is a statistical metric space (see Appendix, Definition A. 1). Indeed, the only property that is not immediate is the generalized triangle inequality (Definition A. 1, IV). But this follows from the fact that if it is certain that the distance between p and q is less than x (i. e., $F_{pq}(x) = 1$), and equally certain that the distance between q and r is less than y (i. e., $F_{qr}(y) = 1$), then in view of the ordinary triangle inequality in E^n , it is certain that the distance between p and r is less than $x+y$ (i. e., $F_{pr}(x+y) = 1$). Accordingly we are led to the following:

¹ Throughout this paper, boldface symbols will denote vectors. Moreover, with the exception of w , which denotes a vector in E^{2n} , all vectors are in E^n .