through P_0, A_n, C_n tends to $R_0 = \frac{1}{2}(1-\gamma)/(1+\gamma) < \frac{1}{2}$ as $n \to \infty$. We have proved before that the radius of the circumference passing through P_0, A_n, B_n tends to $\frac{1}{2}$. Therefore the radius of the circumference passing through points P_0, P_1, P_2, where $P_1 \epsilon \Gamma$, $P_2 \epsilon \Gamma$, $P_1 \neq P_0$, $P_2 \neq P_0$, $P_1 \neq P_2$, has no limit as $P_1 \to P_0$, $P_2 \to P_0$. This completes the proof that curve Γ does not possess the Alt curvature at P_0.

It is evident from (2), (7) that the curve Γ has no osculatory plane at P_0.

REFERENCES

[1] F. Alt, *Über eine metrische Definition der Krümmung einer Kurve*, Dissertation, Wien 1931.

[2] S. Gołąb, *Sur quelques propriétés des lignes géodésiques*, Annales Polonici Mathematici 8 (1960), p. 91-103.

[3] K. Menger, *Untersuchungen über allgemeine Metrik IV. Zur Metrik der Kurven*, Mathematische Annalen 103 (1930), p. 466-501.

MATHEMATICAL INSTITUTE OF THE POLISH ACADEMY OF SCIENCES

Reçu par la Rédaction le 24. 1. 1961

COLLOQUIUM MATHEMATICUM

VOL. IX 1962 FASC. 1

INFORMATION WITHOUT PROBABILITY

BY

R. S. INGARDEN AND K. URBANIK (WROCŁAW)

1. Introduction. Since the first definition of the notion of information given in its full generality by C. E. Shannon in 1948 (1), many mathematical investigations have been concerned with this notion (2). The general tendency of these investigations (initiated by Shannon himself (3)) has been to separate the definition of information, say H, from the explicit formula

$$H = -\sum_{i=1}^{n} p_i \log p_i, \tag{1}$$

adopted by Shannon from statistical physics (Boltzmann's formula for entropy). Here p_i denotes the probability of the i-th elementary event ($i = 1, \ldots, n$; we consider first a finite, or at any rate discrete, probability scheme, convergence of the sum in the case $n = \infty$ being assumed). It was felt from the beginning that such a formula as (1) should be rather a result than a starting point of the theory. Moreover, some investigators, as e. g. Rényi (4), considered (1) as too narrow to cover all possible applications of information theory and tried to generalize this formula. Of course, to get such a generalization in a natural way, it is necessary to have an abstract definition of information, i. e. by means of a set of axioms (this set may be subsequently diminished in the generalization process). Many such sets of axioms have so far been proposed (5) and their consequences as well as mutual interrelations have been investigated. All axiomatic definitions of information known to the present authors are equivalent to formula (1) (except Rényi's gene-

(1) Cf. Shannon [11]. The numbers in square brackets refer to the list of literature given at the end of this paper, p. 149-150.

(2) Cf., e. g., Khintchine [8], Feinstein [4], Rényi [10], where further references may be found.

(3) Cf. [11], p. 392, and Appendix 2, p. 419.

(4) Cf. Rényi [10].

(5) Cf. Shannon [11], Khintchine [8], Faddeev [3], Rényi [10].

ralization just mentioned) and all are essentially based on the notion of probability, i. e. the probability distribution of some set of events is considered as given.

Before we formulate our problem, let us discuss shortly the relation of information theory to probability theory. According to (1), the range of information theory is only a proper part of the range of probability theory, because of the impossibility of a direct application of (1) to the case of continuous probability distributions. In fact, transition in (1) from the sum to an integral is not unique (it depends on units chosen, cf. the discussion below). Naturally enough the question arises how to extend information theory to the continuous case. The first indication in this direction was given by Shannon (6) himself and his idea was subsequently developed by Kolmogorov (7). As is well known, stochastic events $a, b, \ldots$ form a Boolean algebra, or ring (8), say $\mathfrak{a}$. Every subalgebra (subring) $A, B, \ldots$ of $\mathfrak{a}$ may be considered as a trial, e. g., a measurement of some physical quantity. We denote by $a_i, b_j, \ldots$ elementary events (atoms) (9) of $A, B, \ldots$, respectively. If we have two trials A, B belonging to the same $\mathfrak{a}$, i. e. if we have two compatible trials (this condition is important in view of applications to quantum physics) and if both subalgebras are finite (or at least discrete, in the infinite case assuming convergence, as before) we may form the well-defined expression (6)

$$H(A, B) = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{p_i p_j}, \tag{2}$$

where p_{ij} denotes probability of the event consisting in the result a_i of the first trial and the result b_j of the second one. Kolmogorov (7) generalized formula (2) for the case of arbitrary (in particular, continuous) A and B by putting

$$I(A, B) = \sup_{A_1 \subset A, B_1 \subset B} H(A_1, B_1), \tag{3}$$

where the supremum is taken over all finite subalgebras $A_1 \subset A$ and

(6) Cf. [11], Appendix 7, p. 655.

(7) Cf. Gelfand, Kolmogorov and Yaglom [5].

(8) Cf., e. g., Sikorski [12], p. 2, 46. The point of view of Boolean algebras is connected rather with set theory, that of Boolean rings rather with algebra, but both are equivalent (in some sense, although in another sense they are slightly different). In the sequel we shall use the ring point of view as a little more convenient for our purposes. Boolean operations (used by us explicitly) will be denoted as follows: union (joint) of elements by $\cup$, intersection (meet) by $\cap$, difference (ordinary, non-symmetric) by $\setminus$. 0 will denote the zero element.

(9) An element $a \in A$ is said to be an *atom* of A provided $a \neq 0$ and every element $b \in A$ contained in a ($b \subset a$, i. e. $b \cup a = a$ or $b \cap a = b$) is equal either to a or to 0.

$B_1 \subset B$. If A and B are finite, then $I(A, B)$ is real and non-negative; in the general case it may take, besides, only the value $I = +\infty$. Since definition (3) covers all possible cases considered in probability theory, in contradistinction to (1), some authors (10) try to base information theory on notion (3) only and not on (1). But such a radical point of view does not seem very rational, especially from the point of view of applications to physics. In fact, mathematical intuitions connected with (1) and (3) are rather different, in spite of the similarity of formulae (1) and (2) and of some inner relationship between the two notions. This may easily be seen by considering the list of principal properties of $I(A, B)$ (11):

1. $I(A, B) = I(B, A)$,

2. $I(A, B) = 0$ whenever A and B are stochastically independent, i. e. $p_{ij} = p_i p_j$ for all i, j.

3. If (A_1, B_1) and (A_2, B_2) are stochastically independent, then $I((A_1, A_2), (B_1, B_2)) = I(A_1, B_1) + I(A_2, B_2)$,

4. If $A_1 \subset A$, then $I(A_1, B) \leqslant I(A, B)$.

In the light of these properties the description of $I(A, B)$ as: "the amount of information contained in the trial A with respects to the trial B" (12) does not seem very expressive. The term "correlational information" or "informational coefficient of correlation" between A and B (in contradistinction to, e. g. Pearson's correlational coefficient) seems a more appropriate term, especially because of property 2. Such a meaning has rather little to do with the meaning of H (or $H(A)$) (1) which gives a measure of the degree of uncertainty connected with the trial A (this uncertainty corresponds roughly to dispersion in usual statistics, since we have just seen an approximate correspondence between $I(A, B)$ and statistical coefficients of correlation).

Let us see, however, whether it is possible to retain the intuitions connected with (1) for the continuous case, especially in view of their important applications in physics (13).

The problem is to define information not for two but for one trial (or, in a slightly different approach, not for two, but for one random variable, e. g., physical quantity). It may be remarked first that Kol-

(10) Cf. [2], p. 4. Also Shannon [11] writes at the beginning of Appendix 7, p. 655: "The following will indicate a more general and more rigorous approach to the central definitions of communication theory."

(11) Cf. [5], p. 746, or [2], p. 30.

(12) Cf. [5], p. 745.

(13) Cf., e. g., Brillouin's presentation [1].

mogorov's method of generalization from the finite to the general case is here useless. Indeed,

$$J_1(A) = \sup_{A_1 \subset A} H(A) \tag{4}$$

(where the supremum is taken over all finite subalgebras $A_1 \subset A$) is always infinite for continuous distributions, in contrary to $I(A, B)$ which may be finite in the latter case. Whereas (3) may be correctly written in this case as [14]

$$I(A, B) = \int_A \int_B p(da, db) \log \frac{p(da, db)}{p(da)p(db)}, \tag{5}$$

the analogous expression

$$J(A) = -\int_A p(da) \log p(da) \tag{6}$$

is evidently incorrect and does not give $J_1(A)$ (see (4)). Instead, the expression

$$H_q(A) = \int_A p(da) \log \frac{q(da)}{p(da)} = \int_A \varrho(a) \log \frac{\sigma(a)}{\varrho(a)} da \tag{7}$$

is correct and, in general finite (the latter if there exist "densities": $\varrho(a)$ and $\sigma(a)$ such that $p(a) = \varrho(a)da$, $q(a) = \sigma(a)da$). We may call $H_q(A)$ the relative information contained in the trial A with respect to the "grain distribution" q (or σ) [15] in contradistinction to the absolute information $H(A)$ in the finite (or discrete) case. Grain distribution $q(a)$ is, alongside with $p(a)$, the second measure defined on the Boolean ring A. It may be considered either as "a distribution of our interest" [16] or as something objectively determined. An example of the latter case is the well-known entropy of (half-) classical statistical mechanics [17]

$$S = -\int \varrho(P, Q) \log[h^s \varrho(P, Q)] dP dQ, \tag{8}$$

[14] Cf. [5], p. 745.

[15] This notion is closely related to the expression $I_\varphi(\xi)$ defined by Vincze [14], p. 683. The terminology "coarse grained entropy" and "fine grained entropy" is used in physics, cf., e. g., Tolman [13]. Our term "relative information" should not be confused with the term "conditional information", cf., e. g., [8], p. 5, of the English translation.

[16] Cf. Vincze [15], p. 682.

[17] Cf. Landau and Lifschitz [9], p. 39. Another physical example may be found in optics where entropy has been defined by Ingarden, cf. [6], p. 179 (we have there the wave length λ of light instead of the constant h in (8)).

where P and Q denote, respectively, s-dimensional manifolds of momentum and position coordinates of the mechanical system (with s degrees of freedom) and h is the Planck action constant. In this case σ is constant and equal to h^{-s}, i. e. to the density of "elementary cells" in the phase space (or Heisenberg's "cells of uncertainty") [18].

As a further example of (7) we consider the important quantity

$$I(A) = -\int_A p(da) \log\left[\frac{Vp(da)}{da}\right] = -\int_A \varrho(a) \log[V\varrho(a)] da, \tag{9}$$

where

$$V = \int_A da, \tag{10}$$

i. e. the volume of the continuous space considered. Now $\sigma = V^{-1}$. Comparing (9) with (6) we may write

$$I(A) = J(A) - \log V; \tag{11}$$

i. e. $I(A)$ represents the deviation of the "absolute information" $J(A)$ from its maximal possible value (e. g. for physical systems occurring in the equilibrium state.) This deviation is, according to (9), independent of the system of units used and gives, therefore, a well-defined quantity. Although the right-hand side of (11) is incorrect to the same degree as (6), it transmits the very idea that was the physical purpose for which Boltzmann introduced his famous "H"-quantity. In the finite case we may take instead of V (see (10))

$$n = \sum_{i=1}^{n} 1 \tag{12}$$

and we get in place of (9)

$$I(A) = -\sum_{i=1}^{n} p_i \log np_i = H(A) - \log n \, ^{(19)}, \tag{13}$$

which completes our general definition of the quantity $I(A)$.

Summing up we may say that there are two fundamental informational notions, $I(A)$ and $I(A, B)$, which are uniquely determined in all the fields of probability theory, whereas quantities $H(A)$ and $H_q(A)$

[18] If we take $\sigma = 1 \times$ units of a^{-1}, e. g. $\sigma = 1\ \text{cm}^{-3}$ etc., we formally obtain formula (6), but the units chosen must be held in mind (which is frequently forgotten), and with $\dfrac{p(da)}{da}$ instead of $p(da)$.

[19] Our $-I(A)$ in (13), corresponds to I_n considered by Vincze [15], p. 682, as the "information of the system of events".

may be defined only in cases of discrete and continuous probability distributions, respectively. (In the sequel we shall consider only $H(A)$ and $I(A, B)$, which are historically as well as practically the most important. The others may be constructed in a similar way.)

Now we can formulate the central problem of our paper. It consists in the question whether the notion of information may or may not be separated not only from the formula (1), or (5), but also from the notion of probability itself. In other words: can probability theory be considered as a branch of information theory, and not vice versa, as has been held hitherto? The answer to this question is positive and we shall show below how it is possible to define probability by means of the notion of information. The situation is in a high degree analogous to that regarding the notion of mathematical expectation (expectation value). Expectation value may be determined by means of probability distribution, and vice versa, probability distribution may be constructed (by the solution of the s. c. momentum problem) from some expectation values.

The possibility of reversing the-hitherto-usual direction of deduction seems interesting not only from a purely logical and mathematical point of view, but also from the point of view of the philosophy of mathematical and physical notions. Indeed, information seems intuitively a much simpler and more elementary notion than that of probability. It gives more a cruder and global description of some situations physical or other) than probability does. Therefore, information represents a more primary step of knowledge than that of cognition of probabilities (just as probability description is cruder and more global than deterministic description). Furthermore, a principal separation of notions of probability and information seems convenient and useful from the point of view of statistical physics. In physics there prevail situations where information is known (e. g. etnropy of some macroscopic system) and may be measured with a high degree of accuracy, whereas probability distribution is unknown and practically cannot be measured at all (since the number of degrees of freedom of such systems is of the order 10^{23}). A more detailed discussion of this physical problem is given in a separate paper [20], which in other respects is a summary of the present one. Finally, it may be remarked that a new axiomatic definition of information, free of the inessential connection with probability, clears the way for possible future generalizations of this notion.

2. Preliminary notions. Boolean rings will be denoted by A, B, $C, \ldots$, their elements by $a, b, c, \ldots$ In this paper we shall consider non-

[20] Cf. Ingarden and Urbanik [7].

trivial Boolean rings only, i. e. rings which contain at least one element different from 0. Moreover, non-trivial subrings will be briefly called subrings.

Let $\mathscr{H}$ be a class of finite Boolean rings satisfying the following conditions:

(*) If $A \epsilon \mathscr{H}$ and B is a subring of A, then $B \epsilon \mathscr{H}$.

(**) For any $A \epsilon \mathscr{H}$ there exists a ring $B \epsilon \mathscr{H}$ such that A is a proper subring of B.

Let F be a real-valued function defined on $\mathscr{H}$. We say that two rings A and B from $\mathscr{H}$ are *F-equivalent*, in symbols $A \underset{F}{\sim} B$, if there exists an isomorphism φ of A onto B such that $F(C) = F(\varphi(C))$ for any subring C of A. Obviously, the class $\mathscr{H}$ is decomposed into disjoint sets of F-equivalent rings.

For any pair A and B of isomorphic rings from $\mathscr{H}$ we put

$$\delta_F(A, B) = \min_{\varphi} \max_{C} |F(C) - F(\varphi(C))|,$$

where C is running over all subrings of A and φ over all isomorphisms of A onto B. It is easy to verify that the function ϱ_F defined for any $A, B \epsilon \mathscr{H}$ by means of the equality

$$\varrho_F(A, B) = \begin{cases} 1 & \text{if } A \text{ and } B \text{ are non-isomorphic,} \\ \dfrac{\delta_F(A, B)}{1+\delta_F(A, B)} & \text{if } A \text{ and } B \text{ are isomorphic} \end{cases}$$

makes $\mathscr{H}$ a pseudometric space [21]. Moreover, $\varrho_F(A, B) = 0$ if and only if A and B are F-equivalent.

A ring A from $\mathscr{H}$ is said to be *F-homogeneous* if for every automorphism ψ of A and for every subring B of A we have the equality $F(B) = F(\psi(B))$.

Let $a \epsilon A$ and $a \neq 0$. By $a \frown A$ we shall denote the subring of A consisting of all elements of A which are contained in a.

LEMMA 1. *Let A be an F-homogeneous ring from $\mathscr{H}$ and $a \epsilon A$, $a \neq 0$. Then the subring $a \frown A$ is also F-homogeneous.*

Proof. Let ψ be an arbitrary automorphism of $a \frown A$. Setting $\psi_0(b) = \psi(a \frown b) \smile (b \setminus a)$ we get the extension of ψ to an automorphism ψ_0 of the whole ring A. For any subring B of $a \frown A$ we have the equality $F(B) = F(\psi_0(B)) = F(\psi(B))$, which implies the F-homogeneity of $a \frown A$.

Let $a_1, a_2, \ldots, a_n$ be a system of elements of a ring A and $A_1, A_2, \ldots$ $\ldots, A_k$ be a system of subrings of A. By $[a_1, a_2, \ldots, a_n, A_1, A_2, \ldots, A_k]$

[21] Pseudometric $\varrho(x, y)$ differs from the metric only in the absence of the assumption that $\varrho(x, y) = 0$ implies the equality $x = y$.

we shall denote the least subring of A containing all elements $a_1, a_2, \ldots, a_n$ and all subrings $A_1, A_2, \ldots, A_k$. For any $A \in \mathscr{H}$, 1_A will denote the unit element of A, i. e. an element such that $1_A \cap a = a$ for every $a \in A$. The existence of the unit element of a finite Boolean ring is evident; it is simply the union of all atoms.

LEMMA 2. *Let A be an F-homogeneous ring from $\mathscr{H}$ and let $a_1, a_2, \ldots, a_n$ be a system of disjoint elements of A such that $a_1 \cup a_2 \cup \ldots \cup a_n = 1_A$ and*

$$(14) \qquad a_j \cap A \underset{F}{\simeq} a_k \cap A \quad (j, k = 1, 2, \ldots, n).$$

Then the subring $[a_1, a_2, \ldots, a_n]$ is also F-homogeneous.

Proof. Let ψ be an arbitrary automorphism of $[a_1, a_2, \ldots, a_n]$. For any index j there exists an index k_j such that $\psi(a_j) = a_{k_j}$ $(j = 1, 2, \ldots, n)$. From the assumption (14) it follows that there exists an isomorphism φ_j of $a_j \cap A$ onto $a_{k_j} \cap A$ such that $F(C) = F(\varphi_j(C))$ for any subring C of $a_j \cap A$. Setting for any $a \in A$

$$\psi_0(a) = \varphi_1(a \cap a_1) \cup \varphi_2(a \cap a_2) \cup \ldots \cup \varphi_n(a \cap a_n)$$

we get the extension of ψ to an automorphism ψ_0 of A. Hence, for every subring B of $[a_1, a_2, \ldots, a_n]$ we have the equality $F(B) = F(\psi_0(B)) = F(\psi(B))$. Thus the subring $[a_1, a_2, \ldots, a_n]$ is F-homogeneous.

By $\mathscr{H}_F$ we shall denote the class consisting of all F-homogeneous rings from $\mathscr{H}$ and their subrings.

A real-valued function F on $\mathscr{H}$ is said to be *regular* if $\mathscr{H}_F$ is a dense subset of the pseudometric space $\mathscr{H}$, i. e. for any $A \in \mathscr{H}$ there exists a sequence $A_1, A_2, \ldots$ of rings belonging to $\mathscr{H}_F$ such that $\lim_{n\to\infty} \varrho_F(A_n, A) = 0$.

In the sequel by $N(A)$ we shall denote the number of atoms of the ring A.

3. Definition of information. Now we shall give the definition of information. *A real-valued regular function H defined on $\mathscr{H}$ is called information if it has the following properties:*

I. Connection between information of rings and their subrings. Let $A \in \mathscr{H}$ and let a, b be a pair of disjoint elements of A, $a \neq 0$, $b \neq 0$. Setting $A_1 = a \cap A$, $A_2 = b \cap A$, $A_3 = (a \cup b) \cap A$, $\tilde{A}_1 = [a, (1_A \setminus a) \cap A]$, $\tilde{A}_2 = [b, (1_A \setminus b) \cap A]$, $\tilde{A}_3 = [a \cup b, (1_A \setminus (a \cup b)) \cap A]$, we have the equality

$$(15) \quad \big(H(A) - H(\tilde{A}_3)\big)H(A_1)H(A_2) = \big(H(A) - H(\tilde{A}_1)\big)H(A_2)H(A_3) + \big(H(A) - H(\tilde{A}_2)\big)H(A_1)H(A_3).$$

II. Local character of information. Let $A, B \in \mathscr{H}$ and A_1, B_1 be a pair of subrings of A and B, respectively, such that $A_1 \underset{H}{\simeq} B_1$ and

$$[1_{A_1} \cap A, 1_A \setminus 1_{A_1}] \underset{H}{\simeq} [1_{B_1} \cap B, 1_B \setminus 1_{B_1}].$$

Then, setting $A_0 = [1_{A_1}, (1_A \setminus 1_{A_1}) \cap A]$, $B_0 = [1_{B_1}, (1_B \setminus 1_{B_1}) \cap B]$, we have the equality

$$H(A) - H(A_0) = H(B) - H(B_0).$$

III. Monotoneity. If B is a proper subring of A, then $H(B) < H(A)$.

IV. Indistinguishability. Isomorphic H-homogeneous rings are H-equivalent.

V. Normalization. If $N(A) = 2$ and A is H-homogeneous, then $H(A) = 1$.

Remark. We shall show how it is possible to define correlational information $I(A, B)$ for arbitrary rings A, B (not necessarily finite), which are subrings of a ring C, by means of the notion of information defined above. Namely, for any pair A_1, B_1 of subrings of a ring from $\mathscr{H}$ we put

$$H(A_1, B_1) = H(A_1) + H(B_1) - H([A_1, B_1]).$$

Further, for any pair A, B of subrings of an arbitrary Boolean ring C (not necessarily finite) such that every finite subring of C belongs to $\mathscr{H}$, we put

$$I(A, B) = \sup_{A_1 \subset A, B_1 \subset B} H(A_1, B_1),$$

where, as in (3), the supremum is taken over all finite subrings A_1, B_1 of A and B, respectively.

4. Connection between information and probability. We can now state the following fundamental

THEOREM. *Let H be an information on $\mathscr{H}$. Then for every $A \in \mathscr{H}$ there exists one and only one strictly positive probability measure p_A defined on A such that*

$$(16) \qquad p_B(b) = \frac{p_A(b)}{p_A(1_B)} \quad (b \in B),$$

for every subring B of A and

$$(17) \qquad H(A) = -\sum_{j=1}^{n} p_A(a_j) \log p_A(a_j),$$

werre here and henceforth the base of the logarithm is 2 and $a_1, a_2, \ldots, a_n$ are all the atoms of A.

It is easy to verify that for any family p_A ($A \in \mathscr{H}$) of strictly positive probability measures satisfying (16) the function defined by (17) satisfies conditions I-V. The H-homogeneous rings coincide with the rings having uniform probability distribution.

The proof of the Theorem will be carried out in a series of Lemmas in which all Boolean rings are supposed to belong to $\mathscr{H}$. Since information is a regular function, every ring from $\mathscr{H}$ is isomorphic to a subring of an H-homogeneous ring from $\mathscr{H}$. Consequently, according to (**), for any integer n there exists an H-homogeneous ring $A \in \mathscr{H}$ such that $N(A) \geqslant n$. Hence, by Lemma 1, there exists an H-homogeneous subring A_0 of A satisfying the equality $N(A_0) = n$.

LEMMA 3. *If* $N(C) = 1$, *then* $H(C) = 0$.

Proof. It is clear that all one-atomic rings are H-homogeneous and, consequently, by property IV of information, are H-equivalent. Let A be a two-atomic H-homogeneous ring with the atoms a and b. Let $A_1, A_2, A_3, \tilde{A}_1, \tilde{A}_2$ and $\tilde{A}_3$ have the same meaning as in condition I of information. Since $N(A_1) = N(A_2) = 1$, we have $H(A_1) = H(A_2) = H(C)$. Further, $A_3 = \tilde{A}_1 = \tilde{A}_2 = A$, which, according to (15), implies the equality $\big(H(A) - H(\tilde{A}_3)\big) H^2(C) = 0$. But $\tilde{A}_3$ is a proper subring of A and, by property III, $H(\tilde{A}_3) < H(A)$. Thus the last equality implies the assertion of the Lemma.

LEMMA 4. *Let A be an H-homogeneous ring and a_1, a_2, b_1, b_2 a system of disjoint elements of A such that $a_1 \cup a_2 \cup b_1 \cup b_2 = 1_A$, $a_1 \cap A \underset{H}{\sim} a_2 \cap A$ and $b_1 \cap A \underset{H}{\sim} b_2 \cap A$. Then*

$$H([a_1, b_1]) = H([a_1 \cup a_2, b_1 \cup b_2]).$$

Proof. Let $c_1, c_2, \ldots, c_k$ and $d_1, d_2, \ldots, d_k$ be the systems of all atoms of the rings $(a_1 \cup b_1) \cap A$ and $(a_2 \cup b_2) \cap A$, respectively. We may assume that $a_1 = c_1 \cup \ldots \cup c_s$, $a_2 = d_1 \cup \ldots \cup d_s$, $b_1 = c_{s+1} \cup \ldots \cup c_k$, $b_2 = d_{s+1} \cup \ldots \cup d_k$. By Lemma 1, the ring $(a_1 \cup b_1) \cap A$ is H-homogeneous. Further, by Lemma 1, all the rings $(c_1 \cup d_1) \cap A$, $(c_2 \cup d_2) \cup A, \ldots, (c_k \cup d_k) \cap A$ are H-homogeneous. Since they are two-atomic rings, we have, in virtue of property IV

$$(c_1 \cup d_1) \cap A \underset{H}{\sim} (c_2 \cup d_2) \cap A \underset{H}{\sim} \cdots \underset{H}{\sim} (c_k \cup d_k) \cap A.$$

Hence, by Lemma 2, the ring $[c_1 \cup d_1, c_2 \cup d_2, \ldots, c_k \cup d_k]$ is H-homogeneous. Both rings $(a_1 \cup b_1) \cap A$ and $[c_1 \cup d_1, c_2 \cup d_2, \ldots, c_k \cup d_k]$ are k-atomic. Consequently, according to IV,

$$(a_1 \cup b_1) \cap A \underset{H}{\sim} [c_1 \cup d_1, c_2 \cup d_2, \ldots, c_k \cup d_k].$$

We now define an isomorphism φ of $(a_1 \cup b_1) \cap A$ onto $[c_1 \cup d_1, c_2 \cup d_2, \ldots, c_k \cup d_k]$ by the equality $\varphi(c_j) = c_j \cup d_j$ $(j = 1, 2, \ldots, k)$. Taking

into account the H-homogeneity of both rings we get the equality $H(B) = H(\varphi(B))$ for any subring B of $(a_1 \cup b_1) \cap A$. The equality $\varphi([a_1, b_1]) = [a_1 \cup a_2, b_1 \cup b_2]$ completes the proof of the Lemma.

LEMMA 5. *For any H-homogeneous ring A and for any element $a \in A$ $(a \neq 0)$ the equality*

$$H(A) = H\big([a, (1_A \setminus a) \cap A]\big) + \frac{N(a \cap A)}{N(A)} H(a \cap A)$$

holds.

Proof. First let us suppose that $N(a \cap A) = 1$. Then, by Lemma 3, $H(a \cap A) = 0$ and the assertion of the Lemma is a direct consequence of the equality $A = [a, (1_A \setminus a) \cap A]$. Now let us assume that $N(a \cap A) \geqslant 2$. By Lemma 1, $a \cap A$ is an H-homogeneous ring and, consequently, contains a two-atomic H-homogeneous subring. Thus, according to III, IV and V, $H(a \cap A) \geqslant 1$. From H-homogeneity of A it follows that $H\big([a_1, (1_A \setminus a_1) \cap A]\big) = H\big([a_2, (1_A \setminus a_2) \cap A]\big)$ and $H(a_1 \cap A) = H(a_2 \cap A)$, whenever $N(a_1 \cap A) = N(a_2 \cap A)$. Thus the expression

$$\frac{H(A) - H\big([a, (1_A \setminus a) \cap A]\big)}{H(a \cap A)} \tag{18}$$

depends only on $N(a \cap A)$ and $N(a)$. Let us denote the expression (18) by $f\big(N(a \cap A), N(A)\big)$. To prove our Lemma it is sufficient to show that

$$f(k, n) = \frac{k}{n} \qquad (k = 2, 3, \ldots, n;\ n = 2, 3, \ldots). \tag{19}$$

Since $[1_A, (1_A \setminus 1_A) \cap A] = 1_A$, we have $N\big([1_A, (1_A \setminus 1_A) \cap A]\big) = 1$, which, in view of Lemma 3, implies $H\big([1_A, (1_A \setminus 1_A) \cap A]\big) = 0$. In other words, we have proved the equality

$$f(n, n) = 1 \qquad (n = 2, 3, \ldots). \tag{20}$$

Now let a and b be a pair of disjoint elements of A such that $N(a \cap A) \geqslant 2$ and $N(b \cap A) \geqslant 2$. Setting $A_1 = a \cap A$, $A_2 = b \cap A$, $A_3 = (a \cup b) \cap A$, $\tilde{A}_1 = [a, (1_A \setminus a) \cap A]$, $\tilde{A}_2 = [b, (1_A \setminus b) \cap A]$, $\tilde{A}_3 = \big[a \cup b, (1_A \setminus (a \cup b)) \cap A\big]$, we have the equalities

$$f\big(N(a \cap A), N(A)\big) = \frac{H(A) - H(\tilde{A}_1)}{H(A_1)},$$

$$f\big(N(b \cap A), N(A)\big) = \frac{H(A) - H(\tilde{A}_2)}{H(A_2)},$$

$$f\big(N((a \cup b) \cap A), N(A)\big) = \frac{H(A) - H(\tilde{A}_3)}{H(A_3)}.$$

Thus, dividing both sides of equality (15) by $H(A_1)H(A_2)H(A_3)$, we obtain the formula

$$f\big(N((a\cup b)\cap A), N(A)\big) = f\big(N(a\cap A, N(A))\big)+f\big(N(n\cap A), N(A)\big).$$

Finally, taking into account the equation $N((a\cup b)\cap A) = N(a\cap A)+ +N(b\cap A)$, we get the equality

(21) $f(k_1+k_2, n) = f(k_1, n)+f(k_2, n),$ whenever $k_1 \geqslant 2,\ k_2 \geqslant 2$ and $k_1+k_2 \leqslant n$.

Let $n \geqslant 6$. Setting $f(2, n) = 2p$ we have, in view of (21), $f(6, n) = 2f(3, n)$ and $f(6, n) = 3f(2, n) = 6p$, which implies $f(3, n) = 3p$. Every integer $k \geqslant 2$ can be expressed in the form $k = 2r+3s$, where $r = 0, 1, 2, \ldots$ and $s = 0, 1$. By (21), we have the equality

$$f(k, n) = rf(2, n)+sf(3, n) = (2r+3s)p = kp \qquad (k = 2, 3, \ldots, n).$$

Hence and from (20) we get $np = 1$ and, consequently,

$$f(k, n) = \frac{k}{n} \qquad (k = 2, 3, \ldots, n;\ n = 6, 7, \ldots). \tag{22}$$

Thus, for $n \geqslant 6$ equality (19) is proved.

Before proving the last equality for $n < 6$ we shall prove some auxiliary formulae.

Let B be an H-homogeneous ring and a_1, a_2, b_1, b_2 be a system of disjoint elements of B such that $N(B) = 2n$, $N(a_1\cap B) = N(a_2\cap B) = n-1$, $N(b_1\cap B) = N(b_2\cap B) = 1$ and $n \geqslant 3$. From (22) it follows that

$$f\big(N((a_1\cup b_1)\cap B), N(B)\big) = f(n, 2n) = \tfrac{1}{2}.$$

Hence, by the definition of the function f, we get the equality

$$H(B) = H\big([a_1\cup b_1, (a_2\cup b_2)\cap B]\big)+\tfrac{1}{2}H\big((a_1\cup b_1)\cap B\big). \tag{23}$$

The ring $C = (a_2\cup b_2)\cap B$ is a subring of both rings B and $D = [a_1\cup b_1, (a_2\cup b_2)\cap B]$. Since $[1_C\cap B, 1_B\setminus 1_C] = D = [1_C\cap D, 1_D\setminus 1_C]$, we have, in view of property II of information,

$$H(B)-H\big([(a_1\cup b_1)\cap B, a_2\cup b_2]\big) = H(D)-H([a_1\cup b_1, a_2\cup b_2]). \tag{24}$$

From the equality $N((a_1\cup b_1)\cap B) = N((a_2\cup b_2)\cap B)$ and from H-homogeneity of B we obtain the relation $[(a_1\cup b_1)\cap B, a_2\cup b_2] \underset{H}{\simeq} [(a_2\cup b_2)\cap B, a_1\cup b_1]$. Thus

$$H(D) = H\big([a_1\cup b_1, (a_2\cup b_2)\cap B]\big) = H\big([(a_1\cup b_1)\cap B, a_2\cup b_2]\big). \tag{25}$$

Moreover, the rings $(a_1\cup b_1)\cap B$ and $(a_2\cup b_2)\cap B$ are H-equivalent. Using Lemma 2, we infer that the two-atomic $[a_1\cup b_1, a_2\cup b_2]$ is H-homogeneous and, by property V, $H([a_1\cup b_1, a_2\cup b_2]) = 1$. Hence and from (23), (24) and (25) we obtain the formula

$$H(B) = H\big((a_1\cup b_1)\cap B\big)+1. \tag{26}$$

In the same way we may prove the equality

$$H\big((a_1\cup a_2)\cap B\big) = H(a_1\cap B)+1. \tag{27}$$

Further, from (22) we get the equalities

$$f\big(N((a_1\cup a_2)\cap B), N(B)\big) = f(2(n-1), 2n) = \frac{n-1}{n},$$

$$f\big(N(b_1\cup b_2)\cap B), N(B)\big) = f(2, 2n) = \frac{1}{n},$$

which, by the definition of f, imply

$$H(B) = H\big([a_1\cup a_2, (b_1\cup b_2)\cap B]\big)+\frac{n-1}{n}H\big((a_1\cup a_2)\cap B\big)$$

and

$$H(B) = H\big([b_1\cup b_2, (a_1\cup a_2)\cap B]\big)+\frac{1}{n}H\big((b_1\cup b_2)\cap B\big).$$

Since, by Lemma 1, the two-atomic ring $(b_1\cup b_2)\cap B$ is H-homogeneous, we have, in view of V, $H\big((b_1\cup b_2)\cap B\big) = 1$. Thus, according to (27), the last equalities may be written in the form

$$H(B) = H\big([a_1\cup a_2, (b_1\cup b_2)\cap B]\big)+\frac{n-1}{n}H(a_1\cap B)+\frac{n-1}{n}; \tag{28}$$

$$H(B) = H\big([b_1\cup b_2, (a_1\cup a_2)\cap B+]\big)\frac{1}{n}. \tag{29}$$

Consider the rings B and $E = [a_1\cup a_2, (b_1\cup b_2)\cap B]$. They both contain the ring $B_1 = (b_1\cup b_2)\cap B$. It is very easy to verify that

$$[1_{B_1}\cap B, 1_B\setminus 1_{B_1}] = E = [1_{B_1}\cap E, 1_E\setminus 1_{B_1}].$$

Taking into account the equalities

$$[1_{B_1}, (1_B\setminus 1_{B_1})\cap B] = [b_1\cup b_2, (a_1\cup a_2)\cap B],$$
$$[1_{B_1}, (1_E\setminus 1_{B_1})\cap E] = [a_1\cup a_2, b_1\cup b_2],$$

we obtain, in view of property II,

$$H(B)-H\big([a_1\cup a_2, (b_1\cup b_2)\cap B]\big) = H\big([b_1\cup b_2, (a_1\cup a_2)\cap B]\big)-H([a_1\cup a_2, b_1\cup b_2]).$$

Hence and from (26), (28) and (29) we get the formula

$$(30) \qquad H([a_1\cup a_2, b_1\cup b_2]) = H\big((a_1\cup b_1)\cap B\big) - \frac{n-1}{n}H(a_1\cap B).$$

By Lemma 4 we have

$$(31) \qquad H([a_1, b_1]) = H([a_1\cup a_2, b_1\cup b_2]).$$

Finally, setting $A = (a_1\cup b_1)\cap B$, we have $N(a_1\cap A) = n-1$, $N(A) = n$, $[a_1, (1_A\setminus a_1)\cap A] = [a_1, b_1]$ and, according to (30) and (31) $f(N(a_1\cap A), N(A)) = \frac{n-1}{n}$. Thus we have proved the equality

$$(32) \qquad f(n-1, n) = \frac{n-1}{n} \quad \text{for} \quad n\geqslant 3.$$

Now we shall prove equality (19) for $n = 2, 3, 4$, and 5. From (20) and (32) we get $f(2,2) = 1$, $f(2,3) = \frac{2}{3}$, $f(3,3) = 1$, $f(3,4) = \frac{3}{4}$, $f(4,4) = 1$, $f(4,5) = \frac{4}{5}$ and $f(5,5) = 1$ by equality (21), $2f(2,4) = f(4,4)$, $2f(2,5) = f(4,5)$, $f(2,5)+f(3,5) = f(5,5)$. Hence $f(2,4) = \frac{1}{2}$, $f(2,5) = \frac{2}{5}$ and $f(3,5) = \frac{3}{5}$, which completes the proof of the Lemma.

LEMMA 7. *Let A be an H-homogeneous ring and let $b_1, b_2, \ldots, b_n$ be a system of disjoint elements of A such that $b_k\neq 0$ $(k = 1, 2, \ldots, n)$ and $b_1\cup b_2\cup\ldots\cup b_n = 1_A$. Then*

$$(33) \qquad H([b_1, b_2, \ldots, b_n]) = \sum_{k=1}^{n}\frac{N(b_k\cap A)}{N(A)}\big(H(A)-H(b_k\cap A)\big).$$

Proof. Let us introduce the notation

$$B_0 = A, B_k = \big[b_1, b_2, \ldots, b_k, \big(1_A\setminus(b_1\cup b_2\cup\ldots\cup b_k)\big)\cap A\big] \quad (k = 1, 2, \ldots, n),$$

$$A_k = [b_k, (1_A\setminus b_k)\cap A] \quad (k = 1, 2, \ldots, n).$$

By Lemma 5, we have the equality

$$(34) \qquad H(A) = H(A_k) + \frac{N(b_k\cap A)}{N(A)}H(b_k\cap A) \quad (k = 1, 2, \ldots, n).$$

Let us consider the rings A and B_{k-1} $(k = 1, 2, \ldots, n)$. The ring $C_k = b_k\cap A$ is a subring of both A and B_{k-1}. It is easy to see that

$$[1_{C_k}\cap A, 1_A\setminus 1_{C_k}] = [C_k, 1_A\setminus 1_{C_k}],$$

$$[1_{C_k}\cap B_{k-1}, 1_{B_{k-1}}\setminus 1_{C_k}] = [C_k, 1_A\setminus 1_{C_k}],$$

$$[1_{C_k}, (1_A\setminus 1_{C_k})\cap A] = A_k, \qquad [1_{C_k}, (1_{B_{k-1}}\setminus 1_{C_k})\cap B_{k-1}] = B_k.$$

Thus, by property II of information

$$H(A)-H(A_k) = H(B_{k-1})-H(B_k) \quad (k = 1, 2, \ldots, n).$$

Hence and from (34) we get the equalities

$$H(B_{k-1})-H(B_k) = \frac{N(b_k\cap A)}{N(A)}H(b_k\cap A) \quad (k = 1, 2, \ldots, n),$$

which imply

$$H(A)-H([b_1, b_2, \ldots, b_n]) = H(B_0)-H(B_n) = \sum_{k=1}^{n}\frac{N(b_k\cap A)}{N(A)}H(b_k\cap A).$$

Taking into account the equality

$$\sum_{k=1}^{n}\frac{N(b_k\cap A)}{N(A)} = 1,$$

we obtain formula (33).

LEMMA 8. *Let A be a ring from $\mathscr{H}$, $N(A)\geqslant 3$ and let p_1 and p_2 be two strictly positive probability measures on A. If for any triplet a_1, a_2, a_3 of disjoint elements of A satisfying the condition $a_1\cup a_2\cup a_3 = 1_A$ we have the equality*

$$p_1(a_1)\log p_1(a_1)+p_1(a_2)\log p_1(a_2)+p_1(a_3)\log p_1(a_3)$$
$$= p_2(a_1)\log p_2(a_1)+p_2(a_2)\log p_2(a_2)+p_2(a_3)\log p_2(a_3), \text{ [22]}$$

then $p_1 = p_2$.

Proof. Contrary to this statement let us suppose that there exists an atom $a\in A$ such that

$$(35) \qquad p_1(a)\neq p_2(a),$$

Consider the triplet $a_1 = a$, $a_2 = 1_A\setminus a$, $a_3 = 0$. Then

$$p_1(a)\log p_1(a)+\big(1-p_1(a)\big)\log\big(1-p_1(a)\big) = p_2(a)\log p_2(a)+\big(1-p_2(a)\big)\log\big(1-p_2(a)\big).$$

Since the function $x\log x+(1-x)\log(1-x)$ $(0\leqslant x\leqslant 1)$ is convex and symmetric with respect to $x = \frac{1}{2}$, the last equality and (35) imply

$$(36) \qquad p_1(a) = 1-p_2(a).$$

[22] We define the indeterminate form $0\cdot\log 0$ as having the value zero.

Since $N(A) \geqslant 3$, there exist two disjoint elements b and c such that $b \neq 0$, $c \neq 0$ and $b \cup c = 1_A \setminus a$. First let us assume that $p_1(b) = p_2(b)$ and $p_1(c) = p_2(c)$. Then we have the equality

$$p_1(a) = 1 - p_1(b) - p_1(c) = 1 - p_2(b) - p_2(c) = p_2(a),$$

which contradicts inequality (35). Consequently, $p_1(b) \neq p_2(b)$ or $p_1(c) \neq p_2(c)$. By symmetry, we may suppose that $p_1(b) \neq p_2(b)$. Now identical reasoning to that which led to equality (36) gives $p_1(b) = 1 - p_2(b)$. Hence and from (36), according to the strict positivity of p_2, we get

$$p_1(c) = 1 - p_1(a) - p_1(b) = p_2(a) + p_2(b) - 1 = -p_2(c) < 0,$$

which is impossible. The Lemma is thus proved.

Proof of the theorem. We know that the class $\mathscr{H}$ contains a sequence $C_1, C_2, \ldots$ of H-homogeneous rings such that $N(C_n) = n$ $(n = 1, 2, \ldots)$. Put $L(n) = H(C_n)$ $(n = 1, 2, \ldots)$. Let a be an atom of C_{n+1}. Then, by Lemma 1, the ring $(1_{C_{n+1}} \setminus a) \cap C_{n+1}$ is H-homogeneous and, of course, n-atomic. Thus, by property IV of information, $(1_{C_{n+1}} \setminus a) \cap C_{n+1} \underset{H}{\simeq} C_n$, which implies

$$H\big((1_{C_{n+1}} \setminus a) \cap C_{n+1}\big) = L(n).$$

Since $(1_{C_{n+1}} \setminus a) \cap C_{n+1}$ is a proper subring of C_{n+1}, we have, in virtue of III,

$$L(n) < L(n+1) \quad (n = 1, 2, \ldots). \tag{37}$$

Now let $a_1, a_2, \ldots, a_m$ be a system of disjoint elements of $C_{n\cdot m}$ such that $a_1 \cup a_2 \cup \ldots \cup a_m = 1_{C_{n\cdot m}}$ and

$$N(a_j \cap C_{n\cdot m}) = n \quad (j = 1, 2, \ldots, m). \tag{38}$$

Obviously, $a_j \cap C_{n\cdot m} \underset{H}{\simeq} a_k \cap C_{n\cdot m}$ $(j, k = 1, 2, \ldots, m)$, and, by Lemma 2, the ring $[a_1, a_2, \ldots, a_m]$ is H-homogeneous and, by (38), n-atomic. Thus

$$H(a_j \cap C_{n\cdot m}) = L(n) \quad (j = 1, 2, \ldots, m). \tag{40}$$

Using Lemma 7, we obtain the relation

$$H([a_1, a_2, \ldots, a_m]) = \sum_{k=1}^{m} \frac{N(a_k \cap C_{n\cdot m})}{N(C_{n\cdot m})} \big(H(C_{n\cdot m}) - H(a_k \cap C_{n\cdot m})\big).$$

Hence, in view of (38), (39) and (40), we have the equation

$$L(m) + L(n) = L(n\cdot m) \quad (n, m = 1, 2, \ldots).$$

It is well known ([23]) that every solution of this equation satisfying condition (37) is of the form $L(n) = \mu \log n$, where μ is a positive constant. By property V, $1 = L(2) = \mu$. Thus

$$L(n) = \log n. \tag{41}$$

Now let A be an H-homogeneous ring and B_0 its subring. Assume that $1_B = 1_A$ and denote by $b_1, b_2, \ldots, b_n$ all the atoms of B_0. Obviously, $B_0 = [b_1, b_2, \ldots, b_n]$ and, by Lemma 1, all the rings $b_j \cap A$ $(j = 1, 2, \ldots, n)$ are H-homogeneous. Thus, by (41), we have

$$H(A) = \log N(A), \quad H(b_j \cap A) = \log N(b_j \cap A) \quad (j = 1, 2, \ldots, n).$$

Further, applying Lemma 7, we get

$$H(B_0) = \sum_{k=1}^{n} \frac{N(b_k \cap A)}{N(A)} \log \frac{N(b_k \cap A)}{N(A)}. \tag{42}$$

We define a probability measure p_{B_0} on B_0 by means of the formula

$$p_{B_0} = \frac{N(a \cap A)}{N(A)} \quad (a \in B_0). \tag{43}$$

It is easy to verify that if B_0 is a subring of C $(C \in \mathscr{H}_H)$, then

$$p_{B_0}(a) = \frac{p_C(a)}{p_C(1_{B_0})} \quad (a \in B_0).$$

Moreover, by (42),

$$H(B_0) = -\sum_{k=1}^{n} p_B(b_k) \log p_B(b_k), \tag{44}$$

where $b_1, b_2, \ldots, b_n$ are atoms of B_0. Thus we have defined a probability measure p_{B_0} for any ring B_0 from $\mathscr{H}_H$ in such a way that equalities (16) and (17) hold.

Now let A be an arbitrary ring from $\mathscr{H}$ and $N(A) \geqslant 3$. By regularity of information, there exists a sequence $A_1, A_2, \ldots$ of rings belonging

([23]) Cf., e. g., [8], p. 9, 10.

to $\mathscr{H}_H$ such that $\lim\limits_{v\to\infty} \varrho_H(A_r, A) = 0$. We may assume that $\varrho_H(A_r, A) < 1$, i. e. the rings A and A_r $(r = 1, 2, \ldots)$ are isomorphic. By the definition of ϱ_H for every integer r there exists an isomorphism φ_r of A_r onto A such that for any subring B of A

$$\lim_{r\to\infty} H\big(\varphi_r^{-1}(B)\big) = H(B). \tag{45}$$

Let $a_1, a_2, \ldots, a_n$ be the system of all atoms of A. We shall prove that for any j $(j = 1, 2, \ldots, n)$ the sequence

$$p_{A_r}\big(\varphi_r^{-1}(a_j)\big) \qquad (r = 1, 2, \ldots) \tag{46}$$

is convergent. Since the sequence (46) is bounded, each of its subsequences contains a convergent subsequence. Denoting the limit of such a subsequence by $p(a_j)$, we have $\sum\limits_{j=1}^{k} p(a_j) = 1$ and $p(a_j) \geqslant 0$ $(j = 1, 2, \ldots, n)$. Moreover, setting $p(a_{i_1} \cup a_{i_2} \cup \ldots \cup a_{i_k}) = \sum\limits_{s=1}^{k} p(a_{i_s})$ $(a_{i_s} \cap a_{i_m} = 0$ if $s \neq m)$, from (43), (44) and (45) we get the equality

$$H([a, b, c]) = -p(a)\log p(a) - p(b)\log p(b) - p(c)\log p(c),$$

where a, b, c is an arbitrary triplet of disjoint elements of A and $a \cup b \cup c = 1_A$. Hence, by Lemma 8, $p(a_j)$ is the limit of any convergent subsequence of (46). Thus, the sequence (46) itself is convergent to $p(a_j)$. Setting for any element $a = a_{i_1} \cup a_{i_2} \cup \ldots \cup a_{i_k}$ $(a_{i_s} \cap a_{i_m} = 0$ for $s \neq m)$

$$p_A(a) = \sum_{s=1}^{k} p(a_{i_s}),$$

we get a probability measure on A such that, according to (44) and (45),

$$H(A) = -\sum_{j=1}^{n} p_A(a_j)\log p_A(a_j).$$

Moreover, for any subring B of A the formula

$$p_B(b) = \frac{p_A(b)}{p_A(1_B)} \qquad (b \in B)$$

determines a probability measure on B such that, according to (43), (44) and (45),

$$H(B) = -\sum_{j=1}^{m} p_B(b_j)\log p_B(b_j),$$

where $b_1, b_2, \ldots, b_m$ are atoms of B. Hence and from property III, it follows that all the measures p_A are strictly positive. Thus we have defined probability measures for any ring A satisfying the inequality $N(A) \geqslant 3$ and for any its subring. These probability measures satisfy conditions (16) and (17) and, consequently, by (16), every probability measure on a subring B of A is uniquely determined by the probability measure on A. By (**), the class of all rings A from $\mathscr{H}$ satisfying the inequality $N(A) \geqslant 3$ and all their subrings coincides with the whole class $\mathscr{H}$. Thus to prove the uniqueness of p_A $(A \in \mathscr{H})$ it is sufficient to prove this for rings satisfying the condition $N(A) \geqslant 3$. But the last statement is a direct consequence of Lemma 8. In fact, for any triplet a_1, a_2, a_3 of disjoint elements of such a ring A satisfying the equality $a_1 \cup a_2 \cup a_3 = 1_A$ we have

$$H([a_1, a_2, a_3]) = -p_A(a_1)\log p_A(a_1) - p_A(a_2)\log p_A(a_2) - p_A(a_3)\log p_A(a_3).$$

On the other hand, for every probability measure $\tilde{p}_A$ on A satisfying (16) and (17) we have the same equality:

$$H([a_1, a_2, a_3]) = -\tilde{p}_A(a_1)\log \tilde{p}_A(a_1) - \tilde{p}_A(a_2)\log \tilde{p}_A(a_2) - \tilde{p}_A(a_3)\log \tilde{p}_A(a_3).$$

Consequently, in view of Lemma 8, $p_A = \tilde{p}_A$. The Theorem is thus proved.

Added in proof. The term "informational coefficient of correlation" (see p. 133) was also proposed by E. H. Linfoot in his paper *An informational measure of correlation*, Information and Control 1 (1957), p. 85-89.

REFERENCES

[1] L. Brillouin, *Science and Information Theory*, Academic Press, New York 1956.

[2] Р. Л. Добрушин, *Общая формулировка основной теоремы Шеннона в теории информации*, Успехи Математических Наук 14 (1959), No. 6, p. 3-104.

[3] Д. К. Фаддеев, *К понятию энтропии конечной вероятностной схемы*, ibidem 11 (1956), No. 1, p. 227-231; German translation: *Zum Begriff der Entropie eines endlichen Wahrscheinlichkeitsschemas*, Arbeiten zur Informationstheorie I, Deutscher Verlag der Wissenschaften, Berlin 1957, p. 85-90.

[4] A. Feinstein, *Foundations of Information Theory*, New York 1958.

[5] М. М. Гельфанд, А. Н. Колмогоров и А. М. Яглом, *К общему определению количества информации*, Доклады Академии Наук СССР 111 (1956), p. 745-752.

[6] R. S. Ingarden, *The Problem of Optimum Optical Systems*, Optica Acta 7 (1960), p. 179-183.

[7] R. S. Ingarden and K. Urbanik, *Information as a Fundamental Notion of Statistical Physics*, Bulletin de l'Académie Polonaise des Sciences, Série des sciences math. astr. et phys. 9 (1961), No. 4, p. 313-316.

[8] А. Я. Хинчин, *Понятие энтропии в теории вероятностей*, Успехи Математических Наук 8 (1955), No. 3, p. 3-20; English translation in: A. I. Khintchine, *Mathematical Foundations of Information Theory*, New York 1957, p. 1-28.

[9] Л. Ландау и Е. Лифшиц, *Статистическая физика*, Гостехиздат, Москва 1951. English translation: *Statistical Physics*, London 1958.

[10] A. Rényi, *On measures of entropy and information*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. I, to appear.

[11] C. E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal 27 (1948), p. 379-423 and 623-656.

[12] R. Sikorski, *Boolean Algebras*, Berlin 1960 (Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, H. 25).

[13] R. C. Tolman, *Principles of Statistical Mechanics*, Oxford, 1938.

[14] I. Vincze, *An Interpretation of the I-Divergence of Information Theory*, Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague 1960, p. 545-556.

INSTITUTE OF PHYSICS, POLISH ACADEMY OF SCIENCES
MATHEMATICAL INSTITUTE OF THE POLISH ACADEMY OF SCIENCES

Reçu par la Rédaction le 16. 1. 1961

COLLOQUIUM MATHEMATICUM

VOL. IX 1962 FASC. 1

ON PAIRS OF INDEPENDENT RANDOM VARIABLES WHOSE QUOTIENTS FOLLOW SOME KNOWN DISTRIBUTION

BY

I. KOTLARSKI (WARSAW)

1. Introduction. Some important probability distributions encountered in mathematical statistics are defined as probability distributions of a quotient of two independent random variables. So are the distributions of Student as well as those of Fisher's variance ratios. A question suggests itself if the distributions of nominators and denominators of the quotients in question are determined uniquely, up to a multiplication by a constant factor or to a passage to reciprocals of the random variables involved by the distribution of the quotient. The simplest problem is connected with the Cauchy distribution, the probability density of which is

$$\frac{1}{\pi}\frac{1}{1+x^2} \quad (-\infty < x < \infty).$$

It turns out to be the distribution of a quotient of two independent random variables having the same normal distribution symmetrical about zero, or, in other words, to be Student's distribution with one degree of freedom. This problem has been studied by Mauldon [7], Laha [4, 5, 6], Steck [8], Kotlarski [3] to the effect that there is no such uniqueness — there are many non-normal distributions symmetrical about zero such that a quotient of two independent random variables having such distribution has Cauchy distribution. Another case — where nominators and denominators have gamma distributions — has been considered by Mauldon [7]. He has shown that also in this case there is no uniqueness of the above mentioned kind and thus has shown the ambiguity phenomenon for Fisher's, or, as some say, Snedecor's F distributions.

In this paper we are considering the more general case of quotients $U = X_1^{q_1}:X_2^{q_2}$, where X_1 and X_2 are independent random variables having gamma distributions and q_1, q_2 are real numbers not equal to 0.