Vol. 2 (1971) *REPORTS ON MATHEMATICAL PHYSICS* No. 3

WEIGHTED ENTROPY

Silviu Guiașu

Mathematical Institute of Academy; Mathematical Faculty of University,
Bucharest, Rumania

(*Received January 15, 1971*)

Weighted entropy is the measure of information supplied by a probablistic experiment whose elementary events are characterized both by their objective probabilities and by some qualitative (objective or subjective) weights. The properties, the axiomatics and the maximum value of the weighted entropy are given.

§ 1. Introduction

The notion of information entropy plays a role of prime importance in the statistical physics. Many problems were clarified by means of information entropy which is a measure both of the uncertainty and of the information supplied by a probabilistic experiment.

Having its origin in the famous Boltzmann H-function, the Shannon entropy rapidly became a useful tool in several domains, especially in communication theory (Shannon [19], Feinstein [5]), statistical physics (Ingarden [8], [10], Jaynes [13]), nuclear reaction theory (Bloch [2]), learning theory (Watanabe [20]), mathematical statistics (Kullback [14]), and measurement theory (Majernik [15]), without mentioning other domains as linguistics, music or social sciences.

Excepting the applications of classical information theory in all these fields, several generalizations of the Shannon entropy were proposed and, having no pretension to give a complete list of these, let us notice here the Kolmogorov ε-entropy, the Rényi α-entropy, the Kullback relative information, the Perez–Csiszar f-entropy, the Weiss objective-subjective entropy or the general axiomatics of information without probability given by Ingarden and Urbanik ([12], [9]), which puts the notion of information before that of probability from the point of view of importance and generality.

Underlining the importance of Shannon's entropy, it is necessary to notice at the same time that this formula gives us the measure of information as function of the probabilities with which various events occur, only. But there exsist many fields dealing with random events where it is necessary to take into account both these probabilities and some qualitative characteristics of events. For instance, in a two-handed game one should keep in mind both the probabilities of different variants of the game (i.e. the random strategies of the players) and the wins corresponding to these variants.

Often, in a physical experiment it is very difficult to neglect the subjective aspects related to the various goals of the experimenter. At the same time, the possible states of a physical system may differ considerably from the view-point of a given qualitative characteristic. In statistical physics, all elementary events usually have the same importance, i.e. they are physically equivalent, but this is not the general situation. In order to describe the latter, it is necessary to associate with every elementary event both the probability with which it occurs and its qualitative weight. A criterion for a qualitative differentiation of the possible events of a physical experiment is represented by the relevance, the significance, or the utility of the information they carry with respect to a goal with respect to a qualitative characteristic. The occurrence of an event removes a double uncertainty: the quantitative one, related to the probability with which it occurs, and the qualitative one, related to its utility for the attainment of the goal or to its significance with respect to a given qualitative characteristic.

Of course, the qualitative weight of an event may be independent of the objective probability with which it occurs, for instance, an event of small probability can have a great weight, likewise, an event of great probability can have a very small weight. Naturally, the ascription of a weight to every elementary event is not a thing just so easy to be done. These weights may be either of objective or subjective character. Thus, the weight of one event may express some qualitative objective characteristic, but it also may express the subjective utility of the respective event with respect to the experimenter's goal. The weight ascribed to an elementary event may be also related to the subjective probability with which respective events occur, and this does not always coincide with the objective probability.

We shall suppose that these qualitative weights are non-negative, finite, real numbers as the usual weights in physics or as the utilities in decision theory. Also, if one event is more relevant, more significant, more useful (with respect to a goal or from a given qualitative point of view) than another, the weight of the first event will be greater than that of the second one.

How to evaluate the amount of information supplied by a probability space, i.e. by a probabilistic experiment, whose elementary events are characterized both by their probabilities and by some qualitative (objective or subjective) weights? In particular, what is the amount of information supplied by a probabilistic experiment when the probabilities calculated by the experimenter (i.e. the subjective probabilities) do not coincide with the objective probabilities of these random events?

In the present paper, we shall give a formula for the entropy as a measure of uncertainty or information supplied by a probabilistic experiment depending both on the probabilities of events and on qualitative (objective or subjective) weights of the possible events. This entropy will be called the *weighted entropy*. In the following paragraphs the properties, the axiomatic treatment and, finally, the extremal property of the weighted entropy will be given.

§ 2. Definition and properties of the weighted entropy

Consider a probablistic physical experiment whose corresponding probability space has a finite number of elementary events $\omega_1, \ldots, \omega_n$ with the objective probabilities of these events given respectively by the numbers

$$p_k \geqslant 0 \quad (k=1, \ldots, n); \qquad \sum_{k=1}^{n} p_k = 1.$$

The different elementary events ω_k depend more or less relevantly upon the experimenter's goal or upon some qualitative characteristic of the physical system taken into consideration; that is, they have different (objective or subjective) weights. The weight of an event may be either independent or dependent of its objective probability.

In order to distinguish the events $\omega_1, \ldots, \omega_n$ of a goal-directed experiment according to their importance with respect to the experimenter's goal or according to their significance with respect to a given qualitative characteristic of the physical system taken into consideration, we shall ascribe to each event ω_k a non-negative number $w_k \geqslant 0$ directly proportional to its importance or significance mentioned above. We shall call w_k the *weight* of the elementary event ω_k.

We define the *weighted entropy* by the expression

$$I_n = I_n(w_1, \ldots, w_n; p_1, \ldots, p_n) = -\sum_{k=1}^{n} w_k p_k \log p_k. \tag{1}$$

Let us notice briefly some obvious properties of the weighted entropy. The proofs of the first six properties are immediate.

PROPERTY 1.

$$I_n(w_1, \ldots, w_n; \; p_1, \ldots, p_n) \geqslant 0.$$

PROPERTY 2. *If* $w_1 = \ldots = w_n = w$, *then*

$$I_n(w_1, \ldots, w_n; \; p_1, \ldots, p_n) = -w \sum_{k=1}^{n} p_k \log p_k = H_n(p_1, \ldots, p_n),$$

where H_n *is the classical Shannon entropy* (which is determined uniquely up to an arbitrary multiplicative constant).

PROPERTY 3. *If* $p_{k_0} = 1$, $p_k = 0$ $(k=1, \ldots, n; k \neq k_0)$, *then*

$$I_n(w_1, \ldots, w_n; p_1, \ldots, p_n) = 0$$

whatever are the weights $w_1, \ldots, w_n$.

The last property illustrates the obvious fact that an experiment for which only one event is possible does not supply any information. In this case, the Shannon entropy H_n is also equal to zero. Therefore, we are really interested only in the probabilistic experiment having at least two possible events.

PROPERTY 4. *If* $p_i=0$, $w_i \neq 0$ *for every* $i \in I$ *and* $p_j \neq 0$, $w_j=0$ *for every* $j \in J$, *where*

$$I \cup J = \{1, 2, \dots, n\}, \quad I \cap J = \emptyset,$$

then

$$I_n(w_1, \dots, w_n;\ p_1, \dots, p_n) = 0.$$

This property illustrates the intuitive fact that an experiment whose possible events are useless or non-significant, and whose useful or significant events are impossible, supplies a total information equal to zero even if the corresponding Shannon entropy $H_n(p_1 \dots, p_n)$ is different from zero, provided the set J has at least two elements. In particular, when all events have zero weights we get the total information $I_n=0$ even if the Shannon entropy H_n is not null, i.e. if there exist $0<p_k<1$.

PROPERTY 5.

$$I_{n+1}(w_1, \dots, w_n, w_{n+1};\ p_1, \dots, p_n, 0) = I_n(w_1, \dots, w_n;\ p_1, \dots, p_n),$$

whatever are the weights $w_1, \dots, w_n, w_{n+1}$ *and the complete system of probabilities* $p_1, \dots, p_n$.

PROPERTY 6. *For every non-negative, real number* λ *we have*

$$I_n(\lambda w_1, \dots, \lambda w_n;\ p, \dots, p_n) = \lambda I_n(w_1, \dots, w_n;\ p_1, \dots, p_n).$$

Till now we did not impose any restriction on the weights ascribed to the elementary events of the physical experiment (except that they are non-negative real numbers). Let us suppose that the weight of the union of two incompatible events is the mean value of the weights of the respective events, i.e.

$$w(E \cup F) = \frac{p(E)\, w(E) + p(F)\, w(F)}{p(E) + p(F)} \tag{2}$$

for any incompatible events E, F, where $w(E)$ is the weight of the event E and $p(E)$ is the probability of the same event E; in particular, if E and F are complementary events, then

$$w(E \cup F) = p(E)\, w(E) + (1 - p(E))\, w(F).$$

PROPERTY 7. *If the rule* (2) *for the weights holds, then*

$$I_{n+1}(w_1, \dots, w_{n+1}, w', w'';\ p_1, \dots, p_{n-1}, p', p'')$$
$$= I_n(w_1, \dots, w_n;\ p_1, \dots, p_n) + p_n I_2\left(w', w'';\ \frac{p'}{p_n}, \frac{p''}{p_n}\right),$$

where

$$w_n = \frac{p'w' + p''w''}{p' + p''}, \quad p_n = p' + p''.$$

Proof: Taking into account the definition of the weighted entropy and writing

$$w_n = \frac{p'w' + p''w''}{p' + p''}, \quad p_n = p' + p'',$$

we have

$$
\begin{aligned}
I_{n+1}&(w_1, \dots, w_{n-1}, w', w''; p_1, \dots, p_{n-1}, p', p'') \\
&= -\sum_{k=1}^{n-1} w_k p_k \log p_k - w'p' \log p' - w''p'' \log p'' \\
&= -\sum_{k=1}^{n-1} w_k p_k \log p_k - w_n p_n \log p_n + w_n p_n \log p_n - w'p' \log p' - w''p'' \log p'' \\
&= I_n(w_1, \dots, w_n; p_1, \dots, p_n) + (w'p' + w''p'') \log p_n - w'p' \log p' - w''p'' \log p'' \\
&= I_n(w_1, \dots, w_n; p_1, \dots, p_n) + p_n \left(-w' \frac{p'}{p_n} \log \frac{p'}{p_n} - w'' \frac{p''}{p_n} \log \frac{p''}{p_n} \right) \\
&= I_n(w_1, \dots, w_n; p_1, \dots, p_n) + p_n I_2 \left(w', w''; \frac{p'}{p_n}, \frac{p''}{p_n} \right).
\end{aligned}
$$

Q.E.D.

Here we intend to give two examples:

a) Let us consider the weighted entropy (1) and put

$$w_k = -\frac{p_k}{\log p_k} \quad (k = 1, \dots, n). \tag{3}$$

In this case the weight of every elementary event has an *objective character* representing the ratio of the objective probability of this event to the amount of information it supplies. In this case we obtain the following expression for the weighted entropy

$$I_n = \sum_{k=1}^{n} p_k^2,$$

i.e. the Onicescu information energy (see [16]) introduced in the information theory by an analogy to the kinetic energy from mechanics.

b) Consider a probabilistic experiment whose elementary events have the objective probabilities $q_1, \dots, q_n$. Denote by $p_1, \dots, p_n$ the subjective probabilities of the same events established by an experimenter. If we ascribe to every elementary event the *subjective* weight

$$w_k = \frac{q_k}{p_k} \quad (k = 1, \dots, n), \tag{4}$$

representing the ratio of the objective probability to the subjective probability of the event w_k, then the weighted entropy assumes the form

$$I_n = -\sum_{k=1}^{n} q_k \log p_k. \tag{5}$$

If we put

$$x_k = \frac{p_k}{q_k} \qquad (k=1, \ldots, n),$$

in Jensen's inequality

$$\sum_{k=1}^{n} q_k \log x_k \leqslant \log\left(\sum_{k=1}^{n} q_k x_k\right),$$

we obtain

$$I_n = -\sum_{k=1}^{n} q_k \log p_k \geqslant -\sum_{k=1}^{n} q_k \log q_k = H_n, \tag{6}$$

and this results shows that the subjective-objective measure of uncertainty I_n, in this case, is greater than the measure of objective uncertainty H_n. This is a consequence of the fact that the subjective probabilities do not coincide with the objective ones. This means that the degree of uncertainty of the objective probabilities of events is supplemented by another amount of uncertainty as a consequence of the incomplete estimation of these probabilities. In (6), we have the equality if and only if

$$p_k = q_k \qquad (k=1, \ldots, n).$$

Let us notice that the weights given by (4) satisfy the rule (2), while the weights given by (3) do not satisfy it. Indeed, according to (4), we have

$$w(\omega_i \cup \omega_j) = \frac{q(\omega_i \cup \omega_j)}{p(\omega_i \cup \omega_j)} = \frac{q_i + q_j}{p_i + p_j} = \frac{\frac{q_i}{p_i} p_i + \frac{q_j}{p_j} p_j}{p_i + p_j} = \frac{p_i w_i + p_j w_j}{p_i + p_j},$$

i.e. equality (2).

We point out that the rule (2) is satisfied by many types of weights in physics and also by the utilities in the decision theory according to a well-known von Neumann axiom. Notice also that the weights ascribed to the elementary events are usually independent of the probabilities of these events.

Finally, if all the weights are equal, i.e. if $w_1 = \ldots = w_n = w$, then the rule (2) is obviously satisfied and Property 7 becomes the well-known property of Shannon's entropy

$$H_{n+1}(p_1, \ldots, p_{n-1}, p', p'') = H_n(p_1, \ldots, p_n) + p_n H_2\left(\frac{p'}{p_n}, \frac{p''}{p_n}\right),$$

where $p_n = p' + p''$.

§ 3. Axiomatics for the weighted entropy

We are interested here in the uniqueness problem of the weighted entropy. In the proof of the uniqueness theorem that will be given here, we shall use the following known lemma:

LEMMA. *If L: $N \to R^+$ is a non-negative function defined on the set of natural numbers and if*

$$L(mn) = L(m) + L(n),$$

$$\lim_{n \to \infty} [L(n) - L(n-1)] = 0,$$

then

$$L(n) = \lambda \log n, \tag{7}$$

where λ is a positive constant.

For the proof see Rényi [18] or book [7].

Throughout this paragraph we shall also suppose the weights ascribed to the elementary events to satisfy equality (2).

Now let us prove the following uniqueness theorem:

THEOREM 1. *Consider the sequence of non-negative real-valued functions:*

$$(I_n(w_1, \ldots, w_n; p_1, \ldots, p_n))_{1 \leq n < \infty}$$

where every $I_n(w_1, \ldots, w_n; p_1, \ldots, p_n)$ is defined on the set

$$w_k \geq 0, \quad p_k \geq 0 \quad (k = 1, \ldots, n); \quad \sum_{k=1}^{n} p_k = 1.$$

Suppose that the following four axioms hold:

(A_1) *$I_2(w_1, w_2; p, 1-p)$ is a continuous function of p on the interval* [0, 1].

(A_2) *$I_n(w_1, \ldots, w_n; p_1, \ldots, p_n)$ is a symmetric function with respect to all pairs of variables (w_k, p_k) $(k = 1, \ldots, n)$.*

(A_3) *If*

$$w_n = \frac{p'w' + p''w''}{p' + p''}, \quad p_n = p' + p'',$$

then

$$I_{n+1}(w_1, \ldots, w_{n-1}, w', w''; p_1, \ldots, p_{n-1}, p', p'') = I_n(w_1, \ldots, w_n; p_1, \ldots, p_n) + p_n I_2\left(w', w''; \frac{p'}{p_n}, \frac{p''}{p_n}\right). \tag{8}$$

(A_4) *If all the probabilities are equal, then*

$$I_n\left(w_1, \ldots, w_n; \frac{1}{n}, \ldots, \frac{1}{n}\right) = L(n) \frac{w_1 + \ldots + w_n}{n},$$

$L(n)$ being a positive number for every $n > 1$.

Then we have

$$I_n(w_1, \ldots, w_n; p_1, \ldots, p_n) = -\lambda \sum_{k=1}^{n} w_k p_k \log p_k \tag{9}$$

where λ is an arbitrary positive constant.

Notice that axioms (A_1)–(A_2) are very natural.

Axiom (A_3) is just Property 7 mentioned in § 2. Finally, the last axiom (A_4) states that if all the probabilities are equal, then the weighted entropy is proportional to the mean value of weights.

Proof of Theorem 1: a) From (A_3) we have

$$I_3(w_1, w_2, w_3; \tfrac{1}{2}, \tfrac{1}{2}, 0) = I_2(w_1, w_2; \tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{2} I_2(w_2, w_3; 1, 0).$$

But (A_2) and (A_3) imply

$$\begin{aligned} I_3(w_1, w_2, w_3; \tfrac{1}{2}, \tfrac{1}{2}, 0) &= I_3(w_3, w_2, w_1; 0, \tfrac{1}{2}, \tfrac{1}{2}) \\ &= I_2(w_3, \tfrac{1}{2}(w_1+w_2); 0, 1) + I_2(w_1, w_2; \tfrac{1}{2}, \tfrac{1}{2}). \end{aligned}$$

Thus

$$I_2(w_2, w_3; 1, 0) = 2 I_2\big(\tfrac{1}{2}(w_1+w_2), w_3; 1, 0\big),$$

whatever are the weights w_1, w_2, w_3. In particular, if we put $w_1 = w_2$, we obtain

$$I_2(w_2, w_3; 1, 0) = 2 I_2(w_2, w_3; 1, 0)$$

for every pair w_2, w_3; therefore

$$I_2(w', w''; 1, 0) = 0, \tag{10}$$

whatever are the weights w', w''.

b) Applying (A_3) and equality (10), we obtain

$$\begin{aligned} I_{n+1}(w_1, \dots, w_n, w_{n+1}; p_1, \dots, p_n, 0) &= I_n(w_1, \dots, w_n; p_1, \dots, p_n) + p_n I_2(w_n, w_{n+1}; 1, 0) \\ &= I_n(w_1, \dots, w_n; p_1, \dots, p_n). \end{aligned} \tag{11}$$

c) We also have the equality

$$\begin{aligned} &I_{n+m-1}(w_1, \dots, w_{n-1}, w'_1, \dots, w'_m; p_1, \dots, p_{n-1}, p'_1, \dots, p'_m) \\ &\qquad = I_n(w_1, \dots, w_n; p_1, \dots, p_n) + p_n I_m\left(w'_1, \dots, w'_m; \frac{p'_1}{p_n}, \dots, \frac{p'_m}{p_n}\right), \end{aligned} \tag{12}$$

where

$$w_n = \frac{p'_1 w'_1 + \dots + p'_m w'_m}{p_n}, \qquad p_n = p'_1 + \dots + p'_m.$$

We shall prove this fact by induction with respect to m. Indeed, for $m=2$ we just get axiom (A_3). Suppose equality (12) has been verified for m. We shall prove its validity for $m+1$. Taking into account equality (11), we may suppose that $p'_i > 0$ for every $i = 1, \dots, m$. Then, (A_3) and (11) imply

$$\begin{aligned} &I_{n+m}(w_1, \dots, w_{n-1}, w'_1, \dots, w'_{m+1}; p_1, \dots, p_{n-1}, p'_1, \dots, p'_{m+1}) \\ &\qquad = I_{n+1}(w_1, \dots, w_{n-1}, w'_1, w''; p_1, \dots, p_{n-1}, p'_1, p'') + \end{aligned}$$

$$+p''I_m\left(w'_2,\dots,w'_{m+1};\frac{p'_2}{p''},\dots,\frac{p'_{m+1}}{p''}\right)=I_n(w_1,\dots,w_n;p_1,\dots,p_n)+$$

$$+p_nI_2\left(w'_1,w'';\frac{p'_1}{p_n},\frac{p''}{p_n}\right)+p''I_m\left(w'_2,\dots,w'_{m+1};\frac{p'_2}{p''},\dots,\frac{p'_{m+1}}{p''}\right)\tag{13}$$

where

$$w''=\frac{p'_2w'_2+\dots+p'_{m+1}w'_{m+1}}{p''},\qquad p''=p'_2+\dots+p'_{m+1},$$

and

$$w_n=\frac{p'_1w'_1+p''w''}{p'_1+p''}=\frac{p'_1w'_1+p'_2w'_2+\dots+p'_{m+1}w'_{m+1}}{p'_1+p'_2+\dots+p'_{m+1}}$$

$$=\frac{p'_1w'_1+\dots+p'_{m+1}w'_{m+1}}{p_n},\qquad p_n=p'_1+\dots+p'_{m+1}.$$

However, we supposed that (12) is true for m fixed. Thus

$$p_nI_{m+1}\left(w'_1,\dots,w'_{m+1};\frac{p'_1}{p_n},\dots,\frac{p'_{m+1}}{p_n}\right)$$

$$=p_nI_2\left(w'_1,w'';\frac{p'_1}{p_n},\frac{p''}{p_n}\right)+p_n\frac{p''}{p_n}I_m\left(w'_2,\dots,w'_{m+1};\frac{p'_2}{p''},\dots,\frac{p'_{m+1}}{p''}\right).\tag{14}$$

From (13) and (14) we get

$$I_{n+m}(w_1,\dots,w_{n-1},w'_1,\dots,w'_{m+1};p_1,\dots,p_{n-1},p'_1,\dots,p'_{m+1})$$

$$=I_n(w_1,\dots,w_n;p_1,\dots,p_n)+p_nI_{m+1}\left(w'_1,\dots,w'_{m+1};\frac{p'_1}{p_n},\dots,\frac{p'_{m+1}}{p_n}\right).$$

Therefore, equality (12) is true for $m+1$, and hence for arbitrary m.

d) Applying equality (12) several times, we obtain

$$I_{m_1+\dots+m_n}(w'_{11},\dots,w'_{1m_1},\dots,w'_{n1},\dots,w'_{nm_n};p'_{11},\dots,p'_{1m_1},\dots,p'_{n1},\dots,p'_{nm_n})$$

$$=I_n(w_1,\dots,w_n;p_1,\dots,p_n)+\sum_{i=1}^{n}p_iI_{m_i}\left(w'_{i1},\dots,w'_{im_i};\frac{p'_{i1}}{p_i},\dots,\frac{p'_{im_i}}{p_i}\right),\tag{15}$$

where

$$p_i=p'_{i1}+\dots+p'_{im_i}>0\qquad(i=1,\dots,n),$$

$$w_i=\frac{p'_{i1}w'_{i1}+\dots+p'_{im_i}w'_{im_i}}{p_i}\qquad(i=1,\dots,n).$$

e) Let us apply the last equality in the case where $m_1 = \ldots = m_n = m$. We obtain

$$I_{mn}(w'_{11}, \ldots, w'_{1m}, \ldots, w'_{n1}, \ldots, w'_{nm}; p'_{11}, \ldots, p'_{1m}, \ldots, p'_{n1}, \ldots, p'_{nm})$$
$$= I_n(w_1, \ldots, w_n; p_1, \ldots, p_n) + \sum_{i=1}^{n} p_i I_m\left(w'_{i1}, \ldots, w'_{im}; \frac{p'_{i1}}{p_i}, \ldots, \frac{p'_{im}}{p_i}\right).$$

If we take

$$p'_{ij} = \frac{1}{mn} \quad (i = 1, \ldots, n;\ j = 1, \ldots, m),$$

we get

$$p_i = \frac{1}{n} \quad (i = 1, \ldots, n),$$

and therefore

$$I_{mn}\left(w'_{11}, \ldots, w'_{1m}, \ldots, w'_{n1}, \ldots, w'_{nm}; \frac{1}{mn}, \ldots, \frac{1}{mn}\right)$$
$$= I_n\left(\frac{w'_{11} + \ldots + w'_{1m}}{m}, \ldots, \frac{w'_{n1} + \ldots + w'_{nm}}{m}; \frac{1}{n}, \ldots, \frac{1}{n}\right) +$$
$$+ \sum_{i=1}^{n} \frac{1}{n} I_m\left(w'_{i1}, \ldots, w'_{im}; \frac{1}{m}, \ldots, \frac{1}{m}\right).$$

Taking into account axiom (A_4), we obtain

$$L(mn)\frac{1}{mn}\sum_{i=1}^{n}(w'_{i1} + \ldots + w'_{im})$$
$$= L(n)\frac{1}{n}\sum_{i=1}^{n}\frac{w'_{i1} + \ldots + w'_{im}}{m} + L(m)\frac{1}{n}\sum_{i=1}^{n}\frac{w'_{i1} + \ldots + w'_{im}}{m},$$

or

$$L(mn) = L(n) + L(m). \tag{16}$$

f) Using equality (12), we get

$$I_n\left(w_1, \ldots, w_n; \frac{1}{n}, \ldots, \frac{1}{n}\right)$$
$$= I_2\left(w_1, \frac{w_2 + \ldots + w_n}{n-1}; \frac{1}{n}, \frac{n-1}{n}\right) + \frac{n-1}{n} I_{n-1}\left(w_2, \ldots, w_n; \frac{1}{n-1}, \ldots, \frac{1}{n-1}\right).$$

Applying axiom (A_4), we find

$$L(n)\frac{w_1 + \ldots + w_n}{n} = I_2\left(w_1, \frac{w_2 + \ldots + w_n}{n-1}; \frac{1}{n}, \frac{n-1}{n}\right) + \frac{n-1}{n}\cdot\frac{w_2 + \ldots + w_n}{n-1} L(n-1),$$

for every $w_1, w_2, \ldots, w_n$.

Let us put $w_1=0$. Then

$$I_2\left(0, \frac{w_2+\ldots+w_n}{n-1}; \frac{1}{n}, \frac{n-1}{n}\right) = \frac{w_2+\ldots+w_n}{n}\left[L(n)-L(n-1)\right]$$

whatever are the values of $w_2, \ldots, w_n$. Let us take $w_2=\ldots=w_n=w$. Then

$$0 \leqslant I_2\left(0, w; \frac{1}{n}, \frac{n-1}{n}\right) = \frac{n-1}{n} w\left[L(n)-L(n-1)\right],$$

and, according to axiom (A_1) and equality (10), we obtain

$$0 = I_2(0, w; 0, 1) = \lim_{n\to\infty} I_2\left(0, w; \frac{1}{n}, \frac{n-1}{n}\right)$$

$$= \lim_{n\to\infty} \frac{n-1}{n} w\left[L(n)-L(n-1)\right] = w \lim_{n\to\infty}\left[L(n)-L(n-1)\right],$$

for every non-negative real number w, i.e.

$$\lim_{n\to\infty}\left[L(n)-L(n-1)\right]=0. \tag{17}$$

g) Equalities (16) and (17) and lemma mentioned at the beginning of this paragraph imply

$$L(n)=\lambda \log n, \tag{18}$$

where λ is an arbitrary positive constant.

h) Let us substitute in equality (15)

$$n=2, \quad m_1=r, \quad m_2=s-r, \quad p'_{ij}=\frac{1}{s}.$$

Then

$$p_1=p'_{11}+\ldots+p'_{1r}=\frac{r}{s},$$

$$p_2=p'_{21}+\ldots+p'_{2,s-r}=\frac{s-r}{s},$$

$$w_1=\frac{p'_{11}w'_{11}+\ldots+p'_{1r}w'_{1r}}{p'_{11}+\ldots+p'_{1r}}=\frac{1}{r}\sum_{i=1}^{r} w'_{1i},$$

$$w_2=\frac{p'_{21}w'_{21}+\ldots+p'_{2,s-r}w'_{2,s-r}}{p'_{21}+\ldots+p'_{2,s-r}}=\frac{1}{s-r}\sum_{j=1}^{s-r} w'_{2j}.$$

Taking into account all these values, we obtain from (15)

$$I_s\left(w'_{11}, \ldots, w'_{1r}, w'_{21}, \ldots, w'_{2,s-r}; \frac{1}{s}, \ldots, \frac{1}{s}\right)$$

$$= I_2(w_1, w_2; p_1, p_2) + p_1 I_r\left(w'_{11}, \ldots, w'_{1r}; \frac{1}{r}, \ldots, \frac{1}{r}\right) +$$

$$+ p_2 I_{s-r}\left(w'_{21}, \ldots, w'_{2,s-r}; \frac{1}{s-r}, \ldots, \frac{1}{s-r}\right). \tag{19}$$

However, according to (A_4) and (18), we have

$$I_s\left(w'_{11}, \ldots, w'_{1r}, w'_{21}, \ldots, w'_{2,s-r}; \frac{1}{s}, \ldots, \frac{1}{s}\right) = \lambda \frac{1}{s}\left(\sum_{i=1}^{r} w'_{1i} + \sum_{j=1}^{s-r} w'_{2j}\right)\log s,$$

$$I_r\left(w'_{11}, \ldots, w'_{1r}; \frac{1}{r}, \ldots, \frac{1}{r}\right) = \lambda \frac{1}{r}\sum_{i=1}^{r} w'_{1i}\log r,$$

$$I_{s-r}\left(w'_{21}, \ldots, w'_{2,s-r}; \frac{1}{s-r}, \ldots, \frac{1}{s-r}\right) = \lambda \frac{1}{s-r}\sum_{j=1}^{s-r} w'_{2j}\log(s-r),$$

and thus, from (19) we get

$$I_2(w_1, w_2; p_1, p_2) = \lambda \frac{1}{s}\left(\sum_{i=1}^{r} w'_{1i} + \sum_{j=1}^{s-r} w'_{2j}\right)\log s -$$

$$-\lambda \frac{r}{s}\frac{1}{r}\sum_{i=1}^{r} w'_{1i}\log r - \lambda \frac{s-r}{s}\cdot\frac{1}{s-r}\sum_{j=1}^{s-r} w'_{2j}\log(s-r)$$

$$= -\lambda \frac{r}{s}\left(\frac{1}{r}\sum_{i=1}^{r} w'_{1i}\right)\log\frac{r}{s} - \lambda \frac{s-r}{s}\left(\frac{1}{s-r}\sum_{j=1}^{s-r} w'_{2j}\right)\log\frac{s-r}{s}$$

$$= -\lambda w_1 p_1 \log p_1 - \lambda w_2 p_2 \log p_2 .$$

e) In h) we have just proved formula (9) for $n=2$. Suppose it is true for n. We shall see that (9) holds for $n+1$, too. Indeed, from (8) we obtain

$$I_{n+1}(w_1, \ldots, w_{n-1}, w', w''; p_1, \ldots, p_{n-1}, p', p'')$$

$$= -\lambda \sum_{i=1}^{n} w_i p_i \log p_i - p_n\left(\lambda w' \frac{p'}{p_n}\log\frac{p'}{p_n} + \lambda w'' \frac{p''}{p_n}\log\frac{p''}{p_n}\right)$$

$$= -\lambda \sum_{i=1}^{n-1} w_i p_i \log p_i - \lambda w_n p_n \log p_n - \lambda w' p' \log p' - \lambda w'' p'' \log p'' + \lambda w' p' \log p_n + \lambda w'' p'' \log p_n$$

$$= -\lambda w_1 p_1 \log p_1 - \ldots - \lambda w_{n-1} p_{n-1} \log p_{n-1} - \lambda w' p' \log p' - \lambda w'' p'' \log p'',$$

because

$$w'p'+w''p''=w_n p_n .$$

Therefore, equality (9) is true for arbitrary n. Q.E.D.

Remark: Let us take $w_1=\ldots=w_n=1$. Then, as we have seen already, the weighted entropy is just the Shannon entropy. At the same time, axiom (A_4) is obviously satisfied in this case. It assumes the form

$$I_n\left(1, \ldots, 1; \frac{1}{n}, \ldots, \frac{1}{n}\right)=H_n\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)=L(n),$$

and (A_1), (A_2), (A_3) are just the well-known Faddeev axioms for the Shannon entropy (see [4]):

(A_1) *$H_2(p, 1-p)$ is a continuous function of $p \in [0, 1]$;*

(A_2) *$H_n(p_1, \ldots, p_n)$ is a symmetric function in all variables p_i;*

(A_3) *we have*

$$H_{n+1}(p_1, \ldots, p_{n-1}, p', p'')=H_n(p_1, \ldots, p_n)+p_n H_2\left(\frac{p'}{p_n}, \frac{p''}{p_n}\right),$$

where

$$p_n=p'+p''.$$

§ 4. Principle of maximum information

The principle of maximum information established independently by Ingarden [8], Jaynes [13] and Kullback [14] has many important applications in statistical physics and in mathematical statistics. Many important random distributions may be obtained as consequences of this principle. Taking into account some constraints, it is possible to obtain by maximizing information, the equal probability distribution (the finite discrete case, with no constraint), the geometrical distribution (continuous or discrete case where the constraint is the mean value of some fixed random variable), the Gauss distribution (continuous case, the constraints being the mean value and the variance of some fixed random variable). Recently, Ingarden and Kossakowski (see [11]) obtained even the Poisson distribution in the countable discrete case by maximizing the relative information, the constraint being the mean value of the random variable $f(\omega_n)=n$ $(n=0, 1, 2,\ldots)$ and the initial measure being

$$m(\omega_n)=\frac{1}{n!} \quad (n=0, 1, 2, \ldots).$$

Let us now derive the expression for the probability distribution, maximizing the weighted entropy. Dealing with natural logarithms, we shall prove the following theorem (principle of maximum information with no constraint):

THEOREM 2. *Consider the probability distribution*

$$p_i \geqslant 0 \quad (i=1, \ldots, n); \qquad \sum_{i=1}^{n} p_i = 1 \tag{20}$$

and the weights $w_i \geqslant 0$ $(i=1,\ldots,n)$. *The weighted entropy*

$$I_n = I_n(w_1, \ldots, w_n; p_1, \ldots, p_n) = -\sum_{i=1}^{n} w_i p_i \log p_i$$

is maximum if and only if

$$p_i = e^{-(\alpha/w_i)-1} \qquad (i=1, \ldots, n)$$

where α *is the solution of the equation*

$$\sum_{i=1}^{n} e^{-(\alpha/w_i)-1} = 1,$$

$\alpha + \sum_{i=1}^{n} w_i e^{-(\alpha/w_i)-1}$ *being exactly the maximum value of* I_n.

Proof: Because $-x \log x \leqslant 1/e$ for every $x \geqslant 0$ and $-x \log x = 1/e$ if and only if $x = 1/e$, we obtain, by using the Lagrange multipliers method,

$$I_n - \alpha = \sum_{i=1}^{n} w_i p_i \log \frac{1}{p_i} - \alpha \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} p_i \left(w_i \log \frac{1}{p_i} - \alpha \right)$$

$$= \sum_{i=1}^{n} w_i e^{-\alpha/w_i} (p_i e^{\alpha/w_i} \log p_i e^{\alpha/w_i}) \leqslant \sum_{i=1}^{n} w_i e^{-(\alpha/w_i)-1}.$$

The equality holds if and only if

$$p_i = e^{-(\alpha/w_i)-1} \qquad (i=1, \ldots, n),$$

and then $I_n = \alpha$. These probabilities must verify the relation (20), i.e.

$$\sum_{i=1}^{n} e^{-(\alpha/w_i)-1} = 1.$$

Q.E.D.

Remark 1: If all events have the same weight $w_1 = \ldots = w_n = 1$, then

$$p_i = \frac{1}{n} \qquad (i=1, \ldots, n),$$

i.e. we obtain the equal probability distribution.

Remark 2: The definition and some properties of the weighted entropy were given briefly in paper [1]. The axiomatic, i.e. the whole § 3, and the essential Property 7 of the weighted entropy are presented here for the first time.

Remark 3: In § 4, we did not suppose that the weights satisfy rule (2).

Remark 4: During the author's visit in Budapest, Professor I. Vincze suggested to investigate the relative weighted entropy

$$\log n \sum_{i=1}^{n} w_i p_i - I_w = \sum_{i=1}^{n} w_i p_i \log n p_i ,$$

instead of the weighted entropy. The former permits to pass very naturally to the continuous case.

REFERENCES

[1] Beliş, M., and S. Guiaşu, IEEE Trans. Inform. Theory **IT-14** (1968), 593.
[2] Bloch, G., Summer School, Varenna, 1966.
[3] Daróczy, Z., Acta Math. Acad. Sci. Hung. **15** (1964), 203.
[4] Faddeev, D., K., Soviet Math. Uspekni **11** (1956), 227.
[5] Feinstein, A., *Foundations of Information Theory*, McGraw, N.Y., 1958.
[6] Guiaşu, S., C. R. Acad. Sci. Paris **261** (1965), 1179.
[7] Guiaşu, S., and R. Theodorescu, *La théorie mathématique de l'information*, Dunod, Paris, 1968.
[8] Ingarden, R. S. Fortsch. Physik **12** (1964), 567; **13** (1965), 755.
[9] —, Prace Mat. **9** (1965), 273.
[10] —, Acta Phys. Polon. **36** (1969), 855.
[11] Ingarden, R. S. and A. Kossakowski, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astr. Phys. **19** (1971).
[12] Ingarden, R. S., and K. Urbanik, Colloq. Math. **9** (1962), 281.
[13] Jaynes, E. T., Phys. Rev. **106** (1957), 620; **108** (1957), 171.
[14] Kullback, S., *Information Theory and Statistics*, Wiley, N.Y., 1959.
[15] Majernik, V., Commun. Math. Phys. **12** (1969), 233.
[16] Onicescu, O., Stud. Cerc. Mat. **18** (1966), 1419.
[17] Perez, A., Trans. Third Prague Conf. Inform. Theory, Prague (1964), 499.
[18] Rényi, A., Mathematica **1(24)** (1959), 341.
[19] Shannon, C. E., Bell Syst. Techn. J. **27** (1948), 379, 623.
[20] Watanabe, S., *Knowing and Guessing*, Wiley, N.Y., 1969.