ZASTOSOWANIA MATEMATYKI
VI (1963)

C. RAJSKI (Warszawa)

ON THE NORMED INFORMATION RATE OF DISCRETE RANDOM VARIABLES

List of symbols

ξ, η — discrete random variables, A_i — an element of the sample space of ξ, B_j — an element of the sample space of η, $z_{ij} = \Pr\{\xi = A_i,\ \eta = B_j\}$, $x_i = \Pr\{\xi = A_i\}$, $y_j = \Pr\{\eta = B_j\}$, $H_{\xi\eta} = -\sum_{ij} z_{ij}\log z_{ij}$, $I_{\xi\eta} = \sum_{ij} z_{ij}\log z_{ij} - \sum_{ij} z_{ij}\log x_i y_j$, $R_{\xi\eta} = \dfrac{I_{\xi\eta}}{H_{\xi\eta}}$, $0 < \Theta_{ij} < 1$.

Summary. The properties of the quotient of information rate and joint entropy of two discrete random variables are discussed. It is shown that the quotient in question has some advantages over the correlation coefficient and thus may be used instead.

Posing of the problem. The most widely used measure of interdependence between two random variables is the correlation coefficient. Its chief disadvantage is that the knowledge of its value being equal to zero does not permit one to infer that the variables are independent. Besides, if the values of the sample spaces are non-numerical, the correlation coefficient can not be computed except for the case when the number of values taken by each variable is equal to two. It seems desirable, therefore, to define a measure of stochastic interdependence deprived of the disadvantages of the correlation coefficient. This paper is confined to the presentation of one particular solution and to the proof of its properties.

Solution and proof. We shall consider a quantity called subsequently the normed information rate and defined by the formula

$$R_{\xi\eta} = \frac{I_{\xi\eta}}{H_{\xi\eta}}, \qquad H_{\xi\eta} \neq 0. \tag{1}$$

We shall prove that it has the following properties:

A. The normed information rate does not depend upon sample spaces.

B. Its value ranges between 0 and 1.

C. If ξ and η are independent, then the normed information rate is equal to zero.

D. If the normed information rate is equal to zero, then ξ and η are independent.

E. If ξ and η are functionally dependent, then the normed information rate is equal to one.

F. If the normed information rate is equal to one, then ξ and η are functionally dependent.

We call two random variables *functionally dependent* if and only if they fulfil the following relation. For every A_i there exists such a B_j that $\xi = A_i$ implies $\Pr\{\eta = B_j\} = 1$.

Property A is obvious. Property B follows from the fact that both $H_{\xi\eta}$ and $I_{\xi\eta}$ ([1], p. 348) are non-negative and besides

$$I_{\xi\eta}-H_{\xi\eta} = \sum_{ij} z_{ij}\log z_{ij}^2 - \sum_{ij} z_{ij}\log x_i y_j < 0 \tag{2}$$

on account of the inequality

$$z_{ij}^2 \leqslant \Big(\sum_j z_{ij}\Big)\Big(\sum_i z_{ij}\Big) = x_i y_j. \tag{3}$$

Thus B is proved.

If ξ and η are independent, then

$$x_i y_j = z_{ij}, \tag{4}$$

and this makes the information rate and the normed information rate equal to zero, whence C.

If $R_{\xi\eta} = 0$, we shall make use of the following expansion ([1], p. 350).

$$I_{\xi\eta} = \sum_{ij} (z_{ij} - x_i y_j)^2 \frac{1-\Theta_{ij}}{(1-\Theta_{ij})x_i y_j + \Theta_{ij} z_{ij}}.$$

As each summand on the right-hand side is non-negative the sum can be equal to zero if and only if each summand is equal to zero and this implies (4). Thus property D is proved.

It follows from the definition given before that two random variables are functionally dependent if and only if the matrix of joint probabilities contains in each row and in each column only a single term different from zero. Let z_{pq} denote such an element. This permits us to write $x_p = y_q = z_{pq}$. The substitution into the formulas defining $H_{\xi\eta}$ and $I_{\xi\eta}$ yields $H_{\xi\eta} = I_{\xi\eta}$ and hence property E.

Conversely, if it is assumed that $R_{\xi\eta} = 1$, equation (2) turns into

$$I_{\xi\eta}-H_{\xi\eta} = \sum_{ij} z_{ij}\log z_{ij}^2 - \sum_{ij} z_{ij}\log x_i y_j = 0.$$

Restricting the summation to those pairs of indices p, q which satisfy the condition $z_{pq} \neq 0$ we get

$$\sum_{pq} z_{pq} \log \frac{z_{pq}}{x_p} \cdot \frac{z_{pq}}{y_q} = 0.$$

In view of (3) each summand is non-positive. However, since the sum is zero, each summand must be equal to zero as well, and this may occur if and only if $z_{pq} = x_p$ and $z_{pq} = y_q$. Using (3) once more we get

$$\sum_{n\neq q} z_{pn} = 0 \quad \text{and} \quad \sum_{m\neq p} z_{mq} = 0.$$

In view of the fact that the terms z_{ij}, as probabilities, are non-negative, the last two equations are equivalent to the statement that in each row and in each column there is at most one non-zero term. This in turn is equivalent to the statement that ξ and η are functionally dependent. Thus we have proved the last property, property F.

One may consider it to be a drawback of the normed information rate that it is non-negative, and thus carrying to some extent less information than the correlation coefficient, which can be negative as well. This, however, is not a particular property of the normed information rate but a necessary attribute of any measure of interdependence of two random variables with non-numerical sample spaces.

Reference

[1] S. Goldman, *Information Theory*, New York 1954.

Praca wpłynęła 31. 7. 1961

C. RAJSKI (Warszawa)

O UNORMOWANYM ILORAZIE INFORMACJI DYSKRETNYCH ZMIENNYCH LOSOWYCH

STRESZCZENIE

Zbadano właściwości ilorazu informacji przez entropię łączną dwóch zmiennych losowych dyskretnych.

Iloraz ten ma pewne cechy korzystniejsze od współczynnika korelacji i może być używany zamiast niego.

Ч. РАЙСКИЙ (Варшава)

НОРМИРОВАННОЕ ЧАСТНОЕ ИНФОРМАЦИИ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ПЕРЕМЕННЫХ

РЕЗЮМЕ

В работе рассматриваются свойства частного, получаемого в результате деления информации на совместную энтропию двух случайных дискретных переменных.

Это частное обладает некоторыми более полезными свойствами, чем коэффициент корреляции и может употребляться вместо него.