ADAM PŁOCKI

Comments on a definition of entropy

In their monograph [1] Yaglom and Yaglom introduced the concept of the entropy of a stochastic experiment, which is described by a finite space of elementary events. Let us assume that a stochastic experiment α describe a probabilistic system (E_α, S, P). If the space E_α is at most denumerable, then $S = 2^{E_\alpha}$, and in order to define the functional P it suffices to indicate the sequence of the probabilities of all monoelementary random events in S.

Let

$$E = \{e_1, e_2, ...\} \text{ and let } p_i = P(\{e_i\}) \quad \text{for} \quad i = 1, 2, ...$$

and

$$\sum_i p_i = 1$$

The entropy of a random experiment α is given by the expresion:

$$H(\alpha) = -\sum_i p_i lbp_i, \tag{1}$$

where lbx denotes $\log_2 x$. It is assumed that for $p_k = 0$, $p_k lbp_k = \lim_{p \to 0+} plbp = 0$. In [1] the entropy definition is restricted to a finite space; however, the relation (1) also includes the case of a denumerable set comprising all the experimental results possible a priori.

In [5] (pag. 530–531), and also in [4] (pag. 62) the concept of entropy of an at most denumerable set is introduced in the sense of the entropy related to the choice of one element of this set. It is easily seen that this concept coincides with the concept of the experimental entropy, and that the space of elementary events of the experiment is the set in question; the probability p_i of the choice of an element e_i from this set is simply the probability of the stochastic event $\{e_i\}$.

In the same book [5] (pag. 534–537) the concept of entropy of a discrete and continuous distribution is introduced. This concept is the counter part of the concept of entropy of a random variable found in other publications whose subject are the basic conceptions of the theory of information (v. [3] and [2]).

In this connection the question arises concerning the relation between the concept of the entropy of a random experiment and the concept of entropy of a random variable (alias the entropy of the distribution of this random variable) mapping the space of elementary events describing this random experiment.

If E_α is an at most denumerable space of elementary events — then every functional ξ mapping E_α into the set of real numbers is a measurable functional in the system (E_α, S, P). Suppose $\xi(E_\alpha) = \{a_1, a_2, \ldots\}$, and let us regard $\xi(E_\alpha)$ as a new space of elementary events; then $S_\xi = 2^{\xi(E_\alpha)}$ is the σ-field of random events. The functional P_ξ is defined by the indication of the sequence including the probabilities of all the monoelementary random events.

$$P_\xi(\{a_i\}) = P(\{e \in E_\alpha\colon\ \xi(e) = a_i\}) .$$

Thus the random variable ξ leads from the probabilistic space (E_α, S, P) to a new probabilistic space $(\xi(E_\alpha), S_\xi, P_\xi)$. It is easily seen that if the random variable is many-valued then the probabilistic system $(\xi(E_\alpha), S_\xi, P_\xi)$ does not differ from a formal point of view from the system (E_α, S, P).
In the case of a many-valued random variable ξ

$$P_\xi(\{\xi(e_i)\}) = P(\{e_i\}) ,$$

Thus in this case the entropy of the experiment α — in the sense of definition (1) — is identical with the entropy of a many valued random variable ξ mapping E_α.

$$H(\alpha) = H(\xi) .$$

This entropy does not depend on values assumed by this random variable. In the formation of this entropy are engaged only the probabilities in the distribution of the random variable, and the sets of numbers which are the probabilities in the distribution of all many-valued variables mapping the same space are all still equal.

It is obvious that $H(\xi) = H(\xi + c)$, where c is an arbitrary constant. For $b \neq 0$ the coresponding relation is

$$H(\xi) = H(b\xi + c) .$$

More generally, if f is a Borel many-valued function (it ensures the measurability of composition $f(\xi)$ — then

$$H(\xi) = H(f(\xi)) .$$

However, it is not true that "two random variables have the same entropies, one of the variables being the product of the other by an arbitrary constant" — this statement stemms from [2], pag. 195.

Let us assume that the distribution ξ is a two points uniform distribution; thus $H(\xi) = 1$. The random variable 0ξ has a one-point distribution, whence $H(0\xi) = 0$. The constant number mentioned on page 195 in [2] must be different from zero.

For a continuous random variable whose density is f_ξ the entropy $H(\xi)$ is defined by the formula below:

(2) $$H(\xi) = \int_{-\infty}^{\infty} f_\xi(x)\, lb f_\xi(x)\, dx .$$

The main difference between the properties of the entropy of a discrete random variable and the entropy of a continuous random variable consists in the fact that the entropy of a continuous variable can assume negative values.

In (3) the following modification of the entropy of a continuous random variable is proposed:

$$(3) \qquad H(\xi) = \int_{-\infty}^{\infty} f_\xi(x)\, lb\, l_\xi f_\xi(x)\, dx, \text{ where } l_\xi \text{ is a constant}.$$

Let us examine the family of all continuous random variables whose density functions are bounded. If for the random variable ξ of this family q is a number for which

$$\bigwedge_x f_\xi(x) < q\,,$$

then, assuming $l_\xi = \frac{1}{q}$ we have $l_\xi f_\xi(x) < 1$, whence:

$$\bigwedge_x f_\xi(x)\, lb\, l_\xi f_\xi(x) \leqslant 0\,,$$

thus $H(\xi) \geqslant 0$.

Let assume that ξ_1 and ξ_2 are continuous random variables whose densities are respectively $f_{\xi_1}(x_1)$ and $f_{\xi_2}(x_2)$, and that

$$\bigwedge_{x_1} l_1 f_{\xi_1}(x_1) < 1 \quad \text{and} \quad \bigwedge_{x_2} l_2 f_{\xi_2}(x_2) < 1\,.$$

Moreover we assume that $f(x_1, x_2)$ is a two-dimensional density of the random vector $[\xi_1, \xi_2]$, and that the random variables ξ_1 and ξ_2 are independent.

The entropy of the random vector $[\xi_1, \xi_2]$ is the counter part of the entropy of the composition of two experiments α and β; their respective spaces of elementary events E_α and E_β map the many-valued and continuous random variables ξ_1 and ξ_2.

$$(4) \qquad H([\xi_1, \xi_2)] = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\, lb\, lf(x_1, x_2)\, dx_1 dx_2\,.$$

The random variables are independent, whence it follows

$$H([\xi_1, \xi_2]) = H(\xi_1) + H(\xi_2)$$

This equality, however, requires in turn the relation $l = l_1 \cdot l_2$. Let us consider the random vector $[\xi_1, \xi_2]$, whose two-dimensional density is defined in the following way:

$$f(x_1, x_2) = \begin{cases} c & \text{for} \quad (x_1, x_2) \in Q \\ 0 & \text{for} \quad \text{the remaining points of the plane} \end{cases}$$

where $Q = \left\{(x_1, x_2)\colon\ |x_1 - x_2| \leqslant \frac{\varepsilon}{\sqrt{2}} \text{ and } x_2 \geqslant -x_1,\ x_2 \leqslant -x_1 + \varepsilon\sqrt{2}\right\}$. From the fact that $f(x_1, x_2)$ is the density of the random vector it follows that $c = \frac{1}{2\varepsilon}$.

We have for $i = 1, 2$

$$f_{\xi_i}(x_i) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } x_i \in \left[\frac{\varepsilon}{2\sqrt{2}}, \sqrt{2} - \frac{\varepsilon}{2\sqrt{2}}\right] \\ g_1(x_1) & \text{for } |x_i| < \frac{\varepsilon}{2\sqrt{2}}, \text{ and} \\ & \quad |x_i| - \sqrt{2} < \frac{\varepsilon}{2\sqrt{2}} \\ 0 & \text{for the remaining } x_i, \end{cases}$$

where $g_i(x_i)$ is a function whose values are smaller than $\frac{1}{\sqrt{2}}$. For every ε value $f_{\xi_i}(x_i) < 1$, whence it follows that for $l_1 = l_2 = 1$

$$H(\xi_i) = -\int_{-\infty}^{\infty} f_{\xi_i}(x_i)\, lb\, f_{\xi_i}(x_i)\, dx_i \geqslant 0$$

For $l = l_1 \cdot l_2 = 1$ we have:

$$H([\xi_1, \xi_2]) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\, lb\, f(x_1, x_2)\, dx_1\, dx_2 .$$

In the set $Qf(x_1, x_2) = \frac{1}{2\varepsilon}$; hence it results that for $\varepsilon < \frac{1}{2}$ we have

$$lb\, f(x_1, x_2) > 0 ;$$

in turn it follows from this relation that

$$H([\xi_1, \xi_2]) < 0 .$$

Thus the above proposed modification of the definition of entropy for random variables belonging to a family of variables with limited densities does not remove the difference mentioned above between the entropies of discrete and continuous variables.

References

[1] A. M. Jagłom i I. M. Jagłom: *Prawdopodobieństwo i informacja.* KiW, 1963.

[2] J. Nowakowski, W. Sobczak: *Teoria informacji.* WNT, Warszawa.

[3] W. S. Pugaczew: *Teoria funkcji przypadkowych i jej zastosowanie do zagadnień sterowania automatycznego.* MON, 1960.

[4] N. W. Smirnow, I. W. Dunin-Barkowski: *Kurs rachunku prawdopodobieństwa i statystyki matematycznej.* PWN 1969.

[5] M. Warmus: *Wykłady z probabilistyki.* Tom II, PWN, Warszawa 1973.