

Entropy—A Universal Concept in Sciences

Vladimír Majerník

Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovakia
Email: Eva.Majernikova@savba.sk

Received 24 February 2014; revised 24 March 2014; accepted 31 March 2014

Copyright © 2014 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Entropy represents a universal concept in science suitable for quantifying the uncertainty of a series of random events. We define and describe this notion in an appropriate manner for physicists. We start with a brief recapitulation of the basic concept of the theory probability being useful for the determination of the concept of entropy. The history of how this concept came into its to-day exact form is sketched. We show that the Shannon entropy represents the most adequate measure of the probabilistic uncertainty of a random object. Though the notion of entropy has been introduced in classical thermodynamics as a thermodynamic state variable it relies on concepts studied in the theory of probability and mathematical statistics. We point out that whole formalisms of statistical mechanics can be rewritten in terms of Shannon entropy. The notion “entropy” is differently understood in various science disciplines: in classical physics it represents the thermodynamical state variable; in communication theory it represents the efficiency of transmission of communication; in the theory of general systems the magnitude of the configurational order; in ecology the measure for bio-diversity; in statistics the degree of disorder, etc. All these notions can be mapped on the general mathematical concept of entropy. By means of entropy, the configurational order of complex systems can be exactly quantified. Besides the Shannon entropy, there exists a class of Shannon-like entropies which converge, under certain circumstances, toward Shannon entropy. The Shannon-like entropy is sometimes easier to handle mathematically than Shannon entropy. One of the important Shannon-like entropy is well-known Tsallis entropy. The application of the Shannon and Shannon-like entropies in science is really versatile. Besides the mentioned statistical physics, they play a fundamental role in the quantum information, communication theory, in the description of disorder, etc.

Keywords

Probability, Uncertainty, Shannon Entropy, Shannon-Like Entropies

1. Introduction

At the most fundamental level, all our further considerations rely on the concept of *probability*. Although there

is a well-defined mathematical theory of probability, there is no universal agreement about the meaning of probability. Thus, for example, there is the view that probability is an objective property of a system and another view that it describes a subjective state of belief of a person. Then there is the frequentist view that the probability of an event is the relative frequency of its occurrence in a long or infinite sequence of trials. This latter interpretation is often employed in the mathematical statistics and statistical physics. The probability means in everyday life the degree of ignorance about the outcome of a random trial. This is why the probability is commonly interpreted as degree of the subjective expectation of an outcome of a random trial. Both subjective and statistical probability are “normed”. It means that the degree of expectation that an outcome of a random trial occurs, and the degree of the “complementary” expectation, that it does not, is always equal to one [1]¹.

Although the concept of probability is here covered in a sophisticated mathematical language, it expresses only the commonly familiar properties of probability used in everyday life. For example, each number of spots at the throw of a simple die represents an elementary random event to which a positive real number is associated called its probability (relation (i)). The probability of two (or more) numbers of spots at the throw of a simple die is equal to the sum of their probabilities (relation (iii)). The sum of probabilities of all possible numbers of spots is normed to one (relation (iv)).

The word “entropy”² was first used in 1984 by Clausius in his book *Abhandlungen über Wärmetheorie* to describe a quantity accompanying a change from the thermal to mechanical energy and it continued to have this meaning in thermodynamics. Boltzmann [2] in his *Vorlesungen über Gastheorie* presented the statistical interpretation of the thermodynamical entropy. He linked the thermodynamic entropy with the molecular disorder. The general concept of entropy as a measure of uncertainty was first introduced by Shannon and Wiener. Shannon is also credited for the development of a quantitative measure of the amount of information [3]. Shannon entropy may be considered as a generalization of entropy, defined by Hartley, when the probability of each event is equal. Nyquist [4] was the first author who introduced a measure of information. His paper has largely remained unnoticed. After publication of Shannon seminal paper in 1948 [3], the use of entropy as measure of uncertainty grew rapidly and was applied with various successes in most area of human endeavor.

Mathematicians were attracted to the possibility of providing axiomatic structure of entropy and to the ramification thereof. The axiomatic approach to the concept of entropy attempts to find a system of postulates which provides a unique mathematical characteristic of entropy³ and which adequately reflects the properties asked from the probabilistic uncertainty measure in a diversified real situation. This has been very interesting and thought-provoking area for scientists. Khinchin [5] was the first who gave a clear and rigorous presentation of the mathematical foundation of entropy. A good number of works have been done to describe the properties of entropy. An extensive list of works in this field can be found in the book of Aczél and Daróczy [6].

The fundamental concept for the description of random processes is the notion of the *random trial*. A random trial is characterized by a set of its outcomes (values) and the corresponding probability distribution. A typical random trial is the throw of a single dice characterized by the following scheme (S_1, S_2, \dots, S_6 are the positions of dice after its throwing)

S	S_1	S_2	S_3	S_4	S_5	S_6
P	1/6	1/6	1/6	1/6	1/6	1/6
x	1	2	3	4	5	6

¹The concept of probability was mathematically clarified and rigorously determined about sixty years ago. The probability is interpreted as a complete measure on the σ -algebra γ of the subsets S_1, S_2, \dots, S_n of the set of the elementary random events B . The probability measure P fulfils following relations:

(i) $P(S_i) \geq 0, S_i \in \gamma$

(ii) From $S_i \subset S_j$ it follows $P(S_i) \leq P(S_j), S_i, S_j \in \gamma$.

(iii) If S_1, S_2, \dots, S_n are such elements of σ -algebra γ , for which $S_i S_j = 0, i \neq j$, then it holds the following equation:

$$P(S_1 + S_2 + \dots + S_n) = \sum_{k=1}^n P(S_k).$$

(iv) $P(B) = 1$.

The σ -algebra, on which the set function P is defined, is called the Kolmogorov probability algebra. The triplet $[B, \gamma, P]$ denotes the probability space. Under a random variable \tilde{x} we understand each real-valued measurable function defined on the elementary random events B [1].

²The word “entropy” stems from the Greek word “ $\tau\rho\theta\pi\eta$ ” which means “transformation”.

³Entropy is sometimes called “missing information”.

To any random trial it is assigned a *random variable* \tilde{x} which represents a mathematical quantity assuming a set of values with the corresponding probabilities (see, e.g. [1]).

There are two measures which express the uncertainty of a random trial:

(i) The *moment* measures containing in its definition *both* the values assigned to trial outcomes and the set of the corresponding probabilities. The moment uncertainty measures are given as a rule by the *higher statistical moments* of a random variable \tilde{x} . As it is well-known, the k -th moment about zero (uncorrelated moment) x^k , and the central moment of the k -th order x_c^k assigned to a discrete random variable \tilde{x} with the probability distribution $P = \{P(x_1), P(x_2), \dots, P(x_i)\}$, is defined as

$$x^{(k)} = \sum_i P(x_i) (x_i)^k$$

and

$$x_c^{(k)} = \sum_i P(x_i) (x_i - \bar{x}_0)^k, \quad \bar{x}_0 = \sum_i P(x_i) x_i,$$

respectively.

The statistical moments of a random variable are often used as the uncertainty measures of the random trial, especially in the experimental physics, where, e.g., the standard deviation of measured quantities characterizes the accuracy of a physical measurement. The moment uncertainty measures of a random variable are also used by formulating the uncertainty relations in quantum mechanics [7].

(ii) The *probabilistic* or *entropic* measures of uncertainty of a random trial contain in their expressions *only* the components of the probability distribution of a random trial.

To determine the notion of entropy we consider quantities, called as *partial uncertainties* which are assigned to individual probabilities $P_i, i = 1, 2, \dots, n$. A partial uncertainty we denote by symbol H_i . In any probabilistic uncertainty measures, a partial uncertainty is function *only* of the corresponding probability $H_i = f(P_i)$, $i = 1, 2, \dots, n$. The requirements asked from a partial uncertainty $H(P_i)$ are the following [8]: (see Appendix):

- (i) It is a monotonously decreasing continuous and unique function of the corresponding probability;
- (ii) The common value of the uncertainty of a certain outcome of two statistically independent trials $H(P_i, P_j)$ is *additive*, i.e.

$$H(P_i, P_j) = H(P_i) + H(P_j), \quad (1)$$

where P_i and P_j are the probability of the i -th and j -th outcome, respectively;

- (iii) $H(P_i = 1/e) = 1$.

It was shown that the only function which satisfies these requirements has the form [8]

$$H(P_i) = -\log P_i.$$

The mean value of the partial uncertainties H_1, H_2, \dots, H_n is

$$\tilde{H} = -\sum_i^n P_i \log P_i. \quad (2)$$

The quantity \tilde{H} is called information-theoretical or Shannon entropy. We denote it by symbol S . Shannon entropy is a real and positive number. It is a function only of the components of the probability distribution $P \equiv \{P_1, P_2, \dots, P_n\}$ assigned to the set of outcomes of a random trial.

Shannon entropy satisfies the following demands (see Appendix): (i) If the probability distribution contains only one component, e.g. $P_i = 1, i = 1, 2, \dots, n$, and the rest components are equal to zero, then $S(P) = 0$. In this case, there is no uncertainty in a random trial because an outcome is realized with certainty.

(ii) The more spread is the probability distribution P , the larger becomes the entropy S .

(iii) For a uniform probability distribution P_u , $H(P_u)$ becomes maximal. In this case, the probabilities of all outcomes are equal, therefore the mean uncertainty of such a random trial becomes maximum.

One uses for the characterization of a random trial a random scheme. If \tilde{x} is a discrete random variable assigned to a random trial then its random scheme has the form

S	S_1	S_2	S_n
P	$P(x_1)$	$P(x_2)$	$P(x_n)$
X	x_1	x_2	x_n

S_1, S_2, \dots, S_n are the outcomes of a random trial (in quantum physics, e.g. the *quantum states*), $P(x_1), P(x_2), \dots, P(x_n)$ are their probabilities and x_1, x_2, \dots, x_n are the values defined on S_1, S_2, \dots, S_n (in quantum physics, e.g. the *eigenvalues*). A probability distribution, $P \equiv \{P_1, P_2, \dots, P_n\}$, is the complete set of probabilities of all individual outcomes of a random trial.

We note that there is a set of the probabilistic uncertainty measures defined by means of other functions than $H(P_i) = -\log P$. They are called nonstandard or Shannon-like entropies. We shall deal with them in the next sections.

2. Entropy as a Qualificator of the Configurational Order

Since the simple rule holds that the smaller the order in a system the larger its entropy, the entropy appears to be the appropriate quantity for the expression of the measure of the configurational order (organization). The orderliness and entropy of a physical system are related to each other inversely so that any increase in the degree of configurational order must necessarily result in the decrease of its entropy. The measure of the configurational order constructed by using entropies is called Watanabe measure and is defined as follows [9]:

configurational order of a system = (sum of entropies of the parts of the system) – (entropy of the whole system).

The Watanabe measure for configurational order is related to the other measure of configurational organization well-known in theory of information, called redundancy. Both measures express quantitatively the property of the configurationally organized systems to have order between its elements, which causes that the system as a whole behaves in a more deterministic way than its individual parts. If a system consists only of elements which are statistically independent, the Watanabe measure for the configurational organization becomes zero. If the elements of a system are deterministically dependent, its configurational organization gets the maximum value. A general system has its configurational organization between these extreme values. To the prominent systems which can be organized configurationally belong physical statistical systems (*i.e.*, about all, Ising systems of spins) [10]. High configurational organization is exhibited especially by systems which have some spatial, temporal or spatio-temporal structures that have arisen in a process which takes place far from thermal equilibrium (e.g. laser, fluid instabilities, etc.) [11]. These systems can be sustained only by a steady flow of energy and matter, therefore they are called open systems [12]. A large class of systems, which are generally organized configurationally as well as functionally, comprises the so-called string systems which represent sequences of elements forming finite alphabets. To these systems belong, e.g., language, music, genetic DNA and various bio-polymers. Since many of such systems are goal-directed and have a functional organization as well, they are especially appropriate for the study of the interrelation between the configurational and functional organization [10].

3. The Concept of Entropy in Thermodynamics and Statistical Physics

A remarkable event in the history of physics was the interpretation of the phenomenological thermodynamics in terms of motion and randomness. In this interpretation, the temperature is related to motion while the randomness is linked with the Clausius entropy. The homeomorphous mapping of the phenomenological thermodynamics on the formalism of mathematical statistics gave rise to two entropy concepts: the Clausius thermodynamic entropy as a thermodynamic state variable of a thermodynamic system and the Boltzmann statistical entropy as the logarithm of probability of state of a physical ensemble. The fact that the thermodynamic entropy is a state variable means that it is completely defined when the state (pressure, volume, temperature, etc.) of a thermodynamic system is defined. This is derived from mathematics, which shows that only the initial and final states of a thermodynamic system determine the change of its entropy. The larger the value of the entropy of a particular state of a thermodynamic system, the less available is the energy of this system to do work.

The statistical concept of entropy was introduced in physics when seeking a statistical quantity homeomorphous with the thermodynamic entropy. As it is well-known, the Clausius entropy of a thermodynamic system S_i is linked with ensemble probability W by the celebrated Boltzmann law, $S_i = K_B \log W$, where W

is so-called “thermodynamic” probability determined by the configurational properties of a statistical system and K_B is the Boltzmann constant⁴. The Boltzmann law represents the solution to the functional equation between S_i and W . Let us consider a set of the isolated thermodynamic systems $\Sigma_1, \Sigma_2, \dots, \Sigma_n$. According to Clausius, the total entropy of this system is an additive function of the entropies of its parts, *i.e.*, it holds

$$S_i(\Sigma_1 + \Sigma_2 + \dots + \Sigma_n) = S_i(\Sigma_1) + S_i(\Sigma_2) + \dots + S_i(\Sigma_n). \quad (3)$$

On the other side, the joint “thermodynamic” probability of system (3) is

$$W(\Sigma_1 + \Sigma_2 + \dots + \Sigma_n) = \prod_{i=1}^n W(\Sigma_i). \quad (4)$$

To obtain the homomorphism between Equations (3) and (4), it is sufficient that

$$S_i = K_B \log W. \quad (5)$$

which is just the Boltzmann law [2].

We give some remarks regarding the relationship between the Clausius, Boltzmann and Shannon entropies:

(i) The thermodynamic probability W in the Boltzmann law is given by the number of the possibilities how to distribute N particles in n cells having different energies E_1, E_2, \dots, E_n

$$W = \log \frac{N!}{\prod_1^n N_i!}.$$

We show that the physical entropy given by Boltzmann’s law is equal to the sum of Shannon entropies of energies taken as random variables defined on the individual particles, *i.e.*,

$$S_{Boltz} = K_B N S_{Shann}. \quad (6)$$

The probability P_i that a particle of the statistical ensemble has the i -th value of energy is given by the ratio N_i/N . Inserting the probabilities

$$P(x_1) = N_1/N, P(x_2) = N_2/N, \dots, P(x_n) = N_n/N \quad (7)$$

into Boltzmann’s entropy formula we have

$$S_{Boltz} = K_B \log \frac{\log N!}{\prod_1^n (NP_i)!} = K_B \left[\log N! - \sum_{i=1}^n \log (NP_i)! \right]. \quad (8)$$

Supposing that the number of particles in a statistical ensemble is very large, we can use the asymptotic formula

$$\log N! = N \log N - N + O(N^{-1})$$

which inserted in Boltzmann’s entropy yields

$$\begin{aligned} S_{Boltz} &= K_B \left[N \log N - N - \sum_{i=1}^n NP_i \log (NP_i) + \sum_{i=1}^n NP_i + O(N^{-1}) \right] \\ &= -NK_B \sum_{i=1}^n P_i \log P_i + O(N^{-1}). \end{aligned} \quad (9)$$

For very large N , the second term in Equation (9) can be neglected and we find

⁴The probability as well as the Shannon entropy is dimensionless quantities. On the other side, the thermodynamical entropy has the physical dimension equal to $[J \cdot K^{-1}]$. Therefore, in order to get the correct physical dimension for the thermodynamic entropy we must multiply the Shannon entropy by the Boltzmann constant, which has the value $K_B = 1.38 \times 10^{-23} [J \cdot K^{-1}]$.

$$S_{Boltz} = K_B N S_{Shann} = -K_B N \sum_1^n P_i \log P_i.$$

We see that the Boltzmann entropy of an ensemble with large N is equal to the sum of Shannon entropies of the individual particles. The asymptotical equality between Boltzmann and Shannon entropies for the large N makes it possible to use the Shannon entropy also for describing an statistical ensemble. The pioneer on this field was E. Jaynes who published, already in fifties, works in which only Shannon entropy was used to formulate statistical physics [13]. However, many authors advocating the use of Shannon entropy in statistical physics do not fully realized the difference between Boltzmann's and Shannon entropy. The use of Shannon entropy can be only justified if one considers the physical ensemble as a system of random objects on which energy (or other physical quantity) is taken as a random variable. Then, the total entropy of the whole ensemble is given as the sum of Shannon entropies of individual statistical elements (e.g., particles). While the Boltzmann's entropy loses its sense for an ensemble containing only a few particles, Shannon entropy is defined also for an "ensemble" with even one particle. Boltzmann's entropy is typical ensemble concept while Shannon entropy is a probabilistic concept. This is not only the change of the methodology when treating statistical ensemble but it has also long-reaching conceptual and even pedagogical consequences.

According to Jaynes [14], the equilibrium probability distribution of the particle energy of a statistical ensemble should maximize the Shannon entropy

$$S = -K_B \sum_{i=1}^n P(E_i) \log P(E_i) \quad (10)$$

subject to given constraints. For example, by taking the mean energy per particle as the constraint at the extremizing procedure, we obtain the following probability distribution for the particle energy

$$P(E_i) = \exp[-\lambda - \mu E_i],$$

where the constants λ and μ are to be determined by substituting $P(E_i)$ into constraint's equations. We see how easily and quickly we obtain results forming the essence of the classical statistical mechanics. The use of Shannon entropy in statistical physics makes it possible to rewrite it in terms of modern theory of probability where a statistical ensemble is treated as a collection of the mutually interacting random objects [13].

4. The Shannon-Like Entropies

Recently, there is an endeavour in the applied sciences (see, e.g. [15]) to employ entropic measures of uncertainty having similar properties as information entropy, but they are simpler to handle mathematically. The classical measure of probabilistic uncertainty which has dominated in the literature since it was proposed by Shannon, is the *information or Shannon entropy* defined for a discrete random variable according by the formula

$$S = -\sum_i^n P_i \log P_i. \quad (11)$$

Since Shannon has introduced his entropy, several other classes of probabilistic uncertainty measures (entropies) have been described in the literature (see, e.g., [16]). We can broadly divide them into two classes:

(i) The Shannon-like uncertainty measures which for a certain value of the corresponding parameters converge towards the Shannon entropy, e.g., Rényi's entropy

(ii) The Maassen and Uffink uncertainty measures which converges also, under certain conditions, to the Shannon entropy,

$$M_r = \left[\sum_i P(x_i)^{(1+r)} \right]^{1/r}. \quad (12)$$

(iii) The uncertainty measures having no direct connection to Shannon entropy, e.g., information "energy" defined in information theory as [16]

$$E_e(\tilde{x}) = \sum_i (P_i)^2 \quad (13)$$

and called Hilbert-Schmidt norm in quantum physics. The most important uncertainty measures of the first class are:

- (i) The Rényi entropy defined as follows [17]

$$H_\alpha(\tilde{x}) = \frac{1}{1-\alpha} \log \left\{ \sum_{i=1}^n P(x_i)^\alpha \right\} \quad \alpha \neq 1. \quad (14)$$

- (ii) The Havrda-Charvat entropy (or α -entropy)⁵ is defined as [18]

$$S_\beta(\tilde{x}) = \frac{1}{1-\alpha} \sum_{i=1}^n \left(P(x_i)^\alpha - 1 \right) \quad \alpha \in (0, \infty) \quad \alpha \neq 1. \quad (15)$$

For the sake of completeness, we list some other entropy-like uncertainty measures presented in the literature [19]:

- (i) The trigonometric entropy is defined as [10]

$$H_\gamma^{(M)} = \frac{1}{\pi(\gamma-1)} \cos \left(\pi/2 \sum_{i=1}^n \left(P(x_i)^\gamma \right) \right) \quad \gamma \neq 1. \quad (16)$$

- (ii) The R-norm entropy $H_R(\tilde{x})$ defined by the formula

$$H_R(\tilde{x}) = \frac{R}{R-1} \left\{ 1 - \sum_{i=1}^n \left(P(x_i)^R \right)^{1/R} \right\} \quad R \neq 1.$$

All the above-listed Shannon-like entropies converge towards Shannon entropy if $\alpha \rightarrow 1$, $b \rightarrow 1$, $R \rightarrow 1$, $\gamma \rightarrow 1$ and $\beta \rightarrow 1$. In some instances, it is simpler to compute $H_\alpha, H_R, H_\beta, H_\gamma$ and H_b and then recover by taking limits $b, \alpha, \beta, \gamma, R \rightarrow 1$.

A quick inspection shows that all five Shannon-like entropies listed above are all mutually functionally related. For example, each of the Havrda-Charvat entropies can be expressed as a function of the Rényi's entropy, and vice versa

$$S_\beta(\tilde{x}) = \frac{1}{1-\beta} \left(\exp((1-\beta)H_\alpha(\tilde{x})) - 1 \right).$$

There are six properties which are usually considered desirable for a measure of a random trial: (i) symmetry, (ii) expansibility, (iii) subadditivity, (iv) additivity, (v) normalization, and (vi) continuity. The only uncertainty measure which satisfies all these requirements is Shannon entropy. Each of the other entropies violates at least one of them, e.g. Rényi's entropy violates only the subadditivity property, Havrda-Charvat's entropy violates the additivity property, the R-norm entropies violate both subadditivity and additivity. More details about the properties of each entropies can be found elsewhere (e.g., [15]). The Shannon entropy satisfies all above requirements put on uncertainty measure and it exact matches the properties of physical entropy⁶. All these classes of entropies represent the probabilistic uncertainty measures which have similar mathematical properties as Shannon entropy.

The best known Shannon-like probabilistic uncertainty measure is the Havrda and Charvat entropy [18] which is more general than Shannon measure and much simpler than Renyi's measure. It depends on a parameter α which is from the interval $\alpha \in (0, \infty)$. As such, it represents a family of uncertainty measures which includes information entropy as a limiting case when $\alpha \rightarrow 1$. We note that in physics the Havrda Charvat entropy is known as Tsallis entropy [20]. All the mentioned entropic measures of uncertainty are functions of the components of the probability distribution of a random variable \tilde{x}_A and they have three important properties: (i) They assume their maximal values for the uniform probability distribution of \tilde{x} . (ii) They become zero for the probability distributions having only one component. (iii) They express a measure of the spread of a probability distribution. The larger this spread becomes, the smaller values they assume. These properties qualify them for be-

⁵The Havrda-Charvat α -entropy exactly matches the Tsallis [20] non-extensive entropy of statistical physics.

⁶The Tsallis entropy [20] which is mathematically identical with the Havrda-Charvat α -entropy [18], was introduced by two mathematicians Havrda and Charvat in sixties, violates the additivity property which is considered as an essential property of physical entropy. This is why the other Shannon-like entropies could be more suitable for formulation of an alternative statistical physics.

ing the measures of uncertainty (inaccuracy) in the physical theory of measurement.

The entropic uncertainty measures for a discrete random variable are, in the frame of theory of probability, exactly defined. The transition from the discrete to the continuous entropic uncertainty measures is, however, not always unique and has still many open problems. A continuous random variable \tilde{x}_c is characterized by the function of its probability density $p(x)$. The moment and probabilistic uncertainty measures exist also for the continuous random variables. The typical moment measure is the k -th central moment of \tilde{x}_c . The classical probabilistic uncertainty measure of a continuous random variable \tilde{x}_c is the corresponding Shannon entropy $H(\tilde{x}_c)$. It is a function of the probability density $p(x)$ and consists of two terms⁷

$$H(\tilde{x}_c) = H^{(1)}(p) + H^{(2)}$$

$$H^{(1)}(p) = -\int p(x) \log p(x) dx$$

$$H^{(2)} = -\lim_{\Delta x \rightarrow 0} \log \Delta x.$$

$H(\tilde{x}_c)$ with both terms $H^{(1)}(p)$ and $H^{(2)}$ always diverges. Usually, one “renormalizes” $H(\tilde{x}_c)$ by taking only the term $H^{(1)}(x)$ (called differential entropy or the Shannon entropy functional $H(x)$) for the entropic uncertainty measure of a continuous random variable. This functional is well known to play an important role in probability and statistics. We refer to [15] for applications of the Shannon entropy functional to the theory of probability and statistics.

As it is well-known, the Shannon entropy functionals of some continuous variables represent complicated integrals which often are difficult to compute analytically or even numerically. Everybody, who tried to calculate analytically the differential entropies of the continuous variables, became aware how difficult it may be. From the purely mathematical point of view, the differential entropy can be taken as a formula for expressing the spread of any standard single-valued function (the probability density belongs to this class of functions). Generally, the Shannon entropy functional assigns to a probability density function (belonging to the class of functions $L^2(\mathbb{R}^1)$) a real number H through a mapping $L^2(\mathbb{R}_1) \rightarrow H$. H is a monotonously increasing function of the degree of “spreading” of $p(x)$, *i.e.* the larger H becomes, the spread is $p(x)$.

The Shannon entropy functional was studied just at the beginning of information theory [17]. Since that time, besides the Shannon entropy functional, several other entropy functionals were introduced and studied in the probability theory. The majority of them are dependent on certain parameters. As such, they form a whole family of different functionals (including the Shannon entropy functional as a special case). In a sense, they are a generalization of the Shannon entropy functional. Some of them can equally well express the spread of the probability density functions as differential entropy and are considerably easier to handle mathematically. These include:

(i) The Rényi entropic functional [17]

$$H_\alpha^{(R)} = \frac{1}{1-\alpha} \log \left(\int [p(x)]^\alpha dx \right) \quad \alpha \in \mathbb{R}. \quad (17)$$

⁷In order to apply the formula for the Shannon entropy for the continuous random variable \tilde{x}_c with the probability density function $p(x)$, we divide the x -axis into n equidistant intervals. The probability that \tilde{x}_c assumes value from the interval $\langle x_i, x_i + \Delta x \rangle$ is $P(x_i) \approx p(\bar{x}_i) \Delta x$, where \bar{x}_i is from i -th interval. Inserting $P(x_i)$ into the Shannon entropy we have

$$H(\tilde{x}_c) \approx -\sum_{i=1}^n p(\bar{x}_i) \Delta x \log(p(\bar{x}_i) \Delta x) = -\sum_{i=1}^n p(\bar{x}_i) \log p(\bar{x}_i) \Delta x - \sum_{i=1}^n p(\bar{x}_i) \log \Delta x,$$

Passing to the infinitesimal interval, we obtain

$$H(\tilde{x}_c) = H^{(1)}(p) + H^{(2)},$$

where

$$H^{(1)}(p) = -\int p(x) \log p(x) dx$$

and

$$H^{(2)} = -\lim_{\Delta x \rightarrow 0} \log \Delta x.$$

$H^{(1)}(p)$ is the Shannon entropy functional.

(ii) The Havrda-Charvat entropic functional [18]

$$H_{\alpha}^{(HC)} = \frac{1}{1-\alpha} \left(\int [p(x)]^{\beta} dx - 1 \right) \quad \alpha \in R. \quad (18)$$

(iii) The trigonometric entropic functional [10]

$$H_{\gamma}^{(M)} = \frac{2}{\pi(\gamma-1)} \cos\left(\pi/2 \int [p(x)]^{\gamma} dx\right) \quad \gamma \in [1, \infty). \quad (19)$$

Note that $H_{\alpha}^{(R)}$, $H_{\beta}^{(HC)}$ and $H_{\gamma}^{(M)}$ tend to $H(p)$ as α, β, γ tend to 1. Again, in some instances, it is simpler to compute $H_{\alpha}^{(R)}$, $H_{\beta}^{(HC)}$ and $H_{\gamma}^{(M)}$ and then recover $H(p)$ by taking limits $\alpha, \beta, \gamma \rightarrow 1$. In treating $H_{\alpha}^{(R)}$, $H_{\beta}^{(HC)}$ and $H_{\gamma}^{(M)}$, it is often enough to study the properties of the functional

$$G_{\alpha}(p) = \int [p(x)]^{\alpha} dx.$$

As it is known long ago, the entropy functionals $H(p)$, $H_{\alpha}^{(R)}$ and $H_{\gamma}^{(M)}$ have a mathematical shortcoming connected with the dimension of the physical probability density function. In contrast with the probability which is a dimensionless number, the probability density function has a dimension so that its appearance behind the logarithm and cosine in the entropy functionals $H(p)$, $H_{\alpha}^{(R)}$ and $H_{\gamma}^{(M)}$ is mathematically inadmissible. This brings complication when calculating $H(p)$, $H_{\alpha}^{(R)}$ and $H_{\gamma}^{(M)}$ for a physical random variable (see, e.g., [21])^{8,9}.

5. Conclusions

From what has been said so far it follows:

(i) The concept of entropy is inherently connected with the probability distribution of outcomes of a random trial. The entropy quantifies the probability uncertainty of a general random trial.

(ii) There are two ways how to express the uncertainty of a random trial:

The *moment* and probabilistic measure. The former measure includes in its definition *both* values assigned to trial outcomes and their probabilities. The latter measure contains in its definition *only* the corresponding probabilities. The moment uncertainty measures are given as a rule by the *higher statistical moments* of a random variable whereas the probabilistic measure is expressed by means of entropy. The most important probabilistic uncertainty measure is the Shannon entropy defined by the formula

$$E = -\sum_1^n P_i \log P_i,$$

where P_i is the probability of i -th outcome of a random trial.

(iii) By means of Shannon entropy it is possible to quantify the configurational order in the set of elements of a general system. The corresponding quantity is called the Watanabe measure of configurational order and is defined as follows

$$\text{configurational order of a system} = (\text{sum of entropies of the parts of the system}) - (\text{entropy of the whole system}).$$

This measure expresses quantitatively the property of a configurationally organized systems to have order between its elements, which causes that the system as a whole behaves in a more deterministic way than its individual parts.

(iv) The asymptotical equality between the Boltzmann and Shannon entropies for the statistical systems with large particles makes it possible to use the Shannon entropy for describing statistical ensembles.

(v) Besides the Shannon entropy there exists a class of so-called Shannon-like Entropies. The most important

⁸It is typical for the entropy and even for the probability itself that their uncertainty measures are determined through a set of certain reasonable requirements (axioms), therefore they are more "abstract" than the measures, e.g., for work or energy in physics. Physical quantities are mostly derived from the concepts of motion or field which are more concrete than the concept of probability and entropy.

⁹In many textbook and also in some advanced books, Shannon entropy and information are used so as if were synonymous. This can be confusing and may lead to the some conceptual shortcomings. Especially, the Brillouin's use of the concept of information, which he used in his celebrated work "Science and Information Theory" [22], has been often criticized (see, e.g. [23]).

Shannon-like entropies are (a) The Rényi entropy Equation (14); (b) The Havrda-Charvat entropy Equation (13). The well-known Tsallis entropy is mathematically identical with Havrda-Charvat entropy.

In conclusion, we can state that the concept of entropy is inherently connected with the probability theory. The application of Shannon entropy in science is really versatile. Besides the mentioned statistical physics, Shannon entropy is used in metronomic, in biological physics, in quantum physics and even in cosmology. Entropy expresses the extent of the randomness of a probabilistic (statistical) system and, therefore, it belongs to the important quantities for describing the natural phenomena. This is why entropy represents in physics a fundamental quantity next the energy.

References

- [1] Feller, W. (1968) An Introduction to Probability Theory and Its Applications. Volume I., John Wiley and Sons, New York
- [2] Boltzmann, L. (1896) Vorlesungen über Gastheorie. J. A. Barth, Leipzig.
- [3] Shannon, C.E. (1948) A Mathematical Theory of Communication. *The Bell System Technical Journal*, **27**, 53-75. <http://dx.doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- [4] Nyquist, H. (1924) Certain Factors Affecting Telegraph Speed. *Bell System Technical Journal*, **3**, 324-346. <http://dx.doi.org/10.1002/j.1538-7305.1924.tb01361.x>
- [5] Khinchin, A.I. (1957) Mathematical Foundation of Information Theory. Dover Publications, New York.
- [6] Aczél, J. and Daróczy, Z. (1975) On Measures of Information and Their Characterization. Academic Press, New York.
- [7] Merzbacher, E. (1967) Quantum Physics. 7th Edition, John Wiley and Sons, New York.
- [8] Faddejew, D.K. (1957) Der Begriff der Entropie in der Wahrscheinlichkeitstheorie. In: *Arbeiten zur Informationstheorie I. DVdW*, Berlin.
- [9] Watanabe, S. (1969) Knowing and Guessing. John Wiley and Sons, New York.
- [10] Majerník, V. (2001) Elementary Theory of Organization. Palacký University Press, Olomouc.
- [11] Haken, H. (1983) Advanced Synergetics. Springer-Verlag, Berlin.
- [12] Ke-Hsuch, L. (2000) Physics of Open Systems. *Physics Reports*, **165**, 1-101.
- [13] Jaynes, E.T. (1957) Information Theory and Statistical Mechanics. *Physical Review*, **106**, 620-630. <http://dx.doi.org/10.1103/PhysRev.106.620>
- [14] Jaynes, E.T. (1967) Foundations of Probability Theory and Statistical Mechanics. In: Bunge, M., Ed., *Delaware Seminar in the Foundation of Physics*, Springer, New York. http://dx.doi.org/10.1007/978-3-642-86102-4_6
- [15] Ang, A.H. and Tang, W.H. (2004) Probability Concepts in Engineering. *Planning*, **1**, 3-5.
- [16] Vajda, I. (1995) Theory of Information and Statistical Decisions. Kluwer, Academic Publisher, Dordrecht.
- [17] Rényi, A. (1961) On the Measures of Entropy and Information. *4th Berkeley Symposium on Mathematical Statistics*, 547-555.
- [18] Havrda, J. and Charvát, F. (1967) Quantification Method of Classification Processes. Concept of Structural α -Entropy. *Kybernetika*, **3**, 30-35.
- [19] Majerník, V., Majerníková, E. and Shpyrko, S. (2003) Uncertainty Relations Expressed by Shannon-Like Entropies. *Central European Journal of Physics*, **6**, 363-371. <http://dx.doi.org/10.2478/s11534-008-0057-6>
- [20] Tsallis, C. (1988). Possible Generalization of Boltzmann-Gibbs Statistics. *Journal of Statistical Physics*, **52**, 479-487. <http://dx.doi.org/10.1007/BF01016429>
- [21] Majerník, V. and Richterek, L. (1997) Entropic Uncertainty Relations. *European Journal of Physics*, **18**, 73-81. <http://dx.doi.org/10.1088/0143-0807/18/2/005>
- [22] Brillouin, L. (1965) Science and Information Theory. Academic Press, New York.
- [23] Flower, T.B. (1983) The Notion of Entropy. *International Journal of General Systems*, **9**, 143-152.
- [24] Guiasu, S. and Shenitzer, A. (1985) The Principle of Maximum Entropy. *The Mathematical Intelligencer*, **7**, 42-48. <http://dx.doi.org/10.1007/BF03023004>
- [25] Khinchin, A.I. (1957) Mathematical Foundation of Information Theory. Dover Publications, New York.
- [26] Chaitin, G.J. (1982) Algorithmic Information Theory. John Wiley and Sons, New York.
- [27] Kolmogorov, A.N. (1965) Three Approaches to the Quantitative Definition of Information. *Problems of Information Transmission*, **1**, 1-7.

Appendix

The Essential Mathematical Properties of Entropy

We ask from the Shannon entropy the following desirable properties [24]:

- (i) $H(\tilde{x}) \geq 0$, is a continuous function of all components of a probability distribution of a random variable \tilde{x} which is invariant under any permutation of the indices of the probability components.
- (ii) If probability distribution of \tilde{x} has only one component which is different from zero then $H(\tilde{x}) = 0$
- (iii) For $H(\tilde{x})$ it holds

$$H_m(P(x_1), P(x_2), \dots, P(x_m)) = H_{m+1}(P(x_1), P(x_2), \dots, P(x_m), 0). \quad (\text{A1})$$

(iv)

$$H_m(P(x_1), P(x_2), \dots, P(x_m)) \leq H_m(1/m, 1/m, \dots, 1/m), \quad (\text{A2})$$

with equality if and only if $P(x_k) = 1/m$, $k = 1, 2, \dots, m$.

(v) If $\Pi^* = \{\Pi_{1,1}, \Pi_{1,2}, \dots, \Pi_{m,n}\}$ is a joint probability distribution whose marginal probability distributions are $P' = \{P(x_1)', P(x_2)', \dots, P(x_m)'\}$ and $Q' = \{Q(y_1)', Q(y_2)', \dots, Q(y_n)'\}$, respectively, then

$$\begin{aligned} H_{mm} &= (\Pi_{1,1}, \Pi_{1,2}, \dots, \Pi_{m,n}) \\ &= H_m(P(x_1), P(x_2), \dots, P(x_m)) + \sum_{k=1}^m P(x_k) H_n \left(\frac{\Pi_{k,1}}{P(x_k)} + \dots + \frac{\Pi_{k,n}}{P(x_k)} \right), \end{aligned} \quad (\text{A3})$$

where the conditional entropy, defined as $H_n \left(\frac{\Pi_{k,1}}{P(x_k)} + \dots + \frac{\Pi_{k,n}}{P(x_k)} \right)$, is computed only for those values of k

for which $P(x_k) \neq 0$.

(vi) Using notation given above it holds

$$\sum_{k=1}^m P(x_k) H_n \left(\frac{\Pi_{k,1}}{P(x_k)} + \dots + \frac{\Pi_{k,n}}{P(x_k)} \right) \leq H_n(Q(y_1), Q(y_2), \dots, Q(y_n)). \quad (\text{A4})$$

The equality in (A4) is valid if and only if

$$\Pi_{k,l} = P(x_k)Q(y_l) \quad k = 1, 2, \dots, m \quad l = 1, 2, \dots, n,$$

in which case (A3) becomes

$$H_{mm}(\Pi^*) = H_m(P) + H_n(Q).$$

All these properties can be proved in an elementary manner. Without entering into the technical details, we note that properties (i)-(iii) are obvious while property (v) can be obtained by a straightforward computation taking into account only the definition of entropy. Finally, from Jensen's inequality

$$\sum_{k=1}^m a_k f(b_k) \leq f \left(\sum_{k=1}^m a_k b_k \right)$$

applied to the concave function $f(x) = -x \log x$, we obtain property (iv) by putting, $a_k = 1/m$, $b_k = P(x_k)$, $k = 1, 2, \dots, m$, and the inequality (A4) by putting $a_k = P(x_k)$, $b_k = \Pi_{k,l}/P(x_k)$, $k = 1, 2, \dots, n$, and, in the last case, summing the resulting n inequalities.

Interpretation of the above properties agrees with common sense, intuition, and the reasonable requirements that can be asked from a measure of uncertainty. Indeed, a random experiment which has only one possible outcome (that is, a strictly deterministic trial) contains no uncertainty at all; we know what will happen before per-

forming the experiment. This is just property (ii). If to the possible outcomes having the probability zero, the amount of uncertainty with respect to what will happen in the trial remains unchanged (property (iii)). Property (iv) tells us that in the class of all probabilistic trials having m possible outcomes, the maximal uncertainty is contained in the special probabilistic trial whose outcomes are equally likely. Before interpreting the last properties let us consider two discrete random variable \tilde{x} and \tilde{y} , whose ranges contain m and n numerical values, respectively. Using the notations as in property (v), suppose that Π^* is the joint probability distribution of the pair (\tilde{x}, \tilde{y}) , and P and Q are the marginal probability distributions of \tilde{x} and \tilde{y} , respectively. In this case equality iii may be written more compactly

$$H(\tilde{x}, \tilde{y}) = H(\tilde{x}) + H(\tilde{y}|\tilde{x}), \quad (\text{A5})$$

where

$$H(\tilde{x}, \tilde{y}) = H_m(\Pi_{1,1}, \Pi_{1,2}, \dots, \Pi_{m,n}) \quad H(\tilde{x}) = H_m(P(x_1), \dots, P(x_m))$$

and

$$H(\tilde{y}|\tilde{x}) = \sum_{k=1}^m P(x_k) H_n\left(\frac{\Pi_{k,1}}{P(x_k)} + \dots + \frac{\Pi_{k,n}}{P(x_k)}\right).$$

Here $H(\tilde{y}|\tilde{x})$ is the conditional entropy of \tilde{y} given \tilde{x} . According to (v), the amount of uncertainty contains in a pair of random variables (or, equivalently, in compound- or product-probabilistic trial) is obtained by summing the amount of uncertainty contained in \tilde{x} and the uncertainty contained in \tilde{y} conditioned by random variable \tilde{x} . Similarly, we get

$$H(\tilde{x}, \tilde{y}) = H(\tilde{y}) + H(\tilde{x}|\tilde{y}), \quad (\text{A6})$$

where

$$H(\tilde{y}) = H_n(Q(y_1), Q(y_2), \dots, Q(y_n))$$

and

$$H(\tilde{x}|\tilde{y}) = \sum_{l=1}^n Q(y_l) H_m\left(\frac{\Pi_{1,l}}{Q(y_l)}, \dots, \frac{\Pi_{m,l}}{Q(y_l)}\right).$$

Here

$$H_m\left(\frac{\Pi_{1,l}}{Q(y_l)}, \dots, \frac{\Pi_{m,l}}{Q(y_l)}\right)$$

is the conditional entropy of \tilde{x} given the l -th value of \tilde{y} . H_m is defined only for those values of l for which $Q(y_l) > 0$. From (A5) and (A6) we get

$$H(\tilde{x}) - H(\tilde{x}|\tilde{y}) = H(\tilde{y}) - H(\tilde{y}|\tilde{x})$$

which is the so-called ‘‘uncertainty balance’’, the only conservation law for entropy.

Finally, property (vi) shows that some data on \tilde{x} can only decrease the uncertainty on \tilde{y} , namely

$$H(\tilde{y}|\tilde{x}) \leq H(\tilde{y}) \quad (\text{A7})$$

with equality if and only if \tilde{x} and \tilde{y} are independent. From (A5) and (A7) we get

$$H(\tilde{x}, \tilde{y}) \leq H(\tilde{x}) + H(\tilde{y})$$

with equality if and only if \tilde{x} and \tilde{y} are independent.

Fortunately this inequality holds for any number of components. More generally, for s random variables

with arbitrary finite range we can write

$$H(\tilde{x}_1, \dots, \tilde{x}_s) \leq H(\tilde{x}_1) + \dots + H(\tilde{x}_s)$$

with equality if and only if $\tilde{x}_1, \dots, \tilde{x}_s$ are globally independent. Therefore

$$W(\tilde{x}_1, \dots, \tilde{x}_s) = \sum_{s=1}^s H(\tilde{x}_i) - H(\tilde{x}_1, \dots, \tilde{x}_s) \geq 0 \quad (\text{A8})$$

measures the global dependence between the random variables $\tilde{x}_1, \dots, \tilde{x}_s$, that is, the extent to which the system $(\tilde{x}_1, \dots, \tilde{x}_s)$, due to interdependence, makes up “something more” than the mere juxtaposition of components. In particular, $W = 0$ if and only if $\tilde{x}_1, \dots, \tilde{x}_s$ are independent.

Note that the difference between the amount of uncertainty contained by the pair (\tilde{x}, \tilde{y}) and the amount of dependence between the components \tilde{x} and \tilde{y} , namely

$$d(\tilde{x}, \tilde{y}) = H(\tilde{x}, \tilde{y}) - W(\tilde{x}, \tilde{y})$$

or, equivalently,

$$d(\tilde{x}, \tilde{y}) = 2H(\tilde{x}, \tilde{y}) - H(\tilde{x}) - H(\tilde{y}) = H(\tilde{x}|\tilde{y}) + H(\tilde{y}|\tilde{x}),$$

is the *distance* between the random variables \tilde{x} and \tilde{y} , with the two random variables considered identical if either one completely determines the other, or if $H(\tilde{x}|\tilde{y}) = 0$ and $H(\tilde{y}|\tilde{x}) = 0$. Therefore, the “pure randomness” contained in the pair (\tilde{x}, \tilde{y}) , *i.e.*, the uncertainty of the whole, minus the dependence between the components, measured by $d(\tilde{x}, \tilde{y})$, is a distance between \tilde{x} and \tilde{y} .

Khinchin [25] proved that properties (i), (iii), (iv) and (v), taken as axioms imply uniquely the formula of the Shannon entropy (except an numerical factor). It is worthy to remark that there is also another way to determine the uncertainty of a probabilistic object. The Shannon entropy is a measure of the degree of uncertainty of random object whose probability distribution is given. In algorithmic theory, the primary concept is that of the information content of an individual object, which is a measure of how difficult it is to specify or describe and how to construct or calculate that object. This notion is also known as information-theoretical complexity. The information content $I(s)$ of a binary string s is defined to be the size in bits of the smallest program for a canonical universal computer to calculate s [26] [27].

The work was supported by the Slovak Academy Grant Agency VEGA-2/0059/12.