ENTROPY

T. DOWNAROWICZ AND P. FREJ

Abstract. The term "entropy" appears in thermodynamics, probability theory, information theory, theory of dynamical systems (including theory of stochastic processes, ergodic theory and topological dynamics), and moreover it has its popular meaning. Here we review and compare different definitions of entropy, putting special emphasis on the interpretation of entropy as a measure of chaos or randomness.

1. Thermodynamical origins of entropy

The idea of entropy was introduced by a German physicist Rudolf Clausius in the middle of nineteenth century (see [C]). It was closely connected with the second law of thermodynamics formulated earlier by a French engineer and physicist, Sadi Carnot.

Briefly, a change in entropy in a volume element caused by an infinitesimal change in the conditions equals the ratio of the heat gain to temperature:

$$dS = \frac{dQ}{T}.$$

The total change of entropy in the system is defined as the proper definite integral over the phase space. This allows us to calculate the entropy of any system to an accuracy of the universal constant C. One determines C according to the third law of thermodynamics, but we will forget about it and accept a constant which is convenient for us.

The second law of thermodynamics states that the entropy of any closed system, e.g. the whole universe, is a non-decreasing function of time. Equivalently, heat cannot flow by itself from a cold to a hot object. This principle is independent of the first law of thermodynamics, the so called conservation of energy, which does not involve any time-dependence. Most often examples in thermodynamics concern gas contained in one or several reservoirs, subject to compression or expansion and emitting or deriving heat from the environment. To illustrate the importance of the second law we also provide an example of this kind.

Let us consider a closed cylinder with a mobile piston placed at position p (as in Figure 1), filled by gas on both sides of the piston. We assume that there is no friction.

At the beginning, the pressure and the temperature on both sides of the piston and outside are equal. Denote them by P_0 and T_0, respectively, and the original state by **A**. Then, we slowly move the piston to the center of the cylinder,

This article is partially based on a survey "Entropia" (in Polish), [D].

allowing the heat to flow between the cylinder and its environment, so that the temperature is constant. The pressure on the right side of the piston increases.

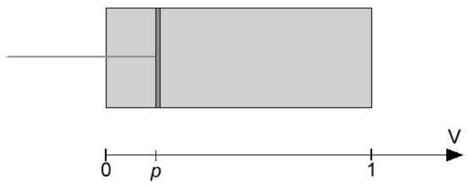


Fig.1. State **A**

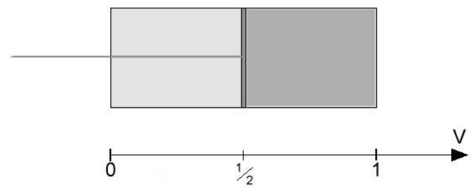


Fig.2. State **B**

The force acting on the piston at position x is proportional to the difference between the pressures:

$$F = c\left(P_0\frac{1-p}{1-x} - P_0\frac{p}{x}\right).$$

Thus, the work done while moving the piston equals:

$$\int_p^{\frac{1}{2}} F\,dx = cP_0[(1-p)\log(1-p) + p\log p + \log 2].$$

The law of conservation of energy implies that this work equals the heat emitted by the system. Due to the temperature being constant, the change in entropy is equal to:

$$\Delta S = c\frac{P_0}{T_0}[-(1-p)\log(1-p) - p\log p - \log 2].$$

Assume that the initial pressure and temperature were chosen so that $c\dfrac{P_0}{T_0} = 1$. Taking the appropriate C, we can also assume that the entropy of state **A** equals $S_{\frac{1}{2}} = \log 2$. Then the entropy in state **B** equals

$$S_p = -(1-p)\log(1-p) - p\log p.$$

Obviously, the function S_p is nonnegative and for $p = \frac{1}{2}$ attains its maximum.

Now, isolate the cylinder and release the piston. Due to the difference between pressures, the piston returns towards position p in the original state **A**, but no energy (neither heat nor work) is emitted or derived from the environment. Obviously, the temperature on the left side of the piston increases. Since the piston is not isolated, the heat is continuously exchanged between separate parts of the cylinder, so at the end the piston is at position p, the temperature on both sides of the piston is again equal and the entropy attains its maximal value $\log 2$. It shows that closed systems automatically achieve the state of the highest entropy. Though it does not contradict the law of conservation of energy, an isolated system never passes to a state of lower entropy, e.g. state **B**, without any influence from outside.

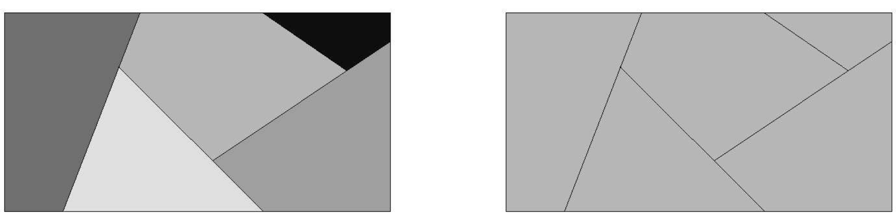
Another excellent physicist from Vienna University, Ludwig Boltzmann, working at the turn of the century, put entropy into a probabilistic setup (see [B]). This gave a new meaning to our interpretation of entropy. Let us remove the piston from our cylinder and calculate the probability that the system reaches the state thermodynamically equivalent to state **B** by itself. Assuming that there are N uniformly distributed particles, independently wandering inside the reservoir, we can calculate that the logarithm of the probability that there are $pN \in \mathbb{N}$ particles in the left half (and the rest $(1-p)N$ in the right half) equals

$$\log\left(\frac{N!}{(pN)!((1-p)N)!}\right) \approx N[-(1-p)\log(1-p) - p\log p] - N\log 2$$

(we used Stirling's formula: $\log n! \approx n\log n - n$). Note that $N\log 2$ denotes the logarithm of the number of "all states", and $N[-(1-p)\log(1-p) - p\log p]$ is the logarithm of the number of "states" corresponding to **B**. One can see that for $p = \frac{1}{2}$ the number of "states" leading to equilibrium in both parts is (for large N) so close to the number of "all states", that the probability of such an equilibrium is almost one. But for $p \neq \frac{1}{2}$ this probability is rather small. Hence, the probabilistic meaning of the second law of thermodynamics is that the system does not achieve state **B** by itself, because its probability is close to zero.

The considerations above result in a popular idea of entropy being an increasing function of the "probability" of observing certain event. It is understood as a measure of disorder and chaos in the system investigated.

We can illustrate it in the following way: let us divide the cylinder into several pieces and assign to the particles in each piece a different color. Then let the particles move freely. After some time every piece of the cylinder consists of the same proportion of colors. The probability of returning to the initial state is negligibly small.



`Fig.3.` Partitions of the cylinder

We can apply this principle not only to gas diffusion but also to propagation of species, illnesses or even our personal things inside our room. Disorder increases spontaneously and it needs some expenditure of efforts to restore the order.

2. Entropy in ergodic theory and stochastic processes

The popular notion of entropy mentioned above leads us to a formal definition used in the theory of stochastic processes. The main concept of entropy in dynamical systems is due to Andrei Kolmogorov ([K]) and Yakov Sinai ([Si]), though "non-physical" entropy appeared earlier in information theory in papers by Claude Shannon ([S]).

Let (X, Σ, μ) be a probability space, and $\alpha = \{A_1, \dots, A_k\}$ be *a finite measurable partition of* X, i.e. a finite family of disjoint measurable sets (events) satisfying $\mu(A_i) > 0$ for every $i = 1, \dots, k$ and $\mu(\bigcup_i A_i) = 1$. Denote $\mu(A_i) = p_i$ for $i = 1, \dots, k$.

Definition 2.1. The *information function* of a partition α is the simple function $I_\alpha : X \to [0, \infty)$ given by the formula:

$$I_\alpha(x) = -\sum_i \mathbf{1}_{A_i}(x) \log p_i.$$

This definition expresses the following intuition: the less probable the event we observe the more precious information we get. Example: the name Bernoulli does not identify a single mathematician but the name de la Vallée-Poussin does. To identify Bernoulli we need some additional information.

Definition 2.2. The *entropy of the partition* α is defined as the integral of the information function I_α:

$$H(\alpha) = -\sum_i p_i \log p_i.$$

For two measurable partitions $\alpha = \{A_1, \dots, A_k\}$ and $\beta = \{B_1, \dots, B_l\}$ we define their *refinement* as the partition

$$\alpha \vee \beta = \{A_i \cap B_j : i = 1, \dots, k;\ j = 1, \dots, l;\ \mu(A_i \cap B_j) \neq 0\}.$$

We also say that β is *finer* than α ($\alpha \preccurlyeq \beta$) if every $B \in \beta$ is contained in some $A \in \alpha$. Some basic properties of entropy are listed below.

Proposition 2.3. *Let* $\alpha = \{A_1, \dots, A_k\}$, $\beta = \{B_1, \dots, B_l\}$ *be finite measurable partitions. Then:*

(1) $0 \leqslant H(\alpha) \leqslant \log k$,
(2) $H(\alpha) = 0$, *if and only if* α *is a trivial partition,*
(3) *entropy attains its maximal value* $\log k$ *for the partition consisting of* k *sets of equal measures,*
(4) $H(\alpha \vee \beta) \leqslant H(\alpha) + H(\beta)$ *and equality holds only for stochastically independent partitions (i.e. such that every pair of sets* $A \in \alpha$, $B \in \beta$ *is independent).*

It can be shown that these conditions determine the formula for the entropy of a partition.

Let $(X_n)_{n \geqslant 1}$ be a stationary stochastic process on some probability space $(\Omega, \mathcal{F}, P)$, with values in X and each coordinate X_n having distribution μ. The space Ω may be thought of as the space of all trajectories of the process with some measure $\tilde{\mu}$ whose one-dimensional marginal measures are equal to μ. For a given partition α of X random variables $X_1, \dots, X_N$ induce the partition α^N of the space Ω by:

$$\alpha^N = X_1^{-1}(\alpha) \vee \cdots \vee X_N^{-1}(\alpha).$$

One of the basic facts is that the sequence $H(\alpha^N)$ is subadditive. Hence the limit in the following definition exists and may be replaced by infimum.

Definition 2.4. *The entropy of the process (X_n) with respect to the partition α* is given by the formula:

$$h(\alpha) = \lim_{N\to\infty} \frac{1}{N} H(\alpha^N).$$

The entropy of a partition is defined as the expectation of the information coming from an experiment, whose result falls into one of separate classes being the elements of the partition. So the latter may be interpreted as the asymptotic average information obtained from one step of a process consisting of sequential repetitions of the same experiment.

If (X_n) is a sequence of independent random variables then $h(\alpha) = H(\alpha)$, since the appropriate counter-images of the partition α are stochastically independent. Now we go back to the example. As it was already calculated, putting N particles uniformly and independently into a cylinder separated into two parts as in state **A** (but without a barrier), we will almost surely obtain a distribution of particles close to $pN : (1-p)N$. Furthermore, the logarithm of the number of such configurations equals approximately $NS_p = NH(\alpha) = H(\alpha^N)$. Obviously, the number of all possible configurations given some bisection is always 2^N, but the probability of configurations leading to other distributions is negligibly small. Thus, $\exp(H(\alpha^N))$ is the number of configurations combining almost the whole support of the probability measure, and further, the information function I_{α^N} divided by N is close to $h(\alpha)$ on a big set. This is one of the most important theorems about entropy, proved by the father of information theory American Claude Shannon in 1948 (see [S]) and improved by Brockway McMillan (1953, [M]) and Leo Breimann (1957, [Br]). In practice, the theorem says that one can compute the entropy considering just one randomly chosen trajectory. There is also no need for the strong assumption of the independence of the process. It is enough to assume that the process is *ergodic*, i.e. the space Ω is not a sum of subsets, on which the process evolves separately.

Theorem 2.5. *(Shannon-McMillan-Breimann)*
If the process (X_n) on $(\Omega, \mathcal{F}, P)$ is ergodic, then

$$\lim_{N\to\infty} \frac{1}{N} I_{\alpha^N}(\omega) = h(\alpha) \qquad \text{a.e.}$$

In other words, a typical set in the partition α^N has size (measure) approximately equal $\exp(-Nh(\alpha))$.

A stochastic process is a probabilistic approach to dynamics. Instead of it one can consider a dynamical system being a measure-preserving map of the measure space (X, μ), i.e. a map T satisfying $\mu(T^{-1}E) = \mu(E)$ for every measurable $E \subset X$. The entropy of T with respect to a given partition α of X is defined by the same formula as the entropy of the stochastic process (see Definition 2.4) with $\alpha^N = \alpha \vee T^{-1}\alpha \vee \cdots \vee T^{-(N-1)}\alpha$. This entropy is usually denoted by $h(T, \alpha)$, as it depends on both the map T and the partition α, but we will keep on suppressing T in order to use the same notation for dynamical systems and stochastic processes. Note that in the case of stochastic processes we refer in the definition to the space of events Ω, while for dynamical systems everything is done in X. As one would expect, Shannon-McMillan-Breimann theorem holds for *ergodic* dynamical systems, i.e. the systems containing no nontrivial invariant ($T^{-1}E = E$) subsets.

In order to be able to talk about the entropy of a process or a dynamical system, one has to get rid of the fixed partition α.

Definition 2.6. *The entropy of a process* (X_n) *(a dynamical system* T*)* is the supremum of all $h(\alpha)$ over all finite measurable partitions α.

One of the most important features of entropy is its invariance under measure-theoretical isomorphism. From the practical viewpoint, taking the supremum is very inconvenient and ineffective. But in some cases, the famous theorem of Kolmogorov and Sinai ([K],[Si]) offers some help. A measurable partition α is called *a generator*, if the smallest σ-algebra containing $\alpha^{-N} \vee \cdots \vee \alpha^{N}$ for all N coincides with the whole σ-algebra of the underlying probability space (or of the space of events in the case of stochastic process).

Theorem 2.7. *(Kolmogorov-Sinai)*
If the partition α *is a generator, then the entropy of the process (dynamical system) equals* $h(\alpha)$.

In 1970 Wolfgang Krieger proved the theorem informing about the range of application of the preceding result (see [Kr]).

Theorem 2.8. *(Krieger)*
If a process (dynamical system) is ergodic and has finite entropy then there exist a finite generator.

One can also deduce that a process satisfying the assumptions of Krieger's theorem is isomorphic to the process having finite number of states. Then the space of events Ω may be identified with the set of all sequences over a finite alphabet, whose cardinality is equal to the cardinality of the generator. The process is interpreted as a measure $\tilde{\mu}$ on the set of trajectories, and stationarity of the process is equivalent to the measure being invariant under the action of the *shift* $\sigma : \Omega \to \Omega$ defined by $(\sigma\omega)(n) = \omega(n+1)$. In the case of independent process the measure $\tilde{\mu}$ is a product measure induced by μ (the distribution of X_n's) and the dynamical system $(\Omega, \tilde{\mu}, \sigma)$ is called *Bernoulli shift*. The entropy of such Bernoulli shift over a k-element alphabet Λ with the probability of symbols given by the stochastic vector $\mu = (p_1, \ldots, p_k)$ is equal to $-\sum_{i=1}^{k} p_i \log p_i$, since for independent random variables $h(\alpha) = H(\alpha)$ (just let α be the partition $\{\omega \in \Lambda^{\mathbb{N}} : \omega(0) = \lambda\}$, where λ admits all elements of Λ). The most spectacular result exploiting the notion of entropy is the famous theorem due to Donald Ornstein (1970; [O1], [O2]): two Bernoulli shifts are isomorphic if and only if they have the same entropy (i.e. entropy is *a complete invariant* in the class of Bernoulli systems).

3. Entropy in information theory and data compression

Assume that somebody sends us information as a sequence of signs from a finite k-element alphabet, e.g. we get a very long enciphered telegram. What we called a "trajectory" in the theory of stochastic processes is now referred to as a "signal", and the "process" is now called the "source". Assume that the received signal is "typical" for a given source, which means that we consider a trajectory from the set of full measure, mentioned in the Shannon-McMillan-Breiman theorem. The entropy of a source is a parameter describing the way in which long but finite subsequences (blocks) appear in the signal. If we made a list of all k^N blocks of length N and assigned to each of them its frequency, then entropy would inform us how far this distribution is from the uniform distribution $\left(\frac{1}{k^N}, \dots, \frac{1}{k^N}\right)$. Maximal entropy $\log k$ means that we observe a totally random signal, which we are not able to unscramble. On the other hand, zero entropy reveals simplicity of the language (e.g. just one block is repeated again and again). To illustrate this notion: a speech of an active party member is of high entropy, if the speaker often surprises us by his arguments. While listening to a low entropy speech, we can predict what the speaker is going to say in the future.

Entropy also has a very practical application. Consider a computer data file x having the form of a long sequence of 0's and 1's. Denote its length by n and divide it into equal pieces of length N, so called *words*. Assume for simplicity that n is a multiple of N, $n = mN$. We can think of x as a sequence $\underline{x}$ having length m and consisting of symbols contained in the set $\mathcal{B}$ of all possible words B over 0 and 1 of length N. To each such word we assign its frequency in $\underline{x}$, denoted by $p(B)$, i.e. the number of appearances divided by m. We take the entropy of the sequence (file) x, or rather the Nth approximation of entropy of x, to be the quantity:

$$h(x) = -\frac{1}{N} \sum_{B \in \mathcal{B}} p(B) \log_2 p(B)$$

(in information theory one uses $\log_2$ instead of the natural logarithm—the maximal value of entropy is now equal to 1). There exists an algorithm of data compression, which allows to map injectively all sequences of length n into sequences of length approximately $nh(x)$. So we can interpret the entropy also as a quantity describing the degree of optimal compression. Such algorithm can be obtained as follows, in two neighbouring columns write: in the first one—all sequences of length n in order of increasing entropy, and in the second—all binary words in order of their length (shorter words are put on top of the column). To every sequence x we assign its neighbour from the same row, till no x's are left. It can be shown that this is the desired algorithm of compression. Notice, that one cannot expect any better data compression, since the number of all sequences having constant entropy h is equal more or less to 2^{mNh}, which is the number of all sequences of length nh. One has to be aware of the fact that globally we have hardly made any compression, since the majority of sequences reach maximal entropy and the gain on the required memory is minimal. However, this majority of sequences are completely useless as (uncompressed) computer files. Most computer files are highly ordered, so their entropy is low and the compression is effective. After the compression they become much shorter and usually have maximal entropy, so repeating of the algorithm would not be effective.

Let us end this section with a theorem due to Ornstein and Weiss. When we observe a process of entropy h, then the trajectories of almost all x behave in the following way: an initial word of length N does not occur again for a time of order $\exp(Nh)$. This statement clarifies the famous paradox lying in the inconsistency between the second law of thermodynamics and one of the basic theorems of ergodic theory—the Poincaré recurrence theorem. According to this theorem, an isolated system returns spontaneously to a state close to its initial condition (including states of low entropy). An explanation is easy: no physical system is ideally isolated, so the corresponding process is not precisely stationary. A theoretical stationary process with positive entropy returns to the original state after a time long enough, so that the real evolution of the system is drawn away from the theoretical one. Thus, Poincaré recurrence theorem does not work for complex physical systems.

4. Entropy in topological dynamics

A topological dynamical system consists of the topological space X and a continuous transformation $T : X \to X$. We will follow the usual assumptions of X being compact and metric. Topological entropy was introduced by Roy Adler, A.G. Konheim and M.H. McAndrew in 1965 (see [AKM]). Defining the entropy we replace partitions by open covers. Let $\mathcal{A}$ be an open cover of X and denote by $N(\mathcal{A})$ the smallest cardinality of a subcover of $\mathcal{A}$. Because of the lack of a measure, we treat every element of such a minimal subcover as "equally probable" obtaining:

Definition 4.1. *The entropy of an open cover* $\mathcal{A}$ is given by

$$H_{\mathrm{top}}(\mathcal{A}) = \log N(\mathcal{A}).$$

Similarly to the measure-theoretical case, one defines the refinement of open covers $\mathcal{A}$, $\mathcal{B}$ by the formula $\mathcal{A} \vee \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$. The appropriate order relation may also be introduced for covers. Namely, $\mathcal{A} \preccurlyeq \mathcal{B}$ if for every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that $A \supset B$. Just as before $T^{-1}\mathcal{A}$ (and $T\mathcal{A}$ in case of homeomorphism) is an open cover of X. We will also use the symbols $\mathcal{A}^n = \mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-(n-1)}\mathcal{A}$ and $\mathcal{A}^n_m = T^{-m}\mathcal{A} \vee \cdots \vee T^{-(n-1)}\mathcal{A}$.

Definition 4.2. *The entropy of the transformation T with respect to a cover* $\mathcal{A}$ is defined by the formula

$$h(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H_{\mathrm{top}}(\mathcal{A}^n).$$

The existence of the above limit follows from the facts analogous to those in Proposition 2.3.

Definition 4.3. *The topological entropy* $h_{\mathrm{top}}(T)$ of the transformation T is equal to the supremum of $h(T, \mathcal{A})$ taken over the set of all open covers of X.

It is worth mentioning that in 1971 Rufus Bowen introduced entropy based on Kolmogorov's and Tihomirov's concepts of ε-entropy and ε-capacity of sets (see [Bo], [KT]) in the case of metric space (suppressing the compactness assumption) with uniformly continuous map T. His definition is equivalent to the one above when restricted to the compact case and has an immediate interpretation as the measure of variety in the behavior of points (it somehow measures the rate of dispersal of trajectories starting at close points).

Topological dynamical systems are not scantier than measure-theoretical ones or stochastic processes. The classical Krylov-Bogoliubov theorem (see [KrB]) states that every such compact system has a non-empty set of Borel invariant measures (moreover, this set is a compact Choquet simplex). Hence, we can introduce an invariant measure into the topological system, obtaining a classical dynamical system or, which is even more common, treat it as a system admitting many invariant measures at the same time. Every such measure has its own measure-theoretical entropy. Note that the latter is often called metric entropy, though it has nothing to do with the metric structure of X. In that sense topological systems can be even richer than measure-theoretical ones.

The most important theorem relating both concepts of entropy is called the Variational Principle (1969, Goodwyn [Go]; 1970, Dinaburg [Di]; 1971 Goodman [G]).

Theorem 4.4. *(The Variational Principle)*
Let $T : X \to X$ be a continuous transformation of a compact metric space. Then

$$h_{\text{top}}(T) = \sup_{\mu\in\mathcal{M}_T} h_\mu(T),$$

where $\mathcal{M}_T$ denotes the set of all T-invariant measures on X and $h_\mu(T)$ stands for the metric entropy of the dynamical system (X, μ, T). Moreover, the supremum may be taken over the set containing only ergodic invariant measures.

If the topological system admits only one invariant measure, then the topological entropy equals the metric entropy.

Generally, topological entropy is much more "rough" than metric entropy, but sometimes it is easy to compute. For instance, let σ be the shift defined on the product space $\{0,1\}^{\mathbb{N}}$ by $(\sigma x)(n) = x(n + 1)$, let X be a shift invariant, closed and nonempty subset of $\{0,1\}^{\mathbb{N}}$. Then (X, σ) is a topological dynamical system (so called *a subshift* or *a symbolic system*). It follows from the definition of entropy that if $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ is an increasing sequence of finite covers such that for every finite cover $\mathcal{B}$ of X we have $\mathcal{B} \preccurlyeq \mathcal{A}_n$ for some index n then

$$h_{\text{top}}(\sigma) = \lim_{n\to\infty} h(\sigma, \mathcal{A}_n).$$

Let $\mathcal{A}_n$ consist of the n-dimensional cylinder sets of the form

$$\{x \in X : x_{-n} = i_{-n}, \dots, x_0 = i_0, \dots, x_n = i_n\},$$

where $i_{-n}, i_{-n+1}, \dots, i_n \in \{0,1\}$. Clearly, the sequence $\{\mathcal{A}_n\}$ satisfies the preceding assumptions. Notice that taking $\mathcal{A} = \mathcal{A}_0$, we obtain $\mathcal{A}_n = \mathcal{A}_{-n}^{n}$. In the forthcoming calculation we will use the following properties of entropy:

Lemma 4.5. (1) $\mathcal{A} \preccurlyeq \mathcal{B} \Rightarrow h(\sigma, \mathcal{A}) \leqslant h(\sigma, \mathcal{B})$,
(2) $H_{\text{top}}(\sigma^{-1}\mathcal{A}) = H_{\text{top}}(\mathcal{A}) = H_{\text{top}}(\sigma\mathcal{A})$,
(3) $H_{\text{top}}(\mathcal{A} \vee \mathcal{B}) \leqslant H_{\text{top}}(\mathcal{A}) + H_{\text{top}}(\mathcal{B})$.

For every n we have:

$$\begin{aligned} h(\sigma,\mathcal{A}) &\leqslant h(\sigma,\mathcal{A}_n) = \lim_{k\to\infty}\frac{1}{k}H_{\mathrm{top}}((\mathcal{A}_n)^{k-1}) \\ &= \lim_{k\to\infty}\frac{1}{k}H_{\mathrm{top}}(\mathcal{A}_{-n}^{n+k-1}) \\ &= \lim_{k\to\infty}\frac{1}{k}H_{\mathrm{top}}(\mathcal{A}_{-n}^{n-1}\vee\mathcal{A}_{n}^{n+k-1}) \\ &\leqslant \lim_{k\to\infty}\frac{1}{k}\left[H_{\mathrm{top}}(\mathcal{A}_{-n}^{n-1})+H_{\mathrm{top}}(\mathcal{A}_{n}^{n+k-1})\right] \\ &= \lim_{k\to\infty}\frac{1}{k}H_{\mathrm{top}}(\mathcal{A}^{k-1}) = h(\sigma,\mathcal{A}). \end{aligned}$$

Hence the entropy of σ equals

$$h_{\mathrm{top}}(\sigma) = \lim_{n\to\infty} h(\sigma,\mathcal{A}_n) = \lim_{n\to\infty} h(\sigma,\mathcal{A}) = h(\sigma,\mathcal{A}) = \lim_{m\to\infty}\frac{H_{\mathrm{top}}(\mathcal{A}^{m-1})}{m}.$$

Since $\mathcal{A}^{m-1}$ is the cover of X consisting of all cylinder sets having fixed its m first coordinates, the number of nonempty elements of $\mathcal{A}^{m-1}$ is equal to the number of blocks of length m appearing in elements of X. Hence the entropy of the system (X,σ) is given by the formula

$$h_{\mathrm{top}}(\sigma) = \lim_{m\to\infty}\frac{\log\operatorname{card}(\mathcal{B}_m(X))}{m},$$

where a block B belongs to $\mathcal{B}_m(X)$ if and only if B consists of m symbols and occurs in some $x\in X$.

We conclude this survey by mentioning some relation between topological entropy and chaos, more precisely, Li-Yorke chaos. Li-Yorke chaos is defined as the existence of an uncountable subset D of X, such that every pair of its points come arbitrarily close to each other and go far from each other infinitely many times during the evolution of a process. Formally, for any different $x,y\in D$ the following holds: $\liminf_i \rho(T^i x, T^i y) = 0$ and $\limsup_i \rho(T^i x, T^i y) > 0$. The connection between topological entropy and Li-Yorke chaos has been studied for quite a long time. It is well known that there exist chaotic systems of zero entropy. It shows that the kinds of randomness exist that are not recognizable by entropy. The question about the opposite implication was open for a long time and it was solved recently by Blanchard, Glasner, Kolyada and Maass ([BGKM]): positive topological entropy always implies Li-Yorke chaos.

The authors would like to emphasize that this paper is only a brief review of the topic. For the full course of the theory of entropy see any of the classical monographs on ergodic theory, e.g. [DGS], [P], [Sm], [W].

References

[AKM] Adler, R.L., Konheim, A.G., McAndrew, M.H., *Topological entropy,* Trans. Amer. Math. Soc. 114 (1965), 309–319

[B] Boltzmann, L., 1877, Wiener. Ber. 75, 67 and Wiener. Ber. 76, 373

[BGKM] Blanchard, F., Glasner, E., Kolyada, S., Maass, A., *On Li-Yorke pairs,* preprint.

[Bo] Bowen, R., *Entropy for group endomorphisms and homogeneous spaces,* Trans. Amer. Math. Soc. 153 (1971), 401–413

[Br] Breiman, L., *The individual ergodic theorem of information theory,* Ann. Math. Stat 28 (1957), 809–811. Correction, ibid. 31 (1960), 809–810

[C] Clausius, R., 1865, Ann. Phys. 125, 353

[D] Downarowicz, T., *Entropia,* Matematyka—Społeczeństwo—Nauczanie 27(2001), Zeszyty XXIV Szkoły Matematyki Poglądowej, Grzegorzewice.

[DGS] Denker, M., Grillenberger, C., Sigmund, K., *Ergodic Theory on Compact Spaces,* Lecture Notes in Mathematics 527, Springer-Verlag, New York, 1976

[Di] Dinaburg E.I., *The relation between topological entropy and metric entropy,* Dokl. Akad. Nauk SSSR 190 (1970), 13–16

[G] Goodman, T.N.T., *Relating topological entropy and measure entropy,* Bull. London Math. Soc. 3 (1971), 176–180

[Go] Goodwyn, L.W., *Topological entropy bounds measure-theoretic entropy,* Proc. Amer. Math. Soc. 23 (1969), 679–688

[K] Kolmogorov, A.N. *A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces,* Dokl. Akad. Nauk SSSR 119 (1958), 861–864

[KT] Kolmogorov, A.N., Tikhomirov, V.A. *ε-entropy and ε-capacity of sets in functional spaces,* Uspekhi Mat. Nauk.SSSR 14 Nr.2 (1959), 3–86

[Kr] Krieger, W., *On entropy and generators of measure-preserving transformations,* Trans. Amer. Math. Soc. 149 (1970), 453–464

[KrB] Krylov, N., Bogoliouboff, N., *La Théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire,* Ann. of Math. 38 (1937), 65–113

[M] McMillan, B., *The basic theorems of information theory,* Ann. Math. Stat 24 (1953), 196–219

[O1] Ornstein, D., *Bernoulli shifts with the same entropy are isomorphic,* Adv. in Math. 4 (1970), 337–352

[O2] Ornstein, D., *Two Bernoulli shifts with infinite entropy are isomorphic,* Adv. in Math. 5 (1970), 339–348

[P] Petersen, K., *Ergodic Theory,* Cambridge Univ. Press, Cambridge, 1989

[S] Shannon, C., *A mathematical theory of communination,* Bell System Tech. J. 27 (1948) 379–423, 623–656

[Si] Sinai, Ya.G., *The notion of entropy of a dynamical system,* Dokl. Akad. Nauk SSSR 125 (1959), 768–771

[Sm] Smorodinsky, M., *Ergodic Theory, Entropy,* Lecture Notes in Mathematics 214, Springer-Verlag, New York, 1971

[W] Walters, P., *An Introduction to Ergodic Theory,* Springer-Verlag, New York, 1982

Institute of Mathematics, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

E-mail address: `downar@im.pwr.wroc.pl`