The first half century of entropy: the most glorious number in dynamics

A. Katok
Penn State University

This is an expanded version of the invited talk given on June 17, 2003 in Moscow at the conference "Komogorov and contemporary mathematics".

The word *entropy*, amalgamated from the greek words *en*ergy and *trop*os (meaning "turning point"), was introduced in an 1864 work of Clausius, who defined the change in entropy of a body as heat transfer divided by temperature and postulated that overall entropy does not decrease, the second law of thermodynamics.

Entropy was clarified by Botlmann who built on work of Maxwell. One can describe the state of a gas of particles with finitely many data by partitioning the phase space into finitely many pieces and denoting the proportion of particles in the ith piece by p_i. Then it turns out that the most probable states can be found by maximizing $\sum_i p_i \log p_i$ (discrete Maxwell–Boltzmann law).

Prehistory (entropy in information theory)

1948–50 Shannon information theory

Shannon considered finite alphabets whose symbols have known probabilities p_i and, looking for a function to measure the uncertainty in choosing a symbol from among these, determined that up to scale $\sum_i p_i \log p_i$ is the only continuous function of the p_i that increases in n when $(p_1, \ldots, p_n) = (1/n, \ldots, 1/n)$ and behaves naturally with respect to making successive choices.

Accordingly, the *entropy* of a finite or countable measurable partition ξ of a probability space is given by

$$H(\xi) =: H_\mu(\xi) =: -\sum_{C \in \xi} \mu(C) \log \mu(C) \geq 0,$$

where $0 \log 0 =: 0$. For countable ξ the entropy may be infinite.

1956 Khinchin's Uspehi survey.

Khinchin gave a very elegant rigorous treatment of information theory with the entropy of a stationary random process as a centerpiece. He developed the basic calculus of entropy which in retrospect looks as a contemporary introduction to the subject of entropy in ergodic theory. For a measure preserving transformation T define the *joint partition* by

$$\xi_{-n}^{T} =: \bigvee_{i=1}^{n} T^{1-i}(\xi),$$

where $\xi \vee \eta =: \{C \cap D : C \in \xi,\, D \in \eta\}$. Now

$$h(T,\xi) =: h_\mu(T,\xi) =: \lim_{n\to\infty} H(\xi_{-n}^{T})/n$$

is called the *metric entropy* of the transformation T *relative* to the partition ξ. (It is easy to see that the limit exists.)

Via *c*onditional entropies

$$H(\xi|\eta) =: -\sum_{D\in\eta}\mu(D)\sum_{C\in\xi}\mu(C|D)\log\mu(C|D),$$

where $\mu(A|B) =: \mu(A\cap B)/\mu(B)$, the entropy of a finite state random process (or, equivalently, the entropy of a measure-preserving transformation with respect to a given finite partition) can also be defined as the average amount of information obtained on one step given complete knowledge of the past (meaning the sequence of partition elements to which preimages of a given point belong), i.e.

$$h(T,\xi) = H(\xi|\xi^{T}_{-\infty})$$

.

Kolmogorov's discovery 1958–59

The same transformation can be coded by many different partitions and entropies of the corresponding random processes may differ. Kolmogorov realized that this can be used to define a quantity that describes the *intrinsic complexity* of a measure-preserving transformation. i.e. an *invariant*.

A partition ξ with finite entropy is a *generator* for a measure-preserving transformation T if the set of partitions subordinate to some $\bigvee_{i=-n}^{n} T^{-i}(\xi)$ is dense in the set of finite entropy partitions endowed with the *Rokhlin metric* $d_R(\xi,\eta) =: H(\xi|\eta) + H(\eta|\xi)$. Kolmogorov noted that all generators for a measure preserving transformation T have the same entropy and defined the *e*ntropy of T to be this common value if T has a generator and ∞ otherwise.

Development of basic theory

1958 Kolmogorov's original work

1959 Sinai found a natural way to make this notion better behaved by observing that generators maximize entropy among partitions and defining the entropy of T as

$$h(T) =: h_\mu(T) =: \sup\{h(T,\xi) : H(\xi) < \infty\}.$$

1960 Rokhlin's Uspehi survey.

Summary of the first explosive phase of development (Sinai, Abramov, Rokhlin, Pinsker in Moscow; also Adler, Parry in the West). Calculations for many examples.

1967 Rokhlin's lectures (written several years earlier) present the core of entropy theory in a definitive form; serve as the model for later textbooks and monographs.

K (Kolmororov)–systems and π (Pinsker)–partitions:

K–property, also introduced by Kolmogorov in 1958, is an isomorphism invariant version of earlier regularity notions for random processes: present becomes asymptotically independent of a all sufficiently long past. It was proved to be equivalent to *completely positive entropy*:

$$h(T,\xi) > 0 \text{ for any partition } \xi \text{ with } H(\xi) > 0.$$

The *π–partition* is the crudest partition (minimal σ–algebra) which refines every partition with zero entropy.

Both are the additional key isomorphism invariants stemming from the notion of entropy.

Entropy and isomorphism: from Kolmogorov's work through the seventies

Kolmogorov's stated motivation for the introduction of entropy was to provide a new isomorphism invariant for measure preserving transformations and flows, more specifically to split those with countable Lebesgue spectrum into continuum non-isomorphic classes. In particular *Bernoulli shifts* (independent stationary random processes) with different entropies (such as $(1/2, 1/2)$ and $(1/3, 1/3, 1/3)$) are not isomorphic.

Two new central problems were formulated:

- Are Bernouill shifts *with the same entropy* isomorphic?
- Are K–systems with the same entropy isomorphic?

Isomorphism of Bernoulli shifts:

1959–63	Early special cases: Meshalkin and Blum–Hansen. Example: Bernoulli shifts $(1/4, 1/4, 1/4, 1/4)$ and $(1/2, 1/8, 1/8, 1/8, 1/8)$ are isomorphic.
1962–64	Sinai: Any two Bernoulli shifts with equal entropy are *weakly isomorphic*: each is isomorphic to a *factor* of the other.
1967	Ornstein: Isomorphism of Bernoulli shifts with equal entropy; end of the dominance of the Moscow school.
1968–72	Ornstein and collaborators: development of isomorphism theory. Efficient necessary and sufficient conditions for isomorphism to a Bernoulii shift (very weak Bernoulli).

Amazing consequences of Ornstein's work: Bernoulli flows, group automorphisms, tremendous variety of smooth systems as well as systems coming from probability and number theory. Bernoulli structure is very soft and seems to be everywhere.

Huge gap between Bernoulii and K–systems: Not only there are many nonisomorphic K–systems with the same entropy but K property can be achieved in unexpected ways, e.g. by changing time in any ergodic positive entropy flow inducing (taking the first return map of any ergodic positive entropy transformation on a certain set (Ornstein, Rudolph, B. Weiss).

Topological entropy and thermodynamical formalism

1964	Parry: Maximal entropy measures for topological Markov chains (subshifts of finte type).
1965	Adler et al: Topological entropy.
1968–70	Variational principle for entropy $$h_{top} = \sup_\mu h_\mu$$ (Goodwin, Dinaburg, Goodman).
1967–70	Maximal entropy measures for hyperbolic systems (Sinai, Margulis, Bowen).

Comparison of entropies:

Topological entropy represents the exponential growth rate of the *total number* of orbit segments distinguishable with arbitrarily fine but finite precision and describes in a crude but suggestive way the total exponential complexity of the orbit structure with a single number.

In the case of an ergodic measure, metric (Kolmogorov) entropy can be characterized as the exponential growth rate for the number of *statistically significant* distinguishable orbit segments. Note that this is clearly never more than the topological entropy.

Maximal entropy measures:

Parry construction: uniform distribution of cylinders.

Sinai construction: reduction to topological Markov chains via *Markov partitions*.

Bowen constuction: asymptotic distribution of periodic points/orbits.

Margulis construction: product of asymptotic distributions of volumes for stable & unstable manifolds.

1972–75 Thermodynamical formalism.

Variational principle for *pressure* for the potential φ:

$$P_\varphi = \sup_\mu (h_\mu + \int \varphi d\mu)$$

(Walters, Misiurewicz).

Theory of Gibbs measures & equilibrium states for Hoelder potentials (Sinai, Ruelle). The Gibbs measure μ_φ is unique and is characterized by

$$P_\varphi = (h_{\mu_\varphi} + \int \varphi d\mu_\varphi).$$

Gibbs measures are constructed via *weighted* versions of Parry, Sinai, Bowen and Margulis methods. SRB (Sinai, Ruelle, Bowen) measures for Anosov systems & hyperbolic attractors are special Gibbs measures with the logarithm of the Jacobian in the unstable (stable) direction as the potential.

Entropy and non-uniform hyperbolicity

1975–77 Pesin theory:

Entropy, Lyapunov characteristic exponents and stochastic behavior. Pesin entropy formula:

$$h_\mu = \int \sum \chi_i^+ d\mu$$

(μ absolutely continuous).

1980–82 A. Katok: Periodic points, homoclinic points and horseshoes. In dimension 2:

$$h_{top} \leq \limsup_{n\to\infty} \frac{\log P_n(f)}{n}.$$

$P_n(f) = \#\{$periodic points for f of period n.$\}$

Upper semicontinuity of h_{top}.

Conclusions from Pesin theory: Hyperbolicity is the only source of entropy in smooth systems. For an absolutely continuous invariant measure entropy gives exact rate of infinitesimal exponential expansion. π partitions are characterized through families of stable and unstable manifolds. In particular, if exponents do not vanish (hyperbolic measures) the system has at most countably many ergodic components and they are Bernoulli up to a finite permutation. In particular, weak mixing, mixing, K, and Bernoulli properties are all equivalent.

On the topological side: Any (non-uniformly) hyperbolic *measure* can be approximated in various senses by uniformly *hyperbolic sets*; in particular, their topological entropies approximate the entropy of the measure.

1985 Ledrappier–Young: Entropy and dimension,

$$h_\mu = \sum d_i \chi_i^+ \ (0 \le d_i \le 1), \quad (\mu \text{ ergodic})$$

converse to the Pesin entropy formula:

$$d_i = 1 \Leftrightarrow \mu \text{ SRB type measure.}$$

1988–89 Yomdin, Newhouse: Entropy and volume growth.

Lower semicontinuity of entropy for C^∞;

continuity of h_{top} in dim 2.

Implies *Shub Entropy Conjecture* for C^∞ maps:

$$h_{top}(f) \ge \log s(f_*);$$

$s(f_*)$ is the spectral radius of the induced map in the total homology.

Entropy, asymptotics of periodic orbits and rigidity

1970 Margulis: Multiplicative asymptotic.

Mixing Anosov flows (e.g. geodesic flows on compact manifolds of (variable) negative curvature).

$$\lim_{t\to\infty} P_t \cdot e^{-t\cdot h_{top}} \cdot t \cdot h_{top} = 1;$$

$P_t = \#\{\text{periodic orbits of period} \le t\}$.

1982 A. Katok: Conformal estimates for surfaces of genus ≥ 2:

$$h_{top} \geq \rho\left(\frac{-2\pi E}{v}\right)^{1/2}, \quad h_\lambda \leq \rho^{-1}\left(\frac{-2\pi E}{v}\right)^{1/2};$$

E – Euler characteristic, v – area,

$\rho \geq 1$ – conformal coefficient, λ–the Liouville measure.

Entropy rigidity for surfaces of genus ≥ 2:

$$h_{top} = h_\lambda \Leftrightarrow \text{constant curvature.}$$

Conjecture: For a negatively curved metric

$$h_{top} = h_\lambda \Leftrightarrow \text{locally symmetric space.}$$

1993 Besson–Courtois–Gallot:

Rigidity of topological entropy for locally symmetric spaces of rank one: locally symmetric metrics strictly minimize h_{top} among metrics of fixed volume.

Later Flaminio: h_λ may increase for fixed volume perturbations of constant curvature metrics in dim 3.

1996–2000 Knieper: Multiplicative bounds for manifolds of non-positive curvature of geometric rank one

$$c_1 t^{-1} e^{t h_{top}} \le P_t \le c_2 t^{-1} e^{t h_{top}}.$$

Uniqueness of maximal entropy measure.

1998 Dolgopyat: Exponential error term for mixing Anosov flows with C^1 foliations (e.g. geodesic flows on surfaces of negative curvature).

2002 Gunesch: Margulis (multiplicative) asymptotic for rank one manifolds of non-positive curvature

$$\lim_{t\to\infty} P_t \cdot e^{-t \cdot h_{top}} \cdot t \cdot h_{top} = 1.$$

Actions of $\mathbb{Z}_+^k, \mathbb{Z}^k, \mathbb{R}^k,\ k \geq 2$; rigidity of positive entropy measures

Basic examples:

• $\times 2$, $\times 3$ (Furstenberg, 1967):

$$E_2 : S^1 \to S^1 \quad x \mapsto 2x, \quad (\mod 1)$$

$$E_3 : S^1 \to S^1 \quad x \mapsto 3x, \quad (\mod 1).$$

• Commuting toral automorphisms: $A, B \in SL(3, \mathbb{Z})$, $AB = BA$, $A^k = B^l \Rightarrow k = l = 0$, A, B hyperbolic. The $\mathbb{Z}^2$ action generated by automorphisms of the torus $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$:

$$F_A : x \mapsto Ax, \quad (\mod 1)$$

$$F_B : x \mapsto Bx, \quad (\mod 1).$$

• Weyl chamber flow (WCF): $M = SL(n, \mathbb{R})/\Gamma$, $n \geq 3, \Gamma$ a lattice in $SL(n, \mathbb{R})$, D positive diagonals isomorphic to $\mathbb{R}^{n-1}$. WCF:

The action of D on M by left translations.

Entropy function for an A action α; $\mathbb{Z}_+^k, \mathbb{Z}^k, \mathbb{R}^k$:

μ– an α-invariant measure; by the natural extension and suspension can always reduce the situation to the case of $\mathbb{R}^k$.

- *Lyapunov exponents:* $\chi_i \in (\mathbb{R}^k)^*$;
- *Lyapunov hyperplanes:* $\operatorname{Ker} \chi_i$;
- *Weyl chambers.* Connected components of $\mathbb{R}^k \setminus \bigcup_i \operatorname{Ker} \chi_i$;
- *Entropy function:*

$$h(t) =: h_\mu(t) =: h_\mu(\alpha(t)).$$

Properties of the entropy function:

- $h(\lambda t) = |\lambda| \cdot h(t)$;
- $h(t+s) \leq h(t) + h(s)$;
- h is linear in each Weyl chamber.

Follow from Ledrappier–Young entropy formula.

1989 Rudolph: $\mu \times 2,\ \times 3$ invariant, ergodic, h_μ does not vanish $\Rightarrow\ \mu$ Lebesgue.

1992-95 A.Katok–Spatzier: Geometric approach, rigidity results for algebraic actions including basic examples, see also Kalinin–A.Katok, 2001.

1999 A. Katok: First measure rigidity results for the non–algebraic non–uniform case

2002 A. Katok, S.Katok, K. Schmidt: $\mathbb{Z}^2$ actions by Bernoulli automorphisms; commuting toral automorphisms with identical entropy functions, weakly isomorphic but not isomorphic (contrast with Sinai, Ornstein).

2003 E. Lindenstrauss: Arithmetic quantum unique ergodicity via measure rigidity via measure rigidity for actions of higher rank abelian groups.

Uses methods and results from earlier work of Ratner.

2003 Einsiedler, A. Katok, E. Lindenstrauss, A partial solution of the Littlewood conjecture in Diophantine approximation via measure rigidity.

Theorem: μ on $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$ is WCF invariant, ergodic, h_μ does not vanish $\Rightarrow$ μ Haar.

Corollary: Let $(u,v) \in \mathbb{R}^2$, $\langle x \rangle = \operatorname{dist}(x,\mathbb{Z})$

$$C(u,v) := \liminf_{n\to\infty} n\langle nu\rangle\langle nv\rangle = 0$$

except for maybe a set of Hausdorff dim 0.

(Littlewood conjecture: $C(u,v) = 0$).