Entropy in Dynamical Systems & Ergodic Theory (A little Glimpse)

Jean-René Chazottes

Centre de Physique Théorique, CNRS & Ecole polytechnique

About Entropy in Large Classical Particle Systems

Clay Institute, 25 september 2023

Introduction

In the context of dynamical systems :

Entropy measures the ***rate of increase in dynamical complexity*** as the system evolves with ***time***.

Entropy is an **invariant under isomorphism** of measure-preserving dynamical systems.

This talk is aimed at nonexperts.

Plan

- Dynamical systems and ergodic theory in a nutshell
- Kolmogorov-Sinai entropy
- Shannon-McMillan-Breiman theorem
- Brin-Katok formula
- Ornstein's theorem(s)

Slides that I will not have time to show, but that you would have liked to see 😉 :

- Entropy via recurrence times (Ornstein-Weiss)
- Topological entropy
- Equilibrium states
- Margulis-Ruelle inequality and Pesin's formula
- Brudno's theorem (Kolmogorov complexity of orbits)
- Krieger's generator theorem
- Entropy is the only finitely observable invariant
- Weak Pinsker conjecture (solved in 2018)

DYNAMICAL SYSTEMS

AND

ERGODIC THEORY

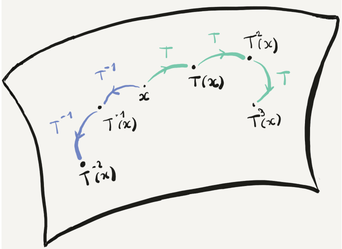
IN A NUTSHELL

Dynamical systems (in this talk)

A dynamical system is :

- A (finite-dimensional) state space X ;
- A map $T : X \to X$ (evolution rule).

T can be invertible or non-invertible.



Discrete-time, *deterministic* evolutions, *e.g.*, in physics and biology.

Models to understand 'chaotic' systems.
('Chaotic' means sensitivity to initial condition *and* strong recurrence.)

Measure-preserving dynamical system :

- $(X, \mathfrak{F}, \mu)$ is a (standard) probability space ;
- $T : X \to X$ a (measurable) map such that $\mu(T^{-1}E) = \mu(E)$, $\forall E \in \mathfrak{F}$.

▷ A general *existence* result : when X is a *compact* topological space and T is continuous, there exists at least one T-invariant measure (Krylov-Bogolioubov).

▷ *Some invariant measures for free* :
If X contains a p-periodic point x ($\exists p \geq 1$ s.t. $T^p(x) = x$), then $\frac{1}{p}\sum_{i=0}^{p-1}\delta_{T^i(x)}$ is T-invariant ('periodic-orbit measure').

(Nota bene : in this talk : measure = probability measure.)

Suppose μ is ergodic, that is, if $T^{-1}(A) \overset{(\mathrm{mod}\,\mu)}{=} A$, then either $\mu(A) = 0$ or $\mu(A) = 1$ (or, equivalently, if $f \circ T \overset{(\mathrm{mod}\,\mu)}{=} f$, then $f \overset{(\mathrm{mod}\,\mu)}{=}$ const, for any measurable $f : X \to \mathbb{R}$).

(Ergodic measures are the 'building blocks' of T-invariant measures.)

Theorem.

Let $(X, \mathfrak{F}, \mu)$ be ergodic.
Let $g : X \to \mathbb{R}$ be in $\mathcal{L}^1(\mu)$.
Then, for μ-almost every x (and also in $\mathcal{L}^1(\mu)$),

$$\frac{1}{N}\sum_{n=0}^{N-1} g(T^n(x)) \xrightarrow[N\to+\infty]{} \int_X g\,\mathrm{d}\mu.$$

In particular, the mean sojourn time of μ-almost every x to $E \subset X$ exists :

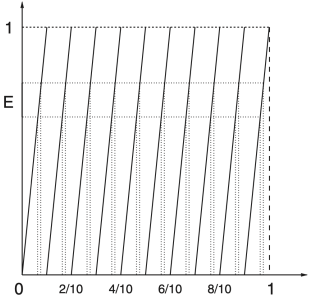
$$\lim_{N\to+\infty} \frac{1}{N}\sharp\{0 \le i < 1 : T^i(x) \in E\} = \mu(E).$$

(A **stronger property** is mixing : $\mu\left(E \cap T^{-N}E'\right) \xrightarrow[N\to+\infty]{} \mu(E)\,\mu(E'), \forall E, E' \in \mathfrak{F}$.)

Decimal expansion as a dynamical system :

$X = [0, 1]$
$T(x) = 10x - \lfloor 10x \rfloor$
If $x = 0.a_0a_1a_2a_3\ldots$ with $a_i \in \{0, 1, 2, \ldots, 9\}$ then

$$T^n(x) = 0.\cancel{a_0 \ldots a_{n-1}}\, a_n a_{n+1} a_{n+2} \ldots$$



(Prototype of a uniformly expanding map.)

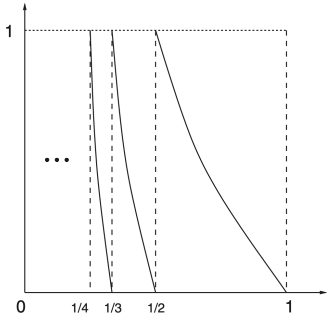
Lebesgue measure is invariant and mixing.

Gauss map and continued fractions :

$X = [0, 1]$
$T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$ if $x \in (0, 1]$, $T(0) = 0$.
For irrational $x \in (0, 1)$,

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n + T^n(x)}}}}$$

where $a_i = \lfloor 1/T^{i-1}(x) \rfloor \geq 1$ for $i \leq n$.



$\mu(E) = \frac{1}{\log 2} \int_E \frac{1}{1+x} \mathrm{d}x$
for any measurable $E \subset [0, 1]$.

This dynamical system is mixing.

Circle rotations :

$X = S^1 = \mathbb{R}/\mathbb{Z}$
Given $\theta \in \mathbb{R}$, let $T_\theta(x) = x + \theta \pmod 1$.

If $\theta = p/q$ (rational), then $T_\theta^q = \mathrm{Id}$, so every orbit is periodic.
If θ is irrational, then every orbit is dense in S^1.

Lebesgue measure is T_θ-invariant.

If θ is *rational* then this system is **not ergodic**.
If θ is *irrational* then it is *ergodic* (but **not mixing**).

(Generalization : rotations on tori.)

Shift spaces, Bernoulli shifts, and Markov shifts :

Fix $k > 1$.
Let $X = \{1, \ldots, k\}^{\mathbb{N}}$ (one-sided sequences)
or $X = \{1, \ldots, k\}^{\mathbb{Z}}$ (two-sided sequences).
Given a one-sided or a two-sided sequence $x = (x_i)_i$, the shift map is

$$(T(x))_i = x_{i+1}.$$

The 'shift' map T is invertible when $X = \{1, \ldots, k\}^{\mathbb{Z}}$.

These shift spaces are compact topological spaces in the product topology and the (Borel) sigma-algebra is generated by cylinder sets

$$[s_m, s_{m+1}, \ldots, s_n] \stackrel{\text{def}}{=} \{x : x_i = s_i, m \leq i \leq n\}, s_i \in \{1, \ldots, k\}, m < n \text{ in } \mathbb{N} \text{ or } \mathbb{Z}.$$

(Full shifts are the simplest example of 'symbolic dynamical systems'.)

Bernoulli shifts :

Take $X = \{1, \dots, k\}^{\mathbb{N}}$ and put a measure $(p_1, \dots, p_k)$ on $\{1, \dots, k\}$. Then let $\mu \stackrel{\text{def}}{=} (p_1, \dots, p_k)^{\mathbb{N}}$ (product measure); it is obviously shift-invariant. Every Bernoulli shift is mixing.

Markov shifts (a.k.a. finite-state space Markov chains) :

Consider an *irreducible* stochastic matrix $Q : \{1, \dots, k\}^2 \to \{0, 1\}$. There is a *unique* $q = (q_1, \dots, q_k)$ such that $qQ = q$ and the measure

$$\mu([s_0, \dots, s_{n-1}]) \stackrel{\text{def}}{=} q_{s_0} Q_{s_0, s_1} \cdots Q_{s_{n-2}, s_{n-1}}$$

is shift-invariant and ergodic. It is mixing if Q is irreducible and aperiodic.

Remarks.

- There are plenty of other shift-invariant measures (*e.g.*, periodic-orbit measures) on these spaces!
- Generalization : Gibbs measures/equilibrium states on (sub-)shifts of finite type modelling 'uniformly hyperbolic systems' (Axiom A diffeomorphisms).

Linear automorphisms of the torus :

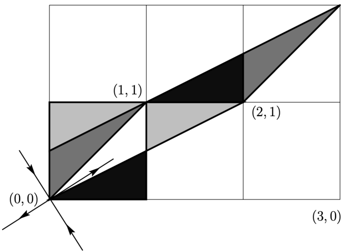
$X = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$

(unit square $[0, 1] \times [0, 1]$ with opposite sides identified)

$T(x_1, x_2) = (2x_1 + x_2, x_1 + x_2)$ mod 1.

It corresponds to the linear map of $\mathbb{R}^2$ given by the matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$



The image of the torus under *A*.

Eigenvalues : $\lambda = \left(3 + \sqrt{5}\right)/2 > 1$ and $1/\lambda < 1$,

so the map expands by a factor of λ in the direction of the eigenvector $v_\lambda = \left(1 + \sqrt{5}\right)/2, 1)$, and contracts by $1/\lambda$ in the direction of $v_{1/\lambda} = \left(1 - \sqrt{5}\right)/2, 1)$.

The determinant is 1.

This (invertible) dynamical system preserves the Lebesgue measure on $\mathbb{T}^2$ and is ergodic (and mixing).

Remark. This is a prototype of a *uniformly hyperbolic system.*

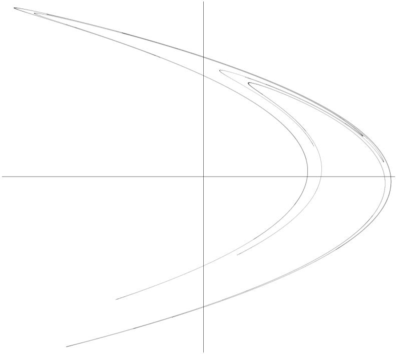
(Uniform expansion and contraction in complementary directions at every point.)

Hénon's map (attractor) :

$X = \mathbb{R}^2$, $T(x, y) = (1 - ax^2 + y, bx)$
where a, b are real parameters.

If $b \neq 0$ the map is invertible, and it changes area by a factor of $|b|$.
For, *e.g.*, $a = 1.4$, $b = 0.3$, there is a trapping region homeomorphic to a disk.

Benedicks and Carleson proved (hard proof!) that if $b \ll 1$, $\exists P_b \subset]2 - \epsilon, 2[$ s.t. $\mathrm{Leb}(P_b) > 0$, s.t. $\forall a \in P_b$, there is a T-invariant measure which is a Sinai-Ruelle-Bowen measure (which is mixing).

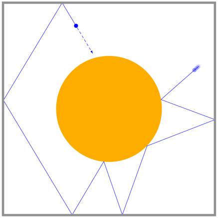


Hénon's attractor for $a = 1.4$, $b = 0.3$ ('fractal' structure).

(This is an example of a *non-uniformly* hyperbolic system.)

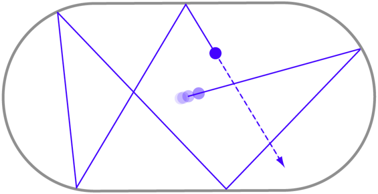
Planar billiards (maps and flows) :

Sinai's billiard



(mixes exponentially fast)

Bunimovich stadium billiard



(mixes very slowly)

(These are *non-uniformly* hyperbolic systems.)

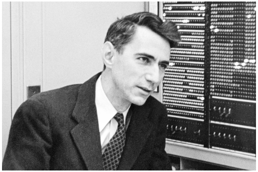
Now, entropy!

Let Z be a random variable taking values in $\{1, \ldots, \Bbbk\}$ ($\Bbbk$ finite and fixed).
Let $(p_1, \ldots, p_{\Bbbk})$ its law ($p_i \geq 0$, $\sum_{i=1}^{\Bbbk} p_i = 1$).

For instance, it can model an 'yes-or-no' random experiment with $p_1 = p$ and $p_2 = 1 - p$ (Bernouilli).

Then

$$H(p_1, \ldots, p_{\Bbbk}) \stackrel{\text{def}}{=} -\sum_{i=1}^{\Bbbk} p_i \log p_i.$$



Claude Shannon

Two extreme cases :

- if $p_i = 1$ for some i, then $H(p_1, \ldots, p_{\Bbbk}) = 0$;
- $H(p_1, \ldots, p_{\Bbbk}) \leq \log \Bbbk$ with equality if and only if $p_i = \frac{1}{\Bbbk}$ for $i = 1, \ldots, \Bbbk$.

Consider a stationary stochastic process $\mathcal{Y} = (Y_i)_{i \in \mathbb{N}\,(\text{or}\,\mathbb{Z})}$ where the random variables Y_i take values in $\{1, \ldots, k\}$.

Stationarity means :

$$\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}) = \mathbb{P}(Y_\ell = y_{0+\ell}, Y_{1+\ell}, \ldots, Y_{n-1} = y_{n-1+\ell})$$

for all n, ℓ and $(y_0, y_1, \ldots, y_{n-1}) \in \{1, \ldots, k\}^n$.

Shannon entropy of the joint law of $Y_0, \ldots, Y_{n-1}$:

$$-\sum_{(y_0,\ldots,y_{n-1})\in\{1,\ldots,k\}^n} \mathbb{P}(Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}) \log \mathbb{P}(Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}).$$

The Shannon entropy of the process is

$$h(\mathcal{Y}) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{\text{what precedes}}{n}.$$

Remark. $\mathcal{Y} = (Y_i)_{i \in \mathbb{N}\text{ or }\mathbb{Z}}$ is a measure-preserving dynamical system with $X = \{1, \ldots, k\}^{\mathbb{N}\,(\text{or}\,\mathbb{Z})}$, $T =$ shift map, a certain shift-invariant measure μ.

Reprinted with corrections from *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number. Doubling the time roughly squares the number of possible messages, or doubles the logarithm, etc.
2. It is nearer to our intuitive feeling as to the proper measure. This is closely related to (1) since we intuitively measures entities by linear comparison with common standards. One feels, for example, that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information.
3. It is mathematically more suitable. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities.

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits,* a word suggested by J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2^N and $\log_2 2^N = N$. If the base 10 is used the units may be called decimal digits. Since

$$\log_2 M = \log_{10} M / \log_{10} 2 \\ = 3.32 \log_{10} M,$$

[1]Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal,* April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A.I.E.E. Trans.,* v. 47, April 1928, p. 617.
[2]Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal,* July 1928, p. 535.

1st page of Shannon's 1948 paper

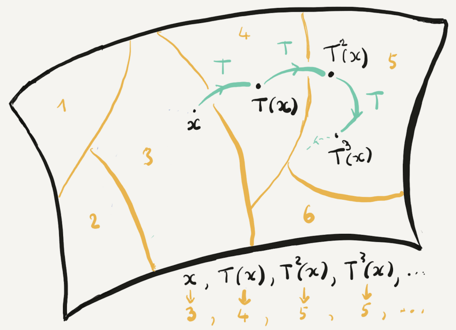
Famous anecdote (urban legend ?) of Shannon, retold so many times :

> *My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage."*

Kolmogorov reads Shannon

Basic observation :

A finite partition of a measure-preserving dynamical system generates a stationary stochastic process with values in the set $\{1, \ldots, k\}$ of labels of the partition.



Here $k = 6$.

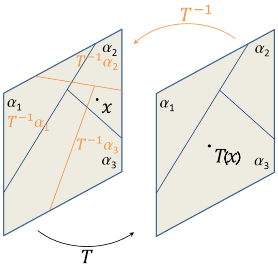
Kolmogorov in the 1940's

Let $(X, \mathfrak{F}, \mu, T)$ be a (non-invertible) measure-preserving dynamical system.
Let $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_k\}$ be a finite (measurable) partition of X ($k \geq 2$ fixed)
(that is, $\alpha_i \in \mathfrak{F}$, $X \stackrel{(\text{mod } \mu)}{=} \alpha_1 \cup \cdots \cup \alpha_k$, $\mu(\alpha_i \cap \alpha_j) = 0$ for $i \neq j$).

Given two partitions $\boldsymbol{\alpha}$ and $\widetilde{\boldsymbol{\alpha}}$, let

$$\boldsymbol{\alpha} \vee \widetilde{\boldsymbol{\alpha}} \stackrel{\text{def}}{=} \{\alpha \cap \tilde{\alpha} : \alpha \in \boldsymbol{\alpha}, \tilde{\alpha} \in \widetilde{\boldsymbol{\alpha}}\}.$$

So, if we let $T^{-1}\boldsymbol{\alpha} \stackrel{\text{def}}{=} \{T^{-1}\alpha_1, \ldots, T^{-1}\alpha_k\}$,
we can form the partition $\boldsymbol{\alpha} \vee T^{-1}\boldsymbol{\alpha}$.



$(k = 3)$

More generally, for each n, define the partition

$$\boldsymbol{\alpha}_n \stackrel{\text{def}}{=} \bigvee_{j=0}^{n-1} T^{-j}\boldsymbol{\alpha}$$

which are sets (atoms) of the form

$$\left\{x \in X : x \in \alpha_{i_0}, T(x) \in \alpha_{i_1}, \ldots, T^{n-1}(x) \in \alpha_{i_{n-1}}\right\}$$

for some $(i_0, i_1, \ldots, i_{n-1})$, where $i_m \in \{1, \ldots, k\}$.

Now let

$$H_\mu(\boldsymbol{\alpha}) \stackrel{\text{def}}{=} H\big(\mu(\alpha_1), \ldots, \mu(\alpha_k)\big).$$

Definition.
Let $(X, \mathfrak{F}, \mu, T)$ be a measure-preserving dynamical system.
Let $\boldsymbol{\alpha}$ be a finite partition.
The entropy of the system w.r.t. to $\boldsymbol{\alpha}$ is

$$h_\mu(T, \boldsymbol{\alpha}) \stackrel{\text{def}}{=} \lim_{n \to +\infty} \frac{1}{n} H_\mu\left(\bigvee_{j=0}^{n-1} T^{-j} \boldsymbol{\alpha}\right).$$

The entropy of the system then is

$$h_\mu(T) \stackrel{\text{def}}{=} \sup\big\{h_\mu(T, \boldsymbol{\alpha}) : \boldsymbol{\alpha} \text{ is a finite partition of } X\big\}.$$

Remarks.

The above limit always exists by Fekete's lemma, and $h_\mu(T, \boldsymbol{\alpha}) \geq 0$ for all $\boldsymbol{\alpha}$.

If T is invertible, then $h_\mu(T^n) = |n| \, h_\mu(T)$ for all $n \in \mathbb{Z}$.

Computing the sup over *all* partitions seems to be a formidable task, but...

Theorem.

Let $(X, \mathfrak{F}, \mu, T)$ be a measure-preserving dynamical system.
Let $\boldsymbol{\alpha}$ be a partition that is 'generating', that is :

$$\bigvee_{j=0}^{\infty} T^{-j}\boldsymbol{\alpha} \overset{(\mathrm{mod}\,\mu)}{=} \mathfrak{F}.$$

Y. Sinai in the 1970's

Then $h_\mu(T) = h_\mu(T, \boldsymbol{\alpha})$.

Let $\alpha_n(x)$ be the element of $\bigvee_{i=0}^{n-1} T^{-i}\boldsymbol{\alpha}$ which contains x.

Corollary.

Let X be a metric space and μ a Borel probability measure.
Let $\boldsymbol{\alpha}$ be a partition such that $\operatorname{diam} \alpha_n(x) \xrightarrow[n\to+\infty]{} 0$ for μ-almost every x.
Then $h_\mu(T) = h_\mu(T, \boldsymbol{\alpha})$.

Examples

(Note that, in general, there is no hope to compute explicitly the value of the entropy of a given measure-preserving dynamical system.)

1. The entropy of any periodic-orbit measure is zero.

2. $X = [0, 1]$, $T(x) = 10x \bmod 1$, μ is Lebesgue measure :
$\boldsymbol{\alpha}$ partition of $[0, 1]$ into $\left(\frac{j-1}{10}, \frac{j}{10}\right]$, $j = 1, \ldots, 10$.
$\boldsymbol{\alpha}_n$ partition of $[0, 1]$ into $\left(\frac{i-1}{10^n}, \frac{i}{10^n}\right]$, $i = 1, \ldots, 10^n$.

$$H_\mu(\boldsymbol{\alpha}_n) = -\sum_{i=1}^{10^n} 10^{-n} \log 10^{-n} = n \log 10,$$

hence $h_\mu(T) = \log 10$.

3. $X = (0, 1]$, $T(x) = \frac{1}{x} \bmod 1$ (Gauss map), $\mathrm{d}\mu(x) = \dfrac{1}{\log 2} \dfrac{\mathrm{d}x}{1+x}$.
Partition of $(0, 1)$ into the subintervals $(1/(m+1), 1/m)$, $m \geq 1$
(countably infinite partition).

$$h_\mu(T) = \int_0^1 \log |DT| \, \mathrm{d}\mu = \frac{\pi^2}{6 \log 2}.$$

4. Circle rotations have entropy 0.

($X = S^1$, $T_\theta(x) = x + \theta \bmod 1$, μ is Lebesgue measure.)

Proof (sketch) :

- If $\theta = p/q \in \mathbb{Q}$, then $T_\theta^q = \mathrm{Id}$, so $h_\mu(T_\theta) = (1/q) h_\mu(T_\theta^q) = (1/q) h_\mu(\mathrm{Id}) = 0$.
- If $\theta \in \mathbb{R} \backslash \mathbb{Q}$, then let $\alpha^{(N)}$ be a partition into N intervals of equal length. Then $\sharp \alpha_n = nN$, so $H_\mu(\alpha^{(N)}) \le \log(nN)$, and thus $h_\mu(T_\theta, \alpha^{(N)}) = \lim_n (\log(nN))/n = 0$.

  Therefore $h_\mu(T_\theta) = 0$, since the collection of partitions $\alpha^{(N)}$, $N \ge 1$, is generating.

5. Shift spaces on k symbols : one recovers Shannon entropy for stationary stochastic processes.

The partition into the 1-cylinders $[s]$, $s \in \{1, \ldots, k\}$, generates the sigma-algebra. In particular :

- Entropy of a Bernoulli shift : $-\sum_{i=1}^{k} p_i \log p_i$.
  ($\leq \log k$ which is the entropy of $(p_1, \ldots, p_k) = (\frac{1}{k}, \ldots, \frac{1}{k})$.)
- Entropy of a Markov shift (Q, q) : $-\sum_{i=1}^{k} q_i \sum_{j=1}^{k} Q_{i,j} \log Q_{i,j}$.

6. Linear automorphism T_A of $\mathbb{T}^2$ from the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$,
μ is Lebesgue measure :

There is a nice partition (Markov partition) which is generating and $h_\mu(T_A) = \log((3 + \sqrt{5})/2)$.

two interpretations

of

Kolmogorov-Sinai entropy :

Shannon-McMillan-Breiman theorem and Brin-Katok formula.

First, a **basic example** :
Bernoulli shift T on 2 symbols ($k = 2$) with Bernoulli measure $(1-p, p)$.

For any $x \in \{1,2\}^{\mathbb{N}}$ and $n \in \mathbb{N}$

$$\mu([x_0 \cdots x_{n-1}]) = (1-p)^{\sharp\{x_i=1 : i \le n-1\}} p^{\sharp\{x_i=2 : i \le n-1\}},$$

thus

$$-\frac{1}{n}\log \mu([x_0 \cdots x_{n-1}]) =$$
$$-\log(1-p) \times \frac{\sharp\{x_i = 1 : i \le n-1\}}{n} - \log p \times \frac{\sharp\{x_i = 2 : i \le n-1\}}{n}.$$

By the strong law of large numbers (Birkhoff's theorem), for μ-almost every x,

$$-\frac{1}{n}\log \mu([x_0 \cdots x_{n-1}]) \xrightarrow[n\to\infty]{} -\log(1-p) \times \overbrace{\mu([1])}^{=1-p} - \log p \times \overbrace{\mu([2])}^{=p} = h_\mu(T).$$

This is in fact a very general behavior.

Let $(X, \mathfrak{F}, \mu, T)$ be ergodic.
Let $\boldsymbol{\alpha}$ be a finite partition.
Given $x \in X$, let $\alpha_n(x)$ be the element of $\bigvee_{i=0}^{n-1} T^{-i}\boldsymbol{\alpha}$ which contains x.

Theorem (Shannon-McMillan-Breiman).
For μ-almost every $x \in X$, one has

$$h_\mu(T, \boldsymbol{\alpha}) = \lim_{n \to +\infty} -\frac{1}{n} \log \mu\big(\alpha_n(x)\big).$$

The convergence is also in $\mathcal{L}^1(\mu)$.

(If $\boldsymbol{\alpha}$ is generating, then we can replace $h_\mu(T, \boldsymbol{\alpha})$ by $h_\mu(T)$.)

Interpretation : aymptotic equipartition property

Fix ε (small enough).
Abbreviate $h_\mu(T, \boldsymbol{\alpha})$ (supposed to be > 0) as h.

Then there exists N_ε such that $\forall n \geq N_\varepsilon$, there exists $G_n \subset X$ such that :

- $\mu(G_n) > 1 - \varepsilon$
- G_n is the union of $e^{(h \pm \varepsilon)n}$ elements of $\bigvee_{i=0}^{n-1} T^{-i}\boldsymbol{\alpha}$, each having μ-measure $e^{-(h \pm \varepsilon)n}$.

'Through μ-glasses', X is (almost) partitioned into equally sized 'atoms', each with mass $\approx e^{hn}$, if we let the dynamics run long enough ($n \gg 1$).

Brin-Katok formula

Here X is a **compact** metric space with distance d (and $\mathfrak{F}$ is the Borel sigma-algebra).
Let $x \in X$, $\varepsilon > 0$ ('resolution'), and $n \geq 1$. Define the 'dynamical ball'

$$B(x, \varepsilon, n) \stackrel{\text{def}}{=} \left\{ y \in X : d(T^i(x), T^i(y)) \leq \varepsilon, \ \forall i = 0, 1, \ldots, n-1 \right\}.$$

This is the set of points that are indistinguishable from x at resolution ε in n iterates.

Theorem.
For μ almost every $x \in X$, one has

$$h_\mu(T) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} -\frac{1}{n} \log \mu\big(B(x, \varepsilon, n)\big).$$

ENTROPY AS INVARIANT

OF

ERGODIC EQUIVALENCE

Let (X, T, μ), (X', T', μ') be two measure-preserving dynamical systems.
(For simplicity, I drop the sigma-algebras $\mathfrak{F}, \mathfrak{F}'$.)

Definition (Ergodic equivalence).
They are ergodically equivalent if

- $\exists N \subset X$ such that $\mu(N) = 0$, $\exists N' \subset X'$ such that $\mu'(N') = 0$;
- There exists a measurable bijection $\phi : X \setminus N \to X' \setminus N'$ with measurable inverse;
- $\mu \circ \phi^{-1} = \mu'$ and $\phi \circ T = T' \circ \phi$.

Properties like ergodicity or mixing are invariants of ergodic equivalence.

There are other notions of equivalence, like spectral equivalence (Koopman operator in $\mathcal{L}^2$), that I won't discuss here.

Example

Consider : $T' : [0, 1] \to [0, 1]$, $T'(x) = 10x \bmod 1$, μ' is Lebesgue measure. Writing $x = 0.a_0a_1a_2\ldots$, with $a_i \in \{0, 1, \ldots, 9\}$, we saw that

$$T'(x) = 0.\not{a_0}a_1a_2\ldots$$

so let

$$\phi : \{0, 1, 2, \ldots\}^{\mathbb{N}} \to [0, 1], \quad \phi((a_n)_n) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \frac{a_n}{10^{n+1}}.$$

One can check that :

- Up to numbers having finite decimal expansion (*i.e.*, such that all but finitely many digits are 0), which form a countable set (hence of Lebesgue measure 0), ϕ is a bijection ;
- $\mu \circ \phi^{-1} =$ Lebesgue, where μ product of Bernoulli measures $\left(\frac{1}{10}, \ldots, \frac{1}{10}\right)$ on $\{0, 1, 2, \ldots, 9\}$;
- $\phi \circ T((a_n)_n) = T' \circ \phi((a_n)_n)$, where T is the shift map.

Von Neumann had asked if the 2-sided Bernoulli shifts

$$T : \{1,2\}^{\mathbb{Z}} \to \{1,2\}^{\mathbb{Z}}, \quad \mu = \left(\frac{1}{2}, \frac{1}{2}\right)^{\mathbb{Z}}$$

and

$$T' : \{1,2,3\}^{\mathbb{Z}} \to \{1,2,3\}^{\mathbb{Z}}, \quad \mu' = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^{\mathbb{Z}}$$

were ergodically equivalent.

Remarks. They cannot be *topologically* conjugate (T has two fixed points and T' has three fixed points). All 2-sided Bernoulli shifts are spectrally equivalent.

Kolmogorov rightly introduced entropy to solve this question. Indeed :

Entropy is an invariant of ergodic equivalence.

Sketch of proof. Let (X, T, μ), (X', T', μ') be two ergodically equivalent measure-preserving dynamical systems. If $\boldsymbol{\alpha}'$ is a partition of X', then $\phi^{-1}(\boldsymbol{\alpha}')$ is a partition of X, and $\mu \circ \phi^{-1} = \mu'$. Hence $H_{\mu'}(\boldsymbol{\alpha}') = H_{\mu}(\phi^{-1}(\boldsymbol{\alpha}))$. $\square$

Since entropy is an invariant of ergodic equivalence :

TWO BERNOULLI SHIFTS CAN BE ERGODICALLY EQUIVALENT
ONLY IF THEY HAVE THE SAME ENTROPY.

In particular :
The above 2-sided Bernoulli shifts cannot be ergodically equivalent.

WHAT ABOUT THE CONVERSE ? THAT IS :
DOES EQUALITY OF ENTROPY ENFORCES ERGODIC EQUIVALENCE ?

Can't be true in general : for instance, all circle rotations have entropy 0 but an irrational rotation (which is ergodic) cannot be ergodically equivalent to a rational rotation (not ergodic).

Meshalkin proved in 1959 that the 2-sided Bernoulli shifts generated by the Bernoulli measures

$$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \quad \text{and} \quad \left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right),$$

which both have entropy $\log 4$, are ergodically equivalent (he wrote down explicitly ϕ, the 'coding map').

Ornstein achieved the following breakthrough.

☕☕☕☕☕**Theorem.**
Two 2-sided Bernoulli shifts generated by, say, $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_{k'})$ are ergodically equivalent **if and only if** they have the same entropy, that is, if and only if

$$-\sum_{j=1}^{k} p_j \log p_j = -\sum_{i=1}^{k'} q_i \log q_i.$$

D. Ornstein in 1970

Ornstein went much further :

He also gave necessary and sufficient conditions for a measure-preseving dynamical system to be ergodically equivalent to a Bernoulli shift (generating partition which is 'very weak Bernoulli', or being 'finitely determined'). (Not discussed here.)

Examples :

- Any irreducible and aperiodic Markov chain on a finite set is ergodically equivalent to a Bernoulli shift.
- Automorphisms of $\mathbb{T}^d$, A is a $n \times n$ matrix with integer coefficients and $|\det(A)| = 1$; we saw above the case $d = 2$ and $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

If none of the eigenvalues $\{\lambda_i\}_{i=1}^d$ of A are roots of unity, then T_A ergodically equivalent to a Bernoulli shift with entropy $\sum_{i=1}^d \max(0, \log \lambda_i)$.

Closing remarks

Generalizations :

- The above theory of entropy works for countable partitions.
- One can define Bernoulli shifts for arbitrary standard probability spaces instead of finite sets with Bernoulli measures.

Trichotomy in ergodic theory according to entropy :

- Zero-entropy (or ‘fully deterministic’) systems (like rotations on the circle. There are many other examples).
  Such systems have subexponential growth of orbit complexity with time ;
- Finite positive entropy (exponential growth of orbit complexity), as in most examples of this talk ;
- Infinite entropy. Includes many important classes of stationary random processes, such as Wiener or Gaussian, with absolutely continuous spectral measure, and various infinite-dimensional dynamical systems.

Slides that I had no time to show

but that you would like to see 😉

ℹ

What follows is very sketchy and only intended to point out other topics involving entropy.

Ornstein-Weiss theorem : entropy via recurrence times

Let $(X, \mathfrak{F}, \mu, T)$ be an ergodic dynamical system.
Let $\boldsymbol{\alpha}$ be a finite partition.

Let $B \subset X$ such that $\mu(B) > 0$ and $\mathcal{T}_B(x) = \inf\{j \geq 1 : T^j(x) \in B\}$.

Theorem. For μ almost every x one has

$$\lim_{n \to +\infty} \frac{1}{n} \log \mathcal{T}_{\alpha_n(x)}(x) = h_\mu(T, \boldsymbol{\alpha}).$$

Remark. By Kač's formula, for every $x \in X$

$$\int \mathcal{T}_{\alpha_n(x)}(y)\, \mathrm{d}\mu_{\alpha_n(x)}(y) = \frac{1}{\mu(\alpha_n(x))}.$$

Then, for μ almost every x, Shannon-McMillan-Breiman theorem implies

$$-\frac{1}{n} \log \int \mathcal{T}_{\alpha_n(x)}(y)\, \mathrm{d}\mu_{\alpha_n(x)}(y) \to h_\mu(T, \boldsymbol{\alpha}).$$

Topological entropy

Here (X, d) is a compact metric space and $T : X \to X$ continuous (a homeomorphism in the invertible case).

Denote by $S_d(\varepsilon)$ the minimal number of balls of radius ε which covers X. Then

$$h_{\mathrm{top}}(T) \stackrel{\mathrm{def}}{=} \lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{\log S_{d_n^T}(\varepsilon)}{n}$$

where

$$d_n^T(x, y) \stackrel{\mathrm{def}}{=} \max \left\{ d(x, y), d(T(x), T(y)), \ldots, d(T^{n-1}(x), T^{n-1}(y)) \right\}.$$

(If d' is another distance defining the same topology as d, then we get the same number.)

Roughly speaking, topological entropy measures the maximal complexity and Kolmogorov-Sinai entropy measures the statistical complexity statistics based on a given measure).
Topological entropy is nvariant by topological conjugacy.

Theorem (Variational principle).
If $T : X \to X$ is a continuous on the compact metric space (X, d), then

$$h_{\text{top}}(T) = \sup\{h_\mu(T) : \mu \ T\text{-invariant}\}.$$

If the supremum is attained, one speaks of *measure of maximal entropy*. What about uniqueness? Hard questions in general.

- Expansive maps of compact metric spaces have a measure of maximal entropy.
- All transitive topological Markov chains ($\equiv$ "skeletons" of irreducible Markov chains), or hyperbolic toral automorphisms have a unique measure of maximal entropy.
- Link with periodic points : for shifts of finite type, one has

$$\lim_{n \to +\infty} \frac{1}{n} \log \sharp \mathrm{Fix}(T^n) = h_{\text{top}}(T).$$

- For the full shift on k symbols : $\sharp \mathrm{Fix}(T^n) = k^n$, hence $h_{\text{top}}(T) = \log k$. The corresponding measure is $\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)^{\mathbb{N} \text{ (or } \mathbb{Z})}$.

Setting :

- $(X, \mathfrak{F}, T)$ where X is a compact metric space and T is continuous ;
- $\varphi : X \to \mathbb{R}$ continuous ('potential').

One can define (not done here) the 'pressure' $P_T(\varphi)$ of a φ wrt T.
(All this jargon comes from statistical physics.)

Theorem (Variational principle).

$$P_T(\varphi) = \sup \left\{ \hbar_\nu(T) + \int \varphi \, \mathrm{d}\nu : \nu \ T\text{−invariant} \right\}.$$

By definition, T-invariant measures achieving the supremum (if any) are called *equilibrium states* of φ.

When $\varphi = 0$, $P_T(0) = \hbar_{\text{top}}(T)$. (So the previous theorem generalizes the theorem above.)

If T is expansive, then every φ admits some equilibrium state.

A few remarks :

- For ‘nice’ dynamical systems (*e.g.*, uniformly expanding maps of the interval, shifts of finite type, Axiom A diffeomorphisms, etc), and φ regular enough (at least Hölder), unique equilibrium states exist and are unique, and are also ‘Gibbs measures’.
- Invertible systems in the previous class are ergodically equivalent to a Bernoulli shift.

Margulis-Ruelle inequality and Pesin's formula

The context is **differentiable dynamics** :
$T : X \to X$ is a C^2-diffeomorphism of a compact Riemannian manifold X and μ is T-invariant ergodic Borel probability measure.
Let $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ denote the distinct Lyapunov exponents of (T, μ) and let E_i be the linear subspaces corresponding to λ_i, so that $\dim(E_i)$ is the multiplicity of λ_i. (There's a big theorem behind this, namely the multiplicative ergodic theorem of Oseledets.)
Lyapunov exponents measure the rates at which nearby orbits diverge.

Theorem (Margulis-Ruelle).

$$h_\mu(T) \leq \sum_{i=1}^{r} \max(0, \lambda_i) \dim(E_i).$$

Theorem (Pesin).
If μ is equivalent to the Riemannian measure on X, then

$$h_\mu(T) = \sum_{i=1}^{r} \max(0, \lambda_i) \dim(E_i).$$

Krieger's generator theorem

Theorem.
Let $(X, \mathfrak{F}, \mu, T)$ be an ergodic invertible dynamical system. Assume that μ is not a periodic-orbit measure and $h_\mu(T) < +\infty$.
Then, there exists a generating partition $\boldsymbol{\alpha}$ with

$$e^{h_\mu(T)} \leq \sharp\boldsymbol{\alpha} \leq e^{h_\mu(T)} + 1.$$

Refined version : A necessary and sufficient condition for the existence of a measure-preserving embedding of (X, T, μ) into $(\{1, \ldots, k\}^{\mathbb{Z}}, T_{shift})$ is

$$h_\mu(T) < \log k \quad \text{or} \quad (X, T, \mu) \text{ is a Bernoulli shift with } h_\mu(T) = \log k.$$

(Measure-preserving embedding : there is a Borel injective map $b : X' \to \{1, \ldots, k\}^{\mathbb{Z}}$ with $\mu(X') = 1$ and $b \circ T = T_{shift} \circ b$.)

Brudno's theorem (Kolmogorov complexity of orbits and entropy)

Let $(X, \mathfrak{F}, \mu, T)$ be an ergodic dynamical system and $\boldsymbol{\alpha}$ a finite partition of X.
Let $\omega_0^{n-1}(x) = (\omega_0(x), \ldots, \omega_{n-1}(x))$ where $\omega_j(x) = i$ if $T^j(x) \in \alpha_i$.
Given a Turing machine M, the Kolmogorov complexity $K_M(\omega_0^{n-1}(x))$ of the 'word' $\omega_0^{n-1}(x)$ is the length of the shortest algorithm which outputs $\omega_0^{n-1}(x)$.
Kolmogorov proved that there exists a universal Turing machine U (*i.e.*, it can simulate any other Turing machine) such that

$K_U(\omega_0^{n-1}(x)) \leq K_M(\omega_0^{n-1}(x)) + C$ (where C is a constant depending only on U and M).

Theorem.
For μ-almost every x

$$\lim_{n\to+\infty} \frac{K_U(\omega_0^{n-1}(x))}{n} = h_\mu(T, \boldsymbol{\alpha}).$$

Entropy is the only finitely observable invariant

This is new (and very surprising) characterization of entropy.
Consider stationary, ergodic, stochastic processes $\mathcal{X} = (X_n)_n$ where the X_n's take values in a finite alphabet.

Roughly speaking, $J(\mathcal{X})$ is finitely observable if there is some sequence of functions $\big(S_n(x_0, \ldots, x_{n-1})\big)_n$ that converges to $J(\mathcal{X})$ for almost every realization $x_0, x_1, \ldots$ of the process $\mathcal{X}$, for all ergodic processes.
Basic example : $J(\mathcal{X}) = \mathbb{E}(X_0)$ and $S_n(x_0, \ldots, x_{n-1}) = (x_0 + \cdots + x_{n-1})/n$; convergence by Birkhoff's theorem.

Theorem (Ornstein-Weiss, 2006).
If J is a finitely observable function, defined on all ergodic finite-valued processes, *which is an invariant of ergodic equivalence*, then J is a continous function of the entropy.

Remark. There are several different estimators converging to the entropy (*e.g.* , using return times, see above).

Weak Pinsker conjecture

Long-standing conjecture solved in 2018.

Theorem (T. Austin, 2018).
Every ergodic invertible dynamical system $(X, \mathfrak{F}, T, \mu)$ has the '*weak Pinsker property*' :
For every $\varepsilon > 0$, it splits into a direct product of a Bernoulli shift and a system of entropy less than ε.

Remark. For $\varepsilon = 0$ ('Pinsker conjecture'), the statement is false (counter-examples by Ornstein).

A non-exhaustive list of references

- M. Viana, K. Oliveira, Foundations of Ergodic Theory, Cambridge University Press, 2016. (This is an excellent recent book that covers many topics of this talk.)
- T. Downarowicz, Entropy in Dynamical Systems, Cambridge University Press, 2011. (See Section 4 in Chapter 3 for Ornstein-Weiss theorem on recurrence times and entropy.)
- A. Katok, Fifty years of entropy. (Available freely online.)
- R. Bowen. Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms 2nd edn (Lecture Notes in Mathematics 2nd edn ed J.-R. Chazottes (Berlin : Springer) vol 470) with a preface by David Ruelle, 2008.
- C. Hoffman, Isomorphism Theory in Ergodic Theory. In : Meyers, R. (eds) Mathematics of Complexity and Dynamical Systems. Springer, 2012.
- P. Shields, The theory of Bernoulli shifts, online reprint of the 1973 University of Chicago Press edition.
- D. Ornstein, B. Weiss, Entropy is the only finitely observable invariant, Journal of Modern Dynamics, 2007, 1(1) : 93–105.
- T. Austin, Measure concentration and the weak Pinsker property, Publications mathématiques de l'IHÉS volume 128, pages1–119 (2018).