Entropies

Order / Disorder

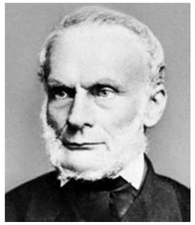
from Time Series

*Any method involving the notion of **entropy**, the very existence of which depends on the second law of thermodynamics, will doubtless seem to many far-fetched, and may repel beginners as obscure and difficult of comprehension.*

Willard Gibbs

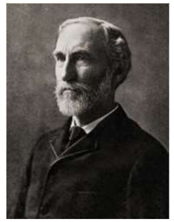
Graphical Methods in the Thermodynamics of Fluids (1906)

fundamental concept
in thermodynamics
and statistical mechanics
(1850s – 1880s)


R. Clausius


J.C. Maxwell


L. Boltzmann


J. Gibbs


H. v. Helmholtz

entropy → expression of the disorder, or randomness of a system

- macroscopically:

$$S = k_B \ln \Omega \quad [J/K]$$
$$\Omega \text{ denotes number of microstates}$$
$$k_B \approx 1.38 \cdot 10^{-23} \quad [J/K]$$

- microscopically:

$$S = -k_B \sum_i p_i \ln p_i$$
$$p_i = \frac{1}{\Omega} \text{ for microcanonical ensemble}$$

phase transitions, entropy-driven order (Landau theory); adiabatic demagnetization; …

fundamental concept
in information theory
(1940-1950)


C. Shannon


A. Rényi


A. Kolmogorov


Y. Sinai

entropy → amount of information needed to specify the full microstate of the system X (Shannon entropy)

$$S(X) = -\sum_i p(x_i) \ln p(x_i)$$

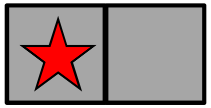
extensions and generalizations useful for time series analysis:

Rényi entropies → diversity, uncertainty, or randomness of a system

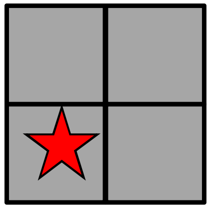
Kolmogorov-Sinai entropies → chaoticity of a system

entropy and Information

observing a system (measurement) is source of information



system with 2 states has maximum information content: 1 bit



system with 4 states has maximum information content: 2 bits

system with M states has maximum information content: $I = \log_2 M$

entropy and Information

measuring statistical events and average information gain

given a priori knowledge:
M events (M system states) will appear (will be taken) with probabilities $\{p_i\}$, with $\sum_i p_i = 1$

measurement:
if you learn that event j ($j \in M$) appeared (system state j has been taken) then you will gain "average information" (through many measurement repetitions) as

$$I = -\sum_i p_i \log_2 p_i$$

(denoted as Shannon information)

entropy and Information

measuring statistical events and average information gain

example: coin flipping; head (p_1) or tail (p_2)?
equal probability for outcome: p_1 = p_2 = 0.5

measurement → head → information gain $I = 1$

and with probabilities:

$$I = -\left(0.5\log_2 0.5 + 0.5\log_2 0.5\right) = -\left(-0.5 - 0.5\right) = 1$$

linear methods for estimating entropies

recall: Fourier transform and Parseval's theorem (see Linear Methods) with normalized power spectrum

$$\hat{P} = \sum_{k=1}^{N} |\hat{v_k}|^2 \stackrel{!}{=} 1$$

we can estimate the entropy S of the relative spectral density as:

$$S = -\sum_{k=1}^{N} \hat{P}(k) \log_2 \hat{P}(k)$$

S characterizes homogeneity of power spectrum:

S is minimum for line spectra (single Fourier component)
S is maximum for broad-band spectra (white noise)
S for chaotic dynamics? (looks like white noise)

need other methods to characterize entropy of chaotic dynamics

entropy and Information

Given:

- measured data follows some probability distribution
- transitions between successive data points occur with well-defined probabilities

Qs:

- if you have performed exactly one measurement, how much do you learn about the state of a system?
- if you have observed the entire past of a system, how much information do you have about future observations?

As:

can be found with generalized Rényi entropies

generalized entropies static distributions

order-q Rényi entropies

… characterize the amount of information needed to specify the value of an observable with a certain precision if only the probability density is known that observable has value **x**.

Idea:
- partition phase space into M *disjoint* hypercubes (boxes) of side length ε (set of all these hypercubes is called a partition $\mathcal{P}_\varepsilon$)
- estimate probability p_j to find state **x** in box j
- define order-q Rényi entropy for partition $\mathcal{P}_\varepsilon$ as:

$$\tilde{H}_q(\mathcal{P}_\epsilon) = \frac{1}{1-q} \ln \sum_{j=1}^{M(\epsilon)} p_j^q$$

generalized entropies — static distributions

order-q Rényi entropies

for q = 1, we derive (L'Hôpital's rule) the *Shannon entropy*:

$$\tilde{H}_1(\mathcal{P}_\epsilon) = -\sum_j p_j \ln p_j$$

which is the only Rényi entropy that is additive:

the Rényi entropy of a joint process is the sum of the entropies of the independent processes

(cf. mutual information)

generalized entropies

static distributions

example: Rényi entropy of a uniform distribution

given: probability density $\mu(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{else} \end{cases}$

partition the unit interval into N partitions of length $\epsilon = \frac{1}{N}$

we find: $\tilde{H}_q(\epsilon) = \frac{1}{1-q} \ln \left(N \epsilon^q \right) = -\ln \epsilon = \ln N$

- all order-q entropies are the same
  (due to the homogeneity of the uniform distribution)

- the better you resolve the real numbers by the partition, the more information you gain

generalized entropies and dimensions

relationship: order-q entropies and order-q dimensions

$$\tilde{H}_q(\mathcal{P}_\epsilon) = \frac{\ln \sum_{j=1}^{M(\epsilon)} p_j^q}{1-q} \qquad D_q := \lim_{\epsilon \to 0} \frac{\ln\left(\sum_{i=1}^{M(\epsilon)} p_i^q\right)}{(q-1)\ln(\epsilon)}$$

- disjoint vs. non-disjoint partitioning

dimensions are the scaling exponents of the Rényi entropies computed for equally-sized partitions as functions of ε and in the limit $\varepsilon \to 0$.

generalized entropies

so far: entropies for *static* distributions

- can characterize attractor “as a whole”
- similar to dimension → no further gain of information
- no information about *dynamics on the attractor*

idea:

- consider entropies for *transition probabilities*
- characterize flow of information from small to large scales (typical for chaotic systems)

generalized entropies **Kolmogorov-Sinai entropy**

- partition m-dimensional phase space into M *disjoint* hypercubes (boxes) of side length ε^m

- let $p_{i_1,\ldots,i_m}$ denote the *joint probability* that state $\mathbf{X}(t=1)$ is in box i_1, state $\mathbf{X}(t=2)$ is in box i_2, etc., and that state $\mathbf{X}(mt)$ is in box i_m

- define ***block-entropies*** of block-size m as:

$$H_q(m, \mathcal{P}_\epsilon) = \frac{\ln \sum_{i_1,\ldots,i_m}^{M(\epsilon)} p_{i_1,\ldots,i_m}^q}{1-q}$$

generalized entropies **Kolmogorov-Sinai entropy**

for $m \to \infty$, block-entropies are related to order-q entropies as:

$$h_q = \sup_{\mathcal{P}} \lim_{m\to\infty} \frac{1}{m} H_q(m, \mathcal{P}_\epsilon)$$
$$h_q = \lim_{m\to\infty} H_q(m+1, \mathcal{P}_\epsilon) - H_q(m, \mathcal{P}_\epsilon)$$
$$\text{with}$$
$$h_q(0, \mathcal{P}_\epsilon) := H_q(0, \mathcal{P}_\epsilon)$$

the supremum indicates: maximize over all possible partitions $\mathcal{P}$, and implies the limit $\varepsilon \to 0$

h_0 is called *topological entropy* (also abbreviated with K_0)
h_1 is called *Kolmogorov-Sinai entropy* (also abbreviated with K_1)

generalized entropies — Kolmogorov-Sinai entropy

what do order-q entropies and order-q dimensions characterize?

topological entropy and Hausdorff dimension

- h_0 (or K_0) counts number of different orbits
- D_0 counts number of non-empty boxes

Kolmogorov-Sinai entropy and information dimension

- h_1 (or K_1) is a measure for the average rate of loss of information loss about a system state
- D_1 is a measure for a gain of information when findings a state in a given box

entropies from time series

entropies provide important information on topology of folding processes, disorder, chaoticity, and predictability

estimating order-q entropies from data is hard, particularly for high-dimensional systems (require more data than dimensions or Lyapunov exponents)

taking the limit $m \to \infty$ is difficult

box-counting (evaluate m-dimensional histograms) is most direct approach but turned out to be impractical

alternative ansatz: ***importance sampling***

entropies from time series correlation entropy

idea:

- instead of using uniformly distributed partitions of phase space center partitions (boxes with fixed ε) on phase-space vectors

- use correlation sum (see Dimensions) to derive *correlation entropy* K_2

Grassberger, P. & Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A*, *28*(4), 2591. 1983

entropies from time series — **correlation entropy**

with order-q correlation sum

$$C_q(\epsilon) := \frac{1}{N}\sum_i \left(\frac{1}{N}\sum_j \Theta\left(\epsilon - |\vec{v_i} - \vec{v_j}|\right)\right)^{q-1}$$

we find for $q = 2$

$$C_2(\epsilon) \propto \text{const.}\ \epsilon^{D_2}$$

in general, we have for $q > 1$

$$C_q(\epsilon) \propto \epsilon^{(q-1)D_q} \mathrm{e}^{(1-q)H_q(m)}$$

if the systems exhibits a scaling region, we have $\epsilon^{D_q} \approx \text{const.}$
we can then find correlation entropy from

$$\begin{aligned} h_q &= \lim_{m\to\infty} H_q(m+1,\epsilon) - H_q(m,\epsilon) \\ &= \lim_{m\to\infty} \ln\left(\frac{C(m,\epsilon)}{C(m+1,\epsilon)}\right) =: K_2 \end{aligned}$$

entropies from time series **correlation entropy**

pros and cons of correlation entropy

- conceptually easy
- quickest to calculate

- requires existence of scaling region (independent on ε)
  (if you can't find a scaling region do not apply this method!)
- needs lots of data
  (you loose ε^{-h} neighbors when going from m to $m+1$)

→

check robustness
constancy for a range of ε values and embedding dimensions m

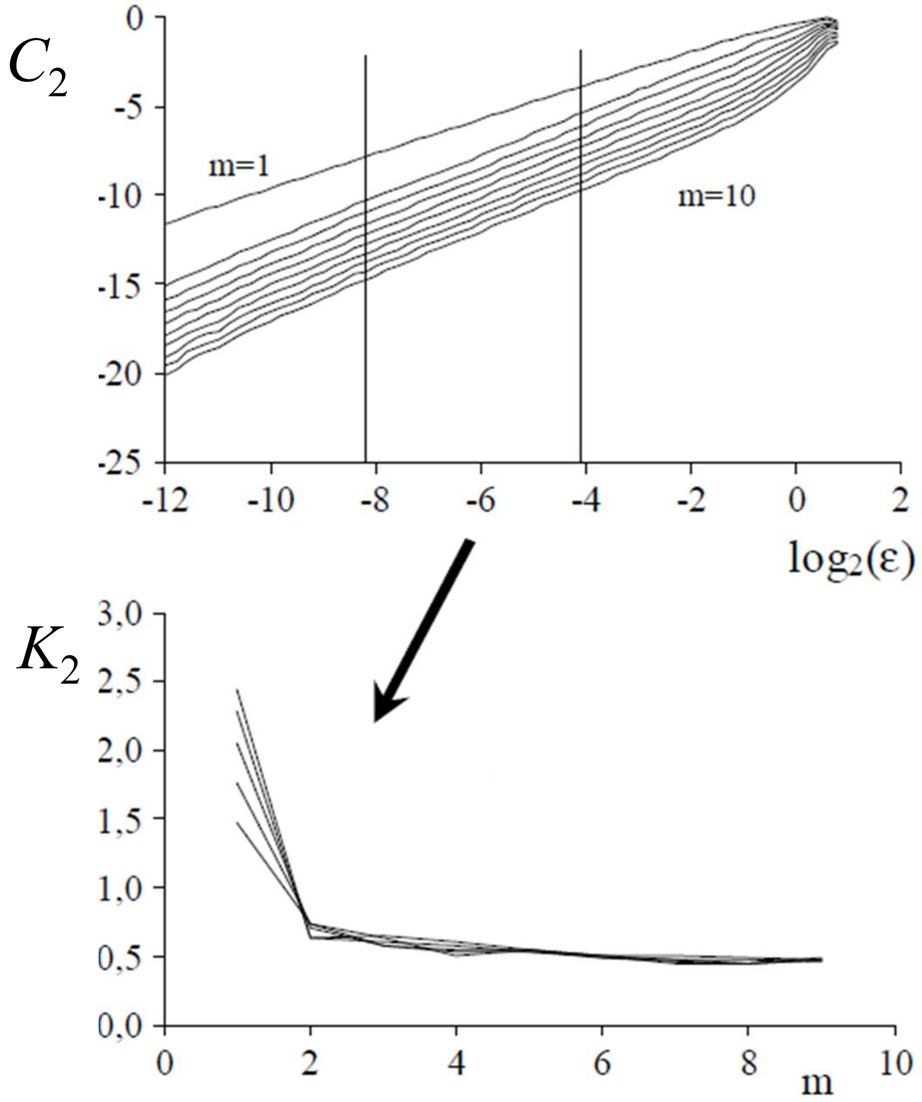
entropies from time series

example: Hénon map

$$x_{n+1} = 1 - a x_n^2 + y_n$$
$$y_{n+1} = b x_n$$

where

$a = 1.4; b = 0.3$



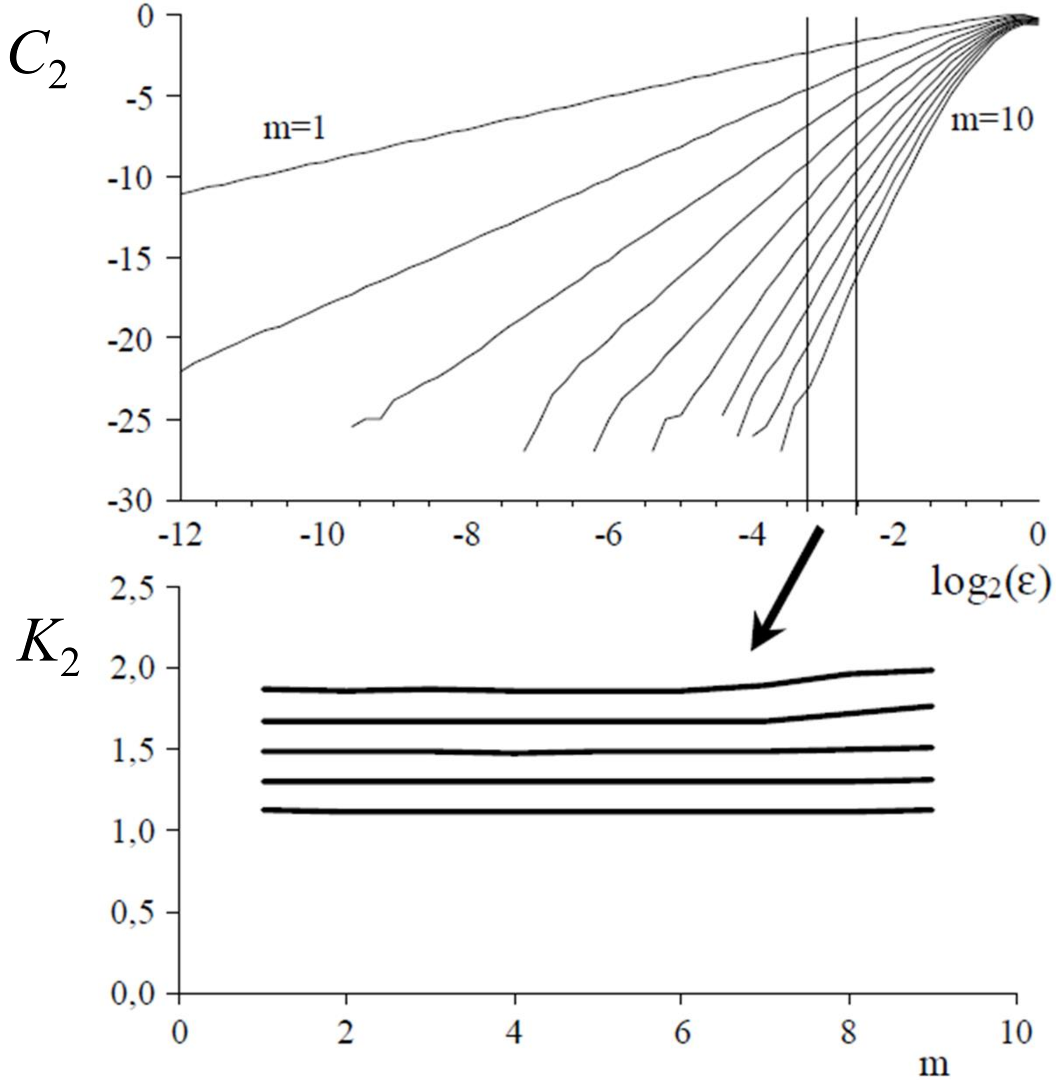
literature ($m \to \infty$):

$K_2 \sim 0.33$

entropies from time series **example: white noise**

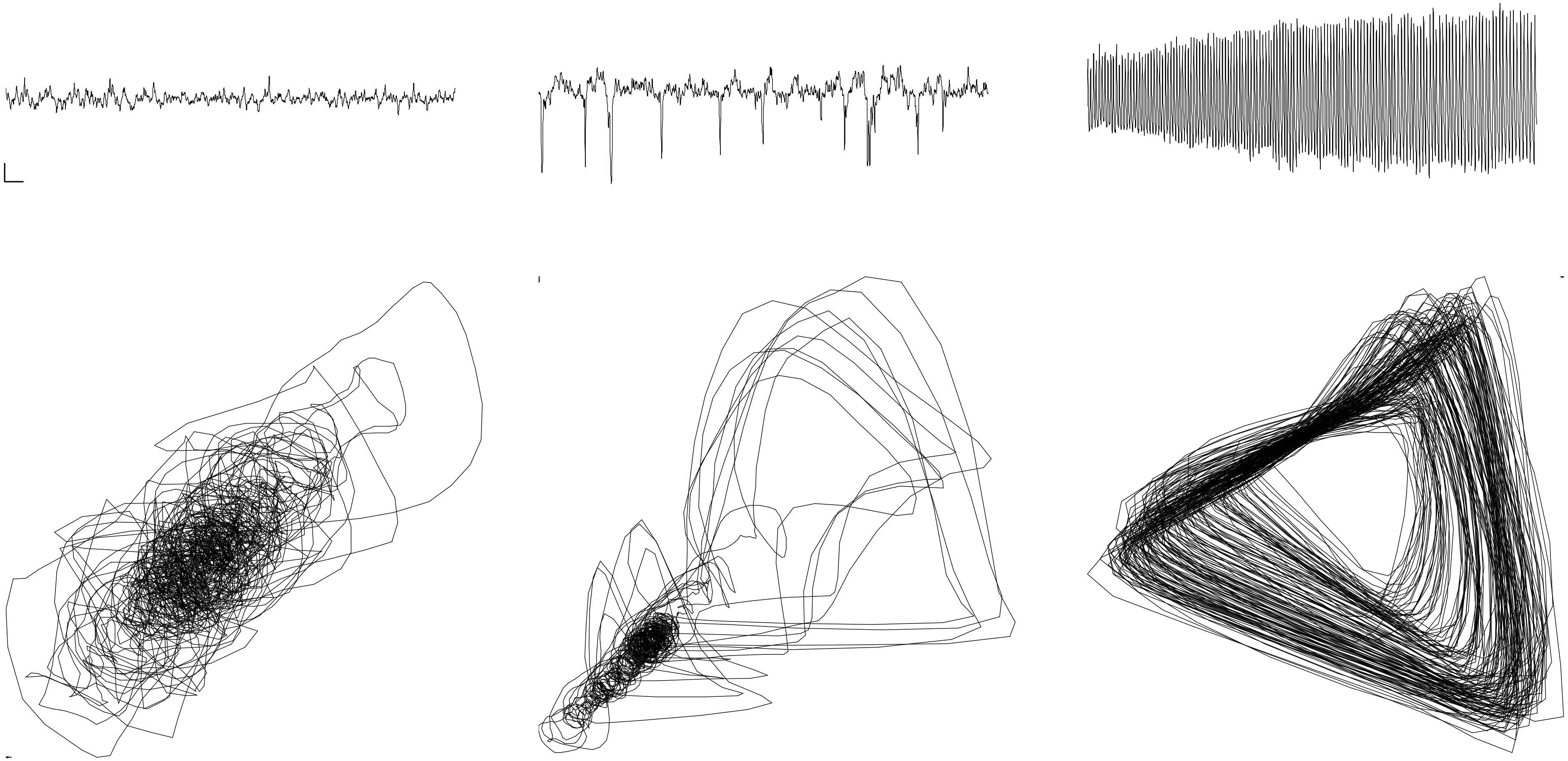
difficult to identify
scaling region

no constancy
for range of ε values



entropies from time series

example: EEG data

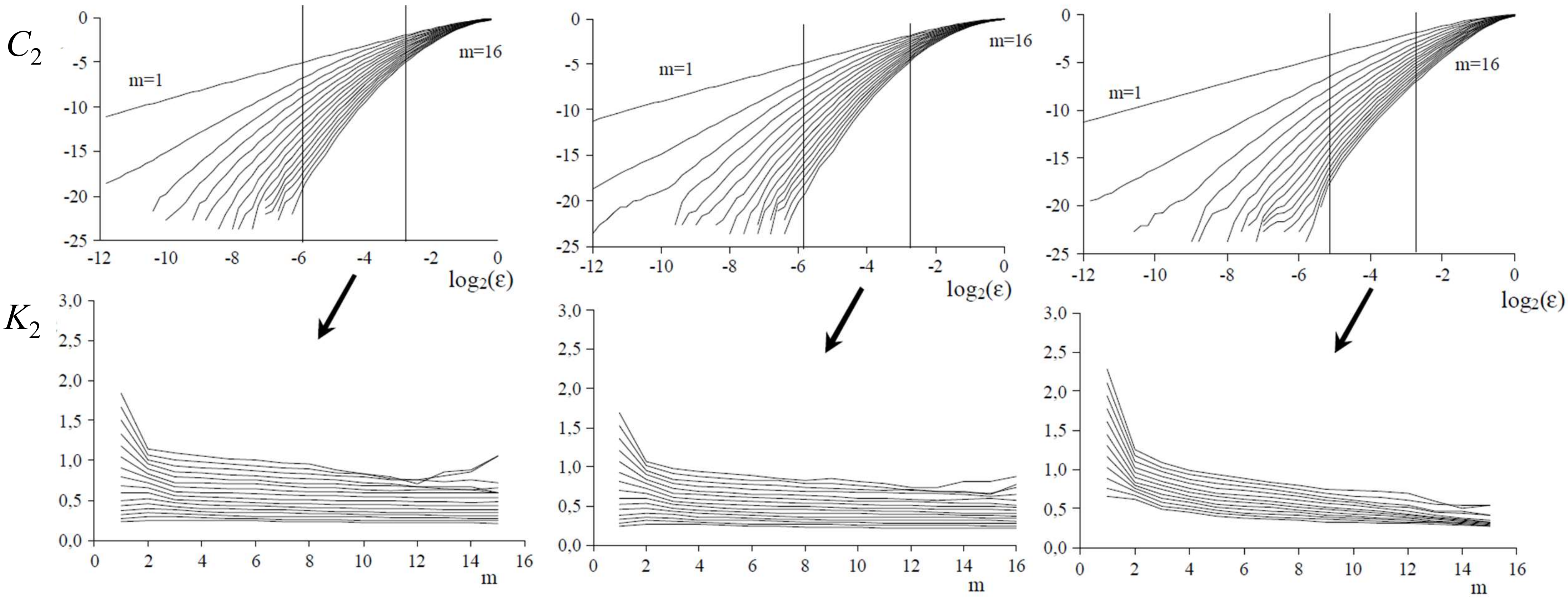


healthy subject

epilepsy patient seizure-free interval

epilepsy patient seizure

entropies from time series example: EEG data



entropies **what can go wrong?**

field applications

- number of data points ($\lim N \to \infty$ and $m \to \infty$)
- data precision
  adopt to requirement of small ε-neighborhood
- strong correlations in data (sampling interval)
  use Theiler correction (see Dimensions)
- noise, filtering
  similar impact as with Dimensions and Lyapunov exponents
- identifiable scaling region

entropies Interpretation

- in general, we have: $K_{q'} \leq K_q$ for $q' > q$

- disorder, chaoticity of a system and type of the dynamics:

  $K > 0$: chaos, unstable dynamics
  $K = 0$: regular dynamics
  $K = \infty$: noise

- average rate of loss of information due to action of nonlinearity

- prediction horizon:

$$T_\mathrm{p} \approx \frac{-\ln(\rho)}{K}$$

where:
ρ denotes accuracy of measurement (initial state)

Pesin's identity

relationship between entropy and Lyapunov exponents

- entropy characterizes average rate of loss of information loss about a system state

- Lyapunov exponents characterize exponential divergence of initially close system states

Pesin's identity: $$K_1 = \sum_{i,\lambda_i>0} \lambda_i$$

Ya. B. Pesin, "Characteristic Lyapunov exponents and smooth ergodic theory", *Uspekhi Mat. Nauk*, **32**:4(196), 1977

Pesin's identity

relationship between entropy and Lyapunov exponents

consistency checks for time-series analysis

estimate K_1 from sum over all positive Lyapunov exponents

note that $K_1 = \sum_{i,\lambda_i>0} \lambda_i$

due to $K_{q'} \leq K_q$ for $q' > q$

we have $K_2 = \sum_{i,\lambda_i>0} \lambda_i$

compare with K_2 estimate from correlation sum