Time series analysis

Dimensions and entropies

Eckehard Olbrich
Max Planck Institute for Mathematics in the Sciences

Leipzig

22.06.2021

Plan for this lecture

1. Deterministic chaos
2. Entropies
3. Dimension

Ressources:

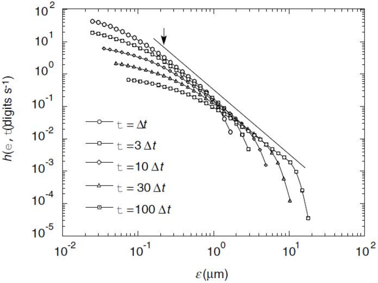
- https://personal-homepages.mis.mpg.de/olbrich/ - in particular the lecture on "Data analysis and modeling".
- Entropy rate: Anatole Katok and Boris Hasselblatt: Introduction to the modern theory of dynamical systems
- Scale dependent entropies: Pierre Gaspard and Xiao-Jing Wang, Noise, chaos, and (ϵ,τ)-entropy per unit time, Physics Reports, **235** (1993), 291 - 343.
- Dimensions: Yakov Pesin and Howerd Weiss, The multifractal analysis of Gibbs measures: Motivation, mathematical foundation, and examples, Chaos **7**, 89 (1997).

Experimental evidence for microscopic chaos

P. Gaspard*, M. E. Briggs†, M. K. Francis‡, J. V. Sengers‡, R. W. Gammon‡, J. R. Dorfman‡ & R. V. Calabrese‡

* *Université Libre de Bruxelles, CP 231-Campus Plaine, Boulevard du Triomphe, B-1050 Brussels, Belgium*
† *University of Utah, Salt Lake City, Utah 84112, USA*
‡ *University of Maryland, College Park, Maryland 20742, USA*

Many macroscopic dynamical phenomena, for example in hydrodynamics and oscillatory chemical reactions, have been observed to display erratic or random time evolution, in spite of the deterministic character of their dynamics—a phenomenon known as macroscopic chaos[1–5]. On the other hand, it has been long supposed that the existence of chaotic behaviour in the microscopic motions of atoms and molecules in fluids or solids is responsible for their equilibrium and non-equilibrium properties. But this hypothesis of microscopic chaos has never been verified experimentally. Chaotic behaviour of a system is characterized by the existence of positive Lyapunov exponents, which determine the rate of exponential separation of very close trajectories in the phase space of the system[6]. Positive Lyapunov exponents indicate that the microscopic dynamics of the system are very sensitive to its initial state, which, in turn, indicates that the dynamics are chaotic; a small change in initial conditions will lead to a large change in the microscopic motion. Here we report direct experimental evidence for microscopic chaos in fluid systems, obtained by the observation of brownian motion of a colloidal particle suspended in water. We find a positive lower bound on the sum of positive Lyapunov exponents of the system composed of the brownian particle and the surrounding fluid.



Brownian motion:

$$h(\epsilon,\tau) \leq A\frac{D}{\epsilon^2}$$

Deterministic systems:

$$h(\epsilon,\tau) \leq h_{KS} = \sum_{\lambda_i>0} \lambda_i$$

Probability space $(\Omega, \mathcal{A}, P)$

Set of possible events Ω: Set of outcomes of an random experiment — in the case of a coin toss $\Omega = (heads, tails)$. Elements denoted by $\omega \in \Omega$.

σ-algebra of subsets $\mathcal{A}$: Set of subsets of Ω.

Probability measure P: Each set of events $A \subseteq \mathcal{A}$ has a probability $0 \leq P(A) \leq 1$. $P(\Omega) = 1$.

Random variable X

Measureable function $X : (\Omega, \mathcal{A}) \to S$ to a measurable space S (frequently taken to be the real numbers with the standard measure). The probability measure $PX^{-1} : S \to \mathbb{R}$ associated to the random variable is defined by $PX^{-1}(s) = P(X^{-1}(s))$. A random variable has either an associated probability distribution (discrete random variable) or probability density function (continuous random variable).

A random variable X is said to be *discrete* if the set $\{X(\omega) : \omega \in \Omega\}$ (i.e. the range of X) is finite or countable.

Alphabet: Set $\mathcal{X}$ of values of the random variable X.

Probability: $p(x) = P(X = x), x \in \mathcal{X}$.

Normalization:

$$\sum_{x \in \mathcal{X}} p(x) = 1$$

Expectation value of X:

$$E_P[X] = \sum_{x \in \mathcal{X}} x p(x)$$

A random variable X is said to be continuous if it has a cumulative distribution function which is absolutely continuous.

Probability density $p(x)$

$$P(a \leq X \leq b) = \int_a^b p(x)dx \ .$$

Cumulative distribution

$$P_{\leq}(x) = P(X \leq x) = \int_{-\infty}^{x} p(y)dy$$

Normalization

$$\int_{x_{min}}^{x_{max}} p(x)dx = 1 \ .$$

Change of variable $y = f(x)$ (f invertible)

$$p(x)dx = q(y)dy \quad \Rightarrow \quad q(y) = \left.\frac{p(x)}{df/dx}\right|_{x=f^{-1}(y)}$$

- Shannon 1948: How much choice is involved in the selection of an event with n possibilities and probabilities $p_1, \ldots, p_n$?
- If we have a random variable X with a probability distribution $p(x)$ the uncertainty about the outcome x of a measurement of X is given by the *entropy*
- Entropy of a discrete random variable

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \ .$$

- Entropy can be considered as a measure of variety or disorder ("objective") or as a measure of uncertainty ("subjective")
- Information reduces uncertainty, i.e. it can be quantified by differences between uncertainties, that is: entropies.
- The entropy can be considered as the expectation value of $\log 1/p(x)$:

$$H(X) = E_P[\log \frac{1}{p(x)}] \ .$$

Are there other functions, which are suitable as a measure of uncertainty?

Theorem: The following three conditions determine the function $H(p_1, \ldots, p_n)$ up to a multiplicative constant, whose value serves only to determine the size of the unit of information.

1. $H(p, 1-p)$ is a continuous function of $p \in [0, 1]$.
2. $H_n(p_1, \ldots, p_n)$ is a symmetric function of all of its arguments.
3. If $p_n = q_1 + q_2 > 0$ then

$$H(p_1, p_2, p_3, \ldots, q1, q2) = H(p_1, p_2, p_3, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) .$$

The last property called "additivity" is dropped for some entropies such as the *Renyi entropies*.

Conditional entropy

- Knowing Y might reduce the uncertainty about X if both are not statistically independent.
- The uncertainty of X having already observed $Y = y$ can be expressed as

$$H(X|Y=y) = -\sum_{x \in \mathcal{X}} p(x|y) \log p(y|x) \; .$$

- This can be averaged also over Y giving

$$H(X|Y) = H(X,Y) - H(Y) \; .$$

$H(X|Y)$ is called *conditional entropy*.

- Chain rule:

$$H(X,Y) = H(X) + H(Y|X) \; .$$

- A **stochastic process** is indexed sequence of random variables. The process is characterized by joint probabilities

$$Pr\{(X_1, X_2, ..., X_n) = (x_1, x_2, \ldots, x_n)\} = p(x_1, \ldots, x_n)$$

with $(x_1, \ldots, x_n) \in \mathcal{X}^n$.

- A stochastic process is said to be **stationary** if the joint distribution of any subset of random variables is invariant with respect to shifts in the time index; that is

$$\begin{aligned} &Pr\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} \\ &= Pr\{X_{1+l} = x_1, X_{2+l} = x_2, \ldots, X_{n+l} = x_n\} \end{aligned}$$

for every n and every shift l and for all $x_1, x_2, \ldots, x_n \in \mathcal{X}$.

Block entropy $H(X_1, X_2, \ldots, X_n)$ is the of the probaility distribution on sequences of length n. Entropy rate as **entropy per symbol:**

$$h_\infty = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n)$$

Entropy rate as **conditional entropy given the past:**

$$h'_\infty = \lim_{n\to\infty} H(X_n | X_{n-1}, \ldots, X_1)$$

Theorem: For a stationary stochastic process the limits exists and are equal.

Can be proven using
Theorem: (*Cesáro mean*) If $a_n \to a$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, then $b_n \to a$.

- Deterministic dynamical system with continuous state variables x.

$$x_{n+1} = F(x_n)$$

- Invariant measure

$$\mu(F^{-1}A) = \mu(A) \quad \forall\, A \in \mathcal{X}\ .$$

- A collection of measurable subsets, $\xi = \{C_\alpha \in \mathcal{X} | \alpha \in \mathcal{I}\}$ is called a **measurable partition** of X if
  1. $\mu(X \setminus \cup_{\alpha\in I} C_\alpha) = 0$, i.e. the partition "contains" the whole measure.
  2. $\mu(C_{\alpha_1} \cap C_{\alpha_2}) = 0$ if $\alpha_1 \neq \alpha_2$, i.e. the cells C_α of the partition are disjoint.

  The entropy of μ with resepect to the partition ξ is then

$$H(\xi) := H_\mu(\xi) = -\sum_{\alpha\in I} \mu(C_\alpha) \log \mu(C_\alpha) \geq 0\ .$$

- By observing the time series $\{x_i\}$ with partition ξ we are generating a symbol sequence α_i
- Because $p(\alpha) = \mu(C_\alpha)$ we can write the entropy of the partition also as

$$H(\xi) = H(\alpha) = -\sum_{\alpha \in \mathcal{I}} p(\alpha) \log p(\alpha)$$

- Joint partition for two partitions $\xi = \{C_\alpha | \alpha \in I\}$ and $\eta = \{D_\beta | \beta \in J\}$

$$\xi \vee \eta := \{C \cap D | C \in \xi, D \in \eta, \mu(C \cap D) > 0\}$$

- Joint partition of ξ and its preimages under F

$$\xi_{-n}^{F} := \xi \vee F^{-1}(\xi) \vee \ldots \vee F^{-n+1}(\xi) \ .$$

- What corresponds then to ξ_{-n}^{F}? Being in a cell of this partition means that the trajectory was at time n in C_{α_n}, at $n-1$ in $C_{\alpha_{n-1}}$ and so on.

- Thus the measure of one cell of this partition corresponds to the joint probability $p(\alpha_n, \alpha_{n-1}, \ldots, \alpha_1)$.
- The **metric entropy** of the transformation F relative to the partition ξ

$$h(F,\xi) := h_\mu(F,\xi) := \lim_{n\to\infty} \frac{1}{n} H(\xi^F_{-n})$$

- The *KS-entropy* of F with respect to μ is then defined as the supremum over all partitions:

$$h_{KS}(F) := h_\mu(F) := \sup_{\xi, h(\xi)<\infty} h_\mu(F,\xi) \; .$$

A **generating partition** ξ_g is a partition for which the metric entropy is maximal, i.e.

$$h(F,\xi_g) = h_{KS}(F) \; .$$

- No general algorithm to find generating partitions for arbitrary dynamical systems.
- For 1-dimensional maps it is known how to find them and for 2-d also an algorithm exists, which allowed to determine the generating partitions for some well known systems, including the Henon map and the standard map.
- Consider a sequence of partitions ξ_i with diam$(\xi_i) \to 0$. diam$(\xi_i) := \sup_{C\in\xi}$ diam(C). Then $h(F, \xi_i) \to h_{KS}(F)$
- Two limits: Infinite sequence length and infinite resoultion.
- From Lyapunov exponents:

$$h_{KS} = \sum_{\lambda_i > 0} \lambda_i$$

- Starting point: Time series $\{x_k\}$ with N data points.
- Simplest idea: Partition the state space into hypercubes of size ϵ transform youd data into a symbol sequence $\{s_k\}$.
- Estimate an empirical probability distribution by counting the points in the hypercubes

$$p_i = \frac{n_i}{N}$$

- Estimating the conditional entropies

$$\begin{aligned} h(m,\epsilon) =& H(S_m | S_{m-1}, \ldots, S_0) \\ =& H(S_m, S_{m-1}, \ldots, S_1, S_0) - H(S_{m-1}, \ldots, S_1, S_0) \end{aligned}$$

- Starting point: Time series $\{x_k\}$ with N data points.
- Simplest idea: Partition the state space into hypercubes of size ϵ transform youd data into a symbol sequence $\{s_k\}$.
- Estimate an empirical probability distribution by counting the points in the hypercubes

$$p_i = \frac{n_i}{N}$$

- Estimating the conditional entropies

$$\begin{aligned} h(m,\epsilon) =& H(S_m | S_{m-1}, \ldots, S_0) \\ =& H(S_m, S_{m-1}, \ldots, S_1, S_0) - H(S_{m-1}, \ldots, S_1, S_0) \end{aligned}$$

- Problem:

$$\lim_{\epsilon \to 0} H(S_m, \ldots, S_1, S_0) = \log N$$

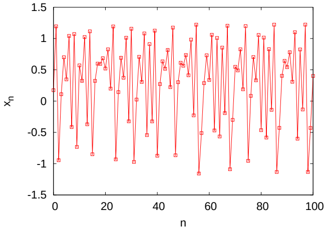
$$\lim_{\epsilon \to 0} h(m, \epsilon) = 0$$

- Henon map

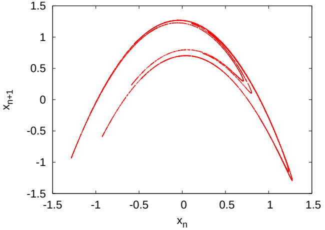
$$x_{n+1} = 1.4x_n^2 - 0.3x_{n-1}$$

- In the following 10 000 data points

Time series (part)

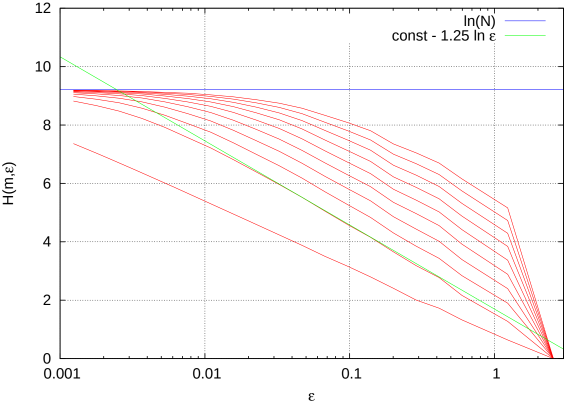


Delay plot



Block entropies

$$H(m,\epsilon) = H(S_m, S_{m-1}, \ldots, ..., S_1)$$

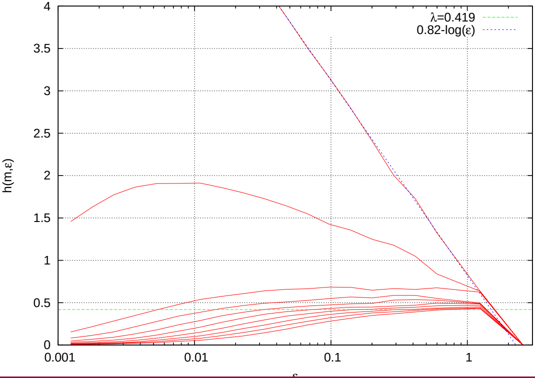


Block entropies

$$H(m,\epsilon) = H(S_m, S_{m-1}, \ldots, ..., S_1)$$

Conditional entropies

$$h(m,\epsilon) = H(m+1,\epsilon) - H(m,\epsilon)$$



Two problems:

- Finite sample bias. There is no genera (but many specific) solutions to resolve this problem for estimating the Shannon entropy (rate).
- Limits $m \to \infty$ and $\epsilon \to 0$ cannot be performed with finite data.
  1. A positive entropy rate for finite m does not mean, that it remains non-zero for $m \to \infty$ until the system is Markovian.
  2. A positive value for finite ϵ does not imply a non-zero value for $\epsilon \to 0$.

Practical solution:

- Looking for convergence in m, i.e. approximate Markovianity
- Looking for plateaus wrt ϵ.

Two problems:

- Finite sample bias. There is no genera (but many specific) solutions to resolve this problem for estimating the Shannon entropy (rate).
- Limits $m \to \infty$ and $\epsilon \to 0$ cannot be performed with finite data.
  1. A positive entropy rate for finite m does not mean, that it remains non-zero for $m \to \infty$ until the system is Markovian.
  2. A positive value for finite ϵ does not imply a non-zero value for $\epsilon \to 0$.

Practical solution:

- Looking for convergence in m, i.e. approximate Markovianity
- Looking for plateaus wrt ϵ. Why is this useful?

Two problems:

- Finite sample bias. There is no genera (but many specific) solutions to resolve this problem for estimating the Shannon entropy (rate).
- Limits $m \to \infty$ and $\epsilon \to 0$ cannot be performed with finite data.
  1. A positive entropy rate for finite m does not mean, that it remains non-zero for $m \to \infty$ until the system is Markovian.
  2. A positive value for finite ϵ does not imply a non-zero value for $\epsilon \to 0$.

Practical solution:

- Looking for convergence in m, i.e. approximate Markovianity
- Looking for plateaus wrt ϵ. Why is this useful?

$\Rightarrow$ Let's look at dimensions!

- The box counting dimension or capacity of a set S in a metric space:

$$D_0 := -\lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \epsilon}$$

with $N(\epsilon)$ being the umber of boxes of side length ϵ that is required to cover the set.

- Note that partitioning the space phace into hypercubes of side length ϵ created such boxes.
- Renyi dimensions are a generalization of the box counting dimension $(q = 0)$:

$$D_q = -\lim_{\epsilon \to 0} \frac{\log \sum_i p_i^q}{(1-q) \log \epsilon}$$

Information dimension

- Applying l'Hospital's rule to the case $q = 1$:

$$D_1 = \lim_{\epsilon \to 0} \frac{\sum_i p_i \log p_i}{\log \epsilon}$$
$$= - \lim_{\epsilon \to 0} \frac{H(\epsilon)}{\log \epsilon}$$

Information dimension

- Applying l'Hospital's rule to the case $q = 1$:

$$D_1 = \lim_{\epsilon \to 0} \frac{\sum_i p_i \log p_i}{\log \epsilon}$$
$$= - \lim_{\epsilon \to 0} \frac{H(\epsilon)}{\log \epsilon}$$

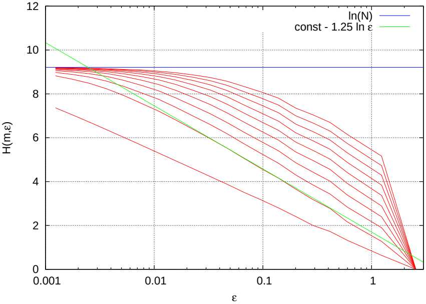
- With finite data $\rightarrow$ looking for "scaling regions"

Information dimension

- Applying l'Hospital's rule to the case $q = 1$:

$$D_1 = \lim_{\epsilon \to 0} \frac{\sum_i p_i \log p_i}{\log \epsilon} = -\lim_{\epsilon \to 0} \frac{H(\epsilon)}{\log \epsilon}$$

- With finite data $\to$ looking for "scaling regions"



Dimension and entropie

- Asymptotic behavior of the entropy controlled by the dimension

$$H(\epsilon) \approx (\text{const}) - D_1 \log \epsilon$$

- We can do the same also for the other Renyi dimensions by defining Renyi entropies

$$H_q = \frac{1}{1-q} \log \sum_i p^q$$

Note that the Renyi-entropies for $q \neq 1$ do not have the "additivity" property.

- Using coverings instead of partitions gives also the Renyi dimensions, but defines another set of entropies (Hentschel and Procaccia 1983)

$$H'_q(\epsilon) = \frac{1}{1-q} \log \left(\frac{1}{N} \sum_i \mu(B(x_i, \epsilon))^{q-1} \right)$$

Correlation dimension and entropy

- Correlation sum: Number of pairs of points in phase space with a distance $\leq \epsilon$.

$$C(m,\epsilon,N) = \frac{2}{(N-m)\cdot(N-m-1)} \sum_{i=1}^{N-m} \sum_{j=i+1}^{N} \Theta(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||)$$

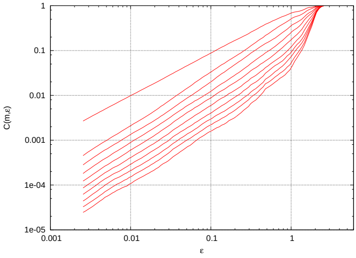
- Correlation dimension for finite $epsilon$

$$\begin{aligned} D_2(m,\epsilon) &= \frac{d \log C(m,\epsilon)}{d \log \epsilon} \\ &= \lim_{\Delta\to 0} \lim_{N\to\infty} \frac{\epsilon}{C(m,\epsilon,N)} \frac{C(m,\epsilon+\Delta,N) - C(m,\epsilon,N)}{\Delta} \\ D_2 &= \lim_{\epsilon\to 0} D_2(m,\epsilon) \end{aligned}$$

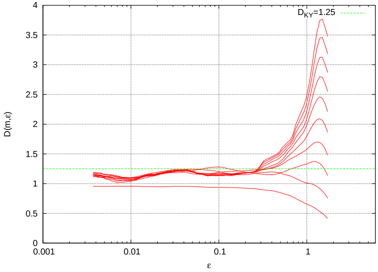
- Correlation entropy

$$H'_{q=2}(m,\epsilon) = -\log C(m,\epsilon)$$

Correlation sum

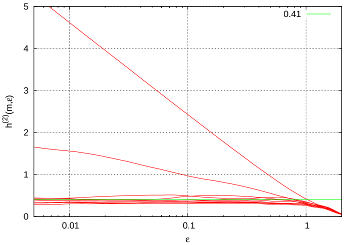


Correlation dimension

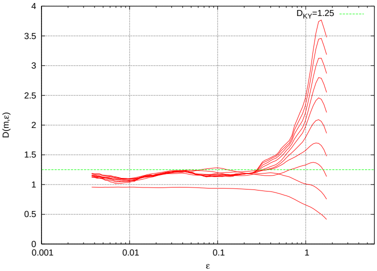


- No finite sample bias

Conditional entropy

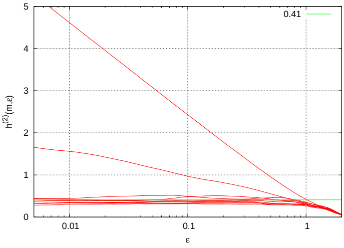


Correlation dimension

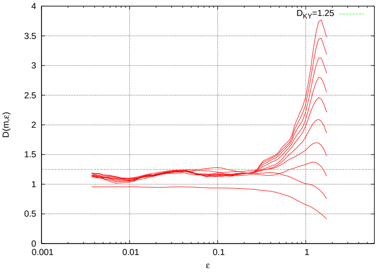


- No finite sample bias
- Can be also used to estimate an entropy rate using the correlation entropy

Conditional entropy

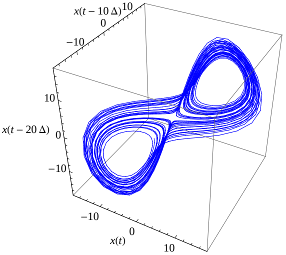


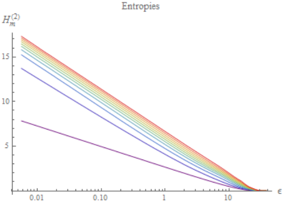
Correlation dimension



- No finite sample bias
- Can be also used to estimate an entropy rate using the correlation entropy
- Requires again the identification of a scaling range

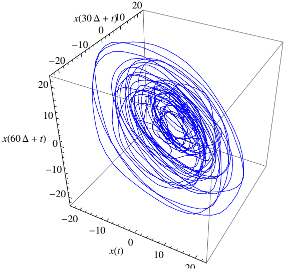
Deterministic: Lorenz attractor

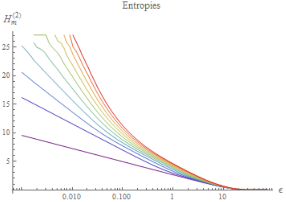




$$H_m(\epsilon) \approx \text{const} - D \log \epsilon$$
for $m > D$

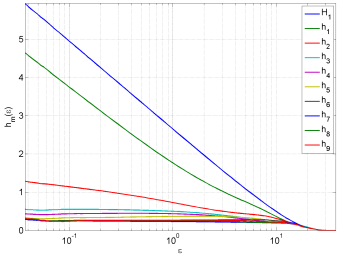
Stochastic: AR(2) model





$$H_m(\epsilon) \approx H_m^C - m \log \epsilon$$
$$H_m^C = -\int \rho(\mathbf{x}) \log \rho(\mathbf{x}) d\mathbf{x}$$

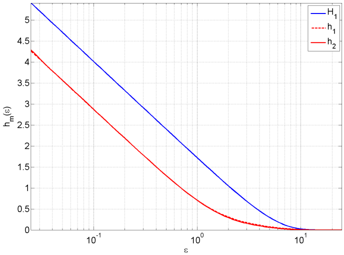
Deterministic: Lorenz system



- Entropy rate asymptotically not depending on ϵ

Stochastic: AR(1)

$$x_{n+1} = a x_n + \xi_{n+1}$$



- Entropy rate $\propto -\log \epsilon$
- Asymptotic behavior from differential entropies

	Deterministic non-chaotic	Deterministic chaotic	Stochastic
Dimension	finite	finite	embedding dimension
Entropy rate	zero	finite	diverges

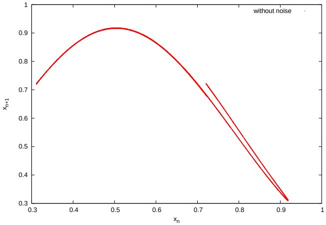
Noisy Gauss map

$$x_{n+1} = \exp(-a(x_n - 0.5)^2) - bx_{n-1} + \sigma\xi$$

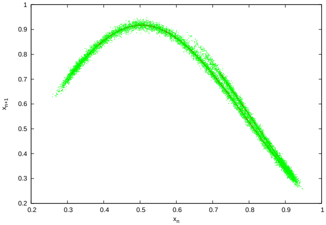
with $a = 5.8 \quad b = 0.1$ and $\sigma = 0.01$.

Delay plot

Without noise



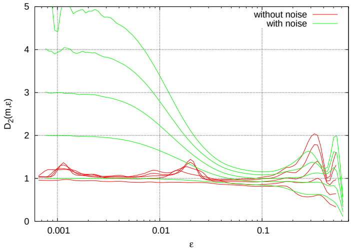
with noise



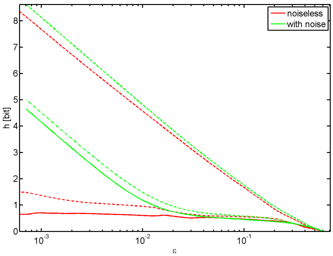
Noisy Gauss map

$$x_{n+1} = \exp(-a(x_n - 0.5)^2) - bx_{n-1} + \sigma\xi$$

Correlation dimension



Correlation entropies



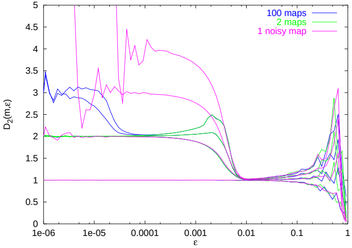
$\Rightarrow$ Stochastic system that looks deterministic on large scales

Unidirectionally coupled tent maps:

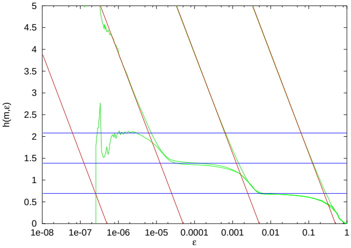
$$x_i(n+1) = (1-\sigma)f(x_i(n)) + \sigma f(x_{i-1}(n))$$

with $f(x) = 1 - |2(x - 1/2)|$ being the tent map.

Dimension

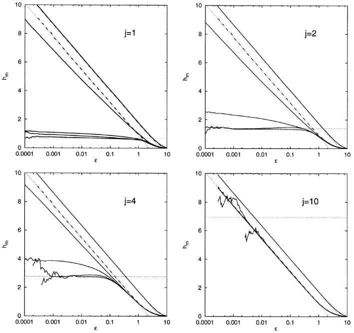


Entropy



E. Olbrich, R. Hegger and H. Kantz, Analysing local observations of weakly coupled maps, Physics Letters A244 (1998), 538-544.

Deterministic system behaving stochastically on large length scales



H. Kantz and E. Olbrich, The transition from deterministic chaos to a stochastic process, Physica A 253 (1998), 105-107.

$$x_{n+1} = 1 - 2x_n^2$$
$$y_{n+1} = \lambda y_n + \nu x_n$$

with $\lambda = e^{-\gamma\tau}$ and $\nu = \sqrt{\tau}$ converges for $\tau \to 0$ to the Ornstein-Uhlenbeck process

$$dY = -\gamma Y\,dt + dW$$

if observed at constant sampling time $\Delta = j\tau$ with j being the time delay for observing y_n.

Relevant for random number generators.

- Piecewise linear, but discontinous maps with an absolute value of the slope of the pieces < 1.
- Related phenomenon: stable chaos - systems that are linearly stable, but appear to be chaotic for finite size perturbations, e.g. neural net models. Can be characterized by finite size Lyapunov exponents (FSLE).
- Coupling a large number of harmonic oscillators with the correct amplitudes and frequencies could generate the behavior of Brownian motion or other dissipative processes on large scales.

M. Cencini, M. Falcioni, E. Olbrich, H. Kantz, and A. Vulpiani. Chaos or noise: Difficulties of a distinction. Physical Review E 62.1 (2000): 427.

Summary

- Dynamical systems can show different behavior on different scales

Summary

- Dynamical systems can show different behavior on different scales
- If we analyze time series from real-world systems and not from mathematical models it is not useful to ask, whether the system **is** deterministic or stochastic, chaotic or non-chaotic. Instead one can ask, how the system is **behaving** on different scales.

Summary

- Dynamical systems can show different behavior on different scales
- If we analyze time series from real-world systems and not from mathematical models it is not useful to ask, whether the system **is** deterministic or stochastic, chaotic or non-chaotic. Instead one can ask, how the system is **behaving** on different scales.
- Brownian motion appears as a stochastic process on the observational accessible scales. Such behvior can be created by stochastic, deterministic chaotic or deterministic non-chaotic (but high-dimensional) models. These differences would be only visible on much smaller scales.

Summary

- Dynamical systems can show different behavior on different scales
- If we analyze time series from real-world systems and not from mathematical models it is not useful to ask, whether the system **is** deterministic or stochastic, chaotic or non-chaotic. Instead one can ask, how the system is **behaving** on different scales.
- Brownian motion appears as a stochastic process on the observational accessible scales. Such behvior can be created by stochastic, deterministic chaotic or deterministic non-chaotic (but high-dimensional) models. These differences would be only visible on much smaller scales.
- There practical limits for the estimation of entropies and dimensions, in particular for high-dimensional systems, because the rquirements for the number of points increase exponentially.