Dynamical systems, information and time series

Stefano Marmi
Scuola Normale Superiore
http://homepage.sns.it/marmi/

Lecture 4- European University Institute

October 27, 2009

- Lecture 1: An introduction to dynamical systems and to time series. Periodic and quasiperiodic motions. (Sept 18)
- Lecture 2: A priori probability vs. statistics: ergodicity, uniform distribution of orbits. The analysis of return times. Kac inequality. Mixing (Sep 25)
- Lecture 3: Shannon and Kolmogorov-Sinai entropy. Randomness and deterministic chaos. Relative entropy and Kelly's betting. (Oct 9)
- Lecture 4: Time series analysis and embedology: can we distinguish deterministic chaos in a noisy environment? (Tuesday, Oct 27, 11am-1pm)
- Lecture 5: Fractals and multifractals. (Nov 6, 3pm-5pm)

Today’s references:

- Daniel Kaplan and Leon Glass: "Understanding Nonlinear Dynamics" Springer (1995) Chapter 6
- Sauer, Yorke, Casdagli: Embedology. J. Stat. Phys. 65 (1991) 579-616
- Michael Small "Applied Nonlinear Time Series Analysis" World Scientific
- Holger Kantz and Thomas Schreiber “Nonlinear Time Series Analysis” Cambridge University Press (2004)

The slides of all lectures will be available at my personal webpage: http://homepage.sns.it/marmi/

An overview of today's lecture

- Entropy of Bernoulli schemes and of Markov chains
- Lyapunov exponent of one-dimensional maps
- Stochastic and deterministic random time series: examples
- Takens theorem
- Embedology

Entropy

In probability theory, *entropy* quantifies the uncertainty associated to a random process

Consider an experiment with mutually esclusive outcomes $A=\{a_{1,} \ldots, a_k\}$

- Assume that the probability of a_i is p_i , $0 \leq p_i \leq 1$, $p_1 + \ldots + p_k = 1$
- If a_1 has a probability very close to 1, then in most experiments the outcome would be a_1 thus the result is not very uncertain. One doea not gain much information from performing the experiment.
- One can quantify the "surprise" of the outcome as

$$\text{information} = -\log(\text{probability})$$

- (the intensity of a perception is proportional to the logarithm of the intensity of the stimulus)

Entropy

The entropy associated to the experiment is

$H=-\sum p_i \log p_i$

Since

information = - Log (probability)

entropy is simply the expectation value of the information produced by the experiment

Entropy, coding and data compression

What does entropy measure?

Entropy quantifies the information content (namely the amount of randomness of a signal)

Entropy : a completely random binary sequence has entropy= $\log_2 2 = 1$ and cannot be compressed

Computer file= infinitely long binary sequence

Entropy = *best possible compression ratio*

Lempel-Ziv algorithm (Compression of individual sequences via variable rate coding, IEEE Trans. Inf. Th. 24 (1978) 530-536): does not assume knowledge of probability distribution of the source and achieves asymptotic compression ratio=entropy of source

Entropy of a dynamical system (Kolmogorov-Sinai entropy)

Given two partitions $\mathcal{P}$ and $\mathcal{Q}$

$\mathcal{P} \vee \mathcal{Q}$ the **join** of $\mathcal{P}$ and $\mathcal{Q}$

$B \cap C$ where $B \in \mathcal{Q}$ and $C \in \mathcal{Q}$

$T : X \to X$ measure preserving

$$\mathcal{P}_n = \mathcal{P} \vee T^{-1}\mathcal{P} \vee \cdots \vee T^{-(n-1)}\mathcal{P}$$

$$h(T,\mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}_n) \qquad h(T) = \sup_{\mathcal{P}} h(T,\mathcal{P})$$

Entropy of Bernoulli schemes

Let $N \geq 2$, $\Sigma_N = \{1, \dots N\}^{\mathbb{Z}}$.

$$d(x,y) = 2^{-a(x,y)} \quad \text{where } a(x,y) = \inf\{|n|\,,\, n \in \mathbb{Z}\,,\, x_n \neq y_n\}$$

$$\textit{shift } \sigma \;:\; \Sigma_N \to \Sigma_N \qquad \sigma((x_i)_{i\in\mathbb{Z}}) = (x_{i+1})_{i\in\mathbb{Z}}$$

The topological entropy of (Σ_N, σ) is $\log N$

$$(p_1, \dots, p_N) \in \Delta^{(N)} \qquad \nu(\{i\}) = p_i$$

Definition 4.26 *The* Bernoulli scheme $BS(p_1, \dots, p_N)$ *is the measurable dynamical system given by the shift map* $\sigma : \Sigma_N \to \Sigma_N$ *with the (product) probability measure* $\mu = \nu^{\mathbb{Z}}$ *on* Σ_N.

Proposition 4.27 *The Kolmogorov–Sinai entropy of the Bernoulli scheme* $BS(p_1, \dots, p_N)$ *is* $-\sum_{i=1}^{N} p_i \log p_i$.

Topological Markov chains or subshifts of finite type

$$\Sigma_A = \{x \in \Sigma_N \,,\, (x_i, x_{i+1}) \in \Gamma \, \forall i \in \mathbb{Z}\} \qquad \Gamma \subset \{1, \ldots N\}^2$$

Σ_A is a compact shift invariant subset of Σ_N

$A = A_\Gamma$ the $N \times N$ matrix with entries $a_{ij} \in \{0, 1\}$

$$a_{ij} = \begin{cases} 1 & \iff (i,j) \in \Gamma \\ 0 & \text{otherwise} \end{cases}$$

The restriction of the shift σ to Σ_A is denoted σ_A

$A^m = (a_{ij}^m)$ and $a_{ij}^m > 0$ for all i, j

Entropy of Markov chains

Theorem 4.35 (Perron–Frobenius, see [Gan]) If A is primitive then there exists an eigenvalue $\lambda_A > 0$ such that :

(i) $|\lambda_A| > \lambda$ for all eigenvalues $\lambda \neq \lambda_A$;

(ii) the left and right eigenvectors associated to λ_A are strictly positive and are unique up to constant multiples ;

(iii) λ_A is a simple root of the characteristic polynomial of A.

the topological entropy of σ_A is $\log \lambda_A$ (clearly $\lambda_A > 1$ since all the integers $a_{ij}^m > 0$)

Let $P = (P_{ij})$ be an $N \times N$ matrix such that

(i) $P_{ij} \geq 0$ for all i, j, and $P_{ij} > 0 \iff a_{ij} = 1$;

(ii) $\sum_{j=1}^{N} P_{ij} = 1$ for all $i = 1, \ldots, N$;

(iii) P^m has all its entries strictly positive.

Such a matrix is called a *stochastic matrix*. Applying Perron–Frobenius theorem to P we see that 1 is a simple eigenvalue of P and there exists a normalized eigenvector $p = (p_1, \ldots, p_N) \in \Delta^{(N)}$ such that $p_i > 0$ for all i and

$$\sum_{i=1}^{N} p_i P_{ij} = p_j \ , \quad \forall\, 1 \leq i \leq N \ .$$

We define a probability measure μ on Σ_A corresponding to P prescribing its value on the cylinders :

$$\mu\left(C\begin{pmatrix} j_0, \ldots, j_k \\ i, \ldots, i+k \end{pmatrix}\right) = p_{j_0} P_{j_0 j_1} \cdots P_{j_{k-1} j_k} \ ,$$

for all $i \in \mathbb{Z}$, $k \geq 0$ and $j_0, \ldots, j_k \in \{1, \ldots, N\}$. It is called the *Markov measure* associated to the stochastic matrix P.

: the subshift σ_A preserves the Markov measure μ.

$$h_\mu(\sigma_A) = -\sum_{i,j=1}^{N} p_i P_{ij} \log P_{ij} \qquad\qquad h_\mu(\sigma_A) \leq h_{top}(\sigma_A)$$

Lyapunov exponent for a map of an interval

- Assume that T is a piecewise smooth map of I=[0,1]
- By the chain rule we have

$$\frac{1}{n}\log|T^n(x)-T^n(y)| \approx \frac{1}{n}\sum_{i=0}^{n-1}\log|T'(T^i x)| \,.$$

- If μ is an ergodic invariant measure for a.e. x the limit exists and it is given by $\int_0^1 \log|T'|\,\mathrm{d}\mu$

it is also called the Lyapunov exponent of T

Expanding maps and Rokhlin formula

If T is expanding then it has a unique a.c.i.p.m. μ and the entropy h of T w.r.t. μ is equal to the Lyapunov exponent

$$h = \int_0^1 \log |T'(x)| \, \mathrm{d}\mu$$

Examples of time-series in natural and social sciences

- Weather and climate measurements (temperature, pressure, rain, wind speed, …)
- Earthquakes
- Lightcurves of variable stars
- Sunspots
- Macroeconomic historical time series (inflation, GDP, unemployment,…)
- Financial time series (stocks, futures, commodities, bonds, …)
- Populations census (humans or animals)
- Physiological signals (ECG, EEG, …)

Stochastic or chaotic?

- An important goal of time-series analysis is to determine, given a times series (e.g. HRV) if the underlying dynamics (the heart) is:
  - Intrinsically random
  - Generated by a deterministic nonlinear chaotic system which generates a random output
  - A mix of the two (stochastic perturbations of deterministic dynamics)

CHAOS 18, 030201 (2008)

Announcement: A new feature—"Controversial Topics in Nonlinear Science: Is the Normal Heart Rate Chaotic?"

Leon Glass
Department of Physics, McGill University, Montréal, Québec H3G 1Y6, Canada

(Received 16 June 2008; published online 6 August 2008)

The normal heart rhythm in humans is set by a small group of cells called the sinoatrial node. Although over short time intervals, the normal heart rate often appears to be regular, when the heart rate is measured over extended periods of time, it shows significant fluctuations. There are a number of factors that affect these fluctuations: changes of activity or mental state, presence of drugs, presence of artificial pace- makers, occurrence of cardiac arrhythmias that might mask the sinoatrial rhythm or make it difficult to measure. Following the widespread recognition of the possibility of deterministic chaos in the early 1980s, considerable attention has been focused on the possibility that heart rate variability might reflect deterministic chaos in the physiological control system regulating the heart rate. A large number of papers related to the analysis of heart rate variability have been published in Chaos and elsewhere. However, there is still considerable debate about how to characterize fluctuations in the heart rate and the significance of those fluctuations. There has not been a forum in which these disagreements can be aired. Accordingly, Chaos invites submissions that address one or more of the following questions:

- Is the normal heart rate chaotic?
- If the normal heart rate is not chaotic, is there some more appropriate term to characterize the fluctuations e.g., scaling, fractal, multifractal?
- How does the analysis of heart rate variability elucidate the underlying mechanisms controlling the heart rate?
- **Do any analyses of heart rate variability provide clinical information that can be useful in medical assessment e.g., in helping to assess the risk of sudden cardiac death**. If so, please indicate what additional clinical studies would be useful for measures of heart rate variability to be more broadly accepted by the medical community.

Chaotic brains at work!

C.R. Acad. Sci. Paris, Sciences de la vie / Life Sciences 324 (2001) 773–793
© 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés
S0764446901013774/REV

Point sur / Concise review

Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation

Philippe Faure, Henri Korn*

Biologie cellulaire et moléculaire du neurone (Inserm V261), Institut Pasteur, 25 rue Docteur Roux, 75724 Paris Cedex 15, France

Received 18 June 2001; accepted 2 July 2001

Communicated by Pierre Buser

Abstract – In the light of results obtained during the last two decades in a number of laboratories, it appears that some of the tools of nonlinear dynamics, first developed and improved for the physical sciences and engineering, are well-suited for studies of biological phenomena. In particular it has become clear that the different regimes of activities undergone by nerve cells, neural assemblies and behavioural patterns, the linkage between them, and their modifications over time, cannot be fully understood in the context of even integrative physiology, without using these new techniques. This

networks and in the study of higher brain functions, will be critically reviewed. It will be shown that the tools of nonlinear dynamics can be irreplaceable for revealing hidden mechanisms subserving, for example, neuronal synchronization and periodic oscillations. The benefits for the brain of adopting chaotic regimes with their wide range of potential behaviours and their aptitude to quickly react to changing conditions will also be considered. © 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT•

C. R. Biologies 326 (2003) 787–840

COMPTES RENDUS BIOLOGIES

Neurosciences

Is there chaos in the brain? II. Experimental evidence and related models

Henri Korn *, Philippe Faure

'Récepteurs et Cognition', CNRS 2182, Institut Pasteur, 25, rue du Docteur-Roux, 75724 Paris cedex 15, France

Received 16 September 2003; accepted 17 September 2003

Presented by Pierre Buser

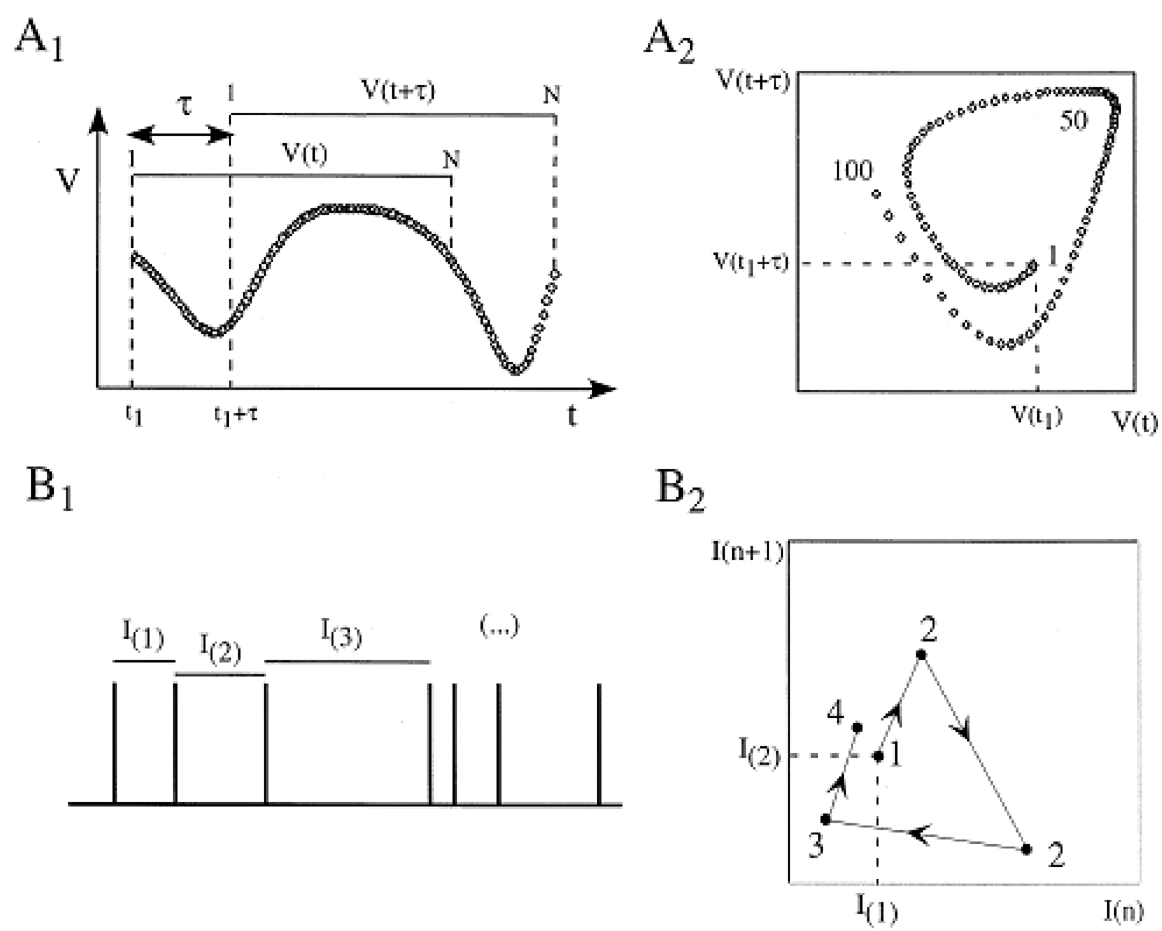


Figure 9. Reconstruction of phase spaces with the delay method. (A1-A2) Case of a continuous signal, as for example the recording of membrane potential, V. (A1) The time series is subdivided into two sequences of measurements of the same length N (here equal to 100 points). Their starting point is shifted by the time lag τ. (A2) The trajectory in a two dimensional phase space is obtained by plotting, for each point of the time series, V_t against $V_{t+\tau}$. (B1-B2) In the case of a discrete signal, such as time intervals between action potentials in a spike train (B1), the same procedure is applied to time intervals $I_1, I_2 \ldots I_N$ (B2).

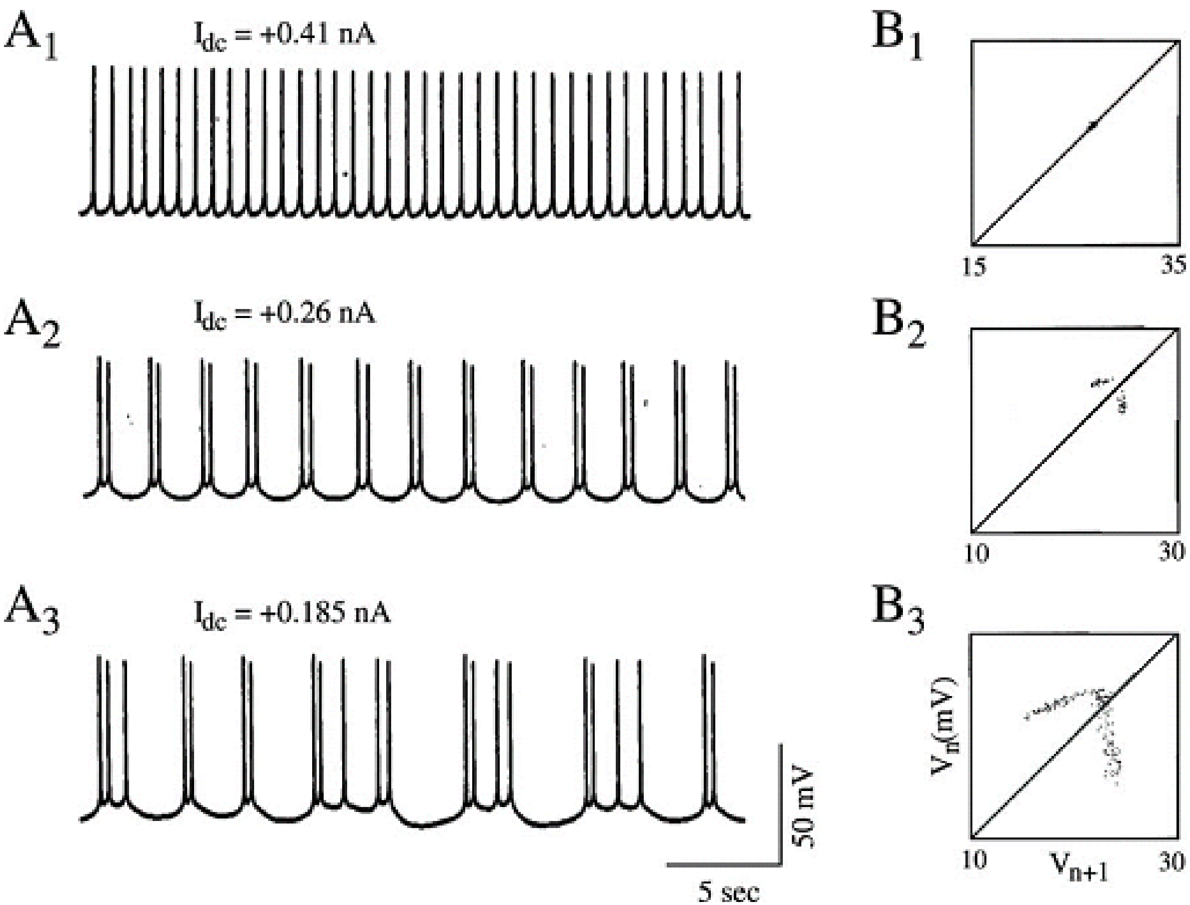


Fig. 5. Discharge patterns of a pacemaker neuron caused by a dc current (**A1–A3**) representative samples of the recorded membrane potential. (**B1–B3**) One-dimensional Poincaré maps of the corresponding sequence of spikes constructed using the delay method (see [1] for explanations). (**A1–B1**) Regular discharges of action potentials. (**A2–B2**) Periodic firing with two spikes per burst. (**A3–B3**) Chaotic bursting discharges. (Adapted from [45], with permission of the Journal of Theoretical Biology.)

THE JOURNAL OF FINANCE • VOL. XLVI, NO. 5 • DECEMBER 1991

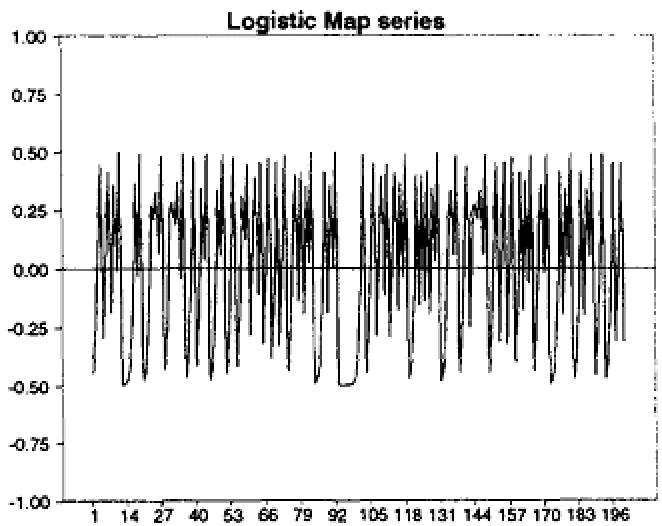
Chaos and Nonlinear Dynamics: Application to Financial Markets

DAVID A. HSIEH*

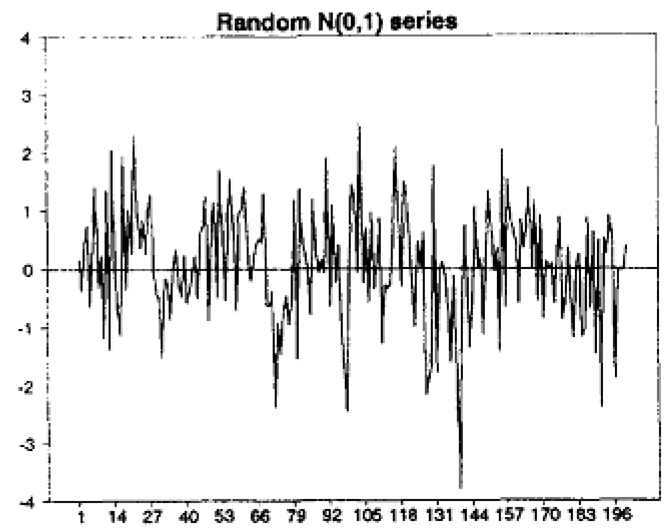
ABSTRACT

After the stock market crash of October 19, 1987, interest in nonlinear dynamics, especially deterministic chaotic dynamics, has increased in both the financial press and the academic literature. This has come about because the frequency of large moves in stock markets is greater than would be expected under a normal distribution. There are a number of possible explanations. A popular one is that the stock market is governed by chaotic dynamics. What exactly is chaos and how is it related to nonlinear dynamics? How does one detect chaos? Is there chaos in financial markets? Are there other explanations of the movements of financial prices other than chaos? The purpose of this paper is to explore these issues.

Logistic map series (adjusted with mean)



Random N(0,1) series



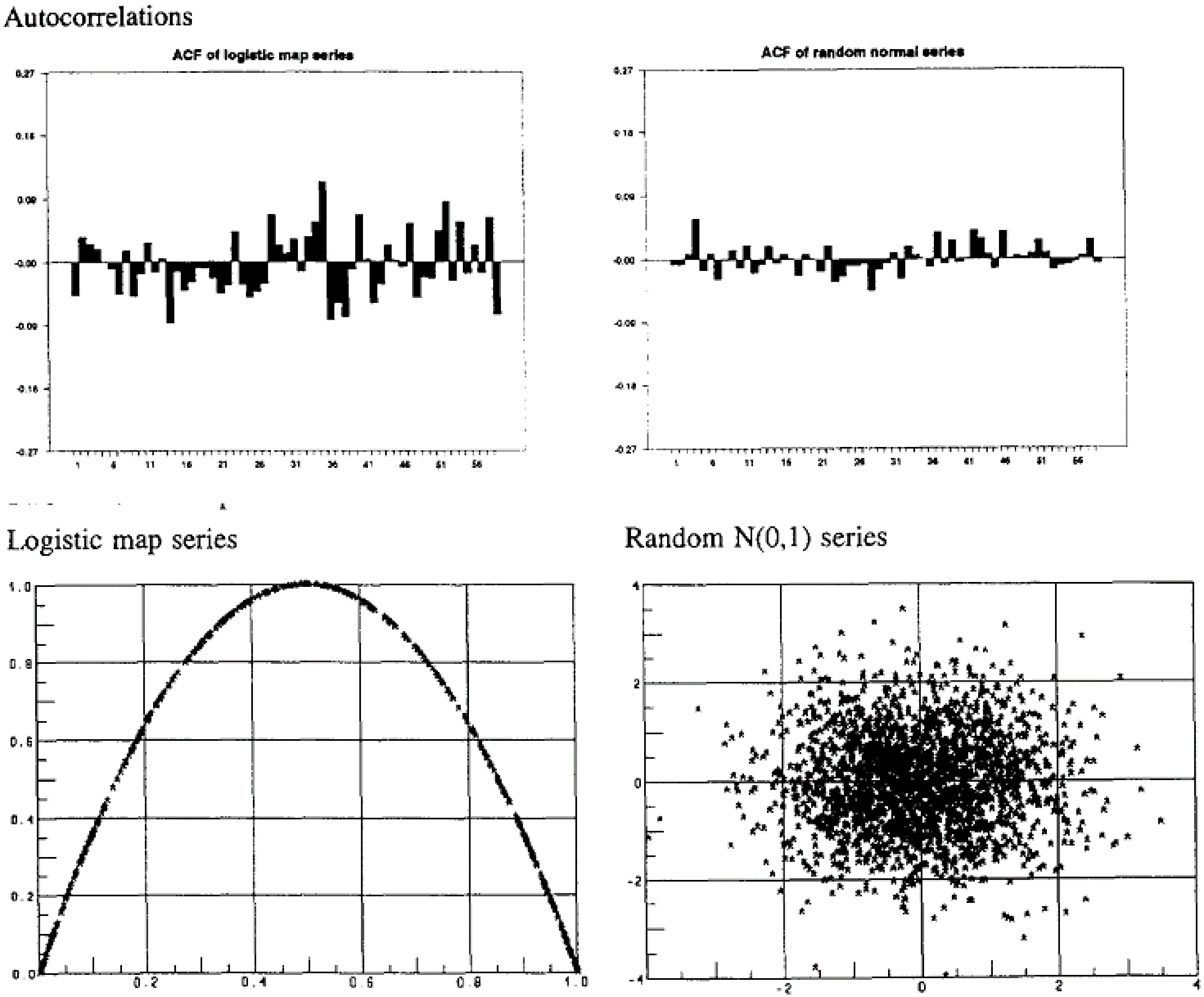


Fig. 2. Comparison of logistic map and random series.

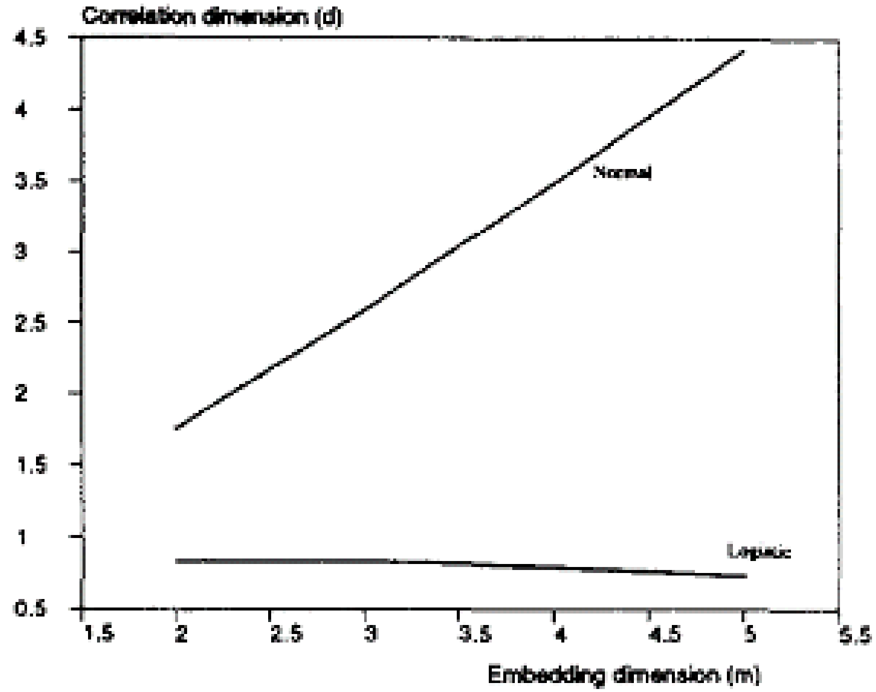
K. Takala, M. Virén / European Journal of Operational Research 93 (1996) 155–172

Embedding dimension = m

$$\boxed{C_m(\varepsilon)=\lim_{N\to\infty}\frac{1}{N^2}\#\left\{(x_{m,i},x_{m,j}),\left\|x_{m,i}-x_{m,j}\right\|<\varepsilon\right\}}$$

$$d(m)=\lim_{\varepsilon\to 0}\frac{\log C_m(\varepsilon)}{\log(\varepsilon)}$$

Correlation dimensions of logistic map and random normal processes



III. What Do We Find in the Stock Market?

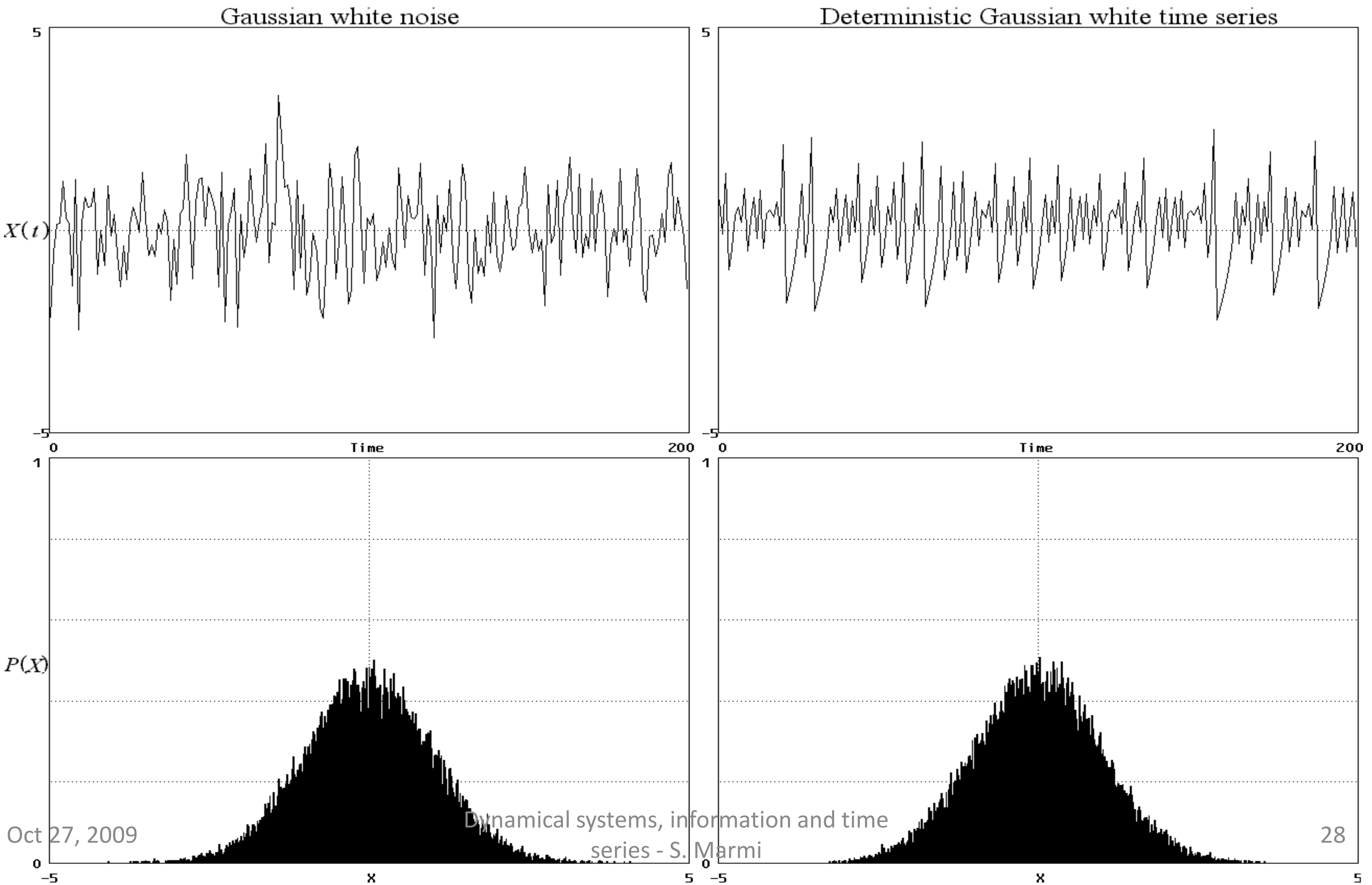
Scheinkman and LeBaron (1989) used the Grassberger-Procaccia plots and calculated the correlation dimension of weekly stock returns. They found that the slope of $\log C_n(\epsilon)$ versus $\log \epsilon$ appears to be around 6, even for dimensions as high as 25. They, however, noted that this is not sufficient evidence of chaos in stock returns, because there are a number of problems with this graphical procedure.

First, Scheinkman and LeBaron (1989) pointed out that some nonlinear stochastic model, such as Engle's (1982) autoregressive conditional heteroskedasticity (ARCH) model, exhibit "dependence" similar to that of chaotic maps. Using data from the ARCH model, they showed that the slopes of the graphs of $\log{}_n C(\epsilon)$ versus $\log \epsilon$ increase at a rate slower than n.

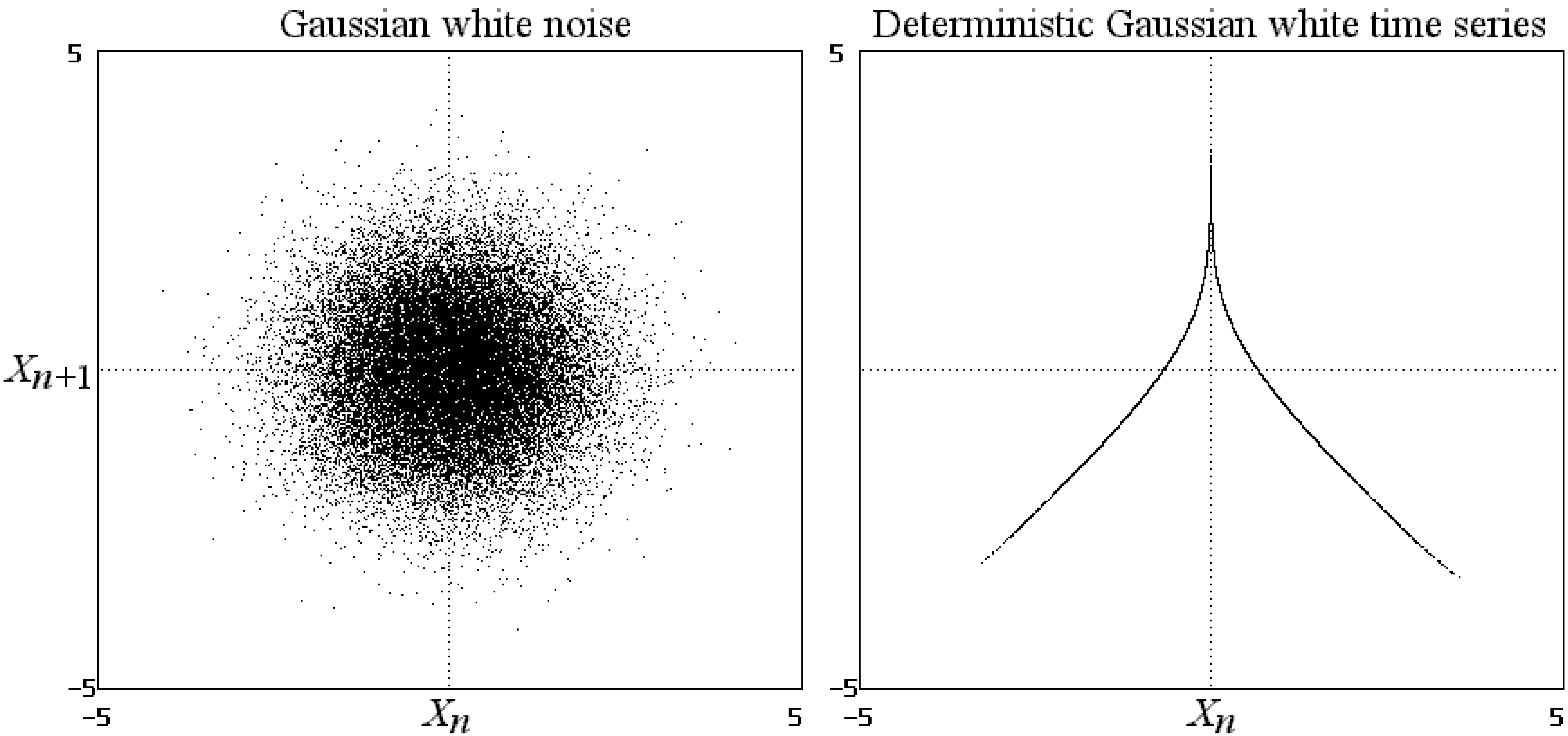
Second, there is no way to verify that a process has an infinite correlation dimension using a finite amount of data. Scientists typically use 100,000 or more data points to detect low dimensional chaotic system. Financial economists have substantially fewer points. The largest data sets generally have 2,000 observations. If we use the imbedding dimension of 10, we have only 200 nonoverlapping 10-histories. It is very hard to say whether 200 10-histories "fill up" a 10-dimensional space.

Hsieh, JOF 1991

Deterministic or random? Appearance can be misleading…



Time delay map



Source: **sprott**.physics.wisc.edu/lectures/**tsa**.ppt

Logit and logistic

The logistic map $x \rightarrow L(x)=4x(1-x)$ preserves the probability measure $d\mu(x)=dx/(\pi\sqrt{x(1-x)})$

The transformation $h:[0,1] \rightarrow \mathbf{R}$, $h(x)=\ln x-\ln(1-x)$ conjugates L with a new map G

$$h \quad L=G \quad h$$

definined on **R.** The new invariant probability measure is $d\mu(x)=dx/[\pi(e^{x/2} + e^{-x/2})]$

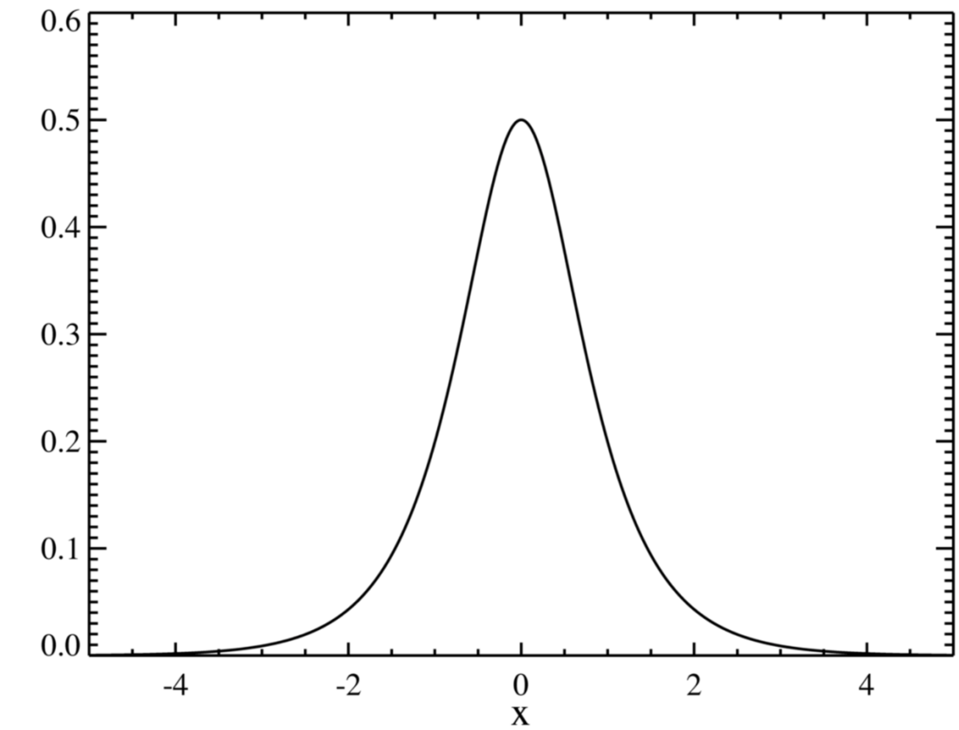
G and L have the same dynamics (the only difference is a coordinates change)

Statistical analysis of a time series: moments of the probability distribution

mean	$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$
variance	$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$
standard deviation	σ
skewness	$\zeta = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma}\right)^3$
kurtosis	$\kappa = -3 + \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma}\right)^4$

Higher moments: simmetry of the distribution and fat tails

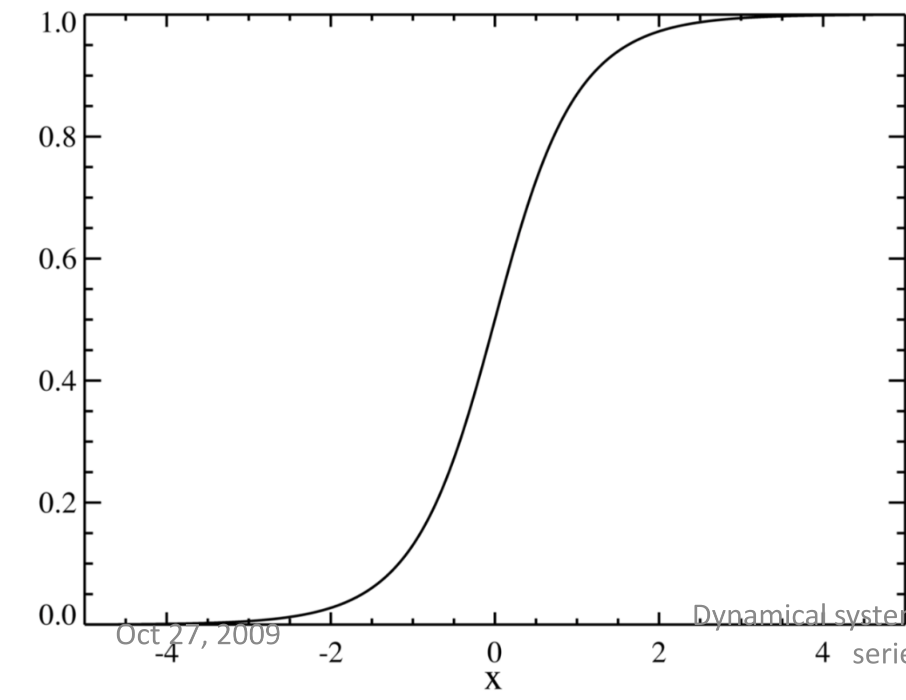
- Skewness: measures simmetry of the data about the mean (third moment)
- Kurtosis: peakedness of the distribution relative to the normal distribution (hence the -3 term)
- Leptokurtic distribution (fat tailed): has positive kurtosis



Hyperbolic secant distribution

Source: wikipedia

Parameters	*none*
Support	$x \in (-\infty, +\infty)$
Probability density function (pdf)	$\frac{1}{2}\operatorname{sech}(\frac{1}{2}\pi x)$
Cumulative distribution function (cdf)	$\frac{2\arctan(\exp(\frac{1}{2}\pi x))}{\pi}$
Mean	0
Median	0
Mode	0
Variance	1
Skewness	0
Excess kurtosis	2
Entropy	$4/\pi\, G \approx 1.16624$



G = 0.915 965 594 177 219 015 054 603 514 932 384 110 774... Catalan's constant

Takens theorem

- $\phi : X \rightarrow X$ map, $f : X \rightarrow R$ smooth observable
- Time-delay map (reconstruction of the dynamics from periodic sampling):
- $F(f,\phi) : X \rightarrow R^n$ n is the number of delays
- $F(f,\phi)(x) = (f(x), f(\phi(x)), f(\phi \;\, \phi(x)), \ldots, f(\phi^n (x)))^{-1}$
- Under mild assumptions if the dynamics has an attractor with dimension k and n>2k then for almost any choice of the observable the reconstruction map is injective

Immersions and embeddings

- A smooth map F on a compact smooth manifold A is an immersion if the derivative map DF(x) (represented by the Jacobian matrix of F at x) is one-to-one at every point x∊A. Since DF(x) is a linear map, this is equivalent to DF(x) having full rank on the tangent space. This can happen whether or not F is one-to-one. Under an immersion, no differential structure is lost in going from A to F(A).
- An embedding of A is a smooth diffeomorphism from A onto its image F(A), that is, a smooth one-to-one map which has a smooth inverse. For a compact manifold A, the map F is an embedding if and only if ,F is a one- to-one immersion.
- The set of embeddings is open in the set of smooth maps: arbitrarily small perturbations of an embedding will still be embeddings!

Embedology (Sauer, Yorke, Casdagli, J. Stat. Phys. 65 (1991)

Whitney showed that a generic smooth map ,F from a d-dimensional smooth compact manifold M to R^n , n>2d is actually a diffeomorphism on M. That is, M and F(M) are diffeomorphic. We generalize this in two ways:

- first, by replacing "generic" with "probability-one" (in a prescribed sense),
- second, by replacing the manifold M by a compact invariant set A

contained in some Rk that may have noninteger box-counting dimension (boxdim). In that case, we show that almost every smooth map from a neighborhood of A to R^n is one-to-one as long as n>2 * boxdim(A)

We also show that almost every smooth map is an embedding on compact subsets of smooth manifolds within l. This suggests that embedding techniques can be used to compute positive Lyapunov exponents (but not necessarily negative Lyapunov exponents). The positive Lyapunov exponents are usually carried by smooth unstable manifolds on attractors.

Embedology (Sauer, Yorke, Casdagli, J. Stat. Phys. 65 (1991)

Takens dealt with a restricted class of maps called delay-coordinate maps: these are time series of a single observed quantity from an experiment. He showed (F. Takens, Detecting strange attractors in turbulence, in Lecture Notes in Mathematics, No. 898 (Springer-Verlag, 1981) that if the dynamical system and the observed quantity are generic, then the delay-coordinate map from a d-dimensional smooth compact manifold M to R^n , n>2d is a diffeomorphism on M.

- we replace generic with probability-one
- and the manifold M by a possibly fractal set.

Thus, for a compact invariant subset A under mild conditions on the dynamical system, almost every delay-coordinate map to R^n is one-to-one on A provided that n>2.boxdim(A). Also, any manifold structure within I will be preserved in F(A).

- Only C^1 smoothness is needed.;
- For flows, the delay must be chosen so that there are no periodic orbits with period exactly equal to the time delay used or twice the delay

Embedding method

- Plot $x(t)$ vs. $x(t-\tau)$, $x(t-2\tau)$, $x(t-3\tau)$, …
- $x(t)$ can be any observable
- The embedding dimension is the # of delays
- The choice of τ and of the dimension are critical
- For a typical deterministic system, the orbit will be diffeomorphic to the attractor of the system (Takens theorem)

Choice of Embedding Parameters

Theoretically, a time delay coordinate map yields an valid embedding for any sufficiently large embedding dimension and for any time delay when the data are noise free and measured with infinite precision.

But, there are several problems:

(i) Data are not clean
(ii) Large embedding dimension are computationally expensive and unstable and require long time series ($10^5 – 10^6$ points …)
(iii) Finite precision induces noise

Effectively, the solution is to search for:

(i) Optimal time delay t
(ii) Minimum embedding dimension d

or

(i) Optimal time window t_w

There is no one unique method solving all problems and neither there is a unique set of embedding parameters appropriate for all purposes.

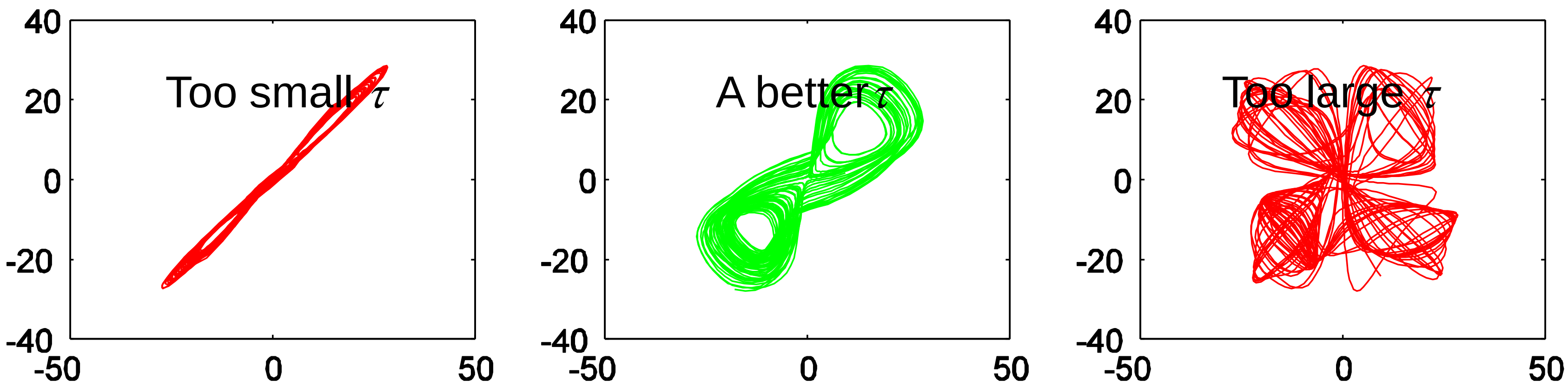
The Role of Time Delay τ

If τ is too small, $x(t)$ and $x(t-\tau)$ will be very close, then each reconstructed vector will consist of almost equal components → *Redundancy (τ_R)*

→ The reconstructed state space will collapse into the main diagonal

If τ is too large, $x(t)$ and $x(t-\tau)$ will be completely unrelated, then each reconstructed vector will consist of irrelevant components → *Irrelevance (τ_I)*

→ The reconstructed state space will fill the entire state space.



http://www.viskom.oeaw.ac.at/~joy/March22,%202004.ppt

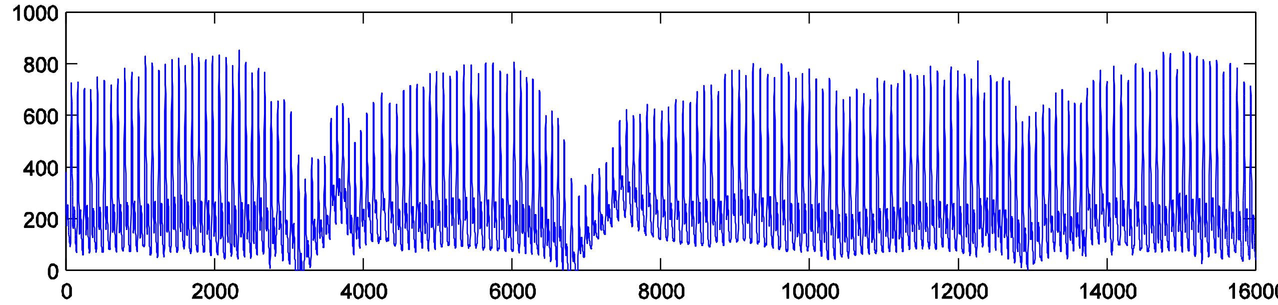
A better choice is:

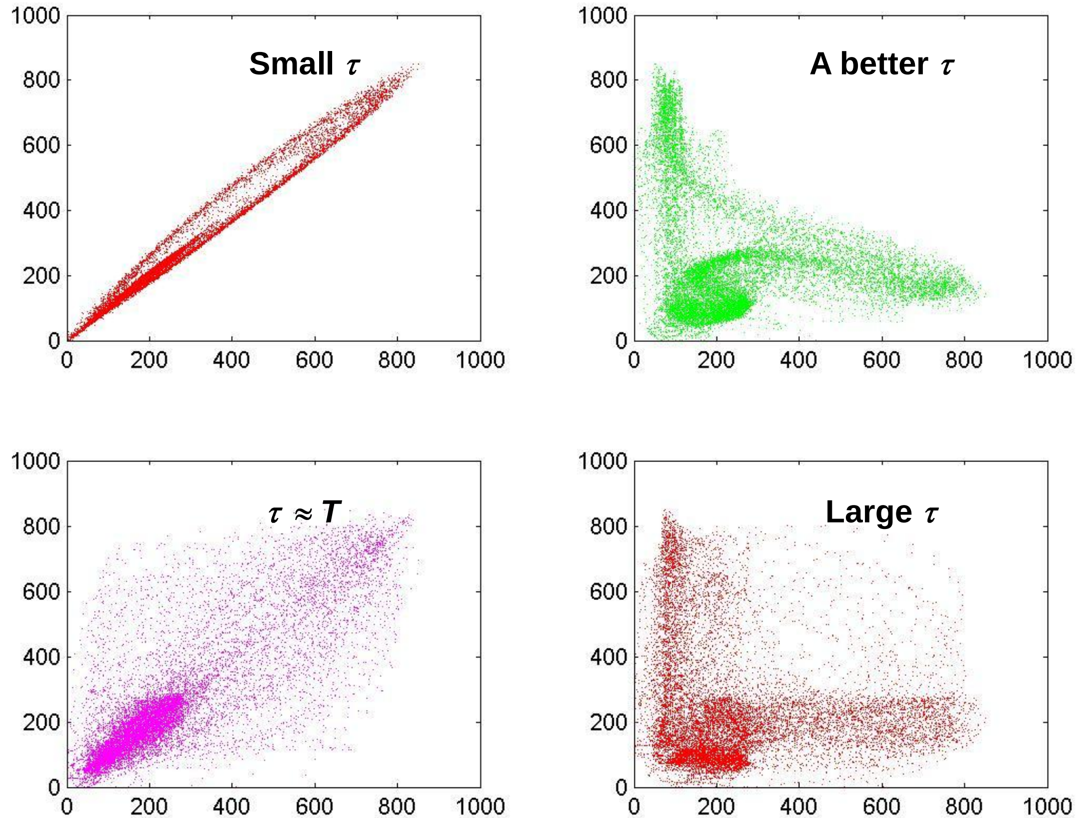
$\tau_R < \tau_w < \tau_I$

Caution: τ should not be close to main period

Collapsing of state space

Blood Pressure Signal





http://www.viskom.oeaw.ac.at/~joy/March22,%202004.ppt

Some Recipes to Choose τ_{opt}

Based on Autocorrelation

Estimate autocorrelation function:

$$C(\tau) = \frac{1}{N-\tau-1}\sum_{t=0}^{N-\tau-1} x(t)x(t+\tau) = \langle x(t)x(t+\tau)\rangle$$

Then, $t_{opt} \approx C(0)/e$ ***or*** first zero crossing of $C(t)$

Modifications:

1. Consider minima of higher order autocorrelation functions, $<x(t)x(t+\mathrm{t})x(t+2t)>$ and then look for time when these minima for various orders coincide.
2. Apply nonlinear autocorrelation functions: $<x^2(t)x^2(t+2t)>$

http://www.viskom.oeaw.ac.at/~joy/March22,%202004.ppt

Some Recipes to Choose τ_{opt}

Based on Time delayed Mutual Information

The information we have about the value of $x(t+\tau)$ if we know $x(t)$.

1. Generate the histogram for the probability distribution of the signal $x(t)$.
2. Let p_i is the probability that the signal will be inside the i-th bin and $p_{ij}(t)$ is the probability that $x(t)$ is in i-th bin and $x(t+\tau)$ is in j-th bin.
3. Then the mutual information for delay τ will be

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \log p_{ij}(\tau) - 2\sum_{i} p_i \log p_i$$

For $\tau \rightarrow 0$, $I(\tau) \rightarrow$ Shannon's Entropy

$\tau_{opt} \approx$ First minimum of $I(\tau)$