Dynamics and time series: theory and applications

Stefano Marmi
Scuola Normale Superiore

Lecture 1, Nov 7, 2011

Dynamics and time series: theory and applications

- Introduction to dynamical systems (deterministic and stochastic) and their use in the study of time series, especially economics and financial time series
- Introduction to dynamical systems and time series. Stationary states. Periodic and quasiperiodic motions. Ergodicity, uniform distribution of orbits. Return times, Kac inequality. Mixing. Shannon entropy. Kolmogorov-Sinai entropy. Lyapunov exponents. Entropy and information theory. Markov chains. Mutual information, relative entropy. Reconstruction of attractors from time series: Takens' theorem. Gambling, probabilistic games, risk management and Kelly criterion.
- Stochastic processes, autoregressive models, random walks, Brownian motion. Ordinary Least Squares and Maximum Likelihood. Granger causality. Correlation and autocorrelation. Stylized facts for financial time series. Volatility, heterosckedasticity, ARCH and GARCH.
- Introduction to R programming and its use for time series analysis

- ***Lectures on OLS, Maximum Likelihood, AR, ARMA, ARCH, GARCH: Fulvio Corsi (Swiss Finance Institute)***
- ***Introduction to R programming language and laboratory: Luigi Bianchi (SNS)***
- ***Dynamical Systems:*** lecture notes; Michael Brin and Garrett Stuck: *Introduction to Dynamical Systems*, Cambridge University Press 2003
- ***Information Theory:*** Thomas Cover, Joy: Thomas *Elements of Information Theory*, 2nd edition, Wiley 2006
- ***Time series:*** Holger Kantz and Thomas Schreiber: *Nonlinear Time Series Analysis*, Cambridge University Press 2004; Peter Brockwell and Richard Davies: *Time Series: Theory and Methods*, Springer 2nd ed. 2006; Lambert Koopmans: *The Spectral Analysis of Time Series*, Academic Press 1974; Johnatan Cryer and Kung-Sik Chan: *Time Series Analysis with Applications in R*, Springer 2008
- ***Mathematical models in finance and time series analysis:*** John Campbell, Andrew Lo and Craig MacKinlay: *The Econometrics of Financial Markets*, Princeton University Press, 1997; Stephen Taylor: *Asset Price Dynamics, Volatility, and Prediction* Princeton University Press 2005.

Dynamical systems

- A dynamical system is a couple (X phase space, time evolution law: either a map $T:X\to X$ or a flow $g_t :X\to X$, here t is time)
- The phase space X is the set of all possible states (i.e. initial conditions) of our system
- Each initial condition uniquely determines the time evolution (determinism)
- The system evolves in time according to a fixed law (iteration of a map T, flow g_t for example arising from solving a differential equation, etc.)
- Often (but not necessarily) the evolution law is not linear
- Observables are simply scalar functions $\phi:X\to\mathbf{R}$
- Time series naturally arise from the time evolution of the observables:

$\phi(x)$, $\phi(T(x))$, $\phi(T\circ T(x))$, $\phi(T^3 (x))$, ….. Here $T^{n+1}(x)=T \circ T^n (x)$

Examples of dynamical systems in natural and social sciences

- The Solar System
- Atmosphere (meteorology)
- Human body (heart, brain cells, lungs, ...)
- Ecology (dynamics of animal populations)
- Epidemiology
- Chemical reactions

Dynamical systems not necessarily deterministic

- Stockmarket
- Electric grid
- Internet

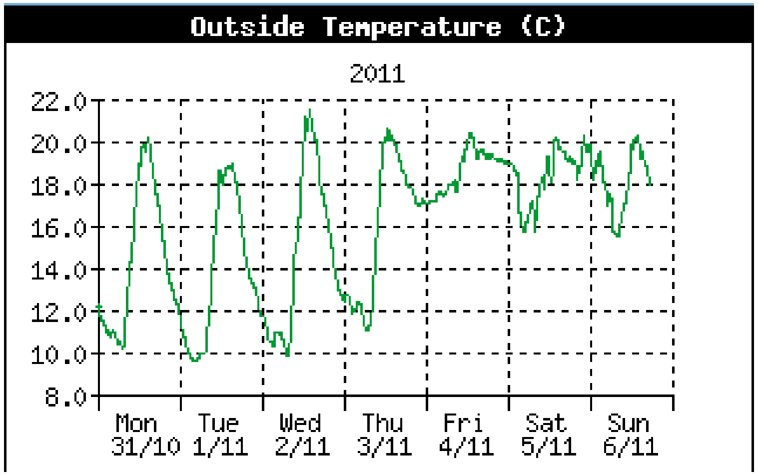
Examples of time-series in natural and social sciences

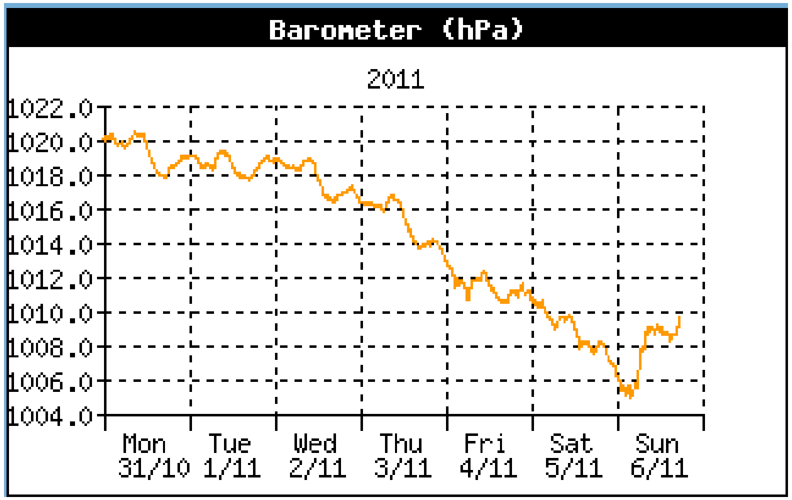
- Weather measurements (temperature, pressure, rain, wind speed, …) . If the series is very long …climate
- Earthquakes
- Lightcurves of variable stars
- Sunspots
- Macroeconomic historical time series (inflation, GDP, employment,…)
- Financial time series (stocks, futures, commodities, bonds, …)
- Populations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

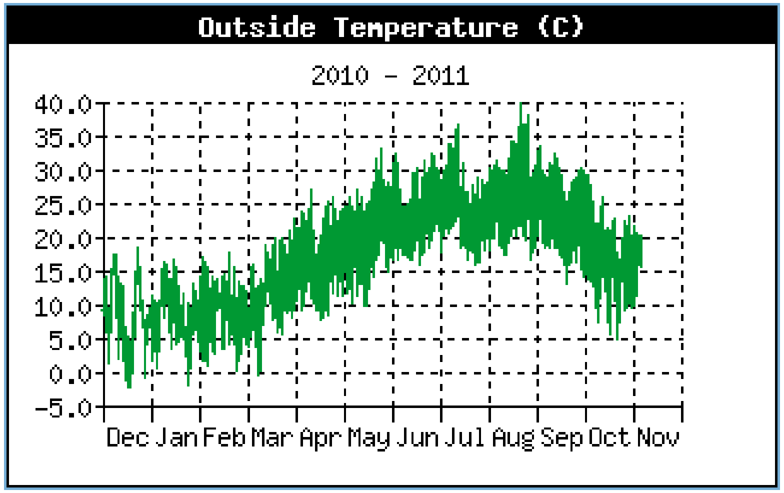
- **Free sources of financial time series (see my homepage)**
- **Reuters/Jefferies CRB Index (commodities)**
- **Oanda.com: FXHistory®: historical currency exchange rates**
- **Yahoo finance page (indexes, stocks, etfs, mutual funds)**
- **Kenneth R. French data library (fantastic!)**
- **Robert Shiller online data (includes montlhy historical S&P500 index value, earnings, long term interest rates, CPI since 1871)**
- **Hedge funds indexes ……………………………….and much more!**

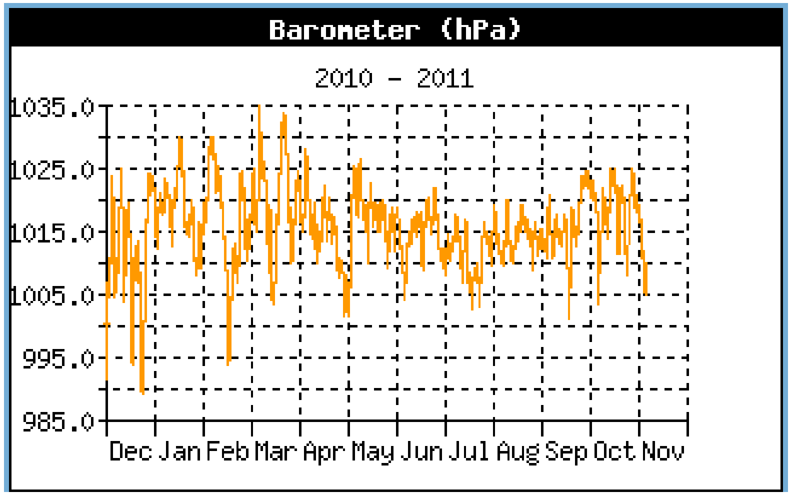
- **High frequency (intraday, 1 min) financial data: ETFs and US stocks**
- **Economic time series:**
- **http://research.stlouisfed.org/fred/**
- **http://www.economicsnetwork.ac.uk/links/data_free.htm**
- **Free sources of metereological data: e.g. meteopisa.it**
- **Free sources of physiological time series: physionet.org**
- **…..**

- Temperature and pressure time series from www.meteopisa.it









Ergodic theory

- The focus of the analysis is mainly on the asymptotic ditribution of the orbits, and not on transient phenomena.
- Ergodic theory is an attempt to study the statistical behaviour of orbits of dynamical systems restricting the attention to their asymptotic distribution.
- One waits until all transients have been wiped off and looks for an invariant probability measure describing the distribution of typical orbits.

Measure theory vs. probability theory

Table 1.1. Comparison of terminology

Measure Theory	Probability Theory
a probability measure space X	a sample space Ω
$x \in X$	$\omega \in \Omega$
a σ-algebra $\mathcal{A}$	a σ-field $\mathcal{F}$
a measurable subset A	an event E
a probability measure μ	a probability P
$\mu(A)$	$P(E)$
a measurable function f	a random variable X
$f(x)$	x, a value of X
a characteristic function χ_E	an indicator function 1_E
Lebesgue integral $\int_X f \, \mathrm{d}\mu$	expectation $E[X]$
almost everywhere	almost surely, or with probability 1
convergence in L^1	convergence in mean
convergence in measure	convergence in probability
conditional measure $\mu_A(B)$	conditional probability $\Pr(B\|A)$

Stochastic or chaotic?

An important goal of time-series analysis is to determine, given a times series if the underlying dynamics 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)

Randomness and the physical law

- It may well be that the universe itself is completely deterministic (though this depends on what the “true” laws of physics are, and also to some extent on certain ontological assumptions about reality), in which case randomness is simply a mathematical concept, modeled using such abstract mathematical objects as probability spaces. Nevertheless, the concept of *pseudorandomness*- objects which “behave” randomly in various statistical senses - still makes sense in a purely deterministic setting. A typical example are the digits of π=3.14159265358979323 85…this is a deterministic sequence of digits, but is widely believed to behave pseudorandomly in various precise senses (e.g. each digit should asymptotically appear 10% of the time). If a deterministic system exhibits a sufficient amount of pseudorandomness, then random mathematical models (e.g. statistical mechanics) can yield accurate predictions of reality, even if the underlying physics of that reality has no randomness in it.

http://terrytao.wordpress.com/2007/04/05/simons-lecture-i-structure-and-randomness-in-fourier-analysis-and-number-theory/

Probability, statistics and the problem of induction

- The probability of an event (if it exists) is almost always impossible to be known a-priori
- The only possibility is to replace it with the frequencies measured by observing how often the event occurs
- The problem of backtesting
- The problem of ergodicity and of typical points: from a single series of observations I would like to be able to deduce the invariant probability
- Bertrand Russell’s chicken (turkey nella versione USA)

Bertrand Russel

(The Problems of Philosophy,
Home University Library, 1912. Chapter VI On Induction) Available at the page http://www.ditext.com/russell/rus6.html

Domestic animals expect food when they see the person who feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.

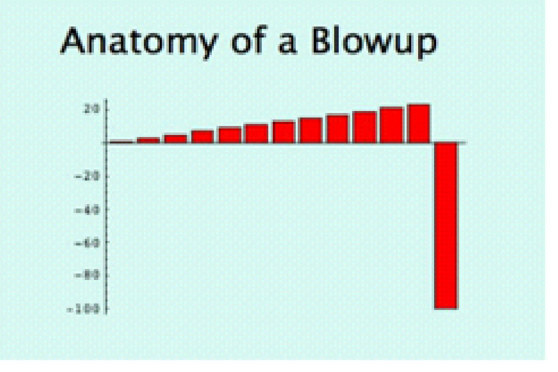


Figure 1 *My classical metaphor: A Turkey is fed for a 1000 days—every days confirms to its statistical department that the human race cares about its welfare "with increased statistical significance". On the 1001st day, the turkey has a surprise.*

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

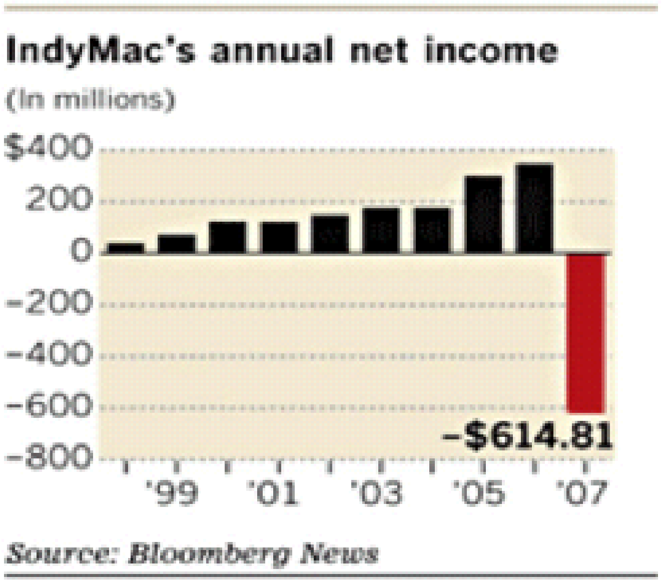


Figure 2 *The graph above shows the fate of close to 1000 financial institutions (includes busts such as FNMA, Bear Stearns, Northern Rock, Lehman Brothers, etc.). The banking system (betting AGAINST rare events) just lost > 1 Trillion dollars (so far) on a single error, more than was ever earned in the history*

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

Payoff from mildly OTM UK Sterling Short Option, 1988-2008

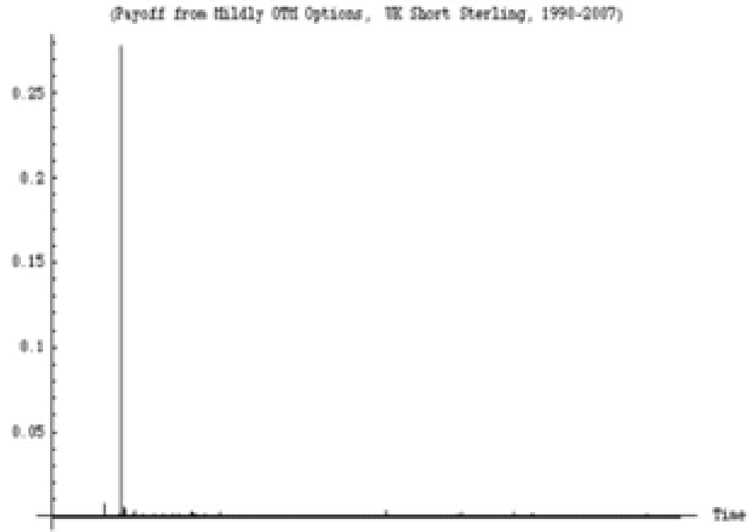
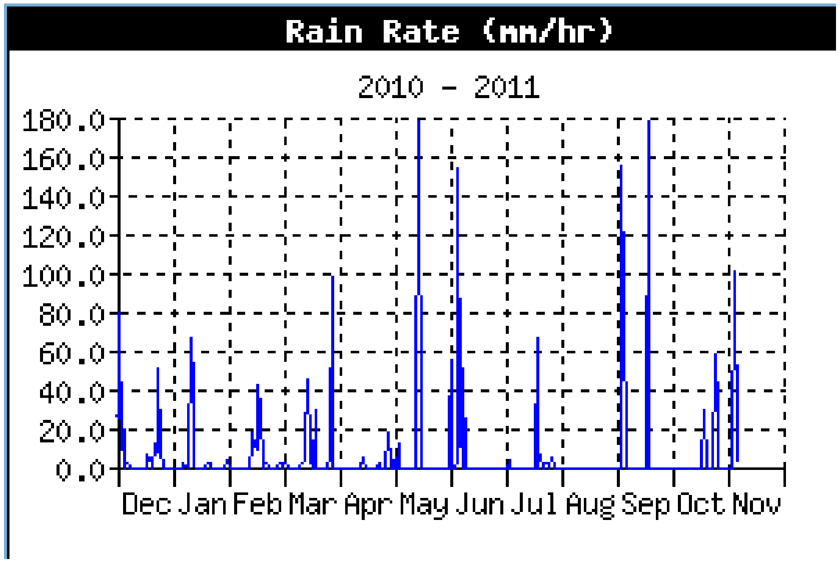
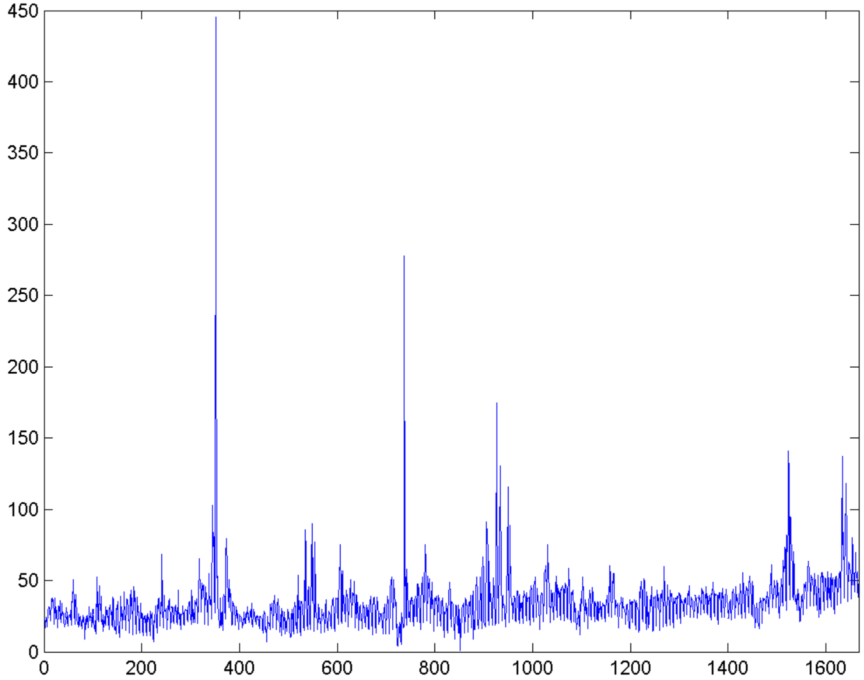


Figure 3 The graph shows the daily variations a derivatives portfolio exposed to U.K. interest rates between 1988 and 2008. Close to 99% of the variations, over the span of 20 years, will be represented in 1 single day—the day the European Monetary System collapsed. As I show in the appendix, this is typical with ANY socio-economic variable (commodity prices, currencies, inflation numbers, GDP, company performance, etc.). No known econometric statistical method can capture the probability of the event with any remotely acceptable accuracy (except, of course, in hindsight, and "on paper"). Also note that this applies to surges on electricity grids and all manner of modern-day phenomena.

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html





Rain rate time series from meteopisa.it,
German electricity prices time series from http://www.econ-pol.unisi.it/~reno/spots.html

Measure-preserving transformations

X phase space, μ probability measure

$\Phi:X \to \mathbf{R}$ observable (a measurable function, say L^2).

Let A be subset of X (event).

$\mu(\Phi) = \int_X \Phi \, d\mu$ is the expectation of Φ

$T:X \to X$ induces a time evolution

on observables: $\Phi \to \Phi \circ T$

on events: $A \to T^{-1}(A)$

T is measure preserving if $\mu(\Phi)= \mu(\Phi \circ T)$ i.e. $\mu(A)=\mu(T^{-1}(A))$

Birkhoff theorem and ergodicity

Birkhoff theorem: if T preserves the measure μ then with probability one the time averages of the observables exist (statistical expectations). The system is ergodic if these time averages do not depend on the orbit (statistics and a-priori probability agree)

$$\frac{1}{N}\sum_{0}^{N-1}\varphi\circ T^{i}(x) := \frac{1}{N}S_N\varphi(x) \longrightarrow \int_X \varphi(t)\,d\mu(t)$$

$$\frac{1}{N}\#\left\{i\in[0,N), T^{i}(x)\in A\right\} \longrightarrow \mu(A)$$

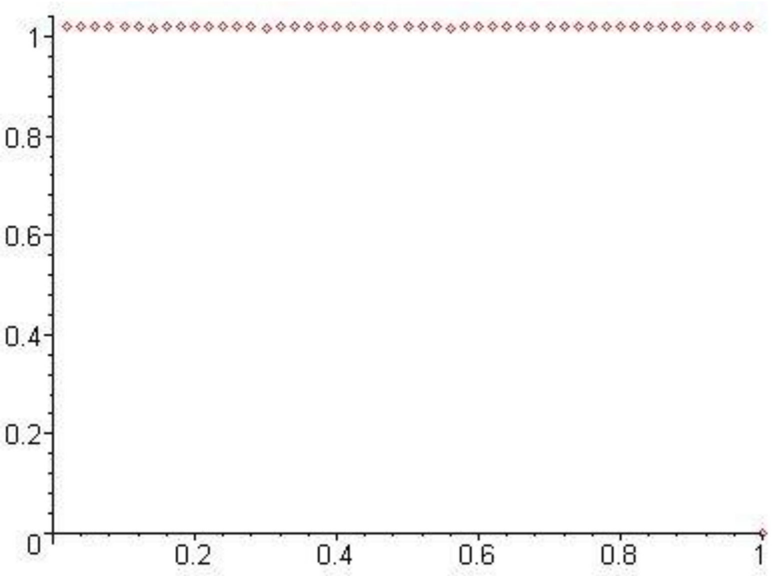
Law of large numbers:
Statistics of orbits = a-priori probability

"Historia magistra vitae" or the mathematical foundation of backtesting

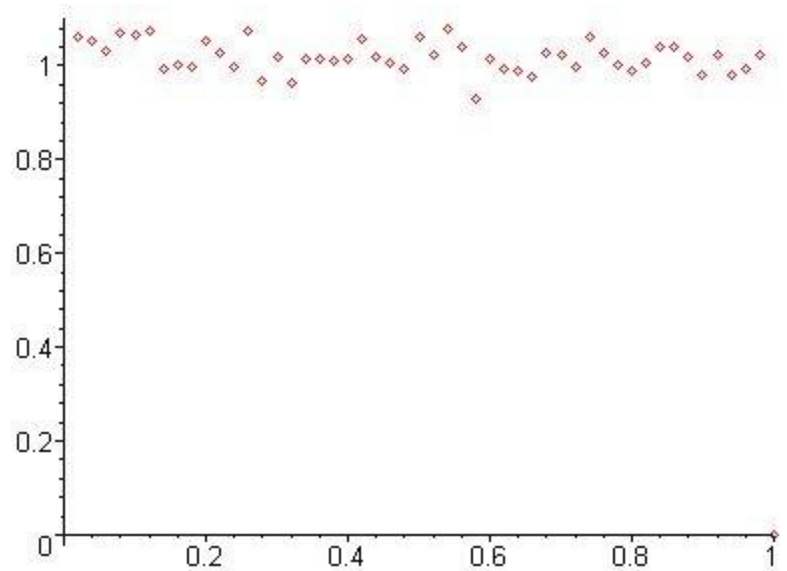
- Without assuming ergodicity, Birkhoff theorem shows that:

- Time averages exist and they give rise to an experimental statistics to compare with theory
- Past and future time averages agree almost everywhere

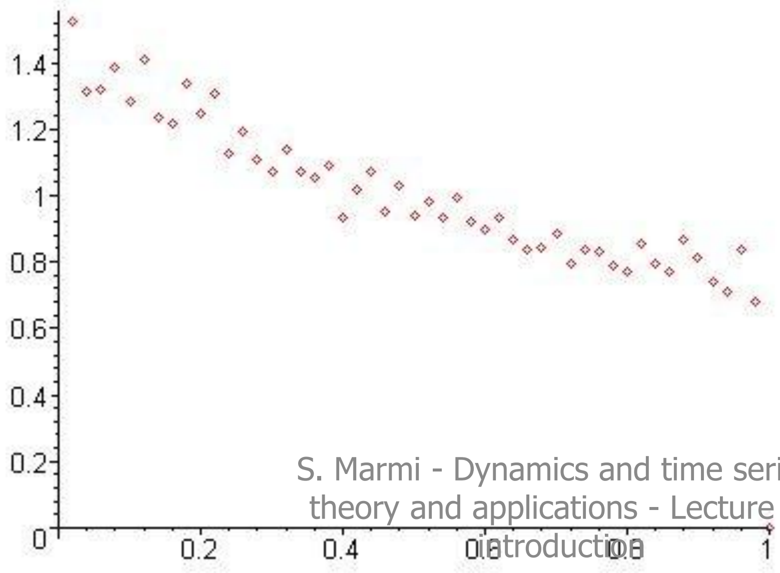
Statistical distribution of frequencies of vists



Rotation

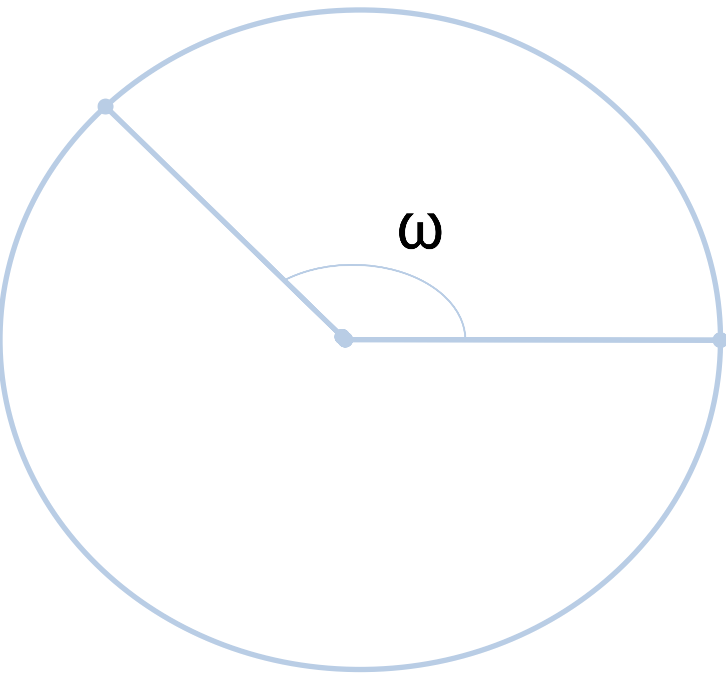


Doubling map



Gauss map

The simplest dynamical systems



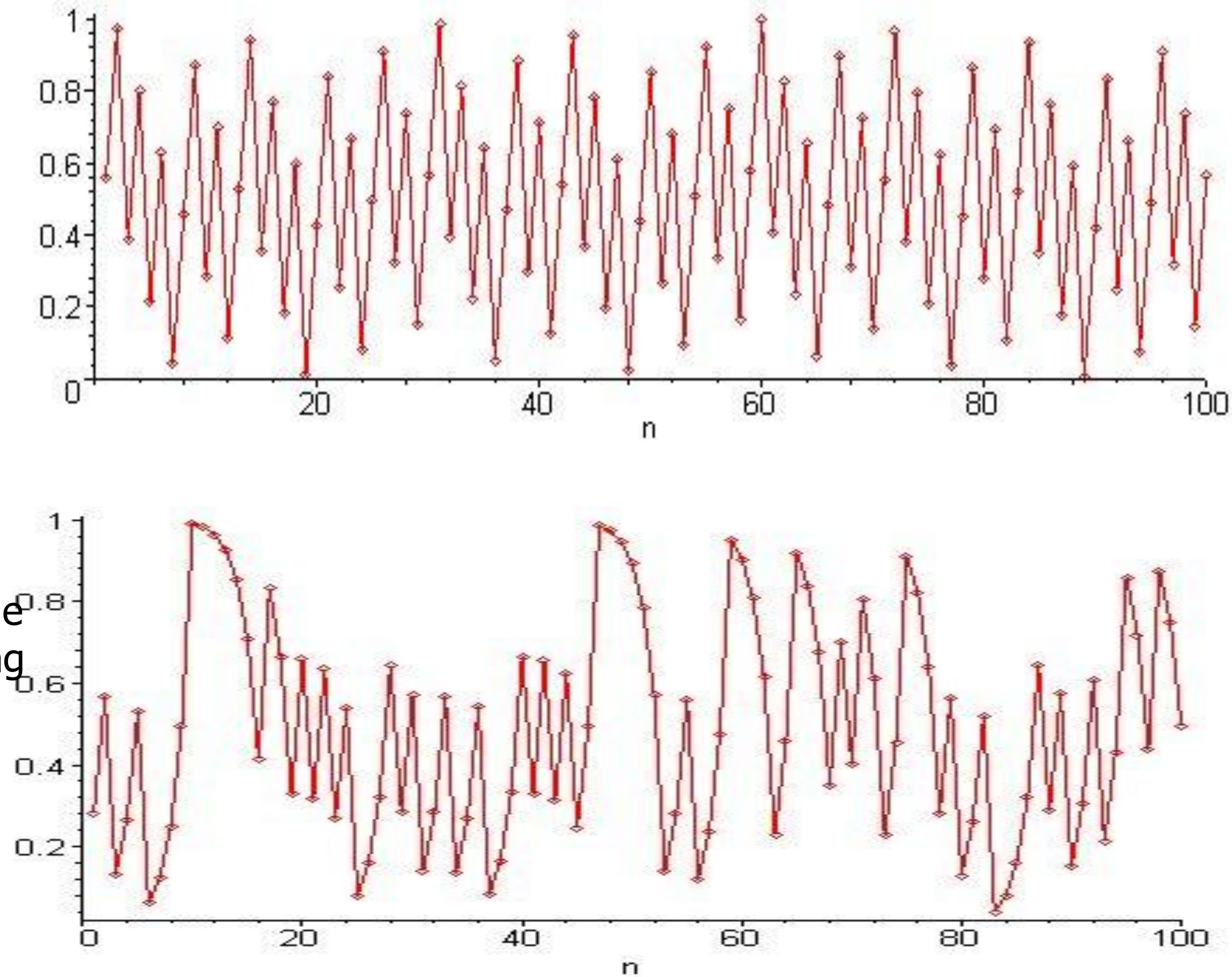
- The phase space is the circle: **S=R/Z**
- Case 1: quasiperiodic dynamics

  $\theta(n+1)=\theta(n)+\omega \pmod 1$

  (ω irrational)
- Case 2: chaotic dynamics

  $\theta(n+1)=2\theta(n) \pmod 1$

100 iterates of the two dynamical systems: the quasiperiodic circle rotation and the chaotic doubling map



Quasiperiodic dynamics

- Quasiperiodic = periodic if precision is finite, but the period $\to\infty$ if the precision of measurements is improved
- More formally a discrete time dynamics f is quasiperiodic if

$$f^{n_k} \to \mathrm{id}$$ For some sequence $n_k \to \infty$

$$f^{n_k+1} \approx f$$

Renormalization approach

$$n_k + 1 \to \infty$$ return times

Sensitivity to initial conditions

For, in respect to the latter branch of the supposition, it should be considered that the most trifling variation in the facts of the two cases might give rise to the most important miscalculations, by diverting thoroughly the two courses of events; very much as, in arithmetic, an error which, in its own individuality, may be inappreciable, produces at length, by dint of multiplication at all points of the process, a result enormously at variance with truth. (Egdar Allan Poe, The mistery of Marie Roget)

For the doubling map on the circle (case 2) one has

$\theta(N)-\theta'(N)=2^N(\theta(0)-\theta'(0))$ ➡ even if the initial datum is known with a 10 digit accuracy, after 40 iterations one cannot even say if the iterates are larger than ½ or not

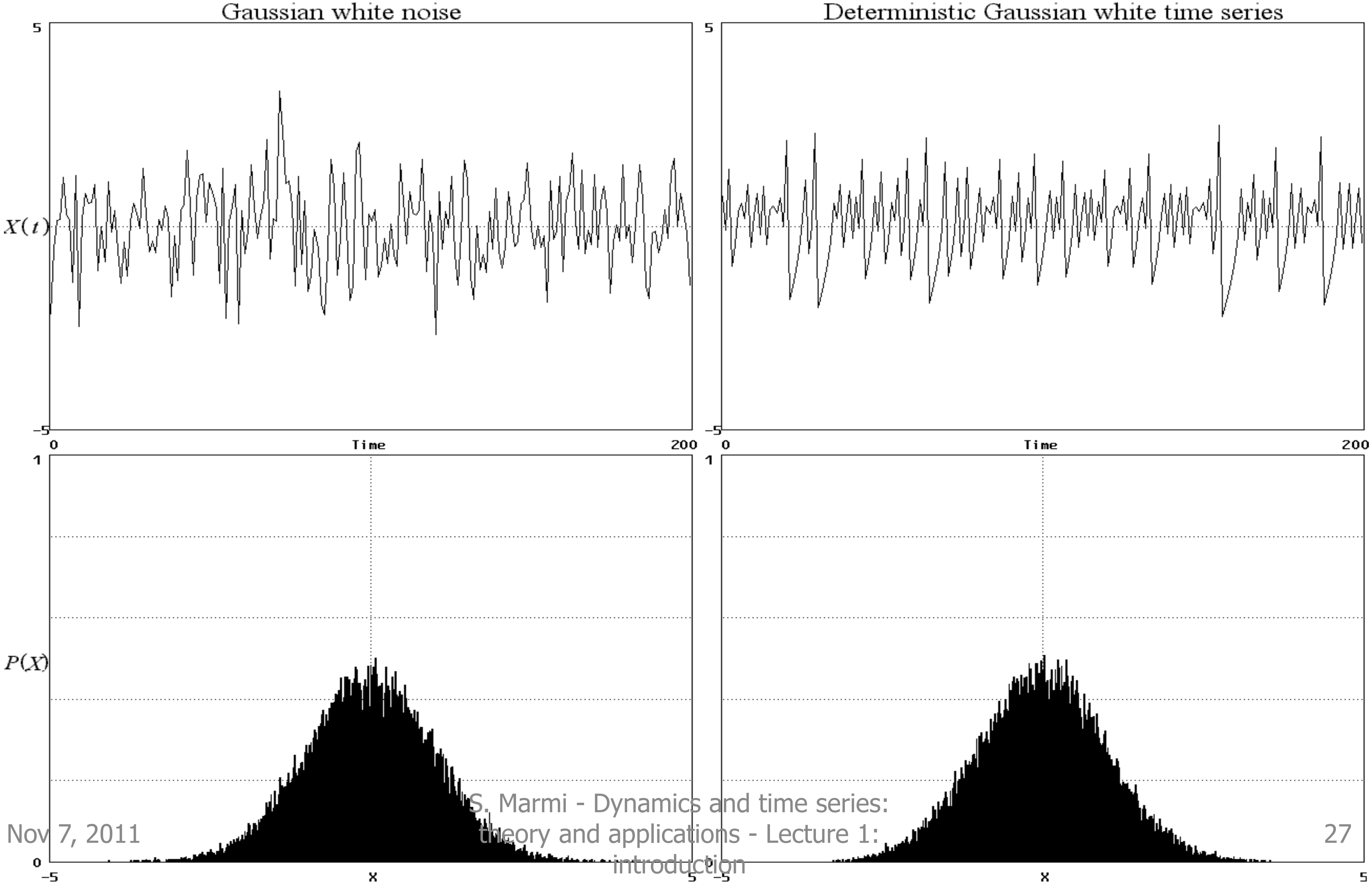
In quasiperiodic dynamics this does not happen: for the rotations on the circle one has $\theta(N)-\theta'(N)=\theta(0)-\theta'(0)$ and long term prediction is possible

Stochastic or chaotic?

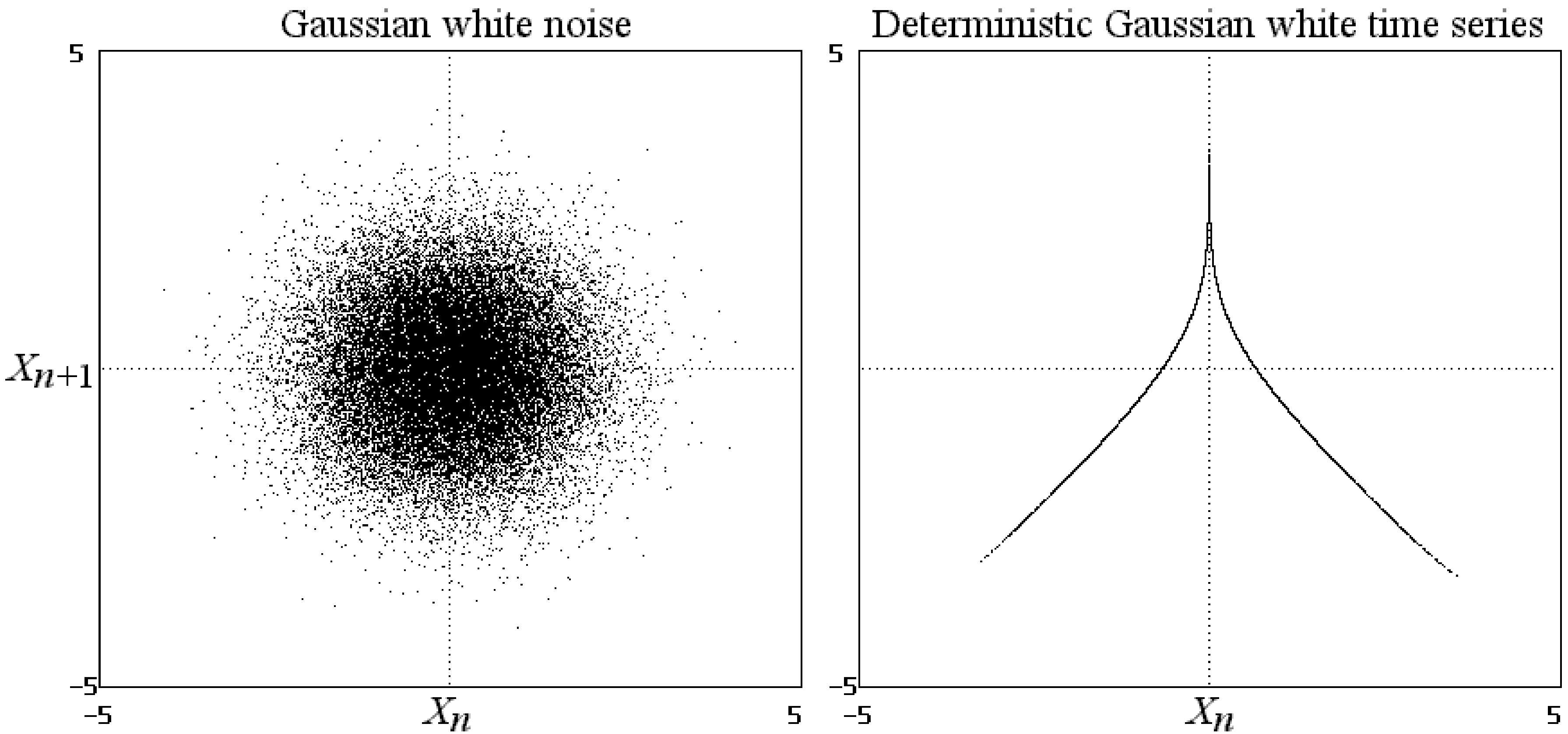
An important goal of time-series analysis is to determine, given a times series (e.g. HRV) if the underlying dynamics (the heart) is:

- Intrinsically random
- Generated by a deterministic nonlinear chaotic system which generates a random output
- A mix of the two (stochastic perturbations of deterministic dynamics)

Deterministic or random? Appearance can be misleading…



Time delay map



Logit and logistic

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

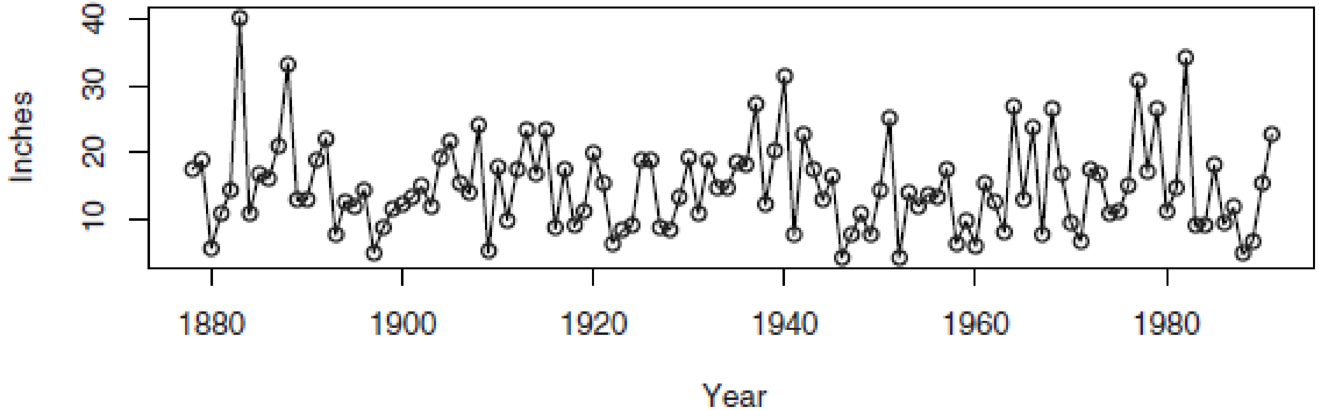
The transformation $h:[0,1] \rightarrow \mathbf{R}$, $h(x)=\ln x-\ln(1-x)$ conjugates L with a new map G defined on **R:**

$$h \circ L = G \circ h$$

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

Clearly G and L have the same dynamics (the differ only by a coordinates change)

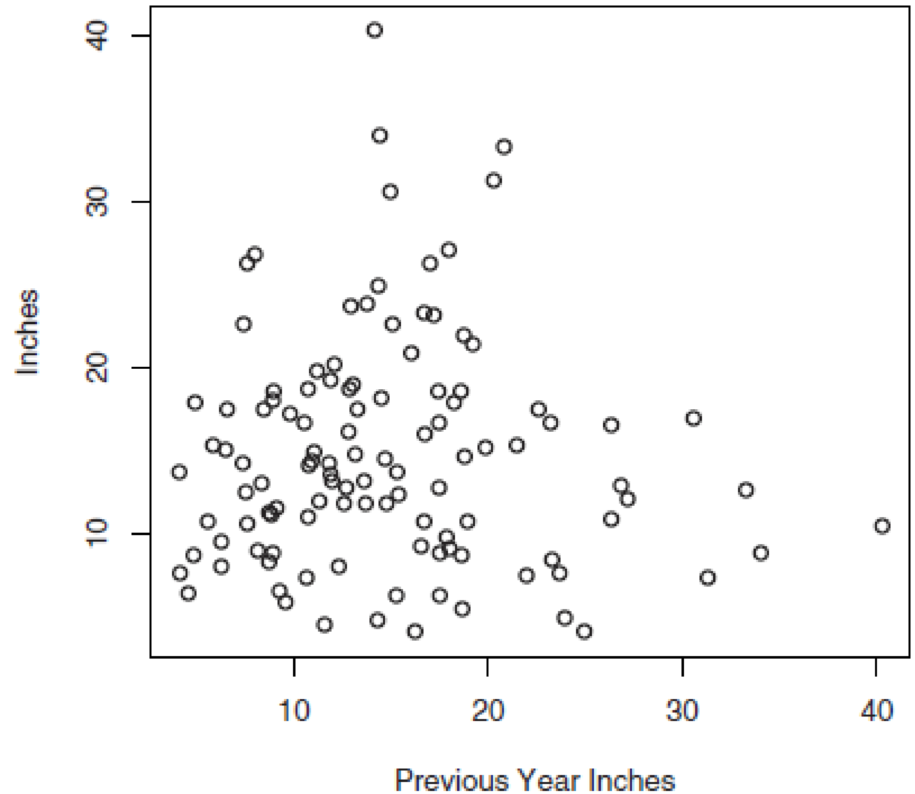
Exhibit 1.1 **Time Series Plot of Los Angeles Annual Rainfall**




```
> library(TSA)
> win.graph(width=4.875, height=2.5,pointsize=8)
> data(larain); plot(larain,ylab='Inches',xlab='Year',type='o')
```

From Cryer and Kung-Sik Chan: *Time Series Analysis with Applications in R*

Exhibit 1.2 Scatterplot of LA Rainfall versus Last Year's LA Rainfall




```
> win.graph(width=3,height=3,pointsize=8)
> plot(y=larain,x=zlag(larain),ylab='Inches',
   xlab='Previous Year Inches')
```

From Cryer and Kung-Sik Chan: *Time Series Analysis with Applications in R*

Embedding method

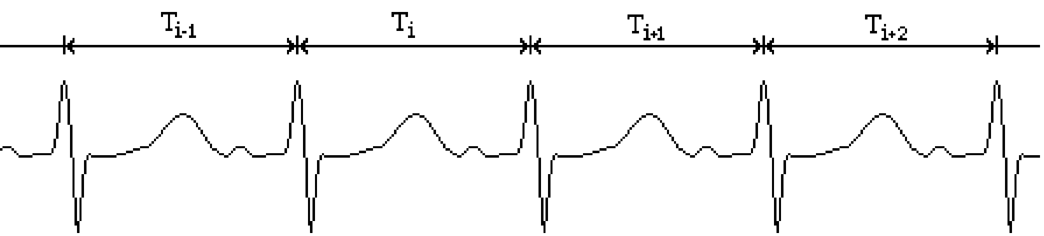
- Plot $x(t)$ vs. $x(t-\tau)$, $x(t-2\tau)$, $x(t-3\tau)$, …
- $x(t)$ can be any observable
- The embedding dimension is the # of delays
- The choice of τ and of the dimension are critical
- For a typical deterministic system, the orbit will be diffeomorphic to the attractor of the system (Takens theorem)

Time series analysis of physiological signals

Physiological signals are characterized by extreme variability both in healthy and pathological conditions. Complexity, erratic behaviour, chaoticity are typical terms used in the description of many physiological time series.

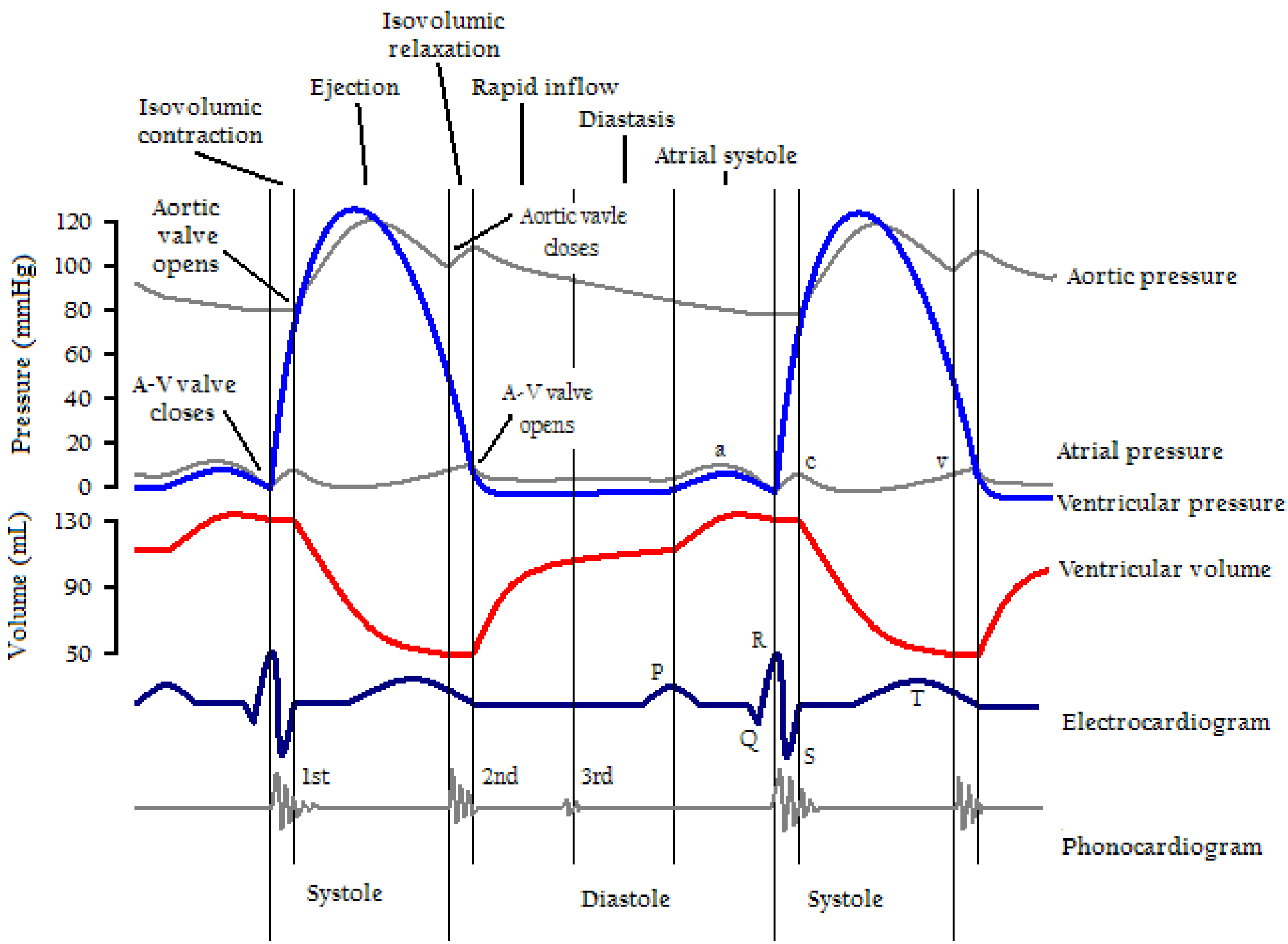
Quantifying these properties and turning the variability analysis from qualitative to quantitative are important goals of the analysis of time-series and could have relevant clinical impact.

From ECG to heart rate variability time series

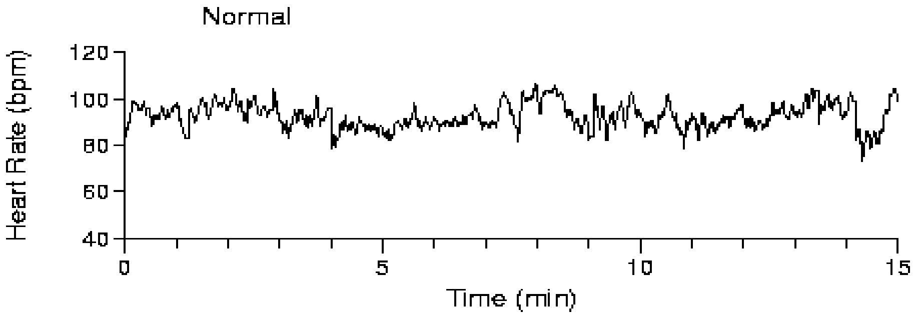


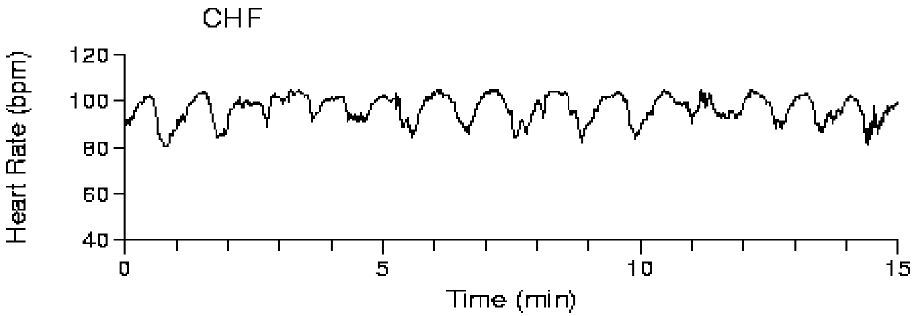
- Example of ECG signal
- The time interval between two consecutive R-wave peaks (R-R interval) varies in time
- The time series given by the sequence of the durations of the R-R intervals is called heart rate variability (HRV)

The heart cycle and ECG

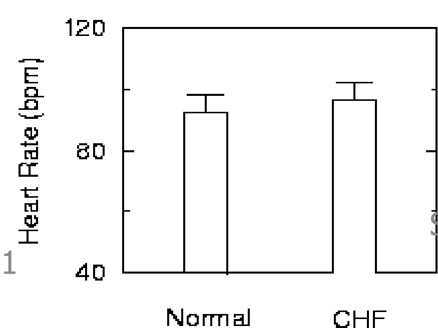


Healthy? Statistical vs. dynamical tools for diagnosis

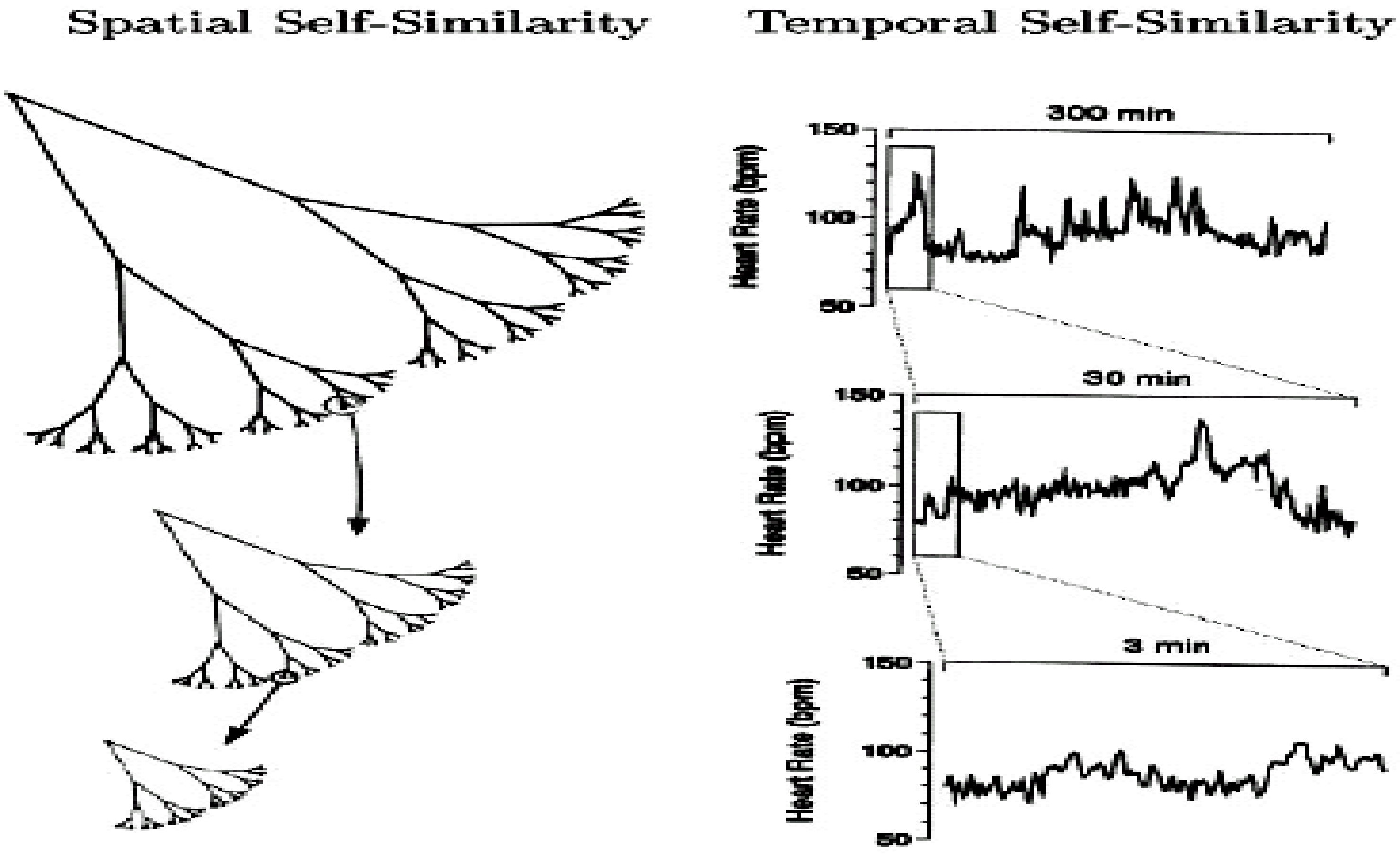




- The HRV plots of an healthy patient show a very different dynamics from those of a sick patient but the traditional statistical measures (mean and variance) are almost the same.
- www.physionet.org

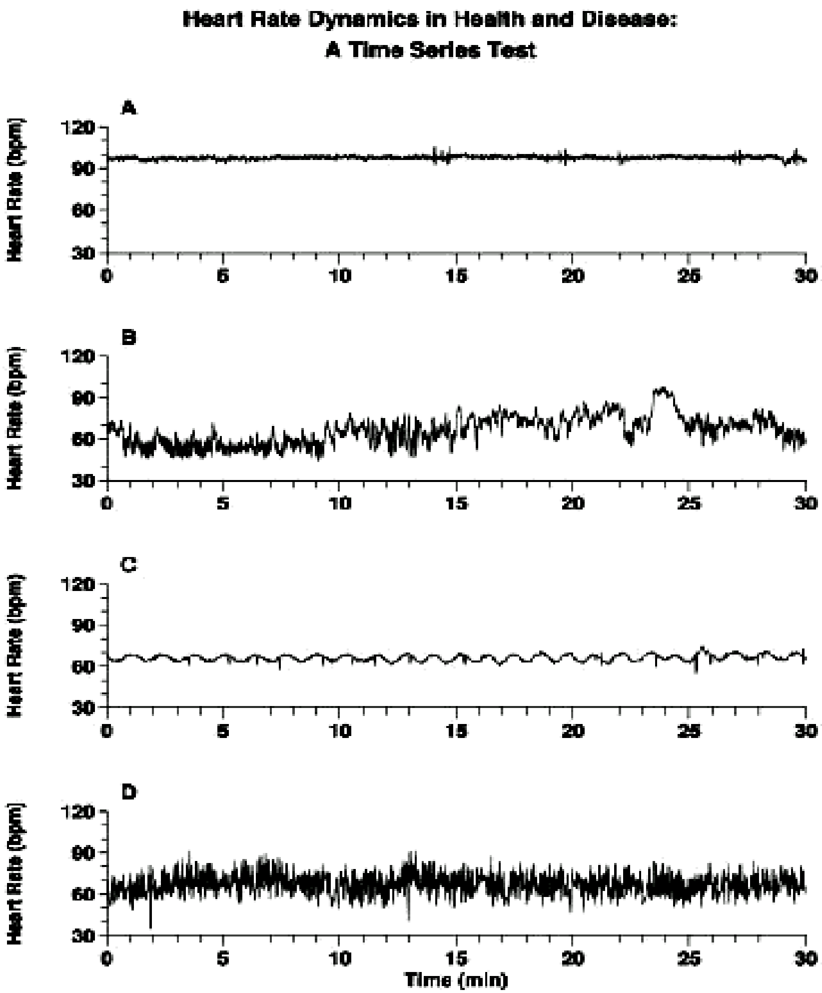


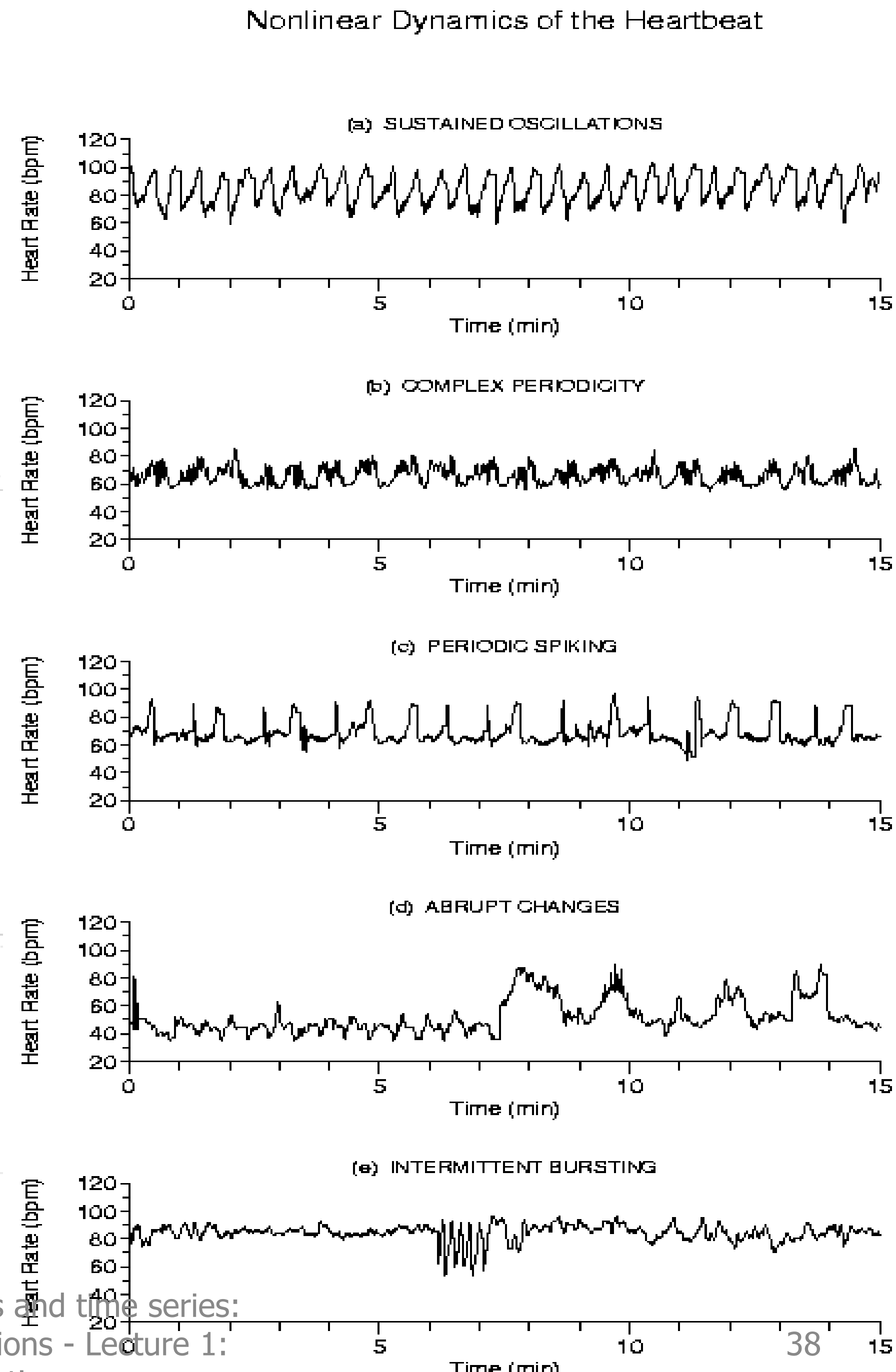
Time series and self-similarity



www.physionet.org

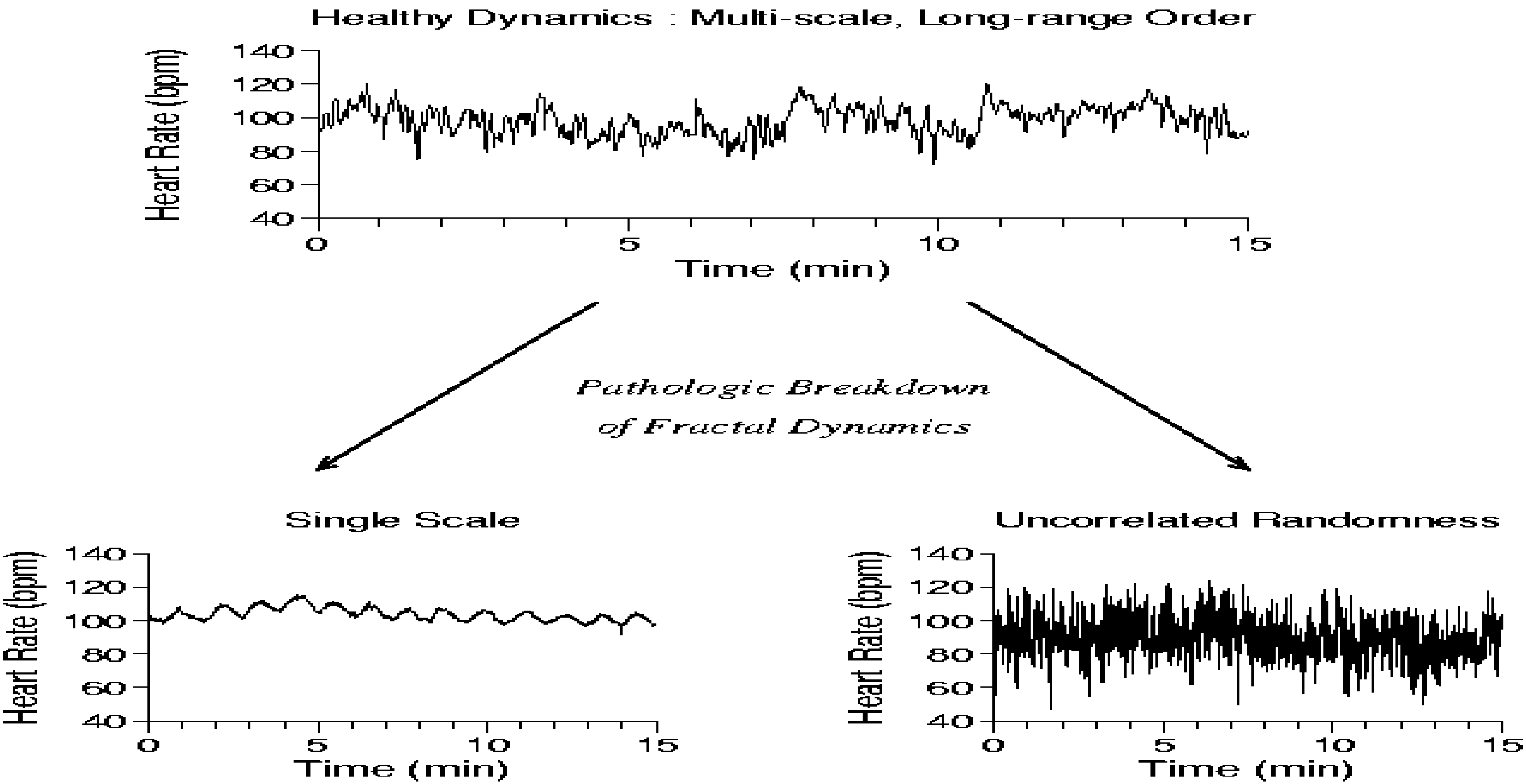
Healthy ?





www.physionet.org

Healthy or not?



(Adapted from Goldberger AL. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *Lancet* 1996;347:1312-1314.)

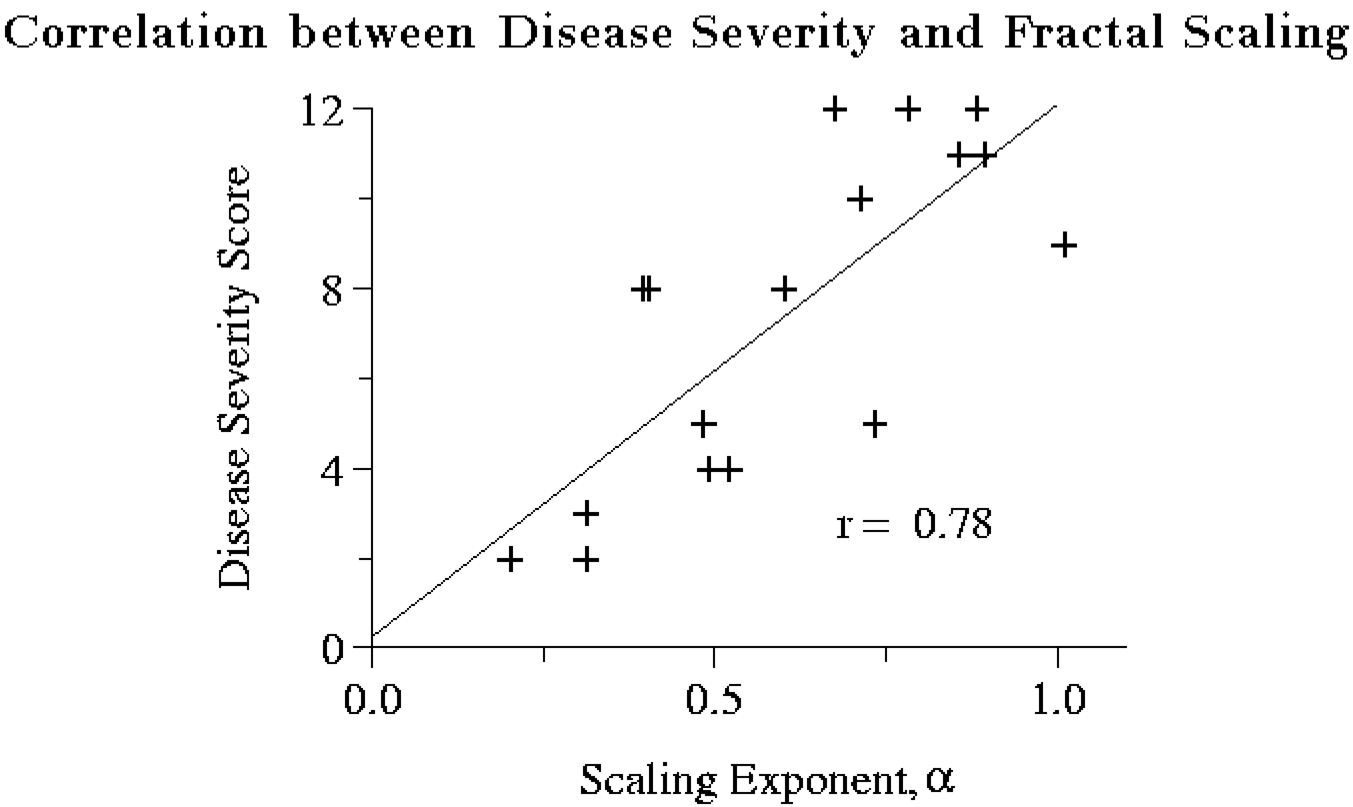
Healthy or not?

FRACTAL DYNAMICS OF HEART RATE AND GAIT

	FRACTAL HEART DYNAMICS	**FRACTAL GAIT DYNAMICS**
Features	Extends over thousands of beats	Extends over thousands of steps
In Health	Persists during different activities (asleep or awake)	Persists regardless of gait speed (slow, normal or fast)
Potential	Altered with advanced age	Altered with advanced age
Diagnostic & Prognostic	Altered with cardiovascular disease (e.g. Heart Failure)	Altered with nervous system disease (e.g. Parkinson's D.)
Utility	Helps predict survival	May predict falls among elderly

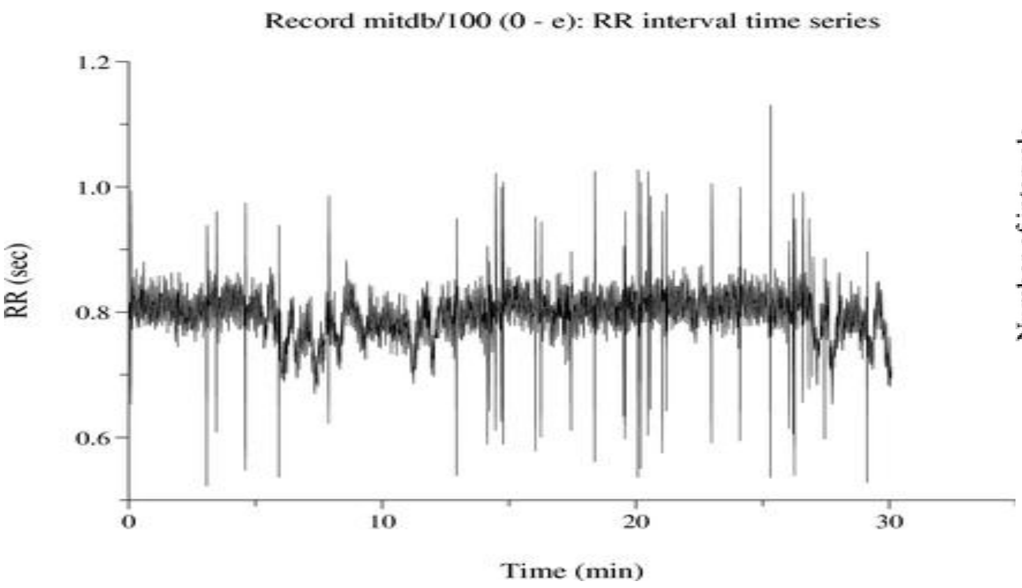
Source: *htpp://www.physionet.org*

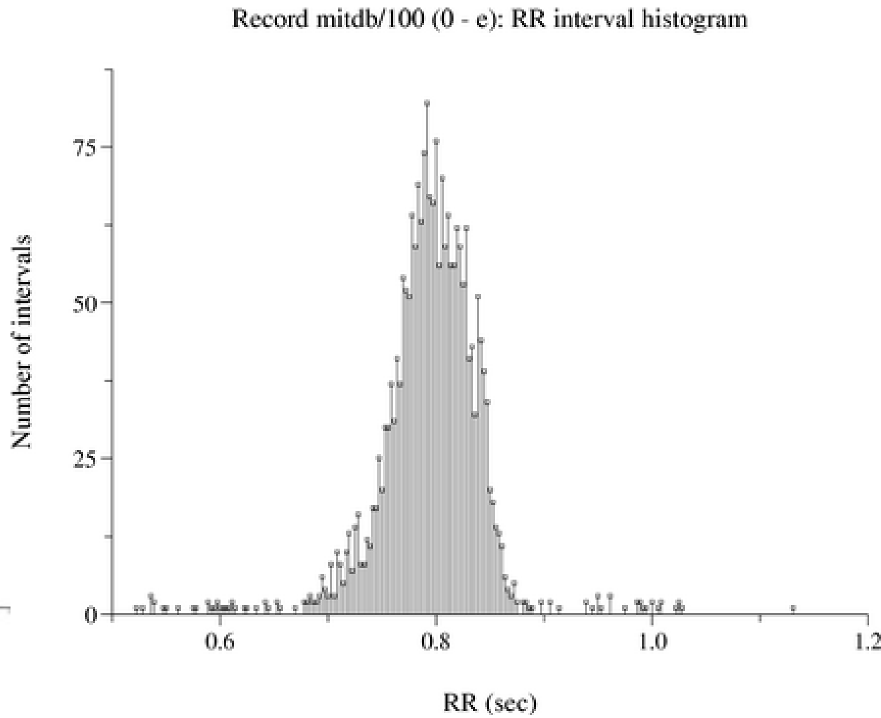
Correlation between disease severity and fractal scaling exponent



www.physionet.org

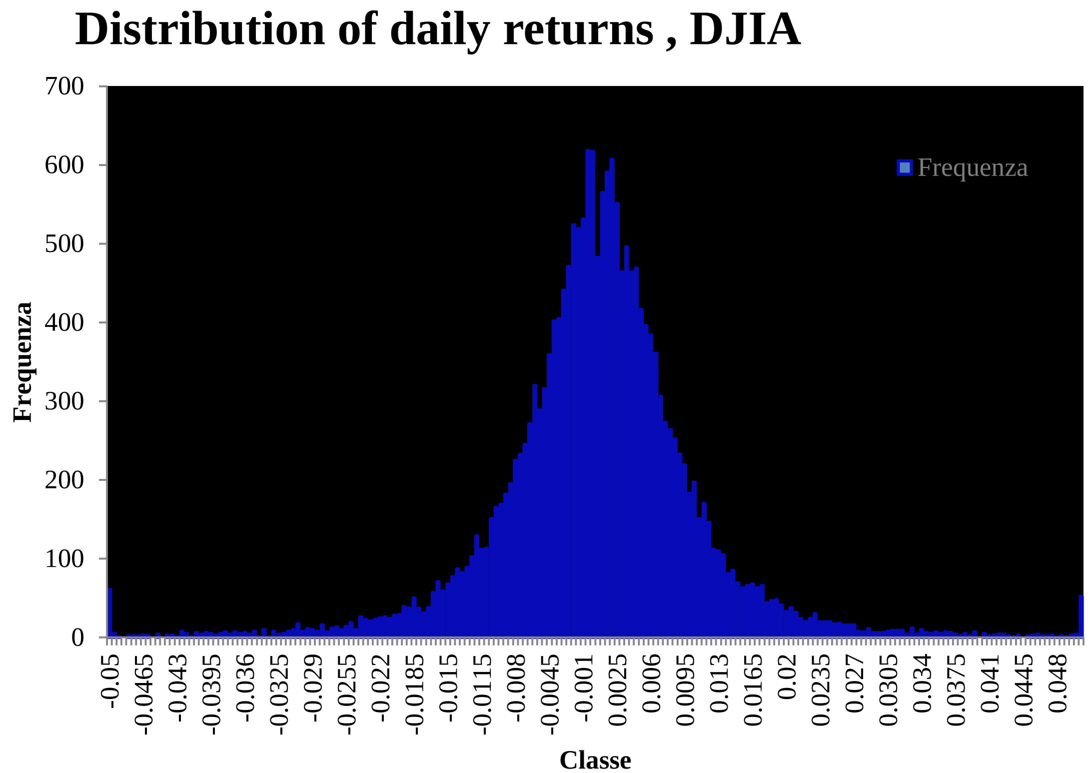
Distribution of R-R intervals

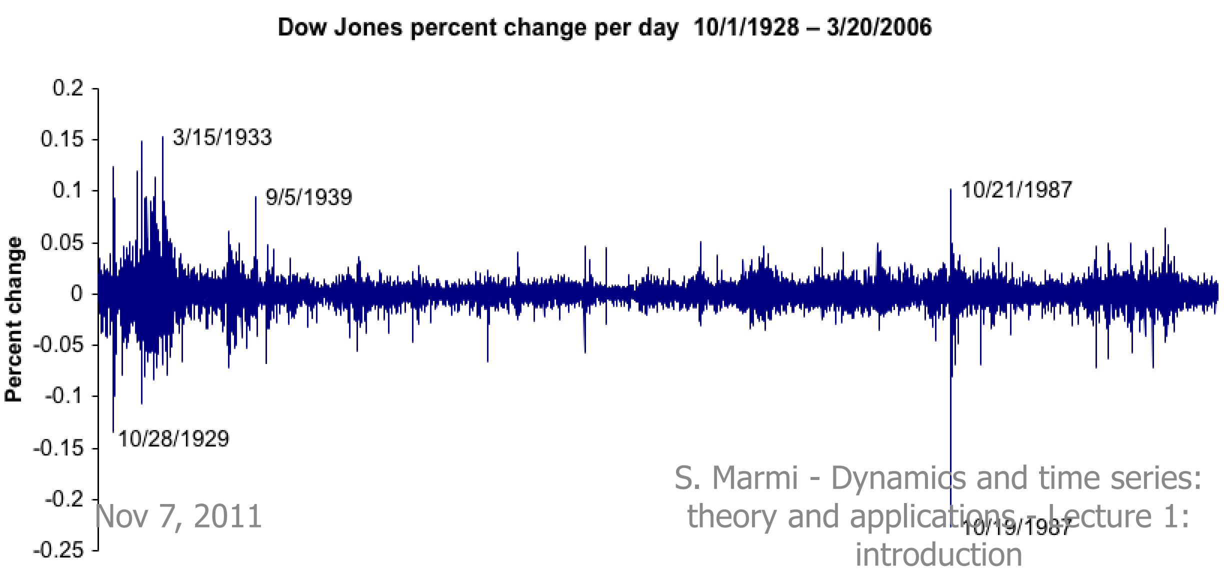




www.physionet.org

Distribution of daily returns, Dow Jones 1928-2007



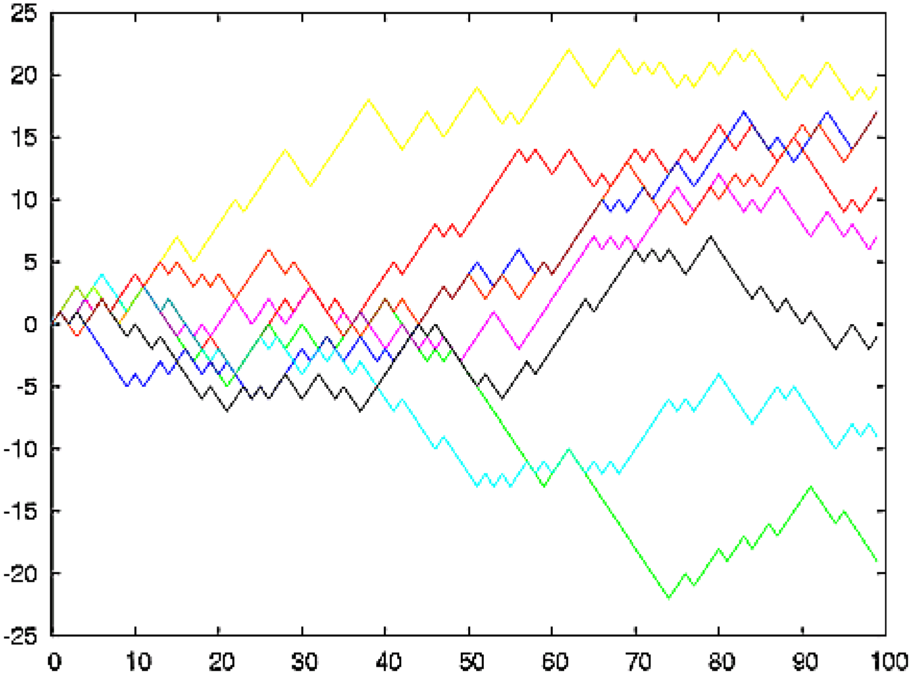


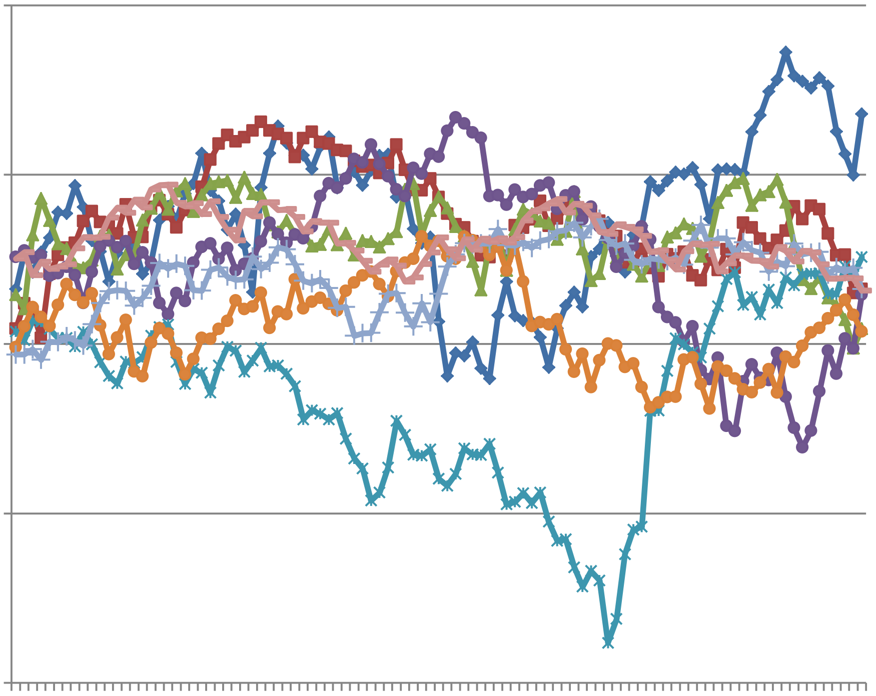
Geometric daily return at time t = (Price of index at time t / Price of index at time t-1)-1

Here the index is the Dow Jones Industrial Average, t is integer and counts only open market days

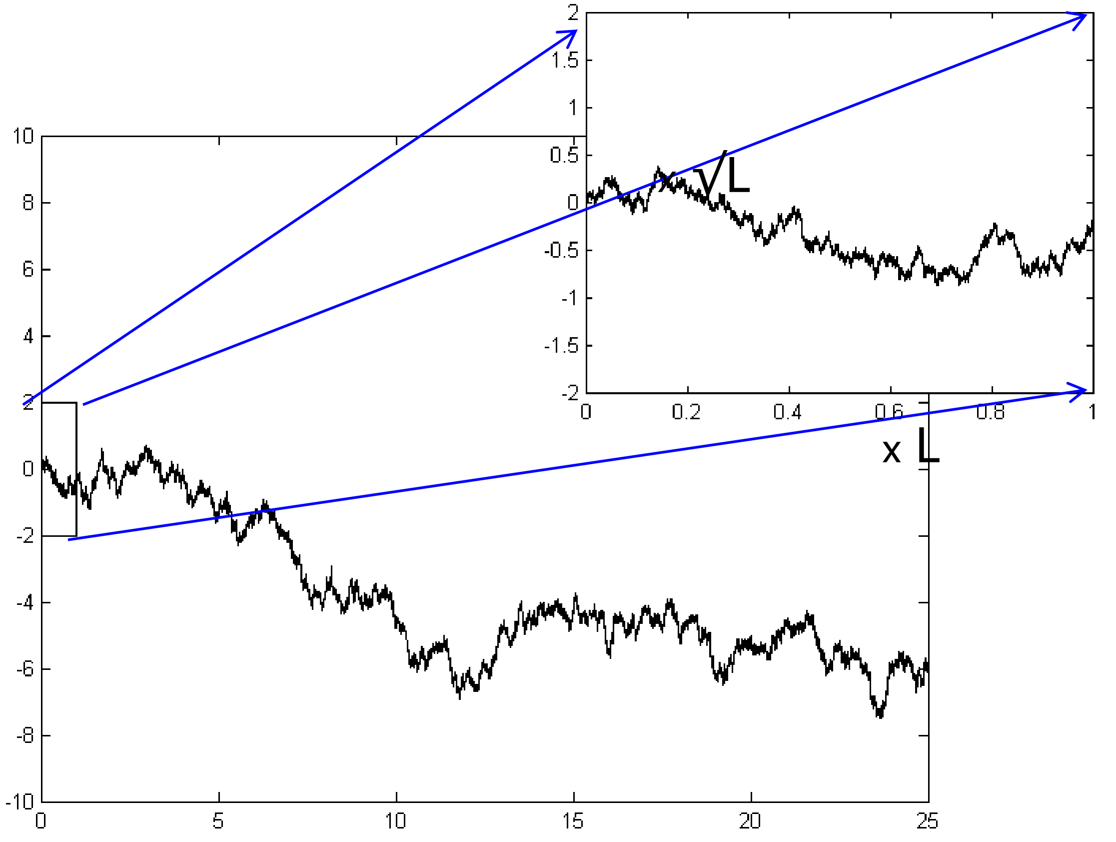
The value of the index is at the close

Random walks vs. Dow Jones





Selfsimilarity



The normal distribution

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R},$$

$$\varphi_{0,t}(x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right). \qquad \frac{\partial}{\partial t}\varphi_{0,t}(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\varphi_{0,t}(x).$$

Do daily returns follow a normal distribution?

Class	*Observed Frequency*	Theoretical Frequency
x< -0.05	67	0.093902
-0.05<x<-0.045	19	0.567355
-0.045<x<-0.04	41	3.207188
--0.04<x<0.035	51	14.9652
-0.035<x<-0.03	78	57.64526
-0.03<x<-0.025	117	183.3153
-0.025<x<-0.02	247	481.2993
-0.02<x<-0.015	484	1043.367
-0.015<x<-0.01	1111	1867.6
-0.01<x<-05	2433	2760.391
-0.05<x<0	4879	3369.05
0<x<05	5119	3395.468
05<x<0.01	2881	2825.84
01<x<0.015	1219	1941.987
0.015<x<0.02	539	1102.011
0.02<x<0.025	241	516.3589
0.025<x<0.03	105	199.7674
0.03<x<0.035	77	63.8089
0.035<x<0.04	43	16.82651
0.04<x<0.045	27	3.662964
0.045<x<0.05	20	0.658208
x> 0.05	50	0.110887

Mean	00204
Median	00411
Moda	0
Standard deviation	0.011355
Varianza campionaria	00129
Kurtosis	26.84192
Asymmetry	-0.67021
Intervallo	0.399044
Minimum	-0.25632
Maximum	0.142729
Sum	4.058169
Number of observations	19848

Theoretical and observed frequency of outliers in the history of 15 stockmarkets

Exhibit 4: Outliers – Expected and Observed

This exhibit shows, for the indexes and sample periods in Exhibit 2, the expected (Exp) and observed (Obs) number of daily returns three standard deviations (SD) below and above the arithmetic mean return (AM); the ratio between the number of these observed and expected returns; and the total number of expected (TE) and observed (TO) returns more than three SDs away from the mean. 'Exp' figures are rounded to the nearest integer.

	Lower Tail				Upper Tail						
Market	AM–3-SD	Exp	Obs	Ratio	AM+3-SD	Exp	Obs	Ratio	TE	TO	Ratio
Australia	–2.46%	17	73	4.4	2.52%	17	53	3.2	33	126	3.8
Canada	–2.48%	11	73	6.9	2.55%	11	43	4.1	21	116	5.5
France	–3.11%	13	79	6.2	3.19%	13	61	4.8	25	140	5.5
Germany	–3.51%	16	85	5.3	3.57%	16	76	4.8	32	161	5.1
Hong Kong	–5.53%	12	77	6.2	5.67%	12	80	6.5	25	157	6.4
Italy	–3.82%	12	71	6.0	3.91%	12	48	4.0	24	119	5.0
Japan	–3.12%	19	132	6.8	3.19%	19	112	5.8	39	244	6.3
New Zealand	–2.51%	12	61	4.9	2.56%	12	57	4.6	25	118	4.7
Singapore	–3.12%	14	90	6.4	3.18%	14	86	6.1	28	176	6.3
Spain	–3.22%	11	52	4.8	3.31%	11	61	5.6	22	113	5.2
Switzerland	–2.74%	13	101	7.9	2.79%	13	62	4.8	26	163	6.4
Taiwan	–4.55%	15	103	6.8	4.65%	15	81	5.3	30	184	6.0
Thailand	–4.40%	10	62	6.0	4.48%	10	81	7.8	21	143	6.9
UK	–3.00%	13	69	5.3	3.07%	13	60	4.6	26	129	5.0
USA	–3.35%	28	180	6.4	3.40%	28	173	6.1	56	353	6.3
Average	**–3.39%**	**14**	**87**	**6.0**	**3.47%**	**14**	**76**	**5.2**	**29**	**163**	**5.6**

Estrada, Javier: Black Swans and Market Timing: How Not to Generate Alpha. Available at SSRN: http://ssrn.com/abstract=1032962