An Information-based tool for inferring the nature of deterministic sources in real data

R. Riera – Dep. Física – PUC-Rio

Complex Systems - Foundations and Applications
CBPF October 29 to November 1 – 2013

Goals

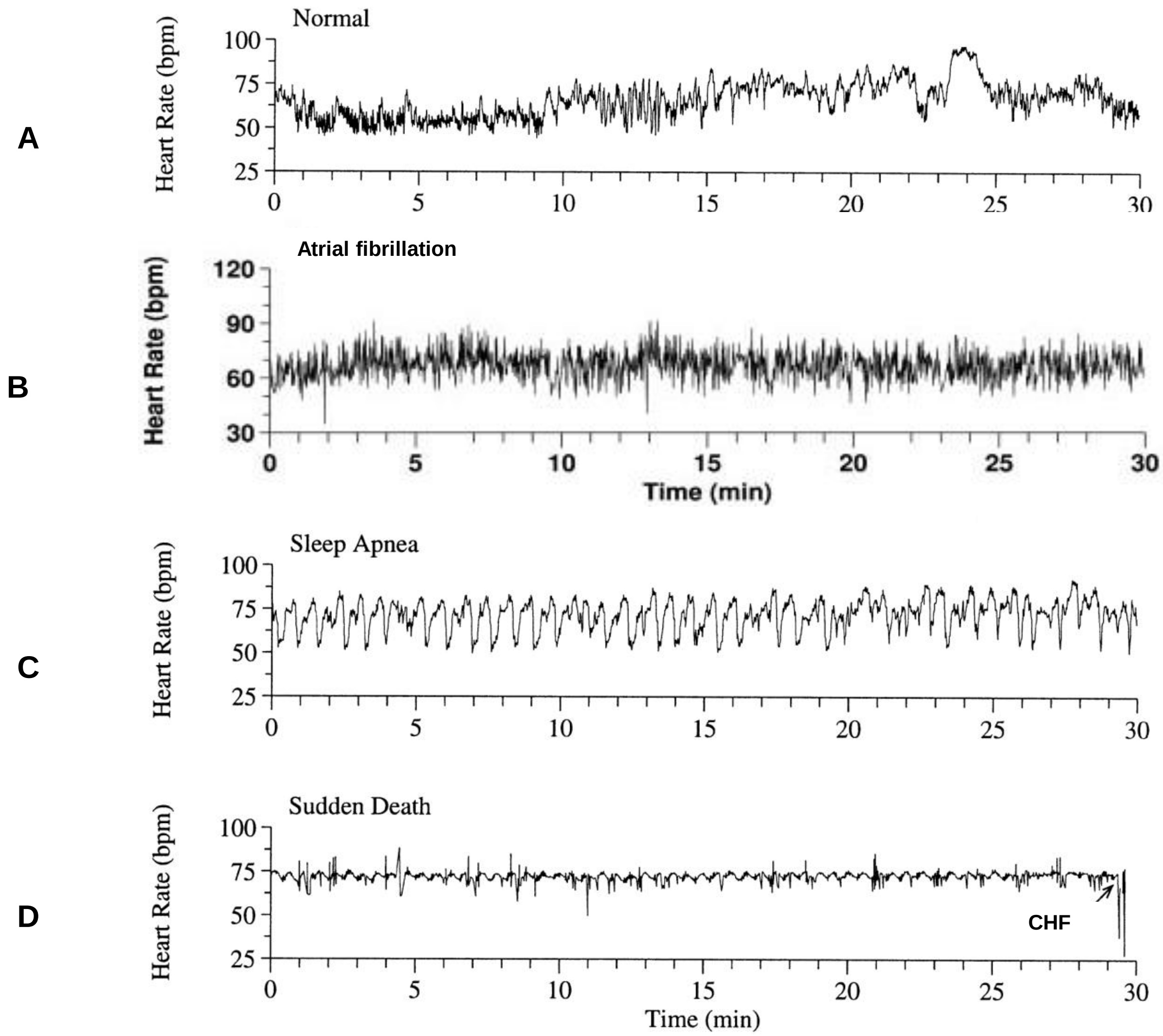
- finding signatures of the forces controlling complex systems from historical time-series

- discriminate empirical signals

Method

- information measures that encompasses the degree of irregularity as well as the correlation structures of the generating process

Representative physiological fluctuations.



Heart rate recordings of 30 minute-time series: (**A**) healthy, (**B**) with atrial fibrillation (AF) - (**C**) with obstructive sleep apnea, and (**D)** with congestive heart failure (CHF).

Goldberger A L et al, Circulation ,**101**:e215-e220 (2000) & PNA **99**:2466 (2002).

Empirical analysis

- The heartbeat time series in disease is associated with the emergence of uncorrelated randomness (B) or regularity (C)

- The healthy heartbeat time series exhibits an intermediate behavior

- A healthy biological system needs to exhibit processes which run on several time scales to be able to respond to unpredictable stimuli and stresses.

- Multi-scaling and complexity degrade with aging and disease reduces the adaptability of the organism.

Entropy measures – I

Consider the PDF *P(x)* of a variable X that assumes values x belonging to a finite alphabet $\mathbb{A}$ with N classes: $P \equiv \{p_i\ ;\ i=1, \ldots, N\}$

Shannon Entropy (quantifies randomness at a global level):

$$S[P] \equiv -\sum_{x \in \mathbb{A}} P(x) \ln P(x) \qquad H[P] = S[P] / S_{\max} \; with \; S_{\max} = \ln N$$

Consider a time-series of real values $\{s_{t}: t=1,\ldots.T\}$ and the n-dimensional sequence vectors $u^{(n)}(i) \equiv \{s_i\ , s_{i+1}\ ,\ldots, s_{i+n-1}\ \}$

- n=1 : amplitude statistics or histograms
- n > 1 : captures local temporal structures

→ Symbolic dynamics - a symbol Π is assigned to a trajectory's portion.
→ define a mapping onto a set of "orthogonal" states

Entropy measures - II

Consider all n! possible sorts of amplitude ordering of n real values and identify each ordering as an ordinal pattern Π

As a result, any sequence vector $u^{(n)}(i)$ is mapped onto :

$$\Pi_i, \ i=0,1,\ldots(n!-1)$$

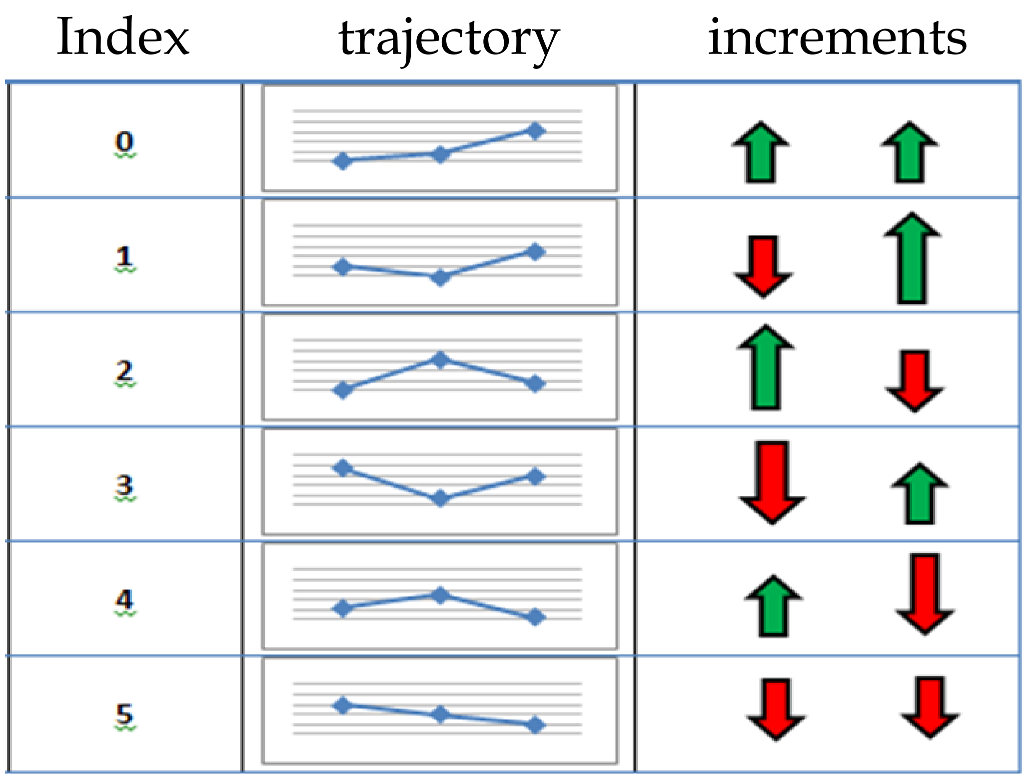
Consider the relative frequency P(Π) of the ordinal pattern Π.

Permutation Entropy (PE) :

$$S[P] \equiv -\sum_{\Pi \in A} P(\Pi) \ln P(\Pi)$$

$$H[P] = S[P] / S_{\max} \ with \ S_{\max} = \ln n!$$

Ordinal Patterns



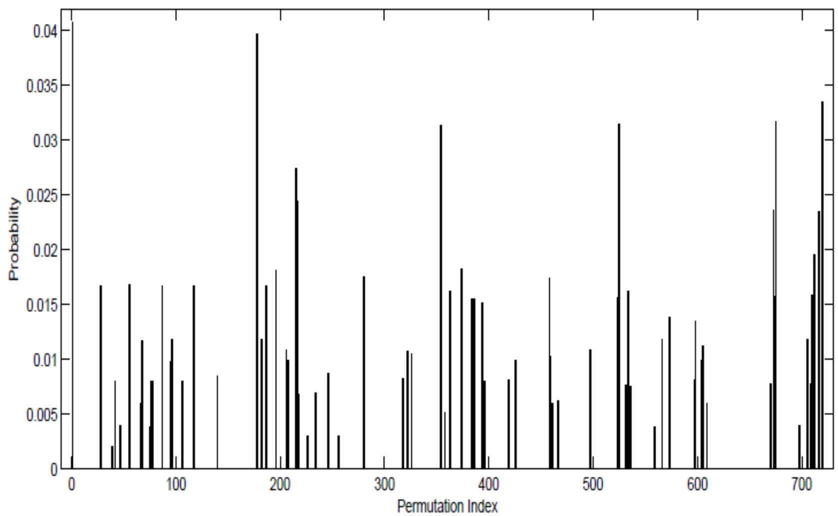
Ordinal Patterns Π_i, i=0,1, …5 for n=3 length scale

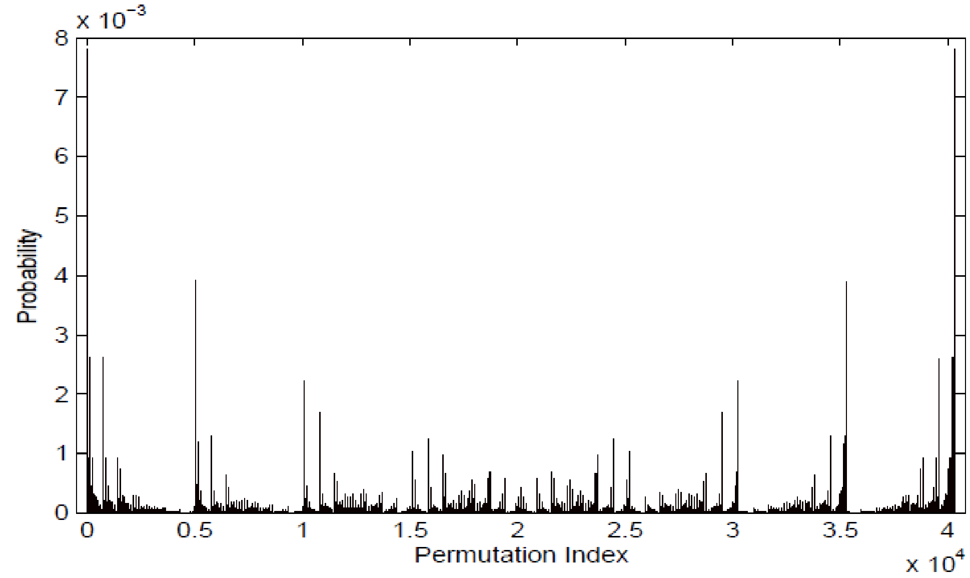
Ordinal Patterns

Symbol Index	1st Increment	2nd Increment
0	$\Delta x_1 > 0$	$\Delta x_2 > 0$
1	$\Delta x_1 < 0$	$\Delta x_2 > 0, \lvert\Delta x_2\rvert > \lvert\Delta x_1\rvert$
2	$\Delta x_1 > 0$	$\Delta x_2 < 0, \lvert\Delta x_2\rvert < \lvert\Delta x_1\rvert$
3	$\Delta x_1 > 0$	$\Delta x_2 < 0, \lvert\Delta x_2\rvert > \lvert\Delta x_1\rvert$
4	$\Delta x_1 < 0$	$\Delta x_2 > 0, \lvert\Delta x_2\rvert < \lvert\Delta x_1\rvert$
5	$\Delta x_1 < 0$	$\Delta x_2 < 0$

Ordinal Patterns Π_i, i=0,1, …5 for n=3 length scale

Distinguishing deterministic and stochastic dynamics





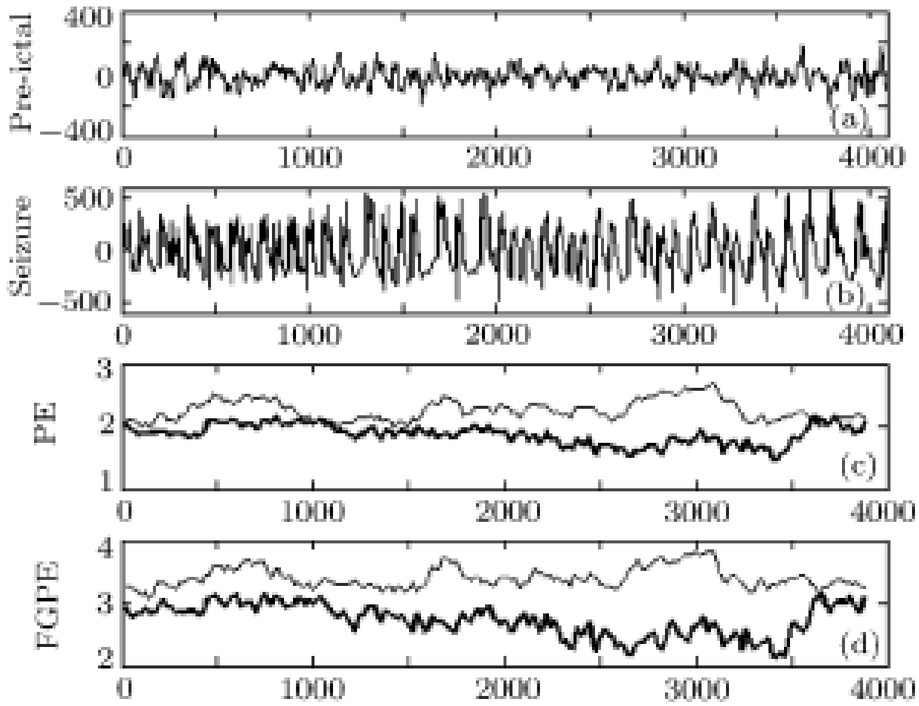
Permutation histogram with = 6
for the fully chaotic logistic map

Permutation histogram with = 8
for Brownian Motion

- Deterministic processes display absent permutation patterns

- Stochastic processes display higher permutation entropy values than the deterministic ones.

Detection of dynamical changes in epileptic data using PE and FGPE



- Accurate detection of transitions from normal to pathological states may improve diagnosis
- PE turns into a promising complexity quantifier

Electroencephalography (EEG) signal in normal state (a) and seizure state (b); PE(c) and FGPE (d) for order n=4 in moving windows of length 200: normal state (thin line) and seizure state (bold line).

Xiao-Feng L and Yue W, Chinese Physics 18, 1674 (2009)

Statistical Complexity

Uncorrelated random signals are highly unpredictable, but, at a global level, they admit very simple statistical description and therefore , are not really complex

Jensen-Shannon Divergence: $J[P,U] = S\left[\frac{P+U}{2}\right] - \frac{S[P]}{2} - \frac{S[U]}{2}$

where U is the uniform distribution, taken as a reference

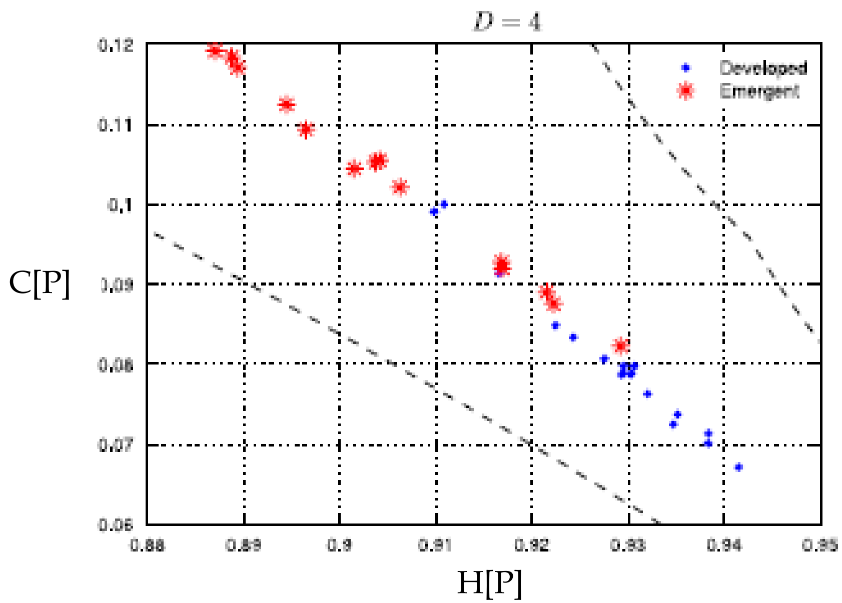
$$Q[P] = J[P,U] / J_{\max} \ \textit{with} \ J_{\max} = \ln 2$$

Statistical Complexity: $C[P] = H[P] * Q[P]$

- The C[P] measure ascribes null value of complexity for both regular and completely random series. Therefore, systems exhibiting a behavior in-between these two extremes, although encompassing intermediate content of information, are more complex.

PE - PSC Plane

H[P] and C[P] evaluated for P[Π] improves the performance of the information quantifiers



Localization of daily world stock indices time series for permutation order *n*=*4*. Zunino L et al, Physica A 389, 1891 (2010)

- The complexity-entropy plane discriminates emergent and develop markets
- In fact, financial processes become unpredictable and lose its memory structure due to enlarged economic activities in developed markets
- Market inefficiency enhances complexity

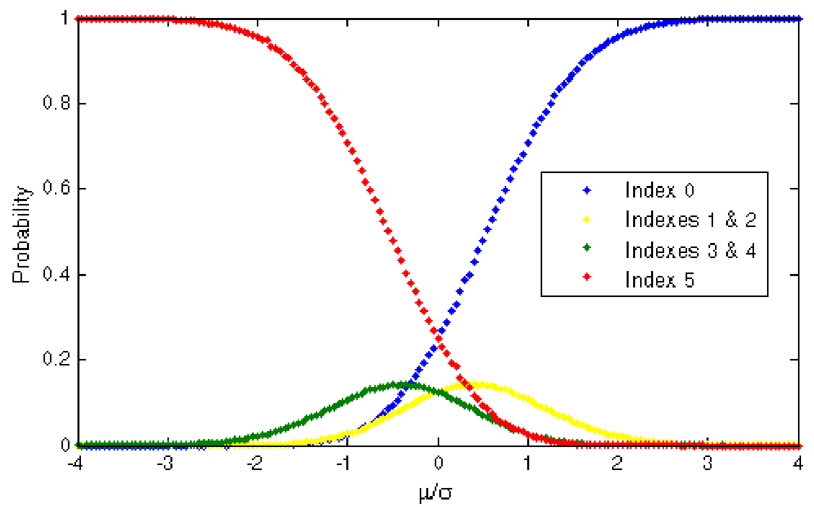
Classes of representative stochastic processes

Due to the invariance properties of the ordinal patterns under non linear monotonous transformations, the results go beyond the specific analyzed models.

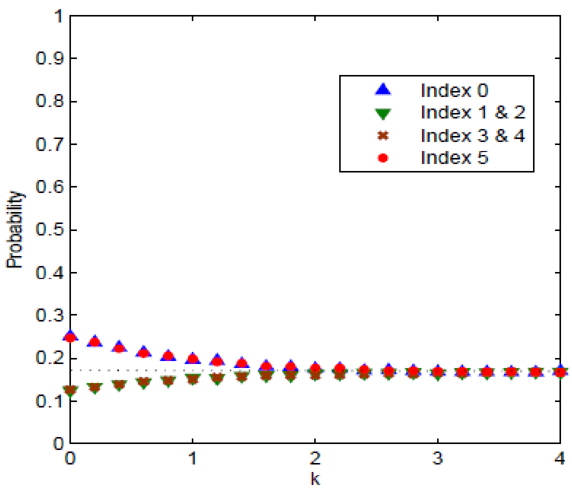
Consider stochastic processes described by: $dx = \mu(x)dt + \sigma(x)dW$

- class of drifting processes: $\mu = const$ and $\sigma = const$
- class of mean-reverting processes: $\mu(x) = -kx$ and $\sigma = const$

P(Π_i, i=0,...5) of order n=3 in the full range of relevant parameters.



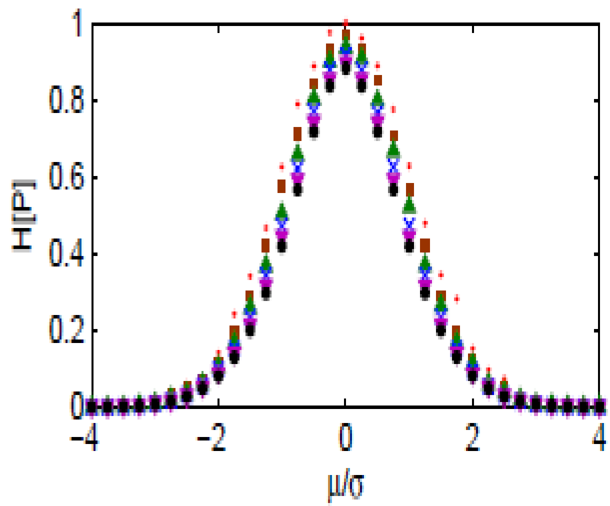
(drifting)



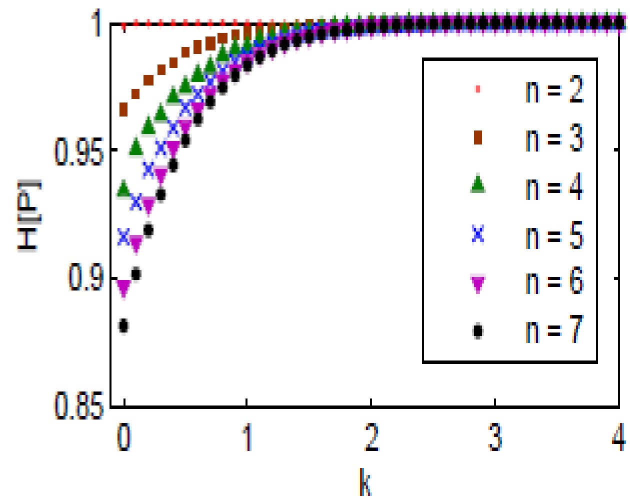
(reverting)

- *The set of symbolic states Π are identified as generalized states , carrying out information about the major nature of the deterministic sources.*

Distinguishing the nature (drift or reverting) of the deterministic forces



(drifting)



(reverting)

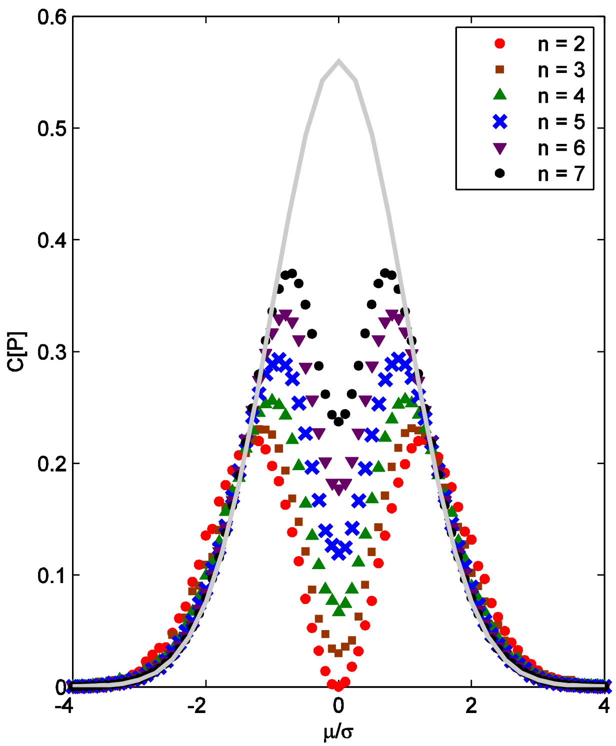
- PE exhibits new relevant parameters as compared to standard sample entropy

- PE ranges for both processes are complementary: for each order n≥3, the class of mean reverting processes furnishes PE values larger than the class of drifting processes; the threshold value corresponds to the Brownian Motion

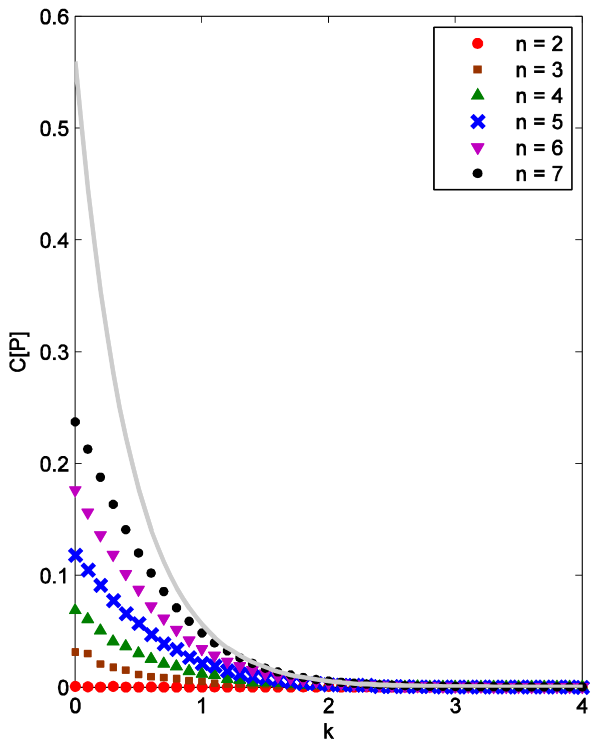
Multi-scale PSC

- New profiles are found in all scales, suggesting that new information is revealed as n increases

PSC curve for several orders *n* and the limiting shape for arbitrary large scales.

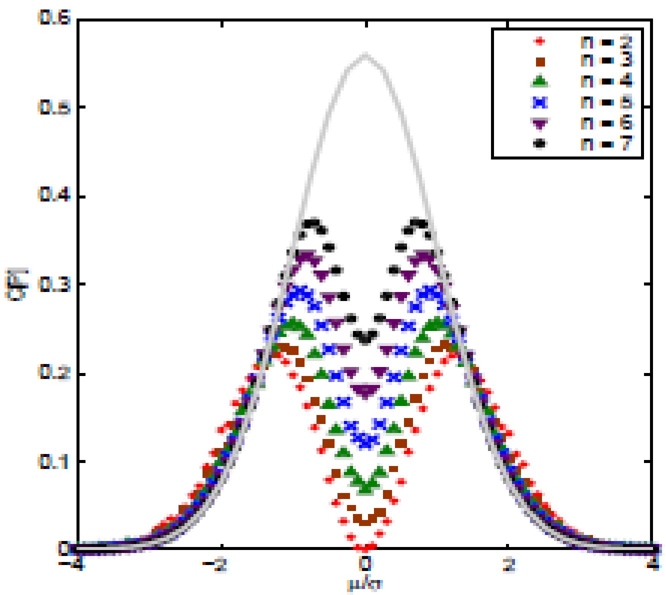


(drifting)



(reverting)

Multi-scale PSC curve for drift processes



PSC curve for several orders *n* and the limiting shape for arbitrary large scales.

➢ Brownian motion is a scale invariant process at a global level, and therefore, are rather complex.

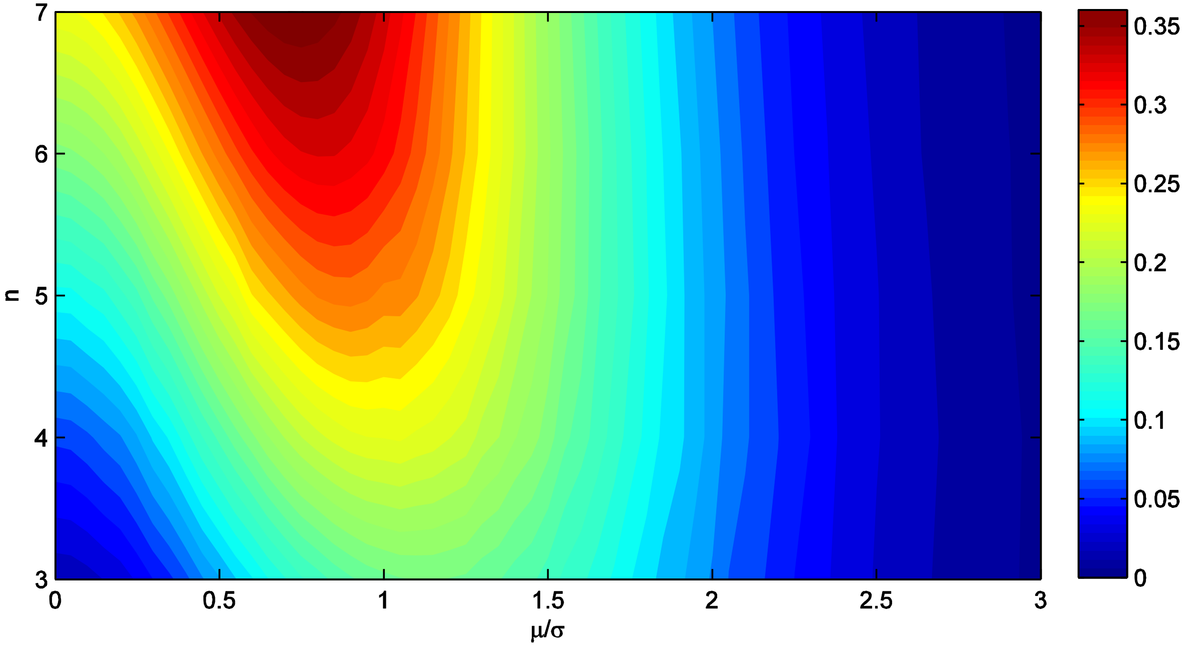
➢ At small-scale factors, the Brown noise has PSC values smaller than the noise with drift, but at larger scales, it produces larger complexity.

→ global trends are associated with complexity loss, indicating forecast opportunities

Multi-scale PSC for drift processes

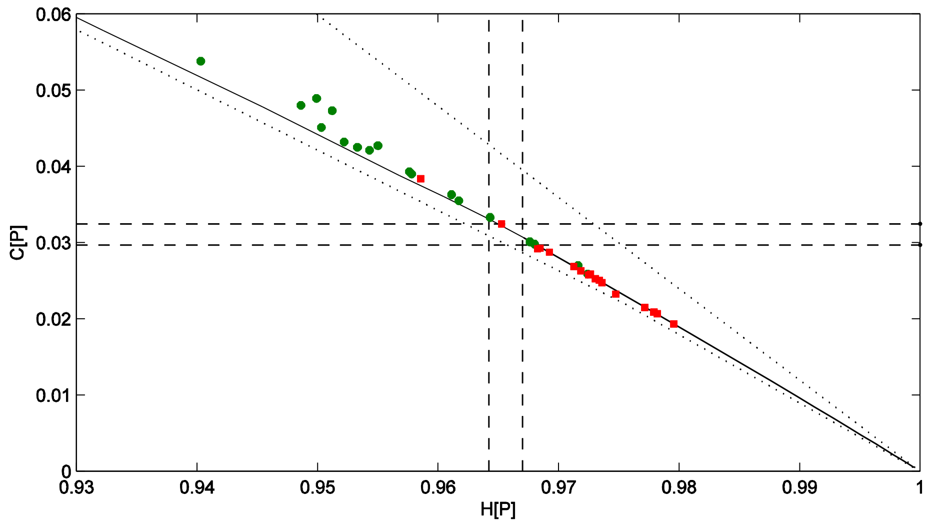
➢The complexity measure PSC depends crucially on the permutation order under consideration

➢systems show different signatures of the PSC dependency on *n*, giving rise to crossovers



Color map of PSC values for drift processes
Ribeiro A and Riera R, Physica A **392**, 5053 (2013)

Application: daily log-price time series



Representation of Brazilian (green circles) and American (red squares) stocks in the PSC-PE plane for permutation order n=3.
The theoretical Drift and MR lines are also shown (solid lines)

- Most American stocks fall in the MR domain

- Brazilian stocks are more subjected to stable trends, but displays residual complexity compared to its drift-class counterpart.

- series in the drift-class domain have $0<\mu/\sigma <0.22$ which do not exhibit crossovers amongst them up to n=7

→ complexity is a sign of forecast opportunities in short time horizons.

Concluding remarks

- We consider the PE and PSC quantifiers to access the temporal dynamics of empirical records.

- As control data sets we consider stochastic processes described by drifting and mean-reverting deterministic forces.

- We evaluated the PE and PSC measures for length scales (or permutation orders) $3 \leq n \leq 7$.

- Due to the invariance properties of the ordinal patterns under non linear monotonous transformations, the results go beyond the analyzed models.

- They come into being a parsimonious tool to capture the nature of the deterministic forces governing the examined signal, without reference to a specific model

Concluding remarks

- New relevant parameters arise as compared to standard entropy measures , meaning that H[P] and C[P] evaluated onto P[Π] present new content of information.

- A key n-invariant outcome arise, that is, for any n , the H[P] values for both class of models keep complementary.

- This result impart to PE the power of discriminating real data according to the nature (drifting or reverting) of the deterministic forces.

Concluding remarks

- For the drift-class of processes, PE measures are rather stable with increasing n, while the PSC measures show non-trivial dependence on n. This means that the new information contents are captured most efficiently by C[P] as n increases.

- As a corollary of the strong n-dependence of C[P], similar drifting systems (with close parameters) may exhibit crossovers as n increases.

- Some time series may look complex in short time intervals whereas the actual signal is not → empirical analysis based on low-order complexity measures should be taken with caution.

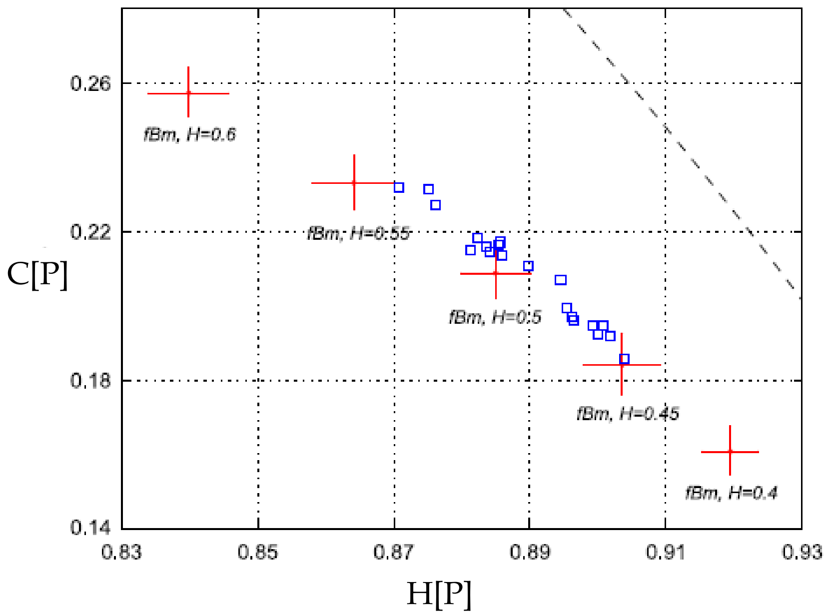
Fine-grained Permutation Entropy (FGPE)

Permutation Entropy focuses only on the order of the elements in the time series, **disregarding the magnitude of the difference** between neighboring elements.

A factor q is introduced to **quantify the major difference between them** and it is appended to the permutation pattern as an **additional element**.

In the calculation of FGPE, one considers the relative frequency of this new defined permutation pattern . In this way, the **FGPE distinguish two sequences according to the magnitude of the fluctuations.**

PE - PSC Plane



Commodity indexes (blue squares) and numerical realizations of FBM for n=6
Zunino et al, Physica A 390, 876 (2011)

- Commodity markets of several sectors are closer to FBM with Hurst exponent in the range [0.45,0.55] signaling weak long range order

- Conversely, in our approach, predictability is associated with the presence of local trends