Ayrton Soares Ribeiro

Inferring the nature of deterministic sources of real series through Permutation Entropy and Statistical Complexity measures

Dissertação de Mestrado

Thesis presented to the Programa de Pós-Graduação em Física of the Departamento de Física do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Mestre.

Advisor: Prof. Rosane Riera Freire

Rio de Janeiro
April 2013

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Ayrton Soares Ribeiro

**Inferring the nature of deterministic sources
of real series through permutation entropy
and statistical complexity measures**

Thesis presented to the Programa de Pós-Graduação em Física of the Departamento de Física do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Mestre.

Profa. Rosane Riera Freire
Advisor
Departamento de Física – PUC-Rio

Prof. Helio Côrtes Vieira Lopes
Departamento de Informática – PUC-Rio

Prof. Marcio Argollo Ferreira de Menezes
UFF

José Eugenio Leal
Coordinator of the Centro Técnico Científico – PUC-Rio

Rio de Janeiro, April 3rd, 2013.

Ayrton Soares Ribeiro

Ayrton Soares Ribeiro graduated from the Universidade Federal do Pará (Belém, Brazil) in Física. During that time he researched in dynamic systems and method of control of chaos. In addition to numerical skills, he learned programming in VBA, C and MATLAB, which were fostered during his research in stochastic processes and econophysics at PUC– Rio.

Bibliographic data

To God and my family
which always support me.

Acknowledgments

Abstract

Ribeiro, Ayrton Soares; Freire, Rosane Riera (Advisor). **Inferring the nature of deterministic sources of real series through Permutation Entropy and Statistical Complexity measures.** Rio de Janeiro, 2013. 53p. Dissertação de Mestrado — Departamento de Física, Pontifícia Universidade Católica do Rio de Janeiro.

The scope of this dissertation is to infer the character of the forces controlling complex systems modeled by Langevin equations, by recourse to information-theory quantifiers. We evaluate in detail the permutation entropy (PE) and the permutation statistical complexity (PSC) measures for two classes of similarity of stochastic models characterized by drifting and reversion properties, respectively, employing them as a framework for the inspection of real series. We found new relevant model parameters for PE and PSC measures as compared to standard entropy measures. We determine the PE and PSC curves according to these parameters for different permutation orders n and infer the limiting measures for arbitrary large order. Although the PSC measure presents a strongly scale-dependent behavior, a key n-invariant outcome arises, enabling one to identify the nature (drifting or reversion) of the deterministic sources underlying the complex signal. We conclude by investigating the presence of local trends in stock price series.

Keywords

Resumo

Ribeiro, Ayrton Soares; Freire, Rosane Riera. **Inferindo a natureza das fontes determinísticas de séries reais através de medidas de Entropia de Permutação e Complexidade Estatística.** Rio de Janeiro, 2013. 53p. Dissertação de Mestrado — Departamento de Física, Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo dessa dissertação é inferir o caráter das forças que governam os sistemas complexos modelados por equações de Langevin, utilizando quantificadores provenientes da teoria de informação. Avaliamos em detalhes as medidas de entropia de permutação (PE) e de complexidade estatística de permutação (PSC) para duas classes de similaridade de modelos estocásticos, caracterizadas por propriedades de arrasto ou de reversão, respectivamente, empregando-as como referência para a inspeção de séries reais. Encontramos novos parâmetros relevantes dos modelos para as medidas de PE e PSC, em relação a medidas tradicionais de entropia. Determinamos as curvas de PE e PSC de acordo com esses parâmetros para diferentes ordens de permutação n e inferimos as medidas limites para uma ordem arbitrariamente grande. Apesar de a medida PSC apresentar comportamento fortemente dependente da ordem de permutação considerada, encontramos um importante resultado n-invariante, que permite identificar a natureza (de arrasto ou de reversão) das fontes determinísticas subjacentes ao sinal complexo. Concluímos investigando a presença de tendências locais em séries de preços de ações.

Palavras–chave

Sistemas Complexos; Processos Estocásticos; Entropia; Econofísica; Teoria da Informação.

Contents

List of Figures

List of Tables

–Você sabe quem foi o primeiro a explicar a verdadeira origem do arco-íris? – perguntei.

–Foi Descartes – ele respondeu. Depois de um momento, me olhou nos olhos. –E qual você acha que foi a característica do arco-íris que mais se destacou aos olhos de Descartes para inspirá-lo na sua análise matemática? – perguntou.

–Bem, o arco-íris na verdade é o pedaço de um cone que surge como um arco das cores do espectro quando gotas d'água são iluminadas pelo sol atrás do observador.

–E?

–Suponho que sua inspiração tenha sido a compreensão de que o problema podia ser analisado a partir de uma única gota d'água e da geometria da situação.

–Você está deixando de lado uma característica fundamental do fenômeno – ele disse.

–Tá legal, desisto. Para você, o que teria inspirado a teoria dele?

–Eu diria que sua inspiração veio pelo fato de ele achar que os arco-íris eram lindos.

Leonard Mlodinow, *O arco-íris de Feynman.*

1
Introduction

When it comes to the study of complex systems, uncertainty plays an ubiquitous and significant role. Any approach that can provide tools which allow scientists to measure and manipulate it is also a contribution to a better description of the complex phenomena.

There has been considerable interest in quantifying the complexity of empirical time-series, despite the ill-defined concept of complexity. Amongst the approaches proposed to date, there are many entropy-based methods (1, 2, 3). Nevertheless, classical entropy measures neglect temporal relationship between neighboring values of time series, so that possible temporal patterns in the process are not accounted for.

Recently, it was introduced an entropy-based complexity measure, namely, the permutation entropy (4), which was designed to characterize the irregularity of the local temporal structure of the dynamic variables. In the course of time, other authors have proposed generalized complexity measures (5, 6, 7). In particular, one can distinguish the entitled statistical complexity measure, which accomplished to discriminate different dynamical behaviors such as periodic, chaotic or stochastic (8).

Here, we will be concerned on the composition of the permutation entropy and the statistical complexity measures to quantify not only diversity but also the correlational structures presented in the hidden dynamics of real-world time series.

Our analysis is based on the assumption that the observed signal evolve according to a deterministic source (predictable from the past) plus unpredictable stochastic fluctuations due to internal degrees of freedom or the environment. In many applied fields such as biomedical sciences, eco-systems and social-economic systems, the characterization of such sources is essential. For example, in the biomedical domain, the analysis of physiologic data from entropy-based complexity measures allows one to distinguish between normal and pathological recordings (9, 10). Another example, in the realm of eco-systems, the complexity of daily flow rate of river systems shows changes along time due to human intervention on the soil (11).

However, it is considerably difficult to find the nature of the deterministic forces describing the evolution of complex systems, which provides the optimal framework for predicting its behavior. In this work we will follow this issue by analyzing in detail some information-based quantifiers.

In order to introduce the reader to this technique, we first present in chapter 2 some measures based on information theory. We define and analyze the main properties of Shannon's entropy, Kullback-Leibler divergence and the Jensen-Shannon divergence.

Next, in chapter 3 we combine the previous measure to build the statistical complexity measure and apply it on the logistic map via amplitude statistics and symbolic analysis. At this point, the reader should be ready to understand the need of a special methodology for constructing an histogram which captures the transition structure present in time series of a given process.

Therefore, in chapter 4 we explain in details the methodology proposed by Bandt & Pompe (4), known as permutation entropy, and illustrate its powerfulness by arriving at the best results for the logistic map: the permutation entropy and statistical complexity measure are capable of distinguishing noise from chaos.

Then, we proceed to chapter 5 where we investigate, via the presented technique, two representative classes of stochastic models showing drifting or reversion properties. The framework of our approach is to consider a parsimonious number of similarity classes, which encompass diverse models, and employ them as a reference tool for the inspection of determinism in real world noisy data sets.

After, in chapter 6, these measures are computed for financial data, namely, price series from American and Brazilian markets. By examining the levels of permutation entropy and complexity measures of both empirical and theoretical models, we achieve important clues about the nature of the underlying market forces. Finally, we summarize the results and findings in chapter 7.

2
Principles of Information Theory

In order to decrease the uncertainty of a given process, one can collect more information about it. However,

> "More often than not, more knowledge may increase the uncertainty and therefore we cannot claim that the so-called information explosion has led to a proportional diminution of ignorance in the world we live."

> J. N. Kapur (12)

In this work, probabilistic uncertainty is investigated. We shall be concerned only with processes that are appropriate to probabilistic modeling and which events are well defined such that there is no ambiguity concerning them. For example, one may analyze uncertainties about a football team winning a championship, or which brand of chocolate Brazilians prefer, or what grade a student's work will receive.

2.1
Principle of Maximum Entropy

Consider a stochastic process with N possible outcomes. The only information known is that for any set of probabilities $\{p_i, i = 1, ..., N\}$ there is a normalization constraint, that is

$$\sum_{i=1}^{N} p_i = 1 \tag{2-1}$$

Different probability distributions have different uncertainties associated with them. For instance, when comparing the probability distribution of winning a lottery ($[\epsilon, 1 - \epsilon]$ with $\epsilon << 1$) with the probability distribution for an unbiased coin ($[0.5, 0.5]$), the latter is much more uncertain. The uncertainty associated with probability distribution P is known as probabilistic uncertainty. Since the work of Shannon (13), it is called entropy[1] and is denoted by $S[P]$.

---

[1]The connection with thermodynamics will be explained in the next section.

On the other hand, uncertainty associated to a process can be minimized using all the information available. Consider for instance an experiment with unknown behavior. First, when looking at the literature one discovers that the experiment can produce at each realization one of six different outcomes only, leading to probability distributions limited to $(p_1, p_2, \ldots, p_6)$. In addition, if mean and variance are given, the uncertainty about the process is reduced even more, because a smaller set of distributions can show the given statistical moments.

Consider the set of probability distributions satisfying the given constrains and the respective associated entropy values. For this set there is a range of entropy $[S_{min}, S_{max}]$. Then, when extra constraints are revealed, the set of distributions is reduced and S_{max} decreases (or at least it doesn't increase) while the S_{min} increases (or at least it doesn't decrease).

If one finds a total of six linear independent constraints, the probability distribution P^* will be uniquely determined, leading to the uncertainty measure $S[P^*]$ for the process. However, what probability distribution one shall assign if not enough constraints are given? The answer is furnished by the *Principle of Maximum Entropy*:

Out of all probability distributions consistent with a given set of constraints, choose the one that has maximum uncertainty.

This statement is known as the principle of ancient wisdom and also the principle of scientific objectivity and honesty, as any other choice of distribution would imply the use of information that was not given.

2.2
Shannon's Entropy

In order to determine the most uncertain probability distribution one needs a measure of uncertainty. One famous candidate is the Shannon's information measure. In the derivation of his measure, Shannon(13) considered the properties that this measure should satisfy. They are:

1. $S(p_1, p_2, \ldots, p_N)$ should be continuous function of $p_1, p_2, \ldots, p_N$;

2. $S(1/N, 1/N, \ldots, 1/N)$ should be a monotonic increasing function of N;

3. $S(p_1, p_2, \ldots, p_N) = S(p_1 + p_2, p_3, \ldots, p_N) + (p_1 + p_2)S(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

Shannon proved that the only expression which satisfies these properties is

$$S(p_1, p_2, \ldots, p_N) = -k \sum_{i=1}^{N} p_i \log p_i \qquad (2\text{-}2)$$

which is mathematically equivalent to the Boltzmann classical entropy. In order to maintain property 2, k has to be a positive constant and for simplicity one may assign $k = 1$. In addition, this measure satisfy other important properties, such as

- non-negativity: $S(p_1, p_2, \ldots, p_N) \geq 0$
- permutation symmetry: $S(p_1, p_2, \ldots, p_N) = S(p_2, p_1, \ldots, p_N)$
- differentiability: $\frac{\partial S}{\partial p_i} \exists$, for $0 < p_i \leq 1$
- concavity: $\frac{\partial^2 S}{\partial p_i^2} < 0$, for $0 < p_i \leq 1$

In particular, the concavity property ensures that a local maximum is always a global maximum. This property turns the Shannon's measure suitable for constrained optimizations.

As a first application, consider that the only information available about the probability distribution is the normalization constraint. Then, in order to maximize Shannon's entropy given just this constraint, one can use Lagrange's Method of multipliers.

First, construct the Lagrangian functional using constraint equations:

$$L = -\sum_{i=1}^{N} p_i \log p_i - \lambda \left(\sum_{i=1}^{N} p_i - 1 \right) \qquad (2\text{-}3)$$

The probability distribution is obtained by maximizing the above functional:

$$\delta L = 0 \Rightarrow p_i = e^{-(\lambda+1)} \qquad (2\text{-}4)$$

Therefore, by taking the normalization constraint into account:

$$\sum_{i=1}^{N} p_i = 1 \Rightarrow p_i = e^{\lambda+1} \qquad (2\text{-}5)$$

$$\text{or} \quad P \equiv U = \left\{ \frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N} \right\} \qquad (2\text{-}6)$$

Thus, it is shown that the distribution which maximizes the Shannon's entropy is the uniform distribution. This in agreement with Laplace's principle

of insufficient reason, which states that in the absence of any reason to the contrary, one should take all outcomes to be equally likely. In this case, the uncertainty measure is

$$S_{max} = S[U] = \log N \tag{2-7}$$

In contrast, $S_{min} = 0$ is obtained for processes with maximum order, i.e. one p_i is equal to unity while the others are null. In such cases, one have complete predictability.

$$S_{min} = S(1, 0, \dots, 0) = 0 \tag{2-8}$$

In Appendix A we illustrate the optimization procedure for processes described by continuous probability distribution constrained by the the first two moments.

2.3
Principle of Minimum Directed Divergence

This principle focuses on a "distance" of a probability distribution P from another distribution Q. When extra information is given about the system, a different probability distribution shall be designated.

Out of all probability distributions consistent with a given set of constraints, select the one that is the most similar to the uniform distribution.

This is called the *Principle of Minimum Divergence*. Naturally, it demands a measure of dissimilarity, distance or divergence in the probability distribution space.

An elaborate measure which features some desired properties was developed by Kullback and Leibler and is defined by

$$D_K[P||Q] = \sum_{i=1}^{N} p_i \log \frac{p_i}{q_i} \tag{2-9}$$

In fact, the Kullback-Leibler measure is a directed divergence, i.e. for any $P \neq Q$, $D_K[P||Q] \neq D_K[Q||P]$. In addition, it can be proved that $D_K[P||Q]$ is non-negative. Here is the proof:

Let $q_i = p_i(1 + \epsilon_i)$ where $(1 + \epsilon_i) > 0 \,\forall\, i$. By computing the sum over all

i, one gets

$$\sum_{i=1}^{N} p_i \epsilon_i = 0 \tag{2-10}$$

Then,

$$
\begin{aligned}
D_K[P||Q] &= \sum_{i=1}^{N} p_i \log \frac{p_i}{p_i(1+\epsilon_i)} \\
&= -\sum_{i=1}^{N} p_i \log (1+\epsilon_i) \\
&= \sum_{i=1}^{N} p_i(\epsilon_i - \log (1+\epsilon_i))
\end{aligned}
\tag{2-11}
$$

The function $f(x) = x - \log(1+x)$, defined for $(1+x) > 0$, is positive and has just one root at $x = 0$. Therefore,

$$D_K[P||Q] > 0 \text{ for } P \neq Q \tag{2-12}$$

and

$$D_K[P||Q] = 0 \text{ iff } P = Q \tag{2-13}$$

The use of the Kullback-Leibler measure in constrained optimization methods enables one to write the p_i as exponential functions (as the maximization of Shannon's entropy). Thus, the optimal probability values are never negative.

A central result is provided by this divergence when the uniform distribution U, which is the most uncertain distribution, is taken as reference: the closer a distribution is to U, the greater is its uncertainty:

$$D_K[P||U] = S[U] - S[P], \tag{2-14}$$

which means that minimum divergence is equivalent to minimum cross-entropy. Moreover, the above expression links the principles of maximum entropy and minimum cross-entropy, and sets the bounds of $D_K[P||U]$:

$$0 \leq D_K[P||U] \leq \log N \tag{2-15}$$

On the other hand, there is no proof of uniqueness for the divergence measure (as there is for information/entropy given by Shannon). This makes the choice of the divergence arbitrary and one may choose other ones instead

of Kullback-Leibler measure. On the following, we present three different measures of divergence for pairs of probability distributions.

First, consider the simplest measure of divergence, the Euclidean norm:

$$D_E[P,Q] = \sum_{i=1}^{N}(p_i - q_i)^2 \tag{2-16}$$

Its simplicity comes from the fact that it disregards the stochastic nature of $\{p_i\}$. In contrast, the Wooters measure of statistical distance is defined as (14):

$$D_W[P,Q] = cos^{-1}\left(\sum_{i=1}^{N}(p_i)^{\frac{1}{2}} \cdot (q_i)^{\frac{1}{2}}\right) \tag{2-17}$$

Both measures are also non-negative but, differently from Kullback measure, they are symmetric divergences. It happens that the Kullback measure of directed divergence can be used to form a symmetric measure. Consider the *Jensen-Shannon divergence* (15):

$$J^\alpha[P||Q] = D_K[\alpha P||\alpha P + (1-\alpha)Q] + D_K[(1-\alpha)Q||\alpha P + (1-\alpha)Q] \tag{2-18}$$

with $0 < \alpha < 1$. In the special case ($\alpha = 1/2$), one gets a symmetric measure.

The Jensen-Shannon divergence is non-negative because it consists of a sum of non-negative terms (Kullback measures). In addition, one can write the above expression in terms of Shannon's entropies:

$$J^\alpha[P||Q] = S[\alpha P + (1-\alpha)Q] - \alpha S[P] - (1-\alpha)S[Q] \tag{2-19}$$

which, reinforce that the Shannon's measure has negative concavity. Moreover, it leads to the Jensen inequality (16):

$$S[\alpha P + (1-\alpha)Q] \geq \alpha S[P] + (1-\alpha)S[Q] \tag{2-20}$$

A generalization of this measure can be made to compare more than 2 probability distributions (15, 17) and there is also a quantum version of it (15).

For the symmetric measure with $\alpha = 1/2$, we denote $J^{\frac{1}{2}}[P||Q]$ as $J[P,Q]$ and write

$$J[P,Q] = S\left[\frac{P+Q}{2}\right] - \frac{S[P]}{2} - \frac{S[Q]}{2}, \tag{2-21}$$

which is the divergence measure used in this work. The bounds of $J[P,Q]$ for any $N \geq 2$ are (15, 17):

$$0 \leq J[P,Q] \leq \log 2 \tag{2-22}$$

In addition, when setting Q to uniform distribution, one achieves at:

$$J[P,U] = S\left[\frac{P+\frac{1}{N}}{2}\right] - \frac{S[P]}{2} - \frac{\log N}{2}, \tag{2-23}$$

For this measure, for each N one obtains different values. One must set $P = \{1, 0, ..., 0\}$, in order to get the maximum of $J[P,U]$, which is

$$J_{max} = -\frac{\left(\frac{N+1}{N}\right)\log(N+1) - \log N - 2\,\log 2}{2} \tag{2-24}$$

Note that when $N \to \infty$, J_{max} converges to maximum $J[P,Q]$.

3
Complexity Measures

3.1
Complexity Vs Randomness

Several generalized entropy measures have been proposed in the literature to quantify complexity (1, 2, 3). These information measures have in common an increasing behavior with disorder. However, random data would exhibit rather low complexity in their statistical description. In the following, we present an information measure that catch the endeavor required to model statistically the data.

3.2
Statistical Complexity Measures

An information measure $I[P]$ characterizes a given probability distribution P. Considering that P is associated to the possible states or patterns of a physical process, one may achieve maximal knowledge if $I[P] = 0$ is obtained. On the opposite side, for $I[P] = I_{max}$ maximal ignorance is reached. This two extreme circumstances (maximum foreknowledge and maximum randomness) can be regarded as "trivial" ones. It follows that to quantify complexity, one should use measures not fully dependent to the degree of randomness. After that, one considers the statistical complexity measure:

$$C[P] = H[P] * Q[P], \tag{3-1}$$

where $H[P] = S[P]/S_{max}$ is the normalized Shannon entropy and $Q[P]$ is a measure of disequilibrium (6), defined as:

$$Q[P] = Q_0 * J[P, U]. \tag{3-2}$$

where $Q_0 = 1/J_{max}$ is a normalization factor.

Note that from Eq. (2-23), the disequilibrium $Q[P]$ (3-2) is different from zero if there exist privileged states among the others, so that the $C[P]$ measure (3-1) will ascribe null value of complexity for both regular (totally predictable) and random (totally unpredictable) series. Therefore, systems composed by a mixture of regular and stochastic processes, exhibiting a behavior in-between these two extremes, although encompassing intermediate content of information, are more complex.

An important issue emerges, that is the characterization the complexity measure bounds for a given value of entropy. They can be obtained by considering the extreme distributions (6) with probability values $p_j, 1 \leq j \leq N$ of the form:

$$0 \text{ for } 1 \leq j \leq m \tag{3-3}$$

$$p \text{ for } m+1 \leq j \leq m+n$$

$$(1-pn)/(N-m-n) \text{ for } m+n+1 \leq j \leq N$$

By evaluating the entropies and complexities of the above distributions, one arrives at significant results for the complexity bounds. The results are illustrated in Fig. 3.1, which shows the complexity bounds in the *Complexity-Entropy*(C-H) plane for $N = 6$ and $N = 1000$.
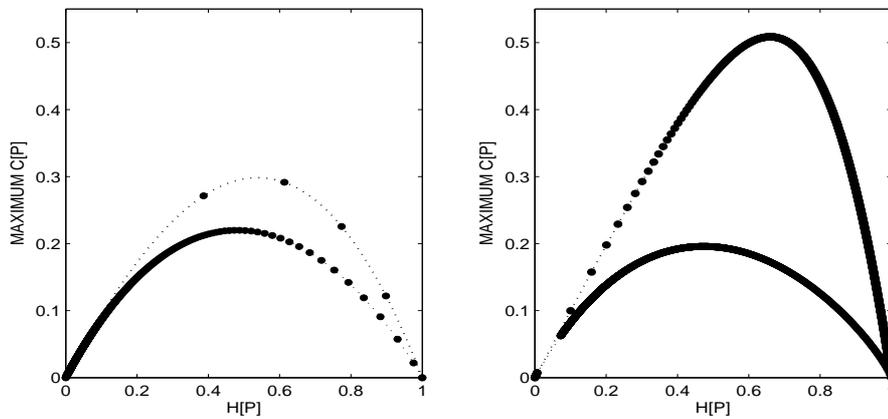
Figure 3.1: Upper and lower complexity bounds for $N = 6$ (left panel) and $N = 1000$ (right panel). Full dark circles indicate the calculated points. A polynomial interpolation was done in order to get the maximum curve for $N = 6$.

When $m = 0$ and $n = 1$, one has $P_{low} = \left\{p_1 = p; p_j = \frac{1-p}{N-1}, 2 \leq j \leq N\right\}$ which furnishes the minimum value $C[P_{low}]$ for the complexity measure Eq. (3-1), considering the set of distributions with entropy value $H = H[P_{low}]$. By varying p, we vary P_{low} and thus, we can get an infinite number of points for the lower complexity bound in the C-H plane. In addition, when $n = 0$, for each

$0 < m \leq N$, one has $P_{upp} = \left\{ p_j = 0, 1 \leq j \leq m; p_j = \frac{1}{N-m}, m+1 \leq j \leq N \right\}$ which furnishes the maximum value $C[P_{upp}]$ for the complexity measure Eq. (3-1), considering the set of distributions with entropy value $H = H[P_{upp}]$. By varying m, we vary P_{upp}, leading to N points for the upper complexity bound in the C-H plane. In Fig. 3.1, we interpolated these points to get a smooth curve for $N = 6$. Note that the limiting upper and lower bounds curves are valid for all distributions with the same number N of partitions.

3.3
Complexity-entropy plane: application to logistic map

In order to illustrate the applicability of the statistical complexity measure, we consider the well known logistic map

$$x_{n+1} = b \; x_n(1 - x_n) \tag{3-4}$$

with b varying in the interval $[3.5, 4.0]$. This system can show periodic or chaotic behavior depending on the value of b. When looking at Fig. 3.2, one can see that for regions of periodic behavior the Lyapunov exponent decreases.
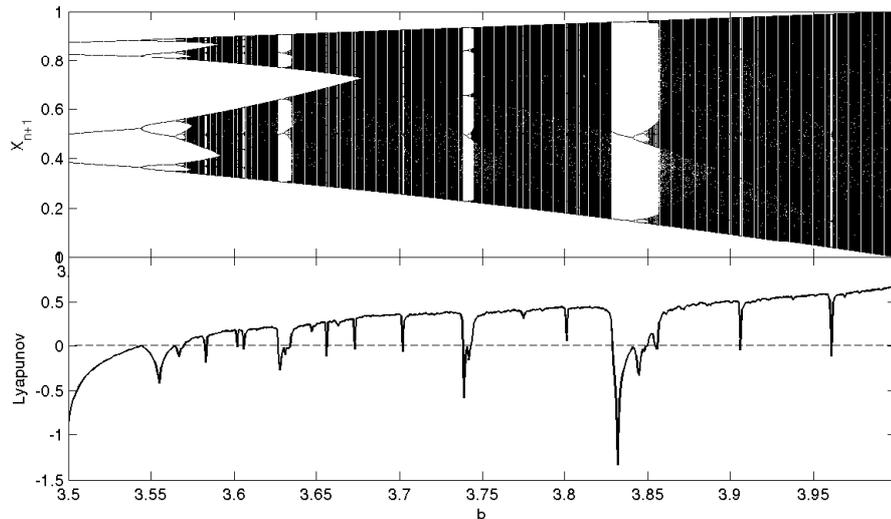
Figure 3.2: Bifurcations Diagram (top panel) and Lyapunov exponent (bottom panel) for the logistic map. Each bifurcation corresponds to a period doubling.

One crucial step is the definition of the discrete probability density function (PDF) associated to the process. A simple approach consists of obtaining it via amplitude statistics, dividing the interval of x_i values into a finite number N of non overlapping subintervals. One then employs the usual histogram-method, based on counting the relative frequencies of the time series values within each subinterval.

For the logistic map, we reproduced the results of (7), by generating 10^6 data values for each specific b value. We varied the logistic map parameter b in the range $[3.5, 4.0]$ with size step $\Delta b = 0.0005$, forming 10^3 time series. The range of data values belongs to the interval $[0, 1]$ and was divided in 10^3 subintervals for the construction of the histogram. In Fig. 3.3, we show the entropy and statistical complexity measures according parameter b. It is observed an "envelop function" in which the entropy measure increases globally with b. The rapid growth at $b = 3.5695$ identifies the exponential multiplication of periodic behaviors. Moreover, the many local falls of the entropic values correspond to the periodic windows where the logistic map presents regular trajectories. In addition, at the full chaotic configuration ($b = 4$), entropy is very close to unity. This happens because, in this case, the logistic map has an almost uniform probability density function, given by

$$p(x) = \frac{1}{\pi \sqrt{x \, (1 - x)}}$$

for $0 < x < 1$. In Fig. 3.4, we show the empirical distribution for this configuration, obtained numerically from Eq. (3-4).
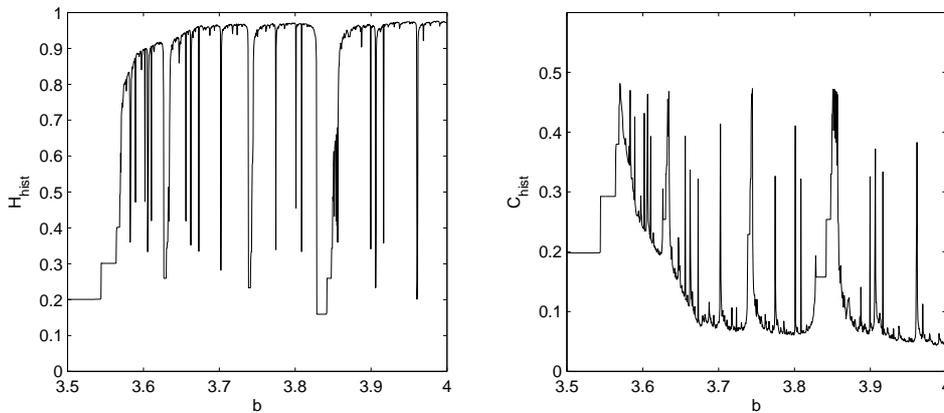
Figure 3.3: Entropy and statistical complexity measures for the logistic map based on amplitude statistics with $N = 10^3$.

On the other hand, when analyzing the statistical complexity measure, although one can observe an abrupt growth for the same parameter $b = 3.5695$, this measure decreases as parameter b increases further, and, instead of multiple drops in the periodic windows, several bursts are observed. This behavior is not acceptable given that periodic trajectories are less complex than the chaotic counterparts.

For a better overview of the relationship between entropy and statistical complexity it is useful to draw the *complexity-entropy plane*. By comparing
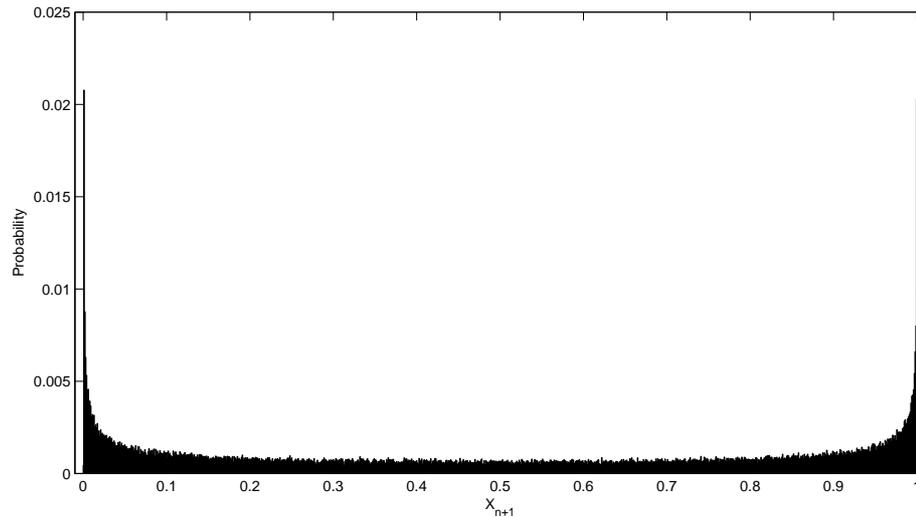
Figure 3.4: Probability density function for the logistic map with $b = 4$ based on amplitude statistics with $N = 10^3$.

the range of entropy values related to periodic windows and with the chaotic regime, one finds that periodic behaviors are located in the left side of the plane ($H < 0.5$) and do not reach maximum complexity (see Fig. 3.5). This property, enables the identification of regular dynamics (i.e. periodic) underlying a given process, even if the model parameters are unknown (7). However, it is not clear how to discriminate chaotic or stochastic dynamics, because both have similar entropy/complexity values. This issue will be addressed in the next chapter.

3.4
Symbolic Analysis

As an alternative to histograms based on single amplitude values, one might consider PDFs related to patterns of a sequence of values. In order to accomplish it, a symbolization technique is required to transform a sequence of real data values into a sequence of symbols belonging to a finite size alphabet. Although there is not just one way of coding a time series, a popular and simple symbolization technique is the so-called binary treatment that translates each data value to 0 if $x \leq x^*$ or to 1 if $x > x^*$. Considering strings of length L, one has 2^L symbol sequences associated to different the binary patterns, for which one constructs the histogram of the relative frequencies of occurrence.

In this approach, we transformed the original time series generated from the logistic map Eq. (3-4) into binary sequences, with $x^* = 0.5$. We considered strings of length $L = 10$ and built the histogram representing the frequency of occurrence of the $N = 1024$ states. This PDF was used for calculating both entropy and statistical measures. By looking at Fig. 3.6, one observes
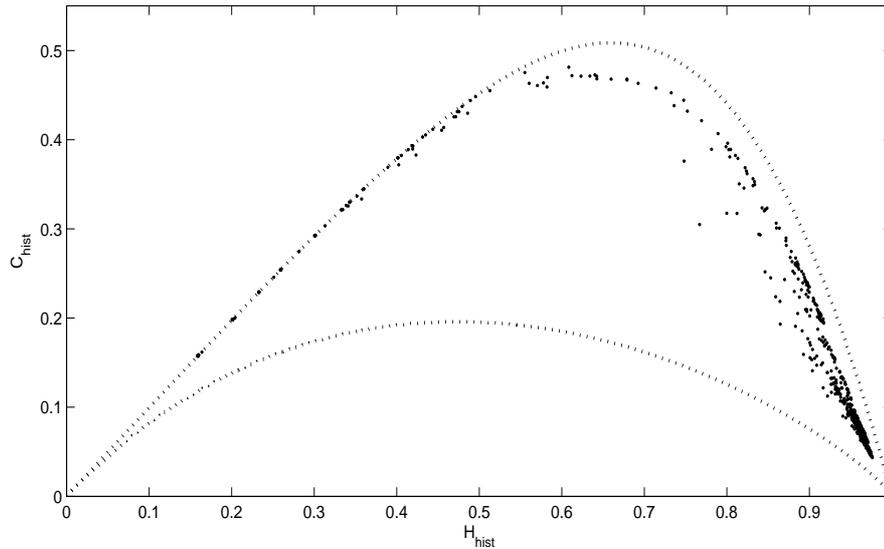
Figure 3.5: *Complexity-entropy plane* for the logistic map based on amplitude statistics with $N = 10^3$. The dotted curves are the upper and lower complexity bounds for $N = 10^3$.

that the binary-based entropy values present almost the same features of the amplitude-based approach. On the other hand, complexity shows a different behavior according to b, with a parabolic-like envelop shape and with falls in the periodic windows (6, 7), which represents an improvement compared to the amplitude-based approach. However, at $b = 4$, as the amplitude histogram is symmetric (see Fig. 3.4), it leads to a uniform binary histogram, implying similar entropy (close to unity) and complexity (close to zero) values. Thus, chaos and noise are yet indistinguishable through this procedure.
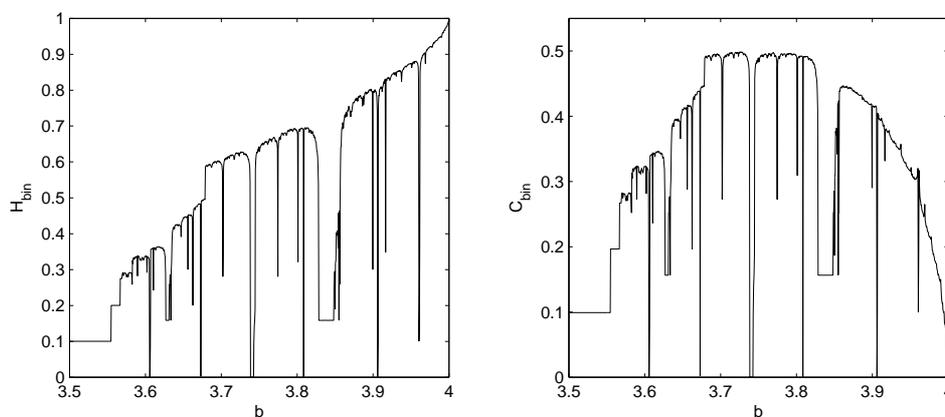
Figure 3.6: Entropy and statistical complexity for the logistic map based on binary symbolic dynamics with $N = 1024$.
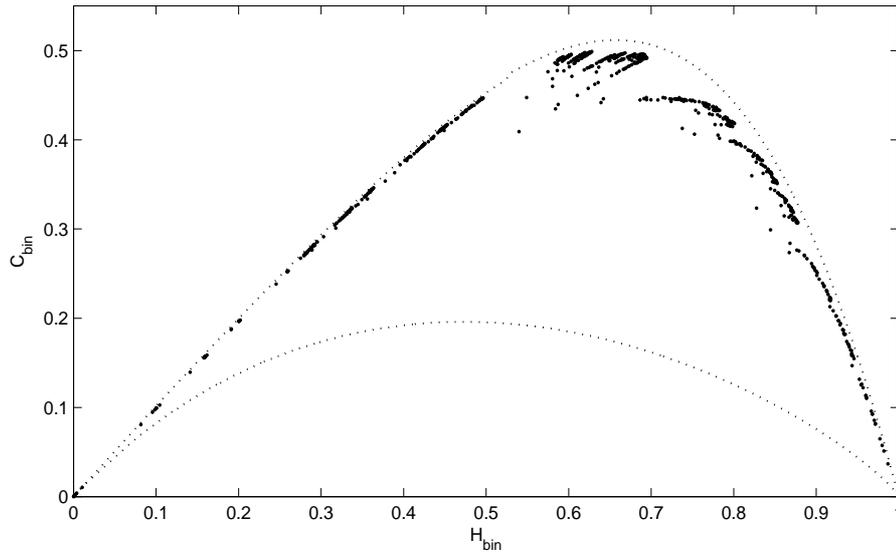
Figure 3.7: *Complexity-entropy plane* for the logistic map based on symbolic dynamics with $N = 1024$.

In sum, from both approaches, periodic behavior are restricted to the region of low entropy in the *complexity-entropy plane* and the chaotic regime is located in the complementary range of entropy. This fact turns difficult the discrimination between irregular deterministic and random process. In order to improve the results from the symbolization technique, a PDF which encompasses time causality should be adopted and is the starting point of the next chapter.

4
Permutation Entropy

There are many techniques used to distinguish regular, chaotic and stochastic behaviors, such as, entropies, fractal dimensions and Lyapunov exponents. They are all defined for typical orbits of presumably ergodic dynamical systems, and there are profound relations between these quantities. However, these measures are not defined for arbitrary time series and may give rise to misleading results when the data is not noise free. This is an important drawback when one deals with real-world applications. Moreover, noise filtering requires careful processing of the data and fine tuning of parameters, and the results cannot be reproduced without specifying details of the method.

To overcome these issues, Christoph Bandt and Bernd Pompe designed a technique called *permutation entropy* (4), which can be applied to any time series, even if there is no knowledge about the system which had generated it.

4.1
Definition

Given a time-series $\{s_t : t = 1, ..., T\}$, we assign a symbol to a string of n consecutive values $[s_{(t-n+1)}, s_{(t-n+2)}, ..., s_{(t-1)}, s_t]$ at each time $t \geq n$. One considers all $n!$ possible sort of ordering the values of the n variables and identifies each ordering as an ordinal pattern. We associate to each ordinal pattern a symbol $\Pi_i, i = 0, 1, ..., (n! - 1)$.

A good way of sorting Π is described in (18) for a sequence of n non-negative integers. Here, we extend this procedure for sequences of n real values . Basically, one starts defining index $i = 0$ for the increasing sequence $s_{(t-n+1)} < s_{(t-n+2)} < ....s_{(t-1)} < s_t$. The index i of the ordinal pattern Π_i is defined according to the hierarchy of interchanges (permutations) necessary to perform over the n neighboring values within the string to attain the reference order $i = 0$. In this way, for each n-size string one associate a permutation symbol Π of order n. In Table 4.1 we illustrate the procedure for $n = 3$.

Ordinal Pattern	Symbol Index	1st Increment	2nd Increment
012	0	$\Delta x_1 > 0$	$\Delta x_2 > 0$
102	1	$\Delta x_1 < 0$	$\Delta x_2 > 0, |\Delta x_2| > |\Delta x_1|$
021	2	$\Delta x_1 > 0$	$\Delta x_2 < 0, |\Delta x_2| < |\Delta x_1|$
120	3	$\Delta x_1 > 0$	$\Delta x_2 < 0, |\Delta x_2| > |\Delta x_1|$
201	4	$\Delta x_1 < 0$	$\Delta x_2 > 0, |\Delta x_2| < |\Delta x_1|$
210	5	$\Delta x_1 < 0$	$\Delta x_2 < 0$

Table 4.1: Ordinal patterns associated to permutation symbols $\Pi_i, i = 0, 1, ...5$ for block size $n = 3$. The strings of integer numbers 0,1 and 2 label the relative level of the values of the variables

.

Note from Table 4.1 that, by inverting the order of increments, index 1 and 2 are interchanged, the same occurring with the pair of indexes 3 and 4. However, indexes 0 and 5 are invariant under this operation because they are twofold, comprising the two situations: $|\Delta x_1| > |\Delta x_2|$ and $|\Delta x_1| < |\Delta x_2|$. Therefore they play a significant role when analyzing processes with independent increments[1].In particular, for symmetric sample distributions, their frequencies are two times larger than the frequencies of other indexes.

The original Bandt-Pompe methodology was conceived for continuous distributions, in which equal values are unlikely to appear. However, for low resolution real-world time series, one should have a special care to process strings that contains values with negligible differences. Simple solutions to this issue are (a) to break equalities by adding statistical noise to the data or (b) to interpret equalities as positive increments (4). Another approach is the modified permutation algorithm (10) which introduces additional patterns associated with strings containing equal values. Although these methodologies are easy to compute, we have adopted an alternative unbiased algorithm, which, for permutations of order $n = 3$ (which is used therein this work) reads as follows.

Whenever $s_t = s_{t+k} = s^*$, we assign $s_t < s_{t+k}$ if s^* is smaller than the string average value; otherwise we assign $s_t > s_{t+k}$. In addition, strings which three equal values are neglected. This leads to a balanced population between symbol indexes with coarse movement up (0,1 and 2) and down (3,4 and 5) shown in Table 4.1.

Next, we proceed to compute the permutation probability $P(\Pi)$. To ascertain that, in the limit of an infinite time series, the stationary distribution is reached, the underlying stochastic process must fulfill a very weak stationary condition: the probability that $s_i < s_j$ for $(j - i) \leq n$ should not depend on t (4).

---

[1]Two events A and B are independent iff $p(A \cap B) = p(A)p(B)$.

The permutation entropy of order $n \geq 2$ is defined as in Eq. (2-2), where the sum runs over all $N = n!$ permutations Π of order n. It is clear that $0 \leq S[P] \leq \log N$, where the lower bound is attained for monotonic sequence of values and the upper bound, for a completely random system. Intermediate $S[P]$ values indicate that the process presents some sort of temporal structure.

We remark that, as the permutation analysis look over the relative values of the elements, not the absolute values, the associated probability measure is scale-invariant. This means that the permutation entropy may be thought as a complementary measure to standard entropy, once the latter is affected by the magnitude of fluctuations.

An important consequence of handling the permutation probability $P(\Pi)$ compared to standard sample distribution $P(x)$ is that one improves the performance of the information-based quantifiers, capturing not only randomness but also the memory structures. In this sense, when the normalized Shannon entropy $H[P] = S[P]/S_{max}$ is measured over the permutation probability distribution $P(\Pi)$, the statistical complexity measure Eq. (3-1) reveals new aspects of the dynamical process: now, while $H[P]$ measures the diversity of the observed ordinal patterns, $C[P]$ measures the endeavor required to reproduce statistically the data flow at time scale n.

4.2
Application to logistic Map

Following the work of Bandt&Pompe, we applied the symbolic permutation methodology to the logistic map with $n = 6$, which is big enough to access the dynamics of the process.

It is worthwhile to inspect the permutation histogram for the fully developed chaos within the logistic map. Indeed, By looking at Fig. 4.1, one can see a highly complex structure. The histogram is rather non-uniform and non-continuous, i.e. there are permutation indexes that never occur in the fully chaotic map, as consequence of a deterministic process.

After that, we calculate the permutation entropy and the statistical complexity using the permutation histogram and compare the results with the approaches of the previous chapter. As we have $n! = 720$ possible patterns, for each value of b we iterated the map $10^3 n!$ times, which is close to the previous statistics made for amplitude and symbolic analysis. As it is shown in Fig. 4.2, permutation entropy behaves similarly to other entropies, exhibiting an overall increasing behavior with parameter b with falls in the periodic windows. On the other hand, for $b = 4$, it is not close to its maximum, indicating that this measure can distinguish chaos (deterministic process) from
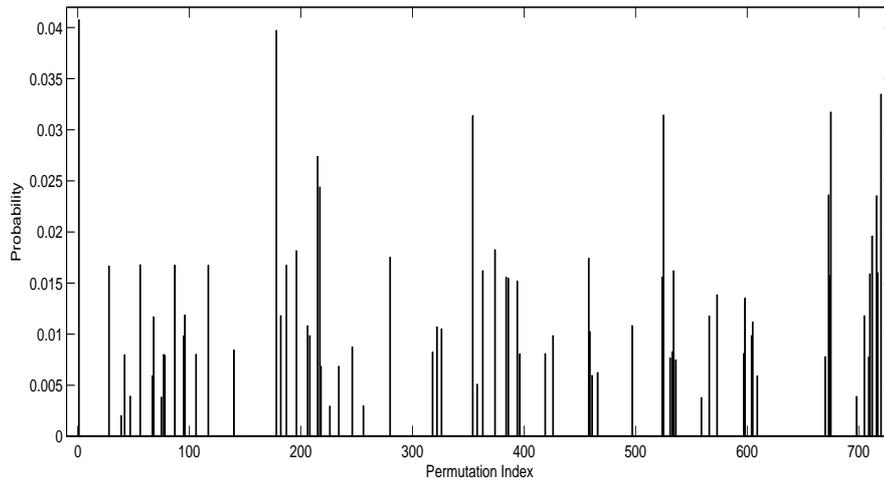
Figure 4.1: Permutation histogram for the full chaotic logistic map ($b = 4$) with $N = 6! = 720$.

noise (stochastic process). This means that the Shannon's entropy measure based on permutation histograms may be able to quantify complexity(4). For recent reviews of the application of ordinal patterns and permutation entropy, the reader is referred to (19) and (20).
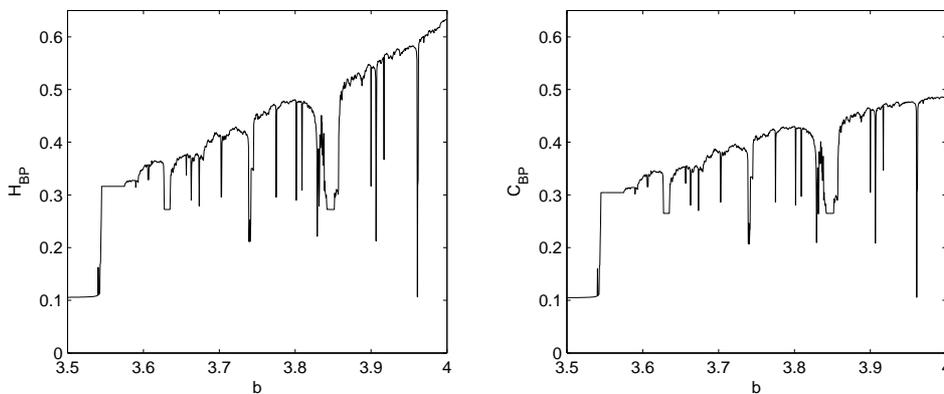
Figure 4.2: Entropy and statistical complexity measures for the logistic map based on permutations with $N = 720$.

As shown in Fig. 4.3, the *complexity-entropy plane*, now presents qualitative different processes located in distinct regions. Regular,i.e. periodic, dynamics is restricted to the left side ($H < 0.35$), while chaos exhibits intermediary permutation entropy and high statistical complexity, Therefore one expects that random processes must be on the right side. This last affirmative will be checked on the next chapter for some classes of stochastic processes.
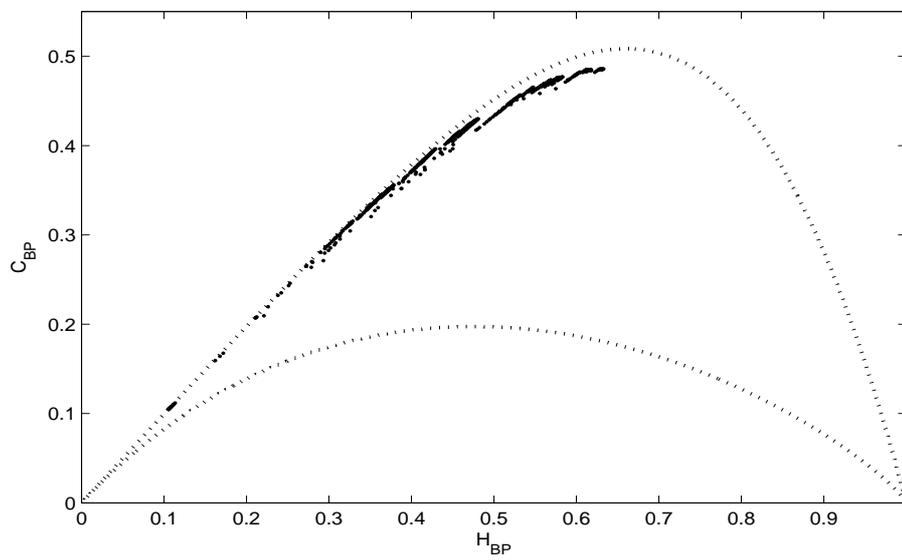
Figure 4.3: *Complexity-entropy plane* for the logistic map based on permutations with $N = 720$.

5
Inferring the nature of deterministic sources

5.1
Classes of Stochastic Processes

In physics, Brownian Motion is regarded as the motion of small particles in suspended liquid and is modeled by a *Langevin* equation:

$$m\ddot{x} = -\alpha\dot{x} + f_{ext}(t) + f(t) \tag{5-1}$$

Where x and m are the position and mass of one particle, respectively, α is the damping constant of the medium, f_{ext} is an external force applied to the particle and f represents the random shocks due to other particles.

By assuming that the second order term is not appreciable when compared to the others (this happens for example in over-damped media), we can write:

$$\frac{dx}{dt} = \frac{f_{ext}(t)}{\alpha} + \frac{f(t)}{\alpha}$$

$$dx = \mu(t)dt + \sigma(t)dW \tag{5-2}$$

where dW is the standard *Wiener process*. The last equation is called an *Ito process* and is composed by two parts: the deterministic one $\mu(t)dt$ and the stochastic one $(\sigma(t)dW)$. Different choices of $\mu(t)$ and $\sigma(t)$ lead to different models, some of them are well known. On the following, we consider three models that will be used as test cases for permutation entropy.

We start with $\mu(t) = 0$ and $\sigma(t) = \sigma$, describing the Brownian Motion (BM), which is a continuous form of a random walk. This process is mathematically equivalent to the Wiener Process, which shows some important properties (21): independent and identically Gaussian distributed increments, self-similarity, non-differentiability and non-limited variation.

The PDF based on amplitude statistics for BM at a fixed time t is a Gaussian function with zero mean and variance $\sigma^2 t$ ($\mathcal{N}(0, \sigma^2 t)$). Therefore,

one gets an increasing entropy value as function of time (12):

$$S(x;t) = \int_{-\infty}^{+\infty} -p(x) * \log(p(x)) dx$$

$$S(x;t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2 t}} \left(\log\sqrt{2\pi\sigma^2 t} + \frac{x^2}{2\sigma^2 t} \right) dx$$

$$S(x;t) = \int_{-\infty}^{+\infty} \frac{\log\sqrt{2\pi\sigma^2 t}}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2 t}} dx + \int_{-\infty}^{+\infty} \frac{x^2}{2\sigma^2 t} e^{-\frac{1}{2}\frac{x^2}{\sigma^2 t}} dx$$

$$S(x;t) = \log\sqrt{2\pi\sigma^2 t} + \frac{1}{2}$$

$$S(x;t) = \frac{1}{2}\log\left(2\pi e\sigma^2 t\right) \tag{5-3}$$

Thus, in order to obtain a stable entropy measure for BM, one needs to deal with other associated PDF, for instance, the permutation histogram.

For general $\mu = const$ and $\sigma = const$, one has the Arithmetic Brownian Motion (ABM), belonging to the class of drifting processes (therein this work, we call it Drift, for simplicity). Its trajectory gets smoother for high $|\mu/\sigma|$, and its PDF is described by $\mathcal{N}(\mu t, \sigma^2 t)$.

For $\mu(t) = -k(x - x^*)$ and $\sigma = const$, one has the Arithmetic Ornstein-Uhlebeck process, belonging to the class of Mean Reverting processes (therein the work, we call it MR, for simplicity). It is asymptotically stationary and Gaussian, with variance of the fluctuations varying with σ^2/k. Thus its PDF is described by $\mathcal{N}(x^*, \sigma^2/k)$.

With the purpose of comparing the three models we set $\sigma = 1.0$ and $\mu = k = 0.5$ and plotted one sample path for each process in Fig. 5.1. One can see that the Drift process perform a global increase while the MR and BM process fluctuate around zero. However, MR is restricted to a tighter range of values and shows smaller fluctuations.

We point out that, the transformation of variable x in Eq. (5-2) by any non-linear function with a monotonous growing character keeps the relative ordering of the variables, leading to invariant $P(\Pi)$[1]. Thus, from the *Itô*'s lemma , the present analysis of the models Drift and MR represents, in fact, a broaden class of models, including for instance, the Geometric Brownian Motion and the Exponential O-U process, respectively. In this way, one can map a rather diverse model-generated temporal patterns $\{x_t\}$ onto two classes that we identify as "generalized ordinal states". Here, we shall use them as a

[1]A monotonic decreasing transformation inverts the order of the indexes, but maintain the same entropy measures due to p_i symmetry.
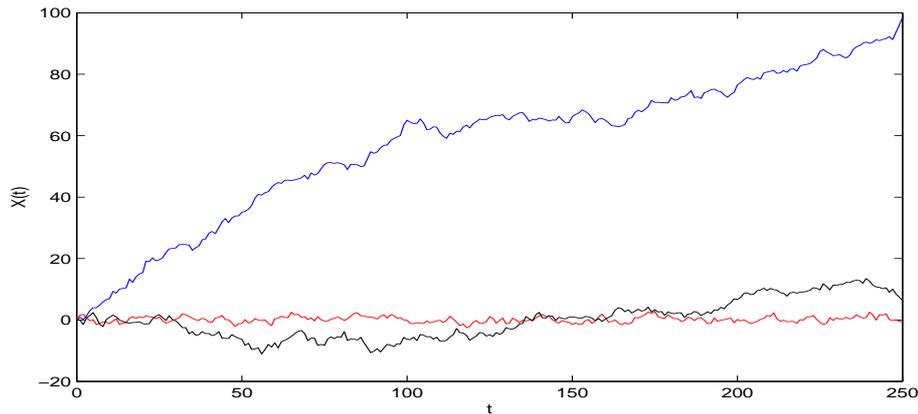
Figure 5.1: Sample paths of Drift (blue), BM (black) and MR (red) with $\mu = k = 0.5$ and $\sigma = 1.0$

parsimonious tool to capture the nature of the deterministic forces governing the examined signal.

5.2
Assessing Permutation Entropy and Statistical Complexity

Artificial time series for the quoted stochastic processes were generated by using the Marsaglia's Mother-of-All random number generator[2] (23).

We compute the set of probability $P(\Pi_i)$ of index $i = 0, ...(n! - 1)$ for permutation order $3 \leq n \leq 8$. In each case, the results are based on the statistics of blocks of n consecutive outcomes along synthetic data of total length $2^{10}n!$.

Its worthwhile to mention that one may discriminate specific time scales by considering a string of n non-consecutive (embedding delays $\tau = 2, 3, ...$) elements of the time series. This approach is particularly useful to identify intrinsic time scales of the dynamics under study (24). Our approach considers consecutive (embedding delay $\tau = 1$) elements only, however, our focus is comparing different systems under a unique basis, including invariant systems under scaling of time (Brownian motion).

New interesting features arise when looking to the relative frequency of the permutation indexes underneath the entropy and disequilibrium measures. In Fig. 5.2, we show the set of BM permutation probability $P(\Pi_i)$ of index i for $n = 8$. For this simplest class of stochastic process, one gets already complex patterns, which anticipate non-trivial information measures.

[2]it has period length of 2^{158} and passes trough all TestU01 statistical tests of randomness (22)
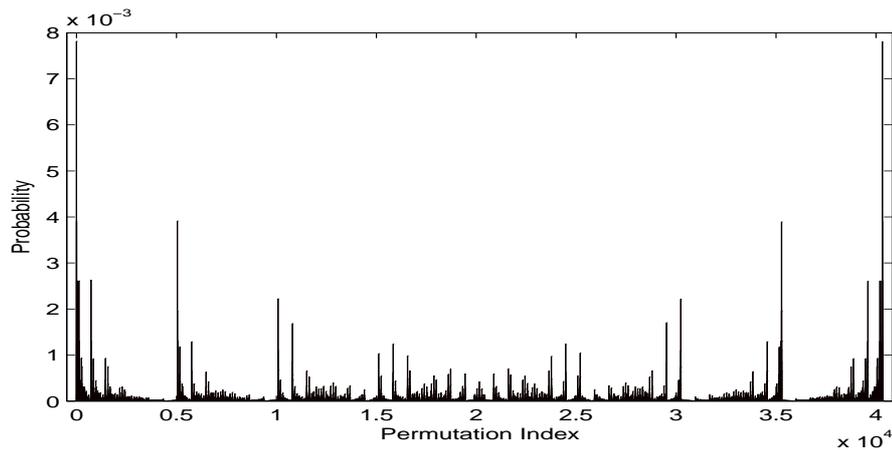
Figure 5.2: Permutation histogram for Brownian Motion with $N = 8! = 40320$.

Moreover, one may catch the relevant parameter of each model, as far as the ordinal permutation pattern is concerned. In Fig. 5.3, we illustrate this finding for Drift and MR classes, for $n = 3$. As expected for symmetric distributions of independent steps, for $\mu = 0$ or $k = 0$ respectively, one gets $p_0 = p_5 = 1/4$ and $p_1 = p_2 = p_3 = p_4 = 1/8$ corresponding to the Brownian Model. This leads to $S_{BM} = 2.5 \log 2$, that will be used as a reference permutation entropy value. Clearly, when adding a drift force (Drift class), the probabilities become asymmetric, while for a restoring force (MR class), the symmetry is sustained.
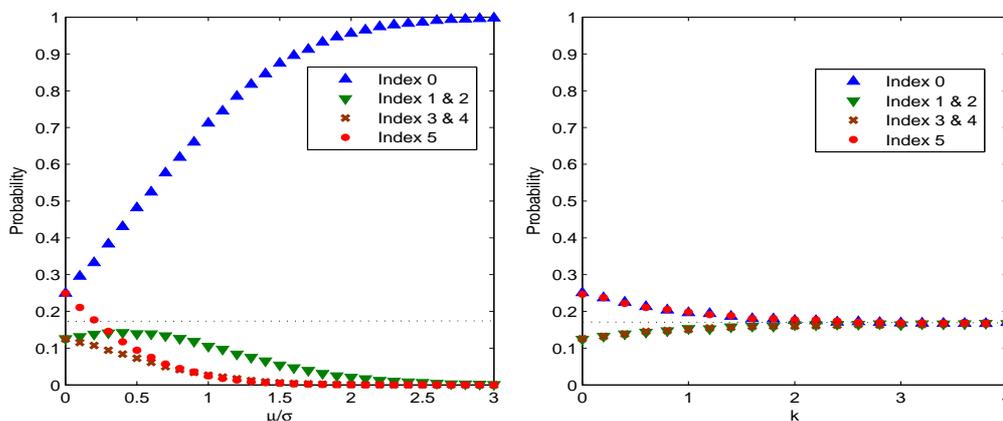
Figure 5.3: Probability $p(\Pi_i)$ of index $i = 0, ...5$ according to the respective relevant parameters $x = \mu/\sigma$ and $x = k$ for Drift processes (left panel) and MR processes (right panel). In the present case, $n = 3$.

In Fig.5.4, we present the normalized permutation entropy $H[P]$ for each analyzed process according to the respective relevant parameter. It is shown that the entropy of Drift-processes follows a q-Gaussian curve according to the order/disorder parameter μ/σ. Such behavior is in contrast with the sample
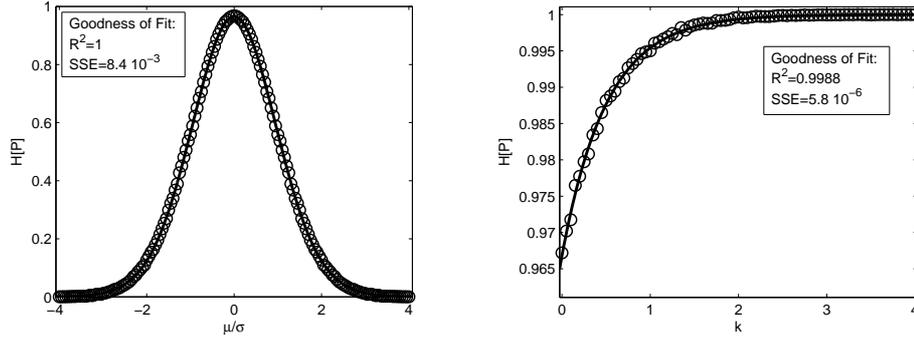
Figure 5.4: Normalized Permutation Entropy (open circles) for Drift and MR processes according to the respective relevant parameters $x = \mu/\sigma$ and $x = k$. The observed values are well fitted (lines) by a q-Gaussian curve $H_{BM}[1 + (q - 1)(x/x_0)^2]^{-1/(q-1)}$ with $q = 1.06$ and $x_0 = 1.35$ (Drift) and by an exponential curve $1 - (1 - H_{BM})exp(-(x/x_0))$ with $x_0 = 0.51$ (MR).

Shannon's entropy, that is independent of μ (similar calculation that led to Eq. (5-3), gives the same result when applied to $\mathcal{N}(\mu t, \sigma^2 t)$).

In the case of MR process, one gets an exponential growing curve according to k, with σ playing no role for the permutation entropy. This means that σ acts only as a scale factor for the MR process, not affecting the relative magnitude ordering of the variables. This behavior is also in contrast with the sample Shannon's entropy, which varies according to $\log [\sigma/\sqrt{k}]$ (similar calculation that led to Eq. (5-3), when applied to $\mathcal{N}(0, \sigma/\sqrt{k})$, gives this last result). The spring constant k acts as a restarter of the process, limiting the relative growth of the variables. In sum, the permutation entropy of both classes present properties not carried by the standard entropy, exhibiting new relevant parameters. This feature distinguishes it as an alternative information measure.

Another important finding for empirical analysis applications is that the range of the normalized permutation entropy of both models are complementary: the maximum value for Drift-processes is $H_{BM} = 2.5 \log 2 / \log 6$ (≈ 0.9671), corresponding to the BM case, while the MR model assumes still bigger values.

In attempt to attain the full description of randomness and memory structure from $H[P]$, $Q[P]$ and $C[P]$, we examined the Drift and MR models, for different permutation order n. This procedure account for the multi-scale structures embedded in time series.

In Fig. 5.5, we plot the normalized permutation entropy $H[P]$ and the normalized disequilibrium $Q[P]$ for both models. It clearly shows that the $H[P]$ and $Q[P]$ measures have opposite behaviors according to the respective model relevant parameters. As expected, Drift-processes covers a wider range

of $H[P]$ values. Note also that selecting $n = 2$, the $H[P]$ peak for Drift-processes is equal to unity, turning MR and BM processes indistinguishable. Thus, one should use $n \geq 3$ for meaningful information measures. Searching for invariant outcomes, the key finding is that $H[P]$ measures for both models keep complementary for any n. In what follows, we address the issue of the relevance of the complexity measure C[P] comparing to H[P]. To this aim, we scrutinize the information measures for Drift and MR processes.
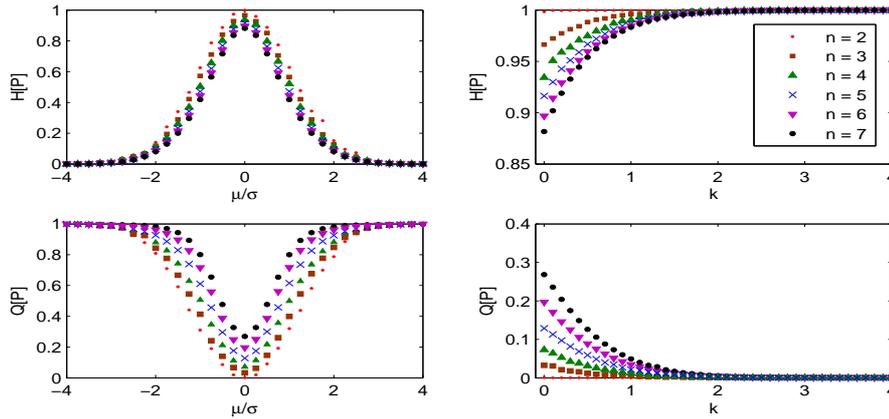
Figure 5.5: Normalized Permutation Entropy and Disequilibrium according to the relevant parameter of Drift (left panels) and MR (right panels) processes. Each color represents a different permutation order n (see inset).

First, we search for possible dependence between H and Q. In Fig. 5.6, a close inspection reveals a strong linear relationship between the two measures for small n, but a non-linear dependence emerges with increasing length scale. This non-trivial relationship according to n endorses that $H[P]$ and $Q[P]$ furnish different information contents.

The independence of both information measures implies that $C[P]$ also provides extra information when compared to $H[P]$ and we proceed with its evaluation. In Fig. 5.7, we show the complexity measure $C[P]$ according to the Drift (μ/σ) and MR (k) parameter for several permutation order n. New structures were found at all scales, which means that new information is revealed as n increases (25).

One natural question arises: once the information measures change with n, can one assign any true statistical complexity value to these model-generated series? We start this analysis by extrapolating the maximum value of $H[P]$ and $C[P]$ for Drift processes with n arbitrary large. The results displayed in Fig. 5.8 shows a well behaved tendency, furnishing $H_{max} \cong 0.80$ and $C_{max} \cong 0.56$. Notice that both extrapolated maximum values are assigned to BM ($\mu = 0$).
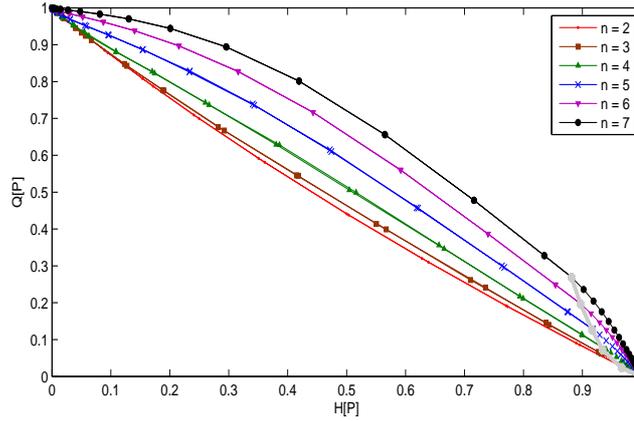
Figure 5.6: Normalized Disequilibrium Normalized versus Permutation Entropy. The gray line represents the values for the BM model, which splits the Drift (left) and MR (right) values. Each color represents a different permutation order n.
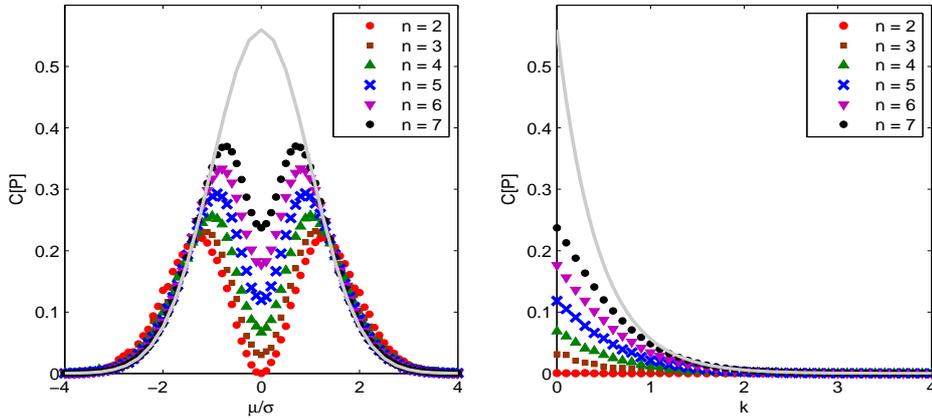
Figure 5.7: Complexity curve for different permutation orders n (see inset) for Drift processes (left panel) and MR processes (right panel). The inferred limiting shape for n arbitrary large is shown in gray.

By inspection of the left panel of Fig. 5.7, notice the faster converge of the tails as n increases. This allows to infer a limiting shape of $C[P]$ of Drift processes for n arbitrary large, shown in gray line. Seeing the Drift limiting behavior as a Gaussian function, we find the asymptotic maximum value 0.56 ± 0.06, consistent with the previous findings shown in Fig.5.8. On the other hand, from the analysis of the right panel of Fig. 5.7, for MR processes we get an exponential behavior for the limiting shape of C[P], shown in gray line, starting from the asymptotic value assigned to BM ($k = 0$).

The limiting shape of $C[P]$ for Drift-processes leads to a remarkable outcome, that is the BM model furnishes low complexity values at short timescales n, but produces the largest value at the longest scale. The large increase
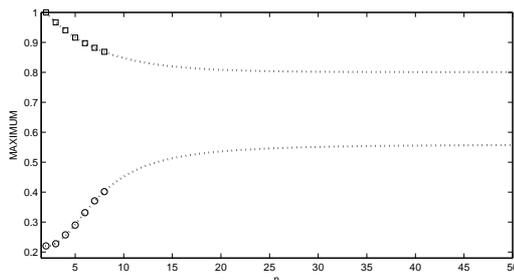
Figure 5.8: Behavior of global maximum of H[P] (open squares) and C[P] (open circles) for Drift processes according to n. Both data are well fitted with exponential curves (lines).

of the statistical complexity for BM goes along with the decreasing value of the maximum normalized entropy with n, as evidenced in Fig. 5.8. This specific case indicates that the permutation order n considered in the analysis of the time series is crucial for ranking purposes. Fig. 5.9 shows a color map of the C[P] magnitudes according to the Drift relevant parameter and permutation order n. From this figure, one can access the crossovers among the systems as n increases. This means that, the finite-scale complexity measures for Drift-like processes may significantly differ from the true values. In this case, general low-order estimates should be taken with care.
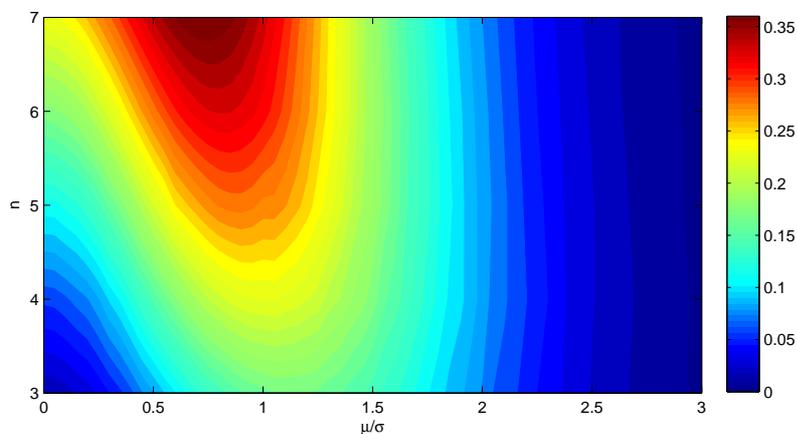
Figure 5.9: Color map of statistical complexity measure according to relevant parameters μ/σ and permutation orders n.

6
Application to Financial Series

Now we proceed with the representation of the analyzed models in the *complexity-entropy plane*, to match up to with real data. We consider daily log-price time series of the Brazilian and American stock markets. All investigated series comprises $N > 2^{10}$ data collected from January 2002 to December 2007 (before the 2008 financial crisis). Several economic sectors are represented with at least 3 companies each : Banking/Financial(B), Consumer/Industrial Goods(G), Energy(E), Mining / iron and steel industry(M), Oil & Gas(O) and Technology(T). The selected companies includes 10 companies amongst the FT Global Top 35 companies at Dec. 2007 (26), with Market Capitalization ranging from \$511.9 to \$155.9 B.

In Fig. 6.1, we display one representative time series of each market. While the Brazilian stock, USIM5, shows a strong positive trend, the American stock, IBM, fluctuates widely in the same period.
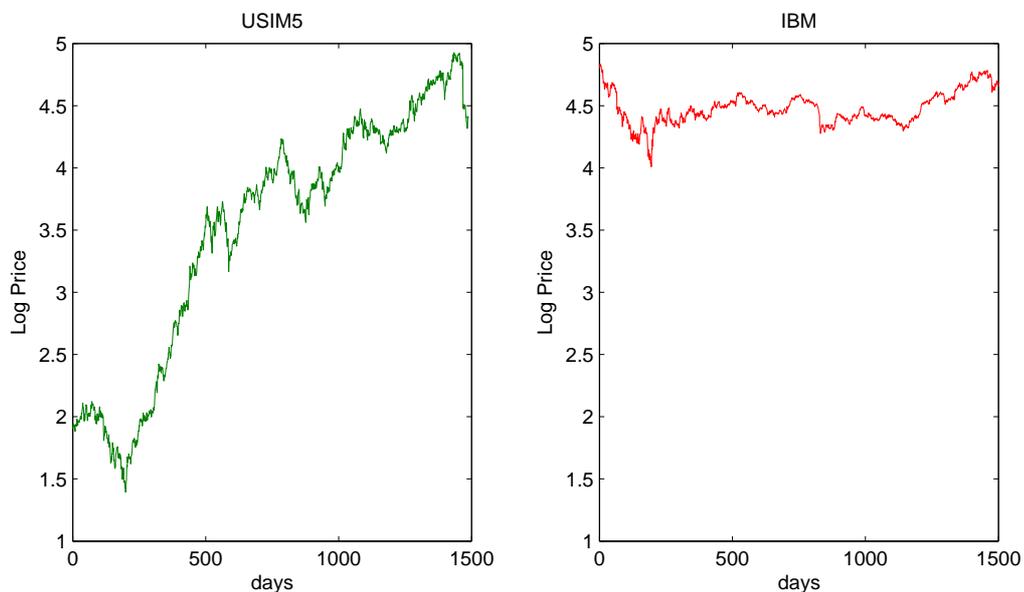
Figure 6.1: Log-price series of two stocks: USIM5 (left panel), quoted in the Brazilian market and IBM (right panel), quoted in the American market.

We also display in Fig.6.2 the permutation histogram for $n = 3$ of the analyzed data. Permutation indexes 0 and 5 exhibits higher frequencies,

with similar levels, a feature that is shared by the BM and MR processes. These results anticipate that the associated model parameters μ/σ or k of the empirical data shall be null or rather low.
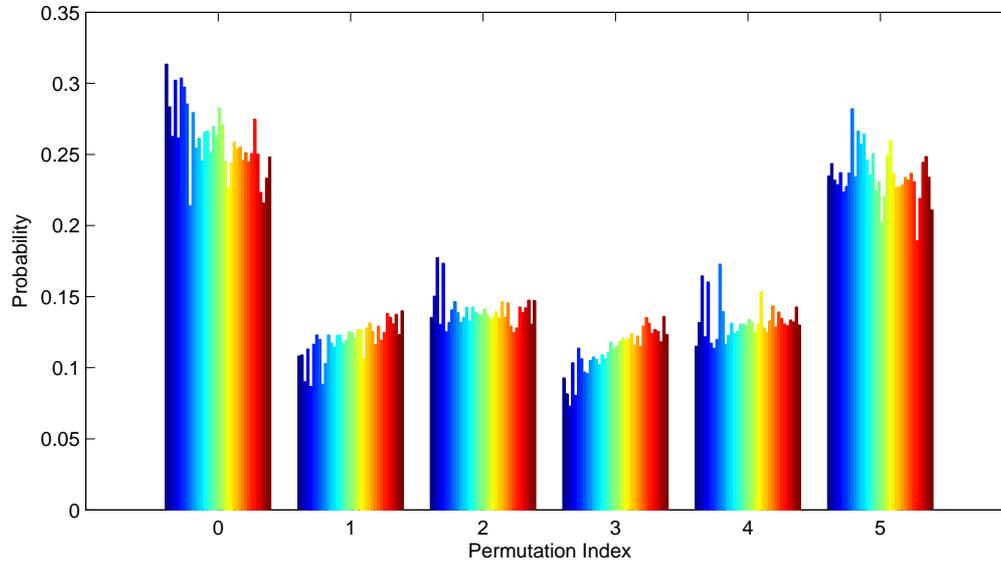
Figure 6.2: Permutation histograms of the analyzed time series. Bars are organized from left to right in ascending order of entropy.

In Table 6.1, we quote the examined stocks and show the Permutation Entropy (PE) and the Permutation Statistical Complexity (PSC) measures.

To set the precision of the threshold values H_{BM} and C_{BM} for finite samples, we consider 100 synthesized series with length 2^{10} from Eq. (5-2). For strings of length $n = 3$, we got $H_{BM} = 0.9656 \pm 0.0014$ and $C_{BM} = 0.0311 \pm 0.0014$, within the 95% confidence interval. It follows that one can detect the character (drifting or restoring) of the deterministic forces underlying the process with a precision at the second decimal digit in the PE measure. This fix the model indeterminacy for the present empirical analysis.

Fig. 6.3 shows the representation of the empirical data in the *complexity-entropy plane* contrasted with the Drift and MR lines for permutation order $n = 3$ (25).

We found that most American stock values fall in the MR domain while the Brazilian stocks fall in the Drift domain. This means that Brazilian stocks are more subjected to local trends than the American counterparts. These results are according previous findings for developed and emerging market indexes (27). However, the Brazilian series displays more complexity than the theoretical Drift-like processes, which means that there is a residual amount of memory structure at the level of this class approximation.

Stock	PE	PSC	Model	Stock	PE	PSC	Model
VALE5(M)	0.9403	0.0538	Drift	AA(M)	0.9585	0.0384	Drift
LAME4(G)	0.9486	0.0480	Drift	INTC(T)	0.9652	0.0325	BM
RAPT4(G)	0.9499	0.0489	Drift	CVX(O)	0.9682	0.0292	MR
USIM5(M)	0.9503	0.0451	Drift	MCD(G)	0.9685	0.0293	MR
POMO4(M)	0.9512	0.0473	Drift	IBM(T)	0.9692	0.0288	MR
CSNA3(M)	0.9522	0.0432	Drift	GE(G)	0.9712	0.0269	MR
PETR4(O)	0.9533	0.0425	Drift	HPQ(T)	0.9718	0.0263	MR
GGBR4(M)	0.9543	0.0421	Drift	DIS(G)	0.9725	0.0258	MR
BOBR4(G)	0.9550	0.0427	Drift	JPM(G)	0.9726	0.0259	MR
ITAU4(B)	0.9576	0.0393	Drift	MMM(G)	0.9730	0.0253	MR
ELET3(E)	0.9578	0.0390	Drift	JNJ(G)	0.9733	0.0251	MR
PCAR4(G)	0.9611	0.0363	Drift	KO(G)	0.9736	0.0248	MR
BBDC4(B)	0.9617	0.0355	Drift	XOM(O)	0.9747	0.0233	MR
BBAS3(B)	0.9643	0.0333	BM	PG(G)	0.9771	0.0215	MR
AMBV4(G)	0.9676	0.0301	BM	AXP(B)	0.9779	0.0209	MR
CMIG3(E)	0.9680	0.0298	BM	WMT(G)	0.9779	0.0209	MR
EMBR3(G)	0.9716	0.0270	MR	MSFT(T)	0.9782	0.0207	MR
TRPL4(E)	0.9724	0.0259	MR	BAC(B)	0.9795	0.0193	MR

Table 6.1: Permutation entropy (PE) and statistical complexity (PSC) for daily log-price series of Brazilian and American stocks (economic sectors are represented in parentheses). Columns 4 and 8 indicates the model domain. The standard Brownian Motion (BM) is ascribed for the entropy range H $\in [0.9642, 0.9670]$.

Moreover, it was observed that stocks belonging to the same economic sector and country differ little from each other. For instance, stocks from the Brazilian Banking sector (ITAU4, BBDC4 and BBAS3) showed PE values in the range $[0.9576, 0.9643]$, while stocks JPM, AXP and BAC, which belong to the same sector but are traded within the American market, presented entropy in $[0.9726, 0.9795]$.

At last, from the entropy values of the empirical data shown in Table 6.1, we calculated the model parameters of the respective domain according to the expressions obtained in the last chapter. As expected, both Drift and MR parameters are rather low:

$$0 < |\mu/\sigma| < 0.22 \tag{6-1}$$

and

$$0 < k < 0.23 \tag{6-2}$$
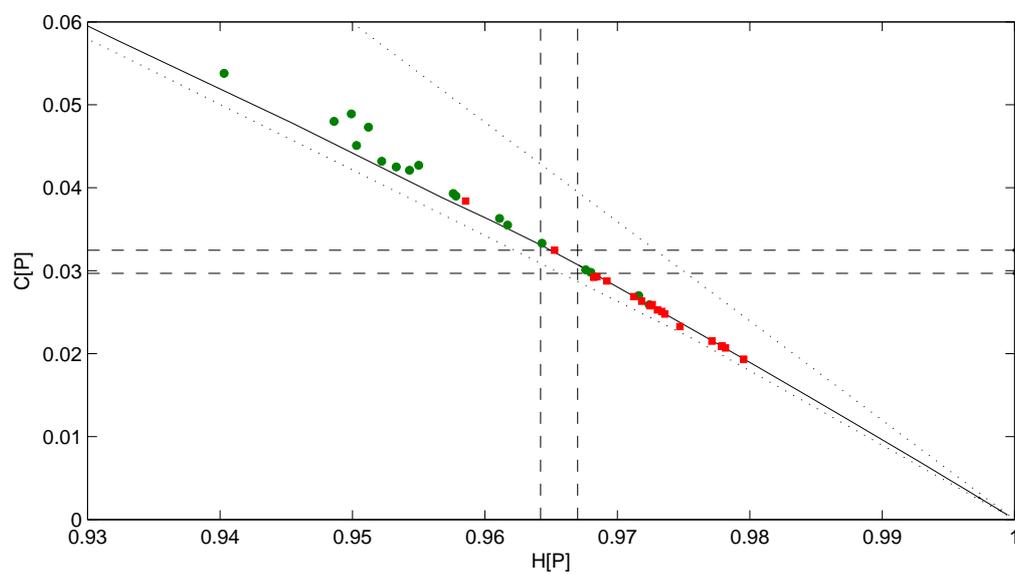
Figure 6.3: *Complexity-entropy plane* for Brazilian (green circles) and American (red squares) stocks for permutation order $n = 3$. The theoretical Drift and MR lines are also shown (solid lines), and the bounds of complexity are indicated by the dotted curves. The error bars of the threshold values H_{BM} and C_{BM} are represented by vertical and horizontal dashed lines and set the model indeterminacy.

7
Conclusion

The objective of this work is develop an information-based tool that enables one to discriminate drifting and reversion behaviors in real series.

As control data sets, we have analyzed the local time behavior of representative stochastic processes characterized by drifting or mean reverting deterministic forces. We computed the frequency of occurrence of synthesized n-data strings according to the relative values of the neighboring variables. Each configuration was associated to an ordinal pattern Π of order n which were identified as "generalized states" carrying information about the character of the deterministic forces. We considered pattern histograms over many time scales n to account for the multi-scale structures embedded in time series. Then, we evaluated the permutation entropy as well as the statistical complexity measures for this pattern histograms and determined the relevant parameters for such measures.

For the class of drift-processes, we found a striking result: notwithstanding the fact that the functions $H[P]$ are very similar, the complexity $C[P]$ depend crucially on the permutation order under consideration, as evidenced in Fig. 5.7. These results support that with increasing n, new informational structures are captured by $C[P]$, but not by $H[P]$.

As a corollary of the strong n-dependence of $C[P]$, similar systems show different signatures of its own complexity dependency on n giving rise to crossovers, as illustrated in Fig. 5.9. Therefore, one particular time-series may look more complex than others in short intervals n whereas the actual full signal is not. This result warns that empirical analysis based on low-order complexity measure should be taken with caution.

However, notice that due to the finite size of real series, large permutation order may lead to non-significant statistical results. The series investigated here are related to drift-class relevant parameter values given by Eq. (6-1). According to Fig. 5.9, these systems do not exhibit crossovers within empirically accessible time scales (which generates statistically meaningful results). This validates the present analysis retrieved from the $n = 3$ results.

We attained the permutation entropy for Drift and MR processes in the

full range of their respective relevant parameters. We found that, no matter the parameter values, for each order n, the class of drift processes furnish entropy values bigger than the class of mean reverting process. This sets an entropy threshold value which impart to PE the power of discriminating real data according to the nature (drift or reverting) of the deterministic forces underlying complexity.

This work is also an attempt to characterize financial data by means of PE and PSC measures. Like healthy physiological systems, a " healthy " market needs to exhibit processes that runs over several different time scales, to be able to adapt to an ever changing economic environment. From the analysis based on these parsimonious number of classes of models, the dominance of single characteristic time scales like in the MR processes causes a complexity loss compared to the invariant scale MB process. In addition, the appearance local trends is an indication of complexity gain, as suggested by the results for Drift processes. In this sense, complexity is a sign of forecast opportunities in short time horizons.

As a first application, we scrutinized log-price series of American and Brazilian stocks. Comparing the empirical values of permutation entropy with those of the studied models, we found that American stocks are less subjected to local trends than Brazilian stocks. In fact, financial processes become more random or unpredictable and loses its structure and correlation due to the enlarged and efficient economic activities in developed markets.

However, mapping the empirical data onto the *complexity-entropy plane*, we found that, while the American stocks are very close to the MR domain, Brazilian stocks are more complex than its Drift-class counterpart. Such result suggests further development of this methodology for the analysis of subtle differences between real paths and general stochastic models. In principle, the efficiency of the theoretical tools devised to estimate financial risks depends on the adequacy of the stochastic modeling of the market fluctuations. Here, we have shown that PE and PSC measures can unravel the nature of the coherent market forces, without reference to a specific model.

Bibliography

[1] FELDMAN, D. P.; CRUTCHFIELD, J. P. Measures of Statistical Complexity: Why? **Phys. Lett. A**, v.238, p.244-252, feb/1998.

[2] CRUTCHFIELD, J. P.; YOUNG, K. Inferring Statistical Complexity.**Phys. Rev. Lett.**, v.63, n.2, p.105, jul/1989.

[3] COSTA, M.; GOLDEBERGER, A. L.; PENG, C.-K. Multiscale Entropy Analysis of Complex Physiologic Time Series.**Phys. Rev. Lett.**, v.89, n.6, aug/2002.

[4] BANDT, C.; POMPE, B. Permutation Entropy: A Natural Complexity Measure for Time Series. **Phys. Rev. Lett.**, v.88, n.17, apr/2002.

[5] SHINER, J. S.; DAVINSON, M.; LANDSBERG, P. T. Simple measure for complexity. **Phys. Rev. E**, v.59, n.2, p.1459, feb/1999.

[6] MARTIN, M. T.; PLASTINO, A.; ROSSO, O. A. Generalized statistical complexity measures: Geometrical and analytical properties. **Physica A**, v.369, p.439462, feb/2006.

[7] ROSSO, O. A. et. al. Generalized Statistical Complexity Measure. **International Journal of Bifurcation and Chaos**, v.20, n.3, p.775-785, 2010.

[8] ROSSO, O. A. et al. Distinguishing Noise from Chaos. **Phys. Rev. Lett.**, v.99, oct/2007.

[9] RANJIT, A. T.; GOTTWALD, G. A. On multiscale entropy analysis for physiological data. **Physica A**, v.366, p.323-332, jul/2006.

[10] BIAN, C. et al. Modified permutation-entropy analysis of heartbeat dynamics. **Phys. Rev. E**, v.85, feb/2012.

[11] ZHONGWEI, L.; ZHANG, Y-K. Multi-scale entropy analysis of Mississippi River flow. **Stoch Environ Res Risk Assess**, v.22, p.507-512, jun/2008.

[12] KAPUR, J. N.; KESAVAN, H. K. **Entropy optimization principles with applications**. Boston: Academic Press, 1992.

[13] SHANNON, C. E. A Mathematical Theory of Communication. **The Bell System Technical Journal**, v.27, p.379-423/623-656, oct/1948.

[14] WOOTERS, W. K. Statistical distance and Hilbert space. **Phys. Rev. D**, v.23, n.2, jan/1981.

[15] BRIET, J.; HARREMOES, P. Properties of classical and quantum Jensen-Shannon divergence. **Phys. Rev. A**, v.79, may/2009.

[16] ASH, R. B. **Basic Probability Theory**. New York: Wiley, 1970.

[17] LIN, J. Divergence Measures based on Shannon Entropy. **IEEE Trans. Inf. Theo.**, v.37, n.1, jan/1991.

[18] CAMPBELL, W. H. Indexing Permutations. **Journal of Computing Sciences in Colleges**, v.19, p.296-300, jan/2004.

[19] J. M. AMIGÓ. **Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That**, Berlin: Springer, 2010.

[20] M. ZANIN et al. Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. **Entropy**, v.14, p.1553-1577, aug/2012.

[21] MIKOSCH, T. **Elementary stochastic calculus with finance in view.** Singapore: World Scientific Publishing Co. Pte Ltd., 1998.

[22] L'ECUYER, P.; SIMARD, R. TestU01: A C Library for Empirical Testing of Random Number Generators. **ACM Transactions on Mathematical Software**, v.33, n.4, 2007.

[23] FOG, A. **Pseudo random number generators**. Available at <http://www.agner.org/random>. Accessed on 05/06/2012.

[24] ZUNINO, L.; SORIANO, M.C.; ROSSO, O. A. Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. **Phys. Rev. E**, v.86, oct/2012.

[25] RIBEIRO, A. S.; FREIRE, R. R. An Information-based tool for inferring the nature of deterministic sources in real data. **Physica A**, Submitted on jan/2013.

[26] **Financial Times Global 500**. Available at
<http://media.ft.com/cms/813c979e-0faa-11dd-8871-
0000779fd2ac.pdf>. Accessed on 08/09/2012.

[27] ZUNINO, L. et al. Complexity-entropy causality plane: A useful approach
to quantify the stock market inefficiency. **Physica A**, v.389, p.1891-1901,
jan/2010.

A
Gaussian Distribution through Principle of Maximum Entropy

Consider that besides the normalization constraint, only information available is that a real random variable X has mean μ and variance σ^2. That is

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \tag{A-1}$$

$$\int_{-\infty}^{+\infty} x^2 f(x)dx = \sigma^2 \tag{A-2}$$

So, for obtaining the probability distribution $f(x)$ which encompasses all the given knowledge of X one should maximize the Shannon's entropy subject to the above constraints:

$$L = \int_{-\infty}^{+\infty} -f(x)\log f(x)dx - \lambda_0 \left(\int_{-\infty}^{+\infty} f(x)dx - 1 \right) - \lambda_1 \left(\int_{-\infty}^{+\infty} x^2 f(x)dx - \sigma^2 \right)$$

$$\delta L = 0 \Rightarrow \log f(x) = -\lambda_0 - 1 - \lambda_1 x^2$$
$$\Rightarrow f(x) = e^{-(\lambda_0+1)}e^{-\lambda_1 x^2} \tag{A-3}$$

Using constraints

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Rightarrow e^{\lambda_0} = \int_{-\infty}^{+\infty} e^{-\lambda_1 x^2} dx \rightarrow e^{\lambda_0} = \sqrt{\frac{\pi}{\lambda_1}} \tag{A-4}$$

$$\int_{-\infty}^{+\infty} x^2 f(x)dx = \sigma^2 \Rightarrow \int_{-\infty}^{+\infty} x^2 e^{-\lambda_1 x^2} dx = \sqrt{\frac{\pi}{\lambda_1}}\sigma^2 \rightarrow \lambda_1 = \frac{1}{2\sigma^2} \tag{A-5}$$

Thus,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}$$

which is the Gaussian probability disribution with zero mean.

B
Analytical determination of ABM permutation probabilities

The permutation probability $P(\pi_i)$ of order $n = 3$ for Drift class of models can be analytically calculated. First note that, from Table 4.1, by changing the signal of μ, the role of the pairs of permutation indexes $0 - 5$, $1 - 4$ and $2 - 3$ interchange. Thus, due to this symmetry, the four curves shown in the left panel of Fig. 5.3 in the range $0 \leq \mu/\sigma$ collapse onto four curves in the full range of the relevant parameter $-\infty \leq \mu/\sigma \leq \infty$. The following figure reproduces this result.
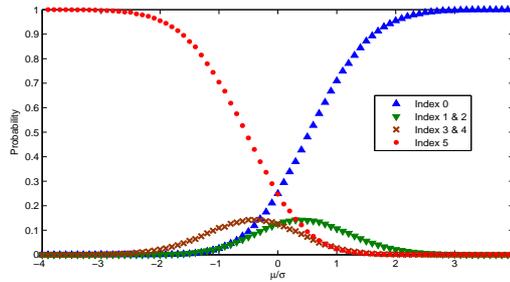

Figure B.1: $P(\pi_i), 1 = 0, ...5$ of order $n = 3$ for Drift model in the full range of μ/σ . Note that some indexes collapse onto the same curve.

In order to derive the analytical expressions for $P(\pi_i)$ according to μ, we previously introduce the definition of the Error function $erf(\mathrm{x}) = (2/\sqrt{\pi}) \int_0^x exp(-t^2)dt$ and the Complementary Error Function $erfc(x) = 1 - erf(x)$. The Normal Cumulative Distribution Function $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x exp(-t^2/2)dt$ is essentially identical to the Complementary Error Function: $\Phi(x) = (1/2)erfc(-x/\sqrt{2})$.

In what follows, we consider Drift-parameter $\sigma = 1$ and general μ. From Eq. (5-2), steps δx are independent and obey the Gaussian distribution $P(\delta x) = (1/\sqrt{2\pi})exp(-(\delta x - \mu)^2/2)$. According to Table 4.1, permutation index 0 comprises consecutive positive steps with arbitrary sizes, then,

$$P_\mu(\pi_0) = [\frac{1}{\sqrt{2\pi}} \int_0^\infty exp[-(\delta x - \mu)^2/2]d(\delta x)]^2 \qquad (A.1)$$

which leads to

$$P_\mu(\pi_0) = \frac{1}{4}[erfc(-\mu/\sqrt{2})]^2 \tag{A.2}$$

Now we proceed to calculate $P(\pi_1)$. According to Table 4.1, this index corresponds to an arbitrary negative step δx followed by a positive one of larger size. Then, given the first step, the occurrence of second step is conditional:

$$P_\mu(\delta x_2|\delta x_1) = \frac{1}{\sqrt{2\pi}} \int_{-\delta x_1}^{\infty} exp[-(\delta x_2 - \mu)^2/2]d(\delta x_2) \tag{A.3}$$

or

$$P_\mu(\delta x_2|\delta x_1) = \frac{1}{2}[1 + erf(\delta x_1 + \mu)/\sqrt{2})] \tag{A.4}$$

Therefore,

$$P_\mu(\pi_1) = \int_{-\infty}^{0} P(\delta x_1)P(\delta x_2|\delta x_1)d(\delta x_1) \tag{A.5}$$

or

$$P_\mu(\pi_1) = \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{0} P(\delta x_1)[1 + erf(\delta x_1 + \mu)/\sqrt{2})]d(\delta x_1) \tag{A.6}$$

The probabilities for the other indexes are deducible along the same lines. One can easily show that $P_\mu(\pi_5) = P_{-\mu}(\pi_0)$ and $P_\mu(\pi_4) = P_{-\mu}(\pi_1)$. The equivalence $P_\mu(\pi_1) = P_\mu(\pi_2)$ and $P_\mu(\pi_3) = P_\mu(\pi_4)$ due to independence of step order completes the analysis. One can also settle the constraints:

$$P_\mu(\pi_1) + P_\mu(\pi_3) = P_\mu(\pi_2) + P_\mu(\pi_4) = \frac{1}{4}[erfc(+\mu/\sqrt{2})erfc(-\mu/\sqrt{2})] \tag{A.7}$$