Entropy of Difference

Pasquale Nardone*

Physics Department, Université Libre de Bruxelles and

50 av F. D. Roosevelt, 1050 Bruxelles, Belgium

(Dated: November 5, 2014)

Here, we propose a new tool to estimate the complexity of a time series: the entropy of difference (ED). The method is based solely on the sign of the difference between neighboring values in a time series. This makes it possible to describe the signal as efficiently as prior proposed parameters such as permutation entropy (PE) or modified permutation entropy (mPE), but (1) reduces the size of the sample that is necessary to estimate the parameter value, and (2) enables the use of the Kullback-Leibler divergence to estimate the distance between the time series data and random signals.

I. INTRODUCTION

Permutation entropy (PE), introduced by Bandt and Pompe[1], as well as its modified version[2], are both efficient tools to measure the complexity of chaotic time series. Both methods propose to analyze time series: $X = \{x_1, x_2, \cdots x_k \cdots\}$ by first choosing an embedding dimension m to split the original data in a subset of m-tuples: $\{\{x_1, x_2 \cdots x_m\}, \{x_2, x_3, \cdots x_{1+m}\}, \cdots\}$, then to substitute to the m-tuples values by the rank of the values, resulting in a new symbolic representation of the time series. For example, consider the time series $X = \{0.2, 0.1, 0.6, 0.4, 0.1, 0.2, 0.4, 0.8, 0.5, 1., 0.3, 0.1, \cdots\}$. Choosing, for example, an embedding dimension $m = 4$, will split the data in a set of 4-tuples: $X_4 = \{\{0.2, 0.1, 0.6, 0.4\}, \{0.1, 0.6, 0.4, 0.1\}, \{0.6, 0.4, 0.1, 0.2\}, \cdots\}$. The Bandt-Pompe method will associate the rank of the value with each 4-tuples. Thus, in $\{0.2, 0.1, 0.6, 0.4\}$ the lowest element 0.1 is in position 2, the second element 0.2 is in position 1, 0.4 is in position 4 and finally 0.6 is in position 3. Thus the 4-tuple $\{0.2, 0.1, 0.6, 0.4\}$

*Electronic address: pnardon@ulb.ac.be

is rewritten as $\{2,1,4,3\}$. This procedure thus results in each X_4 to be rewritten as a symbolic list:$\{\{2,1,4,3\},\{1,4,3,2\},\{3,4,2,1\}\cdots\}$. Each element is then a permutation π of the set $\{1,2,3,4\}$. Next, the probability of each permutation π in X_m is then computed: $p_m(\pi)$, and finally the PE for the embedding dimension m, is defined as $\mathrm{PE}_m(X) = -\sum_\pi p_m(\pi)\log(p_m(\pi))$. The modified permutation entropy (mPE) just deals with those cases in which equal quantities may appear in the m-tuples. For example for the m-tuple $\{0.1,0.6,0.4,0.1\}$, computing PE will produce $\{1,4,3,2\}$ while computing mPE will associate $\{1,1,3,2\}$[14]. Both methods are widely used due to their conceptual and computational simplicity[3, 4, 9–12]. For random signals, PE leads to a constant probability $q_m(\pi) = 1/m!$, which does not make it possible to evaluate the "distance" between the probability found in the signal: $p_m(\pi)$ and the probability produced by a random signal: q_m, with the Kullback-Leibler (KL) divergence[6, 8]: $\mathtt{KL}_m(p\|q) = \sum_\pi p_m(\pi)\log_2(p_m(\pi)/q_m(\pi))$. Furthermore, the number $\mathtt{K}_m$ of m-tuples are $m!$ for PE and even greater for mPE[2], thus requiring then a large data sample to perform significant statistical estimation of p_m.

II. ENTROPY OF DIFFERENCE-METHOD

The entropy of difference (ED) method proposes to substitute to the m-tuples with strings s containing the sign ("+" or "-"), representing of the difference between subsequent elements in the m-tuples. For the same X_4: $\{\{0.2,0.1,0.6,0.4\},\{0.1,0.6,0.4,0.1\},\{0.6,0.4,0.1,0.2\},\cdots\}$ this leads to the representation : $\{``-+-", ``+--", ``--+", \cdots\}$. For an m value, we have 2^{m-1} strings from "$++ + \cdots +$" to "$- - - \cdots -$". Again we compute, in the time series, the probability distribution $p_m(s)$ of these strings s and define the entropy of difference of order m as : $\mathrm{ED}_m = -\sum_s p_m(s)\log p_m(s)$. The number of elements: $\mathtt{K}_m$ to be treated, for an embedding m, are smaller for ED compared with the number of permutations π in PE or to the elements in mPE (see table I).

Furthermore the probability distribution for a string s, in a random signal : $q_m(s)$ is not constant and could be computed through the recursive equation[15] (in the following equations $\mathtt{x}$ and $\mathtt{y}$ are strings):

$$q(+) = q(-) = \frac{1}{2}$$

TABLE I: $\mathtt{K}$ values, for different m-embedding

m	3	4	5	6	7
$\mathtt{K}_{PE}$	6	24	120	720	5040
$\mathtt{K}_{mPE}$	13	73	501	4051	37633
$\mathtt{K}_{ED}$	4	8	16	32	64

$$
\begin{aligned}
q(\underbrace{+,+,+,\cdots,+}_{m}) &= \frac{1}{(m+1)!} \\
q(-,\mathtt{x}) &= q(\mathtt{x}) - q(+,\mathtt{x}) \\
q(\mathtt{x},-) &= q(\mathtt{x}) - q(\mathtt{x},+) \\
q(\mathtt{x},-,\mathtt{y}) &= q(\mathtt{x})q(\mathtt{y}) - q(\mathtt{x},+,\mathtt{y})
\end{aligned}
\tag{1}
$$

leading to a complex probability distribution. For example for $m = 9$ we have $2^8 = 256$ strings with the highest probability for the "$+-+-+-+-$" string (and its symmetric "$-+-+-+-+$"): $q_9(\mathtt{max}) = \frac{62}{2835} \approx 0.02187$ (see Fig. I). These probabilities $q_m(s)$ could then be used to determine the KL-divergence between the time series probability $p_m(s)$ and the random signal.

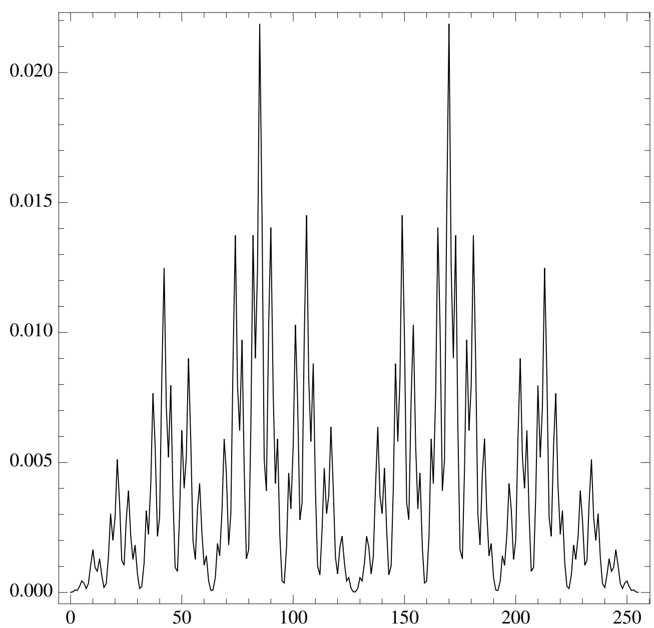

FIG. 1: The 2^8 values for the probability of $q_9(s)$, from $s = ---... \equiv 0$ to $s = +++... \equiv 255$

Despite the complexity of $q_m(s)$, the Shannon entropy for a random signal : $-\sum_s q_m(s) \log_2 q_m(s)$ increases linearly with m, with a slope ≈ 0.905.

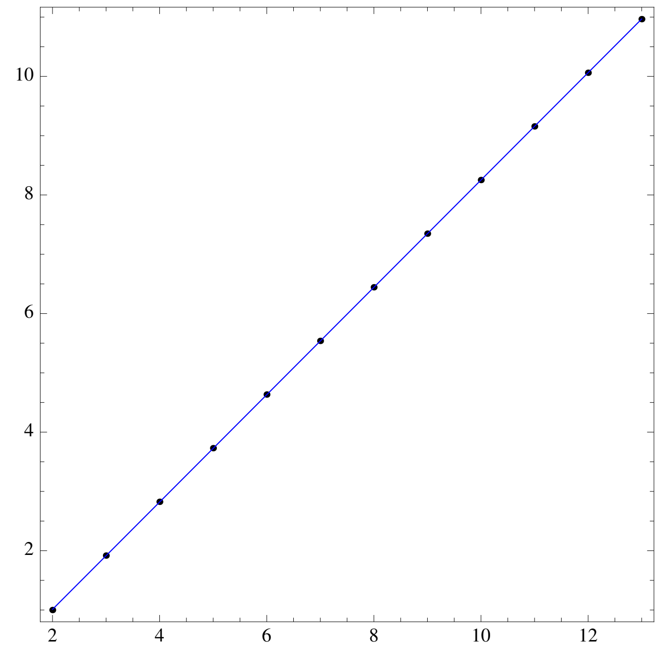

FIG. 2: The Shannon entropy of $q_m(s)$ increases linearly with m, the fit $-0.799574 + 0.905206\, m$ gives a sum of squared residuals of 1.7 10^{-4} and a p-value=1.57 10^{-12} and 1.62 10^{-30} on the fit parameter respectively.

III. CHAOTIC LOGISTIC MAP EXAMPLE

Let us illustrate the use of ED on the well know logistic map[7] $\texttt{Lo}(x, \lambda)$ driven by the parameter λ.

$$x_{n+1} = \texttt{Lo}(x_n, \lambda) = \lambda x_n(1 - x_n) \tag{2}$$

It is obvious that for a range of values of λ where the time series reaches a periodic behavior (any cyclic oscillation between n different values), the ED will remain constant. The evaluation of the ED could thus be used as a new complexity parameter to determine the behavior of the time series (see FIG. 3).

For $\lambda = 4$ we know that the data are randomly distributed with a probability density given by[5]

$$p_{\texttt{Lo}}(x) = \frac{1}{\pi\sqrt{(1-x)x}} \tag{3}$$

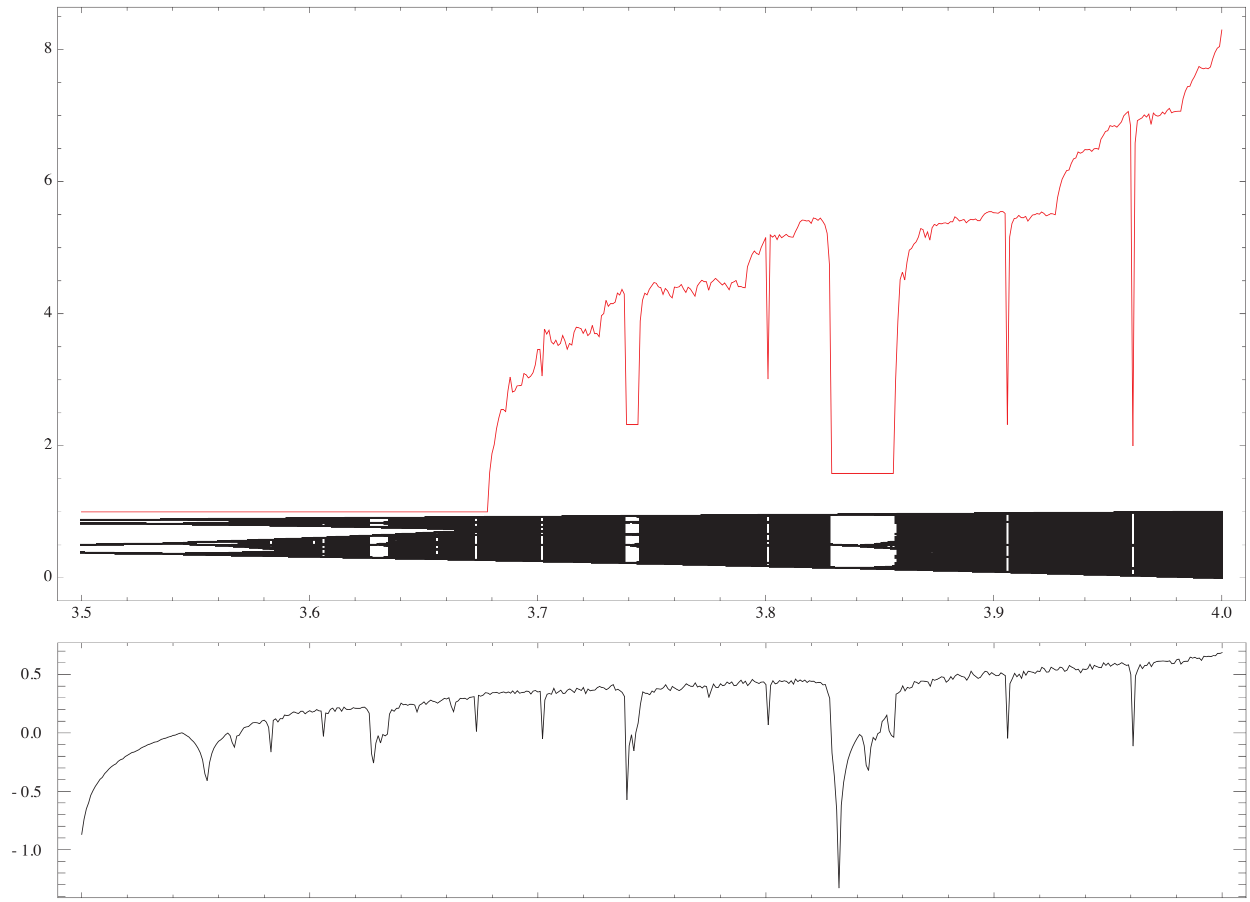

FIG. 3: The $\texttt{ED}_{13}$ (strings of length 12) is plotted versus λ, with the bifurcation diagram, and the value of the Lyapunov exponent respectively. The constant value appears when the logistic map enter into a periodic regime.

We can then compute exactly the $\texttt{ED}$ for an m-embedding, and the $\texttt{KL}$-divergence from a random signal. For example, for $m = 2$, we can determine the p_+ and p_- by solving the inequality $x < \texttt{Lo}(x)$ and $x > \texttt{Lo}(x)$ respectively which implies that $0 < x < 3/4$ and $3/4 < x < 1$, and then

$$p_+ = \int_0^{3/4} dx\; p_{\texttt{Lo}}(x) = \frac{2}{3} \quad p_- = \int_{3/4}^{1} dx\; p_{\texttt{Lo}}(x) = \frac{1}{3} \tag{4}$$

In this case the logistic map produces a signal that contains twice as many increasing pairs

" + " than decreasing pairs " − ". So:

$$\mathtt{ED}_2 = -(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}) = \log_2\frac{3}{2^{2/3}} \approx 0.918 \quad \mathtt{KL}_2 = \frac{1}{3}\log_2\frac{32}{27} \approx 0.082 \tag{5}$$

For $m = 3$ and $m = 4$ we can perform the same calculation:

$$p_3(++) = \frac{1}{3} \quad p_3(+-) = \frac{1}{3} \quad p_3(-+) = \frac{1}{3} \tag{6}$$
$$\rightarrow \mathtt{ED}_3 = \log_2 3 \approx 1.58 \quad \mathrm{KL}_3 = \frac{1}{3} \approx 0.33$$

Effectively the logistic map with $\lambda = 4$ forbids the string "- -" where $x_1 > x_2 > x_3$. For strings of length 3 we also have also the non zero values:

$$p_4(+++) = p_4(++-) = p_4(-++) = p_4(-+-) = \frac{1}{6} \quad p_4(+-+) = \frac{2}{6}$$
$$\rightarrow \mathtt{ED}_4 = \log_2 108^{\frac{1}{3}} \approx 2.25 \quad \mathtt{KL}_4 = \log_2\left(\frac{16384}{1125}\right)^{1/6} \approx 0.64 \tag{7}$$

The probability of difference $p_m(s)$ for some string length m versus s the string binary value, where "+"$\rightarrow$ 1 and "-"$\rightarrow$ 0, give us the "spectrum of difference" for the distribution p (see FIG. 4).

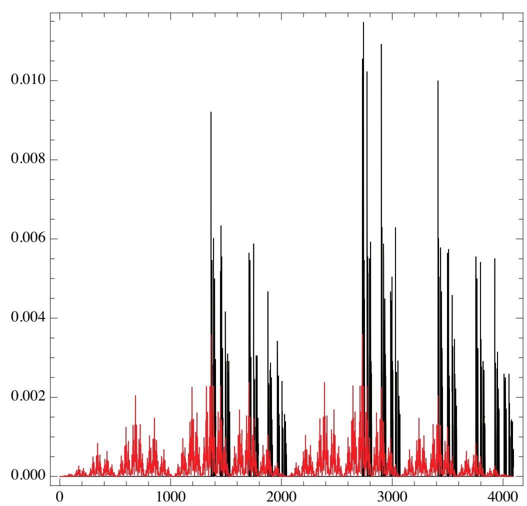

FIG. 4: The spectrum of p_{13} versus the string binary value (from 0 to $2^{12} - 1$) for the logistic map at $\lambda = 4$ and the one from a random distribution q_{13}

IV. $\mathtt{KL}_m(p|q)$ DIVERGENCES VERSUS m ON REAL DATA AND ON MAPS

The manner in which the $\mathtt{KL}_m(p|q)$ evolves with m is another parameter of the complexity measure. $\mathtt{KL}_m(p|q)$ measures the loss of informations when the random distribution q_m is used to predict the distribution p_m. Increasing m introduces more bits information in the signal and the behavior versus m shows how the data diverges from a random distribution.

The graphics (see FIG. 5) shows the behavior of $\mathtt{KL}_m$ versus m for two different chaotic maps and for real financial data[13] : the opening value of the `nasdaq100`, `bel20` everyday from 2000 to 2013. For maps, the logarithmic map $x_{n+1} = \ln(a|x_n|)$ and logistic map are shown.

For maps the simulation starts with a random number between 0 and 1, then first iterate 500 times to avoid transients. Starting with that seeds, 720 iterates where kept on which the $\mathtt{KL}_m$ where computed. It can be seen that the Kullback-Leibler divergence from the logistic map at $\lambda = 4$ to the random signal is fitted by a quadratic function of m: $\mathtt{KL}_m = -0.4260 + 0.2326\ m + 0.0095\ m^2$ (p-value$\approx 2\ 10^{-7}$ for all the parameter), while the logarithmic map behavior is linear in the range $a \in [0.4, 2.2]$. Financial data are also quadratic $\mathtt{KL}_m(\mathtt{nasdaq}) = 0.1824 - 0.0973\ m + 0.0178\ m^2$, $\mathtt{KL}_m(\mathtt{bel20}) = 0.1587 - 0.0886\ m + 0.0182\ m^2$ with a higher curvature than the logistic map due to the fact that the spectrum of the probability p_m is compatible with a constant distribution (see FIG. 6) rendering the prediction of increase or decrease signal completely random, which is not the case in any true random signal.

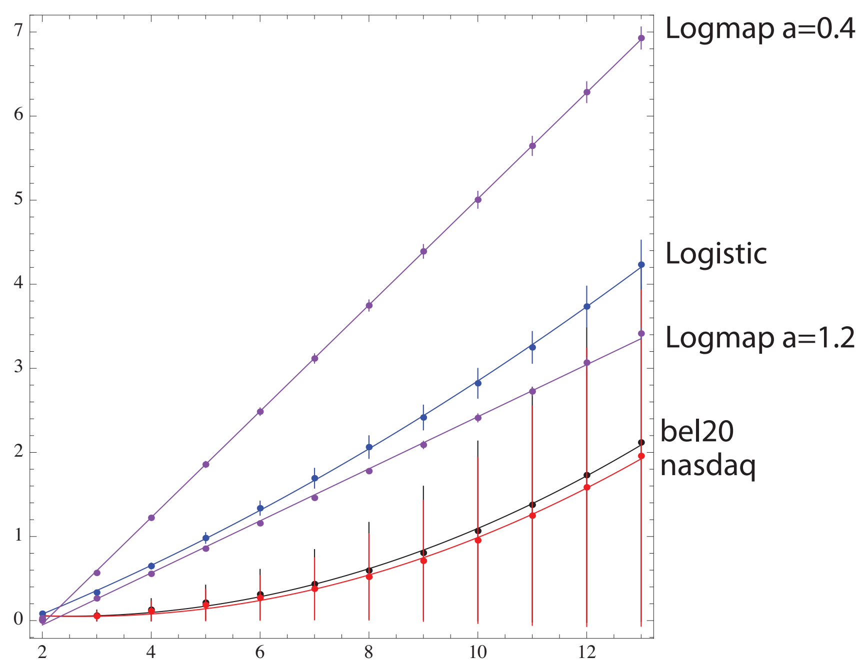

FIG. 5: The KL-divergence for the data

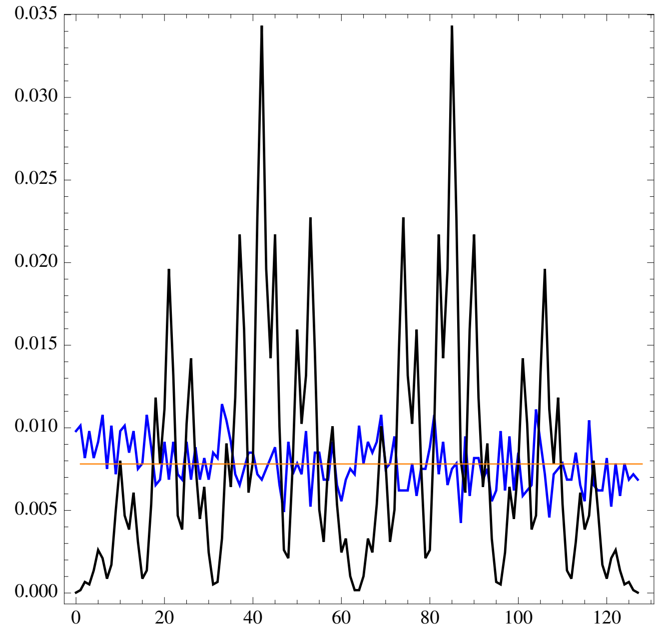

FIG. 6: The spectrum of p_8 versus the string binary value (from 0 to $2^7 - 1$) for the bel20 financial data

V. CONCLUSIONS

The simple property of increases or decreases in a signal makes it possible to introduce the entropy of difference $\mathtt{ED}_m$ as a new efficient complexity measure for chaotic time series. The probability distribution of string q_m for random signal is used to evaluate the Kullback-Leibler divergence versus the number of data m used to build the difference string. This $\mathtt{KL}_m$ shows different behavior for different types of signal and can also be used also to characterize the complexity of a time series.

Appendix: 1

The Mathematica program for m-embding, PE and mPE are simple:

```
mEmbedding[Xlist_,m_]:=Partition[Xlist,m,1];
PE[mList_]:=Ordering[mList];
mPE[mList_]:=Flatten[Map[First[Position[mList, #]] &, Sort[mList]]];
```

Appendix: 2

The Mathematica program for the probability $q(s)$:

```
P["+"]= P["-"] = 1/2;
P["-", x__] := P[x] - P["+", x];
P[x__, "-"] := P[x] - P[x, "+"];
P[x__, "-", y__] := P[x] P[y] - P[x, "+", y];
P[x__] :=1/(StringLength[StringJoin[x]] + 1)!
```


[1] Christoph Bandt and Bernd Pompe. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102, Apr 2002.

[2] Chunhua Bian, Chang Qin, Qianli D. Y. Ma, and Qinghong Shen. Modified permutation-entropy analysis of heartbeat dynamics. *Phys. Rev. E*, 85:021906, Feb 2012.

[3] U. Schneider B. Frank, B. Pompe and D. Hoyer. Permutation entropy improves fetal behavioural state classification based on heart rate analysis from biomagnetic recordings in near term fetuses. *Medical and Biological Engineering and Computing*, 44:179–187, March 2006.

[4] J. W. Sleigh E. Olofsen and A. Dahan. Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *Br. J. Anaesth.*, 101:810–821, December 2008.

[5] M. V. Jakobson. Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Communications in Mathematical Physics*, 81:39–88, 1981.

[6] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22:79–86, March 1951.

[7] Robert M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467, 1976.

[8] Édgar Roldán and Juan M. R. Parrondo. Entropy production and kullback-leibler divergence between stationary trajectories of discrete systems. *Phys. Rev. E*, 85:031129, Mar 2012.

[9] O. A. Rosso, L. Zunino, D. G. Pérez, A. Figliola, H. A. Larrondo, M. Garavaglia, M. T. Martí n, and A. Plastino. Extracting features of gaussian self-similar stochastic processes via the bandt-pompe approach. *Phys. Rev. E*, 76:061114, Dec 2007.

[10] S. Cui X. Li and L. J. Voss. Using permutation entropy to measure the electroencephalographic effects of sevoflurane. *Anesthesiology*, 109:448–456, September 2008.

[11] Douglas A. Richards Xiaoli Li, Gaoxian Ouyang. Predictability analysis of absence seizures with permutation entropy. *Epilepsy research*, 77:70–74, October 2007.

[12] L. Zunino, D.G. Perez, M.T. Martin, M. Garavaglia, A. Plastino, and O.A. Rosso. Permutation entropy of fractional brownian motion and fractional gaussian noise. *Physics Letters A*, 372(27-28):4768 – 4774, 2008.

[13] data are provided by `http://www.wessa.net/`.

[14] see appendix 1

[15] see appendix 2