

PAPER • OPEN ACCESS

Sample entropy applied to the analysis of synthetic time series and tachograms

To cite this article: A Muñoz-Diosdado *et al* 2017 *J. Phys.: Conf. Ser.* **792** 012062

View the [article online](#) for updates and enhancements.

You may also like

- [An open source benchmarked toolbox for cardiovascular waveform and interval analysis](#)

Adriana N Vest, Giulia Da Poian, Qiao Li et al.

- [Correlation between RR, inter-systolic and inter-diastolic intervals and their differences for the analysis of spontaneous heart rate variability](#)

D P M Dias, L E V Silva, P L Katayama et al.

- [Separating sets of term and pre-term uterine EMG records](#)

A Smrdel and F Jager



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Sample entropy applied to the analysis of synthetic time series and tachograms

A Muñoz-Diosdado¹, G G Gálvez-Coyt and E Solís-Montufar

¹Basic Sciences Department,UPIBI, Instituto Politécnico Nacional, Mexico City, Mexico

E-mail: amunozdiosdado@gmail.com

Abstract. Entropy is a method of non-linear analysis that allows an estimate of the irregularity of a system, however, there are different types of computational entropy that were considered and tested in order to obtain one that would give an index of signals complexity taking into account the data number of the analysed time series, the computational resources demanded by the method, and the accuracy of the calculation. An algorithm for the generation of fractal time-series with a certain value of β was used for the characterization of the different entropy algorithms. We obtained a significant variation for most of the algorithms in terms of the series size, which could result counterproductive for the study of real signals of different lengths. The chosen method was sample entropy, which shows great independence of the series size. With this method, time series of heart interbeat intervals or tachograms of healthy subjects and patients with congestive heart failure were analysed. The calculation of sample entropy was carried out for 24-hour tachograms and time subseries of 6-hours for sleepiness and wakefulness. The comparison between the two populations shows a significant difference that is accentuated when the patient is sleeping.

1. Introduction

Currently, cardiovascular diseases are the leading cause of death worldwide and more people die for this reason than for any other reason [1, 2]. Factors such as obesity, the consumption of tobacco, physical inactivity, high blood pressure, hyperlipidemia, diabetes, and inadequate diet, are risk factors for cardiovascular disease, even more when these factors arise together [1]. Age is a factor that affects the heart and vessels, operating on the structures and systems of the body in general, so that it favours the emergence of cardiovascular diseases.

The use of non-linear analysis for the study of complex systems with fractal properties, allows more accurate descriptions of their behavior. The human body consists of different complex systems that have fractal and multifractal behaviours in healthy conditions [3], and it has been reported a decrease in these fractal trends when the systems are in a pathological condition [5].

Congestive Heart failure (CHF) is a syndrome in which patients typically have lack of air both at rest and during exercise; they show signs of fluid retention, and objective evidence of a structural or functional cardiac disturbance at rest [5]. The heartbeat interval time series obtained from ECG records of patients with CHF have been studied with various techniques of Nonlinear Dynamics, especially fractal and multifractal analysis [3, 4] and found significant differences between healthy and CHF subjects. Complexity of these signals seems to be different, so we propose that computational sample entropy could help to differentiate the two types of series, and thus be able to support the diagnosis of patients with heart failure.



2. Methodology

In many cases, the information obtained from the measuring devices provides temporary signals $x(i)$, where $i = 1, \dots, N$. The temporary registration of data obtained from measurements allows predicting the behaviour of the systems [6]. Time series can be characterized by four basic types of variation, which produce the changes observed in the series in a certain time interval, and give the series an erratic behaviour. These components are: secular trend, seasonal variation, cyclical variation and irregular variations, which are short term variations that are unpredictable and non-recurrent.

Frequently, it seems to be that a time series does not have useful information, but a more careful analysis can show the existence of correlations within the series, that is, the existence of a relation between the value at a moment i with the value at the moment j if $j > i$. When there is no relation between the value at the moment i with the value at the moment j ($i \neq j$) it is said that the time series is a succession of independent samples, this sort of time series is called white noise. In the case of white noise time series, a random ordering of its components would be again a white noise.

On the other hand, when the value of the variable i influences the future values $j > i$, it is said that there is dependence and therefore there is correlation. For example, a time series obtained from the accumulation of independent events of white noise is called Brownian noise; this type of noise corresponds to short range correlations [7].

We used a database of physiological signals from the web page Physionet [8] to obtain beat to beat intervals time series of 24 hours, this kind of time series is called tachogram. It was used the Normal Sinus Rhythm RR Interval Database, consisting of a population of 54 healthy individuals (30 men aged between 28 and 76) and 24 women aged between 58 and 73. It was also used two populations of CHF patients, one of them corresponding to the BIDMC Congestive Heart Failure RR Interval Database, which consists of 15 patients (11 men aged between 22 and 71) and 4 women aged between 54 and 63 classified in the NYHA scale 3-4, and a last database named Congestive Heart Failure RR Interval, with 29 patients (which includes 8 men and two women aged 34 to 79 and other 17 individuals of unknown gender) of different NYHA classifications [9]. The NYHA index was proposed by the Nueva York Heart Association as a method to classify CHF patients, the value IV corresponds to the patients in worst conditions.

The 24-hour signals were segmented into series of 6 hours for wakefulness and 6 hours for sleep intervals. Discrimination was possible thanks to the significant reduction in heart rate that individuals suffer while they are asleep, and the contrast with the beat to beat interval of the awake individual, which demands a greater flow of blood and this is reflected in the time between R and R [10] in the electrocardiogram signal.

To apply computational entropy analysis it is required to ensure that the used method provide values that do not depend on the series size, which is a problem with this type of methodologies, also is required a proper calibration of the method. Time series generated with a suitable methodology were used for this. For this reason were used time series with a certain value of β spectral power were generated, these series were obtained using the method proposed by Gálvez-Coyt et al. [7]. In this method a Gaussian white noise with N data is generated, and then the discrete transform of Fourier is applied. The result is the flat spectrum corresponding to a $\beta = 0$ spectral power, where the amplitude $|X(m)|$ is the same for the coefficients with $m = 1, 2, 3, \dots, N$.

Subsequently new coefficients of Fourier transform $X'(m)$ were built, from those obtained from the Fourier transform of the Gaussian white noise:

$$X'(m) = \left(\frac{m}{N}\right)^{-\beta/2} X(m) \quad (1)$$

With $m = 1, 2, 3, \dots, N$. Now the magnitude of the filtered data is taken $|X'(m)|$ and a phase θ_m uniform and random is generated for each one in the $[0, 2\pi]$ interval. Then we get the Fourier coefficients for the magnitudes and phases previously calculated by using:

$$X''(m) = |X'(m)|(\cos\theta_m + i\sin\theta_m), \quad (2)$$

$n = 1, 2, 3, \dots, N$. The inverse Fourier transform of $|X''(m)|$ is obtained,

$$x(n) = \frac{1}{N\delta} \sum_{m=1}^N X''(m) e^{2\pi i n m / N}, \quad (3)$$

With the purpose of obtaining a noise Gaussian with a certain value of β , the real part of the series [19] is considered. In Figures 1, 2 and 3 are shown the self-affine time series with $N = 10000$ data for β values 0, 1, and 2.

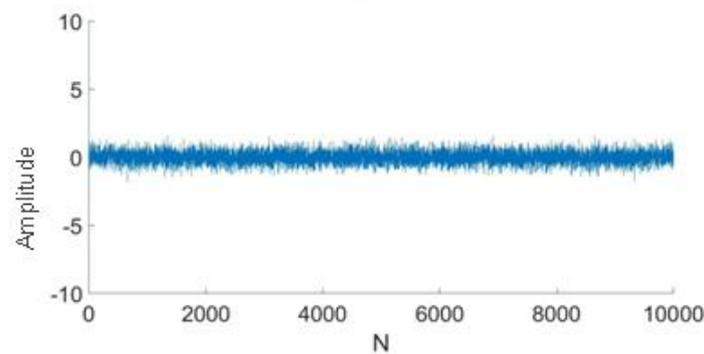


Figure 1. Graph obtained for a time series generated by the algorithm described in the text with power spectra exponent $\beta = 0$.

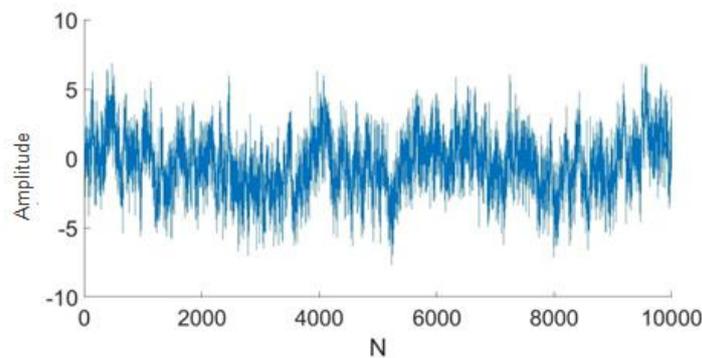


Figure 2. Graph obtained for a time series with power spectra exponent $\beta = 1$.

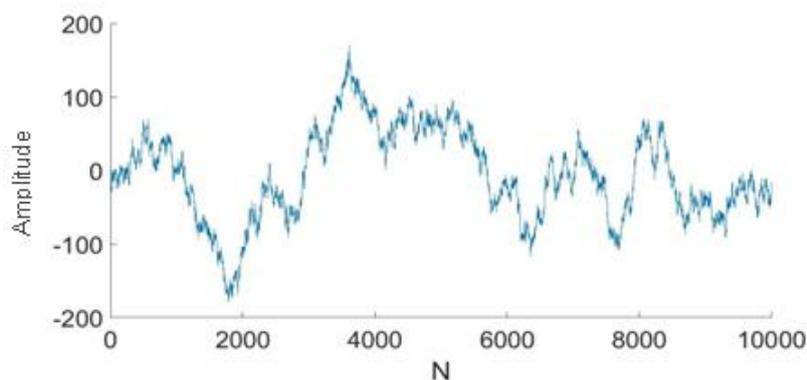


Figure 3. Graph obtained for a time series with power spectra exponent $\beta = 2$.

The method was implemented to generate a group of signals with $N = 10000$ data with known values of β , with β ranging from $\beta = 0$ to $\beta = 2$ with 0.1 increments between each series. Then we proceed to evaluate different types of entropy, but ultimately only the sample entropy was used because the best results were obtained for this type of entropy. The method proposed by Richman and Moorman was used for the calculation of the sample entropy (SampEn) [11]. Taking into account the time series, vectors of length m are defined by: $X_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\}$, for $1 \leq i \leq N - m + 1$. These vectors represent m consecutive values of x starting at point i , subsequently the distance between the vectors $X_m(i)$ and $X_m(j)$ is defined as the absolute value of the maximum difference between its scalar components:

$$d_m[X_m(i), X_m(j)] = \max[X_m(i+k) - X_m(j+k)], 0 \leq k \leq m-1 \quad (4)$$

Subsequently, for a given $X_m(i)$, the number of j ($1 \leq j \leq N - m, j \neq i$) is counted, such that the distance between $X_m(i)$ and $X_m(j)$ is less than or equal to r . Taking into account that $1 \leq i \leq N - m$:

$$B_i^m(r) = \frac{1}{N-m-1} B_i \quad (5)$$

$B^m(r)$ is defined as:

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \quad (6)$$

Then m is changed to $m + 1$ and $A_i^m(r)$ is calculated as:

$$A(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \quad (7)$$

As a result, $B^m(r)$ is defined as the probability that two sequences match in m points, and thus $A^m(r)$ corresponds to the probability that two sequences match on $m + 1$. So the sample entropy can be defined as:

$$\text{SampEn}(m, r, N) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \quad (8)$$

The values of r and m are decisive for the precision of the method. In this case, values of $r = 0.2$ and $m = 3$ were chosen, which proved to be suitable for this type of time series analysis, allowing a greater efficiency in the method and reducing the loss of information [10].

3. Results

For the characterization of the sample entropy method, it was applied to the theoretical time series previously generated, calculating the value of entropy for series of $N = 10000$ data and values of power spectral exponent ranging from $\beta = 0$ to $\beta = 2$ with increments of 0.1 between each series. Figure 4 shows the graph obtained for this calculation, being the abscissas axis the corresponding β values and the ordinates axis the entropy value. The method shows a linear trend extending from the values of $\beta = 0.7$ to $\beta = 2.0$, the human beat to beat intervals time series belongs to this interval.

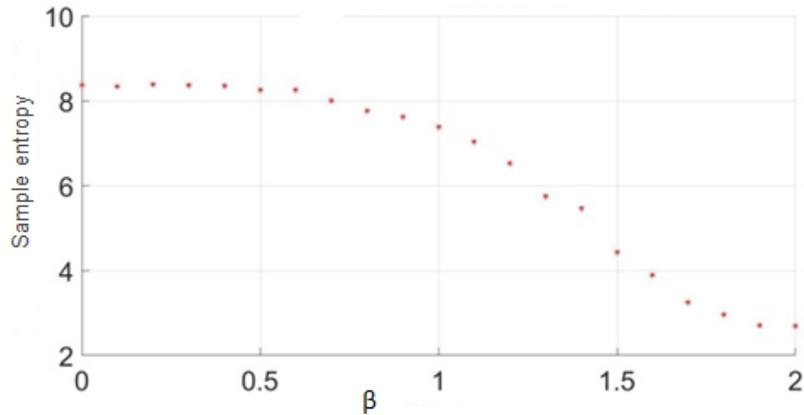


Figure 4. Sample entropy for different values of β .

Once characterized the method with time series of known values of β , it followed the analysis of the time series of the CHF database and healthy people for the three types of series (total 24 hours, sleep 6 hours and wake 6 hours). We proceeded to perform the analysis with the method of sample entropy. In the following figures, the graphic representation of the values sample entropy for each one of the samples, and the average value of both populations is represented as continuous lines. On dispersion graphs shown below, blue colour corresponds to healthy people, and red colour to heart failure patients.

Likewise, the statistical test for each type of study, i.e., complete series, series of 6 hours of sleep and series of 6 hours of wakefulness was made by using T-Student with a significance value of 0.5 and the hypothesis that were used for the previous study: the null hypothesis was $H_0: SE_{healthy} = SE_{CHF}$ and the alternative hypothesis $H_1: SE_{healthy} \neq SE_{CHF}$ (SE is sample entropy).

In the comparison carried out when they are sleeping, it was found a considerable difference between both averages, this difference is statistically significant. Whereas, for the series in which the individuals are awake, it can be seen a very small gap between the two average values, and even the average entropy of healthy people has a lower value than the patients with CHF, which does not correspond with the values obtained for the sleep period.

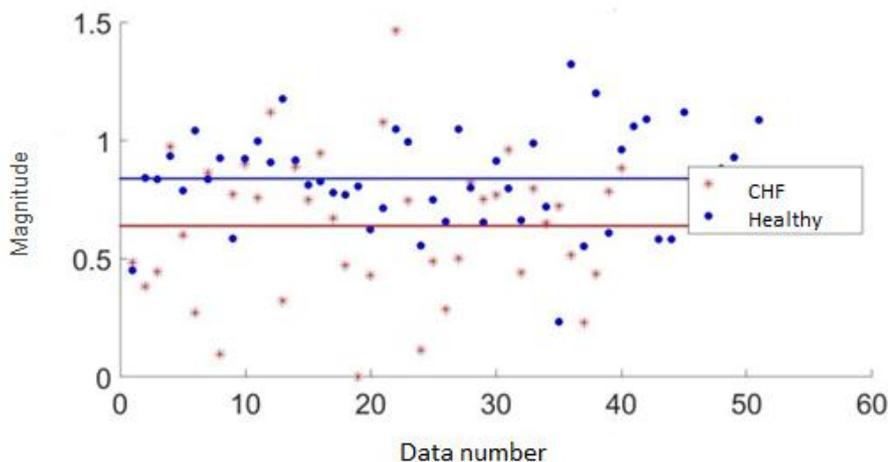


Figure 5. Sample entropy for healthy persons and CHF patients in time series of 6 hours when they were sleeping.

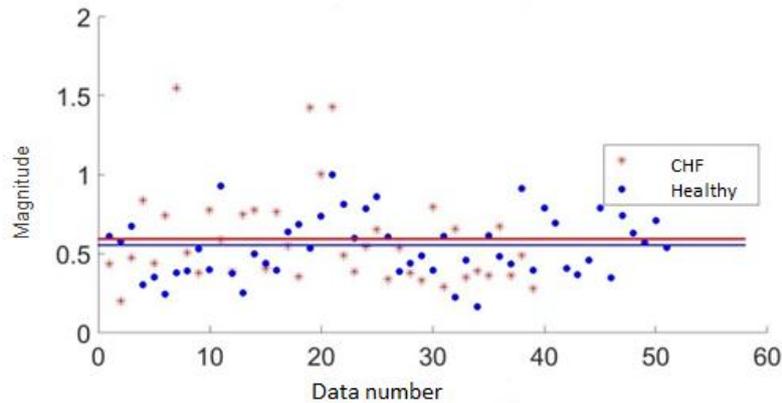


Figure 6. Sample entropy of patients and healthy people in 6 hour time series in the awoken phase.

4. Conclusions

Unlike other methods for evaluating the entropy, sample entropy does not present alterations in their results by the length of the series, however, the processing time increases exponentially as the size of series increases. The use of the $r = 0.2$ value gave reliable results in the calculation of the sampling entropy. We recommend the use of value $m = 3$ which was taken from the literature, which is recognized as a value that allows a satisfactory study of entropy and bringing precision to the method, without losing information. Applying the method to the tachogram analysis of healthy people and CHF patients it was possible to differentiate both groups, but the best results were obtained when people are sleeping, it is believed that this is because the condition of patients with CHF is aggravated by night.

5. References

- [1] Murray C J L and Lopez A D 1997 *Lancet* **349** 1269
- [2] Alberti K G M, Eckel R H, Grundy S M, Zimmet P Z, Cleeman J I, Donato K A, Fruchart J C, James W P T, Loria C M and Smith S C 2009 *Circulation* **120** 1640
- [3] Stanley H E, Amaral L A N, Goldberger A L, Havlin S, Ivanov P Ch and Peng C K 1999 *Physica A* **270** 309
- [4] Muñoz-Diosdado A, Angulo-Brown F and Del Río-Correa J L 2005 *Proceedings IEEE EMBS Annual International Conference* **27** 6981
- [5] McKee PA, Castelli W P, McNamara P M and Kannel W B 1971 *N. Engl. J. Med.* **285** 1441
- [6] Falk M 2012 *A First Course on Time Series Analysis* (Germany: University of Würzburg)
- [7] Gálvez-Coyt G, Muñoz-Diosdado A, Del Río-Correa J L and Angulo-Brown F 2010 *Fractals* **18** 235
- [8] Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P Ch, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 *Circulation* **101** e215
- [9] Bennett J A, Riegel B, Bittner V, Nichols J (2002) *Heart and Lung: Journal of Acute and Critical Care* **31** 262
- [10] Muñoz-Diosdado A, Ramírez-Hernández L, Aguilar-Molina A M, Zamora-Justo J A, Gutiérrez-Calleja R A and Virgilio-González C D 2014 *AIP Conf. Proc.* **1626** 155
- [11] Richman J S, Moorman J R 2000 *Heart and Circulatory Physiology Am. J. Physiol* **278** 2039
- [12] Wessel R, Koch C and Gabbiani F 1996 Coding of time-varying electric field amplitude modulations in a wave-type electric fish *J. Neurophysiol.* **75** 2280–93