Scientific Bulletin of the „Petru Maior" University of Tîrgu Mureş
Vol. 14 (XXXI) no. 2, 2017
ISSN-L 1841-9267 (Print), ISSN 2285-438X (Online), ISSN 2286-3184 (CD-ROM)

TIME SERIES ANALYSIS USING ENTROPY METRICS

Zoltán GERMÁN-SALLÓ[1]
Horațiu-Ștefan GRIF[2]
"Petru Maior" University of Tîrgu-Mureș, Romania
Nicolae Iorga Street, no.1, 540088 Tîrgu-Mureș, Romania
[1]zoltan.german-sallo@ing.upm.ro
[2]horatiu.grif@ing.upm.ro

Abstract

Entropy metrics are valuable tools for use in systems science and engineering. Both of domains are assumed to work with signals, to perform testing and evaluation tasks through specific signals. Entropy metrics indicate the degree of order/disorder in any given system, so they are capable to distinguish deterministic features from noise or chaos. The motivation of this paper is to present the main components and actual developments in entropy metrics used to analyze discrete time signals. The entropy metrics and measurements developed in this work try to find regularity in discrete time series in order to inspect them or predict future events.

Key words: discrete time series, entropy, entropy metrics, sample entropy, approximation entropy

1. Introduction

The purpose of this paper is to identify how entropy metrics can provide information about the complexity of discrete time series, reviewing theoretical elements and current developments. This paper performs the study using different signals as 1/f noise, white noise, and long term heart rate variability (HRV) signals. Starting from Clausius' first definition in thermodynamics through Shannon's entropy [1] of information in communication systems, a lot of another metrics have been developed lately [6], [9].

The present work introduces these lately developed metrics, the approximate, the sample and the multiscale entropies to discern between regularity and randomness in different test signals.

The approximate entropy represents a simple index for the overall complexity and predictability of time series [2], quantifying the likelihood that runs of patterns, which are close, remain similar for subsequent incremental comparisons.

The sample entropy improves the approximate entropy by excluding self-matches, quantifies the conditional probability that two sequences of m successive data points that are similar to each other (within a given tolerance r) will remain similar when one consecutive point is included [3], [8].

The multiscale entropy measurements differ from previous entropy techniques by including multiple time scales of measurement, computing the sample entropy at different scales [4], [5], [6], [10], [11].

2. Entropy metrics

The approximate entropy (ApEn) was derived from the theory of the Kolmogorov–Sinai entropy [12] for signals that included both information and noise. Its definition is presented below (relation 1).

$$ApEn(m,r,N) = \varphi^{m}(r) - \varphi^{m+1}(r) \qquad (1)$$

The parameter m is the distance between time series points to be compared, N is the length of the time series, r is the domain of similarity (or tolerance) and φ is the probability that points m distance apart would be within the distance r. The purpose for the similarity domain term, r, is to identify a range in which variations are to be considered similar. High values of approximate entropy indicate high irregularity and complexity in time-series data.

The most important limitations of ApEn are that it requires noise-free stationary data, an inherent bias exists and self-matches are counted too. Also, it strongly depends on the discrete signal's length and evaluates regularity on one scale only.

The sample entropy (SampEn) which is an improved version of approximate entropy is defined as:

$$SampEn = \lim_{N\to\infty}\left\{-\ln\left[\frac{\varphi^{m+1}(r)}{\varphi^{m}(r)}\right]\right\} \qquad (2)$$

Is the natural logarithm of of an estimate of the conditional probability that subseries of length m that match pointwise within a tolerance r also match at the next point. The semnifications of parameters are the same as in relation (1), the sign minus assures positive values [6]. The SampEn is estimated by the statistic

$$SampEn(N,m,r) = -\ln\left[\frac{\varphi^{m+1}(r)}{\varphi^{m}(r)}\right] \quad (3)$$

SampEn is more useful than the ApEn at identifying changes in the point-to-point variations in discrete time signals. It has limitations too, as stationary signal requirement; a higher pattern length requires an enormous number of data points and evaluates regularity on one scale only. A distance between two sequences as the absolute maximum difference between their scalar components, can be defined as follows:

$$r = \max_{k=0,\dots m-1}\left(\left|x(i+k)\right| - \left|x(j+k)\right|\right) \quad (4)$$

The multiscale entropy (MsEn) is used to compute the corresponding sample entropy over a sequence of scale factors. For a one-dimensional time series, $x=\{x_1, x_2, \dots, x_N\}$ the so-called coarse-grained time series $y^{(\tau)}$, can be constructed at a scale factor of τ, according to the following equation [5]:

$$y_j^{\tau} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i \qquad 1 < j < \frac{N}{\tau} \quad (5)$$

The coarse-grained procedure is illustrated in fig. 1.

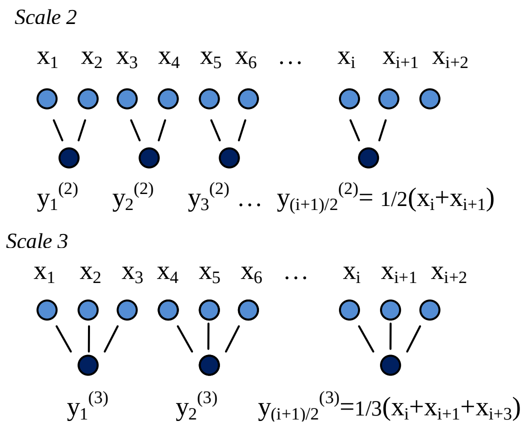

Fig. 1: Schematic illustration of coarse-grained procedure

As shown in fig. 1, the coarse-grained time series is divided into non-overlapping sequences of length τ, and the data points inside each sequence are averaged. The MsEn value is then defined as the entropy measurement of each coarse-grained time series.

$$MsEn(\mathbf{x},\tau,m,r) = SampEn(\mathbf{y}_1^{(\tau)},m,r) \quad (6)$$

By including multiple time scales of measurement, the MsEn allows to identify the time scales at which the change in complexity occurs.

In the composit multiscale entropy (CmsEn) algorithm, the k-th coarse-grained time series for a scale factor of τ, $y_k^{(\tau)}=\{y_{k,1}^{(\tau)}, y_{k,2}^{(\tau)}, \dots, y_{k,p}^{(\tau)}\}$is defined as:

$$y_{k,j}^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau+k-1} x_i, \quad 1 < j < \frac{N}{\tau}, \quad 1 \le k \le \tau \quad (7)$$

In the Composite Multiscale Entropy (CmsEn) algorithm, at a scale factor of τ, the sample entropies of all coarse-grained time series are calculated and the CmsEn value is defined as the means of τ entropy values [12], [13].

$$CmsEn(\mathbf{x},\tau,m,r) = \frac{1}{\tau}\sum_{k=1}^{\tau} SampEn(\mathbf{y}_k^{(\tau)},m,r) \quad (8)$$

The CmsEn procedure is presented on fig. 2, where are two and three coarse-grained time series obtained from the original sequence for scale factors of 2 and 3.

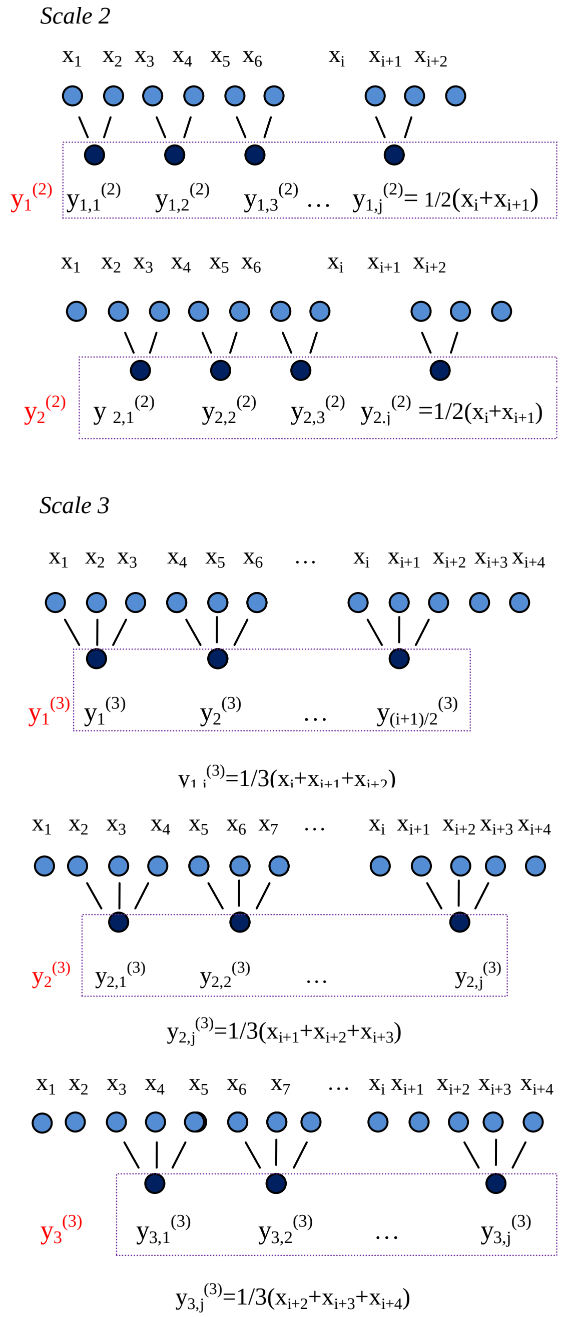

Fig. 2: Schematic illustration of the CmsEn procedure

3. Entropy measurements results

This work uses test signals in order to evaluate the different entropy metrics. The used signals are white noise, 1/f noise, and long-term heart rate variability signals. The evaluation is performed under Matlab environment, using Signal Processing Toolbox to calculate the entropy metrics The mentioned noise signals were developed and generated also in Matlab, the others are obtained from specific databases [7].

The next figures present the used signals with their most important parameters (histograms, spectrum)

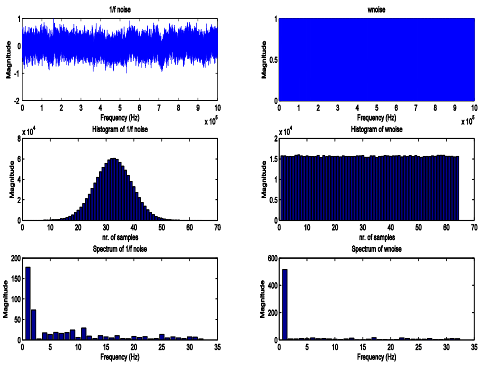

Fig. 3: Noise test signals

The 1/f noise is a largely used signal, its power spectral density is inversely proportional to the frequency of the signal. In in this work it was obtained by filtering random noise through a FIR filter that has a 1/f pass-band [14].

Figure 4 presents two heart rate variability signals with their main properties, histogram and spectrum. The signal were obtained from Physionet database [7].

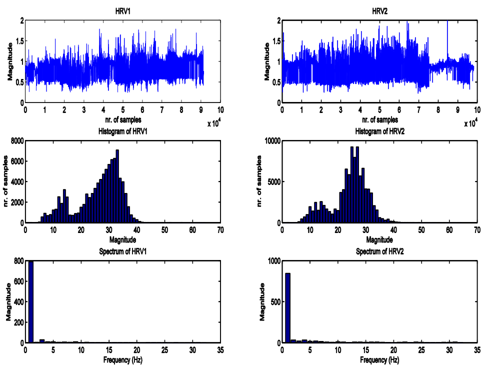

Fig. 4: HRV test signals

These signals were analyzed through entropy metrics and the previously mentioned parameters were calculated. In the case of approximate and sample entropy, usual the value for sensibility domain was chosen 0.2 times the standard deviation of the data [9] and computation were made for different embedding lengths as can be seen in figures 5,6,7.

Figure 5 presents the obtained approximate entropy and the sample entropy for white and 1/f noise for different lengths of signal with embedding length of 2.

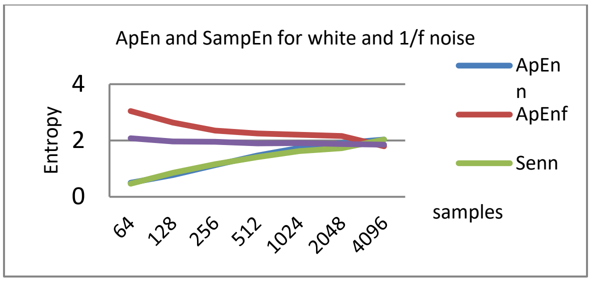

Fig. 5: White and 1/f noise entropies vs length

The corresponding entropies for HRV signals are presented on figure 6. The obtained values are smaller due to increased level of regularity.

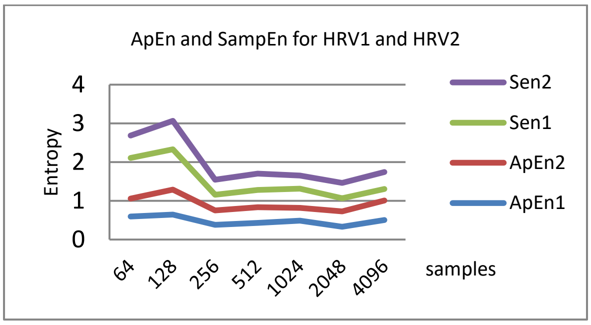

Fig. 6: HRV signal entropies vs signal length

The highest value of approximate entropy and sample entropy indicates a highest degree of disorder, a more accentuate randomness. As it can be seen, the sample entropy offers a more accurate insight into the signal regularity. The multiscale and multiscale composite entropies are computed according to the algorithms presented in [5]. The obtained values are presented in figure 7 together to emphasize the obvious differences. Ew and Ef are the multiscale composite entropies for white noise and 1/f noise and E1 E2 are for HRV1 and HRV2 signals. The used lengths are 4096 (samples), the sample entropies for each time scale were computed with a threshold of 0.15 times the standard deviation.

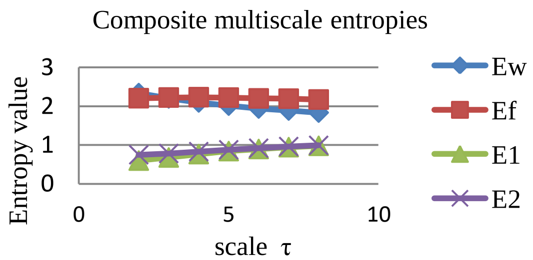

Fig. 7: Composite multiscale entropies for time scales 2 to 8 for the test signals

At the multiscale algorithms the larger the scale factor is, the shorter the coarse-grained sequence is. Therefore, the variance of the entropy of the coarse-grained series estimated by sampling entropy increases as a time scale factor increases.

4. Conclusions

The effectiveness of the entropy metrics is evaluated through two generated noise signals and two real heart rate variability data set provided by Physionet database. The analysis of white noise, 1/f noise and two heart rate variability signals shows that the multiscale entropies provide a more accurate estimation of regularity, self-repeatability than the approximate and sample entropies. These entropy measurements depend on the length of time-series. In case of the multiscale statistics each coarse-grained time series is equal to that of the original time series divided by the scale factor, τ, then the variance of entropy values grows as the length of coarse-grained time series is reduced. These entropy metrics computations need a lot of computing resources due to the signal length, the multiscale entropy metrics bring better results at shorter sequences.

References

[1] Richman J.S., Moorman J.R., *Physiological time-series analysis using approximate entropy and sample entropy*, (2000),American Journal of Physiology, Heart Circulation Physiology, 278(6):H2039-49.

[2] García-Martínez, B., Martínez-Rodrigo, A., Cantabrana, Z., R., García, J., M., P., Alcaraz, R. (2016), *Application of Entropy-Based Metrics to Identify Emotional Distress from Electroencephalographic Recordings*, Entropy 2016, 18(6), 221; doi:10.3390/e18060221.

[3] Eckmann J.P., Ruelle D., *Ergodic theory of chaos and strange attractors.* Rev Mod Phys 1985;57:617–56.

[4] Busa, M., Emmerik, R. (2016), *Multiscale entropy: A tool for understanding the complexity of postural control*, Journal of Sport and Health Science 5, pp. 44–51.

[5] Pincus S., M. (1991), *Approximate entropy as a measure of system complexity*. Proc Natl Acad Sci USA 1991;88:2297–301.

[6] Goldberger A.L., Amaral LAN, Glass L., Hausdorff J.M., Ivanov P.Ch., Mark R.G., Mietus J.E., Moody G.B., Peng C-K, Stanley H.E. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals.* Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/content/101/23/e215.full]; 2000 (June 13).The IEEE website (2002). [Online]. Available: http://www.ieee.org/

[7] Guzmán-Vargas, L., Ramírez-Rojas, A.; Angulo-Brown, F. (2008), *Multiscale entropy analysis of electroseismic time series.* Nat. Hazards Earth Syst. Sci. 2008, vol. 8, pp. 855–860.

[8] Pan, Y.H., Lin, W.Y., Wang, Y.H., Lee, K.T., *Computing multiscale entropy with orthogonal range search. J. Mar. Sci. Technol*, 2011, vol. 19, pp. 107–113. Karnik, A. (1999), *Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP*, M. Eng. thesis, Indian Institute of Science, Bangalore, India.

[9] Pincus, S., Gladstone, I., Ehrenkranz, R. (1991). *A regularity statistic for medical data analysis*, Journal of Clinical Monitoring, 7 (4), pp. 335-345.

[10] Costa, M.; Goldberger, A.L.; Peng, C.K. (2005) *Multiscale entropy analysis of biological signals.* Phys. Rev. E Stat. Nonlin Soft Matter Phys. 2005, 71, 021906.

[11] Christoph Bandt, Bernd Pomp, (2002) *Permutation Entropy: A Natural Complexity Measure for Time Series*, Phys. Rev. Lett. 88, 174102.

[12] Costa M, Goldberger AL, Peng CK, (2002) *Multiscale entropy analysis of complex physiologic time series*, Phys Rev Lett., 2002.

[13] Wei, Q., Liu, D.H., Wang, K.H., Liu, Q., Abbod, M.F., Jiang, B.C., Chen, K.P., Wu, C., Shieh, J.S. (2012) *Multivariate multiscale entropy applied to center of pressure signals analysis: an effect of vibration stimulation of shoes*. Entropy, 14, pp. 2157–2172.

[14] Litak, G.; Syta, A.; Rusinek, R. Dynamical changes during composite milling: recurrence and multiscale entropy analysis. Int. J. Adv. Manuf. Technol. 2011, vol. 56, pp. 445–453.

[15] Lin, J.L.; Liu, J.Y.C.; Li, C.W.; Tsai, L.F.; Chung, H.,Y.,(2010) *Motor shaft misalignment detection using multiscale entropy with wavelet denoising*. Expert Syst. Appl. 37, pp. 7200–7204.