Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Characterization of Sample Entropy in the Context of Biomedical Signal Analysis

Mateo Aboy, David Cuesta-Frau, Daniel Austin and Pau Micó-Tormos

Abstract— **Sample Entropy (SampEn) has been proposed as a method to overcome limitations associated with approximate entropy (ApEn). The initial paper describing the SampEn metric included a characterization study comparing both ApEn and SampEn against theoretical results and concluded that SampEn is both more consistent and agrees more closely with theory for known random processes than ApEn. SampEn has been used in several studies to analyze the regularity of clinical and experimental time series. However, questions regarding how to interpret SampEn in certain clinical situations and its relationship to classical signal parameters remain unanswered. In this paper we report the results of a characterization study intended to provide additional insights regarding the interpretability of SampEn in the context of biomedical signal analysis.**

I. INTRODUCTION

Approximate Entropy (ApEn) is one of the most popular metrics used to estimate complexity and regularity in the field of biomedical signal analysis [1]–[5]. This metric compares patterns within the time series and estimates the regularity of the signal. ApEn has been successfully applied to analyze physiologic signals in diverse applications. For instance, ApEn has been used to study intracranial hypertension episodes in pediatric patients with traumatic brain injury [6], [7], to analyze temperature registers with the objective of predicting outcome [8], [9], to analyze time series generated by schizophrenic patients [10], and to study heart rate variability in disease and due to aging [11].

Despite its popularity, ApEn has known shortcomings including bias, consistency, and dependance on sample length [12]. These shortcomings have led to the development of a related measure, sample entropy (SampEn) [12]. Theoretically, SampEn reduces the ApEn bias by avoiding counting self-matches, is independent of the time series length, and is more consistent than ApEn. Additionally, SampEn is easier to calculate than ApEn.

The algorithm used to measure the complexity of a time series with SampEn is similar to that used by ApEn. Three parameters are needed to calculate SampEn: m, r, and N. Although entropy is defined as m approaches infinity and as r approaches 0, there are no general guidelines to set the values of these parameters. Experimental tests indicate that a good choice for the pattern length parameter m is $m = 1$ or $m = 2$. On the other hand, it is common to set the threshold parameter r to be some percentage of the standard deviation of the time series in order not to depend on the absolute amplitude of the signal. Recommendations lead to the use of r between 0.1 and 0.25. If r is too small, noise affects the SampEn measure. If r is too large, some changes of the signal are not detected [13], [14].

Recent studies have used SampEn to analyze physiologic signals in diverse applications such as heart rate variability analysis [11] and EEG analysis [15]. However, contrary to ApEn or Lempel–Ziv complexity where characterization studies have been conducted to aid their interpretability [7], [16], these studies have not been reported on SampEn. In this paper we report the results of a study aimed at providing better interpretability of SampEn in the context of biomedical signal analysis.

M. Aboy is with the Department of Electrical Engineering at the Oregon Institute of Technology (Portland Campus), Portland, OR, USA mateoaboy@ieee.org

D. Austin is with the Department of Electrical Engineering at University of Southern California, Los Angeles, CA, USA daniel@metodlabs.com

D. Cuesta-Frau and P. Micó Tormos are with the Department of Computer Science, Polytechnic University of Valencia, Alcoi Campus, Alcoi, 03801, Spain dcuesta@disca.upv.es

II. METHODS

A. Calculation of Sample Entropy

The sample entropy of a time series $\langle x(n) \rangle$ of length N, $SampEn(m, r, N)$ is computed as follows:

1) Take m vectors $X_m(1), X_m(2), \ldots, X_m(N-m+1)$, defined as $X_m(i) = [x(i), x(i+1), \ldots, x(i+m-1)]$, for $1 \leq i \leq N-m+1$. These vectors are m consecutive values of x, commencing at the ith sample.
2) The distance between vectors $X_m(i)$ and $X_m(j)$, $d[X_m(i), X_m(j)]$ is defined as:
$$d[X_m(i), X_m(j)] = max(|x(i+k) - x(j+k)|) \quad (1)$$
For a given $X_m(i)$, count the number of $j (1 \leq j \leq N-m, j \neq i)$, such that $d[X_m(i), X_m(j)] \leq r$. This number is denoted as B_i. For $1 \leq i \leq N-m$, two new values are defined and computed, $B_i^m = \frac{1}{N-m-1} B_i$ and $B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r)$.
3) Length is increased to $m = m+1$, and previous steps are repeated to obtain the counterpart of B with this new value of m, $A_i^m = \frac{1}{N-m-1} A_i$ and $A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r)$, where B^m is the probability that two sequences coincide for m points, and A^m is the probability that coincide for $m+1$ points.
4) Finally, compute SamEn as $\mathrm{SampEn}(m, r) = lim\{-\log[\frac{A^m(r)}{B^m(r)}]\}$. Since the time series length is

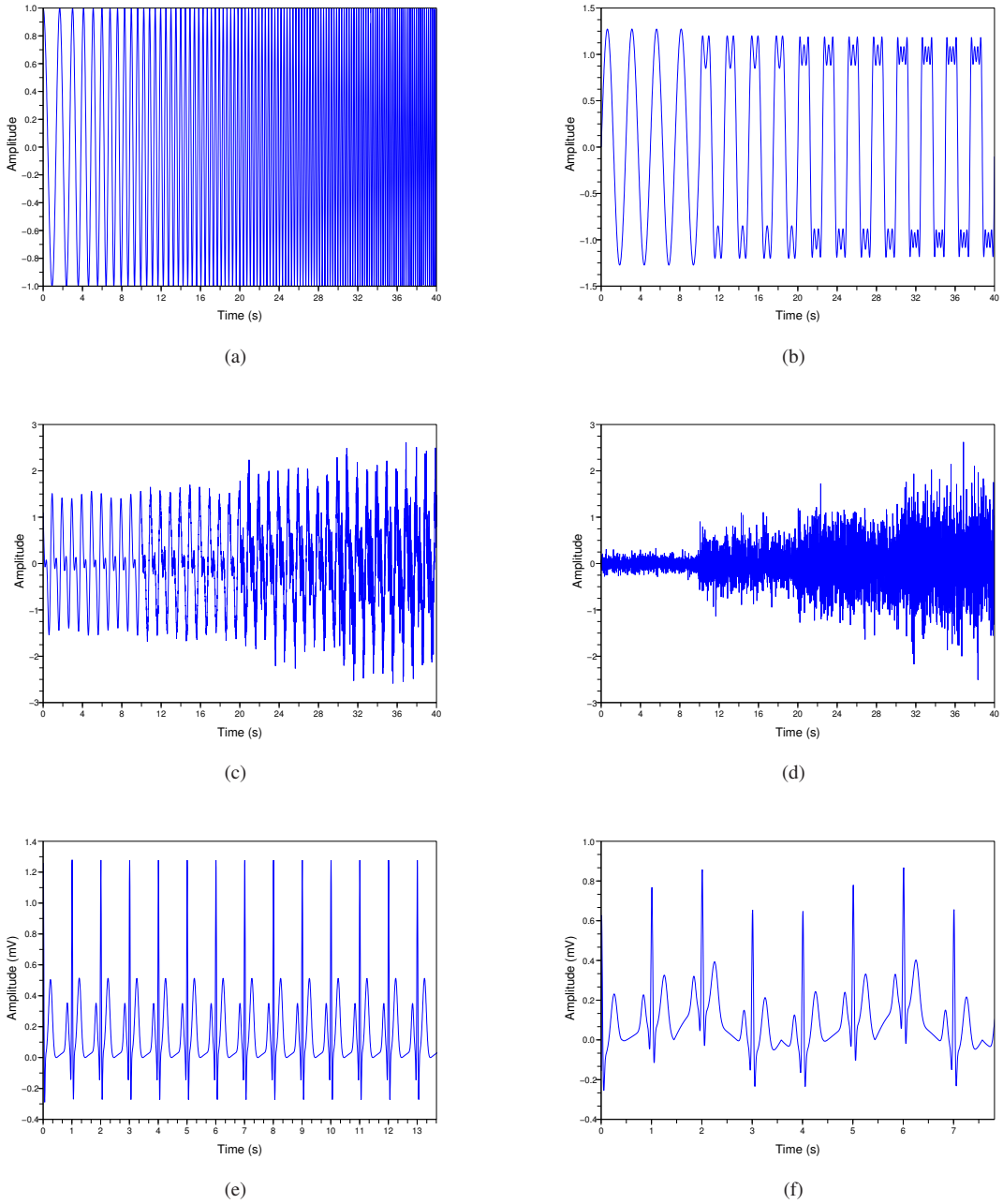

Fig. 1. Synthetic signals used in this study: (a) Chirp signal. (b) Signal with growing number of harmonics. (c) Quasiperiodic signal with different noise levels. (d) White Gaussian noise with step increases in power. (e) Synthetic normal ECG. (f) Synthetic normal ECG with baseline wander.

finite, SampEn is estimated as $\mathrm{SampEn}(m, r, N) = -\log[\frac{A^m}{B^m}]$.

B. Synthetic test signals

In order to characterize SampEn we used a subset of the test signals that have been previously employed to study ApEn and Lempel-Ziv complexity [7], [16] . These synthetic signals include:

1) SampEn versus frequency. Chirp signal whose frequency was increased linearly from 0.5 Hz to 5 Hz in 5s.
2) SampEn versus frequency content. Four concatenated periodic signals of 10s with 1, 2, 5 and 7 frequency components.
3) SampEn versus quasi-periodic signal plus noise. An amplitude modulated harmonic quasi–periodic signal with white Gaussian noise of different power levels.
4) SampEn versus noise power.

In addition to these general synthetic signals, we used a synthetic electrocardiogram in order to study how the different ECG parameters affected SampEn.

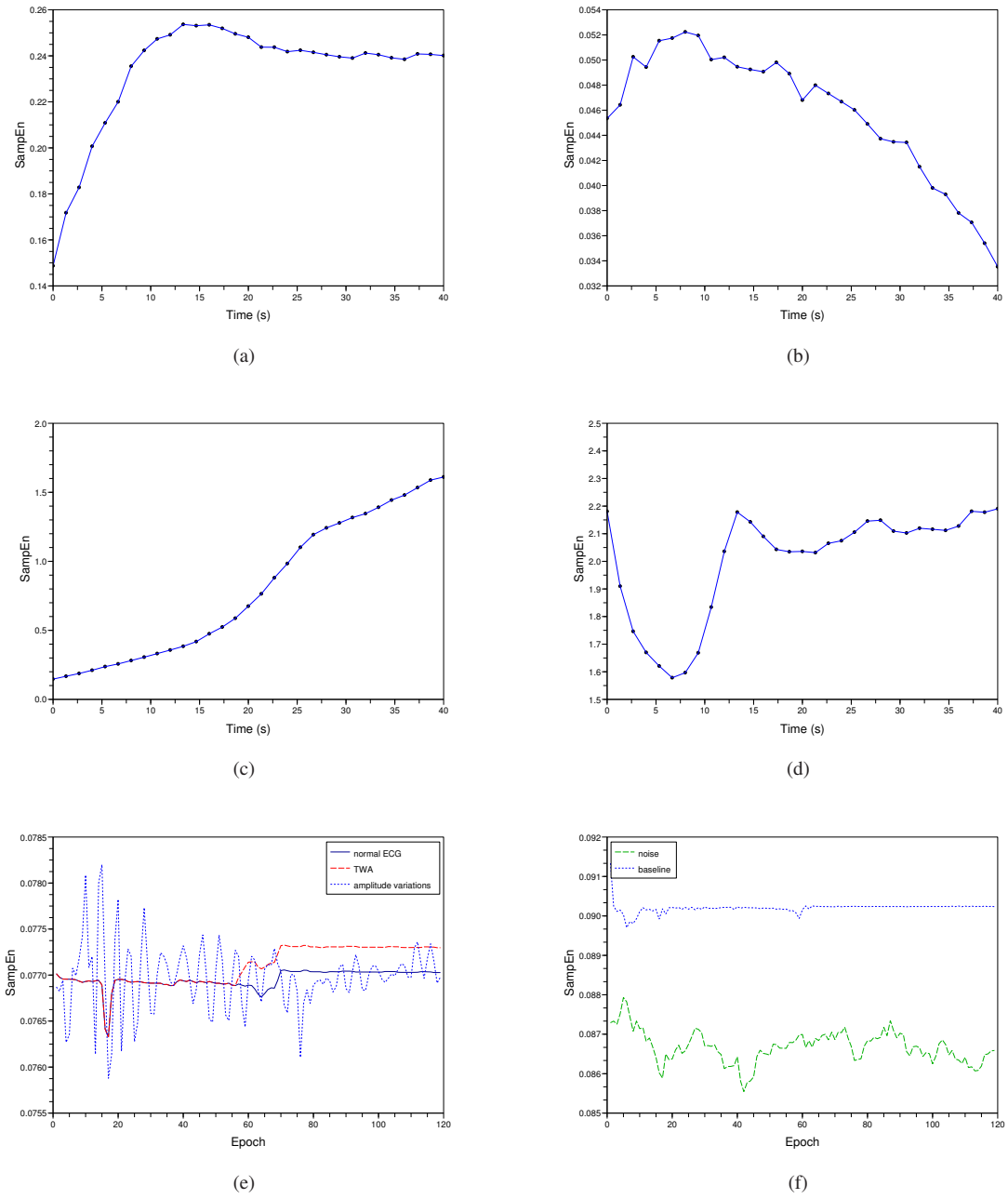

Fig. 2. Results of the simulation study: (a) Chirp signal. (b) Signal with increasing number of harmonics. (c) Quasiperiodic signal with different noise levels. (d) White Gaussian white noise with step increases in power. (e) Synthetic normal ECG. (f) Synthetic normal ECG with baseline wande.

III. RESULTS AND DISCUSSION

SampEn was measured using SampEn$(m = 2, r = 0.2, N)$, where the length N was 5120 for ECG signals (20s), and 1250 for the general signals (10s). The data was normalized (mean subtraction and division by the standard deviation) prior to the SampEn computation. We used an overlapping window of 90% between consecutive analyzed signal epochs.

The results of the test are shown in Fig. 2. Each plot shows the results corresponding to the input signals depicted in Fig.1. Fig. 2–(a) shows the relationship between SampEn and frequency changes. This test shows that SampEn increases as the frequency increases up to a saturation point. The saturation point depends on the rate of change of the frequency and the relationship between the maximum frequency and the Nyquist rate. Thus, for oversampled signals where the frequency changes linearly, it is expected that SampEn will increase also. However, this result does not apply in the case of undersampled signals. The results of the second test (Fig. 2–(b)) illustrate this point. Note that SampEn initially increases as the number of harmonics increases but decreases

TABLE I
PARAMETERS USED FOR THE SYNTHETIC ECG GENERATION

Parameter	Value
Number of beats	256
Sampling frequency (fs)	256 Hz
Beat rate	60 bpm

later as the frequency of the harmonics approaches half the sample frequency. This type of behavior often leads to interpretation problems in biomedical signal analysis; researchers should be cautioned about it since different sample frequencies may result in different SampEn results.

The relationship between white noise and SampEn is shown in Fig. 2–(c) and Fig. 2–(d). We can see that for high SNR involving quasi-periodic signals the SampEn increases as the SNR decreases (i.e. SampEn is positively correlated with the noise power) as shown in Fig. 2–(c). However, in situations of low SNR, SampEn is not so dependent on the noise power. Note how after the initial transient shown in Fig. 2–(d), SampEn is nearly independent of the noise power.

Fig.2–(e) shows the SampEn measures obtained for a normal (ideal) ECG with constant wave amplitude and RR interval, for an ECG with wave amplitude fluctuations (average amplitude remains the same), and for an ECG with T wave alternancy starting at beat 128 (in the middle of the signal) of 50μV in amplitude. In the first case (normal ECG), the SampEn remained constant around a value of 0.0770 (this happened also for N=2s, 8s, and 32s and for N=0.25s, it was 0.0940, that is, epoch length was too short to obtain a correct value). When ECG wave amplitude was allowed to fluctuate (std=1.0), the average SampEn measured was the same but it caused a SampEn variance increase. In the last case – when the T wave alternancy begins, SampEn increases by 1%. These results indicate that a few heartbeats suffice to compute the SampEn (fast convergence). Additionally, centered fluctuations do not affect the global SampEn average (only local SampEn measurements are affected), and even small changes in ECG complexity (T wave alternancy) are captured by SampEn.

Fig.2–(f) shows the SampEn results obtained when the normal ECG was corrupted by uniform random noise (10μV amplitude) or by sinusoidal baseline wandering (50μV amplitude). The SampEn in the first case (noisy ECG) yielded smaller values than those of ECG with baseline drifts. As for the case of wave amplitude fluctuations, it seemed information provided by SampEn was present in both the measurement average and in its variance. Additionally, both ECG disturbances caused the SampEn average to raise by 13% or 17%, which may mask other SampEn changes.

IV. CONCLUSIONS

We performed a characterization study of SampEn aimed at providing additional insights regarding the interpretation of this complexity metric in the context of biomedical signal analysis. Our results indicate that SampEn is dependent on the rate of change of frequency and the relationship between the maximum frequency of the signal and the Nyquist rate. The dependance of SampEn on noise power depends on the SNR. For quasi-periodic signals with high SNR the SampEn increases as the power of the noise increases. However, for very low SNR or in the case of pure noise signals, the SampEn is not so clearly dependent on the noise power. Finally, we also study how different ECG parameters affect SampEn. Further studies are needed to provide a complete characterization of SampEn.

REFERENCES

[1] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis." *J Clin Monit*, vol. 7, no. 4, pp. 335–345, Oct 1991.

[2] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proc Natl Acad Sci U S A*, vol. 88, no. 6, pp. 2297–2301, Mar 1991.

[3] ——, "Greater signal regularity may indicate increased system isolation." *Math Biosci*, vol. 122, no. 2, pp. 161–181, Aug 1994.

[4] S. Pincus and R. E. Kalman, "Not all (possibly) "random" sequences are created equal." *Proc Natl Acad Sci U S A*, vol. 94, no. 8, pp. 3513–3518, Apr 1997.

[5] S. Pincus and B. H. Singer, "A recipe for randomness." *Proc Natl Acad Sci U S A*, vol. 95, no. 18, pp. 10 367–10 372, Sep 1998.

[6] R. Hornero, M. Aboy, D. Abasolo, J. McNames, W. Wakeland, and B. Goldstein, "Complex analysis of intracranial hypertension using approximate entropy." *Crit Care Med*, vol. 34, no. 1, pp. 87–95, Jan 2006.

[7] R. Hornero, M. Aboy, D. Abásolo, J. McNames, and B. Goldstein, "Interpretation of approximate entropy: analysis of intracranial pressure approximate entropy during acute intracranial hypertension." *IEEE Trans Biomed Eng*, vol. 52, no. 10, pp. 1671–1680, Oct 2005.

[8] M. Varela, L. Jimenez, and R. Fariña, "Complexity analysis of the temperature curve: new information from body temperature." *Eur J Appl Physiol*, vol. 89, no. 3-4, pp. 230–237, May 2003. [Online]. Available: http://dx.doi.org/10.1007/s00421-002-0790-2

[9] M. Varela, J. Churruca, A. Gonzalez, A. Martin, J. Ode, and P. Galdos, "Temperature curve complexity predicts survival in critically ill patients." *Am J Respir Crit Care Med*, vol. 174, no. 3, pp. 290–298, Aug 2006. [Online]. Available: http://dx.doi.org/10.1164/rccm.200601-058OC

[10] R. Hornero, D. Abásolo, N. Jimeno, C. I. Sánchez, J. Poza, and M. Aboy, "Variability, regularity, and complexity of time series generated by schizophrenic patients and control subjects." *IEEE Trans Biomed Eng*, vol. 53, no. 2, pp. 210–218, Feb 2006.

[11] M. M. Platisa and V. Gal, "Dependence of heart rate variability on heart period in disease and aging." *Physiol Meas*, vol. 27, no. 10, pp. 989–998, Oct 2006. [Online]. Available: http://dx.doi.org/10.1088/0967-3334/27/10/005

[12] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy." *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2039–H2049, Jun 2000.

[13] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability." *Am J Physiol Regul Integr Comp Physiol*, vol. 283, no. 3, pp. R789–R797, Sep 2002. [Online]. Available: http://dx.doi.org/10.1152/ajpregu.00069.2002

[14] D. E. Lake, "Renyi entropy measures of heart rate gaussianity." *IEEE Trans Biomed Eng*, vol. 53, no. 1, pp. 21–27, Jan 2006.

[15] D. Abásolo, R. Hornero, P. Espino, D. Alvarez, and J. Poza, "Entropy analysis of the eeg background activity in alzheimer's disease patients." *Physiol Meas*, vol. 27, no. 3, pp. 241–253, Mar 2006. [Online]. Available: http://dx.doi.org/10.1088/0967-3334/27/3/003

[16] M. Aboy, R. Hornero, D. Abásolo, and D. Alvarez, "Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis." *IEEE Trans Biomed Eng*, vol. 53, no. 11, pp. 2282–2288, Nov 2006.