Cross-Entropy of Systolic Blood Pressure – Pulse Interval: Automatic Threshold and its Reliability

Tamara Ćeranić[1], Tatjana Lončar-Turukalo[1], Branislav Milovanović[2], Nina Japundžić-Žigon[2], Dragana Bajić[1]

[1]Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
[2]Faculty of Medicine, University of Belgrade, Belgrade, Serbia

Abstract

This paper introduces a set of formulae for automatic evaluation of threshold level r_{TEOR} for which the cross-entropy of systolic blood pressure (SBP) and pulse interval (PI) (Cross-ApEn) reaches its maximum. A mathematical method for estimating the level of consistency of entropy estimates is established as well. These two methods jointly determine a steady working point for consistent entropy estimates.

1. Introduction

Cross-entropy (*CrossEn*) method quantifies asynchrony in two parallel time series [1]. Despite the importance of understanding the relationship between two related signals in medical applications (epidemiology, hormone study, cardiovascular study), the method have been discussed in a few papers [1-4].

Cross-ApEn is an extension of widely applied approximate entropy (*ApEn*) [1] with the difference that the signal sub-segments are compared to templates from another time series. *Cross-ApEn* is based on estimation of the conditional probability that sub-segments of two observed time series remain similar if the length of sub-segment is increased. To calculate *Cross-ApEn,* some pre-determination criteria are required, the same ones as for *ApEn*. One is the threshold *r*: two sub-segments are considered as similar if the distance between their respective elements is below the threshold. Besides the mentioned criteria, it is also necessary to define the length of sub-segments (*m*). Our recent work on *ApEn* has shown that the greatest instability in entropy estimation is caused by uncritical selection of threshold *r*. [5].

Pincus, who proposed *AppEn* and *Cross-ApEn* [4], recommended that value of *r* should be in range of 0.1 – 0.2 times the standard deviation of time series. Chon et al. noted that the recommended range does not always corresponds to the maximum value of *ApEn* and proposed formulae for automatic selection of the threshold *r* that yields maximum value of *ApEn* ($ApEn_{MAX}$) [6,7]. In [2] the unreliability of entropy estimation is pointed out, depending on the choice of the value of threshold *r*.

The aim of this paper is to propose formulae for automatic selection of threshold *r* (r_{TEOR}) for *Cross-ApEn* maximum value of systolic blood pressure (SBP) and pulse interval (PI), and to check the consistency of *Cross-ApEn* estimation. Also, the threshold value that improves the reliability of entropy estimate – *r* (r_{CON}) – is proposed.

2. Experimental data

2.1. Data collection

The experimental data includes two sets of SBP and PI signals, short ones and longer ones. Short signals have been collected from 41 healthy volunteers, different in gender and age. Task Force Monitor was used for both recording ECG and arterial blood pressure (ABP) (fs= 1000Hz) and extraction of SBP and PI. Patients did not consume any type of medication before recording. Only last 500 samples of time series were analyzed, because it is considered that during this period patients were fully adapted to the recording conditions.

The second data set (long time series) was extracted from ABP recorded from laboratory animals (fs= 1000Hz). Six normotensive Wistar rats (NRM) and six borderline hypertensive rats (BHR) were exposed to two types of stress (shaker and restraint). Shaker stress involved shaking platform (200 cycles per minute) on which rats were placed. For the second type of stress, animals were placed in a Plexiglas restraint tube for around 60 minutes. The data set includes the signal recorded in baseline condition, before exposure to stress (BS) and after first stress (FS). Both groups of rats (BHR and NRM) were exposed to both types of stress, under the same conditions. The experiments on laboratory animals were performed in accordance with European Communities Council directive of 24/1986. The description of the experiment in details was given in [5].

2.2. Preprocessing data

After a careful visual examination of time series, the artefacts were removed. Thereafter, the time series were detrended by method proposed in [8].

The wide sense stationarity of detrended time series was checked by stationarity test described in [9]. All signals were normalized and centralized, in order to enable the comparison of different signals.

3. Methods

3.1. Cross approximate entropy

The *Cross-ApEn* estimation, as applied in [3], is described in five steps. Two time series are denoted as $u=\left(u(1),u(2),..u(N)\right)$ and $v=\left(v(1),v(2),...v(N)\right)$, N is the length of time series.

Step 1: Divide both time series into ($N - m + \tau$) patterns, for given m.

$$X_m^{(i)}=\left[u(i),u(i+\tau),....,u(i+m-\tau)\right] \quad i=1,...,N-m+\tau \quad (1)$$

$$Y_m^{(j)}=\left[v(j),v(j+\tau),....,v(j+m-\tau)\right] \quad j=1,...,N-m+\tau. \quad (2)$$

Step 2: Calculate the distance between patterns defined as

$$d_m\left(X_m^{(i)},Y_m^{(j)}\right)=\max_{k=0,..,m-1}\left[\left|u(i+k-\tau)-v(j+k-\tau)\right|\right] \quad (3)$$

The two patterns are similar to each other if the distance between the respective elements in patterns is less than predefined threshold r.

Step 3: The probability of occurrence of patterns that are similar to the template patterns is estimated as

$$C_i^m(r)(v\,\mathrm{P}u)=\frac{B_m^r(i)}{N-(m-1)\tau}, \quad (4)$$

where $B^r{}_m(i)$ is a number of patterns for which the distance is below r, v template vector and u matching vector.

Step 4: The procedure is repeated for patterns of the length m+1.

Step 5: *Cross-ApEn* is defined as

$$Cross-ApEn(m,r,\tau)=\frac{1}{N-(m-1)\tau}\sum_{i=1}^{N-(m-1)\tau}\ln\left(C_i^m(r)\right)-\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\ln\left(C_i^{(m+1)}(r)\right). \quad (5)$$

Cross-ApEn(m,r,τ) is an asymmetric measure, so it is necessary to observe both time series in the role of template/matching vectors to get a realistic image of the relationship between them. The bias caused by self-matching (comparison of the pattern with itself) inherent to *ApEn* is not a characteristic of *Cross-ApEn*.

Authors usually opt for $(0.1-0.2)\times std$ of time series as a range for the threshold r, for m = 2, as suggested by Pincus [4], and for τ = 1. Influence of selected parameters m, r on entropy estimation was described in [10], and influence of parameter τ in [11]. The largest instability in entropy estimation is induced by the threshold r [5].

3.2. Automatic selection of the threshold value r (r_{TEOR}) for cross-entropy maximum value

The maximum value of *Cross-ApEn* corresponds to the largest deviation between the two given time series [6]. To find it, the safest but time-consuming solution would be to estimate *Cross-ApEn* for a wide range of values of r (usually 0 - 1) and to select the maximum value of cross entropy.

Cross-ApEn (PI‖SBP) of three NRM rats for a range of threshold values (0 – 0.5) is shown in Fig. 1. It shows that the conclusion about the complexity of time series can be completely different for different threshold values. If we observe the complexity of the signals for the recommended values r = 0.1 and r = 0.2 (values usually chosen in literature), the relationship between complexity time series would be completely misleading. For the first value, the first signal would show the lowest complexity, while for the second value, the third signal would have the lowest complexity.

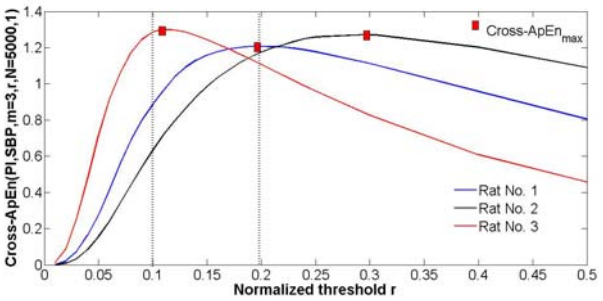

Figure 1. *Cross-ApEn* (PI‖SBP) three NRM rats are shown. The filled rectangle indicates maximum of *Cross-ApEn*. Dotted vertical lines are boundaries for the recommended range of threshold r values

3.3. Formulae and analysis of parameters that determine the r_{TEOR}

To develop formula for automatic selection of r that corresponds to $Cross\text{-}ApEn_{MAX}$, the relationship between the two automatically selected thresholds r_{TEOR} for $ApEn_{MAX}$ of individual signals is examined. Chon and et al. proposed two sets of formulae for automatic selection of r_{TEOR} for $ApEn_{MAX}$ in [6,7]. Slight difference in the formulae was explained as a consequence of differences that appeared in generating random values for experimental data [7]. For experimental data, the formula proposed in [6] demonstrated higher precision in the assessment $ApEn_{MAX}$.

In order to establish the threshold value r that

corresponds to *Cross-ApEn*$_{MAX}$, we tried first arithmetic mean, and then geometric mean of the two r_{TEOR} values for individual signals obtained according to formulae in [6,7]. The results were far from satisfactory. However, we have noticed that threshold r_{TEOR} of the signal that served as template is closer to the true value of r_{MAX} for *Cross-ApEn*$_{MAX}$ than the threshold r_{TEOR} of the signal that served as matching signal in *Cross-ApEn* estimation.

It was noted in [6] that the threshold r that corresponds to *ApEn*$_{MAX}$ for single time series depends on standard deviation of the differential series $u(i)$-$u(i-\tau)$ data length and standard deviation of observed time series $u(i)$. In this paper, we analyze the dependence of the threshold r_{MAX} for *Cross-ApEn*, as a function of all the mentioned parameters that are derived from the two observed time series.

To perform an analysis of signals of diverse complexity, SBP and PI of laboratory animals are observed in baseline (BL) and first stress (FS) condition. To determine the effect of different lengths of time series, we observe subset of time series starting from 1000 samples to 6000, with step 1000. As in the case of *ApEn* in [10], r_{MAX} of *Cross-ApEn*$_{MAX}$ decreases with the increasing length of the data (N). It is noted that r_{TEOR} depends on the standard deviation of differentials of both the series (standard deviation of $u(i)$-$u(i-\tau)$ is denoted as std_u and standard deviation of $v(i)$-$v(i-\tau)$ is denoted as std_v). Figure 2 illustrates the dependence of r_{MAX} of *Cross-ApEn*$_{MAX}$ on the average value of both standard deviation (std_u and std_v) and length N for m=3.

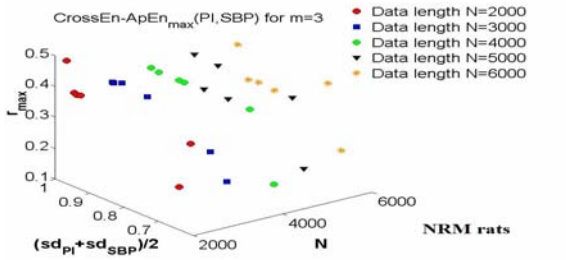

Figure 2. Plot r_{MAX} for *Cross-ApEn*$_{MAX}$ (PI‖SBP) as function of $\frac{(std_u + std_v)}{2}$, and data length N.

As the final result, we propose the expression that describe the dependence of the threshold r_{MAX} that corresponds to *Cross-ApEn*$_{MAX}$ of the aforementioned parameters, for m = 2 and m = 3:

$$m=2:\ r_{TEOR} = T_{max}(x) + \left| -0.02 + 0.23 \cdot \left(\sqrt{\frac{(std_x + std_y)}{2}} \Big/ \sqrt[4]{\frac{N}{1000}} \right) \cdot 0.1 \right| \quad (6)$$

$$m=3:\ r_{TEOR} = T_{max}(x) + \left| \left(-0.06 + 0.43 \cdot \sqrt{\frac{(std_x + std_y)}{2}} \right) \Big/ \sqrt[4]{\frac{N}{1000}} \cdot 0.1 \right| \quad (7)$$

In the Eqs (6) and (7) x,y $\in$ {SBP,PI}, *std* is standard deviation of the differential time series and $T_{max}(x)$ is threshold of a single series (i.e. r_{MAX} for which *ApEn*$_{MAX}$ is evaluated) according to [6].

3.4. Consistency of estimates *Cross-ApEn*

In our recent work [5], it was pointed that flip-flop effect (exchange of the complexity measure of two signals due to different selection of parameters) was present in assessment *ApEn*(r_{TEOR}) of the same signals when their length N is changed. The suspicion that the flip-flop effect is not only a consequence of self-matching has been confirmed in the case of *CrossEn*, where self-matching does not exist.

Fig. 3 (a) shows *Cross-ApEn* (SBP‖PI) in BL and FS condition for the first N = 3000 samples of recorded signal, Fig. 3 (b) presents *Cross-ApEn* (SBP‖PI) for the first 6000 samples of the same signal. Note that maximum value of *Cross-ApEn* (SBP‖PI) in FS condition exceeds maximum value of entropy in BL condition when N is equal to 3000 points, and for N=6000 results is just opposite.

It was also noted in [5] that assessment of the conditional probability that forms a core for *ApEn* estimates, for small values of r was done on a very small number of samples. Jeruchim et al. argued in [12] that if the number of samples that were used for estimation was smaller then $100/p$ (p -probability), it would result in uncertainty of the estimate. Fig. 4 illustrates fulfilment of mentioned criteria in case of a conditional probability estimation necessary to assess the *Cross-ApEn*$_{MAX}$(r_{TEOR}).

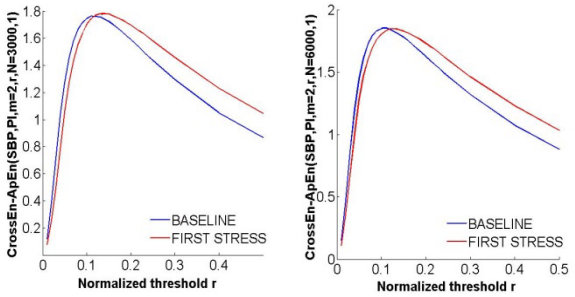

Figure 3. Flip flop effect *Cross-ApEn* (SBP‖PI) of BHR rats

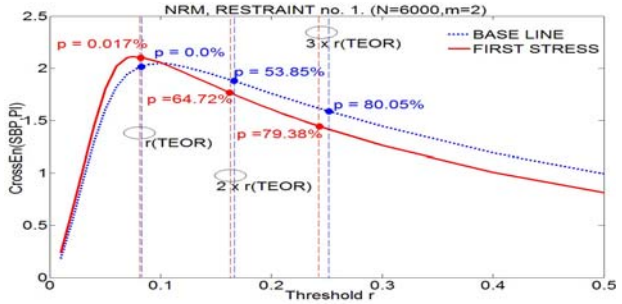

Figure 4. Percentage of statistically reliable probability estimates for threshold values: r_{TEOR}, $2 \cdot r_{TEOR}$ and $3 \cdot r_{TEOR}$.

From Fig. 4 it is obvious that if threshold increases from r_{TEOR} to $3 \cdot r_{TEOR,}$ the percentage of reliable probability estimates increases from 0% to 80%.

The low level of compliance with the conditions indicates that it is necessary to move threshold level towards the higher values in order to get the reliable estimates. The proposed value for threshold r for which the estimates are reliable (r_{con}) is equal to $r_{con}=3 \cdot r_{TEOR}$.

4. Results

To check the accuracy of the proposed formulae for automatic threshold selection for maximum value of *Cross-ApEn* entropy estimation is performed for the range of the value of r (0 – 0.5). Error probability of estimation of *Cross-ApEn* maximum value is calculated as:

$$P_{err} = \frac{CrossApEn_{MAX}(r_{MAX}) - CrossApEn_{MAX}(r_{TEOR})}{CrossApEn_{MAX}(r_{MAX})} \quad (8).$$

The accuracy of the formulae is verified on the long time series of laboratory animals (Table 1).

Table 1. Estimation error P_{err}±SE [%] of *Cross-ApEn* (PI‖SBP) and *Cross-ApEn* (SBP‖PI) for different series length N (long data sets).

N	Number of time series	P_{err} ±SE [%] *Cross-ApEn* (PI‖SBP)	P_{err} ±SE [%] *Cross-ApEn* (SBP‖PI)
2000	96	1.82 ± 0.35	2.35±0.35
3000	94	1.58 ± 0.33	1.42±0.25
4000	72	2.34 ± 0.39	1.55±0.25
5000	68	2.27 ± 0.47	2.17±0.33
6000	64	2.38 ± 0.55	1.85±0.30

Considering the short signals of healthy volunteers (N = 500), the average value of P_{err} for 41 time series pairs is equal to 1.42% for *Cross-ApEn* (PI‖SBP) and 2.34% for *Cross-ApEn* (SBP‖PI).

For threshold values $r_{TEOR,}$ $2 \cdot r_{TEOR}$, and $3 \cdot r_{TEOR.}$, the percentage of probability estimates that fulfilled the criteria for statistical reliability were 5.32% ± 1.81, 63.95% ± 3.94 and 81.37% ± 2.17 respectively for long signals, and 0%, 14.37% ± 2.1 and 52.4% ± 2.69 respectively for short signals.

5. Conclusion

Automatic threshold for cross-entropy is of the same quality as established thresholds for single time series. The calculated percentage of statistically reliable probabilities that form a core for entropy estimates, shows that the threshold level must be shifted towards the higher values.

Acknowledgements

This paper is supported in part by grant TR32040 of Serbian Ministry of Science.

References

[1] Pincus SM, Mulligan T, Iranmanesh A, Gheorghiu S, Godschalk M, et al. Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males. Proc Natl Acad Sci USA 1996;93:14100–14105.

[2] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;278:2039–2049.

[3] Pincus SM, Singer BH. Randomness and degrees of irregularity. Proc Natl Acad Sci USA 1995;93:2083–2088.

[4] Pincus SM. Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA 1991;88:2297–2301.

[5] Boskovic A, Loncar-Turukalo T, Sarenac O, Japundzic-Zigon N, Bajic D. Unbiased entropy estimates in stress: A parameter study. Computers in Biology and Medicine 2012;42:667-679.

[6] Lu S, Chen X, Kanters JK, Solomon IC, Chon KH. Automatic selection of the threshold value r for approximate entropy. IEEE Trans Biomed Eng 2008;55:1966–1972.

[7] Chon KH, Scully CG, Lu S. Approximate entropy for all signals. IEEE Eng Med Biol 2009;28:18–23.

[8] Tarvainen MP, Ranta-aho PO, Karjalainen PA. An advanced detrending approach with application to HRV analysis. IEEE Transaction on Biomedical Engineering 2002;42:172-174.

[9] Carvajal R. Dimensional analysis of HRV in hypertrophic cardiomyopathy patients. IEEE Eng Med Biol 2002;21: 71–78.

[10] Chen X, Solomon IC, Chon KH. Comparison of the use of approximate entropy and sample entropy: application to neural respiratory signal. In: Proceedings of the 27th IEEE EMBS Annual Conference. China: 2005;4212–4216.

[11] Kaffashi F, Foglyano R, Wilson CG, Loparo K. The effect of time delay on approximate and sample entropy calculations. Physica D 2008;237:3069–3074.

[12] Jeruchim MC, Balaban P, Shanmugan KS, In: Wolf JK, Simulation of Communication System Modeling, Methodology, and Techniques. New York: Kluwer academic publishers, 2000:1-907.