УНИВЕРЗИТЕТ У НОВОМ САДУ

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА

ЕНЕРГЕТИКА, ЕЛЕКТРОНИКА И ТЕЛЕКОМУНИКАЦИЈЕ

ПРОЦЕНА РЕГУЛАРНОСТИ И СИНХРОНИЗМА У ПАРАЛЕЛНИМ БИОМЕДИЦИНСКИМ ВРЕМЕНСКИМ НИЗОВИМА

ДОКТОРСКА ДИСЕРТАЦИЈА

Ментор: проф. др Драгана Бајић

Кандидат: Омер Али Мохамоуд

Нови Сад, 2016. године

UNIVERSITY OF NOVI SAD

FACULTY OF TECHNICAL SCIENCES

POWER ENGINEERING, ELECTRONICS AND TELECOMMUNICATIONS

ESTIMATION OF REGULARITY AND SYNCHRONISM IN PARALLEL BIOMEDICAL TIME SERIES

PhD THESIS

Mentor: prof. dr Dragana Bajic

Candidate: Omer Ali Mohamoud

Novi Sad, 2016.

УНИВЕРЗИТЕТ У НОВОМ САДУ

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj: RBR	
Identifikacioni broj: IBR	
Tip dokumentacije: TD	Monografska dokumentacija
Tip zapisa: TZ	Tekstualni štampani materijal
Vrsta rada (dipl., mag., dokt.): VR	Doktorska disertacija
Ime i prezime autora: AU	Omer Ali Mohamoud
Mentor (titula, ime, prezime, zvanje): MN	Prof. dr Dragana Bajić, redovni profesor
Naslov rada: NR	Procena regularnosti i sinhronizma u paralelni biomedicinskim vremenskim nizovima
Jezik publikacije: JP	Engleski

Jezik izvoda: JI	srp. / eng.
Zemlja publikovanja: ZP	Srbija
Uže geografsko područje: UGP	Vojvodina, Novi Sad
Godina: GO	2016.
Izdavač: IZ	autorski reprint
Mesto i adresa: MA	Trg Dositeja Obradovića 6, Novi Sad
Fizički opis rada: FO	broj poglavlja 7 / stranica 89/ slika 21/ tabela 3/ referenci 71/ priloga 0
Naučna oblast: NO	Elektrotehničko i računarsko inženjerstvo
Naučna disciplina: ND	Telekomunikacije i obrada signala
Predmetna odrednica, ključne reči: PO	Entropija, kardiovaskularni vremenski nizovi
UDK	
Čuva se: ČU	U biblioteci
Važna napomena:	

VN	
Izvod: IZ	**Cilj:** Snimanje sopstvenih zdravstveih prametara je postalo deo koncepta mobilnog '*crowdsensing*-a' prema kojem učesnici sa nakačenim senzorima skupljaju i dele informacije, na ličnu ili opštu dobrobit. Međutim, ograničenja u prenosu podataka dovela su do koncepta lokalne obrade (na licu mesta). To je pak nespojivo sa uobičajenim metodama za koje je potrebno da podaci koji se obrađuju budu stacionarni i bez artefakata. Ključni deo ove teze je opis procesorski nezahtevne binarizovane unakrsne aproksimativne entropije (X)*BinEn* koja omogućava analizu kardiovaskularnih podataka bez prethodne predobrade, u uslovima ograničenog napajanja i procesorskih resursa. **Metoda:** (X)*BinEn* je nastao razradom postojećeg postupka unakrsne entropije ((X)*ApEn*). Definisan je nad binarnim diferencijalno kodovanim vremenskim nizovima, razdeljenim u binarne vektore dužine *m*. Za procenu razmaka između vektora koristi se Hemingovo rastojanje, a sličnost vektora se ne procenjuje između svakog vektora pojedinačno, već između skupova vektora. Procedura je testirana nad laboratorijskim pacovima izloženim različitim vrstova stresova i upoređena sa postojećim rezultatima. **Rezultati:** Broj potrebnih procesorskih operacija je značajno smanjen. (X)*BinEn* registruje promene entropije slično (X)*ApEn*. Beskonačno klipovanje je gruba kvantizacija i za posledicu ima smanjenu osetljivost na promene, ali, sa druge strane, prigušuje binarnu asimetriju i nekonzistentnan uticaj parametara. Za određeni skup parametara (X)*BinEn* je ekvivalentna Šenonovoj entropiji. Uslovna binarna m=1 entropija automatski se dobija kao uzgredni product binarizovane entropije, i može da se iskoristi kao komplementarna dinamička mera. **Zaključak:** (X)*BinEn* može da se koristi za jedan vremenski niz, kao auto-entropija, ili, u opštem slučaju, za dva vremenska niza kao unakrsna entropija. Namenjena je mobilnim uređajima sa baterijskim napajanjem za individualne korisnike, to jest za korisnike sa ograničenim napajanjem i procesorskim resursima.
Datum prihvatanja teme od strane Senata: DP	

Datum odbrane: DO	
Članovi komisije: (ime i prezime / titula / zvanje / naziv organizacije / status) KO	predsednik: dr Željen Trpovski, vanredni profesor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu član: dr Nina Japundžić-Žigon, redovni profesor, Medicinski fakultet, Univerzitet u Beogradu član: dr Vojin Šenk, redovni profesor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu član: dr Živko Bojović, docent, Fakultet tehničkih nauka, Univerzitet u Novom Sadu mentor: dr Dragana Bajić, redovni profesor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu

University of Novi Sad

Faculty of technical sciences

Key word documentation

Accession number: ANO	
Identification number: INO	
Document type: DT	Monograph documentation
Type of record: TR	Textual printed material
Contents code: CC	PhD Thesis
Author: AU	Omer Ali Mohamoud
Mentor: MN	Prof. dr Dragana Bajić
Title: TI	Full professor
Language of text: LT	English
Language of abstract: LA	eng. / srp.

Country of publication: CP	Serbia
Locality of publication: LP	Vojvodina, Novi sad
Publication year: PY	2016.
Publisher: PU	
Publication place: PP	Novi Sad

Physical description: PD	Chapters 7 / Pages 89/ Figures 21/ Tables 3/ References 71/ Appendices 0
Scientific field SF	Electrical engineering and computer science
Scientific discipline SD	Communications and Signal Processing
Subject, Key words SKW	Entropy, Cardiovascular time series
UC	
Holding data: HD	
Note: N	

Abstract: AB	**Objectives:** Self-monitoring in health applications has already been recognized as a part of the mobile crowdsensing concept, where subjects, equipped with adequate sensors, share and extract information for personal or common benefit. Limited data transmission resources force a local analysis at wearable devices, but it is incompatible with analytical tools that require stationary and artifact-free data. The key objective of this thesis is to explain a computationally efficient binarized cross-approximate entropy, (X)*BinEn*, for blind cardiovascular signal processing in environments where energy and processor resources are limited. **Methods:** The proposed method is a descendant of cross-approximate entropy ((X)*ApEn*). It operates over binary differentially encoded data series, split into *m*-sized binary vectors. Hamming distance is used as a distance measure, while a search for similarities is performed over the vector sets, instead of over the individual vectors. The procedure is tested in laboratory rats exposed to shaker and restraint stress and compared to the existing (X)*ApEn* results. **Results:** The number of processor operations is reduced. (X)*BinEn* captures entropy changes similarly to (X)*ApEn*. The coding coarseness has an adverse effect of reduced sensitivity, but it attenuates parameter inconsistency and binary bias. A special case of (X)*BinEn* is equivalent to Shannon entropy. A binary conditional *m*=1 entropy is embedded into the procedure and can serve as a complementary dynamic measure. **Conclusion:** (X)*BinEn* can be applied to a single time series as auto-entropy or, more generally, to a pair of time series, as cross-entropy. It is intended for mobile, battery operated self-attached sensing devices with limited power and processor resources.
Accepted on Senate on: AS	
Defended: DE	
Thesis Defend Board: DB	president: dr Željen Trpovski, associate professor, Faculty of Technical Sciences, University of Novi Sad

	member: dr Nina Japundžić-Žigon, full professor,, School of Medicine, University of Belgrade member: dr Vojin Šenk, full professor, Faculty of Technical Sciences, University of Novi Sad member: dr Živko Bojović, assistant professor, Faculty of Technical Sciences, University of Novi Sad mentor: dr Dragana Bajić, full professor, Faculty of Technical Sciences, University of Novi Sad

CONTENTS

ABSTRACT

Health issues are recently included into the mobile crowdsensing concept, in applications both at personal and at community level. The transmission of increased amount of recorded signals is a subject to bandwidth and energy constraints. As a part of computing vs. energy/bandwidth trade-off, crowdsensing applications propose a local analysis, performed in a wearable device. Such an analysis implies computationally efficient, blind and unaided processing, while most of the sophisticated analytical tools, including approximate entropy, have pronounced complexity and require active pre-processing to obtain stationary, noise- and artifacts - free data.

A key component of this thesis is an explanation of a robust binarized modification of approximate entropy that ensures a speedy, energy efficient and blind implementation. Method is tested using the signals from laboratory animals exposed to stress. Same signals have already been used for entropy study, so a standard to which proposed method can be compared to exists, while the physiological aspects of this experiment are well documented. Stress is a major adverse factor in healthy population, and healthy population is the most likely consumer of self-monitoring devices and corwdsensing applications.

The proposed (X)*BinEn* method descends from cross-approximate entropy ((X)*ApEn*). It operates over binary differentially encoded data series, split into the streams of *m*-sized binary vectors. The alternations included to (X)*BinEn* are as follows: the vectors are binary, the distance is measured by Hamming distance and template matching (finding the similar vectors) is performed over the vector sets (cardinality 2^m) instead of over the individual vectors (cardinality *N*). The realization of (X)*BinEn* can be cross-*BinEn* and auto-*BinEn*.

Binary differential encoding is insensitive both to non-stationarities and to a reasonable amount of artifacts, so a blind analysis without the pre-processing is sufficient. It is also an extremely coarse procedure, where signal space is reduced to a simple binary decision whether the next sample increases or decreases in respect to the current one, while the absolute value of signal change is lost. Applying both (X)*ApEn* and (X)*BinEn* to the same signal sets and comparing the results, it was shown that the information located in binary direction of signal changes is sufficient for (X)*BinEn* estimation. Loss of information stored within the signal amplitude reflects in decreased (X)*BinEn* sensitivity, i.e. in inability to capture a small amount of changes, but it also has a positive effect of filtering out unwanted binary bias and parameter instability.

The complexity of (X)*BinEn* procedure is linear, it operates with 2^m histograms and pre-processing is not necessary. Considering (X)*ApEn*, the complexity is quadratic, it operates with N real vectors and pre-processing includes trend and artifacts removal and stationarity check.

It was proven that a special case of auto-*BinEn* is equivalent to Shannon unconditional binary entropy. It was also shown that binary conditional entropy is embedded into the auto-*BinEn* procedure and its appending as a complementary option would be at no processing cost. It enables a dynamic observation of complexity changes related to inter-samples dependency, in respect to the complexity of the same signal but with assumed statistically independency.

(X)*BinEn*, however, is not a substitute for (X)*ApEn*, nor for any of its follow-ups. It is intended for a quick, robust and blind local analysis in unprofessional wearable devices where (X)*ApEn*-like procedures cannot be applied due to their quadratic computational complexity and requirements for semi-automatic pre-processing.

SAŽETAK

Koncept mobilnog '*crowdsensing*'-a uključuje i merenje fizioloških parametara, za individualnu primenu al i za opštu statistiku. Mobilno praćenje svih (ne samo zdravstvenih) parametara podrazumeva prenos velike količine podataka, odnosno zauzeće propusnog opsega i utrošak energije. Aplikacije koje koriste ove usluge su prešle na lokalnu analitiku, rukovodeći se zaključkom da se obradom na licu mesta postižu uštede. Podrazumeva se da obrada na licu mesta može da se radi automatski, da troši malo procesorskog vremena i ne zahteva predobradu snimljenih parametara. Na nesreću, analitičke metode koje se koriste u obradi biomedicinskih podataka (a jedna od njih je aproksimativna entropija) zahtevaju stacionarne signale, bez šuma i artefakata a dodatno imaju visoku kompleksnost.

Ova teza je posvećena binarizovanoj modifikaciji aproksimativne entropije. Modifikacija omogućava brzu, energetski efikasnu i automatsku primenu. Testovi su sprovedeni na laboratorijskim pacovima u stresnim uslovima, za koje referentne vrednosti već postoje. Stres je izabran namerno, jer zdrava ljudska populacija na koju su i uređaji za samostalno permanentno merenje i crowdsensing aplikacije usmereni, po pravilu nema drugih zdravstvenih problema sem stresa.

(X)*BinEn* je nastao razradom postojećeg postupka unakrsne entropije ((X)*ApEn*). Definisan je nad binarnim diferencijalno kodovanim vremenskim nizovima, razdeljenim u binarne vektore dužine *m*. Za procenu razmaka između vektora koristi se Hemingovo rastojanje, a sličnost vektora se ne procenjuje između svakog vektora pojedinačno, već između skupova vektora.

Binarno diferencijalno kodovanje nije osetljivo na nestacionarnosti in a artefakte (u razumnim količinama). Dozvoljava automatsku analizu bez predobrade. Međutim, to je veoma grubo kodovanje koje svodi proctor podataka na jednostavnu binarnu odluku da li naredni odmerak raste ili opada u odnosu na prethodni. Poredeći rezultate (X)*ApEn* i (X)*BinEn* analize istih signala, pokazalo se da je informacija zaostala nakon binarnog diferencijalnog kodovanja dovoljna da registruje (relativne) promene entropije. Neizbežan gubitak informacije (zbog binarizacije) očitava se i smanjenju (X)*BinEn* osetljivosti, tj. u nesposobnosti da registruje sitne promene. Međutim, binarizacija ima i pozitivne efekte, upravo smanjena osetljivost 'filtrira' nepoželjno forsiranje nula i sprečava nestabilnost rezultata pri promeni parametara.

Kompleksnost (X)*BinEn* procedure je linearna, dok je kompleksnost (X)*ApEn* procedure kvadratna.

Za poseban skup parametara, auto-*BinEn* je ekvivalentna Šenonovoj entropiji.

(X)*BinEn,* nije zamena ni *za* (X)*ApEn*, ni za i jedan od njenoh sledbenika. Njena primena je u uređajima za samo-očitavanje vitalnih parametara, gde bi bilo korisno primeniti (X)*ApEn* i slične procedure, ali to njihova visoka kompleksnost i potreba za vizualnim podešavanjem parametara sprečava.

1. Introduction

Approximate entropy (*ApEn*) is among the most exploited nonlinear techniques to quantify the complexity of a time series. The introductory *ApEn* contributions have reached an amazing number of almost 4000 citations. The fascinating increase of *ApEn* quotation rate started to be graphically highlighted in scientific publications. The role of *ApEn* as a supporting tool for biomedical data analysis has been approved by thousands of pre-clinical and clinical studies: e.g. for neural signals – in sleep, in Alzheimer's disease and in epileptic seizures – but more prominently, it was used for long and short cardiovascular signals of various pathologies.

In spite of vast and firmly established implementation, the activities on *ApEn* improvement have never ceased, splitting the research efforts into the two lines. The first line is devoted to the variations of entropy estimation tools, like *SampEn*, *CApEn* , *KNNCE*, *FuzzyEn*, *MultiScaleEn*, *ApEn* based on wave mode and also to the algorithms for speeding up he estimation process. The other research line is dedicated to the problem of parameter choice that influence the value of *ApEn* estimates and may induce biased and inconsistent results. These contributions are predominantly devoted to the threshold selection problems and to the temporal lags and scaling.

The aim of this thesis is to analyze Binarized approximate entropy which is proposed for novel applications, in a surrounding when a constant monitoring of physiological parameters become a way of life, and not only during the sporting activities.

The Thesis is organized as follows: Chapter 2 gives a brief but necessary introduction to cardiovascular signals and the corresponding derived time series. Chapter 3 introduces various forms of Entropy as a regulatory measure, outlining the roles of famous approximate and sample entropies (*ApEn* and S*ampEn*). It also explains the known problems of entropy estimates. Without the pretension to present all of countless entropy improvements, Chapter 4 discusses the major ones, Fuzzy, Multiscale and Conditional entropy and briefly, in Chapter 5, introduces a concept of Joint Symbolic Dynamic and the entropies defined within it.

The Chapter 6 is a core of this thesis, as it discusses the concept of Binarized entropy in crowdsensing and individual environment, and compares it to the classical entropy measures. The concept is tested using artificial data, as well as pulse interval and systolic blood pressure signals recorded from the laboratory rats.

The thesis is a segment of broader study that considers entropy and cardiovascular signals, under the shield of TR32040 research project. A brief recap of the previous contributions are included within the Chapters 2 and 3.

2. Cardiovascular signals and systems

Homeostasis is a term that usually appears in a description of a certain system, regardless of whether the system is opened or closed. The term comes from the phrase *milieu interieur* that was generated in 1865 by Claude Bernard, and it was used in a connection to a living organism. Today, homeostasis can be thought of as the element in the organ system which regulates its inside, trying to keep the system in balance at all times.

When it comes to the system regulation, a lot of changes in balance, and a lot of regulation mechanisms are necessary in order to facilitate homeostasis. Based on the given life system parameter, an organism can be a *regulator*, whose job is to keep the parameter stable despite the variations in the environment, or a *conformer*, whose job is to let the environment determine the parameter. What is good when it comes to the

homeostasis is that it enables an organism to work effectively in a variety of different environmental conditions. Some of the functions that are controlled by homeostasis are blood pressure, heart rate, breathing and temperature.

There are at least three separate elements present in each homeostatic mechanisms. The first element is called *sensory receptor*, and its function is to monitor and react to any kind of changes in the environment. When a change in the environment does occur, the receptor sends the information to the *control center (integrators)* that determines the range at which a variable is kept. Integrators can be seen as control points that collect the information from one or more sensory receptors, and extract the correct response from the effectors. Control centers are usually found in the brain for most of the homeostatic mechanisms. When it comes to muscles and glands, brain is the control center that excites them. Once the *effectors* pick up a signal they influence a change that corrects the deviation that has occurred through enhancing the deviation with positive feedback, or through depressing the deviation with negative feedback.

The purpose of a positive feedback mechanism is to accelerate or enhance the output that has occurred as a result of an already activated stimulus i.e., they are created in order to push levels out of normal ranges. In order to do this, there is a series of events that initiates a cascading process which enlarges the effect of the stimulus. However, in spite of the advantages, this process is not used by the body very often because the acceleration might become uncontrollable. Positive feedback mechanisms are present during the childbirth, when they release oxytocin which intensifies contractions. On the other hand, negative feedback mechanisms are responsible for the reduction of the output or activity of an organ or system to its regular scope of functioning. They occur within systems such as blood pressure regulation.

The disturbance of homeostasis is called homeostatic imbalance, and it can lead to a disease. The decrease of the function of control systems results in destabilization of the internal environment, which is why a system becomes more susceptible to illnesses. The decrease of the function of control systems can also be connected to aging, and physical changes that are usually associated with aging are in return connected to homeostatic imbalance. In some (the most severe) cases, homeostatic imbalance leads to death.

Patients who suffer a heart failure belong to this group because their negative feedback mechanisms get overwhelmed by the destructive power of positive feedback mechanisms.

2.1. Central and peripheral nervous system

The nervous system is an important organ that consists of a network of neurons. The neurons receive, integrate, and transmit the information that coordinates all the actions and body functions. The nervous system is divided into the central nervous system (CNS), and peripheral nervous system (PNS). The CNS consists of the brain, spinal cord, and retina, whereas the PNS consists of sensory neurons, ganglia, and nerves that connect them to each other and to the CNS. The complex neural pathways are responsible for the connections between neural regions, and they allow normal functioning, and information integration and transfer. The enteric nervous system, one of the PNS's subsystems, is one of the few that can function independently, and it controls the gastrointestinal system.

The main function of the peripheral nervous system is to connect the central nervous system to the limbs and organs. It is exposed to all kinds of toxins and mechanical injuries due to the fact that it is not protected by bone or blood-brain barrier like the CNS. It is divided into the somatic nervous system and the automatic nervous system (ANS).

The basic unit of the nervous system is neuron. Neurons usually group together forming nuclei and ganglia, and sheets called laminae which are found in the grey matter or specialized ganglia of the PNS. Neurons transmit information through their axon (a thin fiber that conducts neuron's action potentials to the synaptic junctions, which then transmit the effect through neurotransmitters), which leads to the modulation of the activity of the organ that receives the synaptic information and manifests as inhibition or excitation. Physical stimuli activate the sensory neurons and send a signal to the central nervous system informing the system about the overall body condition and the external environment; motor neurons are responsible for the connections between the nervous system and effector organs, and they are located in the CNS or in peripheral ganglia. The

highest percentage of neurons belongs to the central neurons, which make input and output connections only with other neurons. The interactions of all the neurons from the neural circuit create an organism's perception of the world and its surrounding, and determine organism's behavior.

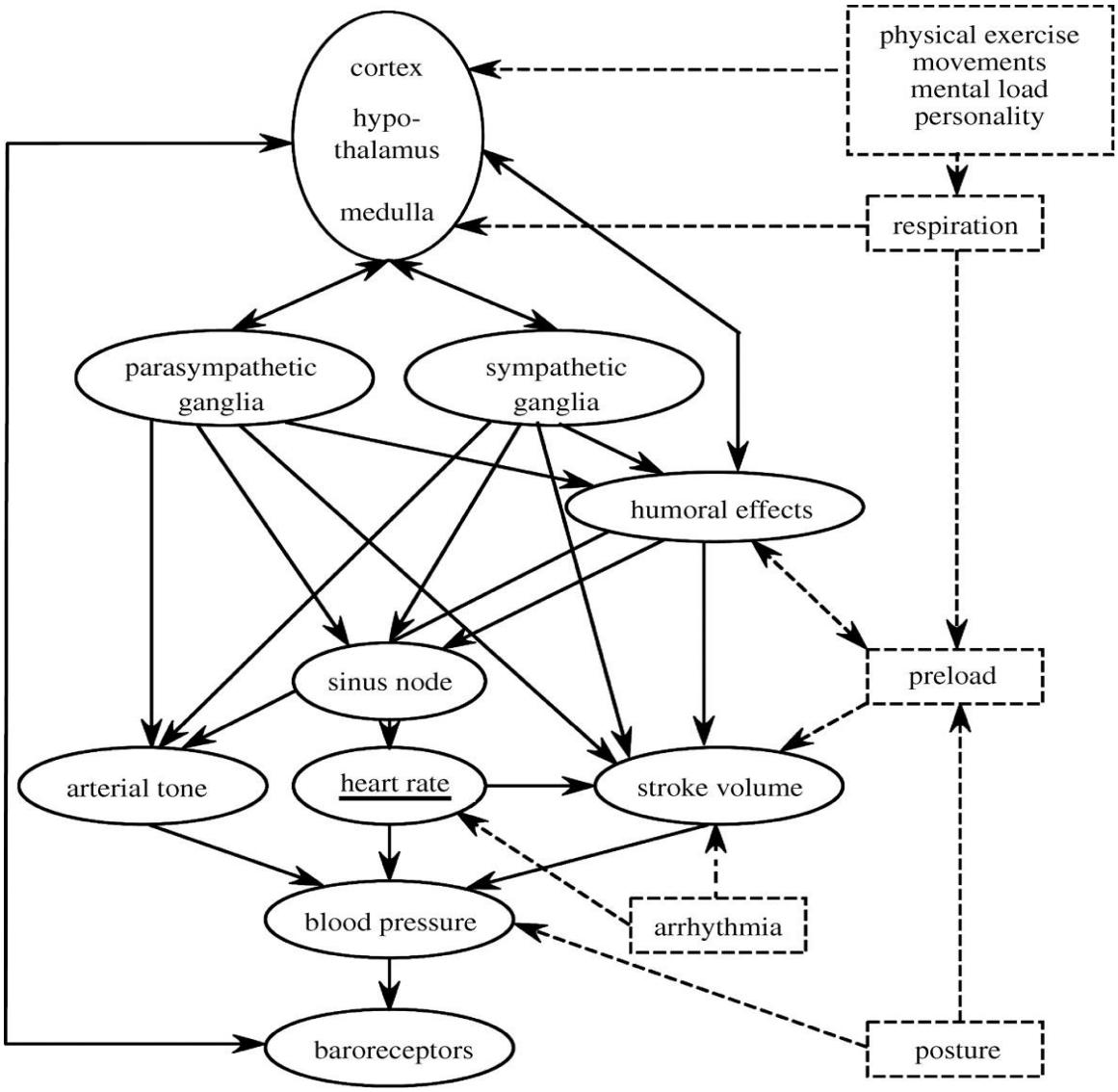


Figure 1.1: Graphic representation of the nervous system

The way in which neurons function can be looked at as firing sequences of action potentials. The action potentials are always treated as identical stereotyped events in

neural coding studies, despite the fact that they can vary in their duration, amplitude or shape. The difference between thin and thick neurons and axons is that thin need less metabolic expense to produce and carry action potentials than thick axons, but thick axons in return convey impulses faster. A lot of neurons have insulating sheaths of myelin around their axons in order to make a balance between the minimal metabolic expense and fast conduction. Peripheral nerves usually have the myelin sheath along the axon in sections about 1 mm long, punctuated by unsheathed nodes of Ranvier, which has a high density of voltage-gate ion channels. The sheath is the part of a neuron that allows fast saltatory action potential propagation from one Ranvier node to the next one. When it comes to the nerve conduction velocity, it can vary from 45-70 m/s in nerve fibers, through 0.2-0.4 m/s in the heart muscle, to 0.013-0.05 m/s in time-delays fibers found between atria and ventricles in the heart.

The information which is transmitted by neurons is assumed under the pattern of action potential production. Measuring and characterization of the way stimulus attributes[1] are represented by neuron action potentials or spikes is studied within neural coding. There are suppositions that temporal (the exact timing of action potential) and also rate coding (the number of action potential produced in time unit) contribute to the nervous system's ability to differentiate between complex objects and produce graceful movements. Due to this, even the smallest change in inter-spike-intervals generates differences in rate and timing of the neurons output, which can be identified as unwanted noise. In the Figure 1.2., you can see the performance of a neuron with low and high amount of noise of the input. The data given in the figure come as a result of the neural model simulation.

2.2. Neurocardiology and autonomic nervous system

Autonomic nervous system (ANS) is divided into sympathetic and parasympathetic (vagal). The main function of the ANS is to preserve homeostasis through the regulation of vial functions. The sympathetic and parasympathetic

[1] Stimulus attributes are light or sound intensity, or motor actions like direction of an arm or leg movement

subsystems have adverse effect on the organs they control, and the balance between them changes in order to achieve the optimal working point depending on the internal and external stimuli. The actual effect the two subsystems have on the organs they control is not universal, but represents the stimulation of one component which leads to the effects that are opposite to those produced by the stimulation of the other component. The parasympathetic nervous system (PSNS) is responsible for keeping the body function stable, whereas the sympathetic nervous system (SNS) is responsible for dealing with stress and it activates fight or fight response.

In the Figure 1.2., you can see the illustrative examples of the neural behavior induced by different levels of noise at the input. Fig.1.2.a, for example, shows convergence of signals onto a single neuron. The noise levels scale in proportion to the number of signals or to its square root, if the noise of signal is perfectly correlated or not, respectively. Fig.1.2.b shows the signal propagation through a series of neurons. In this case, noise levels increase in proportion to the square root of the number of successive neurons. Fig.1.2.c. shows that the network loops result in the build-up of correlated noise.

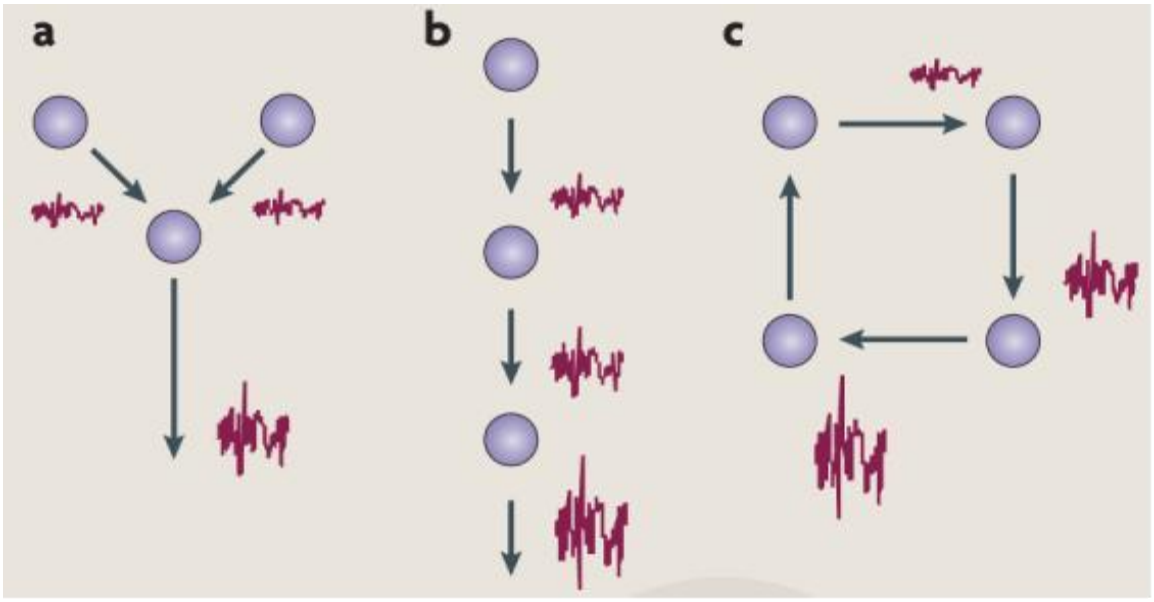


Figure 1.2: The neural network induced by different levels of noise

The SNS is the system that stimulates organs' functioning in most of the organs such as heart. When there is an increase in sympathetic outflow, it influences the increase

in average heart rate, bronchodilation, pupillary dilation, decreased muscle fatigue, elevated blood glucose, etc. The sympathetic nervous system can react in stress induce Figure 1.2: Variability in neuronal firingd situations with global or localized effect. Contrary to this, the PSNS constrains organs' functioning, which is why the increased parasympathetic stimulation leads to a decrease of the average heart rate and stroke volume, systemic vasodilatation, constriction of the pupils, and increased secretion in the digestive tract.

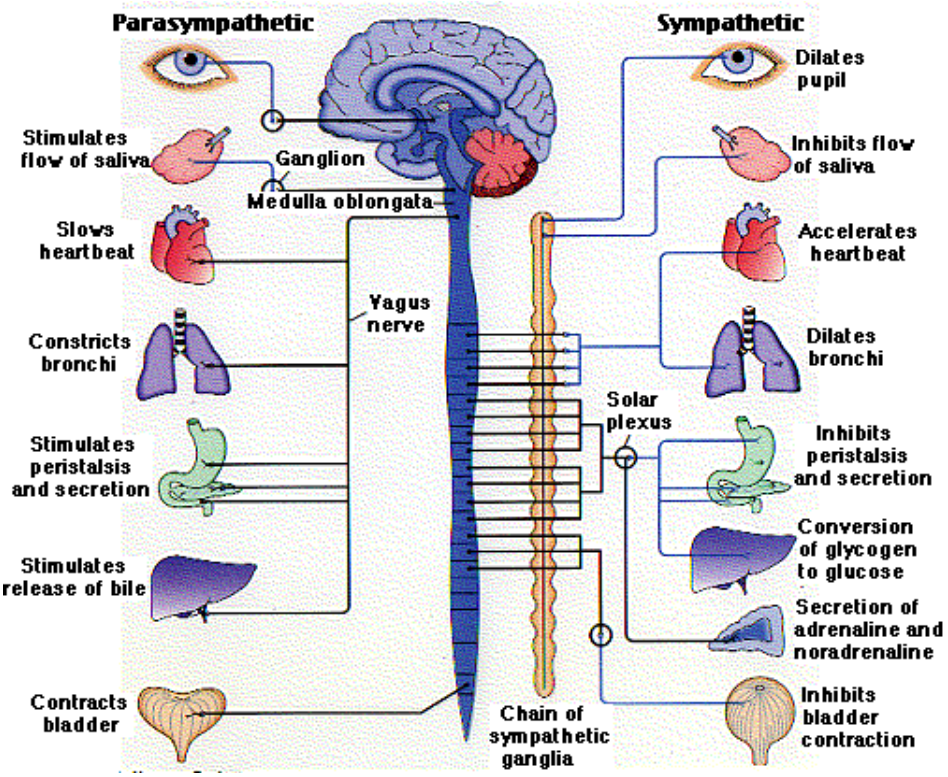


Figure 1.3: Heart innervation

The primary place in brain that regulates sympathetic and parasympathetic outflow to the heart and blood vessels is called medulla, and it is found in the brainstem above the spinal cord. All the autonomic influences that are directed to heart start here. What receives sensory input from different systemic and central receptors such as baroreceptors and chemoreceptors is the nucleus tractus solitaries, or NTS, but medulla also gets information from other parts of the brain such as hypothalamus. The hypothalamus and higher centers are very important for the stimulation of cardiovascular

responses to emotion and stress, because they modify the activity of medullary centers. The heart itself is stimulated by vagal and sympathetic fibers (Figure 1.3). The right vagus nerve stimulates the sino-atrial (SA) node, and the left vagus nerve stimulates the atrio-ventricular (AV) node. Nevertheless, the overlaps in the anatomical distribution may sometimes be great. Atrial muscle is also stimulated by vagal efferent, unlike the ventricular myocardium which is rarely stimulated by vagal efferent. On the other hand, the sympathetic efferent nerves are present in the atria (especially in the SA node), ventricles, and the conduction system of the heart. Heart's reaction time to sympathetic stimulation takes a few seconds and can be considered slow, contrary to the almost instantaneous reaction time to the parasympathetic modulation.

Cardiovascular oscillations can be seen as a window into cardiovascular regulation and underlying regulatory mechanism, but other factors such as vasculature's response, CNS and reflex pathways with arterial baroreflex also influence cardiovascular variabilities (CVV) and frequency bands. Since non-autonomic and nonbaroreflex pathways have an important effect on cardiovascular variabilities, it can be difficult to correlate the changes in the strength of certain oscillation to the change in autonomic control. CVV indicate the neural cardiovascular modulation dynamics, more than average autonomic tone, which is why a simple assessment of CVV parameters will not give a specific measure of the average circulation tone. Nevertheless, the CVV analysis can indicate the features of overall CV regulation, which is measured through adequate approach to the complex interactions between the mechanisms that are involved.

2.3. Cardiovascular oscillations and variability

The cardiovascular system is important for every living organism due to its function of removing the waste materials produced in the cell's working cycle, and delivering the necessary supplies, thus enabling the maintenance of the vital functions. In order to do this, the cardiovascular system consists of highly specialized subsystems which interact with each other, and compete for resources in situations such as the maintenance of the cerebral circulation during a serious hemorrhage. Each of these subsystems has a local regulatory mechanism, but they have to interact with humoral

factors and reflex neural commands, in order to enable the efficient blood delivery. The reflex neural commands occur as result of a change in some of the controlled variables, like blood pressure. The rhythm of the regulatory mechanisms is visible in beat-to-beat recordings, and they produce constant adjustments of cardiovascular variables. These constant variations are called cardiovascular variability (CVV) or cardiovascular oscillations, and they appear in a range from very slow rhythms, to those that are faster than the heart rate.

There are a lot of cardiovascular parameters, but only some of them are easily measured noninvasively or inexpensively. Parameters such as ECG and BP are measured during the standard clinical practice, and they show the short-term variables of heart rate and blood pressure. What is important when it comes to the ECG and BP measuring is that HRV and BPV can be easily derived. The mentioned parameters can be determined using almost any of the available ECG and PB monitoring devices, which is why they are used in every research and clinical study. However, in spite of being easily attainable the two parameters are very complex, and the interpretation and analysis of their values should be done carefully.

Instantaneous heart rate is usually described as reciprocal of the period between two successive heart beats (RR interval, Figure 1.3), and in order to obtain a proper identification of HR, the suggested ECG resolution is at least 4ms, sometimes reduced to 1ms. In certain cases, such as good parameterization of ECG waveform, or measurement of the patterns in ECG that reflect the atria activities (P wave) and ventricular repolarization (T wave), it is necessary to do more accurate measurements. R peak is most often used to represent the fiducial point of HR extraction, which is why the produced time series is referred to as RR-series. The RR series is the duration of

Figure 1.3: ECG (upper line) and the corresponding blood pressure (BP) waveforms (lower line). Distance between the ECG peaks (R peaks) are RR intervals. Local maxima and minima in BP signal correspond to systolic and diastolic BP (SBP and DBP). Note a delay of R peak in respect to the corresponding SBP peak.

successive heart beats and its reciprocal time series. The formula HR=60/RP[s] represents the series of instantaneous heart rate per minute – [bpm]. The heart rate is measured in beats. The linear quantification of heart rate variability can depend on the choice of the observed signal, which comes as a result of nonlinear relationship between HR and RR signals.

When it comes to the BP monitoring, it can be done in a number of ways – through invasive methods, by the use of intra-arterial catheter, or through noninvasive techniques that are based on the Penaz principle which says that "a force exerted by a body can be determined by measuring an opposing force that prevents physical disruption''. BP devices that are used to measure the time series noninvasively are those such as Finapres (FINger arterial PRESsure). They were first used in 1980s, and they were the first devices that allowed reliable noninvasive measurements of beat-to-beat blood pressure signal. Along with the wide-spread use of the finger cuffs, many studies dealing with the accuracy and precision of this kind of measurements started to appear. These studies have proved that the use of finger cuffs is limited, due to the fact that even the slightest changes in finger cuff's positioning and tightness can result in wide variations in readings in the same patient. Because of this, finger cuffs are typically not used as an absolute assessment of blood pressure levels, but they can give a reliable assessment of beat-to-beat BP variability, which is why they are routinely used for monitoring BP regulation during the standard reflex tests in cardiovascular laboratories.

There are three elements that are often extracted from the BP waveform such as systolic blood pressure (SBP), diastolic blood pressure (DBP), and mean blood pressure (MBP). The SBP represents the maximum BP value during each beat; the DBP represents the minimal BP during each beat; whereas the MBP represents the mean pressure between two diastolic time instants. Since the regulation mechanisms of SBP and DBP are somewhat different, they allow a more detailed BP parameterization. In those cases when it is not possible to get anything but the BP, the BP value should be used for estimating the intervals between the consecutive hear beats. This kind of estimation is done through the determination of the intervals between consecutive maxima of the BP wave and resulting time series, which is known as pulse interval series (PI).

The variations in the heart beat that appear during the resting conditions are, in fact, the adjustments of the beat-to-beat control mechanisms. Efferent sympathetic and vagal activities that go to the sinus node are regulated by central and peripheral oscillators, which produce rhythmic fluctuations in efferent neural discharge. This discharge manifests itself in short-term and long-term oscillations in the heart period. The careful analysis of the oscillations can give some information regarding the state and function of the central oscillators, sympathetic and vagal efferent activity, humoral factors, and the sinus node.

The analysis of the frequency at which the HRV and BPV oscillations appear serves to show that HRV's and BPV's periodic oscillation tends to combine into several frequency bands. There are two frequency bands present in the heart rate and blood pressure in humans and animals. Frequency band for the upper respiration is between 0.15 Hz to 0.4 Hz, and it is transmitted through the parasympathetic autonomic branch. The other frequency band is slower, with the range from 0.05 Hz to 0.15 Hz, and it encompasses the dominant 0.1 Hz component in humans. The frequency content can be described as specious dependent because the body's position and size determine the respiration and heart rate. As a result of this, the slow component in dogs is 0.14-0.3 Hz, and 0.4 Hz in rabbits and rats. The slow component is also called as 'Mayer waves" despite the fact that Mayer had actually observed much slower oscillations. In 1951, Gayton and Harris introduced the term *vasomotor waves* to describe these slow oscillations, unrelated to respiration. It is still questionable where these waves exactly come from, but it is clear that they are connected to the synchronous oscillations of efferent sympathetic nervous activity.

In addition to these two dominant frequency bands, there are even slower frequency bands present in humans – very low frequencies (VLF) with the range from 0.0003 Hz to 0.05 Hz, and ultra low frequencies (ULF), but they are not studied much. The VLF usually reflect thermoregulatory cycles and plasma renin activity, however, it still has not been studied enough, and it is not advisable to interpret its values when they are taken from recordings shorter than five minutes. On the other hand, the ULF include circadian rhythms that reflect determinants like changes in activity, posture, breathing, autonomic outflow, state of personal arousal and a number of behavioral variables.

However, despite the possible clinical applications, slower rhythms have not been described fully, and it is still unknown where they originate and what their physiological correlates are.

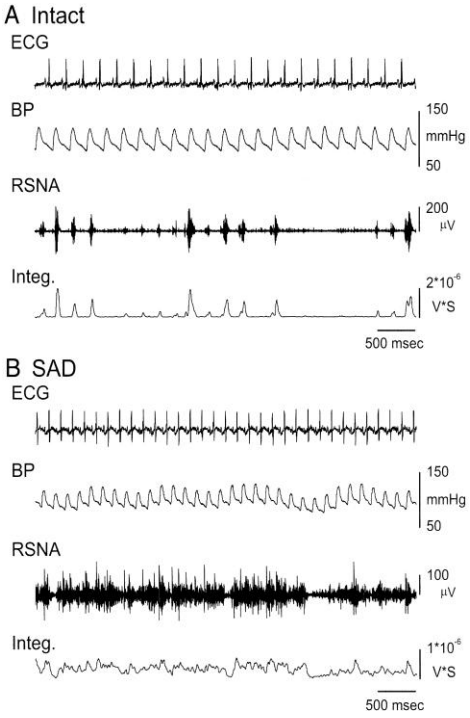


Figure 1.4: ECG and BP waveforms coupled by renal sympathetic neuronal activity.

2.4. References

[1] A. Aldo Faisal, Luc P. J. Selen, and Daniel M. Wolpert, Noise in the nervous system. Nat Rev Neurosci. 2008 Apr; 9(4): 292–303.

[2] Andreas Voss, Steffen Schulz, Rico Schroeder, Mathias Baumert and Pere caminal, Methods derived from nonlinear dynamics for analysing heart rate variability. Phil. Trans. R. Soc., Vol. A367, pp. 277–296, 2009.

[3] Simon C. Malpas, Neural influences on cardiovascular variability: possibilities and pitfalls. Am J Physiol Heart Circ Physiol 282: H6–H20, 2002.

[4] T. Kunitake, H. Kannan, Discharge pattern of renal sympathetic nerve activity in conscious rabbits, Journal of neurophysiology, Vol. 84, no 6, pp 2859-2867, 2000.

[5] Tatjana Loncar Turukalo, Modeling the Temporal Sequence Parameters in Surrogate Data of Cardiovascular Oscillations. PhD thesis, Faculty of Technical Sciences, University of Novi Sad. 2011.

3. Entropy as a regulatory measure

3.1. Estimating entropy from a time series

There are partly random chains of symbols, such as s_1, s_2, s_3…. that can come up in almost all sciences. These chains are drawn from a finite alphabet (like spins in one-dimensional magnets, written texts, DNA sequences, geological records of orientation of the magnetic field of the earth, bits of storage and transmission of digital data), and here we will talk only about them. However, it should be noted that our conclusions can also be applied to countable alphabets, too. The most important thing when it comes to these finite alphabets is the degree to which the sequences can be 'compressed' without the loss of important information, which was first noticed by Shannon (1949) in a connection to a probabilistic context. Shannon (1949) was the one who managed to prove that *entropy* (or average information content) h represents the relevant quantity, and that it corresponds to the thermodynamic entropy of the spin degrees of freedom in the case of magnets. If we are talking about complex and long range correlations, the estimation of the entropy can be very demanding. In order to do the estimations well, it is necessary to fully understand the correlations for optimal compression and entropy estimation, since the h estimates will also be the measure of the degree to which the structure of the sequence is understood.

We would also like to see how this applies to chaotic dynamical systems. To do this, we start with the time series x_t, t-1, …..,N, where time is discretized, but where x_t is continuous. For reducing this to the case mentioned above, it is necessary to discretize x_t by defining a partition $\mathcal{P}_\epsilon$ in the phase space where all elements have diameter of $< \epsilon$ The time series are represented with the s_1, s_2 . . . , s_t, . . . , Here, $S_t = \sigma$, which means that x_t is in the σ-th element of $\mathcal{P}_\epsilon$, which causes 'symbolic dynamics'. This kind of entropy converges to a finite non-zero value h_{KS} (metric entropy) for $\epsilon \longrightarrow 0$, in those cases when the system that generates the time series is chaotic (Kolmogorov , 1958, 1959; Sinai,

1959). All of this makes the measuring of the entropy of a symbolic dynamics extremely important if we want to see whether a system is chaotic or not. What is more, Kolmogorov (1958, 1959) and Sinai (1959) claim that it is not even necessary to take the $\epsilon \longrightarrow 0$ limit, due to the fact that there are 'generating' partitions (eventually infinite, but countable) whose entropy is h_{KS}. However, in most cases, generating partitions are not known for chaotic systems. The exceptions to this rule are $1 - d$ maps where all the partitions into monotonic laps are generating. In the case of Henon map $(x,y)\rightarrow(1.4 - x_2 + 0.3y,x)$, there is no strict construction of a generating partition, in spite of the fact that there are strong indications of a heuristic argument based on homoclinic tangencies that leads to a correct result.

Estimating an entropy can be very demanding, even in those cases when there are no problems in finding a good partition. The difficulties always arise when there are strong long range correlations, but they in turn help us get higher compression rates, due to the fact that they reduce the uncertainty of yet unseen symbols. However, finding those correlations, and taking them into account can turn out to be difficult due to the exponential increase of the number of different blocks ("words") of symbols with the block length.

This problem can be somewhat overcome in the case of natural languages through the use of subjective modes, which work because of the fact that humans know the structure of their own language well enough to have an opportunity the guess most of the information that is given on single missing letters. In the past few years, algorithms that are based on this idea have been improved a lot, but they are still not reliable, or fast, and they are still dependent to natural languages.

The most truthful and objective method that can be used is comprised of counting the frequencies of all blocks up to a certain length and estimating the probabilities from them. Given an alphabet with d symbols, for example, it most often breaks down when $d^m \approx N$ (in which case n represents the block length), which gives $n \approx 3 - 4$ for written English. It is clear that there significantly longer correlations in this type of cases (orthographic, syntactic, semantic) that cannot possibly be taken into account in this way, and which further lead to the overestimation of h in case that sufficient case is used. In

case that we do not use a lot of care, it is possible to get underestimations, as you will see below.

New and enhanced methods that into account that what one wants to know is in a connection to those long blocks, which have high probability. Because of this, the correlations are to be regarded selectively, depending on how important they are. Methods based on Lempel-Ziv coding are the best known methods that work in this way. In them, them string is coded only by breaking it into non-overlapping "words", and the code length used for specifying the string of words is an upper bound for Nh. The most efficient way of doing this is by preparing a "dictionary" of words in the form of a prefix tree. The prefix tree is actually a rooted tree where every single relevant word is attached to a leaf in a way that branch that is common to any two leaves corresponds only to their longest common prefix. In other words, word y is a prefix of another word x only if the x word can be obtained by concatenating one or more letters to y. Grassberger (1989) and Sheilds et all (1992) have also studied an estimator of h that is based on similar ideas and uses similar trees, but, unlike here, does not give upper bounds.

The last type of method that will be mentioned here is related to a method for earning maximal long time profit in a game where one has d different options at any time step. Because of this, these methods are usually referred to as "gambling" methods. The capital is placed on option i, then multiplied by d, in the case that the last option is actually realized, but the money placed on all the other options is lost. The optimal strategy of this method consists in sharing the total capital K_1 at time t among all options a according to their probability $p(a)$. The expected gain in this method depends on the entropy - a factor $d|e^h$ per time step.

At this moment, it would be necessary to take a step back in order to show that what has been mentioned above is not just a superficial coincidence, but a real efficient algorithm that can be used for the estimation of h. As it can be found in the literature, Shannon theory is entirely based on probabilistic concepts and deals only with *average* code lengths. Modern literature on information theory is, on the other hand, mostly concerned with *individual* sequences and the estimation of their shortest codes. The that main flaw of Shannon theory seems to be the fact that it does not take into account the

information that is needed for the description of the probability distribution itself. However, this is completely irrelevant in Shannon theory whose main purpose is the transmission of very long sequences with moderately complex constrains, because the description of the distribution is significantly shorter than the description of the actual string. However, this claim is not correct.

Kolmogorov, Chaitin and others have been led by their desire to completely exclude probabilistic ides from the equation, which is how they came to algorithmic information theory. Based on the same ideas, but a more practical point of view was given by Rissanen himself. His theory is called minimum description length (MDL) principle, and he states that: "a 'good' encoding of a string should be one which minimizes the *total* code length" (Rissanen, J., 1994). This theory corresponds to *Occan's razor* - a good theory is a short theory, when it is applied to a string which itself is a description of some physical phenomenon.

Rissanen applied the MDL principle to entropy estimation and coding, and termed the resulting method - context algorithm. His point of view will later be discussed in more details. At this point, it is just worth mentioning that there is no need to be as profound as Chaitin and Rissanen, and that it is actually possible to merge these with probabilistic ideas. Bell et all (1990) offer the best example of this. In fact, this type of algorithms is implemented in most modern text compression routines.

3.2. Block entropies

Here, we will look at the one-sided infinite sequences $s_1, s_2, \ldots$ where $s_t \in \{0,1,\ldots,d-1\}$. In the majority of examples, we plan to deal with d=2, but everything is also true for d>2 with only small modifications. It is assumed that these are realizations of a stochastic process $\mathbf{s}_1$, $\mathbf{s}_2$,. . . with probabilities

$$p_t(s_1,\ldots,s_n) = prob\{s_{t+1} = s_1,\ldots,s_{t+n} = s_n\} \quad (3.1)$$

These probabilities are often considered as stationary, in which case we can put the index t on $p_t(s_1, \ldots, s_n)$ and define *block entropies*,

$$H_n = -\sum_{s_1,\ldots,s_n} p(s_1,\ldots,s_n)\log p(s_1,\ldots,s_n) \quad (3.2)$$

These entropies measure the average amount of information contained in a word of length n.

The different entropies,

$$h_n = H_n - H_{n-1} = -\sum_{s_1,\ldots,s_n} p(s_1,\ldots,s_n)\log p(s_n|s_1,\ldots,s_{n-1}) \quad (3.3)$$

Provide us with some new information about the n-th symbol in case that the preceding (n -1) symbols are known. Here, $p(s_n|s_1,\ldots,s_{n-1})$ represnts the conditional probability for $\mathbf{s}_n$ being s_n, conditioned on the previous symbols $s_1, \ldots, s_{n-1}$. The Shannon entropy is

$$h = \lim_{n\to\infty} h_n \quad (3.4)$$

In case that all correlations and constrains are taken into account, the equation measures the average amount of information per symbol. The approach to the limit is monotonic, that is all h_n are upper bounds on h.

The first step of calculating h from a finite sequence of length N is usually the estimation of all word probabilities $p(s_1, \ldots, s_n)$ up to some fixed n through the standard likelihood estimate,

$$\hat{p}(s_1,\ldots,s_n) = \frac{n_{s_1\ldots s_n}}{N} \quad (3.5)$$

In this equation $n_{s1\ldots sn}$ represents the number of occurrences of the word $s_1, \ldots, s_n$. Technically speaking, the denominator should only be $N - n + 1$, but this difference is trivial. From the equations above, it is further possible to calculate $\widehat{H}_n$ by including them into the equation $H_n = -\sum_{s_1,\ldots,s_n} p(s_1,\ldots,s_n)\log p(s_1,\ldots,s_n)$. In the end, the estimator $\hat{h}$ can be calculated by computing h_n and extrapolating, or as $\hat{h} = \lim_{n\to\infty} \widehat{H}_n / n$.

In practice, a lot of difficulties can arise if N is not very large and the limit in the fourth equation is not reached fast. These difficulties usually appear due to the fact that, in those cases when it is necessary to determine $p(s_1, \ldots, s_n)$ correctly, the number of different possible words of length n exponentially increases along with n, as well as the necessary minimum length N of the sample sequence. What this means is that the probability estimates $\hat{p}(s_1,\ldots,s_n)$ (used for the determination of $\widehat{H}_n$ estimates), already go through strong oscillations for moderate block lengths n. This usually leads to the underestimation of $\widehat{H}_n$estimates. These calculations can be easily understood by looking

at the expected value of $\widehat{H}_n$. Using the Kullback-Leibler (1951) inequality, we come to the following calculations:

$$\langle \widehat{H}_n \rangle = \langle -\sum_{s_1,\ldots,s_n} \frac{n_{s_1,\ldots,s_n}}{N} \log \frac{n_{s_1,\ldots,s_n}}{N} \rangle \leq \langle -\sum_{s_1,\ldots,s_n} \frac{n_{s_1,\ldots,s_n}}{N} \log p(s_1,\ldots,s_n) \rangle =$$

$$-\sum_{s_1,\ldots,s_n} \frac{\langle n_{s_1,\ldots,s_n} \rangle}{N} \log p(s_1,\ldots,s_n) = H_n \qquad (3.6)$$

A systematic underestimation of $\widehat{H}_n$ will exist as long as there are oscillations.

Harris (1975) gives us a detailed calculation of the expectation value of $\widehat{H}_n$ up to second order in *N*:

$$\langle \widehat{H}_n \rangle = H_n - \frac{M-1}{2N} + \frac{1}{12N^2} \times \left(1 - \sum_{p(s_1,\ldots,s_n)>0} \frac{1}{p(s_1,\ldots,s_n)}\right) + \mathcal{O}(N^{-3}) \qquad (3.7)$$

In this particular equation *M* represents the number of blocks (s_1, . . . ,s_n) with $p(s_1, \ldots, s_n)>0$. It is a direct way to correct for the leading $\mathcal{O}(1/N)$ bias in $\widehat{H}_n$ (the second term on the rhs), since the number of different observed words usually turns out to be a good predictor for *M*. Herzel (1988), and Caswell and Yorke (1986) also use $\mathcal{O}(1/N)$ correction term. $1/N^2$ includes unknown probabilities $p(s_1, \ldots, s_n)$, and it can never be calculated accurately. In this term, it is not sufficient to replace the probabilities by $\hat{p}(s_1,\ldots,s_n)$.

Grassberger (1988) tried to make an alternative approach where only observables appear in the correction terms. He assumed that each $n_{s_1,\ldots,s_n}$ is a random variable in itself, and that it should follow a Poisson distribution in case $p(s_1,\ldots,s_n) \ll 1$. This further leads to an asymptotic series in which higher order terms can be used only for increasingly large *N*. Modified Grassberger's equation, based on what has just been mentioned, is:

$$\widehat{H}_n = \sum_{j=1}^{M} \frac{n_j}{N} \left(\log N - \psi(n_j) - \frac{1}{n_j} \frac{(-1)^{n_j}}{n_j+1}\right) \qquad (3.8)$$

In this equation, index *j* counts the blocks (s_1, . . . ,s_n) for which $n_{s_1,\ldots,s_n}>0$, and $\psi(x)$ represents the logarithmic derivative of the gamma function. As it can be seen, the leading $\mathcal{O}(1/N)$ correction is the same as in the equation 2.7.

What we wanted to do next is to apply the calculations from the equation 2.8 and the native estimate based on the equations 2.3 and 2.5 to the Henon map with standard parameters $x_{n+1} = 1 + 0.3x_{n-1} - 1.4x_n^2$. In order to convert this equation into a bit sequence, we used a binary partition like Grassberger and Kantz (1985). We then compared the results to the Lzapunov exponent, which is determined by iterating the dynamics. (Thanks to Pesin's identity (Ruelle, 1981), we knew that the positive Lyapunov exponent of the Henon map corresponds to the entropy.) As it can be seen in Figure 2.1 (below), the convergence of the truncated entropy $\hat{h}_n = \hat{H}_n - \hat{H}_{n-1}$ is much faster than the more conservative estimate $\hat{H}_n/n$. When it comes to long block lengths *n*, the underestimation is very significant for small data sets, even in those cases when equation 2.8 is used.

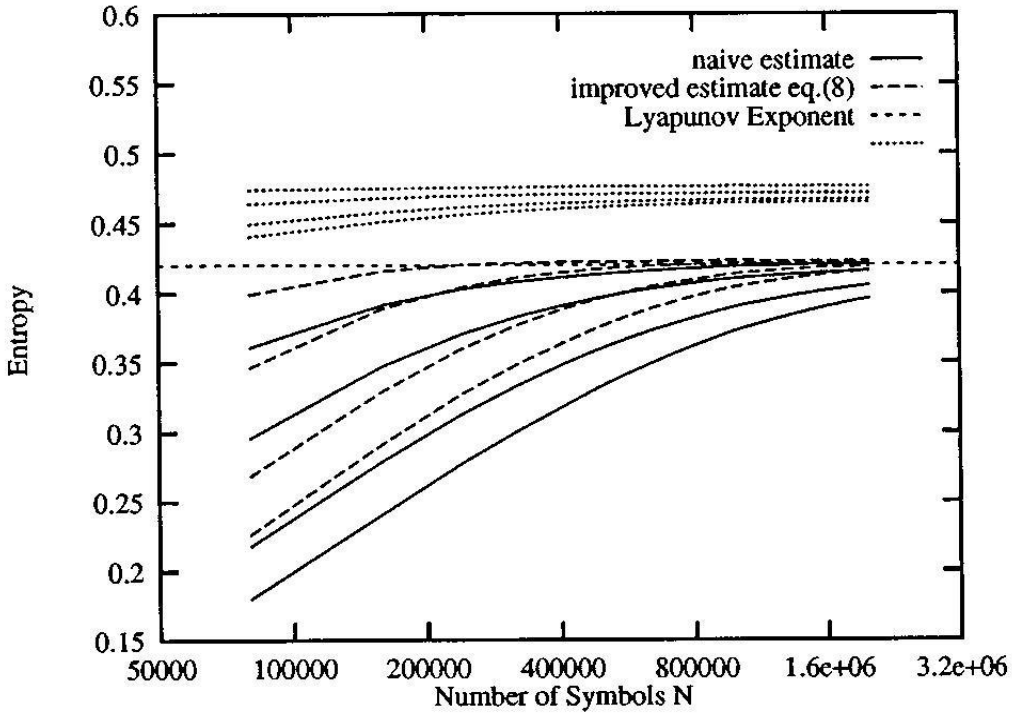


Figure 3.1: The convergence of entropy $\hat{h}_n = \hat{H}_n - \hat{H}_{n-1}$

Source: T. Schu¨rmann and P. Grassberger: Entropy estimation

In those cases where non-zero frequencies $n_{s_1,\ldots,s_n}$ are $\gg 1$ in average, that is, in those cases where $N \gg M$, it can be useful to use corrections made to the native block entropy estimators given in the equation 2.8. Other estimators that are supposed to work for $N \gg M$ have been proposed by Ebeling et al (1992, 1993, 1994). In these papers authors make explicit assumptions regarding $p(s_1, \ldots, s_n)$ that are insufficiently big to be estimated through their occurrences. However, despite the fact that these assumptions are motivated by the McMillan theorem, we do not believe that this method is reliable enough, given the fact that there are no means of checking the assumptions, and the results actually depend on the checkup. Kantz and Schumann (1996) propose a similar, but safer approach. They have found that it is possible to get surprisingly robust results by simply neglecting the information in the small $n_{s_1,\ldots,s_n}$.

The neglect of the information needed to determine the probability of distribution is related to the fact that $\widehat{H}_n$ underestimates H_n. However, this relationship is not that straightforward. In those cases where the distribution is simply given in the form of the integers $n_{s_1,\ldots,s_n}$, the necessary code length can be calculated as $\leq d^n \log N$, where we consider all $n_{s_1,\ldots,s_n}$ as $\leq N$ Including all these into the total information will give us a safe upper bound for H_n. However, this bound will not work well in a broader sense, due to the fact that we did not make any kind of effort to encode the probability distribution efficiency. For this reason, we will not talk about this particular subject anymore. Instead of that, we will focus on the alternative methods that can also give us upper bounds for h, because they can give us all the information needed for unique coding's.

In the end, we just like to mention that it is also possible to replace the likelihood estimator $\hat{p}$ by some other estimator. One of those other possible estimators is Laplace's successor rule:

$$\hat{p}(s_1, \ldots, s_n) = \frac{n_{s_1,\ldots,s_n}+1}{N+d^n} \qquad (3.9)$$

By including this calculation into the equation 2.2 will always give us a bigger estimate of H_n than the likelihood estimate, equation 2.5. On the other hand, it is not always closer to the correct value, especially in those cases when the true probabilities are far from the equipartition. Similary to equation 2.5, the equation 2.9 also gives a

biased estimator of H_n, ($\langle \widehat{H}_n \rangle \neq H_n$). Because of this, it seems that the best calculation consists of using the equation 2.3 with $p(s_n|s_1, \ldots s_{n-1})$ replaced by the Laplace estimator, but keeping the likelihood estimator for $p(s_1, \ldots, s_n)$:

$$\hat{p}(s_n|s_1, \ldots s_{n-1}) = \frac{n_{s_1,\ldots s_{n-1}}+1}{n_{s_1,\ldots s_{n-1}}+d} \tag{3.10}$$

However, this equation is also not biased, and it does not give upper and lower bounds on h_n.

In those cases when we have a good prior estimate for the (distribution of) $p(s_1, \ldots, s_n)$, it is possible to use the Bayesian estimate of h_n developed in Wolpert and Wolf (1995). Nevertheless, this equation can also lead to systematic errors in either direction if bad prior estimates are used.

3.3.Approximate and Sample entropy

An important and effective approach used for the understanding of biological systems is Nonlinear Dynamical Analysis. However, the problem with this approach is that it usually requires extensive and almost unobtainable data sets for its calculations. In his articles (1991, 1995) Pincus gives us the theory and method for a measure of regularity which is closely connected to the Kolmogorov entropy. In this theory, he talks about the rate of appearance of new information, and how it can be used on typically short and noisy time series of clinical data. A branch of statistics called approximate entropy (ApEn) is based on the works of Grassberger and Procaccia (1983), and Eckmann and Ruelle (1985), and it has been used in clinical cardiovascular studies for certain calculations.

The approximant entropy deals with time series for similar epochs where more frequent and more similar epochs cause lower ApEn values. According to this method, if we have the N points, the family of statistics (M, r, N) is almost equal to the negative average natural logarithm of the conditional probability that two sequences with similar m points remain in the same relation to each other, within a tolerance r, at the next point. This means that low ApEn value denotes a high degree of regularity. What is important here is that, thanks to the work of Eckmann and Ruelle, the algorithm counts each

sequence as matching itself, avoiding the occurrence if ln (0) in the calculations. This fact raised many debates regarding the bias of the ApEn, and the practice shows that the bias certainly causes problems because it leads to the omission of two important properties in the ApEn – it makes the approximate entropy highly dependent on the record length and lower that it is expected for short records, and it lacks relative consistency. What this means is that if ApEn of one data set is higher than another, it ought to stay higher for every condition tested, but this does not happen. The lack of this kind of consistency is greatly important because the approximant entropy is meant to be used as a relative measure for the comparison of two data sets.

In order to scale down the bias, a new branch of statistics called sample entropy (SampEn) has been devised. It is based on the theories and approaches of Grassberger abd his co-workers, and in contrast to approximant entropy SampEn does not count self-matches. SamEn (*m, r,* N) represents the negative natural algorithm of the conditional probability that two sequences with similar *m* points remain in the same relation to each other at the next point, where self-matches are not included in the calculation of the probability. This means that lower SampEn value points to the more self-similarity in time series. The sample entropy algorithm is also simpler than the approximant entropy algorithm, and it requires 50% less time to do the calculations. SampEn is an algorithm independent of record length, and is mostly consistent in those circumstance where ApEn lacks consistency.

In recent years, a new technique for the analysis of two related time series has been introduced. The cross-ApEn technique is used for measuring the degree of asynchrony, and the only difference between the ApEn and cross-Ap-En is the fact that cross-ApEn compares sequences from one series with those of the second, which is how it avoids the appearance of bias that comes as a result of self-matches. But, it creates a different kind of problem, since it requires that each template generates a defined, nonzero probability. This means that each template needs either to find minimum one match for $m + 1$ points, or a probability is assigned to it based on the "correction" strategy. The problem with this is that conclusions about relative synchrony of pairs of time series depend on the unguided selection of analysis schemes. In contrast to this, cross-SampEn is defined as

long as one template finds a match, and it stays more or less consistent in those conditions where cross-ApEn lacks the required consistency.

Compared to the ApEn, SampEn statistics have reduced bias, and they have been devised specifically for the reason of eliminating the bias caused by self-matching. The SampEn is used for time series data that are sampled from a continuous process, which is how it got its name. In addition to reducing the bias, the SampEn algorithm can also be applied to sample statistics for the evaluation of results, in the way explained below.

The two most important differences between the ApEn and SampEn are the facts that SampEn does not count self-matches, nor a template-wise approach during the estimation of conditional probabilities (it only needs one template to find a match of legth m + 1). When it comes to the discounting of self-matches, it was based on the fact that entropy assumed as a measure of the rate of information production, what makes the data comparison pointless. In addition to this, Grassberger and his co-workers also dismiss the self-matches in their later work.

Grassberger and Procaccia defined the average of $C_i^m(r)$ as $C^m(r) = (N - m + 1)^{-1} \sum_{j=1}^{N-m+1} C_i^m(r)$. The difference between $C_i^m(r)$ and $\phi^m(r)$ is in the fact that $\phi^m(r)$ represents the average of the natural logarithms of the $C_i^m(r)$, whereas Grassberger and Procaccia suggest approximating the Kolmogorov entropy of a process that can be seen a time series by $\lim_{r\to 0} \lim_{n\to\infty} \lim_{N\to\infty} -\ln[{C^{m+1}(r)}/{C^m(r)}]$. They also count the self-matches and ${C^{m+1}(r)}/{C^m(r)} = (N - m + 1) \sum_{i=1}^{N-m} A_i / (N - m) \sum_{i=1}^{N-m+1} B_i$.

However, when limits are counted in this way, they become inadequate for the analysis of finite time series with noise. In order to resolve this issue, two alterations have been adapted. The first step was to follow limits' later practice in calculating correlation integrals, without the use of self-matches the computation of $C^m(r)$, and then to consider only the first $N - m$ vectors of length m, making sure that for $1 \le i \le N - m, x_m(i)$ and $x_{m+1}(i)$ were defined.

In this work, the $B_i^m(r)$ was defined as $(N-m-1)^{-1}$ times the number of vectors $x_m(j)$ within the tolerance *r* of $x_m(i)$, where *j* ranges from 1 to $N-m$, and $j \neq i$ in order to eliminate the self-matches. After that, the $B_i^m(r)$ was defined as $B^m(r) = (N-m)^{-1}\sum_{i=1}^{N-m} B_i^m(r)$, and $A_i^m(r)$ was defined as $(N-m-1)^{-1}$ times the number of vectors $x_{m+1}(j)$ within *r* of $x_{m+1}(i)$, where *j* can have a range from 1 to $N-m$ $j \neq i$, and set $A^m(r) = (N-m)^{-1}\sum_{i=1}^{N-m} A_i^m(r)\,.\,B^m(r)$. In this case $B_i^m(r)$ represents the probability that two sequences will match for *m* points, and $A_i^m(r)$ represents the probability that two sequences will match for *m* + 1 points. This leads to the definition of the parameter SampEn(*m, r*) as = $\lim_{N\to\infty}\{-\ln[A^m(r)/B^m(r)]\}$, which is estimated by the statistic $\mathrm{SampEn}(m,r,N) = -\ln[A^m(r)/B^m(r)]$. In those case where parameter *r* and the length *m* of the template vector are clear, B = $\{[(N-m-1)(N-m)]/2\}B^m(r)$ and *A* = $\{[(N-m-1)(N-m)\}/2\}A^m(r)$, where *B* represents the total number of template matches of length *m* and *A* represents the total number of forward matches of length *m* + 1. It should also be mentioned that $A/B = [A^m(r)]/B^m(r)]$, which means that SampEn(*m, r, N*) can be expressed as –ln (*A/B*).

The *A/B* quantity is the conditional probability that two sequences within a single tolerance *r* for *m* points remain within *r* of each other at the next point. Contrary to the ApEn(*m, r, N*) which uses a template-wise way for the calculation of probabilities, SampEn(*m, r, N*) calculates the negative logarithm of a probability associated the time series as a whole. The only time when Samp(*m, r, N*) is not defined, is when *B* = 0, in which case the regularity could not be identified, or when *A* = 0 that corresponds to a conditional probability of 0 and an infinite value of SampEn(*m, r, N*). $2[(N-m-1)(N-m)]^{-1}$ represents the lower nonzero conditional probability that the algorithm can report. This means that the upper bound of the statistic SampEn(*m, r, N*) is ln (*N –m*) + ln (*N – m – 1*) – ln, almost doubling ln (*N – m*) , the dynamic range of ApEn(*m, r, N*).

Approximate and sample entropy have been thoroughly explained overall in the work (1-3), and will be only summarized here. When it comes to the time series $[x(j)], j = 1, \dots, N$, the length vector (m) and a distance between any two vectors $X_m^{(i)}$ and $X_m^{(j)}$ are determined in the next way:

$$X_m^{(i)} = [x(i), x(i+\tau), \dots, x(i+(m-1)\tau)] \qquad , i = 1, \dots, N-(m-1)\tau \qquad (3.11)$$

$$d_m\left(X_m^{(i)}, X_m^{(j)}\right) = \max_{k=0,\dots,m-1}[|x(i+k\tau) - x(j+k\tau)|] \quad (3.12)$$

The formula for calculating the probability $C_i^m(r)$ of whether a vector $X_m^{(j)}$ is within the distance *r* from the template $X_m^{(i)}$ is:

$$C_i^m(r) = \frac{B_m^r(i)}{N-(m-1)\tau} \quad (3.13)$$

In this formula, $B_m^r(i)$ represents a number of vectors whose distance is lower than the predetermined threshold value *r*, and where I{.} represents an indicator function:

$$B_m^r(i) = \sum_{j=1}^{N-(m-1)\tau} I\{d_m(X_m^{(i)}, X_m^{(j)}) < r\}, \qquad i = 1, \dots, N-(m-1)\tau \quad (3.14)$$

After this, the same procedure is done again in order to calculate the vector of the length *m+1*, and the approximate entropy (ApEn) is defined as:

$$ApEn(m, r, N, \tau) = \frac{1}{N-(m-1)\tau}\sum_{i=1}^{N-m\tau} \ln(C_i^m(r)) - \frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau} \ln(C_i^{m+1}(r)) \quad (3.15)$$

The difference between the approximate and sample entropy is in the fact that, in order to avoid the logarithm of zero, the ApEn allows self-matches (j=i in Eq. (3.14)) which causes the bias in estimates, whereas the SampEn prevents the bias. The bias is prevented thanks to a) the exclusion of self-matches from the $B_m^r(i)$, b) ecquilazation of the number of sliding window comparison for template length vectors *m* and *m + 1*, and c) the place substitution of summation and logarith in Eq (3.15):

$$SampEn(m, r, N, \tau) = \ln(\sum_{i=1}^{N-m\tau}(B_m^r(i) - 1)) - \ln(\sum_{i=1}^{N-m\tau}(B_{m+1}^r(i) - 1)) \quad (3.16)$$

3.4. Approximate entropy as a measure of system complexity

Shaw (1981) was the one who managed to see the measure of the rate of information generation of a chaotic system as an important parameter. Later on, inspired by the K-S entropy, Grassberger and Procaccia (1983) came up with the formula which was used as a way of calculating rate from time series data. The formula was first altered by Takens (1983) who included the distance metric in it (as shown in the second equation), and then by Eckmann and Ruelle (1985) who used Takens's formula as the

basis and changed it in the way which allowed a 'direct' calculation of the K-S entropy for the *physically* invariant measure, which is supposed to be underlying the data distribution. Nowadays, these formulas are used with time-series data as 'standard' entropy measures. The Eckmann-Ruelle (E-R) entropy formula is given below.

This kind of formula and its variations can be very important for the classification of low-dimensional chaotic system. However, the formula is usually not applicable in other contexts because of the statistical flaws mentioned in the Introduction. In order to use the E-R entropy formula with experimental data, it is necessary to make an approximation of equation 3.17, and to establish a meaningful range of "*r*" (vector comparison distance). The reason for this is the fact that E-R entropy is infinity for a process with superimposed noise of a magnitude. The changes in the formula results in the loss of deterministic system, as well as its inability to differentiate some processes that appear to differ in complexity, such as the E-R entropy for the MIX process which equals infinity (for all $p \neq 0$).

Fix m and r in equation 2.16 and define

$$\phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \tag{3.17}$$

$$E-R\ entropy = \lim_{r\to 0} \lim_{m\to\infty} \lim_{N\to\infty} [\phi^m(r) - \phi^{m+1}(r)] \tag{3.18}$$

Note that

$$\phi^{m+1}(r) - \phi^m(r) = \text{average over i of } \log[conditional\ probability\ that\ |u(j+m) - u(i+m)| \leq r \text{ , given that } |u(j+k) - u(i+k)| \leq r \text{ for } k = 0, 1, 2, \cdots, m-1] \tag{3.19}$$

$$ApEn(m,r) = \lim_{N\to\infty} [\phi^m(r) - \phi^{m+1}(r)] \tag{3.20}$$

$$ApEn(m,r,N) = \phi^m(r) - \phi^{m+1}(r) \tag{3.21}$$

Having N data points, we will apply this formula by defining the statistic

What E-R entropy and ApEn measure is the probability that runs of close patterns to stay close on the incremental comparisons. ApEn can be used for the calculation of any

of the time series, chaotic or not. The basis of the ApEn says that if joint probability measures (for the constructed m vectors) that describe each of two systems are different, their marginal distribution on a fixed partition will probably be different, too.

In contrast to the ApEn which cannot confirm a chaotic system, a nonzero value used for E-R entropy makes sure that a known deterministic system is chaotic. This is the first advantage of E-R entropy over the ApEn. In addition to this, the ApEn (m, r) was not developed as an approximate value of E-R entropy. In some cases, like when we have a large number of points, a low-dimensional attractor, and a large m, the two parameters can prove to be almost the same. Nevertheless, it is of essential importance to look at the ApEn(m, r) as a part of formulas, and ApEn(m, r, N) as a part of statistics; fixed m and r are used for the comparison of the systems.

3.5.Cross entropy of parallel time series

Cross-Ap entropy has been only recently introduced as a possible way of comparing two different time series in order to assess their degree of asynchrony or dissimilarity. The definitions of cross-ApEn and ApEn are, in fact, very similar to each other. If we have two time series of N points $\{u(j)\colon 1 \le j \le N\}$ and $\{v(j)\colon 1 \le j \le N\}$, from the vectors $\mathbf{x}_m(i) = \{U(i + k)\colon 0 \le k \le m - 1\}$ and $y_m(i) = \{v(i + k)\colon 0 \le k \le m - 1\}$; and the distance between these kinds of vectors is defined as $d[\mathbf{x}_{\mathrm{m}}(i), \mathbf{y}\mathrm{m}(i)] = \max\{|u\ (i + k) - v(j + k)|\colon 0 \le k \le m - 1\}$. First, it is necessary to define $C_i^m(r)(v\|u)$ as the number of $\mathbf{Y}_m(j)$ within r of $\mathbf{x}_m(i)$ divided by $(N - m + 1)$, and then define $\Phi^m(r)(v\|u) = (N - m + 1)^{-1} \Sigma_{i=1}^{N-m+1} \ln [C_i^m(r)\ (v\|u)]$, and cross-ApEn $(m, r, N)(v\|u) = \Phi^m(r)(v\|u) - \Phi^{m+1}(r)(v\|u)$. The given definition almost completely coincides with that of the statistic ApEn, except for those templates that are chosen from the series u and compared with vectors from v. This means that there is an obvious directionality with this analysis, and the series that contributes the templates the template series and the series with which they are compared are termed as target series in this work, which is why it is possible to refer to cross-ApEn(m, r, N)(target||template).

Two things have been noticed here. Firstly, there are no self-matches due to the fact that templates cannot be compared with themselves. Because of this, $C_i^m(r)(v\|u)$ can

equal 0, and it cannot be guaranteed that cros-ApEn(m, r, N)(v||u) will be defined. Secondly, there is the existence of a "direction dependence" of cross-ApEn analysis. In order to avoid these possible problems, the cross-SampEn is defined.

It should be emphasized that cross-ApEn is not always defined. As it has already been mentioned earlier, there are no self-matching in cross-ApEn, which is why it does not suffer from the same bias as ApEn. Nevertheless, there is still the issue of the necessity for each template to generate a defined, nonzero conditional probability. In this way, each template has to find minimum one match for m + 1 points, or a probability has to be assigned to it, and so far, there has not been suggested any solutions for this problem. In contrast to this, cross-SampEn needs only one pair of vectors in the two series match for m + 1 points.

Pincus (Pincus, SM., 1991) talks about the family of MIX(P) stochastic processes which allowed a testing ground for cross-ApEn. MIX(P) time series of N points, in which the value of P is somewhere between 0 and 1, is usually a sine wave, where N points N X P randomly chosen points have been replaced with random noise. We have managed to determine cross-ApEn(1, r, 250) for the pair {MIX (Q)||MIX(P)] and its direction conjugate [MIX(P)||MIX(Q)] for 16 realizations of each of the six combinations of P=0.1, 0.2, 0.3 and Q=0.5, 0.7 over a range of values of r from 0.01 to 1.0. Here, cross-ApEn(l, r, 250) [MIX(Q)||MIX(P)] was not defined for the 96 pairs for $r \leq 0.16$ and was defined, and was defined only for $r \geq 0.50$. Also, the cross-ApEn (1, r, 250) [MIX(P)||MIX(Q)] was not defined for the values of $r \leq 0.32$ and was defined only for r = 1.0.

As a way of expanding the conditions for which cross-ApEn was defined, it was necessary to introduce a correction factor into the algorithm; and as a way of avoiding ln (0) whenever $C_i^m(r)(v\|u) = 0$, or when $C_i^{m+1}(r)(v\|u) = 0$, they were redefined to be positive and nonzero. Their effect on the overall calculation was also reduced through their inclusion only in those cases when it was necessary to make sure that cross-ApEN was defined. On the other hand, this approach causes the introduction of bias, which can be a problem. Due to this, cross-ApEN with two different correction strategies has been investigated. The first strategy, called bias 0, proved to be similar to self-matching.

$C_I^m(r)(v\|u) = 0$ or $C_I^{m+1}(r)(v\|u)$ were set to 1. In this way, if a template did not match any other, it was assigned a conditional probability of 1, just like the original description of ApEn. However, in cases when $C_I^m(r)(v\|u) \neq 0$, but $C_I^{m+1}(r)(v\|u) = 0$, they were redefined $C_I^{m+1}(r)(v\|u) = (N-m)^{-1}$ so that the probability assigned would be the lowest possible, nonzero probability given the nonzero value of $C_I^m(r)(v\|u)$.

The other approach, called bias max, is also concerned with the modification of $C_I^m(r)(v\|u)$ and $C_I^{m+1}(r)(v\|u)$ functions whose value would otherwise have been 0. In these modifications $C_I^m(r)(v\|u)$ was redefined to be 1, and $C_I^{m+1}(r)(v\|u)$ was redefined to be $(N - m + 1)^{-1}$, like for bias 0.

The one discrepancy between these two strategies is that bias max assigns to a template a probability of $(N - m)^{-1}$, which represents the lowest nonzero probability that is allowed by the length of time series, without causing any matches. This means that the bias 0 sets the bias toward a cross-ApEn value of 0 in cases when there are no matches. Contrary to this, bias max sets bias toward the highest observable value of cross-ApEn.

Unlike the cross-SampEn, cross-ApEn is direction dependent. The logarithms inside the summation do not generally allow $\Phi^m(r)(v\|u)$ to be equal to $\Phi^m(r)(u\|v)$. Due to this, cross-ApEn(m, r,N)(v||u) and its direction conjugate cross-ApEn(m, r, N)(u||v) are not equal in most of the cases.

In an attempt to define cros-SampEn, $B_I^m(r)(v\|u)$ was set as $(N - m)^{-1}$ times the number of vectors $\mathbf{y}_m(j)$ within r of $\mathbf{x}_m(i)$, where j goes from 1 to $N - m$. The definition is then $B^m(r)(v\|u) = (N-m)^{-1} \Sigma_{I=1}^{N-m} B_I^m(r)(v\|u)$. Similarly to this, $A_I^m(r)(v\|u)$ was set as as $(N - m)^{-1}$ times the number of vectors $\mathbf{y}_{m+1}(j)$ within r of $\mathbf{x}_{m+1}(i)$, where j goes from 1 to $N - m$. The definition of this would then be $A^m(r)(v\|u) = (N-m)^{-1} \Sigma_{I=1}^{N-m} A_I^m(r)(v\|u)$. In the end, cros-SampEn $(m, r, N)(v\|u) = -\ln [A^m(r)(v\|u)] / [B^m(r)(v\|u)]$ was set. Looking at the direction dependence in the previous definition, it can easily be observed that $(N - m) B_I^m(r)(v\|u)$ represents the number of vectors from v within r of the ith template of the series u; and looking at all the templates together, it can be seen that $\Sigma_{I=1}^{N-m} (N - m) B_I^m(r)(v\|u)$ counts just the number of pairs of vectors from the two series that match within r. The number of overall pairs I not dependant on the series, nor the

target. As it can be seen, the last summation equals $(N-m)^2\ B^m(r)(v\|u)$, which leads to the conclusion that $B^m(r)(v\|u)$ is also direction independent, making cross-SampEn(*m, r, N*)(*v*||*u*) equal to cross-SampEn(*m, r, N*)(*u*||*v*). It has also been observed that cross-SampEn can be defined if $A^m(r)(v\|u) \neq 0$, but cross-SampEn requires that one pair of vectors in the two series match for *m* + 1 points.

3.6.Inconsistency of entropy estimations

3.6.1 Flip-flop effects

There are two sources for unstable results in entropy estimates: the threshold r and time series length N (Boskovic 2011).

An initial preference for the threshold values were rMAX and rTHEOR. The first one is a threshold value for which ApEn reaches its maximum and it is obtained separately for each time series (for a range of thresholds a series of ApEn were generated, called a "threshold profile"; the maximal entropy value of a threshold profile corresponds to the threshold rMAX). The second one is a theoretical estimate of the threshold value for maximal value of ApEn, calculated according to the formulae from literature:

$$m=2:\quad r_{\mathrm{THEOR}}=(-0.02+0.23\sqrt{\hat{\mathrm{sd}}_1/\hat{\mathrm{sd}}_2})/\sqrt[4]{N}/1000\ , \tag{3.22}$$

$$m=3:\quad r_{\mathrm{THEOR}}=(-0.06+0.43\sqrt{\hat{\mathrm{sd}}_1/\hat{\mathrm{sd}}_2})/\sqrt[4]{N}/1000\ . \tag{3.23}$$

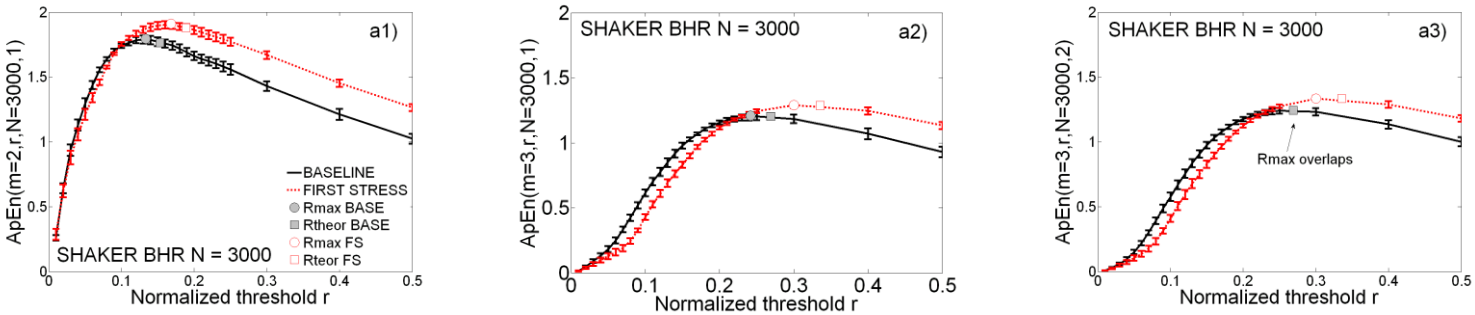
The terms $\hat{\mathrm{sd}}_1$ and $\hat{\mathrm{sd}}_2$ in Eqs. (3.22) and (3.23) can be regarded as a short and a long term variability of a (bounded) signal. The first one is an estimate of standard deviation of differential series x(i)-x(i-τ), while $\hat{\mathrm{sd}}_2$ is an estimate of standard deviation of bounded PI time series. The formulae (7, 8) are derived for human time series, but the accordance between theoretical and empirical values is excellent for signals taken from rats as well.

There is no explicit constraints regarding the signal stationarity within the entropy estimate definitions and this subject has seldom been addressed. Applications especially

in neurological signal analysis note that stationarity is not a prerequisite. Others implicitly disagree, performing the first difference as a standard method to stationarize time series. Pincus et al. state that ApEn is insensitive to artifacts (outliers).

An implicit statement that a time series need to be stationary arises from the normalized threshold *r* definition: to ensure unbiased distance measure $d_m(\mathbf{x}_m^{(i)}, \mathbf{x}_m^{(j)})$, a normalized threshold is specified for each signal separately as a fixed portion of the standard deviation estimated from the signal. A time average estimates of the statistical moments are meaningful for the stationary signals only [4], so signals for an entropy analysis need to be at least wide sense stationary.

Entropy estimates are sensitive to time series lengths. To check the influence of this experimental environment, the signal length N was gradually shortened from N=6000 to N=1000, remaining within the recommended limits (N = 10^m to 20^m). Unexpectedly, a change in the experimental outcome occurred. If N=3000, maximal entropy values ApEn(rTHEOR) for STRESS exceeded the ones in BASELINE conditions; if N=6000, the results were just opposite, Pincus et al in indeed warned that two processes may exchange the complexity measure if different parameter choice is applied (flip-flop effect). It is usually accounted for bias induced by self-matching, which is a dominant but not the only factor, since the same effect, in less extent, is noticed in SampEn analysis of the same signals. This is not an outcome of a possible animal adaptation to stress. The flip-flop effect persisted if the first N=3000 samples were replaced with the last N=3000 samples, when the rats were supposed to be already adapted (Fig. 3.2).



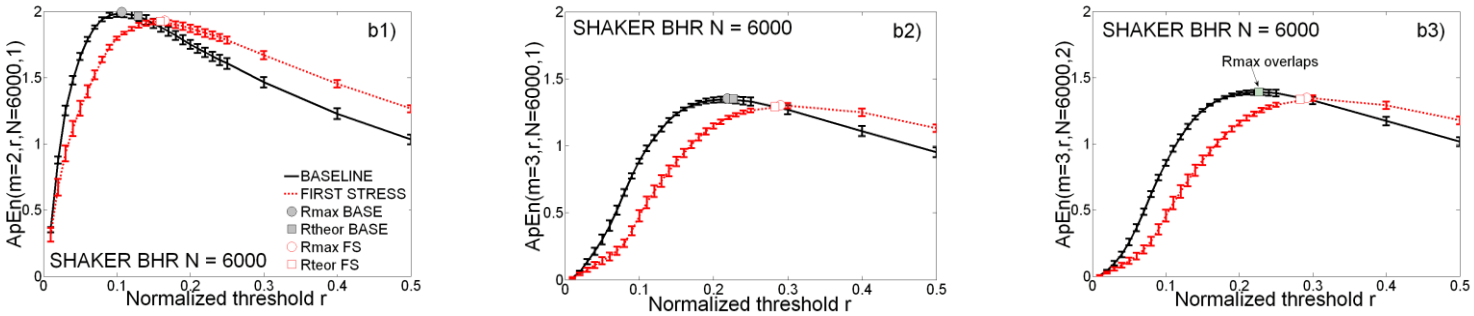


Figure 3.2. Flip-flop effect: mean *ApEn* ± SE (standard error) for BHR rats in SHAKER stress

a) N=3000 > 10^m ; maximal *ApEn* is <u>*higher*</u> in STRESS,

a1) m=2, τ=1; a2) m=3, τ=1; a3) m=3, τ=2

b) N=6000 > 10^m ; maximal *ApEn* is <u>*lower*</u> in STRESS, b1) m=2, τ =1; b2) m=3, τ=1; b3) m=3, τ=2)

Source: A. Boskovic et al. / Computers in Biology and Medicine 42 (2012) 667–679

If the threshold r is changed so that it exceeds the proposed values rMAX and rTHEOR, all the experimental results become consistent, regardless of the record lengths (Fig.3.2). An explanation of unstable behavior at ApEn maximums could be given using the circular plots that, for a fixed template $\mathbf{X}_m^{(i)}$, depicts all the distances d(i,j)=dm($\mathbf{X}_m^{(i)}$, $\mathbf{X}_m^{(j)}$). The coordinates of circular plots are:

$$x_j = d(i,j)\cdot\cos(j\cdot 2\cdot\frac{\pi}{N}); \quad y_j = d(i,j)\cdot\sin(j\cdot 2\cdot\frac{\pi}{N}); \tag{3.24}$$

Fig. 3.3 shows circular plot for a randomly chosen rat and template $\mathbf{X}_m^{(80)}$ (Fig. 3.3a) and for the template that is at the greatest distance from $\mathbf{X}_m^{(80)}$ (Fig. 3.3b). The radius of the circle in the centre of the plots is equal to rTHEOR. The zoomed part of both plots show that the number of matches $B_m^{r_{\mathrm{THEOR}}}(i)$ (Eq. (3,3)) is insignificant as compared to the total number of points N=6000. It is equal to 14 in Fig. 3.3a and only one (self-match) in Fig 3.3b.

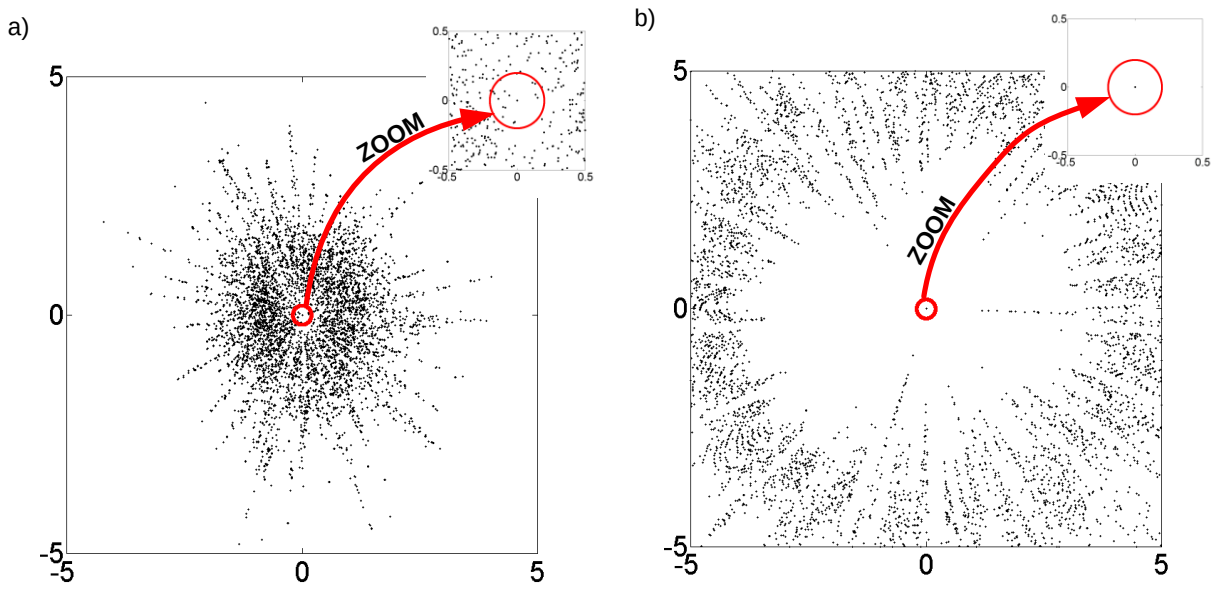


Figure 3.3: Distance circular plots of a) randomly chosen template no. 80 of rat 1, BHR SHAKER; b) template at greatest distance from template 80; The diameter of central circle is equal to rTHEOR

Source: A. Boskovic et al. / Computers in Biology and Medicine 42 (2012) 667–679

It might be suspected that a slightest change of $B_m^{r_{\text{THEOR}}}(i)$, e.g. induced by noise, would considerably alter the entropy estimate. To verify this assumption, a set of experiments is made: each number $B_m^r(i)$ in Eq. (3.3) is randomly altered, gradually increasing r:

$$B_m^r(i)_z = B_m^r(i) + z, \quad i = 1,\ldots, N - \tau \cdot (m-1) \tag{3.25}$$

where $z \in \{-1, 0, 1\}$ is an uniform random variable with the constraint that $B_m^r(i)_z$ can be neither zero nor negative. Then a new, experimental, ApEnEXPER is evaluated and its relative difference in respect to the initial ApEn is expressed as:

$$DIFF_{\text{EXPER}} = \frac{ApEn_{\text{EXPER}} - ApEn}{ApEn} \cdot 100\,[\%] \tag{3.26}$$

This difference, averaged over all experimental animals, is presented in Fig. 3.4, together with the median value of all the rTHEOR estimated over all the experimental data. It can be clearly seen that ApEn estimates are sensitive to minor changes of $B_m^r(i)$ for lower values of the threshold level. However, if the threshold level increases, the estimates become stable and insensitive to the noise.

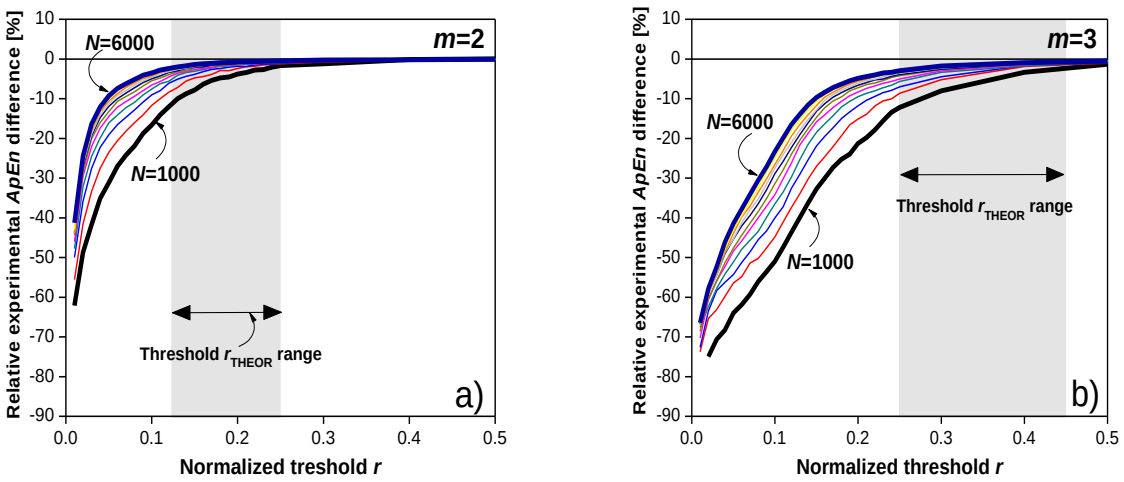


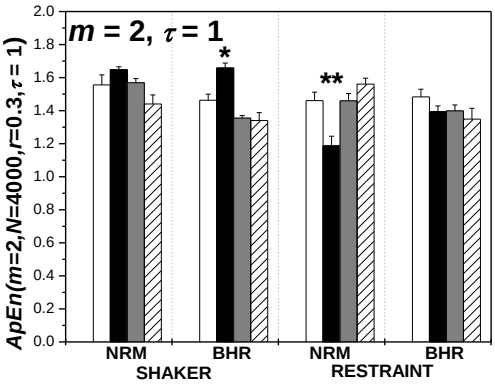
Figure 3.4. Relative difference of the experimental ApEnEXPER in respect to the original ApEn. Gray segments show the theoretical threshold range.

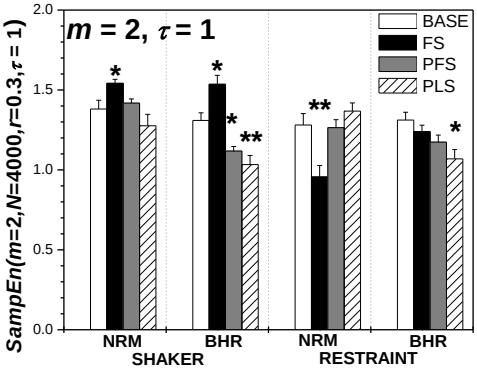
Source: A. Boskovic et al. / Computers in Biology and Medicine 42 (2012) 667–679

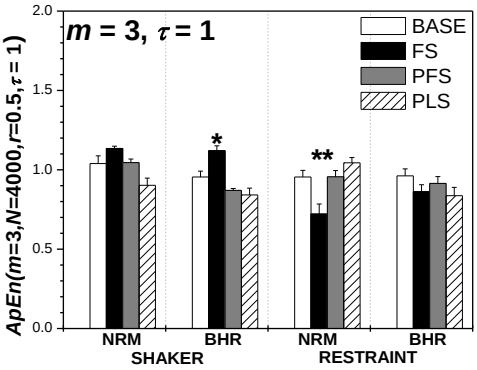
So the threshold values obtained by the automatic selection cannot be applied straightforwardly, since the estimated entropy values are not stable and lead to incorrect physiological interpretation. However, these formulae are excellent guidelines. From Fig. 3.4 it is clear that, for the threshold level that is a double of median of the theoretically evaluated rTHEOR of all the experiments in the data set, the induced noise caused the entropy distortion of less than 1%, which is quite a reasonable stability.

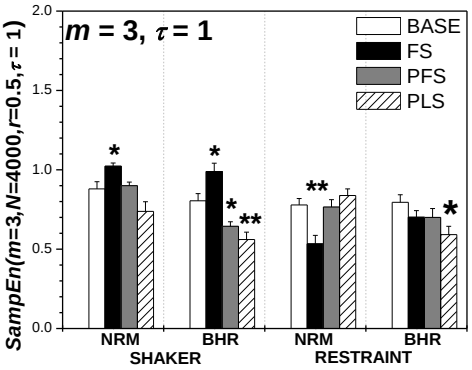
For this reason, we propose the following correction to automatic threshold evaluation. The threshold level is evaluated for each time series separately, according to Eqs. (3.22) and (3.23). Then a median of the obtained results (125 threshold values in our case) is found. The double value of the evaluated median is a normalized threshold value that ensures stable estimates.

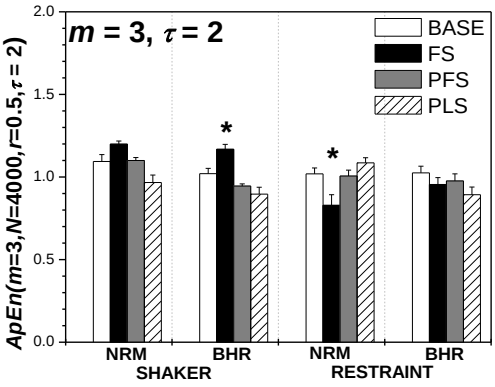
This method yields the normalized threshold values as (rounded) r=0.3 for m=2 and r=0.5 for m=3, regardless of τ and N.











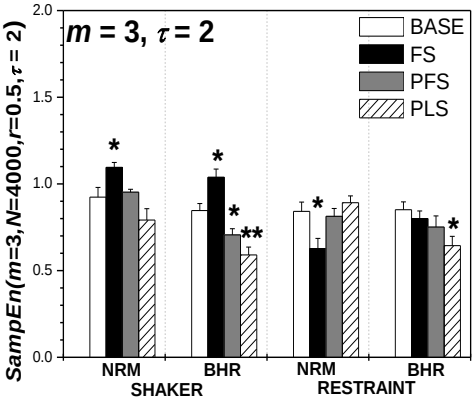


Figure. 3.5. Absolute values of *ApEn* and *SampEn* for a sample size *N*=4000; results are given as mean+SE, the statistical significance was assessed using repeated measures ANOVA test at levels $p<0.05$ (*), $p<0.01$ (**)

Source: A. Boskovic et al. / Computers in Biology and Medicine 42 (2012) 667–679

3.6.2. Effects of auto correlation and cross correlation

It should be said that in those situations where these measures are applied, the autocorrelation of the time series data tend to deteriorate quickly, thus minimizing the effect of long range linear correlations on the computations.

When it comes to the signals with long range linear correlations, the autocorrelation function deteriorates slowly and represents a good method for quantifying the linear stochastic predictability in time series. What unity time delay can do when it is used in the calculation of Approximate and Sample Entropy is to mask the ability of the algorithms to quantify the complexity in the time series the come as a consequence of non-linear influences. In order to get higher values of complexity for respiratory time series, Chen and his colleagues (Chen et al, 2005) lowered the sample size by factors 2, 4, and 8, which lead to the adjustment of the time delay parameter in the computation. The method of down-sampling leads to the reduction of linear correlation between consecutive samples, and to the increase of Approximate Entropy values. In addition to this, the down-sampling influences the reduction of the frequency content in the signal, which further eliminates important features in the data. The calculation of Approximate and Sample Entropy measures as a function of the delay τ would supposedly provide new information and insight into the analysis.

The lack of probability can come as a consequence of numerous different sources, such as stochastic and nonlinear deterministic sources. The investigation of the role which time delay parameter (τ) of Approximate and Sample Entropy can have is done in an attempt to see how these measures can be used as a way of making meaningful comparisons across data sets. Autocorrelation or the PSD can be used for the comparison of predictability obtained from linear stochastic sources across data sets. For this reason,

the use of non-unity time delay influences the reduction of the effects of autocorrelation, which allows that comparisons between data sets to be made based on the influence of nonlinearity on the predictability of the data. The first thing that was done in the study was to see the effects of autocorrelation of a time series, regarding the computation of the Approximate and Sample Entropy measures.

In the past, the effects of autocorrelation have been researched in a connection to other nonlinear time series analysis techniques, like correlation dimension. Within this kind of research, it was discovered that long range correlation property of a signal adds a shoulder to the logarithmic plot of the correlation integral, which causes inaccurate and spurious dimension estimation (Theiler, 1990). In order to solve this problem, the Theiler window was introduced as a part of the calculation of the correlation integral (Theiler, 1986). A lot of research has been dedicated to the effects which long range correlation has on the computation of correlation dimension, and a number of techniques like false nearest neighbors, used for picking an appropriate embedding dimension, and the first minimum or zero crossing of the autocorrelation function, or mutual information used for the selection of an appropriate delay, have been used. Nevertheless, there are still many ongoing debates when it comes to the linear stochastic versus nonlinear deterministic features that contribute to the complexity of a signal measured by correlation dimension. As an attempt to solve these debates, surrogate data methods have been purposed in order to us a statistical framework within which similar can be studied thoroughly.

Time delay is used as way of reducing the effects of linear and nonlinear correlations in a signal, and here are two ways in which it can be determined; it is possible to use either a delay that corresponds to the first minimum or zero crossing of the autocorrelation function, or the mutual information. In those cases when surrogate data analysis is used as a way of establishing statistical significance of the computational results, a time delay based on the first minimum or zero crossing should be used. This is done when original and surrogate data sets have the same autocorrelation function.

3.7.Reference

[1] Thomas Schu¨rmann and Peter Grassberger: Entropy estimation of symbol sequences Department of Theoretical Physics, University of Wuppertal, D-42097 Wuppertal, Germany.

[2] Aleksandar Boskovic, Tatjana Loncar-Turukalo, Olivera Sarenac, Nina Japundzic-Zigon, Dragana Bajic, Unbiased entropy estimates in stress: A parameter study. Computers in Biology and Medicine 42 (2012) 667–679.

[3] Joshua S. Richman And J. Randall Moorman, Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 278: H2039–H2049, 2000.

[4] Steven M. Pincus, Approximate entropy as a measure of system complexity. Proc. Nati. Acad. Sci. USA Vol. 88, pp. 2297-2301, March 1991 Mathematics.

[5] Farhad Kaffashi, Ryan Foglyano, Christopher G. Wilson and Kenneth A. Loparo, The effect of time delay on Approximate & Sample Entropy calculations. Physica D 237 (2008) 3069-3074.

[6] A.L. Goldberger, S.M. Pincus, Physiological time-series analysis: What does regularity quantify?, Am J Physiol(Heart Circ Physiol), 266 (1994) H1643-H1656.

[7] D. E. Lake, J.S. Richman, Sample entropy analysis of neonatal heart rate variability, Am J Physiol Regul Integr Comp Physiol, 283 (2002) R789-R797

[8] X.Chen, I.C.Solomon, K.H.Chon, Comparison of the use of approximate entropy and sample entropy: application to neural respiratory signal, in Proc. of the 27th IEEE EMBS Ann. Conf (2005) 4212-4216

[9] S.Lu, X.Chen, J.K.Kanters, I.C.Solomon, K.H.Chon, Automatic selection of the threshold value r for approximate entropy, IEEE Trans. Biomed. Eng. 55, 8 (2008) 1966-1972

[10] K. H. Chon, C.G. Scully, S. Lu, Approximate Entropy for all signals, IEEE Engineering in Medicine and Biology 28, 6 (2009) 18-23,

[11] P Castiglioni, M Di Rienzo, How the Threshold ”R” Influences Approximate Entropy Analysis of Heart-Rate Variability, Computers in Cardiology 35 (2008) 561-564

[12] M.B. Kennel, R. Brown, Determining embedding dimension for phase-space reconstruction using geometrical reconstruction, Physical Review A Gen. Phys. 45,6 (1992) 3403-3411

[13] R.B. Govindan, J.D.Wilson, H. Eswaran, C.B. Lowery, H. Preisl, Revisiting sample entropy analysis, Physica A, 376 (2007) 158-164

[14] F. Kaffashi, R.Foglyano, C.G.Wilson, K.Loparo, The effect of time delay on approximate and sample entropy calculations, Physica D 237 (2008), 3069-3074;

[15] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, “An advanced detrending approach with application to HRV analysis, IEEE Trans. Biomed. Eng. 42(2) (2002) 172-174.

[16] R. Karvajal, Dimensional analysis of HRV in hypertrophic cardiomyopathy patients, IEEE Engineering in Medicine and Biology, **21** (4) (2002) 71-78

[17] J.S.Bendat, A.G. Piersol, Random data analysis and measurement procedures, Wiley Series in Probability and Statistics, New York (1986)

[18] A. Papoulis, Probability, random variables and stochastic processes, McGraw-Holl International Edition, (1984)

[19] A.G. Hudetz, J.D. Wood, J.P. Kampine, Cholinergic reversal of isoflurane anesthesia in rats as measured by cross-approximate entropy of the electroencephalogram, Anesthesiology 99 (2003) 1125–31

[20] L. Huang, Y. Wang, J. Liu, J. Wang, Approximate entropy of EEG as a measure of cerebral ischemic injury, Proceedings of the 26th Annual International Conference of the IEEE EMBS (2004)

[21] G.M. Souza, R.V. Ribeiro, M.G. Santos, H.L. Ribeiro, R.F. Oliveira, Approximate entropy as a measure of complexity in SAP flow temporal dynamics of two tropical tree species under water deficit, Annals of the Brazilian Academy of Sciences76(3) (2004) 625-630

[22] A.M. Frazer, H.L. Swinney, Independent coordinates for strange attractors from mutual information, Physical Review A, 33(2), (1986), pp 1134-1140

[23] R. Carvajal, J.J. Zebrowski, M. Vallverdú, R. Baranowski, L. Chojnowska, W. Poplawska, P. Caminal, Dimensional Analysis of HRV in Hypertrophic Cardiomyopathy Patients, IEEE Engineering In Medicine And Biology, (2002), pp 71-78

[24] G.P. King, R. Jones and D.S. Broomhead, Phase portraits from a time series: A singular system approach, Nucl. Phys., vol. B2, (1987) p 379

[25] T.Loncar Turukalo, D.Bajic, O.Sarenac, N.Japundzic-Zigon, A.Boskovic, Environmental stress: approximate entropy approach revisited, Proc. of the 31th IEEE EMBS Ann. Conf (2009) 1804-1807

[26] T. Loncar Turukalo, O. Sarenac, N.Japundzic Zigon, D.Bajic, Multiscale and moving average antropy: statistical dependance in stressed subjects, Proc. of the 7th International Symposium on Intelligents Systems and Informatics (2009)

[27] T.Schreiber, A.Schmitz, Surrogate time series, Physica D 142 (2000) 346-382

[28] A.P. Blaber, Y. Yamamoto, R.L. Hughson, Methodology of spontaneous baroreflex relationship assessed by surrogate data analysis,. Am J Physiol 268 (1995) H1682-H1687

[29] A. Porta, R. Furlan, O. Rimoldi, M. Pagani, A. Malliani, P. van de Borne, Quantifying the strength of the linear causal coupling in closed loop interacting cardiovascular variability signals, Biol Cybern; 86 (2002) 241-251.

3. Approximate entropy improvements

3.1. Fuzzy Entropy approximate entropy for chaotic and natural complexity

When it comes to the physical world, it is often hard to determine if an input pattern completely belongs to a class, due to the fact that boundaries between classes can be very ambiguous. Zadeh was the one who came up with the concept of "fuzzy set" in order to characterize input-output relations in an environment of imprecision. The theory he proposes gives us a mechanism with which we can measure the degree to which a pattern belongs to a given class. What Zadeh actually did was to use the "membership degree" as a part of a fuzzy function u_c(x), where each point x is connected to a real number in the range [0,1]. The closer the u_c(x) value to the unity is, the higher the membership grade of x in the set C is. In order to calculate a fuzzy measurement of the similarity between X_i^m and X_j^m based on their shapes, the fuzzy membership function $u(d_{ij}^m, r)$ is used in fApEn. This new similarity index causes the softening of the hard boundary of the Heaviside function, as well as the approach of the points that become more similar. Based on all this, it can be said that the definition of the fuzziness approximate entropy states that for the same time series T, the vector sequence takes a form similar to that of the definition of ApEN:

$X_i^m = \{u(i), u(i+1), \ldots, u(i+m-1)\} - u0(i) \quad (i = 1, \ldots, N-m+1)$

But X_j^m is still generalized by removing a baseline $u0(i) = \frac{1}{m}\sum_{j=0}^{m-1} u(i+j)$, and the distance d_{ij}^m between the two vectors X_i^m and X_j^m is defined as

$d_{ij}^m = d\left|X_i^m, X_j^m\right| = \max_{k\in(0,m-1)} |u(i+k) - u0(i) - (u(j+k) - u0(j))| \quad (i, j = 1 \sim N-m+1, j \neq i)$

When *r* is known, the similarity degree D_{ij}^m between X_i^m and X_j^m can be calculated by a fuzzy membership function $D_{ij}^m = u(d_{ij}^m, r)$

The function B_r^m is calculated with $C_r^m(i) = \frac{1}{N-m+1}\sum_{j=1, j\neq i}^{N-m+1} D_{ij}^m$ and $\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} \ln[C_r^m(i)]$ functions.

In addition to this, it is also possible to define the measure *f*ApEn(*m*, *r*) of the time series as $fApEn(m,r) = \lim_{N\to\infty}[\phi^m(r) - \phi^{m+1}(r)]$ from the vector sequence $\{X_i^{m+1}\}$ and the function $\phi^{m+1}(r)$.

In the end, the finite datasets of the *f*ApEn is estimated from the statistic $fApEn(m,r,N) = \phi^m(r) - \phi^{m+1}(r)$.

Many functions like Gaussian, Sigmund, bell shape function, or any other membership function which has the desired properties can be used as a description of the two vectors' similarities in practice. The two necessary properties are 1) continuity, which prevents the similarity from changing too fast, and 2) convex, which sets the self-similarity as the maximum. The authors of the present study have used the Gaussian function $u\left(d_{ij}^m, r\right) = \exp(-d_{ij}^2/r)$ as the fuzzy membership function for the *f*ApEn calculations. The vector dimension *m* was set to 2, in order to calculate *f*ApEn, as well as ApEn.

3.2.Multiscale Entropy

It has been found that traditional single-scale entropy estimates usually influence lower entropy in time series of physiologic data like inter-beat (RR) interval series

compared to the surrogate series which are formed through the shuffling of the original physiologic data. The reason for this is the fact that shuffled data display more irregularity and thus are less predictable than the original series, which tend to have correlations at many time scales. The very process used for the generation of surrogate data destroys the correlations and degrades the information content of the original signal. It should also be mentioned that greater entropy is not a characteristic of greater complexity. On the other hand, a different method – MSE, tells us that the original time series are more complex compared to the surrogate time series, by showing us the dependence of entropy measures on scale.

There are two operations within the MSE method:

1. First of all, time series go through the process of "coarse-graining". Here, multiple coarse-grained time series are created for each given time series. This is done through the averaging of data points within non-overlapping windows of increasing length, τ, as it is shown in the picture below.

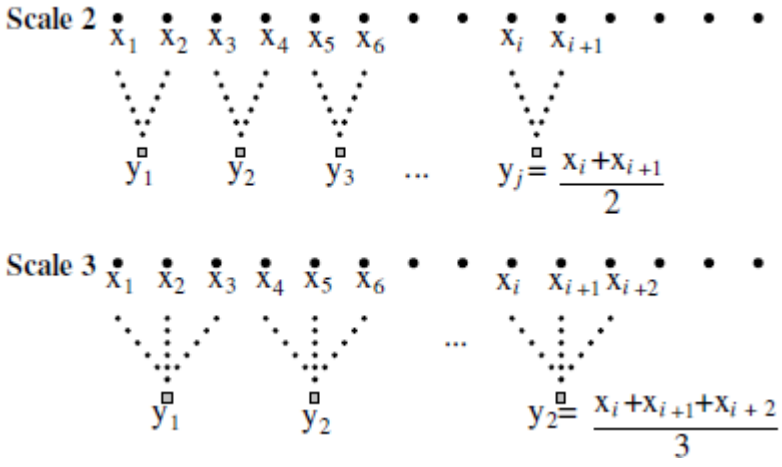


Figure 4.1: The coarse-graining procedure for scales two and three (Costa, M. et all, 2002)

The equation used for the calculation of each element of coarse-grained time series $y_i^{(\tau)}$ is:

$$y_i^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} x_i$$

In this equation, τ is the scale factor and $1 \leq j \leq N/\tau$, and the length of the series isN/τ.The first scale of the in the coarse-grained time series is the same as the original series, which is why it is not shown here.

2. The second process in the MSE method is concerned with the calculation of the SampEn. After SampEn has been calculated for each coarse-grained time series, it is than plotted as a function of the scale factor. Sample Entropy represents the regulatory statistic, and it looks for patterns in time series. Its job is also to determine a certain pattern's level of predictability/regularity.

3.3.Conditional entropy

It has been proven that it is possible to find different levels of synchronization between the RR interval and respiration, between coupled oscillators describing biped animals, between cardiac-related discharges of sympathetic nerves with different spinal inputs. Due to this, quantification of the degree of synchronization, such as the coupling strength, between two signals should be thought of as one of the most important goals. The usual method used for the quantification is the coherence function, however, based on cross-spectral analysis, it cannot be used when non-linear coupling is present. It has been proven that when N:M periodic dynamics is found (with the exception of a 1:1 coordination), the coherence is low thanks to the fact that the dominant rhythmicity of the forcing and spontaneous activities do not have the same frequency nor linear cross-correlation. What is more, sliding dynamics can lead to uncertain changes in the coherence function, based on the amount of phase jitters and changes in the interaction scheme.

Because of this, a new function, the one that measures the degree of uncoupling between two signals, has been worked out. It investigates the ability of the conditional entropy determined over two signals in order to give the amount of information carried by one of the signals in those cases when the samples of the other are known. It has been suggested that an estimator of the uncoupling function should calculate reliable estimates even on short data segments (around a few hundred samples). This is an extremely

important characteristic since stable coordination in experimental preparations cannot be kept without being corrupted by non-stationarities.

There are two methods of conditional entropy, Cross-conditional entropy and Corrected cross-conditional entropy

A normalised process $y = \{y(i), i = 1, \ldots, N\}$ is calculated by subtracting the mean and dividing by the standard deviation, $N - L + 1$ patterns $y_L(i)$ of length L. From this equation, it is possibleyo extract $(y(i), y(i-1), \ldots, y(i-L+1))$, and to represent a length L pattern $y_L(i)$ as the union of a lemgth $L-1$ pattern plus one more sample: $y_L(i) = (y(i), y_{L-1}(i-1))$. The conditional entropy

$$CE(L) = -\sum_{L-1} p(y_{L-1}) \sum_{i/(L-1)} p\left(\frac{y(i)}{y_{L-1}}\right) \times \log p\left(\frac{y(i)}{y_{L-1}}\right) \quad (4.1)$$

represents the joint probability of the pattern $y_{L-1}(i)$ and $p(y(i) / y_{L-1})$ probability of the sample $y(i)$ given the pattern $y_{L-1}(i)$; it also represents the amount of information carried by the most recent point $y(i)$ of the pattern $y_L(i)$in cases when the previous $L-1$ samples $y_{L-1}(i)$ are know. In the equation 3.1, the log function performs the natural logarithm. Since $\log_e$ is used, CE is determined in nats (natural ogarithm). After the Shannon entropy of length L patterns had been introduced

$$E(y_L) = -\sum_L p(y_L) \log p(y_L) \quad (4.2)$$

havin $p(y_L)$ for the joint probability of thepattern $y_L(i)$, the CE can be obtain as

$$CE(L) = E(y(i), y_{L-1}) - E(y_{L-1}) \quad (4.3)$$

The equation is used for calculating the information carried bz the new sample $y(i)$given the $L-1$ previous onesto the information increment that is connectedto the enlargement of the patternlength from $L-1$ to L. CE(1) is the same as E(y), that is to the entropy of the signal y in cases when we do not have any information about its past. The equation 3.3 has a number of computational advantages. First of all, the equation is based on a count of the frequency of different patterns of length $L-1$, which makes it very simple. What is more, a proper and efficient sorting algorithm has the ability to minimase the comptational cost in the calculation of CE. Like other entropy rates, CE is usually used for measuring the regularity of the signal y, and while increasing L, the CE can:

1) reache zero in cases when the signal *y* is periodic (predictable);

2) remain high when there is an irregular signal like white noise;

3) decrease to a value somwhere between the two extremes, based on the abilty of past samples to predict the future demands.

4.3.1. *Cross-conditional entropy:*

At this point, we will analyse two normalised signals: *u={u(i), i = 1, . . ., N}* and *y = {y(i), i =1,..., N}*. In these equations a CE of *y* given a pattern of *u* is defined as

$$CE_{y/u}(L) = -\sum_{L-1} p(u_{L-1}) \sum_{i/(L-1)} p(\frac{y(i)}{u_{L-1}}) \times \log p(\frac{y(i)}{u_{L-1}}) \quad (4.4)$$

in a way that is similar to the definition of CE in the first equation. Again, the CE is the amount of information carried by the sample *y(i)* in cases when the pattern $u_{L-1}(i)$ is assigned. To obtain the formulation of $CE_{y/u}$ in terms of an increment in Shannon entropy similar to the one in the equation 3.3, it is necessary to introduce a mixed pattern which is formed by *L-1* samples of *u* and the present sample of *y* : *(y(i), u(i), . . . , u (i – L + 2)) = (y(i), $u_{L-1}(i)$).* After making these changes, the $CE_{y/u}$ becomes

$$CE_{y/u}(L) = E(y(i), u_{L-1}) - E(u_{L-1}) \quad (4.5)$$

This means that the practical computation of $CE_{y/u}$ *is* actually the mixture of sorting and counting and *u* patterns. In this equation $CE_{y/u}(1)$ is equal to *E(y)*. In other words, it is equal to the entropy of *y* in cases when we do not have any information about *u*. As *L* is slowly increased, the $CE_{y/u}$ can: 1) reach a zero if a sufficient number of samples of *u* allows one to completely predict the*y*; 2) remain steady in cases when *u* and *y* are independent processes; 3) decrease towards a value somewhere between the two extremes, in cases when the knowledge of *u* is useful for partial estimation of *y*.

4.3.2. *Corrected cross-conditional entropy:*

As it has already been mentioned by some authors (Porta et al, 1998b), in those cases when a conditional entropy is determined from a limited amount of samples, it

always decreases to zero when L is increased. This comes as a result of the approximation of the conditional probabilities through the conditional sample frequencies. In addition to this, the cross-conditional entropy $CE_{y/u}$ gets affected by this bias. In fact, the frequency distribution of $y(i)$ with the pattern $u_{L-1}(i)$, can give us an erroneous certainty in cases when the conditioning pattern is found only once in the series u, resulting in an underestimate of $CE_{y/u}$. This also occurs when $u(i)$ is predicted by $y_{L-1}(i)$. This induces an false decrease in $CE_{y/u}$. Porta et al (1998b) gave us basis for the corrective term. Due to this, corrected $CE_{y/u}$ and $CE_{y/u}$ ($CCE_{y/u}$ and $CCE_{u/y}$) are defined as

$$CCE_{y/u}(L) = C\hat{E}_{y/u}(L) + perc_{y/u}(L) \cdot \hat{E}(y) \quad (4.6)$$

and

$$CCE_{u/y}(L) = C\hat{E}_{u/y}(L) + perc_{u/y}(L) \cdot \hat{E}(u) \quad (4.7)$$

In this equation, $perc_{y/u}$ and $perc_{u/y}$ represent the ratios of mixed patterns found only once over the total number of mixed patterns, $C\hat{E}_{y/u}(L), C\hat{E}_{u/y}(L), \hat{E}(y)$ and $\hat{E}(u)$are the estimates of the $CE_{y/u}(L), CE_{u/y}(L), E(y)$ and $E(u)$, based on the considered limited set data. The function of $perc_{y/u}(L) \cdot \hat{E}(y)$ and $perc_{u/y}(L) \cdot \hat{E}(u)$ is to prevent the decrease to zero of the $C\hat{E}_{y/u}(L)$ and $C\hat{E}_{u/y}(L)$, respectively. They do this by substituting the false certainty produced by a length L pattern that is found only once in the data with the maximum uncertainty that can be detected in the set of data (quantified by , $\hat{E}(y)$ and $\hat{E}(u)$). In this way, it is possible to get a maximum entropy estimate of the cross-conditional entropies. The increase of L, we decrease $CCE_{y/u}$ and $CCE_{u/y}$ to a minimum, based on the ability of u to predict y and vice versa. When it comes to long patterns (compared to the length of series), the correction pushes $CCE_{y/u}$ and $CCE_{u/y}$ to a maximum uncertainty, and during the analysis of independent processes, $CCE_{y/u}$ and $CCE_{u/y}$ both remain high and constant at , $\hat{E}(y)$ and $\hat{E}(u)$.

4.4.Reference

[1] Hong-Bo Xie, Jing-Yi Guo and Yong-Ping Zheng, Approximate entropy analysis of surface electromyography for assessing local muscle fatigue. Annals of Biomedical Engineering 2010.

[2] Madalena Costa, Ary L. Goldberger and C.-K. Peng, Multiscale Entropy Analysis (MSE). Beth Israel Deaconess Medical Center, Boston, USA 2005.

[3] A. Porta, G. Baselli, F. Lombardi, N. Montano, A. Malliani, and S. Cerutti, Conditional entropy approach for the evaluation of the coupling strength. Biol. Cybern. 81, 119-129 (1999).

5.Joint Symbolic Dynamics and Shannon entropy

Another quite frequently used concept of entropy in especially cardiovascular and respiratory time series analysis is a Shannon entropy of Joint symbolic dynamics (JSD) symbols. Given the systolic blood pressure (SBP) and pulse interval (PI) time series x_n^{SBP} and x_n^{PI}, n=1,…,N, new binary time series are b_n^{SBP} and b_n^{PI}, n=1,…,N-1, are derived, following the rule [8]:

$$b_n^{\mathrm{H}} = \begin{cases} 0 & x_{n+1}^{\mathrm{H}} - x_n^{\mathrm{H}} \le 0 \\ 1 & x_{n+1}^{\mathrm{H}} - x_n^{\mathrm{H}} > 0 \end{cases}, n = 1, \cdots, N-1, \mathrm{H} \in \{\mathrm{SBP}, \mathrm{PI}\} \qquad (5.1.)$$

Symbols are formed observing the bits of binary stream through a sliding window that comprise m bits (window size, closely related to the vector length m in a concept of approximate entropy).

The m bits that form a symbol (or word, as an alternative expression) need not be contiguous: the level of their spread (rake factor or factor of decorrelation τ the default is τ =1 for contiguous symbol) shows their distribution (Fig. 5.1a). While sliding, the window does not necessarily slide bit-by-bit. The number of bits the window “jumps” is called inter symbol interval (ISI, Fig 5.1a).

The temporal spread of signal templates has been already introduced in estimates at signal level. It was shown that downsampling the data (i.e. increased spread) resulted in higher signal complexity.

Symbols from the two time series need not be observed simultaneously. An offset for PI symbols, positive or negative, in respect to SBP symbols can be introduced (Fig. 5.1b).

The requirements of a statistically sufficient representation of word types limit the length of the words. In order to determine the number of word types (histogram classes), we used the $\sqrt{N}$ approximation for histogram construction of N observations. It was determined that for the 30-min recordings (with mean human heart rate of 80bpm) there were no more than 64 different word types feasible. If we look at all four different symbol combinations within **S** (the alphabets of **PI** as well as **SBP** consist of two elements), it can be seen that words with a maximum length of three are realizable ($2^3 * 2^3 = 64$). This means that we can use this approach to map the dynamics of **PI** and **SBP** within four consecutive heart beats (or three **PI**), which corresponds to a three-dimensional embedding, if we look at it from the perspective of phase space. This embedding is more pragmatic and not a faithful reconstruction of the system, however, a number of studies could show that a three-dimensional embedding of **HRV** is good for mapping of the short-term dynamics. In addition to this, in order to avoid spurious results, the embedding dimension should not exceed $\log_{10}(N)$. The mapped short-term fluctuations we are referring to here are predominantly influenced by respiratory activity, and they can result in high frequency (**HF**) oscillations in the power spectrum and respiratory sinus arrhythmia.

First of all, we will look at the dynamics of **PI** and **SBP** within the word distribution density matrix **W**. This matrix contains the frequency of each of the 8 x 8 possible combinations of **PI** and **SBP** patterns:

$$W = \begin{bmatrix} PI_{000}, SBP_{000} & \cdots & PI_{000}, SBP_{111} \\ \vdots & \ddots & \vdots \\ PI_{111}, SBP_{000} & \cdots & PI_{111}, SBP_{111} \end{bmatrix} \tag{5.2}$$

Relative frequencies of PI symbols, SBP symbols, $P_{\mathrm{PI}}(i)$ and $P_{\mathrm{SBP}}(i)$, i=1,…,2^m, as well as of joint symbols $P_{\mathrm{JSD}}(i,j)$, i,j=1,…,2^m, can then be estimated. Consequently, Shannon entropy is evaluated as:

$$\mathrm{H}_{Sh} = -\sum_{i=1}^{2^m} P_{\mathrm{H}}(i) \cdot ld(P_{\mathrm{H}}(i)), \quad \mathrm{H} \in \{\mathrm{SBP,PI}\}$$

$$JSD_{Sh} = -\sum_{i=1}^{2^m}\sum_{j=1}^{2^m} P_{\mathrm{JSD}}(i,j) \cdot ld(P_{\mathrm{JSD}}(i,j)) \tag{5.3}$$

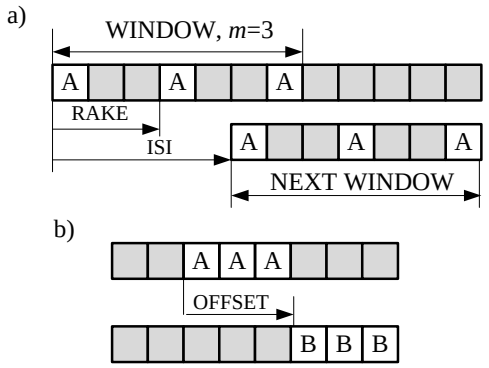


Figure 5.1. Symbol parameters, A,B∈{0,1};: a) Window size *m*, rake factor τ and ISI; b) Offset between the time series (e.g. SBP and PI)

Shannon entropy is a limit obtained under the assumption that all the symbols are statistically independent. This is not the case in physiological signals, and the reasons are twofold. The first one arises from the symbol definition: if the sliding window through which the successive symbols are observed overlaps, the result is a Markov source, shown using a state-diagram of 2^m states, where symbol length m=3, ISI=1 and τ=1 (Fig. 5.2). Even if the window does not overlap its previous contents (ISI>$r\cdot(m$-1)), the symbols would still be statistically dependent, since the adjacent symbols of biomedical time series exhibit strong correlation.

The transition probability $P(i|j)$ $i,j=1,\ldots,2^m$, must be taken into account and estimated from the time series, provided that the data length is sufficient. The entropy of a Markov source is then defined as:

$$\mathrm{H}_{Markov} = -\sum_{i=1}^{N}\sum_{j=1}^{N} P_{\mathrm{H}}(i) \cdot p(i \mid j) \cdot ld(p(i \mid j)), \cdots \mathrm{H} \in \{\mathrm{SBP,PI}\}$$

$$JSD_{Markov} = -\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} P_{\mathrm{JSD}}(i,j) \cdot p(kl \mid ij) \cdot ld(p(kl \mid ij)) \tag{5.4}$$

The transition probabilities $\mathbf{\Pi}=[p(i|j)]$ ($\mathbf{\Pi}=[p(kl|ij)]$) and the relative frequencies vector $\mathbf{P_H}$ (and $\mathbf{P_{JSD}}$) satisfy the following relations:

$$\mathbf{P_H} = \mathbf{P_H} \cdot \Pi; \quad \sum_{i=1}^{N} P_{\mathrm{H}}(i) = 1; \quad \sum_{j=1}^{N} p(i \mid j) = 1; i = 1, \cdots N. \tag{5.5}$$

Shannon entropy is an asymptotic value of Markov entropy, obtained for the statistically independent symbols.

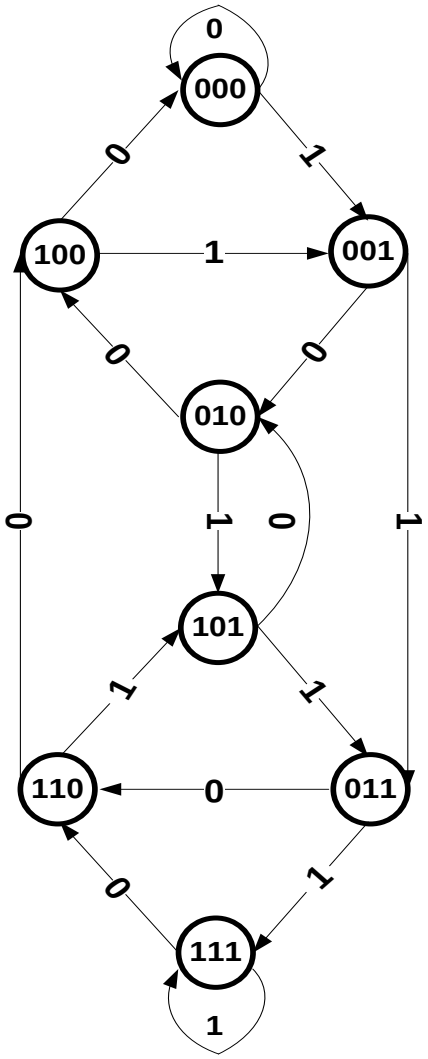
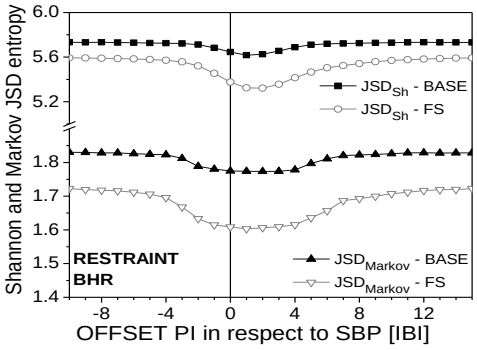
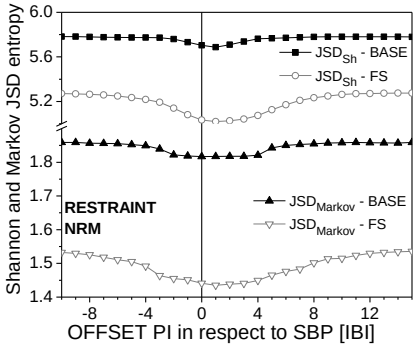
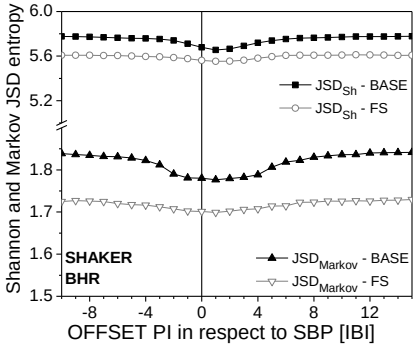


Figure 5.2. State transition diagram for symbols m=3, ISI=1 (sliding bit-by-bit) and τ =1 (contiguous symbols)

The levels of freedom considering the symbol entropy estimates (window length m, rake factor τ, ISI and offset), and the experiment itself including two strains of animals and two types of stress offers multiple possibilities for investigation.


Shannon and Markov JSD entropy
6.0
5.6
5.2
1.8
1.7
1.6
1.5
1.4
JSDSh - BASE
JSDSh - FS
RESTRAINT
BHR
JSDMarkov - BASE
JSDMarkov - FS
-8
-4
0
4
8
12
OFFSET PI in respect to SBP [IBI]


Shannon and Markov JSD entropy
6.0
5.6
5.2
1.8
1.7
1.6
1.5
1.4
JSDSh - BASE
JSDSh - FS
RESTRAINT
NRM
JSDMarkov - BASE
JSDMarkov - FS
-8
-4
0
4
8
12
OFFSET PI in respect to SBP [IBI]


Shannon and Markov JSD entropy
6.0
5.6
5.2
1.8
1.7
1.6
1.5
JSDSh - BASE
JSDSh - FS
SHAKER
BHR
JSDMarkov - BASE
JSDMarkov - FS
-8
-4
0
4
8
12
OFFSET PI in respect to SBP [IBI]

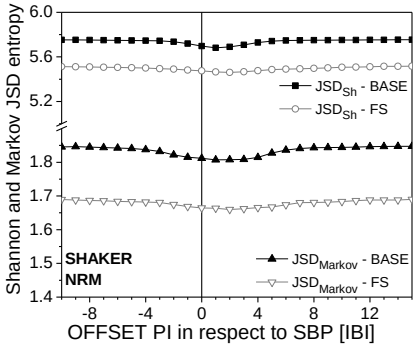

Shannon and Markov JSD entropy
6.0
5.6
5.2
1.8
1.7
1.6
1.5
1.4
JSDSh - BASE
JSDSh - FS
SHAKER
NRM
JSDMarkov - BASE
JSDMarkov - FS
-8
-4
0
4
8
12
OFFSET PI in respect to SBP [IBI]

Fig. 5.3. Markov and Shannon entropy in respect to the offset between the PI and SBP time series

Figure 5.3. presents the measures of SBP-PI coupling in respect to their mutual offset, i.e. delay time of PI series in respect to SBP series. It is generally assumed that, in rats, the SBP-PI coupling is the strongest for the offset 3 to 5. This is important if, for example, barroreflex sensitivity is evaluated using the sequence method. The decrease of entropy, showing increased dependence between the series, is clear for wider range of both the positive and the negative offsets. Furthe on, the decrease of entropy is almost non-existent in NRM animals, and if the animal is exposed to shaker stress, the decrease also diminishes, as if the shaker stress destroys the correlation between the SBP and PI series.

5.1. Reference

[1] Mathias Baumert, VicoBaier, Sandra Truebner, Alexander Schirdewan, and Andreas Voss, Short- and Long-Term Joint Symbolic Dynamics of Heart Rate and Blood Pressure in Dilated Cardiomyopathy. IEEE Transactions on Biomedical Engineering, 2005; 52 (12):2112-2115.

[2] Thomas Schurmann, Bias Analysis in Entropy Estimation. Research, Westdeutsche Genossenschafts-Zentralbank eG, Ludwig-Erhard-Allee 20, 40227 D¨usseldorf, Germany, 2008.

[3] O. Mohamoud, T. Loncar-Turukalo, N. Japundzic-Zigon, D. Bajic: On Symbol Entropy of Joint Time Series, Proceedings of the 7th ESGCO 2012, april 22-15, 2012, Kazimierz Dolny, Poland

6.Binarized approximate entropy

Continuous monitoring of patient's vital parameters has long ceased to be a privilege of medical doctors. A rapid development of sensors, communications devices and protocols, followed by fair channel distribution regulations, increased the availability of monitoring equipment and made it accessible to a diversity of users. The permanent self-monitoring - during fitness, sport, walk, sleep, work - became a way of life. It has already been recognized as a part of the new mobile crowdsensing concept, where subjects, equipped with adequate sensors, share and extract information to measure and map various phenomena for personal or common benefit, including the ones considering the health. Such a huge amount of signals, recorded at daily basis, presents a gold mine of valuable diagnostic and prognostic information awaiting to be explored. With a proper legislative permissions, it could become a part of databases available to the research society, but also used by health authorities for statistical purposes and further beneficial activities.

The transmission of recorded signals, however, is a subject to bandwidth and energy constraints. A part of computing vs. energy/bandwidth trade-off would be a local analysis performed in a wearable device. Unfortunately, most of the analytical methods are not compatible with such concept: they require stationary and artifact-free signals that an unprofessional self-acquisition - i.e. devices loosely attached or displaced and a subject in harsh movement - cannot provide. An example is (cross)-approximate entropy (X)*ApEn* one of the most valuable methods in biomedical research. It is widely accepted for assessing the complexity of biomedical signals: the complexity change may indicate an increase of adverse occurrences prior to alternation of any other parameter. This explains an amazing number of (X)*ApEn* citations in open literature. It also explains the number of (X)*ApEn* improvements, alternations and adjustments to specific applications, to mention just a few ones.

A reliable (X)*ApEn* estimate requires artifact-free and stationary signals and its quadratic

computational complexity is not fit for a local analysis concept. We propose a robust modification of (cross)-approximate entropy that ensures a speedy, energy efficient and blind implementation. It is motivated by a long-known property that infinitely clipped speech signal preserves the intelligibility, i.e. property that the majority of information content is carried by signal changes. We set a hypothesis that a binary differentially encoded biomedical signals preserve sufficient information to allow approximate entropy estimation, in spite of the coding coarseness. Differential coding is known to make the signals stationary, whereas the binary representation attenuates the effect of artifacts. The binarization also allows a considerable reduction of the computational complexity, reducing the cardinality of vector sets required for the entropy estimation procedure.

Binary differentially encoded signals, coupled with discrete unconditional Shannon entropy, have already found a vast implementation in joint symbolic dynamic (JSD) studies, briefly explained within the Chapter 5.

The aims of this section is 1) to introduce and explain binary cross-approximate entropy (X)*BinEn* as a method that considerably increases the computational efficiency of entropy estimation; 2) to establish a theoretical relationship between (X)*BinEn* and classical Shannon entropy; 3) to prove that (X)*BinEn* estimates are comparable to (X)*ApEn* results and also to point out the limitations; 4) to explore the (X)*BinEn* consistency regarding the binary bias and regarding the entropy parameters known to cause (X)*ApEn* instability; 5) to append a dynamic complementary measure, based on elements embedded in (X)*BinEn* procedure.

6.1. Experimental protocol and signal acquisition

The method is evaluated using the signals from laboratory rats exposed to stress. An extensive entropy study has already been performed over these signals, so the standard to which (X)*BinEn* results can be compared already exists (Boskovic 2011). Physiological aspects of these signals are elaborated and known (Sarenac 2010). The target group for this method – the mobile crowdsensing subjects, or individual subjects with wearable sensors – are most likely to be healthy, so stress is the major adverse factor they suffer from.

Signals for this study are derived from blood pressure waveforms, recorded from the outbred male normotensive Wistar rats (NRM), as well as from the Borderline Hypertensive rats

that are F1 offspring of Wistar dames and Spontaneous Hypertensive sires (BHR), each rat weighing 330 ± 20 g. A pressure sensor with wireless transmitter (TA11PA-C40, DSI, Transoma Medical) was implanted in abdominal aorta under combined ketamine and xylazine anesthesia, along with gentamicin, followed by metamizol injections for pain relief. The recording started 10 days after the surgery, to allow the recovery. The arterial blood pressure (BP) signal was digitized at 1000Hz and relayed to a PC equipped with Dataquest A.R.T. 4.0. software for analysis of cardiovascular signals. Systolic blood pressure signal (SBP) is defined as a local blood pressure waveform maximum and pulse interval (PI) signal is extracted as a time interval between the successive points of maximal pressure.

During the experiments, the animals were exposed to two types of stress: shaker stress, with rats positioned on a platform shaking at 200 cycle/min, and restraint stress, with rats placed in a Plexiglas restrainer tube (ID 5.5 cm with pores) in the supine position. Prior to any stress exposure, baseline signals were recorded (BASE).

The number of animals per experimental group was 6 or 7, satisfactory according to the variability of the parameters in the control group rats (statistical software "Power Sample Size Calculation"). All experimental procedures in this study confirmed to European Communities Council directive of 24 November 1986 (86/609/ECC) and the School of Medicine, University of Belgrade Guidelines on Animal Experimentation.

Each one of the raw SBP and PI time series is provided with a set of fifty artificial control signals that included isodistibutional surrogate time series (randomly permuted raw signal samples) and pseudo-random time series with uniform and normal distribution, with the same mean and standard deviation as the signal they accompany. It should be noted that pre-processing, necessary for (X)*ApEn* but not for the proposed (X)*BinEn*, includes the visual inspection of time series, artifacts removal and slow signal component elimination by a filter designed specifically for cardiovascular signals. A stationarity test must also be performed: cross-*ApEn* requires normalized and centralized signals and the estimation of necessary statistical parameters needs signals stationary at least in a wide sense.

The statistical significance of the obtained results was assessed using repeated measures ANOVA test, marked with "*" for $p<0.05$ and "**" for $p < 0.01$.

6.2. XBinEn procedure

The cross-entropy of binary differentially modulated signal, or (X)*BinEn*, is a procedure that gives a blind entropy estimate without a need to preprocess the signal. It can be applied as an auto-*BinEn* that, in a single time series, estimates a likelihood of an observed pattern to preserve the similarities in the next incremental step. It can also be applied for estimating the relationship existing between a related pair of time series – cross-entropy (X)*BinEn*. The latter case is more general so it would be used for the procedure explanation.

While the (X)*BinEn* follows the basic steps of (X)*ApEn* [3], it also induces substantial modifications that are in accordance with the specific properties of the binary time series.

The (X)*BinEn* procedure is applied to two initial time series. The first one is known as a "*master signal"* $sx_i \in \mathbf{S_X}$, $i=1,\ldots,N$. The cross-entropy is estimated in respect to the "*signal – follower"* $\mathbf{S_Y}$, $sy_j \in \mathbf{S_Y}$, $j=1,\ldots,N$. The binary differential encoding is performed according to the established procedure (also given in the previous section):

$$x_i = \begin{cases} 0 & sx_{i+1} - sx_i \le 0 \\ 1 & sx_{i+1} - sx_i > 0 \end{cases} \quad i = 1,\cdots,N-1, \quad x_i \in \mathbf{X}, y_i \in \mathbf{Y}. \qquad (6.1)$$
$$y_i = \begin{cases} 0 & sy_{i+1} - sy_i \le 0 \\ 1 & sy_{i+1} - sy_i > 0 \end{cases}$$

The new binary master series **X** and its binary follower **Y** are then divided into the binary vectors of length *m*. The "*master vector"* $\mathbf{X}_m^{(i)}$ (obtained from the master binary signal **X**) is also called a "*template vector*" and the $\mathbf{Y}_m^{(j)}$ vector obtained from binary signal-follower **Y** is its "*vector- follower*":

$$\begin{aligned} \mathbf{X}_m^{(i)} &= [x_i, x_{i+\tau}, ..., x_{i+(m-1)\cdot\tau}], \quad i = 1,\cdots, N\text{-}(m\text{-}1)\cdot\tau, \\ \mathbf{Y}_m^{(j)} &= [y_j, y_{j+\tau}, ..., y_{j+(m-1)\cdot\tau}], \quad j = 1,\cdots, N\text{-}(m\text{-}1)\cdot\tau. \end{aligned} \qquad (6.2)$$

The purpose of the parameter τ (time lag) is to spread (distribute) vector elements, thus performing a vector de-correlation. Although the most applications use the contiguous $\tau=1$ vectors, for the sake of the complete description, the general $\tau \ge 1$ form is used. Time lag is a key component for adding the dynamic features into the (X)*BinEn* estimation.

In a classical (X)*ApEn* procedure the vectors are real, each one comprising *m* original real signal samples. Each vector is unique so the cardinality of set of different vectors is $N\text{-}(m\text{-}1)\cdot\tau$.

The (X)*BinEn* vectors are binary and thus of limited diversity. The cardinality of set of different *m*-bit vectors is equal to 2^m. To each binary vector $\mathbf{X}_m^{(i)} = [x_i, x_{i+\tau}, ..., x_{i+(m-1)\cdot\tau}]$ a decimal counterpart *k* can be associated, evaluated as:

$$k = \sum_{n=0}^{m-1} x_{i+n\cdot\tau} \cdot 2^n . \tag{6.3}$$

Vector histograms $N_{\mathrm{X}}^{(m)}(k)$ and $N_{\mathrm{Y}}^{(m)}(n)$ show the number of occurrences of a particular vector within the series **X** and **Y**, respectively:

$$\begin{aligned} N_{\mathrm{X}}^{(m)}(k) &= \sum_{i=1}^{N-(m-1)\cdot\tau} \mathrm{I}\left\{\sum_{l=0}^{m-1} x_{i+l\cdot\tau} \cdot 2^l = k\right\}, \quad k = 0, 1, \cdots, 2^m - 1, \\ N_{\mathrm{Y}}^{(m)}(n) &= \sum_{j=1}^{N-(m-1)\cdot\tau} \mathrm{I}\left\{\sum_{l=0}^{m-1} y_{j+l\cdot\tau} \cdot 2^l = n\right\}; \quad n = 0, 1, \cdots, 2^m - 1. \end{aligned} \tag{6.4}$$

Here I{.} denotes an indicator function that is equal to 1 if the condition it indicates is fulfilled (the observed sum is equal to *k* or to *n*) [24]. Obviously, $\sum_{k=0}^{2^m-1} N_{\mathrm{X}}^{(m)}(k) = \sum_{n=0}^{2^m-1} N_{\mathrm{Y}}^{(m)}(n) = N - (m-1)\cdot\tau$. The probability mass functions of vectors in **X** and **Y** can be estimated as:

$$\hat{P}_{\mathrm{X}}^{(m)}(k) = \frac{N_{\mathrm{X}}^{(m)}(k)}{N-(m-1)\cdot\tau}, \quad \hat{P}_{\mathrm{Y}}^{(m)}(n) = \frac{N_{\mathrm{Y}}^{(m)}(n)}{N-(m-1)\cdot\tau}, \quad k, n = 0, 1, \cdots, 2^m - 1, \tag{6.5}$$

where "^" denotes an estimate.

Following the (X)*ApEn* as a source procedure, the next step would be to quantify a difference between each pair of master-follower vectors. In structures defined over the binary field, Hamming distance is almost exclusively applied as a distance measure:

$$d(\mathbf{X}_m^{(i)}, \mathbf{Y}_m^{(j)}) = \sum_{k=0}^{m-1} x_{i+k\cdot\tau} \oplus y_{j+k\cdot\tau} = \sum_{k=0}^{m-1} \mathrm{I}\{x_{i+k\cdot\tau} \neq y_{j+k\cdot\tau}\}, \quad i, j = 1, \cdots, N-(m-1) \tag{6.6}$$

The sign $\oplus$ in (6) denotes an ex-or logical function, while I{.} denotes an indicator function. Both functions are equal to one if $x_{i+k\cdot\tau} \neq y_{j+k\cdot\tau}$, and equal to zero otherwise. The distance between binary vectors defined with (6) is a discrete variable that can get one out of *m*+1 values, i.e. $d(\mathbf{X}_m^{(i)}, \mathbf{Y}_m^{(j)}) \in \{0,1,\cdots,m\}$ Since the number of different binary vectors of length *m* is equal to 2^m, the number of different template-follower vector pairs is equal to $2^m \cdot 2^m$. Their corresponding Hamming distance can be expressed in a form of $2^m \cdot 2^m$ matrix denoted $\mathbf{H}^{(m)}$. Its element $h_{kn}^{(m)}$ is a Hamming distance between a template vector with a decimal value *k* and its vector-follower with a decimal value *n*. Equation (6.7) shows an example of a Hamming distance

matrix for m=2; the possible two-bit vectors from **Y** and from **X** are written above the matrix, and in the right, respectively:

$$\mathbf{H}^{(m)} = [h_{kn}^{(m)}] = \begin{array}{c} \begin{matrix} 00 & 01 & 10 & 11 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} 00 \\ 01 \\ 10 \\ 11 \end{matrix} \qquad m = 2, \quad k, n = 0, 1, 2, 3. \tag{6.7}$$

In equation (6.7), k and n denote the decimal values of the corresponding template and follower vectors.

For each m, a unique matrix $\mathbf{H}^{(m)}$ exists, so it can be calculated just once, stored and retrieved when necessary. Note that distance calculation for (X)*ApEn* requires m real subtractions and comparisons per each one of the $[N\text{-}(m\text{-}1)\cdot\tau]^2$ master-follower real vector pairs:

$$d_{\mathrm{REAL}}(\mathbf{X}_{XApEn\,m}^{(i)}, \mathbf{Y}_{XApEn\,m}^{(j)}) = \max_{k=0,..,m-1} \left[|sx(i+k\cdot\tau) - sy(j+k\cdot\tau)| \right], \quad i, j = N\text{-}(m\text{-}1)\cdot\tau\,. \tag{6.8}$$

The next step is estimating a probability $\hat{p}_k^m(r)$ that a vector-follower is within the distance r from a particular vector-template $\mathbf{X}_m^{(k)}$, i.e. that a vector-follower "*match*" the observed template. For this estimation it is sufficient to know a vector-follower histogram $N_{\mathrm{Y}}^{(m)}(n)$ and Hamming distance matrix $[h_{kn}^{(m)}]$. The matrix gives an information which vectors, out of 2^m possible ones, are within the distance r from the template vector $\mathbf{X}_m^{(k)}$, while a sum of their histograms gives the number of vectors that satisfy this criterion:

$$\hat{p}_k^m(r) = \Pr\{d(\mathbf{X}_m^{(k)}, \mathbf{Y}_m) \le r\} = \frac{1}{N-(m-1)\cdot\tau} \sum_{n=0}^{2^m-1} N_{\mathrm{Y}}^{(m)}(n) \cdot \mathrm{I}\{h_{kn}^{(m)} \le r\} = \sum_{n=0}^{2^m-1} \hat{P}_{\mathrm{Y}}^{(m)}(n) \cdot \mathrm{I}\{h_{kn}^{(m)} \le r\}. \tag{6.9}$$

The corresponding probabilities in (X)*ApEn* are evaluated as follows:

$$\hat{p}_{XApEni}^m(r) = \frac{1}{N-(m-1)\cdot\tau} \sum_{j=1}^{N-(m-1)\cdot\tau} \mathrm{I}\{d_m(\mathbf{X}_{XApEn\,m}^{(i)}, \mathbf{Y}_{XApEn\,m}^{(j)}) \le r\}, \quad i = 1, \cdots, N-(m-1)\cdot\tau. \tag{6.10}$$

From the Eqs. (6.7-6.10) the computational reduction achieved by (X)*BinEn* becomes obvious: the number of probabilities (Eq. (6.10)) in (X)*ApEn* is equal to $N-(m-1)\cdot\tau$ (tens of thousands) and for each one $m\cdot(N-(m-1)\cdot\tau)$ amplitude subtractions and comparisons (Eq. (6.8)) must be done; (X)*BinEn*, on the other hand, operates with histograms so the number of probabilities is equal to 2^m (Eq. (6.9)) with distances retrieved from the Hamming distance matrix (6.7).

The final step in (X)*ApEn*-like procedures is forming a summand Φ that averages logarithms of the estimated probabilities over all 2^m templates. Each template k within the series **X** occurs $N_{\mathrm{X}}^{(m)}(k)$ times:

$$\hat{\Phi}^{(m)}(r,N,\tau)=\frac{1}{N-(m-1)\cdot\tau}\sum_{k=0}^{2^m-1}N_{\mathbf{X}}^{(m)}(k)\cdot\ln(\hat{p}_k^m(r))=\sum_{k=0}^{2^m-1}\hat{P}_{\mathbf{X}}^{(m)}(k)\cdot\ln(\hat{p}_k^m(r)). \quad (6.11)$$

The complete procedure is than repeated for the vectors of length m+1, yielding the estimate:

$$\hat{\mathrm{X}}BinEn(m,r,N,\tau)=\hat{\Phi}^{(m)}(r,N,\tau)-\hat{\Phi}^{(m+1)}(r,N,\tau). \quad (6.12)$$

For (X)*ApEn*, summand Φ requires $(N-(m-1)\cdot\tau)$ logarithms and summations:

$$\hat{\Phi}_{\mathrm{X}ApEn}^{(m)}(r,N,\tau)=\frac{1}{N-(m-1)\cdot\tau}\cdot\sum_{i=1}^{N-(m-1)\cdot\tau}\ln\left(\hat{p}_i^m(r)\right). \quad (6.13)$$

If a single time series **X** is observed, than (X)*BinEn* procedure converts to its auto *BinEn* variant, where the template vector and the follower vector are both from the same signal. Formally, it requires the changes in equations (6.6), (6.8) and (6.9) where the notation referring

Phase	(X)*ApEn*	(X)*BinEn*
Pre-processing	Visual inspection and semi-automatic artifacts removal; Trend removal; Stationarity test;	None;
Start	Standard deviation (for *r*): *N*·(summation, subtraction and multiplication); Normalization and centralization (for cross-entropy only): *N*·(division, summation and subtraction);	Binary differential encoding: *N*·(subtraction and comparison);
Vector histogram and distribution	None;	Eq. 6.4: 2·*Q*·(*m* summations and *m* multiplications, comparison, division);
Difference	Eq. 6.8*:* $m\cdot Q^2$·(subtraction and comparison);	None; for a given *m* Hamming matrix is always the same;
Probability	Eq. 6.10*:* Q^2 comparisons;	Eq. 6.9: $2^m\cdot 2^m$·(multiplication, summation and comparison);
Summand Φ	Eq. 6.13: Q·(logarithm and summation);	Eq. 6.11: 2^m·(logarithm, multiplication and summation).

Table 6.1: Number of operations required for one and one estimation

to the "**Y**" time series should be replaced with the notations that correspond to the "**X**" time series.

Table 6.1 summarizes the necessary operations required by (X)*BinEn* and by (X)*ApEn*. Within this table, the abbreviation $Q = N-(m-1)\cdot\tau$ is used. Time series length N (and number of vectors Q) are measured in thousands of samples, while the vector length m is typically set to 2 or 3. The complexity of (X)*ApEn* is quadratic, while (X)*BinEn* has a linear complexity considering N (and Q). Additional reduction is achieved by substituting the number of different real vectors Q by number of different binary vectors 2^m within the two last phases in Table 6.1.

6.3. Relation to Shannon entropy

The core of (X)*ApEn* is an Eckmann-Ruelle (E-R) entropy of a single time series, for which an auto-*ApEn* is an empiric approximation (thus the name – approximate entropy). E-R entropy implies infinite data sets N; infinite vectors m and a threshold r that is equal to zero. None of these bounds can be reached in (X)*ApEn* that is defined over the real-valued data series. But, due to the limited cardinality of the (X)*BinEn* vectors, the distance between the template and the corresponding vectors followers, defined by (6.6), is discrete, including zero as a possible threshold value. Threshold r equal to zero occurs only if a template vector and a follower vector are equal, i.e. when $n=k$ in (6.9). Then the probability $\hat{p}_k^m(r=0)$ (6.9) is equal to the probability of a particular vector with decimal value k:

$$\hat{p}_k^m(r=0) = \sum_{n=0}^{2^m-1} \hat{P}_{\mathbf{Y}}^{(m)}(n)\cdot \mathrm{I}\{h_{kn}^{(m)}=0\} = \hat{P}_{\mathbf{Y}}^{(m)}(k)\cdot \mathrm{I}\{h_{kk}^{(m)}=0\} = \hat{P}_{\mathbf{Y}}^{(m)}(k). \tag{6.14}$$

If, additionally, $\mathbf{Y} = \mathbf{X}$ (auto-*BinEn*), $\hat{P}_{\mathbf{Y}}^{(m)}(k)$ converts to $\hat{P}_{\mathbf{X}}^{(m)}(k)$ and the expression (6.11) becomes:

$$\hat{\Phi}^{(m)}(r=0,N,\tau) = \sum_{k=0}^{2^m-1} \hat{P}_{\mathbf{X}}^{(m)}(k)\cdot \ln(\hat{P}_{\mathbf{X}}^{(m)}(k)) = -H_{Sh}^{(m)}. \tag{6.15}$$

Equation (6.15) states that a zero-threshold summand Φ and Shannon entropy are closely related. Equation (6.15) also states that auto-*BinEn* corresponds to the difference of Shannon entropies estimated for the vectors of incremental lengths (vectors of length m and m+1), as follows:

$$BinEn(r=0,N,m,\tau)=\hat{\Phi}^{(m)}(r=0,N,\tau)-\hat{\Phi}^{(m+1)}(r=0,N,\tau)=$$
$$=\sum_{k=0}^{2^m-1}\hat{P}_{\mathrm{X}}^{(m)}(k)\cdot\ln(\hat{P}_{\mathrm{X}}^{(m)}(k))-\sum_{k=0}^{2^{m+1}-1}\hat{P}_{\mathrm{X}}^{(m+1)}(k)\cdot\ln(\hat{P}_{\mathrm{X}}^{(m+1)}(k))=\quad H_{Sh}^{(m+1)}-H_{Sh}^{(m)}. \tag{6.16}$$

The relations (6.15) and (6.16) do not hold for a general (X)*BinEn* case, as all the comprising probabilities must be estimated from the same time series, which is true in auto-entropy case only (Eqs. (6.9) and (6.11)).

A known property of discrete Shannon entropy is that the entropy difference is equal to the entropy of incremental symbol set, i.e. to the entropy of m=1 vectors :

$$H_{Sh}^{(m+1)}-H_{Sh}^{(m)}\quad=\quad H_{Sh}^{(1)},\quad\forall m>0. \tag{6.17}$$

Proof: Suppose that m-bit symbols (or m-bit vectors) are marked S_i, i=0,...,2^m-1, with the corresponding symbol probability $P(S_i)$; a bit longer symbols are V_j, j=0,...,2^{m+1}-1, with the corresponding symbol probability $P(V_j)$; an m=1 vector comprise a single bit b_k, k=0,1 with the corresponding bit probability $P(b_k)$. Than, for statistically independent bits and symbols, it holds:

$$H(V)=-\sum_{i=0}^{2^{m+1}-1}P(V_{\mathrm{i}})\cdot\ln(P(V_{\mathrm{i}}))=\sum_{k=0}^{1}\sum_{ij=0}^{2^m-1}P(b_k)\cdot P(S_j)\cdot\ln(P(b_k)\cdot P(S_j))=$$
$$=\sum_{k=0}^{1}\sum_{ij=0}^{2^m-1}P(b_k)\cdot P(S_j)\cdot\ln(P(b_k))\;+\sum_{k=0}^{1}\sum_{ij=0}^{2^m-1}P(b_k)\cdot P(S_j)\cdot\ln(P(S_j))=$$
$$=\left(\sum_{j=0}^{2^m-1}P(S_j)\right)\cdot\sum_{k=0}^{1}P(b_k)\cdot\ln(P(b_k))+\left(\sum_{k=0}^{1}P(b_k)\right)\cdot\sum_{j=0}^{2^m-1}P(S_j)\cdot\ln(P(S_j))=$$
$$=1\cdot H(b)+1\cdot H(S)=H(S)+H(b).$$

Q.E.D.

From Eq. (6.17) it could be concluded that auto-*BinEn* (6.16) would always remain the same if the threshold is equal to zero, but it is true for statistically independent data only. If data are dependent, Shannon entropy (6.15) is just an upper bound of the true entropy value.

6.4. Binary bias as a possible source of errors

The differential encoding defined by (6.1) forces the appearance of binary symbol “0”. It is negligible in SBP signals, but it causes 6% of the adjacent PI signal samples to be exactly the same (Table 6.2).

EXPERIMENT	[%] of SBP	[%] of PI
BHR SHAKER	0.35±0.13	5.62±2.08
BHR RESTRAINT	0.35±0.11	5.79±1.52
NRM SHAKER	0.41±0.19	6.10±1.52
NRM RESTRAINT	0.36±0.14	6.64±2.06

Results are expressed as mean ± standard deviation.

Table 6.2. Binary bias - percentage of adjacent equivalent samples

Binary bias is a consequence of temporal resolution $\Delta T=1$ ms that, within a physiological range of PI signals in rats, allows a limited number of different PI values. This seemingly contradicts the sampling theorem that guarantees a perfect waveform reconstruction. The alleged paradox is clarified recalling that a strict implementation of the sampling theorem requires a low-pass filtering that increases temporal resolution, as a simple example in Fig. 6.1 shows. An increase of temporal resolution can be done by interpolation, or by associating a random binary symbol to each occurrence of equal adjacent samples in (6.1).

To explore whether the binary bias affects the entropy estimate, (X)*BinEn* is applied to three types of signals. Within the first type PI signals were derived from the original time series; within the second type, PI signals were derived from the filtered time series. Within the third type, a random correction was done during the binarization procedure (1) and a bit associated to the equal successive PI sample got a random binary value. The results are presented in Table 6.3, showing that there is no significant differences between the three types of results, so no binary bias correction is actually necessary.

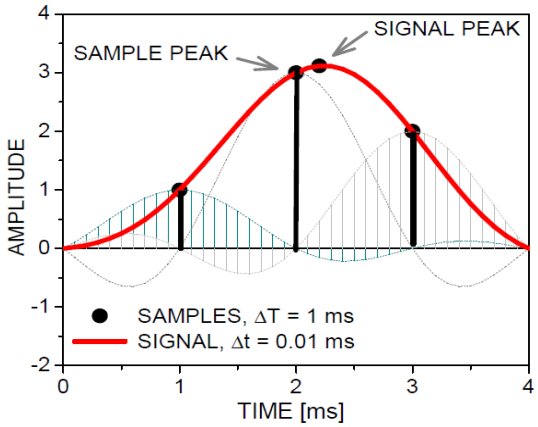


Fig 6.1. Binary bias eliminated by NF filtering: the maximum of the filtered signal (thick red line) is shifted 0.02 ms in respect to the position of maximal signal sample provided by the acquisition equipment (black vertical lines)

EXPERIMENT	(X)*BinEn* ORIGINAL SIGNAL	NF FILTER	RANDOM CORRECTION
BHR SHAKER BASE	0.602±0.015	0.605±0.014	0.603±0.014
BHR SHAKER STRESS	0.566±0.024	0.566±0.020	0.567±0.022
BHR RESTRAINT BASE	0.575±0.026	0.580±0.026	0.582±0.022
BHR RESTRAINT STRESS	0.521±0.032	0.544±0.048	0.547±0.048
NRM SHAKER BASE	0.571±0.035	0.584±0.033	0.582±0.035
NRM SHAKER STRESS	0.539±0.022	0.544±0.024	0.542±0.021
NRM RESTRAINT BASE	0.601±0.016	0.606±0.016	0.608±0.015
NRM RESTRAINT STRESS	0.480±0.089	0.513±0.095	0.508±0.091

Parameters: $m = 2$, $r = 1$, $\tau = 1$, $N = 3000$; results are presented as mean ± SE

Table 6. 3. Influence of binary bias: (X)*BinEn* is estimated from the time series derived from the original signal, from the NF-filtered signal, and from the randomly corrected time series.

6.5. Entropy estimated by (X)*BinEn* and (X)*ApEn*

The source signal for (X)*BinEn* estimation is a coarsely coded signal with two amplitude levels per sample. The source PI and SBP signals for (X)*ApEn* estimation are derived from the

output of 12-bit A/D convertor with 4096 amplitude levels. The assumption is that binary coding preserves a sufficient amount of information, and that the results of (X)B*inEn* would be comparable to (X)*ApEn*. This assumption follows a speech signal studies, where it was shown that binary coding preserves signal information, and that speech remains comprehensible although the voice becomes harsh. To prove this assumption, (X)*BinEn* is estimated from the same signals from laboratory rats submitted to stress, for which (X)*ApEn* results already exist.

The relative (X)*BinEn* and (X)*ApEn* changes induced by stress in respect to baseline conditions are presented in Fig. 6.2. While (X)*BinEn* was applied to the raw signals, prior to (X)*ApEn* analysis signals had to be checked, the artifacts removed, slow-varying trend component filtered out, the stationarity test applied and data normalized and centralized.

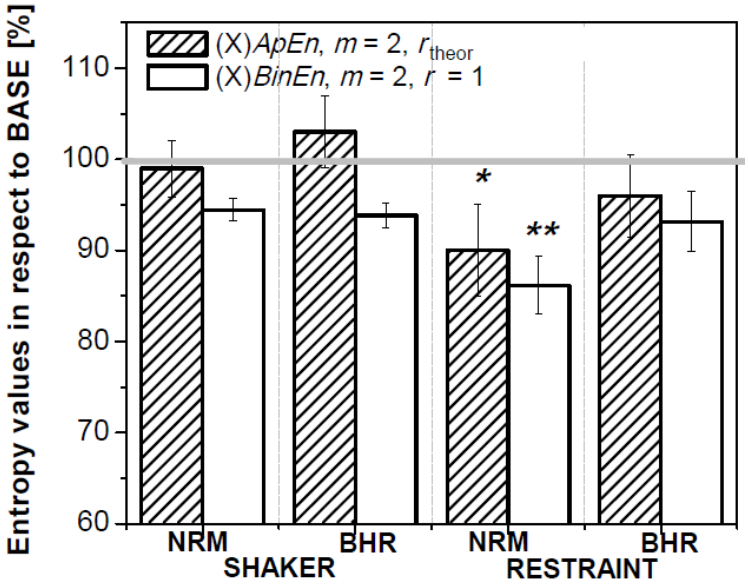


Fig. 6.2. (X)*ApEn* and (X)*BinEn* for N=5000 and τ = 1; the results are presented relative to baseline (horizontal gray line)

A formal statistical comparison yielded no significant differences between the (X)*BinEn* and (X)*ApEn* results. When tested in respect to baseline conditions (thick gray line in Fig. 6.2), both methods exhibited a statistically significant relative entropy decrease in normotensive rats exposed to restraint stress only. In other cases and estimated by both methods, relative entropy change in respect to baseline remained without the significant change. Mutually comparing

relative (X)*ApEn* and (X)*BinEn* changes, no statistically significant difference between the methods is obtained. A visual inspection shows a slight, but insignificant, increase of (X)*ApEn* in BHR rats exposed to shaker stress, and a slight, but insignificant, decrease of (X)*BinEn* applied to the same signal. This discrepancy is an outcome of parameter instability of (X)*ApEn* and would be addressed to in the next subchapter.

However, the binary differential coding is coarse, so reduced levels of freedom (number of amplitude levels decreased from 4096 to 2) must elicit some consequences. And indeed, the major drawback of (X)*BinEn* can be seen in Fig. 6.3, where the entropy changes (in respect to

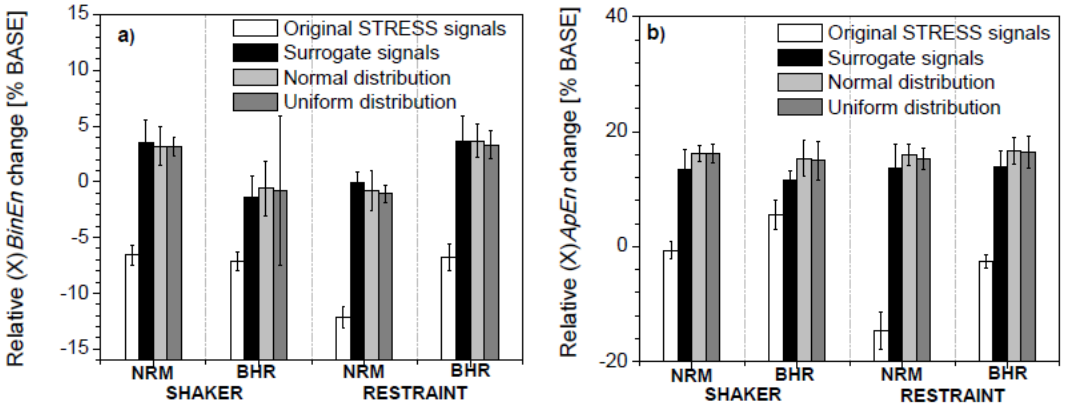


Fig. 6.3. Relative entropy changes in respect to the baseline signal a)(X)*BinEn*; b) (X)*ApEn*;, In both panels N = 3000. Other parameters are according to the general guidelines - m = 2, r = 0.2 and τ = 1;

baseline) are estimated for the signals in stress, and for three types of accompanying artificial signals with the same mean and standard deviation: isodistributional surrogate data, pseudorandom signals with normal distribution and pseudorandom signals with uniform distribution. In Fig. 6.3, left panel (a) corresponds to (X)*BinE*n estimates, and right panel (b) corresponds to (X)*ApEn* estimates.

At the first glance, both panels are similar in appearance and display similar behavior. Only after observing the scale of y-axis a substantial difference can be noticed: the evoked (X)*BinEn* changes correspond to the dynamic range of 15%, while the changes evoked in (X)*ApEn* correspond to the dynamic range of 40%. It follows that (X)*BinEn* would be insensitive to fine signal changes that (X)*ApEn* would notice, i.e. (X)*BinEn* is a rough measure. This also explains the inability of (X)*BinEn* to notice the small changes induced by binary bias: signal space is reduced to a simple binary decision whether the next sample increases or decreases in

respect to the current one, but the absolute value of change is lost. So the coarse binary coding does preserve the majority of signal information located in binary direction of signal changes. This enables (X)*BinEn* to be comparable to (X)*ApEn*. However, information stored within an amplitude of signal change is lost and cannot be reflected to (X)*BinEn* estimates, resulting in decreased (X)*BinEn* sensitivity.

6.6. Threshold r and time series length N as possible sources of inconsistency

Approximate entropy is always implemented as a relative measure, as its absolute value is influenced by parameters. It is known, for example, that (X)*ApEn* increases with the time series length, asymptotically reaching a plateau that requires a large number of template-follower matching pairs to be reliably achieved. However, parameters can also induce an experimental instability. Threshold *r* is shown to be its dominant source and it was subject to numerous studies. It was pointed out that it can even alter experimental outcome and lead to the physiologically incorrect conclusions (a flip-flop effect).

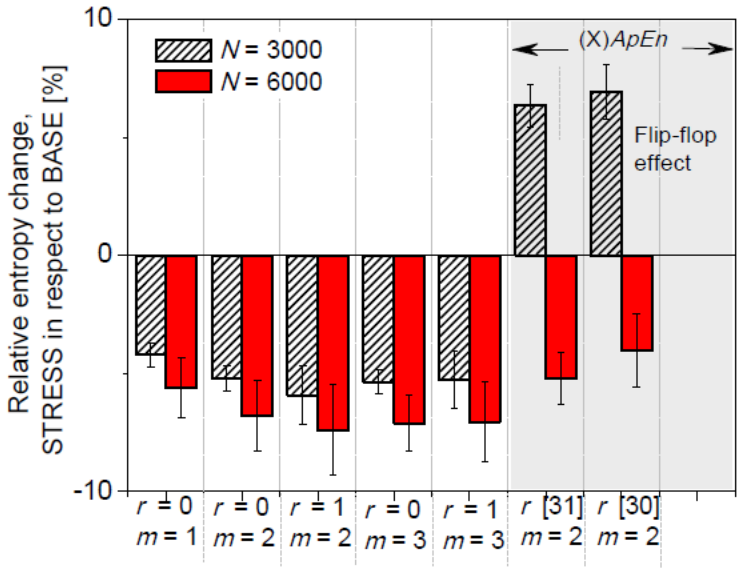


Fig. 6.4. (X)*BinEn* changes induced by SHAKER stress in BHR rats. The experimental outcomes are consistent for different thresholds *r*, vector lengths *m*; and time series lengths; (X)*ApEn* outcomes are incosistent, increasing in stress for *N*=3000 and decreasing for *N* = 6000 (gray part of figure).

To test whether the instabilities observed in (X)*ApEn* remain in (X)*BinEn*, both methods were applied to a set of signals where a flip-flop effect has already been recorded – signals from BHR rats exposed to shaker stress. The results are presented in Fig. 6.4. (X)*BinEn* was tested for all possible parameters and the results consistently show that entropy decreases in stress in respect to baseline entropy. (X)*ApEn* estimates (in the gray part of Fig. 6.4) show that the experimental outcome depends on parameter choice, and, depending on parameter choice, and that entropy in stress may both increase and decrease.

An influence of time series length *N* is shown in Fig. 6.5. Contrary to (X)*ApEn* that monotonously increase within the given range of lengths *N*, the absolute values of (X)*BinEn* estimates does not show any fluctuations, and they are also insensitive to a slow signal component (trend) – panels (a) and (b). Panels (c) and (d) show that the relation between the BHR and NRM rats remain the same regardless both if estimated by (X)*ApEn* and (X)*BinEn.*

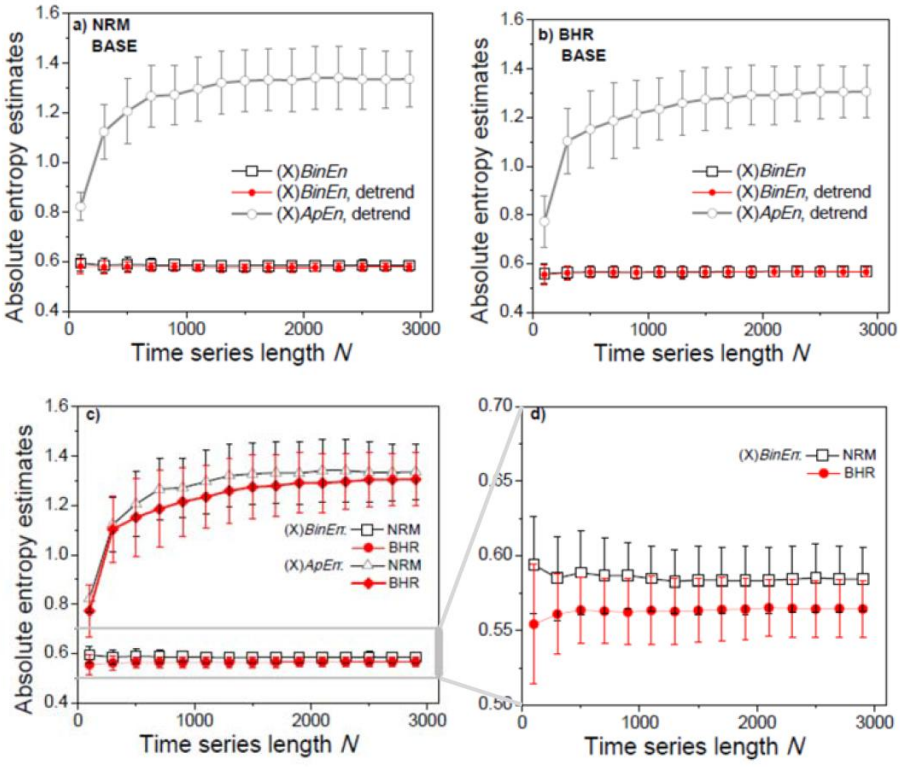


Fig. 6.5. Influence of time series length. (X)BinEn(m=2,r=1,τ=1) is not influenced by the signal trend (panels a) and b)). Estimated (X)ApEn(m=2,r=0.15, τ =1) values are smaller in BHR rats, but without the statistical significance (panel c)); the same relation is preserved in (X)BinEn (panel d)).

As already stated, (X)*ApEn* requires a large number of template-follower vector matching pairs, otherwise the probabilities (6.10) are underestimated and sum (6.13) is not reliable. This problem is attenuated by binary coding, as template-vectors have at most 2^{m+1} different realizations (as compared to $(N-(m-1)\cdot\tau)$ individual ones in (X)*ApEn* estimation), accomplishing a sufficient number of matching vector pairs for. Furthermore, for discrete vectors theoretical guidelines can be used for finding the minimal time series length necessary for reliable estimation:

$$N_{\mathrm{MIN}} = \frac{10^{l}}{\left(\min(p_0, p_1)\right)^{m+1}}, \quad p_0 + p_1 = 1, \quad l \in \{1,2\} \tag{6.18}$$

Here p_0 and p_1 are probabilities of 0 and 1 respectively and the choice of l distinguishes whether the event would appear at least 10 (l=1) or 100 (l=2) times during the experiment. Equation (6.18) considers the worst-case, when a template vector consists of m+1 identical and the least probable binary symbols so the denominator in (6.18) has the lowest value.

An additional test of (X)*BinEn* stability regarding the data length is performed using the artificially generated random binary time series with time series length N and probability of zero p_0 as parameters. The results are presented in Fig. 6.6.

The left panel (a) shows the estimated (X)*BinEn* for m=2, r=1 and τ=1, averaged over 1000 simulation runs and presented as mean ± standard deviation. The right panel (b) shows standard deviation (in respect to entropy mean value) with the values for N given by (6.18) emphasized. As biomedical signals are under physiological constraints and can neither increase nor decrease infinitely, the corresponding binary probability is in the vicinity of 0.5. So the reliable (X)*BinEn* estimation of cardiovascular time series is ensured with N = 500 signal samples.

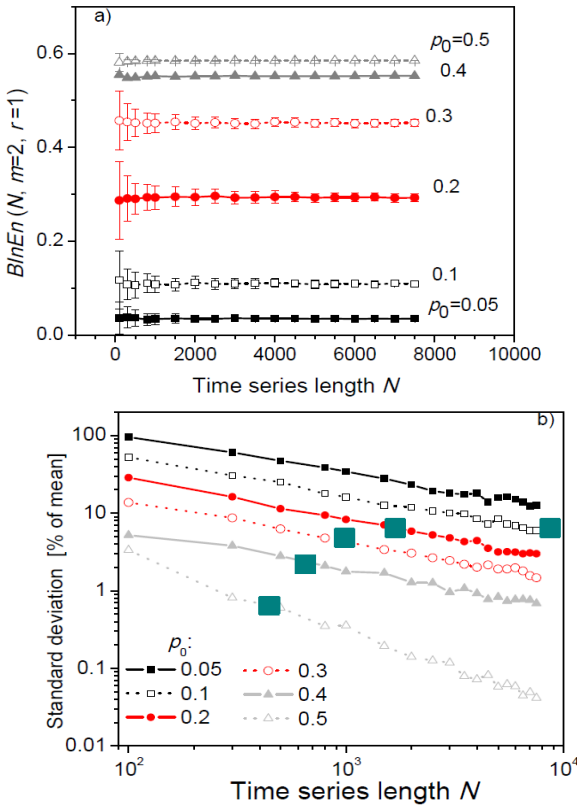


Fig. 6.6. a) Estimated (X)*BinEn*(*N*,*m*=2,*r*=1) of artificially generated random binary time series (1000 runs per point); parameter is probability of binary zero p_0; b) standard deviations from panel a) expressed in % of mean; squares mark the length for which N=N_{MIN} (Eq. 6.18).

6.7. Components for the dynamic measures

All (X)*ApEn*–based methods yield a result expressed as a single number. Such a compact outcome is an advantage through which (X)*ApEn* earned its meritorious status. This convenience, however, requires an extensive averaging which conceals the subtle inter-sample relationships and makes (X)*ApEn* a static measure, as was indeed the original intention of the authors: "*the intuition motivating* (X)*ApEn*" is actually designed to compare the joint and marginal distribution functions, therefore not the sample inter-relationships.

A necessity to include the inter-sample dependence initiated the follow-up studies, widely accepted to complement (X)*ApEn*. The underlying theory is based on conditional entropy. With Shannon entropy already known to be equal to (auto)*BinEn*(*r*=0), (Chapter 6.3), appending a conditional entropy component to a static (auto)*BinEn* measure is computationally effortless: the

elements for conditional entropy $H^{(m=1)}(\mathbf{X}\,|\,\mathbf{X})$ are already embedded into the equations (6.4) and (6.5) as follows:

$$
\begin{aligned}
&H^{(m=1)}(\mathbf{X}\,|\,\mathbf{X}) = -\sum_{k=0}^{1} P(x_k)\cdot\sum_{n=0}^{1} P(x_n\,|\,x_k)\cdot\ln(P(x_n\,|\,x_k)) = -\sum_{k=0}^{1}\sum_{n=0}^{1} P(x_n,x_k)\cdot\ln(P(x_n\,|\,x_k))\\
&P(x_k) \quad = \quad N_{\mathrm{X}}^{(1)}(k)/N, \quad k \;=\; 0,1 \qquad (\text{Eqs. (6.4) and (6.5) for } m=1)\\
&P(x_n,x_k) \quad = \quad N_{\mathrm{X}}^{(1)}(l)/(N-\tau), \quad k,n \;=\; 0,1; \quad l=k+2n \quad (\text{(6.4) and (6.5) for } m=2 \text{ and time lag } \tau)\\
&P(x_n\,|\,x_k) \quad = P(x_n,x_k)/P(x_k), \quad k,n \;=\; 0,1.
\end{aligned}
\tag{6.19}
$$

In (6.19) $P(x_n\,|\,x_k)$ denotes a conditional probability.

A key component for dynamic measures is a decorrelation time lag τ. It induces a temporal distance between the observed *m*=2 vector elements (Fig. 6.7), thus enabling the complexity measure related to samples at a specific distance along the time axis.

The proposed measure is *m*=1 conditional entropy $H^{(m=1)}(\mathbf{X}\,|\,\mathbf{X})$ normalized by *BinEn*(*r*=0). It shows complexity changes induced by inter-sample relationships at particular time lags, in respect to the complexity of the same signal, but with assumed statistically independency. This measure has a theoretical background for auto-entropy case only, as it requires all the probabilities to be estimated from the same time series.

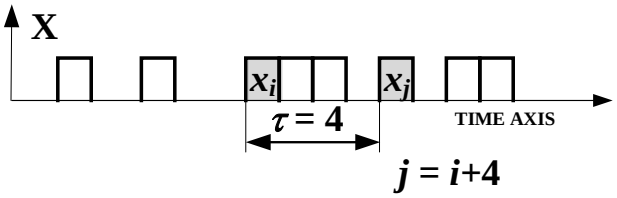


Fig. 6.7. An *m*=2 vector; its bits (gray) are at the time lag $\tau = 4$

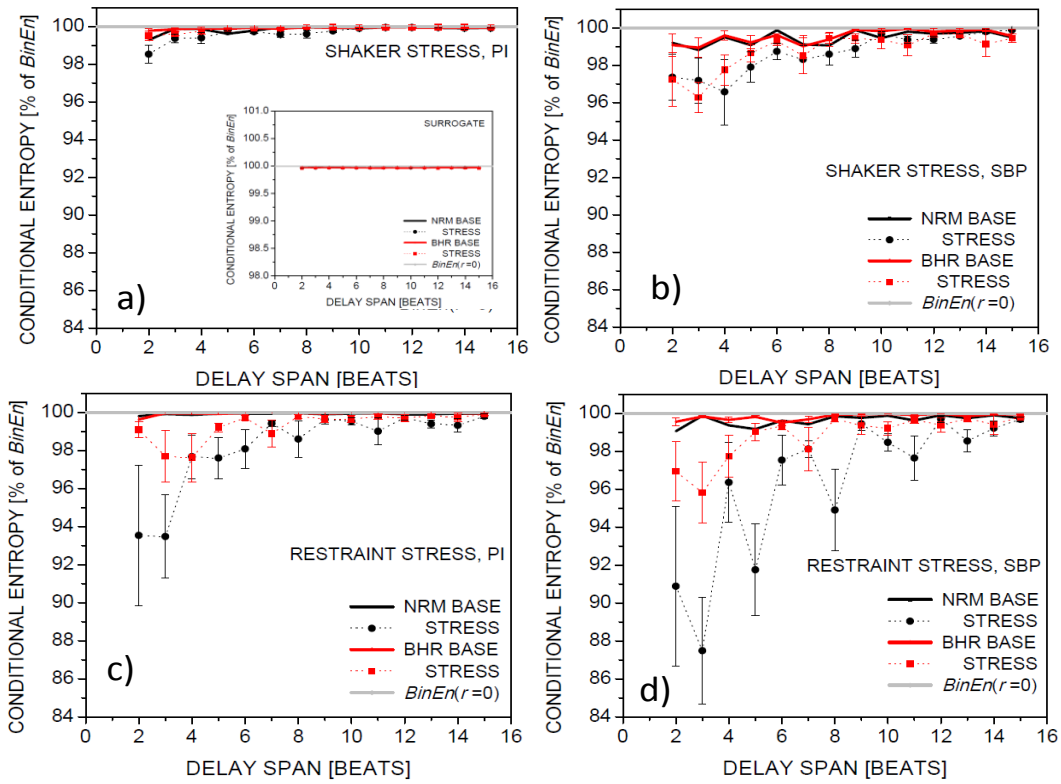


Fig. 6.8. Normalized conditional entropy at different time lags: PI signals in shaker (a) and restraint (c) stress, and SBP signals in shaker (b) and restraint (d) stress

Figure 6.8 presents an application example, a normalized conditional entropy applied to signals recorded from stressed subjects. Results for $\tau = 1$ are not presented, as the statistical dependence of adjacent samples is a consequence of memory induced by differential coding, and therefore incompatible to the other results.

At baseline conditions, the physiological regulators maintain in the perfect order and there is either no difference between the conditional entropy and *BinEn* (PI signals), or the difference is less than 1% (SBP signals). Shaker stress does not affect conditional entropy of PI signals, but an entropy decrease in SBP signals indicates that the amplitude of 10 successive beats become dependent. The major changes – up to 15% - are observed in restraint stress, one of the strongest stress an animal can be exposed to. Increased statistical dependence and lessening the effect of regulations affects – in SBP case – 15 adjacent beats. It can be observed, both in PI and in SBP, that effects of stress are more indicative in normotensive, than in border-line hypertensive rats. The inset in Fig. 6.5a shows the conditional entropy estimated from the

isodistributional surrogate data set applied as control for PI signals in shaker stress. The results are in a form of flat lines that overlap, as a random permutation of signal samples perfectly destroys inter-sample dependency and entropy always yields the same estimate.

The entropy behavior - decrease in stress and stability at baseline - is in accordance with the findings that stress attenuates regulatory mechanisms in rats. In baseline conditions, regulations are firm and SBP and PI are constantly adjusted, achieving a dynamic equilibrium. In stress, the regulation is loose, signal ramps (successive increase or decrease of signal samples) are not interrupted causing the samples to become more dependent. The dependence is not bounded to neighboring samples only, it can spread to the samples that are at time lags of ten or more heart-beats.

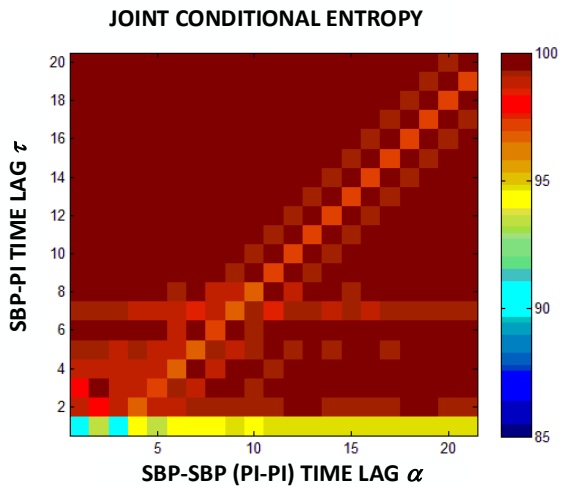


Fig. 6.9. Normalized joint conditional entropy at different PI-PI (and SBP-SBP) delay spans (horizontal axis), plotted for various SBP-PI time lags.

Another illustrative example is presented in Fig. 6.9. reveals the influence of both the delay spans to the joint conditional binary entropy. In spite of coding coarseness, the SBP-PI dependences known to exist at time lags are noticed, shown by lighter horizontal lines at lags of 2, 4, 5 and 7 [heart-beats]. However, a yellow horizontal line is an artifact (due to the differential coding), while the diagonal lines are an outcome of statistical dependence caused by overlapping bits of symbols and should be considered as artifacts as well. It is an occurrence similar to symbol overlapping in Joint Symbolic Dynamics mentioned in Chapter 5.

6.8. References

[1] Ganti, R.K. Mobile crowdsensing: current state and future challenges, IEEE Communications Magazine, 2011, Volume 49, No. 11, pp. 32 – 39 doi: 10.1109/MCOM.2011.6069707

[2] Chessa, S., Corradi, A., Foschini L., Girolami, M. Empowering Mobile Crowdsensing through Social and Ad Hoc Networking IEEE Communications Magazine, 2016, Volume 54, No. 7 pp. 108-114, doi: 10.1109/MCOM.2016.7509387

[3] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc Nat. Acad. Sci. USA, 88 (1991) 2297-2301,

[4] A.L. Goldberger, S.M. Pincus, Physiological time-series analysis: What does regularity quantify?, Am J Physiol(Heart Circ Physiol), 266 (1994) H1643-H1656.

[5] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am J Physiol Heart Circ Physiol, 278(6) (2000) H2039-H2049.

[6] D. E. Lake, J.S. Richman, Sample entropy analysis of neonatal heart rate variability, Am J Physiol Regul Integr Comp Physiol, 283 (2002) R789-R797

[7] M. Costa, A. Goldberger, C-K. Peng: Multiscale entropy analysis of biological signals, Phys. Rev. E, Vol.71(2), pp.021906 1-17 1-4, 2005.

[8] H.B. Xie, J.Y. Guo, Y.P. Zheng: Fuzzy approximate entropy analysis of chaotic and natural complex systems: detecting muscle fatigue using electromyography signals, Ann Biomed Eng. Vol,.38(4), pp 1483-96, 2010.

[9] A. Porta, G. Baselli, F. Lombardi, N. Montano, A. Malliani, S. Cerutti, Conditional entropy approach for the evaluation of the coupling strength, Biol. Cybern. 81, 119-129 (1999)

[10] M. Baumert, V. Baier, S. Truebner, A. Schirdewan, A. Voss: Short- and Long-Term Joint Symbolic Dynamics of Heart Rate and Blood Pressure in Dilated Cardiomyopathy, IEEE Transactions on Biomedical Engineering, Vol. 52, no. 12, December 2005, pp 2112-2115.

[11] O. Mohamoud, T. Loncar-Turukalo, N. Japundzic-Zigon, D. Bajic: On Symbol Entropy of Joint Time Series, Proceedings of the 7th ESGCO 2012, april 22-15, 2012, Kazimierz Dolny, Poland

[12] R.B. Govindan, J.D.Wilson, H. Eswaran, C.B. Lowery, H. Preisl, Revisiting sample entropy analysis, Physica A, 376 (2007) 158-164

[13] F. Kaffashi, R.Foglyano, C.G.Wilson, K.Loparo, The effect of time delay on approximate and sample entropy calculations, Physica D 237 (2008), 3069-3074;

[14] X.Chen, I.C.Solomon, K.H.Chon, Comparison of the use of approximate entropy and sample entropy: application to neural respiratory signal, in Proc. of the 27th IEEE EMBS Ann. Conf (2005) 4212-4216

[15] A. Papoulis, Probability, random variables and stochastic processes, McGraw-Holl International Edition, (1984)

[16] T. Buchner, G. Grzyb, P. Krzesinsky, M. Grzeda: Time frequency analysis of the tilt test recordings discussion, NEUROCARD 2010. Abstract book and discussion

[17] J. Oosting, HAJ Struijker-Boudier, BJA Janssen BJA. 1997. Validation of a continuous baroreceptor reflex sensitivity index calculated from spontaneous fluctuations of blood. J Hypertens 15:391-399, 1997.

[18] D. Bajic, T. Loncar-Turukalo, S. Stojicic, O. Sarenac, T. Bojic, D. Murphy, J. F. R. Paton, N. Japundzic-Zigon: Temporal analysis of the spontaneous baroreceptor reflex during mild emotional stress in the rat, STRESS-The International Journal On The Biology Of Stress, March 2010; Vol 13, No. 2, pp 142–154

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

[20] Pincus, S. M. Approximate entropy as a measure of system complexity. Proc. Natl Acad. Sci. U.S.A. 88:2297−2301, 1991.

[21] Pincus, S. M. Approximate entropy (ApEn) as a complexity measure. Chaos 5:110–117, 1995.

[22] Pincus, S. M. and B. H. Singer. Randomness and degrees of irregularity. Proc. Natl Acad. Sci. U.S.A. 93:2083−2088, 1996.

[23] Yentes, J. M., N. Hunt, K. K. Schmid, J. P. Kaipust, D. Mcgrath, and N. Stergiou. The appropriate use of approximate entropy and sample entropy with short data sets. Ann. Biomed. Eng. 41: 349−365, 2013.

[24] Richman, J. S., and J. R. Moorman. Physiological time series analysis using approximate entropy and sample entropy. Am. J. Physiol. Heart Circ. Physiol. 278:H2039–H2049, 2000.

[25] Boskovic A., T. Loncar-Turukalo, O. Sarenac, N. Japundzic-Zigon, and D. Bajic. Unbiased entropy estimates in stress: a parameter study. Comput. Biol. Med. 42: 667–679, 2012. (Computers in Biology and Medicine-proveriti skracenicu casopisa)

[26] Kaffashi, F., R. Foglyano, C. G. Wilson, and K. A. Lopar. The effect of time delay on approximate and sample entropy calculations. Physica D 237:3069–3074, 2008.

[27] AL Fawe Interpretation of infinitely clipped speech properties Audio and Electroacoustics, IEEE Transactions on (Volume:14 , Issue: 4) Pp 178 - 183 ISSN : 0018-9278 Dec 1966

[28] RW Wolf: Stochastic modeling and theory of queues, Prentice Hall, Englewood Hills, New Jersey, 1989, pp 35-38 ISBN-13: 978-0138466923

[29] Sarenac, O., M. Lozic, S. Drakulic, D. Bajic, J. F. Paton, D. Murphy and N. Japundzic-Zigon. Autonomic mechanisms underpinning the stress response in borderline hypertensive rats. Exp. Physiol. 96:574-589, 2011. (Experimental Physiology-)

[30] Omer Mohamoud, Tamara Skoric, Nina Japundzic Zigon and Dragana Bajic: Entropy relations in infinitely clipped differential time series, 9th conference of the European Study Group on Cardiovascular Oscillations ESGCO 2016, Lancaster, UK.

[31] Omer Mohamoud, Tamara Skoric, Branislav Milovanovic, Nina Japundzic Zigon and Dragana Bajic: Binarized Cross-Approximate Entropy in Crowdsensing Environment, accepted for publication in Computers in Biology and Medicine (Elsevier, M22, IF 1.509).

[32] Tatjana Tasic, Sladjana Jovanovic, Omer Mohamoud, Tamara Skoric, Nina Japundzic-Zigon, Dragana Bajic: Dependency structures in differentially coded cardiovascular time series, submitted.

7. CONCLUSION

The aim of this contribution was to adjust the approximate entropy estimation to applications where no pre-processing and human interaction is possible and where power and processor capacities are poor. It is motivated by mobile crowdsensing that recognizes health as a part of its platforms and forces local analysis as an engineering compromise to transmit the results and not the source data, thus saving the transmission resources. But, a local analysis implies blind processing, while approximate entropy, as well as many other sophisticated tools, require active pre-processing to obtain stationary, noise-free and artifacts - free data.

Binary differential encoding is insensitive both to non-stationarities and to a reasonable amount of artifacts, but the coding is coarse and reduces the number of amplitude levels per sample from 4096 to two. In spite of coding coarseness, amount of information remain sufficient to capture entropy changes in different experimental conditions. The conclusion is achieved comparing the results of (X)*ApEn* (where prior to analysis trend and artifacts were removed, signals visually inspected and stationarity checked) and (X)*BinEn* (where the analysis was blind).

The coarseness of data, however, reduces the sensitivity: within the same experiment, the range of (X)*ApEn* and (X)*BinEn* changes were 40% and 15% respectively. This means that smaller amount of signal changes would remain unnoticed and that (X)*BinEn* is actually a rough measure.

The reduced sensitivity, although undesirable, has a positive effect – it filters out other unwanted phenomena. One of them is binary bias, a problem of forcing the binary zeros due to the rules of differential coding. A more serious (X)*ApEn* problem, the instability of experimental outcomes, is also attenuated: the reduction of different binary template vectors increases the number of matching vectors-followers and ensures the stable probability estimation.

Another consequence of binarization is a reduced computational complexity: (X)*BinEn* operates with 2^m different binary vectors and have a linear complexity, while (X)*ApEn* operates with ~N vectors and has a quadratic complexity. A detailed elaboration is presented in Table 1.

It was shown that auto-*BinEn*(r=0) is equivalent to Shannon m=1 entropy. It was also shown that the elements of binary conditional m=1 entropy are embedded into the described procedure, so its addition as a complementary option is computationally effortless. This feature

enables a dynamic observation of complexity changes related to the inter-samples dependency at a particular time lag, in respect to the complexity of the same signal, but with assumed statistically independency.

(X)*BinEn* is not a substitute for (X)*ApEn*, nor for any of its follow-ups. It is intended for a quick, robust and blind local analysis in wearable devices where (X)*ApEn*-like procedures cannot be applied due to their quadratic computational complexity and requirements for semi-automatic pre-processing.

Aknowledgment: to the authors who let me check and re-check their patent application.