The Case of Ties in Bubble Entropy

George Manis[1], Dimitrios Platakis[1], Roberto Sassi[2]

[1]University of Ioannina, Dept. of Computer Science and Engineering, Ioannina, Greece
[2]Università degli Studi di Milano, Dipartimento di Informatica, Milan, Italy

Abstract

Bubble entropy (bEn), like Permutation entropy (pEn), belongs in the family of entropy definitions, which exploit sorting for the estimation of the complexity in an m-dimensional embedding space. In this paper, we investigate how ties (successive RR intervals with equal values, obtained from ECG recorded at sampling rates smaller than 500Hz) affect the method. Although researchers have claimed that in pEn ties have no significant effect and might even be beneficial in the context of classification, a common practical work-around is to add small perturbations. We quantify how much bEn is affected by ties, analyzing HRV series from Holter ECG data sets (Normal Sinus Rhythm and Congestive Heart Failure). The original sampling rates are between 128Hz and 250Hz. We further increase the number of ties in the HRV series by replicating values at random and compare, with statistical ANOVA F-Test, the discriminating capability of the metrics before and after the distortion. An increasing addition of ties inevitably reduces the discriminating capability of bEn, likely due to the loss of information. However, bEn still discriminates the populations efficiently and almost consistently presents the best F-value among the metrics we consider. Surprisingly, pEn slightly increases F-value for some values of m, possibly due to an initial low discriminating capability and a subsequent random behavior.

1. Introduction

Entropy, as a concept in Information Theory, was introduced by Claude Shannon in 1948 [1] to express randomness of a source of data. Since then, several definitions have been introduced and welcomed by the scientific community, making entropy an invaluable measure to analyze time series. In cardiovascular signal analysis, the use of entropy has been popularized in investigating complexity of the physiological series. Two of the most popular definitions are Approximate [2] and Sample entropy [3], both expressing the conditional probability two subsequences of size m to remain similar at the next point. Applications in biomedical engineering include many physiological signals such as Heart Rate Variability series (HRV) [4], Electroencephalogram (EEG) [5], Electrocardiogram (ECG) [6] and more.

The main disadvantage of both Approximate and Sample entropy is the dependence on two parameters: m, which defines the embedding dimension and r, a real parameter to express the similarity threshold. Values of m are practically limited to the range $m = 1 \dots 4$. However, the second parameter r is a real one and, not only the domain set is infinite, but also any interval subset is an infinite domain set as well. This makes the exhaustive examination of all values impossible.

Bandt and Pompe introduced Permutation entropy [7] in 2002, providing an entropy definition without r. Permutation entropy has been widely accepted by the research community and has been used in diverse research areas [8–11]. The core idea was to measure the entropy of the ordinal patterns observed in the embedding space. In 2017 [12], Bubble entropy was proposed as an entropy definitions which, not only does not need r, but also does not rely much on m. Bubble entropy has been named after the *bubble sort* algorithm, since it uses bubble sort to order the elements in each embedded vector. Even though Bubble entropy is considered similar to Permutation entropy, Bubble entropy follows a completely different depth of thought, based on measuring the entropy of the amount of swaps needed to sort each embedded vector. It considers the distribution of the task of sorting, as something that expresses physical work. In this paper, we will explore and compare how Permutation and Bubble entropy interact with ties. *Ties* are equal values appearing successively in a time series. Our purpose is to illuminate the difference in the behavior of the two definitions. We chose to use HRV signals for this purpose, since, as already noted in [13] and as we confirm later in this paper, ties are quite frequent in such signals. We used time series from subjects in Normal Sinus Rhythm (NSR) and patients with Congestive Heart Failure (CHF) [14]. Firstly, we will try to study how regularity is influenced by ties. Next, and more importantly, we will analyze the effect of ties on Permutation and Bubble entropy. The investigation will be focused on statistical analysis and various tests. Methods and data sets used will

be presented in section 2 and the results in section 4.

2. Methods and Datasets

In this section we will presented the definitions of entropy we consider and the data sets we employ in our experiments.

2.1. Bubble Entropy

Bubble entropy, proposed in 2017 [12], is computed by the following steps :

1. For each embedded vector, count and store in a list the amount of swaps performed by bubble sort to order this vector.
2. Find the probability distribution of the number of swaps.
3. Compute $SwapEn(m)$, defined as the Renyi entropy of order 2 on the distribution.
4. Repeat the above steps for $m{+}1$ to compute $SwapEn(m{+}1)$.
5. Finally, compute and report as result:

$$bEn(m) = \frac{SwapEn(m+1) - SwapEn(m)}{\log {}^{m+1}/_{m-1}}.$$

The amount $\log {}^{m+1}/_{m-1}$ is a normalization factor. For more on normalization alternatives please see [15].

2.2. Permutation and Sorting Entropy

Permutation and Sorting Entropy are also based on sorting. After embedding the time series in $N{-}m{+}1$ embedded vectors, the ordinal pattern of each vector is computed. The probability p_j each ordinal pattern to appear is given by:

$$p_j = \frac{\text{\# the amount of times pattern j appeared}}{N - m + 1}.$$

Using the distribution and Shannon entropy we compute Permutation entropy:

$$PE_m = -\frac{1}{\log_2 m!} \sum_{i=0}^{m!} p_j \log_2 p_j.$$

Sorting entropy is computed by subtracting PE_m and PE_{m+1}:

$$SortEn_m = PE_m - PE_{m+1}.$$

2.3. Data Sets

To conduct our experiments, we employed four data sets with HRV signals [14]. The first two (NSR and NSR2) include recordings in Normal Sinus Rhythm. NSR consists of 18 long-term recordings, while NSR2 comprises 54, totally giving 72 recordings. CHF and CHF2 include patients with Congestive Heart Failure. CHF consists of long-term HRV recordings from 15 subjects and CHF2 of 29 subjects, totally giving 44 recordings. The sampling rates of all data sets are between 128Hz and 250Hz.

3. Data Preparation

Starting off, we will investigate how frequent ties are in the original HRV signals. As referenced by other researchers, ties in HRV signals are quite common [13]. From each signal, we kept the first 70,000 samples to make all signals having equal length. Then, we computed the mean amount of ties, which are reported in table 1. The large number of ties makes the investigation of the problem more interesting.

Table 1. Description of Data Sets.

Data Set	Mean Ties (#)	Mean Ties (%)
Nsr	8965	12.80
Nsr2	11999	17.14
Chf	11308	16.15
Chf2	16207	23.15

In order to increase the amount of ties, and create several signals for our experiments, we broke each signal down to consecutive non-overlapping windows, each consisting of 10 samples. For each window w_i we picked randomly one sample and replicated it n times, creating a new artificial signal with increased number of ties. Repeating this process for every signal in every data set, we generated the data we used for our experiments. In table 2, we see how ties have been increased for different values of n:

Table 2. Mean Ties (%) for each Value of n

Databases	n=1	n=2	n=3	n=4
Nsr	12.84	21.59	30.33	39.04
Nsr2	17.00	25.35	33.66	42.00
Chf	16.55	24.98	33.36	41.77
Chf2	23.39	31.12	38.83	46.55

We also used a second method to generate signals with increased number of ties. Using the Bernoulli probability mass function we defined the probability for each sample to be equal to the previous one. Parameter p expresses the probability of success, so that the function $f(p)$ returns 1 in case of success and 0 in case of failure:

$$x_i = f(p) * x_{i-1} + (1 - f(p)) * x_i. \quad (1)$$

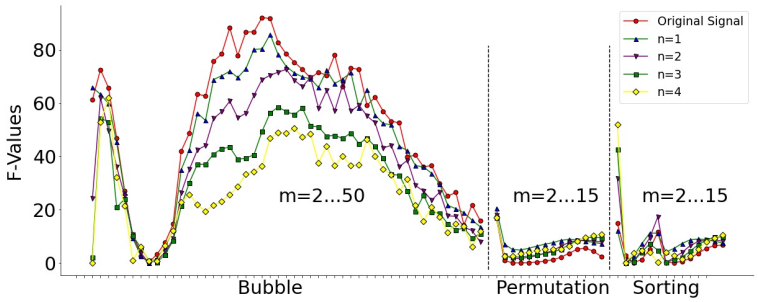

Figure 1. F-values for signals generated with different values of n.

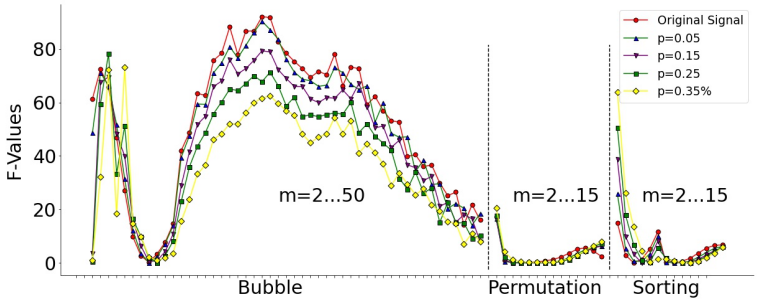

Figure 2. F-values for signals generated with different values of p.

In our experiment we used p=5%, p=15%, p=25% and p=35%. The mean percentage of the number of ties are shown on table 3.

Table 3. Mean Ties (%) for each Value of p

Databases	p=0.05	p=0.15	p=0.25	p=0.35
Nsr	17.07	25.55	34.07	42.63
Nsr2	21.14	29.22	37.28	45.41
Chf	20.24	28.56	36.79	45.14
Chf2	26.97	34.57	42.17	49.82

4. The Effect of Ties

Bubble and Permutation entropy are inherently dependent on the order of the samples making the problem more interesting. When ties are frequent, these definitions could present interesting behavior. The point of our experiments was to investigate how the discriminating capabilities of the methods change when we increase the number of ties. To explore the discriminating abilities we used the F-values of the ANOVA F-test on each generated data set. We investigated how the F-value had changed between the original signal and those with the increased number of ties. We created artificial data sets with increased number of ties using the methods outlined in section 3.

The results using the first of those methods are depicted on figure 1. In this figure we see the F-values computed for each examined method and for $n=1...4$. For Bubble entropy, the red line of the original signals is almost always on the top, meaning that it discriminates better the two classes. Under the red line, we see how the F-value is decreased for each artificial data set. More precisely, as n increases (an increased n leads to larger number of ties, see table 2), the two classes become less successfully discriminated. Permutation and Sorting entropy, however, present an unexpected behavior, which is also observed and reported by [13]. We can see in both entropy definitions that an increased amount of ties could contribute to the discrimination ability, a phenomenon that was characterized as *stochastic resonance* [13].

In our second experiment, we generated signals using the equation (1), for $p=0.05, p=0.15, p=0.25$ and $p=0.35$. In table 3 we can see that an increased value of p leads to an increased number of ties.

Again, we can observe the same behavior. In Bubble entropy, the original signal, represented by the red line, is, at its peak ($m=18...22$), on top of all other lines, indicating that discrimination is clear. As the probability of replication p increases, the F-values gradually drop, indicating once again, that two signals with increased number of ties are more difficult to be discriminated by their entropy.

5. Conclusions

Ties refer to two consecutive equal values in a time series. In physiological signals, this phenomenon is quite

common and especially in HRV series with low sampling rate. In this paper, we compare how ties affect Bubble entropy, Permutation entropy and Sorting entropy, three entropy definitions which have many in common. We chose physiological data sets with Holter recordings, one class with subjects with Normal Sinus Rhythm and one class with patients with Congestive Heart Failure. We used two techniques to artificially increase the number of ties in the signals. Next by using ANOVA F-Test, we evaluated how the increase of the ties affect the discriminating capabilities of the examined definitions. We see that, as the amount of ties increases, the classification of the two classes becomes less clear. We also noted that both Permutation and Sorting entropy presented an unexpected behavior, in which the distorted time series is possible to present clearer classification than the original time series.

.

References

[1] Shannon CE. A mathematical theory of communication. The Bell System Technical Journal 1948;27:379–423.

[2] Pincus S. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America 04 1991;88:2297–301.

[3] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology Heart and Circulatory Physiology 2000;278(6):H2039–H2049.

[4] Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. American Journal of Physiology Regulatory Integrative and Comparative Physiology 2002;283(3):R789–R797. PMID: 12185014.

[5] Sabeti M, Katebi S, Boostani R. Entropy and complexity measures for eeg signal classification of schizophrenic and control participants. Artificial Intelligence in Medicine 2009;47(3):263–274. ISSN 0933-3657.

[6] Li T, Zhou M. Ecg classification using wavelet packet entropy and random forests. Entropy 2016;18(8). ISSN 1099-4300.

[7] Bandt C, Pompe B. Permutation entropy: A natural complexity measure for time series. Phys Rev Lett 2002; 88:174102.

[8] Carricarte Naranjo C, Sanchez-Rodriguez LM, Brown Martínez M, Estévez Báez M, Machado García A. Permutation entropy analysis of heart rate variability for the assessment of cardiovascular autonomic neuropathy in type 1 diabetes mellitus. Computers in Biology and Medicine 2017;86:90–97. ISSN 0010-4825.

[9] Xia Y, Yang L, Zunino L, Shi H, Zhuang Y, Liu C. Application of permutation entropy and permutation min-entropy in multiple emotional states analysis of rri time series. Entropy 2018;20(3). ISSN 1099-4300.

[10] Gao Y, Villecco F, Li M, Song W. Multi-scale permutation entropy based on improved lmd and hmm for rolling bearing diagnosis. Entropy 2017;19(4). ISSN 1099-4300.

[11] Morabito FC, Labate D, La Foresta F, Bramanti A, Morabito G, Palamara I. Multivariate multi-scale permutation entropy for complexity analysis of alzheimer's disease eeg. Entropy 2012;14(7):1186–1202. ISSN 1099-4300.

[12] Manis G, Aktaruzzaman M, Sassi R. Bubble entropy: An entropy almost free of parameters. IEEE Trans Biomed Eng 2017;64:2711–2718.

[13] Cuesta–Frau D, Varela–Entrecanales M, Molina–Picó A, Vargas B. Patterns with equal values in permutation entropy: Do they really matter for biosignal classification? Complexity 2018;2018(1):1324696.

[14] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[15] Manis G, Bodini M, Rivolta MW, Sassi R. A two-steps-ahead estimator for bubble entropy. Entropy 2021;23(6). Article 761.

Address for correspondence:

George Manis
University of Ioannina
Dept. of Computer Science & Engineering
Ioannina, 45110, Greece
manis@cs.uoi.gr