2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

# Quantifying partition-based Kolmogorov-Sinai Entropy on Heart Rate Variability: a young vs. elderly study

Andrea Scarciglia[1], Vincenzo Catrambone[1], Claudio Bonanno[2], and Gaetano Valenza[1]

***Abstract*— In the last decades, a considerable effort has been devoted to quantify complexity in physiological time series, with a particular focus on heart rate variability (HRV). To this end, exemplary quantifiers including Approximate Entropy and Sample Entropy have successfully been applied by leveraging on statistical approximation and further parametrization through the definition of tolerance and embedding dimension, among others. In this study, we investigate the use of the Algorithmic Information Content, which is estimated through an effective compression algorithm, to quantify partition-based Kolmogorov-Sinai (K-S) entropy on HRV series. We test such a K-S estimate on real data gathered from the *Fantasia* database, aiming to discern young vs. elderly complex dynamics. Experimental results show that elderly people are associated with a lower HRV complexity and a more predictable behavior, with significantly lower partition-based K-S entropy than the young adults. We conclude that partition-based K-S entropy may effectively be used to investigate pathological conditions in the cardiovascular system, complementing state-of-the-art methods for complexity assessment.**

## I. INTRODUCTION

Kolmogorov-Sinai (K-S) entropy is a nonlinear quantifier of dynamical systems and corresponds to the maximum amount of information needed to describe the systems' behavior [1]. From K-S entropy estimates it is possible to discern between deterministic and random systems, with the former being characterized by a finite K-S entropy, and the latter associated with an infinite K-S entropy. Positive or vanishing K-S entropy further distinguish between chaotic and regular systems.

From a theoretical viewpoint, the K-S entropy of a time-series can be obtained by considering the supremum among all the values of K-S entropy associated with different partitions of the codomain that the time series might cover [1]. In other words, the information produced by the ergodic dynamical system is computed with respect to partitions, i.e. divisions of a space into different sets. If the space of a dynamical system is divided through a fixed partition, any orbit representing a time series [2] may be described in terms of the sequence of parts visited in time. Thus, the K-S entropy associated with the partition quantifies the mean information needed to identify that specific sequence. In this context, the behavior of an orbit might be intended as its peculiarity of being predictable, so the possibility to find repetitive patterns in the sequence of the visited sets. In this case, the higher is the series predictability, the lower is the entropy value. Grassberger, Procaccia [3], [4], and Renyi [5] achieved the estimation of a lower bound of the K-S entropy, but they did not reach a factual K-S entropy quantification.

It has been proven that there exist special partitions able to reach the supremum of K-S entropy among all the partitions, which, by the definition, is the K-S entropy of the entire time-series; these have been defined as *generating partitions* [6], [7]. Since it is quite hard to find an analytical or explicit expression for those *generating partitions* in case of real time series, i.e., when the equation of the dynamical system is unknown, it is convenient to use K-S entropy relative to finite uniform partitions, i.e. formed by a finite number of sets of the orbit's codomain of equal length. It is known that the use of uniform partitions with a vanishing diameter is a good way to approximate K-S entropy in suitable conditions [6].

The use of partition-based K-S entropy is effective if the system under study is perturbed by random noise because it avoids situations where the supremum among all the partitions is infinity [8], [9]. To our knowledge, K-S entropy has not been applied to characterize complexity in cardiovascular variability series; therefore, the proposed entropy estimation may effectively complement the available set of complexity estimates for HRV series, aiming to identify and characterize different cardiac conditions in health and disease [10]–[12].

In this study, we use the Algorithmic Information Content (AIC) and the Complexity of symbolic strings [13]–[15] to compute the partition-based K-S entropy. Briefly, AIC uses a symbolic string $s$ of finite length whose elements are in a finite alphabet and takes the shortest binary program that is executable by a universal Turing machine, which gives the string $s$ as an output. The Complexity extends the notion of AIC for infinite symbolic sequences.

Since the AIC represents a theoretical limit and so is a not-computable function, we exploited a data compression algorithm to estimate the AIC and, consequently, the partition-based K-S entropy [16] in real HRV series gathered from the publicly available database *Fantasia* [17]. Previous studies suggest that complexity in heartbeat dynamics decreases with age [18], [19], but this has not been investigated in terms of K-S entropy yet. As mentioned above, the proposed computation of K-S entropy through the AIC might constitute an important step towards a more robust and effective quantification of complexity in cardiovascular time series, complementing quantifiers such as sample and approximate entropy [20], [21], their multiscale estensions, Lempel-Ziv

[1]Bioengineering and Robotics Research Center E.Piaggio & Department of Information Engineering, School of Engineering, University of Pisa, Italy;
[2]Department of Mathematics, University of Pisa, Italy;
*Corresponding author. Email at: *vincenzo.catrambone@ing.unipi.it*
The research leading to these results has received partial funding by the University of Pisa in the frame of the project PRA 2020, and by Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Departments of Excellence).

compression [22], and other complexity estimates adopting techniques involving data compression [23] or not involving data compression [24].

Note that the proposed K-S entropy quantification is not bounded to preliminary choices of statistical parameters such as the embedding dimension, the tolerance, and the possibility of counting self-matches. Details follow below.

## II. Materials and Methods

### A. Experimental Setup and Data Preprocessing

HRV series were gathered from the database *Fantasia* [17], publicly available at *https://physionet.org/content/fantasia/1.0.0/*. Briefly, twenty young (21 - 34 years old) and twenty elderly (68 - 85 years old) healthy subjects underwent 120 minutes of continuous supine resting while continuous electrocardiographic (ECG) signals were collected at 250Hz. Each subgroup of subjects consists in the same number of man and women. All subjects remained in a resting state while watching the movie *Fantasia* (Disney, 1940) to help maintain wakefulness [17]. For this pilot study, a total of 10 ECG recordings from the young and 10 recordings from the elderly groups were retained for further analyses. The HRV series were derived by using the well-known Pan-Tompkins algorithm [25], and then verified by an expert through visual inspection.

### B. K-S Entropy Related to a Partition

Given an ergodic dynamical system $(X, \mu, f)$ described by the continuous function $f : X \to X$ on a metric space $X$ and preserving the ergodic measure $\mu$, we consider a finite measurable partition $Z = \{I_i\}_{i=1...N}$ which splits the space $X$ into a $\mu$ disjoint finite collection of sets $I_i$. The definition of entropy related to a partition $Z$ is:

$$H_\mu(Z) = -\sum_{i=1}^{N} \mu(I_i)\log(\mu(I_i)) \quad (1)$$

if we consider the partition $Z_n$ given by all the possible intersections between the counter images $f^{-j}(I_i)$ for all $i = 1, \ldots N$ and for $j = 0, \ldots, n-1$, the Kolmogorov-Sinai entropy $h_\mu(f, Z)$ related to the partition $Z$ can be defined as

$$h_\mu(f, Z) = \lim_{n\to\infty} \frac{1}{n} H_\mu(Z_n) \quad (2)$$

The partition-based K-S entropy is a quantifier of the information (with respect to the measure $\mu$ and the chosen partition) needed to describe each point of the orbit of a generic dynamical system.

### C. AIC and Complexity

Let $\Omega$ be the set formed by infinite strings with elements in a finite alphabet $\mathcal{A} = \{1, \ldots, N\}$. As mentioned before, the AIC notion of a finite symbolic string $s$ relates to the length of the smallest binary computer program that produces $s$ as output: in other words, this program might be thought of as a code which contains all the minimal information that describes $s$. We define the complexity $K$ for infinite strings as

$$K(\omega) = \limsup_{n\to\infty} \frac{\mathrm{AIC}(\omega^n)}{n}, \quad \omega \in \Omega \quad (3)$$

where $\omega^n$ is the string $\omega$ truncated at the $n^{th}$ element. It can be interpreted as the length of the binary program identifying each term of the string. Lower values of complexity are associated with strings that show repetitive patterns and hence necessitate less information to be described.

### D. Symbolic Dynamics

By identifying the partition $Z = \{I_i\}_{i=1...N}$ with the finite alphabet $\mathcal{A} = \{1 \ldots N\}$, any orbit of the ergodic dynamical system is converted into a string of symbols by assigning for each point of the orbit a specific symbol associated with the set of the partition it is visiting. Through this process, the complexity $K(x, Z)$ related to a partition $Z$ of the orbit with initial condition $x \in X$ is identified by the complexity of the associated symbolic string.

In the case of ergodic dynamical systems, Brudno [13] showed that:

$$K(x, Z) = h_\mu(f, Z), \quad x \in X \quad (4)$$

and for all finite measurable partitions $Z$ of $X$. Eq. 4 guarantees the computation of the K-S entropy for a partition through the complexity $K(x, Z)$ and, consequently, the AIC.

Unfortunately, the computational procedure associated with the AIC quantification is not performed by any algorithm, and a compression algorithm should then be exploited. In this study, we applied the CASToRe compression algorithm to approximate the AIC [15], especially because of its speed of convergence in weakly chaotic systems [15]. Moreover, the CASToRe algorithm is not bounded to preliminary choices of statistical parameters such as the embedding dimension, the tolerance and the possibility of counting self-matches.

### E. Complexity Estimation Procedure

Because of the different subjects' heart rate, the original HRV series have been interpolated at a sampling frequency of $2Hz$ to obtain series of equal length. Supposing that each HRV series of the dataset *Fantasia* corresponds to an observable of the ergodic dynamical system described by the cardiac cycles, each series was converted into a symbolic series with respect to a finite measurable uniform partition; then the CASToRe algorithm was applied to compute complexity $K(x, Z)$, thus obtaining K-S entropy $h_\mu(f, Z)$.

More in detail, Fig. 1 graphically shows the conversion of the time series into a symbolic series, in which the interval $I = [0.6, 1.79]$ represents the subject-wise range of the heartbeat series. In other words, interval $I$, on which the partition of all subjects has been performed, has been chosen as the complete range between the minimum and maximum inter-beat interval across all subjects and time series. The partitions considered on $I$ are uniform, i.e. formed by sets of the same length, so each set is associated with a symbol, and the points of the time series are converted in a symbolic

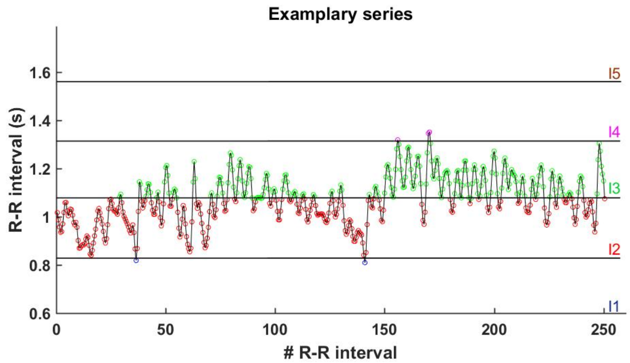


Fig. 1: Conversion of an exemplary time series to symbols. The space of all the possible measurements has been divided here according to a uniform partition of $N = 5$ sets, i.e., $I_N = I1, I2, I3, I4, I5$. All the points of the series falling in the same interval, here identified by the same color, are converted into the same symbol.

string by taking into account the set of partition they belong to. The CASToRe algorithm is then applied [15] and the K-S entropy is evaluated from equations (3), and (4).

### F. Statistical analysis

K-S entropy values extracted from artifact-free HRV series in the two experimental groups (i.e., elderly, and young people) have been compared by employing non-parametric Mann-Whitney tests for unpaired samples, with the null hypothesis of equal median between the two populations. The statistical comparison has been repeated 15 times, for partitions $Z = \{I_{i=1...N}\}$ with increasing $N$, going from 2 to 17. The statistical significance was set to $\alpha = 0.01$.

## III. Experimental Results

Fig. 2 shows the K-S entropy values related to the uniform partitions for the two groups of subjects (elderly and young people) as a function of the partition. The finer is the partition, the higher the values that are shown in the experimental series of the two groups. Notably, young subjects show consistently higher K-S entropy (close to $K - S = 2$ at higher partitions) than elderly subjects.

The lower cardiac complexity associated with the elderly is clearly shown in Fig. 3, where the K-S entropy group-wise median is depicted for the two groups. Note that the higher is the cardinality of the uniform partition considered (i.e. the number of sets), the wider the difference between the two groups. Such a difference is statistically significant (i.e., $p < 0.01$) at almost every cardinality higher than 6.

## IV. Discussion And Conclusion

In this preliminary study, the partition-based K-S entropy has been estimated on real HRV series gathered from the publicly available database *Fantasia*, considering 20 series from 10 young and 10 elderly subjects gathered while watching a movie. The K-S entropy computation has been performed through the approximation of the AIC with the lossless data compression CASToRe algorithm [15]. Results confirmed that the proposed partition-based K-S entropy can profitably be exploited as a potential biomarker quantifying the complexity in heartbeat dynamics series.

From a methodological viewpoint, even if the partition-based K-S entropy computation requires a sufficient large number of samples (see eq. (3)-(4)), the CASToRe algorithm speed of convergence confirms the robustness of the AIC approximations [15]. While other compression algorithms may be exploited to estimate K-S entropy [22], [23], the proposed partition-based algorithm may provide better estimates using real data. Indeed, CASToRe is able to compress any finite time series with a lower amount of information than other algorithms, e.g. Lempel-Ziv data compression, especially in presence of repetitive patterns (it has been used to study weakly chaotic dynamical systems) [15]. Eventually, it results in a better approximation of the AIC, which, by definition, represents the theoretical minimum [13]. Moreover, the CASToRe nonlinear approach has the advantage of being not supported by preliminary choices of statistical parameters, such as embedding dimension, tolerance, or possibility of counting self-matches.

Our experimental results showed that K-S entropy values tend to increase for all subjects as the cardinality of the partition grows. This is due to the finiteness of the time series considered: the more the number of intervals (and hence symbols), the less the possibility to find similar or predominant patterns in a finite symbolic time series [16]. Results confirmed that the complexity level, as quantified through K-S entropy, is statistically higher in young adults than in the elderly subjects, and the difference between the two groups becomes statistically significant when more than six intervals of the partition are considered. The lower values of partition-based K-S entropy obtained for elderly subjects prove the less amount of information needed to describe the

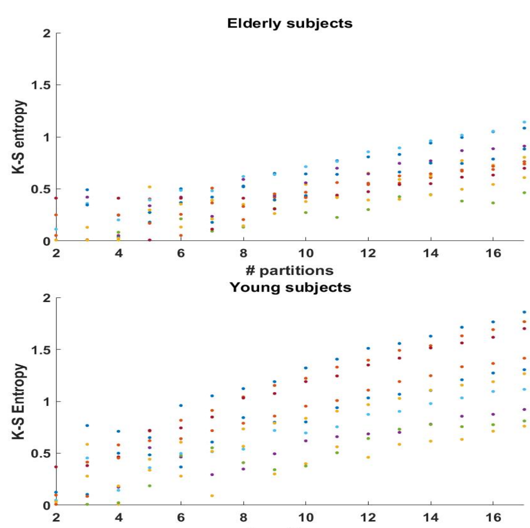


Fig. 2: The K-S entropy value related to uniform partitions for elderly and young subjects. The x-axis refers to the cardinality of partitions (range between 2 and 17), and the y-axis refers to the K-S entropy values.

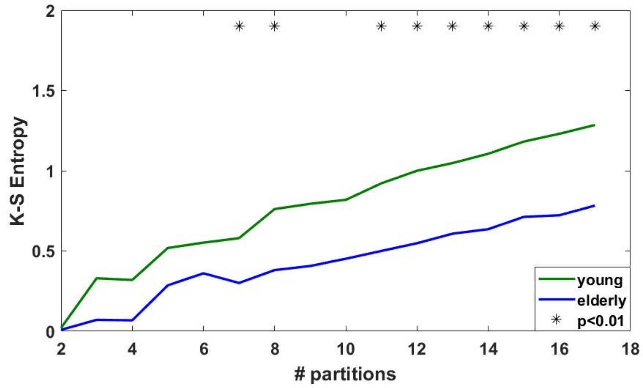


Fig. 3: Median across subjects of partition-based K-S entropy for the two groups, i.e., elderly (blue) and young (green). Cardinality of partitions considered ranges between 2 and 17. Asterisks indicate statistically significant comparisons ($p < 0.01$).

behavior of the associated heartbeat dynamics, as well as the symbolic strings related to the HRV series report more repetitive patterns when compared to younger subjects. More specifically, it was found that partitions composed by at least six uniform intervals are needed to detect statistically significant differences among the two groups. This might be due to the fact that when the partition is too coarse (i.e. composed by few sets), all the heartbeat series tend to occupy the same number of intervals, whereas in finer partitions the time series of the young group tend to visit a higher number of intervals and with faster variations from one interval to the other, with respect to the elderly group. In this sense, the HRV series gathered from elderly subject's show more predictability, leading to lower K-S entropy values.

From the physiological viewpoint, these results are in agreement with previous studies suggesting that aging reduces the variability of the cardiac time series, while stress and pathologies may accentuate this phenomenon [12], [26]–[28]. Indeed, the proposed method confirms these results needing less parameters such as time scaling [22] or the embedding dimension [28]. Of note, it can be appreciated that in our results the difference between the two groups of subjects are stable with respect to the number of partition, whereas other methods showed more variations [22].

Limitations of this work might be associated with the relatively low number of subjects involved in the study, and future developments should be directed towards a systematic characterization of K-S entropy estimates in larger cohorts and in different patho-physiological conditions.

To conclude, the proposed exploitation of the partition-based K-S entropy computed through the approximation of the AIC with the lossless data compression CASToRe algorithm can be considered as a potential biomarker of heartbeat dynamics, complementing state-of-the-art methods for cardiac complexity assessment.

## REFERENCES


[1] K. Keller *et al.*, "On the relation of ks entropy and permutation entropy," *Physica D: Nonlinear Phenomena*, vol. 241, no. 18, pp. 1477–1481, 2012.

[2] T. Sauer *et al.*, "Embedology," *Journal of statistical Physics*, vol. 65, no. 3, pp. 579–616, 1991.

[3] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical review letters*, vol. 50, no. 5, p. 346, 1983.

[4] P. Grassberger and I. Procaccia, "Estimation of the kolmogorov entropy from a chaotic signal," *Physical review A*, vol. 28, no. 4, p. 2591, 1983.

[5] A. Rényi *et al.*, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, The Regents of the University of California, 1961.

[6] M. Brin and G. Stuck, *Introduction to dynamical systems*. Cambridge University Press, 2002.

[7] W. Krieger, "On entropy and generators of measure-preserving transformations," *Transactions of the American Mathematical Society*, vol. 149, no. 2, pp. 453–464, 1970.

[8] Y. Kifer, "Random perturbations of dynamical systems," 1988.

[9] C. Bonanno, "The algorithmic information content for randomly perturbed systems," *Discrete & Continuous Dynamical Systems-B*, vol. 4, no. 4, p. 921, 2004.

[10] U. Acharya *et al.*, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[11] M. Nardelli *et al.*, "Multichannel complexity index (mci) for a multi-organ physiological complexity assessment," *Physica A: Statistical Mechanics and its Applications*, vol. 530, p. 121543, 2019.

[12] R. Barbieri *et al.*, *Complexity and nonlinearity in cardiovascular signals*. Springer, 2017.

[13] A. Brudno, "Entropy and the complexity of the trajectories of a dynamical system," *Trans. Moscow Math. Soc.*, vol. 2, pp. 127–151, 1983.

[14] F. Argenti *et al.*, "Information and dynamical systems: a concrete measurement on sporadic dynamics," *Chaos, Solitons & Fractals*, vol. 13, no. 3, pp. 461–469, 2002.

[15] V. Benci *et al.*, "Dynamical systems and computable information," *Discrete and Continuous Dynamical Systems-Series B*, no. 4, pp. 935–960, 2004.

[16] G. Chaitin, *Information, randomness & incompleteness: papers on algorithmic information theory*, vol. 8. World Scientific, 1990.

[17] N. Iyengar *et al.*, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, pp. R1078–R1084, 1996.

[18] D. Simpson and R. Wicks, "Spectral analysis of heart rate indicates reduced baroreceptor-related heart rate variability in elderly persons," *Journal of Gerontology*, vol. 43, no. 1, pp. M21–M24, 1988.

[19] H. Tsuji *et al.*, "Reduced heart rate variability and mortality risk in an elderly cohort. the framingham heart study.," *Circulation*, vol. 90, no. 2, pp. 878–883, 1994.

[20] S. Pincus, "Approximate entropy as a measure of system complexity.," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[21] J. Richman and J. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, 2000.

[22] M. Baumert *et al.*, "Compression based entropy estimation of heart rate variability on multiple time scales," in *2013 35th Annual International EMBC*, pp. 5037–5040, 2013.

[23] M. Baumert *et al.*, "Estimating the complexity of heart rate fluctuations: an approach based on compression entropy," *Fluctuation and Noise Letters*, vol. 5, no. 04, pp. L557–L563, 2005.

[24] A. Porta *et al.*, "Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1282–1291, 2001.

[25] J. Pan and W. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[26] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, p. 258, 2017.

[27] J. Thayer *et al.*, "Stress and aging: A neurovisceral integration perspective.," *Psychophysiology*, pp. e13804–e13804, 2021.

[28] J. Hu *et al.*, "Multiscale analysis of heart rate variability: a comparison of different complexity measures," *Annals of biomedical engineering*, vol. 38, no. 3, pp. 854–864, 2010.