Parametric Estimation of Entropy Using High Order Markov Chains for Heart Rate Variability Analysis

Corrado Ameli[1], Roberto Sassi[1]

[1] Università degli Studi di Milano, Milan, Italy

Abstract

The aim of this study is to investigate the parametric estimation of entropy and entropy rate of Heart Rate Variability (HRV) series, through the usage of Higher Order Markov Chain (HOMC) models. In HOMCs, the dynamic depends on an arbitrary number of previous steps, and not just the present state as in traditional Markov chains.

After obtaining the transition probabilities, entropy and entropy rate were derived in terms of the stationary distribution. First, we empirically confirmed the convergence of the estimated values to the theoretical ones, by creating synthetic signals from HOMCs with known characteristics. Then, we tested the methodology on HRV series derived from long-term recordings of 44 patients affected by congestive heart failure and 54 normal controls. After quantization of RR series with three different strategies, metrics were estimated varying the HOMC order (up to 7) and the number of samples. As no gold standard was available, we measured the capability of entropy and entropy rate of discriminating among the two populations considered, using a support vector machine model ($k = 5$ fold validation).

On synthetic series, the estimation error was marginal when $N > 200$ and smaller when the MCs were tightly connected. The classification averagely scored an accuracy of about 80% in distinguishing normal and CHF patients, with a maximum value of 86.7% (AUC=0.92).

1.   Introduction

Cardiovascular signals provide relevant information on the state of the heart and the autonomic nervous system. Heart Rate Variability (HRV) has been extensively characterized quantifying its entropy and entropy rate [1]. However, recent studies suggest that the use of parametric estimators of entropy rate, based on autoregressive (AR) models might have advantages, e.g., when dealing with noisy signals [2]. Methods performing parametric estimation of entropy measures are also referred to as *model-based* [3].

The aim of this study is to investigate the parametric estimation of entropy and entropy rate, in HRV series,

through the usage of Higher Order Markov Chain (HOMC) models [4]. While the future evolution of a Markov chain depends only on its present state, in HOMCs the dynamic depends on an arbitrary number of previous steps.

2.   Methods

Markov Chains (MCs) are stochastic processes such that the probability of each event depends only on the state attained previously. Thanks to this property called *Markov*, MCs are able to encode dependencies of events which are closely related in time. MCs are fully parameterized by a square $n \times n$ transition matrix $\mathbf{P}$, where n is the cardinality of the discrete state space . The probability of transiting from a state i to a state j is denoted by the corresponding entry P_{ij}. The sum of each row is thus 1, that denotes the probability of moving from state i to any state .

The stationary distribution μ of a MC solves the eigenvector problem: $\mu\mathbf{P} = \mathbf{P}$. In particular, μ is a $1 \times n$ vector which represents the probability of being in a particular state i at any time t.

The (Shannon) entropy $H(\bullet)$ quantifies the minimum descriptive complexity of a random variable (average information). For a given string of data, it is related to the length of its shortest binary representation. For the realization of a MC X, it is evaluated as:

$$H(X) = -\sum_i \mu_i \, log_2 \, \mu_i, \tag{1}$$

where the sum is extended to all the states. On the other hand, the entropy Rate $H_r(\bullet)$ describes the time-average conditional entropy of the stochastic process. In other words, it encodes the (average) information available about the future state, given the present (and past, for HOMC described in section 2.1) outcomes. For the MC X:

$$H_r(X) = -\sum_{ij} \mu_i P_{ij} \, log_2 \, P_{ij} \tag{2}$$

2.1.   Higher Order Markov Chains

The order of a MC denotes the number of states considered for transiting from the current state to the next one.

For classical MCs, when the Markov property holds, the process order is 1. On the other hand, an order 2 MC assigns the probability of X_{i+1} based on the current and previous outcomes of the process (X_i, X_{i-1}). As the name suggests, HOMCs are MCs whose order is greater than 1.

An order k HOMC is fully parameterized by a $n^k \times n$ transition matrix $\hat{\mathbf{P}}$, where each element denotes the probability of transiting from the ordered set of k states $i = (X_i, \ldots, X_{i+k-1})$ to a new outcome $j = X_{i+1}$. Not being simply Markov anymore, equations (1) and (2) must be adapted as well, losing their computational simplicity. An alternative solution, is to consider a $n^k \times n^k$ square transition matrix $\mathbf{P}$ describing the probability of transiting from the ordered set of k states $i = (X_i, \ldots, X_{i+k-1})$ to the ordered set $j = (X_{i+1}, \ldots, X_{i+k})$ (lagged state vectors). For instance, let a element of $\hat{P}_{ij} \neq 0$ for a third order HOMC with $i = (2, 0, 1)$ and $j = 5$; then the corresponding element in $\mathbf{P}$ will link $i = (2, 0, 1)$ and $j = (0, 1, 5)$. The price of encoding a longer time-memory in a square transition matrix is clearly its larger size (and memory footprint in numerical computations).

2.2. Encoding RR series into HOMC

RR series are values of distances in time between nearby beats and each value $RR_i \in \mathbb{R}$. To model them with HOMC, RR_i values need to be aggregated into states. This process is formally called quantization and the number of states (N) is a parameter of the process. In this work three types of quantization have been considered:

• Uniform: N equally big partitions are built spanning from RR_{MIN} to RR_{MAX};

• Gaussian: a gaussian distribution is first fit to the sample distribution of the series (maximum likelihood). Then N not-overlapping partitions of equal probability are created;

• Minimization of Mean Square Distortion (MSD): the dimension of the series is quantized into N partitions, obtained by solving an optimization problem that minimizes the overall quantization error: $Q_{\mathrm{err}} = \sum_i (RR_i - RR_i^q)^2$ where RR_i^q are the RR_i values after quantization.

In the following, the states will be numbered with integers, starting from 0, in order to obtain a sequence S, of length L (as the RR series), composed by N different symbols. To estimate the transition matrix for the sample sequence S, a frequency matrix $\mathbf{F}$ is built and then normalized (Algorithm 1). With $\mathbf{P}$ available, the stationary distribution μ is computed and the entropy metrics estimated.

2.3. Synthetic Sequences

We first verified the convergence of the estimated metrics to the correct theoretical values of entropy and entropy rate, using synthetic sequences generated from known

---

Algorithm 1 Transition matrix estimation.

1: Let k be the order of the HOMC, N the number of states, S the sequence and L the length of S.
2: $\mathbf{F} = \mathbf{zeros}(\mathbf{N^k}, \mathbf{N^k})$
3: **for** $i = k + 1$ to L **do**
4: $\quad in = S_{(i-k):(i-1)}$
5: $\quad out = S_{(i-k+1):(i)}$
6: $\quad \mathbf{P_{in,out}}$ ++
7: **end for**
8: $\mathbf{P} = \mathbf{zeros}(\mathbf{N^k}, \mathbf{N^k})$
9: **for** $j = 1$ to N^k **do**
10: $\quad \mathbf{P_{j,:}} = \mathbf{F_{j,:}}/\mathbf{sum}(\mathbf{F_{j,:}})$
11: **end for**

---

stochastic processes. Three MCs have been taken into consideration. Each MC is composed by the same number of states $(N = 6)$ but the density of the connections changes among the processes. Thus, we identify a loosely connected MC (7 arcs), an averagely connected MC (11 arcs) and a tightly connected MC (25 arcs, 15 of them with probability less or equal to 0.1). Each of the MC had a single stationary distribution.

The first experiment we performed consisted in generating sample sequences from these processes and in estimating the entropy metrics as their length varied from 50 to 1000 points. We then measured the difference from the theoretical value to the estimated value normalized by the former. For statistical convergence, the procedure was repeated 100 times and the results averaged. We then evaluated how noise affects the estimates by artificially inducing error by letting each state of the sequence to be wrongly re-labelled as one of its neighbors with an error probability varying from 0 to 0.1 (*i.e.*, the probability of remaining in the original state was 0.8). This was meant to mimic what happens with quantization.

Finally, we further considered three HOMCs $(N = 6)$, of order 2, 3 and 4, to assess the errors produced by the wrong choice of the order's parameter.

2.4. HRV Data

We considered the long-term RR series of 98 subjects: 54 with a normal cardiovascular activity and 44 affected by Congestive Heart Failure (CHF). The signals were obtained [5] from the Normal Sinus Rhythm RR Interval Database, the Congestive Heart Failure RR Interval Database (29 subjects) and the BIDMC Congestive Heart Failure Database (15 subjects) . The sampling rates were 128 Hz for the first two databases and 250 Hz for the third.

Given the fact that for real sequences reference entropy values are not available, we verified how effective are entropy and entropy rate in distinguishing subjects from the two populations. Quantizations methods presented in sec-
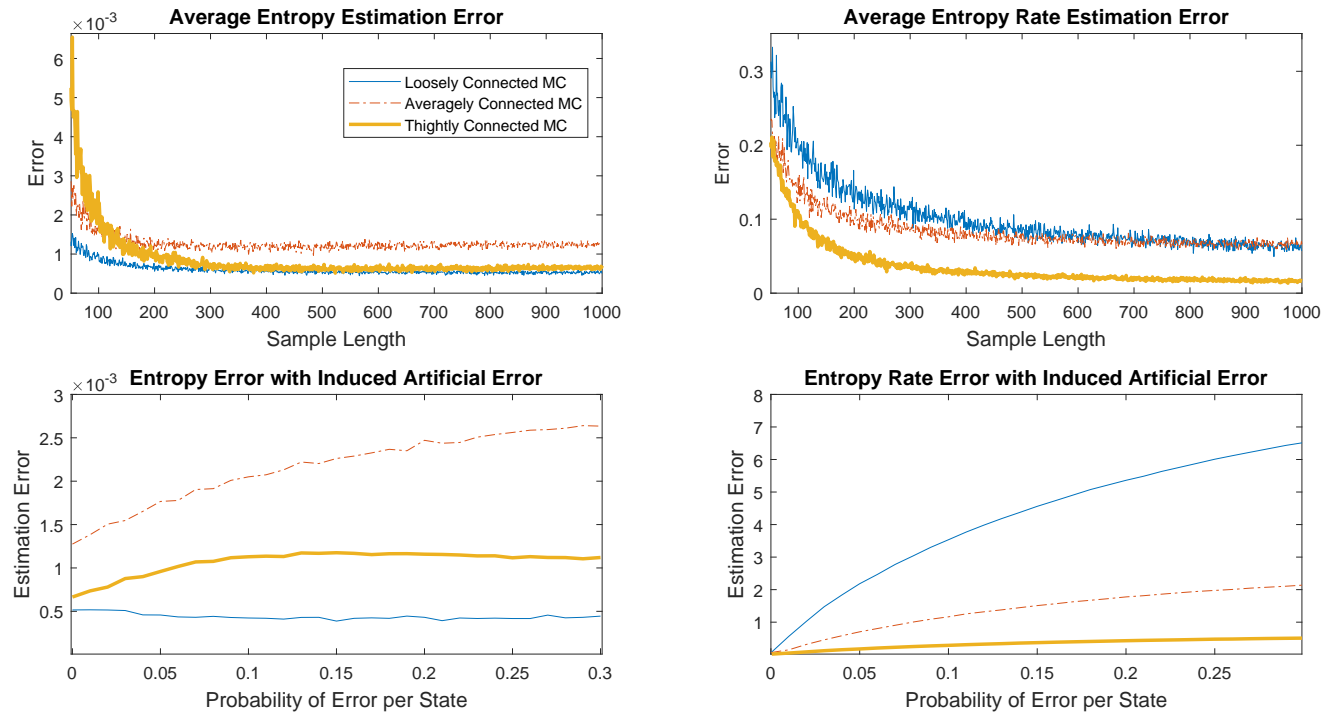
Figure 1. On top: relative errors (averaged over 100 runs) for synthetic sequences, as the sample length increases (MC of order 1). On bottom: relative errors (averaged over 100 runs) for synthetic sequences when $L = 1000$, as the noise increases (MC of order 1).

Table 1. Configurations which reached the largest AUC, at each sequence length.

L	Diff.	Quantiz.	N	Order	AUC	Acc.
100	Yes	Gauss.	12	1	0.79	78.6
200	Yes	Gauss.	13	2	0.77	75.5
500	Yes	Gauss.	10	2	**0.84**	**83.7**
1000	Yes	Gauss.	9	2	0.81	79.6
75000	No	MSD	10	1	**0.92**	**86.7**

tion 2.2 were all tested, with a number of states N up to 14. Then, different lengths of the HRVs (100, 200, 500, 1000 and 75000) and different parameterizations of the HOMCs ($k \leq 7$) were employed. Given the non-stationarity of long-term RR series, we also considered the first-difference (*Diff.*) series $I_i = RR_{i+1} - RR_i$. For each configuration of the parameters, the average Area Under the Curve (AUC) of a Support Vector Machine (SVM) classifier was assessed.

3.   Results

The results of the simulations performed on synthetic sequences, while varying their length, are presented in Fig. 1. The error regarding entropy is almost negligible, while entropy rate requires at least 100 sample points to achieve a relative error below 0.1, in the tightly connected MC. The impact of noise is also reported in Fig. 1 when $L = 1000$.
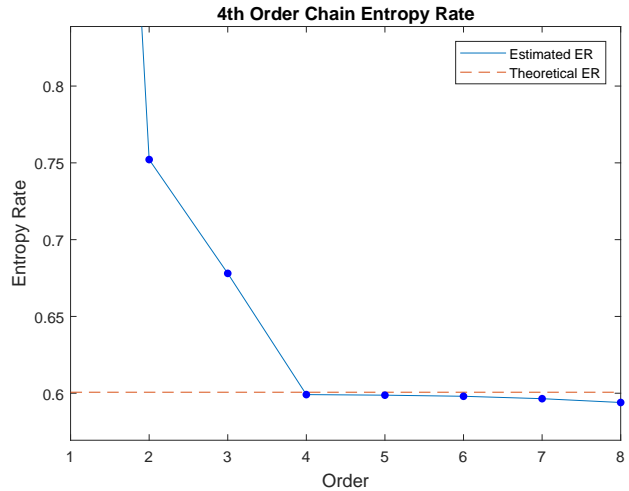

Figure 2. Estimates of entropy rate for series generated by an order 4 HOMC, while varying the order of the model.

Entropy is very resilient to this kind of noise, while entropy rate estimates are more severely affected (except the tightly connected MC, where the impact is milder).

The results obtained on synthetic series produced by HOMC showed that, once the order of the model is equal or larger than the order of the HOMC, the estimated entropy approaches the theoretical one. As shown in Fig. 2, where the series were generated by an order $k = 4$ HOMC, the error dropped when reaching k. These preliminary findings suggest that the order of the model might be cho-
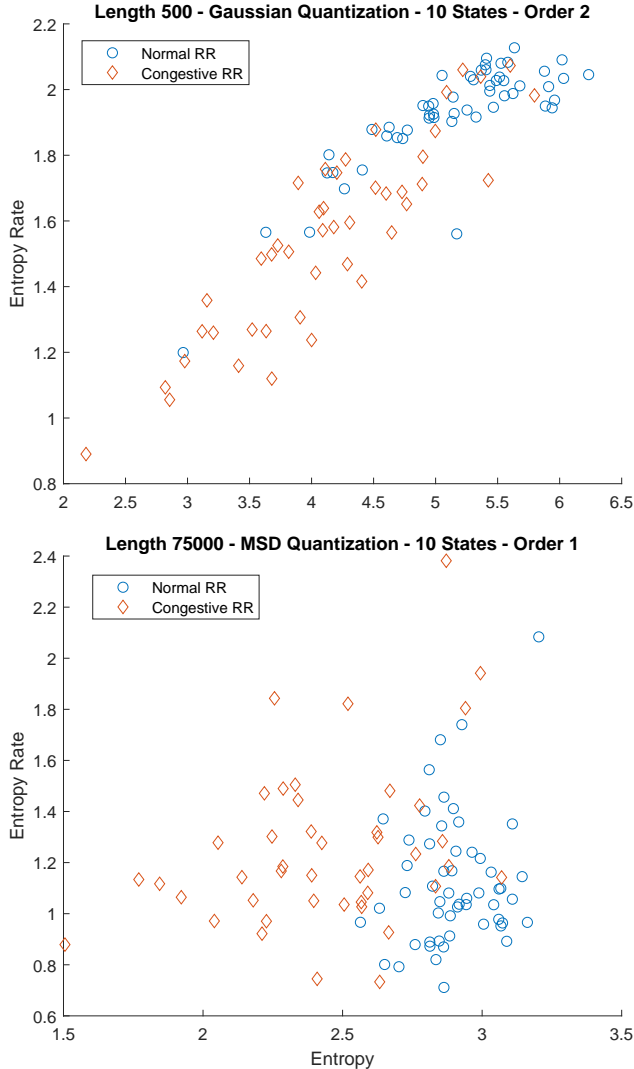
Page 3

Figure 3. Entropy and entropy rate values for each of the 98 normal and CHF subjects considered, computed with two different series lengths: 500 (top) and 75000 (bottom).

sen by picking the one for which the error then plateaus.

Regarding RR series, the best value of AUC, for each of the sequence's length tested, are reported in Tab. 1. For short signals (100-1000), the best results were achieved by considering the first-difference series (*Diff.*) and a Gaussian quantization. As the length of the signal considered grew, MSD and Uniform quantization produced larger AUC. The values of entropy and entropy rate, for the two cases with the largest AUC (bold in table 1), are shown in Fig. 3. Clusters are largely distinguishable, and the classification accuracy was larger for the longest series tested ($L = 75000$). However, with short series entropy rate plays an important role in the discrimination, while with very long series this metric has little to offer in the classification process (the dots are spread nearly horizontally).

While HOMCs up to order 7 were tested, only two previous states proved relevant (order 2). On the other hand, differently than the 4 to 6 states typically employed in symbolic dynamics, optimal classification always happened for value of N close or larger than 10.

4. Conclusions

In this work, we positively verified the possibility of using parametric estimates of entropy measures, based on HOMC models, in HRV series. Entropy proved more robust then entropy rate, to the addition of quantization errors; also, the average estimation error was larger on the second metrics. However, both proved effective, even if comparison with different parametric models are necessary, and will be performed in the future.

In the analysis of real HRV series, entropy rate was more relevant in the classification of short time series. This seems to suggest that non-stationarities, inherent in Holter HRV series and more relevant as the length of the series increases, might affect the estimate and blur the differences between the two populations.

References

[1] Sassi R, Cerutti S, Lombardi F, Malik M, Huikuri HV, Peng CK, Schmidt G, Yamamoto Y, Reviewers D, Gorenek B, Lip GYH, Grassi G, Kudaiberdieva G, Fisher JP, Zabel M, Macfadyen R. Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the EHRA co-endorsed by the APHRS. Europace 2015;17:1341–1353.

[2] Aktaruzzaman M, Sassi R. Parametric estimation of sample entropy in heart rate variability analysis. Biomed Signal Process Control 2014;14:141–147.

[3] Xiong W, Faes L, Ivanov PC. Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations. Phys Rev E 2017;95:062114.

[4] Ching WK, Huang X, Ng MK, Siu TK. Higher-Order Markov Chains. Springer 2013;141–176.

[5] 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:e215–e220.

Address for correspondence:

Corrado Ameli
Dipartimento di Informatica
Università degli Studi di Milano
via Bramante 65, 26013 Crema (CR) Italy
E-mail: corrado.ameli@studenti.unimi.it