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 quadratic discriminant classification model (k = 5 fold validation).

On synthetic series, the estimation error was marginal when N > 200. 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 estimators [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. This property is called Markov and thanks to it, 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 $P$, where $n$ is the cardinality of the discrete state space, and each of its rows and columns denotes a state. The probability of transitioning 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 (including itself).

The stationary distribution $\mu$ of a MC solves the following eigenvector problem:

$$\mu P = \mu$$

(1)

In particular, $\mu$ is a $1 \times n$ vector which represents the probability of being in a particular state $i$ at any time $t$ during the process.

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:

$$H(X) = -\sum_i \mu_i \log_2 \mu_i,$$

(2)

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) amount of information available on the future state given the present (and past for HOMC described in section 2.1) outcomes. For the MC
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 instance, when the Markov property holds (classical MCs), 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.

Let \( k \) be the order of the HOMC. A HOMC is fully parameterized by a \( n^k \times n \) transition matrix \( \hat{P} \), 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 \( j = X_{i+k} \). By letting the process not to be Markov anymore, equations (1), (2) and (3) must be adapted as well, losing their computational simplicity. An alternative solution, meant to maintain a square transition matrix, is to consider a \( n^k \times n^k \) transition matrix \( P \) 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 \( P'_{ij} \neq 0 \) for a HOMC with \( i = (2, 0, 1) \) and \( j = 5 \); then the corresponding element in \( P \) 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 \( RR_{\text{MIN}} \) to \( RR_{\text{MAX}} \);
- 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: 
  \[
  Q_{\text{err}} = \sum_i (RR_i - RR_i^q)^2
  \]
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 \( F \) is built and then normalized, as illustrated in Algorithm 1. With \( P \) available, the stationary distribution \( \mu \) is computed and the entropy metrics estimated.

**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. \( F = \text{zeros}(N^k, N^k) \).
3. for \( i = k + 1 \) to \( L \) do
4. \( \text{in} = S_{(i-k):(i-1)} \).
5. \( \text{out} = S_{(i-k+1):i} \).
6. \( P_{\text{in, out}}++ \).
7. end for
8. \( P = \text{zeros}(N^k, N^k) \).
9. for \( j = 1 \) to \( N^k \) do
10. \( P_{j, i} = F_{i,j} / \text{sum}(F_{i,j}) \).
11. end for

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 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 500 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 also evaluates 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.

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Yes</td>
<td>Gauss.</td>
<td>12</td>
<td>1</td>
<td>0.79</td>
<td>78.6</td>
</tr>
<tr>
<td>200</td>
<td>Yes</td>
<td>Gauss.</td>
<td>13</td>
<td>2</td>
<td>0.77</td>
<td>75.5</td>
</tr>
<tr>
<td>500</td>
<td>Yes</td>
<td>Gauss.</td>
<td>10</td>
<td>2</td>
<td>0.84</td>
<td>83.7</td>
</tr>
<tr>
<td>1000</td>
<td>Yes</td>
<td>Gauss.</td>
<td>9</td>
<td>2</td>
<td>0.81</td>
<td>79.6</td>
</tr>
<tr>
<td>75000</td>
<td>No</td>
<td>MSD</td>
<td>10</td>
<td>1</td>
<td>0.92</td>
<td>86.7</td>
</tr>
</tbody>
</table>

3. Results

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

Figure 1. Relative errors (averaged over 100 runs) for synthetic sequences, as the sample length increases (MC of order 1).

Figure 2. Relative errors (averaged over 100 runs) for synthetic sequences when $L = 1000$, as the noise increases (MC of order 1).

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. All the different quantizations methods presented in section 2.2 were 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 (order up to 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 Curve (AUC) of a set of quadratic discriminant classifier was assessed. Other more powerful classifiers might have been employed for the task; our choice was dictate simplicity (e.g., the lack of tunable parameters, to avoid the need of a validation set).
The best value of AUC, for each of the sequence’s length tested, are reported in table 1. For short signals (100-1000), the best results were achieved by considering the first-difference series (Diff.) of the RR series and a Gaussian quantization. As the length of the signal considered grows, MSD and Uniform quantization produced larger AUC. The values of entropy and entropy rate, for each of the subjects in the two cases with the largest AUC (bold in table 1), are shown in Figure 3. Clusters are largely distinguishable, and the classification accuracy was larger for the longest series tested ($L = 75000$). The biggest difference between the two cases is that with short series entropy rate plays an important role in the discrimination, while as we consider 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 than 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


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