Complexity Analysis of the Fetal Heart Rate for the Identification of Pathology in Fetuses

M Ferrario[1], MG Signorini[1], G Magenes[2]

[1]Politecnico di Milano, Italy
[2]Università di Pavia, Italy

Abstract

Monitoring of fetal conditions is a very important step to assure a healthy newborn state. A special interest is reserved to Intrauterine Growth Restricted (IUGR) fetuses in the very preterm period, when the interpretation of fetal examinations is often very difficult. In order to study the fetal heart rate complexity we have introduced the Lempel-Ziv (LZ) complexity parameter as a global index estimating unpredictability level of the time series. We have compared this new parameter with other regularity estimators: Approximate Entropy and Sample Entropy. The results show that the LZ complexity is a stable parameter and its value significantly discriminates the severe IUGR (preterm delivered) from the not severe IUGR (at term delivered) and normal fetuses. The introduction of the LZ complexity parameter could improve the diagnostic ability of the fetal monitoring systems.

1. Introduction

The objective to thoroughly analyze the biological systems from the signals we can record often appears unfeasible. The knowledge we have about the system behavior is limited by their intrinsic complexity, which is basically produced by numerous interacting mechanisms contributing to the physiological system performance. There is no a unique and rigorous way of consistently defining complexity. The concept of complexity sometimes quantifies the difficulties we have in describing or understanding a signal.

Pincus introduced a family of statistics, the Approximate Entropy (ApEn) [1], with the purpose to measure the signal "regularity" i.e. the presence of similar patterns in a time series, thus allowing the analysis of different systems also corrupted by random noise. Richman and Moorman modified the ApEn to eliminate its imperfections, so they introduced the Sample Entropy (SampEn), which is always a regularity estimator [2]. In the context of these parameters (ApEn, SampEn) the term complexity refers to the predictability of the system state, by knowing the initial conditions. The less predictable the states are, the more complex the system is.

On the other hand, Lempel and Ziv introduced a measure of complexity which evaluates the number of distinct sub strings and the rate of their recurrence: the Lempel Ziv complexity (LZ). This index reflects the gradual increase of new patterns along the given sequence. In this case the word complexity refers to the so-called algorithmic complexity, which is defined according to the Information Theory as the minimum quantity of information needed to define a binary string [3]. In case of random strings, the algorithmic complexity is the length of the string itself. In fact any compression effort will produce an information loss.

In this work we have firstly compared the index performances and we have caught analogies and differences. Successively we have evaluated their ability in the identification of severe Intrauterine Growth Restricted (IUGR) fetuses through the analysis of Heart Rate Variability (HRV) signals.

The fetuses are classified as small for gestational age (SGA) when the fetal biometry measures (fetal weight, abdominal and femur measure) have low values, generally below the 10th percentile. However SGA condition is defined in comparison with population standards, so including in this group also healthy fetuses having a small size. A growth restriction is instead a pathological state: the fetus is at risk of hypoxia. Moreover the IUGR condition is often associated with an increase of perinatal mortality and morbidity [4].

2. Methods

2.1. Parameter description

Approximate Enropy. Given N point series $\{u(i)\}$, the algorithm constructs the sequences $x_m(i) = [u(i),...,u(i+m-1)]$, and it computes, for each $i \leq$N-m+1, the quantity

$$C_i^m(r) = N^{-1}\{ \text{number of } j \leq N-m+1 \mid d[x_m(i), x_m(j)] \leq r \} \quad (1)$$

that measures, with a tolerance r, the regularity of patterns comparing them to a given pattern of length m (m and r are fixed values: m is the detail level at which the signal is analyzed and r is a threshold, which filters out irregularities).

The statistics is defined as :

$$\mathrm{ApEn}(m,r)=\lim_{N\to\infty}[\Phi^m(r)-\Phi^{m+1}(r)], \quad (2)$$

where $\Phi^m(r)=(N-m+1)^{-1}\sum_{i=1}^{N-m+1}\ln C_i^m(r)$.

$\mathrm{ApEn}(m,r,N)=[\Phi^m(r)-\Phi^{m+1}(r)]$ is the estimator of this index for an experimental time series of a fixed length N [5].

The parameter adopted for this work are m=1 and r=0.1, 0.2, m=2 and r=0.15.

Sample Entropy. The differences of SampEn index with respect to ApEn are: (i) self-matches are not counted, (ii) only the first N-m vectors of length m are considered and (iii) the conditional probabilities are not estimated in a template manner: they do not adopt as probability measure the ratio of the logarithmic sums, but they compute directly the logarithm of conditional probability [2][6].

The parameter adopted for this work are m=1 and r=0.1, 0.2, m=2 and r=0.15.

Lempel Ziv Complexity. In the calculation of the LZ complexity $c(n)$ we firstly define an alphabet A, the set of symbols which compose the sequence (for a binary string A is simply {0,1}). We define a finite-length sequence S, where $S=s_1\ s_2\ \ldots s_n$ and $s_i\in A$, and the vocabulary of the sequence $v(S)$, that is the set of all substrings of S. For example if S=10 then $v(S)$={1,0,10}.

Now let S and Q denote, respectively two strings, and SQ be the concatenation of S and Q, while the string $SQ\pi$ is derived from SQ by deleting its last character (π means the deletion of the last symbol).

In a given sequence of length n, at the start, $c(n)$ is 1, $S=s_1$, $Q=s_2$ and $SQ\pi=s_1$. For generalization sake, let's suppose $S=s_1\ s_2\ \ldots s_r$ and $Q=s_{r+1}$. If $Q\in v(SQ\pi)$ then s_{r+1} is a substring of SQ so the new sample does not add information to the actual sequence content. Then we leave S unchanged and we add to Q a new character s_{r+2}. We verify if Q belongs to $v(SQ\pi)$ or not until Q is not a substring of $SQ\pi$. When it occurs, we increase $c(n)$ by one and we concatenate S and Q to form $S=SQ=s_1\ s_2\ \ldots\ s_r\ s_{r+1}\ \ldots s_{r+i}$. We add to Q a new character s_{r+i+1} and we repeat the procedure until Q contains the last character of the sequence. At that time, the number of different substrings of S is $c(n)$, which reflects the new pattern rate arising while the sequence length increases. In this paper, we use the complexity measure in its normalized form [7], i.e. the measure of complexity is normalized by a factor depending on the sequence length. This permits to compare the complexity values of two strings different in length. The maximum value LZ index can assume is thus 1, corresponding to a completely random string.

In order to estimate the complexity measure for the HRV time series, we have transformed the signals in symbolic sequences. The simple increase or decrease of the signal is our encoding criterion, in this way we avoid the dependence of results on superimposed noise [8].

As a coding procedure we adopted both a binary and a ternary code. From an HRV series $\{x_n\}$, we construct a new sequence by mapping the original one through a binary alphabet. We symbolize with 1 a signal increase ($x_{n+1}>x_n$) and with 0 a decrease ($x_{n+1}\leq x_n$). In case of ternary alphabet, 1 denotes the signal increase ($x_{n+1}>x_n$), 0 the decrease ($x_{n+1}<x_n$) and 2 the signal invariance ($x_{n+1}=x_n$).

2.2. Simulations

We have compared the performances of LZ complexity and regularity estimators. We evaluated their behavior for known signals: $1/f^\alpha$ noises, MIX(p) process, Logistic Map.

The $1/f^\alpha$ signals were built by generating a $1/f^\alpha$ power spectral density, with random phases (uniformly distributed on [-2π, 2π]) and by applying the inverse Fast Fourier Transform (FFT). The α values range from 1 up to 2 and the time series generated are 30000 point long.

The MIX(p) process is defined as $(1-z)\cdot x+z\cdot y$, where z is a random variable, that assumes value 1 with probability p and 0 with probability 1-p, x is a sequence generated by the equation $x_j=\sqrt{2}\sin(2\pi j/12)$ and y is a uniformly distributed variable on $[\sqrt{3},-\sqrt{3}]$. The lower is p, the more periodic and regular the signal is.

We have generated 5 sequences of N=30000 samples for each p (0÷0.9).

We have also generated 30000 point long time series from a Logistic Map process $x_i=x_{i-1}\cdot r\cdot(1-x_{i-1})$, by varying the r parameter (r=3÷4).

2.3. Data Collection

We have analyzed HR signals belonging to fetuses, whose gestational age was in the range between the 27th and the 34th gestational week. The HRV signals were recorded by a CTG monitor HP M1351A.

The fetuses were then classified as normal, severe IUGR and not severe IUGR. The normal group includes 17 fetuses without pathologies, delivered by spontaneous labour with a good score at delivery. The severe IUGR group includes 23 SGA fetuses with complications so that they were preterm delivered by a caesarean section. The

not severe IUGR group includes 19 small fetuses classified as IUGR, but delivered at term.

We calculated the parameters over short sequences as suggested in a recent study [9]. In fact during CTG monitoring, the fetus can move and breath as well as contractions can occur. These events dramatically change the fetal heart rate variability signal, demonstrating its non-stationary nature. This methodology would allow to capture the development of signal modifications and, eventually, to understand physiological implications. Each index (LZ complexity, ApEn, SampEn) is thus computed for adjacent 360 point intervals of signal. The HR signal was corrected before the analysis and the signal subsets with insufficient quality were excluded from the computation. The index we considered was obtained as the mean of index values in the entire signal. Successively we modified the method by adopting 512 point length. We repeated also the analysis by introducing a more restrictive criterion: the signal intervals containing zeros were excluded from the analysis (0 is the value that the HP monitor attributes when the signal is unavailable or the preprocessing procedure judges it unacceptable; it has no physiological meaning). We computed again all the parameters by considering also signal subsets with 50% overlap.

3. Results

3.1. Simulations

Simulations test the ability of the indexes in the signal differentiation. Figure 1 illustrates results obtained with time series from a MIX(p) process. The lower p is, the more periodic and regular the signal is. As expected, time series corresponding to low p values show lower values of LZ complexity and entropy indexes: their behavior is similar.

The same performance was obtained also for the Logistic Map, that is LZ and entropy indexes have increasing values for increasing r values.

On the contrary, LZ complexity index has a completely different trend in respect to the entropy estimators in the case of $1/f^{\alpha}$ noises. As the α parameter increases the signal appears more regular so the entropy estimators assign it a low values, whereas the LZ complexity maintains high values (Figure 2). This is in accordance to the unpredictability of 1/f process.

3.2. FHR analysis

The analyses show that only the LZ complexity index (ternary coding) is able to discriminate the severe IUGR from healthy fetuses (P-value ANOVA<1% and P-value Scheffè test <5%).

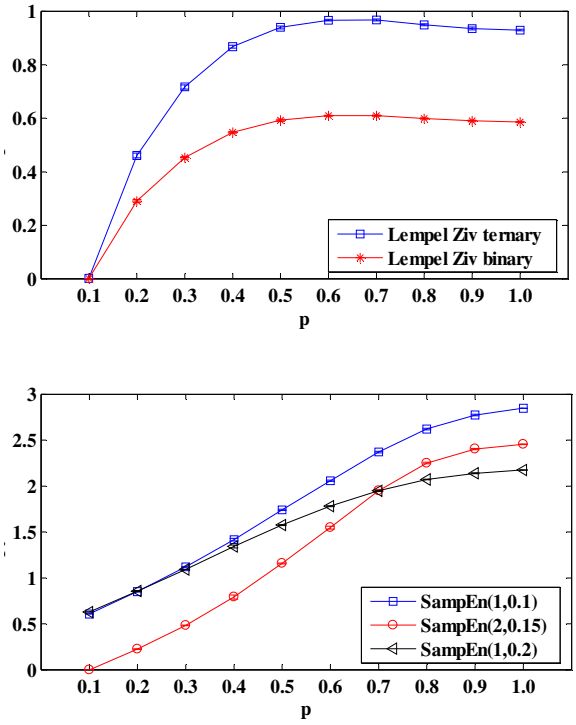

Figure 1. LZ complexity values (upper panel) and sample entropy values (lower panel) computed for a MIX(p) process. The values are plotted as a function of p.

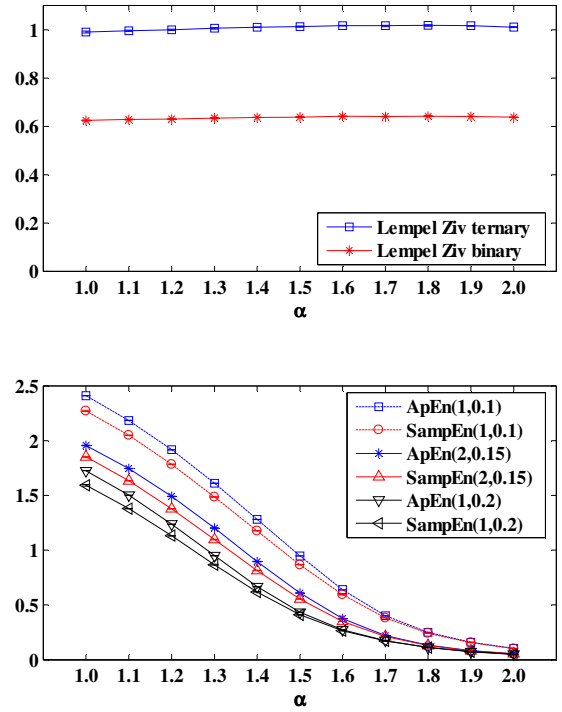

Figure 2. LZ complexity values (upper panels) and entropy values (lower panels) computed for $1/f^{\alpha}$ process. The values are plotted as a function of α.

Table1: Parameters values obtained from the three groups (avg±std). The results refer to the analysis over 360 points FHR overlapping subsets by adopting the more restrictive criterion (absence of zeros). *=P-value ANOVA<1% and P-value Scheffè test <5%.

	LZ (binary)	LZ (ternary)	ApEn(1,0.1)	SampEn(1,0.1)	ApEn(2,0.15)	SampEn(2,0.15)	ApEn(1,0.2)	SampEn(1,0.2)
Healthy	1,013±0,038	**0,887±0,029***	1,220±0,159	1,169±0,219	0,764±0,099	0,790±0,169	0,790±0,159	0,671±0,160
severe IUGR	1,022±0,019	**0,951±0,028***	1,308±0,200	1,304±0,273	0,786±0,094	0,906±0,233	0,862±0,191	0,750±0,193
not sev. IUGR	1,032±0,018	**0,914±0,035***	1,302±0,146	1,276±0,184	0,811±0,082	0,880±0,155	0,845±0,149	0,728±0,140

Moreover, when we analyzed overlapped intervals of 360 and 512 points, by adopting the restrictive criterion (absence of zeros), the Lempel Ziv complexity (ternary coding) discriminates significantly the three groups of fetuses (ANOVA test and Kruskall-Wallis test were performed among the three patient groups, then post-hoc comparisons were made by Scheffè test, P-value ANOVA<1% and P-value Scheffè test <5%). Some results are reported in table I.

The introduction of the entropy estimators did not provide instead a significant improvement in the identification of fetal distress.

Moreover, we wanted also to verify if the indexes perform in the same way or better when computed for long sequences, i.e. over the entire signal. We have considered in this case only time series with less than 250 zeros as sequences of zeros may be interpreted as repetitive patterns. (It must be said that in this way the number of subjects was drastically reduced: 10 healthy, 8 distressed IUGRs, 16 not severe IUGRs.) The analyses show that all indexes, except the binary LZ, discriminate between the IUGR groups and the control group, but no one is able to separate the distressed from the slight IUGR fetuses.

4. Discussion and conclusions

The simulation results demonstrate that the complexity estimators considered in this work provide different information about the time series. In fact in the case of $1/f^{\alpha}$ noise process, we have signals apparently more regular for increasing α, but which are generated by a stochastic process with long term correlation. This can be recovered in the index values. In fact the entropy values have a growing trend for increasing α: the signals are more regular and then the entropy estimators recognize the repetitive patterns. The unpredictable nature of the signals are instead detected by the LZ parameter, that evaluates the arising of new patterns along the time series. In this way the entropy and the LZ complexity indexes can be considered complementary parameters.

Furthermore the analysis about the fetal HRV shows that the introduction of the regularity estimators does not improve the FHR classification. On the contrary, the Lempel Ziv complexity seems to be a possible solution to the clinical problem of the identification of distressed IUGR fetuses. In fact LZ index separates IUGR group from the healthy SGA fetuses.

References

[1] Pincus SM, "Approximate entropy (ApEn) as complexity measure", Chaos, 1995, vol. 5(1): 110-117.

[2] Richman J.S. and Moorman J.R., "Physiological time-series analysis using approximate entropy and sample entropy", Am J Physiol Heart Circ Physiol, 2000, vol. 278: H2039-2049.

[3] Lempel A. and Ziv J., "On the complexity of finite sequences", IEEE Trans. on Inf. Th,, 1976, vol. 22: 75-81.

[4] Malcus P, Antenatal fetal surveillance. Current opinion in Obstetrics and Gynaecology, 2004; vol.16: 123-128.

[5] Pincus S.M., Gladstone I.M., Ehrenkranz R.A., "A regularity statistic for medical data analysis", J Clin Monit, 1991, vol. 7: 335-345.

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

[7] M. Ferrario, M.G. Signorini, S. Cerutti, "Complexity Analysis of 24 Hours Heart Rate Variability Time Series", Proc. of the EMBC 04 Conference, IEEE Press.

[8] Yang A.C-C *et al*., "Linguistic Analysis of the human heartbeat using frequency and rank order statistics", Phys Rev Letters, 2003, vol. 90, no. 10:108103-1—108103-4.

[9] Signorini M.G., Magenes G., Cerutti S., Arduini D., "Linear and Nonlinear Parameters for the Analysis of Fetal Heart Rate Signal from Cardiotocographic Recordings", 2003, IEEE Trans Biom Eng, vol.50(3): 365-375.

Address for correspondence:

Manuela Ferrario
Politecnico di Milano- Dip. Bioingegneria
P.zza Leonardo da Vinci, 32
20133 Milano (MI)-Italy
manuela.ferrario@biomed.polimi.it