Complexity of Continuous Functions and Segmentation of Time Series

Boris Darkhovsky* Alexandra Piryatinska †

Abstract

In many applications time series are sequences of connected, distinct segments which are generated by their own individual mechanisms. To analyze such series it is necessary to split them into these segments. If time series is generated by stochastic mechanisms, then the segmentation problem can be reduced to the classical change-point detection problem. However it is not the case for deterministic or mixed mechanisms. A new approach to this problem based on the novel concept of the complexity of a continuous function is proposed. The complexity of a continuous function is defined as the fraction of the function values necessary to recover the original function via a certain fixed family of approximation methods without exceeding a given error. It is shown that the dependence of the complexity of a function on the reconstruction error can be well approximated in logarithmic coordinates by an affine function. Its parameters, calculated dynamically, are then used as diagnostic sequences to find the change-points of the original time series. The effectiveness of this procedure is verified in the case of several simulated time series and this approach is applied to the EEG data.

Key Words: complexity, time series, non-stationarity, EEG.

Introduction

In many applications we deal with non-homogeneous time series. We call a time series non-homogeneous if it is a sequence of connected distinct segments, which are generated by their own individual mechanisms. The generating mechanism can be stochastic as well as deterministic. A sample path of the non-stationary random process is a time series generated by a stochastic mechanism. An example of the non-homogeneous deterministic time series is a trajectory of a non-linear dynamical system with varying coefficients.

To model such time series it is necessary to split them into certain *homogeneous segments*. For stochastic non-stationary time series the classical change-point detection algorithms can achieve this segmentation.

Over the past 50 years the change-point detection problem of stochastic data spawned a large mathematical statistics literature (see, e.g., reviews[1]). At the end of the 70's Darkhovsky(see[2]) suggested that the detection of changes in any probabilistic characteristics of a random process (field) can be reduced to detection of changes in the mean of some new random sequence (called the *diagnostic sequence*) formed from the original data.

However, in practice, time series are not always stochastic. Moreover, use of probabilistic tools in analysis of time series is often justified by the fact that the true phenomena underlying generation of the data are unknown. Such is the case of the EEG recordings, and the financial time series, where there are no generally accepted phenomenological models.

Even in the case when data are indeed generated by probabilistic mechanisms, there are significant difficulties in the implementation of the segmentation procedures. In particular, if the probabilistic characteristics responsible for the change of the generating mechanism are unknown *a priori*, then, in principle, we have to go through all the (generally infinite) set

*Institute for Systems Analysis RAS, Moscow, Russia
†San Francisco State University, 1600 Holloway Ave, San Francisco, CA 94132

of such characteristics to construct the diagnostic sequence. That, of course, significantly complicates the analysis.

To address this problem we pose the following question: could we find an "inner" characteristic of the time series which, on the one hand, would permit segmentation of the data, and on the other hand, would not depend on the type of the mechanism (probabilistic or not) that generates the data? We believe that the concept of the *complexity* provides an answer to the above question.

In this paper we assume that our data are values of continuous vector-valued functions defined on a compact region in the Euclidean space. The complexity of a continuous function will be the measure used for segmentation purposes.

The idea of the quantitative estimation of the complexity of a continuous function was first proposed by Darkhovsky in [3]. It was then pre-tested successfully on the human EEG data [3]. In this paper we suggest an effective characterization of complexity of a continuous function given on a unite cube in a finite dimensional space. In particular, we show that for Hölder class functions the dependence of the function complexity on the reconstruction error can be well approximated in logarithmic coordinates by an affine function. In other words, complexity of the Hölder class of functions can be characterized by a pair of real numbers. Based on this result we can formulate our main conjecture: *complexity of an individual function from the Hólder class can be characterized by two numbers*. We call these numbers the complexity coefficients. The above conjecture is supported in this paper by a series of simulations. This simple characterization of the complexity of an individual function from the Hölder class allows us to develop a methodology for segmentation of time series *regardless the generating mechanism*. We also conduct new change-point detection experiments for real EEG data using the dynamically calculated complexity parameters as the diagnostic sequences.

The paper is organized as follows. In Section 1 we give a brief review of the known approaches to define complexity. In Section 2 we define the complexity of an individual continuous function as well as the complexity of the functional class on a unit cube in the Euclidean space and give its characterization. In Section 3 we present our main conjecture and verify it on a series of simulations. In Section 4 we use simulations to verify effectiveness of our approach to detection of the change-points and, in particular, we use our segmentation procedure to identify sleep stages of neonates employing EEG records.

1. Complexity of an object - a brief review

The concept of the "complexity" of an object is one of the fundamental scientific paradigms. Numerous attempts have been made to apply it in practice. Let us briefly recall the basic ideas associated with it.

One of the first efforts to provide a quantitative approach to the concept of the "complexity of a physical system" was made in the 1870s by Boltzmann, a physicist who introduced the notion of the entropy in equilibrium statistical physics. Here the entropy H_f is defined via the formula

$$H_f = \lim_{p \to \infty} \frac{\log N(p)}{p},$$

where p is the number of degrees of freedom, and $N(p)$ is the number of different system configurations that have certain properties. The greater the entropy, the more "complicated" the system is.

In the 1940s Shannon developed the concept of entropy to measure the uncertainty of a discrete random variable. The Shannon entropy H_s is defined as follows. Consider

a sequences of n i.i.d. random variables with a finite number of values $x_1, \ldots, x_r$, and corresponding probabilities $p_1, \ldots, p_r$. Then

$$H_s = -\sum_{i=1}^{r} p(x_i) \ln p(x_i).$$

Again, we can say that more complex probability distributions have greater entropy. From this point of view, the most "complex" distribution is the uniform distribution as it generates the greatest uncertainty.

In the 1950s Kolmogorov and Sinai (see, e. g., [5]) introduced the entropy concept to the theory of dynamical systems. In fact, their definition was a generalization of Shannon's entropy. The entropy of a dynamical system is the coefficient of the asymptotic behavior of the logarithm of the number of different types of trajectories of a dynamical system when the time goes to infinity. Again, the entropy of a dynamical system may serve as a measure of "complexity": the more "complex" the system the richer the variety of its trajectories.

The main assumption of nonlinear dynamics (that is, the theory of dynamical systems, including deterministic chaos theory) and, consequently, the basis for the application of the concept of entropy to a dynamical system, is the assumption that the studied dynamical system is stationary (i.e., it does not change its properties over time). Any violation of this assumption makes the above approach inappropriate, but non-stationarity is, perhaps, the main feature of the data generated by different mechanisms.

Note, that even for the stationary case the problem of a reconstruction of the full description of the dynamical system from its observed trajectory is very complicated and there is no perfect solution to it (see, for example[6]).Thus, our conclusion is that the concept of entropy of a dynamical system is not an appropriate tool for a study of non-stationary sequences.

At the beginning of the 1980s, Kolmogorov ([7]) suggested an algorithmic approach to the notion of "complexity" of an object. The main idea of his approach is as follows: *A "complex" object requires a lot of information for its reconstruction and, for "simple" objects, little information is needed*. He formalized this idea in the language of the theory of algorithmic complexity. Roughly speaking, algorithmic complexity measures the length of the program leading to the selection of the object from a set of objects. This idea is the closest one to our approach to the complexity of a continuous function defined on a compact set in a finite dimensional space.

2. Definition of function complexity and its characterization

2.1 Continuous variable

2.1.1 Complexity of an individual function

In this section, without loss of generality, we assume that a continuous vector-valued function is defined on the unit cube $\mathbb{I}$ in the space $\mathbb{R}^k$. Here, we will define complexity as a functional of a vector-valued function. Since we consider only vector-valued functions with a finite number of components, it is sufficient to restrict our attention to a scalar continuous function $x(t)$, $t \in \mathbb{I}$. On the set of such functions we introduce the norm $\|\cdot\|$. To be able to compare the complexity of different functions, it is reasonable to assume that $\|x(t)\|$=1, i.e., to consider $x(t)/\|x(t)\|$ instead of $x(t)$.

Let $\mathbb{Z}_h$ be a k-dimensional grid with spacing h and $\mathbb{I}_h = \mathbb{I} \cap \mathbb{Z}_h$. Assume that we only know the values of $x(t)$ at the points of the set $\mathbb{I}_h$. With what precision can we reconstruct the function $x(t)$ utilizing only this information?

Suppose we are given a set $\mathcal{F}$ of approximation methods. Using this set of methods we recover missing function values using its values at $\mathbb{I}_h$. Consider the approximation error

$$\delta(h) = \inf_{\mathcal{F}} \|x(t) - \hat{x}(t)\|,$$

where $\hat{x}(t)$ is the approximation of the function constructed by one of the allowable methods of approximation. The infimum is taken over the whole set of allowable methods of approximation.

We now construct the graph of the error function $\delta(h)$, increasing the value of h (i.e., reducing the amount of available data). It is clear that the function $\delta(h)$ must increase monotonically, because the increase of the grid spacing means that we discard more and more function values. Hence, its approximation becomes worse and worse assuming that the set of approximation methods is fixed. Therefore, the error function $\delta(h)$ is a monotone nondecreasing positive function of its argument. If we fix a certain "acceptable" (user-specified) level of approximation error, then we can determine the fraction of the function values that could be discarded to reconstruct the original function via a certain fixed family of approximation methods without exceeding a given error.

Note that the error of approximation should be related to the norm of the function, but since we assume that the function is pre-normalized , $\delta(h)$ is actually the relative error.

Let

$$h^*(\epsilon) = \begin{cases} \inf\{h \le 1 : \delta(h) > \epsilon\}, & \text{if} \quad \{h : \delta(h) > \epsilon\} \neq \emptyset \\ 1 & \text{if the set is empty} \end{cases} \tag{1}$$

Thus the value $h^*(\epsilon)$ is the minimum grid spacing for which the error of the function reconstruction from its values on the grid exceeds a given ϵ.

The value $(h^*(\epsilon))^k$ is the fraction (in relation to the volume of the unit cube) of the discarded function values.

It is clear that the value $(1/h^*(\epsilon))^k$ estimates the number of points in the set $\mathbb{I}_{h^*(\epsilon)}$. So, it is natural to use $\frac{1}{h^*(\epsilon)}$ as a measure of the function complexity.

Definition 1. *The number*

$$S(\epsilon, \mathcal{F}, \|\cdot\|) \stackrel{def}{=} S(\epsilon) = \log \frac{1}{h^*(\epsilon)}$$

is called the $(\epsilon, \mathcal{F}, \|\cdot\|)$*-complexity of an individual function* $x(t)$ *(or, briefly, the* ϵ*-complexity).*

Thus, the complexity of a continuous function is the *logarithmic* fraction of the function values necessary to be retained to recover the original function via a certain fixed family of approximation methods with relative error not exceeding ϵ.

We will use this argument to analyze functions defined by their values at a discrete set of points.

It is natural to assume that $\mathcal{F}$ contains a method of approximation utilizing affine functions on R^k. It is easy to see that if $x(t)$ is an affine function then for its error-free recovery it is sufficient to use $(k+1)$ of its values. But $\sharp(\mathbb{I}_h) \ge (k+1)$, for any $0 < h \le 1$. Therefore, according to (1), for an affine function, $h^*(0) = 1$, and therefore, the complexity $S(0)$ of any affine function is equal zero, which is intuitively clear.

It should be noted that the proposed complexity measure is an individual characteristics of a particular function, rather than of a set of functions generated by the same mechanism; the latter situation arises when one considers the entropy of a dynamical system.

Furthermore, this measure is not related to possible mechanisms of function generation (so, for example, it does not depend on whether the function is a sample path of a random

field/process or a trajectory of a dynamical system). In general, the proposed measure is constructed with the view towards its practical usefulness, i.e., it is explicitly constructed to make it easy to work with the discrete information about the function.

2.1.2 Complexity of a functional class

Let

$$\omega_x(h,\theta) = \max_{(t,s)\in\mathbb{I},\|t-s\|\le h} |x(t)-x(s)|$$

be the modulus of continuity of the function $x(t)$. Here, θ denotes a vector of parameters. For example, functions from the Hölder class, with $\omega(h,\theta) = L\cdot h^p$, have the vector parameter $\theta = (L,p)$.

Let us consider a class of functions X_ω with a fixed, up to the parameter θ, modulus of continuity $\omega(h)$ (in what follows we will omit the subscript θ) and investigate the question of its ϵ-complexity $S_{cl}(\epsilon)$.

In accordance with the general idea of the previous section we define the complexity S_{cl} as follows:

Definition 2.

$$S_{cl}(\epsilon) = \log\big(1/h(\epsilon)\big), \tag{2}$$

where $h(\epsilon)$ is the grid spacing such that the maximum (over all functions in the class) error of the function recovery from its values on the grid with this spacing is equal ϵ in the case of the best reconstruction.

Thus, to estimate S_{cl} we have to find the minimax error of the function recovery from the given class using its values on the grid with spacing h.

Theorem 1. *Let the error ϵ of function recovery be measured in the standard norm of the space of continuous functions C. Then, for functions in the class X_ω, the following relationship holds:*

$$S_{cl}(\epsilon) = \log\frac{\sqrt{k}}{2\omega^{-1}(\epsilon)} \tag{3}$$

Proof. To prove the theorem, it is sufficient to find a maximal (over the class X_ω) recovery error for one cell of the grid in the unite cube. Let $t^0 = (t_1,\dots,t_k)$ and let e_i, $i = 1,\dots,m$ be k-dimensional vectors, whose components are equal to zero or one (here, $m = 2^k$ and $e_1 = (0,0,\dots,0),\dots,e_m = (1,1,\dots,1)$). Consider values of the function from this class in a single cell of such grid, i.e., consider the set $A_t^0 \stackrel{def}{=} \{x(t^i)\}_{i=1}^m$, $t^i = t^0 + he_i$, $t^0 \in \mathbb{I}_h$. We pose the problem of estimating the value of the function from our class at an arbitrary point τ inside the cell. In other words, In other words, we consider the problem of finding

$$\inf_{u\in\mathbb{R}} \sup |x(\tau)-u| \tag{4}$$

where the supremum is taken over all values of $x(\tau)\in X_\omega$ and over all admissible (in the class X_ω) values $\{x(t^i)\}$. Let u_{opt} be a solution of the problem (4).

Denote $\|\tau - t^i\| = r_i$, $\imath = 1,\dots,m$. Then the set of all possible values of $x(\tau)$ with the fixed set $\{x(t^i)\}$ is a segment that is formed by the intersection of the segments $[x(t^i)-\omega(r_i), x(t^i)+\omega(r_i)]$, $\quad i = 1,\dots,m$, and the optimal solution for the problem (4) under the same conditions (i.e., the minimax estimate of the function values at τ with the fixed set $\{x(t^i)\}$) is the midpoint of this segment.

Taking into account the symmetry we find that the maximum (over the class X_ω and the set $\{x(t^i)\}$) estimation error of $x(\tau)$ is equal to $\varphi(\tau) \stackrel{def}{=} \min_{1\le i\le m} \omega(\|\tau - t^i\|)$ (unless $x(t^i) = 0, \quad , i = 1, \ldots, m$; in that case in the problem (4), $u_{\text{opt}} = 0$).

Due to the finite number of the grid cells, the maximum (over the class X_ω) error $\delta_{cl}(h)$ of the function's approximation over the set $\mathbb{I}$, in the case of the best reconstruction, is equivalent to the error in one cell in a particular functional norm. Let τ^* be the center of the cell $\mathbb{I}_h$. Then, we can obtain $\delta_{cl}(h) = \varphi(\tau^*) = \omega(\sqrt{k}h/2)$.

□

The following Corollary is an immediate consequence of Theorem 1.

Corollary 2. *For the Hölder class functions, the complexity can be expressed as follows;*

$$S_{cl}(\epsilon) = A + B \log \epsilon. \tag{5}$$

In the case of a sufficiently rich set $\mathcal{F}$, the complexity of an individual function $x(t) \in X_\omega$ satisfies the inequality $0 \le S_x(\epsilon) \le S_{cl}(\epsilon)$. This suggests that the complexity of an individual function from the Hölder class satisfies (5) for some values of the coefficients A and B.

2.2 Discrete variable

In the majority of applications, we have to deal with functions given by their values at a discrete set of points (i.e., a finite sample). We still assume that they are the values of a continuous function on the grid in the unite cube in a k-dimensional space. Let us discuss how the definition of complexity has to be adjusted to this situation.

Let N^k be the number of values of a continuous function $x(t)$ on the k-dimensional grid in the unit cube. We discard a fraction $[(1-\alpha)N^k]$, $0 < \alpha < 1$, of the given values ($[a]$ denotes the integer part of a) and reconstruct the function in the discarded points using the set of approximation methods $\mathcal{F}$ and the values at remaining $[\alpha N^k]$ points (since we work in a finite-dimensional space, the choice of the approximation norm does not matter). Let us consider the quantity $h^*(\epsilon)$ introduced in (1). If a k-dimensional unit cube has N^k values of a continuous function, then the k-dimensional cube with side $h^*(\epsilon)$ has $[h^*(\epsilon)N]^k$ values, i.e., the initial set of N^k values of the function is replaced by $[N^k/[h^*(\epsilon)N]^k]$ values. In other words, the number of values, sufficient for the function reconstruction with the relative error not exceeding ϵ is equal to $n^* = N^k/[h^*(\epsilon)N]^k$.

Therefore, in analogy with the case of the continuous variable, we can formulate the following

Definition 3. *The value*

$$S_N(\epsilon) = \log \frac{N^k}{[h^*(\epsilon)N]^k} \tag{6}$$

is called the ϵ-complexity of an individual function $x(t)$, given by the set of its discrete values.

From (6), it follows that the $\lim_{N\to\infty} S_N$ exists and we denote it $S(\epsilon)$. But the growth of N means the growth in the sampling frequency if the domain of the function is fixed. Therefore, in the case of a sufficiently high sampling frequency of the function, the complexity of the sample calculated over the discrete set of values approaches the true complexity.

Of course, the question arises what should be the sampling frequency to make this difference sufficiently small, but in the case of data obtained with the same sampling frequency, this question is not essential. In any case, we must bear in mind that the comparison

of functions in the case of a discrete set of values can be performed only when the sampling frequency is the same.

In view of the above comments, for the Hölder class functions (compare with (5)) we have the following equality:

$$S_N(\epsilon) = \mathcal{A} + \mathcal{B}\log\epsilon \tag{7}$$

3. Algorithmic procedure to estimate complexity

3.1 The Main Conjecture

Suppose we are given an array of size N of the values of a function. Let us choose a number $0 < \mathbb{S} < 1$, and discard from the array $[(1-\mathbb{S})N]$ values. In the next step we use the remaining $[\mathbb{S}N]$ values to approximate the values of the function for all discarded points using a collection $\mathcal{F}$ of approximation methods, and find the best approximation (the approximation with the smallest error).

Two factors have to be taken into account. First, the remaining points should be distributed relatively uniformly. Second, since the error of the approximation depends on the location of the remaining points, for the sake of the stability of the method it is expedient, for a given percentage of removed points, to choose different selection schemes and average the corresponding minimal approximation errors over them. This allows us to smooth out the unavoidable random errors in the calculations.

Thus, for given values of $\mathbb{S}$ we determine the value of minimal error ϵ of the function recovery. It is obvious that, for any $\mathbb{S} > 0$, the error of the function recovery tends to zero, as $N \to \infty$. On the other hand, if the sample size N is too small, then estimation of the recovery error is affected by the calculation errors even for large values of $\mathbb{S}$.

For this reason, and based on the previous discussion (see (5),(7)), we can state the following main conjecture.

Conjecture 1. *Let $x(t)$ be a function from the Hölder class given by its discrete values. We can chose the range of the data sizes so that within this range there exists an interval $[\alpha, \beta]$, $0 < \alpha \le \mathbb{S} \le \beta < 1$ such that for all $\mathbb{S}$ within this interval the following relationship holds:*

$$\log\epsilon = \mathbb{A} + \mathbb{B}\log\mathbb{S} \tag{8}$$

In what follows we call the parameters $(\mathbb{A}, \mathbb{B})$ *c*omplexity coefficients. Below we present the algorithm of the complexity coefficients estimation and results of the simulations.

3.2 Estimation procedure and simulations to verify conjecture

To verify the affine relationship (8) in the conjecture (1) we perform a series of the experiments(simulations) using finite samples from deterministic continuous functions and simulated stochastic processes.

For each case the experiment is conducted according to the following algorithm:

1. Select $\mathbb{S}$, the fraction of the remaining points as follows: $\mathbb{S}_1 = 50\%$, $\mathbb{S}_2 = 33\%$, $\mathbb{S}_3 = 29\%$, $\mathbb{S}_4 = 25\%$, $\mathbb{S}_5 = 22\%$.

2. Discard the values of the functions at points which are placed uniformly, or almost uniformly according to the following algorithm:.

   Let $x_1, x_2, x_3, \ldots, x_n$ be the values of a function on a grid.

(a) $\mathbb{S}_1 = 50\%$: Values of $x_2, x_4, \ldots, x_{2i}, \ldots$; or $x_1, x_3, \ldots, x_{2i+1}, \ldots$; are discarded. Notice we have two different ways to discard function values;

(b) $\mathbb{S}_2 = 33\%$:Values of $x_1, x_4, x_7, x_{10}, \ldots$; or $x_2, x_5, x_8, x_{11}, \ldots$; or $x_3, x_6, x_9, x_{12}, \ldots$; are discarded. We have three different placements of discarded values;

(c) $\mathbb{S}_4 = 25\%$: Values of $x_1, x_5, x_9, x_{13}, x_{17}, \ldots$; or $x_2, x_6, x_{10}, x_{14}, x_{18}, \ldots$; or $x_3, x_7, x_{11}, x_{15}, x_{19}, \ldots$; or $x_4, x_8, x_{12}, x_{16}, x_{20}, \ldots$; are discarded. We have 4 different placements of discarded values;

(d) The procedures are similar in the case $\mathbb{S}_3 = 29\%$, and $\mathbb{S}_5 = 22\%$.

3. For each $\mathbb{S}_i$, and a fixed placement of the discarded points we reconstruct function values at the discarded points using the polynomial (up to the 5-th degree) approximation, and find the error of approximation for each degree of the polynomial.

4. Choose the method of function reconstruction which gives the smallest error.

5. Consider all possible placements of discard points and find the best reconstruction in each case.

6. Find the mean reconstruction error over all placements. It is our estimation of ϵ_i in the case of $\mathbb{S}_i$. Repeat the procedure for $i = 1, \ldots, 5$.

7. Consider points $(\log(\mathbb{S}_i), \log(\epsilon_i))$, and find the best linear fit $\log \epsilon = \mathbb{A} + \mathbb{B} \log \mathbb{S}$ using least square method.

The estimated values of $\mathbb{A}$, and $\mathbb{B}$ are our estimates for the complexity coefficients.

Figures 1, and 2, present results of such simulations for a number of functions and stochastic time series. In all figures the circles correspond to the points $(\log(\mathbb{S}_i), \log(\epsilon_i))$, and the line is the linear regression line. Also, we provide the relative error of the deviation of the $\log(\epsilon_i)$ from the regression line.

Notice that in our experiments we consider finite samples from continuous functions and, therefore, due to the conjecture 1 we have to choose the right sample frequency and sample size to get the affine dependence between $(\log(\mathbb{S}_i)$ and $\log(\epsilon_i))$. The sample size used in each simulation is indicated in each plot title. The grid spacing of the function argument is one unit (of time).

Below we are presenting some examples of our simulations.

1. *Continuous functions*. Figure 1 (top) shows the result of an experiment for the function $x = \sin(0.1t) + 2\cos(0.2t)$. We can see that the relative error is 0.15% and the data points line up very closely to the straight line.

2. *Continuous function with added noise*.

   (a) Figure 1 (middle) shows the results of simulations for the function $x(t) = \sin(0.01t + 0.01e(t))$. Here, $e(t)$ are i.i.d. random variables with $\mathcal{N}(0, 1)$ (normal with mean zero and variance one) distribution. The selected sample size is 90 points and the relative error is 0.2%.

   (b) Figure 1 (bottom) shows results of simulations for $x(t) = G(t)\sin((0.01 + 0.0001e(t))t)$, $G(t) = 1/(1 + exp(-t))$. The corresponding relative error is 0.2%.

3. *Continuous functions, solutions of non-linear differential equations*.

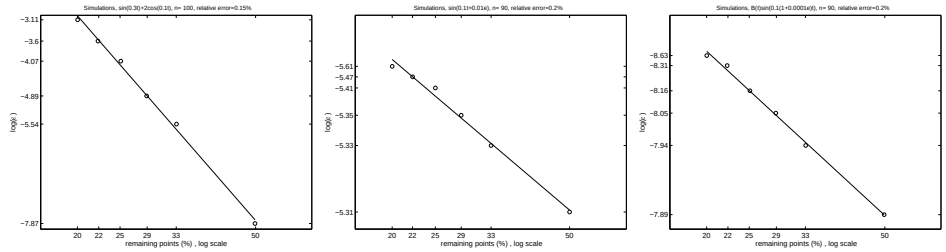

Figure 1: Examples of the verification of the linear relationship 8 for the continuous functions, and functions with added noise

(a) We consider a function $x(t)$ which is a solution of the non-linear differential equation

$$x'' + 0.1x' - x + \mu x^3 = 2\cos t, \tag{9}$$

where $2.5 \leq \mu \leq 4$. In the case $\mu > 4.3$, bifurcations can be observed. Figure 2 (top and middle) shows the results for the solutions of the equation 9 with parameters $\mu = 3.5$, and $\mu = 9$, respectively.

(b) We consider a Bessel function of the second kind which is a solution of the equation

$$tx'' + tx' + (t - \nu)x = 0.$$

Figure 2 (bottom) shows the results of our experiment for the Bessel function of the second kind with $\nu = 9$.

In all cases we observe a good fit of $(\log(\mathbb{S}_i), \log(\epsilon_i))$ by a linear function.

4. *Sample paths of random processes such as ARMA, ARIMA and FARIMA [8]*

(a) Figure 3 (top) shows results of an experiment for the simulated ARMA(p,q) process. with parameters $\varphi = (0.2, 0.3, 0.5)$ and $\theta = (0.1, -0.3, 0.5)$

(b) Figure 3 (middle) corresponds to the results of simulations of the ARIMA(p,d, q) process with parameters $\varphi = (0.1, 0.4, 0.7, 0.2)$, $\theta = 0.01$ and $d = 1$.

(c) Figure 3 (bottom) corresponds to the results of simulations of the FARIMA(p,d,q) process, with parameters $\varphi = (0.10.4)$, $\theta = (0.01, 0.2)$, $d = 0.3$.

The corresponding sample sizes are 100, 100, 90, and the corresponding relative errors are 0.04, 0.04, 0.12.

The results of our simulations are consistent with conjecture 1; the affine relationship (8) holds true.

3.3 Application of function complexity to the segmentation problem

For the sake of simplicity, we restrict ourselves to functions of a scalar argument (the argument of such function can be treated as time). Transition to the multidimensional case does not cause fundamental difficulties.

If the interval of observations is large enough then it can be assumed that the time series is generated by different mechanisms (stochastic as well as deterministic) in different

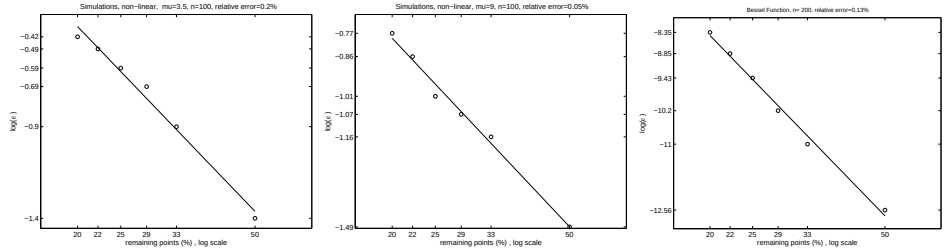

Figure 2: Examples of the verification of the linear relationship 8 for the continuous functions, solutions non-linear differential equations

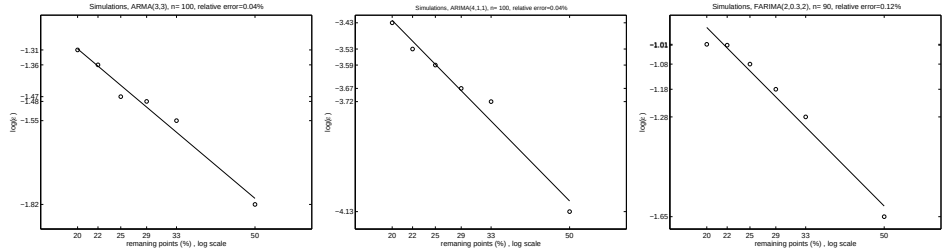

Figure 3: Examples of the verification of the linear relationship 8 for the processes ARMA (top), ARIMA(middle), FARIMA (bottom)

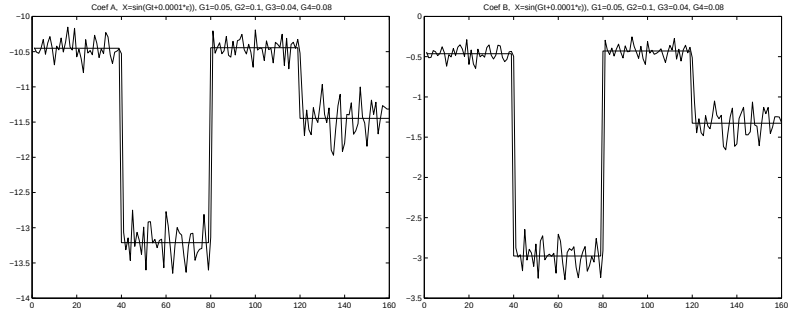

Figure 4: Simulation Results, Change point detection, sin with the added noise in the phase

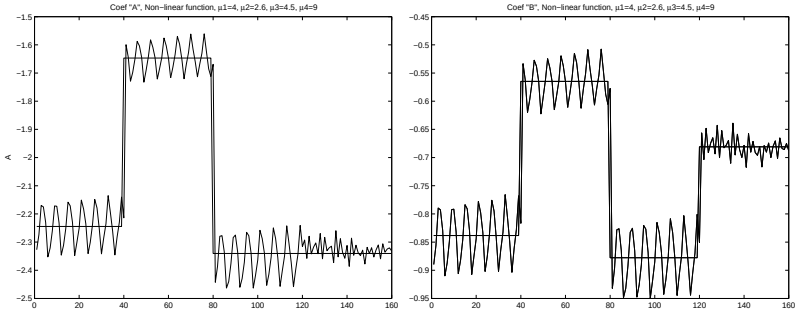

Figure 5: Simulation Results, Change point detection, Solutions of the non-linear equation

subintervals. For example, this is the case in the analysis of long EEG recordings, or long (days, weeks, months) financial time series.

In these cases we can estimate complexity parameters dynamically using a sliding window, or split the record into disjoint segments. Then the sequences $(\mathbb{A}(t), \mathbb{B}(t))$ (see (8)), $t = 1, 2, \ldots, m$ (here m is the number of sliding windows or segments), can be used as

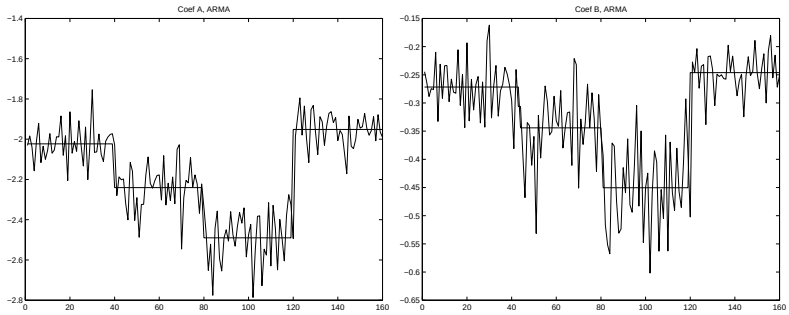

Figure 6: Simulation Results, Change point detection, ARMA processes

diagnostic sequences for the change-point detection algorithm. Since such a diagnostic procedure does not relate to the mechanism of data generation it creates possibilities to devise novel segmentation schemes.

To detect change-points in diagnostic sequences in our experiments we use the non-parametric method described in [2]. This methodology is based upon two main ideas. The *first* idea relies on the observation that detection of changes in any probabilistic characteristic can be reduced (with an arbitrary degree of accuracy) to detection of changes in the mean value of some new, *diagnostic* sequence constructed from the original one.

The *second* idea of the non-parametric approach employs the following family of statistics for detection of change-points in the mean

$$Y_N(n,\delta) = \left[\left(1-\frac{n}{N}\right)\frac{n}{N}\right]^{\delta}\left[\frac{1}{n}\sum_{k=1}^{n} x_k^N - \frac{1}{N-n}\sum_{k=n+1}^{N} x_k^N\right], \tag{10}$$

where $0 \le \delta \le 1$, $1 \le n \le N-1$, $X^N := \left\{x_k^N\right\}_{k=1}^N$, is a diagnostic sequence (see details in [2]).

3.4 Simulations to detect change-points

Here we perform simulations to verify effectiveness of above approach to detect change-points in non-homogeneous time series.

In each experiment we simulate four time series from the same type of processes but with different parameters and concatenate them. The length of each component is 8000 with the total length of the simulated process being 32000. Then we separate each process into segments of the length 200. We obtain 160 segments and estimate the complexity coefficients $\mathbb{A}(t)$ and $\mathbb{B}(t)$, $t = 1, \ldots, 160$. They form our new time series, our *diagnostic sequences*, for which we detect the change-points. Notice that the points 40, 80, 120 of the diagnostic sequence corresponds to the change-points of the original process. We applied the non-parametric change-point detection procedure for these two sequences, $\mathbb{A}(t)$ and $\mathbb{B}(t)$, and find the change-points. Then we plot the diagnostic sequences and means of the diagnostic sequences within the intervals where the processes is homogeneous.

Examples of our simulations are presented below.

1. Figure 4 shows the results for the process $x(t) = sin(G_j t + 0.0001e(t)))$, where $G_1 = 0.05$, $G_2 = 0.1$, $G_3 = 0.04$, $G_4 = 0.08$, and $e(t)$ are i.i.d. random variables from $\mathcal{N}(0,1)$. In this case both diagnostic sequences perfectly detect the change points. Left figure corresponds to the coefficient $\mathbb{A}(t)$, right, to $\mathbb{B}(t)$.

2. Figure 5. Here we find solutions $x_i(t)$, $i = 1, 2, 3, 4$ of the non-linear differential equation (9) with the corresponding parameters $\mu_1 = 4$, $\mu_2 = 2.6$, $\mu_3 = 4.5$, $\mu = 9$.

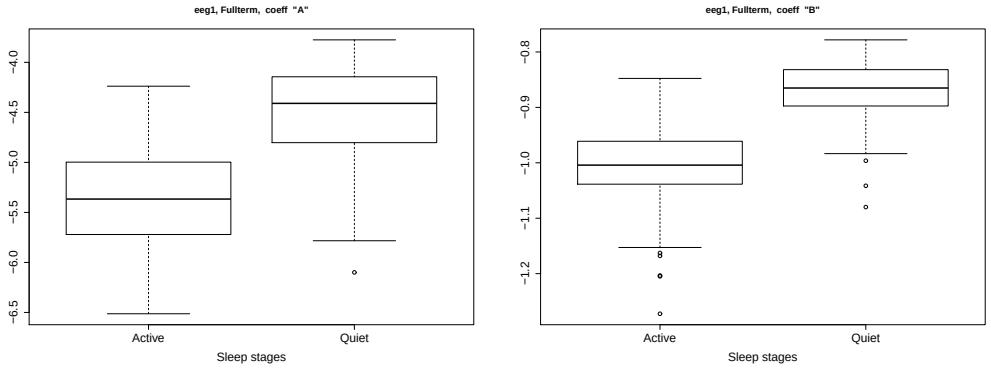

Figure 7: Example of the box-plots for the estimated complexity coefficients $(\mathbb{A}(t), \mathbb{B}(t))$ for fullterm neonate, characteristics A(left) and B (right) in active and quiet sleep stages

The diagnostic sequence $\mathbb{B}(t)$ perfectly detects the change-points, but we are unable to distinguish automatically between cases $\mu_3 = 4.5$ and $\mu = 9$ using the diagnostic sequence $\mathbb{A}(t)$.

3. Figure 4 corresponds to the case of ARMA(3,2) processes with the following parameters $\phi_1 = (-0.1, 0.3, 0.1)$, $\theta_1 = (0.2, 0.1)$; $\phi_2 = (0.5, -0.7, 0.9$, $\theta_2 = (0.5, 0.6)$; $\phi_3 = (0.4, 0.3, 0.4$, $\theta_3 = (0.1, -0.5)$; $\phi_4 = (-0.2, -0.3, -0.8)$, $\theta_4 = (0.2, 0.1)$. We observe that both complexity coefficients allow us to detect change points.

3.5 Applications to the EEG data

As an example of EEG data we use data from the EEG sleep study of neonates. The goal of that study is to identify sleep stages of neonates. We consider two main sleep stages, active and quiet. The records for 20 fullterm, and 16 preterm, healthy neonates at the 40-week post-conceptional age are available. Most of the fullterm neonate recordings were made on the 3rd day after the baby was born; preterm neonate recordings were obtained at the same post-conceptional age. Each recording collected signals on 14 EEG channels at a sampling rate of 64Hz. Minute by minute manual scoring of the sleep stages by a clinician (Dr. M. Scher of the Case Medical School at Case Western Reserve University) is also available. The portions of the signals corresponding to the awake state and those caused by artifacts (such as rapid physical movements) were removed from our data prior to analysis.

In this paper we are going to illustrate our methodology of sleep stage separation using just the single channel 11:$C4 - Cz$. A full analysis of the neonate sleep EEG data will be published elsewhere.

Description of the algorithm and results. Our analysis proceeds in the following three steps: In the *First Step* the signal for each of the 36 neonates in our study is divided into 30-second segments, and the coefficients $(\mathbb{A}(t), \mathbb{B}(t))$ are estimated for each segment thus providing us with what we call *diagnostic sequences*.

Figure 7 shows an example of the box-plots for the estimated coefficients $(\mathbb{A}(t), \mathbb{B}(t))$ for the 30 sec segments of active and quiet sleep stages correspondingly for a fullterm neonate. We can see that the complexity coefficients can be used to separate sleep stages utilizing the mean. Notice the existence of some outliers.

In the *Second Step*, the change-point detection algorithm is used to find the change-points in the means of the above two parameters for each neonate.

We use the sequences $(\mathbb{A}(t); \mathbb{B}(t))$(see (8)), $t = 1, 2, \ldots$ as the diagnostic sequences and the non-parametric change point detection algorithm is applied to each of them. Then

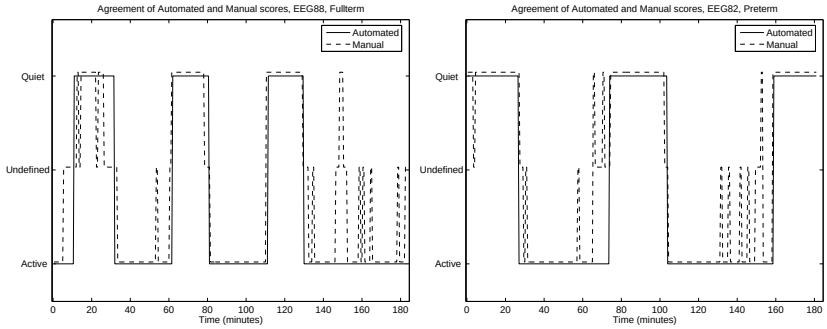

Figure 8: Example of the agreement percentage of the automated and manual scores for fullterm neonate(top) and preterm neonate (bottom).

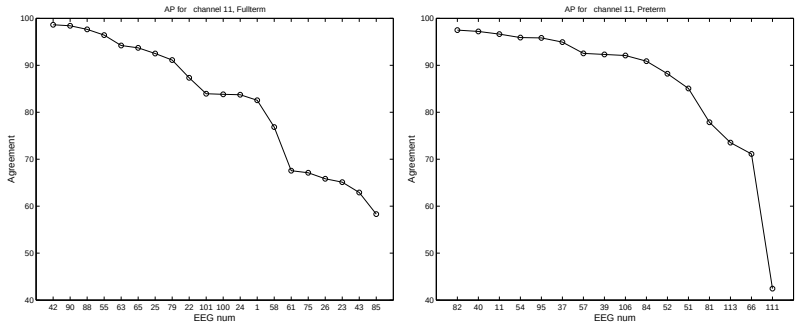

Figure 9: Agreement percentage by neonates, fullterm (top), preterm (bottom)

the original diagnostic sequence is replaced by its local means over the interval where the mean remains constant. We call the latter the *denoised diagnostic sequences*.

In the *Third Step*, the sleep stage separation operation is carried out. We apply the k-mean cluster analysis, with 2 clusters for 2-dimensional denoised diagnostic sequences for each neonate (for details see [9, 10]). We assign the clusters to the active and quiet sleep stages depending on whether $mean(\mathbb{A}(t_{active})) < mean(\mathbb{A}(t_{quiet}))$, or $mean(\mathbb{B}(t_{active})) < mean(\mathbb{B}(t_{quiet}))$. Then the results are compared with the manual scores by a clinician. Figure 8 shows degree of the agreement of the automated and manual scores for the fullterm (left) and preterm (right) neonates.

Figure 9 shows the performance (agreement percentage (AP)) of our automated sleep stage separation algorithm using just one channel for all individuals involved in our study. We observe that the mean agrement percentage (MAP) is 82.4% (standard deviation (sd) is 13.5%) for fullterm neonates, and 86.5% (sd=14.5%) for preterm neonates.The median AP is 83.9% for fullterm and 92.2% for preterm neonates. We observe that 9 out of 20 fullterm neonates and 12 out of 16 preterm neonates have $AP > 85\%$. Notice that the average AP between manual scores provided by different experienced physicians is 85 % ([11]).

4. Conclusions

In this paper we proposed a formal definition of the complexity of a continuous function defined on the unite cube in a finite-dimensional space. This definition is consistent with the idea of Kolmogorov complexity of objects and agrees with intuition. Roughly speaking, the complexity of a continuous function can be estimated by the fraction of the function

values which is required to reconstruct the function with a given accuracy and with a given set of approximation methods.

We show that the complexity has an effective characterization, due to the detected *affine dependance*. Function complexity can be characterized by a pair of real numbers, complexity coefficients. It enables us to develop a novel approach to the important problem of segmentation of time series. If the time series is generated by different mechanisms (either probabilistic, or deterministic) in different time intervals, then we can detect change-points using only the "internal" characteristics of the function (i.e., the complexity), independently of the actual mechanisms of data generation.

Our simulations and preliminary results of the EEG data analysis suggest that the proposed methodology can be widely used.

References

[1] A. Khodadadi, M. Asgharian, "Change-point Problem and Regression: An Annotated Bibliography". COBRA Preprint Series, Working Paper 44, Nov. 2008. Available:`http://biostats.bepress.com/cobra/art44`

[2] B. E.Brodsky and B. S. Darkhovsky, Non-parametric Statistical Diagnosis: Problems and Methods, Kluwer, Dordrecht, 2000.

[3] B. S. Darkhovsky, A. Ya. Kaplan, S. L. Shishkin, "On an Approach to the Estimation of a Complexity of Curves (by an Example of the Electroencephalogram in Humans)", *Automation and Remoute Control*, Vol. 63, no. 3, pp. 468-474, 2002.

[4] A. N. Shyriaev, Probability, Spinger-Verlag, New York Inc, 1995.

[5] Ya. G. Sinai, Introduction to ergodic theory, Fazis, Moscow, Russia, 1996.

[6] E. E. Demidov, Yu. V Darevskaya, O. A Morenkov and A. A Tovchigrechko, "Nonlinear correlation analysis", *Surveys Applied and Industrial Mathematics*, vol.6, issue 1, pp. 4-57, 1999.

[7] A. N. Kolmogorov, "Combinatorial foundations of information theory and probability calculations", *Uspexi math. nauk*, vol. 38, no. 4, pp. 27-36, 1983.

[8] P. J. Brockwell, R. A. Davis, Time Series: Theory and Methods, Springer-Verlag, New York Inc, 1991.

[9] A. Piryatinska, G. Terdik, W. A. Woyczynski, K. A. Loparo, M. S. Scher, and A. Zlotnikm, "Automated detection of neonate EEG sleep stages", *Computer Methods and Programs in Biomedicine* vol 9, pp.31-46, 2009.

[10] A. Piryatinska, W. A. Woyczynski, K. A. Loparo and M. S. Scher, "Optimal channel selection for analysis of EEG-sleep patterns of neonates", *Computer Methods and Programs in Biomedicine*, vol. 106, pp. 14-26, 2012

[11] H. Danker-Hopfe, G. Gruber, G. Klöch, J. L. Lorenzo JL, S. L. Himanien, B. Kemp, T. Penzel, H. Röshke, H. Dorn, A. Schlöl, E. Trenker, G. Dorffner, "Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders", *Journal of Sleep Reseach* vol. 13, pp. 63-69, 2004.

[12] J. G. Snodgrass , M. Vanerwart, ”A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity”, *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6, no. 2, pp. 174-215, 1980.