Analysis of Non-stationary Data for Heart-rate Fluctuations in Terms of Drift and Diffusion Coefficients

F. GHASEMI$^1$, MUHAMMAD SAHIMI$^2$, J. PEINKE$^3$ and M. REZA RAHIMI TABAR$^{4,5}$

$^1$Institute for Studies in Theoretical Physics and Mathematics, P.O.Box 19395-5531, Tehran, Iran; $^2$Department of Chemical Engineering, University of Southern California, Los Angeles, California 90089-1211, USA; $^3$Carl von Ossietzky University, Institute of Physics, D-26111 Oldenburg, Germany; $^4$CNRS UMR 6202, Observatoire de la Côte d’Azur, BP 4229, 06304 Nice Cedex 4, France; $^5$Department of Physics, Sharif University of Technology, P.O. Box 11365-9161, Tehran 11365, Iran

Abstract. We describe a method for analyzing the stochasticity in non-stationary data for the beat-to-beat fluctuations in the heart rates of healthy subjects, as well as those with congestive heart failure. The method analyzes the return time series of the data as a Markov process, and computes the Markov time scale, i.e., the time scale over which the data are a Markov process. We also construct an effective stochastic continuum equation for the return series. We show that the drift and diffusion coefficients, as well as the amplitude of the return time series for healthy subjects are distinct from those with CHF. Thus, the method may potentially provide a diagnostic tool for distinguishing healthy subjects from those with congestive heart failure, as it can distinguish small differences between the data for the two classes of subjects in terms of well-defined and physically-motivated quantities.

Key words: non-stationary time series, heart interbeat, Fokker-Planck equation, drift and diffusion

PACS: 05.10.Gg, 05.40.-a,05.45.Tp, 87.19.Hh

Introduction

Cardiac interbeat intervals fluctuate in a complex manner [1–7]. Recent studies reveal that under normal conditions, beat-to-beat fluctuations in the heart rate may display extended correlations of the type typically exhibited by dynamical systems far from equilibrium. It has been shown [2], for example, that the various stages of sleep may be characterized by long-range correlations in heart rate, separated by a large number of beats.

The analysis of the interbeat fluctuations in heart rate belongs to a much broader class of many natural, as well as man-made, phenomena that are characterized by a degree of stochasticity. Turbulent flows, fluctuations in the stock market prices, seismic recordings, internet traffic, pressure fluctuations in chemical reactors, and
the surface roughness of many materials and rock [8], are but a few examples of such phenomena and systems. A long standing problem has been the development of an effective reconstruction method for such phenomena. That is, given a set of data for certain characteristics of such phenomena (for example, the interbeat fluctuations in the heart rate), one would like to develop an effective equation that can reproduce the data with an accuracy comparable to the measured data. Although many methods have been suggested in the past, and considerable progress has been made, the problem remains, to a large extent, unsolved.

In many cases the stochastic process to be analyzed is non-stationary. If the process also exhibits extended correlations, then deducing its statistical properties by the standard methods of analyzing such processes is very difficult. One approach to analyze such processes was proposed by Stanley and co-workers [1, 3, 5, 20–24] and others [25-29]. They studied data for heart-rate fluctuations, for both healthy subjects and those with congestive heart failure (CHF), in terms of self-affine fractal distributions, such as fractional Brownian motion (FBM). FBM is a non-stationary stochastic process which induces long-range correlations, the successive increments of which are, however, stationary, and follow a Gaussian distribution. The power spectrum of an FBM distribution is given by, \( S(f) \propto f^{-(2H+1)} \), where \( H \) is the Hurst exponent that characterizes the type of the correlations that the data contain. Thus, one may distinguish healthy subjects from those with CHF in terms of the numerical value of \( H \) associated with the data: negative or antipersistent correlations for \( H < 1/2 \), as opposed to positive or persistent correlations for \( H > 1/2 \). The analysis of Stanley and co-workers indicated that there may indeed be long-range correlations in heart-rate fluctuation data that can be characterized by FBM and similar fractal distributions. In addition, the data for healthy subjects seem to be characterized by \( H < 1/2 \), whereas those with CHF by \( H > 1/2 \). This was a significant discovery over the traditional methods of analyzing non-stationary data for heart-rate fluctuations.

However, values of the Hurst exponent \( H \) associated with the two groups of subjects are non-universal. Thus, it would, for example, be difficult to distinguish the two groups of subjects if their associated Hurst exponents are both close to 1/2. In addition, the FBM is a non-self-averaging distribution, i.e., given a fixed Hurst exponent \( H \), each realization of a FBM may be significantly different from its other realizations with the same \( H \). As a result, estimating \( H \) alone and characterizing the data by a FBM cannot enable one to predict the future trends of the data. One may also analyze such data by the deterrend fluctuation analysis [2–5] which, in many cases, is capable of yielding accurate and insightful information about the nature of the data.

Recently, a novel method of analyzing stochastic processes was introduced [9–12]. It was shown that by analyzing stochastic phenomena as Markov processes and computing their Markov time (or length) scale (that is, the time scale over which the process can be thought of as Markov), one may reconstruct the original process with similar statistical properties by constructing an effective equation that governs the
process. The constructed equation helps one to understand the nature and properties of the stochastic process. The method utilizes a set of experimental data for a phenomenon which contains a degree of stochasticity, and constructs a simple equation that governs the phenomenon [9–16]. The method is quite general; it is capable of providing a rational explanation for complex features of the phenomenon in question. More significantly, it requires no scaling feature.

In this paper we describe a method for analyzing non-stationary data, and then utilize it to study the interbeat fluctuations of the heart rate. We show that the application of the method to the analysis of interbeat fluctuations of the heart rate may potentially lead to a novel method for distinguishing healthy subjects from those with CHF.

The plan of this paper is as follows. In the next section, we describe the method. We then utilize the method to analyze data for heart-rate fluctuations in human subjects.

Markov Analysis of Non-Stationary Data

Given a (discrete) non-stationary time series $r_i$, we introduce a quantity $x_i$, called the return of $r_i$, defined by

$$x_i = \ln(r_{i+1}/r_i),$$

where $r_i$ is the value of the stochastic quantity at step $i$. If there are long-range positive correlations in the series, then $r_i$ and $r_{i+1}$ are close in values and, therefore, we expect the series $x_i$ to have very small values for all $i$. For white noise, as well as data that exhibit negative or anti-correlations, $r_i$ and $r_{i+1}$ can be completely different and, therefore, the time series $x_i$ will fluctuate strongly.

Figures 1 and 2 present typical data $r_i$ and the corresponding returns $x_i$ for healthy subjects and those with CHF. The number of data is of the order of 30,000-40,000, taken over a period of about 6 h. It is evident that the return series for the subjects with CHF has small amplitude, implying that the $r_i$ data set has long-range positive correlations, which is consistent with the previous analysis [1]. It can be verified straightforwardly that the series $x_i$ is stationary, by measuring the stability of its average and variance in a moving window (that is, over a period of time which varies over the length of the series). Due to the stationarity of the series $x(t)$, we can construct an effective stochastic equation for the returns series of the two groups of subjects, and distinguish the data for healthy subjects from those with CHF. The procedure to do so involves two key steps:

1. Computing the Markov time scale (MTS) $t_M$ constitutes the first step. $t_M$ is the minimum time interval over which the data can be considered as a Markov process [9–12, 17]. As is well-known, a given stochastic process with a degree
of randomness may have a finite or even an infinite $t_M$. To estimate the MTS $t_M$, we note that a complete characterization of the statistical properties of stochastic fluctuations of a quantity $x(t)$ in terms of a parameter $t$ requires the evaluation of the joint probability distribution function (PDF) $P_n(x_1, t_1; \ldots; x_n, t_n)$ for an arbitrary $n$. 

Figure 1. Interbeat fluctuations of healthy subjects (top), and its return time series (bottom).

Figure 2. Interbeat fluctuations of subjects with congestive heart failure (top), and its return time series (bottom).
the number of the data points. If a stochastic phenomenon is a Markov process, an important simplification can be made as \( P_n \), the \( n \)-point joint PDF, is generated by the product of the conditional probabilities, \( p(x_{i+1}, t_{i+1} | x_i, t_i) \), for \( i = 1, \ldots, n-1 \).

The simplest way to determine \( t_M \) for stationary data is by using the least-squares test. The rigorous mathematical definition of a Markov process is given [18] by

\[
P(x_k, t_k | x_{k-1}, t_{k-1}; \ldots; x_1, t_1; x_0, t_0) = P(x_k, t_k | x_{k-1}, t_{k-1}).
\]  

Intuitively, the physical interpretation of a Markov process is that it “forgets its past.” In other words, only the closest “event” to \( x_k \), say \( x_{k-1} \) at time \( t_{k-1} \), is relevant to the probability of the event \( x_k \) at \( t_k \). Hence, the ability for predicting the event \( x_k \) is not enhanced by knowing its values in steps prior to the most recent one. Therefore, an important simplification made for a Markov process is that the conditional multivariate joint PDF is written in terms of the products of simple two parameter conditional PDFs [18] as

\[
P(x_k, t_k; x_{k-1}, t_{k-1}; \ldots; x_1, t_1 | x_0, t_0) = \prod_{i=1}^{k} P(x_i, t_i | x_{i-1}, t_{i-1}).
\]  

Testing Equation (3) for large values of \( k \) is beyond current computational capabilities. For \( k = 3 \) (three points or events), however, the working equation, given by,

\[
P(x_3, t_3; x_2, t_2; x_1, t_1) = P(x_3, t_3 | x_2, t_2),
\]  

should hold for any value of \( t_2 \) in the interval \( t_1 < t_2 < t_3 \). A process is then Markovian if Equation (4) is satisfied for a certain time separation \( t_3 - t_2 \), in which case, \( t_M = t_3 - t_2 \). Thus, to compute the \( t_M \) we use a fundamental theory of probability according to which we write any three-point PDF in terms of the products of the conditional probability functions as,

\[
P(x_3, t_3; x_2, t_2; x_1, t_1), = P(x_3, t_3 | x_2, t_2; x_1, t_1)P(x_2, t_2; x_1, t_1).
\]  

Using the properties of Markov processes to substitute Equation (5), we obtain,

\[
P_{\text{Markov}}(x_3, t_3; x_2, t_2; x_1, t_1) = P(x_3, t_3 | x_2, t_2)P(x_2, t_2; x_1, t_1).
\]  

We then compare the deviation of \( P_{\text{Markov}} \) from that given by Equation (5). Using the least squares method [10], we write:

\[
\chi^2 = \int d\chi_3 d\chi_2 d\chi_1 \times \frac{[P(x_3, t_3; x_2, t_2; x_1, t_1) - P_{\text{Markov}}(x_3, t_3; x_2, t_2; x_1, t_1)]^2}{\sigma^2 + \sigma_{\text{Markov}}^2}.
\]
where $\sigma^2$ and $\sigma_{\text{Markov}}^2$ are the corresponding variances of terms in the numerator. Thus, one should plot the reduced chi-square, $\chi^2_v = \chi^2 / N$ (with $N$ being the number of degrees of freedom), as a function of the time scale $t_3 - t_2$. Then, $t_M = t_3 - t_2$ for that value of $t_3 - t_2$ for which $\chi^2_v$ either achieves a minimum or becomes flat and does not change anymore; see Figure 3.

On the other hand, a necessary condition for a stochastic phenomenon to be a Markov process is that the Chapman-Kolmogorov (CK) equation,

$$P(x_3, t_3 \mid x_1, t_1) = \int dx_2 P(x_3, t_3 \mid x_2, t_2) P(x_2, t_2 \mid x_1, t_1),$$  \tag{8}$$

should hold for the time separation $t_3 - t_2$, in which case, $t_M = t_3 - t_2$. Therefore, to test whether the time series $x(t)$ is a Markov process, one should check the validity of the CK equation for describing the process using different $x_1$ by comparing the directly-evaluated conditional probability distributions $P(x_3, t_3 \mid x_1, t_1)$ with the one calculated according to right side of Equation (8).

(2) Estimation of the Kramers-Moyal coefficients is the second step in constructing an effective equation for describing the series $x_i$. The CK equation is an evolution equation for the distribution function $P(x, t)$ at any time $t$. When formulated in differential form, the CK equation yields the Kramers-Moyal (KM) expansion [18], given by,

$$\frac{\partial}{\partial t} P(x, t) = \sum_{n=1}^{\infty} \left( -\frac{\partial}{\partial x} \right)^n \left[ D^{(n)}(x) P(x, t) \right].$$  \tag{9}$$
The coefficients $D^{(n)}(x, t)$ are called the KM coefficients. They are estimated directly from the data, the conditional probability distributions, and the moments $M^{(n)}$ defined by,

$$M^{(n)} = \frac{1}{\Delta t} \int d x'(x' - x)^n P(x', t + \Delta t \mid x, t),$$

$$D^{(n)}(x, t) = \frac{1}{n!} \lim_{\Delta t \to 0} M^{(n)}.$$ (10)

According to the Pawula’s theorem, for a process with $D^{(4)} \sim 0$ all the $D^{(n)}$ with $n \geq 3$ vanish, in which case the KM expansion reduces to the Fokker-Planck equation, also known as the Kolmogorov Equation (18):

$$\frac{\partial}{\partial t} P(x, t) = \left[ -\frac{\partial}{\partial x} D^{(1)}(x, t) + \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \right] P(x, t).$$ (11)

Here $D^{(1)}(x, t)$ is the drift coefficient, representing the deterministic part of the process, and $D^{(2)}(x, t)$ is the diffusion coefficient that represents the stochastic part.

We now apply the above method to the fluctuations in the human heartbeats of both healthy subjects and those with CHF. As mentioned in the Introduction, several studies [5, 6, 10–12, 19–21] indicate that, under normal conditions, the beat-to-beat fluctuations in the heart rate may display extended correlations of the type typically exhibited by dynamical systems far from equilibrium, and that the two groups of subjects may be distinguished from one another by a Hurst exponent. We show that the drift and diffusion coefficients (as defined above) of the interbeat fluctuations of healthy subjects and patients with CHF have distinct behavior, when analyzed by the method we propose in this paper, hence enabling one to distinguish the two groups of the subjects.

We analyzed both daytime (12:00 pm to 18:00 pm) and nighttime (12:00 am to 6:00 am) heartbeat time series of healthy subjects, and the daytime records of patients with CHF. Our database includes 10 healthy subjects (7 females and 3 males with ages between 20 and 50, and an average age of 34.3 years), and 12 subjects with CHF (3 females and 9 males with ages between 22 and 71, and an average age of 60.8 years). Figures 1 and 2 present the data. We first estimate the Markov time scale $t_M$ for the return series of the interbeat fluctuations, using the chi-square method described above. In Figure 3 the results for the $\chi^2$ values for a subject with CHF are shown. For the healthy subjects we find the average $t_M$ for the returns, for both the day- and nighttime data, to be (all the values are measured in units of the average time scale for the beat-to-beat times of each subject), $t_M = 10$. On the other hand, for the daytime records of the patients with CHF, the estimated average $t_M$ is $t_M = 20$. Therefore, the data for the healthy subjects are characterized by $t_M$ values that are smaller than that of the patients with CHF by a significant factor of 2.
Figure 4. Test of Chapman-Kolmogorov equation for the time separation between \( t_3 \) and \( t_1 \) equal to the Markov time scale, for \( x_1 = -6 \times 10^{-2}, x_1 = 0, \) and \( x_1 = 6 \times 10^{-2} \). Squares and triangles represent, respectively, the directly-evaluated PDF and that computed according to the right side of Equation (8). For clarity, the PDFs are shifted in the vertical directions.

We then check the validity of the CK equation for several \( x_1 \) triplets by comparing the directly-evaluated conditional probability distributions \( P(x_3, t_3 | x_1, t_1) \) with the ones calculated according to right side of Equation (8). Here, \( x \) represents the returns. In Figure 4, the two differently-computed PDFs are compared. Assuming the statistical errors to be the square root of the number of events in each bin, we find that the two PDFs are statistically identical.

Using Equation (10) directly we calculate the drift and diffusion coefficients, \( D^{(1)}(x) \) and \( D^{(2)}(x) \), for the entire set of data for the healthy subjects, as well as those with CHF. The corresponding \( D^{(1)}(x) \) and \( D^{(2)}(x) \) are displayed in Figure 5. We find that these coefficients provide another important indicator for distinguishing the ill from the healthy subjects: The drift \( D^{(1)} \) and the diffusion coefficients \( D^{(2)}(x) \) follow, respectively, linear and quadratic equations in \( x \) with distinct coefficients for the healthy subjects and patients with CHF. The analysis of the data yields the following estimate for the healthy subjects (averaged over the samples),

\[
D^{(1)}(x) = -0.1x, \quad D^{(2)}(x) = 3.7 \times 10^{-5} - 6.6 \times 10^{-5}x + 0.06x^2, \tag{12}
\]

with \(-0.15 < x < 0.15\), whereas for the patients with CHF we find that,

\[
D^{(1)}(x) = -0.06x, \quad D^{(2)}(x) = 8.6 \times 10^{-6} - 2.7 \times 10^{-5}x + 0.03x^2, \tag{13}
\]

with \(-0.04 < x < 0.04\).
We find two important differences between the heartbeat dynamics of the two classes of subjects:

1. Compared with the healthy subjects, the drift and diffusion coefficients for the patients with CHF are small.

2. The fluctuations of the return time series for healthy subjects are distinct from those with CHF. They also fluctuate over different intervals, indicating that the return data for the healthy subjects fluctuate over a large interval. The fluctuation intervals are $-0.04 < x < 0.04$ and $-0.15 < x < 0.15$ for patients with CHF and healthy subjects, respectively. Hence, we suggest that one may use the drift and diffusion coefficient magnitudes, as well as the fluctuation intervals for the
returns, for characterizing the dynamics of human heartbeats, and to distinguish healthy subjects from those with CHF.

**Discussion**

Lin [30] argued that the daytime heart rate variability of healthy subjects may exhibit *discrete* scale-invariance (DSI). A stochastic process $x(t)$ possesses *continuous* scale-invariant symmetry if its distribution is preserved under a change of variables, $t \to \lambda t$ and $x \to x/\mu$, where $\lambda$ and $\mu$ are real numbers, so that,

$$x(t) = \frac{1}{\mu}x(\lambda t).$$

(14)

If Equation (14) holds only for a countable (discrete) set of values of $\lambda$, $x(t)$ is said to possess DSI, which implies a power-law behavior for $x(t)$ that has a log-periodic correction of frequency $1/\log \lambda$, so that

$$x(t) = t^{\gamma} F(\log t/\log \lambda),$$

(15)

with $\gamma = \log \mu/\log \lambda$, and with $F(x) = F(x+1)$ being a period scaling function. Generally speaking, one may write $x(t) = c(t)t^{\zeta}$, with $\zeta = \gamma + 2n\pi i/\log \lambda$, $n = 1, 2, \ldots$. The existence of log-periodicity was first suggested by Novikov [31] in small-scale energy cascade of turbulent flows. It has been argued [32] that log-periodicity may exist in the dynamics of stock market crashes [33], turbulence [34], earthquakes [35], diffusion in disordered materials [36, 37], and in fracture of materials near the macroscopic fracture point [38]. The log-periodicity, if it exists in the heart rate variability (HRV), implies the existence of a cascade for the multifractal spectrum of HRV, previously reported by others. However, Lin’s method neither provides a technique for distinguishing the HRV of healthy people from those with CHF, nor can it predict the future behavior of HRV based on some data at earlier times.

The method proposed in the present paper is different from such analyses in that, the returns for the data are analyzed in terms of Markov processes. Our analysis does indicate the existence of correlations in the return time series which can be quite extended (and is characterized by the value of the Markov time scale $t_M$).

**Summary**

We distinguish the healthy subjects from those with CHF in terms of the *differences* between the drift and diffusion coefficients of the Fokker-Plank equations that we construct for the return data which, in our view, provide a clearer and more physical way of understanding the differences between the two groups of the subjects. In addition, the reconstruction method suggested in this paper enables one to predict
the future trends in the returns (and, hence, in the original series $r_t$) over time scales that are of the order of the Markov time scale $t_M$.

None of the previous approaches for analyzing the data could provide such a reconstruction method.

We also believe that the computational method that is described in this paper is more sensitive to small differences between the data for healthy subjects and those with CHF. As such, it might eventually provide a diagnostic tool for detection of CHF in patients with small amounts of data and in its initial stages of development.

References