Time series analysis of data from stress ECG

Camillo Cammarota

Dipartimento di Matematica
'La Sapienza' Università di Roma, Italy
cammar@mat.uniroma1.it

Abstract

The heartbeat time series of the electrocardiogram recorded during stress test is non stationary, showing both decreasing and increasing trends and time variability of the variance. The analysis of the extrema is used to investigate the durations of accelerations and decelerations of the residuals obtained subtracting the trend. The time mean of these durations is used as a statistic to test the hypothesis that the residuals are independent and identically distributed (i.i.d.) variables. In this hypothesis the expectation of the statistic is 3/2; the rejection region of the test is computed by numerical simulation. Data analysis performed over the heartbeat series of 14 healthy subjects shows that the mean is significantly greater 3/2 and different in stress and recovery.

Keywords: `time series, extrema, stress ECG.`

1. Introduction

The electrocardiogram (ECG) is the recording on the body surface of the electrical activity generated by the heart. In 1903 Einthoven introduced a method of recording the signal (the magnitude is of the order of mV) the labeling of the various waves and investigated a variety of cardiac abnormalities. The main peak is called R wave, corresponding to the contraction of ventricles (systole) (Figure 1). The RR interval is the time between two consecutive R peaks and it is inversely proportional to the instantaneous heart rate. The RR time series shows a variability reminiscent of complex systems, stochastic or deterministic chaotic, that is known as Heart Rate Variability (HRV). The HRV is mainly due to the autonomic control, to respiratory interaction,to humoral regulation. Spectral analysis [6] is the main tool used to investigate the control of the autonomic system on the heart rate. It is believed that the parasympathetic control is related to high frequency (HR) spectral components of RR sequence (0.15 Hz - 0.4 Hz) and the sympathetic one is related to low frequency (LF) components (0.04 Hz - 0.15 Hz). These results are supported by experiments in humans performed with pharmacological blockade or stimulation such as the tilt test.

It is also conjectured that sympathetic stimulation produces accelera-

tion of heart rate, and parasympathetic one produces deceleration. It is known that these controls act on the scale of a few seconds, reflecting in the dependence of a small number of consecutive elements of the RR time series. In order to investigate this dependence the RR series is coded into a sequence of binary symbols and the frequency of suitable short words is computed (symbolic analysis). This method provides a characterization of the autonomic control similar to the one of spectral analysis in ECG data recorded during tilt test [2]. Binary words of fixed length can be used to code periods of acceleration and deceleration in RR series of 24 hours (Holter recording). Data analysis of normal subjects shows a significant difference between positive and negative accelerations [3], suggesting that sympathetic and parasympathetic controls are not symmetric.
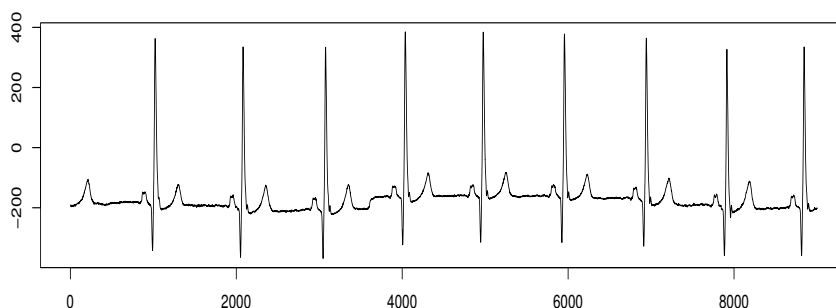

Fig. 1. The electrocardiogram of a normal subject; the sharp peaks (R) correspond to systole.

The ECG stress test is performed to evaluate the presence of myocardial ischaemia. In normal subjects during effort myocardial blood flow increases three or four times compared to rest, as a consequence of increased oxygen request. The RR series extracted from the ECG recorded during a stress test shows: a decreasing trend (stress phase), a global minimum (acme) and an increasing trend (recovery phase) (Figure 2, top panel). It is commonly believed that during stress, in which heart rate is increasing, the heart is prevalently under influence of the sympathetic branch and during recovery, in which heart rate is decreasing, the heart is prevalently under influence of the parasympathetic branch. The recording of RR series during ambulatory ECG stress monitoring is not included in the standard protocols. The analysis of this type of series is new in the HRV literature [4], and can provide new insights into the neuroautonomic control of heart rate. In that paper

2

the "analysis of extrema" was proposed as a method of investigation of non stationary series. The sequence of the lengths of monotonicity intervals was extracted, its properties were investigated and the time mean of the lengths was used as a statistic in the data analysis. This method was applied to the RR series of 14 healthy subjects, providing a clear evidence of the control.

In the present contribution we improve the results obtained in [4] in three directions. First, in the analysis of extrema, after a summary of known results, we give a new one concerning the variance of the statistic. Second we find the rejection region of a test of serial independence, simulating series of independent and identically distributed (i.i.d.) random variables (r.v.). Third we improve data analysis, performing the analysis of the variability after detrending the series, in order to compare the results previously obtained without detrending.
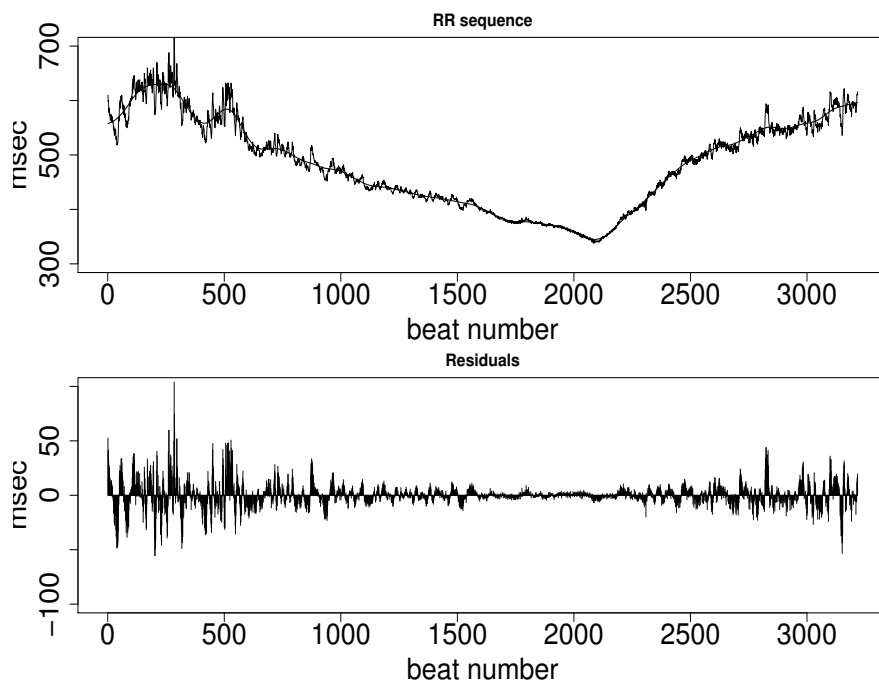
Fig. 2. First panel: RR series extracted from the ECG recorded during stress test of a normal subject. Second panel: Residuals of the RR series after detrending.

2. Analysis of extrema

The analysis of extrema is based on the idea that a relevant information for a time series is contained in the extrema (local maxima and minima). This idea is also used in the " Empirical Mode Decomposition" (EMD) [5],

3

a method of analysis of continuous non stationary signals. Mathematical properties of the extrema of the intrinsic mode functions of EMD are non trivial; in the case of white noise some results have been obtained by using numerical simulations.

The analogous of white noise in time series is a sequence of i.i.d. variables. In this case some results on the extrema can be obtained [4]. We summarize them below.

We model the observed time series $x_1, ..., x_n$ as the realization of a sequence of r. v. $X_1, ..., X_n$, with joint continuous distribution P. We consider the subsets of $\mathbb{R}^3$: $A = \{x_1 < x_2, \ x_2 > x_3\}$ corresponding to maximum condition and $B = \{x_1 > x_2, \ x_2 < x_3\}$ corresponding to minimum condition and the event $E = A \cup B$, corresponding to an extreme (or turning point).

For the sequence $X'_i = (X_{i-1}, X_i, X_{i+1}) \in \mathbb{R}^3$, $i = 2, ..., n - 1$, we consider the sequence of r.v. T_i, $i = 1, ...$ defined as the occurrence times of the event E and the sequence of r.v. defined as the recurrence times

$$H_i = T_{i+1} - T_i, \quad i = 1, ...$$

We refer to the H_i as to the "lengths of monotonicity intervals". We focus on the properties of the H_i's under the assumption that the variables X_i are i.i.d. with continuous distribution. In these assumptions the sequence X'_i, $i \geq 2$ is also stationary. Hence the H_i's form a stationary sequence under the conditional probability $P(\ \ | \ X'_2 \in E)$ [1].

The first result concerns the conditional mean of the H_i's.

Theorem 2.1. *For the i.i.d. sequence $X_1, X_2, ...$ of continuous variables, the variable "length of the monotonicity interval" has a conditional mean*

$$(1) \qquad\qquad E(H_1 \mid X'_2 \in E) = \frac{1}{P(X'_2 \in E)} = \frac{3}{2}$$

A second result is the conditional distribution of H_1, which despite its simplicity was not known.

Theorem 2.2. *For the i.i.d. sequence $X_1, X_2, ...$ of continuous variables, the variable "length of the monotonicity interval" has a discrete density*

$$(2) \quad P(H_1 = s \mid X'_2 \in E) = 3 \left[\frac{1}{(s+1)!} - 2\frac{1}{(s+2)!} + \frac{1}{(s+3)!} \right], \quad s \geq 1$$

We now give a new result concerning the conditional variance of the length of the monotonicity intervals.

Theorem 2.3. *For the i.i.d. sequence $X_1, X_2, ...$ of continuous variables, the variable "length of the monotonicity interval" has a conditional variance*

$$(3) \qquad \mathrm{Var}(H_1 \mid X_2' \in E) = 3\left[2e - \frac{63}{12}\right]$$

Proof.

We compute the variance using in short notations

$$Var(H) = E(H^2) - E(H)^2$$

By the distribution 2.2 we have

$$E(H^2) = 3\sum_{s=1}^{\infty} s^2 \left[\frac{1}{(s+1)!} - 2\frac{1}{(s+2)!} + \frac{1}{(s+3)!}\right]$$

We compute separately the three contributions. After some easy calculations one gets

$$\frac{s^2}{(s+1)!} = \frac{(s+1-1)^2}{(s+1)!} = \frac{1}{(s-1)!} - \frac{1}{s!} + \frac{1}{(s+1)!}$$

$$\frac{s^2}{(s+2)!} = \frac{(s+2-2)^2}{(s+2)!} = \frac{1}{s!} - \frac{3}{(s+1)!} + \frac{4}{(s+2)!}$$

$$\frac{s^2}{(s+3)!} = \frac{(s+3-3)^2}{(s+3)!} = \frac{1}{(s+1)!} - \frac{5}{(s+2)!} + \frac{9}{(s+3)!}$$

Collecting the three summands we get

$$s^2 \left[\frac{1}{(s+1)!} - 2\frac{1}{(s+2)!} + \frac{1}{(s+3)!}\right] =$$

$$= \frac{1}{(s-1)!} - \frac{3}{s!} + \frac{8}{(s+1)!} - \frac{13}{(s+2)!} + \frac{9}{(s+3)!}$$

This quantity can be decomposed into 4 telescopic series and a residual term:

$$\frac{1}{(s-1)!} - \frac{1}{s!} - \frac{2}{s!} + \frac{2}{(s+1)!} + \frac{6}{(s+1)!} - \frac{6}{(s+2)!} - \frac{7}{(s+2)!} + \frac{7}{(s+3)!} + \frac{2}{(s+3)!}$$

5

The sums of the four series take the values 1, -2, 3, -7/6 and the residual is

$$2 \sum_{s=1}^{\infty} \frac{1}{(s+3)!} = 2(e - \frac{8}{3})$$

Finally we have obtained

$$E(H^2) = 3 \left(2e + 2 - \frac{39}{6} \right)$$

Using that $E(H) = 3/2$ we get the result. $\qquad\square$

3. A test of serial independence

For a time series $X_1, X_2, \ldots, X_n$ we consider the lengths of the monotonicity intervals $H_1, H_2, \ldots, H_N$, and the statistic

$$D_n = \frac{1}{N} \sum_{j=1}^{N} H_j$$

i.e. the time mean of the variables $H_1, H_2, \ldots, H_N$. Note that N is random and depends on n. Given n, the distribution of D_n does not depend on the distribution of the X_i if these are independent. By using the ergodic theorem one can see that the statistic D_n converges almost everywhere:

$$\lim_{n \to \infty} D_n = \frac{3}{2}, \quad \text{a.e.}$$

The distribution of D_n depends only on n and can be computed simulating on a computer a large number of i.i.d. sequences. This simulation and the data analysis is performed using the statistical software R [7]. In table 1 two examples of quantiles useful for applications are reported. The lengths of data series $n = 500$ and $n = 1000$ are typical of the present application. These ideas allow to perform a test of serial independence: if the observed value of D_n falls outside an interval centered in 3/2 the i.i.d. hypothesis is rejected. The main interest in this test with respect to other tests is: it does not assume a specific distribution for the data; the statistic has a clear meaning related to acceleration of heart rate.

4. Data analysis

In the data analysis of [4] the analysis of the extrema was performed on the RR time series. Here we perform the same type of analysis on the

Table 1.   Rounded values of the quantiles of the simulate distribution for the statistic D_n, the time mean length of monotonicity intervals, in case of two sequences with n= 500 and n=1000 data.

probability	1%	2.5%	5%	95%	97.5%	99%
quantile n= 500	1.42	1.43	1.44	1.56	1.58	1.59
quantile n=1000	1.43	1.44	1.45	1.55	1.56	1.57

sequence of residuals obtained after detrending the RR series. The trend is estimated using a cubic spline (see Figure 2, first panel) and the residuals are computed (second panel). The residuals are non stationary in variance, but this has a weak effect on the statistic D_n, since the time variation of the variance is very slow if compared to the length of an interval of acceleration that is typically of a few units.

A group of 14 healthy subjects underwent to stress test in the Laboratory of Cardiology. ECG was recorded with the PC-ECG 1200 (Norav Medical Ltd.), which provides output digital signal with an amplitude resolution of 2.441 microV and 500 Hz sampling frequency. The 50 Hz power-line interference and voluntary muscular activity were removed by using a discrete wavelet transform filter. An automated method was used for R peaks detection from the V5 lead. The RR sequence was extracted. All the cases show a profile of the RR series similar to the one of Figure 2(first panel). The filtered series shows a unique global point of minimum (acme); the original series restricted to beat numbers smaller (larger) than acme is called 'stress (recovery)phase'respectively. The analysis of the extrema was applied to the residuals (Figure 2,second panel) for both phases. The results are reported in table 2.

Table 2.   Rounded values of the statistic D_n, mean length of monotonicity intervals of the 14 cases. First column: case number; second column: stress phase; third column: recovery phase.

case	stress	recovery	case	stress	recovery
1	1.80	1.89	8	1.75	2.02
2	1.77	1.81	9	1.92	1.92
3	1.76	2.60	10	2.22	2.26
4	1.94	2.12	11	1.78	2.34
5	2.00	2.04	12	1.68	1.80
6	1.73	2.23	13	1.77	2.03
7	2.13	2.28	14	1.56	1.46

5. Conclusions

Statistical analysis of the data has given the following results.

(i) In all the cases, with exception of case 14, the statistic D_n has values greater than 1.59 (the 0.99 quantile for $n = 500$) both for stress and recovery.

(ii) The values in stress are significantly greater than in recovery, as confirmed by a standard t-test.

Result (i) means that the hypothesis of independence for the residuals has to be rejected. More precisely it suggests that the control system acts prolonging the durations of acceleration and deceleration. In case 14 the hypothesis is not rejected: the residuals can be considered an i.i.d. sequence and this suggests a reduced neuroautonomic control. Therefore the action of the neuroautonomic control reflects not only in the trend of the RR series but also in the residuals.

Result (ii) suggests that during exercise the control of the sympathetic and parasympathetic branch of the neuroautonomic system are of different intensity and that they could be quantified by the durations of acceleration and deceleration in the residuals of the RR series.

Acknowledgments.

REFERENCES

1. L. Breiman , Probability, SIAM, 1992.

2. C. Cammarota , E. Rogora, Spectral and symbolic analysis of heart rate during tilt test, *Phys. Rev. E* **74**,(2006), pp. 042903-1, 042903-4.

3. C. Cammarota , E. Rogora , Time reversal, symbolic series and irreversibility of human heartbeat, *Chaos, Solitons and Fractals* **32** (2007), pp.1649–1654.

4. C. Cammarota, M. Curione, Analysis of extrema of heartbeat time series in exercise test *Mathematical Medicine and Biology* **25** (2008), pp. 87-97.

5. N.E. Huang, Z.Shen ,S.R. Long , M.C.Wu ,H.H. Shih , Q. Zheng , N-C.Yen ,C.C. Tung ,H.H. Liu , The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc.Roy.Soc. London* **A454** (1998), pp. 903–995.

6. A. Malliani , M. Pagani , F. Lombardi , S. Cerutti, Cardiovascular neural regulation explored in the frequency domain, *Circulation* **84** (1991), pp. 482–492.

7. R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.