Non stationary model of the heartbeat time series in atrial fibrillation

Camillo Cammarota[1], Giuseppe Guarini[2], Enrico Rogora[1], and Maria Ambrosini[2]

[1] Department of Mathematics University of Rome La Sapienza
[2] Department of Cardiology Policlinico University of Rome La Sapienza

Abstract. The heartbeat time series of subjects with atrial fibrillation is modeled using a non stationary sequence of random variables. The mean and standard deviation are estimated by using a segmentation of the data. We find that they are linearly related and then we use a lognormal model to analyze the data. The autocorrelation function of the residuals shows to be significantly close to zero.

1 Introduction

The heartbeat time series is the sequence of time intervals between two consecutive R peaks of the ECG, corresponding to the depolarization of the ventriculus(systole). The sequence is extracted from 24 hours Holter ECG and defines a series of about 10^5 terms. Cardiac interbeat intervals usually fluctuate in a complex and apparently erratic manner. These fluctuations can be considered as a superposition of those caused by autonomous dynamics and the those driven by environmental stimuli. Actually a normal heart shows a strong adaptability to external changes. This adaptability relies on two mechanisms. The first one is the influence of the autonomous nervous system on the pacemaker cells of the heart with typical response delay of the order of 1 sec. The second mechanism depends on humoral stimulation for which the response delay is much longer.

Atrial fibrillation is an arrhythmia characterized by rapid randomized contraction of the atrial miocardium causing a totally irregular, often rapid, ventricular rate. In this situation the control of the autonomous nervous system on the RR sequence fades and the heart rhythm should be mainly modulated by long term humoral stimulation.

The fluctuations of the RR sequence are larger than in a normal heart and their correlations appear to be weaker. There is evidence that an associated suitable symbolic dynamics is described by a sequence of independent random variables (Cammarota et al., 2002). Recently it has been found by using multiscale analysis that a suitable notion of complexity, measured as approximate entropy, is smaller for subjects with atrial fibrillation than for normals (Costa et al. 2002).

In this study we analyze the heartbeat time series of subjects with atrial fibrillation. We apply standard tools of analysis of time series, like the autocorrelation function, in order to quantify the correlation between the values

in the series. The main problem is that the heartbeat time series is highly non stationary. Various techniques have been adopted to investigate this kind of series. One can consider the series of the differences and restrict the analysis to a suitable short segment (Cammarota, 2000). In econometrics and finance non stationary time series are modeled as the sum of several components, some deterministic (trend and seasonal variation) and other stochastic (fluctuation). Two types of non stationary behaviors are usually considered: non stationarity in mean, due to the presence of a deterministic level function, and non stationarity in variance. In physical literature the analysis of non stationary series is performed by using the detrended fluctuation analysis (Peng et al. , 1995). In our model we try to use a mixture of the above techniques. We divide the series into short segments where it can be assumed to be approximately stationary. We evaluate mean and standard deviation for each segment and represent these pairs on a scatter plot. The striking fact is that for most of our cases this is approximately linear, whereas for a normal subject this does not happen (see fig 1). This is the main motivation for the choice of our model: a sequence of lognormal variables with time dependent parameters. In this paper we test the above probabilistic model and use it to extract from the RR sequence an essentially stochastic component of short range memory (range zero or one). Our approach is in some sense complementary to the one in (Hayano et al., 1998) where suitable indexes (in our opinion related to the deterministic component of the sequence) have been shown to display a circadian rhythm. Our results on short range memory (stated at the end of section 4) should be compared to the ones obtained in (Stein et al., 1999), where short range predictability in RR series has also been found in some cases.

2 The model

In our model of fibrillating heart time series we assume that non stationarity in mean and in variance are both present but that they are linearly related. This can be achieved by considering a non stationary sequence of random variables X_i, based on lognormal distribution. We denote by ϵ_i a stationary sequence of standard normal random variables and define

$$X_i = a + e^{\mu_i + \sigma\epsilon_i}$$

The sequence of positive numbers μ_i represents the trend, and takes into account environmental stimuli; the two constants a and σ describe the heart response. The random variables ϵ_i can be assumed to be independent (this is a good approximation in some cases) or to form an AR(1) sequence (see below). We can easily compute the expectation

$$E(X_i) = a + e^{\mu_i} E(e^{\sigma\epsilon_i}) = a + e^{\mu_i + \frac{1}{2}\sigma^2}$$

and the variance

$$Var(X_i) = Var(e^{\mu_i + \sigma\epsilon_i}) = e^{2\mu_i} Var(e^{\sigma\epsilon_i}) = e^{2\mu_i + \sigma^2}(e^{\sigma^2} - 1)$$

Therefore these two quantities are related. Actually

$$Var(X_i) = (E(X_i) - a)^2(e^{\sigma^2} - 1)$$

and then

$$\sqrt{Var(X_i)} = \eta(E(X_i) - a), \quad \eta = \sqrt{(e^{\sigma^2} - 1)}$$

We stress that the peculiar feature of our model is that the local mean and the standard deviation are linearly related, as shown in the last equation.

3 Estimation of parameters

We have analyzed the RR time series from Holter ECG of 28 patients of our Cardiology Department with atrial fibrillation. The trend is estimated by the step function T_k defined below. We consider disjoint intervals of l integers

$$I_k = \{kl, ..., kl + l - 1\}, \quad k = 0, 1, ...$$

and put

$$T_k = \frac{1}{l} \sum_{i \in I_k} X_i$$

In our analysis we have chosen l in the range $(100, 500)$ and the results essentially do not depend on the particular choice. The expectation $E(X_i)$ is estimated by T_k when $i \in I_k$. If we put

$$V_k = \frac{1}{l} \sum_{i \in I_k} (X_i - T_k)^2$$

the variance $Var(X_i)$ is estimated by V_k when $i \in I_k$. We also put $S_k = \sqrt{V}_k$ and consider the scatter plot of the pairs (T_k, S_k). The typical shape of this plot for atrial fibrillation and normal heart are shown in fig.1. It is natural therefore to describe plots for atrial fibrillation by using linear functions. One can estimate the parameters a and η by using a linear regression. We have found that the typical values of η are between 0.2 and 0.3 in a subgroup of 10 cases. From η one can compute σ and μ_i by using the equations

$$\sigma^2 = \log(1 + \eta^2)$$

and

$$\mu_i = \log(E(X_i) - a) - \frac{1}{2}\sigma^2$$

Hence for any $i \in I_k$ the numbers μ_i can be estimated by the quantity

$$\mu_i = \log(T_k - a) - \frac{1}{2}\sigma^2.$$

The variability of the sequence μ_i (or equivalently of T_k) is related to the non-stationary behavior of the RR-series (night and day differences for instance), but it is not the object of our study. Instead we are interested in introducing quantities which are intrinsically related to the heart response, namely a and η. These quantities, by definition, are estimated on the whole 24 hours RR series.

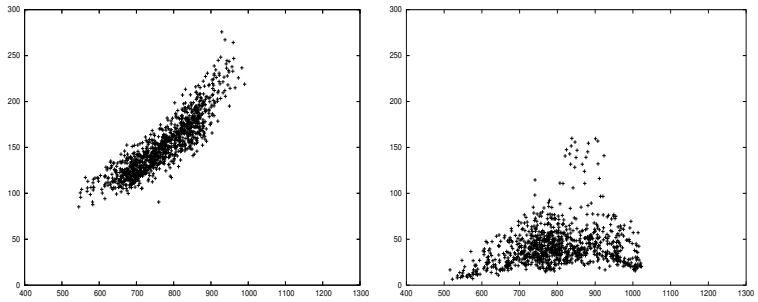

Fig. 1. Scatter plot of the pairs (T_k, S_k) for the interval length $l = 100$ in atrial fibrillation (left figure) and normal heart (right figure).

4 Testing the model

The residuals of our model are defined by

$$\epsilon_i = \frac{\log(X_i - a) - \mu_i}{\sigma}$$

One has to test two things: that the random variables ϵ_i are standard normal and that they are independent.

The histogram of the residuals is much more symmetric than the histogram of the series and at a qualitative level it has a satisfactory normal shape. The autocorrelation function of residuals, supposed to be stationary, is

$$\rho(s) = Cov(\epsilon_0, \epsilon_s)/Var(\epsilon_0)$$

This function can be used to quantify the correlation between the elements of the series. If they were independent this should be equal to 0 for $s \geq 1$. The sample autocorrelation is given by

$$r_s = \sum_{i=0}^{n-s}(\epsilon_i - \overline{\epsilon})(\epsilon_{i+s} - \overline{\epsilon}) / \sum_{i=0}^{n-s}(\epsilon_i - \overline{\epsilon})^2$$

were $\bar{\epsilon}$ is the mean of the series ϵ_i. In the analysis of the residuals it is generally assumed that if their autocorrelation falls outside the range

$$-\frac{2}{\sqrt{N}}, \quad +\frac{2}{\sqrt{N}}$$

where N is the length of the series, the residuals are significantly different from zero. We have found that this function is approximately comprised in this range in many cases. In some of them however it has a significantly non zero value at lag $s = 1$(see fig.2). We improve our model using an AR(1) model for the residuals

$$\epsilon_i = \theta\epsilon_{i-1} + bw_i, \quad w_i \sim N(0,1) \quad \text{i.i.d.}$$

The parameter θ is estimated by r_1, the value of the sample autocorrelation of the sequence ϵ_i at lag 1. The new residuals w_i are now close to zero for any $s \geq 1$ (fig. 2).

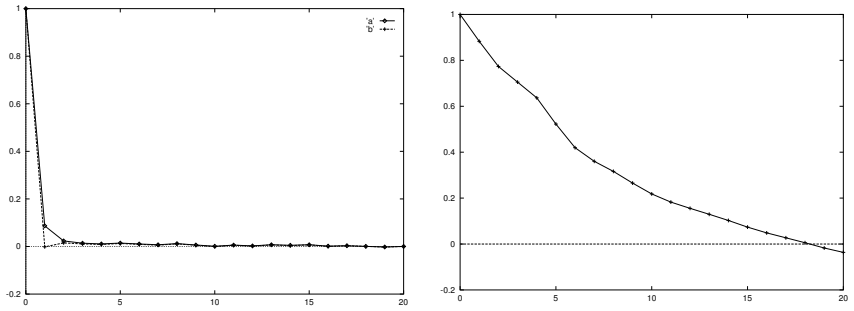

Fig. 2. Left figure: Sample autocorrelation function of the residuals for a) ϵ_i and b) w_i. The lower line (dotted in the figure) represents the autocorrelation of the residuals w_i of the AR(1) model for atrial fibrillation. Right figure: Sample autocorrelation function of the residuals for a normal heart.

5 Conclusion

The main feature of our model is that it essentially describes the stochastic component of fibrillation time series using only two parameters, a and η. These parameters are found by a linear regression of the scatter plot of the mean - standard deviation pairs (T_k, S_k) of the series, where k runs over the segments in which the series has been divided. Actually a linear relationship between mean and standard deviation seems to be a feature of atrial fibrillation, whereas it is not observed in normal heartbeat. The step function T_k, which takes into account the long term external influences, provides useful information when coupled with the function S_k. For some of our series a third

parameter is needed. It is the coefficient of the AR(1) model for the sequence of residuals. This parameter should describe a dependence of range 1 which occurs in some individuals.

Acknowledgments: we thank the referee for useful remarks and bibliographic suggestions.

References

1. M.Costa, A.L.Goldberger, C.K.Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (2002).
2. C. Cammarota, Estimating correlation exponents of the heartbeat time series, Int. Jour. Bif. Chaos, 10, 1513-1520 (2000).
3. C.K.Peng, S.Havlin, H.E. Stanley, A.L.Goldberger, Quantification of scaling exponents and crossover phenomena in non stationary heartbeat time series, Chaos 5, 82-87 (1995).
4. C. Cammarota, G. Guarini, M. Ambrosini, Analysis of stationary periods of heart rate via symbolic dynamics, Third International Symposium on Medical Data Analysis (ISMDA 2002) Rome,October 2002, Lecture Notes in Computer Science 2526, Colosimo, Giuliani, Sirabella Eds., Springer.
5. J. Hayano, S. Sakata, A. Okada, S. Mukai, T. Fujinami, Circadian Rhythms of Atrioventricular Conduction Properties in Chronic Atrial Fibrillation With and Without Heart Failure, J. Am. Coll. Card. 31, 158-166 (1998).
6. K. M. Stein, J. Walden, N. Lippman, B. B. Lerman, Ventricular Response in Atrial Fibrillation: random or deterministic, Am. J. Phisiol., 277, H452-H458 (1999).