International Journal of the Physical Sciences Vol. 6(4), pp. 651-659, 18 February, 2011
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS10.537
ISSN 1992 - 1950 ©2011 Academic Journals

Full Length Research Paper

A bounded random process model and its application in heart rate variability analysis

Ying Wang[1]* and Sheng Lu[2]

[1]Department of Automation, Xiamen University, Fujian, China.
[2]Sterlingtech Software Inc. Rochelle Park, NJ, USA.

Accepted 17 February, 2011

A bounded random process (BRP) model was developed in this paper to interpret why approximate entropy (ApEn) can not describe the stochastic characteristic of Brownian motion time series correctly. Low ApEn value, generally implying presence of determinacy, also existed in Brownian motion series, a stochastic process. The BRP model investigated this phenomenon through quantifying the relationship between ApEn and a parameter of BRP model. BRP model was then applied to analyze electrocardiograph (ECG) time series from 60 healthy subjects and 60 myocardial infarction (MI) patients. ApEn of the healthy group had a close relationship with the parameter of BRP model, while this relationship could not be found in MI patient group. Grounded on combination of BRP and ApEn, a classifier was designed to assist to diagnose the MI patients. ROC curve and classification figures verified the classifier.

Key words: Bounded random process, approximate entropy, Brownian motion, heart rate variability.

INTRODUCTION

Approximate entropy (ApEn), presented by Pincus (1991, 1995), was adopted to analyze the uncertainty or variability of a system. It was frequently applied to biological time series to determine the regularity and complexity of signals (Fusheng et al., 2001; Veldhuis et al., 2001; Pincus and Goldberger, 1994; Xua et al., 2007).

In the present study, we found that low ApEn value did not necessarily imply the presence of deterministic characteristics. For example, Brownian motion time series, although stochastic, had a very low ApEn value, which indicated the determinacy. Thus, we cannot determine the regularity or determinacy of signals only based on ApEn. Goldberger et al. (2002), Xua et al. (2007) also mentioned that ApEn can not be directly applied to study the irregularity of slowly fluctuating curves with broad amplitude, such as electrocardiograph (ECG) signals. So wavelet transformation (Xua et al., 2007) and correlation dimension (D2), Largest Lyapunov exponent (LLE) (Behnia et al., 2008) and other complicated methods had to be adopted to improve the estimation of ApEn value. But basic mechanism behind the inapplicability of ApEn has not been investigated.

So, in this work, we developed a bounded random process (BRP) model to interpret why ApEn can not describe the stochastic characteristic of Brownian motion time series correctly. The small ratio of short-term variability to long-term variability of Brownian time series led to the small value of ApEn, which was confused with deterministic processes. Furthermore, $\log(\varepsilon)$, a parameter of BRP model was combined with ApEn to determine the uncertainty or variability of Brownian motion signals. For Brownian motion and Henon map series, although stochastic and deterministic respectively, they both have low ApEn values. But they had distinct relationships between ApEn and $\log(\varepsilon)$.

Then the BRP model was applied to heart rate variability (HRV) analysis of healthy subjects and myocardial infarction (MI) patients. People tried to

*Corresponding author. E-mail: wangying@xmu.edu.cn. Tel: +86-13850072593. Fax: +86-592-2580258.

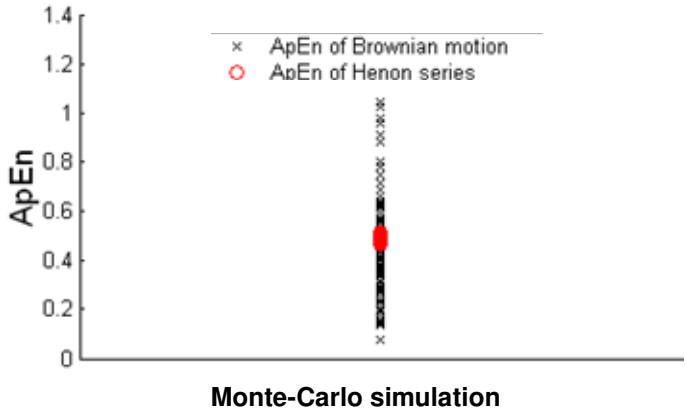


Figure 1. ApEn value for 100 realizations of Brownian motion and Henon series. The X axis were simulation times, the Y axis was the corresponding ApEn value for each Monte-Carlo simulation. All the points having same values in X axis were for the better comparison of the distribution of ApEn values of two time series.

determine the nature of HRV to diagnose heart diseases. Complex features (de Chazal and Celler, 1997b), up to 276 parameters, and multiple neural networks (de Chazal and Celler, 1997a) were adopted to make a classification to assist diagnosis. Complicated dynamic nonlinear techniques were extracted to differentiate between normal and arrhythmia, supraventricular arrhythmia, and congestive heart failure (Nizami et al., 2010). BRP model was applied to analyze HRV for 60 healthy subjects and 60 MI patients with two features and a linear regression function. For the healthy group, there was a close relationship between $\log(\varepsilon)$ and ApEn, which could not be found in MI patient group. A classifier was designed to assist to diagnose the MI patients through linear regression of healthy subjects. The ROC curve and classification figures verified the performance of the classifier.

METHODS

Approximate entropy (ApEn)

ApEn (Pincus, 1991) offered an applicable way to analyze short and noisy time series. A time series containing many repetitive patterns had a relatively small ApEn; A less predictable process had a higher ApEn.

For us to calculate ApEn in the following analysis, we recalled the calculation of ApEn (Pincus, 1995). For a time series $a(n), n = 1, 2, \ldots, N$, vectors with length m, $v(n) = [a(n), a(n+1), \ldots, a(n+m-1)]^T$, were retrieved. $D(i, j)$, the distance between $v(i)$ and $v(j)$, was defined as the maximum difference in the scalar components of $v(i)$ and $v(j)$. If $D(i, j) \le r$, $v(j)$ and $v(i)$ was *similar*. r was a fixed parameter setting the "tolerance" of comparison.

$N^{m,r}(i)$ was the number of vectors $v(j)$ ($j \le N - m + 1$) such that for $v(i)$ ($i \le N-m+1$), $D(i, j) \le r$. $C^{m,r}(i)$ was the probability to find a vector which was *similar* with $v(i)$.

$$C^{m,r}(i) = N^{m,r}(i) / (N-m+1) \tag{1}$$

and

$$F^{m,r} = \frac{\sum_{i=1}^{N-m+1} \log(C^{m,r}(i))}{N-m+1} \tag{2}$$

was the logarithmic average over all the vectors of the $C^{m,r}(i)$ probability.

ApEn was given by:

$$ApEn^{m,r} = F^{m,r} - F^{m+1,r} \tag{3}$$

In this paper, ApEn value was calculated with $m = 2$ and the r equaled 15% of the standard deviation (SD), as suggested by Pincus (1991).

ApEn values in Brownian motion and Hénon map series

ApEn values were calculated for Brownian motion and Hénon map series. The Hénon map was a deterministic discrete-time dynamical system (Hénon, 1976). It was one of the most studied examples of dynamical systems that exhibit chaotic behavior, and the parameters in our work were 1.4 and 0.3. Brownian motion was a stochastic process, while Hénon map was a deterministic series. The two series should have distinct ApEn characteristics if ApEn can catch their uncertainty effectively. Monte-carlo simulation of 100 realizations of Brownian motion series and 100 realizations of Hénon map series were implemented. Brownian motion time series, although stochastic, were found to have some very low ApEn values, just as deterministic signals. The results are shown in Figure 1. There was no statistical significance (0.4230 ± 0.2149 vs. 0.4796 ± 0.0075, $p > 0.05$) between the two processes. Thus, it was impossible to decide determinacy or uncertainty of these two series only by ApEn values.

So, bounded random process (BRP) model was developed to investigate the mechanism behind the phenomenon.

Bounded random process (BRP) model

From the definition of ApEn, we hypothesized that the low ApEn value of Brownian motion series (It can be generated with the integral of white noise and has $1/f^2$ spectrum) might be as a result of the small inter-beat variability comparing with the range time series varies. A bounded random process (BRP) model was constructed to interpret the relationship between ApEn and the ratio of short term variability to the long term variability. The BRP model was defined as following:

$$\begin{aligned} y(i) - y(i-1) &= e(i) \\ Bound^{-} < y < Bound^{+} & \end{aligned} \tag{4}$$

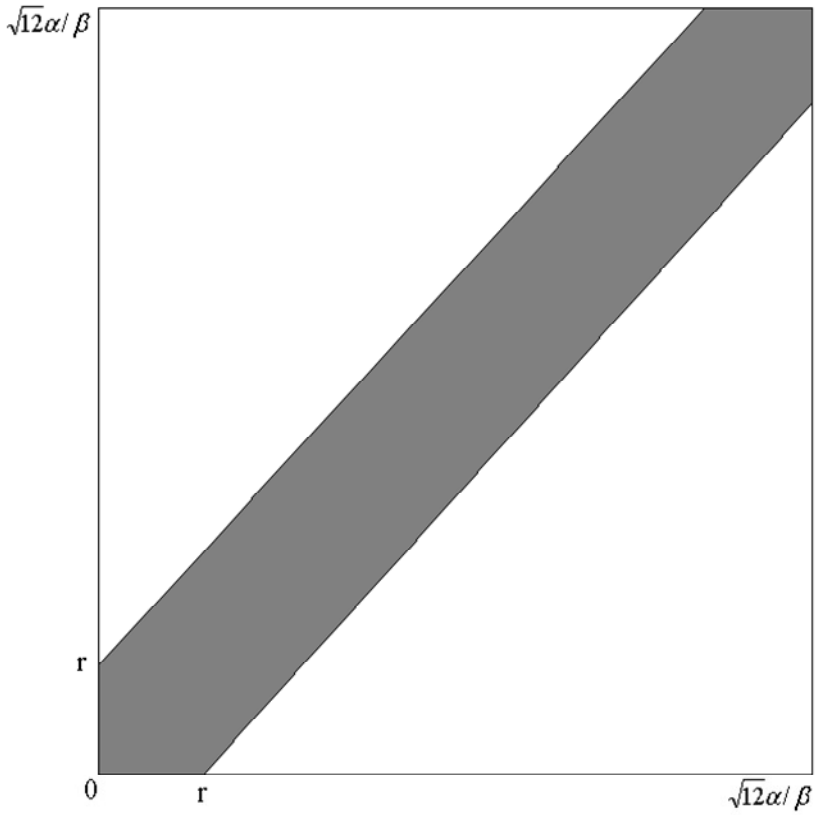


Figure 2. The ApEn representation of the bounded random process.

$e(i)$ described the short-term variability of a time series, where differences between successive points were random processes, with zero mean and α^2 variance, where α was the standard deviation of $e(i)$. The model differed from Brownian motion because of the long-term bound $Bound^- < y < Bound^+$. β stood for the range of signal varies, that is, $\beta = Bound^+ - Bound^-$, and $\varepsilon = \alpha / \beta$. When $y(i)$ exceeded the boundary, new $e(i)$ was produced until $y(i)$ were within the boundary.

Then ApEn was calculated for the time series generated from this BRP model. For simplicity, $e(i)$ was assumed to be uniformly distributed and $\alpha << \beta$. The time series was rescaled into $[0,1]$, and the normalized series was:

$$\begin{cases} y(i+1) - y(i) = e^*(i) \\ 0 < y(i) < 1 \end{cases} \quad (5)$$

where the standard deviation of $e^*(i)$ was $\varepsilon = \alpha / \beta$. The reason we did not use Equation (5) from the beginning was that we wanted to preserve the parameter β and be consistent with the former analysis. The flow to construct y series with BRP model is shown in Appendix A.

In the new time series characterized by Equation (5), we considered two characteristic series with data length m were *similar* to each other. ApEn reflected the probability of the states that $m+1$ of the two series were still within the distance r. Assuming two series as $A(1:m)$ and $B(1:m)$, because of the similarity between A and B, there was:

$$\max \left| A(i) - B(i) \right| < r \quad i = 1...m \quad (6)$$

For the specialty of the BRP model, the state $m+1$ was more related to the state m than earlier states. For simplification, we assumed $A(m) \approx B(m)$, and $A(m+1) - A(m)$ was uniformly distributed. So the range that $A(m+1) - A(m)$ varies was $\sqrt{12}\,\varepsilon = \sqrt{12}\,\alpha / \beta$, so as $B(m+1) - B(m)$. Then if $r < \sqrt{12}\varepsilon$, the probability that $A(m+1)$ and $B(m+1)$ were *similar* was equal to the ratio of the shaded area to the square area in Figure 2 (also Appendix B for the detailed explanation).

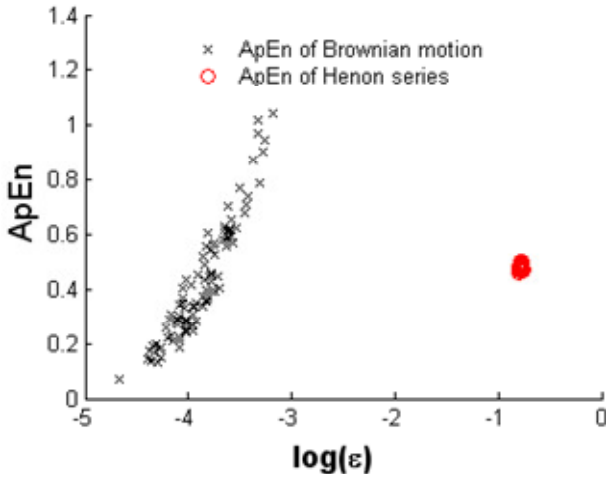


Figure 3. ApEn-$\log(\varepsilon)$ relationship for 100 realizations of Brownian motion and Henon series. X axis were the $\log(\varepsilon)$ values for each realization of Brownian motion; Y axis were the corresponding ApEn value.

$$ApEn \approx \log\left(\frac{12(\alpha/\beta)^2}{12(\alpha/\beta)^2 - (\sqrt{12}\alpha/\beta - r)^2}\right) = -\log\left(\frac{2\sqrt{12}r*\alpha/\beta + r^2}{12(\alpha/\beta)^2}\right) \quad (7)$$

If $r << 2\sqrt{12}\varepsilon$,

$$ApEn \approx -\log\left(\frac{\sqrt{3}r*\alpha/\beta}{3(\alpha/\beta)^2}\right) = -\log\left(\frac{\sqrt{3}r}{3\alpha/\beta}\right) = \log\sqrt{3} + \log(\varepsilon) - \log(r) \quad (8)$$

Equation (8) revealed that ApEn was directly related to both α and β values. When ε was small, that is, $\alpha << \beta$, the ApEn value significantly decreased, which led to the same ApEn character as deterministic signals. That was why Brownian motion and Henon map series can not get distinct ApEn characteristics. In such a case, the deterministic and stochastic signals both had small ApEn values, so it was not possible to discriminate them only based on ApEn.

Then, we tried to investigate the Brownian motion and Henon map series with the combination of ApEn and $\log(\varepsilon)$.

ApEn and $\log(\varepsilon)$ of Brownian motion and Hénon map series

To validate the BRP model, we plotted the Monte-Carlo simulation (ApEn values in Brownian motion and Hénon map series) results with the combination of ApEn and $\log(\varepsilon)$ in Figure 3. We saw the monotonic-increasing linear relationship between ApEn and $\log(\varepsilon)$ for Brownian motion series, which indicated the small ApEn values were caused by the small $\log(\varepsilon)$. While for Hénon map series, there was no such relationship, since it cannot be characterized with BRP model.

Here, although we assumed $e(i)$ was uniformly distributed, $e(i)$ could have other distributions, such as Gaussian distribution. The results of simulation showed that the ApEn had similar characteristics to the uniform distribution shown in Equation (8), and the expression of the equation was more complicated.

APPLICATION IN HEART RATE VARIABILITY (HRV) ANALYSIS

Dataset description

ECGs Data from 60 healthy subjects and 60 MI patients between 17 and 87 years of age were downloaded from PhysioBank(http://www.physionet.org/). The sampling frequency of the ECG data sets was 1 kHz and the resolving power of the data was 16 bits.

ApEn and $\log(\varepsilon)$ analysis for healthy and MI groups

The BRP model was applied to ECG signals to analyze HRV of healthy and MI group. α was approximated by the root mean square of successive differences (RMSSD) between adjacent RR-intervals (the interval from the peak of one QRS complex to the peak of the next) and β was calculated by the range of heart rate varies. To overcome the influence of outlier points, the range β included 98% of the data points. $\varepsilon = \alpha/\beta$, where ε stood for the ratio of short-term variability to long-term variability

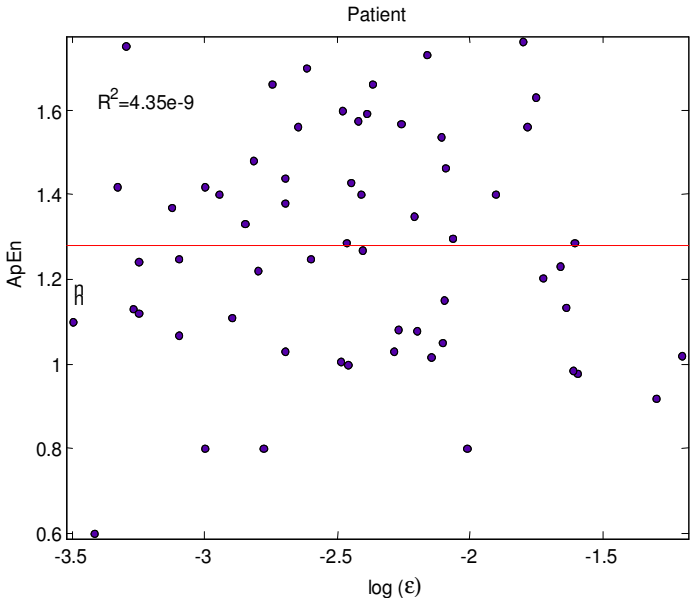


Figure 4. The relationship between log (ε) and ApEn for 60 MI patients. X axis were the $\log(\varepsilon)$ values for each MI patient; Y axis were the corresponding ApEn value.

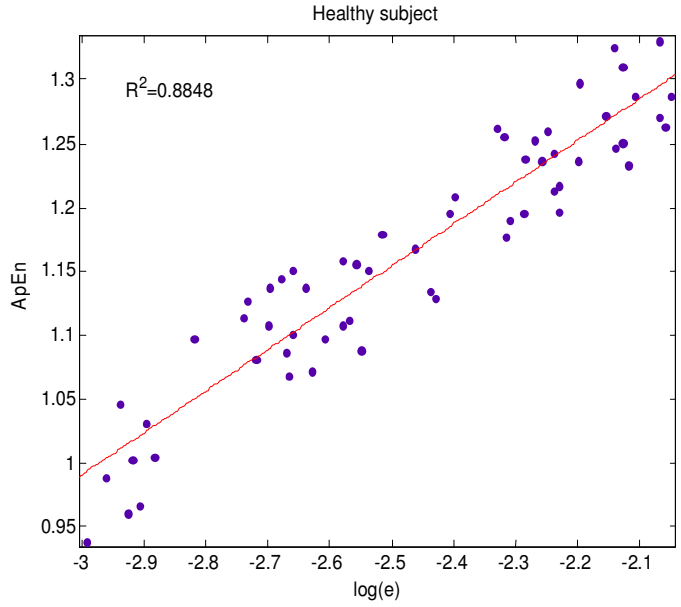


Figure 5. The relationship between $\log(\varepsilon)$ and ApEn for 60 healthy subjects. X axis were the $\log(\varepsilon)$ values for each healthy subject; Y axis were the corresponding ApEn values.

We showed the relationship between $\log(\varepsilon)$ and ApEn of 60 MI patients and 60 healthy subjects in Figures 4 and 5, respectively. For healthy group, the R-square between ApEn and $\log(\varepsilon)$ was 0.8848; while for MI group, the R square was 4.35e-9.

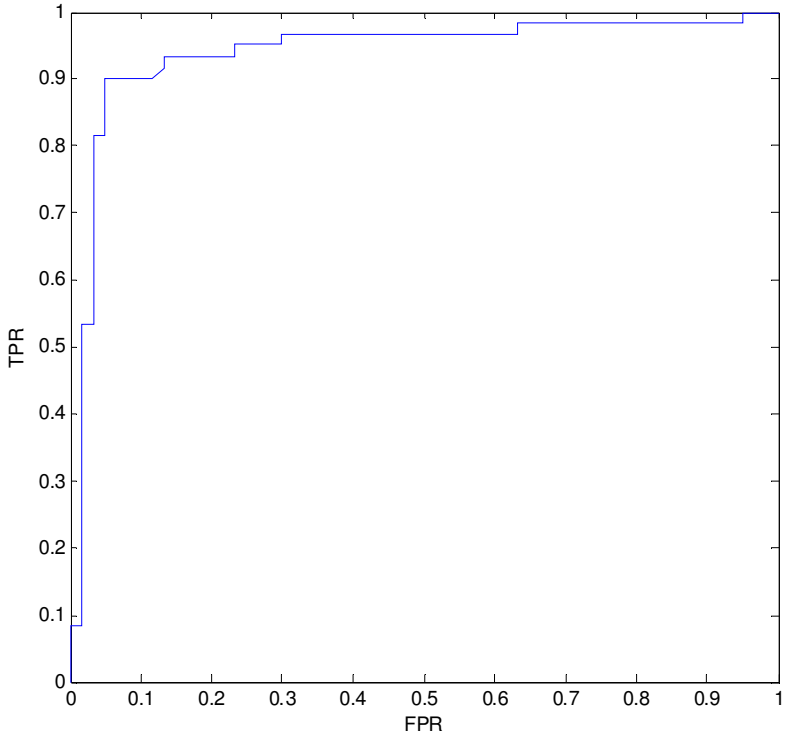


Figure 6. ROC curve for the classification of healthy subjects and MI patients.

So, for healthy people, there was an obvious linear relationship between $\log(\varepsilon)$ and ApEn, and this relationship did not exist in the ECGs of MI patients. Based on the distinct characteristics of MI and health group, a classifier was designed to diagnose MI patients.

A classifier based on ApEn-log(ε)

Based on the different characteristics of healthy subjects and patient group, we extracted the ApEn and $\log(\varepsilon)$ as features to design a classifier, which could assist to diagnose MI patient clinically.

30 subjects were extracted from the 60 healthy subjects as training data to design the classifier. With the linear regression of the 30 training data, we could get a linear function $y = 0.35x + 1.98$. Based on the distance of ApEn-$\log(\varepsilon)$ point to the linear function, each individual was classified to healthy or MI group.

The remained 30 healthy subjects and 60 MI patients were adopted as test data to evaluate the classification. The ROC curve was plotted for the different distance thresholds, as shown in Figure 6. The AUC of the ROC curve was very close to 1, indicating the good performance of the classifier. From ROC curve, the optimal diagnostic point was obtained, that was, the optimal distance threshold in our classifier was 0.107. It meant that for an individual, if the distance between its ApEn-$\log(\varepsilon)$ value and the linear function was no greater than 0.107, it was diagnosed to be healthy; otherwise, it was diagnosed to be MI patient. The classification figure was shown in Figure 7. Points falling into the region between the dot lines were classified into healthy group. The true positive rate was 90% and true negativerate was 95%, respectively. By contrast, if only ApEn values were used for classification, the true positive was 60% and true negative was 65%.

DISCUSSION

In this paper, we developed a bounded random process model to derive the relationship among short-term variability, long-term variability and ApEn value in biological time series. The model was applied to the analysis of HRV in ECGs signals, which can assist to diagnose MI patient subjects based on the ApEn and $\log(\varepsilon)$ relationship. This paper also suggested that the ratio of short-term variability to long-term variability should be considered when we apply ApEn method to effectively reveal the deterministic rhythm.

The relationship between $\log(\varepsilon)$ and ApEn analyzed in the method was consistent with data analysis results in healthy subjects. While in MI patients, this relationship no longer existed. Since BRP model was a statistical time series model, this results revealed that HRV series from healthy subjects can be well characterized by BRP model, which was less deterministic; while HRV series from MI patients were inclined to include deterministic rhythm, such deterministic rhythm may destroy the linear relationship between $\log(\varepsilon)$ and ApEn expressed in Equation (8).

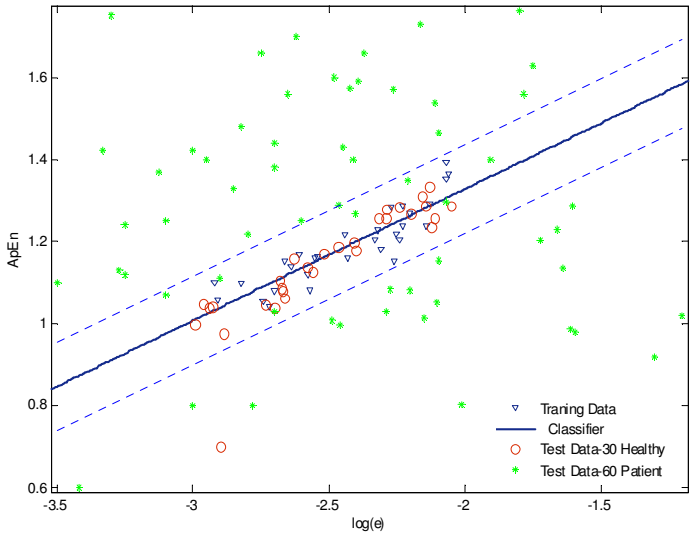


Figure 7. Classification for healthy subjects and patients group under the optimal diagnosis point 0.107.

Table 1. R-square values between ApEn and $\log(\varepsilon)$.

Data length / R-square value	300	400	500	600
Healthy subjects	0.2635	0.4789	0.6796	0.8848
MI patients	3.25e-10	9.01e-10	3.73e-9	4.35e-9

For 60 healthy subject and 60 MI patients, the R-Square value between ApEn and $\log(\varepsilon)$ were calculated for different data length.

The basic mechanism of difference between healthy and patient groups needs further investigation.

Due to the non stationary nature of HRV series, ApEn values were calculated based on different data lengths, which led to different R-Square values between ApEn and $\log(\varepsilon)$, shown in Table 1. We noticed that for HRV series from healthy subjects, data length > 400 was enough to show the significant correlation between ApEn and $\log(\varepsilon)$; while for MI patients, R-square values were low for all different data lengths.

The two parameters used in the BRP model resembled the two time domain parameters in ECGs series, the short-term and the long-term variability or bound measures. The short-term bound, $e(i)$ and long-term bound could limit y from severe fluctuations from one beat to the next as well as limit the heart beating out of the normal condition.

From the analysis in the method, it was also easy for us to qualitatively understand the relationship between ApEn and sample frequency. If the sample frequency was doubled, the ε decreased by half of its amplitude. Assuming variance of new time series was not significantly affected by ε, the ApEn value decreased, which was consistent with the results of paper (Hornero et al., 2005).

ACKNOWLEDGMENT

This work was supported by Natural Science Foundation of Fujian (2009J01304).

REFERENCES

Behnia S, Akhshani A, Mahmodi H, Hobbenagi H (2008). Nonlinear Measure of ECG Time Series: Detection of Cardiac Diseases. Iran. Phys. J., 2: 56-62.

de Chazal P, Celler B (1997a). Improving ECG diagnostic classification by combining multiple neural networks. Comput. Cardiol., 24: 473-476.

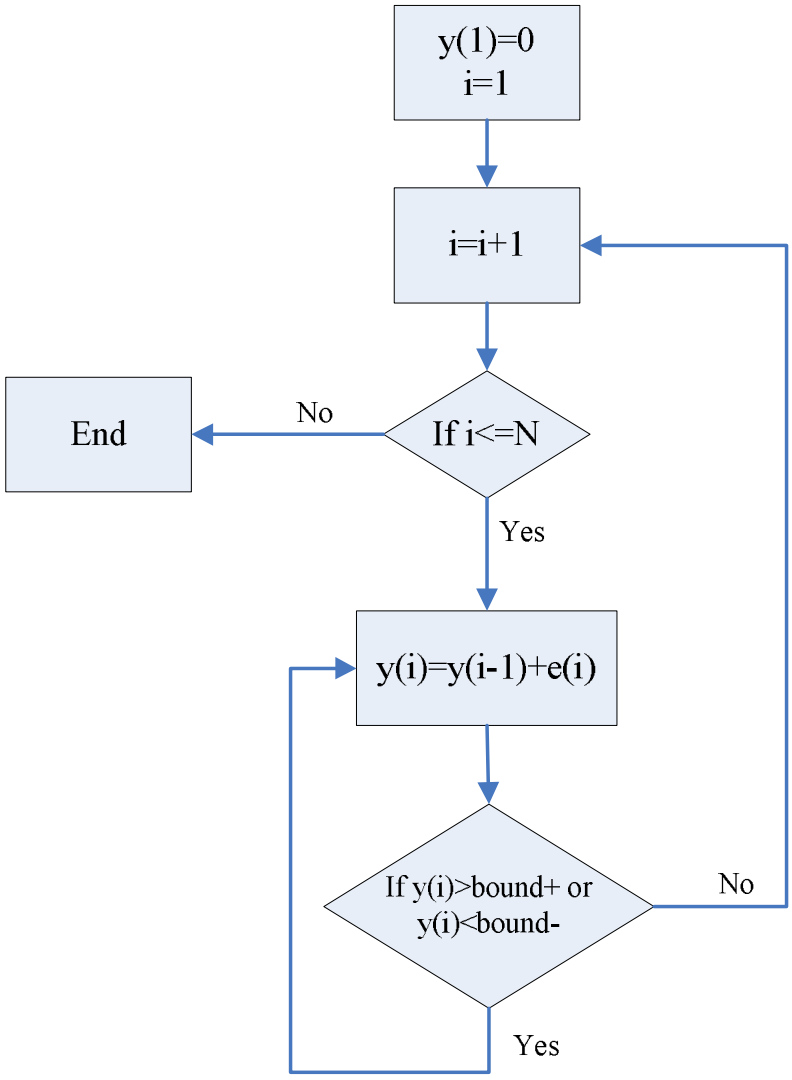
de Chazal P, Celler B (1997b). Selection of optimal parameters for ECG diagnostic classification. Comput. Cardiol., 24: 13-16.

Fusheng Y, Bo H, Qingyu T (2001). Approximate entropy and its application in biosignal analysis. In: Press I Nonlinear Biomedical

Signal Processing Volume II: Dynamic Analysis and Modeling. Biomedical Engineering.
Goldberger AL, Peng C-K, Lipsitz LA (2002). What is physiologic complexity and how does it change with aging and disease? Neurobiol. Aging, 23: 23-26.
Hénon M (1976). A two-dimensional mapping with a strange attractor. Commun. Math. Phys., 50: 69-77.
Hornero R, Aboy M, Abasolo D, McNames J, Goldstein B (2005). Interpretation of approximate entropy: Analysis of intracranial pressure approximate entropy during acute intracranial hypertension. IEEE Trans. Biomed. Eng., 52: 1671-1680. http://www.physionet.org/.
Veldhuis MJ, Veldhuis O, Straume M, Pincus SM (2001). Impact of pulsatility on the ensemble orderliness (approximate entropy) of neurohormone secretion. Am. J. Physiol. (Regulatory Integrative Comp. Physiol.). 281: R1975-R1985,
Nizami S, Green J, Eklund J, McGregor C (2010). Heart disease classification through HRV analysis using Parallel Cascade Identification and Fast Orthogonal Search Medical Measurements and Applications. Proceedings (MeMeA), 2010 IEEE International Workshop.
Pincus SM (1991). Approximate Entropy as a Measure of System Complexity. Proc. Natl. Acad. Sci., 88: 2297-2301.
Pincus SM (1995). Approximate entropy (ApEn) as a complexity measure. Chaos, 5: 110-117.
Pincus SM, Goldberger A (1994). Physiological time series analysis: What does regularity quantify? Am. J. Physiol., (Heart Circ. Physiol.,) 266: H1643-H1656.
Xua Y, Ma X, Ning X (2007). Estimation of ApEn for short-term HFECG based on multi-resolution analysis. Phys. A., 376: 401-408.

APPENDIX

A. The detailed procedure to realize Bounded Random Process model is as follows:



B: The calculation of relationship between ApEn and ε is equivalent as the following problem:

Random variables A and B are uniformed distributed in $(0, \sqrt{12}\varepsilon)$, what is the probability of $|A-B| < r$, while $r < \sqrt{12}\varepsilon$.

The answer is equal to the ratio of shaded area to the square area in Figure 1, noted as λ,

$$ApEn = -\log(\lambda) = \log(1/\lambda).$$