# Characterizing Histograms of Heartbeat Interval Differences with Gaussian Mixture Densities

R Sassi

Dipartimento di Tecnologie dell'Informazione, Università degli Studi di Milano, Italy

## Abstract

*In long-term HRV analysis, it is common choice to study the difference signal $IRR_i = RR_{i+1} - RR_i$. In this work we first verified the fitting of a Lévy stable distribution on the signals $IRR$ obtained from four databases, available on Physionet. They included normal subjects (N) but also individuals suffering from congestive heart failure (CHF) or showing ST segment changes (ST). The study showed that a Lévy stable distribution was generally more appropriate on the series than a Gaussian one (N: $1.70\pm0.19$; CHF: $1.74\pm0.18$; ST: $1.66\pm0.22$). The differences between the populations were not significant ($p > 5\%$).*

*Based on the value of RMSSD on local short intervals, we built a simple Gaussian mixture density for each $IRR$ series. Such mixture densities were able to properly describe the histograms in the databases under analysis. This explanation, which also avoids the necessity of invariant densities with not-finite second moments, might be closer to the physiological situation at hand.*

## 1. Introduction

When analyzing HRV in long recordings, RR series typically show non stationary evidences. For this reason it is common choice to study the difference signal $\mathrm{IRR(i)} = \mathrm{RR(i+1)} - \mathrm{RR(i)}$, which is, by construction, more stationary. Peng and coworkers [1] showed that such difference signal displays a normalized histogram with very long tails which can be properly modeled as a symmetrical Lévy stable distribution. In particular they found that the statistics of healthy subjects and of patients suffering from dilated cardiomyopathy are similar stable distribution with $\alpha \approx 1.7$.

A distribution $F_X$ is called *strictly stable* if, given $N$ mutually independent random variables $X_k$ with common distribution $F_X(x)$, does exist a value $C > 0$ such that

$$c_1X_1 + c_2X_2 + ... + c_NX_N = CX \qquad (1)$$

where $c_k$ are real numbers. The theory of random variables was formalized by Lévy [2]; among other results, he showed that the family of symmetrical distributions

$$f_X(x, \alpha, \gamma) = \frac{1}{\pi}\int_0^{+\infty} e^{-\gamma q^\alpha}\cos(qx)dq \qquad (2)$$

is the only possible symmetrical solution to the functional equation (1) combined with the auxiliary relation

$$c^\alpha = c_1^\alpha + c_2^\alpha + ... + c_N^\alpha. \qquad (3)$$

The value $\alpha \in (0, 2]$ is the characteristic exponent and it defines the spread of probability towards the tails of the density function; $\gamma > 0$ is instead a scaling constant. Such distributions are sometimes called *Lévy stable distributions*. The density $f_X$ can be obtained as inverse Fourier transform of the function $e^{-\gamma|q|^\alpha}$ often called *characteristic function*.

A strictly stable distribution is *fractal* in nature, as the sum of N independent variables extracted from it looks exactly the same as a single variables[1], once adjusted by a scale factor $C$. The Cauchy (Lorentzian) and the Gaussian distribution are particular cases of Lévy stable distributions obtained with $\alpha = 1$ and $\alpha = 2$ respectively. The first moment is finite for $1 < \alpha \leq 2$ while the second only for $\alpha = 2$, that is the normal is the only stable distribution with finite second moment. All superior moments are infinite.

The results of Peng and coworkers thus suggested that the tails in the distribution of the difference signal $\mathrm{IRR(i)}$ were "heavier" than those permitted by a normal distribution, making large values of $\mathrm{IRR(i)}$ more likely. They also suggested that the slow decay of the distribution for large increment might relate to the dynamics of the cardiovascular control system. A detailed study of the importance of the Lévy statistics in physiology was given by West and Deering [3] (see also [4]) who developed a simple linear mathematical model for the fluctuations which can also be applied to $\mathrm{IRR(i)}$. Also, Tsallis et al. [5] using a generalization of the central limit theorem, offered a more general explanation for the ubiquity of Lévy distributions.

[1] It is possible to derive a generalization of the box-counting dimension for a random variable $X$ extracted from a strictly stable distribution $F_X$. It follows that the fractal dimension of $X$ is $D = \alpha$

In this work we first studied the histograms of the difference signals obtained from four different databases of long term HRV recordings, available on Physionet, to verify the findings of Peng et al. on a larger number of cases. While several reasons might induce the presence of a density similar to a Lévy stable distribution, we then tried to verify if a "stable-like" distribution could be qualitatively obtained with a simple Gaussian mixture density, not fitted on the data but based on the short term values of RMSSD.

## 2. Methods

*Dataset*. We analyzed 187 RR series obtained from long-term Holter recordings. The series were selected among the ones available on Physionet [6] to make the study as reproducible as possible. Four databases were considered: (i) the MIT-BIH Normal Sinus Rhythm Database (`nsrdb`, long-term ECG recordings of 18 subjects with no significant arrhythmias, 5 men and 13 women; sampling rate: 128 Hz); (ii) the Normal Sinus Rhythm RR Interval Database (`nsr2db`, beat annotation files for long-term ECG recordings of 54 subjects in normal sinus rhythm, 30 men and 24 women; original sampling rate: 128 Hz); (iii) the Congestive Heart Failure RR Interval Database (`chf2db`, beat annotation files for long-term ECG recordings of 29 subjects with congestive heart failure; original sampling rate: 128 Hz); (iv) the Long-Term ST Database (`ltstdb`, 86 Holter ECG recordings of 80 subjects displaying a variety of events leading to ST segment changes; sampling rate: 250 Hz). Given the fact that the databases already contained beat annotations obtained by labeling software with manual review and corrections, and that for two databases the original recordings were not available, we further analyzes the annotations as provided. Summarizing we studied normal subjects (N, 72 cases), individuals suffering from congestive heart failure (CHF, 29 cases) or with ECG showing ST segment changes (ST, 86 cases).

*Distribution's parameters estimation*. For each RR series, only interval differences obtained from consecutive NN intervals were retained for further processing. The fit of a discrete Lévy stable distribution to each sample was performed using maximum likelihood [7] employing a very robust code (STABLE Matlab toolbox, Robust Analysis Inc [8]). For the computations, intervals in absolute values larger than about 2 seconds were marked as not physiological and excluded. In fact, inter-beat intervals generated by spurious beats and artifacts would have increased anomalously the weight of the tails in the distribution (the number of sample excluded was minimal though, on average $\approx 5$ points per series). Then during the fitting, symmetry of the stable distribution was assumed.

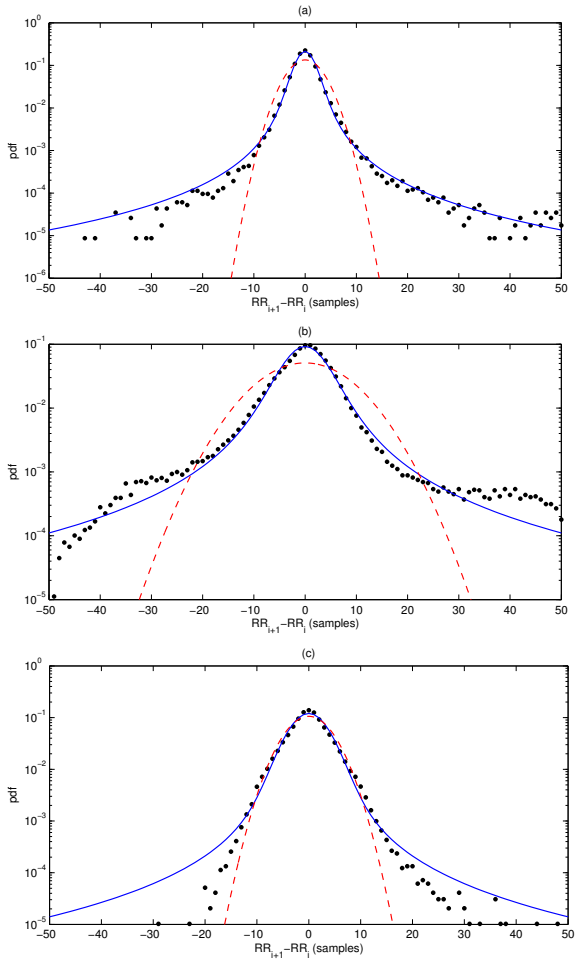


Figure 1. Probability density functions of IRR(i), the differences between adjacent NN intervals, for three different long-term recordings (dots): (a) record `nsr2db/nsr029`; (b) record `nsrdb/16273`; (c) record `nsr2db/nsr004`. The Lévy stable distributions fitted to the IRR(i) series are plotted with continuous lines; the corresponding values of $\alpha$ are 1.63, 1.47 and 1.80 respectively. Also the Gaussian distributions which best fit the same data (maximum likelihood) are reported (sketched lines).

## 3. Results

The fitting of a discrete Lévy stable distribution was performed on each of the 187 series. Only the parameter $\alpha$, characterizing the weight of the tails, was considered. Three sample distributions are showed in figure 1. The fitting was in many cases good as in the case reported in panel (a). For other recordings, for values of IRR(i) larger than 20 samples (about 160 ms), the actual distributions showed departures from a stable-like shape. Panel (b) and (c) report cases in which the tails have a weight, respec-

tively, larger and smaller than expected. On average, the mean value of $\alpha$ across each population was smaller than 2 (N: $1.70 \pm 0.19$; CHF: $1.74 \pm 0.18$; ST: $1.66 \pm 0.22$) but it was not so uncommon to have values of $\alpha$ comprised between 1.9 and 2. We also checked the width of the confidence intervals for the parameter $\alpha$ which was always very small, but this is more the effect of the large number of points included in each series (on average $100325 \pm 1985$) than a proof of goodness-of-fit to a stable-like distribution.

The differences between the populations were not significant ($p > 5\%$, t test for multiple comparisons) and the average value across all the 187 recordings was $1.69 \pm 0.20$. These results confirm the findings of Peng et al. [1] who reported a value of $\alpha$ of 1.7 and did not find any significant differences between a normal subject and a severely heart diseased one. The histogram of the values of $\alpha$ obtained on the whole population is reported in figure 2.

Several reasons might induce the presence of a density similar to a Lévy stable distribution. Recently, Lin and Hughson [9], building on the work of Hausdorff and Peng [10], suggested a model in which the increments $\mathrm{IRR(i)}$ are generated by averaging different Gaussian random walkers (with proper weights). Such model was able to produce series which exhibit distributions with long tails.

A factor which need to be considered though is that the physiological mechanisms which influence the increments signal $\mathrm{IRR(i)}$ are difficultly stationary on the time span on which the series were recorded. Non-stationarity is likely to be the reason underlying the deviation from the fitted distribution in figure 1. To understand this further we repeated the fitting procedure on shorter non-overlapping segments of 500 points, slightly longer than 5 minutes, a time span under which we might assume a larger degree of stationarity. While the histograms were still well described by a stable distribution, the mean value of $\alpha$ across the populations was markedly larger (N: $1.88 \pm 0.10$; CHF: $1.88 \pm 0.11$; ST: $1.84 \pm 0.13$) and closer to a normal distribution (i.e. $\alpha = 2$). Similar results were obtained also considering longer segments (1000 points).

While such discrepancies might be due to numerical problems of convergence (shorter series imply a smaller number of extreme events, on average), in this case it is unlikely to be so. In fact, we verified that once generated 20 random series of 500 points distributed according to a Lévy stable distribution (using the STABLE toolbox and with the same parameters obtained from each HRV series), then a newly fitted stable distribution displayed an average estimation error on $\alpha$ of about $0.048 \pm 0.036$. Therefore the differences are improbably originated from the estimation uncertainties alone.

To explore this further we then built a simple Gaussian mixture density $p(x)$ , which is obtained as a combination of different probability Gaussian density functions $g(x, \theta_j)$, e.g.

$$p(x) = \sum_{j=1}^{c} \omega_j g(x, \theta_j) \tag{4}$$

where $\theta_j$ is the standard deviation which fully characterize the density $g()$ (the mean is set to zero). We arbitrary selected $c = 5$ and for each RR series we constructed the density (4) as follow. First, we estimated the RMSSD[2] on consecutive 500 points intervals. Then an histogram of the RMSSD values was built with $c$ bins. Finally, the central point of each bin provided a value for $\theta_j$ and $\omega_j$ was set to the relative frequency of that bin.

A mixture density was built for each series and then used to produce a surrogate signal. The surrogate signals were still displaying heavy tails thus a Lévy stable distribution was eventually fit on them. Figure 3 compares the histograms of the interval difference series and of the one obtained from the corresponding mixture model. While the values of $\alpha$ obtained were larger than the ones observed on the series, they were still significantly smaller than 2 and smaller than those estimated on shorter non-overlapping segments of 500 points (N: $1.79 \pm 0.14$; CHF: $1.86 \pm 0.14$, ST: $1.78 \pm 0.15$).

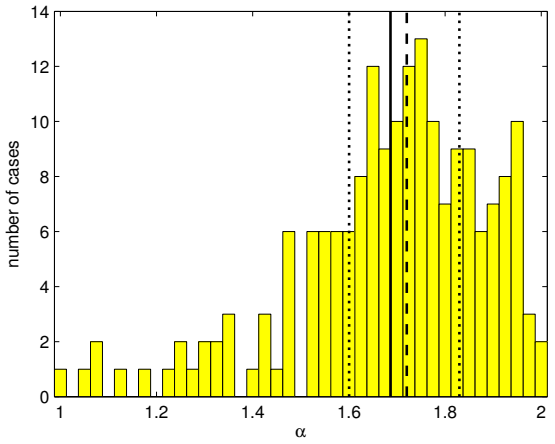


Figure 2. Histogram of the values of $\alpha$ fitted on the 187 interval difference series. The vertical lines represents: mean (continuous), median (sketched), 25% and 75% percentiles (dotted lines).

## 4. Discussion and conclusions

The study confirmed on average the findings of Peng et al. [1], that is long-term HRV recordings display interval difference series which can be reasonably modeled

[2]RMSSD: square root of the mean squared differences of successive NN intervals [11]. RMSSD is a measure of the short-term HRV components, highly correlated with pNN50 and HF power.

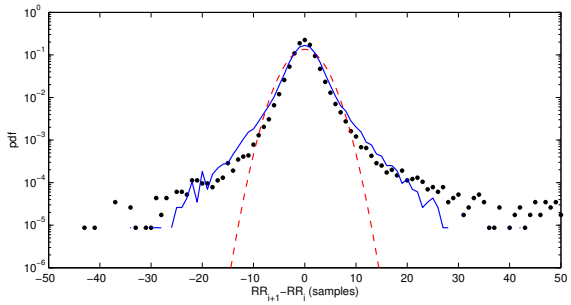


Figure 3. Probability density functions of IRR(i) for subject `nsr2db/nsr029` (dots), along with the sample density obtained from the corresponding mixture density (thick line), and the fitted Gaussian density (sketched line)

with symmetrical Lévy stable distributions. In most cases the model holds well and offers a superior fit with respect to a normal distribution for values of IRR(i) smaller than $\approx 160$ ms. For larger values of IRR(i), departures from a stable distribution might appear, similarly to what shown in figures 1(b) and 1(c). Also, we verified that the value of the parameter $\alpha$, which describes the weight of the tails, did not vary across the three groups we considered. It displayed a mean value of 1.69 very close to what suggested by Peng et al. Said that, we also noticed cases in which either the value of $\alpha$ was much smaller than 1.7 or closer to 2 as figure 2 shows.

The naive Gaussian mixture densities we built, for which we did not perform any actual fit, were able to generate heavy tails and to properly describe the histograms in the databases under analysis. The weights and the variances of the Gaussian components were derived from the behavior of the HRV short-term components (RMSSD). As Nolan noticed [7], skeptics of non-Gaussian stable models often employ time-varying variances and mixture models to explain for the heavy tails found in the data. In this case though, mixture models must be considered a description possibly closer to the physiological mechanisms underlying the generation of the series. In fact the continuous adaptations of the cardiovascular control system would suggest to be cautious before assuming stationarity.

Concluding, this study support the idea that a Lévy stable distribution might be employed, when a concise description of the series IRR(i) is needed. On the other hand, in our opinion, the non-stationarity of the cardiovascular control system hints that the heavy tails might arise from the superposition of Gaussian phenomena, driven by the same input which modulate the HRV short-term components. Thus the long tails seen in the IRR(i) could be modeled also by Gaussian mixture densities. Occam's razor would prefer this simple explanation which also avoids the necessity of invariant densities with not-finite second moments (unlikely in a physiological system).

## References


[1] Peng CK, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL. Long-range anticorrelations and non-gaussian behaviour of the heartbeat. Phys Rev Lett 1993; 70(9):1343–1346.

[2] Lévy P. Théorie de l'Addition des Variables Aléatories. Paris: Gauthier-Villars, 1937.

[3] West BJ, Deering W. Fractal physiology for physicists: Lévy statistics. Physics Reports 1994;246(1-2):1 – 100.

[4] West BJ. Fractal statistics in biology. Physica D Nonlinear Phenomena 1995;86(1-2):12 – 18.

[5] Tsallis C, Levy SVF, Souza AMC, Maynard R. Statistical-mechanical foundation of the ubiquity of Lévy distributions in nature. Phys Rev Lett 1995;75(20):3589–3593.

[6] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000 (June 13);101(23):e215–e220.

[7] Nolan JP. Maximum likelihood estimation of stable parameters. In Barndorff-Nielsen OE, Mikosch T, Resnick SI (eds.), Lévy Processes: Theory and Applications. Boston, Birkhäuser, 2001; 378–400.

[8] Robust Analysis Inc, Takoma Park, MD, USA. `www.RobustAnalysis.com`.

[9] Lin DC, Hughson RL. Using gaussians to model increment distribution of the long-term R-wave interval in healthy humans. Chaos Solitons and Fractals 2001;12:1335–1354.

[10] Hausdorff JM, Peng CK. Multiscaled randomness: A possible source of 1/f noise in biology. Phys Rev E 1996; 54:2154–2157.

[11] Task Force of the European Society of Cardiology, the North American Society of Pacing and Electrophysiology. Heart Rate Variability, Standards of Measurement, Physiological Interpretation and Clinical Use. Circulation 1996; 93:1043–1065.



Address for correspondence:

Roberto Sassi
Dipartimento di Tecnologie dell'Informazione
Università degli Studi di Milano
via Bramante 65
26013 Crema (CR) Italy
E-mail address: roberto.sassi@unimi.it