Robust Estimation of the Scaling Exponent in Detrended Fluctuation Analysis of Beat Rate Variability

Matti Molkkari, Esa Räsänen

Tampere University of Technology, Tampere, Finland

Abstract

Detrended fluctuation analysis (DFA) is a popular method for studying fractal scaling properties in time series. The method has been successfully employed in studying heart rate variability and discovering distinct scaling properties in different pathological conditions. Traditionally the analysis has been performed by extracting two scaling exponents from linear fits, for short- and long-range correlations respectively. The extent of these ranges is subjective and the linear two-range model potentially disregards additional information present in the data.

Here we present an optimization scheme to obtain data-adaptive segmentation of the fluctuation function into approximately linear regimes. Additionally, we present a method based on the Kalman smoother for obtaining a whole spectrum of scaling exponents as a function of the DFA window size. The smoother is resistant to statistical noise in the fluctuation function, while remaining sufficiently sensitive to capture variations in the scaling exponent.

We employ the methods in the analysis of the heart rate variability of patients with different pathological conditions. Their performance is evaluated by the additional information provided to support correct classification of the pathological conditions, revealing more complex structure in the scaling exponent beyond the two-range model.

1. Introduction

The relationship between the autonomic nervous system and cardiovascular health has long been recognized. Their compounded effect is manifested in heart rate variability (HRV), and alterations in HRV may be indicative of various cardiovascular diseases [1]. Detrended fluctuation analysis (DFA), originally developed for studying long range correlations in DNA sequences [2], has been a popular method for studying self-affinity in signals. By computing the mean fluctuations $F(s)$ around a trend at multiple scales s, DFA is applied to assess power law scaling $F(s) \propto s^{\alpha}$ described by the scaling exponent α.

In the context of HRV, DFA is traditionally performed on the signal consisting of interbeat intervals, and the scaling exponent is determined by linear regression on a log-log plot of the fluctuation function $F(s)$. Conventionally the linear regression is applied separately to two regimes to obtain short- and long-range scaling exponents. This approach has been criticized as over-simplification, and alternative methods have been proposed [3, 4].

We propose two methods to complement the analysis. We present an optimization scheme for determining segmentation of the fluctuation function into approximately linear regimes, and a method for obtaining a smooth spectrum of scaling exponents as a function of the scale. The methods are data-adaptive and parameter-free, while providing robust estimates in the presence of statistical noise in the fluctuation function.

2. Detrended fluctuation analysis

In the standard DFA formulation [2], integrated time series is divided into N_s non-overlapping windows of size s, from which the local trend is calculated in each window w by a least squares fit of a low order polynomial. The squared fluctuations $F^2_{s,w}$ in each window are defined as the mean squared differences from the local trend. We compute an error estimate for the fluctuation function as follows.

Let μ_s and ϵ_s denote the mean and its standard error, respectively, of the squared fluctuations

$$\mu_s = \left\langle F^2_{s,w} \right\rangle; \qquad \epsilon_s = \sqrt{\frac{\mathrm{Var}[F^2_{s,w}]}{N_s}}, \tag{1}$$

where the mean and the variance are taken over all windows w of size s. The fluctuation function $F(s)$ and its error estimate $\Delta F(s)$ are then given by

$$F(s) = \sqrt{\mu_s}; \qquad \Delta F(s) \approx \frac{\epsilon_s}{2\sqrt{\mu_s}}. \tag{2}$$

We denote their logarithmic counterparts by tildes, $\tilde{F}(\tilde{s})$ and $\Delta\tilde{F}(\tilde{s})$, respectively.

3. Kalman filter and smoother

Let us consider a linear probabilistic state space model of a system, described by its hidden state $\boldsymbol{x}_k \in \mathbb{R}^n$ and yielding the measurement $\boldsymbol{y}_k \in \mathbb{R}^m$ at the time step k, described by the equations

$$\boldsymbol{x}_k = \mathbf{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{q}_{k-1} \tag{3}$$

$$\boldsymbol{y}_k = \mathbf{H}_k\boldsymbol{x}_k + \boldsymbol{r}_k. \tag{4}$$

where $\mathbf{A}_{k-1} \in \mathbb{R}^{n\times n}$ and $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ are the state transition and measurement model matrices, respectively. The system is disturbed by process noise $\boldsymbol{q}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ and measurement noise $\boldsymbol{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$, modeled by zero mean Gaussian noise with covariance matrices $\mathbf{Q}_{k-1} \in \mathbb{R}^{n\times n}$ and $\mathbf{R}_k \in \mathbb{R}^{m\times m}$.

With a normally distributed prior distribution $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{m}_0, \mathbf{P}_0)$ for the hidden state, the Kalman filter [5] provides a recursive closed-form minimum-mean-squared error solution for the posterior filtering distributions $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \mathcal{N}(\boldsymbol{m}_k, \mathbf{P}_k)$, given all the measurements up to the current time step k. Similarly, the Kalman smoother allows the computation of the smoothing distributions $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T}) = \mathcal{N}(\boldsymbol{m}_k^{\mathrm{s}}, \mathbf{P}_k^{\mathrm{s}})$, that take into account all the T measurements to provide an estimate for the state of the system at the time step $k \leq T$. The smoothing distributions are calculated from the filtering distribution means $\boldsymbol{m}_k$ and covariances $\mathbf{P}_k$ by backwards recursion. For a modern exposition about the subject, see, e.g., Ref. [6].

The scaling exponent $\alpha(s)$ is estimated from the logarithmically transformed quantities. The dynamic model applied in this paper assumes that the exponent remains approximately constant between adjacent window sizes, except for tiny perturbations derived from the data. The hidden state $\boldsymbol{x}_k$ to be estimated then consists of the observed fluctuation function values and its derivative, which corresponds to the scaling exponent:

$$\boldsymbol{x}_k = \begin{bmatrix}\tilde{F}_k & \alpha(\tilde{s}_k)\end{bmatrix}^\top. \tag{5}$$

Here $\tilde{s}_k$ are the utilized logarithmic DFA window sizes in ascending order and $\tilde{F}_k = \tilde{F}(\tilde{s}_k)$. The state transition and measurement model matrices are then

$$\mathbf{A}_k = \begin{bmatrix}1 & h_{k_+}\\ 0 & 1\end{bmatrix}; \qquad \mathbf{H} = \begin{bmatrix}1 & 0\end{bmatrix}, \tag{6}$$

where $h_{k_+} = \tilde{s}_{k+1} - \tilde{s}_k$ is the forward difference of the window size at k. The error estimate of the fluctuation function is readily utilized as the measurement noise covariance by $R_k = \left[\Delta\tilde{F}(\tilde{s}_k)\right]^2$, which is now simply a scalar.

The scaling exponent is gently perturbed by the process noise, allowing it to vary as a function of the window size. The noise covariance is estimated by taking a weighted sample variance of the finite difference approximation [7] of the logarithmic fluctuation function derivative. The derivative is estimated at each window size $\tilde{s}$ by

$$\tilde{F}'_k \approx \frac{h_{k_-}^2\tilde{F}_{k+1} + \left(h_{k_+}^2 - h_{k_-}^2\right)\tilde{F}_k - h_{k_+}^2\tilde{F}_{k-1}}{h_{k_-}h_{k_+}\left(h_{k_+} + h_{k_-}\right)}, \tag{7}$$

along with an estimate for the error arising from the uncertainty in the fluctuation function

$$\Delta\tilde{F}'^2_k \approx \left[\frac{1}{h_{k_-}h_{k_+}\left(h_{k_+} + h_{k_-}\right)}\right]^2 \Bigg[\left(h_{k_-}^2\Delta\tilde{F}_{k+1}\right)^2 + \left(h_{k_+}^2 - h_{k_-}^2\right)^2\Delta\tilde{F}_k^2 + \left(h_{k_+}^2\Delta\tilde{F}_{k-1}\right)^2\Bigg], \tag{8}$$

where $h_{k_-} = \tilde{s}_k - \tilde{s}_{k-1}$ are the backwards differences. The boundaries are considered by assuming that the function continues linearly.

The weighted sample variance $\hat{\sigma}^2$ of the derivatives is computed with the weights taken to be inversely proportional to this squared error estimate. This variance estimates the magnitude of the nudges that the scaling exponent experiences between logarithmic window sizes, leading to the following process noise covariance:

$$\mathbf{Q}_k = \hat{\sigma}^2\begin{bmatrix}\frac{1}{3}h_{k_+}^3 & \frac{1}{2}h_{k_+}^2\\ \frac{1}{2}h_{k_+}^2 & h_{k_+}\end{bmatrix}. \tag{9}$$

4. Integer linear programming

Let us consider the problem of optimally segmenting the logarithmic fluctuation function $\tilde{F}_k$ into approximately linear regimes. First we compute all the possible n linear regression fits to the fluctuation function that cover some desired minimum number of different window sizes. The optimal segmentation is then sought by solving the following linear programming problem

$$\arg\min_{\boldsymbol{x}} \boldsymbol{c}^\top\boldsymbol{x}; \qquad \boldsymbol{c} \in \mathbb{R}^n; \qquad \boldsymbol{x} \in \{0,1\}^n, \tag{10}$$

where the components x_i of the binary vector $\boldsymbol{x}$ indicate whether the i-th fit is utilized in the optimal segmentation. The components of the coefficient vector $\boldsymbol{c}$ are the squared residuals r_i^2 of the linear regression fits: $c_i = r_i^2$.

For expressing the optimization constraints we introduce the auxiliary variables a_{ik}, defined to be equal to unity if the i-th fit covers the k-th window size and zero otherwise. The constraints are then given by the following equations

$$\sum_{i=1}^{n} a_{ik} = 1; \qquad \sum_{i=1}^{n} x_i = N, \tag{11}$$

where the former relations ensure that each point in the fluctuation function is covered by exactly one linear fit,

and the latter relation ensures that the solution consists of the desired number of linear fits N. The problem is then readily solved by any existing integer linear programming suite, such as the open source GNU Linear Programming Kit (GLPK).

Choosing an optimal value for the number of linear segments, N, is a partial and subjective decision. The approach taken here proceeds as follows. Let $R^2(N)$ denote the total squared residual of the optimal segmentation with N segments. This residual is trivially reduced by increasing the number of segments. We seek a solution that consists of as few segments as admissible for equitable segmentation. Therefore, we consider the desirability $D(N) \propto 1/\left[NR^2(N)\right]$ of a solution to be inversely proportional to the number of segments and the total squared residual. The solution is optimal when it maximizes this quantity.

This optimization scheme may be performed separately for each fluctuation function or simultaneously to a group of fluctuation functions, provided that they all have been calculated at compatible window sizes. In the latter case the squared residuals r_{ij}^2 of the i-th fit in the j-th fluctuation function are summed over all the fluctuation functions when computing the coefficients as $c_i = \sum_j r_{ij}^2$.

5. Data and methods

The analysis is performed on the publicly available resources of PhysioBank [8], utilizing the following databases: Normal Sinus Rhythm RR Interval Database (nsr2db), The Long-Term ST Database (ltstdb), The Long-Term AF Database (ltafdb), Congestive Heart Failure RR Interval Database (chf2db), and The BIDMC Congestive Heart Failure Database (chfdb). The `ihr` utility of the WFDB software package [8] is employed for extracting interbeat intervals from the already annotated databases. The utility is executed with the parameters `-i` to include all intervals bounded by QRS annotations and `-d 20` to filter outliers.

Distinguishing intrinsic variations in the heart rate from those induced by extrinsic influences in long term recordings may be perplexing, resulting in unknown trends in the data. While DFA is successful in detrending simple long-term trends, its ability to cope with complex nonlinear trends is disputed [9]. Therefore the analysis is additionally performed by explicitly removing a trend determined by the local median in a moving window of 101 beats wide. In both cases DFA is performed with linear detrending in 45 logarithmically distributed window sizes in the interval of 5–200 beats.

The scaling exponent α is determined from the fluctuation functions with three methods: the traditional linear fits in short (5–16 beats) and long (16–64 beats) range regimes, the optimal linear segmentation applied to the whole dataset at once, and the alpha spectra obtained with the Kalman smoother. These scaling exponents are then utilized as features in the classification task of identifying the different pathological conditions. The classification is performed by various algorithms available in the scikit-learn Python module [10]. Hyperparameters are optimized with a simple grid search.

6. Results

In Fig. 1(a) we show examples of the fluctuation functions together with the Kalman smoother estimates. The Kalman smoother yields smooth continuous estimates for the fluctuation functions, and consequently for the scaling exponents, even in the presence of statistical noise. The optimal linear segmentation is also illustrated (wide transparent lines), and it is found to produce rational segmentations. The optimal segmentation of the chfdb data is portrayed in Fig. 1(b). The four linear regimes discovered suggest that the traditional two-range model with two slopes may be insufficient. Performing the optimal linear segmentation on the whole data set reveals four distinctive segments in the following ranges: 5–12, 13–32, 35–81, and 87–200 beats.

The mean alpha spectra of the detrended data are shown in Fig. 2(a) for different pathological conditions. The average scaling behavior is very similar in healthy individuals and in those suffering from ST episodes. To evaluate the different methods, we attempt to classify the pathological conditions based on the scaling exponents. Typical results are shown in Fig. 2(b). The more detailed picture depicted by the new methods is especially necessary for distinguishing the previously mentioned problematic cases, with the two-range model (right) being moderately worse in the task. The classification results with the optimized linear segmentation are not shown, as they are already very similar to the results obtained with the smooth alpha spectra.

Curiously, the data detrended by the local median yields better classification results, despite the detrending removing long range correlations. This is consistently observed with all three estimation schemes for the scaling exponent, and across all the classification algorithms.

7. Conclusions

We have introduced methods for robustly determining the DFA scaling exponent, and a spectrum of exponents, from noisy fluctuation functions in a parameter-free manner. The methods have been tested by classifying different pathological conditions according to the scaling exponents of interbeat intervals. The results support the hypothesis that the traditional two-range model for the scaling exponents can be insufficient, especially for complex classification tasks.

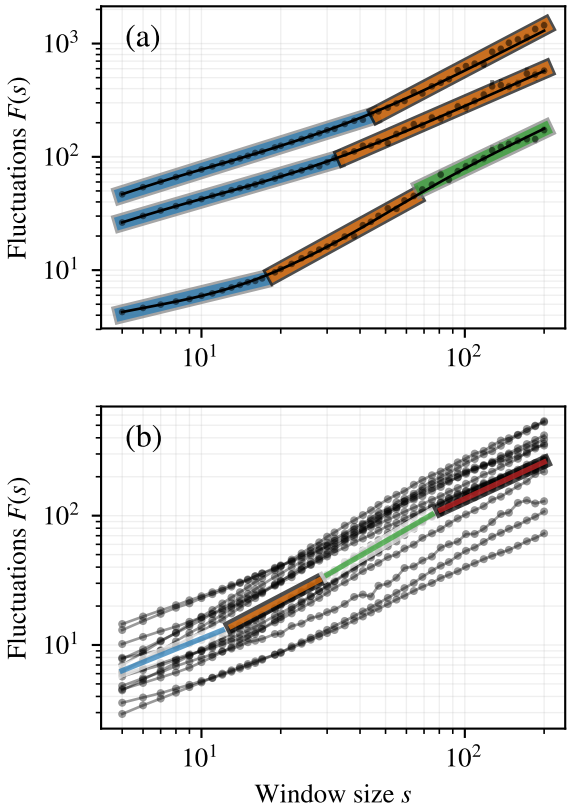

Figure 1. (a) Examples of fluctuation functions from the data (dots with error bars) together with the Kalman smoother estimates (thin black lines) and the optimal segmentation (wide transparent lines). (b) Fluctuation functions of the chfdb data (dotted lines) together with the optimal segmentation (wide transparent lines). The slopes of the optimal segmentation correspond to the averages of all the fluctuation functions.

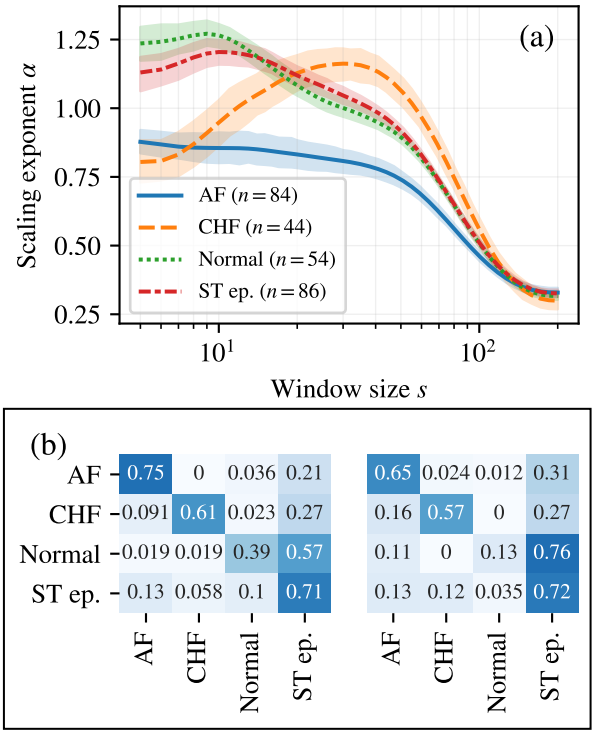

Figure 2. (a) Kalman smoother estimates for the scaling exponents of the detrended data. Shown are the mean scaling exponents and their 95 % confidence intervals for the various pathological conditions. (b) Confusion matrices of the classification task with support vector machines for the Kalman smoother alpha spectra (left) and the traditional short/long range alphas (right) on the detrended data. The results are based on leave-one-out cross-validation.

References

[1] Malik M, et al. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal 1996;17(3):354–381.

[2] Peng C, et al. Mosaic organization of DNA nucleotides. Physical Review E 1994;49(2):1685–1689.

[3] Echeverría JC, et al. Interpretation of heart rate variability via detrended fluctuation analysis and $\alpha\beta$ filter. Chaos An Interdisciplinary Journal of Nonlinear Science 2003; 13(2):467–475.

[4] Xia J, Shang P, Wang J. Estimation of local scale exponents for heartbeat time series based on DFA. Nonlinear Dynamics 2013;74(4):1183–1190.

[5] Kalman RE. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 1960;82(1):35.

[6] Särkkä S. Bayesian filtering and smoothing. Cambridge: Cambridge University Press, 2013.

[7] Fornberg B. Generation of finite difference formulas on arbitrarily spaced grids. Mathematics of Computation 1988; 51(184):699–706.

[8] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[9] Bryce RM, Sprague KB. Revisiting detrended fluctuation analysis. Scientific reports 2012;2(1):315.

[10] Pedregosa F, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011; 12:2825–2830.

Address for correspondence:

Matti Molkkari
Tampere University of Technology, Laboratory of Physics
Korkeakoulunkatu 10, 33720 Tampere, Finland
matti.molkkari@tut.fi