fathon: A Python package for a fast computation of detrendend fluctuation analysis and related algorithms

Stefano Bianchi[1]

1 Department of Mathematics and Physics, Roma Tre University

Summary

`fathon` is a Python package for DFA (detrended fluctuation analysis) and related algorithms.

DFA (fathon.DFA) was first developed by Peng to study memory effects in sequences of DNA (Peng et al., 1993). DFA is useful in recognising long memory processes by means of the scaling properties of the fluctuation function $F(n)$. $F(n)$ is calculated starting from the cumulative sum of the process $X(t)$ (most likely a time series), $Y(t) = \sum_{t'=1}^{t}[X(t') - \langle X \rangle]$, with $\langle X \rangle$ the mean of $X(t)$. $Y(t)$ is then divided into $N_n = \lfloor N/n \rfloor$ non-overlapping time intervals of length n. Since the length of the time series is often not a multiple of n, some data at the end of $Y(t)$ could be excluded. In order to not neglect them, the same procedure can be repeated starting from the end of $Y(t)$, and the total number of time intervals will amount to $2N_n$. The data in each interval s are fitted with a polynomial Y_s^{fit}, and the variance is computed, $F^2(n,s) = \frac{1}{n}\sum_{i=1}^{n}[Y((s-1)n+i) - Y_s^{fit}(i)]^2$ for $s = 1, \ldots, N_n$, and $F^2(n,s) = \frac{1}{n}\sum_{i=1}^{n}[Y(N-(s-N_n)n+i) - Y_s^{fit}(i)]^2$ for $s = N_n + 1, \ldots, 2N_n$. Finally, the fluctuation function is defined as

$$F(n) = \left[\frac{1}{2N_n}\sum_{s=1}^{2N_n} F^2(n,s)\right]^{1/2}$$

$F(n)$ is expected to scale as n^H. H is the Hurst exponent, and its value can indicate if a process is persistent or anti-persistent:

- $H \in [0.0, 0.5) \to$ anti-persistency. The process under study is anti-persistent and tends to decrease (increase) after a previous increase (decrease). An anti-persistent process appears very noisy;
- $H = 0.5 \to$ uncorrelated process;
- $H \in (0.5, 1.0] \to$ persistency. If a process has been increasing (decreasing) for a period T, then it is expected to continue to increase (decrease) for a similar period. Persistent processes show long-range correlations and exhibit relatively little noise;
- $H > 1.0 \to$ nonstationary process, stronger long-range correlations are present.

DFA has found various applications in different fields, such as geophysics (Fraedrich & Blender, 2003; Varotsos, Assimakopoulos, & Efstathiou, 2007) and economics (Ivanov, Yuen, Podobnik, & Lee, 2004; Liu, Gopikrishnan, & Stanley, 1999). In the the geophysics references, DFA has been applied to study the persistency of deseasonalized CO_2 data and the difference in persistency between temperature records over land and ocean in order to validate climate models. In the economics references, DFA has been applied to find a common scaling behavior in the temporal organization of trading, and to study the persistency of economic records within different time periods (hours, days, months).

DFA has since evolved to multifractal detrended fluctuation analysis (MFDFA, fathon.MFDFA) (Kantelhardt et al., 2003) to take into account the variability of time series and the possible multiple scaling properties of the fluctuation function. $F(n)$ can be generalised and computed at different orders q,

$$F_q(n) = \left[\frac{1}{2N_n} \sum_{s=1}^{2N_n} [F^2(n,s)]^{q/2} \right]^{1/q} \sim n^{h(q)}$$

DFA corresponds to MFDFA with $q = 2$, and $h(2) = H$. q can take any real value. For positive values of q, time intervals s with a large variance $F^2(n,s)$ will dominate. Thus, for positive values of q, $h(q)$ describes the scaling behavior of time intervals with large fluctuations. Instead, for negative values of q, time intervals with a small variance $F^2(n,s)$ will dominate. Hence, for negative values of q, $h(q)$ describes the scaling behavior of time intervals with small fluctuations. For $q = 0$, $F_q(n)$ is divergent and can be replaced by an exponential of a logarithmic sum

$$F_0(n) = \exp \left[\frac{1}{4N_n} \sum_{s=1}^{2N_n} \ln(F^2(n,s)) \right] \sim n^{h(0)}$$

If $h(q)$ is approximately constant for all the values of q, the time series is said to be monofractal, i.e., it exhibits the same scaling at all scales. If $h(q)$ instead varies significantly, the time series is multifractal, i.e., it exhibits different scalings at different scales. The different scalings are better described by the multifractal spectrum, defined as

$$f(\alpha) = q[\alpha - h(q)] + 1$$

with $\alpha = h(q) + q\frac{dh(q)}{dq}$. The spectrum is concave down, and the wider it is, the more multifractal the time series is.

Other algorithms have been further developed, such as the time dependent Hurst exponent (fathon.HT) (Ihlen, 2012) that examines the changes of persistency as a function of time intervals. The square root of the variance $F^2(n,s)$ is computed in every interval of length n. However, now the time intervals overlap, spanning the whole time series, and the fluctuations in every window are given by

$$F(n,s) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [Y((s-1)+i) - Y_s^{fit}(i)]^2}$$

with $s = 1, \ldots, N - n + 1$. Then, each value of the fluctuations and the value of $F_q(N)$ (that is the same for every order q) can be connected with a straight line to obtain a local Hurst exponent H_t as the slope of the line, i.e., the Hurst exponent for a single time interval.

Another recent algorithm is the detrended cross-correlation analysis (DCCA, fathon.DCCA) (Podobnik & Stanley, 2008) that studies cross-correlations in terms of persistency between non-linear time series. It needs two time series to be calculated, $X_a(t)$ and $X_b(t)$. They are divided into $N - n$ overlapping time intervals, each one containing $n + 1$ values. The covariance is then computed, $F_{ab}^2(n,s) = \frac{1}{n-1} \sum_{k=s}^{n+s} [Y_{as}(k) - Y_{as}^{fit}(k)][Y_{bs}(k) - Y_{bs}^{fit}(k)]$, where a and b refer to the first and the second time series, respectively, and $s = 1, \ldots, N - n$. Finally, the fluctuation function is calculated summing over all the overlapping $N - n$ time intervals of size $n + 1$,

$$F_{ab}(n) = \left[\frac{1}{N-n} \sum_{s=1}^{N-n} F_{ab}^2(n,s) \right]^{1/2}$$

Recently, a cross-correlation index $\rho(n)$ has been introduced in order to make the interpretation of DCCA clearer (Zebende, 2011),

$$\rho(n) = \frac{F_{ab}^2(n)}{F_{aa}(n)F_{bb}(n)}$$

Similar to the standard cross-correlation coefficient, $\rho(n) = 1.0$ indicates a perfect cross-correlation, while $\rho(n) = -1.0$ means a perfect anti-cross-correlation. If $\rho(n) = 0.0$, there is no cross-correlation between the two time series at the scale n.

`fathon` implements common operations for all the algorithms described above, such as the computation of the fluctuation function and its fit in different ranges, the multifractal spectrum, and the detrended cross-correlation index and its confidence intervals. `fathon` aims at providing a simple and interactive interface to perform multiple analyses on multiple time series. The Python interface is class-based and user-friendly, while the underlying code is written in Cython and C, allowing a fast computation of the algorithms.

`fathon` can be useful to all researchers and data analysts who utilize the methods listed above, especially in the fields of geophysics, economics, and biophysics. Its purpose is to gather all the algorithms together, in order to have a single package that includes different methodologies. They are easy to use in their basic or advanced versions, and the user can easily switch between them. Undoubtedly, codes for each method have already been individually developed, see for instance mf-dfa, dfa, and detrended_fluctuation. These packages implement only a single algorithm of `fathon` and are less flexible. The most comprehensive package is the Ihlen's Matlab package for multifractal detrended fluctuation analysis (Ihlen, 2012). It does not cover DCCA and the computation of the cross-correlation index, is less flexible, and some functions' parameters are not considered, like the possibility of calculating the fluctuation function only forward or both forward and backward. The `fathon` package is intended to be more complete, gathering different algorithms under a single package. Moreover, `fathon` offers more functions and more flexibility in parameters choice.

References

Fraedrich, K., & Blender, R. (2003). Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Physical Review Letters*, *90*(10), 108501. doi:http://dx.doi.org/10.1103/PhysRevLett.90.108501

Ihlen, E. A. (2012). Introduction to multifractal detrended fluctuation analysis in Matlab. *Fractal analysis*, *3*, 97. doi:http://dx.doi.org/10.3389/fphys.2012.00141

Ivanov, P. C., Yuen, A., Podobnik, B., & Lee, Y. (2004). Common scaling patterns in intertrade times of us stocks. *Physical Review E*, *69*(5), 056107. doi:http://dx.doi.org/10.1103/PhysRevE.69.056107

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Bunde, A., Havlin, S., & Stanley, H. E. (2003). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, *316*(1), 87–114. doi:http://dx.doi.org/10.1016/S0378-4371(02)01383-3

Liu, Y., Gopikrishnan, P., & Stanley, H. E. (1999). Statistical properties of the volatility of price fluctuations. *Physical Review E*, *60*(2), 1390. doi:http://dx.doi.org/10.1103/PhysRevE.60.1390

Peng, C. K., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Simons, M., & Stanley, H. E. (1993). Finite-size effects on long-range correlations: Implications for analyzing DNA sequences. *Physical Review E*, *47*(5), 3730. doi:http://dx.doi.org/10.1103/PhysRevE.47.3730

Podobnik, B., & Stanley, H. E. (2008). Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Physical Review Letters*, *100*(8), 084102. doi:http://dx.doi.org/10.1103/PhysRevLett.100.084102

Varotsos, C., Assimakopoulos, M. N., & Efstathiou, M. (2007). Long-term memory effect in the atmospheric CO_2 concentration at Mauna Loa. *Atmospheric Chemistry and Physics*,

7(3), 629–634. doi:https://doi.org/10.5194/acp-7-629-2007

Zebende, G. F. (2011). DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A: Statistical Mechanics and its Applications*, *390*(4), 614–618. doi:http://dx.doi.org/10.1016/j.physa.2010.10.022