A note on the performance of a wavelet-based multiscale estimator

Leandro Santi, Luis Marrone

Universidad de Buenos Aires,
Paseo Colón 850, Ciudad de Buenos Aires, Argentina
lsanti@fi.uba.ar
lmarro@fi.uba.ar
lmarrone@linti.unlp.edu.ar
http://www.fi.uba.ar/

Abstract. Quantitative evidence is presented in order to study the performance of a wavelet-based, second order estimator of relevance to multiscaling phenomena such as telecommunications traffic. Special attention is given to the behavior of the confidence intervals of the scaling exponent under Gaussian assumptions with synthetic long-range dependent trajectories, allowing for a straightforward contrast with the theoretical results presented in the bibliography. The bias-variance tradeoff decision behind the choice of the onset of the scaling region is also revisited, along with its influence on the confidence of the intervals of the exponent.

Keywords: Wavelet Decomposition, Long Range Dependence, Scale Invariance, Multiscale Estimation, Fractional Gaussian Noise, Telecommunications Traffic.

1 Introduction

Stochastic processes with scale invariant features are central to the study of a vast collection of natural and human constructions including hydrology, turbulence, biology, telecommunications traffic and computers. In essence, the concept of scale invariance refers to the absence of a particular, privileged, or reference scale: thus, the paradigm is built around the relations among scales, perhaps within a given range or interval in which the scaling behavior manifests itself.

One of the recurring difficulties in the analysis of scale invariant series is related to the poor performance of the traditional estimation techniques, which are often constructed under stationary and/or short memory hypotheses. Wavelet-based estimation, on the other hand, can often eliminate or mitigate some of these shortcomings, enabling for computationally efficient implementations. It is therefore of particular interest to have appropriate tools for the study of these phenomena.

This article complements earlier work from the authors, where a self-contained implementation of the wavelet estimator was built for studying the statistical

behavior of the estimator: the performance of the scaling onset detection algorithm and the effectiveness of confidence intervals involved in measurements of the scaling exponent. Both of them are discussed below.

The outline of this article is as follows: section 3 provides a brief introduction to two of the most commonly used scale invariant processes, fractional Brownian motion and its increment process, the fractional Gaussian noise. Special attention is given to the properties emerging from the wavelet analysis of these, in order to grasp some of the theory behind the estimation tool. In section 4 we present evidence related to the statistical performance of the estimator, then conclude in section 5.

2 Scope and Objectives

The goal of this work is to provide quantitative evidence related to certain aspects of the behavior of the wavelet-based estimator described in [3]. More specifically, we target the study of the performance of the confidence levels of $\hat{\alpha}$, the scaling exponent estimator with a fully-automated configuration using the transition detection algorithm originally presented in [7].

Additionally, we have restricted the scope of our analysis to the (ideal) case of long memory processes such as fractional Gaussian noise. This has been done on purpose in order to be able to contrast our results with the vast amount of predictions given in the literature.

3 Background

In this section we start with a brief introduction to the main concepts supporting the estimation of long range dependent processes using wavelets. For convenience we have adopted, wherever possible, the naming convention commonly used in the bibliography [3] [7].

3.1 LRD Trajectories

A process $X = \{X(t), t \in \mathbb{R}\}$, wide-sense stationary, presents long-range dependencies or LRD if its correlation takes the following form:

$$c_{XX}(t+\tau, t) \sim c_r |\tau|^{-\beta}, \quad \tau \to \infty \tag{1}$$

where c_r is a constant and $0 < \beta < 1$. Equivalently, the power spectral density satisfies:

$$\Gamma_X(f) \sim c_f |f|^{-\gamma}, \quad f \to 0 \tag{2}$$

with $0 < \gamma = 1 - \beta < 1$ and $c_f = 2(2\pi)\sin((1-\gamma)\pi/2)\Gamma(\gamma)c_r$, where Γ is the usual Gamma function.

The behavior described by the formula (1) determines a very slow decay rate of the autocorrelation function, which is quantified by the exponent β.

Indeed, this decay is so slow that the correlation function is not summable: LRD processes thus have *long memory*, because samples relatively distant from each other have stronger statistical influences. In practice, the presence of the long memory effect is often detrimental to the performance of the analytical tools, for example worsening the convergence of algorithms that employ empirical estimators of higher order moments [10].

As we can see from expressions (1) and (2), the LRD effect is by definition an asymptotic behavior, so it does not provide a clear cutoff frequency after which the dependencies become apparent. As discussed below, this feature (actually the lack of such) has a direct impact in the design of the algorithm for automatic detection of the scaling region.

3.2 On the fBm and fGn processes

Let us now briefly present some features of the family of Gaussian process we use throughout this work: the (normalized) fractional Brownian motion which we refer to as fBm, and its increment process called fractional Gaussian noise, commonly abbreviated as fGn.

About fBm processes. It is perhaps the family of continuous-time processes most commonly used for modeling self-similar processes, characterized by a single parameter $0 < H < 1$, the Hurst exponent.

By definition, a fBm with parameter H is a self-similar Gaussian process, beginning at the origin (i.e. $B_H(0) = 0$) and has zero mean and stationary increments. Additionally, the normalized autocovariance function takes the following form:

$$c_{BB}(t, t') = \frac{1}{2}(|t|^{2H} + |t'|^{2H} - |t - t'|^{2H})$$

This processes can be considered as a generalization of the traditional Brownian movement; i.e., unlike the latter, the fBm incorporates the possibility of displaying non-trivial amounts of self-correlation in its increment process:

$$\mathbb{E}\,(B_H(t+\tau) - B_H(t))(B_H(t) - B_H(t-\tau)) = (2^{2H-1} - 1)|\tau|^{2H}$$

Thus, the special case $H = \frac{1}{2}$ is actually an ordinary Brownian motion, since it provides the necessary (i.e. full) degree of independence.

About fGn processes. Fractional Gaussian noise is defined as the increment of its associated fBm process:

$$G_{H,\delta} \triangleq \frac{1}{\delta}(B_H(t+\delta) - B_H(t))$$

which by definition determines a family of stationary processes. In case of the normalized fBm, the resulting autocorrelation function takes the following form:

$$c_{GG}(t+\tau,t) = \frac{1}{2}|\tau|^{2H}\left(|1+\delta/\tau|^{2H} + |1-\delta/\tau|^{2H} - 2\right)$$

In addition, it can be proven that it verifies:

$$c_{GG}(t+\tau,t) \sim H(2H-1)|\delta|^2|\tau|^{2(H-1)}, \quad \tau \gg \delta \tag{3}$$

That is, fGn processes with $H > \frac{1}{2}$ exhibit long memory since its correlation function shows a sublinear decay rate with exponent $\beta = 2(H-1)$.

3.3 Wavelet analysis of LRD processes

In section 3.1 we briefly discussed an important, well-known consequence of the long memory feature, i.e. the practical issues that complicate the estimation processes using traditional analysis tools. On the other hand, it has been shown [5] [6] that the wavelet decomposition of an LRD process can mitigate these effects: under certain conditions, the transformed processes $d_x(j,k)$ do not exhibit long range dependencies, but instead residual short-range correlations (i.e. SRD).

An additional property from the wavelet decomposition is that the power law that describes the second order behavior, formulae (1) and (2), is also present in the transformed domain: $\mathbb{E}\,|d_X(j,k)|^2 \sim 2^{j\gamma} c_f C(\gamma,\psi_0)\ j \to +\infty$, where c_f is a constant, $C(\gamma,\psi_0) = \int |f|^{-\gamma} |\mathbf{\Psi_0}(f)|^2\, df$, and the symbol ψ_0 identifies the mother wavelet (for a sketch of the proof of this result, see for example [5] or [6]).

Since these behaviors can be generalized to an ample family of scale invariant processes, to summarize the above and following the guidelines used in [4] the properties are formalized as follows: let $X = X(t)$ be a continuous time random process with scaling exponent α. Then,

- **P1**: the process $\{d_X(j,k), k \in \mathbb{Z}\}$ is stationary with $n_\psi \geq (\alpha-1)/2$, and the variance $\mathbb{E}\,|d_X(j,k)|^2$ reproduces the power law of the original process under a given range of octaves: that is, for $j_1 \leq j \leq j_2$ we have $\mathbb{E}\,|d_X(j,k)|^2 = 2^{j\alpha} c_f C(\alpha,\psi_0)$. Here, n_ψ stands for the amount of vanishing moments of ψ_0, the mother wavelet; and $c_f C$ depends on the actual form of the stochastic process $X(t)$.
- **P2**: the details $\{d_X(j,k), k \in \mathbb{Z}\}$ conform a stationary process, and do not exhibit long-range dependencies but instead local short-term correlations, i.e. d_X is an SRD process on the condition that $n_\psi \geq \alpha/2$.

These two properties are central for the conception of the wavelet based estimation tool, which we discuss below.

3.4 Estimation of the scaling exponent

Property **P1** suggests that it is possible to estimate the scaling exponent α using the fact that $\log \mathbb{E}\,|d_X(j,k)|^2 \sim j\alpha + \log c_f C$. Of course, this would require

estimating the argument of the logarithm from a finite-length realization (n samples) of the underlying stochastic process $X(t)$. Following this line of reasoning, a semi-parametric wavelet-based estimator has been proposed in [3]: more specifically,

$$y_j = \log_2(1/n_j \sum_k |d_X(j,k)|^2) - g_j \tag{4}$$

$$\hat{\alpha} = \sum_{j_1}^{j_2} w_j y_j \tag{5}$$

$$\hat{b} = \sum_{j_1}^{j_2} v_j y_j \tag{6}$$

where $n_j = \Theta(n/2^j)$ stands for the amount of available detail coefficients in octave j, and the interval $[j_1, j_2]$ represents the range of octaves spanning the scaling phenomenon, and so the region of interest for the linear regression. The deterministic quantities $g_j = \ominus(n_j/2)/\ln 2 - \log(n_j/2)$ are meant to correct the bias related to the fact that $\log \mathbb{E}\,|d_X(j,k)|^2 \neq \mathbb{E} \log |d_X(j,k)|^2$ under Gaussian assumptions. Here, $\ominus(x) = \Gamma'(x)/\Gamma(x)$ represents the psi or digamma function. Coefficients w_j and v_j come from the standard weighted linear regression technique over the set of points (j, y_j) spanning $j_1 \leq j \leq j_2$, with weights $1/\sigma_j^2$, the inverse of the variances of the y_j in order to provide more influence to the points with a smaller amount of variations.

Due to the lack of a closed expression for the variances of the y_j in the general case, in practice it is common to use the following approximation which is obtained under Gaussian and strict decorrelation hypothesis in the wavelet domain:

$$\sigma_j^2 = \operatorname{var} y_j = \zeta(2, n_j/2)/\ln 2$$

where $\zeta(s,q)$ stands for the Hurwitz zeta function. In practice, this equation is used for both the computation of the linear regression, and to construct confidence intervals for y_j and $\hat{\alpha}$ under Gaussian assumption.

3.5 Automatic detection of the scaling onset

From the perspective of the estimation of scaling, the range of scales over which a scaling regime could be present is *a priori* unknown, and hence needs to be identified within the context of the estimation process. Indeed, the detection of such range, in the log-scale diagram, consists in the identification of the alignment region(s) whose endpoints are j_1 and j_2. Such problem, as stated, cannot possess a generic solution due to the fact that scaling properties are often of asymptotic nature, lacking a clear frontier that would enable us to precisely identify of the frontiers as can be seen from equations (1) and (2).

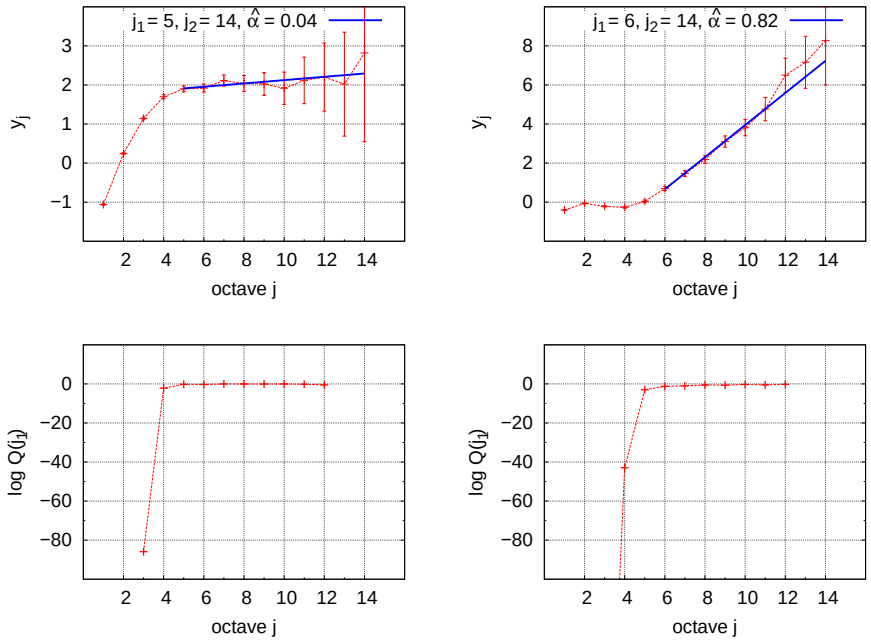

Fig. 1. Behavior of $\hat{Q}(j_1)$ in the transition scale j_1^* detection algorithm for a couple of FARIMA$(1, d, 1)$ realizations. (Left, above) log-scale diagram of a realization of 100,000 samples of a stochastic process with $(\phi, d, \theta) = (0.5,\ 0,\ 0)$. Note the outcome from the algorithm, $j_1^* = 5$. (Left, below) representation of the amount of goodness of fit as a function of the onset octave, where we observe a rapid improvement zone beginning in octave 3. (Right, above) log-scale diagram for 100,000 samples of $(\phi, d, \theta) = (0.3,\ 0.4,\ 0.7)$. (Right, below) evolution of the $\hat{Q}$ metric.

In practice, the selection of the cutoff scales can be done using a chi-squared heuristic techique to measure the goodness of fit of the linear regression [7]. The main idea is to monitor the values of $\hat{Q}$, the statistic for measuring the quality of the fit as a function of the endpoints of the alignment zone, j_1 and j_2:

$$\hat{V}(j_1, j_2) = \sum_{j=j_1}^{j_2} \frac{(y_j - \hat{\alpha} j - \hat{b})^2}{\sigma_j^2} \tag{7}$$

$$\hat{Q}(j_1, j_2) = 1 - F_{J-2}(\hat{V}(j_1)) \tag{8}$$

where F_m is the cumulative probability distribution function of a chi-squared variable with m degrees of freedom, and $J = j_2 - j_1 + 1$ represents the length of the alignment interval. In these expressions, the $J - 2$ degrees of freedom result from considering J points with two restrictions: the determination of the slope

and abscissa. As we can see in (7), the $\hat{V}$ statistic is a measure of the amount of error between the data and the model, in the mean squared sense. This figure is then input into $\hat{Q}$, generating an index in $[0, 1]$ that enables for the determination of the goodness of the fit, since values of $\hat{Q}$ near 1 indicate the adequacy of the model, in opposition improper cases where the index approaches 0.

In practice, in the context of stochastic processes with long range dependencies we have $j_2 \to +\infty$, and these quantities only need to be monitored as a function of j_1 to obtain the threshold indicator of the beginning of the alignment region. More specifically, the logic of the detection algorithm for LRD data has been defined as follows.

Transition scale j_1^* detection algorithm for long memory processes

1. Determine the range $[1, j_{ND}]$ in which $\hat{Q}(j_1)$ is non-decreasing. If $j_{ND} = 1$ then set $j_1^* = 1$, else
2. Calculate the improvement ratios $r_j = \hat{Q}(j)/\hat{Q}(j-1)$ for each $j \in [2, j_{ND}]$. Select a rapid improvement factor fac (defaults to 10) and find the largest j such that $r_j > fac$. If there is such j, set j_1^* equal to it; else set $j_1^* = 1$. Finally, add one unit to the value of j_1^*.

As stated in [7], the design of this algorithm is centered around the premise of minimizing the mean squared error $\mathbb{E}\,(\hat{\alpha} - \alpha)^2$. By using these metrics, the heuristic technique allows us to compute an approximation of the onset octave for the LRD case. The reasoning presented in that work indicates good performance results in numerous aspects and simulations; however, as we shall see below, an (sub)optimal selection of the onset octave, in the MSE sense, can be detrimental to the actual confidence levels of the experiments.

Example. Figure 1 shows the evolution of the goodness of fit metric in the context of two FARIMA(1, d, 1) realizations. In both cases we observe a remarkable increase of $\hat{Q}(j_1)$ in the lower octave range, dubbed *improvement zone* in the references. After this transitional area we notice the presence of the *alignment intervals*: it is here where the long memory feature is present, leading to a more stable behavior of the metric along the remaining portion of the octave range.

4 Discussion

We now present the results of the simulations carried out in the context of this work, which aim to test the performance of scaling onset detection algorithms for LRD trajectories. We commence with a brief introduction to our simulation environment, which has been originally conceived as a test tool for embryonic versions of our own implementation of the wavelet estimator.

4.1 Simulation Environment

One of the key aspects of our software environment is that is has been structured for reproducible analysis, enabling other parties (colleagues, etc.) to obtain the

same results by running their own copy of the software. In order to cope with this, we have adopted some of the advice in [1], in the form of a single master script that builds all the analysis results from input data, which is readily available as well. The source code of the experiments supporting this article is available in [9].

Recall that the statistical properties of the wavelet estimator are derived under Gaussian assumptions: therefore, the experiments below use synthetic samples of fractional Gaussian noise in order to evaluate some performance aspects of the tool on an ideal environment. The meaningful MRA initialization technique [8] is used for proper second order evaluation of discrete data. Unless explicitly noted, in the following simulations we sample the (α, n) space using three different values of n, the input trajectory length: 1,000, 10,000, and 100,000. The `Math::Random::Brownian` [2] implementation of the (exact) Davies and Harte and (approximate) Dieker-Mandjes synthesis algorithms were used to generate two sample sets covering the same portion of the (α, n) plane.

4.2 Confidence intervals for $\hat{\alpha}$

From equation (5) we can see that $\hat{\alpha}$ is a linear combination of the y_j, which have been shown to be asymptotically normally distributed under Gaussian and strict decorrelation assumptions [4],

$$\log \mu_j \overset{d}{\sim} N(j\alpha + \log(c_f C), \frac{2(\log e)^2}{n_j})$$

Because of this, $\hat{\alpha}$ itself can be considered as approximately Gaussian distributed, with variance $\sigma^2_{\hat{\alpha}} = \sum \sigma_j^2 w_j^2$. In practice, confidence intervals for the estimations of the scaling exponent are computed using these results.

Performance evaluation. Empirical estimations of the confidence coefficient of $\hat{\alpha}$ are carried out for different values of (α, n), where n is the amount of samples per synthetic fGn trajectory. Each pair is associated with 1,000 realizations of a fGn process with said length and scaling exponent: by applying the `scalens` software to these [9], we can estimate the confidence level of the experiment by counting the amount of times the interval contains the actual value of the scaling exponent. In every case, the intervals have been generated with the confidence parameter set to 90%.

It is evident from table 1 that a non-trivial amount of observations of the confidence coefficient are well under the value of 90% that was configured to run the experiment. In other words, the intervals generated by the estimation tool should contain, on average, the true value of α 90% of the times, however evidence suggests that this is not the case.

As we explain below, this behavior can be understood by considering the combined effects of the algorithm for automatic detection of the onset of the alignment region on one hand, and the asymptotic nature of the long memory effect on the other.

α	sample length n								
	1k			10k			100k		
	j_1^{MSE}	j_1^*	$\hat{c}c$	j_1^{MSE}	j_1^*	$\hat{c}c$	j_1^{MSE}	j_1^*	$\hat{c}c$
0.0	**2**	**2 (60%)**	76%	2	3 (98%)	89%	2	3 (99%)	91%
0.1	**2**	**2 (59%)**	**75%**	2	3 (98%)	89%	3	3 (95%)	88%
0.2	**2**	**2 (59%)**	**75%**	2	3 (97%)	90%	**3**	**3 (90%)**	**82%**
0.3	**2**	**2 (60%)**	**74%**	2	3 (97%)	88%	**3**	**3 (82%)**	**79%**
0.4	**2**	**2 (59%)**	**72%**	3	3 (97%)	88%	**3**	**3 (70%)**	**76%**
0.5	**2**	**2 (58%)**	**70%**	3	3 (97%)	87%	**3**	**3 (63%)**	**73%**
0.6	**2**	**2 (61%)**	**71%**	3	3 (95%)	87%	**4**	**3 (54%)**	**75%**
0.7	**2**	**2 (58%)**	**69%**	3	3 (96%)	88%	**4**	**4 (54%)**	**75%**
0.8	**2**	**2 (58%)**	**67%**	3	3 (97%)	87%	**4**	**4 (60%)**	**73%**
0.9	**2**	**2 (57%)**	**67%**	3	3 (95%)	86%	**4**	**4 (64%)**	**75%**

Table 1. Performance measurement of the $\hat{\alpha}$ statistic with automatic detection of the onset of the scaling regime. Columns j_1^{MSE} indicate the optimum choice according to the minimum mean squared error criterion, whereas j_1^* represent the octave most frequently selected by the algorithm for a given pair (α, n). The figure in parentheses is the rate of selection of said j_1^*, indicating suboptimal performance. Differences in 5 or more points between the observed confidence coefficient ($\hat{c}c$) and actual the simulation parameter (90%) are highlighted in **bold**.

Effect of the transition scale j_1^* detection algorithm

Recall from section 3.5 that this algorithm, of heuristic nature, has been conceived to detect the initial (i.e., lower) octave of the region in the logscale diagram where the LRD phenomenon is present: this is done by defining the onset scale as the one that gives minimum mean squared error defined as $\mathbb{E}\,(\hat{\alpha} - \alpha)^2$.

From the previous discussion, it is natural to ask ourselves whether the observed effectiveness of the confidence intervals is somehow related to the performance of the detection algorithm. For this reason, we have instrumented the simulation environment to obtain the actual values of j_1^* chosen by the heuristics: as we can see in table 1, in the case of trajectory of 1,000 samples the detection rate is close to 60%, indicating suboptimal performance under the MSE criterion. Realizations of 10,000 units of length exhibit an outcome octave $j_1^* = 3$ more than 95% of the times; however we note that for $0.0 \leq \alpha \leq 0.3$ the optimal threshold is $j_1^{\mathrm{MSE}} = 2$ instead of 3.

Reevaluation with optimal choice $j_1 = j_1^{\mathrm{MSE}}$. We now redo the experiment replacing the heuristics for choosing j_1 with the optimal selection strategy under MSE criterion, using the appropriate (pre-computed) values of j_1^{MSE} for each point in the sampling of the (α, n) plane. A quick glance at table 2 reveals a partial improvement in the performance of some of the confidence intervals.

Nevertheless, these results suggest that an optimal choice of the onset scale, under minimum MSE criterion, does not necessary ensure proper performance

α	sample length n								
	1k			10k			100k		
	j_1^{MSE}	j_1	$\hat{cc}$	j_1^{MSE}	j_1	$\hat{cc}$	j_1^{MSE}	j_1	$\hat{cc}$
0.0	2	2 (100%)	89%	2	2 (100%)	88%	**2**	**2 (100%)**	**74%**
0.1	2	2 (100%)	90%	**2**	**2 (100%)**	**84%**	3	3 (100%)	88%
0.2	2	2 (100%)	90%	**2**	**2 (100%)**	**79%**	**3**	**3 (100%)**	**83%**
0.3	2	2 (100%)	88%	**2**	**2 (100%)**	**71%**	**3**	**3 (100%)**	**80%**
0.4	2	2 (100%)	88%	3	3 (100%)	89%	**3**	**3 (100%)**	**78%**
0.5	2	2 (100%)	85%	3	3 (100%)	88%	**3**	**3 (100%)**	**74%**
0.6	2	2 (100%)	85%	3	3 (100%)	88%	4	4 (100%)	91%
0.7	2	2 (100%)	86%	3	3 (100%)	88%	4	4 (100%)	90%
0.8	**2**	**2 (100%)**	**84%**	3	3 (100%)	87%	4	4 (100%)	88%
0.9	**2**	**2 (100%)**	**84%**	3	3 (100%)	86%	4	4 (100%)	90%

Table 2. Reevaluation of the statistical performance of $\hat{\alpha}$, disabling the heuristics for automatic detection of the scaling onset. Here, we set j_1 to j_1^{MSE}, the optimal choice under MSE criterion instead of the algorithm presented in section 3.5. Compare this with the data from table 1. Again, differences in 5 or more points between the observed confidence coefficient ($\hat{cc}$) and the simulation parameter (90%) are highlighted in **bold**.

of the confidence intervals. As we shall see below, this behavior can be explained by the bias-variance tradeoff relation inherently present in the estimation tool.

Effect of the scaling onset in the performance of the CIs

As mentioned in section 3.4, the estimation of the scaling exponent is computed using the detail coefficients from the DWT operation, $d_X(j,\cdot)$. In particular, we observed an exponential reduction of the amount of useful information at each octave, $n_j = \Theta(n/2^j)$. Because of this, in the context of our discussion the variances of y_j and $\hat{\alpha}$ increase monotonically with j and j_1 respectively.

On the other hand, due to the asymptotic nature of the LRD behavior, equations (1) and (2), we can expect higher deviations from the linear trend in the lowest portion of the wavelet analysis, mainly because of the high frequency content of those octaves. Therefore, it is reasonable to expect a higher amount of estimation bias with low values of the onset octave.

This tradeoff relationship between the mean squared error $\mathbb{E}\,(\hat{\alpha}-\alpha)^2$ and j_1 can be observed in the upper region of figure 2: there, empirical estimations of the mean squared errors are shown for different values of α and trajectory sizes. In every case, we observe the presence of an optimal scale in the sense of the MSE criterion, j_1^{MSE}, to the left of which the error is dominated by estimation bias; whereas variance has stronger influence on the opposite direction of j.

An alternative view is presented in the lower portion of the same figure, where we can clearly see the early detrimental effect of j_1 on the estimation bias, and a progressive increase on the variance of $\hat{\alpha}$ with higher values of the onset threshold, in the form of wider histograms for $\hat{\alpha}$.

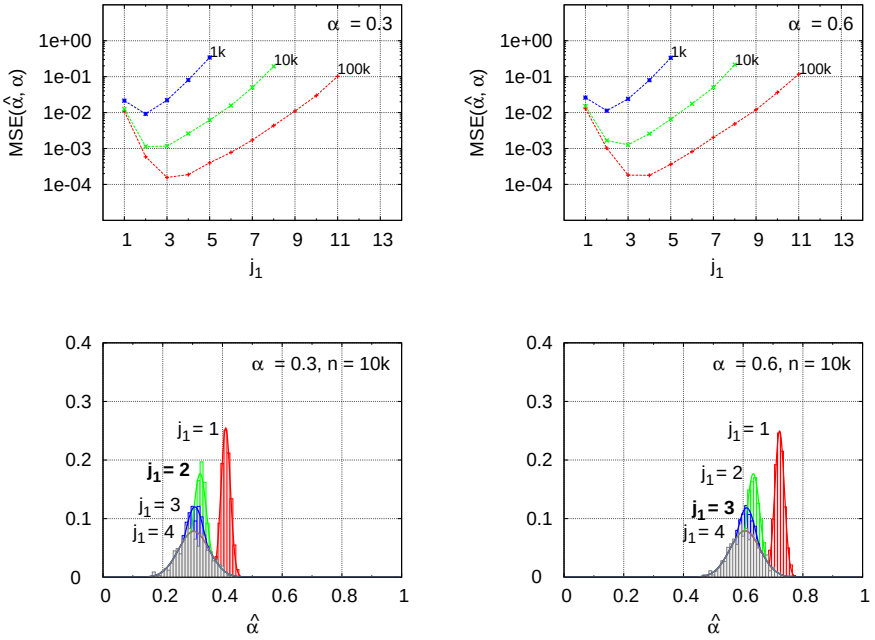

Fig. 2. Influence of the scaling onset parameter, j_1, on the mean squared error. The uppermost plots show an estimation of the MSE as a function of j_1. An alternative view is shown below, in the form of the observed histograms for $\hat{\alpha}$ as a function of j_1. Highlighted in **bold** are the optimal values of the onset scale under the minimum MSE criterion. It is interesting to contrast these with the informations from table 2. In general, and because the LRD power law behavior of fGn of asymptotic nature, we observe lower variance and higher estimation bias in the small-scale region, due to departures from the linear progression. On the opposite direction of j_1 we find increasingly higher variance and wider histograms for $\hat{\alpha}$, due to lower amount of available detail coefficients in the highest octaves of the wavelet analysis.

5 Conclusions

We have presented quantitative evidence aiming to understand the statistical performance of a wavelet-based estimator for scale invariant phenomena. In particular, we have payed attention to the behavior of the scaling exponent statistic, $\hat{\alpha}$, which is of special relevance in the study of computer networks.

This work is actually a detachment from a previous initiative by the authors, where we worked on an independent implementation of the wavelet estimator and contrasted the confidence levels of the intervals. The context of our analysis has been limited to the study of the performance of the tool under ideal conditions, Gaussian LRD trajectories such as fGn. This has been done on purpose in order to contrast our observations with the predictions made in the bibliography.

Within the context of our simulation environment, we have observed significant deviations in the statistical performance of the confidence coefficient of the scaling exponent estimations. This can be attributed to the suboptimal behavior of the algorithm for automatic detection of the scaling onset on one hand, which is constructed around the premise of minimizing the expectation of the mean squared error. On the other hand, we argue that such goal can be detrimental to the actual confidence levels of the intervals, because the inherent bias-variance tradeoff seeking to minimize the MSE can introduce a non-trivial amount of displacement in the actual position of the intervals. Since the confidence regions are built on Gaussian assumptions and are thus symmetrical, we conclude that an optimal strategy for choosing j_1, in the minimum MSE sense, cannot provide sufficient control over the estimation bias for meaningful confidence levels.

The reasoning presented along the lines of our work allows us to argue that, when using the wavelet-based estimation tool to study scaling exponents, an improper detection of the onset of the scaling region can have a significant detrimental effect in the determination of the confidence levels, which, in turn, can compromise the estimation and statistical inference processes.

References

1. Vern Paxson. “Strategies for Sound Internet Measurement”. ACM Internet Measurement Conference, October 2004.
2. Walter Szeliga. `Math::Random::Brownian` Perl module for generating Fractional Brownian and Gaussian Noise, version 0.03. `http://search.cpan.org/~wmszeliga/Math-Random-Brownian-0.03/`
3. Darryl Veitch, Patrice Abry, “A wavelet based joint estimator for the parameters of long-range dependence” (1999). `http://www.cubinlab.ee.unimelb.edu.au/~darryl/Publications/A3.pdf`
4. Patrice Abry, Patrick Flandrin, Murad S. Taqqu, Darryl Veitch, “Wavelets for the analysis, estimation, and synthesis of scaling data”, pp.39-88, Self-Similar Network Traffic and Performance Evaluation, K. Park and W. Willinger, Wiley (2000).
5. Patrick Flandrin, Paulo Goncalvès and Patrice Abry. “Scale invariance and wavelets”. Pp. 71-102, Scaling, Fractals and Wavelets, Digital Signal and Image Processing Series. ISBN 9781848210721. ISTE-Wiley (2009).
6. Patrice Abry, Patrick Flandrin, Murad S. Taqq, Darryl Veitch. “Self-similarity and long-range dependence through the wavelet lens”. Pp. 527-556, Theory and Applications of Long-Range Dependence (2003).
7. Darryl Veitch, Patrice Abry, Murad S. Taqqu. “On the Automatic Selection of the Onset of Scaling” (2003). `http://www.cubinlab.ee.unimelb.edu.au/~darryl/Publications/choosej1_camera.pdf`
8. Darryl Veitch, Murad S. Taqqu, Patrice Abry, “Meaningful MRA initialization for discrete time series”. Pp. 1971-1983, Vol.8, Signal Processing, Elsevier Science (2000).
9. Leandro Santi, Luis Marrone. Source code of the experiments presented in this article. `http://webs.sinectis.com.ar/tesis/jaiio2012/`.
10. Jan Beran. Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability 61. Chapman & Hall/CRC (1994).