Analytical Determination of Fractal Structure in Stochastic Time Series

Fermín Moscoso del Prado Martín

Laboratoire de Psychologie Cognitive (UMR–6146)
CNRS & Aix–Marseille Université I, Marseille, France

(Dated: November 13, 2009)

Abstract

Current methods for determining whether a time series exhibits fractal structure (FS) rely on subjective assessments on estimators of the Hurst exponent (H). Here, I introduce the Bayesian Assessment of Scaling, an analytical framework for drawing objective and accurate inferences on the FS of time series. The technique exploits the scaling property of the diffusion associated to a time series. The resulting criterion is simple to compute and represents an accurate characterization of the evidence supporting different hypotheses on the scaling regime of a time series. Additionally, a closed-form Maximum Likelihood estimator of H is derived from the criterion, and this estimator outperforms the best available estimators.
In this study I introduce a method for deciding whether a time series exhibits fractal scaling (FS; [1, 2, 3]) in a simple, objective, and accurate way. Determining the presence or absence of FS in a time series is an important question for making inferences about the nature of underlying system that generated it. Important conclusions concerning the nature of systems in a wide range of domains rest on assessing whether – and, if so, the extent to which – observed time series exhibit FS. Usually, this is addressed by estimating whether the value of the Hurst exponent (H; [3]) of the series is equal to or different from 1/2. Often, some additional verification is performed by looking for the consistency of multiple estimators, but the criteria for establishing FS retains a fair deal of subjectivity (see, e.g., [4] for a heuristic methodology for combining estimators). However, estimating the value of a parameter does not constitute a valid assessment of FS. Rather, in order to argue that a series is fractal (or, more generally, its H exponent ranges within some hypothesized interval) an explicit contrast between different hypotheses about H is required (see [5], for a discussion distinguishing parameter estimation from hypothesis contrast). The few approaches that consider some hypothesis testing rely on approximative techniques with multiple adjustable parameters and strong assumptions on the underlying model [6, 7, 8, 9].

The elements in a time series can be thought of as the fluctuations of a particle following a diffusion process [10]. From this perspective, for any process \( X(\tau) \) generating values with a stationary distribution \( f_X \), one can consider the diffusion process \( Y(\tau) \) associated to \( X(\tau) \), which is defined by \( dY(\tau)/d\tau = X(\tau) \). Since what is normally available is not the continuous evolution of \( X \), but rather a discretely sampled time series \( x = \{x(1), \ldots, x(N)\} \), one can instead define the corresponding random walk \( y \) on discrete time \( k \)

\[
y(k) = \sum_{i=1}^{k} x(i), \quad k = 1, \ldots, N.
\]  

(1)

If the generating process \( X \) is stationary, the scaling property [11] indicates that, for large times, the probability distribution of the walk approaches

\[
p(y|k, \delta) = \frac{1}{k^\delta} f_L \left( \frac{y}{k^\delta} \right),
\]  

(2)

where \( \delta \) is the scaling exponent associated with the time series and \( f_L \) is a limiting probability density function. The scaling exponent is closely related to the H exponent [11, 12]. In the cases when \( f_X \) corresponds to the normal distribution – fractional Gaussian noise (fGn; [2, 3]) – \( \delta \) is equivalent to H.
Let \( x = \{x(1), \ldots, x(N)\} \) be a stationary time series of \( N \) samples, and \( y_k \) the \( k \)-th order running sum of \( x \), \( y_k(i) = \sum_{j=i}^{i+k-1} x(j) \). The running sums can be taken as samples at time \( k \) of an ensemble of \( N - 1 \) trajectories from the diffusion process (see Refs. [12, 13] and references therein). For large summing orders (i.e., diffusion times), the scaling condition should apply. The simplest case is when two \textit{a priori} candidate values for the scaling exponent, \( \delta_1 \) and \( \delta_2 \) are available. One’s task is to determine which of the two scaling regimes better describes the complexity of the observed series \( x \). This amounts to determining the Bayes factor \( \lambda_{1,2} = \ln[p(H_1|x)/p(H_2|x)] \) between two simple hypotheses \( H_1 \equiv \delta = \delta_1 \), and \( H_2 \equiv \delta = \delta_2 \). If \( \lambda_{1,2} > 0 \), then the data support hypothesis \( H_1 \) more strongly than hypothesis \( H_2 \), and the reverse holds if \( \lambda_{1,2} < 0 \). By Bayes’ Theorem, \( \lambda_{1,2} \) can be decomposed into the sum of the \textit{a priori} information \( \alpha_{1,2} \) (obtained by theoretical predictions or previous observations of the same process) and the evidence provided by the data \( \eta_{1,2}(x) \)

\[
\lambda_{1,2}(x) = \ln \frac{p(H_1)}{p(H_2)} + \ln \frac{p(x|H_1)}{p(x|H_2)} = \alpha_{1,2} + \eta_{1,2}(x). \tag{3}
\]

Exploiting the scaling property in Eq. 2, the likelihood above can be expressed in terms of the \( k \)-th running sum \( y_k \). When doing this, one should consider that the overlap between running sum windows introduces much redundancy in the combined evidence. This imposes rescaling it from its length of \( N - k + 1 \) elements, to its maximum effective length, \( N/k \). The combined evidence is therefore

\[
\eta_{1,2}(x|k) = \frac{N}{(N - k + 1)k} \sum_{i=1}^{N-k+1} \ln \frac{p(y_k(i)|\delta = \delta_1, k)}{p(y_k(i)|\delta = \delta_2, k)}. \tag{4}
\]

Combining the scaling condition of Eq. 2 with the evidence of Eq. 4 results in

\[
\eta_{1,2}(x|k) = \frac{N}{(N - k + 1)k} \sum_{i=1}^{N-k+1} \ln \frac{f_L [y_k(i)/k^{\delta_1}]}{f_L [y_k(i)/k^{\delta_2}]} + \ldots + \frac{N}{k} (\delta_2 - \delta_1) \ln k. \tag{5}
\]

For a time series of length \( N \), there are \( N - 1 \) summing orders than can be investigated. The combined evidence across all orders corresponds to the weighted mean of the evidences across summing orders – note that, as all the running sums were computed from a single realization of the series, a plain sum would exaggerate the evidence –

\[
\eta_{1,2}(x) = \frac{2}{N(N-1)} \sum_{k=2}^{N} (N-k+1) \eta_{1,2}(x|k). \tag{6}
\]
In a more general case when a candidate value for the scaling exponent $\delta$ is not available, but rather a range of possible values is provided, our task consists in comparing the simple hypothesis $H_0 \equiv \delta = \delta_0$ and the composite hypothesis $H_1 \equiv \delta \in [\delta_1, \delta_2]$. As before, the evidence for each hypothesis is expressed in terms of its $k$-order running sum. In the likelihood for the composite Hypothesis $H_1$, $\delta$ is a now a ‘nuisance’ parameter, and the estimation of the likelihood requires integrating it away,

$$p(y|\delta \in [\delta_1, \delta_2], k) = \int_{\delta_1}^{\delta_2} p(y|\delta, k) p(\delta|k) \, d\delta$$  \hspace{1cm} (7a)$$

$$= \frac{1}{\delta_2 - \delta_1} \int_{\delta_1}^{\delta_2} p(y|\delta, k) \, d\delta$$  \hspace{1cm} (7b)$$

Assuming nothing about the value of $\delta$ apart from it falling in the interval $[\delta_1, \delta_2]$, requires the use of an uninformative uniform prior for it, $p(\delta = h|k) = 1/(\delta_2 - \delta_1)$, changing Eq. 7a into Eq. 7b. Combining the scaling condition with the log-likelihoods of the hypotheses, and integrating Eq. 7b, the expression for the evidence provided by summing order $k > 1$ becomes

$$\eta(x|k) = \frac{N}{k} \ln \frac{k^{\delta_0}}{\delta_2 - \delta_1} + \frac{N}{(N-k+1)k} \sum_{i=1}^{N-k+1} \ln \frac{F_L[y(k,i)/k^{\delta_1}]}{y(k,i)f_L[y(k,i)/k^{\delta_0}]} - F_L[y(k,i)/k^{\delta_2}]
\ln k$$  \hspace{1cm} (8)$$

where $F_L(x) = \int_{-\infty}^{x} f_L(z) \, dz$ is the cumulative probability function of $f_L$. Calculating the evidences in Eqs. 5 and 8 requires an analytical expression for the limiting distribution $f_L$. Strictly speaking, the scaling condition applies exactly only in the infinite time limit. The process of convergence to this limit is one by which the original distribution of the data, $f_X$, is gradually being transformed into the limiting distribution $f_L$. Fig. 1 illustrates this point. It plots the departure from normality of the diffusion associated to an fGn as time increases. As expected, one finds a progressive increase in the Kullback-Leibler Divergence (KLD; 11) between $f_X$ and the observed distribution. Even for this simple fGn case, full convergence to $f_L$ is not attained at very large summing orders, but the divergence from normality is rather small.

By the scaling property, if $\delta$ is the hypothesized scaling exponent, the expected log-
likelihood of the sum $y_k$ is \[16\]

$$\langle \ell [y|k, N, \delta] \rangle_{f_L} = \frac{N}{k} \int_{-\infty}^{\infty} \frac{1}{k^\delta} f_L \left( \frac{y}{k^\delta} \right) \ln f_L \left( \frac{y}{k^\delta} \right) dy - \ldots$$

$$- \frac{N}{k} \delta \ln k$$

$$= - \frac{N}{k} \delta \ln k + \frac{N}{k} \int_{-\infty}^{\infty} f_L(z) \ln f_L(z) dz.$$ (9a)

However, using $f_X$ in place of $f_L$ to estimate the likelihood would have resulted in an estimate

$$\langle \ell [y|k, N, \delta] \rangle_{f_X} = - \frac{N}{k} \delta \ln k + \frac{N}{k} \int_{-\infty}^{\infty} f_L(z) \ln f_X(z) dz,$$ (9b)

so that the estimation error would have been

$$\Delta_k = \langle \ell [y|k, N, \delta] \rangle_{f_L} - \langle \ell [y|k, N, \delta] \rangle_{f_X}$$

$$= \frac{N}{k} \int_{-\infty}^{\infty} f_L(z) \ln \frac{f_L(z)}{f_X(z)} dz = \frac{N}{k} K[f_L\|f_X],$$ (11)

where $K[\cdot\|\cdot]$ refers to the KLD between two distributions. Eq. (11) reflects the upper bound of the error in estimating the log-likelihood. In general, for two summing orders $k_1 < k_2$, it should hold that $0 \leq k_1 \Delta(k_1)/N \leq k_2 \Delta(k_2)/N \leq K[f_L\|f_X]$. It is therefore clear that using $f_X$ instead of $f_L$ to estimate the log-likelihood of the running sums should result in an underestimation of the likelihood, with the amount of underestimation bound by the KLD between the two distributions (the KLD between two distributions is always non-negative; [15]). In the fGn case, $f_X$ is the Gaussian distribution and the approximation is exact for $\delta = 1/2$, and otherwise the underestimation increases with $\delta$. Therefore, comparing the approximated log-likelihoods for $\delta > 1/2$ against the hypothesis that $\delta = 1/2$ will be slightly biased in favor of the latter null hypothesis (which is perhaps desirable).

In the case when one has sufficient evidence for the presence of FS, it is often necessary to determine what the scaling exponent is. A simple way to obtain such an estimate, is to find the value $\hat{\delta}$ that maximizes the posterior probability of the exponent given the observed data. If one assumes a uniform uninformative prior on $\delta$, this is equivalent to maximizing the log likelihood of the data given an exponent. Using the scaling condition provides an expression of the log-likelihood for each summing order ($\ell[x|k, \delta]$) as a function of $\delta$. The maxima are found where the first derivative $\partial \ell / \partial \delta$ is zero with a negative second derivative $\partial^2 \ell / \partial \delta^2$. If $\hat{\delta}(k)$ is a maximum of the likelihood then, assuming that that the log-likelihood is peaked around the single maximum $\hat{\delta}(k)$ with a shape that is roughly Gaussian, the variance
of the estimator $\varepsilon(k)^2$ is given by the minus reciprocal of the second derivative evaluated at $\hat{\delta}(k)$. When $f_X$ is a zero-mean Gaussian distribution of variance $\sigma^2$, one finds that the log-likelihood (again approximating $f_L$ by $f_X$) has a single maximum for all values of $k > 1$ with

$$\hat{\delta}(k) = \frac{\ln \left( \sum_{i=1}^{N-k+1} y(k, i)^2 \right) - \ln (\sigma^2 (N - k + 1))}{2 \ln k},$$

(12)

and estimated error

$$\varepsilon(k)^2 = \frac{(N - k + 1) \sigma^2 k^{2\hat{\delta}(k)+1}}{2N \ln^2(k) \sum_{i=1}^{N-k+1} y(k, i)^2}.$$  

(13)

This produces $N-1$ estimators, one for each summing order. From these, the $\hat{\delta}(k)$ with the smallest error $\varepsilon(k)$ is chosen as the estimator.

In order to explore the power and accuracy of the BAS criterion and the associated H estimator, I generated 1,000 artificial fGn series of lengths randomly chosen between 100, 1,000, and 10,000 elements, and real values of $H = \delta$ uniformly sampled from the $(0, 1)$ interval. Fig. 2 plots the results of using the BAS criterion to compare the theory that $\delta \neq 1/2$ (or more precisely, $\delta \in [0,1]$) with the null hypothesis of $\delta = 1/2$. The figure shows that, when the real value of $\delta$ was either lower than .32 or higher than .61 (these limits are plotted by the vertical lines), the BAS clearly detects the FS. However, for values of $\delta$ within the interval [.32, .61], the BAS supports the null hypothesis more strongly. This means that the theory of $\delta = 1/2$ presents a better description of the data than stating that nothing is known about $\delta$. Notice that this does not imply that $\delta = 1/2$ is the ‘correct’ scaling, but rather, that it is the best among the theories compared. Choosing a better specified theory to test against, would enable contrasting hypotheses with higher resolution. This shows
Figure 2. (color online) Performance of the BAS criterion in detecting FS for 1,000 artificially generated fGn series of length 100 (top), 1,000 (middle), and 10,000 (bottom). The inset focuses on the lower values of $H$.

Figure 3. (color online) Comparison of the quality of Hurst exponent estimators for 1,000 artificially generated fGn series of length 100 (top), 1,000 (middle), and 10,000 (bottom).

one of the main advantages of the BAS: It permits the comparison of theories of different complexities, and naturally weighs the theories by their quality (i.e., stating that nothing is known about δ is a very bad theory, so any not too bad approximation should be considered an improvement).

Fig. 3 compares the accuracy of the estimator proposed here with some of the most commonly used estimators, including Rescaled Range Analysis (R/S; [1, 3]), Detrended Fluctuation Analysis (DFA; [10]), the local Whittle estimator [17], and the Scaled Windowed Variance and Spectral Power Density (SWV and SPD; in their respective improved variants SSC and lowSPDw,e as recommended by Ref. [4]). Already for series of mild length, the BAS estimator shows a substantially more accurate estimation than all other methods – both in terms of error and consistency – across all values $H < .9$. Only for $H > .9$ did the performance of the BAS estimator decrease, and even then it was not worse than the other estimators.
The next question that arises is that real world time series are hardly ever ‘clean’ examples of FS, rather, the FS structure is usually corrupted by noise of different types. Of especial importance are short-term correlations, that some time can give the false impression of FS, or can mask it away. To investigate the robustness of the BAS to this type of signal contamination, I generated 1,000 fGn time series of length 1,000 with $H = .75$ that were contaminated by random samples from a simulated ARMA(2,1) process (Autoregressive Moving Average; \cite{18}). The proportion of signal amplitude driven by ARMA was randomly sampled from $(0, 1)$. Fig. 4 plots the resulting evidence (comparing $\delta \in [0, 1]$ with $\delta = 1/2$) and the error in estimating $\delta$ for each series. The BAS robustly detects the presence of FS up to the point when the ARMA component accounts for more than half of the signal power. The contamination drives down $\hat{\delta}$, and when this estimator goes below the .61 limit mentioned above, a plain short-term correlation provides a better account of the data.

I have introduced an accurate method for deciding whether it is justified to claim that an observed series indicates FS. The BAS is fully analytical, does not have parameters that need to be subjectively adjusted, and its results can be computed with at most one level of very constrained approximation (the Gaussian proposal). The main component of the BAS, summarized by Eq. 8, is valid for stationary series of any type, whether they have finite variance, or they produce anomalous scaling of the Lévy type. Additionally, if the values are roughly normally distributed (or can be normalized) one can use the normal approximation and associated H estimator. Combined, the estimator and the BAS provide a powerful tool, enabling the use of the estimator for generating hypotheses on one portion of the data, and the BAS criterion to test them on a different portion. If two theories make different scaling
predictions, the BAS can be applied to decide which theory the data support. The BAS is only valid for stationary ‘noises’. If the series of interest is of the fractional Brownian motion type, differentiation is required before the methods can be applied. For more complex non-stationarity, preprocessing to isolate the stationary components of the series is required (see [13], for a recent proposal). Finally, if it is not clear that the scaling should apply at the earliest times [13], a later reference time $k_0 \geq 2$ can be chosen, and the time used in the equations expressed relative to it $k_r = k/k_0$.

* Contact: fermin.moscoso-del-prado@univ-provence.fr

[1] H. E. Hurst, Trans. Am. Soc. Civ. Eng. 116, 770 (1951)
[2] B. B. Mandelbrot and J. W. van Ness, SIAM Rev. 10, 422 (1968)
[3] B. B. Mandelbrot and J. R. Wallis, Water Resources Research 5, 228 (1969)
[4] A. Eke, P. Hermán, L. Kocsis, and L. R. Kozak, Physiol. Meas. 23, R1 (2002)
[5] E. T. Jaynes, Probability Theory: The Logic of Science (Cambridge University Press, Cambridge, UK, 2003)
[6] R. B. Davies and D. S. Harte, Biometrika 74, 95 (1987)
[7] E.-J. Wagenmakers, S. Farrell, and R. Ratcliff, Psychon. Bul. Rev. 11, 579 (2004)
[8] E.-J. Wagenmakers, S. Farrell, and R. Ratcliff, J. Exp. Psychol. Gen. 134, 108 (2005)
[9] T. L. Thornton and D. L. Gilden, Psychon. Bul. Rev. 12, 409 (2005)
[10] C. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E 49, 1685 (1994)
[11] P. Allegrini, P. Grigolini, and B. J. West, Phys. Rev. E 54, 4760 (1996)
[12] N. Scafetta and P. Grigolini, Phys. Rev. E 66, 036130 (2002)
[13] M. Ignaccolo, P. Allegrini, P. Grigolini, P. Hamilton, and B. J. West, Physica A 336, 595 (2004)
[14] The case comparing two composite hypotheses can be derived trivially from Eq. 8.
[15] S. Kullback and R. A. Leibler, Ann. Math. Stat. 22, 79 (1951)
[16] Eq. 9a is transformed into Eq. 9b by a simple change of variable $z = y/k^\delta$, $dy = k^\delta dz$.
[17] P. M. Robinson, Ann. Stat. 23, 1630 (1995)
[18] E. J. Hannan and J. Rissanen, Biometrika 69, 81 (1982)