The Mystery of the Missing Scales: Pitfalls in the Use of Fractal Renewal Processes to Simulate LRD Processes

Matthew Roughan[1] **Jennifer Yates**[2] **Darryl Veitch**[1]

1 – Software Engineering Research Centre, RMIT University
Level 3, 110 Victoria St, Carlton, Vic 3053, Australia
E-mail: {matt,darryl}@serc.rmit.edu.au

2 – AT&T Laboratories
180 Park Avenue, P.O. Box 971, Florham Park, NJ 07932-0000, USA
Email: jyates@research.att.com

Abstract

It has now been demonstrated in many studies that network traffic exhibits properties consistent with Long Range Dependence (LRD) and self-similarity. While theoretical frameworks are currently being developed to estimate the performance of such systems, simulation will remain a valuable tool for validating these theoretical models, and providing insight into systems which are too complicated to effectively model. Furthermore, when testing real systems, it is desirable to have traffic sources which are realistic, and hence display self-similarity. The Fractal Renewal Process (FRP) and its variants (including On/Off processes and superpositions thereof) have been proposed as models for LRD processes, in particular for network traffic. The FRP is a simple renewal point process with heavy-tailed inter-renewal times. The long-range correlations in the process are directly introduced by the heavy tail of the renewal times. The FRP has the great advantage that the number of computations required to generate a time series is linear and the time series can be generated on-line, facilitating generation of real traffic. However, there are some problems which arise when using such processes to generate LRD traffic. Most notably undersampling of the heavy-tailed random variables used to generate FRPs can lead to a truncation of the sampled autocorrelation that is not consistent with LRD. This problem becomes clear when the processes are investigated using the wavelet based methods of Abry and Veitch which segregate behaviour at different scales. This paper will describe the problem of undersampling, and its effects, and methods for avoiding the problem.

1 Introduction

It has now been demonstrated in many studies, for example [10, 14, 2, 8], that data network traffic exhibits properties consistent with Long Range Dependence and self-similarity. These properties have been shown to have profound effects on network performance [7, 13]. While theoretical frameworks are currently being developed to estimate the performance of such systems, simulation will remain a valuable tool for validating these theoretical models, and providing insight into systems which are too complicated to resolve analytically. Furthermore, when testing real telecommunications systems, it is desirable to have traffic sources which are realistic, and hence display self-similarity [16, 12].

The Fractal Renewal Process (FRP) [11, 15], and variants, have been proposed as models for Long-Range Dependent (LRD) or statistically self-similar data [18, 16, 15, 7, 11, 21]. The FRP is a simple renewal point process with heavy-tailed inter-renewal times. The long-range correlations in the process are directly introduced by the heavy tailed renewal times – the only mechanism available because in a renewal process there can be no connections across renewal points. This process has been proposed as a method for generating traffic for simulation, and for testing switching equipment, for instance by modeling packet or cell arrival times as a FRP.

Variants of the above include renewal reward processes in which 1 or more Markov renewal processes are simulated. At each renewal time a reward is chosen, either dependent on the state (e.g. the On/Off processes described in [7, 21] and elsewhere), or independently (e.g. the Spatial Renewal Process (SRP) described in [17]). The reward specifies the traffic rate between renewal times. There are also classes of doubly stochastic Poisson processes driven by FRPs, for instance the Fractal Shot Noise Driven Poisson Process, and the Fractal Binomial Noise Driven Poisson Process of [11, 15].

Superpositions of the above models may be taken to provide more complex models of traffic, and in fact as the number of superposed sources goes to infinity, in a number of cases, depending on the kind of renormalisation, the superposition approaches either the LRD Fractional Gaussian Noise (FGN) process, or an α-stable process.

These techniques share the important advantage that the number of computations required to generate a time series for simulation is linear in time, and the series can be generated on-line, that is, a point at a time, for as long as necessary. The final point is critical if the method is to be used to generate real test traffic [16].

This report outlines a problem that may be encountered when attempting to use these processes for simulation or generation of LRD traffic. The problem is present in all renewal based models to a greater or lesser extent. We will focus here on the FRP in order to make the problems clear.

The problem lies in the fact that the correlations over long time scales are introduced by the heavy tail of the inter-renewal times. Thus, in order to see correlations over a particular time scale in a particular realisation of the process, there must be a significant number of inter-renewal events of this length or greater present. However, in many cases, the random variables involved are not sampled sufficiently often for the larger values to occur in a typical realisation. Hence the sample autocorrelation function is in effect truncated, it does not provide a valid estimate of the ensemble autocorrelation function at large lags. In theory, the sample autocorrelation function converges to the ensemble autocorrelation function, but in practice the convergence can occur too slowly to be useful.

This property has not, to the authors' knowledge, been understood before, though a number of published results could be due to this effect (e.g. the behaviour of the variance vs. aggregation level plot of the SRP of [17]). This new understanding has arisen through using a sharper tool for investigating self-similar phenomena. We use the wavelet transform (and the related Abry-Veitch estimator), to investigate the FRP. The advantage of this approach (apart from the computational benefits of the algorithm, namely O(n) complexity, and on-line implementation) is that the wavelet transform separates the scaling behaviour into approximately independent coefficients at each scale, allowing investigation of the scaling phenomena of a process at set scales, independent of the behaviour elsewhere.

2 The Fractal Renewal Process

The Fractal Renewal Process (FRP) and its variants described above have been adopted as a means of modeling traffic for two reasons. The first is an empirical study of real data networks [21] which has shown that when traffic is broken into different source/destination flows these flows act as On/Off sources which can be modeled using an alternating renewal process, which is both intuitively simple and computationally efficient. The second reason is that these renewal based processes can easily be made LRD by choosing regularly varying or heavy-tailed on and/or off times, resulting in FRPs, consistent both with the observations of LRD in individual flows, and in aggregate data network traffic measurements. By heavy-tailed here we mean more specifically power-law with infinite variance. Recall also that by LRD we mean slow power-law decay of the autocorrelation function, as discussed later.

For the remainder of the paper we consider a FRP where the inter-renewal time distribution is heavy-tailed with infinite variance, in fact Pareto distributed as described below. Unless stated otherwise we will consider the *equilibrium renewal process*, that is where the time of the first arrival or event has a different distribution from that of the rest of the events – we do not assume an arrival at time 0. By choosing the first arrival time to be the residual life time of the usual distribution, the FRP becomes a stationary process, and is thus more suitable for our purposes.

The Pareto distribution is a commonly cited example of a regularly varying, or heavy-tailed distribution. One form of the Pareto distribution has Probability Distribution Function (PDF)

$$F(x) = 1 - \left(\frac{a+x}{a}\right)^{-\gamma},$$

and density

$$p(x) = \frac{\gamma}{a}\left(\frac{a+x}{a}\right)^{-\gamma-1},$$

for $x \geq 0$. For $\gamma \in (1,2]$ (the case of interest here) this distribution has mean $a/(\gamma-1)$ and infinite variance. We shall use this distribution for purposes of exposition. The results can be extended to other heavy-tailed distributions, and intuitively only the slow decay of the tail is important.

2.1 Tracking the Largest Scales

The question of critical interest to us here is, given n samples from such a distribution, what is the largest value we can reliably hope to see? Thus we consider the survival function of the Pareto distribution:

$$\text{prob}(X \geq x) = \left(\frac{x+a}{a}\right)^{-\gamma}. \tag{1}$$

If we have n independent trials chosen from the Pareto distribution, then

$$\begin{aligned}
&\text{prob}(X \geq x \text{ for at least one of } n \text{ trials}) \\
&= 1 - \text{prob}(X \leq x \text{ for all } n \text{ trials}) \\
&= 1 - [\text{prob}(X \leq x \text{ for a single trial})]^n \\
&= 1 - \left[1 - \left(\frac{x+a}{a}\right)^{-\gamma}\right]^n
\end{aligned} \tag{2}$$

By choosing a fixed $p = \text{prob}(X \geq x$ for at least one of n trials) we can compute the largest sample which will appear in our time series with this probability (or greater), which we denote by $L(n;p)$. Rearrangement of Equation 2 gives

$$\begin{aligned}
L_{\max}(n;p) &= a.\left(1-(1-p)^{\frac{1}{n}}\right)^{-\frac{1}{\gamma}} - a \\
&\approx a.\left(1-(1-p)^{\frac{1}{n}}\right)^{-\frac{1}{\gamma}}
\end{aligned} \tag{3}$$

for n large, $p > 0$. This 'largest event' determines the largest lag and scale on which the (finite length) realisation of the corresponding renewal process can display correlations. We define this largest scale $S_{\max}(n;p)$ to be the logarithm of the size of the largest event, that is

$$\begin{aligned}
S_{\max}(n;p) &= \log_2\left(L_{\max}(n;p)\right) \\
&\simeq \log_2(a) - \frac{1}{\gamma}\log_2\left(1-(1-p)^{\frac{1}{n}}\right),
\end{aligned} \tag{4}$$

for large n.

Next recall the well known limit as $n \to \infty$,

$$\begin{aligned}
\left(1-\frac{x}{n}\right)^n &\to e^{-x} \\
\left(1+\frac{\ln(1-p)}{n}\right)^n &\to (1-p),
\end{aligned}$$

by taking $x = -\ln(1-p)$. Taking a power of $1/n$ then rearranging gives

$$\begin{aligned}
\left(1+\frac{\ln(1-p)}{n}\right) &\to (1-p)^{\frac{1}{n}}, \\
\left(1-(1-p)^{\frac{1}{n}}\right) &\to -\frac{\ln(1-p)}{n}, \\
\frac{1}{\left(1-(1-p)^{\frac{1}{n}}\right)} &\to \frac{n}{-\ln(1-p)},
\end{aligned}$$

The net result is that

$$S_{\max}(n;p) \overset{n\to\infty}{\sim} \log_2(a) + \frac{1}{\gamma}\log_2\left(\frac{n}{-\ln(1-p)}\right). \tag{5}$$

Figure 1 shows $S_{\max}$ vs. $\log_2(n)$ for $p = 0.99$ and a number of values of γ. This largest scale gives a measure of the maximum extent of the correlation structure of an ordinary (non-equilibrium) FRP. Obviously, when constructing a FRP from the inter-renewal times described, the length of the time series $T = \sum_{i=1}^{n} X_i$ will vary. However, we may note that in the figure we have chosen $a = \gamma - 1$ in each case so that the mean inter-renewal time is 1. Thus, for a long time series we can expect from standard renewal theory that the time series will be of length $T \approx n$. Nonetheless, the central limit theorem does not apply in this case as the distribution in question has infinite variance, however for $\gamma \in (1, 2]$ convergence does still occur, albeit much more slowly than we would like. To first order therefore, the cutoff scale identified in *Figure 1* for n Pareto samples can be taken as the cutoff scale for the (ordinary) FRP up to time $T \approx n$.

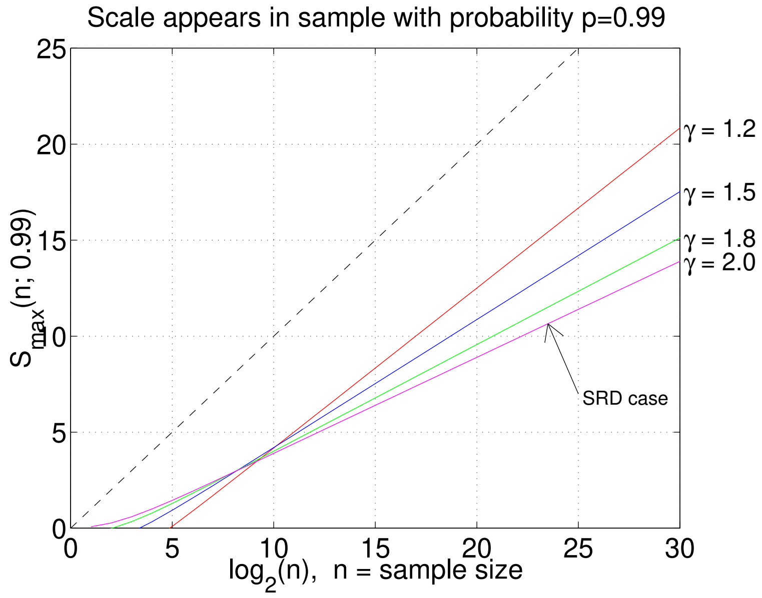


Figure 1: The largest scale which still has a probability $p = 0.99$ of appearing in n Pareto samples X_i, with $E[X_i] = 1$, for four values of the exponent γ ($a = \gamma - 1$). Note that the slope of the lines is less than 1, and is lowest for $\gamma = 2.0$, the Short Range Dependent (SRD) case.

A key observation is that the scale of the largest event is smaller – by several orders of magnitude – than the average length n of the data series. Thus the phenomenon of undersampling of large scales is not due to rare extreme cases where a single inter-arrival dominates the length of the time-series, where T would be far from n. On the contrary, even for realisations where the length T is close to n, scales well below the size of the series are systematically under-represented. Note also that the slope of the lines is less than 1, so that the problem becomes more noticeable for longer time series (rather than better as one might hope).

Also, contrary to intuition, this problem becomes worse as $\gamma \to 2$, that is as we approach the Short Range Dependent (SRD) case. Most problems associated with LRD (slow convergence of simulations etc.) become worse as $\gamma \to 1$. Though the counterintuitive nature of the problem is important, the fact that the problem appears worst in the SRD case is misleading. It is a simple reflection of the fact that large samples are even rarer for variables with a variance than for those without. The undersampling is simply not noticed in the SRD case as large lags are unimportant, and also very small by definition, and so easily obscured by noise.

2.2 A Refinement Through Extreme Value Theory

Extreme value theory [9] can be used to estimate the limiting distribution of the maximum of a series of random variables. The theory leads to three classes of limiting distributions, each with its own region of attraction, that is classes of distributions which correspond to the different extreme classes.

For instance, [9, Ex.1.7.6, p.22] shows that the Pareto distribution lies within the region of attraction of the Extreme Value Distribution of Type II. That is, the limiting distribution for the appropriately scaled maximum of a series of Pareto random variables with PDF[1]

$$F(x) = 1 - \left(\frac{x}{a}\right)^{-\gamma},$$

is an Extreme Value Distribution of type II. The PDF of the maximum

$$M_n = \max_n\{X_n\},$$

of the series X_i of n Pareto random variables, scaled by $1/(an^{\frac{1}{\gamma}})$ obeys

$$P\left\{\frac{M_n}{an^{\frac{1}{\gamma}}} \leq x\right\} \to e^{-x^{-\gamma}},$$

so that, setting $x = \tau^{-1/\gamma}$, the maximum itself has a distribution obeying

$$P\left\{M_n \leq a\left(\frac{n}{\tau}\right)^{\frac{1}{\gamma}}\right\} \to e^{-\tau}.$$

Applying this result by setting constant $1 - p = e^{-\tau}$ and thence $\tau = -\ln(1-p)$ we get

$$P\left\{M_n \geq a\left(\frac{n}{-\ln(1-p)}\right)^{\frac{1}{\gamma}}\right\} \to p$$

Therefore in the limit as $n \to \infty$ we get a similar result to before for the largest sample corresponding to probability p, namely

$$L_{\mathrm{ex}}(n;p) \to a\left(\frac{n}{-\ln(1-p)}\right)^{\frac{1}{\gamma}}$$

and again, taking the logarithm to obtain the corresponding scale, we have

$$S_{\mathrm{ex}}(n;p) \quad \to \quad \log_2(a) + \frac{1}{\gamma}\log_2\left(\frac{n}{-\ln(1-p)}\right). \tag{6}$$

Note that this agrees with our previous asymptotic results $L_{\max}(n;p)$ and $S_{\max}(n;p)$ respectively, obtained by more heuristic arguments.

2.3 Number of events of a given scale

Extreme value theory also shows that the number of events to exceed a threshold approaches a Poisson distribution (under the right circumstances). Thus we can easily predict the largest scale for which at least k larger events will occur with probability p. From [9, Thm 2.1.1, p.32], we note that the probability that S_n out of n samples exceed the level u_n, which in the Pareto case is given by $u_n = a\left(\frac{n}{\tau}\right)^{\frac{1}{\gamma}}$, has distribution

$$P\{S_n = k\} \approx e^{-\tau}\frac{\tau^k}{k!},$$

for $k = 0, 1, 2\ldots, n$, and the mean number of exceedances will be τ.

One could set τ to some value, say 10 to ensure that a reasonable number of events fall above the threshold. The corresponding minimum value of n required to see a certain scale could be found by inverting the expression for the level $u_n(\tau)$ given above.

2.4 Discussion

We have given the distribution of the number of Pareto samples larger than a particular scale, and in particular, the distribution of the largest of these samples. The important fact to note is that the size of this largest event increases more slowly than the number of events n. In the corresponding FRP the size of these inter-arrival events, and their number, are directly related to the correlation scales in the process,

[1] This is a right shifted version of the Pareto distribution used earlier. This change has no effect on the asymptotic extreme value distribution, but makes the mathematics easier.

and the length of the process respectively. Thus as the number n of renewals increases the maximum correlation scale in the process increases more slowly than the total length of the realisation – a fact which should be intuitively obvious but which is obscured by the details of the process.

This property is a general feature of renewal processes and is insensitive to details of the inter-arrival distribution. Indeed, it persists, in fact is more severe, in the 'light-tailed' or SRD case with $\gamma = 2.0$. In the light-tailed case however we do not care if these events are undersampled, because they will have little effect on the overall behaviour of the process. In contrast, in the heavy-tailed case the absence of the large events in the sample can drastically effect its properties. In particular we will see that their absence prevents the sample autocorrelation function from converging to the ensemble autocorrelation function above a cutoff scale which depends on n.

3 Long Range Dependence and the FRP

In this section we will very briefly describe some experimental results. We have conducted other experiments over a range of parameters for verification, but have restricted the set of examples here to make our point more clearly.

The experiment described here involves simulating a FRP for a set integer time $t = N$. In our context the renewals might model packet arrivals from a data link. To study this process we divide time into uniform intervals: $[0, 1)$, $[1, 2)$, $[2, 3) \ldots$, $[N-1, N)$, and we count the number of renewal events to occur during each, and denote them by Y_1, Y_2, $Y_3 \ldots$, Y_N. This forms our basic time series Y_k.

3.1 Long Range Dependence

Long Range Dependence (LRD) refers to the slow decay of the process's autocorrelation function $r_Y(k) = C_Y(k)/C_Y(0)$, where $C_Y(k)$ is the autocovariance function, defined by $C_Y(k) = E[(Y_{n+k} - \bar{Y})(Y_n - \bar{Y})]$, and where $\bar{Y}$ is the mean (note stationarity means that each of these is independent of n). In fact, we are concerned with processes for which the autocorrelation's decay is so slow that the sum of all correlations beyond any given lag is infinite, and therefore significant, even if individually the correlations are small. The past therefore exerts a long term influence on the future, exaggerating variability and rendering statistical estimation problematic. Commonly, a more practical definition is adopted – LRD is defined by a power-law decay in the autocorrelation $r_Y(k) \sim \bar{c}_r |k|^{-(1-\alpha)}$, $\alpha \in (0, 1)$, or equivalently as the power-law divergence at the origin of its power spectrum: $f_Y(\nu) \sim c_f |\nu|^{-\alpha}$, $|\nu| \to 0$. The Hurst parameter describes the (in practice, asymptotic) self-similarity of the cumulative or counting process $\sum_{k=1}^{n} Y_k$ while the LRD parameter α describes the rate process Y_n. It is nonetheless common practice to speak of H in relation to LRD via the relation $H = (1 + \alpha)/2$, and we follow this convention here.

3.2 The FRP and LRD

One of the principal reasons for using FRPs and their variants is that these processes are LRD, which is consistent with data network traffic measurements.

For FRP type models the autocorrelation function is given by [15] as

$$r(k) = \frac{1}{1 + T_0^{\alpha}} \frac{1}{2} \nabla^2 (k^{\alpha+1}),$$

where the central difference operator is given by

$$\nabla^2 \Big(f(k) \Big) = f(k+1) - 2f(k) + f(k-1),$$

and T_0 depends on the exact distribution of the renewal times. For large k this relationship can be approximated by

$$r(k) \sim c k^{\alpha - 1},$$

for some constant c, agreeing with the definition of LRD given above.

For the specific case of the FRP described above $\alpha = 2 - \gamma$, and the Hurst parameter is given by $H = (3 - \gamma)/2$. Note that when $\gamma = 2$ the Hurst parameter is 0.5 corresponding to the SRD case, but that for $\gamma < 2$ the Hurst parameter is > 0.5 leading to LRD.

Note however, that the result above is an ensemble average! That is, the expectation in the definition of the autocorrelation refers to the average over an infinite set of realisations of the process. Typically ensemble averages, for example the mean of the process, are estimated using a time average: the sample mean of one realisation of a time series. Many processes, including FRPs, possess ergodicity properties which guarantee that time averages do indeed converge to ensemble averages. However, ergodicity does not specify how quickly this convergence occurs. We shall see here that the undersampling effects described above lead to the sample autocorrelations at large lags (calculated using a time average) converging too slowly to be useful estimates of the ensemble autocorrelations. Therefore any property of the process dependent on the autocorrelation at high lags (for instance LRD) will also be affected by this lack of convergence.

3.3 The Abry-Veitch estimator

We now briefly introduce the wavelet analysis tools to be used to analyse the data.

In [20, 19] a semi-parametric joint estimator of the spectral domain parameters of LRD: (α, c_f) is described based on the *Discrete Wavelet Transform* (DWT). Wavelet transforms in general can be understood as a more flexible form of a Fourier transform, where $Y(t)$ is transformed, not into a frequency domain, but into a time-scale wavelet domain. The sinusoidal functions of Fourier theory are replaced by wavelet basis functions $\psi_{a,t}(u) \equiv \psi_0(\frac{u-t}{a})/\sqrt{a}$, $a \in I\!R^+$, $t \in I\!R$ generated by simple translations and dilations of the *mother wavelet* ψ_0, a band pass function with limited spread in both time and frequency. The wavelet transform can thus be thought of as a method of simultaneously observing a time series at a full range of different scales a, whilst retaining the time dimension of the original data. Multiresolution analysis theory shows that no information is lost if we sample the continuous wavelet coefficients at a sparse set of points in the time-scale plane known as the *dyadic grid*, defined by $(a, t) = (2^j, 2^j k)$, $j, k \in I\!N$, leading to the DWT with discrete coefficients $d_Y(j, k)$ known as *details*. Intuitively, the dyadic grid samples the wavelet domain at a resolution appropriate to the scale. Henceforth we will deal exclusively with the details of the DWT, which can be calculated in $O(n)$ time. Scale, which can be thought of as the inverse of frequency and similar to lag, is often measured in logarithmic units j, where $a = 2^j$, and k plays the role of time (although a time whose rate varies with j).

For finite data of length n, j will vary from $j = 1$, the finest scale in the data, up to some $j_{\max} \approx \log_2(n)$. The number of coefficients available at octave j is denoted by n_j, and approximately halves with each increase of j.

The main feature of the wavelet approach which makes it so effective for the statistical analysis of scaling phenomenon such as LRD is the fact that the wavelet basis functions themselves possess a scaling property, and therefore constitute an optimal 'co-ordinate system' from which to view such phenomena. The main practical outcome is that the LRD in the time domain representation is reduced to residual **short** range correlation in the wavelet coefficient plane $\{j, k\}$, thus removing entirely the special estimation difficulties. Thus for each fixed j, the series $d_Y(j, \cdot)$ can be regarded as a stationary process with weak short-range dependence, and these series can be regarded as independent of each other. The long-range dependence now manifests itself in a statistically benign way- by controlling the way the variances of the $d_Y(j, \cdot)$ change with scale, rather than modulating the co-variance structure.

To measure the second order properties of the data we compute a set of statistics μ_j using

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_Y(j, k)|^2. \tag{7}$$

The μ_j are non-parametric, unbiased estimators of the variance of the $d_Y(j, \cdot)$ (the means of the details are zero), and can be thought of as a way of concentrating the gross second order behavior of Y at octave j. Furthermore, the μ_j are themselves only weakly dependent, so the analysis of each scale is largely decoupled from that at other scales. Therefore, to analyze the second order dependence of Y_n on scale, we are naturally lead to study μ_j as a function of j. Since we consider LRD to be essentially a power-law behavior of second order moments, this is naturally done in a log-log plot, called the *Logscale Diagram* (see *Figure 2*). The Logscale Diagram is a kind of spectral estimate, where frequency has been replaced by its inverse, scale, in a discrete logarithmically spaced grid.

We can now outline the estimator for the parameters of LRD as consisting of the following four stages:

1. **Wavelet decomposition:** A discrete wavelet transform of the data is performed, generating the details $d_Y(j, k)$ over the dyadic grid.

2. **Detail variance estimation:** At each fixed octave j the details are squared and averaged across 'time' k to produce μ_j, an excellent estimate of the variance of the details, μ_j, the average of the squares of the details at a given j, is an estimate of the variance at that j. For LRD processes these variances follow a power-law in j with exponent α.

3. **Analysis using the Logscale Diagram:** From the plot of $y_j = \log_2(\mu_j)$ against j, the *Logscale Diagram*[2], the scaling range (j_1, j_2) where scaling occurs (alignment observed) is determined.

4. **LRD parameters estimation:** The LRD parameters H and c_f are extracted by performing a weighted linear regression over the scaling region[3,4,5].

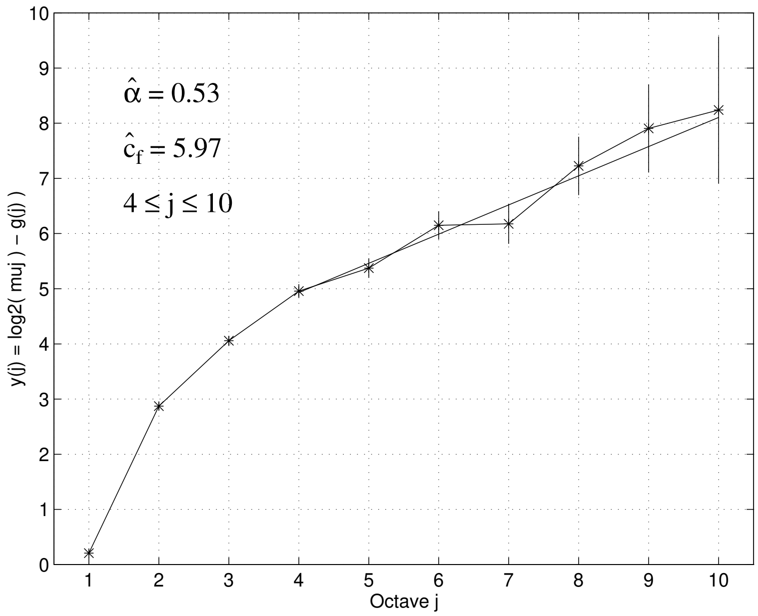


Figure 2: **Stages 3 and 4: estimation from the Logscale Diagram.** An example of the y_j against j Logscale Diagram and regression line for a LRD process with strong SRD. The vertical bars at each octave give 95% confidence intervals for the y_j. The series is simulated fARIMA(0,d,2) with $d = 0.25$ ($\alpha = 0.50$) and $\Psi = [-2, -1]$ implying $c_f = 6.38$. Selecting $(j_1, j_2) = (4, 10)$ allows an accurate estimation despite the strong SRD: $\hat{\alpha} = 0.53 \pm 0.07$, $\hat{c}_f = 6.0$ with 95% confidence interval $4.5 < \hat{c}_f < 7.8$.

An example of the regression fit using a simulated data set is given in *Figure 2*. The 95% confidence intervals for each y_j, shown as vertical lines at each octave j, are seen to increase with j. A plot such as this of y_j against j, complete with confidence intervals about the y_j, has been termed the *Logscale Diagram* [20, 1], and constitutes an effective starting point for the analysis of scaling phenomenon. The estimator, being semi-parametric, requires an analysis phase prior to estimation to determine the scaling range where alignment is observed in the Logscale Diagram (see [1] for further details on the reading of Logscale Diagrams).

The choice of alignment range $j \in (j_1, j_2)$ over which straight line behaviour is observed, and estimation performed, is particularly relevant here. For LRD data we expect that there will be some lower scale $j = j_1$ at which the power-law scaling will 'begin', and the scaling will continue up to the largest scales in the data, because LRD is defined as an asymptotic phenomenon valid at large scales. This is indeed what is observed in *Figure 2*, the log-scale diagram of a simulated fARIMA (fractional Auto-Regressive Integrated Moving Average) process. As we will see however, despite the sure knowledge of LRD in the FRP series we examine, there exists an upper cutoff scale j_2 for the scaling behaviour that must be chosen considerably smaller than $j_{\max} \approx \log_2(n)$. This is the observation that motivated the present work.

[2] In forming Logscale Diagram small corrective terms $g(n_j)$ are in fact subtracted from $\log_2(\mu_j)$ to compute y_j to account for the fact that $E[\log](\cdot) \neq \log(E[\cdot])$.

[3] H is related to the slope of the plot, and c_f to a power of the intercept.

[4] The weights are functions of the known variances of the y_j and do **not** depend on the data.

[5] Confidence intervals for H are derived from the standard variance formulae for weighted linear regression with mutually independent y_j, and so again are **not** functions of the data.

3.4 Numerical Results

In *Figure 3* the Logscale Diagrams of four realisations of the FRP are shown with $H = 0.75$. The figure also shows the length of the time series used ($N = 2^{20}$), the number of renewal events n, and an estimate of the Hurst parameter, based on regression over the scales (j_1, j_2) between the vertical dashed lines.

In addition, superimposed over the Logscale Diagram, and sharing the same axes, is the $\log_{10}$ of a histogram $h(j)$ of the number of events at scale j (crosses). Each Pareto sample x was binned into scale j given by

$$j = \begin{cases} 0, & \text{for } x \leq 1, \\ \lceil \log_2(x) \rceil, & \text{for } x > 1. \end{cases}$$

The histograms show that there is an upper scale, around $j = 13$, above which samples of the Pareto random variable do not occur in these realisations, and that this cutoff scale is far smaller than the length of the data. Also plotted on the graph is a horizontal line at 1, corresponding to 10 events, defining a related cutoff scale, around $j = 10$, where events become too scarce for reliable representation. Note that the previous sections predicted scale at which this cutoff would occur is 10.9, based on the length of the time series (and the estimated number of events the time series would include), or the actual number of events in each of the four time series realisations considered.

As noted previously the y_j (plotted as stars) should, up to the 95% confidence intervals (dark vertical lines), lie on a straight line above a lower cutoff scale j_1 where the LRD 'begins'. We have chosen $j_1 = 7$. The crucial issue here however is the choice of the *upper* cutoff scale. Rather than selecting the largest scale available in the Logscale Diagram, $j = 17$, we are guided by the 'largest scale' $j = 10$ selected above by the cutoff corresponding to having at least 10 events. The LRD parameter estimation is therefore performed over $(j_1, j_2) = (7, 10)$, where excellent alignment is observed in each Logscale Diagram. It is clear from the plots that the y_j at the higher scales do not lie on the dotted extension of this plotted regression line. In fact in each case the systematic deviation of the y_j away from the line begins almost as soon as the histogram values drop below the magical mark of ten, and become more severe at higher scales.

Examining the four different realisations we see that the estimates at higher scales are highly variable as well as systematically biased. This effect is the result of the undersampling of events at these scales. If these scales were used in estimation of the Hurst parameter, then a quite inaccurate estimate might result, but more importantly, the process itself yields sample paths where correlation structure is truncated, or at least strongly distorted.

We should note at this point that although the confidence intervals have been calculated using Gaussian assumptions, which do not hold true for this process, the results seem to indicate a systematic bias regardless of variation.

3.5 Sample Autocorrelation Function

We have used the wavelet transform to investigate directly the properties of the process as a function of scale. As already noted however, there is a direct link in the FRP between scale and correlation length. It is of interest therefore to examine the autocorrelation function directly. We give only one example here, mainly to illustrate that in fact the sample autocorrelation function is a poor tool. While the wavelet based method separates behaviour at different scales well, the sample autocorrelation function below does not. Hence the effect described above is not so clear cut as the distortion does not begin at a particular scale, but rather corrupts a large range of values. In other words, effects due to poor estimation are mixed with those of the missing scales phenomenon.

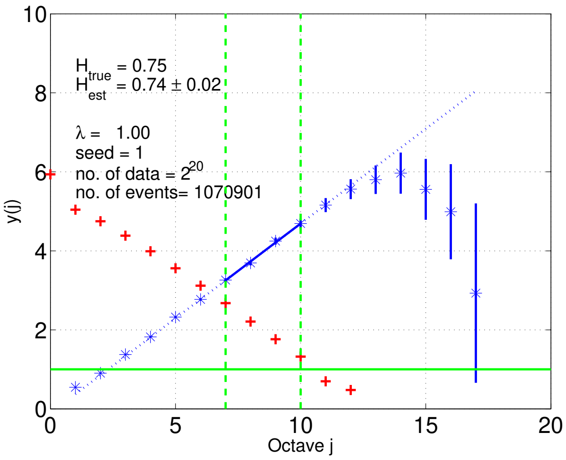
The sample autocovariance [4, p.18] of a sequence of N values can be computed as

$$\hat{C}(k) = \frac{1}{N} \sum_{n=1}^{N-|k|} (Y_n - \bar{Y})(Y_{n+|k|} - \bar{Y}),$$

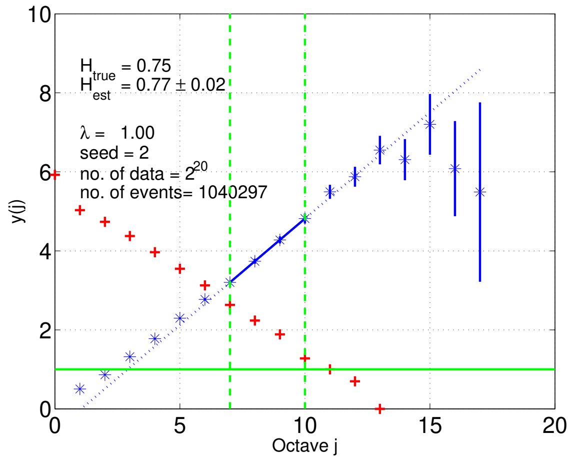
for $-N < k < N$, where $\bar{Y}$ is the sample mean of the process $\bar{Y} = \frac{1}{N} \sum_{n=1}^{N} Y_n$. The sample autocorrelation function may then be estimated using

$$\hat{r}(k) = \hat{C}(k)/\hat{C}(0).$$

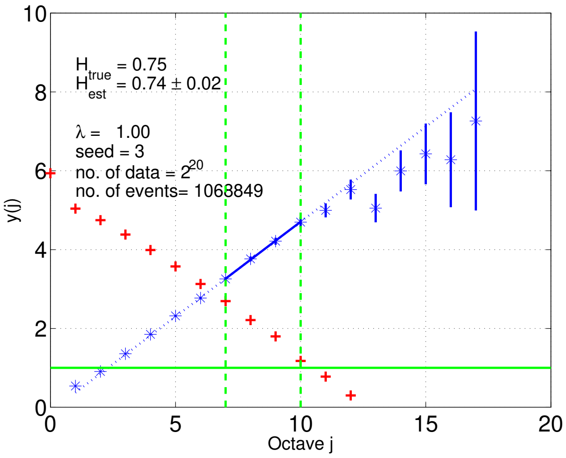
Note that these two estimates are computed from time averages of a single realisation of the process.



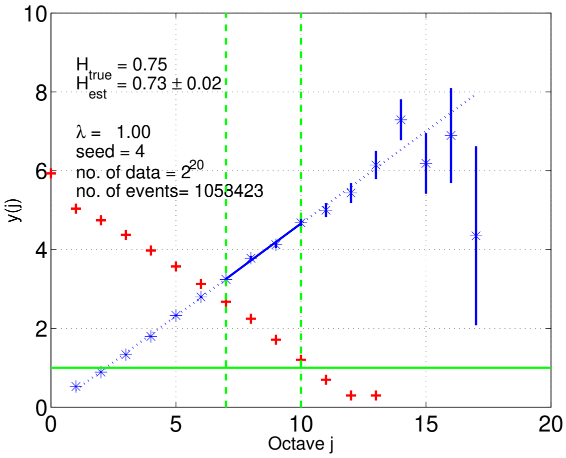
(a) SEED = 1



(b) SEED = 2



(c) SEED = 3



(d) SEED = 4

Figure 3: The log-scale plots for the regression variables y_j (stars) for the FRP. The vertical dashed lines indicate j_1 and j_2, the range of scales used to calculate the regression line, itself shown as a solid (dotted) line within (outside) this range. Actual and estimated Hurst parameter values are shown, with 95% confidence bounds based on Gaussian assumptions. The figures also show as +'s the $\log_{10}$ of the number of events to occur at each scale (set to zero in the case of zero events). The horizontal dashed line at 1 shows when the number of events at a particular scale drops below 10 - and is used to define j_2.

Figure 4 shows an example of the sample autocorrelation for the FRP together with the theoretical ensemble autocorrelation function. A log-log scale is used so that we can easily check for the expected straight line of the ensemble autocorrelation. To facilitate comparison, vertical lines have been plotted at lags which correspond to the scales (j_1, j_2) from the wavelet analysis.

First note that the sample autocorrelation only follows the theoretical value for a range of small lags. At the other extreme, above scale 10 the correlations fall below the horizontal line corresponding to the 95% confidence interval for the sample autocorrelation of an IID Gaussian series [4, p.18]. They therefore cannot be used to infer the shape of the correlation function. The fact that so many scales fall in this range is a measure of the strength of the undersampling present. Below scale 10 however the autocorrelation estimates are above this line and are therefore significant, but are nonetheless far from the theoretical values. It is clear that slopes calculated in this plot over this range, for example between the vertical lines, would yield highly inaccurate estimates. Here the estimation is not strongly affected by the undersampling, but remains poor due to the statistical difficulties of time domain estimation in the presence of LRD. This phenomenon is well understood [3].

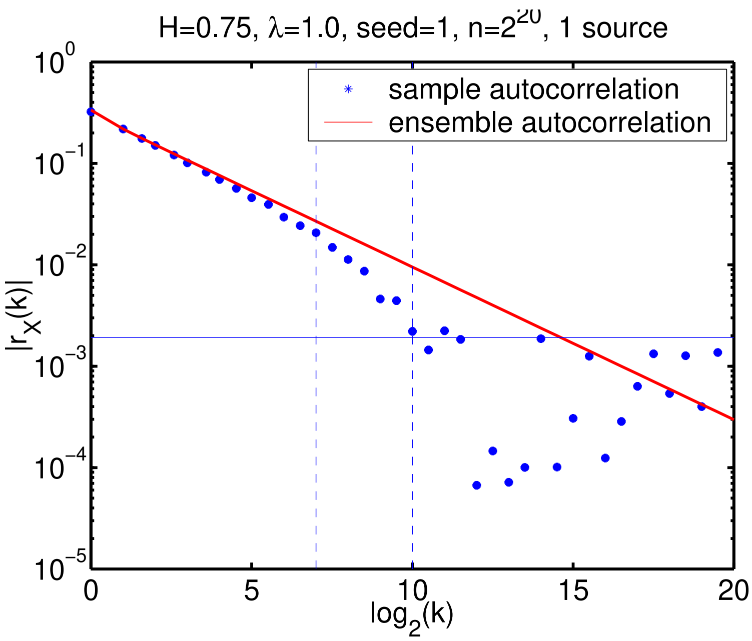


Figure 4: A log-log plot of the absolute value of the sample and ensemble autocorrelation functions (at a representative set of lags). The vertical lines show the same scales used in Hurst estimation using the wavelet method- here estimation in this range is very poor. The horizontal line gives the 95% confidence interval for the sample autocorrelation of an IID Gaussian series. Points above this line indicate a significant correlation, while points below could be due to chance.

4 Implications

We have shown how the sample autocorrelations of an FRP do not reflect the theoretical ensemble autocorrelations at large, or even medium, timescales. This has two major implications. The first is in estimation of the parameters of the process. It is obvious from the autocorrelation plot in Figure 4 that attempting to measure the LRD parameters from this plot would be difficult. However even more advanced estimation methods, such as the AV estimator, suffer from similar problems. The advantage of the AV estimator is that, given an upper cutoff scale below which the truncation effects will have very little effect (and we have shown how to estimate such a cutoff), the wavelet estimation is still good. That is, if the correct scaling behaviour is actually present, then it can be detected cleanly in the wavelet framework, due the scaling properties inherent in the wavelet basis itself. Thus we can still obtain an accurate estimate, provided we know approximately how many renewal events have been measured.

The second major consequence of the work above is in simulation – an area where their use has often be advocated. We are not saying that FRPs and their variants should never be used in simulation, but merely that great care be taken in their use. This is of course the case whenever simulating systems in which heavy tails play a part. It is known for example that heavy-tails can have drastic effects on the convergence of estimates [5, 6]. We have noted here that because the FRP generates correlations through large events, we must keep track of the number of events at each scale in order to know whether correlations on that scale are represented in the data. It follows that if we are interested in measuring any phenomena at all on a particular time scale, we must be sure that that time scale is represented in the data. This applies doubly when trying to measure phenomena which are the result of LRD – an asymptotic property relating different scales. In short, simulations must be run long enough to generate enough events of the type needed, rather than just up to some fixed time, implying far longer simulations, and more careful monitoring.

Note that for particular situations such as the FRP with Pareto inter-arrivals, it is sufficient to simply keep track of the number of Pareto events, and the largest scales available can be computed using the formula provided here.

The careful reader will note that when large numbers of FRP sources are superimposed the problem is diminished because there will be a larger number of events in a simulation of fixed length. In the limit as the number of sources goes to infinity, with the average rate of each source going to zero, and the total rate remaining constant, we may arrive at some limiting process, for example FGN. One could consider generating such limit processes directly, without making use of FRPs, and thereby bypassing the missing scale problem. Furthermore, simulation of a large but finite number of sources requires a large amount of computational power. Though the computations can be easily parallelised, not everyone has a MasPar available for doing traffic simulations (for instance see [21]). Finally, for any finite number of sources the undersampling problem will eventually recur for a long enough traffic sequence, thus necessitating the care advocated above in any case.

5 Conclusion

The problem described above, truncation of the sample correlation structure of a process through undersampling, occurs to a greater or lesser extent in all FRP variants. We have shown how the phenomenon leads to the effective loss of a large number of scales in the estimation of second order quantities such as the correlation function. We have pointed out that this is particularly problematic for the estimation of LRD, which is a phenomenon defined at large scales, causing a difficulty even for the wavelet based estimator of Abry and Veitch. A method for estimating a cutoff scale above which estimates fail to capture the ensemble correlation structure is given. Wavelet based estimates using this scale as an upper limit of the estimation range can recover the correct parameters. In contrast, even in the light of this cutoff scale, time domain estimates are unreliable.

In future work, a more rigorous connection between the ensemble characteristics and the sample path properties for the FRP could be made, using the ergodic theory of renewal processes and other techniques. This could lead to methods of including the information present at larger scales, via bias and variance correction factors which explicitly take into account the highly non-Gaussian nature of the data. More general methods of determining the cutoff scale could be considered, based on some general considerations of tail behaviour of the inter-arrival distribution, generalising the specific Pareto result.

References

[1] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch. Wavelets for the analysis, estimation and synthesis of scaling data. submitted to Self Similar Network Traffic Analysis and Performance Evaluation, K. Park and W. Willinger, Eds., 1999.

[2] A.Feldmann, A.C.Gilbert, W.Willinger, and T.G.Kurtz. The changing nature of network traffic: Scaling phenomena. *Computer Communications Review*, 28(2), 1998.

[3] J. Beran. *Statistics for Long-Memory Processes.* Chapman and Hall, New York, 1994.

[4] Peter J. Brockwell. *Introduction to Time Series Forecasting.* Springer Texts in Statistics. Spinger-Verlag, New York, 1996.

[5] Mark E. Crovella and Lester Lipsky. Long-lasting transient conditions in simulations with heavy-tailed workloads. In S.Andadottir, K.J.Healy, D.H.Withers, and B.L.Nelson, editors, *Proceedings of the 1997 Winter Simulation Conference*, 1997.

[6] Ashok Erramilli, James Gordon, and Walter Willinger. Applications of fractals in engineering for realistic traffic processes. In Jaques Labetoulle and James W. Roberts, editors, *Proceedings of the 14th International Teletraffic Congress - ITC 14*, volume 1a, pages 35–44. Elsevier, Amsterdam, 1994.

[7] Ashok Erramilli, Onuttom Narayan, and Walter Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking*, 4(2):209–223, April 1996.

[8] Judith L. Jerkins and Jonathan L. Wang. A measurement analysis of ATM cell-level aggregate traffic. In *IEEE GLOBECOM'97*, 1997.

[9] M.R. Leadbetter, Georg Lindgren, and Holger Rootzen. *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1983.

[10] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, Feb 1994.

[11] Steven B. Lowen and Malvin C. Teich. Estimation and simulation of fractal stochastic point processes. *Fractals*, 3(1):183–210, 1995.

[12] M.Roughan, D.Veitch, and M.Rumsewicz. An ATM traffic generator for self-similar traffic. Technical Report SERC-0076, SERC, 110 Victoria St, Carlton, VIC 3052, AUSTRALIA, 1998.

[13] I. Norros. A storage model with self-similar input. *Queueing Systems*, 16:387–396, 1994.

[14] V. Paxson and S. Floyd. Wide-area traffic: the failure of Poisson modelling. In *Proceedings of SIGCOMM '94*, 1994.

[15] Bong K. Ryu and Steven B. Lowen. Point process approaches to the modelling and analysis of self-similar traffic - part i: Model construction. In *IEEE INFOCOM'96: The Conference on Computer Communications*, volume 3, pages 1468–1475, San Francisco, California, March 1996. IEEE Computer Society Press, Los Alamitos, California.

[16] Bong K. Ryu and Mahesan Nandikesan. Real-time generation of fractal atm traffic: Model, algorithm and implementation. preprint.

[17] Tarkan Taralp, Michael Devetsikiotis, Ioannis Lambadaris, and Amit Bose. Efficient fractional Gaussian noise generation using the spatial renewal process. In *ICC'98*, 1998.

[18] Darryl Veitch. Novel models of broadband traffic. In *IEEE GLOBECOM'93, Houston, Texas*, 1993.

[19] Darryl Veitch and Patrice Abry. Estimation conjointe en ondelettes des paramètres du phénomène de dépendance longue. In *Proc. 16ième Colloque GRETSI, pp.1451–1454, Grenoble, France*, 1997.

[20] Darryl Veitch and Patrice Abry. A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory special issue on "Multiscale Statistical Signal Analysis and its Applications"*, 45(3), April 1999.

[21] Walter Willinger, Murad S. Taqqu, Robert Sherman, and Daniel V. Wilson. Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level. *Proceedings of the ACM/SIGCOMM'95*, 1995.