Some Remarks on Unexpected Scaling Exponents

Matthew Roughan
School of Mathematical Science
University of Adelaide
SA 5005, Australia
matthew.roughan@adelaide.edu.au

Darryl Veitch*
Department of E&E Engineering
The University of Melbourne
Victoria 3010, Australia
d.veitch@ee.unimelb.edu.au

This article is an editorial note submitted to CCR. It has NOT been peer reviewed. Authors take full responsibility for this article's technical content. Comments can be posted through CCR Online.

Categories and Subject Descriptors

G.3 [**PROBABILITY AND STATISTICS**]: Stochastic processes

General Terms

Traffic Models

Keywords

Fractal Renewal Process, Long-range dependence, wavelet estimator, Hurst parameter, sampling, discontinuities.

1. INTRODUCTION

In recent issues of the ACM Computer Communications Review, editorials have appeared [1, 5] regarding Logscale-Diagram (LD) plots for estimating parameters of Long-Range Dependent (LRD), self-similar, or "heavy-tailed" traffic. Since its discovery in data traffic, LRD has sparked many research papers (for instance see the references contained in [10]), including a stream of papers on how to estimate the parameters of LRD processes. One of the most commonly used approaches is the wavelet based estimation technique by Abry and Veitch [7]. Although it performs well in most circumstances, like any other estimation technique for LRD, it is not perfect. Each of [1, 5] notice that the LD plots fail to measure the correct value of the Hurst parameter for simple long-range dependent processes built from independent heavy-tailed random variables. The use of such processes as LRD traffic models arose some years ago in response both to empirical measurements of Ethernet traffic [11], and asymptotic results showing that superpositions of many such processes would result in aggregates with nice properties. The fractional Gaussian Noise (fGN) [3] is a well known example of such a limiting LRD aggregate.

In this note, we point out that the issues raised by [1,5] are in fact two entirely separate phenomena, each of which has been noted before, and each of which is understood. They are intrinsic to the nature of the processes being studied, rather than artifacts of the LD based estimation. The phenomenon noted by [1] is due to the time-scale scaling signature of singularities in the data, which must be distinguished from the different scaling signature of LRD. The phenomenon noted by [5] is more subtle. It arises through a sampling issue, corresponding to a lack of events at large to medium scales. In each case, correct estimates of the Hurst parameter can be made using the LD if the analysis is made over the appropriate range of scales. Given the space restrictions of this note, we will give an intuitive description here, and direct readers to [6, 9] for a more detailed mathematical treatment.

2. TWO PHENOMENA

A process is long-range dependent when its autocorrelation does not decay quickly in time. That is, points separated by long intervals can still be significantly correlated. LRD has been observed in many network data sets, and is of enduring interest as its presence is robust, and it often generates counter-intuitive properties.

A classical approach for simulating LRD is to 'lay heavy tailed durations end-to-end'. A simple model for traffic using this idea is the *On/Off source*, where mutually independent On and Off periods alternate, for example an On period can represent when a constant bit rate flow is active, and Off when it is not. If the On period duration is taken to be a heavy-tailed random variable such as Pareto (with infinite variance), the rate process becomes LRD. The high variability of such On periods automatically give them a mice/elephant[1] character: very long On periods, although still rare, are now likely enough to actually occur in practice, wreaking havoc with sample statistics.

On/Off sources are particularly useful as they can be easily simulated from their heavy-tailed building blocks, and can be aggregated together to create more realistic LRD models (for example see [3, 4]). The difficulty noticed by [1] was in the context of an aggregated On-Off type model known as the 'M/G/∞' source [2], whereas [5] studied a single On/Off source. We begin with the problem noticed by [5]. To understand it, it suffices to study the simplest process of this type, a *Fractal Renewal Process* (FRP).

2.1 The Mystery of the Missing Scales

A FRP is a renewal process, i.e., a set of points on the time line separated by inter-arrivals which are independent and identically distributed. It is called fractal (and is LRD) when the inter-arrival variable is heavy tailed. In [6] we noted, much as [3, 4] did, that LD plots for such processes do not have the expected behaviour. An example of this from [6] is reproduced in Figure 1. The stars (with

*D.Veitch is with the ARC Special Research Centre on Ultra-Broadband Information Networks(CUBIN), CUBIN is an affiliated program of National ICT Australia (NICTA).

[1] It is ironic that the heavy-*tail* of the flow distribution should have been called an elephant (given that their tails are rather small). We recommend 'kangaroo'.

95% confidence intervals shown as vertical lines) are the LD plot[2] of a FRP. As a LRD process, we expect these to form a straight line over large scales, and the slope of the line to enable the recovery of the Hurst exponent H. However as we see, beyond octave $j = 11$ (the octave is defined as $j = \log_2(\text{scale})$) the linear behaviour is mysteriously absent. The greater mystery is that the problem is intrinsic: it holds true no matter how much data we have, and even if transients (deviations from stationarity) and edge effects are perfectly dealt with.

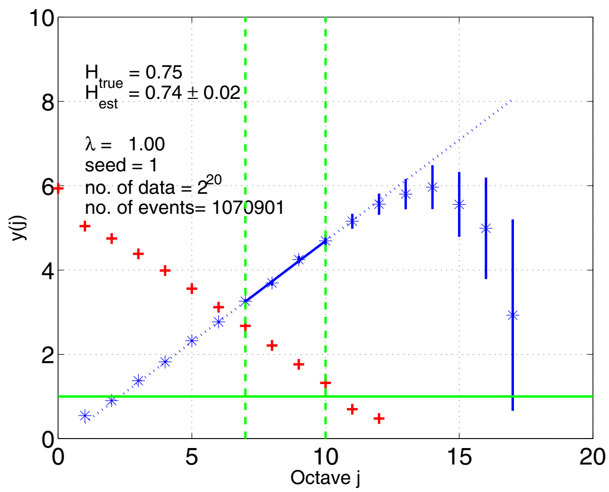

Figure 1: The LD plot (stars) shows linear behaviour only up to $j = 11$. The expected linearity is mysteriously missing at the largest scales. The log-histogram of the number of inter-arrival events at this scale (shown by '+' signs) dives to zero beyond $j = 10$. Estimates made to the left of this can recover H.

In fact the mystery can be easily understood in terms of *under-sampling*. Recall that in the FRP, inter-arrivals are independent of each other. So how can two points in time show correlated behaviour? Such a connection can only be generated when a single inter-arrival interval *spans the distance between the points*. In other words, correlations at some timescale τ are directly dependent on inter-arrivals being present of duration larger than τ. If such events do not occur in a given sample path (or data set), then there will be no correlation to measure at scale τ!

We can see this effect at work in dramatic fashion in Figure 1, where we superimpose (fortuitously the same axes serve) $\log_{10}$ of the histogram of the number of events at octave j (scale 2^j), that is, the number of inter-arrivals of about that duration. We immediately see that the deviation of the LD from the expected straight line occurs precisely when the supply of 'correlation inducing intervals' dries up, that is when the crosses drop below 10 samples (horizontal green line) at $j = 11$ and then to 0 samples at $j = 13$.

Attempts to estimate the slope beyond the scale at which correlations exist are doomed to failure, because these correlations create the slope. Clearly, this is not a problem of the LD itself. The same phenomena would appear in all estimators, and one of the distinct advantages of the LD plot is that it clearly separates phenomena at different scales, and thereby makes the problem obvious whereas other estimators (e.g. the Whittle estimator) don't provide adequate diagnostics to understand the problem, and simply return biased estimates.

The failure of scaling at large scales is problematic because LRD is an asymptotic large-scale phenomenon, and so it is typically taken as a given that the largest scales in a data set should be included in estimates. Indeed usually it is the choice of the *lower* cutoff, the onset scale of LRD [8] which is problematic and requires attention. In this context however, and others like it where the correlations are simply not present at the largest scales, we must also pay close attention to an *upper cutoff* scale, failing which estimator bias will result. In the figure an estimation is performed over octaves $j = [7, 11]$, a range where the number of corresponding inter-arrival events is adequate. The result, as printed inside the plot, is that the true value of H can be recovered, whereas an estimate over the full range of scales would have been heavily biased. In [6] it is shown how the upper cutoff scale can be precisely predicted for the FRP, and related processes.

The above mechanism is not the only one where under-sampling creates a problem. Consider for example a superposition of FRPs. This will increase the supply of events generating correlations, pushing the cutoff to larger scales. For a sufficiently large number of component FRPs, one might believe that the missing scales problem will vanish. However there is a sense in which it can never be entirely eliminated, since if a data set is of duration T, it cannot access scales beyond T. Furthermore, due to edge effects it cannot access scales very close to T either. Such effects are common in traffic analysis, for example when studying flow durations, there is a fundamental problem in measuring flows which began before the trace did, and those which have yet to terminate by trace's end. These latter causes of under-sampling however are less drastic than the chronic lack of events exemplified by the FRP, and can be addressed to a large extent by the calculation of appropriate confidence intervals and bias corrections. These take into account the steady loss of data as one moves to larger scales.

An obvious question is how to deal with the problem in practice. It should be clear from the above discussion that the problem does not go away simply by using longer traces- the largest scales will always be 'missing'. However, if we fix on a given octave $j = j^*$, then by increasing the trace length we will always be able to eventually access the expected theoretical correlation behaviour at that scale. Unfortunately, for both real and simulated data increasing data length is often not a practical option, in which case the determination of an upper (and as ever, a lower) cutoff scale as described is indispensable.

2.2 The Discontinuity Signature

The process studied by [1] consists of On periods which arrive as a Poisson process, and is a particular example of an infinite aggregate of On-Off sources. As On periods now typically overlap, there is a larger supply of events, as so we expect the missing scales problem to be weaker here. In fact [1] notes a drop off at large scales but describes it as an 'edge effect' and immediately removes those scales from consideration. Instead, they focus on the fact that the (very straight) LD plot over all other scales shows an unexpected slope of $\alpha = 2$, corresponding to an Hurst parameter of $H = 1.5$, far from the expected value, in fact larger than 1! It is not even consistent with stationarity! Since Ricciato ignores the largest scales, it follows that the 'mystery of the missing scales' plays no role in the problem discussed in [1], instead a different phenomenon is at work.

[2]Recall that the LD or 'Logscale Diagram' is a wavelet form of an energy spectral density, in log-log coordinates. As a function of the logarithm $j = \log_2(\text{scale})$ of time-scale, it gives the logarithm of the variance of wavelet coefficients, interpreted as the energy present in the signal as a function of scale. Just as for the usual Fourier spectrum, white noise would be a flat line in this plot.

The straight lines in the LD, and the measured slope of $\alpha = 2$, are in fact the expected behaviour of another scaling signal: an isolated discontinuity. Intuitively a discontinuity is scaling because it remains a discontinuity under arbitrary magnification. Next note that since the spectra of independent signals simply add, a superposition of independent singularities would exhibit the same behaviour. Now a sample path of an On-Off source can be viewed as a function with many discontinuities, corresponding to the rising and falling edges of the On period 'blocks'. As explained in [9], these edges can be approximately viewed as independent over a wide range of scales, giving rise to a strong $\alpha = 2$ signature. The scale range in question corresponds to when edges can be regarded as 'isolated'. Very roughly, this will cease to be the case at scales where On periods will be seen as objects in their own right. Beyond this, the power law scaling of the size of On periods, and the associated LRD, will take over.

This above analysis fits the observations in [1], where the average flow duration of 27[sec] corresponds to scale $j = 11$, which is precisely where the slope begins to drop below 2. In other words, in discarding the largest scales a priori, the range of scales where the LRD 'lives' was also discarded, in favour of a range where the very different scaling law, due to discontinuities, dominates.

Put in context, the observations of [1] reduce again to a matter of cutoff scale, in this case the correct identification of the lower onset scale where LRD 'begins' [8]. In this case, parameter values were used which made this scale quite large, close to the total length of the data. Not only does this make the correct identification of the onset scale difficult, but any attempt to measure the LRD exponent will have to cope both with edge effects as well as the missing scales phenomenon. It may be that once both upper and lower scales are taken into account, there may be precious little data left. This however is not a fault of the LD. On the contrary, the LD provides a powerful tool to observe and estimate the various phenomena associated with cutoff scales.

3. CONCLUSION

We hope to have provided a substantial answer to the questions posed in [1], and [5], and placed them in a broader context of the need when using the LD for both lower and upper cutoff scales whose values depend on the signals themselves.

In so doing, we also hope to show that CCR will (at least in part) fill the role spoken of in [1] for an environment for providing answers to questions of interest to the research community, and for furthering research on those topics. In that vein we should note that although the results cited above clearly explain the observations described in [1, 5], there are still gaps in the understanding of such phenomena, in particular on the relationship between edge effects and their observed impact on LRD in more highly aggregated processes, and how this impacts real network measurements.

4. REFERENCES

[1] F.Ricciato. Some remarks to recent papers on traffic analysis. *ACM Sigcomm Computer Communications Review*, 36(3):99–102, July 2006.

[2] M. M. Krunz and A. M. Makowski. Modeling video traffic using M/G/∞ input processes: A compromise between Markovian and LRD models. *IEEE Journal on Selected Areas in Communications*, 16(5):733–748, June 1998.

[3] J. Levy and M. S. Taqqu. Renewal reward processes with heavy-tailed inter-renewal time and heavy-tailed rewards. *Bernoulli*, 6(1):23–44, 2000.

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

[5] D. Malone, K. Duffy, and C. King. Some remarks on LD plots for heavy-tailed traffic. *ACM Sigcomm Computer Communications Review*, 37(1), January 2007.

[6] M.Roughan, J.Yates, and D.Veitch. The mystery of the missing scales: Pitfalls in the use of fractal renewal processes to simulate LRD processes. In *Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, American University, Washington, DC, June 1999.

[7] D. Veitch and P. 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.

[8] D. Veitch, P. Abry, and M. Taqqu. On the Automatic Selection of the Onset of Scaling. *Fractals*, 11(2), June 2003.

[9] D. Veitch, N. Hohn, and P. Abry. Multifractality in TCP/IP Traffic: the Case Against. *Computer Network Journal, special issue "Long-Range Dependent Traffic"*, 48:293–313, June 2005.

[10] W. Willinger, M. Taqqu, and A. Erramilli. *in Stochastic Networks: Theory and Applications, eds. F. P. Kelly and S. Zachary and I. Ziedins*, chapter A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High-Speed Networks, pages 339–366. Clarendon Press (Oxford University Press), Oxford, 1996.

[11] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. *Proceedings of the ACM SIGCOMM'95*, 1995. Available at `http://www.acm.org/sigcomm/sigcomm95/sigcpapers.html`.