

Some Remarks on LD Plots for Heavy-Tailed Traffic

David Malone,
Hamilton Institute
NUI Maynooth, Ireland
david.malone@nuim.ie

Ken Duffy,
Hamilton Institute
NUI Maynooth, Ireland
ken.duffy@nuim.ie

Christopher King,
Department of Mathematics
Northeastern University, USA
king@neu.edu

ABSTRACT

Ricciato [6] poses several questions, including why a particular LD (log-diagram) plot does not give the Hurst parameter predicted by theory? We offer an explanation of his observation and highlight other unusual aspects of LD plots.

1. INTRODUCTION

In [6] Ricciato constructs two simulated data sets corresponding to the number of active flows in a system over time. During one hour 1,000 flows arrive to the system, uniformly distributed throughout the period. Flow durations are Pareto distributed. The two data sets differ in the Pareto tail-index parameter: $\alpha_1 = 1.25$ and $\alpha_2 = 1.6$. When the number of flows is analysed using [7], the estimate of the Hurst parameter for both data sets was found to be approximately 0.76, rather than $H_1 = 0.875$ and $H_2 = 0.7$, predicted by the well-known relation $H = (3 - \alpha)/2$ ([5]). We suggest that the observed discrepancy may be a more general feature of these estimates and explore other unusual aspects of LD plots.

2. LD ANALYSIS OF ON-OFF TRAFFIC

To consider the simplest version of this problem, we treat a single source of traffic. Its on periods are distributed with a heavy tailed Pareto distribution and its off periods are distributed with a geometric distribution. This is a simplification of the traffic considered in [6], but should display similar LD plots.

We use the wavelet estimator described in [1] to estimate the Hurst parameter of these processes. The code provided at [7] is designed to be general purpose and to operate on real data. Instead, we use an implementation that exploits the constant bursts of an on-off process [4]. This allows us to quickly make calculations even in the region where α is small and long bursts are common.

Figure 1 shows estimates for simulated on-off sources. On periods last for $\lfloor u^{-1/\alpha} \rfloor$ where u is uniform $[0, 1]$, an off period is longer than l with probability 2^{l-1} and each run comprised 1,000,000 on-off cycles. The graph shows Hurst parameter estimates obtained by a least-squares fitted slope to the LD plots between octaves 3 and 15. For each different value of α there are 10 runs. Theory predicts the line shown.

As we can see from Figure 1, the estimates display the correct trend. However, the curve is flatter than the theory predicts, and is pulled towards the middle value of 0.75. Even with 1,000,000 on-off cycles, there is large variability.

Figure 2 shows the same experiment but for 1,000 on-

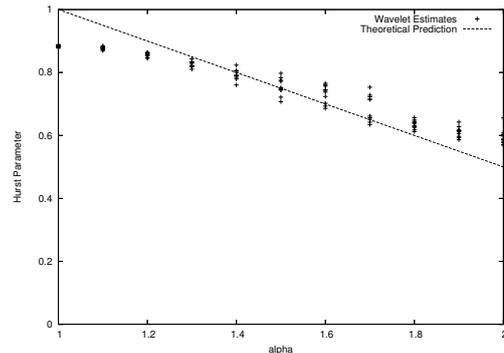


Figure 1: LD plot estimates of the Hurst parameter for Pareto on-off sources with $1 \leq \alpha \leq 2$, 1,000,000 cycles. Ten experiments are shown for each α value.

off cycles, mirroring the amount of data considered in [6]. Here we have fitted the slope over octaves 2 to 6, as this better describes the straight region of the LD plots. By repeating this experiment we find a distribution of Hurst parameter estimates produced by this procedure. Figure 3 shows the results for 10,000 runs with $\alpha_1 = 1.25$ and $\alpha_2 = 1.6$. We can see that the estimates for $\alpha_1 = 1.25$ peak closer to one, confirming the higher Hurst parameter that we expect. We also see a number of estimates outside $[0.5, 1]$, the meaningful range for the Hurst parameter. We include these to demonstrate how noisy these estimates can be.

3. EXPLORING FURTHER

In Figure 1 and Figure 2 there is some indication that if α is smaller then the values we see are more clustered, although the distribution in Figure 3 suggests that we may occasionally get estimates that are far from the peak.

This trend actually continues as we choose α values that are *less than* 1. These processes have on periods drawn from a distribution with infinite mean. They are interesting because, unlike the processes with $1 < \alpha$, no stationary version of them can be constructed. This puts them beyond the range where notions such as the Hurst parameter are usually considered. Regardless, we may simulate such on-off sources and construct their LD plots. The main obstacle to this is the infinite expected burst length associated with $0 < \alpha \leq 1$, which makes the expected run time of such a simulation infinite. However, using the wavelet code adapted for long bursts makes this practical. The resulting

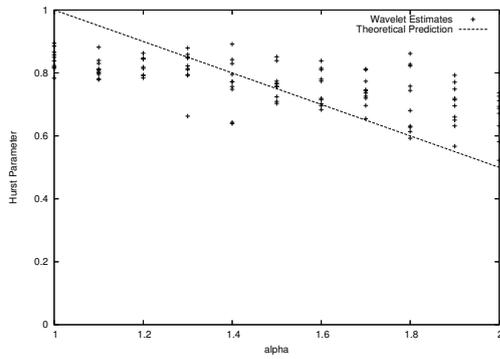


Figure 2: LD plot estimates of the Hurst parameter for Pareto on-off sources with $1 \leq \alpha \leq 2$, 1,000 cycles. Ten experiments are shown for each α value.

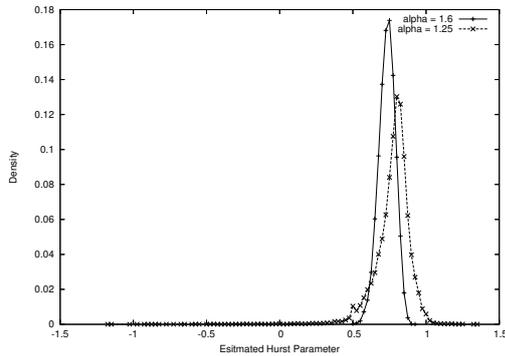


Figure 3: Distribution of LD plot estimates of the Hurst parameter for two α values.

LD plots have linear sections similar to those for $1 < \alpha < 2$. When we estimate the slopes of these lines over octaves 3–15 we get estimates as shown in Figure 4. Theory predicts the ‘hat’ function shown by considering the divergence of a regularised power spectrum [2]. Thus, these estimates do not uniquely identify α , even in theory. Naturally, another statistic such as the tail of the on time distribution obtained by Hill’s estimator [3], could distinguish them. For comparison Figure 5 shows the slopes fitted over octaves 6–18. Note the estimates for small α are closer to the theoretical values, but for larger α they are more scattered.

While processes with infinite mean on time seem exotic, they do arise in networking contexts. For example, an M/M/1 queue whose mean arrival rate equals its mean service rate (i.e. the traffic intensity $\rho = 1$ so the queue is on the boundary of stability), has on-off output with $\alpha = 0.5$ [2].

4. CONCLUSIONS

Using a simple Pareto on-off source we have briefly considered the variability of estimating the Hurst parameter. We have seen that with small numbers of on-off periods, LD estimates can be highly variable. Even with large amounts of data, estimates are flattened towards the middle value of 0.75. This may help to explain the results seen in [6]. We also note that when the α value is extended beyond that

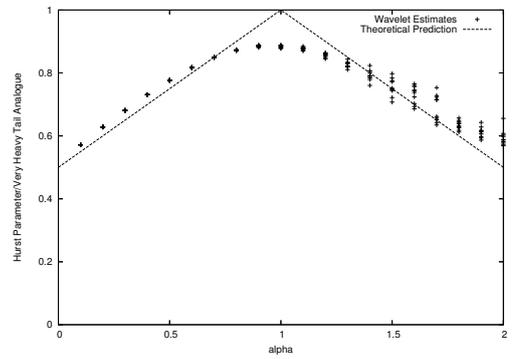


Figure 4: LD plot estimates of the Hurst parameter for Pareto on-off sources with $0 < \alpha \leq 2$, 1,000,000 cycles. Ten experiments are shown for each α value.

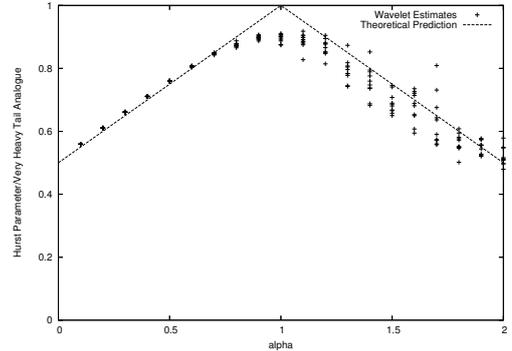


Figure 5: As Figure 4, but fitting over octaves 6–18.

usually considered, the estimator takes an unexpected turn.

5. REFERENCES

- [1] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2–15, 1998.
- [2] K. Duffy, C. King, and D. Malone. Ambiguities in estimates of critical exponents for long-range dependent processes. *Physica A*, to appear.
- [3] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.
- [4] D. Malone. Burst wavelet C code. <http://www.hamilton.ie/dwmalone/wavelet/>.
- [5] K. Park and W. Willinger. *Self-similar Network Traffic and Performance Evaluation*. Wiley, 2000.
- [6] F. Ricciato. Some remarks to recent papers on traffic analysis. *ACM CCR*, 36(3):99–102, 2006.
- [7] D. Veitch and P. Abry. Matlab code for wavelet analysis. http://www.cubinlab.ee.mu.oz.au/~darryl/secondorder_code.html.