Timeseries, Self-Similarity and Network Traffic

Mark Crovella

March 28, 1996

1 Background

1.1 Timeseries

Consider a set of n measurements of some particular aspect of a computer system. These may be considered to be samples or observations of random variables $X_i, 0 \leq i < n$.

One often hears the statement that certain random variables are *idependent,* which means that knowledge of the behavior of one random variable gives one no knowledge of the behavior of any other. In addition, if the random variables are *identically distributed,* then it is impossible to tell one random variable X_i from another $X_j, i \neq j$, so we often just drop any distinction and refer to a set of random variables X that are i.i.d. This is the model that captures the essence of flipping a coin.

However, in practice, it's often the case that random variables are *not* i.i.d. In particular we're going to consider the case in which the random variables have an *ordering.* The ordering will correspond to time passing, and we will assume that past history will influence present events. This is a very natural and useful assumption for computer systems, because after all, any computer's behavior is determined by past events as well as present.

An ordered set of random variables $\{X_t, t \in T\}$ is called a *stochastic process.* Any particular set of measurements is a set of samples from these random variables; these are called *realizations* of the process. We use the term *timeseries* to refer either to a stochastic process or to one of its realizations.

Any realization defines a function on T; The index set T will almost always be the integers Z. Using Z as our index set implies that a realization would correspond to a set of measurements made at equally spaced intervals. However, most of our discussion would apply to more general indexing as well (such as irregularly spaced measurements, or continuous measurements).

1.2 Autocorrelation and Stationarity

Since a timeseries is distinguished by its ordering among random variables, and the possibility of dependence among those random variables, we need to quantify the existence and nature of dependence.

> **Definition.** If $\{X_t, t \in T\}$ is a process such that $\mathrm{Var}(X_t) < \infty$ for each $t \in T$, then the *autocovariance function* $\gamma(.,.)$ of $\{X_t\}$ is defined by
>
> $$\gamma_X(r, s) = \mathrm{Cov}(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], \quad r, s \in T.$$

This gives us a measure of the correlation between two components of the series based on their placement within the series.

Definition. The time series $\{X_t, t \in Z\}$ is said to be *stationary* if

$$E|X_t|^2 < \infty \quad \text{for all } t \in Z,$$

$$EX_t = m \quad \text{for all } t \in Z, \text{ and}$$

$$\gamma_X(r, s) = \gamma_X(r + t, s + t) \quad \text{for all } r, s, t \in Z.$$

This is also referred to as *convariance stationarity* or *wide-sense stationarity.* It means that the series has constant mean, finite variance, and that the autocovariance of the series only depends on the distance between the two random variables selected.

If $\{X_t, t \in Z\}$ is stationary then $\gamma_X(r, s) = \gamma_X(r - s, 0)$ for all $r, s \in Z$. For this reason we redefine the autocovariance function of a stationary process as the function of just one variable, the "lag" k:

$$\gamma_X(k) = \gamma_X(k, 0) = \mathrm{Cov}(X_{t+k}, X_t) \quad \text{for all } t, k \in Z.$$

The autocovariance depends only on the lag, so $\gamma_X(0)$ (which is the variance of any random variable in the timeseries) is constant. So the series has constant variance as well as constant mean.

If a series consists of independent r.v.'s then its autocovariance is everywhere zero, except at lag 0, where it is equal to the variance of the r.v.'s.

Definition. The *autocorrelation function* (acf) of $\{X_t\}$ is defined as the fuction whose value at lag k is

$$\rho_X(k) = \gamma_X(k)/\gamma_X(0) = \mathrm{Corr}(X_{t+k}, X_t) \quad \text{for all } t, k \in Z.$$

1.3 Motion vs. Increment Series

Stationary series are useful because of their simple correlation structure. But notice that in our context, a stationary series is describing a set of measurements that isn't "going anywhere:" its mean and variance are constant, so it's not a good model for many characteristics of computer systems (like the number of users on the system over 24 hours, or the number of pages in the World Wide Web).

However, what makes stationary series so important is that, usually, we consider any set of real measurements to consist of *trend* and *noise.* The trend may be easy to account for – in the case of users on the system, we can assume there is a periodic function with period 24 hours, find its contribution to mean, and subtract it. What we are left with is *noise* – random fluctuations, typically with mean zero.

In the case of network traffic, we assume stationarity in many cases (and that assumption is often open to debate). Typically we may look at an hour's worth of traffic data under the assumption that the average load on the network (the mean of the timeseries) isn't changing too much over the span of an hour. So, often we assume that a traffic timeseries is stationary. This assumption is virtually always made in simulations, where it is called the "steady-state" assumption.

One particular way in which a noise process can relate to a non-stationary process is through its *motion* process. From a timeseries $\{X_t, t \in 0, 1, 2, \ldots\}$ form:

$$Y_t = \sum_{i=0}^{t} X_i.$$

Then Y_t is called the *motion* process of X_t and X_t is called the *increment* process of Y_t. If X_t is stationary then we say the Y_t has *stationary increments.*

2 Self-Similarity

To begin discussing self-similarity we'll start with an important motion process: *Brownian motion.* Brownian Motion can be defined in any number of dimensions, but in our work we'll define it as a timeseries, that is, as a motion process on the line.

The increments of Brownian motion are i.i.d. Gaussian r.v.'s. That is,

$$B(t) = \sum_{i=0}^{t} G_t$$

in which the G_ts are i.i.d. Gaussian.

Brownian motion has a number of interesting properties. It is a motion process with stationary increments. Furthermore, its increments are not just stationary, but independent. Therefore, if we take any point $t_1 > 0$, $B(t)$ $t > t_1$ forms a new Brownian motion process, with a different mean. Also, it can be shown that for any value x (in particular 0), $B(t)$ takes on the value x infinitely often as $t \to \infty$. Thus $B(t)$ "fills" the line. The result is the basis for "Gambler's ruin" which says that a gambler starting with a fixed amount of money, betting on a truly random outcome, will eventually lose all the money with probability 1.

Brownian motion has another curious property, which was the source of its original attraction to scientists: if you look at a sample realization of Brownian motion, it is impossible to tell the scale at which one is observing without knowing the absolute magnitude of the values the process is taking. Brownian motion was first observed (in three dimensions) by the Scottish botanist Robert Brown in the late 1820s, in the motion of pollen grains or other particles suspended in water. A complete explanation for the strange characteristics of Brownian motion was finally provided by Einstein in 1905, who argued that it was the result of collisions between the water molecules and the pollen or other particle, and who gave its first convincing mathematical treatment.

The scale-invariance of Brownian motion is similar to the scale-invariance of *fractals.* A fractal is an object with a well-defined fractal dimension different from its Euclidean dimension. One way of measuring fractal dimension is by measuring a sample path of the obejct; the expected measure of the sample path scales with the size of the radius. We'll not go into fractals too much except to note that a fractal, like the Mandelbrot set that is seen on posters and screensavers, has the property that one cannot determine the scale of observation from the shape of the object: any view could be a look at the whole object, or at some tiny portion thereof that has been magnified.

2.1 Self-Similarity of a Motion Process

For a stochastic process, we can't define self-similarity in the same way we do as for fractals, because stochastic processes don't have fixed realizations. Instead we define self-similarity in terms of *equality in distribution.* This is a way of equating two random variables: their mean, variance, and all higher moments are equal (which means that their distributions are exactly the same).

> **Definition.** Let $\{Y_t\}$ be a stochastic process with continuous time parameter t. Then we say that Y_t is H-self-similar if for any positive stretching factor c, the rescaled process with time scale ct, which is $c^{-H}Y_{ct}$ is equal in distribution to the original process Y_t. That is,
>
> $$Y_t =_d c^{-H} Y_{ct} \quad \text{for all } c > 0. \tag{1}$$
>
> If Y_t has stationary increments then it is H-sssi.

Brownian motion appears to be self-similar; if so, what is its parameter H? Take the case in which the mean of $B(t)$ is 0, and the mean of the increment process is 1. Well, the mean of $B(ct)$ is still 0, so that moment stays constant for any H. Consider the variance of $B(t)$, the expected value of $B^2(t)$. It turns out that $EB^2(t) = t$.[1] For the second moment of $B(t)$ to be equal to the rescaled second moment of $B(ct)$, what must the value of H be?

$$EB^2(t) = E[(c^{-H}B(ct))^2] = E[c^{-2H}B^2(ct)] = c^{-2H}EB^2(ct).$$

Now,

$$\begin{aligned} EB^2(t) &= t, && \text{and,} \\ EB^2(ct) &= ct, && \text{so, substituting:} \\ t &= c^{-2H}ct, && \text{so:} \\ H &= 1/2. \end{aligned}$$

So, Brownian motion behaves like a sequence of random variables that scale with the square root of time. That is, the standard deviation of $B(t)$ is proportional to $\sqrt{t}$.

2.2 Self-Similarity of a Stationary Process

Now, Brownian motion is not stationary, and as we said at the outset, we are mainly interested in stationary processes. As a result, we will create an analagous definition of self-similarity for stationary processes.

Definition. Let $\{X_t\}$ be a stationary process with continuous time parameter t. Then we say that X_t is H-self-similar if for any positive aggregation factor m the series obtained by m-aggregating X_t is equal in distribution to X_t. That is:

$$X_t =_d m^{-H} \sum_{i=(t-1)m+1}^{tm} X_i \quad \text{for all } m \in N \tag{2}$$

Basically, what this definition is saying is that we will call a stationary process self-similar if its corresponding motion process shows scaling in time. Thus, the operation of aggregating or summing a stationary self-similar process brings us into the domain where we can observe the scaling directly.

Since the aggregation of a self-similar process leads to effects that show scaling, we will define the m-aggregation of a stationary process formally:

Definition. If $\{X_t\}$ is a stationary process, we define its m-aggregation, as:

$$X_t^{(m)} = \sum_{i=(t-1)m+1}^{tm} X_i$$

with corresponding autocorrelation function $\rho^{(m)}(k)$.

This means that we form $X_t^{(m)}$ simply by summing X_t in nonoverlapping blocks of size m.

[1] This is actually the key fact to remember about Brownian motion.

2.3 The Correlation Structure of Self-Similar Processes

So far, our only example of a self-similar process has been Brownian motion, which has independent increments. Now we consider the effect of the presence of dependence between the increments of a self-similar process.

Consider an H-self-similar, stationary sequence $\{X_t\}$ in which the X_t's are not independent. Thus the sequence must have an acf $\rho(h)$ which is nonzero for some $h \neq 0$. Now, note that $\{X_t\}$ must have the same acf as $m^{-H}\{X_t^{(m)}\}$, by Equation 2. It's possible to show [1, pp. 51f] that this implies that the resulting form for ρ is a power-law; in particular:

$$\rho(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \;\; k \geq 0$$

which is the exact form of $\rho(k)$; its asymptotic behavior is

$$\lim_{k \to \infty} \rho(k) = H(2H-1)k^{2H-2}$$

Note that for the increment process of Brownian motion, the fact that $H = 0.5$ implies that the acf is everywhere zero, which is what we expect, since the increments of Brownian motion are independent.

However, the important thing to realize about self-similar stationary processes is that if there is *any* correlation between X_t's, then that correlation *must* take the form of a power-law. This is the central connection between self-similarity and long-range dependence.

3 Long-Range Dependence

Now we turn to looking at the effects of dependence in a timeseries. Traditional methods of simulation and analysis of computer networks have either considered independent series, or series with exponentially decaying autocorrelation functions.

Consider the Poisson process, which is the most commonly used model for an arrival process. It models the arrivals in a system with a constant rate, but in which no previous arrival affects any subsequent arrival. The system is "memoryless" and interarrival times are independent; the acf of such a process is zero except at lag 0, where it is 1.

Or consider a process in which each random variable is equal to some uncorrelated noise, plus a fraction δ of the previous value. This has an exponentially decaying acf, proportional to δ^k.

A timeseries with exponentially decaying acf is said to exhibit *short-range dependence.* For large values of k, the dependence between r.v.'s is negligible.

On the other hand, timeseries with power-law decaying acf's are said to exhibit *long-range dependence.* There is a non-negligible correlation at all lags k, and so between all r.v.'s in the timeseries. To be more precise:

> **Definition.** A stationary process is said to be long-range dependent if, for some constant $0 < \beta < 1$ and some finite positive constant c_ρ, its acf obeys
>
> $$\rho(k) = c_\rho |k|^{-\beta} \quad \text{as } |k| \to \infty.$$

For a self-similar series $\rho(k) \sim k^{2H-2}$, so for the self-similar series to be long-range dependent, since $2H - 2 = -\beta$, it must the case that $1/2 < H < 1$.

Long-range dependence has a number of implications for the behavior of the timeseries. First, note the relationship between the variance of an aggregated series, and its mean.

The mean of $X^{(m)}$ grows like m, while the standard deviation grows like m^H. For an uncorrelated self-similar process, this means that the coefficient of variation (stanard deviation divide by mean) declines like $1/\sqrt{m}$. However for $H = 1$, the coefficient of variation is independent of m! That is, there is no "apparent" decrease in variability at all as the series is aggregated. For values of H in between 0.5 and 1, the coefficient of variation declines like m^{-H}. This means that aggregating a stationary self-similar series produces a new series with only slightly decreased variability.

Why is aggregation important? It corresponds to what happens when we measure a traffic flow at increasing degrees of granularity; and it corresponds to considering a system involving flows at increasing timescales. In the case of network traffic, variability is equivalent to "burstiness." If a flow is highly variable, we perceive it as having significant bursts.

More fundamentally, note that both the original series and all of its aggregations have long-range dependence. So correlations persist at all aggregation levels.

This can be seen in another way: in terms of variance of a sample mean.

$$Var(\overline{X}) = \sigma^2 n^{-1}$$

for short-range or independent series, but for long range dependent series:

$$Var(\overline{X}) = \sigma^2 n^{-\beta}$$

3.1 Tests for Long-Range Dependence

1. The Variance-Time Plot.
2. The Periodogram Plot.
3. The R/S Statistic.
4. Whittle's MLE Estimator.

3.2 Examples of LRD Series

1. Nile River data.
2. VBR Video/MPEG data.
3. Ethernet Traffic.

4 Presence of LRD in Network Traffic

Leland Et. Al, SIGCOMM '92 .

5 Implications of LRD for Network Performance

Since LRD implies burstiness on all levels, one conclusion that can be formed is that "smoothing doesn't work" for an LRD series.

In the context of networks, smoothing can be seen as the goal of buffering. Bursts at all timescales implies that buffer sizes must grow without bound as the timeseries grows. This is in contrast to buffering for SRD series, in which a fixed buffer size can eliminate virtually all burstiness (= packet drops).

Looking at the effects of self-similarity on network performance, there are three metrics of interest: packet loss, buffer occupancy (*i.e.,* packet delay), and throughput.

A number of papers have shown:

- With increasing buffer capacity, the cell loss probability decreases very slowly, not exponentially fast as in Markovian models.
- Cell average delay (buffer occupancy) *always* increases with buffer capacity. This is confirmed experimentally for TCP traffic in [6]. This is in contrast to Markovian models where delay does not exceed a fixed limit regardless of buffer size. As a corollary, with an infinite buffer, the expected number of items in the buffer is infinite.

Also, in [6] we show that throughput is not severely degraded by long-range dependence (typically, at most about 5-10% decline in throughput can be attributed to LRD).

5.1 Queueing Network Analysis of Self-Similar Traffic

The most straightforward method of modeling such a network is G/D/1, in which self-similar traffic demand G is serviced by a constant-rate network D.

This is shown to be equivalent to M/G/1, where the G represents a heavy-tailed service time, when the number of sources grows large but the intensity of the arrival process stays constant, and the individual streams are kept off for longer periods to diminish their individual rates (makes sense) [5].

For example, in [5], it is shown that for an M/G/1 queue, the cell loss probability when the queue length is fixed at L, is proportional to $L^{1-\alpha}$. For α close to 1, this means that cell loss declines very slowly with buffer size.

However, in [7], we show that this is not how it actually works in practice. In a real network, instead of moderating traffic by increasing OFF times, rather the transmission rate during the ON time is diminshed, and the ON times are expanded.

6 Causes of LRD in Networks

6.1 File Sizes in the WWW

[2]

6.2 File Sizes in General

[7]

References

[1] Jan Beran. *Statistics for Long-Memory Processes.* Monographs on Statistics and Applied Probability. Chapman and Hall, New York, NY, 1994.

[2] Mark E. Crovella and Azer Bestavros. Self-similarity in World Wide Web traffic: Evidence and possible causes. In *Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, May 1996.

[3] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of ethernet traffic. In *Proceedings of SIGCOMM '93*, pages 183–193, September 1993.

[4] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2:1–15, 1994.

[5] Nikolai Likhanov, Boris Tsybakov, and Nicolas D. Georganas. Analysis of an ATM buffer with self-similar ("fractal") input traffic. In *Proceedings of* INFOCOM '95, pages 985–992, 1995.

[6] Kihong Park, Gi Tae Kim, and Mark E. Crovella. The effects of traffic self-similarity on tcp performance. Technical report, Boston University Computer Science Department, 1996.

[7] Kihong Park, Gi Tae Kim, and Mark E. Crovella. On the relationship between file sizes, transport protocols, and self-similar network traffic. Technical report, Boston University Computer Science Department, 1996.