Presentation Outline

1. Self similarity in nature
2. Quick review of autocorrelation
3. Definition of self-similar discrete process
   - Exactly/asymptotic self-similar
   - Long range vs short range dependence
4. Measures of burstiness
5. Determining presence of self-similarity
6. Implications in computer networks

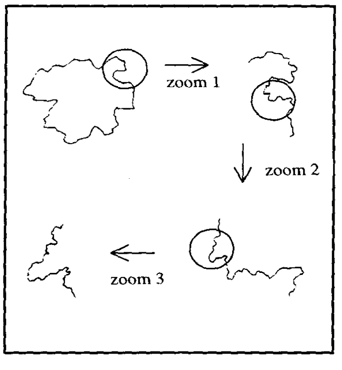
Self-Similarity Defined

- Self-similarity is the unifying concept for the theories of fractals and chaos.
- A phenomenon that is self-similar looks the same or behaves the same when viewed at different degrees of magnification or different scales on a dimension. The dimension can be **space** (length, width) or **time**.

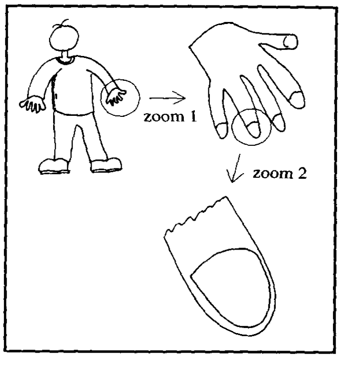
Fractals in Nature

- The term “fractal” was coined by Menoit Mandelbrot.
- A fractal is an object that appears self-similar under varying degrees of magnification. It possess symmetry across scale, with each small part of the object replicating the structure of the whole.
- Fractals can be a mathematical construct, but they also abound in nature.
- We can speak of
  – Statistical self-similarity: coastline (paradox: length boundary is function of measuring unit), crack in wall
  – Exact (geometric) self-similarity: fern, spiral, binary tree
- Fractals in nature do no exhibit self-similarity over all time scales; self-similarity eventually breaks down.

Examples of Fractals

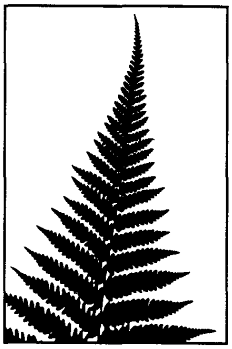


Coastline

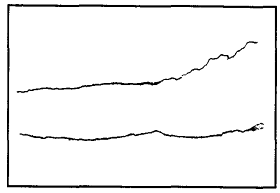


Person

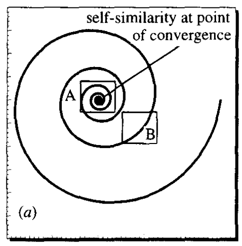
Examples of Fractals



Fern tip



Wall cracks



Review: Autocorrelation

$$E[X1 * X2] = E[X(t) * X(t+\tau)] = Rxx(\tau)$$

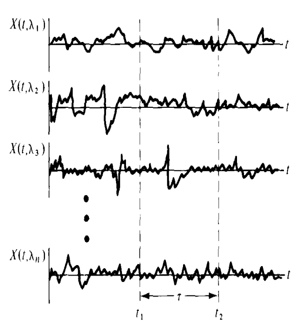
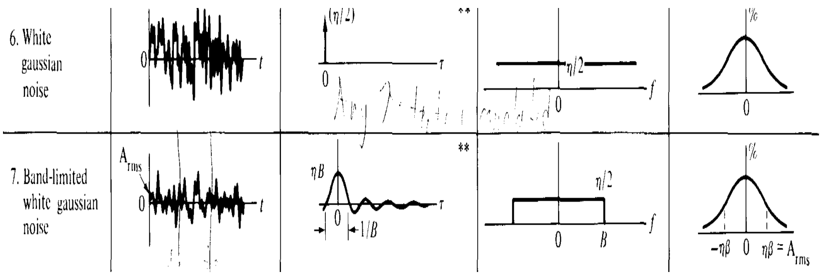


Figure 8.18 An ensemble of continuous random waveforms.

- Autocorrelation: second joint moment of the process. It is a measure of how dependent a particular value of a sample function is on another value that is removed tao units of time.
- Stationary: the mean is independent of time, and the autocorrelation function depends only on the time difference

Review: Autocorrelation Examples

process time trace autocorrelation power density PDF



Definition: Self Similar Process

Let X = (Xt: t = 0,1,2,…) be a *covariance stationary* stochastic process with mean μ, variance σ^2, and autocorrelation function r(k), K>=0. In particular, we assume that X has an autocorrelation function of the form

$r(k) \sim k^{-\beta} L(t)$ as $k \to \infty$, (1)

where $0 < \beta < 1$ and L is slowly varying at infinity, i.e., $Lim_{t\to\infty} L(tx)/L(t) = 1$, for all x>0. For each m=1,2,3,…, let $X^{(m)} = (X_k^{(m)}$: k=1,2,3,…) denote the new covariance stationary time series (with corresponding autocorrelation function $r^{(m)}$) obtained by averaging the original series X over non-overlapping blocks of size m. That is, for each m=1,2,3,…, $X^{(m)}$ is given by $X_k^{(m)} = 1/m(X_{km-m+1} + \ldots + X_{km})$, k>=1.

Definition: Self Similar Process

The process X is called **exactly second-order self-similar** with self-similarity parameter $H = 1 - \beta/2$ if for all m=1,2,..., $\mathrm{var}(X^{(m)}) = \sigma^2 m^{-\beta}$ and $r^{(m)}(k) = r(k)$, k>=0.

X is called **asymptotically second-order self-similar** with self-similary parameter $H = 1 - \beta/2$ if for all k *large enough*

$r^{(m)}(k) \to r(k)$ as $m \to \infty$

with r(k) given by (1).

Definition: Self Similar Process

An interesting feature is that the autocorrelation of the aggregated process does not go to zero as $m \to \infty$. This is in contrast to stochastic processes traditionally used for packet data models where $r^{(m)}(k) \to 0$ as $m \to \infty$; as the level of aggregation increases, the process resembles white noise.

Note that the **variance** of the aggregated process does go to zero, but it does so at a slower rate than a stationary stochastic process. For a stationary stochastic process, B=1 and the variance decays at a rate of $1/m^1$. For a self-similar process, the variance of the aggregated process decays more slowly, at $1/m^{\beta}$. There is **persistence** of the statistical properties across time scales.

Definition: Self Similar Process

Mathematically, self-similarity manifests itself in a number of **equivalent** ways:

(i) slowly decaying variances: the variance of the sample mean decreases more slowly than the reciprocal of the aggregation sample size, m.

$\mathrm{var}(X^{(m)}) \sim a_2 m^{-\beta}$ as $m \to \infty$, with $0 < \beta < 1$

(ii) the autocorrelations decay hyperbolically rather than exponentially fast.

In a short range-dependent process the autocorrelation decays at least as fast as exponential. In a long range dependent process the autocorrelation decays hyperbolically.

(iii) the spectral density obeys a power law (pole) near the origin.

$$S(w) = \sum_{k=-\infty}^{\infty} R(k) e^{-j2kw}, S(0) = \sum_{k=-\infty}^{\infty} R(k)$$

An infinite value results at S(0) if the values of R(k) do not decay sufficiently rapid for large K to form a finite sum. This can be useful for testing for self-similarity.

Measures of Burstiness

- H. E. Hurst (hydrologist) spent a lifetime studying the Nile and other rivers and problems related to water storage. Hurst discovered that levels of the Nile River over an 800 year period obeyed a self-similar pattern. In the short term, there were year-to-year fluctuations. In the long term, there were long periods when droughts were followed by long periods of flooding.
- Hurst examined a number of different phenomena and developed a normalized, dimensionless measure to characterize variablility: R/S statistic

Measures of Burstiness: H

- H, the Hurst parameter, or self-similarity parameter, is a key measure of self-similarity. H is a measure of the persistence of a statistical phenomenon and is a measure of the length of the long-range dependence of a stochastic process.
  - H = 0.5 indicates absence of self-similarity
  - H → 1 indicates the degree of persistence or long-range dependence.

Measures of Burstiness: IDC

- A commonly used measure for capturing the variability of traffic over different time scales is provided by the **index of dispersion for counts (IDC)**.
- For a given time interval L, IDC is given by the variance of the number of arrivals during the interval L, divided by the expected value on the same interval. SS traffic produces a monotonically increasing IDC. For Poisson, IDC is either constant or converges rapidly to a constant.
- Estimating IDC and plotting it in log-log provides a quick and simple approach to test for SS.

Measures of Burstiness: Peak to Mean

- Any possible peak to mean ratio is possible depending on the length of the measurement interval.
- The dependence of the burstiness measure on choice of time interval is undesirable.

Determining Presence of SS

- Three of the most common approaches to determine whether a given time series of actual data is self-similar, and if so, estimate H are
  - Variance-time plots
  - R/S (rescale adjusted range) statistic
  - Frequency domain: periodogram + Whittle's

Determining Presence of SS

- The variance-time plot and R/S plot are heuristic or "eyeballing" methods. These two methods are used to test whether a time series is self-similar and if so to obtain a rough estimate for H. The Whittle Estimator assumes the time series is from a self-similar process of a particular form and provides an estimate of H with confidence intervals.

Determining Presence of SS: variance-time plots

For a self-similar process, the variance of the aggregated process $X^{(m)}$ decays more slowly (at $1/m^{\beta}$) than that of a stationary stochastic process.

Variance-time plots are obtained by plotting log(var($X^{(m)}$) against log(m) ("time") and by fitting a least-square line through the resulting points ignoring those for small m. Values of the estimate B' of the asymptotic slope between −1 and 0 suggest self similarity, and an estimate for the degree of self-similarity is given by H' = 1 − B'/2. Slowly decaying variances (shallower than −1) indicate slowly decaying autocorrelation (self-similarity), or possibly, non-stationarity (caution!).

Recall that $Var(X^{(m)}) = \sigma^2/m^{\beta}$, if we take the log of this

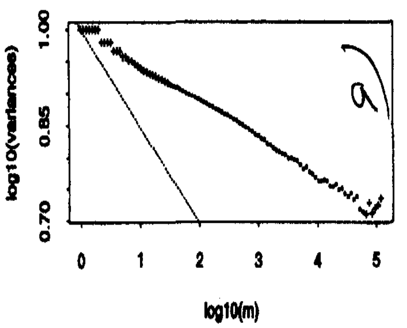
$\log[Var(X^{(m)})] = \log[\sigma^2] - \beta\log[m]$,

the resulting log-log plot as a function of m should be a straight line with a slope of $-\beta$.

Determining Presence of SS: variance-time plots

- From [LELAND 94] for an Ethernet trace collected in 1989 consisting 360k observations:
- Variance-time curve has been normalized by dividing by the sample variance
- Slope is estimated to be about B'=-0.4 resulting in an estimate

H'=1-B'/2~0.8



Determining Presence of SS: R/S statistic

For a given set of observations (Xk: k=1,2,...) the *rescaled adjusted range statistic* (or R/S-statistic) over a time interval N is given by

$$\frac{R(N)}{S(N)} = \frac{\max_{1\le j\le N} \sum_{k=1}^{j} (X_k - M(N)) - \min_{1\le j\le N} \sum_{k=1}^{j} (X_k - M(N))}{\sqrt{\frac{1}{N} \sum_{J=1}^{N} (X_k - M(N))^2}}$$

where M(N) is the sample mean over the time period N.

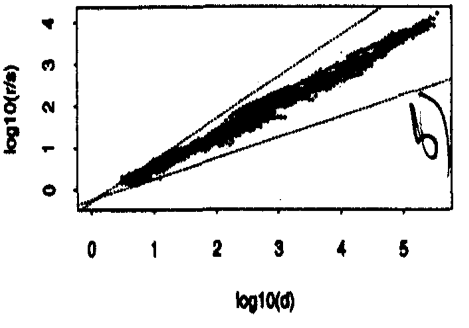
The numerator in this ratio is a measure of the range of the process and the denominator is the sample standard deviation. Note that the range is not the same as the range of the time series. Instead, we can view each of the terms in the numerator as the accumulated amount by which the time series has departed from the mean up to time k. Both the numerator (R) and the denominator (S) measure the variability of the data; R "looks" at the data linearly, whereas S is based on the squared data.

Determining Presence of SS: R/S statistic

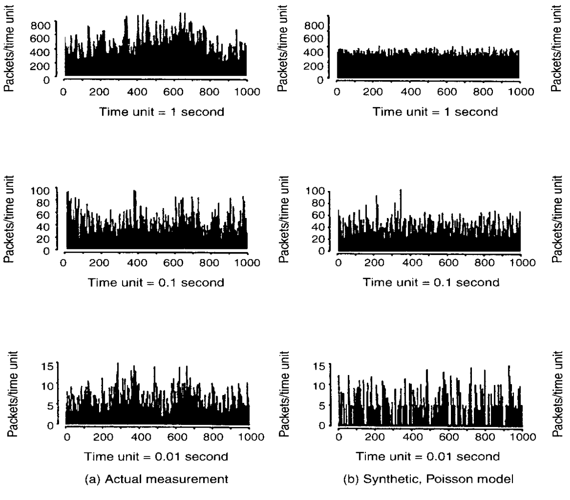
- Graphical R/S analysis consists of taking logarithmically spaced values of N (starting with N~10), and plotting log(R(N)/S(N)) versus log(N) results in the rescaled adjusted range plot. If the data is well defined self-similar, an estimate H' of H is given by the street's asymptotic slope which can take any value between ½ and 1.

Determining Presence of SS: R/S statistic

- From [LELAND 94] for an Ethernet trace collected in 1989 consisting of 360k observations:
- The value of the asymptotic slope of the R/S plot is between ½ and 1 (upper and lower dotted lines)
- Least-square fit results in H~0.79



Implications in Computer Networks

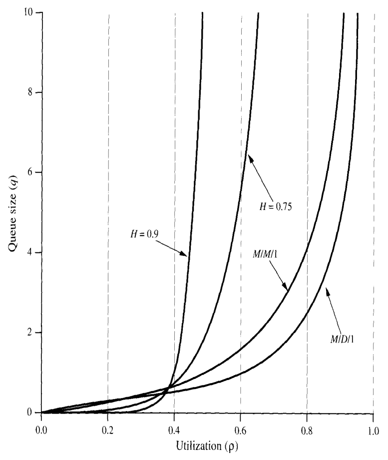


Implications in Computer Networks

- Observations from self-similar packet traffic:
  - The aggregated time-series do not resemble second-order white noise.
  - There is no “natural” burst size. The traffic is bursty over a wide range of time-scales.
  - The auto correlations of the aggregated series remains the same or tend to be the same as that of the original series over a wide range of time scales.
  - The variances of the aggregated processes decrease shallower than 1/m
  - The rescaled adjusted range increases faster than n^0.5

Implications in Computer Networks

- Assuming poisson traffic leads to modest-sized queue buffers because traffic “smooths out” over the long term, despite burstiness in “short-term”. A queue may build up on short run, but over a longer period the buffers are cleared out. However, if traffic bursts are themselves bursty, queue sizes may build up more than would be expected resulting in overflows.



Implications in Computer Networks

- Observations from LELAND (1994) paper based on very high quality traces (20us time resolution over 4 years) and rigorous statistical analysis
  - Ethernet LAN traffic (packet arrival count) is self-similar irrespective of time of measurement.
  - The Hurst parameter, H, is a function of the overall utilization of the Ethernet and can be used for measuring the burstiness of the traffic. The higher the traffic, the higher H.
  - This paper proved that queuing analysis using Poisson traffic assumption is not adequate to model network traffic

References

- On the self-similar nature of Ethernet traffic (extended version), Leland, W.E.; Taqqu, M.S.; Willinger, W.; Wilson, D.V. Networking, IEEE/ACM Transactions on , Volume: 2 Issue: 1 , Feb. 1994. Page(s): 1 -15
- On the relevance of long-range dependence on network traffic, Mattias Grossglauser, Jean-Chrisostome Bolot, IEEE/ACM Transactions on , Volume: 7 Issue: 5 , Oct. 1999 Page(s): 629 -640
- High Speed Networks TCP/IP and ATM design principles, William Stallings, Prentice Hall, 1998, Chapter 8 "Self-similar traffic"
- Introduction to Communication Systems, 3rd Edition, Ferrel Stremler, Addison Wesley, 1990.
- Fractals and Chaos An Illustrated Course, Paul S Addison, Institute of Physics, London, 1997.