Self-similar Distributions

David L. Mills
University of Delaware
http://www.eecis.udel.edu/~mills
mailto:mills@udel.edu

Sir John Tenniel; *Alice’s Adventures in Wonderland*, Lewis Carroll
Minimize effects of serial port hardware and driver jitter

- Graph shows raw jitter of millisecond timecode and 9600-bps serial port
  - Additional latencies from 1.5 ms to 8.3 ms on SPARC IPC due to software driver and operating system; rare latency peaks over 20 ms
  - Latencies can be minimized by capturing timestamps close to the hardware
  - Jitter is reduced using median filter of 60 samples
  - Using on-second format and median filter, residual jitter is less than 50 µs
Measured PPS time error for Alpha 433

Time Offset (us)

Standard error 51.3 ns
Clock filter performance

- Left figure shows raw time offsets measured for a typical path over a 24-hour period (mean error 724 µs, median error 192 µs).
- Right graph shows filtered time offsets over the same period (mean error 192 µs, median error 112 µs).
- The mean error has been reduced by 11.5 dB; the median error by 18.3 dB. This is impressive performance.

27-Apr-04
Measured PPS time error for Alpha 433

![Graph showing time error](image)

Standard error 51.3 ns
Performance with a modem and ACTS service

- Measurements use 2300-bps telephone modem and NIST Automated Computer Time Service (ACTS)
- Calls are placed via PSTN at 16,384-s intervals
Huff & puff wedge scattergram

Offset (ms) vs. Delay (ms)
The traces show the cumulative probability distributions for:

- Upper trace: raw time offsets measured over a 12-day period
- Lower trace: filtered time offsets after the clock filter
Roundtrip delays

- Cumulative distribution function of absolute roundtrip delays
  - 38,722 Internet servers surveyed running NTP Version 2 and 3
  - Delays: median 118 ms, mean 186 ms, maximum 1.9 s(!)
  - Asymmetric delays can cause errors up to one-half the delay
Allan deviation characteristics compared

![Graph showing Allan deviation characteristics for different processors: SPARC IPC, Pentium 200, Alpha 433, and the Resolution limit. The x-axis represents Time Interval (s) on a logarithmic scale, and the y-axis represents Allan Deviation (PPM) on a logarithmic scale. The graph illustrates how these characteristics compare over a range of time intervals.]
Allan deviation calibration

[Graph showing Allan deviation over time intervals]
Self-similar distributions

- Consider the (continuous) process $X = (X_t, \ -\infty < t < \infty)$
- If $X_{at}$ and $a^H(X_t)$ have identical finite distributions for $a > 0$, then $X$ is self-similar with parameter $H$.
- We need to apply this concept to a time series. Let $X = (X_t, \ t = 0, 1, \ldots)$ with given mean $\mu$, variance $\sigma^2$ and autocorrelation function $r(k)$, $k \geq 0$.
- It’s convenient to express this as $r(k) = k^\beta L(k)$ as $k \to \infty$ and $0 < \beta < 1$.
- We assume $L(k)$ varies slowly near infinity and can be assumed constant.
**Self-similar definition**

- For $m = 1, 2, \ldots$ let $X^{(m)} = (X_k^{(m)}, k = 1, 2, \ldots)$, where $m$ is a scale factor.

- Each $X_k^{(m)}$ represents a subinterval of $m$ samples, and the subintervals are non-overlapping: $X_k^{(m)} = 1 / m (X^{(m)}_{(k-1)m} + \ldots + X^{(m)}_{km-1})$, $k > 0$.

- For instance, $m = 2$ subintervals are (0,1), (2,3), \ldots; $m = 3$ subintervals are (0, 1, 2), (3, 4, 5), \ldots

- A process is (exactly) self-similar with parameter $H = 1 - \beta / 2$ if, for all $m = 1, 2, \ldots$, $\text{var}[X^{(m)}] = \sigma^2 m - \beta$ and

  $$r^{(m)}(k) = r(k) = 1 / 2 ([k + 1]^{2H} - 2k^{2H} + [k - 1]^{2H}), k > 0,$$

  where $r^{(m)}$ represents the autocorrelation function of $X^{(m)}$.

- A process is (asymptotically) second-order self-similar if $r^{(m)}(k) \to r(k)$ as $m \to \infty$. 

27-Apr-04
Properties of self-similar distributions

- For self-similar distributions ($0.5 < H < 1$)
  - Hurst effect: the rescaled, adjusted range statistic is characterized by a power law; i.e., $E[R(m) / S(m)]$ is similar to $m^H$ as $m \to \infty$.
  - Slowly decaying variance: the variances of the sample means are decaying more slowly than the reciprocal of the sample size.
  - Long-range dependence: the autocorrelations decay hyperbolically rather than exponentially, implying a non-summable autocorrelation function.
  - $1/f$ noise: the spectral density $f(.)$ obeys a power law near the origin.

- For memoryless or finite-memory distributions ($0 < H < 0.5$)
  - $\text{var}[X^{(m)}]$ decays as to $m^{-1}$.
  - The sum of variances is finite.
  - The spectral density $f(.)$ is finite near the origin.
Origins of self-similar processes

- Long-range dependent $(0.5 < H < 1)$
  - Fractional Gaussian Noise (F-GN)
    \[
    r(k) = \frac{1}{2} ([k + 1]^{2H} - 2k^{2H} + [k - 1]^{2H}), \ k > 1
    \]
  - Fractional Brownian Motion (F-BM)
  - Fractional Autoregressive Integrative Moving Average (F-ARIMA)
  - Random Walk (RW) (discrete Brownian Motion (BM))

- Short-range dependent
  - Memoryless and short-memory (Markov)
  - Just about any conventional distribution – uniform, exponential, Pareto
  - ARIMA
Examples of self-similar traffic on a LAN
R/S plot
Periodogram plot