Traffic Modeling

Including material modified from:
Queuing Theory and Traffic Analysis
Richard Martin, Rutgers University
and
Carey Williamson
Department of Computer Science
University of Saskatchewan

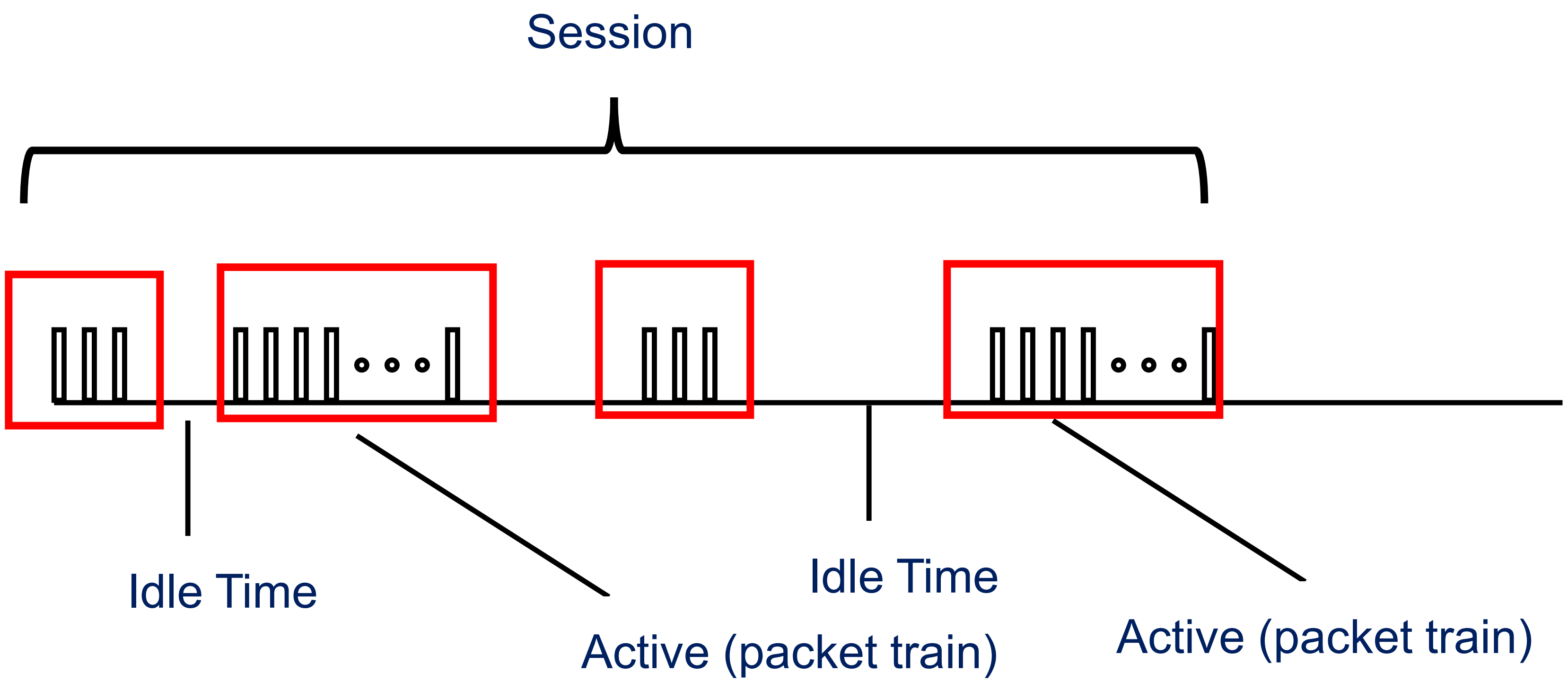
Traffic

- Network supports the transmission of packets

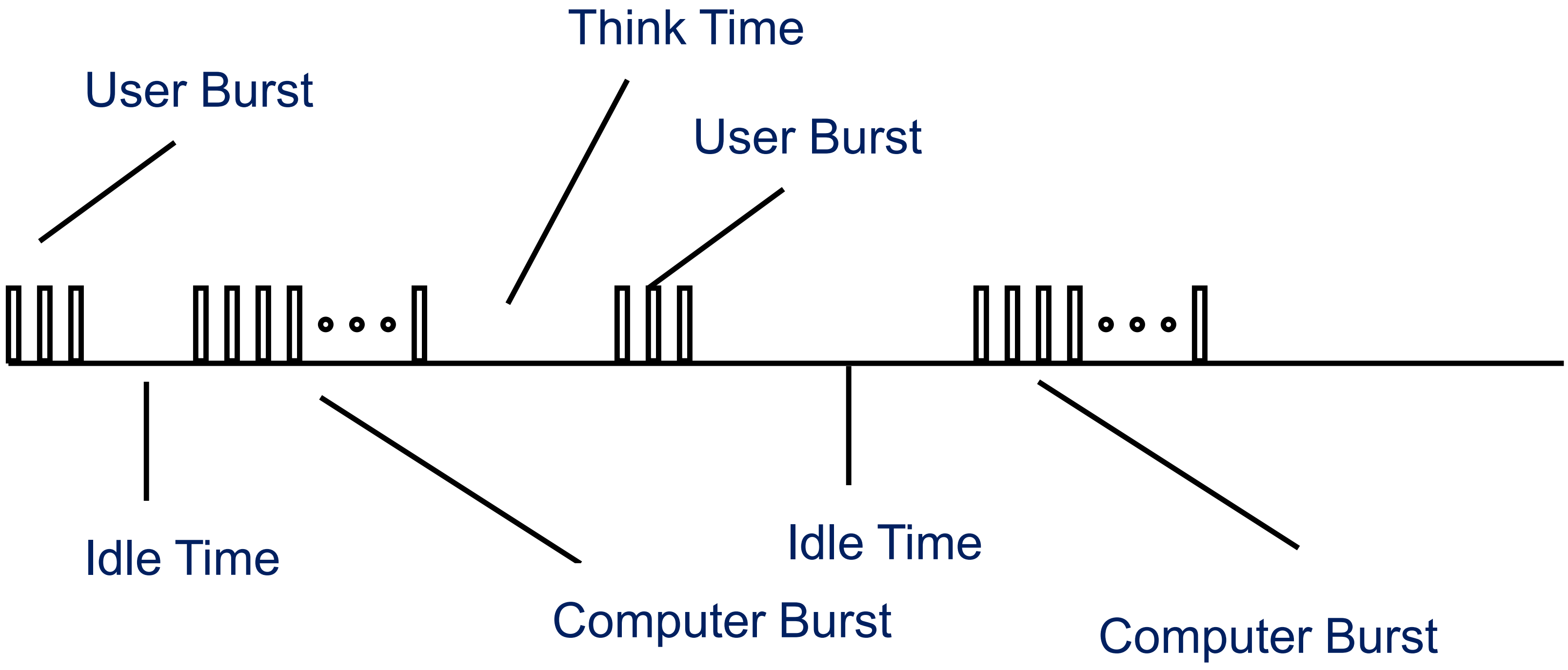
  –traffic-

  from:
  - Many sources
  - Many applications
    - High performance computers
    - IoT
  - Many protocols
    - TCP
    - UDP
  - Levels
    - Session
    - Active/Idle periods within a session
    - Packets during an active period

Traffic levels



Asymmetric Nature of Interactive Traffic



This Asymmetric property has lead to asymmetric services

Flows

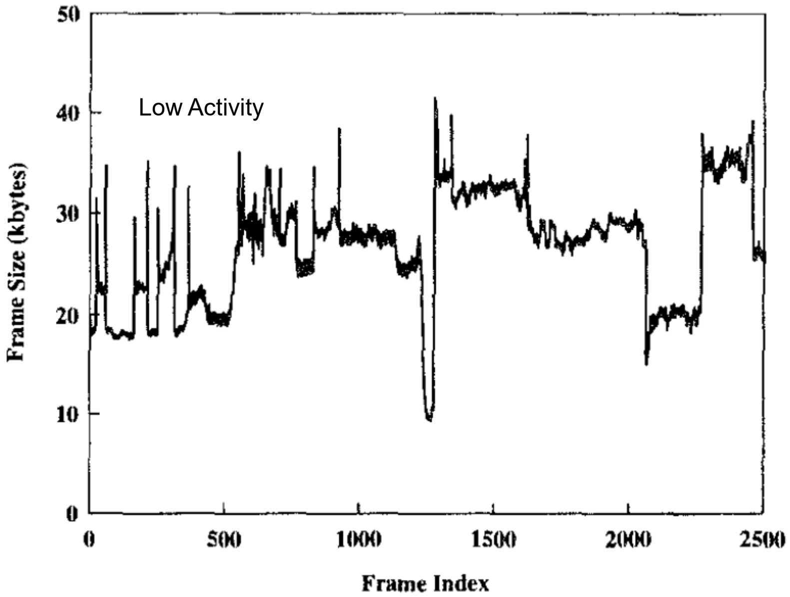
- Set of packets with a common set of properties passing a measurement point.
- Example Properties
  - IP packet header
  - IP & TCP header (e.g. application traffic from a source)
  - MPLS label
  - Layer 2 header (MAC header)
- Control traffic, e.g.,
  - DNS
  - Routing - BGP

Sample Realization of an Traffic Process

Message number	1	2	3	4	5	6	7	8	9	10	11	12
Interarrival time between i+1 and i message (seconds)	2	1	3	1	1	4	2	5	1	4	2	--
Length of i^{th} message (seconds)	1	3	6	2	1	1	4	2	5	1	1	3

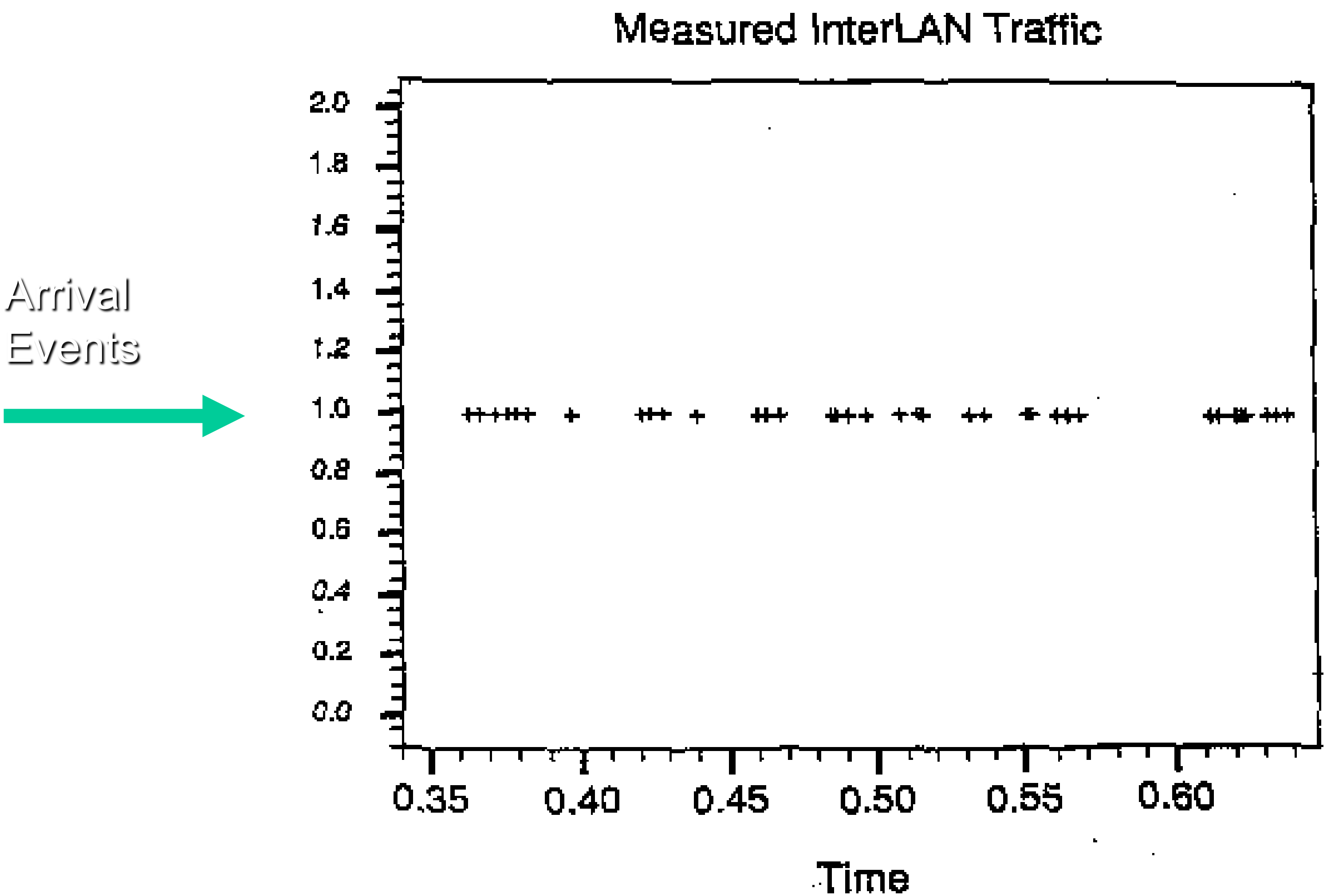
Arrival Events & Lengths

Video Traffic



A TES-based model for compressed "Star Wars" video, B. Melamed. D. Pendarakis, 1994 IEEE GLOBECOM. Communications: Communications Theory Mini-Conference Record

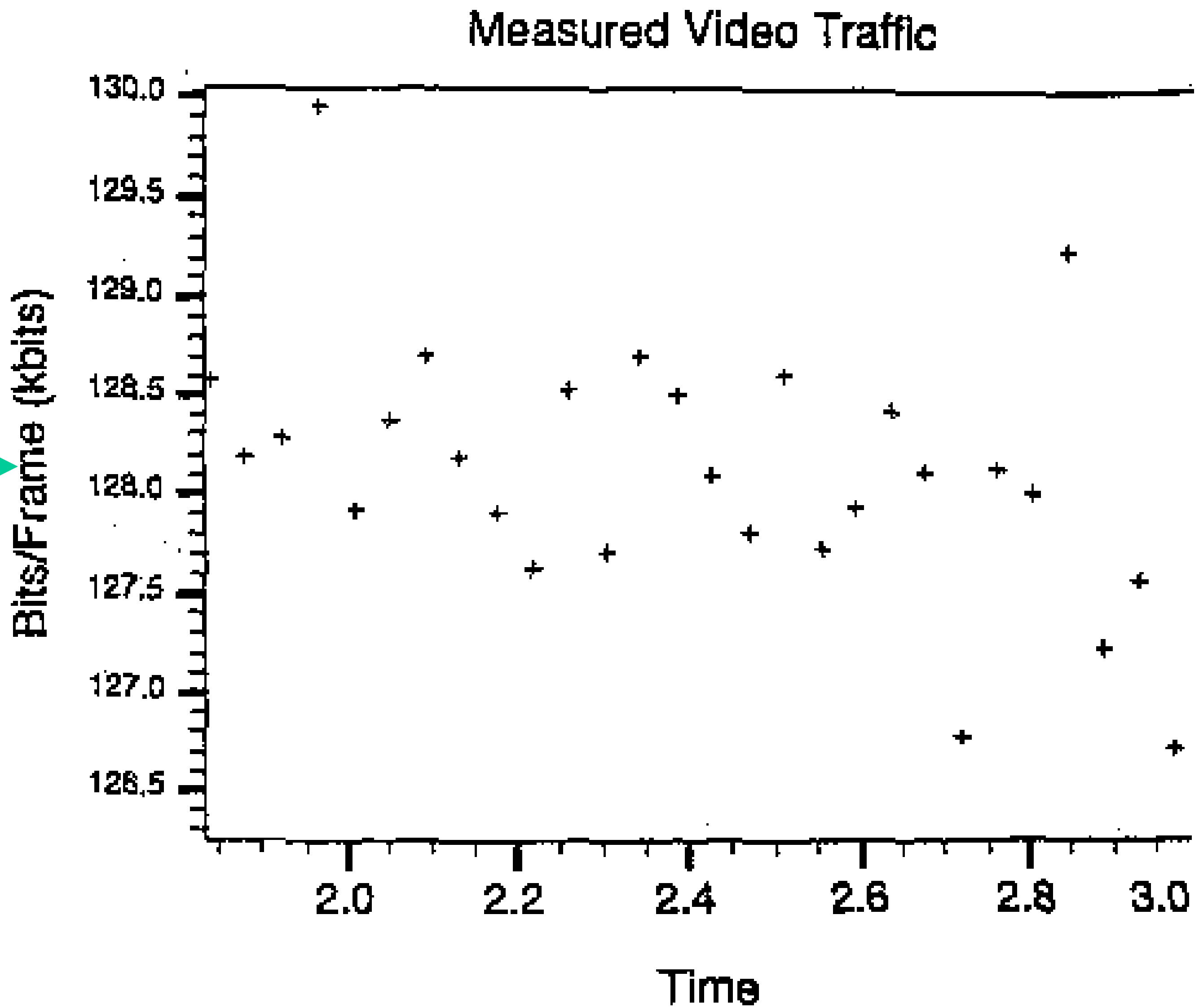
Traffic



Measured InterLAN Traffic
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.35
0.40
0.45
0.50
0.55
0.60
Time
Arrival
Events

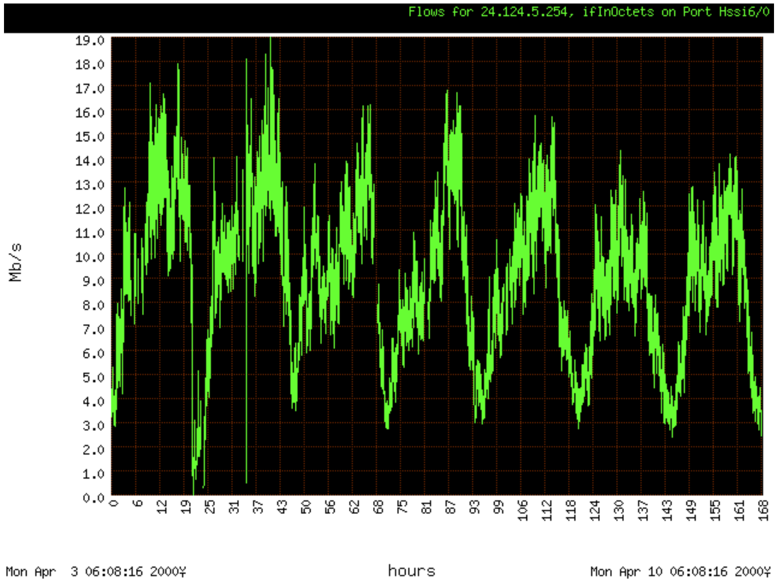
Traffic

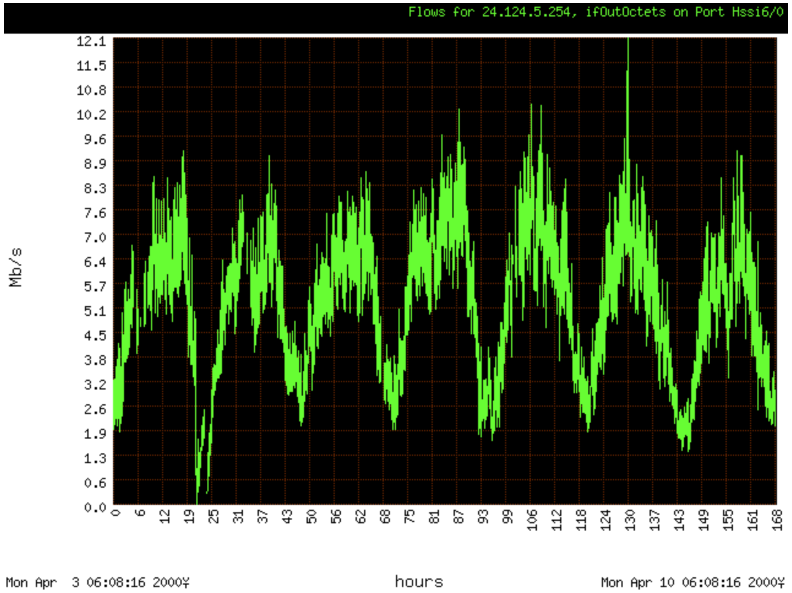
Arrival Events & Lengths



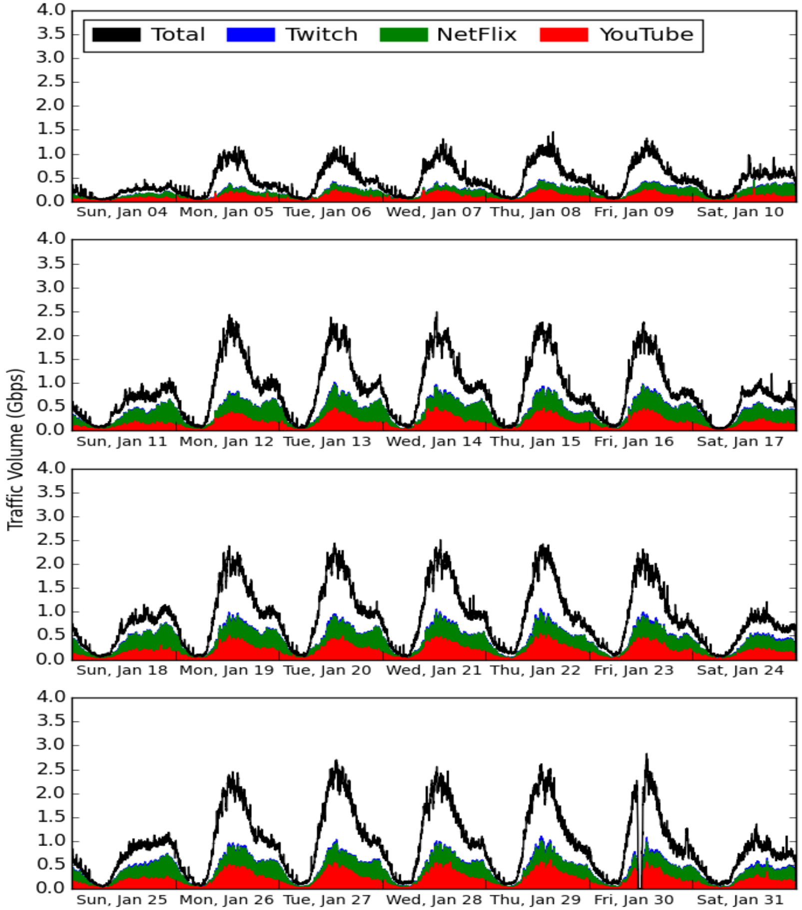
Traffic

Time of day variations





Video Traffic



- January 2015
- Top line (Total) is HTTP+HTTPS
- Red is (HTTPS) YouTube
- Green is NetFlix
- Blue is Twitch

From: NetFlix : Traffic Characterization, Michel Laterman Department of Computer Science, University of Calgary: https://pages.cpsc.ucalgary.ca/~carey/CPSC641/slides.html

Overview

- Traffic consists of single arrivals of discrete entities (packets)
- Arrival times are $T_1, T_2, T_3, .. T_n, \ldots$
- The arrival process is a random process called a Point Process
- The arrival of each discrete entity carries a "length" $L_1, L_2, L_3, .. L_n, \ldots.$
- The interarrival time process is

$$A_1, A_2, A_3, .. A_n..., \text{ where } A_n = T_n - T_{n-1}$$

Overview

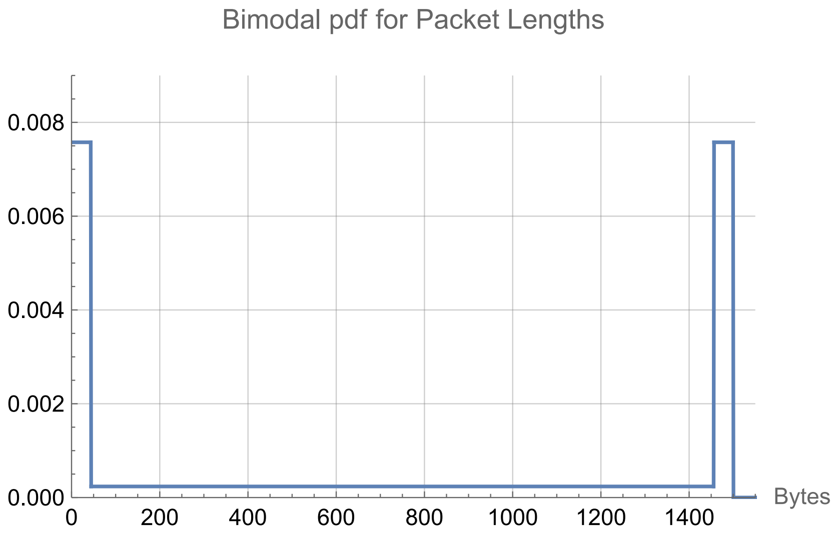
- Typical assumptions
  - Poisson arrival and departure processes results in: exponential distributions for the A_n's and L_n's
  - The sequence of A_n's are statistically independent
  - The sequence of L_n's are statistically independent
  - The sequences of A_n's and L_n's are statistically independent of each other.
- Advanced traffic models do not use some of these assumptions
- Purpose of this discussion is the introduce other traffic models.

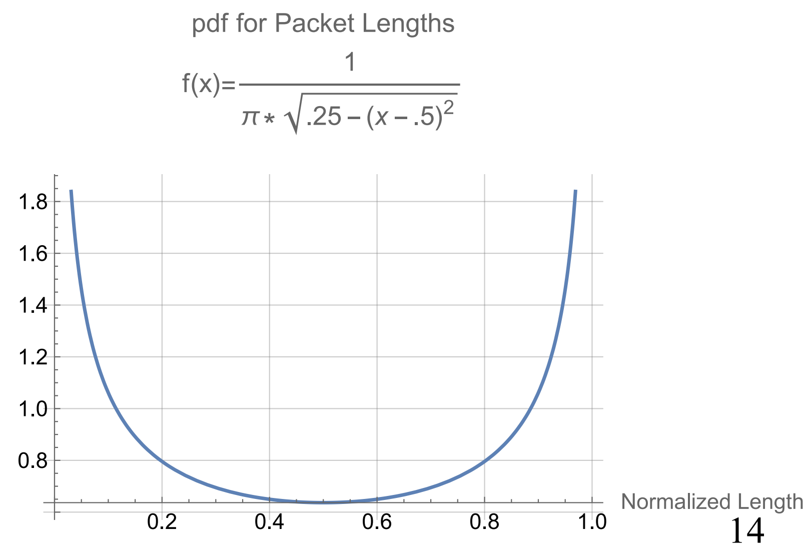
Example Message Length Models

- Fixed length
- Uniform with minimum/maximum packet sizes.
  - Minimum from MAC and Minimum MTU.
  - Maximum from MAC ~1500 bytes.
- Pareto
  - File size distribution (in bytes) using FTP (File Transfer Protocol) fits a Pareto distribution with
- $0.9 < \alpha < 1$ (more later)

Example Message Length Models

- Bimodal
  - 40% are of size smaller than 44 bytes
  - 40% of the packets are between 1400 bytes and 1500 bytes (w/o MAC headers) From: W. John and S. Tafvelin, "Analysis of internet backbone traffic and header anomalies observed," in IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, New York, NY, USA,2007, pp. 111–116.





Correlation

Autocorrelation of a random sequence X_n is

$$\rho_X(k) = \frac{E[(X_n - E[X_n])(X_{n+k} - E[X_{n+k}])]}{\sigma_X^2}$$

k=lag

- Correlation is a statistical measure of the relationship, if any, between two random variables
- Positive correlation: both behave similarly
- Negative correlation: behave as opposites
- No correlation: behavior of one is unrelated to behavior of other

Correlation

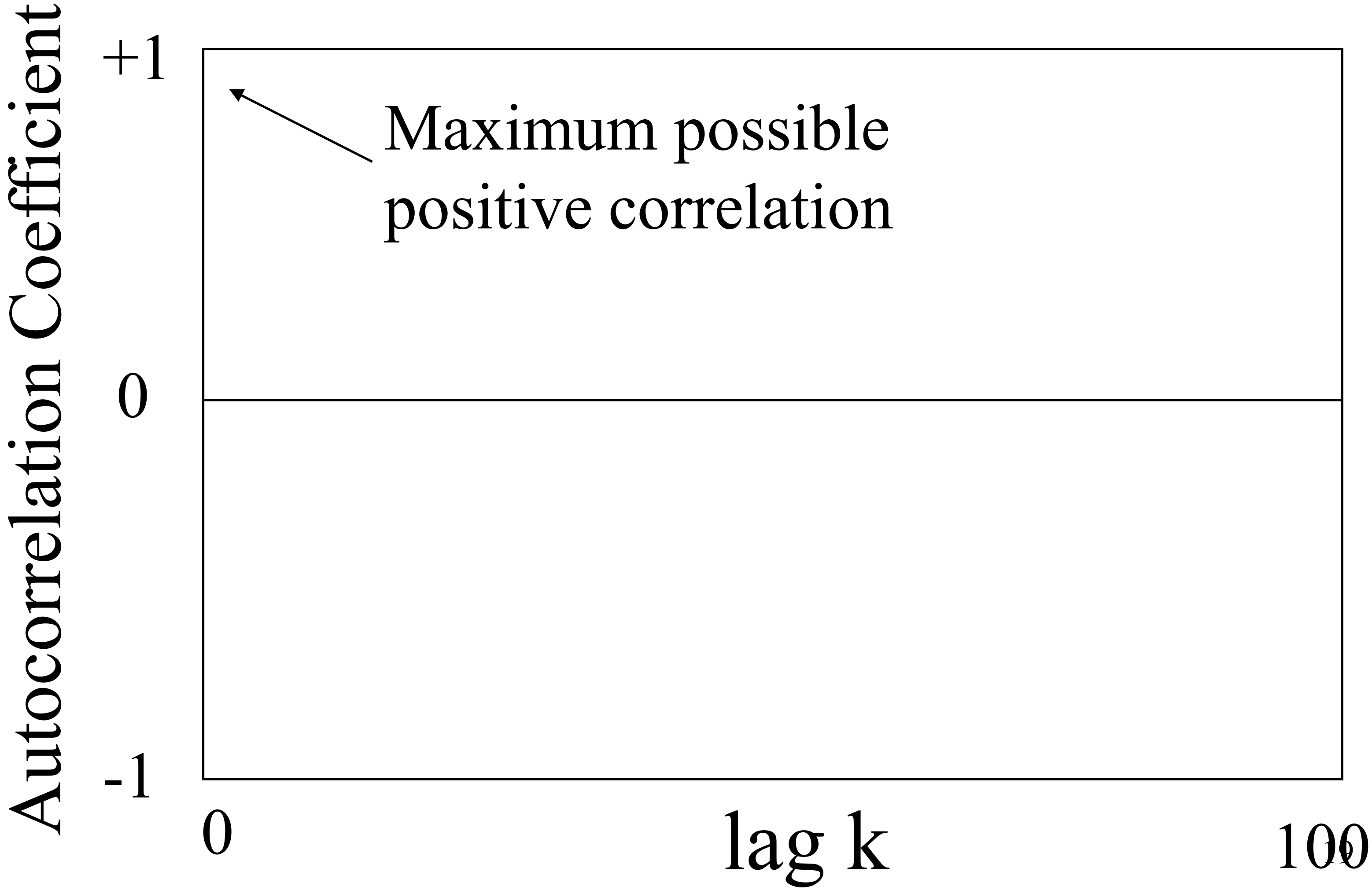
- Autocorrelation is a statistical measure of the relationship, if any, between a random variable and itself, at different time lags
- Positive correlation: big observation usually followed by another big, or small by small
- Negative correlation: big observation usually followed by small, or small by big
- No correlation: observations unrelated

Correlation

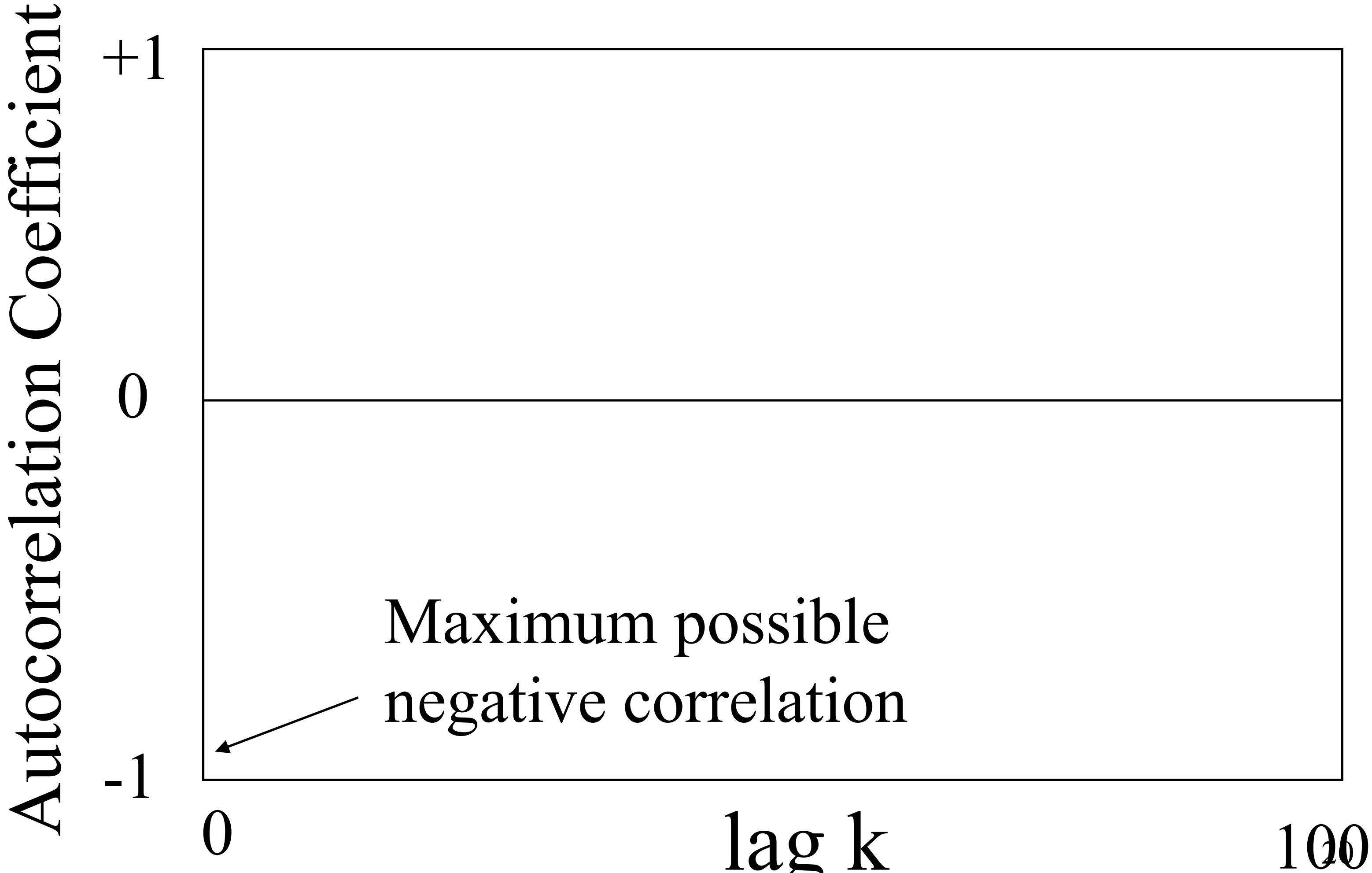
- Autocorrelation coefficient can range between:
  - +1 (very high positive correlation)
  - -1 (very high negative correlation)
- Zero means no correlation
- Autocorrelation function shows the value of the autocorrelation coefficient for different time lags k

Autocorrelation Function

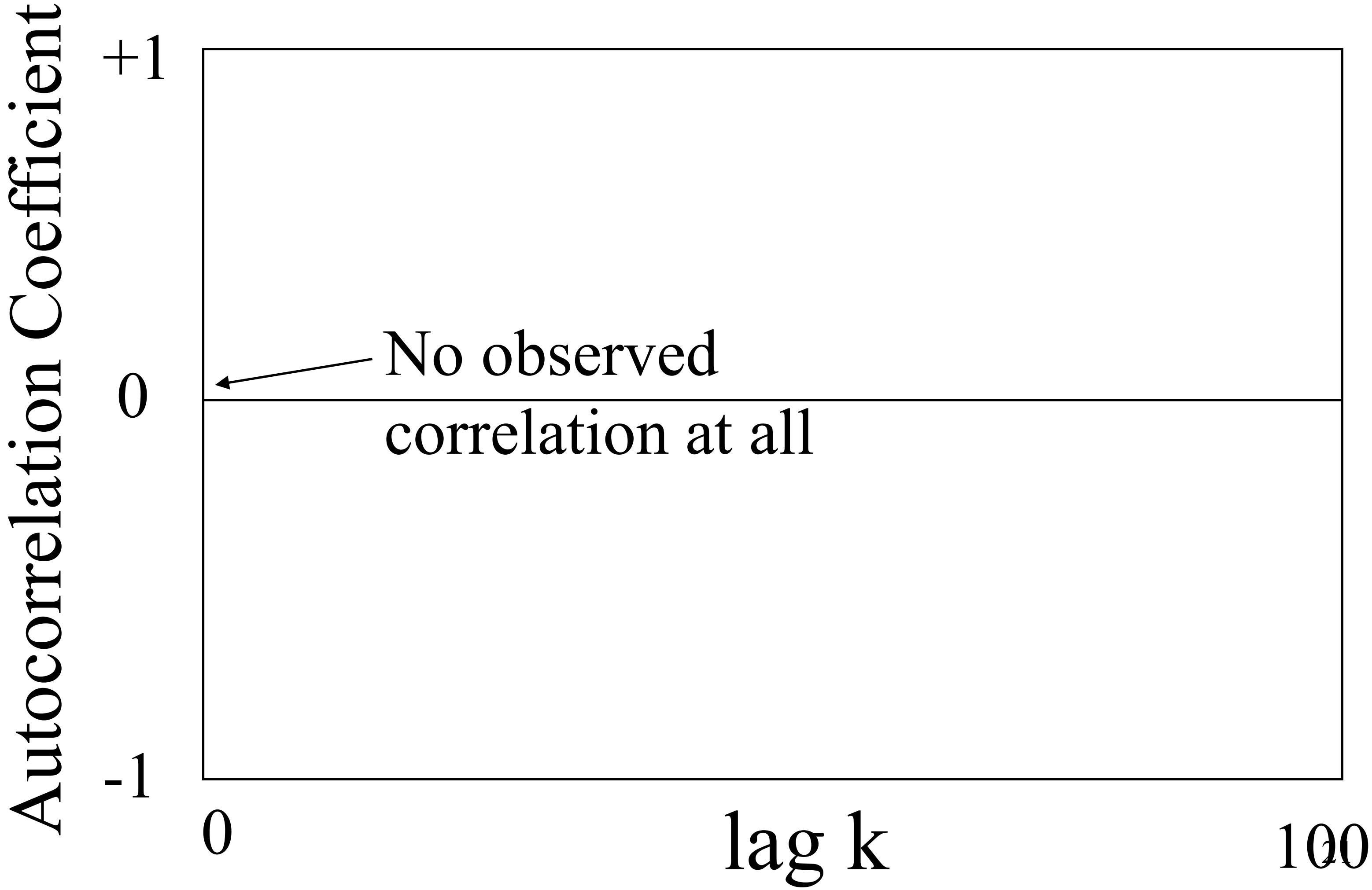
Autocorrelation Function



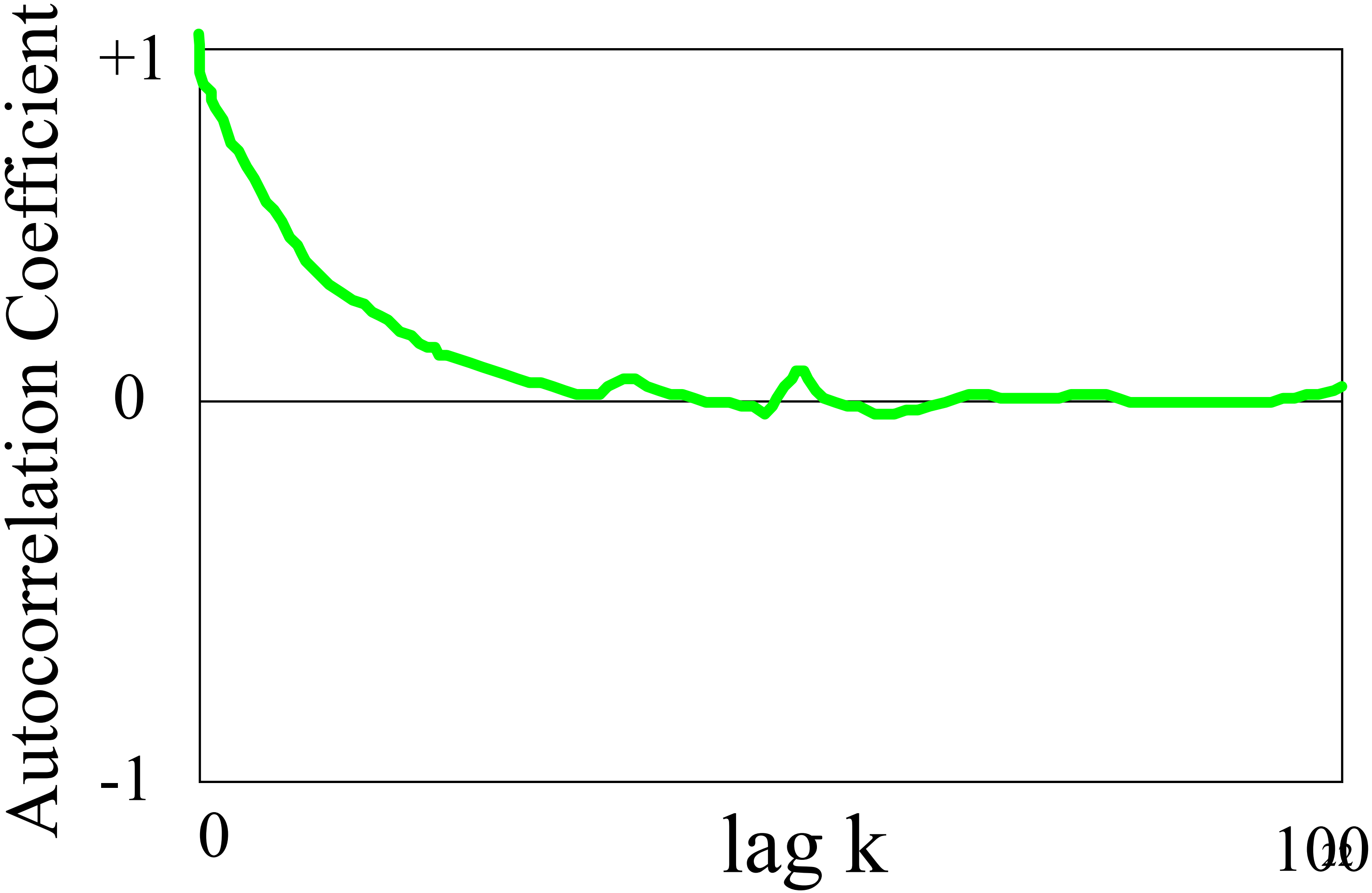
Autocorrelation Function



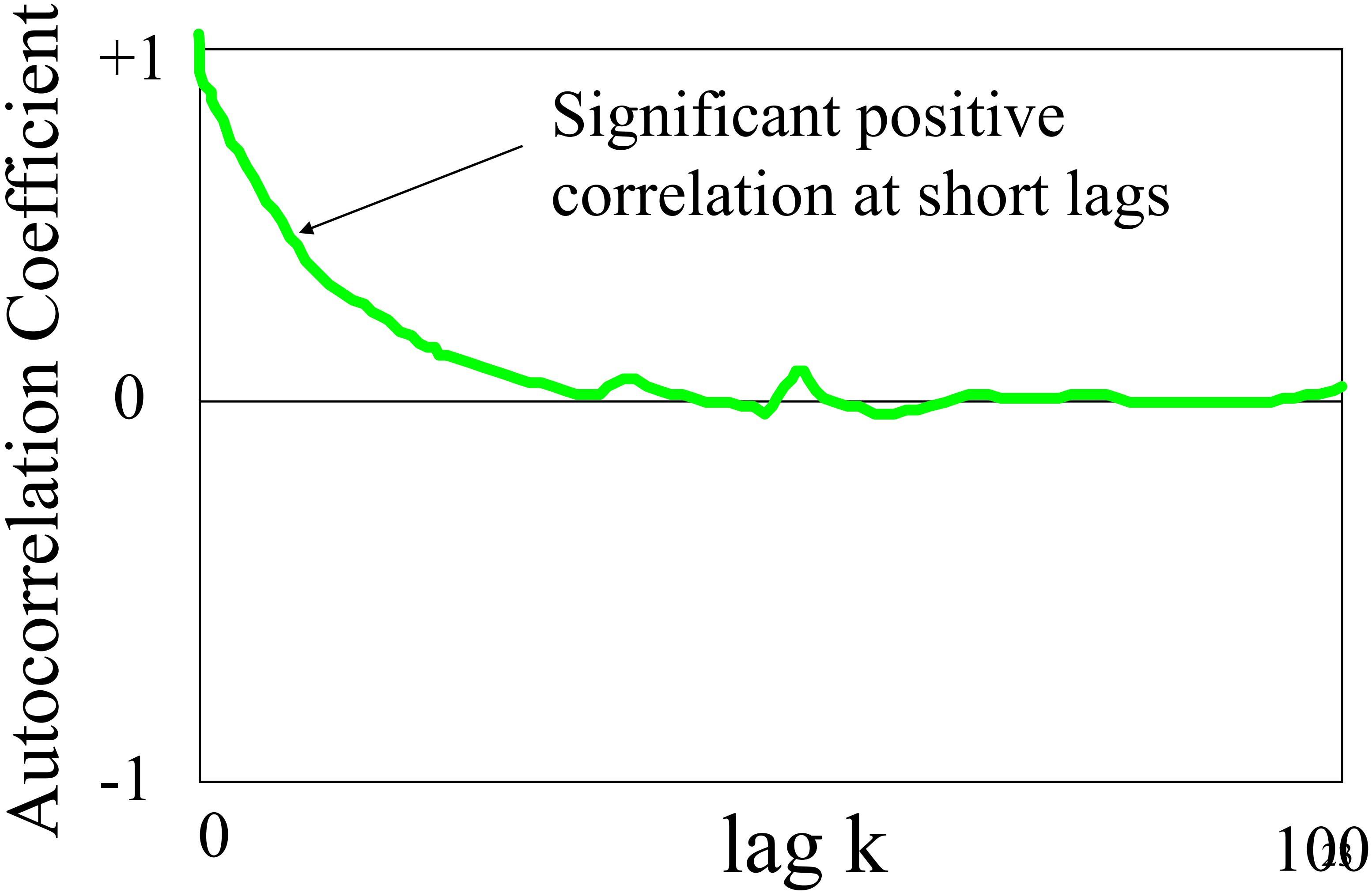
Autocorrelation Function



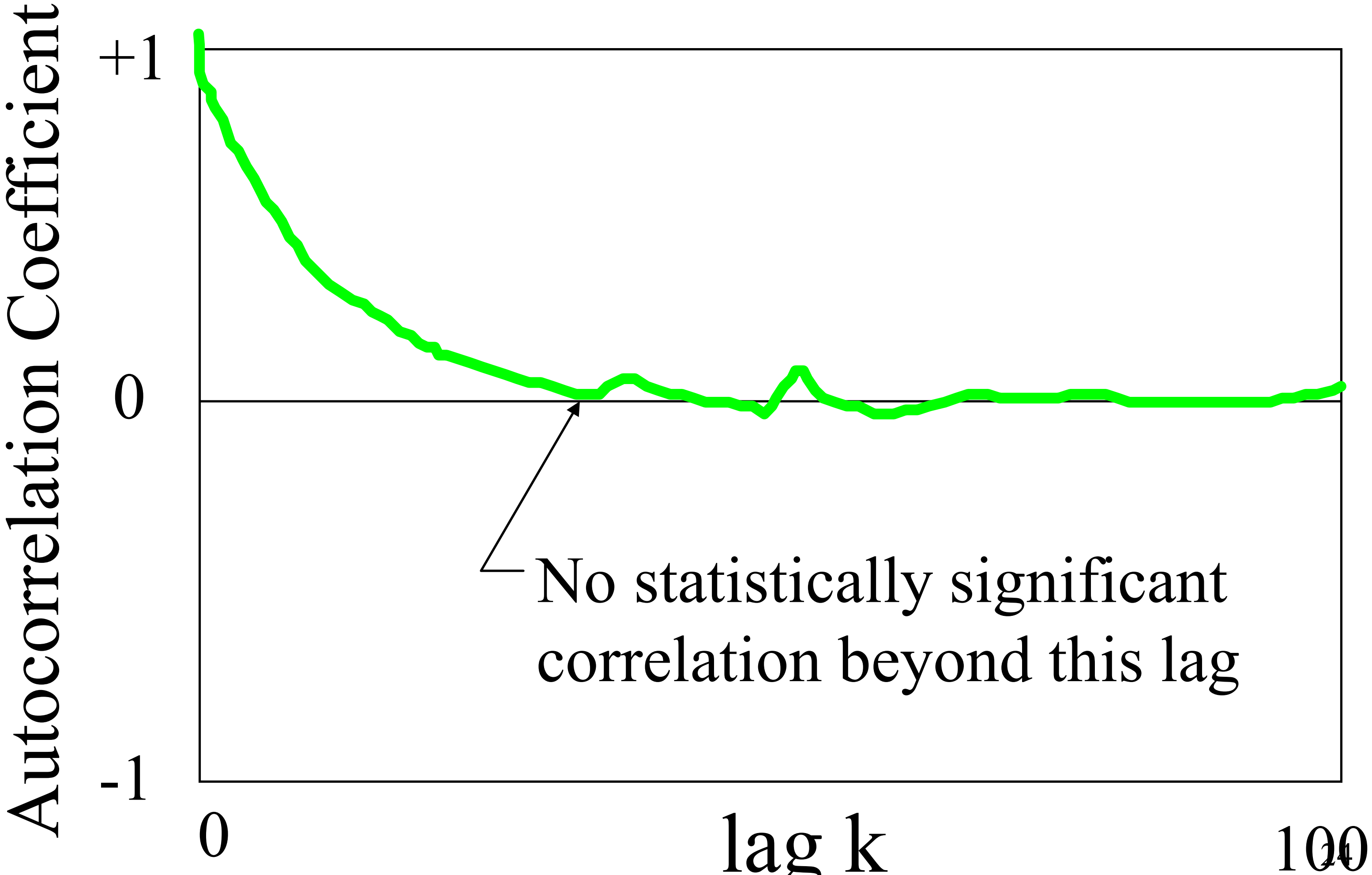
Autocorrelation Function



Autocorrelation Function



Autocorrelation Function



Correlation

- For most processes (e.g., Poisson, or compound Poisson), the autocorrelation function drops to zero very quickly
  - Usually immediately, or exponentially fast
  - Short range time dependence

Example: Poisson processes

- Find the autocorrelation function for the Poisson arrival process

Let $D_n = A_n - E[A_n]$ So $E[D_k] = 0.$

Define the autocorrelation as (k=lag)

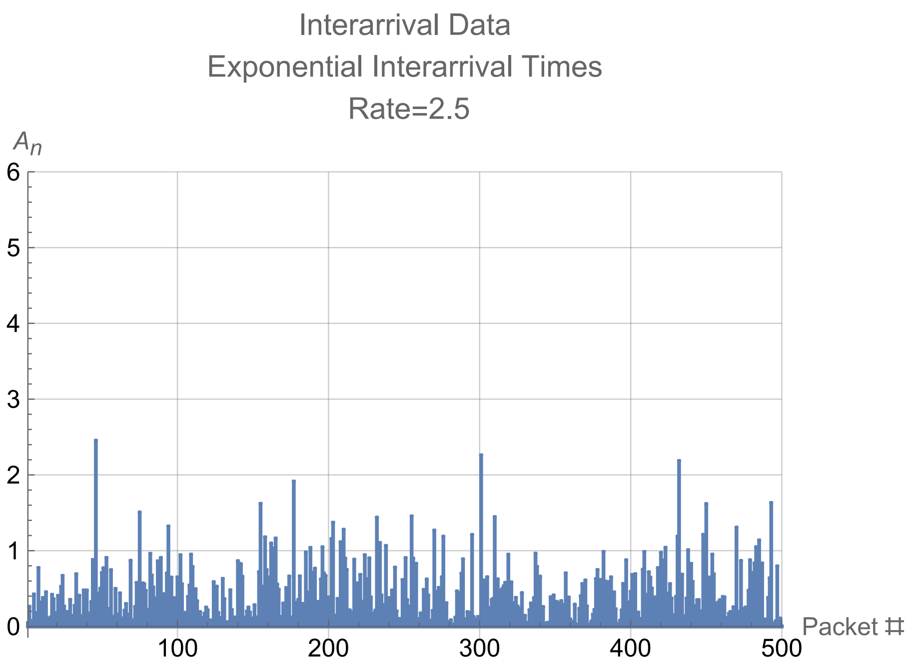
$$\rho_A(k) = \frac{E[D_n D_{k+n}]}{\sigma_A^2}$$

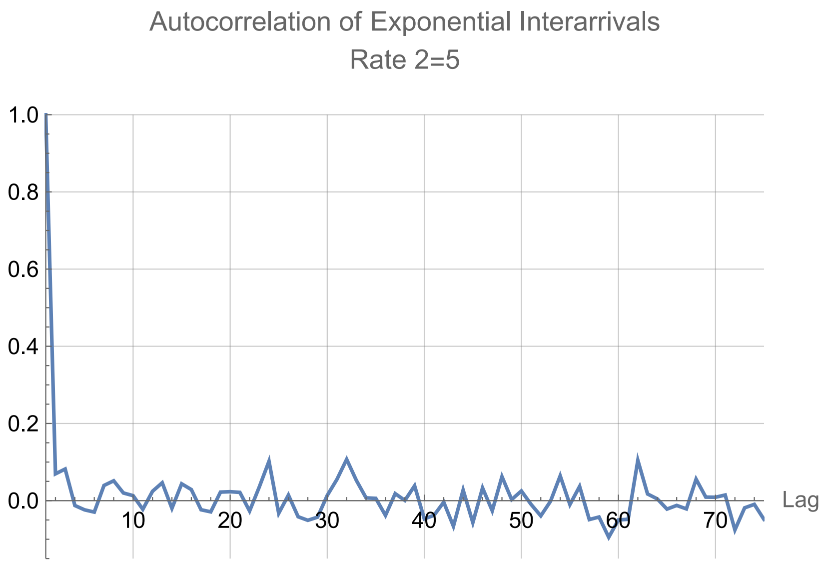
The A_n's are i.i.d resulting in

$$\rho_A(k) = \delta(k) \text{ where } \sigma_A^2 = \frac{1}{\lambda^2}$$

- The autocorrelation function drops to 0 after lag 1

Example: Poisson processes





Short range time dependence

Example: Markov Modulated Poisson Process MMPP

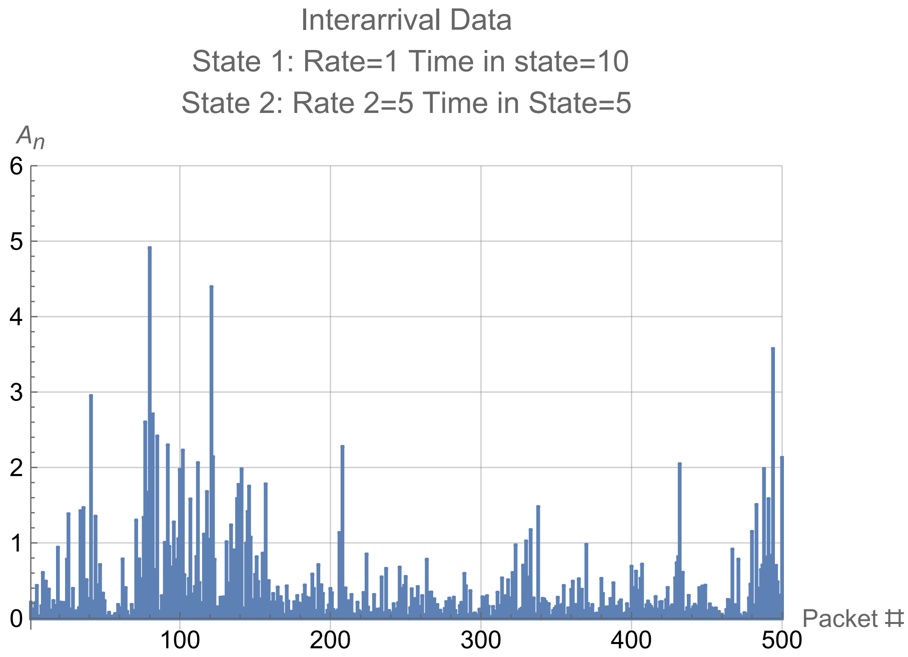
- A source generates packets at a set of discrete rates, $\lambda_1, \lambda_{2\ldots}\ \lambda_M$
- The source can be in state 1…M
- While in state j the source generates packets according to a Poisson Process at rate λ_j
- The time in state j follows an exponential pdf and are i.i.d., i.e., the rate process is Poisson.
- Thus, the state process "modulates" the rate of the source
- The autocorrelation function for this source is not a $\delta(t)$
- See Section 8.1 in Queueing Modeling Fundamentals: With Applications in Communication Networks", 2nd Edition Chee-Hock Ng and Soong Boon-Hee

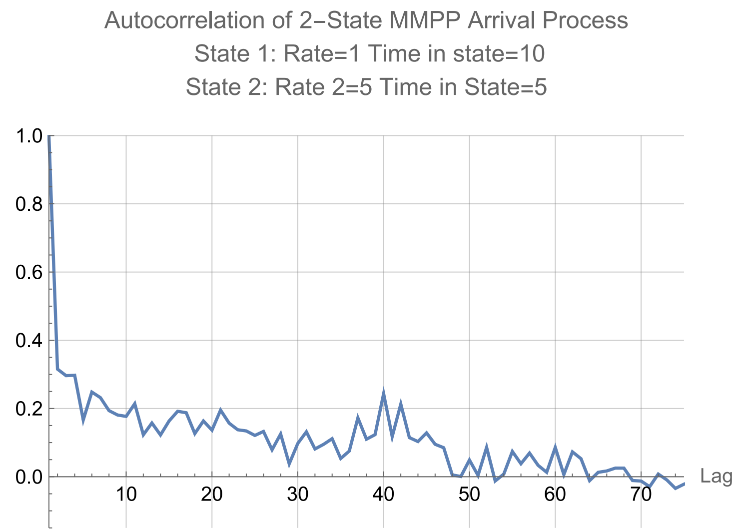
Example: Markov Modulated Poisson Process MMPP

- A source generates packets at two discrete rates,

  $\lambda_1 = 1$ with an average holding time of 10

  $\lambda_2 = 5$ with an average holding time of 5





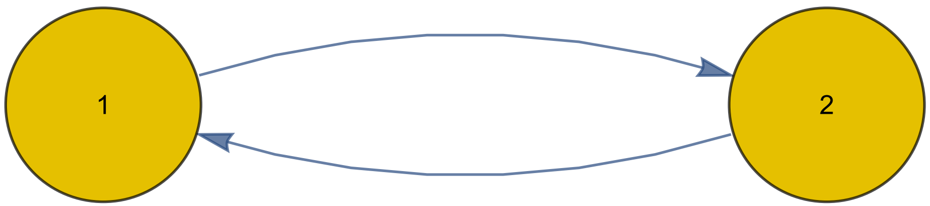
Short range time dependence

Example: Markov Modulated Poisson Process MMPP

- A source generates packets at two discrete rates,

  $\lambda_1 = 1$ with an average time in state =1

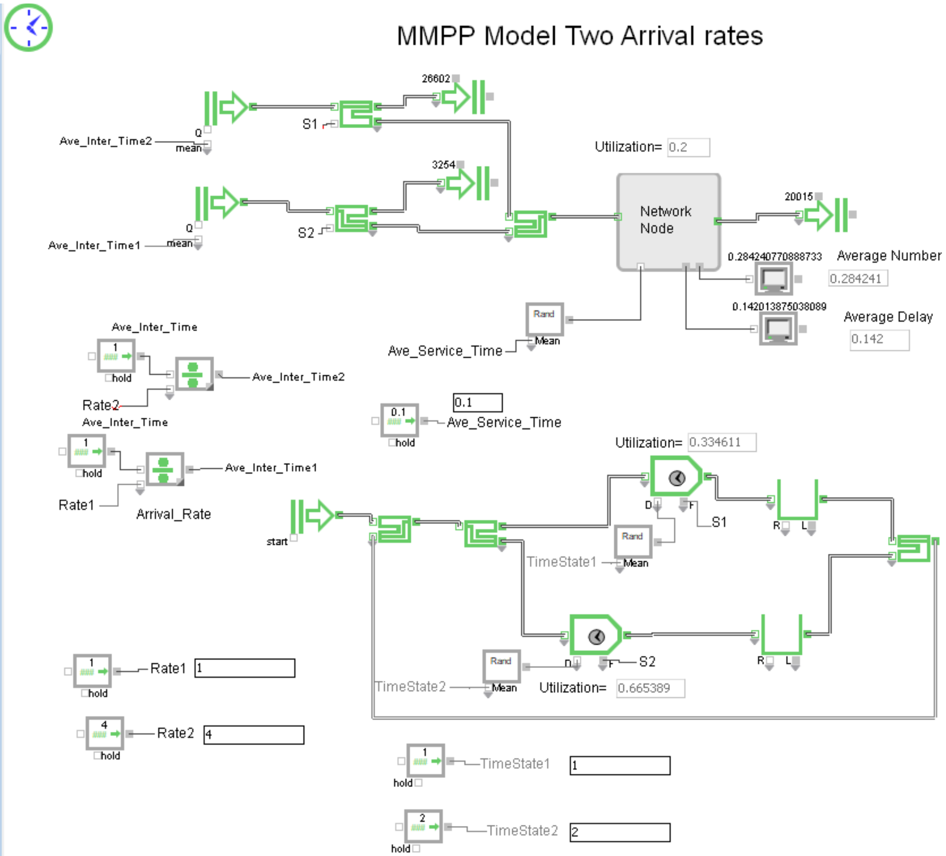
  $\lambda_2 = 4$ with an average time is state =2

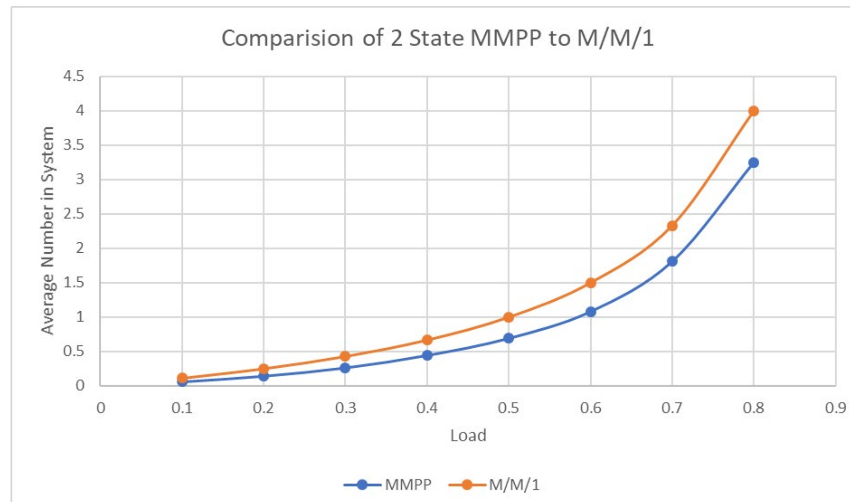


$\lambda_{12}=1$ and $\lambda_{21}=0.5$ then $\pi_1=2/3$ and $\pi_2=1/3$

$\lambda = 1(2/3)+4(1/3)=2$

Example: Markov Modulated Poisson Process MMPP





Example: Markov Modulated Poisson Process MMPP

- A source generates packets at three discrete rates,

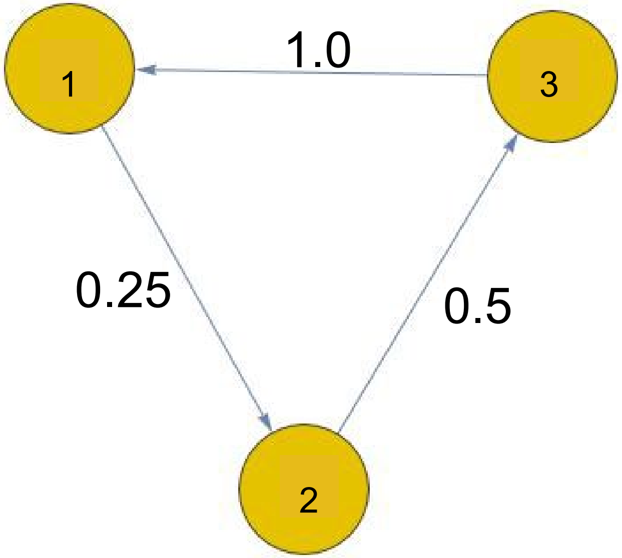
  $\lambda_1 = 1$ with an average time in state = 4

  $\lambda_2 = 2$ with an average time is state = 2

  $\lambda_3 = 3$ with an average time is state = 1

$\pi_1=0.57$ and $\pi_2=0.28$ and $\pi_3=0.14$

$\lambda = 1(0.57)+2(0.28)+3(0.14)=1.57$



Example: Interrupted Poisson Process IPP

- A source generates packets at two discrete rates,

  $\lambda_1 = \lambda$ with an Exponentially distributed holding time with mean $1/\alpha$ (for voice $1/\alpha = 352$ms)

  $\lambda_2 = 0$ with an Exponentially distributed holding time with mean $1/\beta$ (for voice $1/\beta = 650$ms)

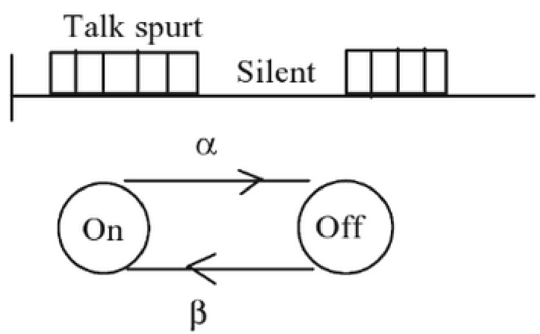


Figure 8.4 Interrupted Poisson Process model for a voice source

From: Queueing Modeling Fundamentals: With Applications in Communication Networks", 2nd Edition Chee-Hock Ng and Soong Boon-Hee

Example: Autoregressive Traffic Models

- The next interarrival time (or message length) model uses an explicit function of previous numbers in the sequence.
- The next number in the sequence X_n determined by

$$X_n = a_0 + \sum_{r=1}^{P} a_r X_{n-r} + \varepsilon_r \; n > 0$$

- With a_r's=real constants and ε_r's zero mean IID random variables.
- Used in modeling sequence of video frame sizes.

Discrete time arrival process Bernoulli Arrival Process

- Time is slotted
- Probability of one arrival in a time slot = p
- Probability of more than on arrival in a time slot = 0
- Probability of no arrival in a time slot = 1- p
- Arrivals in time slot i and time slot j are i.i.d.
- Time between arrivals is now a discrete R.V.
- $P(A_n = j) = p(1-p)^k$, a geometric pdf
- The sequence of A_n's are statistically independent

Non-Traditional Traffic Models

- A classic measurement study has shown that aggregate Ethernet LAN traffic is self-similar [Leland et al 1993]
- A statistical property that is very different from the traditional Poisson-based models
- Here
  - Definition of network traffic self-similarity,
  - Bellcore Ethernet LAN data,
  - Implications of self-similarity
  - Long Range Dependence
  - Heavy Tailed pdfs

Bellcore Measure Methodology

- Collected lengthy traces of Ethernet LAN traffic on Ethernet LAN(s) at Bellcore
- High resolution time stamps
- Analyzed statistical properties of the resulting time series data
- Each observation represents the number of packets (or bytes) observed per time interval (e.g., 10 4 8 12 7 2 0 5 17 9 8 8 2...)

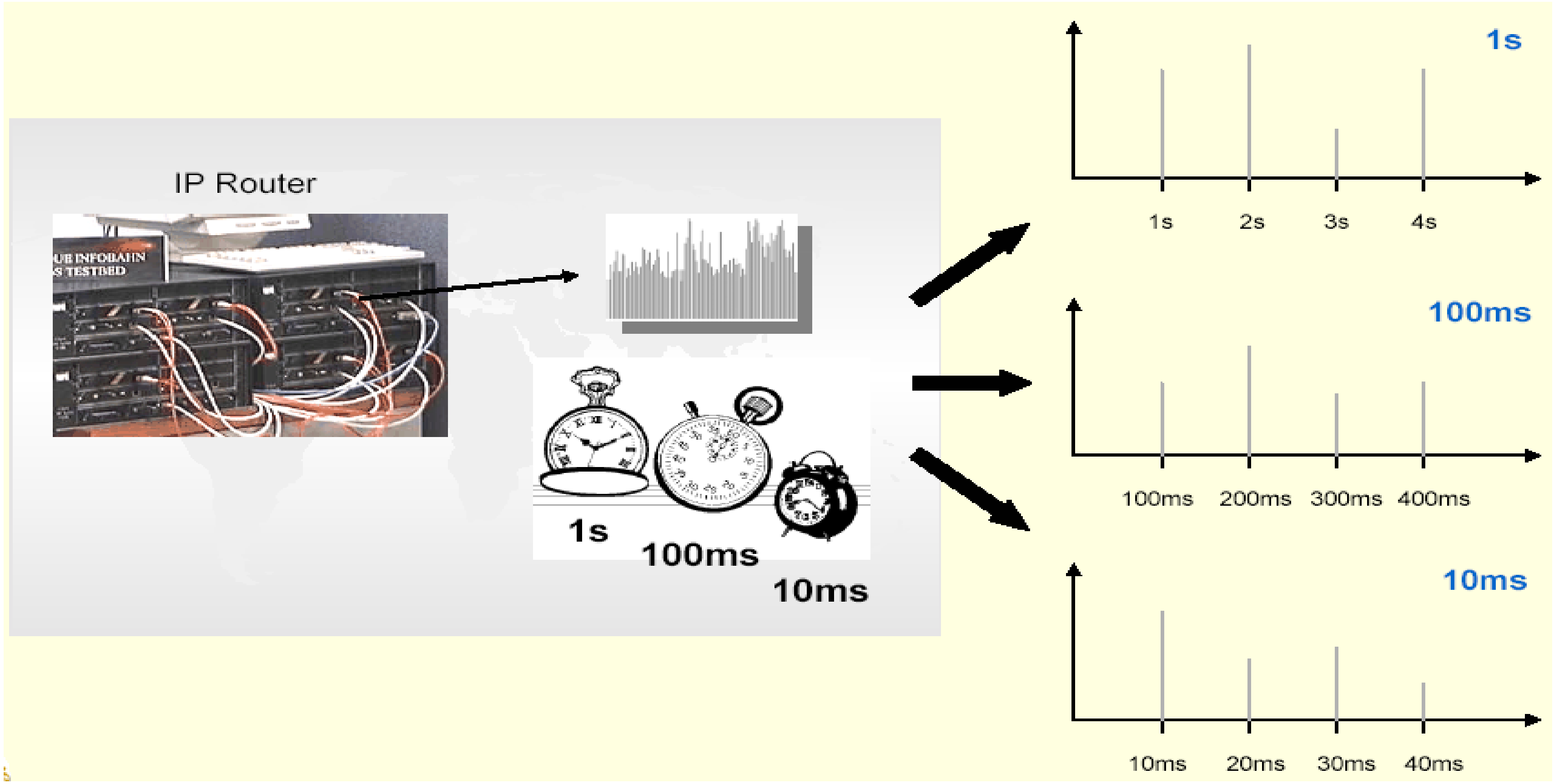
Self-Similarity: The intuition

- If you plot the number of packets observed per time interval as a function of time, then the plot looks "the same" regardless of what interval size you choose
- E.g., 10 msec, 100 msec, 1 sec, 10 sec,...
- Same applies if you plot number of bytes observed per interval of time

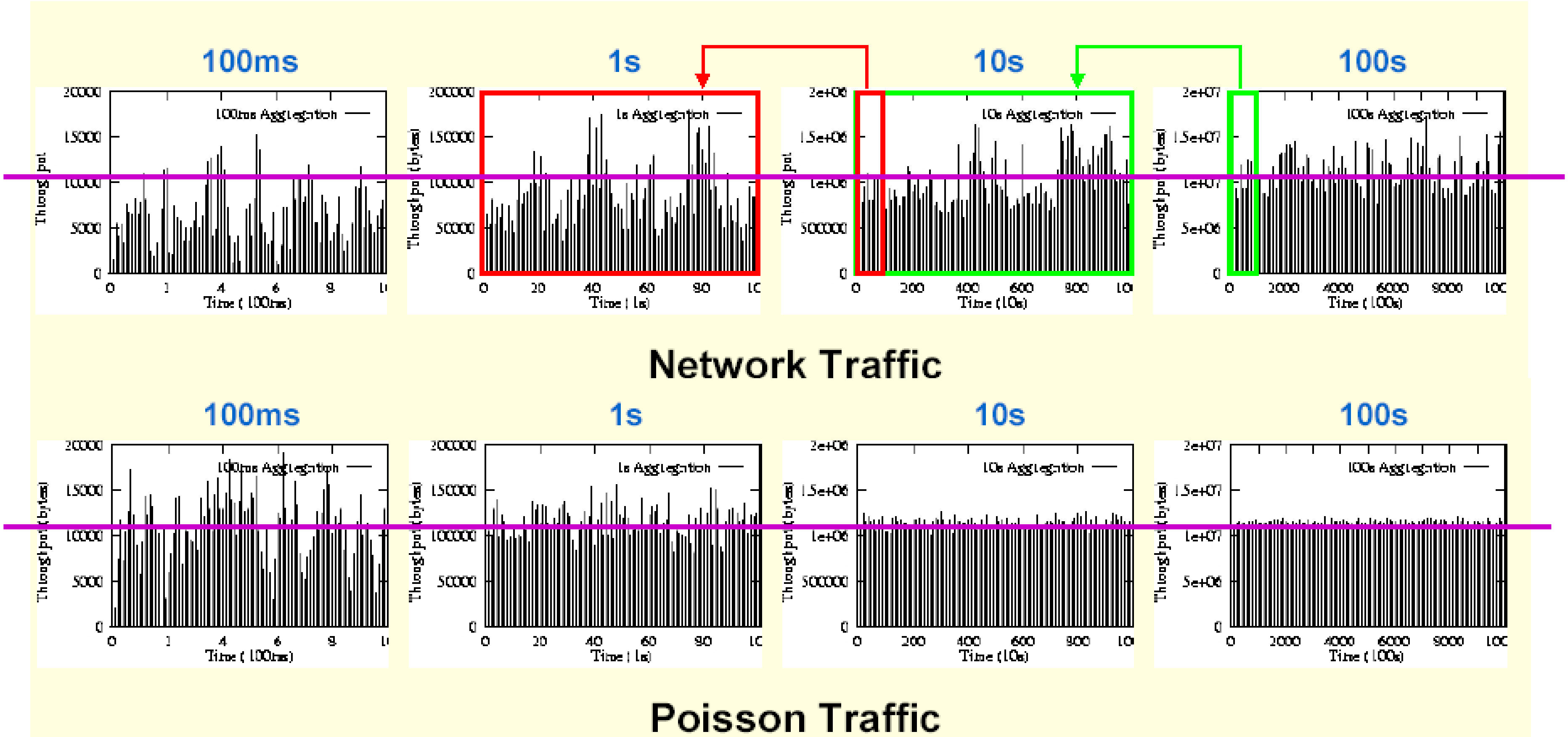
Self-Similarity: The Intuition

- In other words, self-similarity implies a “fractal-like” behavior: no matter what time scale you use to examine the data, you see similar patterns
- Implications:
  - Burstiness exists across many time scales
  - No natural length of a burst
  - Key: Traffic does not necessarily get “smoother” when you aggregate it (unlike Poisson traffic)
  - Self-Similarity can significantly impact queueing performance

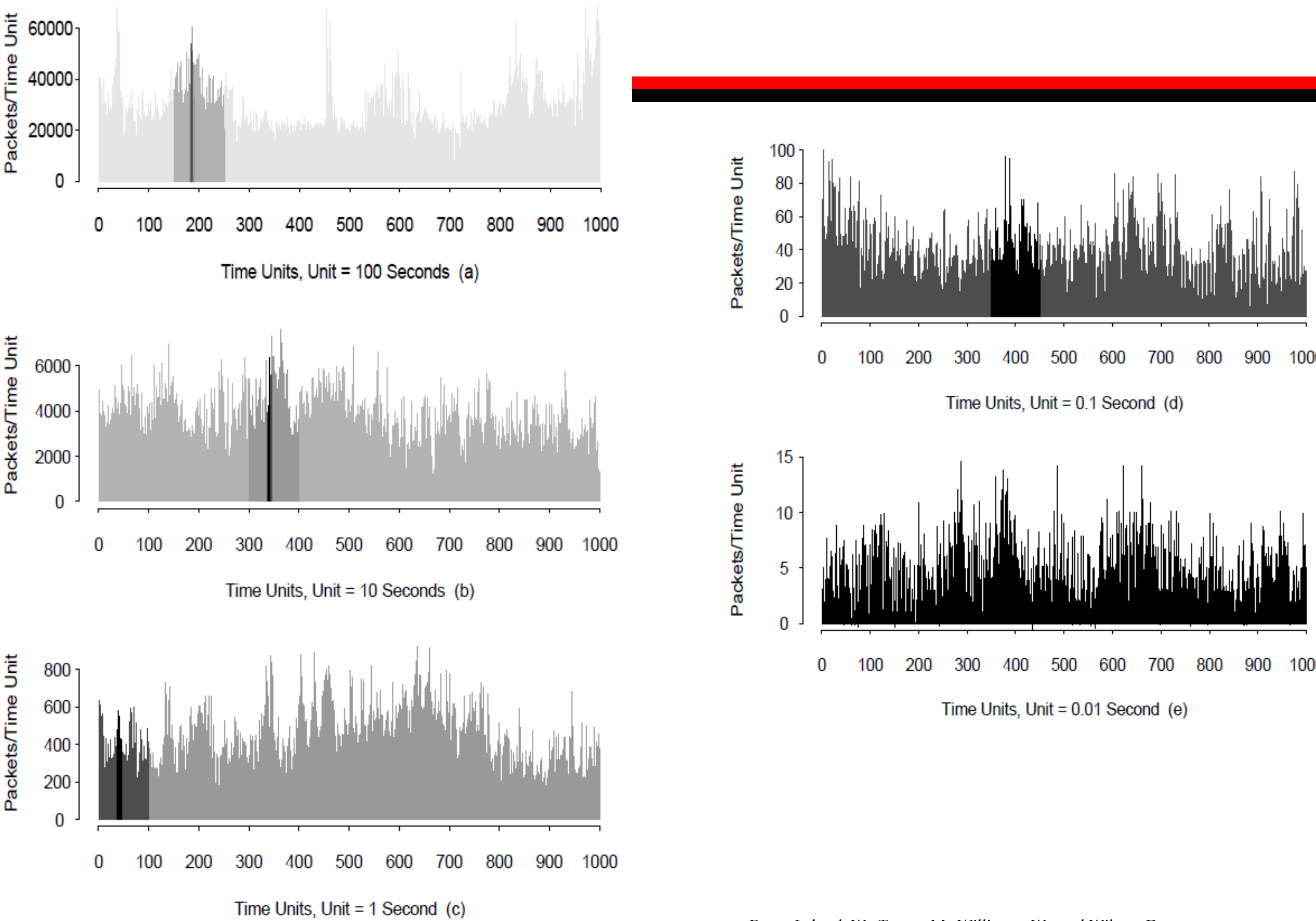
Self-Similarity Traffic Intuition (I)



Self-Similarity in Traffic Measurement II



On the Self-similarity of Network Traffic



From: Leland, W.; Taqqu, M.; Willinger, W.; and Wilson, D.
"On the Self-Similar Nature of Ethernet Traffic (Extended Version)."
IEEE/ACMTransactions on Networking, February 1994

Self-Similarity: The Math

- Self-similarity is a rigorous statistical property
  - (i.e., a lot more to it than just the pretty "fractal-like" pictures)
- Assumes you have time series data with finite mean and variance
  - i.e., covariance stationary stochastic process
- Must be a very long time series
  - infinite is best!
- Can test for presence of self-similarity

Self-Similarity: The Math

- Self-similarity manifests itself in several equivalent fashions:
- Slowly decaying variance
- Long range dependence
- Non-degenerate autocorrelations
- Hurst effect

Self-Similarity: The Math

- Self-similarity

A random process X(t) is self similar if

Y(t) = $a^{-H} X(at)$ $0.5 \le H \le 1$

and Y(t) has the same statistical properties as X(t).

[Remember x(at) is time scaling the time function x(t)]

H=Hurst Parameter

Where "same statistical property" is the autocorrelation function for the aggregated process (time scaled) is indistinguishable from that of the original process.

Long Range Dependence: The Math

- Long Range Dependence

Let C(k) = covariance of X(t) [Scaled autocorrelation function]

If X(t) is short range dependent if

$C(k) \propto a^{-|k|}$ $|a|<1$

That is, the correlation function decays exponentially fast.

X(t) is long range dependent if

$C(k) \propto |k|^{-\beta}$ with $H=1-\frac{\beta}{2}$

C(k) decays much slower.

Technically X(t) can be self-similar and not long range dependent

and X(t) can be long range dependent and not self-similar.

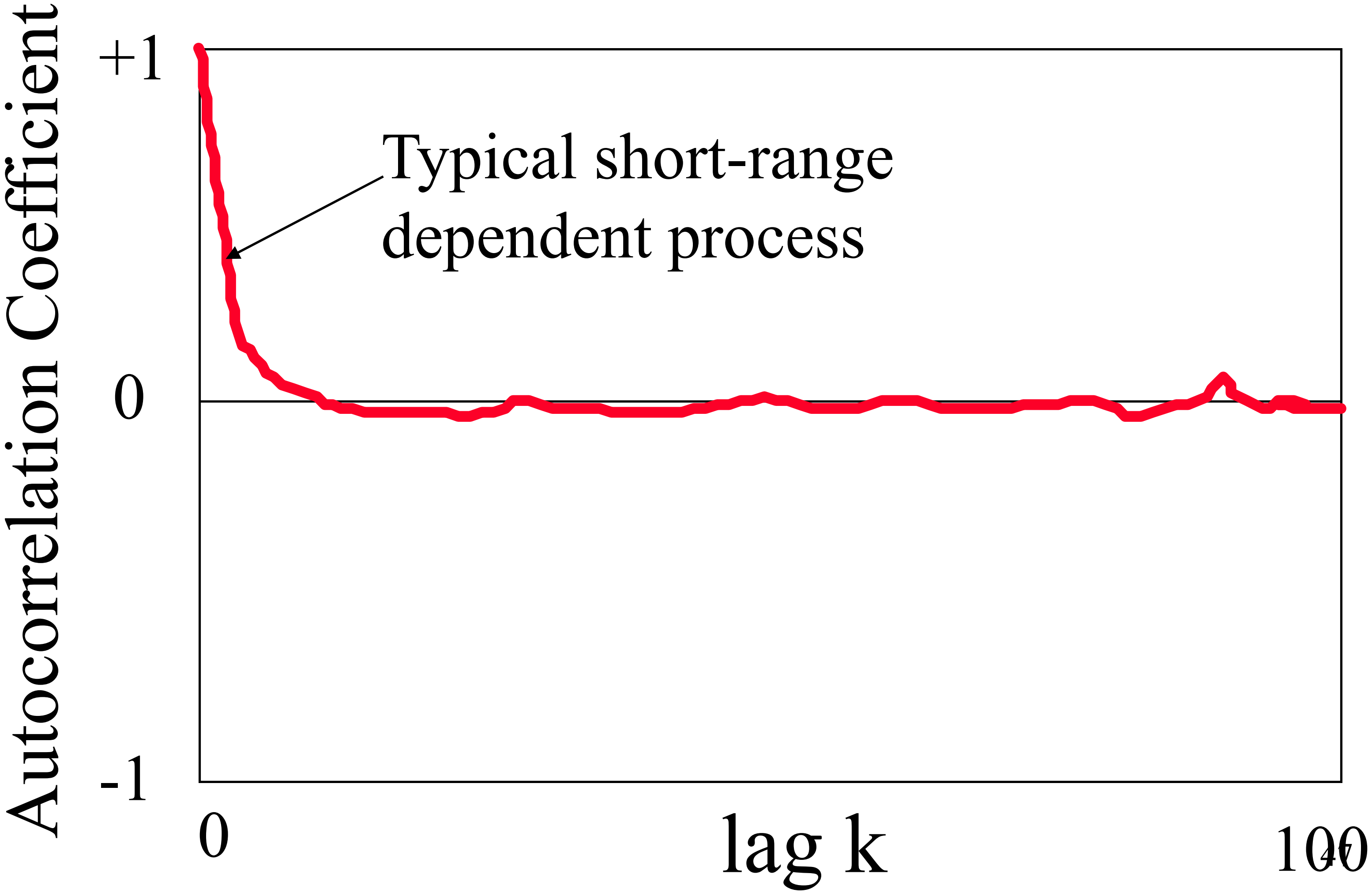
However, for network traffic

self-similar $\rightarrow$ long range dependent

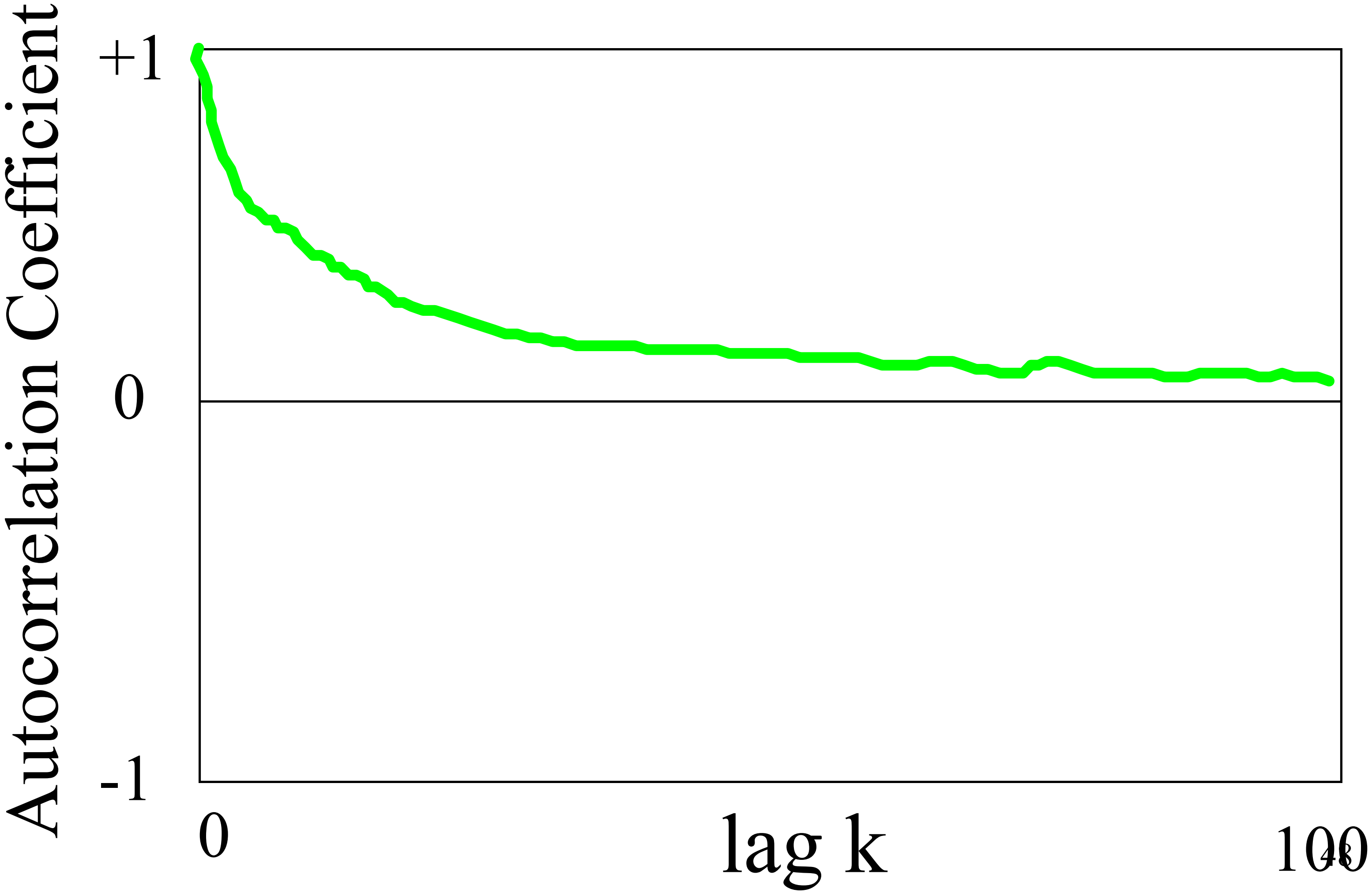
and

long range dependent $\rightarrow$ self-similar

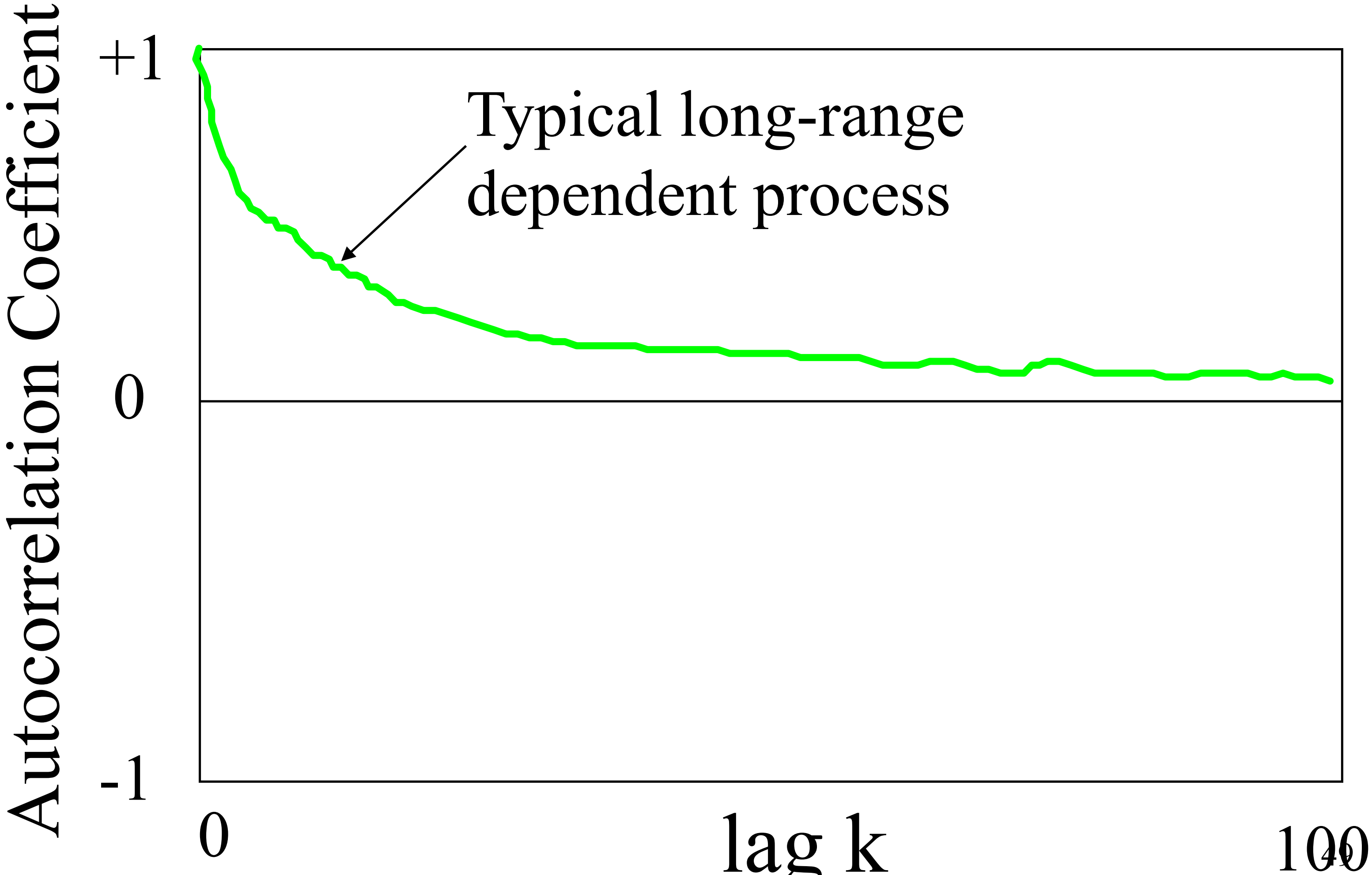
Autocorrelation Function



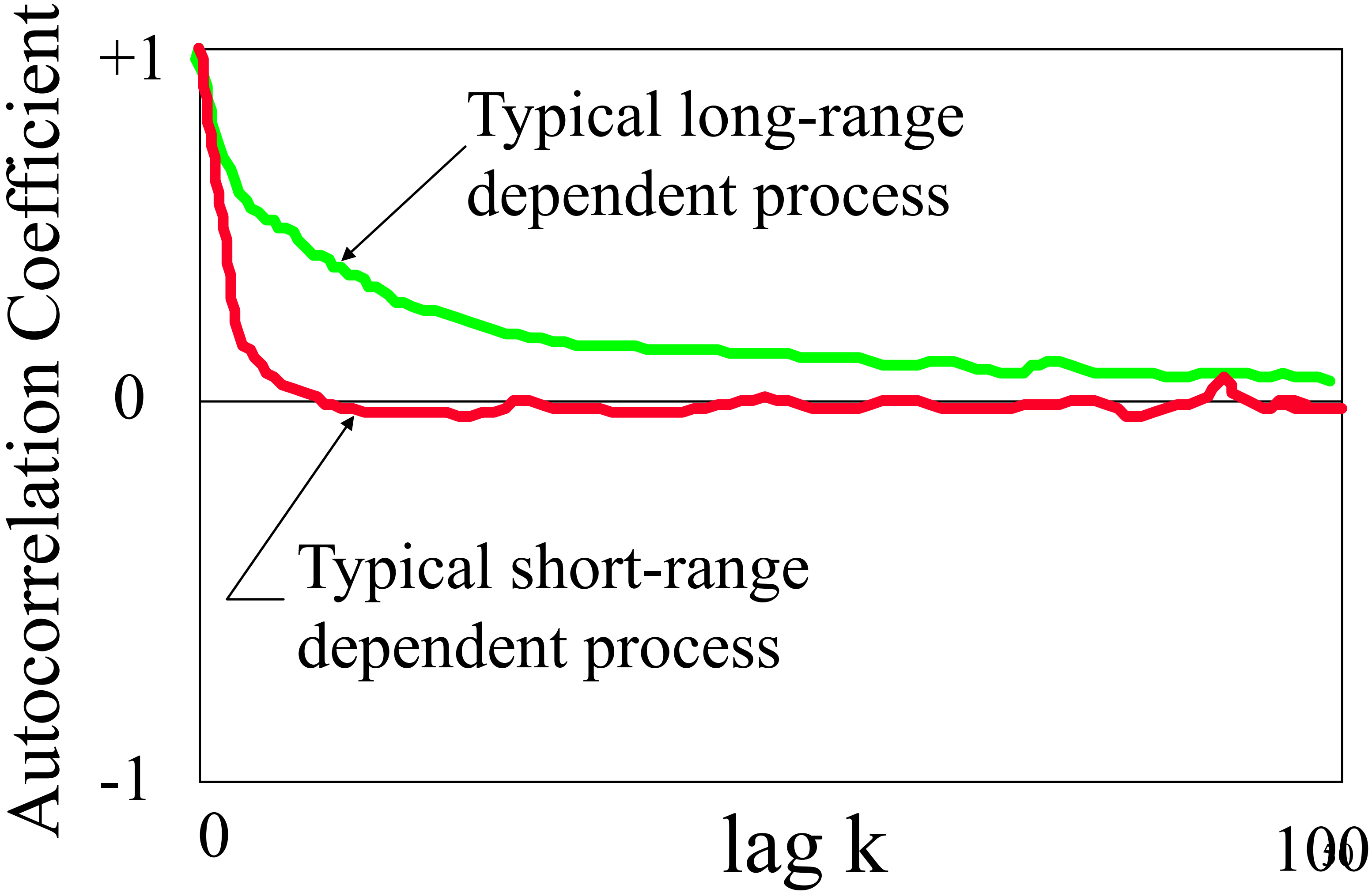
Autocorrelation Function



Autocorrelation Function



Autocorrelation Function



Heavy-Tailed Distributions: The Math

- Heavy-Tailed Distributions

  A random variable, X, is heavy-tailed if

  $$P(X>x) \propto \frac{1}{x^{\alpha}}$$

  Note $\mathrm{Var}[X] \to \infty$

- If X has a heavy-tailed pdf then very large values of X can be observed with nonnegligible probability.
- Heavy-Tailed interarrival times or message lengths can cause long range dependence and self-similarity

Heavy-Tailed Distributions: Example

- Pareto distribution

A random variable, X, has a **Pareto distribution** (Type I) if

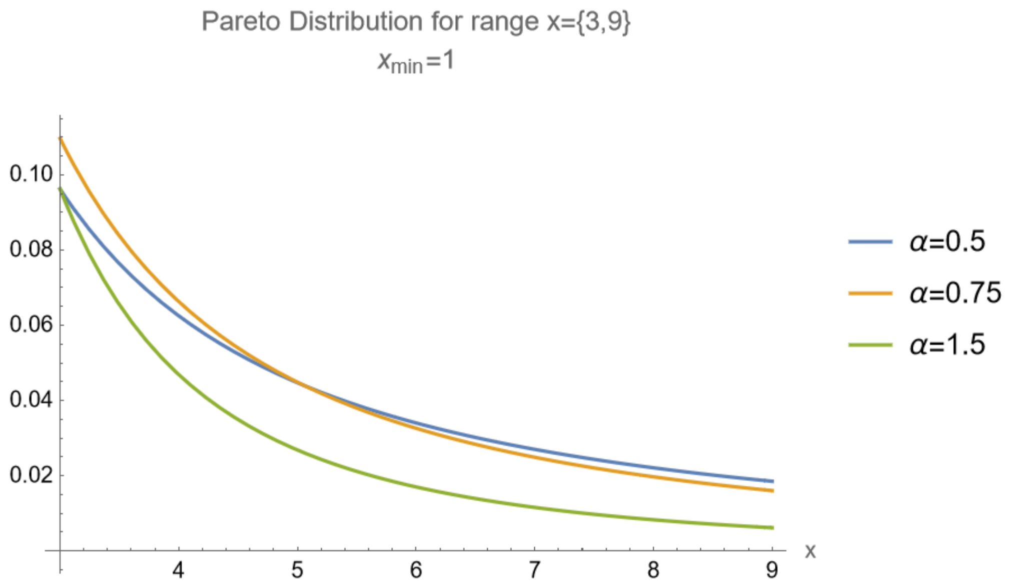
$$f_X(x;\alpha) = \frac{\alpha x_{\min}}{x^{\alpha+1}} \quad \text{for } x > x_{\min}$$
$$= 0 \quad \text{for } x < x_{\min}$$

$$E[X] = \infty \text{ for } \alpha \le 1$$

$$E[X] = \frac{\alpha x_{\min}}{\alpha - 1} \text{ for } \alpha > 1$$

$$\text{Var}[X] = \infty \text{ for } \alpha \le 2$$

$$\text{Var}[X] = \frac{\alpha x_{\min}^2}{(\alpha-1)^2(\alpha-1)} \text{ for } \alpha > 2$$



Heavy-Tailed Distributions: Example

- Pareto distribution
  - Impact on run time
  - Example:
    - M/M/1 with C=1Mb/s, E[L]=6352bits λ=78.7→ ρ=.5, T_{stop}=100 sec/run
      - A simulation:
        - » 10 runs Delay 13ms with 95% CI +/- .7ms (Theory 12.7ms)
  - Example: See ExtendSim Simulation

The story of mice and elephants

- Mice flows as small volume, short-lived flows.
- Elephant flows are just what you would expect, in contrast to mice flows: large, long-lived flows.
- 2010 research suggests that mice flows comprise more than 90% of all flows in a data center network, but carry less than 10% of the total number of bytes transmitted on the network. Mice flows are typically less than 10KB in size and therefore fit into just a few packets. Elephant flows are just the opposite, constituting only 10% of the flows, but carrying 90% of the transmitted bytes.

Measuring non traditional traffic

- R/S statistic
- Let $Y_1, Y_2, Y_3, Y_4 \ldots Y_n$ be n samples of a time series

$$\overline{Y}(n) = \frac{1}{n}\sum_{i=1}^{n} Y_i = \text{sample mean as a function of the sample size n}$$

$$S^2(n) = \frac{1}{n-1}\sum_{i=1}^{n} (Y_i - \overline{Y}(n))^2 = \text{sample variance as a function of the sample size n}$$

Define

$$\text{R(n)} = \text{Max}\{\sum_{i=1}^{k} (Y_i - \overline{\text{Y}}) \text{: } 1 \le \text{k} \le n\} \text{-Min}\{\sum_{i=1}^{k} (Y_i - \overline{\text{Y}}) \text{: } 1 \le \text{k} \le n\}$$

then

R(n)/S(n) = R/S statistic

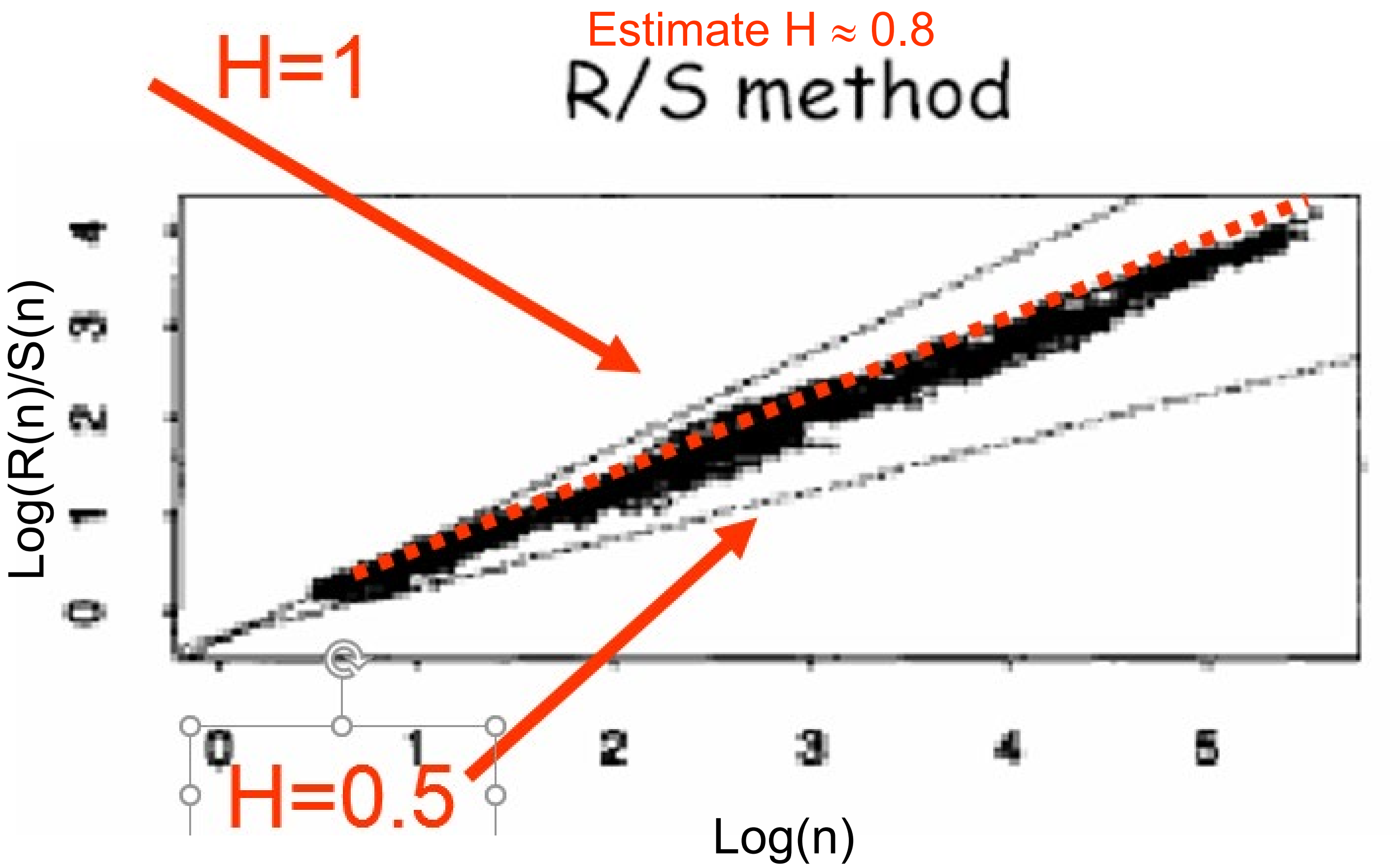
It has been empirically found that

$\text{E[R(n)/S(n)]} \cong a\text{n}^H$

A plot of Log[R(n)/S(n)] vs Log(n) is the linear as Log[R(n)/S(n)] = Log[a]+HLog(n)

The Hurst parameter is the slope of the line.

Visualizing Long Range Dependence



Hurst Effect

- For models with only short range dependence, H is almost always 0.5
- For self-similar processes, 0.5 < H < 1.0
- This discrepancy is called the Hurst Effect, and H is called the Hurst parameter
- Single parameter to characterize self-similar processes

R/S Plot

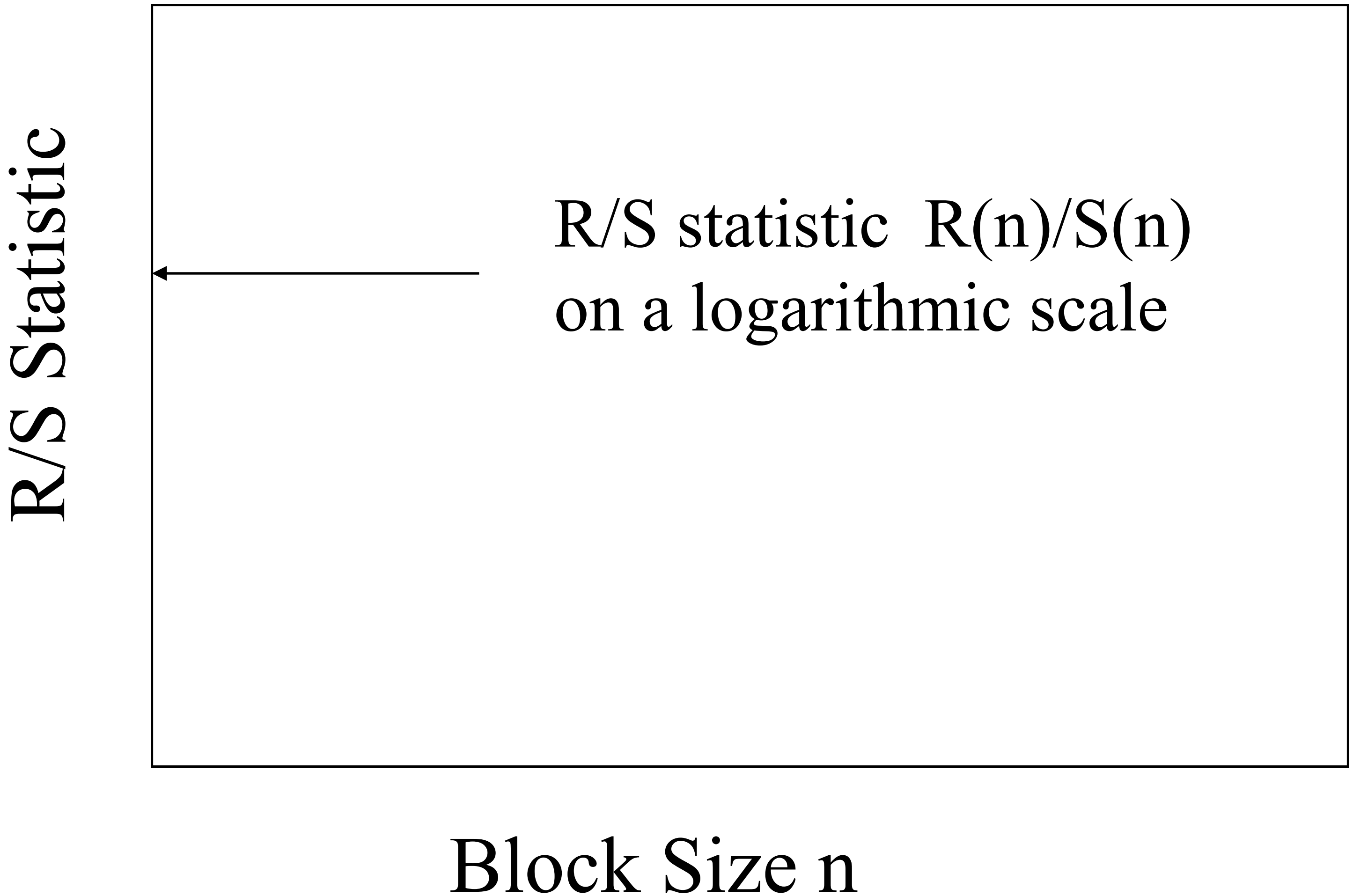
- A way of testing for self-similarity, and estimating the Hurst parameter
- Plot the R/S statistic for different values of n, with a log scale on each axis
- If time series is self-similar, the resulting plot will have a straight line shape with a slope H that is greater than 0.5
- Called an R/S plot, or R/S pox diagram

R/S Pox Diagram

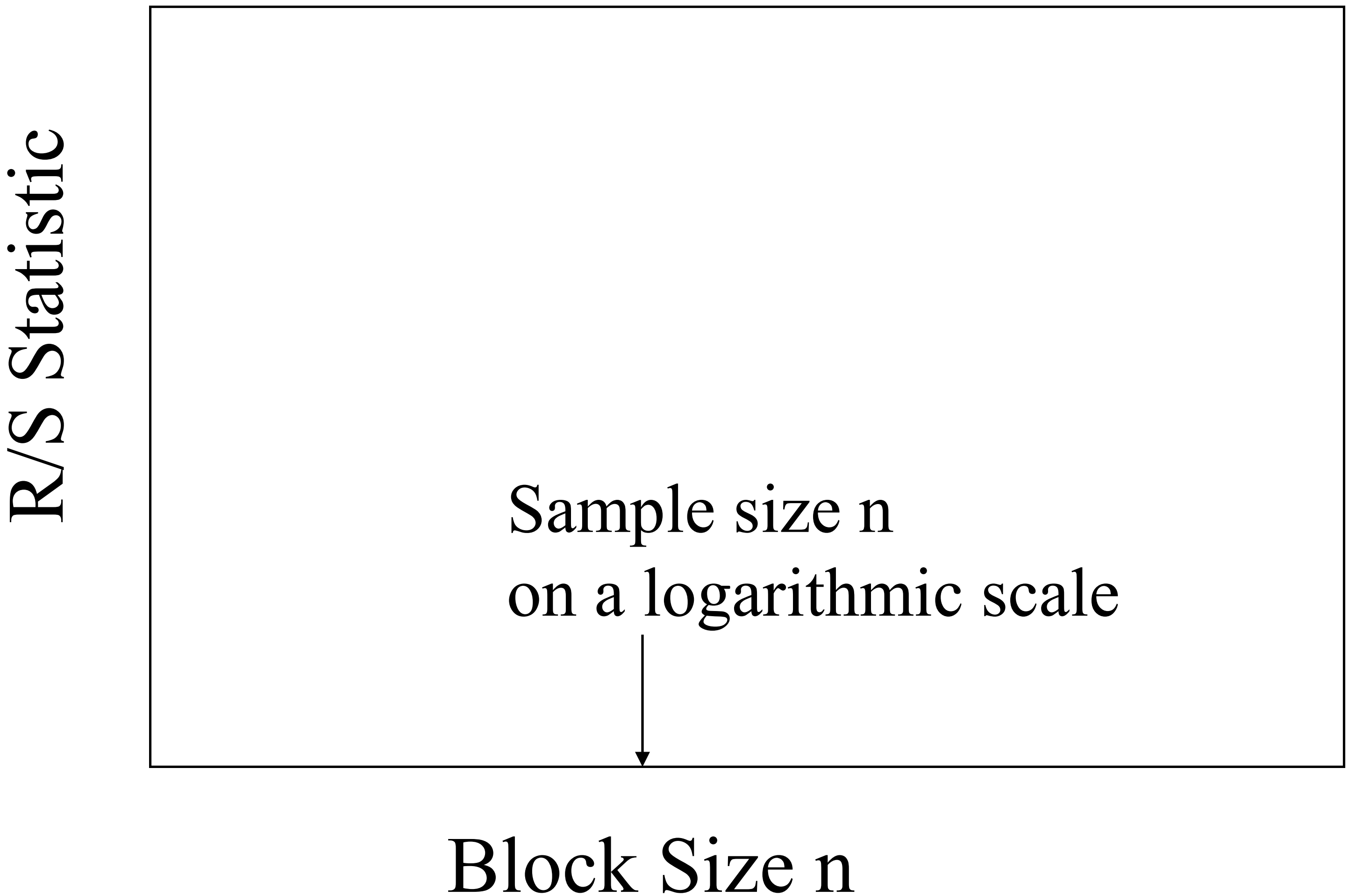
R/S Statistic

Block Size n

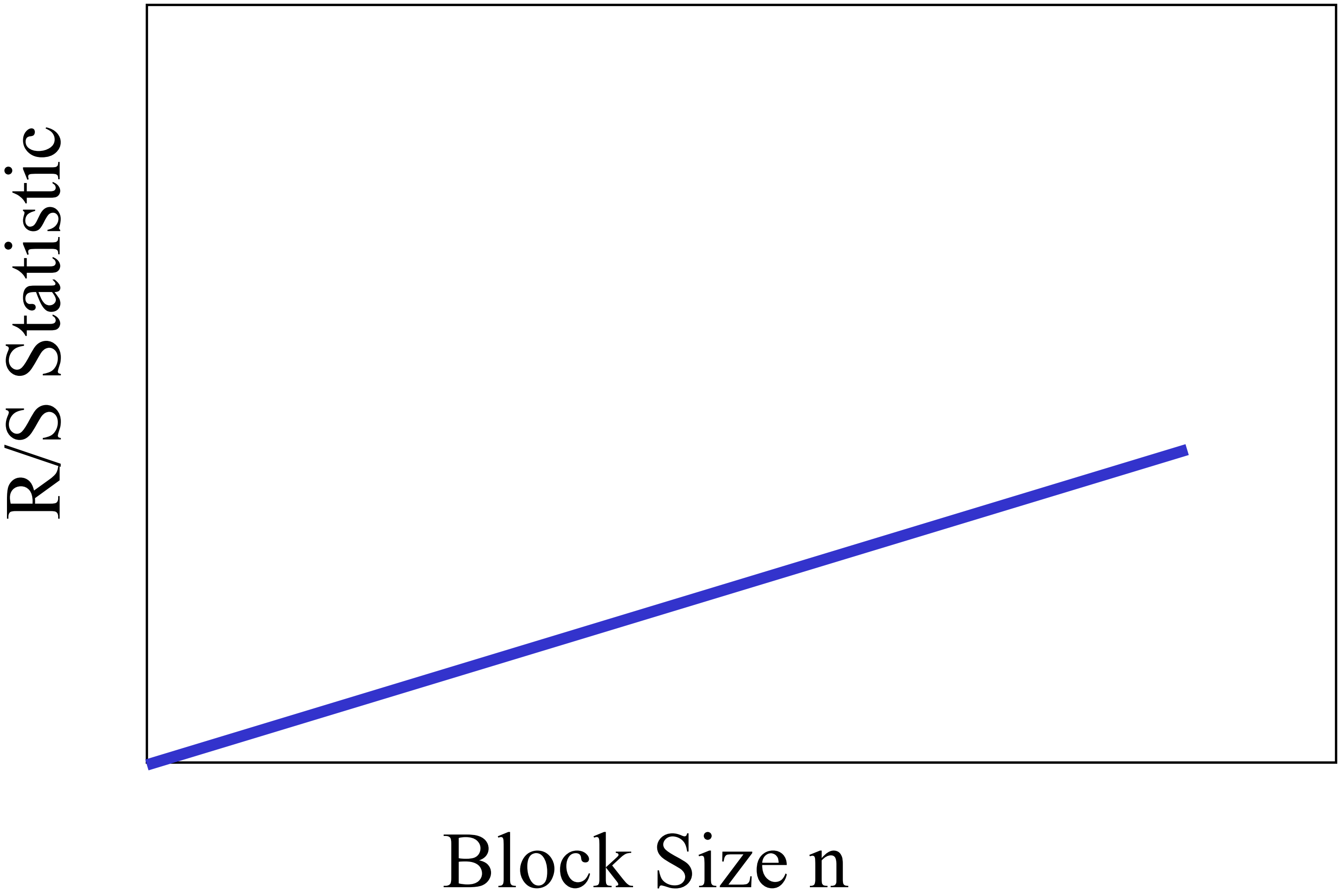
R/S Pox Diagram



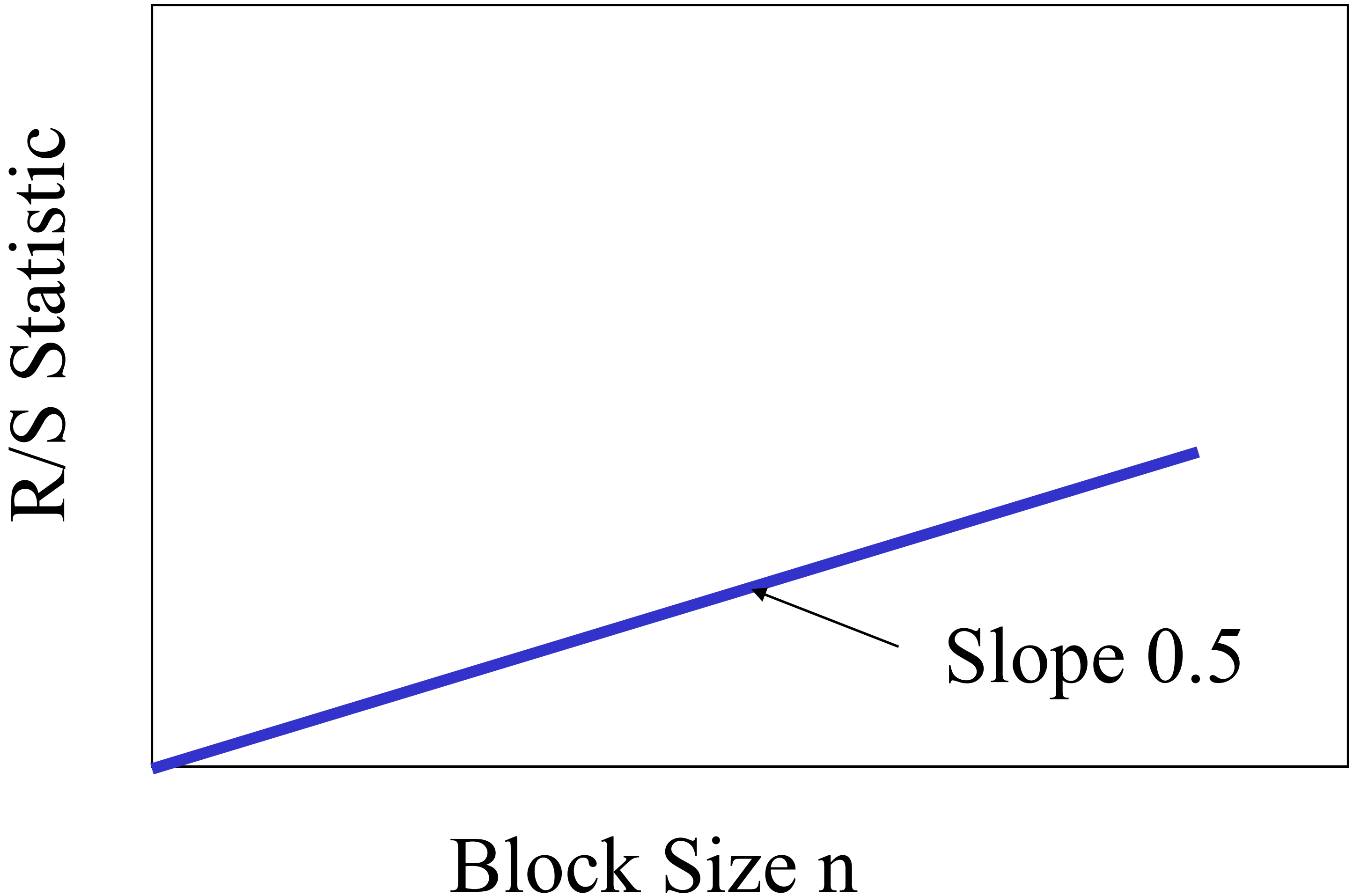
R/S Pox Diagram



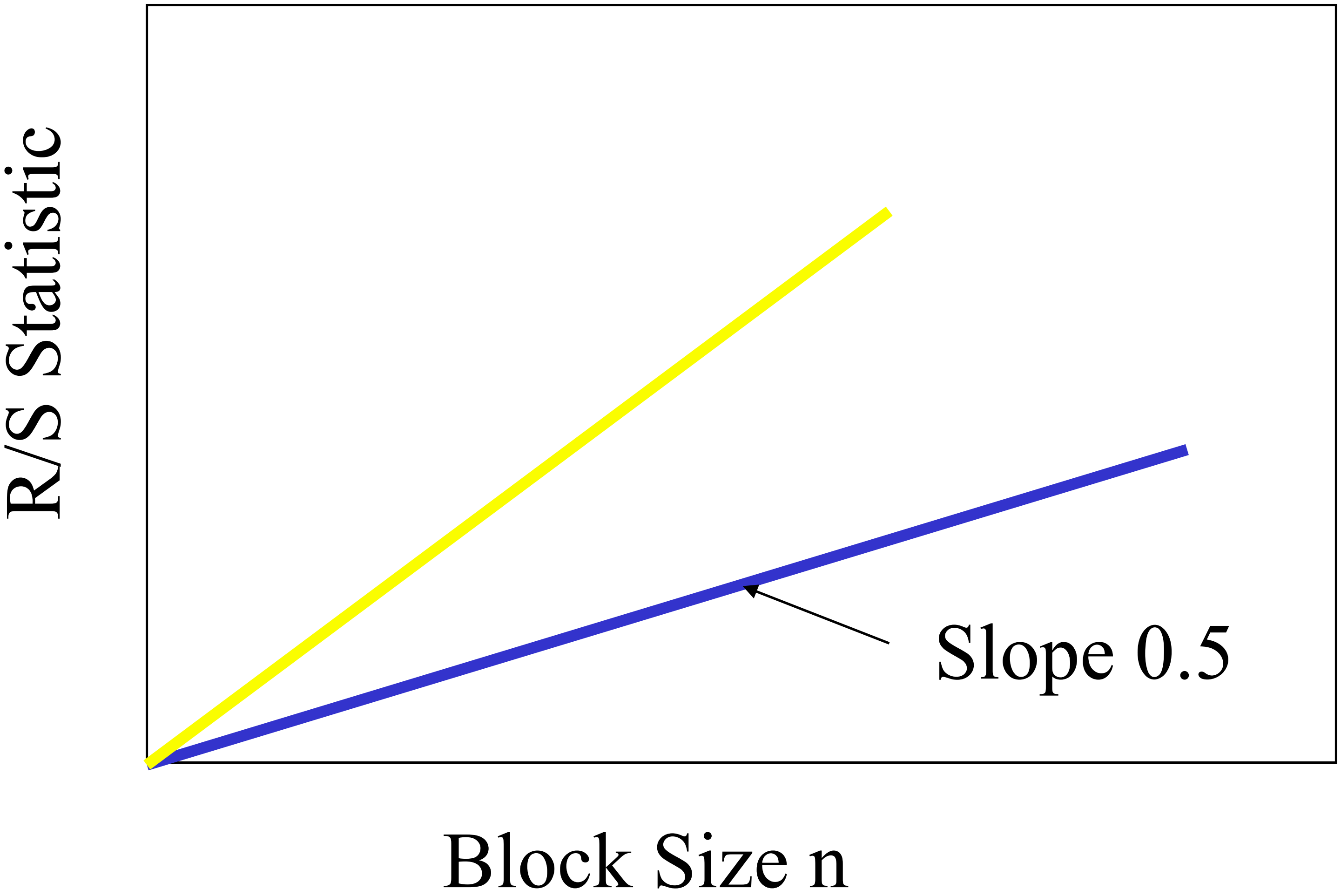
R/S Pox Diagram



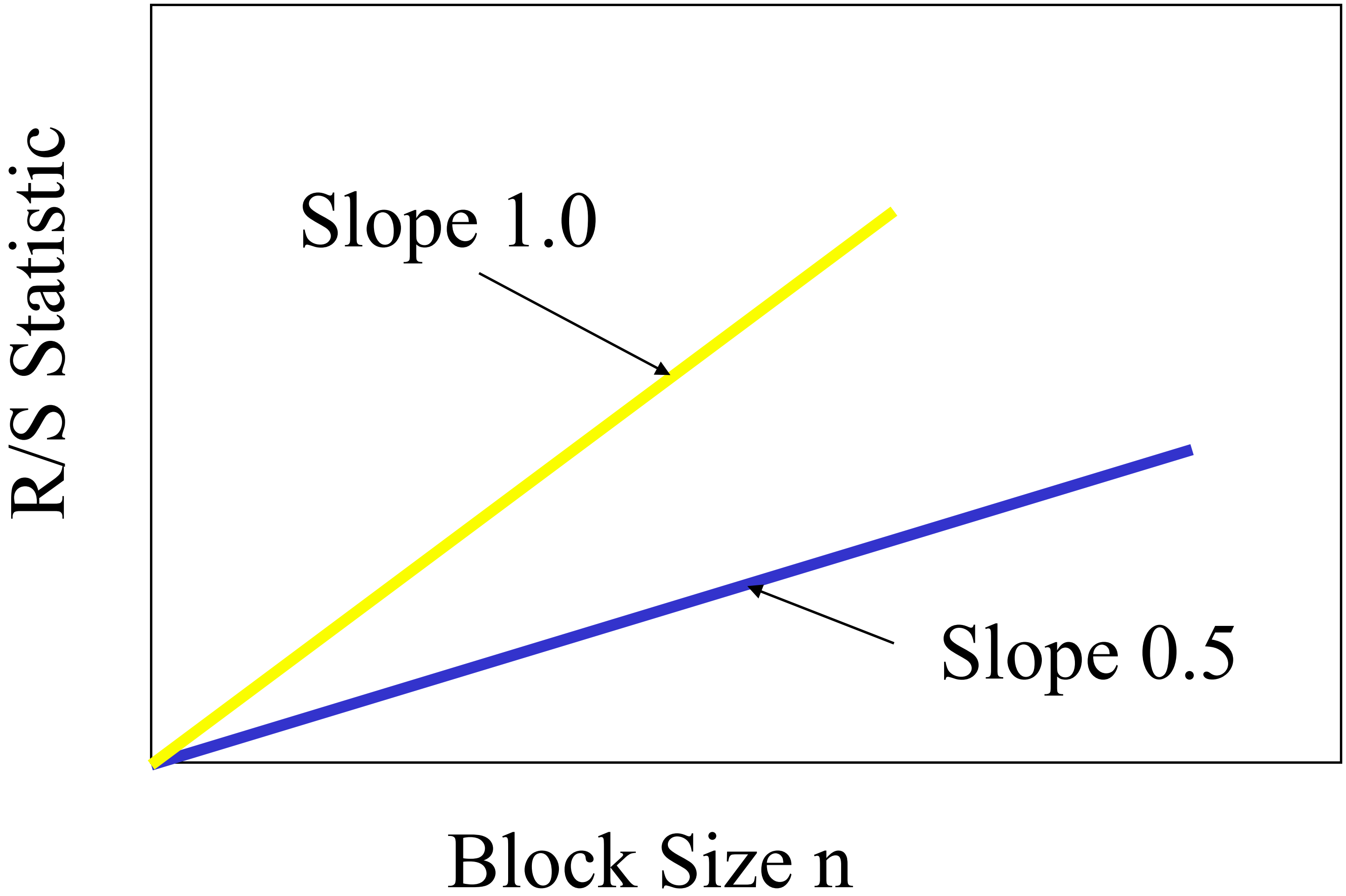
R/S Pox Diagram



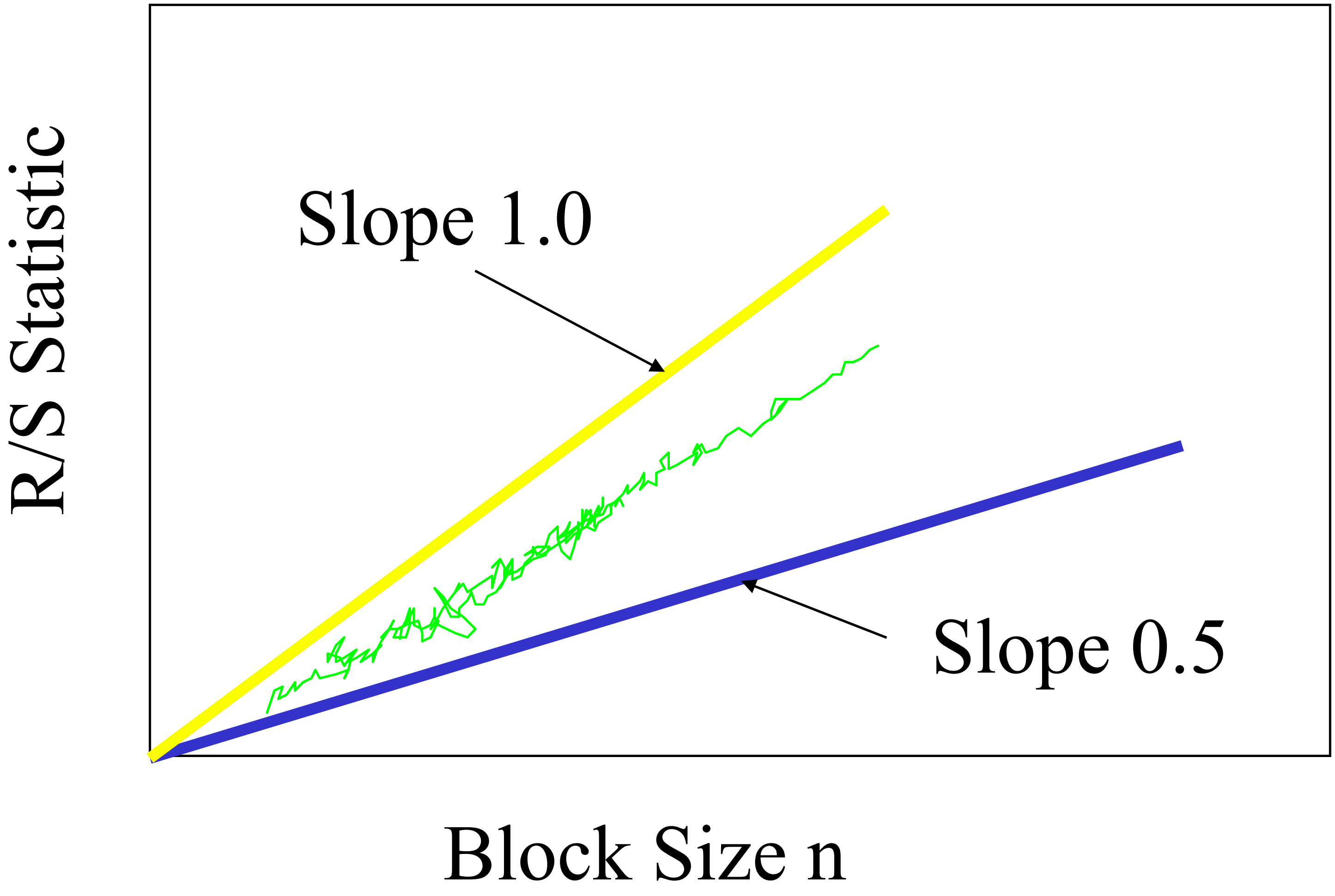
R/S Pox Diagram



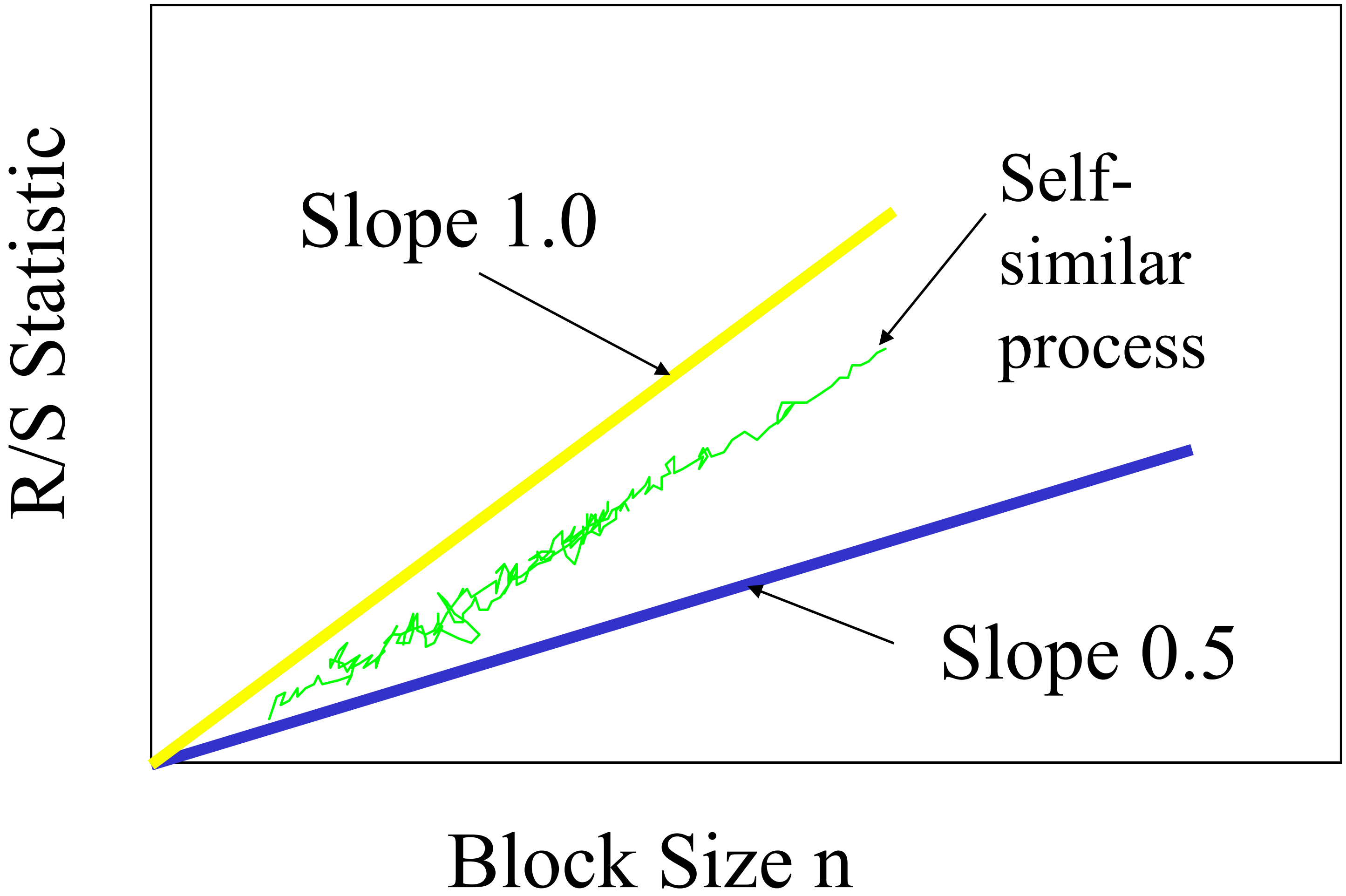
R/S Pox Diagram



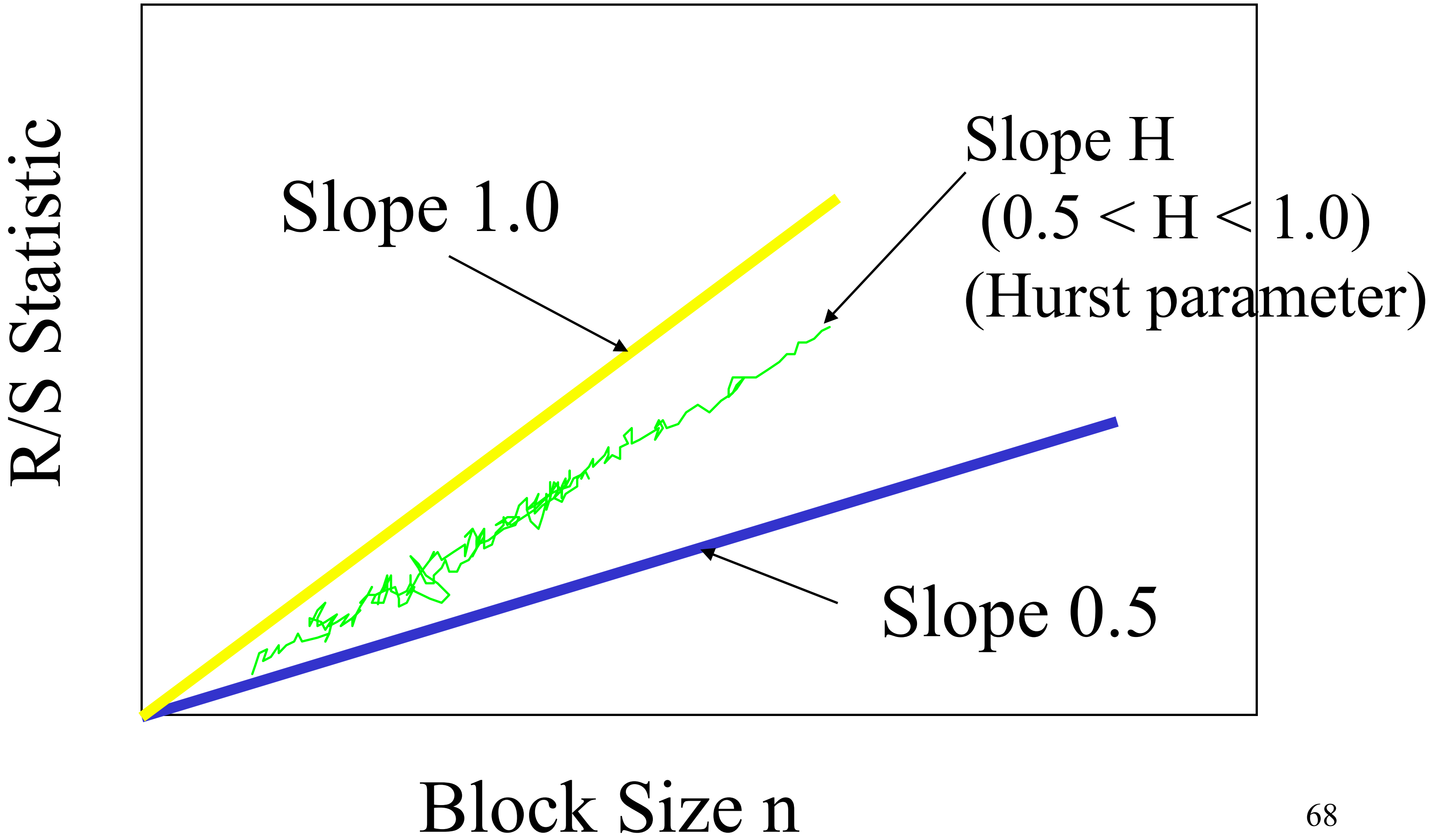
R/S Pox Diagram



R/S Pox Diagram



R/S Pox Diagram



Slowly Decaying Variance

$$\overline{X} = \frac{1}{N}\sum_{k-1}^{N} X_k$$

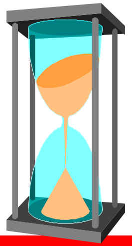
$$Var[\overline{X}] = \frac{\sigma_X^2}{N}$$

- For i.i.d samples the variance of the sample decreases as 1/N

Slowly Decaying Variance

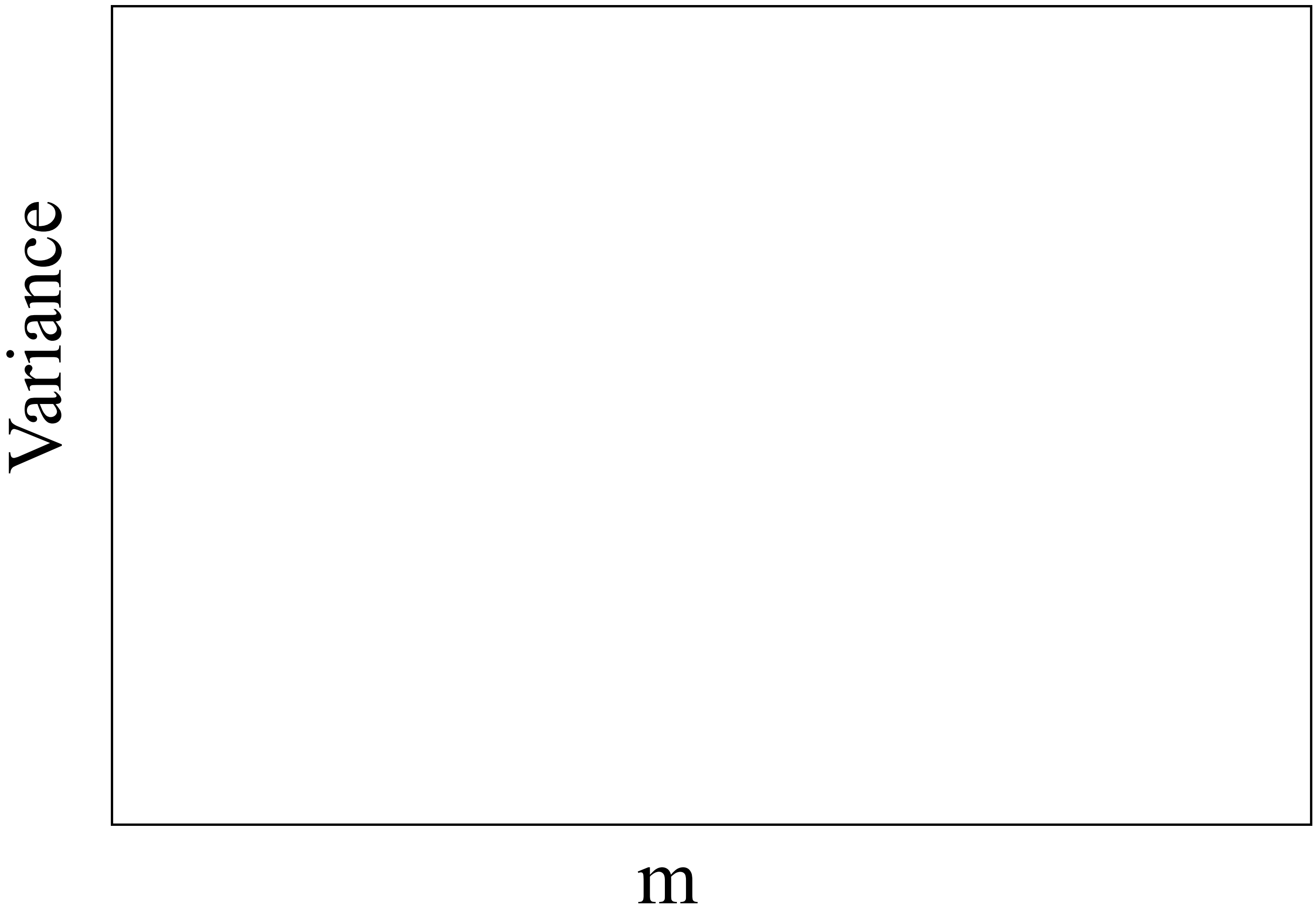
- The variance of the sample decreases more slowly than the reciprocal of the sample size
- For most processes, the variance of a sample diminishes quite rapidly as the sample size is increased, and stabilizes soon
- For self-similar processes, the variance decreases <u>very slowly</u>, even when the sample size grows quite large

Time-Variance Plot

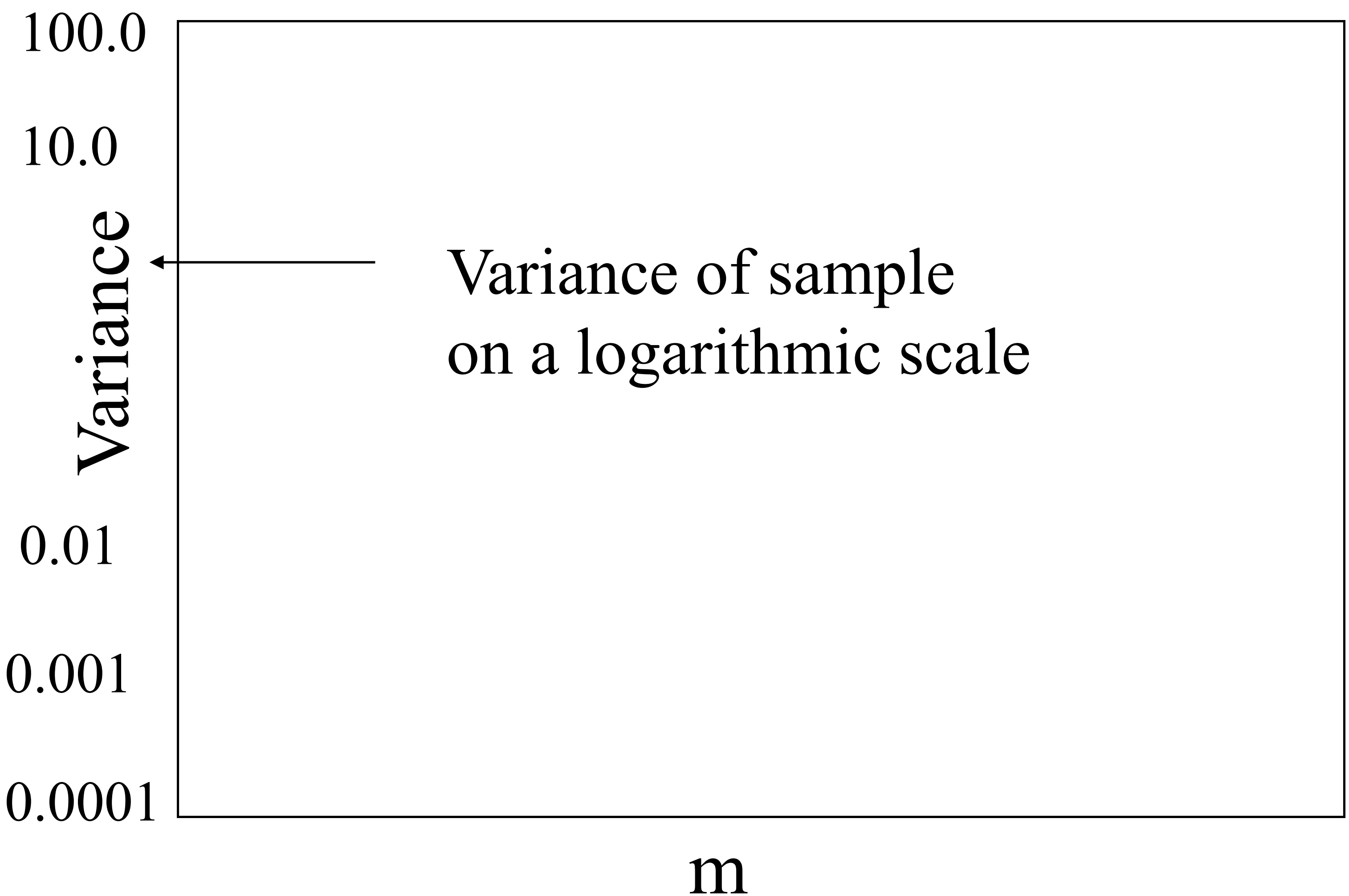


- The "variance-time plot" is one means to test for the slowly decaying variance property
- Plots the variance of the sample versus the sample size, on a log-log plot
- For most processes, the result is a straight line with slope -1
- For self-similar, the line is much flatter

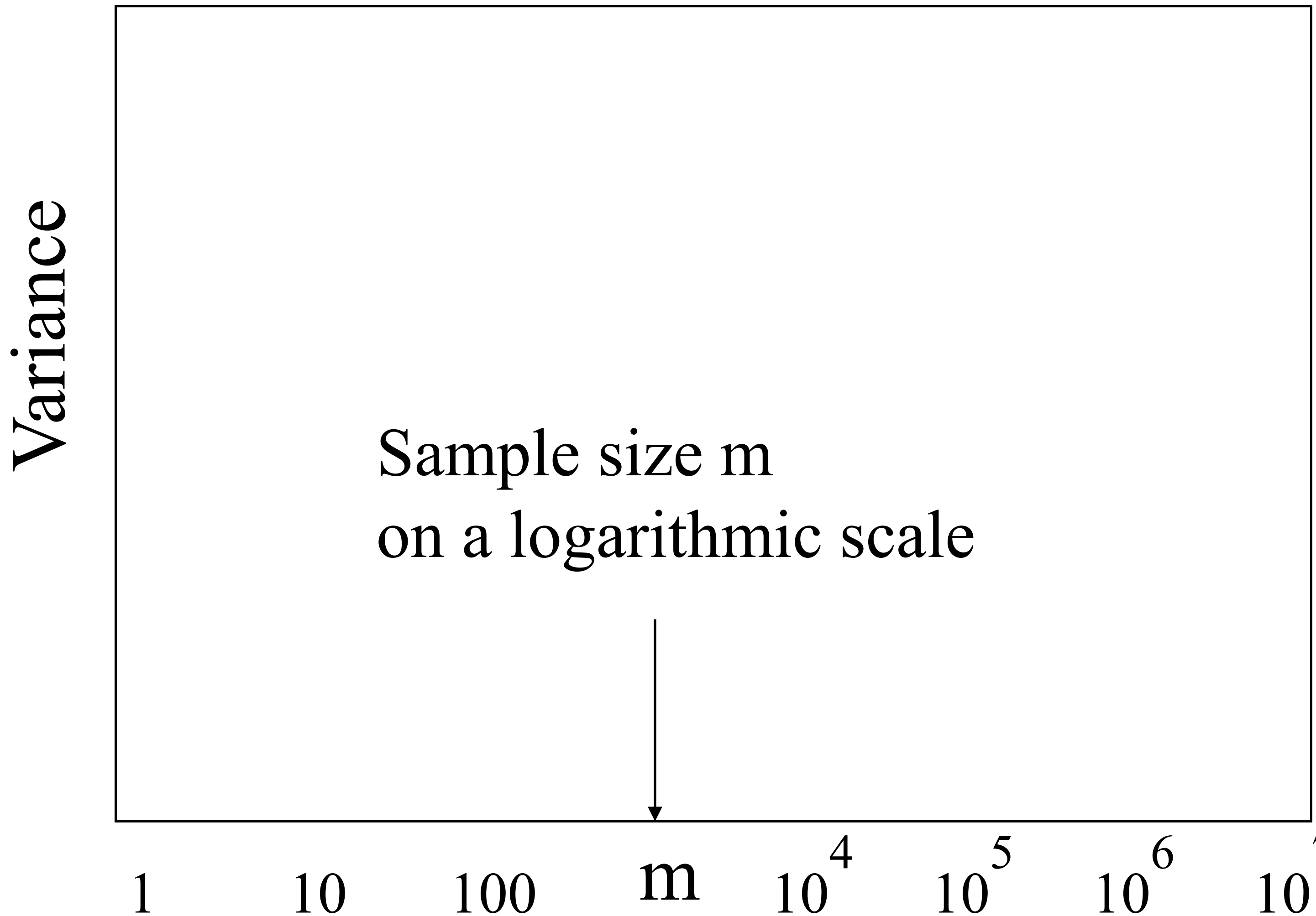
Time Variance Plot



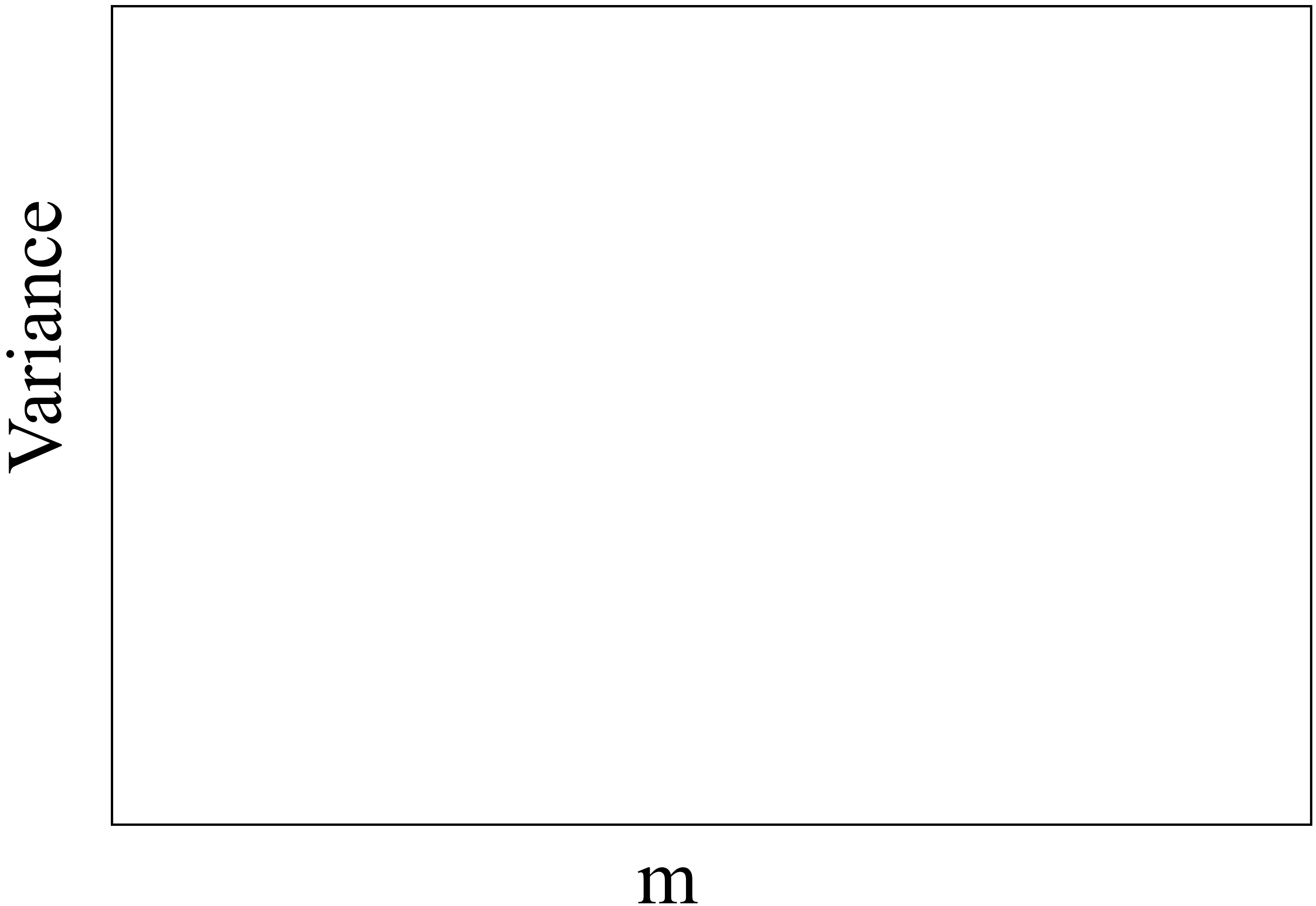
Variance-Time Plot



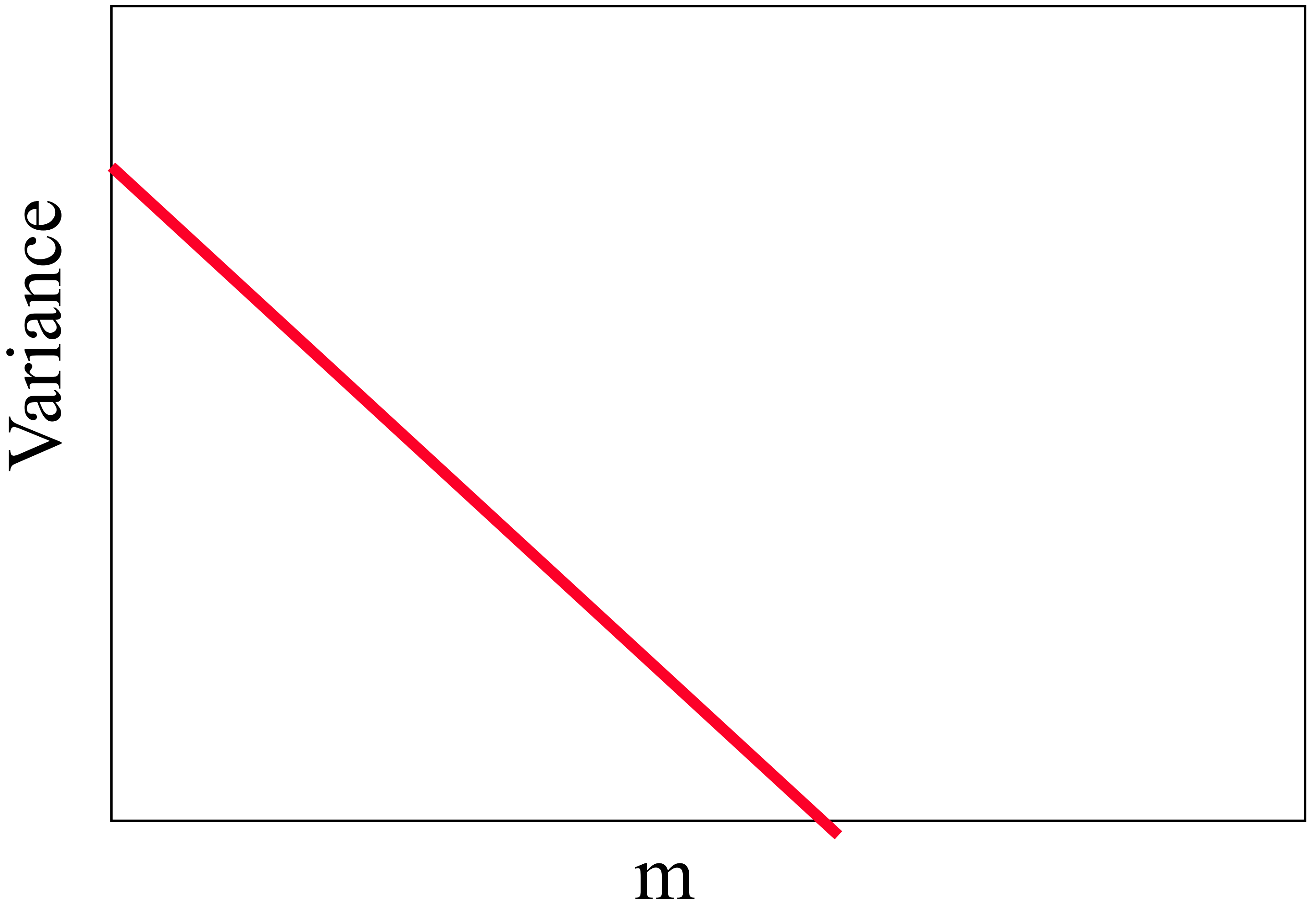
Variance-Time Plot



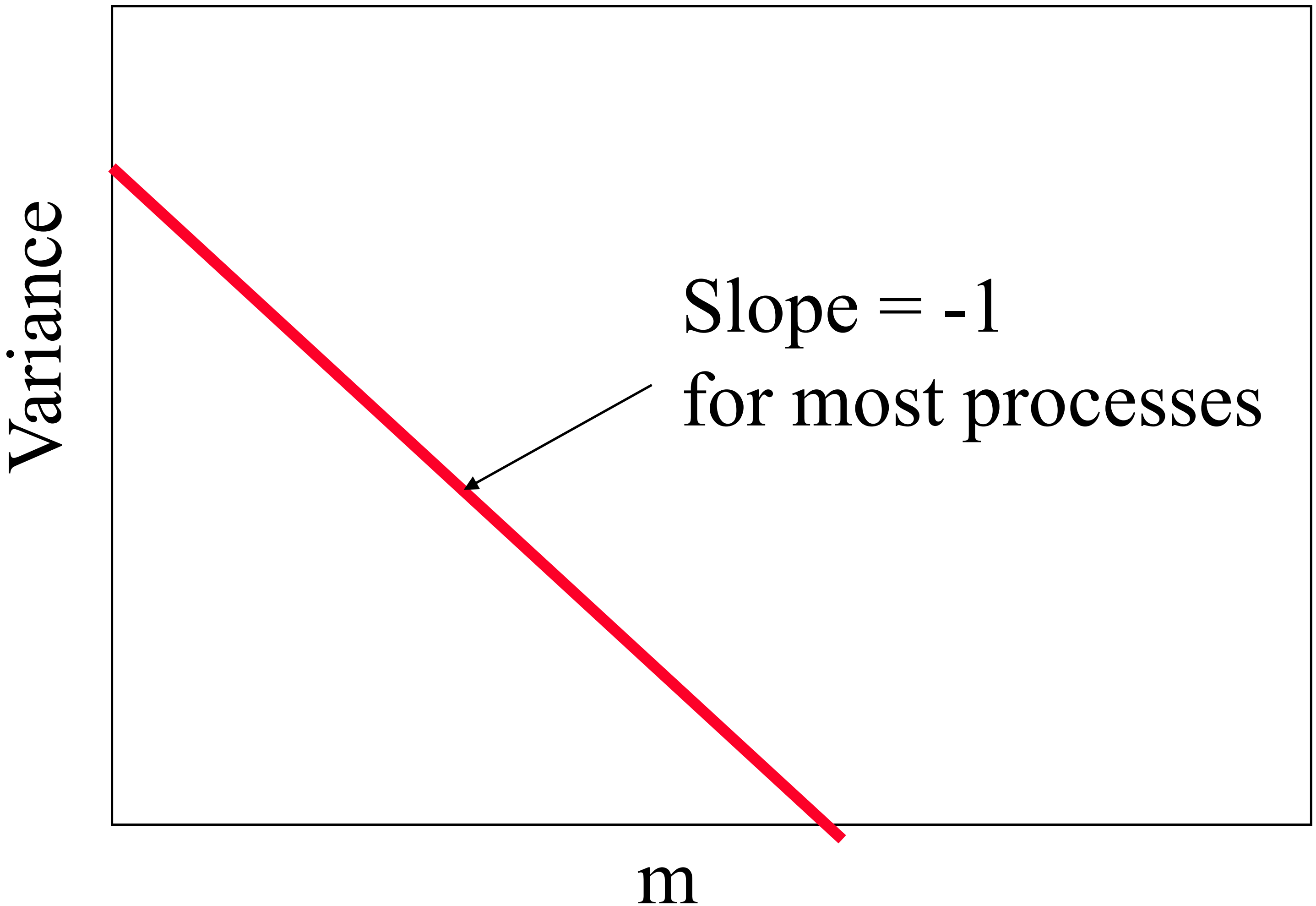
Variance-Time Plot



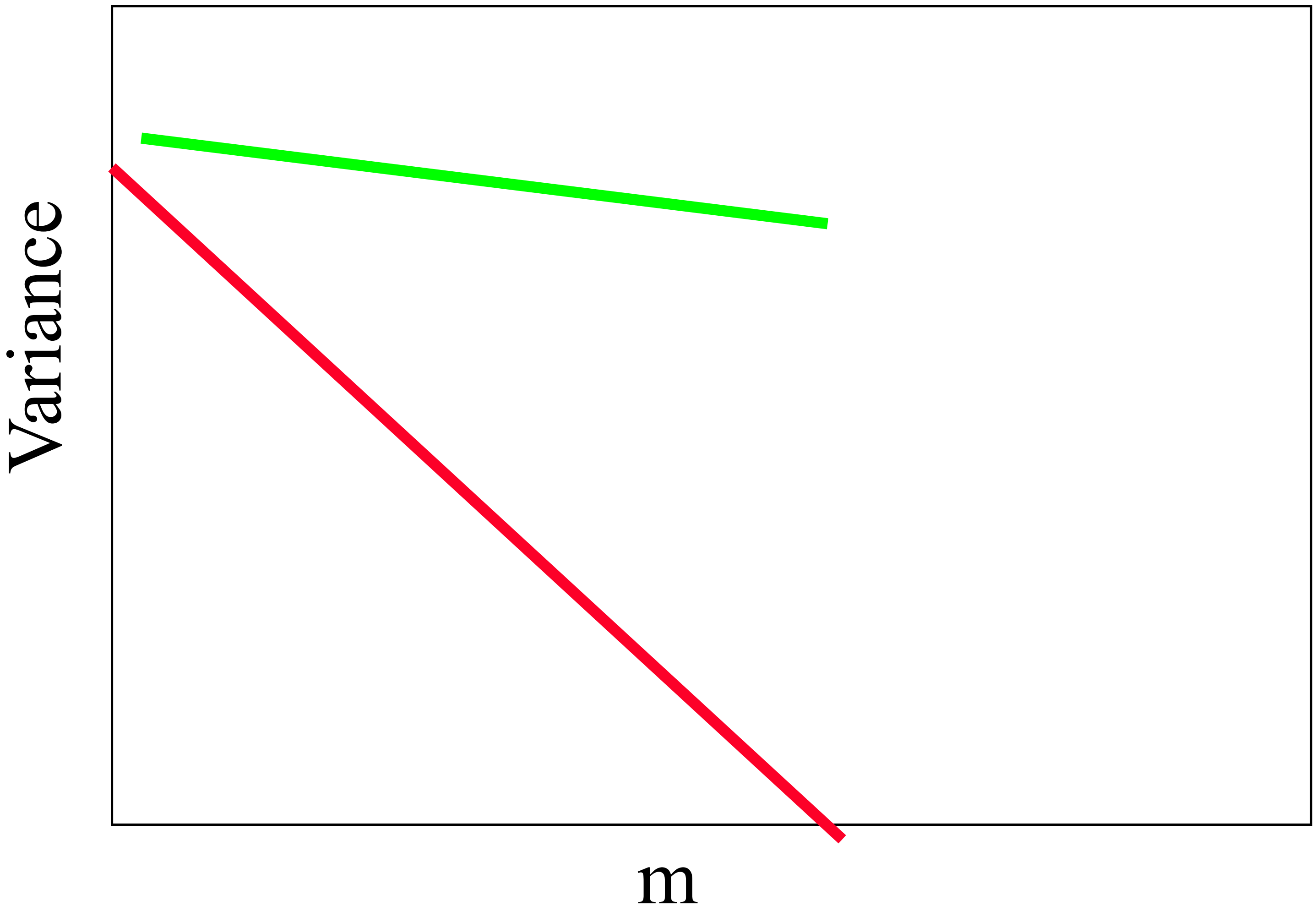
Variance-Time Plot



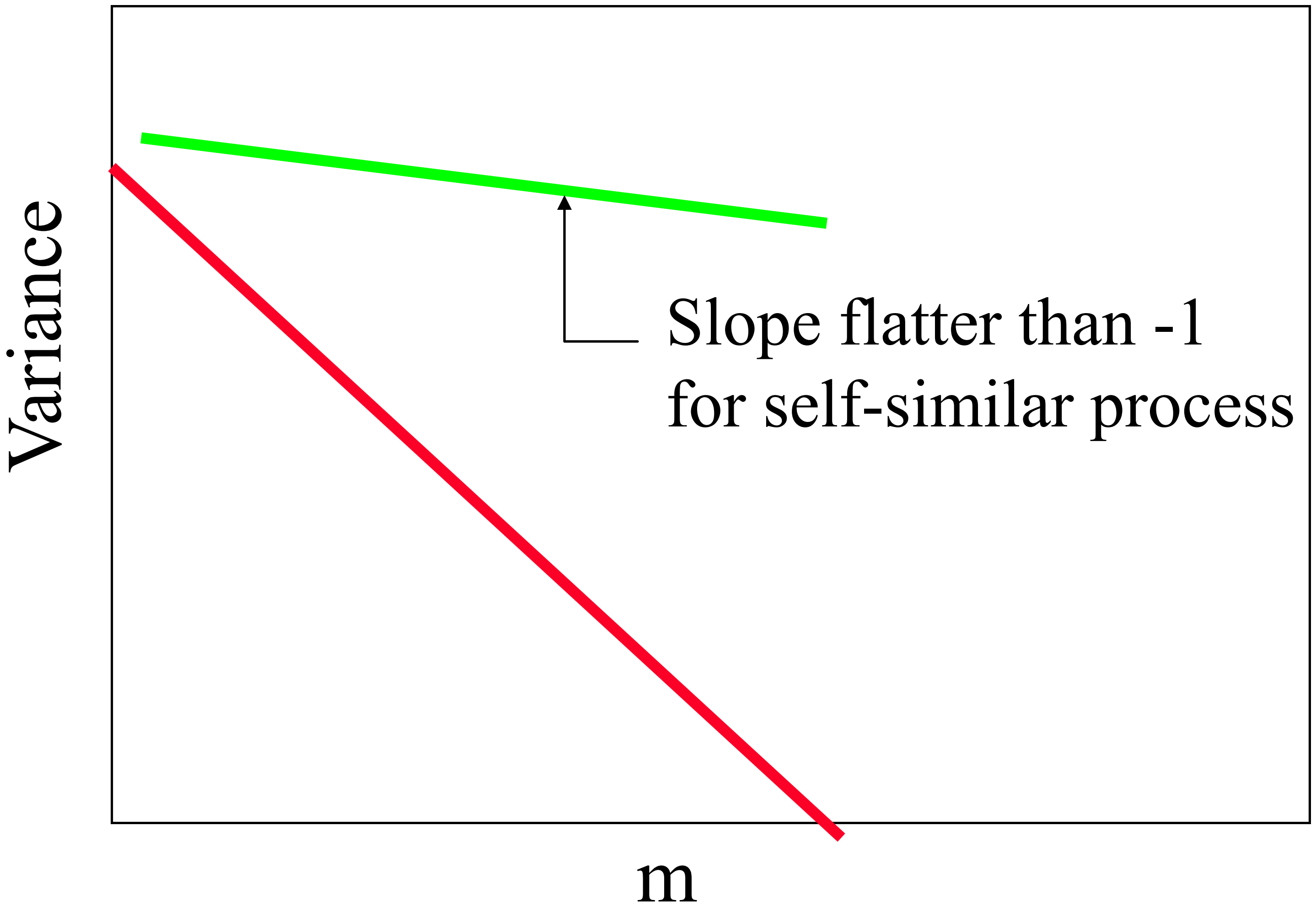
Variance-Time Plot



Variance-Time Plot



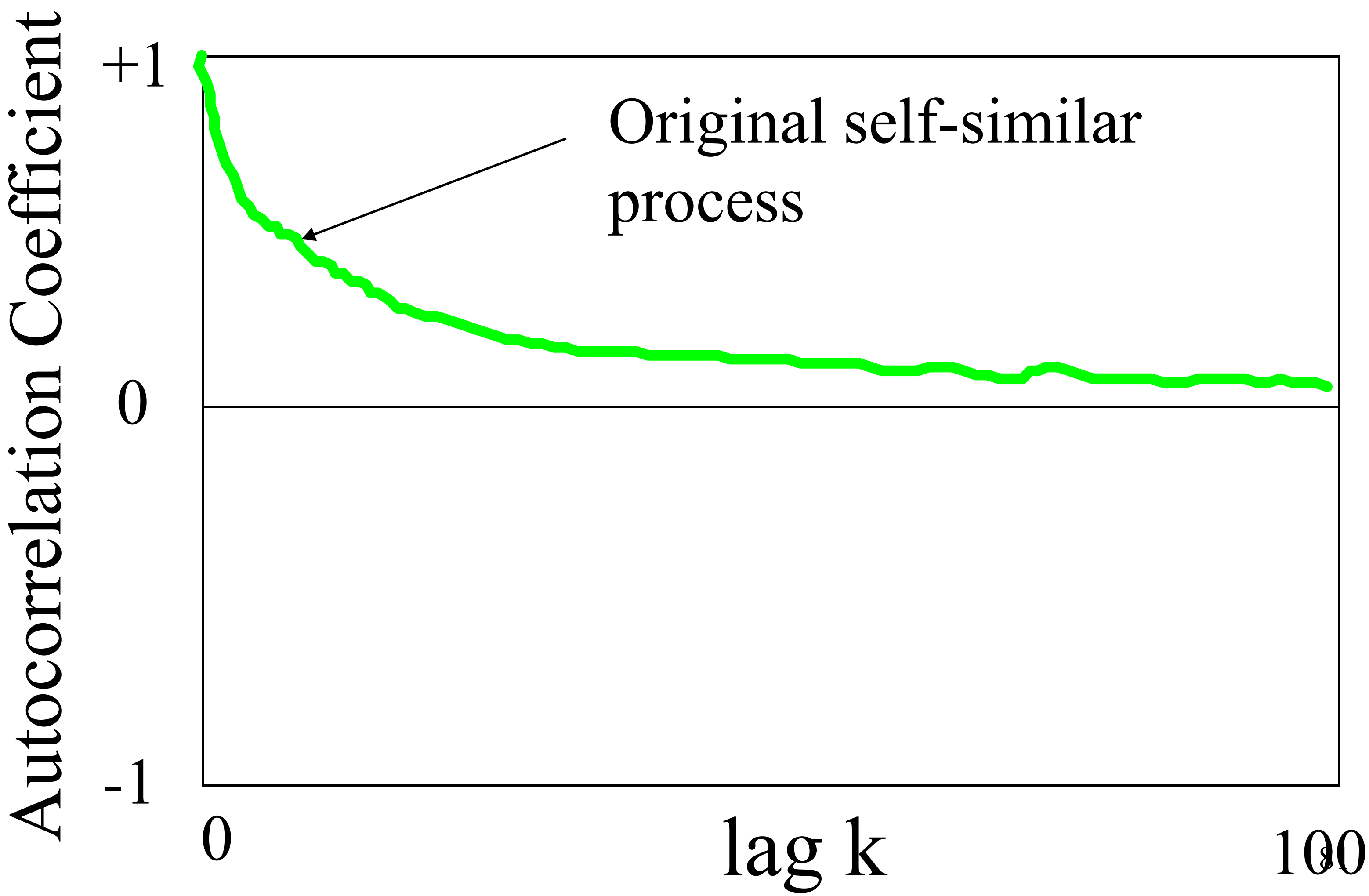
Variance-Time Plot



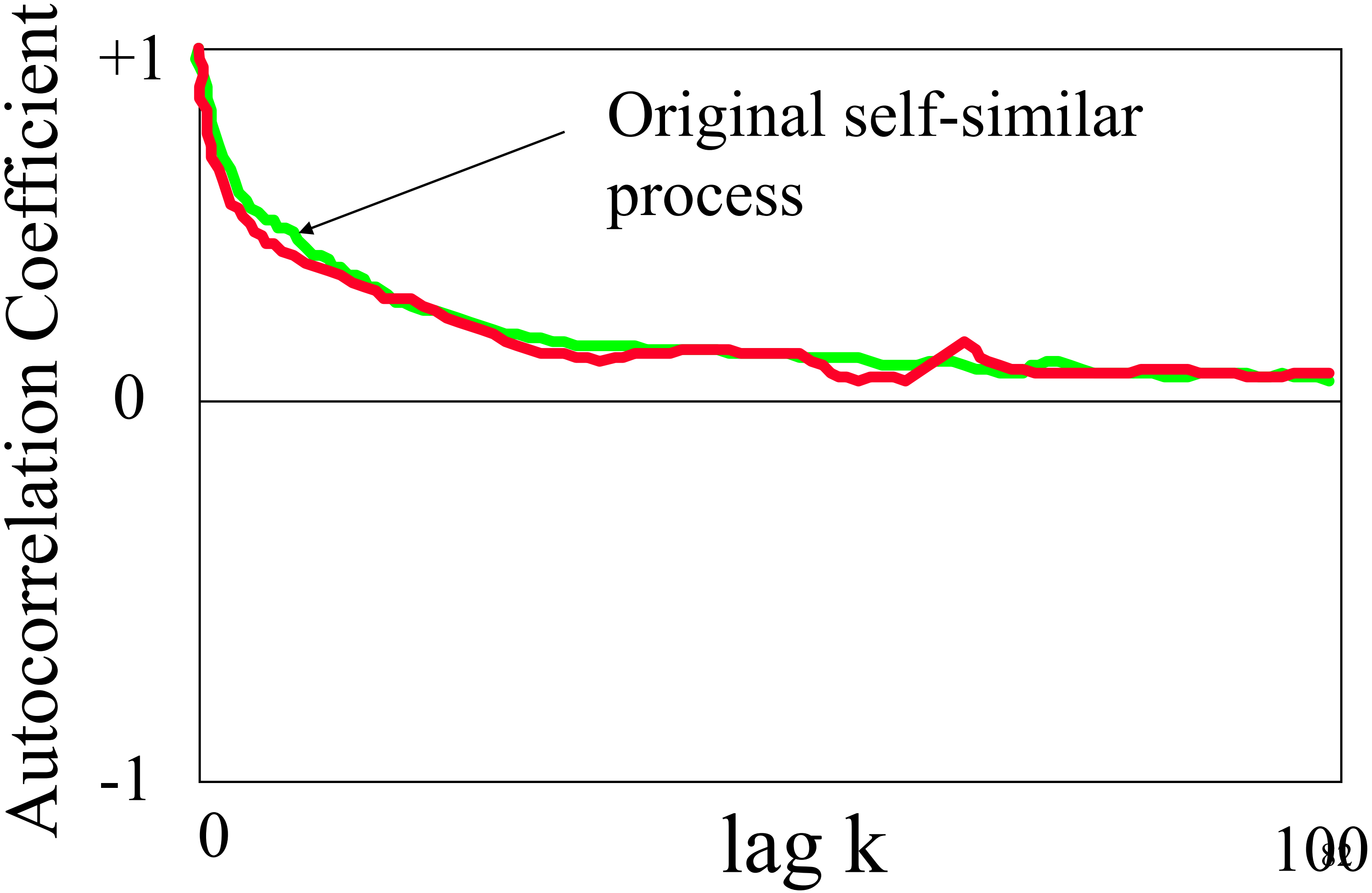
Non-Degenerate Autocorrelations

- For self-similar processes, the autocorrelation function for the aggregated process is indistinguishable from that of the original process
- If autocorrelation coefficients match for all lags k, then called exactly self-similar
- If autocorrelation coefficients match only for large lags k, then called asymptotically self-similar

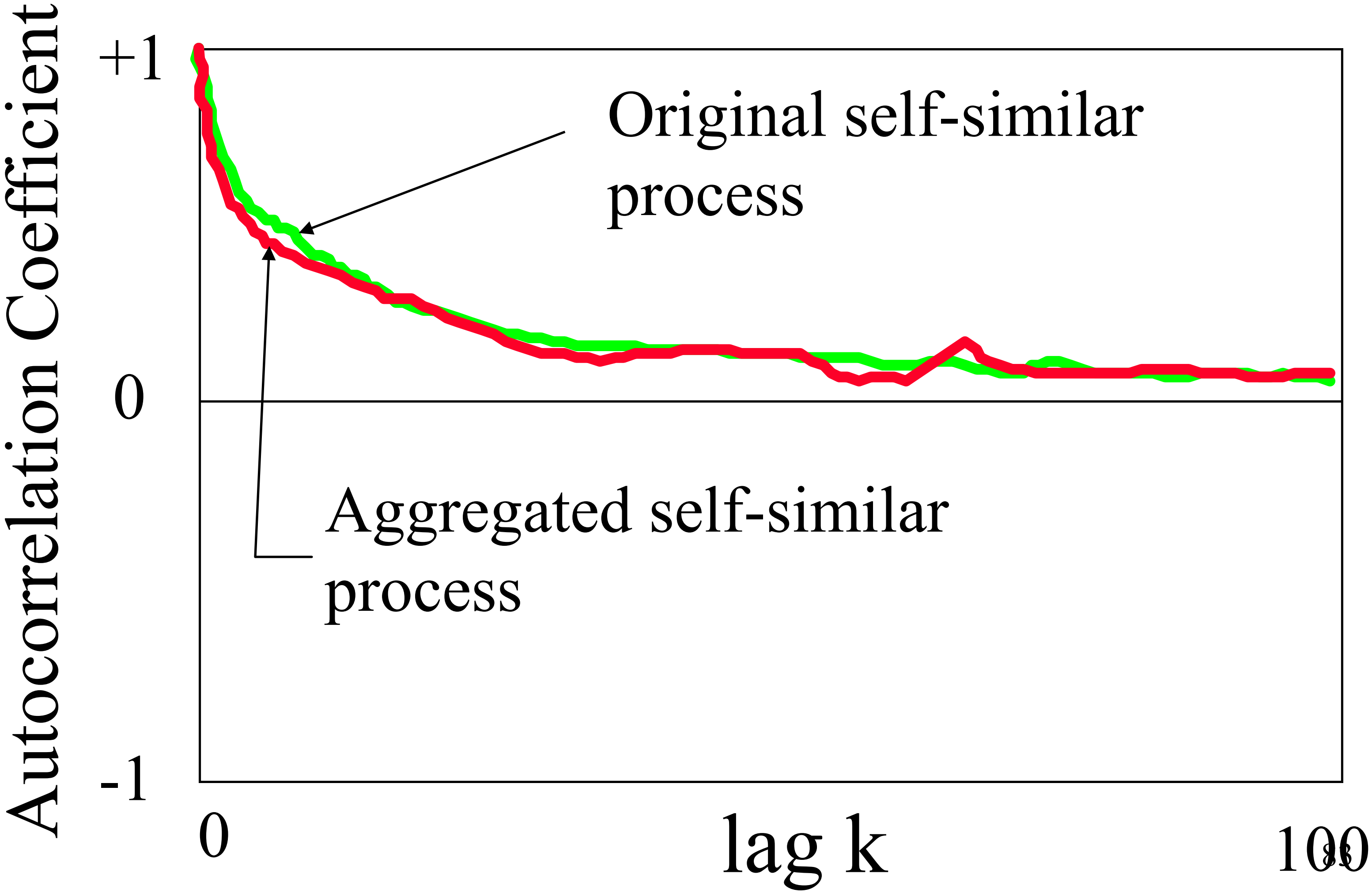
Autocorrelation Function



Autocorrelation Function



Autocorrelation Function



Aggregation

- Aggregation of a time series X(t) means smoothing the time series by averaging the observations over non-overlapping blocks of size m to get a new time series $X_m(t)$

Aggregation Example

- Suppose the original time series X(t) contains the following (made up) values

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

- Then the aggregated series for m = 2 is:

Aggregation Example

- Suppose the original time series X(t) contains the following (made up) values:

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

- Then the aggregated series for m = 2 is:

Aggregation Example

- Suppose the original time series X(t) contains the following (made up) values:

`2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...`

- Then the aggregated series for m = 2 is:

4.5

Aggregation example

- Suppose the original time series X(t) contains the following (made up) values:

2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...

- Then the aggregated series for m = 2 is:

4.5 8.0

Aggregation Example

- Suppose the original time series X(t) contains the following (made up) values:

`2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...`

- Then the aggregated series for m = 2 is:

4.5 8.0 2.5

Aggregation Example

- Suppose the original time series X(t) contains the following (made up) values:

```
2  7  4  12  5  0  8  2  8  4  6  9  11  3  3  5  7  2  9  1...
```

- Then the aggregated series for m = 2 is:

4.5 8.0 2.5 5.0

Aggregation Example

- Suppose the original time series X(t) contains the following (made up) values:

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

- Then the aggregated series for m = 2 is:

4.5 8.0 2.5 5.0 6.0 7.5 7.0 4.0 4.5 5.0...

Aggregation Example

- Suppose the original time series X(t) contains the following (made up) values:

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

Then the aggregated time series for m = 5 is:

Aggregation: An Example

- Suppose the original time series X(t) contains the following (made up) values:

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

Then the aggregated time series for m = 5 is:

Aggregation: An Example

- Suppose the original time series X(t) contains the following (made up) values:

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

Then the aggregated time series for m = 5 is:

6.0

Aggregation: An Example

- Suppose the original time series X(t) contains the following (made up) values:

2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...

Then the aggregated time series for m = 5 is:

6.0 4.4

Aggregation: An Example

- Suppose the original time series X(t) contains the following (made up) values:

2 7 4 12 5	0 8 2 8 4 6	9 11 3 3 5	7 2 9 1...

Then the aggregated time series for m = 5 is:

6.0 4.4 6.4 4.8 ...

Aggregation: An Example

- Suppose the original time series X(t) contains the following (made up) values:

`2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...`

Then the aggregated time series for m = 10 is:

Aggregation: An Example

- Suppose the original time series X(t) contains the following (made up) values:

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

Then the aggregated time series for m = 10 is:

5.2

Aggregation: An Example

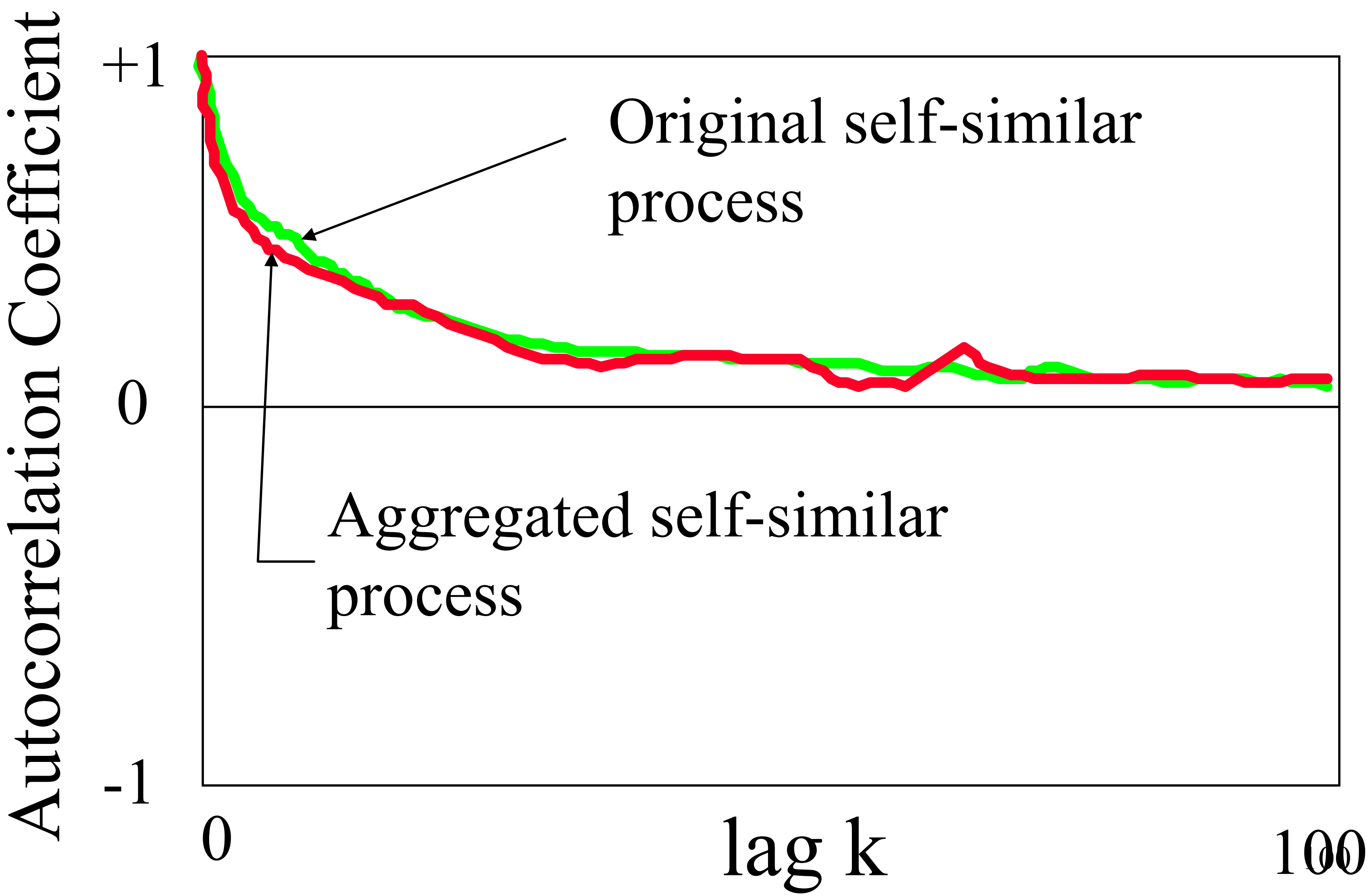
- Suppose the original time series X(t) contains the following (made up) values:

```
2 7 4 12 5 0 8 2 8 4 6 9 11 3 3 5 7 2 9 1...
```

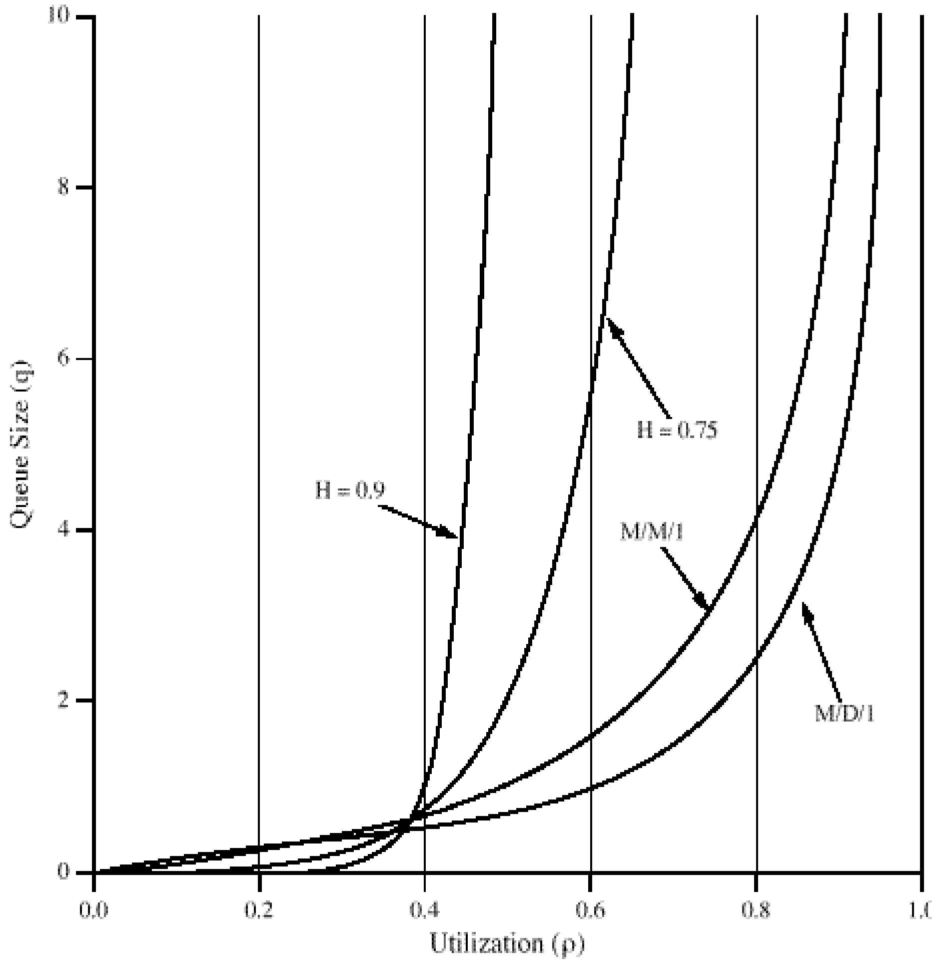
Then the aggregated time series for m = 10 is:

5.2 5.6

Autocorrelation Function



Impact of Self-similarity of Network Performance



From: High-Speed Networks, William Stallings. Prentice Hall, 1998

Self-Similarity Summary

- Self-similarity is an important mathematical property that has recently been identified as present in network traffic measurements
- Important property: burstiness across many time scales, traffic does not aggregate well
- There exist several mathematical methods to test for the presence of self-similarity, and to estimate the Hurst parameter H
- There exist models for self-similar traffic
- Self-Similarity impacts system performance
- Self-Similarity can be difficult to simulate