Traffic Engineering
EN3370

Chathuranga Weeraddana

February 2021

Lecture 9: Modeling Network Traffic II

Why Traffic Modelling?

Logistics of Operating a Network

- continual evaluation of networks' performance
- associated methods include
  - measurements
  - analysis
  - modeling
  - subsequent simulations
- traffic modeling plays a pivotal role

Traffic Modeling

- Markov models
- Poisson models
- Bernoulli models
- autoregressive models
- self-similar statistical models
- stochastic processes with long-range dependence
- heavy-tailed distributions

Traffic Modeling

- we already discussed Markov models, Poisson models, Bernoulli models, and autoregressive models

Key Characteristic (Burstiness) [2]

- many models above do not allow descriptions of bursts
- bursts are distinctive in today's network traffic [1]
- a natural length of a burst is no longer present

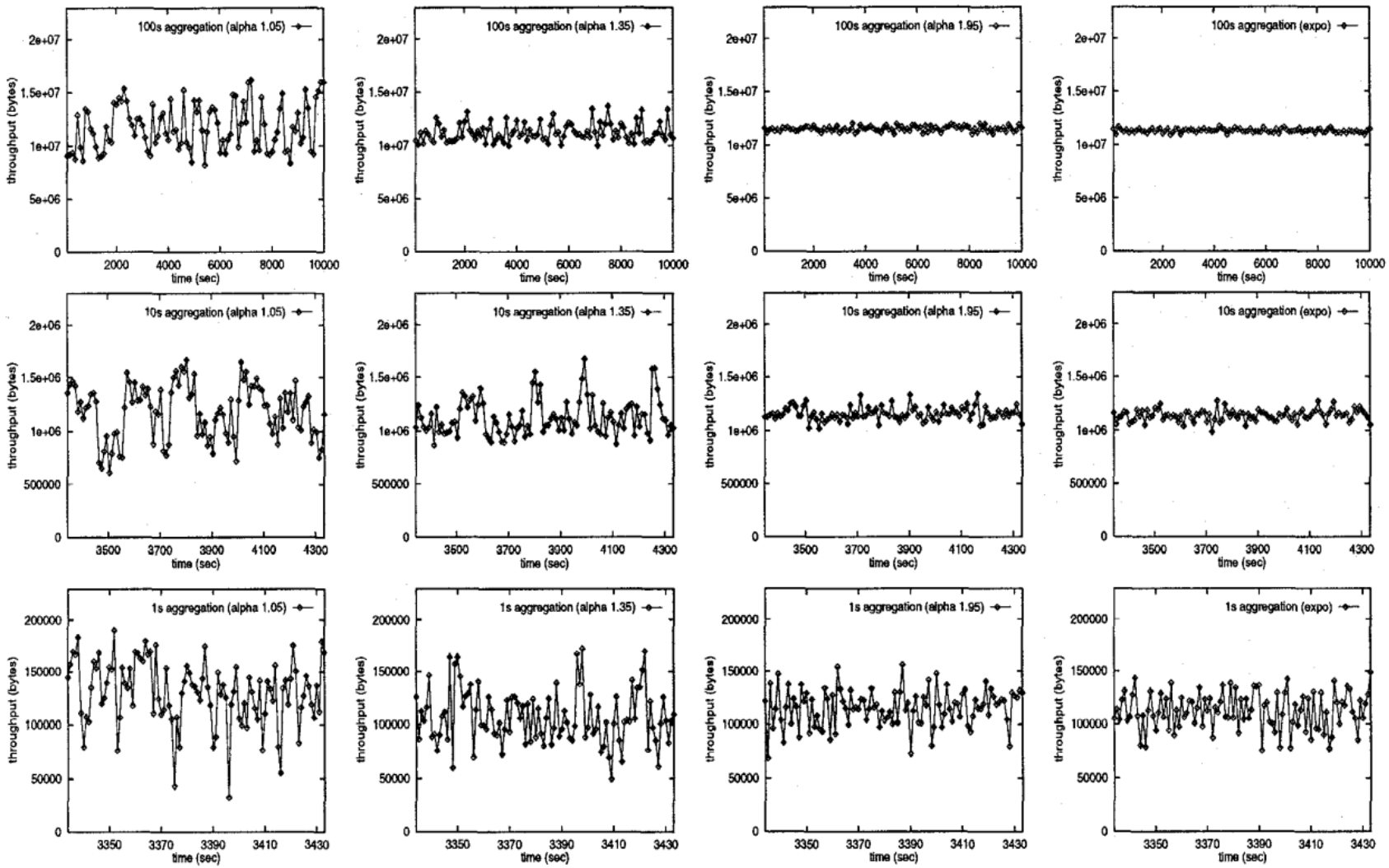
[1] [Leland et al., 1994]: W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," in IEEE/ACM Transactions on Networking, vol. 2, no. 1, pp. 1-15, Feb. 1994.

[2] See § 15 of [Ortiz-2012]: J. H. Ortiz, Telecommunications Networks: Current Status and Future Trends, InTech, 2012.

Key Characteristic (Burstiness)

- empirical evidences suggest that interarrivals of packets preserves burstiness over many time scales, e.g.,
  - wide-area traffic is much burstier than Poisson models predict
- Poisson arrival processes are quite limited in their burstiness, especially when multiplexed to a high degree

Numerical Experiments [3]



[3] [Park et al., 1996]: K. Park, G. T. Kim, and M. E. Crovella, *On the relationship between file sizes, transport protocols, and self-similar network traffic*, ICNP, 1996.

Numerical Experiments

- authors in [Park et al., 1996] showed that
  - the degree of scale-invariant burstiness can be modeled by the a parameter of the Pareto distribution [4]
  - a approaches $1 \rightarrow$ burstiness is preserved even at large time scales $\rightarrow$ the time series possesses long-range dependency
  - $a \approx 2 \rightarrow$ a smoothing effect as the aggregation level is increased $\rightarrow$ indicating a weak dependency structure in the underlying time series
  - exponential $\rightarrow$ cannot effectively model the long-range dependency over a range of time scales

[4] See page 33 of this lecture slides.

Key Characteristic (Burstiness)

- at every time scale ranging from a few milliseconds to minutes and hours, bursts consist of bursty subperiods separated by less bursty subperiods
- the nature of congestion differs drastically from that predicted by classic traffic models
- more care is to be taken, e.g., in design, control, and performance analysis result

Key Characteristic (File Size) [5]

- network traffic is researched from application or data source point of view
- focused on statistics of file sizes
- WWW network traffic $\rightarrow$ file sizes of such traffic are best described by Pareto distribution with shape parameter 1
- FTP traffic $\rightarrow$ Pareto distribution with shape parameter $0.9 - 1.1$

[5] See § 15 of [Ortiz-2012]

Key Characteristic (Inter-Arrival Time) [6]

- inter-arrival time of TCP connections $\rightarrow$ self similar
- can be described by Weibull heavily-tailed distribution

[6] See § 15 of [Ortiz-2012]

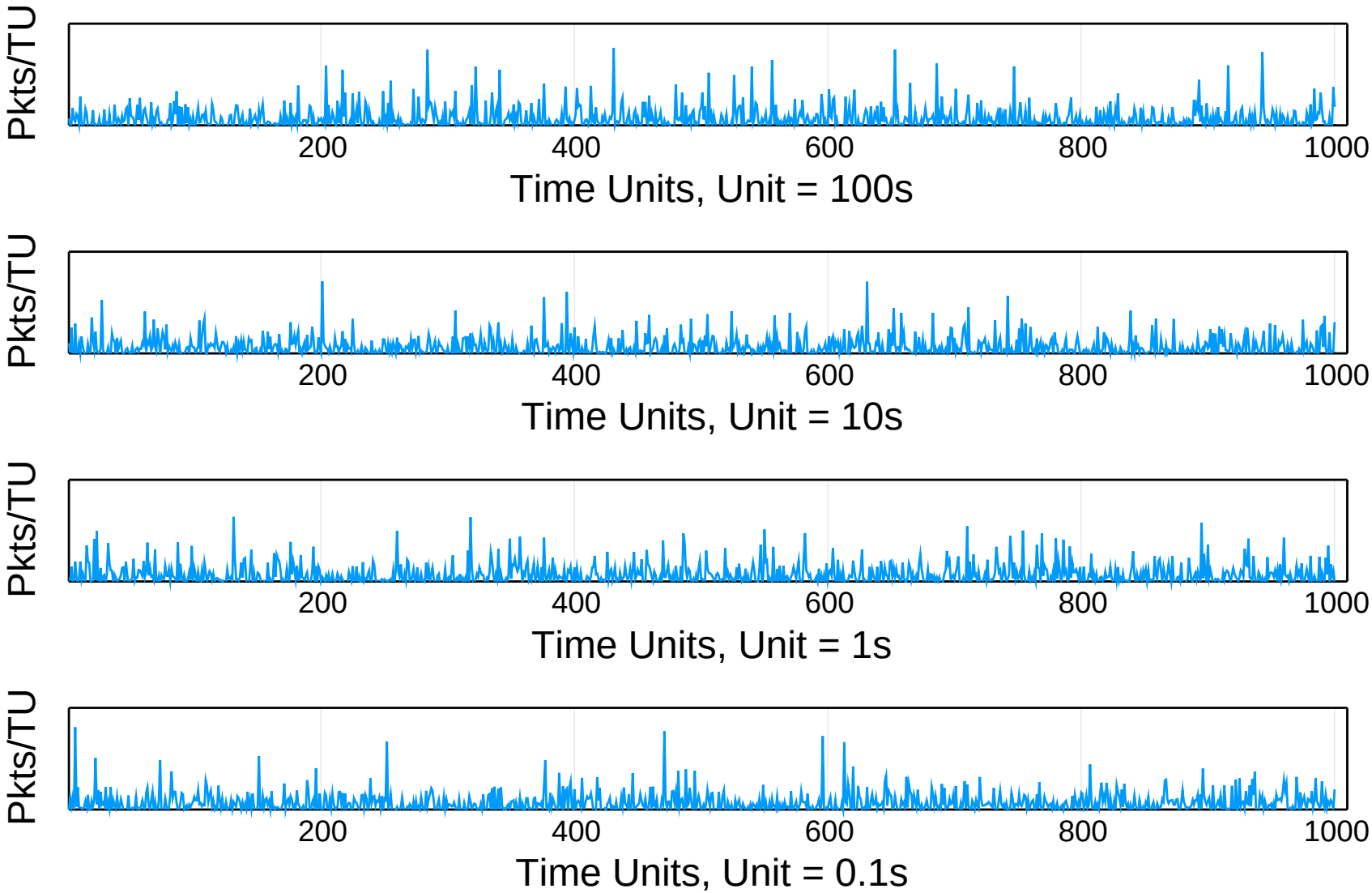
Modeling

- bursts can be described by a self-similarity model
- bursts over a wide-range of time scales can be modeled
- correlated with another very important property: long-range dependence (LRD)
- long-range dependence is also manifested with heavy-tailed probability of density distributions

Modeling

- heavy-tailed probability of density distributions
  - Pareto
  - Weibull
- Pareto's and Weibull's distributions became the most frequently used distributions to describe self-similar network traffic

Self-Similarity



Mathematics of Self-Similarity

Self-Similarity

- not just an intuitive description
- precise concept is captured by the rigorous mathematical definitions [7]

[7] For a rigorous treatment of the subject, the book [Park et al., 2000]: K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*, John Wiley & Sons Inc., 2000 or [Sheluhin et al., 2007]: O. I. Sheluhin, S. M. Smolskiy, and A. V. Osin, *Self-Similar Processes in Telecommunications*, John Wiley & Sons, Ltd, 2007, can be referred.

Autocorrelation Function of X

- let $\{X_t\}_{t=1,2,\ldots}$ be a wide-sense stationary process
- for clarity let us define

$$X \triangleq \{X_t\}_{t=1,2,\ldots}$$

- thus we have
  - constant mean $\mu = \mathbb{E}[X_t]$
  - variance $\sigma^2 = \mathbb{E}[(X_t - \mu)^2]$
  - autocorrelation function r depends only on k

Another Process $X^{(m)}$ from X

- to this end, average the original process X over non-overlapping blocks of size m
- let $m \in \{1, 2, 3, \ldots\}$
- let us define

$$X^{(m)} \triangleq \left\{ X_k^{(m)} \right\}_{k=1,2,\ldots}$$

where

$$X_k^{(m)} \triangleq \frac{1}{m} \left(X_{km-m+1} + \cdots + X_{km} \right)$$

Another Process $X^{(m)}$ from X

- the process $X^{(m)}$ is also called an aggregation process
- m is called the aggregation level
- it can be shown that $X^{(m)}$ is wide-sense stationary
- we denote by $r^{(m)}$ the corresponding autocorrelation function

Definitions for Self-Similarity

Self-Similarity: Definitions [8]

- X is exactly second-order self-similar with Hurst parameter H $(0.5 < H < 1)$ if

$$r(k) = \frac{\sigma^2}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right], \quad \forall\, k \geq 1 \quad (1)$$

- X is asymptotically second-order self-similar with Hurst parameter H $(0.5 < H, 1)$ if

$$\lim_{m\to\infty} r^{(m)}(k) = \frac{\sigma^2}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right], \forall\, k \geq 1 \quad (2)$$

- the second-order self-similarity $\rightarrow$ correlation structure is preserved under time aggregation

[8] See [Park et al., 2000].

Other Point of Views

- consider the relation

$$X =_d \quad m^{1-H} X^{(m)} \tag{3}$$

  where $=_d$ denote the equivalence in finite-dimensional distribution

- $X^{(m)}$ is related to X via a simple scaling relationship involving H

- (3) holds for all m: exactly second-order self-similar

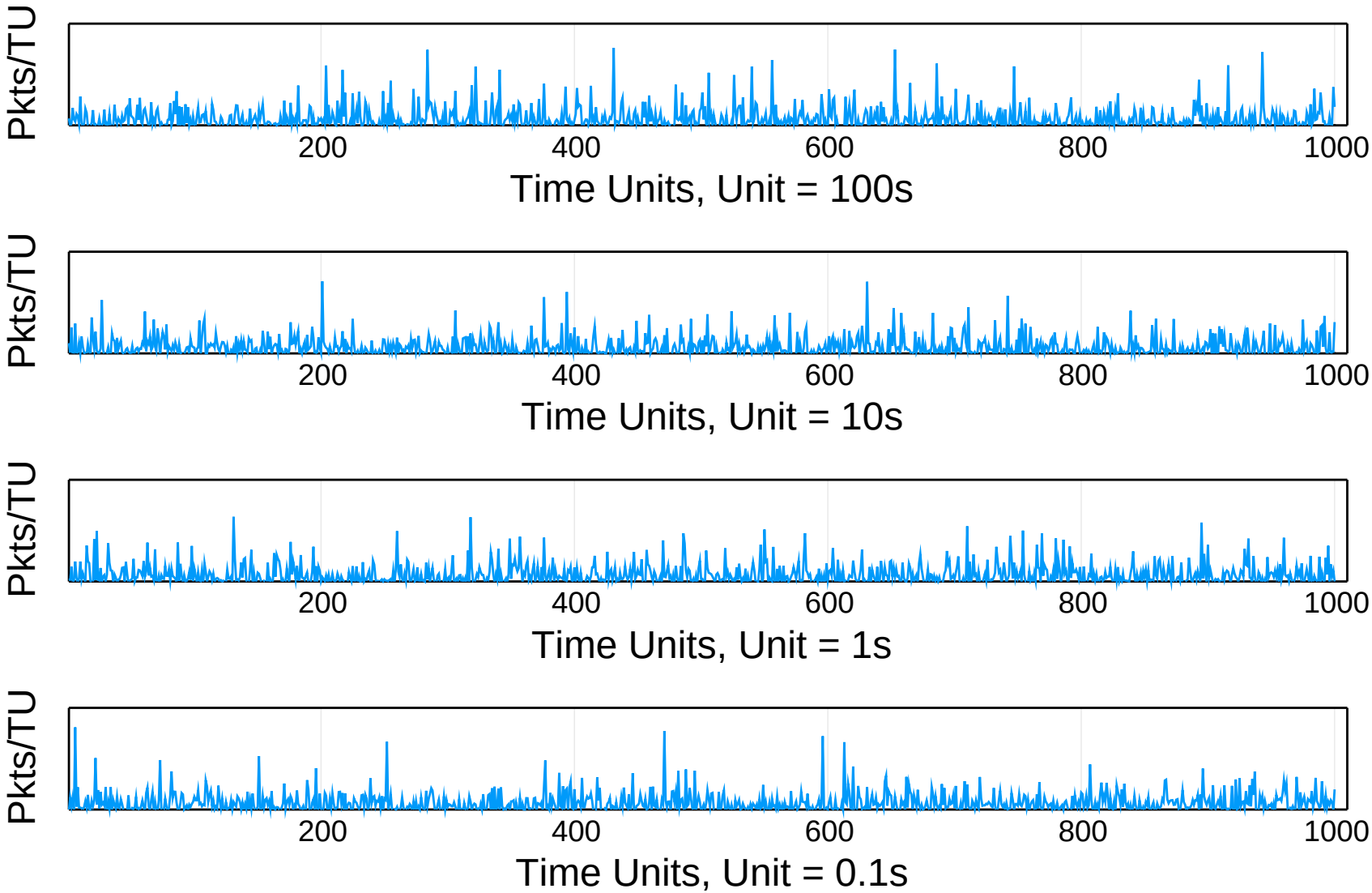
- (3) holds as $m \to \infty$: exactly second-order self-similar

Equivalent Features

Equivalent Features

- the statistical similarity when scaling

Statistical Similarity



Equivalent Features

- the sample variance of aggregated processes decreases more slowly than the magnitude and is inversely proportional to the sample size [9], i.e.,

$$\text{Var}\left(X^{(m)}\right) \sim \sigma^2 m^{-\beta} \quad \text{as } m \to \infty$$

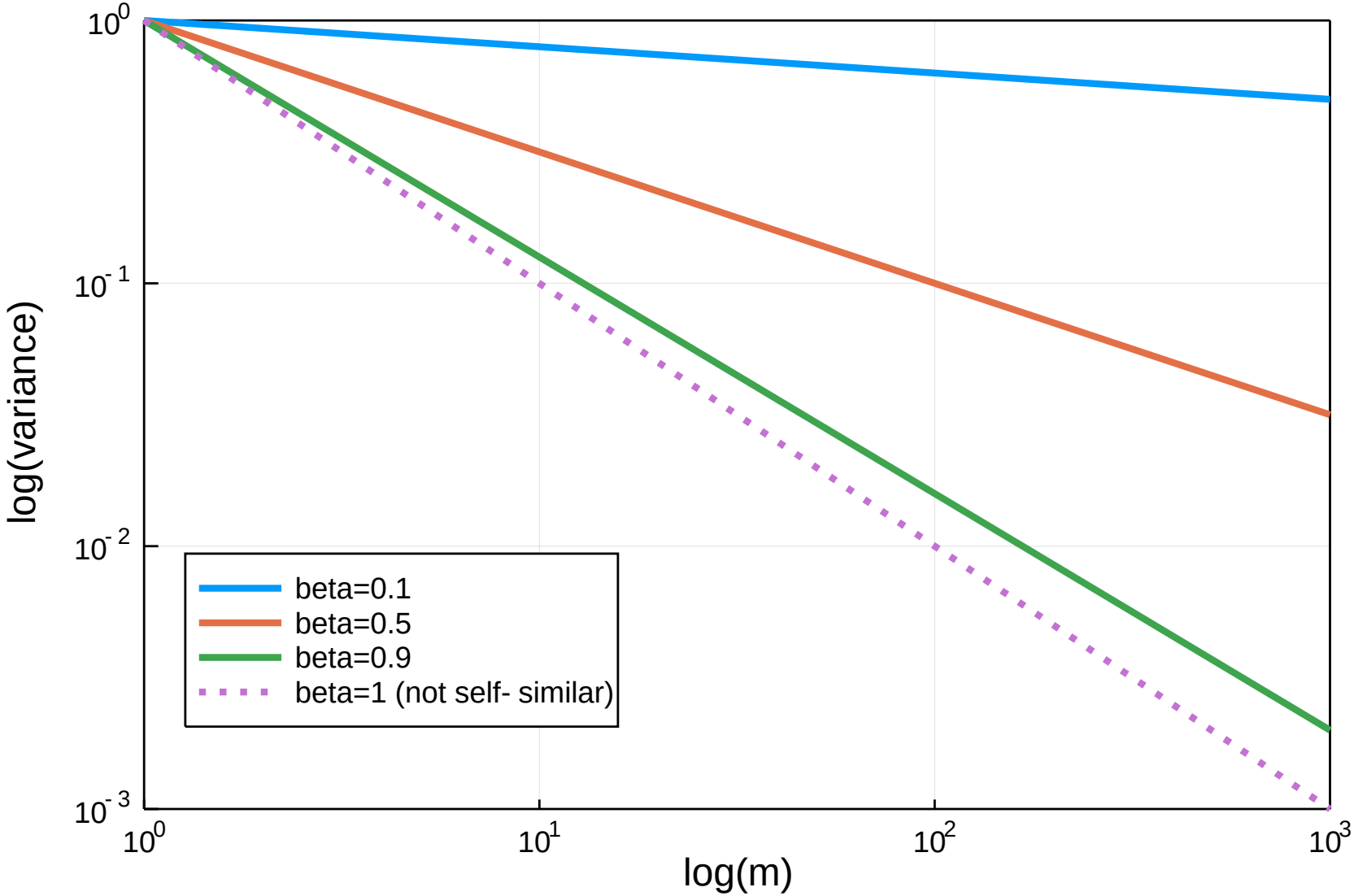
where $\beta = 2 - 2H$

[9] Compare with (3).

Equivalent Features

- recall: $\beta = 2 - 2H$
- $H = 0.5 \rightarrow$ independent samples (then $\beta = 1$)
- $H = 1 \rightarrow$ underlying processes are mathematically pathological
- recall: the range $0.5 < H < 1$ or equivalently $0 < \beta < 1$
- this hints a dependency structure in the "samples" that makes $\text{Var}\left(X^{(m)}\right) \rightarrow 0$ slower than $(1/m)$

Variance of Aggregated Processes



Equivalent Features

- X has an autocorrelation function of the form:

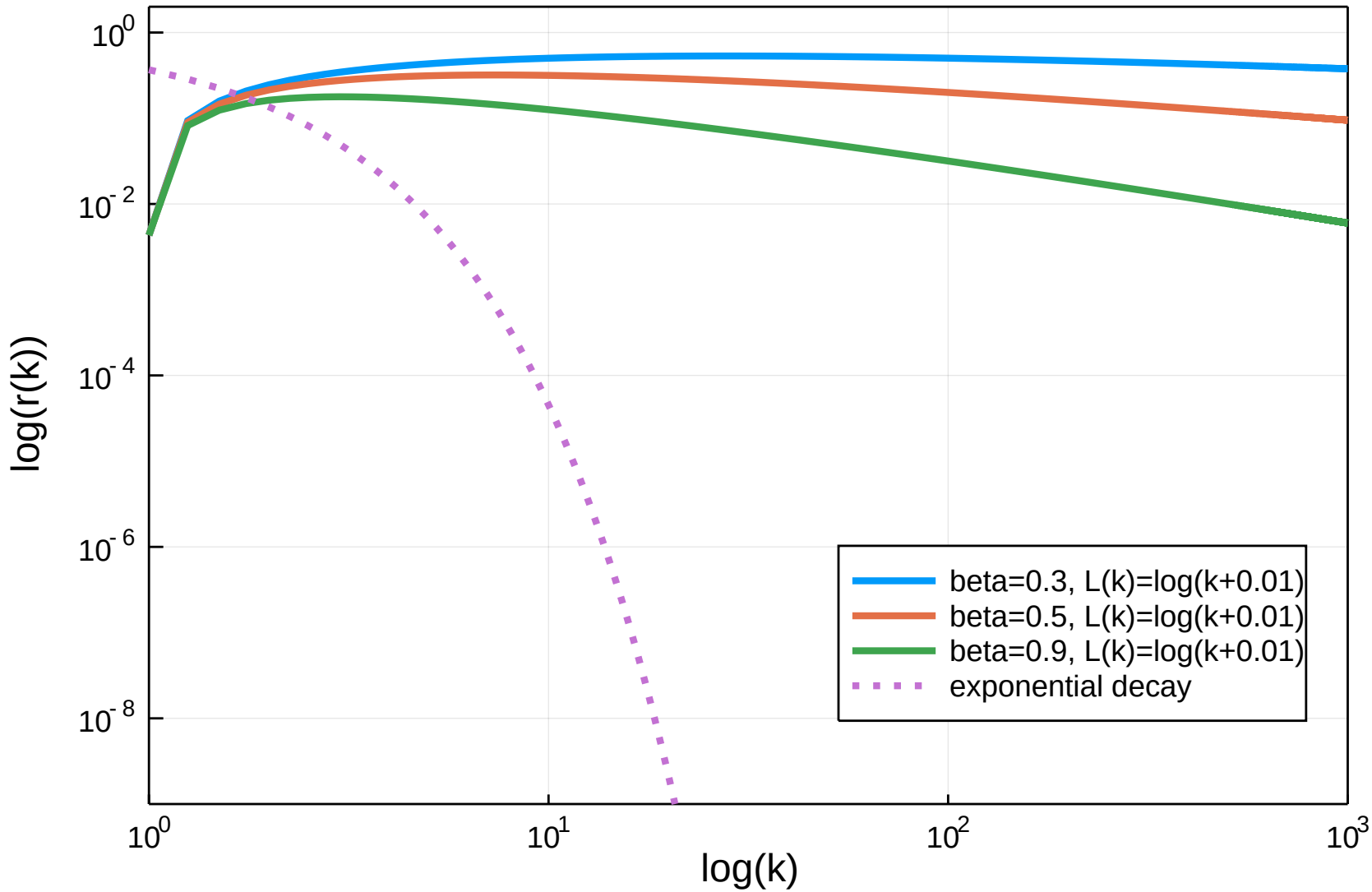
$$r(k) \cong k^{-\beta} L(k), \quad \text{as } k \to \infty \tag{4}$$

  where $0 < \beta < 1$ and $L(k)$ is a slowly varying function [10] at large values of k

- the autocorrelations decay hyperbolically rather than exponentially fast

[10] See Appendix for an example.

Hyperbolic Decay of $r(k)$



Equivalent Features

- the covariance function is nonsummable, i.e.,

$$\sum_k r(k) = \infty$$

- infinite sum is another definition of the LRD

- above is crucial $\rightarrow$ cumulative effect of a wide range of delays can essentially differ from those observed in a short-range dependent process (e.g., Poisson, AR)

Self-Similarity, LRD, and Heavy-Tailed Distributions

Self-Similarity Versus LRD

Self-Similarity Vs LRD

- in general,

$$\text{self-similarity} \not\Longrightarrow \text{LRD}$$
$$\text{LRD} \not\Longrightarrow \text{self-similarity}$$

- however, for $0.5 < H < 1$,

$$\text{self-similarity} \iff \text{LRD} \tag{5}$$

- (5): self-similarity and LRD used interchangeably when the context is clear

LRD Versus Heavy-Tailed Distributions

LRD Vs Heavy-Tailed Distributions

- there is an intimate relation between the two
- lets first have some useful definitions

Heavey-Tailed Distributions

- a RV X has a heavy-tailed distribution if

$$P[X > x] \sim cx^{-\alpha} \text{ as } x \to \infty$$

  where $0 < \alpha < 2$ and c is a positive constant

- α : the tail index or the shape parameter
- tail decays hyperbolically [11]

[11] This is in contrast to the distributions of exponential and Gaussian RVs.

Heavey-Tailed Distributions

- some properties are qualitatively different than those from more commonly encountered exponential
- $0 < \alpha < 2 \implies$ infinite variance
- $0 < \alpha \leq 1 \implies$ infinite mean
- α decreases, a large portion of the probability mass is present in the tail
- i.e., extremely large values are there with non-negligible probability

Pareto Distributions

- one of the frequently used heavy-tailed distributions
- distribution function F_X has the form

$$F_X(x) = P[X \leq x] = 1 - (b/x)^a, \qquad x \geq b$$

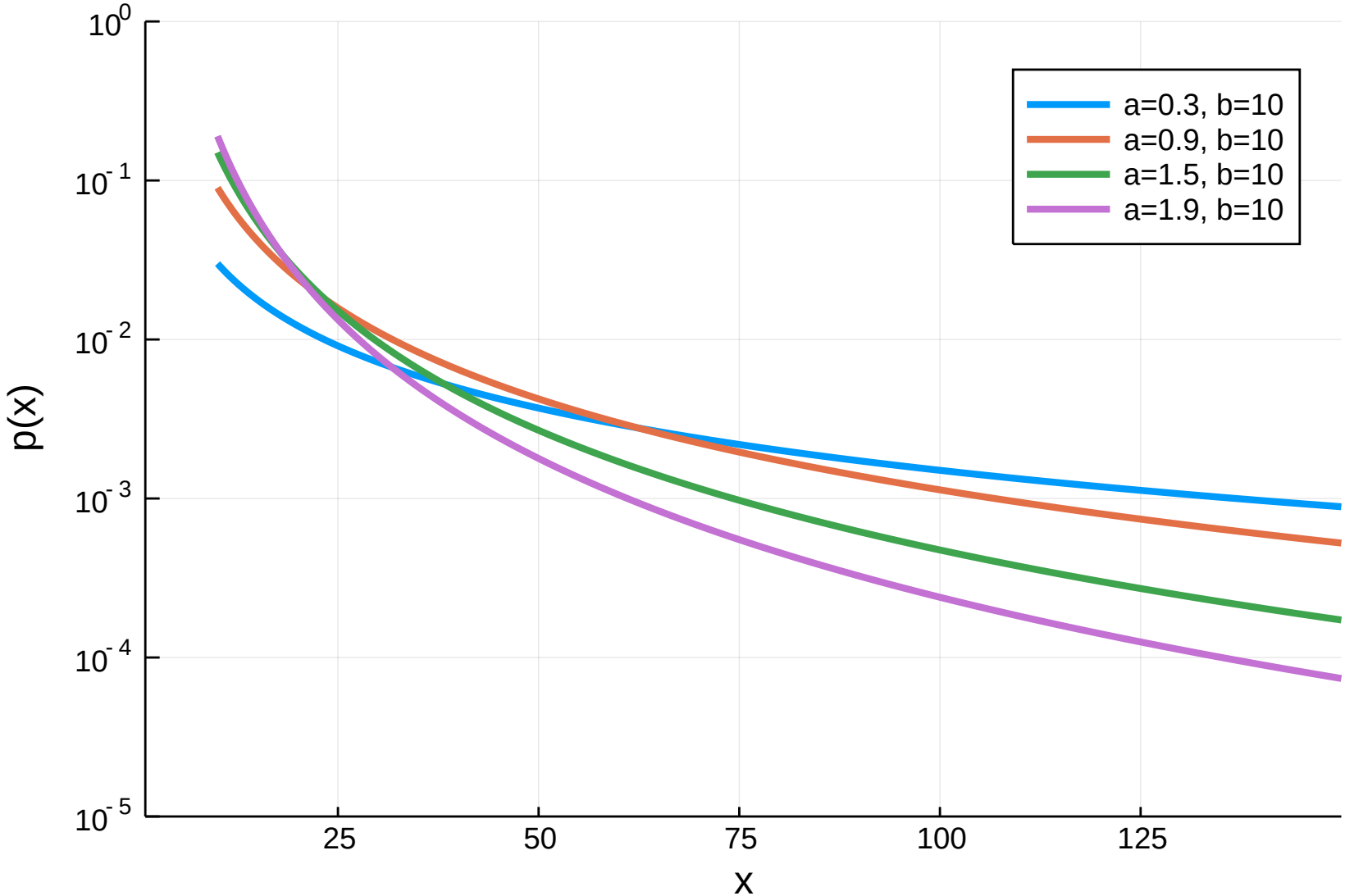
where $0 < a < 2$ is the shape parameter and b is the position parameter

- probability density function is given by

$$p_X(x) = ab^a x^{-a-1}$$

where $x \geq b$

Pareto Distributions



LRD Vs Heavy-Tailed (Cont.)

- heavy-tailedness related to a variable/activity (e.g., file size, connection duration) can be shown to underlie the root cause of LRD (thus of self-similarity)
- technical details should carefully be handled with proper attention

LRD Vs Heavy-Tailed

- however, for an intuitive understanding:
  - there is a notion of predictability associated with heavy-tailed RVs
  - as a result, the longer the period of an observed activity, the higher the certainty that the activity persists into the future
  - thus the prediction error can be made arbitrarily zero conditioned on a sufficiently long past observation
- heavy-tailed $\rightarrow$ predictability $\rightarrow$ LRD

Testing for Self-Similarity

Testing for Self-Similarity

- read § 3.3 and § 4.1 of [Leland et al., 1993]: W. E. Leland, M. S Taqqu, W. Willinger, and D. V. Wilson, *On the Self-Similar Nature of Ethernet Traffic*, ACM SIGCOMM Computer Communication Review, October 1993

Appendix

Slowly Varying Function

- Let L be a slowly varying function, e.g.,
  - $L(k) =$ costant
  - $L(k) = \log(k)$
- we have, for all $y > 0$

$$\lim_{k \to \infty} \frac{L(ky)}{L(k)} = 1$$