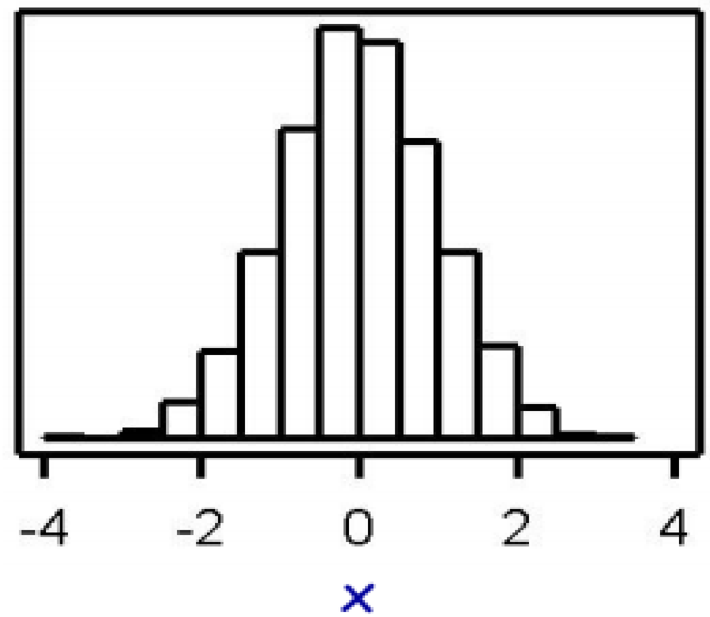
Traffic Modeling(2)

Modeling Traffic as a Stochastic Process

- A good (descriptive) model of network traffic is a stochastic process
- We are generally talking about number of bytes (or packets or flows) per unit time
- A (discrete time) stochastic process is a collection of random variables $\{X_i, i=1, 2, \ldots\}$

Distribution Function

- Given a random variable X, we can fully characterize it by its probability distribution function (pdf):
- i.e. $f(x) = P_X(x)$
- Estimated using a histogram



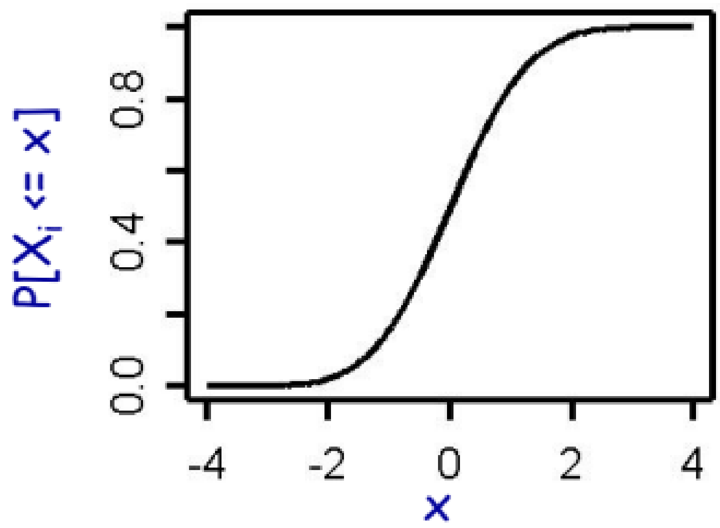
Histograms and CDFs

- A histogram is often a poor estimate of the pdf $\mathbf{f(x)}$ because it involves binning the data
- The CDF $\mathbf{F(x) = P[X_i \leq x]}$ will have a point for each distinct data value; can be much more accurate
- Statistical data binning is a way to group numbers of more or less continuous values into a smaller number of "bins"

Histograms and CDFs

- A histogram is often a poor estimate of the pdf $\mathbf{f}(\mathbf{x})$ because it involves binning the data

- The CDF $\mathbf{F}(\mathbf{x}) = \mathbf{P}[\mathbf{X_i} \leq \mathbf{x}]$ will have a point for each distinct data value; can be much more accurate

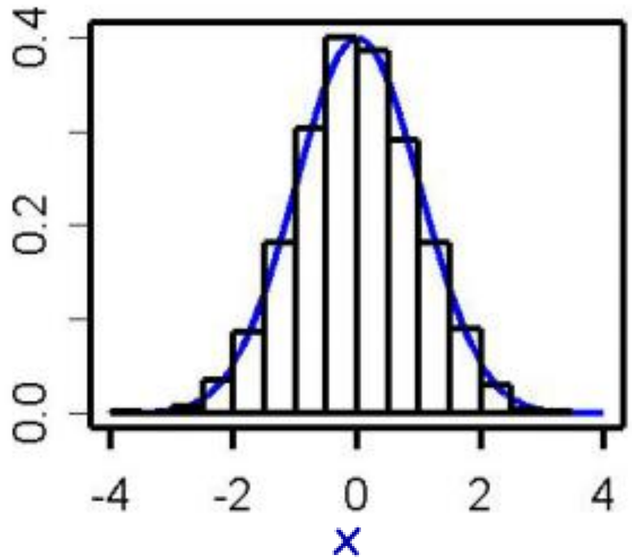


Modeling a Distribution

- We can form a compact summary of a pdf $\mathbf{f}(\mathbf{x})$ if we find that it is well described by a standard distribution – e.g.,
  - Gaussian (Normal)
  - Exponential
  - Poisson
  - Pareto

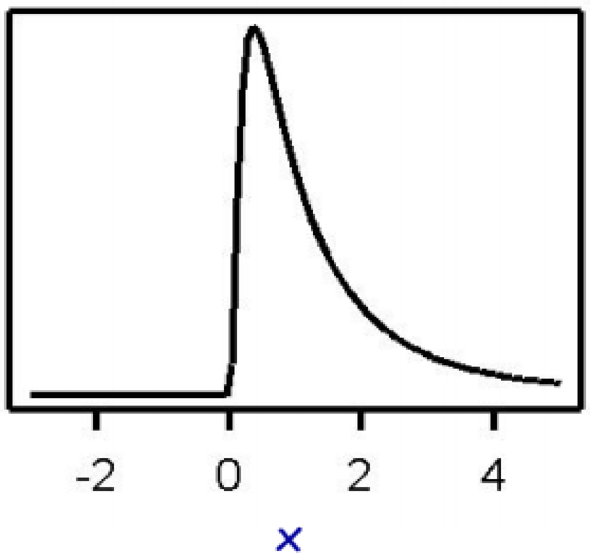
Modeling a Distribution

- Statistical methods exist for asking whether a dataset is well described by a particular distribution
- Estimating the relevant parameters



Distributional Tails

- A particularly important part of a distribution is the (upper) tail
- $P[X > x]$
- Large values dominate statistics and performance
- "Shape" of tail critically important



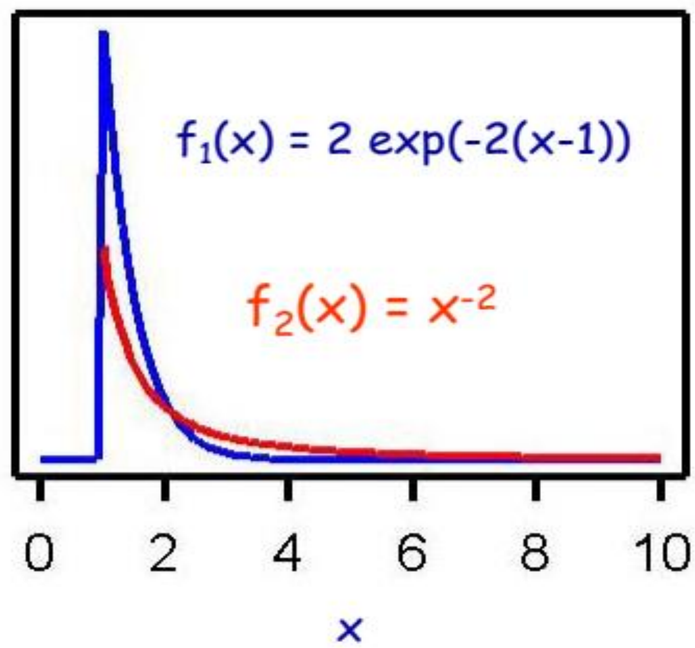
Light Tails, Heavy Tails

- Light tails– Exponential or faster decline

$$f_1(x)$$

- Heavy tails–Slower than any exponential

$$f_2(x)$$



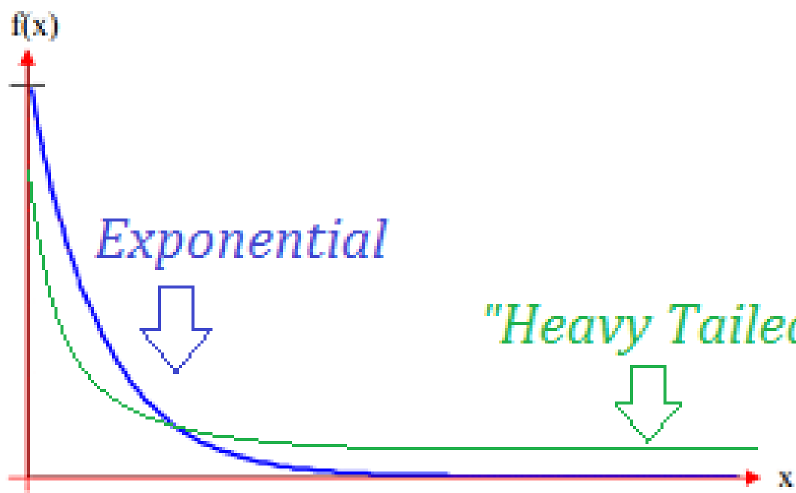
History: Heavy Tails Arrive & Today's traffic

- pre-1985: Scattered measurements note high variability in computer systems workloads
- 1985 – 1992: Detailed measurements note "long" distributional tails
  - File sizes
  - Process lifetimes

History: Heavy Tails Arrive & Today's traffic

- 1993 – 1998: Attention focuses specifically on (approximately) polynomial tail shape: "heavy tails"

- Post-1998: Heavy tails used in standard models



Heavy-tailed

- A distribution is heavy-tailed if the asymptotic shape of the distribution follows a power-law so that

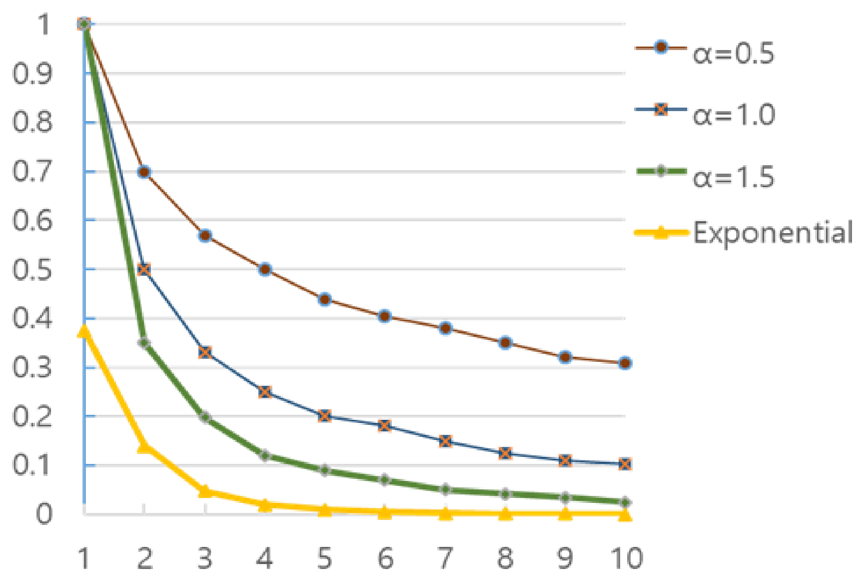
$$\boldsymbol{P[X > x] \cong x^{-\alpha} \text{ as } x \longrightarrow \infty, 0 < \alpha < 2}$$

- The parameter α describes the heaviness of the tail distribution so that as α gets smaller the distribution becomes more heavy-tailed

- Larger portion of the probability mass may be present in the tail of the distribution

The effect of in a heavy-tailed distribution

- The asymptotic (i.e. tail) shape of the distribution is hyperbolic and converges slower than the exponential distribution



The effect of α in a heavy-tailed distribution

A Fundamental Shift in Viewpoint

- Traditional modeling methods have focused on distributions with “light” tails
  - Tails that decline exponentially fast (or faster)
  - Arbitrarily large observations are vanishingly rare

A Fundamental Shift in Viewpoint

- Heavy tailed models behave quite differently
  - Arbitrarily large observations have non-negligible probability
  - Large observations, although rare, can dominate a system's performance characteristics

Use of Heavy-tailed

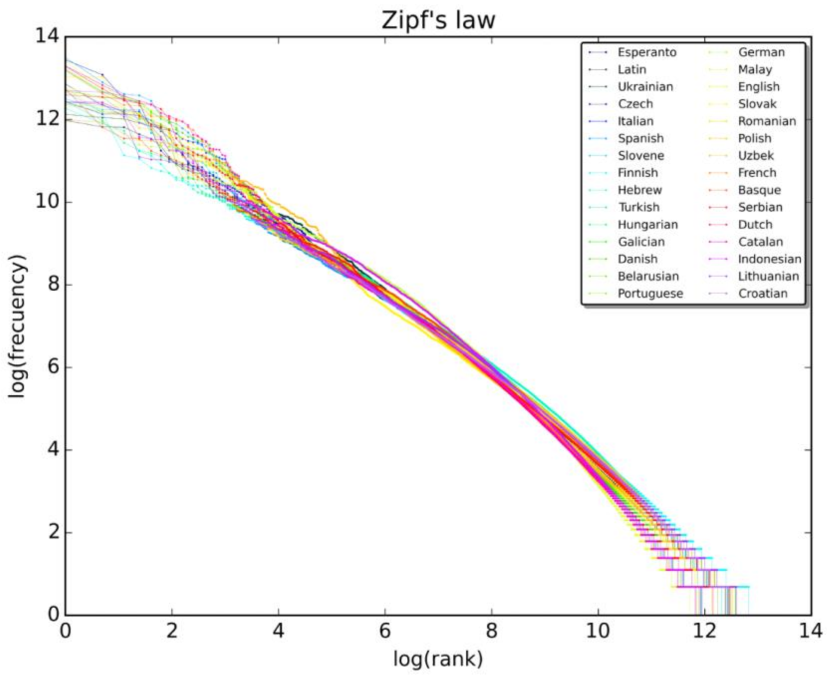
- Sizes of data objects in computer systems
  - Files stored on Web servers
  - Data objects/flow lengths traveling through the Internet
  - Files stored in general-purpose Unix file systems
  - I/O traces of file system, disk, and tape activity

Use of Heavy-tailed

- Process/Job lifetimes
- Node degree in certain graphs
  - Inter-domain and router structure of the Internet
  - Connectivity of WWW pages
- Zipf's Law

Zipf's law

- Zipf's Law is a statistical distribution in certain data sets, such as words in a linguistic corpus, in which the frequencies of certain words are inversely proportional to their ranks.

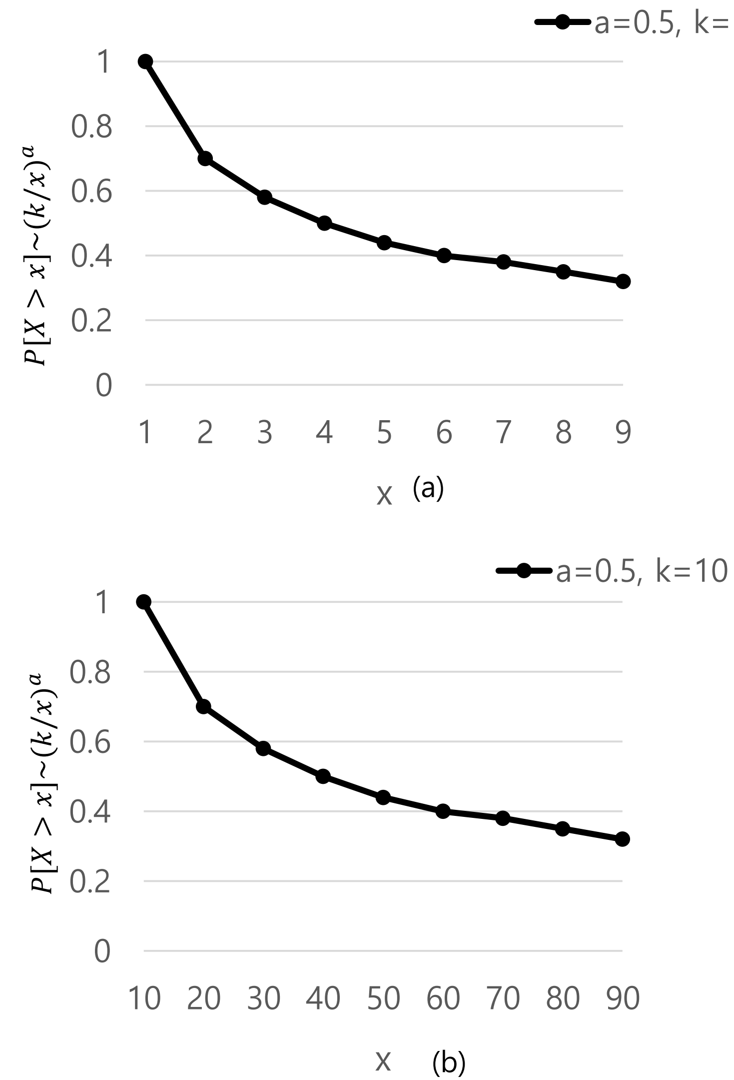


Caution of Heavy-tails

- Workload metrics following heavy tailed distributions are extremely variable
- For example, for heavy tails:
  - When $\alpha \leq 2$, distribution has infinite variance
  - When $\alpha \leq 1$, distribution has infinite mean
- In practice, empirical moments are slow to converge or non-convergent

Pareto distribution

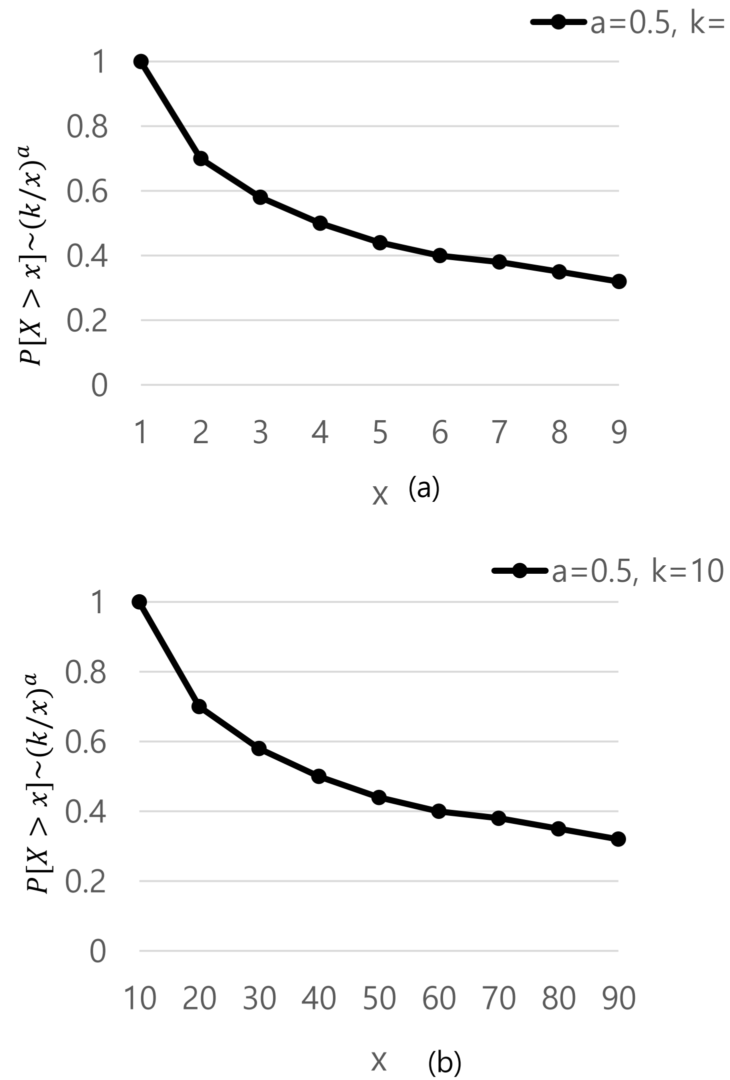
- The Pareto distribution process produces independent and identically distributed(IID) inter-arrival times
- The simplest heavy-tailed distribution
- k is the minimum value of x (simply the scaling factor) and doesn't affect the tail distribution



The effect of $\boldsymbol{k}$ in the Pareto distribution with (a) $\boldsymbol{k}$ =1; and (b) $\boldsymbol{k}$ =10

Pareto distribution

- x is a random variable: a mathematical function that maps outcomes of random experiments to numbers
- α is the heaviness of the tail distribution



The effect of $\boldsymbol{k}$ in the Pareto distribution with (a) $\boldsymbol{k}$ =1; and (b) $\boldsymbol{k}$ =10

Pareto distribution

- The parameters α and k are the shape and location parameters, respectively.
- The Pareto distribution is applied to model self-similar arrival in packet traffic.
- Other important characteristics of the model are that the Pareto distribution has infinite variance, when $\alpha \leq 2$ and achieves infinite mean, when $\alpha \leq 1$.

Pareto distribution

- If X is a random variable with a Pareto distribution, then the probability that X is greater than some number X, i.e. the survival function (also called tail function), is given by

$$\bar{F}(x) = P[X > x] = \begin{cases} (\frac{k}{x})^{\alpha}, & x \geq k \\ 1\,, & x < k \end{cases}$$

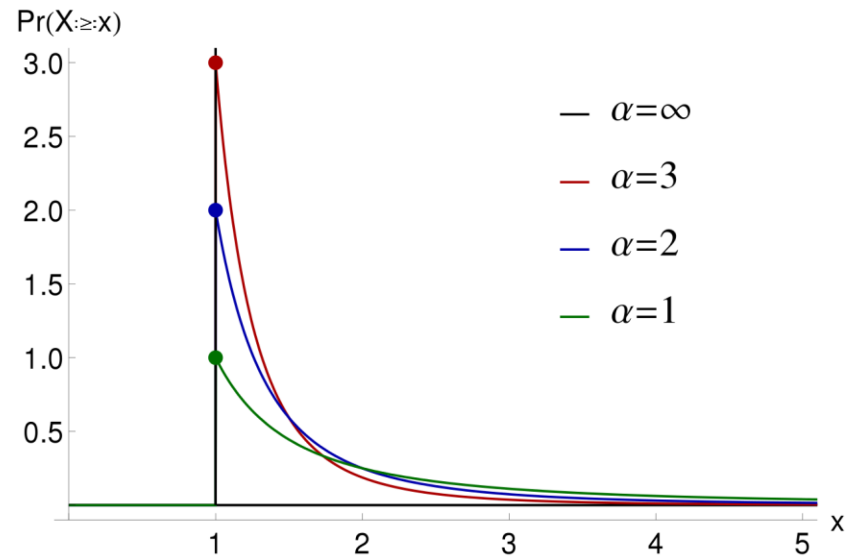
where k is the (necessarily positive) minimum possible value of X, and α is a positive parameter.

Pareto distribution

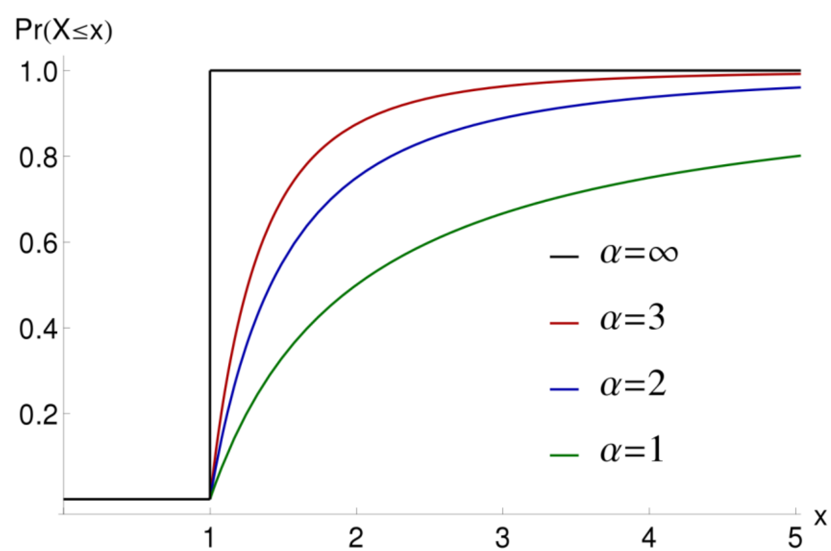
- The Pareto distribution is characterized by a scale parameter k and a shape parameter α, which is known as the tail index.

- CDF of Pareto distribution

$$F_p(x) = 1 - \left(\frac{k}{x}\right)^{\alpha}$$



Pareto probability density functions for various α with $k = 1$.



Pareto cumulative distribution functions for various α with $k = 1$.

Pareto distribution

- $P(X > x) = \left(\frac{k}{x}\right)^{\alpha}$ for all $x \geq k$ where α is a positive parameter and k is the minimum possible value of x

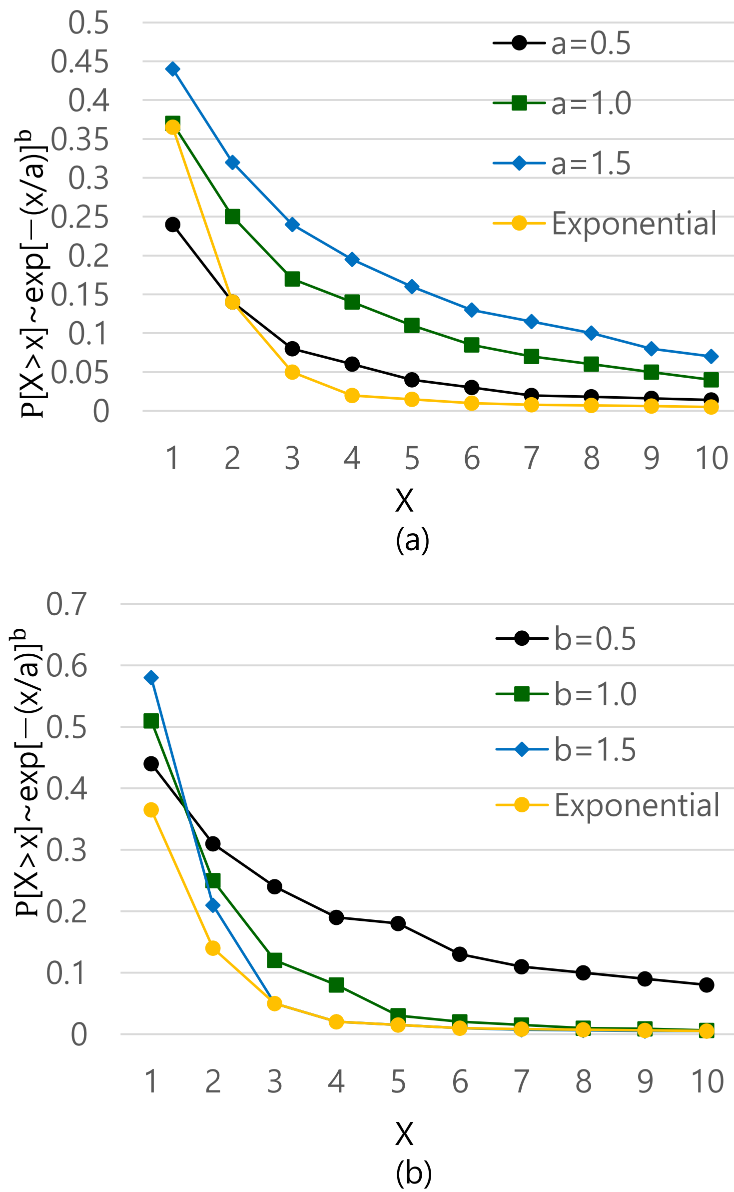
- The probability distribution and the density functions are represented as:

$$F(x) = \int_x^{\infty} f(x)dx = 1 - \left(\frac{k}{x}\right)^{\alpha}$$

where $\alpha,\ k \geq 0,\ x \geq \alpha,\ f(x) = \alpha k^{\alpha} x^{-\alpha-1}$

Weibull distribution

- The Weibull distributed process is heavy-tailed and can model the fixed rate in ON period and ON/OFF period lengths, when producing self-similar traffic by multiplexing ON/OFF sources.

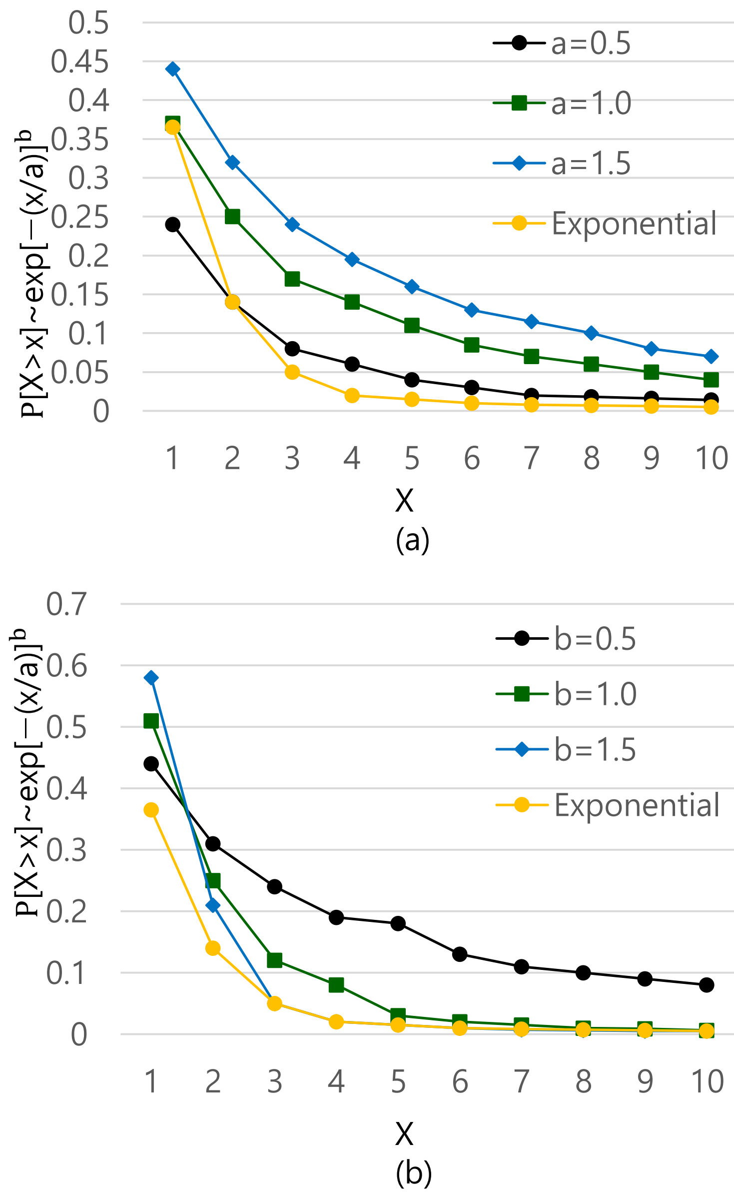


The effect of (a) a; and (b) b in Weibull distribution

Weibull distribution

- Both parameters a and b affect the tail distribution
- More sensitive to the value of b
- CDF of Weibull distribution

$$F_w(x) = 1 - e^{-(x/a)^b}$$



The effect of (a) a; and (b) b in Weibull distribution

Weibull distribution

- The distribution function in this case is given by:

$$F_w(x) = 1 - e^{-\left(\frac{x}{a}\right)^b}, \qquad x \geq 0$$

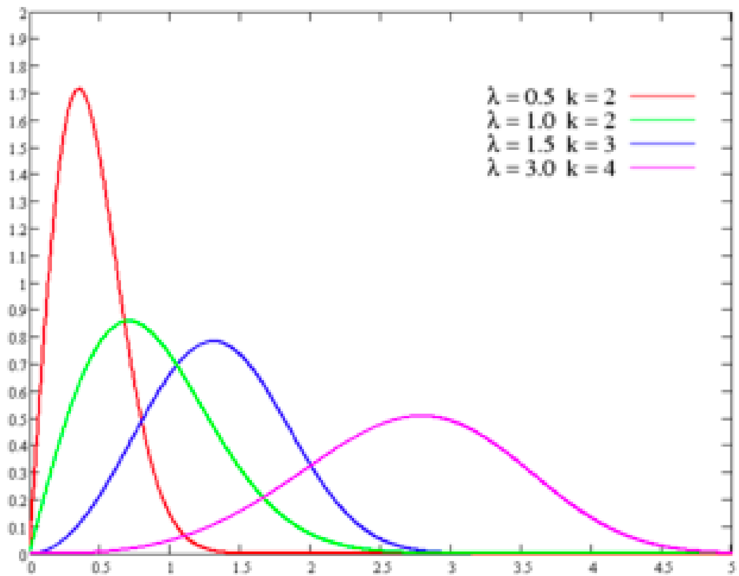
and the density function of the Weibull distribution is given as:

$$f(t) = ba^{-b}x^{b-1}e^{-\left(\frac{x}{a}\right)^b}, \qquad x \geq 0$$

where parameters $a > 0$ and $b > 0$ are the scale and location parameters respectively.

Weibull distribution

- The Weibull distribution is close to a normal distribution.
- For $a \leq 1$ the density function of the distribution is L shaped and for values of $a > 1$, it is bell shaped.



Meaning of heavy-tailed distribution

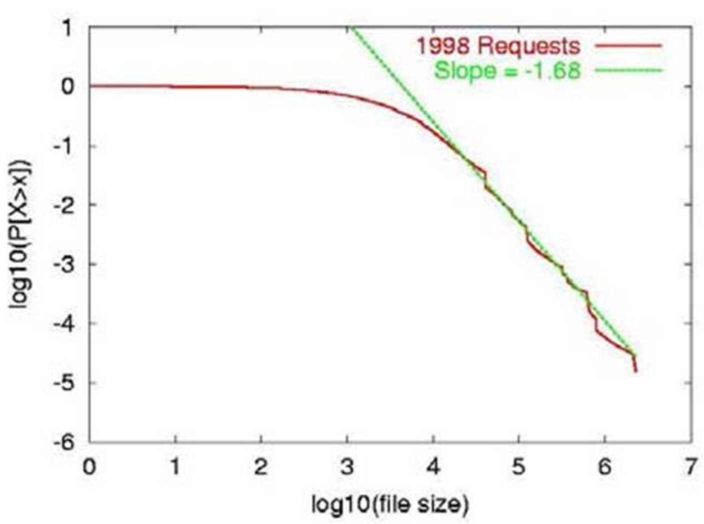
- Usually, a heavy-tailed distribution describes traffic processes such as packet inter-arrival times and burst length
- Heavy tailed distributions tend to have many outliers with very high values. It means that the arrival rate is higher than the service rate.

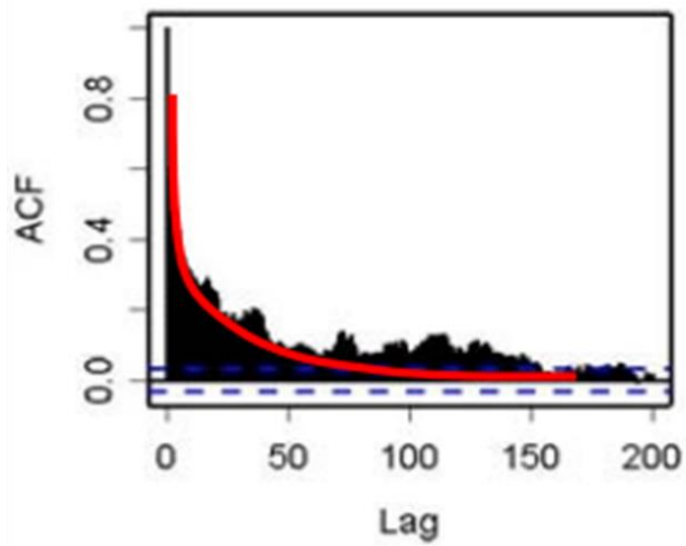
Characterizing a traffic process

- Marginals and Autocorrelation
  - Characterizing a traffic process in terms of these two properties gives you a good approximate understanding of the process, without involving a lot of work, requiring complicated models, or requiring estimation of too many parameters.
- Recent analysis on traffic measurements on packet-data networks such as LAN and WAN, show heavy-tailed, self-similar, fractal, and LRD characteristics.

How Does Self-Similarity Arise?

- Flows → Autocorrelation → Self-similarity
- Distribution of flow lengths has power law tail
  →Autocorrelation declines like a power law

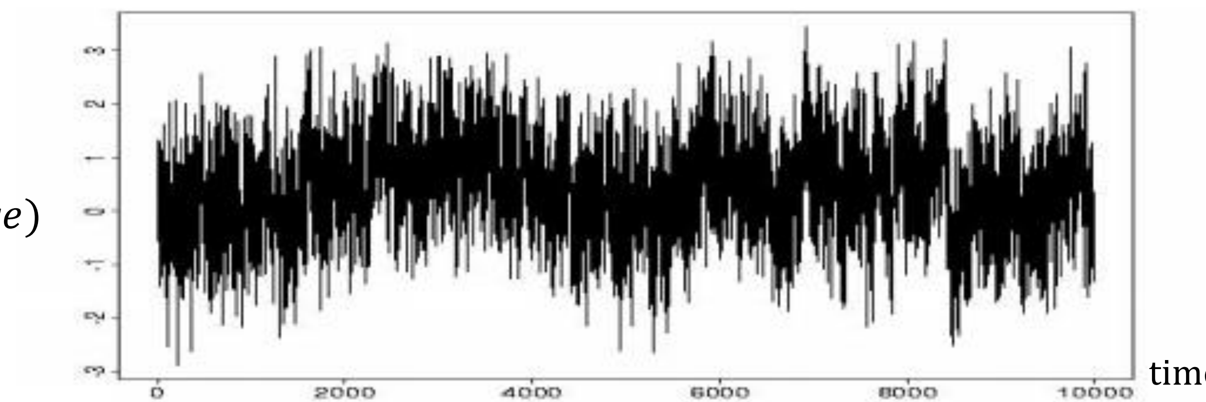




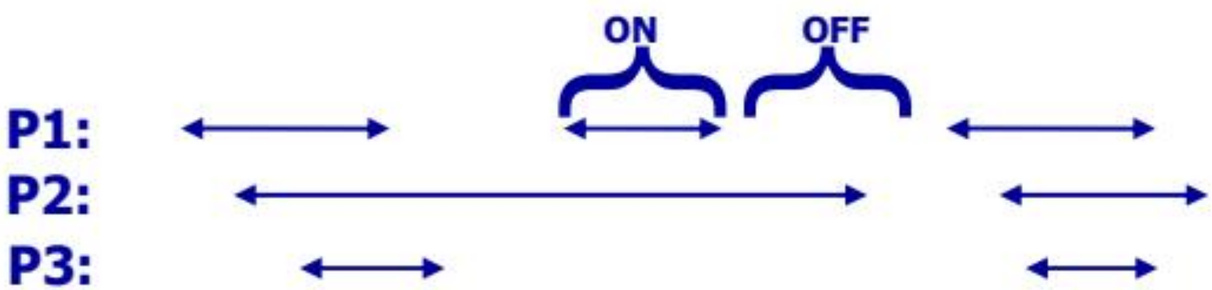
Self-Similarity

- Power Tailed ON/OFF sources →Self-Similarity

$\log(Variance)$



time(1 unit = 500ms)



Self-similarity indicator

- If the aggregate traffic exhibits time correlation over a wide rage of timescales can be characterized by a single parameter called **Hurst parameter (H)**
- Hurst parameter
  - Measure of the degree of self-similarity of the aggregate traffic stream
  - **If H gets closer to 1, the degree of self-similarity increases**

Self-similarity indicator

- Three methods that can measure Hurst parameter
  - Variance vs Time
  - R/S plot
  - Whittle Estimator
- Exactly self-similar ($H = 1$)
- Asymptotically self-similar ($0.5 < H < 1$)

Evidence of Self-similarity

- A recent measurement study has shown that aggregate Ethernet LAN traffic is self-similar
- A statistical property that is very different from the traditional Poisson-based models
- In 1993, a group at Bellcore recorded a large series of highly detailed Ethernet data. By chance, a mathematician specializing in self-similarity was available, and a complete analysis demonstrated the phenomenon beyond any reasonable doubt

Evidence of Self-similarity

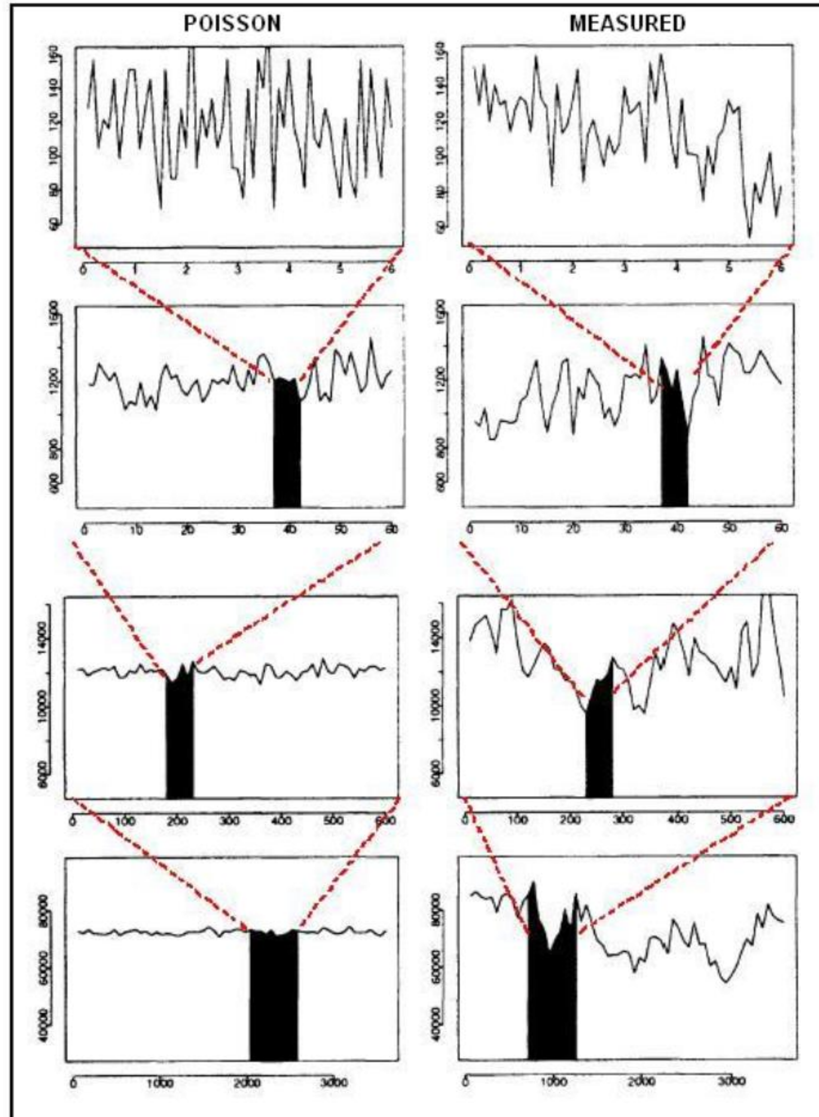
- The proof is best illustrated graphically. The original study provided the best available graphical demonstration of the problem

- Self-Similarity refers to distributions that exhibit the same characteristics at all scales.

- This is clearly not the case for Poisson traffic.

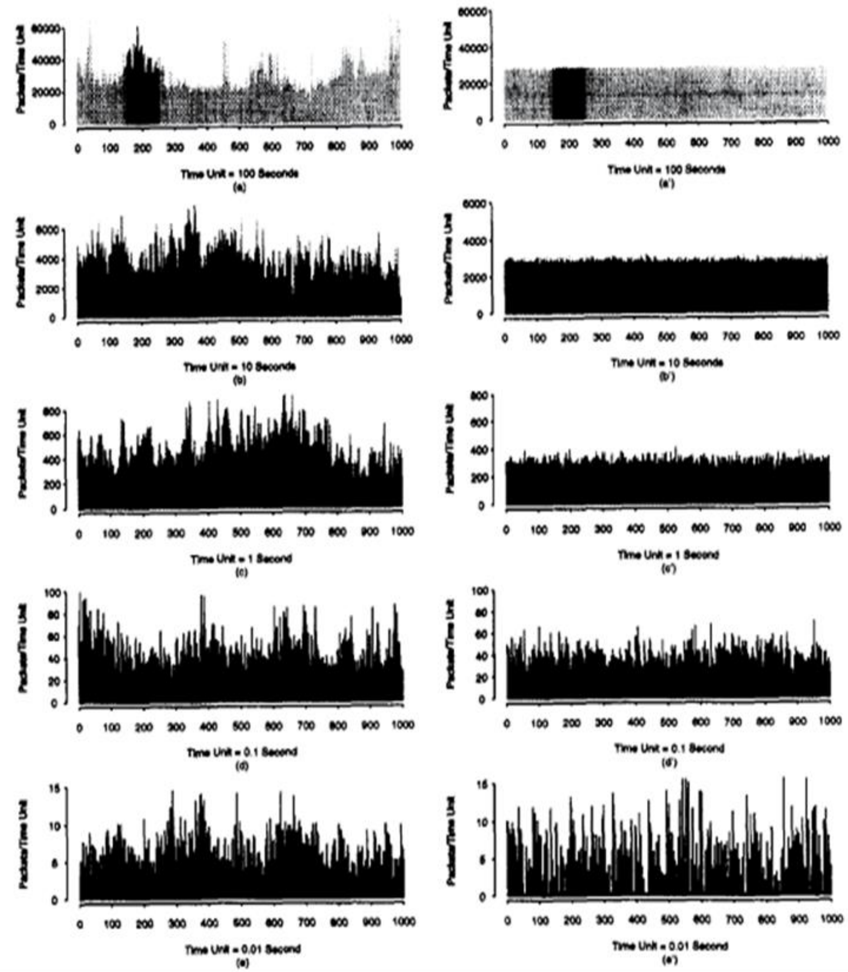
Evidence of Self-similarity

- As bin sizes increase, Poisson traffic will “smooth,” eventually reaching a flat line at the distribution mean.
- Truly self-similar traffic will not; it will continue to show bursts at all scales.



Evidence of Self-similarity

- On the left, we have a real network trace appearing at different time scales.
- On the right, we have a pure Poisson process generating synthetic traffic at the same time scales.



Graphical demonstration of Self Similarity vs. Poisson model

Evidence of Self-similarity

- The packet counts are renormalized to an appropriate scale as the time scale changes. The difference is clearest at the largest time scales.
- Both Poisson processes and self-similar processes are bursty at the correct time scales. However, unlike Poisson processes, self-similar process bursts have no natural length.
- Bursts are evident from the 10ms scale all the way to the 100 seconds scale.

Meaning of Self-similarity

- 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

- No matter what time scale you use to examine the data, you see similar patterns

  E.g., 10 msec, 100 msec, 1 sec, 10 sec,...

Meaning of Self-similarity

i) Burstiness exists across many time scales

ii) No natural length of a burst

iii) Traffic does not necessarily get "smoother" when you aggregate it (unlike Poisson traffic)

Several equivalent fashions of Self-similarity

- Slowly decaying variance
- Long range dependence
- Non-degenerate autocorrelations
- Hurst effect

Slowly decaying variance: Variance-Time Plot

- 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 very slowly, even when the sample size grows quite large

Variance-Time Plot

- Plots the variance of the sample versus the sample size that is changed to the log value m, on a log-log plot:

$$Var\left(X^{(m)}\right) = \sigma^2 m^{-\beta}$$

$$\log Var\left(X^{(m)}\right) = \log \sigma^2 m^{-\beta} = -\beta \log m + \log \sigma^2$$

So, $-\beta$ is slope.

- The "variance-time plot" is a well known technique for testing the behavior of the variance with respect to the time scale.

Variance-Time Plot

- For most processes, the result is a straight line with slope -1
- For self-similar, the line is much flatter

$$H = 1 - \frac{\beta}{2}$$