Introduction to Heavy-Tailed Distributions, Self-Similar Processes, and Long-Range Dependence

Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@cse.wustl.edu

These slides are available on-line at:
http://www.cse.wustl.edu/~jain/cse567-13/
Overview

1. Heavy-Tailed Distributions (HTDs)
2. How to Check for Heavy Tail?
3. Self-Similar Processes
4. Long Range Dependence (LRD)
5. Generating LRD Sequences
6. Self-Similarity vs. LRD
7. Hurst Exponent Estimation
Heavy-Tailed Distributions (HTDs)

- CCDF decays slower than the exponential distribution
  \[ P(X > x) = 1 - F(x) = \bar{F}(x) = e^{-\lambda x} \]
- CCDF = Complementary cumulative distribution function
  \[ \bar{F}(x) = 1 - F(x) \]
- For heavy-tailed distributions, CCDF is slower by some power of \( x \)
  \[ \bar{F}(x) \rightarrow c x^n e^{-\lambda x} \]
- Very large values possible

Washington University in St. Louis  
http://www.cse.wustl.edu/~jain/cse567-13/k_38lrd.htm
©2013 Raj Jain
Examples of HTD Variables

- Many real-world phenomenon have been found to follow heavy tailed distributions.
  - Distribution of wealth.
    One percent of the population owns 40% of wealth.
  - File sizes in computer systems
  - Connection durations
  - CPU times of jobs
  - Web pages sizes

- Significant impact on buffer sizing in switches and routers.
Example 38.1

- Weibull distribution

\[ F(x) = 1 - e^{-\lambda x^\alpha} \]

\[ 1 - F(x) = e^{-\lambda x^\alpha} > e^{-\lambda x} \quad \lambda > 0, 1 > \alpha > 0 \]

- Other examples of heavy tailed distributions are Cauchy, log-normal, and t-distributions.
Power Tailed Distributions

- A subset of heavy tailed distributions
  \[ \bar{F}(x) \Rightarrow \frac{c}{x^\alpha} \]

- CCDF approaches a power function for large \( x \)
  - For such distributions:
    - all moments \( E[x^l] \) for all values of \( l \geq \alpha \) are infinite.
    - If \( \alpha \leq 2 \), \( x \) has infinite variance
    - If \( \alpha \leq 1 \), the variable has infinite mean
    - \( \alpha \) is called the tail index.
Example 38.2

- Pareto distribution:
  \[ F(x) = 1 - x^{-\alpha} \quad 1 \leq x \leq \infty, \alpha > 0 \]
  \[ \bar{F}(x) = 1 - F(X) = \frac{1}{x^\alpha} \]
  \[ f(x) = \frac{d}{dx} F(x) = \alpha x^{-\alpha-1} \]

- pdf: \[ f(x) = \frac{d}{dx} F(x) = \alpha x^{-\alpha-1} \]

- Its \( l^{th} \) moment is:
  For \( l \neq \alpha \):
  \[ E[x^l] = \int_1^\infty x^l f(x) dx = \int_1^\infty \alpha x^{l-\alpha-1} dx = \frac{\alpha}{l-\alpha} x^{l-\alpha}\bigg|_1^\infty = \left\{ \begin{array}{ll} \infty & l > \alpha \\ \frac{\alpha}{\alpha-l} & l < \alpha \end{array} \right. \]
  For \( l = \alpha \):
  \[ E[x^l] = \int_1^\infty \alpha x^{-1} dx = \alpha \ln x\bigg|_1^\infty = \infty \]

- All moments for \( l \geq \alpha \) are infinite.
- For \( 2 \geq \alpha > 1 \) variance and higher moments are infinite.
  For \( 1 \geq \alpha \) variance does not exist.
- For \( 1 \geq \alpha > 0 \), even mean is infinite.
Effect of Heavy Tail

- A random variable with HTD can have very large values with finite probabilities resulting in many outliers.
- Sampling from such distributions results in mostly small values with a few very large valued samples.
- Sample statistics (e.g., sample mean) may have a large variance \( \Rightarrow \) sample sizes required for a meaningful confidence are large.
- Sample mean generally under-estimates the population mean.
- Simulations with heavy-tailed input require very long time to reach steady state and even then the variance can be large.

\[
| \bar{x}_n - \mu | \approx c n^{\frac{1}{\alpha} - 1}
\]

\( c \) is some constant.
Effect of Heavy Tail (Cont)

- The number of observations required to reach k-digit accuracy:
  \[
  \frac{|\bar{x}_n - \mu|}{\mu} \leq 10^{-k} \quad \frac{cn^{\frac{1}{\alpha} - 1}}{\mu} \leq 10^{-k} \quad n \geq 10 \left( \log(c/\mu) + k \right)^{1 - \frac{1}{\alpha}}
  \]

- Assuming c=1, \( \mu=1 \), \( 10^{11} \) observations are required for a single decimal digit accuracy \((k=1)\) if \( \alpha=1.1 \).

- Central limit theorem applies **only** to observations from distributions with finite variances.
  - For heavy-tailed distributions with infinite variance, the central limit theorem does **not** apply.
  - The sample mean does not have a normal distribution even after a large number of samples.
  - Confidence interval formulas mentioned earlier can not be used.
Effect of Heavy Tail (Cont)

- **M/PT/1 queue**: Poisson arrivals and power-tailed service time
  - pdf of queue length $f(n) \rightarrow c(\rho)/n^\alpha$
    where $c(\rho)$ is a function of the traffic intensity $\rho$.
  - If $\alpha \leq 1$, the mean service time is infinite and so are the traffic intensity and the mean queue length.
  - If $\alpha \leq 2$, the service time has infinite variance, and so does the queue length.

- **PT/M/1 queue**:
  - Tail index $\alpha \leq 1$, the mean inter-arrival time is infinite.
  - For $1 < \alpha \leq 2$, the variance of the inter-arrival time is infinite.

- Heavy tailed-ness also implies **predictability**:
  - If a heavy tailed task has run a long time, it is expected to run for an additional long time.
  - $\lim_{x \to \infty} E[X - x|X > x] = \infty$
How to Check for Heavy Tail?

- Make a Q-Q plot on a log-log graph assuming a Pareto distribution

\[ F(x) = 1 - x^{-\alpha} \]

\[ x = (1-F)^{-1/\alpha} \]

- On a log-log graph: \( \ln x = (-1/\alpha) \ln (1-F) \)

- Find \( \alpha \) from the slope of the best-fit line. \( \alpha \geq 1 \Rightarrow \text{Heavy Tailed} \)
Example 38.3

Check if this set of 50 observations has a heavy tail: 2.426, 1.953, 1.418, 1.080, 3.735, 2.307, 1.876, 1.110, 3.131, 1.134, 1.171, 1.141, 2.181, 1.007, 1.076, 1.131, 1.156, 2.264, 2.535, 1.001, 1.099, 1.149, 1.225, 1.099, 1.279, 1.052, 1.051, 9.421, 1.346, 1.532, 1.000, 1.106, 1.126, 1.293, 1.130, 1.043, 1.254, 1.118, 1.027, 1.383, 1.288, 1.988, 1.561, 1.106, 1.256, 1.187, 1.084, 1.968, 1.045, 1.155

\[ \alpha = \frac{1}{\text{slope}} = \frac{1}{0.427} = 2.34 \]

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>Rank</th>
<th>( R_i )</th>
<th>( Q_i = (R_i - 0.5)/n )</th>
<th>( \text{Ln}(x_i) - \ln(1 - Q_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.426</td>
<td>46</td>
<td>0.910</td>
<td>0.886</td>
<td>2.408</td>
</tr>
<tr>
<td>1.953</td>
<td>40</td>
<td>0.790</td>
<td>0.669</td>
<td>1.561</td>
</tr>
<tr>
<td>1.418</td>
<td>36</td>
<td>0.710</td>
<td>0.349</td>
<td>1.238</td>
</tr>
<tr>
<td>1.080</td>
<td>10</td>
<td>0.190</td>
<td>0.077</td>
<td>0.211</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1.084</td>
<td>11</td>
<td>0.210</td>
<td>0.081</td>
<td>0.236</td>
</tr>
<tr>
<td>1.968</td>
<td>41</td>
<td>0.810</td>
<td>0.677</td>
<td>1.661</td>
</tr>
<tr>
<td>1.045</td>
<td>6</td>
<td>0.110</td>
<td>0.044</td>
<td>0.117</td>
</tr>
<tr>
<td>1.155</td>
<td>24</td>
<td>0.470</td>
<td>0.144</td>
<td>0.635</td>
</tr>
<tr>
<td>80.204</td>
<td>17.308</td>
<td>49.654</td>
<td>Sum</td>
<td></td>
</tr>
<tr>
<td>207.810</td>
<td>14.781</td>
<td>95.774</td>
<td>Sum of Sq</td>
<td></td>
</tr>
<tr>
<td>1.604</td>
<td>0.346</td>
<td>0.993</td>
<td>Average</td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{Sum} = 17.308, \quad \text{Sum of Sq} = 14.781, \quad \text{Average} = 0.346, 0.993
\]
Self-Similarity

- When zoomed, the sub objects have the same shape as the original object
- Also called Fractals
- Latin “fractus” = “fractional” or “broken”
  ⇒ Traditional Euclidean geometry can not be used to analyze these objects because their perimeter is infinite.
Self-Similar Processes

- Scaling in time = scaling in magnitude
  \[ x_{at} \sim a^H x_t \quad \forall a > 0 \]

- Statistical similarity \( \Rightarrow \) Similar distributions with similar mean and variance
  \[ E[x_{at}] = a^H E[x_t] \]
  \[ \text{Var}[x_{at}] = a^{2H} \text{Var}[x_t] \]
  - Similar variance \( \Rightarrow \) Self-similar in the second order
  - Similar higher order moments \( \Rightarrow \) Self-similarity of higher orders
  - All moments similar \( \Rightarrow \) strictly self-similar.
Example 38.4

- Consider the white noise process \( e_t \) with zero mean and unit variance: \( e_t = z \sim N(0, 1) \)
  Here \( z \) is the unit normal variate.

- Consider the process \( x_t: x_t = t^H e_t = t^H z \)

- For this process:
  \[ x_{at} = (at)^H z = a^H t^H z = a^H x_t \]
  - Therefore, \( x_t \) is a self-similar process.
  - \( H = \) Hurst exponent
Short Range Dependence (SRD)

- Sum of Autocorrelation function is finite.
- Example: AR(1) with zero mean:
  \[ x_t = a_1 x_{t-1} + e_t \]
  For this process, the autocorrelation function decreases exponentially:
  \[ \text{Cor} [x_t, x_{t-k}] = r_k = a_1^{|k|} \]
- Sum of autocorrelations is finite (provided \( |a_1| < 1 \)):
  \[ \sum_{k=0}^{\infty} r_k = \frac{1}{1 - a_1} \]
Long Range Dependence (LRD)

- Sum of Autocorrelation function is infinite \( \sum r_k = \infty \)

Alternative Definition:

- Limiting tail behavior of the autocorrelation:
  \[ r(k) \to k^{2H-2} L(k) \quad k \to \infty \]
  
  Here, \( L(x) \) is a *slowly varying function* of \( x \).

\( L(ax)/L(x) \) tends to 1 as \( x \) approaches infinity.

Constants and logarithms are examples of slowly varying functions.
Examples of Processes with LRD

- Aggregation of a large number of on-off processes with heavy-tailed on-times or heavy-tailed off times results in long-range dependence.
- File sizes have a long-tailed distribution $\Rightarrow$ Internet traffic has a long range dependence.
- Connection durations have also been found to have a heavy-tailed distribution $\Rightarrow$ traffic has a long range dependence.
- UNIX processes have been found to have a heavy-tailed distribution $\Rightarrow$ resource demands have LRD.
- Congestion and feedback control mechanisms such as those used in Transmission Control Protocol (TCP) increase the range of dependence in the traffic.
Effect of Long Range Dependence

- Long-range dependence invalidates all results for queueing theory obtained using Poisson processes, e.g., Buffer sizes required to avoid overflow may be off by thousands times.
Self-Similarity vs. LRD

- Self-similarity ≠ Long-range dependence
- Self-similar process can be short-range dependent or long-range dependent
- Self-similar processes with $\frac{1}{2} < H < 1$ have long range dependence.
- Self-similar processes with $0 < H \leq \frac{1}{2}$ have short range dependence.
- ARIMA$(p, d, q)$ with integer valued $d$ are SRD.
- FARIMA$(p, d+\delta, q)$ with $-\frac{1}{2} < \delta < \frac{1}{2}$ and $\delta \neq 0$ have long-range dependence.
FARIMA Models and LRD

- Fractional Auto-regressive Integrated Moving Average (FARIMA) processes exhibit LRD for certain values of \( d \).
- Consider FARIMA(0, 0.25, 0): \((1 - B)^{0.25} x_t = e_t\)

\[
x_t = (1 - B)^{-0.25} e_t
\]

\[
= e_t - (-0.25)e_{t-1} + \frac{-0.25(-0.25 - 1)}{(1)(2)} e_{t-2} + \frac{-0.25(-0.25 - 1)(-0.25 - 2)}{(1)(2)(3)} e_{t-3} + \cdots
\]

\[
= e_t + 0.25e_{t-1} + \frac{0.25(0.25 + 1)}{(1)(2)} e_{t-2} + \frac{0.25(0.25 + 1)(0.25 + 2)}{(1)(2)(3)} e_{t-3} + \cdots
\]

\[
= e_t + 0.25e_{t-1} + 0.16e_{t-2} + 0.12e_{t-3} + 0.10e_{t-4} + \cdots + 0.04e_{t-10}
\]

The coefficient of \( e_{t-k} \) is

\[
\frac{0.25(0.25 + 1) \cdots (0.25 + k - 1)}{(1)(2) \cdots (k)} = \frac{\Gamma(0.25 + k)}{\Gamma(0.25)\Gamma(k + 1)}
\]

Here, \( \Gamma() \) is the Gamma function: \( \Gamma(p+1) = p\Gamma(p) \).

It is a generalization of factorial. For integer \( p \), \( \Gamma(p + 1) = p! \)

For example, \( \Gamma(3) = 2 \), \( \Gamma(2) = 1 \), \( \Gamma(1) = 1 \), \( \Gamma(0) = \infty \)
Consider FARIMA(0, $\delta$, 0) with $-1/2 < \delta < 1/2$ and $\delta \neq 0$.

$$x_t = (1 - B)^{-\delta}e_t = \sum_{k=0}^{\infty} b_k e_{t-k}$$

Where:

$$b_k = \frac{\Gamma(k + \delta)}{\Gamma(\delta) \Gamma(k + 1)}$$

Since $e_t$ is Gaussian, $x_t$ is also Gaussian.

$e_t$ is Gaussian Noise, $x_t$ is fractional Gaussian Noise (fGn)
FARIMA and LRD (Cont)

- The autocovariance of the FARIMA(0, δ, 0) sequence is:
  \[
  E[x_t x_{t-k}] = \begin{cases} 
  \sigma^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)} & k = 0 \\
  \sigma^2 \frac{\Gamma(k+\delta)\Gamma(1-2\delta)}{\Gamma(k-\delta+1)\Gamma(\delta)\Gamma(1-\delta)} & k \neq 0 
  \end{cases}
  \]

- Autocorrelation at lag k: \( r_k = \frac{\Gamma(k + \delta)\Gamma(1 - \delta)}{\Gamma(k - \delta + 1)\Gamma(\delta)} \)

- Stirling’s approximation: \( \Gamma(p + 1) \approx \sqrt{2\pi p} \left(\frac{p}{e}\right)^p \)

- For large \( k \), \( r_k \) tends to \( ck^{2\delta-1} \) where \( c = \frac{\Gamma(1 - \delta)}{\Gamma(\delta)} \)

- Recall that for LRD: \( r(k) \rightarrow k^{2H-2}L(k) \) as \( k \rightarrow \infty \)
  \[ 2H-2=2\delta-1, \text{ that is, } H=\delta+1/2. \]
  \[ \Rightarrow \text{A FARIMA}(0,\delta,0) \text{ sequence has LRD if } 0<\delta<1/2. \]
Generating LRD Sequences

- Generate the FARIMA(p, d+δ, q) LRD sequence
- FARIMA(p, d+δ, q) = ARIMA(p, d, q) with $e_t$ replaced by fractional Gaussian noise generated by FARIMA(0, δ, 0)
- ARIMA(p,d,q) is given by
  \[
  \phi(B)(1 - B)^d x_t = a_0 + \psi(B) e_t
  \]
- It can be generated by one of the following two methods:
  1. Using previous values of $x_t$:
     \[
     x_t = a_0 + (1 - \phi(B)(1 - B)^d) x_t + \psi(B) e_t
     \]
Generating LRD Sequences (Cont)

2. Converting the model to a moving average model using a Taylor series expansion:

\[ x_t = \frac{\psi(B)}{\phi(B)(1 - B)^d} e_t = \left( \sum_{i=0}^{m} c_i B^i \right) e_t = \sum_{i=0}^{m} c_i e_{t-i} \]

Here \( c_i \) are coefficients of the Taylor series expansion and \( m \) is selected large enough so that \( c_i \) for \( i > m \) are negligible.

- Generate a white noise sequence \( e_i \sim N(0, 1) \)
- Generate a FARIMA\((0, \delta, 0)\) sequence \( y_i \) using a moving average of a large number \( m \) of \( e_i \):

\[ y_i = \sum_{k=0}^{m} \frac{\Gamma(k + \delta)}{\Gamma(k + 1) \Gamma(\delta)} e_{i-k} \]
Generating LRD Sequences (Cont)

- Generate a FARIMA\( (p, d+\delta, q) \) sequence \( x_i \) by generating a usual ARIMA\((p, d, q)\) as in Step 1 above with the white noise \( e_i \) replaced by \( y_i \)

\[
x_t = a_0 + \frac{\psi(B)}{\phi(B)(1 - B)^d} y(i)
\]

- \( m=100 \) or \( m=1000 \) has been found to provide good results.
Example 38.5

- Generate a FARIMA(0,0.25,0) Sequence
  \[ x_t = (1 - B)^{-0.25} e_t \]
  \[ = e_t + 0.25e_{t-1} + \frac{0.25(0.25 + 1)}{2} e_{t-2} + \frac{0.25(0.25 + 1)(0.25 + 2)}{6} e_{t-3} + \cdots \]
  \[ = e_t + 0.25e_{t-1} + 0.16e_{t-2} + 0.12e_{t-3} + 0.10e_{t-4} + \cdots + 0.04e_{t-10} \]

- Generate 60 N(0,1) random numbers for \( e_9 \) thru \( e_{50} \)
- Use the above equation to get \( x_1 \) through \( x_{50} \)
- The numbers are: -2.16, -0.77, -2.23, 0.61, -0.93, -1.85, -1.68, -1.15, -0.96, -0.27, 0.33, -0.18, 0.32, -0.23, -1.90, 0.54, 0.27, -0.02, 0.04, 1.58, 0.23, -0.43, -0.48, 0.19, 1.16, 0.60, 1.61, -0.23, 1.08, 0.22, 1.63, -0.48, 0.51, -0.51, 0.74, -1.53, -0.63, -0.10, 0.01, -0.19, -1.35, -1.19, -1.36, -0.29, 0.87, 1.70, 0.97, 1.70, 1.18, 0.20
Hurst Exponent Estimation

Variance-time plot (Similar to the method of independent runs)

1. Start with \( m=1 \)

2. Divide the sample of size \( n \) into non-overlapping subsequences of length \( m \). There will be \( j = \left\lfloor n/m \right\rfloor \) such subsequences.

3. Take the sample mean of each subset

\[
\bar{x}_{km} = \frac{1}{m} \sum_{i=(k-1)m+1}^{i=km} x_i \quad k = 1, 2, 3, \ldots, j
\]

4. Compute the overall mean:

\[
\bar{x}_m = \frac{1}{j} \sum_{k=1}^{k=j} \bar{x}_{km}
\]
5. Compute the variance of the sample means

\[ s_m^2 = \frac{1}{j-1} \sum_{k=1}^{j} (\bar{x}_{k,m} - \bar{x}_m)^2 \]

6. Repeat steps 2 through 5 for m=1, 2, 3, …

7. Plot variance \( s_m^2 \) as a function of the subsequence size \( m \) on a log-log graph

8. Fit a simple linear regression to log(var) vs. log \( (m) \).

9. The slope of the regression line is 2H-2.

10. That is, the Hurst exponent is \( 1 + a_1/2 \), where \( a_1 \) is the slope of the regression line.

Note: 1. H estimate using this variance time plot method is biased
2. If a process is non-stationary, it may not be self-similar or have LRD, but may result in Hurst exponent between 0.5 and 1
Example 38.6

Determine the Hurst exponent for the data of Example 38.5

<table>
<thead>
<tr>
<th>m</th>
<th>1.00</th>
<th>2.00</th>
<th>3.00</th>
<th>4.00</th>
<th>5.00</th>
<th>6.00</th>
<th>7.00</th>
<th>8.00</th>
<th>9.00</th>
<th>10.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.21</td>
<td>-0.19</td>
<td>-0.18</td>
<td>-0.13</td>
<td>-0.11</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>Variance</td>
<td>0.99</td>
<td>0.52</td>
<td>0.62</td>
<td>0.44</td>
<td>0.34</td>
<td>0.36</td>
<td>0.35</td>
<td>0.16</td>
<td>0.36</td>
<td>0.11</td>
</tr>
<tr>
<td>ln(m)</td>
<td>0.00</td>
<td>0.69</td>
<td>1.10</td>
<td>1.39</td>
<td>1.61</td>
<td>1.79</td>
<td>1.95</td>
<td>2.08</td>
<td>2.20</td>
<td>2.30</td>
</tr>
<tr>
<td>ln(Variance)</td>
<td>-0.01</td>
<td>-0.65</td>
<td>-0.48</td>
<td>-0.82</td>
<td>-1.06</td>
<td>-1.03</td>
<td>-1.04</td>
<td>-1.82</td>
<td>-1.03</td>
<td>-2.25</td>
</tr>
</tbody>
</table>

Slope = -0.73

\[ H = 1 - \frac{0.73}{2} = 0.635 \]

ARIMA(0,0.25,0) \implies H=0.75
Summary

1. Heavy tailed distributions: CCDF tail higher than exponential distribution
2. Self-Similar Process: $x_{at} \sim a^H x_t$
3. Long Range Dependence: $\sum r_k = \infty$
4. ARIMA(p,d+\delta,q) with $0<\delta<0.5$ can be used to generate LRD sequences
5. Hurst parameter can be estimated with variance-time plots. For LRD $0.5<H<1$. 
References

- Fractals, [http://mathworld.wolfram.com/Fractal.html](http://mathworld.wolfram.com/Fractal.html)