A Brief History of Generative Models for Power Law and Lognormal Distributions

Michael Mizenmacher

Presenter: Bo Jiang

Sept 17, 2010

Basic Definitions

Power law distribution

$$\bar{F}_X(x) = \mathbb{P}\{X > x\} \sim cx^{-\alpha} \quad \text{for } c, \alpha > 0$$

Basic Definitions

Power law distribution

$$\bar{F}_X(x) = \mathbb{P}\{X > x\} \sim cx^{-\alpha} \quad \text{for } c, \alpha > 0$$

Pareto distribution

$$\bar{F}_X(x) = \mathbb{P}\{X > x\} = \left(\frac{x}{k}\right)^{-\alpha} \quad \text{for } k, \alpha > 0$$

Basic Definitions

Power law distribution

$$\bar{F}_X(x) = \mathbb{P}\{X > x\} \sim cx^{-\alpha} \quad \text{for } c, \alpha > 0$$

Pareto distribution

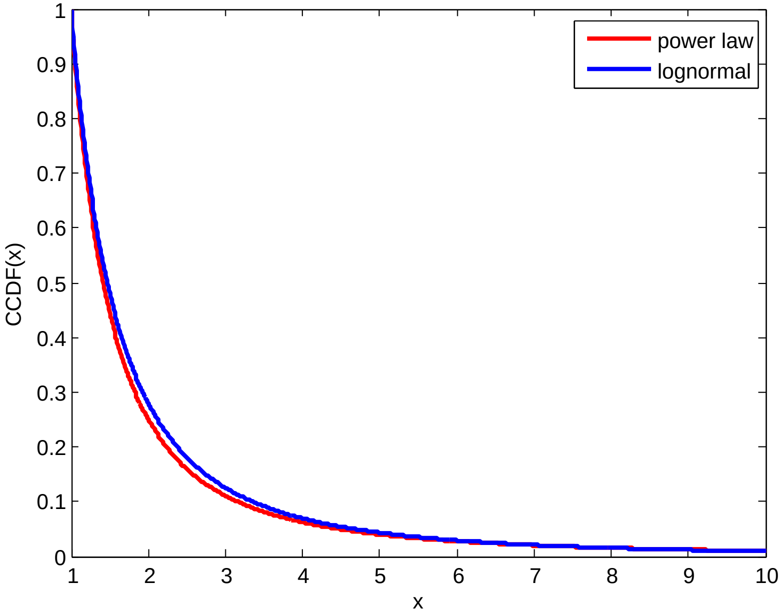
$$\bar{F}_X(x) = \mathbb{P}\{X > x\} = \left(\frac{x}{k}\right)^{-\alpha} \quad \text{for } k, \alpha > 0$$

Lognormal distribution

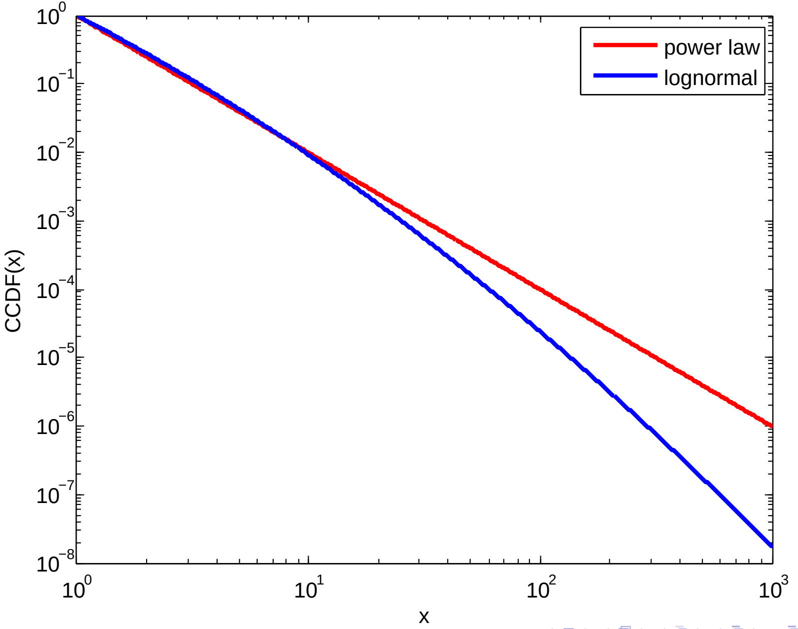
$$\log X \sim \mathcal{N}(\mu, \sigma^2), \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2 / 2\sigma^2}$$

If $\sigma \gg 1$, density approx. linear for large range in log-log scale.

Power Law vs Lognormal



Power Law vs Lognormal



Preferential Attachment

Prototype model

- Start with single node with one self-loop.
- At each time step, add one new node of outdegree 1
  - $\alpha < 1$.
  - With probability α, link to node uniformly at random.
  - With probability $1 - \alpha$, link to node $\propto$ *indegree*.

Preferential Attachment

Prototype model

- Start with single node with one self-loop.
- At each time step, add one new node of outdegree 1
  - $\alpha < 1$.
  - With probability α, link to node uniformly at random.
  - With probability $1 - \alpha$, link to node $\propto$ *indegree*.

Indegree Distribution

- CCDF: $\bar{F}(k) \sim ck^{-1/(1-\alpha)}$, power law
- Simple mean field analysis.
- Can be justified using martingales.

Preferential Attachment (cont'd)

Mean field analysis

- $X_j(t)$: # of nodes of indegree j at time t
- At time t, total # of nodes = sum of indegrees = t, i.e.

$$\sum_j X_j(t) = \sum_j jX_j(t) = t$$

Preferential Attachment (cont'd)

Mean field analysis

- $X_j(t)$: # of nodes of indegree j at time t
- At time t, total # of nodes = sum of indegrees = t, i.e.

$$\sum_j X_j(t) = \sum_j jX_j(t) = t$$

- For $j \geq 1$,

$$\frac{d}{dt}X_j = \alpha\frac{X_{j-1}}{t} + (1-\alpha)\frac{(j-1)X_{j-1}}{t} + \alpha\frac{X_j}{t} + (1-\alpha)\frac{jX_j}{t}$$

- $$\frac{d}{dt}X_0 = 1 - \alpha\frac{X_0}{t}$$
- Solve for $c_j = X_j/t$ in steady state, i.e. $t \to \infty$.

Optimization

Mandelbrot's model for word rank-frequency distribution

- Average amount of information per transmission:

$$\text{entropy:} \quad H = -\sum_{j=1}^{n} p_j \log_2 p_j$$

n = total # of words.

Optimization

Mandelbrot's model for word rank-frequency distribution

- Average amount of information per transmission:

$$\text{entropy:} \quad H = -\sum_{j=1}^{n} p_j \log_2 p_j$$

n = total # of words.

- Optimize average amount of information per unit cost

$$\min A = \frac{C}{H} = \frac{\sum_{j=1}^{n} p_j C_j}{H} \implies p_j = e^{-1} 2^{-HC_j/C}$$

C_j cost for jth word.

Optimization

Mandelbrot's model for word rank-frequency distribution

- Average amount of information per transmission:

$$\text{entropy:} \quad H = -\sum_{j=1}^{n} p_j \log_2 p_j$$

n = total # of words.

- Optimize average amount of information per unit cost

$$\min A = \frac{C}{H} = \frac{\sum_{j=1}^{n} p_j C_j}{H} \implies p_j = e^{-1} 2^{-HC_j/C}$$

C_j cost for jth word.

- $$C_j \sim \log_d j \implies \text{power law for } p_j \sim e^{-1} j^{-H \log_d 2/C}$$

d alphabet size.

Mulplicative Processes

Basic Operation

$$X_j = F_j X_{j-1} \quad \text{for r.v. } F_j$$

X_0, F_j's independent.

Mulplicative Processes

Basic Operation

$$X_j = F_j X_{j-1} \quad \text{for r.v. } F_j$$

X_0, F_j's independent.

Limit Distribution

$$\log X_j = \log X_0 + \sum_{k=1}^{j} \log F_k$$

- F_j's independent lognormal $\implies$ X_j lognormal $\forall j$
- In general, CLT $\implies$ X_j converges to lognormal

Mulplicative Processes

Basic Operation

$$X_j = F_j X_{j-1} \quad \text{for r.v. } F_j$$

X_0, F_j's independent.

Limit Distribution

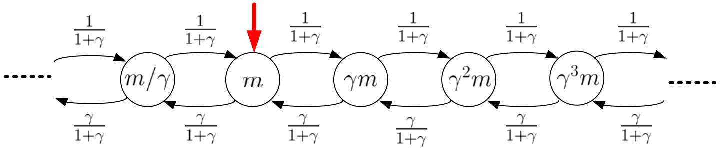
$$\log X_j = \log X_0 + \sum_{k=1}^{j} \log F_k$$

- F_j's independent lognormal $\implies$ X_j lognormal $\forall j$
- In general, CLT $\implies$ X_j converges to lognormal

However, seemingly trivial modifications can lead to power law!

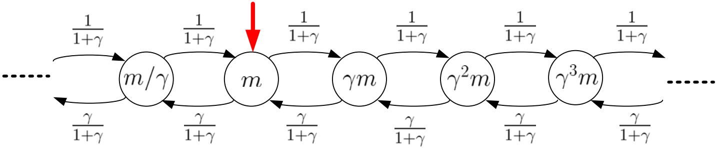
Multiplicative Processes (cont'd)

- $X_0 = m$.
- $\mathbb{P}\{F_j = \gamma\} = 1/(1+\gamma), \mathbb{P}\{F_j = \gamma^{-1}\} = \gamma/(1+\gamma), \gamma > 1$
- X_j lognormal

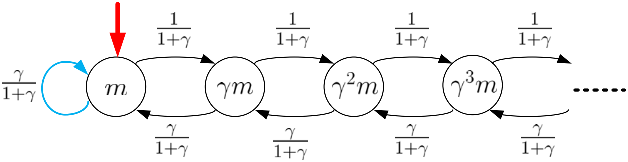


Multiplicative Processes (cont'd)

- $X_0 = m$.
- $\mathbb{P}\{F_j = \gamma\} = 1/(1+\gamma), \mathbb{P}\{F_j = \gamma^{-1}\} = \gamma/(1+\gamma), \gamma > 1$
- X_j lognormal



- But, if $X_j = \max\{F_j X_{j-1}, m\} \implies X_j$ power law!



Monkey Typing Randomly

The model

- Alphabet $\mathcal{X} = \{x_1, \ldots, x_n\}$ plus space
- Hit space with probability q
- Hit x_j with probability $q_j = (1 - q)p_j$, where $\sum_j p_j = 1$
- Space separates words

Monkey Typing Randomly

The model

- Alphabet $\mathcal{X} = \{x_1, \ldots, x_n\}$ plus space
- Hit space with probability q
- Hit x_j with probability $q_j = (1-q)p_j$, where $\sum_j p_j = 1$
- Space separates words

Frequency-rank Distributions

uniform $p_j = \frac{1}{n} \implies$ power law

Monkey Typing Randomly

The model

- Alphabet $\mathcal{X} = \{x_1, \ldots, x_n\}$ plus space
- Hit space with probability q
- Hit x_j with probability $q_j = (1-q)p_j$, where $\sum_j p_j = 1$
- Space separates words

Frequency-rank Distributions

uniform $p_j = \frac{1}{n} \implies$ power law

- length: $L(w)$
- frequency: $f(w) \approx \mathbb{P}(\{w\}) = \left(\frac{1-q}{n}\right)^{L(w)} q$

Monkey Typing Randomly

The model

- Alphabet $\mathcal{X} = \{x_1, \ldots, x_n\}$ plus space
- Hit space with probability q
- Hit x_j with probability $q_j = (1 - q)p_j$, where $\sum_j p_j = 1$
- Space separates words

Frequency-rank Distributions

uniform $p_j = \frac{1}{n} \implies$ power law

- length: $L(w)$
- frequency: $f(w) \approx \mathbb{P}(\{w\}) = \left(\frac{1-q}{n}\right)^{L(w)} q$
- rank: $r(w) \approx n^{L(w)-1} \sim n^{L(w)} \implies L(w) \approx \log_n r(w)$

Monkey Typing Randomly

The model

- Alphabet $\mathcal{X} = \{x_1, \ldots, x_n\}$ plus space
- Hit space with probability q
- Hit x_j with probability $q_j = (1-q)p_j$, where $\sum_j p_j = 1$
- Space separates words

Frequency-rank Distributions

uniform $p_j = \frac{1}{n} \implies$ power law

- length: $L(w)$
- frequency: $f(w) \approx \mathbb{P}(\{w\}) = \left(\frac{1-q}{n}\right)^{L(w)} q$
- rank: $r(w) \approx n^{L(w)-1} \sim n^{L(w)} \implies L(w) \approx \log_n r(w)$

$$\boxed{f \approx q r^{\log_n(1-q)-1}}$$

Monkey Typing Randomly (cont'd)

Distributions

non-uniform p_j: $\exists i \neq j$ s.t. $p_i \neq p_j$

- $\Omega_m =$ set of all distinct words of length m, $|\Omega_m| = n^m$
- $\Omega = \bigcup_{m=0}^{\infty} \Omega_m =$ set of all distinct words

Monkey Typing Randomly (cont'd)

Distributions

non-uniform p_j: $\exists i \neq j$ s.t. $p_i \neq p_j$

- $\Omega_m =$ set of all distinct words of length m, $|\Omega_m| = n^m$
- $\Omega = \bigcup_{m=0}^{\infty} \Omega_m =$ set of all distinct words
- $C_i(w) = i$-th letter of w, $i = 1, 2, \ldots, L(w)$

Monkey Typing Randomly (cont'd)

Distributions

non-uniform p_j: $\exists i \neq j$ s.t. $p_i \neq p_j$

- $\Omega_m =$ set of all distinct words of length m, $|\Omega_m| = n^m$
- $\Omega = \bigcup_{m=0}^{\infty} \Omega_m =$ set of all distinct words
- $C_i(w) = i$-th letter of w, $i = 1, 2, \ldots, L(w)$
- frequency of w,

$$f(w) = \mathbb{P}(\{w\}) = q \prod_{i=1}^{L(w)} q_{C_i(w)}$$

Monkey Typing Randomly (cont'd)

- Let $\mathbb{Q}_m$ be uniform distribution on Ω_m

Monkey Typing Randomly (cont'd)

- Let $\mathbb{Q}_m$ be uniform distribution on Ω_m
- As r.v. on Ω_m, $f : \Omega_m \to [0, 1]$ has CCDF $\bar{G}_m$ w.r.t. $\mathbb{Q}_m$,

Monkey Typing Randomly (cont'd)

- Let $\mathbb{Q}_m$ be uniform distribution on Ω_m
- As r.v. on Ω_m, $f : \Omega_m \to [0, 1]$ has CCDF $\bar{G}_m$ w.r.t. $\mathbb{Q}_m$,
- Rank of w within Ω_m

$$\begin{aligned} r_m(w) &\approx \#\{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \mathbb{Q}_m \{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \bar{G}_m(f(w)) \end{aligned}$$

Monkey Typing Randomly (cont'd)

- Let $\mathbb{Q}_m$ be uniform distribution on Ω_m
- As r.v. on Ω_m, $f : \Omega_m \to [0, 1]$ has CCDF $\bar{G}_m$ w.r.t. $\mathbb{Q}_m$,
- Rank of w within Ω_m

$$\begin{aligned} r_m(w) &\approx \#\{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \mathbb{Q}_m\{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \bar{G}_m(f(w)) \end{aligned}$$

- CLT $\implies$ $\bar{G}_m$ is roughly lognormal
  For fixed length, frequency-rank distribution is lognormal

Monkey Typing Randomly (cont'd)

- Let $\mathbb{Q}_m$ be uniform distribution on Ω_m
- As r.v. on Ω_m, $f : \Omega_m \to [0, 1]$ has CCDF $\bar{G}_m$ w.r.t. $\mathbb{Q}_m$,
- Rank of w within Ω_m

$$\begin{aligned} r_m(w) &\approx \#\{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \mathbb{Q}_m\{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \bar{G}_m(f(w)) \end{aligned}$$

- CLT $\implies$ $\bar{G}_m$ is roughly lognormal
  For fixed length, frequency-rank distribution is lognormal
- Upper bound word length, Anscombe's theorem $\implies$
  lognormal frequency-rank distribution

Monkey Typing Randomly (cont'd)

- Let $\mathbb{Q}_m$ be uniform distribution on Ω_m
- As r.v. on Ω_m, $f : \Omega_m \to [0, 1]$ has CCDF $\bar{G}_m$ w.r.t. $\mathbb{Q}_m$,
- Rank of w within Ω_m

$$\begin{aligned} r_m(w) &\approx \#\{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \mathbb{Q}_m\{w' \in \Omega_m : f(w') > f(w)\} \\ &= |\Omega_m| \bar{G}_m(f(w)) \end{aligned}$$

- CLT $\implies$ $\bar{G}_m$ is roughly lognormal
  For fixed length, frequency-rank distribution is lognormal
- Upper bound word length, Anscombe's theorem $\implies$ lognormal frequency-rank distribution
- No bound on word length $\implies$ power law (very complicated proof involving analytical number theory)

Double Pareto Distribution

- For $t \geq 0$, let $X(t)$ be lognormal, i.e.

$$\log X(t) \sim \mathcal{N}(\log x_0 + \mu t, \sigma^2 t)$$

where $x_0 > 0$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$.

Double Pareto Distribution

- For $t \geq 0$, let $X(t)$ be lognormal, i.e.

$$\log X(t) \sim \mathcal{N}(\log x_0 + \mu t, \sigma^2 t)$$

  where $x_0 > 0$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$.
- Let $T \sim \exp(\lambda)$

Double Pareto Distribution

- For $t \geq 0$, let $X(t)$ be lognormal, i.e.

$$\log X(t) \sim \mathcal{N}(\log x_0 + \mu t, \sigma^2 t)$$

  where $x_0 > 0$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$.
- Let $T \sim \exp(\lambda)$
- Distribution of $Y = X(T)$?

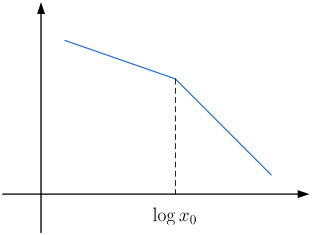
Double Pareto Distribution

- For $t \geq 0$, let $X(t)$ be lognormal, i.e.

$$\log X(t) \sim \mathcal{N}(\log x_0 + \mu t, \sigma^2 t)$$

  where $x_0 > 0$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$.

- Let $T \sim \exp(\lambda)$
- Distribution of $Y = X(T)$? In log-log scale,



Double Pareto Distribution

- If population grows exponentially, age distribution will be truncated exponential.

Double Pareto Distribution

- If population grows exponentially, age distribution will be truncated exponential.
- $X(t)$ is geometric Brownian motion

$$\dot{X} = (\mu + \xi)X \quad \text{or} \quad \mathrm{d}X(t) = X(\mu \mathrm{d}t + \sigma \mathrm{d}W)$$

  ξ white noise, W standard Brownian motion.

  - “noisy” exponential growth
  - “continuous” multiplication

Double Pareto Distribution

- If population grows exponentially, age distribution will be truncated exponential.
- $X(t)$ is geometric Brownian motion

$$\dot{X} = (\mu + \xi)X \quad \text{or} \quad \mathrm{d}X(t) = X(\mu \mathrm{d}t + \sigma \mathrm{d}W)$$

  ξ white noise, W standard Brownian motion.

  - "noisy" exponential growth
  - "continuous" multiplication

- If T is geometric: lognormal body, power law tail

$$X(T) = x_0 \prod_{i=1}^{T} X_i, \quad \text{where } X_i \text{ i.i.d} \sim \text{log-normal}(\mu,\sigma)$$

Double Pareto Distribution

- If population grows exponentially, age distribution will be truncated exponential.
- $X(t)$ is geometric Brownian motion

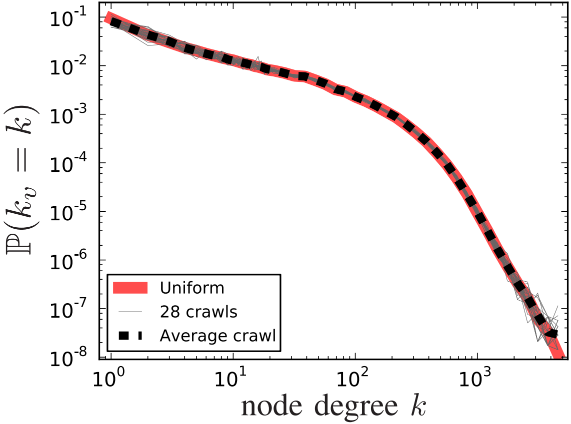
$$\dot{X} = (\mu + \xi)X \quad \text{or} \quad \mathrm{d}X(t) = X(\mu \mathrm{d}t + \sigma \mathrm{d}W)$$

  ξ white noise, W standard Brownian motion.
  - "noisy" exponential growth
  - "continuous" multiplication
- If T is geometric: lognormal body, power law tail

$$X(T) = x_0 \prod_{i=1}^{T} X_i, \quad \text{where } X_i \text{ i.i.d} \sim \text{log-normal}(\mu,\sigma)$$

- Exponential growth + exponential (geometric) stopping $\implies$ power law

Double Pareto Degree Distribution?



Degree distribution of Facebook [Gjoka et al 2010]