Power Laws

Leonid E. Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

20.01.2014

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

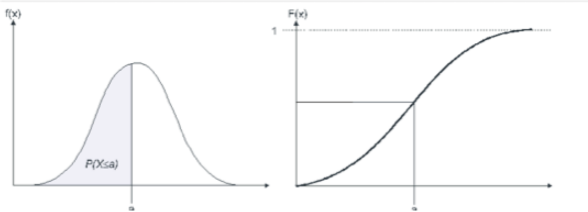
Continues distribution

- Continues random variable X
- Probability density function (PDF) $f(x)$:

$$Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

- $f(x) \geq 0$, $\int_{-\infty}^{\infty} = 1$
- Cumulative distribution function (CDF)

$$F(x) = Pr(X \leq x) = \int_{-\infty}^{x} f(y)dy$$



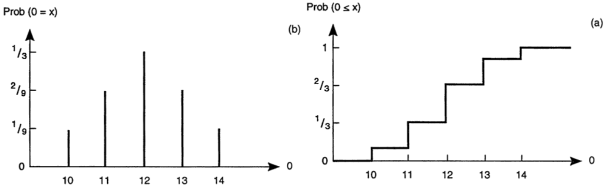
Discrete distribution

- Discrete random variable X_i
- Probability mass function (PMF) $p(x)$:

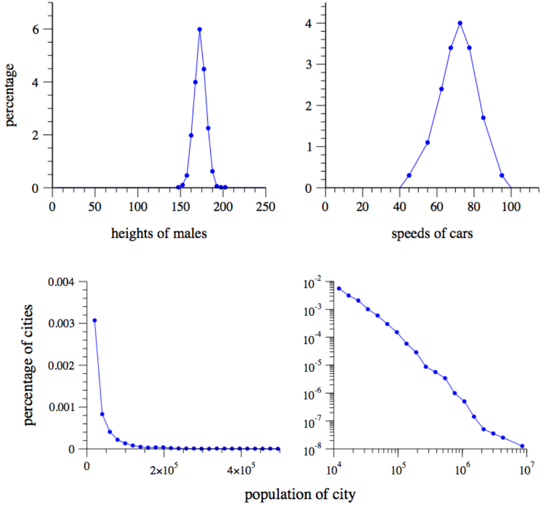
$$p(x) = Pr(X_i = x)$$

- $p(x) \geq 0$, $\sum_x p(x) = 1$
- Cumulative distribution function (CDF)

$$F(x) = Pr(X_i \leq x) = \sum_{y \leq x} p(y)$$

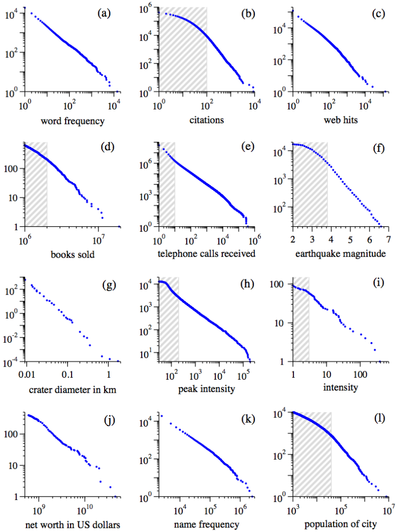


Empirical distributions



Newman et.al, 2005

Empirical distributions



Newman et.al, 2005

Continues approximation

- Power law

$$p(x) = Cx^{-\alpha} = \frac{C}{x^{\alpha}}, \quad \text{for } x \geq x_{min}$$

- Normalization ($\alpha > 1$)

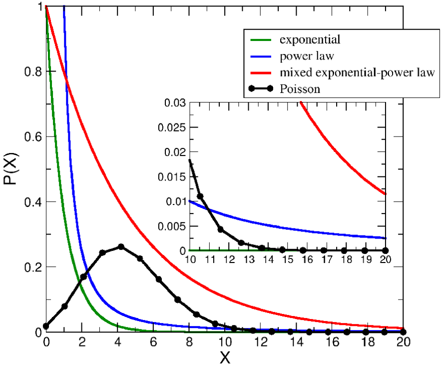
$$1 = \int_{x_{min}}^{\infty} p(x)dx = C \int_{x_{\min}}^{\infty} \frac{dx}{x^{\alpha}} = \frac{C}{\alpha - 1} x_{\min}^{-\alpha+1}$$

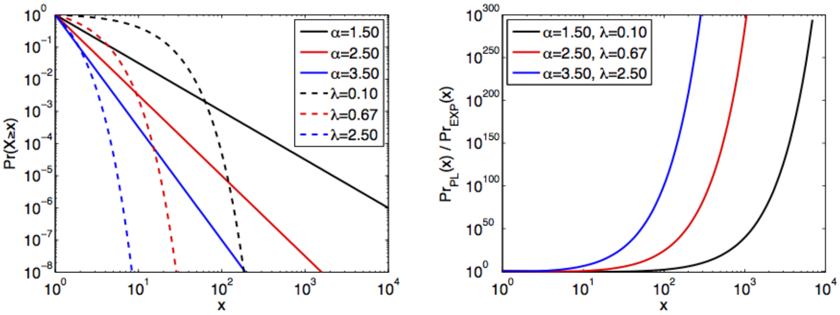
$$C = (\alpha - 1) x_{\min}^{\alpha-1}$$

- Power law PDF

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

Power Laws


exponential
power law
mixed exponential-power law
Poisson
P(X)
X
0
0.2
0.4
0.6
0.8
1
0
2
4
6
8
10
12
14
16
18
20
0.03
0.025
0.02
0.015
0.01
0.005
0
10
11
12
13
14
15
16
17
18
19
20



$$p(x) = Cx^{-\alpha}$$

$$\log p(x) = \log C - \alpha \log x$$

Moments

- PDF

$$p(x) = \frac{C}{x^\alpha}$$

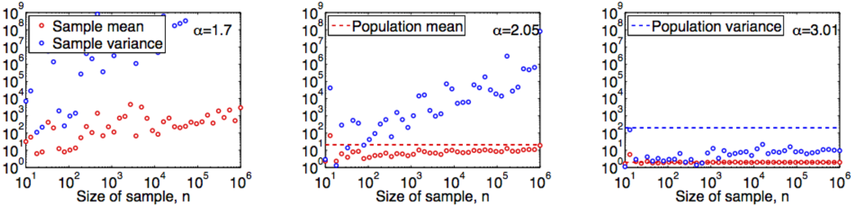
- mean $(\alpha > 2)$

$$\langle x \rangle = \int_{x_{\min}}^{\infty} x p(x) dx = C \int_{x_{\min}}^{\infty} \frac{dx}{x^{\alpha-1}} = \frac{\alpha - 1}{\alpha - 2} x_{\min}$$

- standard deviation $(\alpha > 3)$

$$\langle x^2 \rangle = \int_{x_{\min}}^{\infty} x^2 p(x) dx = C \int_{x_{\min}}^{\infty} \frac{dx}{x^{\alpha-2}} = \frac{\alpha - 1}{\alpha - 3} x_{\min}^2$$

Moments



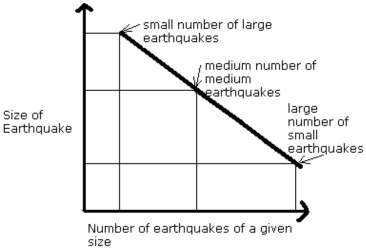
Clauset et.al, 2009

- Scaling of the density

$$p(bx) = C(bx)^{-\alpha} = Cb^{-\alpha}x^{-\alpha} \sim p(x)$$

- Scale invariance

$$\frac{p(10x_2)}{p(10x_1)} = \frac{p(x_2)}{p(x_1)}$$



Power Laws

- Cumulative distribution function CDF

$$F(x) = Pr(X \leq x) = \int_{-\infty}^{x} p(y)dy = \int_{x_{\text{min}}}^{x} p(y)dy$$

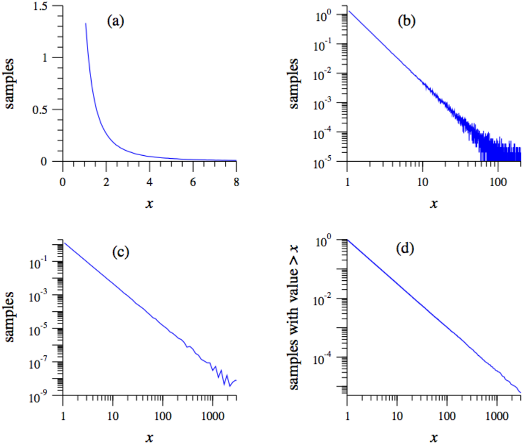
- Complimentary cumulative distribution function cCDF

$$\bar{F}(x) = 1 - F(x) = Pr(X > x) = \int_{x}^{\infty} p(y)dy$$

- Power law $p(x) = Cx^{-\alpha}$

$$\bar{F}(x) = \frac{C}{\alpha - 1} x^{-(\alpha-1)} = \left(\frac{x}{x_{\text{min}}}\right)^{-(\alpha-1)}$$

Power law histograms



Newman et.al, 2005

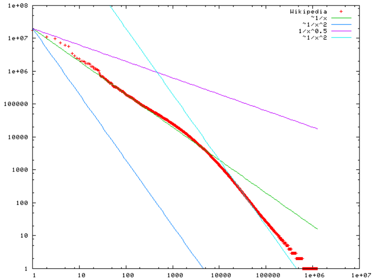
Zips'f law

Zipf's law - the frequency of any word is inversely proportional to its rank in the frequency table: words are sorted by frequency in decreasing order.

- word frequency

$$f(k) \sim 1/k^s$$

- cCDF $\bar{F}(x) = Pr(X > x)$: for n-th word, there are n words with higher frequences sorted before it
- Rank-frequency plot: rank(frequency) , x- frequency, y-rank



Node degree distribution

- Node degrees $k_i = 1, 2, ...k_{\max}$
- Degree distribution $P(k) \equiv P(k_i = k)$

$$P(k) = \frac{n_k}{n} = \frac{n_k}{\sum_k n_k}$$

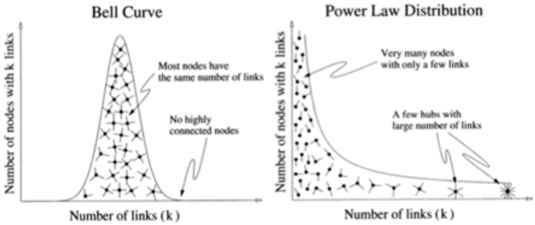
- Power law

$$P(k) = ck^{-\gamma} = \frac{c}{k^\gamma}$$

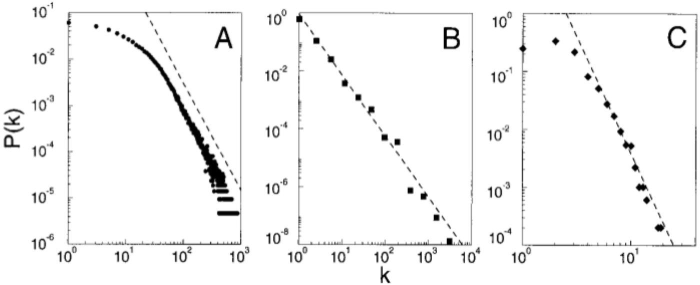
- Logarithmic coordinates

$$\log(P(k)) = -\gamma \log k + \log c$$

Power law networks



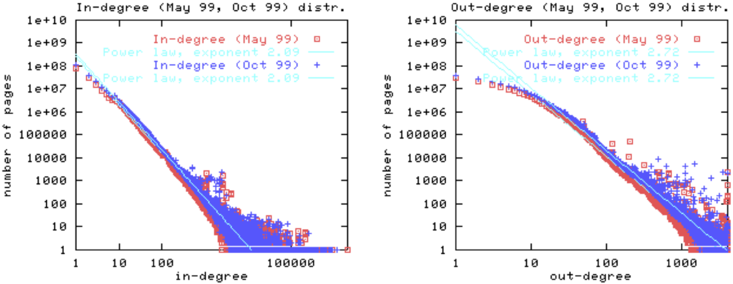
Power law networks



Actor collaboration graph, N=212,250 nodes, $\langle k \rangle = 28.8$, $\gamma = 2.3$
WWW, N = 325,729 nodes, $\langle k \rangle = 5.6$, $\gamma = 2.1$
Power grid data, N = 4941 nodes, $\langle k \rangle = 5.5$, $\gamma = 4$

Barabasi et.al, 1999

Power law networks



In- and out- degrees of WWW crawl 1999

Broder et.al, 1999

Parameter estimation: α

Maximum likelihood estimation of parameter α

- Let $\{x_i\}$ be a set of n observations (points) independetly sampled from the distribution

$$P(x_i) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\alpha}$$

- Probability of the sample

$$P(\{x_i\}|\alpha) = \prod_i^n \frac{\alpha - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\alpha}$$

- Bayes' theorem

$$P(\alpha|\{x_i\}) = P(\{x_i\}|\alpha) \frac{P(\alpha)}{P(\{x_i\})}$$

Maximum likelihood

- log-likelihood

$$\mathcal{L} = \ln P(\alpha|\{x_i\}) = n\ln(\alpha - 1) - n\ln x_{\min} - \alpha \sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}$$

- maximization $\frac{\partial \mathcal{L}}{\partial \alpha} = 0$

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1}$$

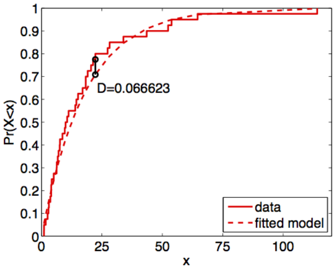
- error estimate

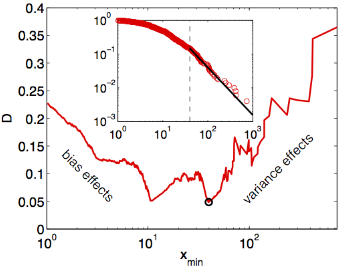
$$\sigma = \sqrt{n}\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1} = \frac{\alpha - 1}{\sqrt{n}}$$

Parameter estimation: x_{min}

- Kolmogorov-Smirnov test (compare model and experimental CDF)

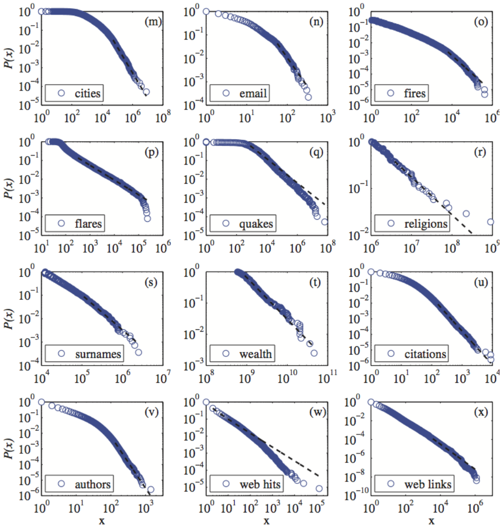
$$D = \max_x |F(x|\alpha, x_{min}) - F_{exp}|$$





Clauset et.al, 2009

Empirical models



Clauset et.al, 2009

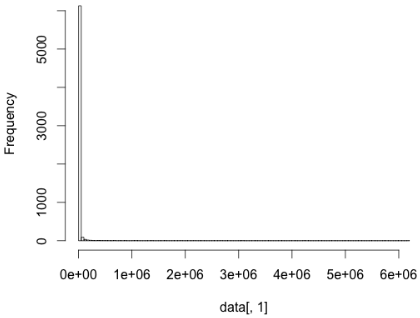
Example

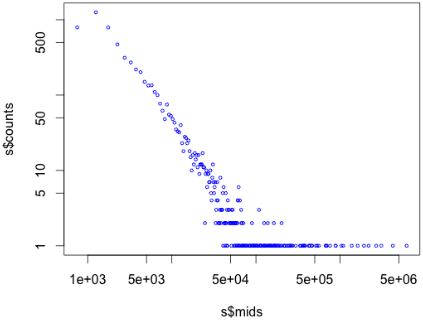
Word count:

```
6187267 the
4239632 be
3093444 of
2687863 and
2186369 a
1924315 in
1620850 to
........
801 incredibly
801 historically
801 decision-making
800 wildly
800 reformer
800 quantum
800 considering
```

Example

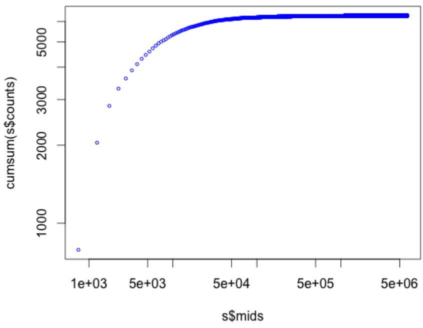
Histogram (PDF)

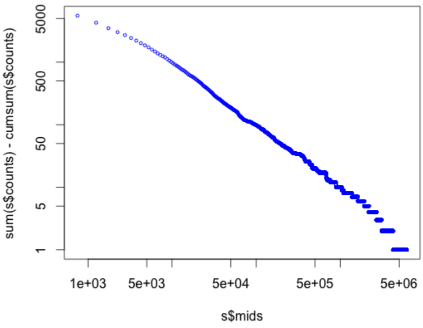




Example

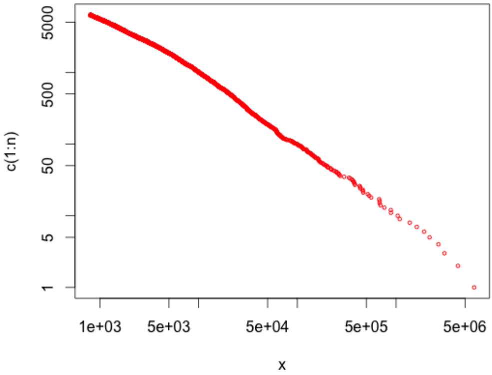
CDF and cCDF





Example

Rank-Frequency plot



References

- Power laws, Pareto distributions and Zipf's law, M. E. J. Newman, Contemporary Physics, pages 323–351, 2005.
- Power-Law Distribution in Empirical Data, A. Clauset, C.R. Shalizi, M.E.J. Newman, SIAM Review, Vol 51, No 4, pp. 661-703, 2009.
- A Brief History of Generative Models for Power Law and Lognormal Distributions, M. Mitzenmacher, Internet Mathematics Vol 1, No 2, pp 226-251.