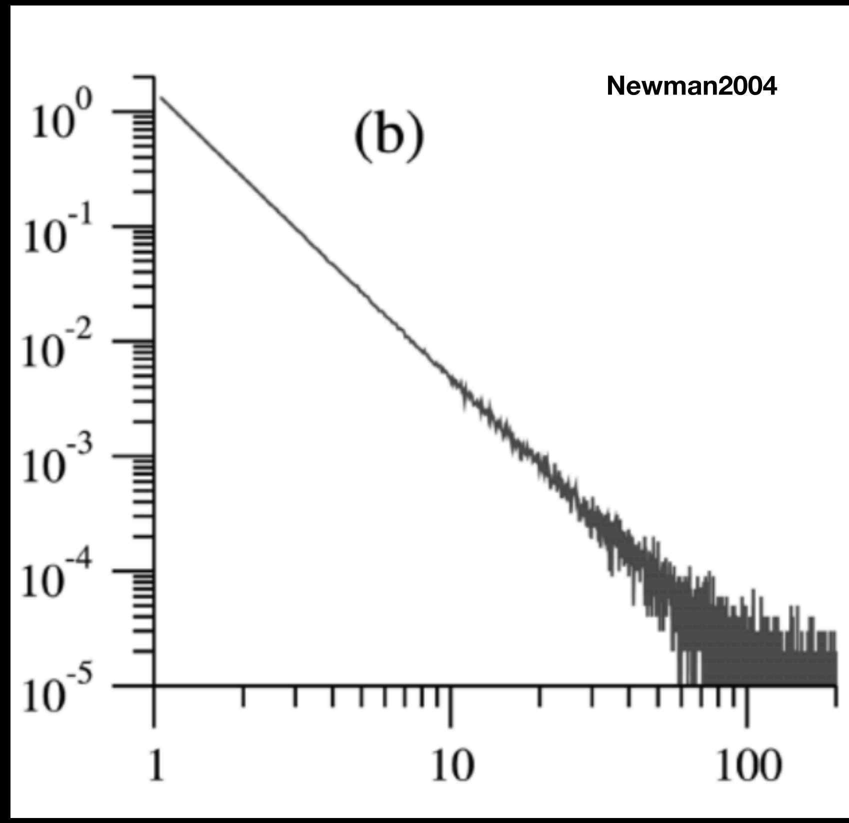
What

- Zipf: $: n \sim r^{-b}$ (descending (in r) frequency distribution) "The ***r*** th largest city has ***n*** inhabitants"
- Pareto $: r \sim n^{-1/b}$ Or $P[X \geq x] \propto x^{-k}$ (CDF) "***r*** cities have ***n*** or more inhabitants"
- Power Law $: p(x) \propto \frac{d}{dx} x^{-k} \propto x^{-\alpha}$ (PDF)
  - Diverges at x = 0 (thus x_{min})
  - $\alpha > 1$ (normalisation)
  - Divergent mean (for $\alpha \leq 2$)
  - Divergent variance (for $\alpha \leq 3$)
  - Scale 'invariant' (sufficient & necessary)



Finite Data in Sample. Extreme/Rare Events in Population.

Data

Naive approach

1. Regression analysis has assumptions (affects normalisation, errors, independence, correlations, etc)
2. Choice of bins (additional parameter)
3. Large R^2 but process even if generative process is not a power law

Robust approach

1. x_{min} (sensitive to sample size n)
   1. marginal-likelihod (BIC)
      1. unfixed-too many parameters
      2. Not for continuous data
   2. Kolmogorov-Smirnov Statistic
2. Fix x_{min}, find α ($\Pi p(x \mid \alpha)$, sensitive to n)
3. Goodness of fit - large p value does not necessarily mean data is from power law

Comparing alternate hypothesis

1. Say alternate hypothesis is an exponential or log-normal distribution or a mixture!
2. Likelihood ratio test: $\frac{p(x \mid hyp1)}{p(x \mid hyp2)}$

Unlikely that one single power law is enough to explain all ranges of x

How/Why

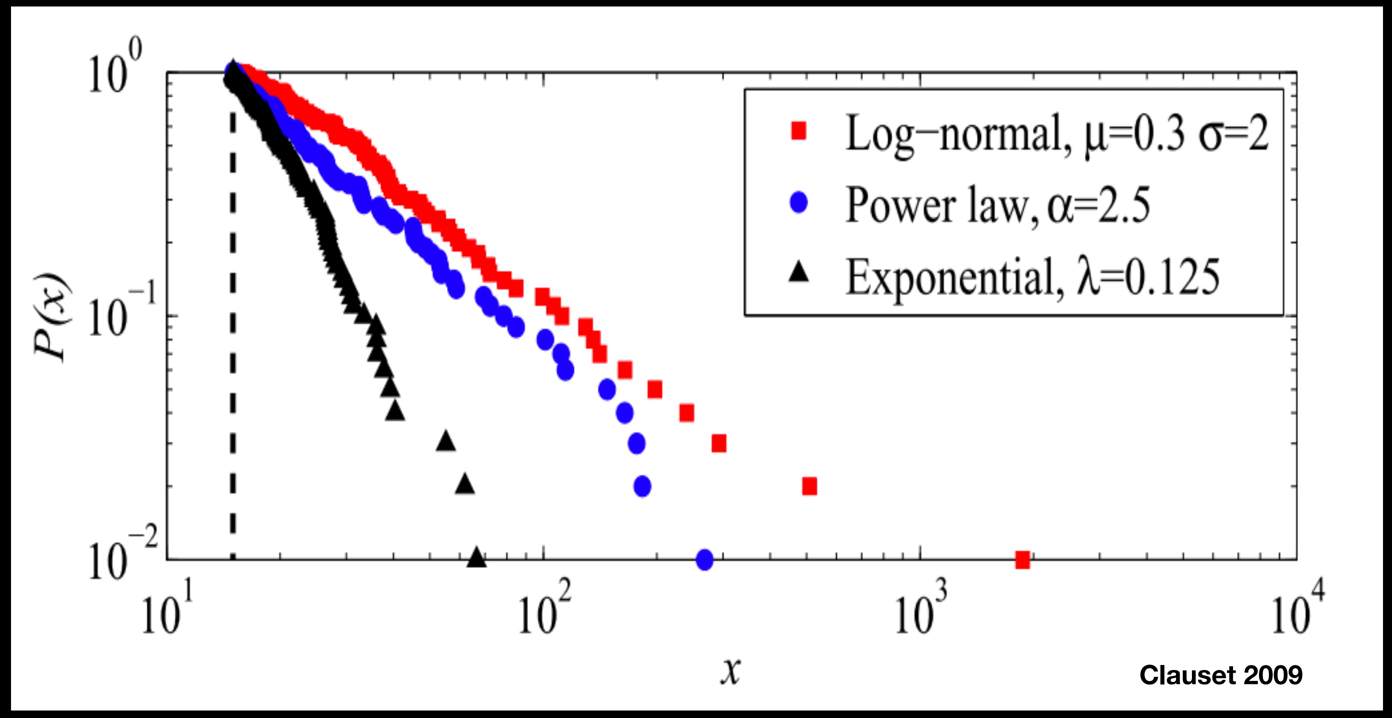
Power laws via

1. Preferential attachment (rich get richer)
2. Optimisation (information/word length)
3. Multiplicative process (X(t+1) = F*X(t)) with minimum
4. Monkey typing randomly with equal letter frequencies
5. Self organised criticality, Critical phenomena/phase transitions (sandpile, forests), etc.

Log normal via

1. Multiplicative process (X(t+1) = F*X(t)) without minimum
2. Monkey typing randomly with unequal letter frequencies

Competing



Rachith Aiyappa

Papers Read

1. M.E.J Newman, Power Laws, Pareto distributions and Zipf's law . Contemporary Physics 46, 323 (2005).

2. Clauset et al., Power-law distributions in empirical data.SIAM Review 51, 661 (2009)

3. M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. Internet Mathematics 1(2), 226-251(2004)

4. L. A. Adamic, Zipf, Power-laws, and Pareto - a ranking tutorial

5. C. R. Shalizi, Power Law Distributions, 1/f Noise, Long-Memory Time Series, Notebooks.