Economic Networks

Theory and Empirics

Giorgio Fagiolo
Laboratory of Economics and Management (LEM)
Sant'Anna School of Advanced Studies, Pisa, Italy

http://www.lem.sssup.it/fagiolo/
giorgio.fagiolo@sssup.it

Lecture 4

This Lecture

- What is a network? Examples of networks
- Why networks are important for economists?
- Networks and graphs
- Measures and metrics on networks
- Distributions of metrics and measures in large networks
- Models of network formation
- Null statistical network models
- Economic applications

Social vs. Complex Network Analysis

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)
  - ✓ Fields: Sociology, Psychology, Economics (see Borgatti et al., Science, Vol. 323, February 2009)

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)
  - ✓ Fields: Sociology, Psychology, Economics (see Borgatti et al., Science, Vol. 323, February 2009)
- Complex Network Analysis (CNA, Newman, 2010)

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)
  - ✓ Fields: Sociology, Psychology, Economics (see Borgatti et al., Science, Vol. 323, February 2009)
- Complex Network Analysis (CNA, Newman, 2010)
  - ✓ Large network size, data often retrieved automatically from large databases (Internet, WTW, biological nets, large social networks, etc.)

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)
  - ✓ Fields: Sociology, Psychology, Economics (see Borgatti et al., Science, Vol. 323, February 2009)
- Complex Network Analysis (CNA, Newman, 2010)
  - ✓ Large network size, data often retrieved automatically from large databases (Internet, WTW, biological nets, large social networks, etc.)
  - ✓ Fields: Physics, Biology, Computer Science, Economics (see Schweitzer, Fagiolo, et al., Science, Vol. 325, July 2009)

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)
  - ✓ Fields: Sociology, Psychology, Economics (see Borgatti et al., Science, Vol. 323, February 2009)
- Complex Network Analysis (CNA, Newman, 2010)
  - ✓ Large network size, data often retrieved automatically from large databases (Internet, WTW, biological nets, large social networks, etc.)
  - ✓ Fields: Physics, Biology, Computer Science, Economics (see Schweitzer, Fagiolo, et al., Science, Vol. 325, July 2009)
- Different goals

Social vs. Complex Network Analysis

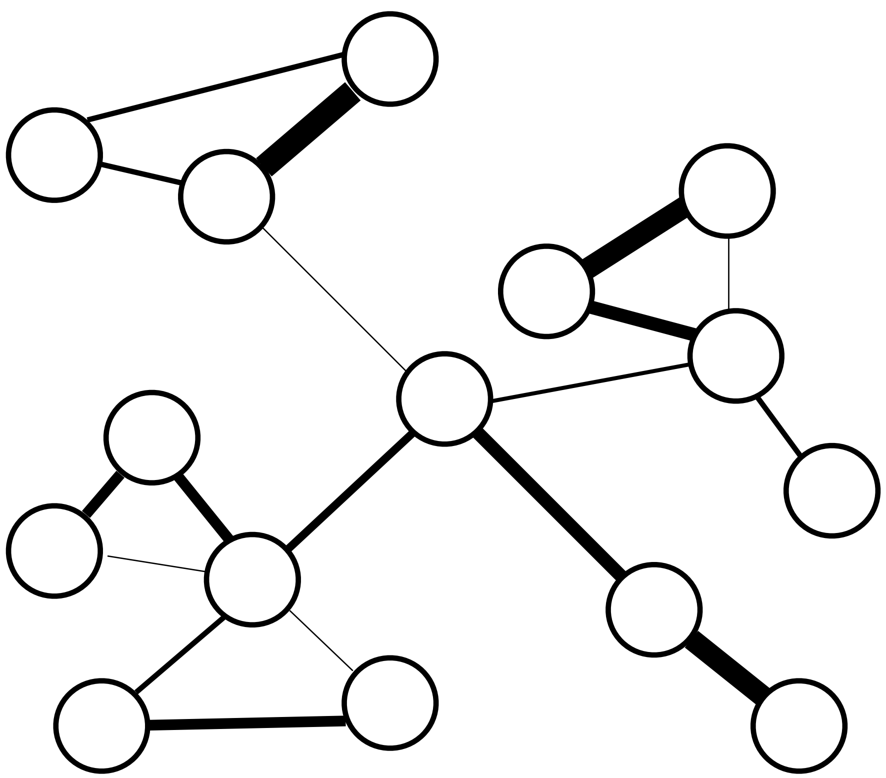
- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)
  - ✓ Fields: Sociology, Psychology, Economics (see Borgatti et al., Science, Vol. 323, February 2009)
- Complex Network Analysis (CNA, Newman, 2010)
  - ✓ Large network size, data often retrieved automatically from large databases (Internet, WTW, biological nets, large social networks, etc.)
  - ✓ Fields: Physics, Biology, Computer Science, Economics (see Schweitzer, Fagiolo, et al., Science, Vol. 325, July 2009)
- Different goals
  - ✓ SNA: node behavior, mostly descriptive analysis (no models)

Social vs. Complex Network Analysis

- Social Network Analysis (SNA: Wasserman and Faust, 1994)
  - ✓ Small networks size, data often obtained through questionnaires or experiments (N<100)
  - ✓ Fields: Sociology, Psychology, Economics (see Borgatti et al., Science, Vol. 323, February 2009)
- Complex Network Analysis (CNA, Newman, 2010)
  - ✓ Large network size, data often retrieved automatically from large databases (Internet, WTW, biological nets, large social networks, etc.)
  - ✓ Fields: Physics, Biology, Computer Science, Economics (see Schweitzer, Fagiolo, et al., Science, Vol. 325, July 2009)
- Different goals
  - ✓ SNA: node behavior, mostly descriptive analysis (no models)
  - ✓ CNA: network statistical properties, mostly quantitative (comparison with models), focus on distributional properties of node- (and link)-specific statistics and their dynamics over time

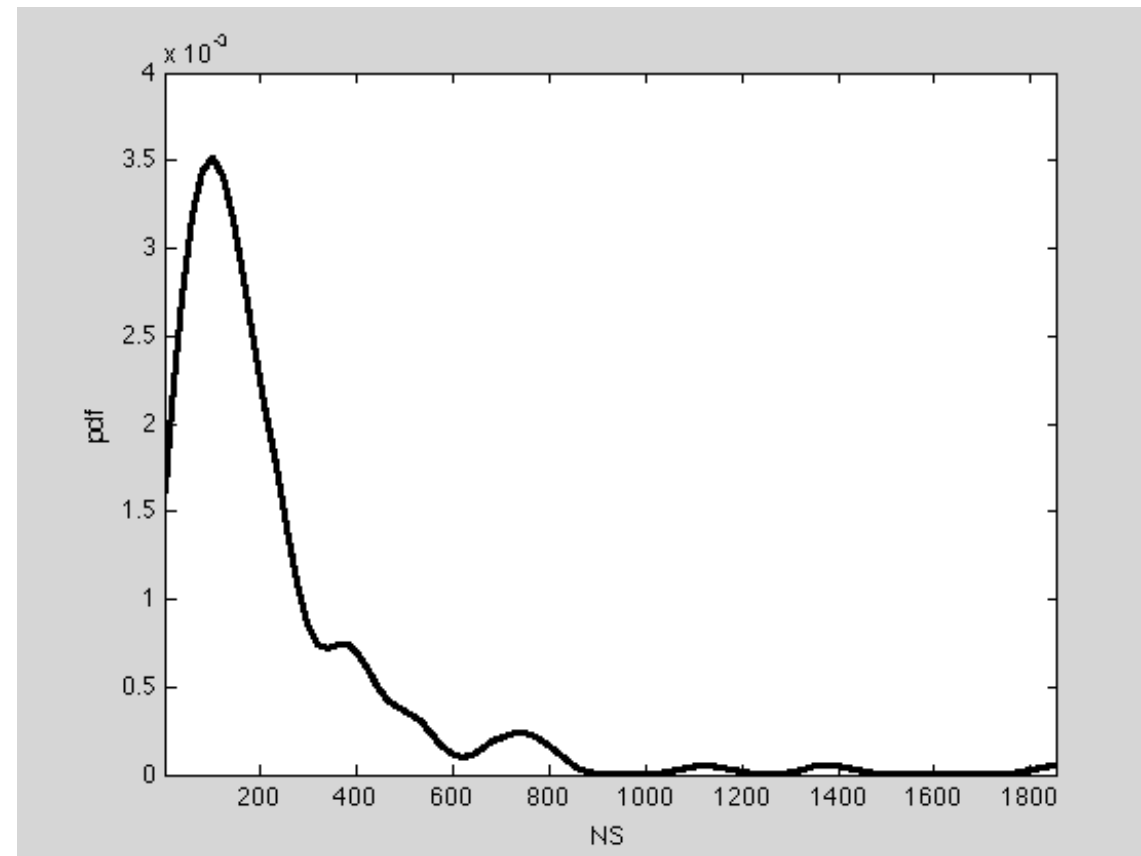
Node-Specific Statistic Distributions

Time t



$X_{i,t}$ $i=1,\ldots,N$

$f(X_t)$



Time-t node-distribution for X

Distribution dynamics : $f(Xt)$, $t=1,2,\ldots,T$

Distribution of a Random Variable

Distribution of a Random Variable

- Let X be a random variable defined over a given probability space

Distribution of a Random Variable

- Let X be a random variable defined over a given probability space
  - ✓ Discrete: p(x)=Prob{X=x} is the probability distribution function (PDF) of X and G(x)=Prob{X≤x} is its cumulative distribution function (CDF)

Distribution of a Random Variable

- Let X be a random variable defined over a given probability space
  - ✓ Discrete: $p(x)=\text{Prob}\{X=x\}$ is the probability distribution function (PDF) of X and $G(x)=\text{Prob}\{X\leq x\}$ is its cumulative distribution function (CDF)
  - ✓ Continuous: $F(x)=\text{Prob}\{X<x\}$ is the CDF of X; $f(x)=dF/dx$: density function, i.e. probability that X is in a neighborhood of x

Distribution of a Random Variable

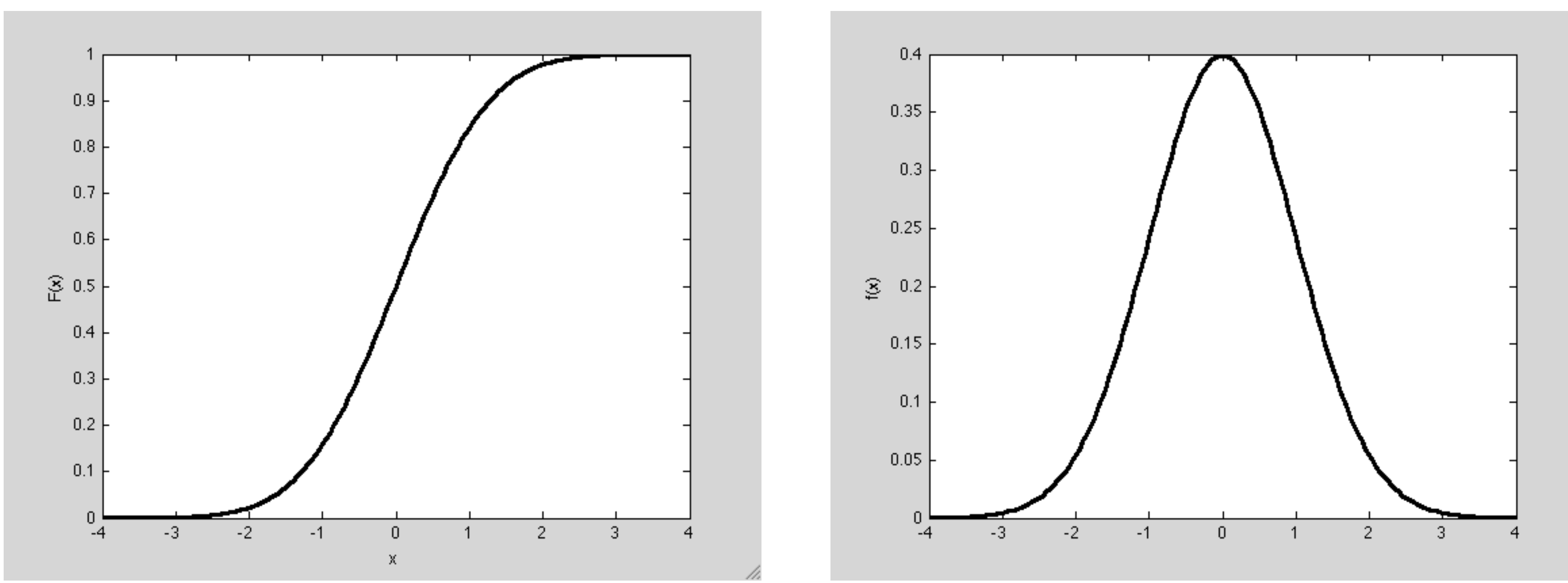
- Let X be a random variable defined over a given probability space
  - ✓ Discrete: $p(x)=\text{Prob}\{X=x\}$ is the probability distribution function (PDF) of X and $G(x)=\text{Prob}\{X\leq x\}$ is its cumulative distribution function (CDF)
  - ✓ Continuous: $F(x)=\text{Prob}\{X<x\}$ is the CDF of X; $f(x)=dF/dx$: density function, i.e. probability that X is in a neighborhood of x
- The CDF of X fully characterizes X (alt: sequence of all moments)

Distribution of a Random Variable

- Let X be a random variable defined over a given probability space
  - ✓ Discrete: $p(x)=\text{Prob}\{X=x\}$ is the probability distribution function (PDF) of X and $G(x)=\text{Prob}\{X\leq x\}$ is its cumulative distribution function (CDF)
  - ✓ Continuous: $F(x)=\text{Prob}\{X<x\}$ is the CDF of X; $f(x)=dF/dx$: density function, i.e. probability that X is in a neighborhood of x
- The CDF of X fully characterizes X (alt: sequence of all moments)
- Economics: Income distribution, consumption distribution, firm-size distribution, etc.

Distribution of a Random Variable

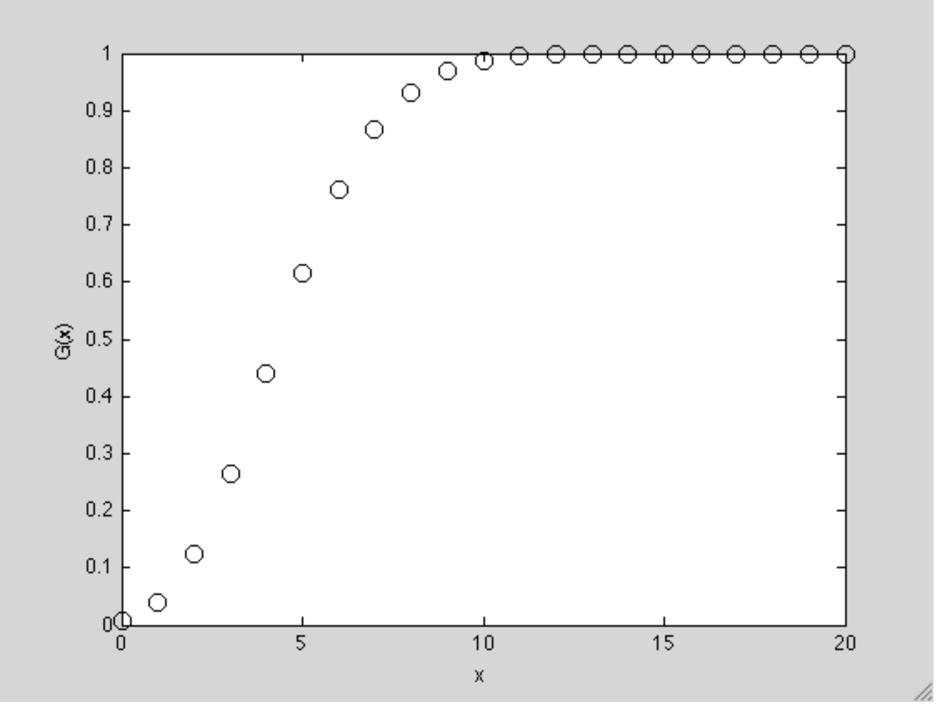
- Let X be a random variable defined over a given probability space
  - ✓ Discrete: p(x)=Prob{X=x} is the probability distribution function (PDF) of X and G(x)=Prob{X≤x} is its cumulative distribution function (CDF)
  - ✓ Continuous: F(x)=Prob{X<x} is the CDF of X; f(x)=dF/dx: density function, i.e. probability that X is in a neighborhood of x
- The CDF of X fully characterizes X (alt: sequence of all moments)
- Economics: Income distribution, consumption distribution, firm-size distribution, etc.

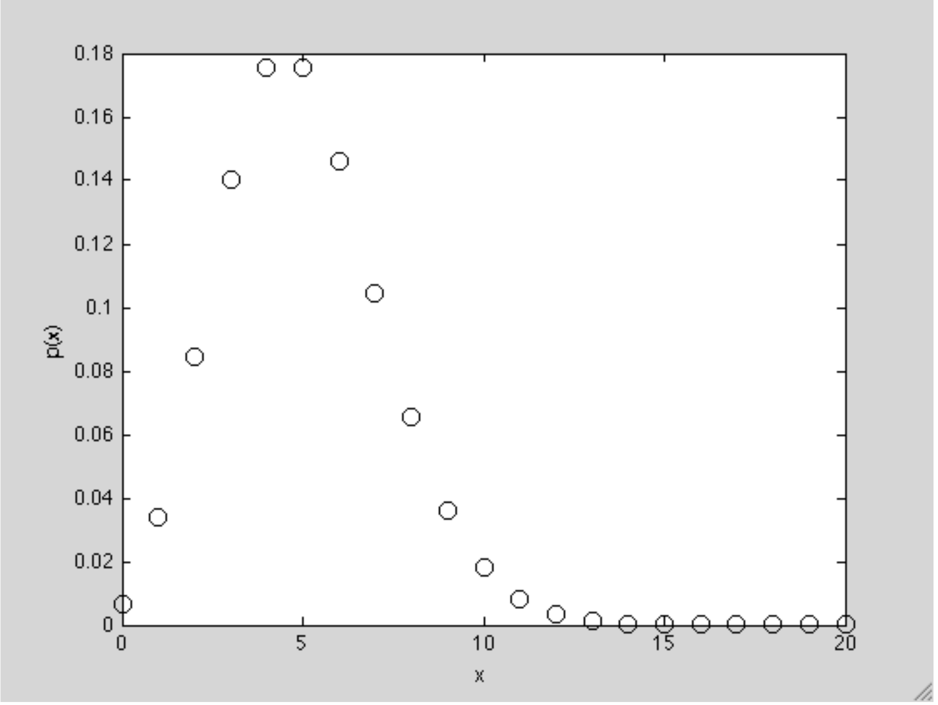


The CDF (left) and the density (right) of a standard normal random variable

Distribution of a Random Variable

- Let X be a random variable defined over a given probability space
  - ✓ Discrete: $p(x)=\text{Prob}\{X=x\}$ is the probability distribution function (PDF) of X and $G(x)=\text{Prob}\{X\leq x\}$ is its cumulative distribution function (CDF)
  - ✓ Continuous: $F(x)=\text{Prob}\{X<x\}$ is the CDF of X; $f(x)=dF/dx$: density function, i.e. probability that X is in a neighborhood of x
- The CDF of X fully characterizes X (alt: sequence of all moments)
- Economics: Income distribution, consumption distribution, firm-size distribution, etc.





The CDF (left) and the PDF (right) of a Poisson random variable ($\lambda=5$)

Discrete Random Variables

- Geometric (p): k=1,2,...
  - ✓ E(X)=1/p; VAR(X)= $(1-p)/p^2$

$$p(k) = p(1-p)^k$$

- Poisson(λ): k=0,1,...
  - ✓ E(X)=VAR(X)= λ

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

- Binomial(n,p): k=0,...,n
  - ✓ E(X)=np; VAR(X)=np(1-p)
  - ✓ Tends to a Poisson as n→∞

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Continuous Random Variables (1)

- Normal (μ,σ): $x \in \Re$
  - ✓ $E(X)=\mu$; $VAR(X)=\sigma^2$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Exponential(λ): $x \in \Re^+$
  - ✓ $E(X)=1/\lambda$; $VAR(X)=1/\lambda^2$

$$f(x) = \lambda e^{-\lambda x}$$

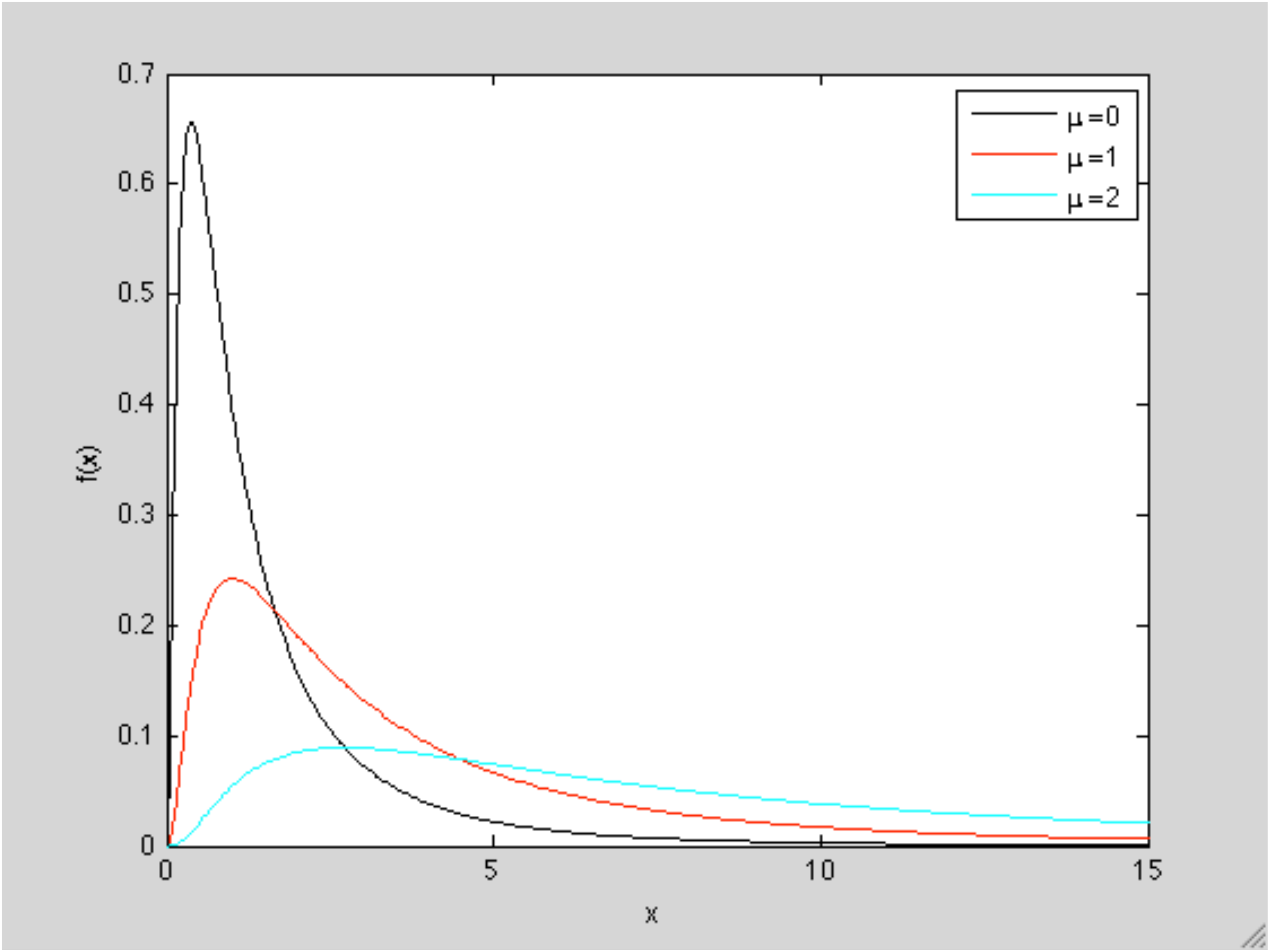
- Beta(a,b): $x \in [0,1]$
  - ✓ $E(X)=a/(a+b)$; $VAR(X)=h(a,b)$
  - ✓ $B(a,b)$=Beta function

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$$

Continuous Random Variables (2)

- Log-Normal (μ,σ): $x \in \Re^+$
  - ✓ X is LogN(μ,σ) ⇔ log(X) is N(μ,σ)
  - ✓ NB: EX≠μ and VAR(X)≠σ

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

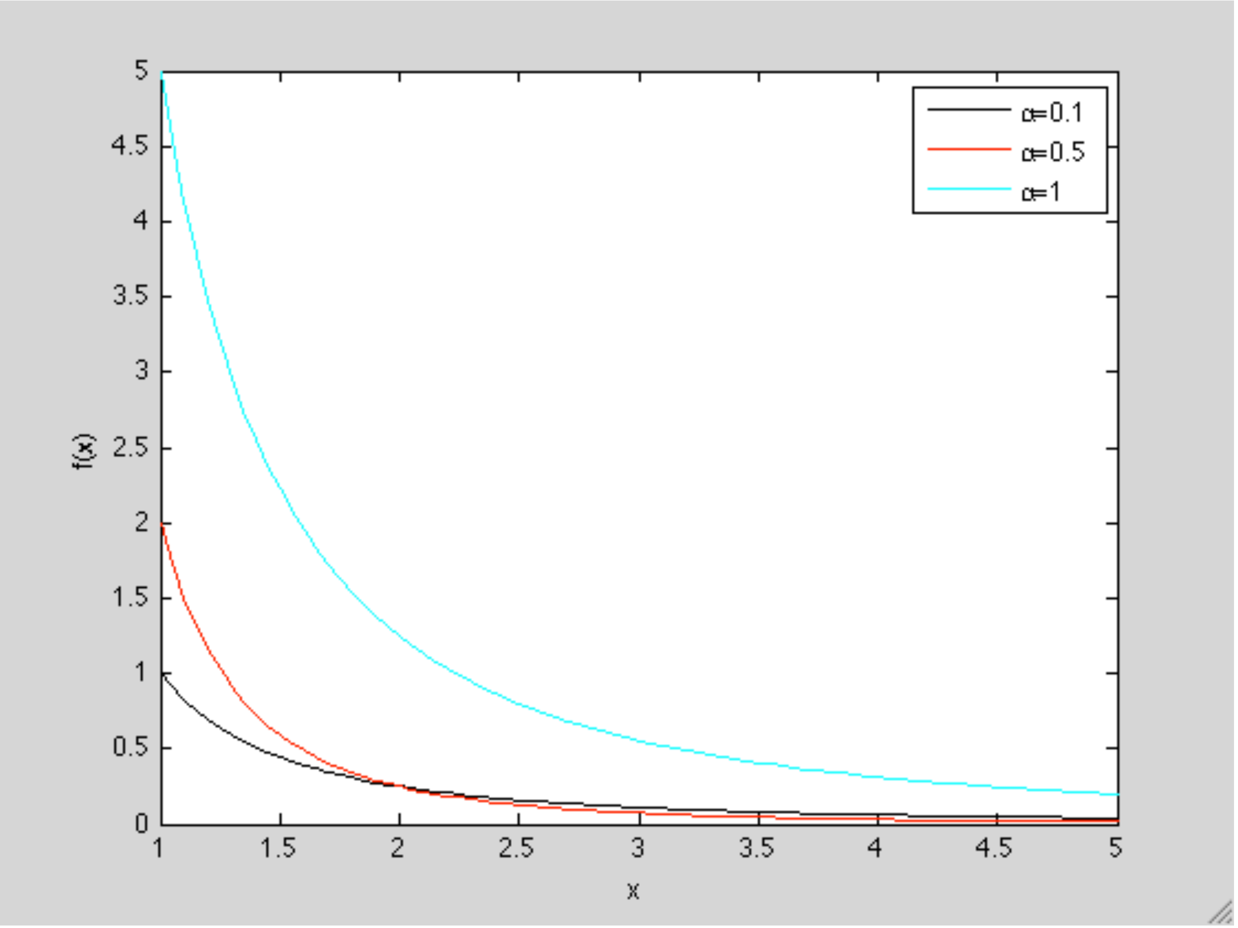


Continuous Random Variables (3)

- Pareto(α,x_m): $x>x_m$
  - ✓ $E(X)=\alpha x_m/(\alpha-1)$
  - ✓ $VAR(X)=E(X)\, x_m/(\alpha-1)(\alpha-2)$
  - ✓ Pareto=Power Law

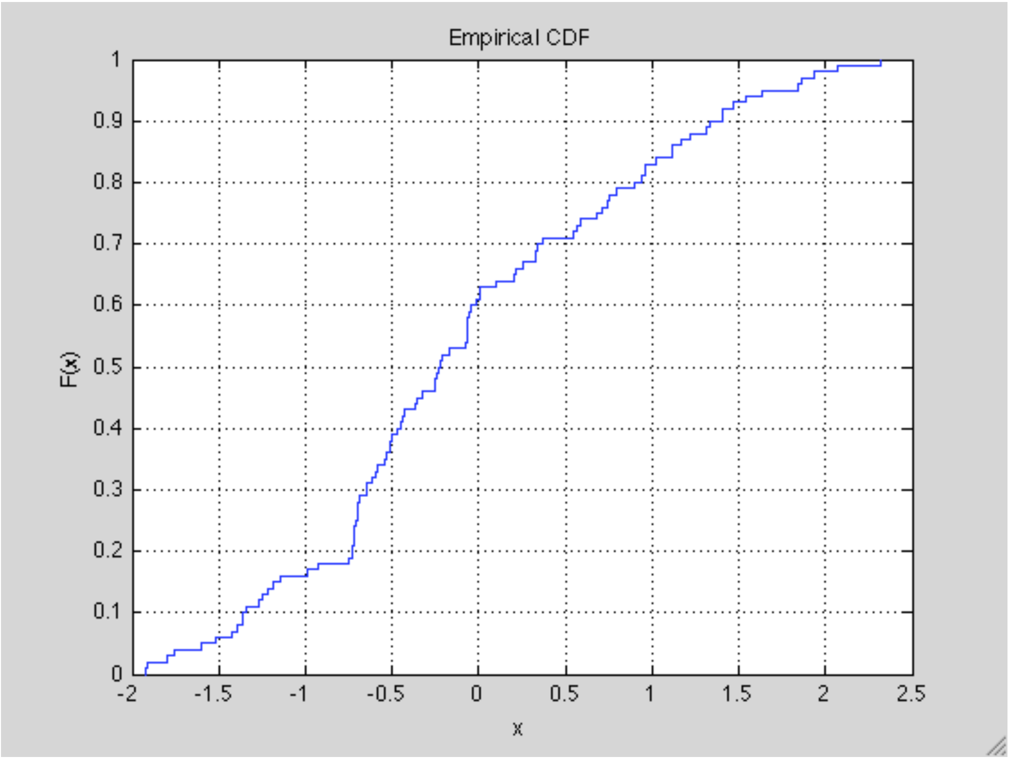
$$f(x) = \frac{\alpha x_m^\alpha}{x^{1+\alpha}}$$

- Power-law distributions
  - ✓ Scale-free: $f(kx)/f(x)=g(k)$
  - ✓ Fat upper tail (thicker than log-normal): much more observations with larger values (orders of magnitude)
  - ✓ If $\alpha<1$ the mean may be meaningless

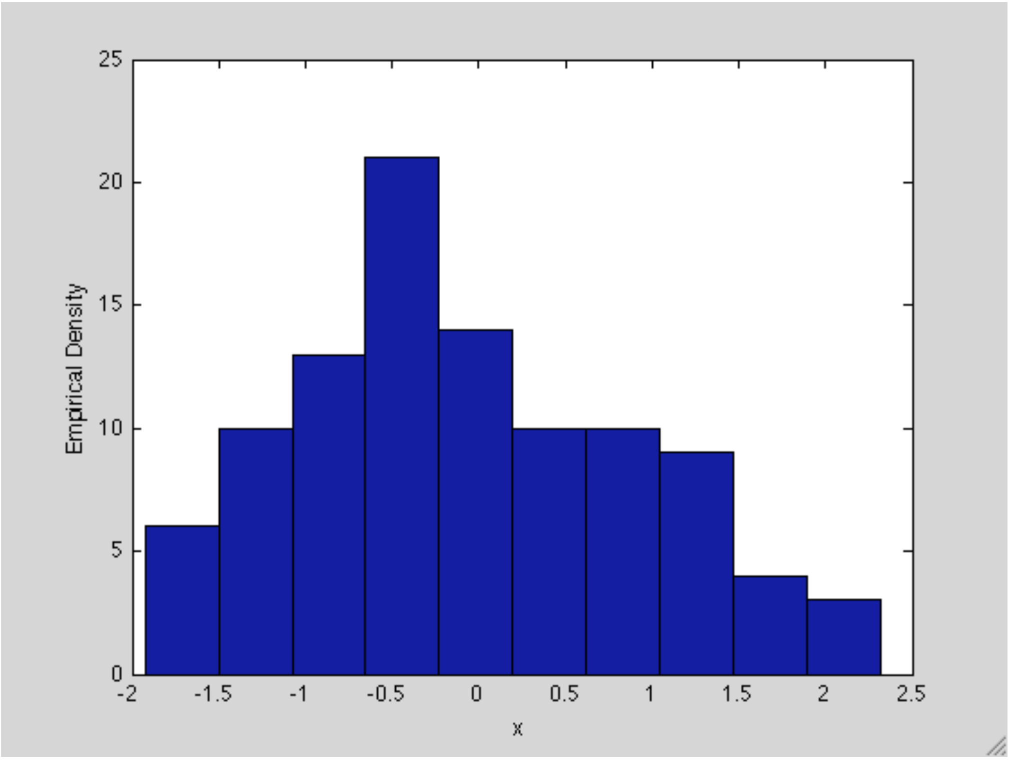


Empirical PDFs and CDFs

- Suppose to have a data-vector $X_N=\{x_1,..,x_N\}$ that we know comes from i.i.d. draws: how can we estimate the CDF of its generating RV?
  - ✓ Empirical CDF: $G(x)=Freq\{x_i \leq x\}$
- What about the PDF or density?
  - ✓ Binned histogram: Given a certain partition of the range $\{\min x_i , \max x_i\}$ (i.e. n equi-spaced intervals, say $\{a_k,a_{k+1}\}$) compute the frequency that data X_N fall in each interval



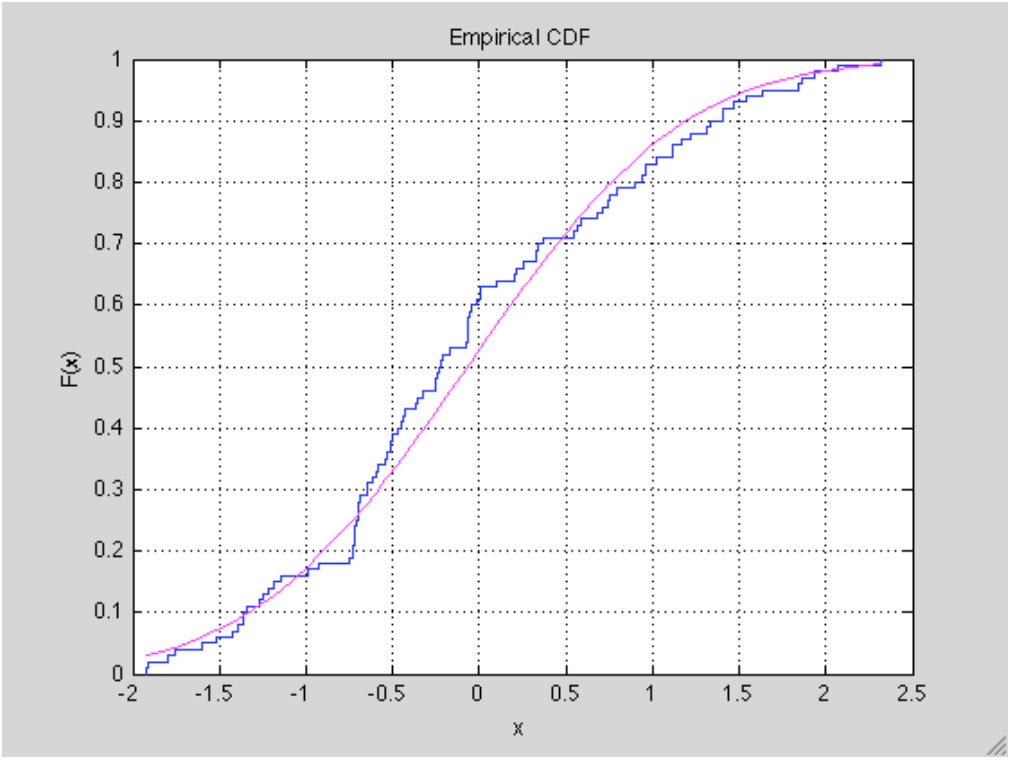
Empirical CDF from 100 values from a N(0,1)



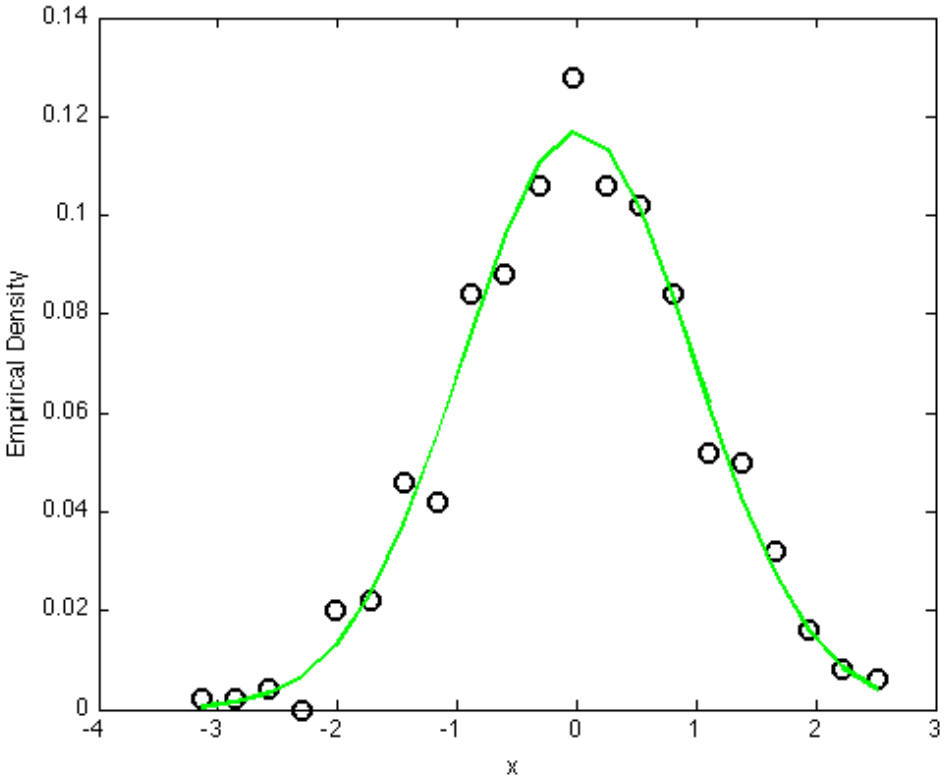
Empirical Density from 100 values from a N(0,1)

Fitting a Distribution to the Data

- Suppose to have a data-vector $X_N=\{x_1,..,x_N\}$ that we know comes from i.i.d. draws: how can we fit data with a known distribution?
  - ✓ Given $F(x,\theta)$, fit via ML the parameters using X_N
  - ✓ Plot ECDF or ED together with empirical data to assess visually GoF (see: stat tests…)
- What about the PDF or density?
  - ✓ Binned histogram: Given a certain partition of the range $\{\min x_i , \max x_i\}$ (i.e. n equi-spaced intervals, say $\{a_k,a_{k+1}\}$) compute the frequency that data X_N fall in each interval



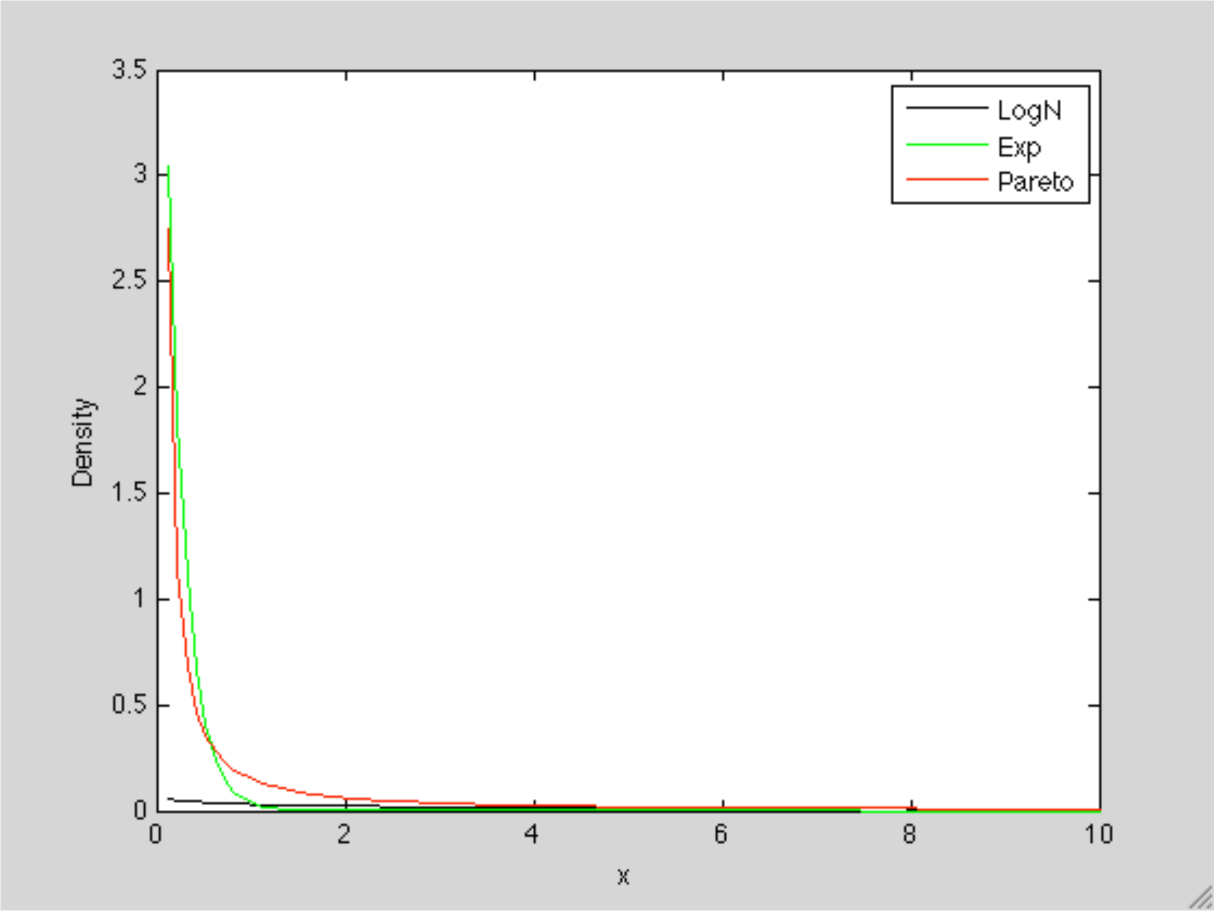
Empirical CDF vs. ML fit: 100 obs. from N(0,1)



Ed vs ML fit: 500 values from a N(0,1)

Skewed Distributions (1)

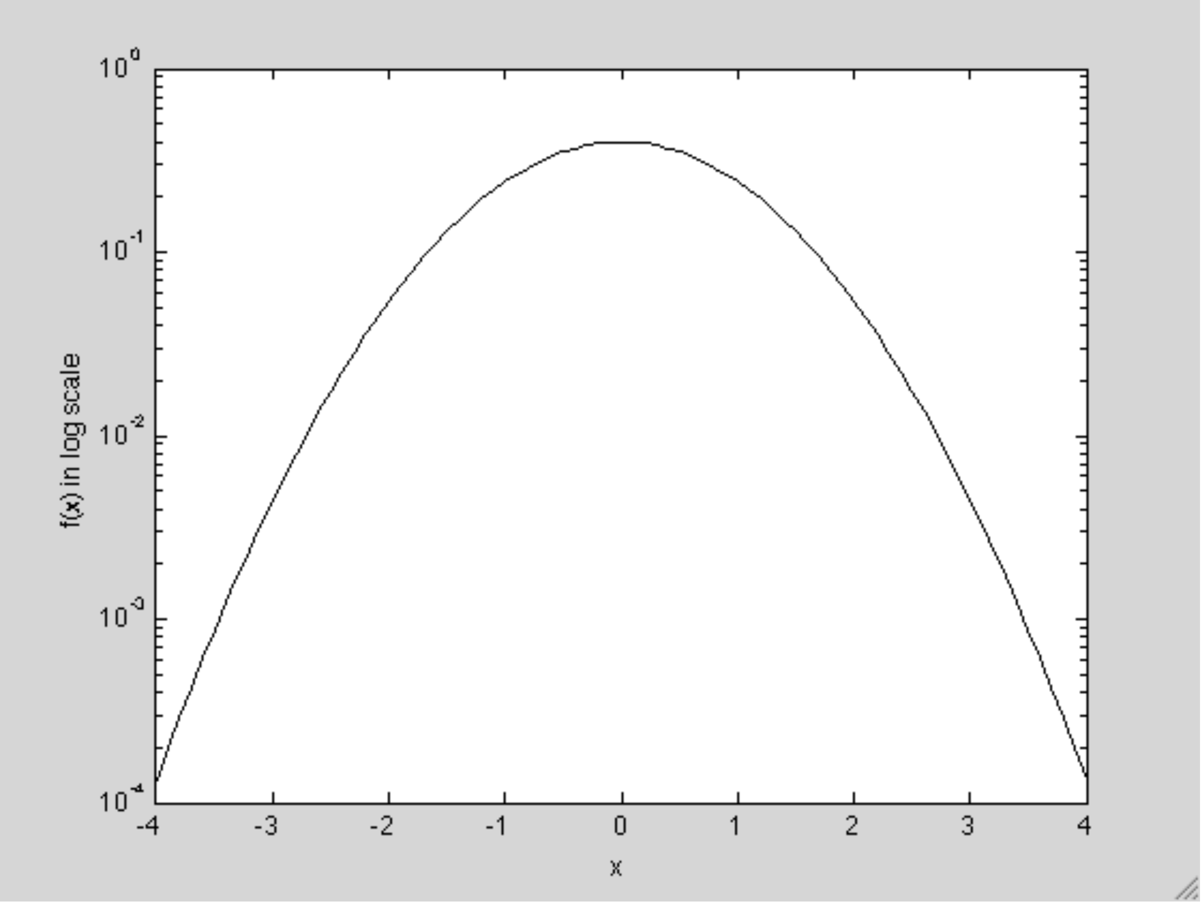
- Many positive-valued distributions are skewed to the left (exponential, log-normal, Pareto, etc.): Many observations with low x-values and a few of the upper (right) tail
- Pareto: 80-20 rule in personal income distribution
- How can we assess the extent to which two skewed distributions differ in their upper tail?



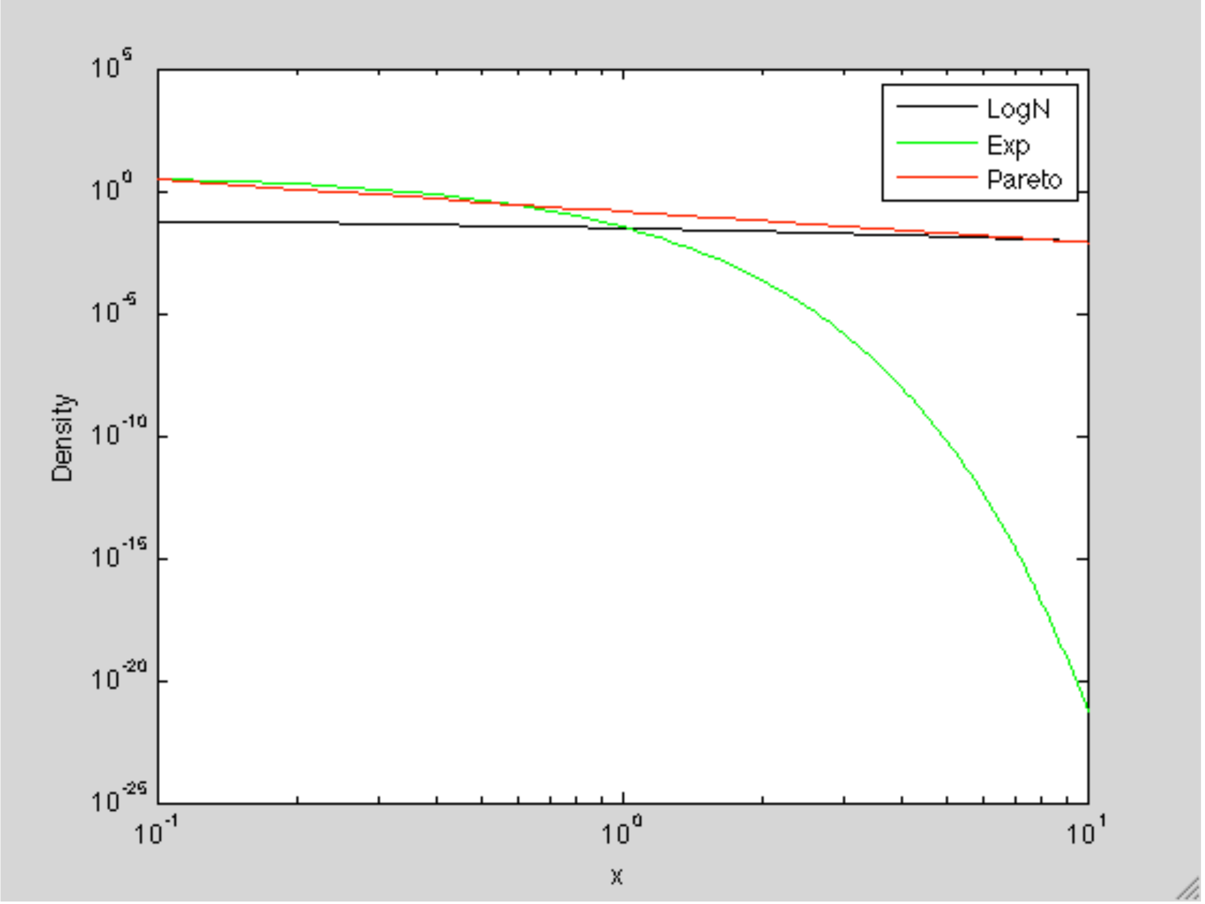
Three skewed distributions with (approx) same mean

Skewed Distributions

- Logarithmic transformations can help.
  - ✓ Linear-Log space (x,logy): Normal densities become parabolas; exponential densities become straight lines
  - ✓ Log-Log space (logx,logy): Pareto distributions are straight lines; log-normals have smoothly curved shapes



A normal density in the linear-log space



Same plot as in last slide but now in log-log space

Rank-Size (Zipf) Plot

- In a RSP we log-log plot the rank of an observation x_i vs its size. Suppose we are given $X_N=\{x_1,..,x_N\}$ and that $x_1 \geq \cdots \geq x_N$ (so that i is actually the rank of i). We have:

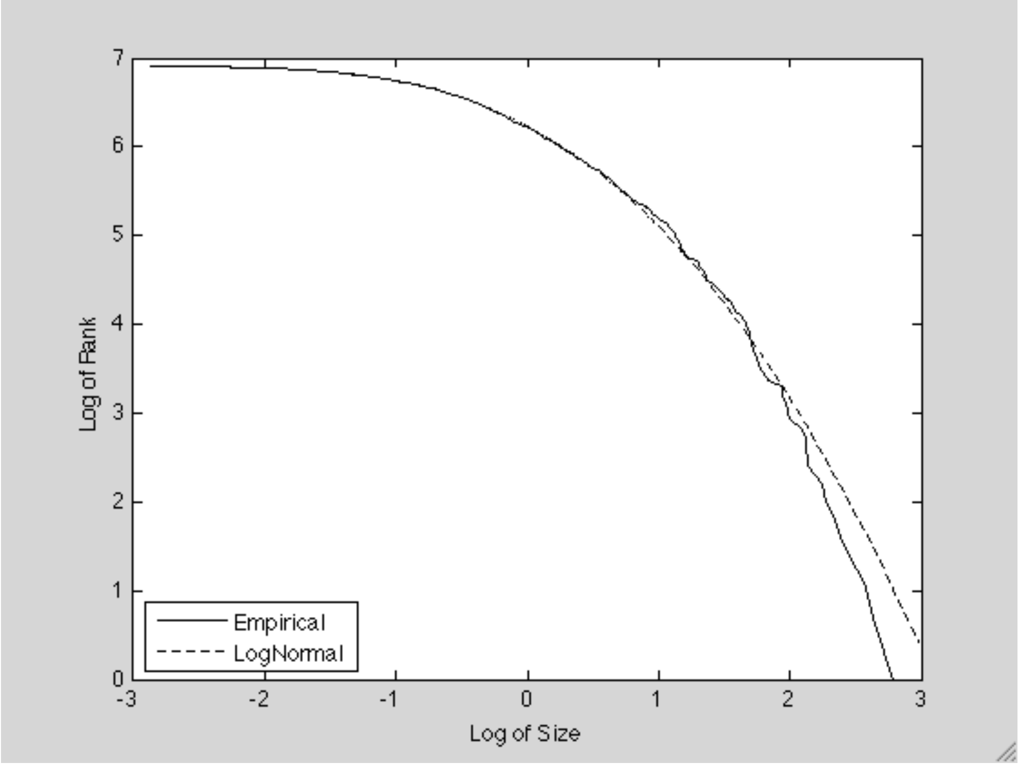
$$i = rank(i) = \#\{X \geq x_i\} \qquad \Longrightarrow \qquad \frac{i}{N} = Freq\{X \geq x_i\} = 1 - F(x_i)$$

Rank-Size (Zipf) Plot

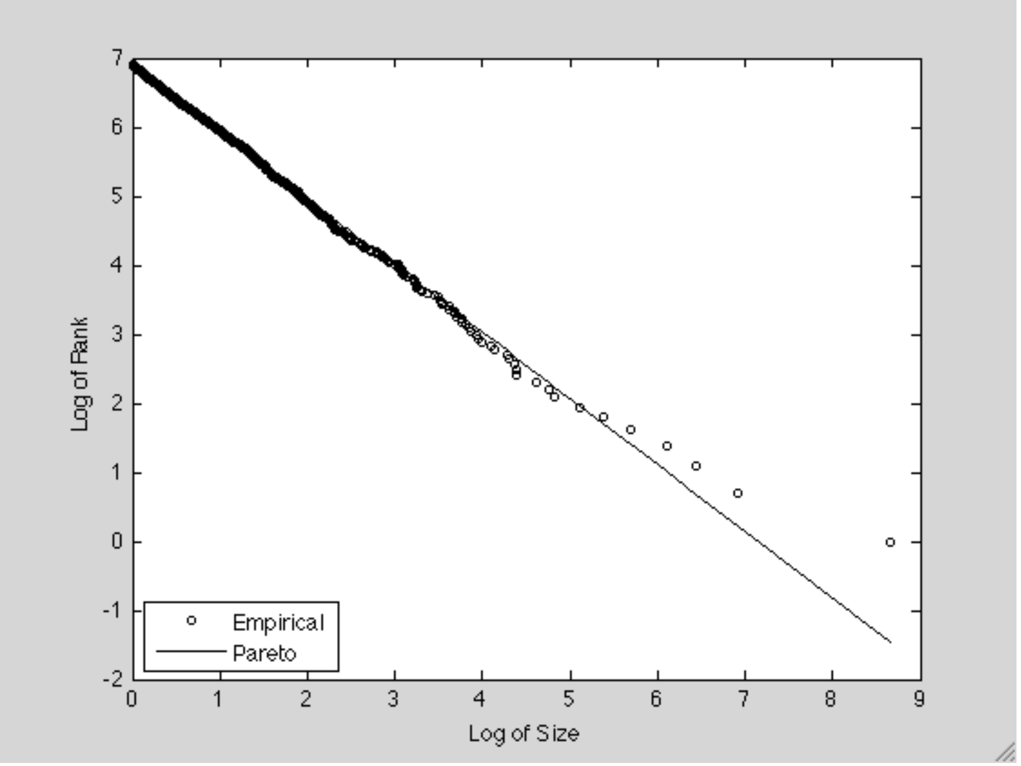
- In a RSP we log-log plot the rank of an observation x_i vs its size. Suppose we are given $X_N=\{x_1,..,x_N\}$ and that $x_1 \geq \cdots \geq x_N$ (so that i is actually the rank of i). We have:

$$i = rank(i) = \#\{X \geq x_i\} \qquad \Rightarrow \qquad \frac{i}{N} = Freq\{X \geq x_i\} = 1 - F(x_i)$$

- A log-log plot of i against x_i is thus a simple plot of log(1-F) against log(x). What it does is to magnify the upper tail and allow for easier comparisons between distributions (see Stanley et al., 1994)



Zipf plot of lognormal data with theoretical fit



Zipf plot of Pareto data with theoretical fit

Distributions in Economics

Distributions in Economics

- Standard assumption: “Everything is Gaussian”

Distributions in Economics

- Standard assumption: "Everything is Gaussian"
  - ✓ No need to look at moments higher than 2: mean/var are enough

Distributions in Economics

- Standard assumption: “Everything is Gaussian”
  - ✓ No need to look at moments higher than 2: mean/var are enough
  - ✓ Gaussian econometrics

Distributions in Economics

- Standard assumption: "Everything is Gaussian"
  - ✓ No need to look at moments higher than 2: mean/var are enough
  - ✓ Gaussian econometrics

- Many economic variables are NOT Gaussian

Distributions in Economics

- Standard assumption: "Everything is Gaussian"
  - ✓ No need to look at moments higher than 2: mean/var are enough
  - ✓ Gaussian econometrics

- Many economic variables are NOT Gaussian
  - ✓ Income, wealth, firm-size distributions, growth distributions at any scale, household consumption levels, etc.

Distributions in Economics

- Standard assumption: “Everything is Gaussian”
  - ✓ No need to look at moments higher than 2: mean/var are enough
  - ✓ Gaussian econometrics
- Many economic variables are NOT Gaussian
  - ✓ Income, wealth, firm-size distributions, growth distributions at any scale, household consumption levels, etc.
  - ✓ Fat and heavy tailed distributions may not have finite moments (cf. Pareto): variance (and sometimes mean) may be meaningless

Distributions in Economics

- Standard assumption: "Everything is Gaussian"
  - ✓ No need to look at moments higher than 2: mean/var are enough
  - ✓ Gaussian econometrics

- Many economic variables are NOT Gaussian
  - ✓ Income, wealth, firm-size distributions, growth distributions at any scale, household consumption levels, etc.
  - ✓ Fat and heavy tailed distributions may not have finite moments (cf. Pareto): variance (and sometimes mean) may be meaningless

- Non-Gaussian economics

Distributions in Economics

- Standard assumption: "Everything is Gaussian"
  - ✓ No need to look at moments higher than 2: mean/var are enough
  - ✓ Gaussian econometrics

- Many economic variables are NOT Gaussian
  - ✓ Income, wealth, firm-size distributions, growth distributions at any scale, household consumption levels, etc.
  - ✓ Fat and heavy tailed distributions may not have finite moments (cf. Pareto): variance (and sometimes mean) may be meaningless

- Non-Gaussian economics
  - ✓ Statistical studies: to characterize heterogeneity in economic distributions is not enough to look at mean and variance, we must know what distribution best fits the data

Distributions in Economics

- Standard assumption: "Everything is Gaussian"
  - ✓ No need to look at moments higher than 2: mean/var are enough
  - ✓ Gaussian econometrics
- Many economic variables are NOT Gaussian
  - ✓ Income, wealth, firm-size distributions, growth distributions at any scale, household consumption levels, etc.
  - ✓ Fat and heavy tailed distributions may not have finite moments (cf. Pareto): variance (and sometimes mean) may be meaningless
- Non-Gaussian economics
  - ✓ Statistical studies: to characterize heterogeneity in economic distributions is not enough to look at mean and variance, we must know what distribution best fits the data
  - ✓ Non-Gaussian econometrics: obtain estimators with non Gaussian errors (LAD regressions, etc.)

Distributions in Complex Networks

- If the network is large enough (many nodes, many links) , one can characterize heterogeneity in
  - ✓ Node-specific statistics: degrees, strengths, etc.
  - ✓ Link-specific statistics: link weights

Distributions in Complex Networks

- If the network is large enough (many nodes, many links) , one can characterize heterogeneity in
  - ✓ Node-specific statistics: degrees, strengths, etc.
  - ✓ Link-specific statistics: link weights

- Underlying homogeneity assumption
  - ✓ Node-specific observations x_i are i.i.d. draws from the same RV
  - ✓ Often not true, but cf. firm size and growth distributions, etc.: homogeneity assumptions are necessary in economics and (time-series) econometrics

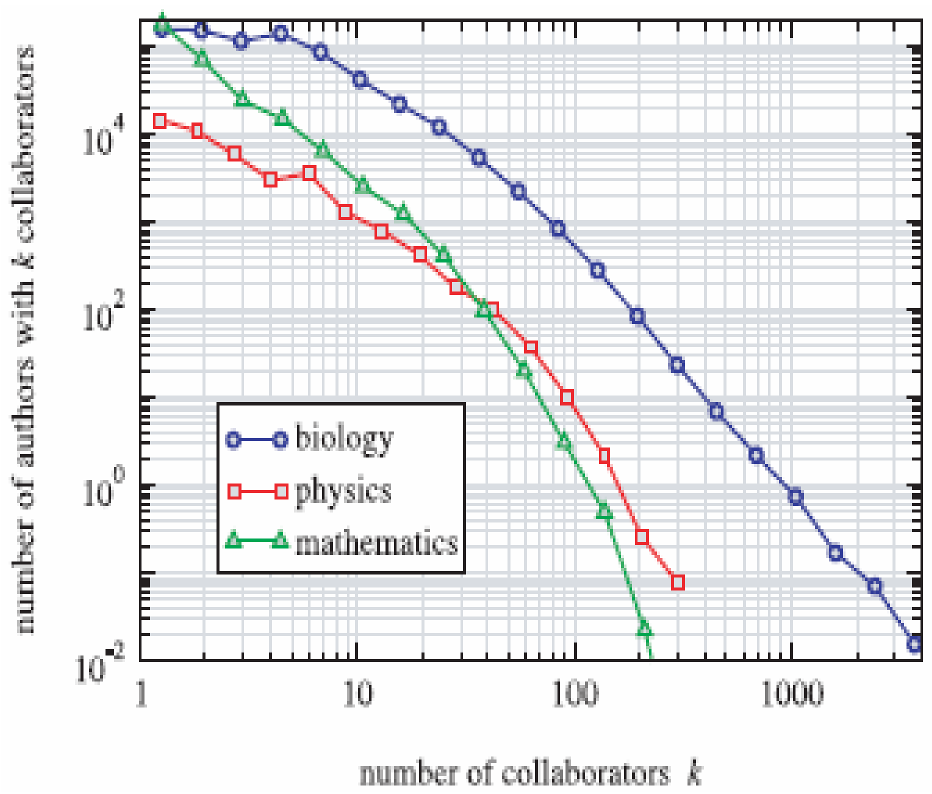
Distributions in Complex Networks

- If the network is large enough (many nodes, many links) , one can characterize heterogeneity in
  - ✓ Node-specific statistics: degrees, strengths, etc.
  - ✓ Link-specific statistics: link weights

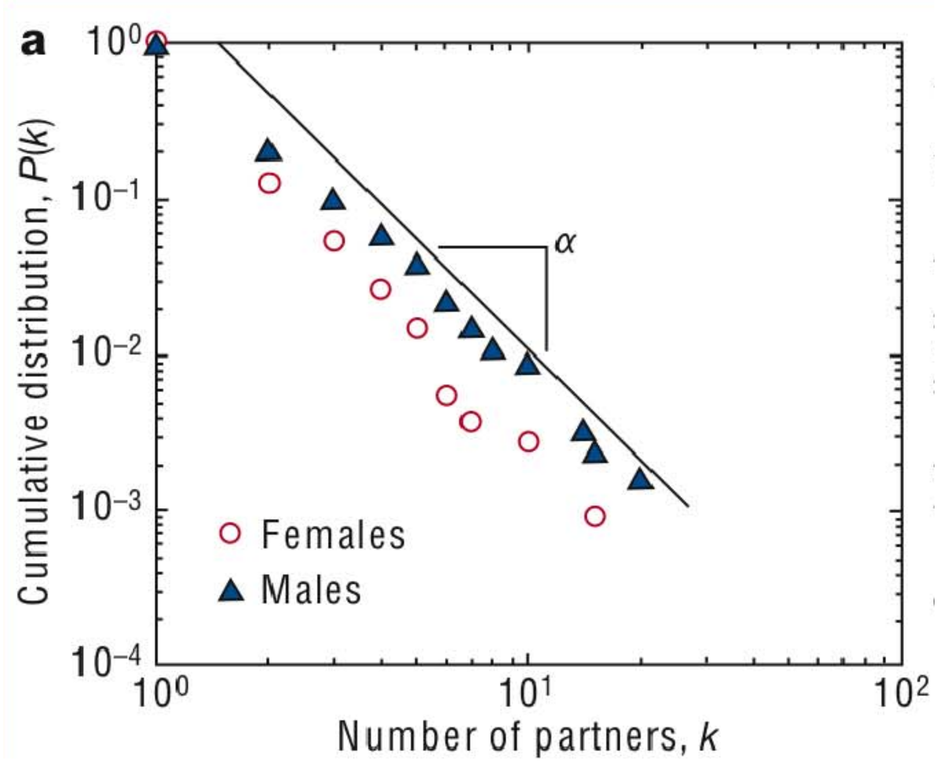
- Underlying homogeneity assumption
  - ✓ Node-specific observations x_i are i.i.d. draws from the same RV
  - ✓ Often not true, but cf. firm size and growth distributions, etc.: homogeneity assumptions are necessary in economics and (time-series) econometrics

- Stylized facts
  - ✓ ND distributions: Poisson (friendship networks, Dunbar's number) vs. Power-Law (Internet, the WWW)
  - ✓ Strength distributions: Log-normal, maybe with power law tail (ITN)

ND Distribution in Real Networks

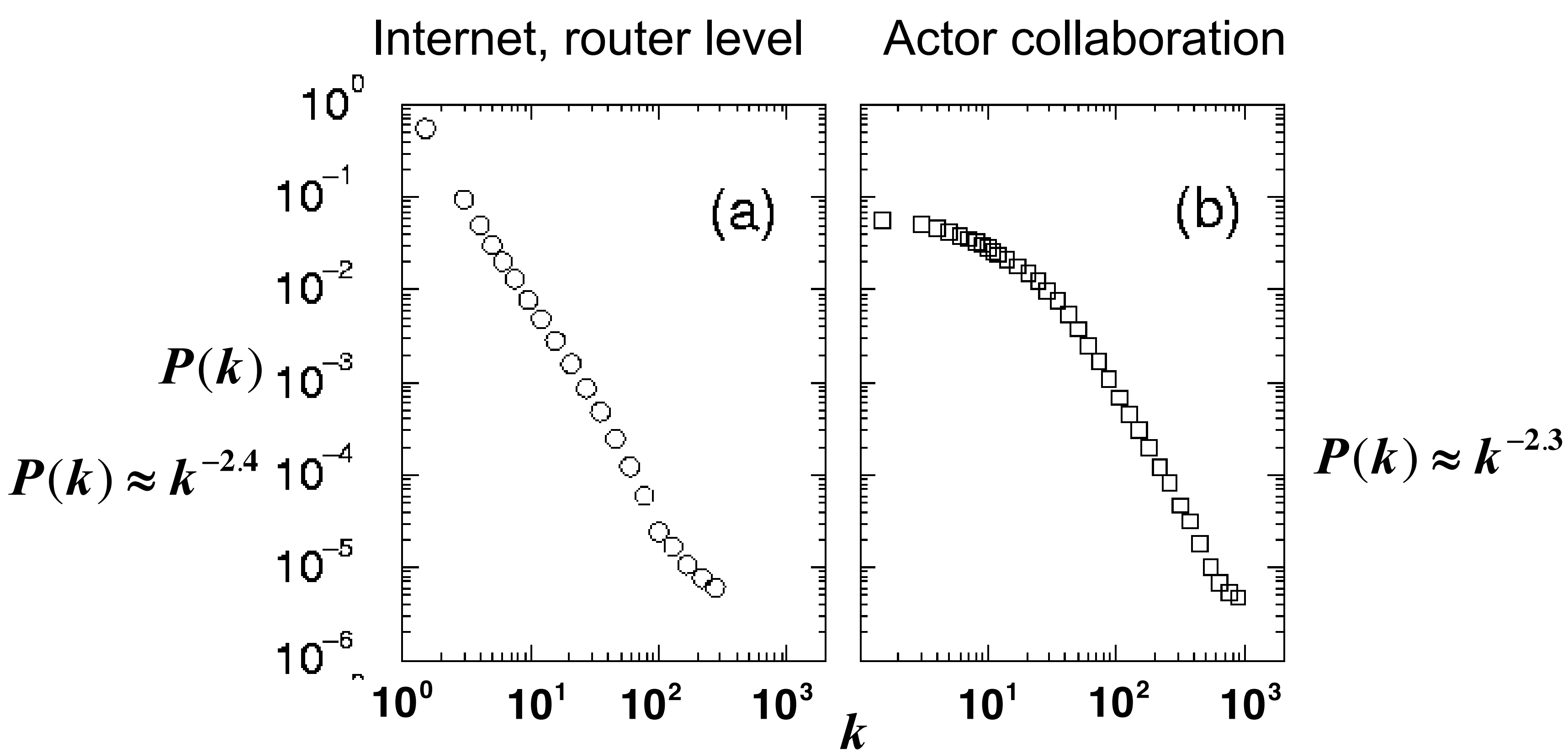


Co-Authorship Data (Newman, Grossman, 1999)



Sexual behavior in Sweden (Liljeros et al, 1999)

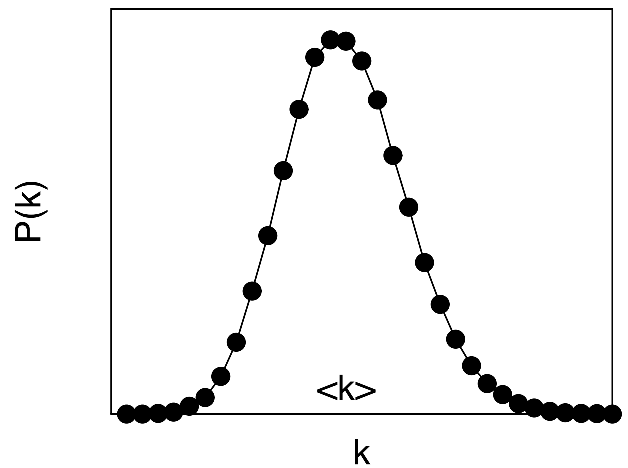
ND Distribution in Real Networks

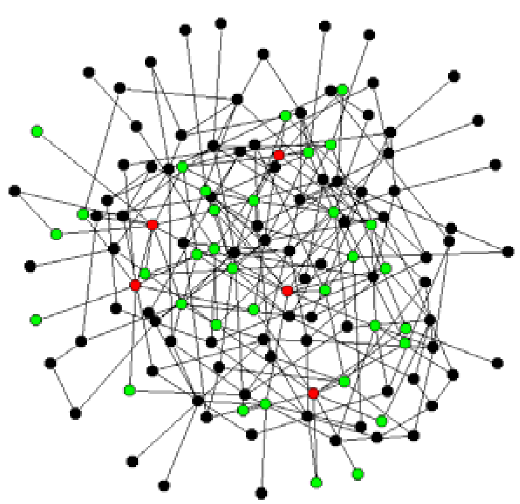


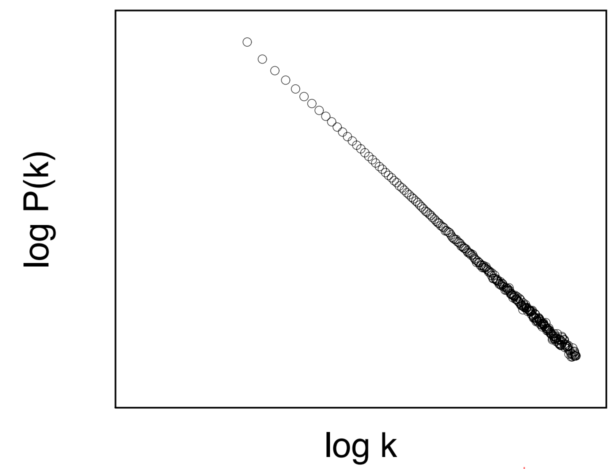
R. Govindan, H. Tangmunarunkit, IEEE Infocom (2000)

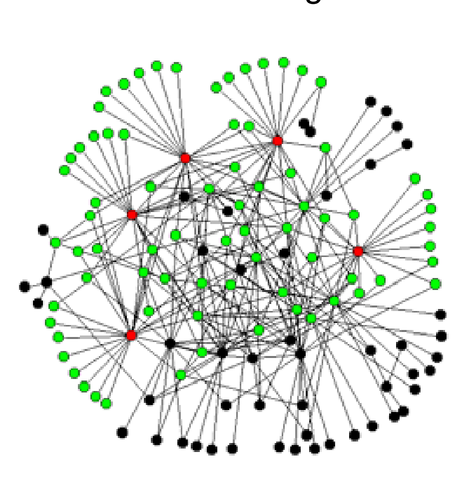
A.-L. Barabási, R. Albert, Science 286, 509 (1999)

ND Distribution and Network Topology









- The average degree gives the characteristic scale (value) of the degree
- All nodes are on average linked to the same number of other nodes

- Large variability, the average degree is not informative, no characteristic scale for the degree (scale-free)
- There are nodes (hubs) that are connected with a number of other nodes that is order-of-magnitudes larger than that of nodes in the left tail

Next Lecture

- What is a network? Examples of networks
- Why networks are important for economists?
- Networks and graphs
- Measures and metrics on networks
- Distributions of metrics and measures in large networks
- Models of network formation
- Null statistical network models
- Economic applications