Theory and Applications of Complex Networks

Classes Six–Eight

College of the Atlantic

David P. Feldman

30 September 2008

`http://hornacek.coa.edu/dave/`

A brief interdisciplinary primer on probability distributions and stochastic processes and then a bunch of stuff about power laws, what they are, what they mean, and what they don't mean

Why Probabilities?

Probabilities are used for somewhat different ends:

Compression:

- To compress a large series of data for a more compact representation.
- E.g., rather than listing the height of every COA student, just assume (or show) that the heights are Gaussian, and give the mean and standard deviation of the heights.
- Tradeoff: a more complicated distribution might fit the data better, but description would be larger, achieving less compression.

Why Probabilities?

Prediction:

- Given a set of observations, we wish to form a probabilistic description that predicts or describes future observations.
- E.g., the distribution of heights of current students could be used to predict the heights of next year's entering class.
- Tradeoff: a more complicated distribution might fit the data better, but it would *generalize* less well.
- In other words, there is a risk of *over-fitting* or fitting to noise.
- Predicting well may not be the same thing as "getting the model right."
- Prediction is, arguably, a form of understanding. But prediction does not necessarily tell us about mechanism or causality.

Why Probabilities?

Causality:

- Given a set of observations, we wish to form a probabilistic description that sheds light on the causes or mechanisms that produced the data.
- Here we are interested in the “right” or best model.
- Establishing this usually necessitates comparing different models, not just finding a good fit to one model.
- Challenge: this is often a difficult inferential task.
- Challenge: there may be very different mechanisms that produce a given probability distribution.

Where do Probabilities come from?

- “Nature does not give us probabilities or probability distributions, it gives us measurements.” (Chris Wiggins)
- Going from measurements to probability distributions is an essential part of statistics and statistical inference.

The Binomial Distribution

- Supposed we have a binary variable. Successive outcomes are uncorrelated and are A with probability p and B with probability $q = (1 - p)$.
- Such a set of random variables are referred to as IID: Identically and independently distributed.
- The probability that we observe n A's in N trials is:

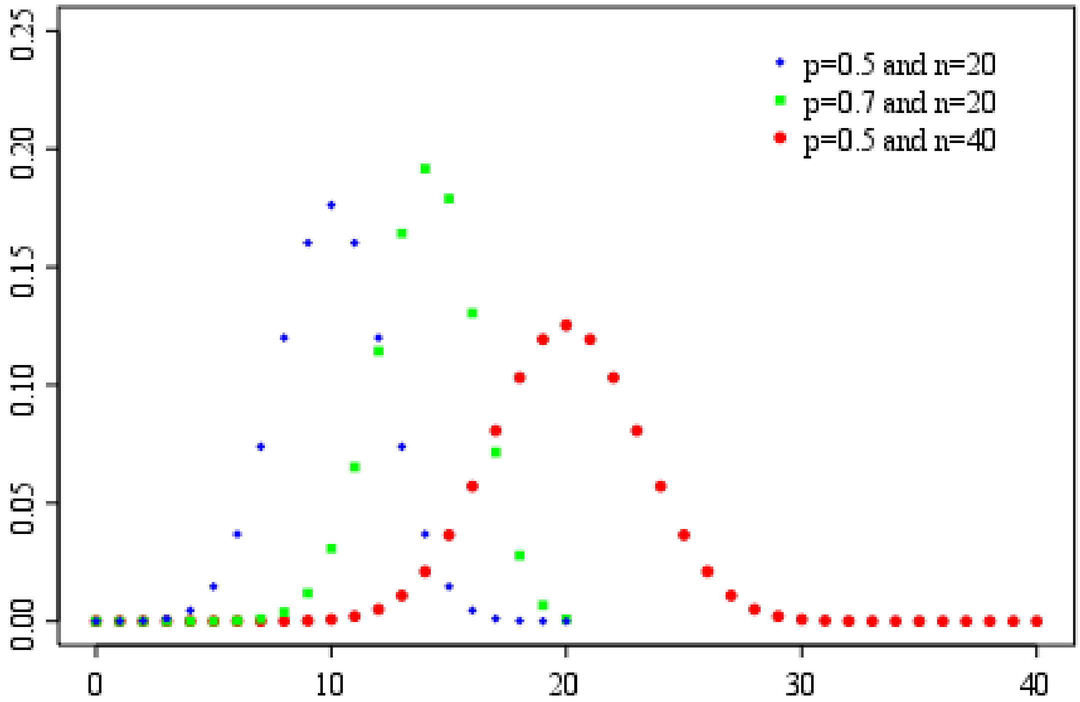
$$\mathrm{P}(n) = \binom{N}{n} p^n q^{n-k} \ . \qquad (1)$$

- This is known as the *binomial distribution*.
- The binomial coefficients are:

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \ . \qquad (2)$$

- This isn't a bad thing to memorize/internalize.

The Binomial Distribution



- Figure source: `http://en.wikipedia.org/wiki/Image:Binomial_distribution_pmf.svg`.
- The mean is np and the variance is $np(1-p)$.
- Standard deviation $= \sqrt{\text{variance}}$. (True for all distributions, not just binomial.)

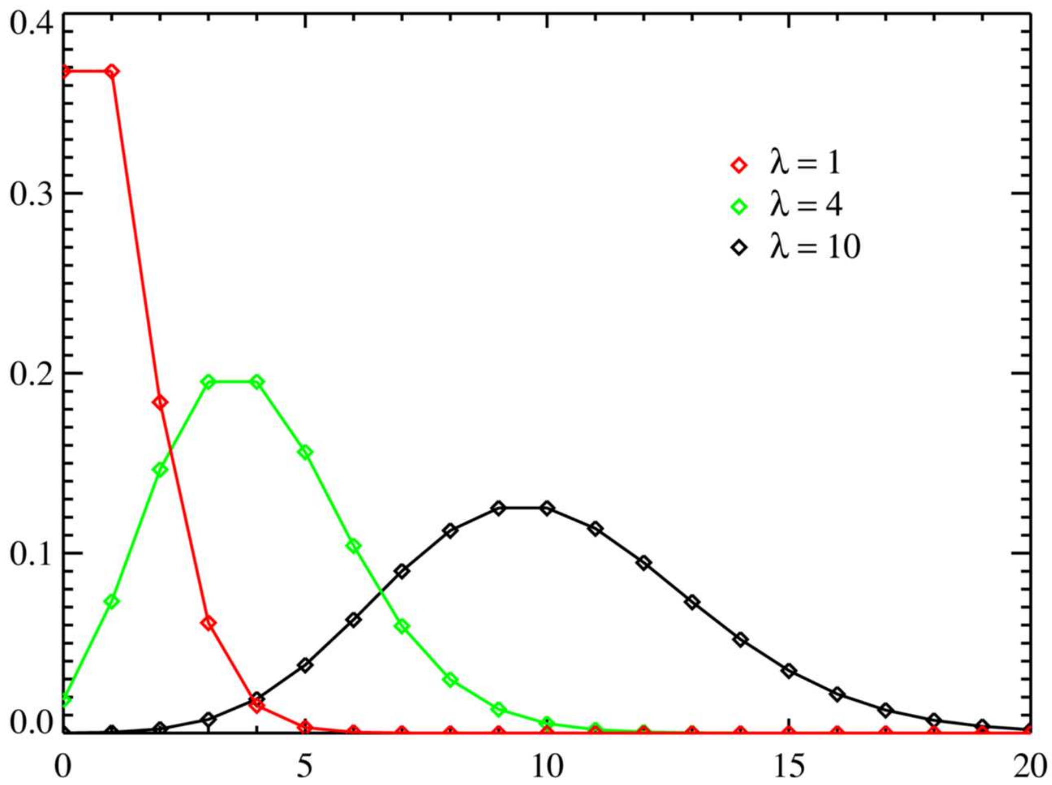
Poisson Distribution

- If Np stays fixed as N gets large, the Binomial distribution converges to the Poisson distribution.
$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!} . \quad (3)$$
where $\lambda = Np$ is the mean.
- For the Poisson distribution, the mean and variance are equal.
- Suppose there is a random event that occurs with fixed probability r per unit time.
- Then in a fixed amount of time, the number occurrence in a fixed time T is Poisson distributed with mean rT.

Poisson Distribution

- A vast number of phenomena are well approximated by Poisson distributions: distribution of roadkill on a highway, mutations in a strand of DNA, typos per page, calls per hour to a call center, number of raindrops falling over a spatial area, number of customers arriving to a store each hour.
- Poisson distributions are generated by a *Poisson process*: a stochastic process where events occur continuously and independently of each other.
- If the rate of occurrence is constant, then we get the Poisson distribution.
- In general, Poisson distributions are common in “counting” situations.
- Note that the Poisson distribution is for a discrete number of events. Those events may, or may not, occur continuously in time.

Poisson Distribution



- Note that the Poisson distribution has only one parameter, λ.

Normal Distributions

- For large n, the Binomial distribution converges to a Gaussian distribution:

$$\mathrm{P}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma}} \,. \tag{4}$$

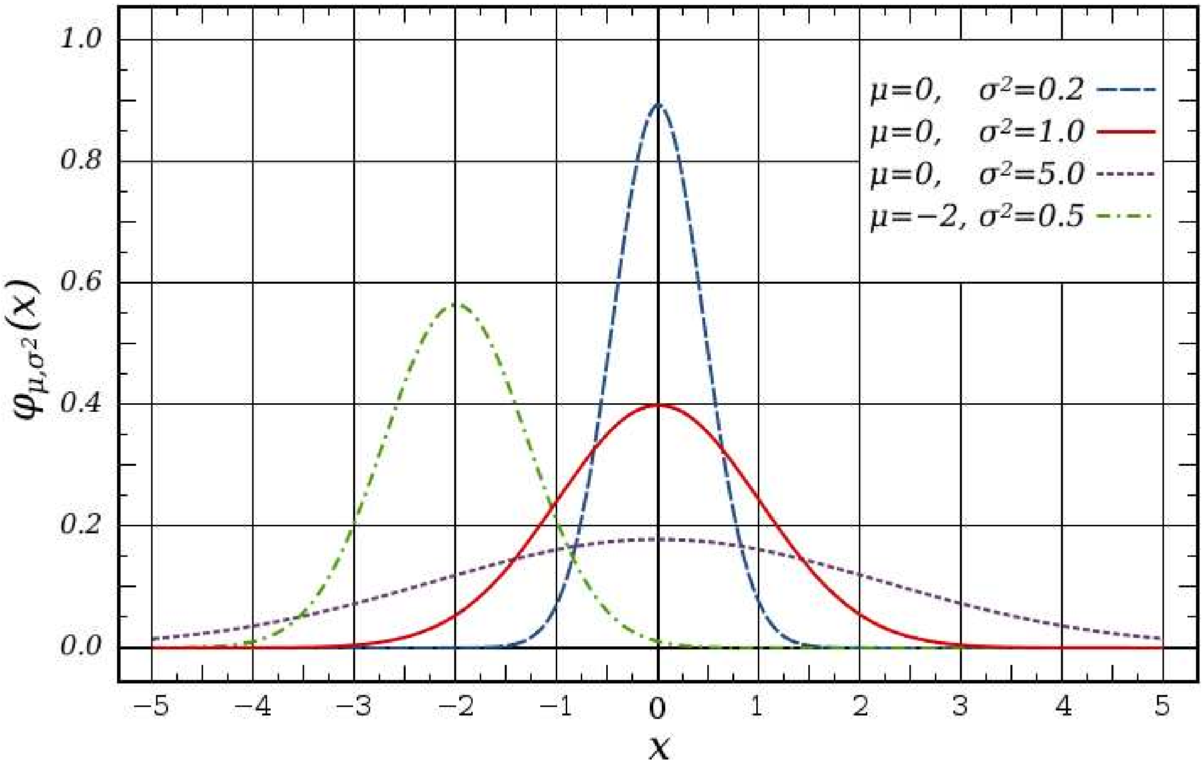
- This is a *continuous* distribution. Hence, $P(x)$ must be interpreted as a *probability density*. This is a different mathematical concept than a discrete probability!

- The mean of the distribution in μ.

- The variance is σ^2.

- The distribution has “thin tails.” The probability that x deviates from its average by more than $2\sigma, 3\sigma, 4\sigma, 5\sigma$ is, respectively: 0.042, 0.003, 6×10^{-5}, and 6×10^{-7}.

Normal Distribution



- A vast number of phenomena are distributed according to the normal distribution, even those which don't originate from a binomial distribution.
- Why?

Central Limit Theorem!

- A sum of random variables with finite mean and variance is Gaussian distributed, *regardless of the distribution of the random variables themselves!*
- Let $X_1, X_2, \ldots X_n$ be a set of N independent random variables.
- Let each X_i have mean 0 and a finite variance σ_i^2.
- Define a a new random variable to the the sum of N random variables:

$$X = \frac{1}{N} \sum_{i=1}^{N} X_i \; . \tag{5}$$

- Define the "mean of the means"

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \mu_i \; . \tag{6}$$

Central Limit Theorem

- And the mean squared error:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \ . \tag{7}$$

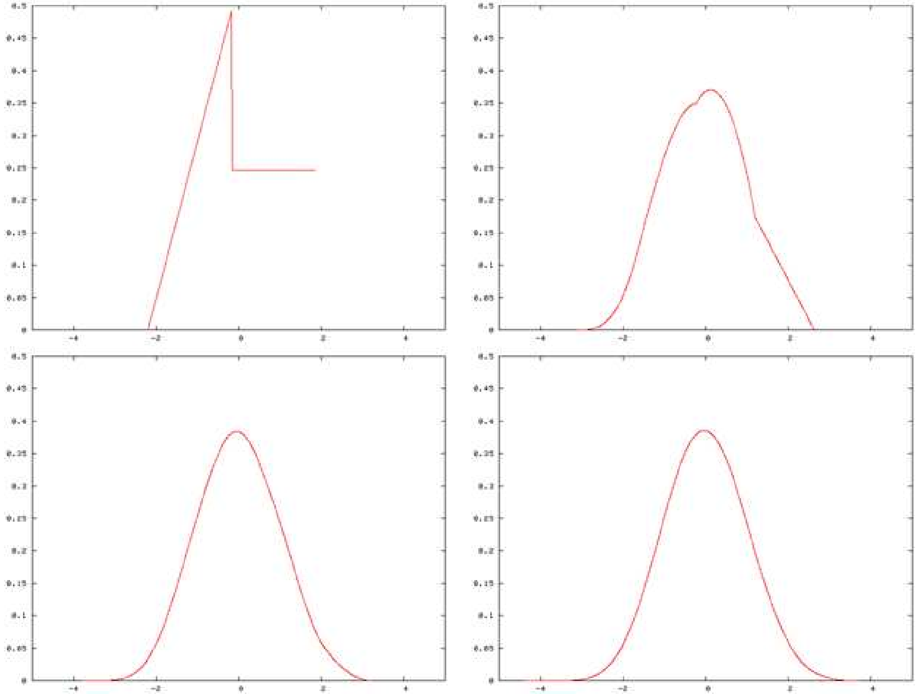
- Then, in the limit of large N, X is normally distributed:

$$\mathrm{P}(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma}} \ . \tag{8}$$

- There is some additional small mathematical print: the sum of the central third moments of the X_i's must be finite, and the cube root of this sum, divided by the square root of the mean squared error, must tend to zero. This is the Lyapunov condition.

- The basic result is that a sum of random variables has a Gaussian distribution, provided that the variance of the random variables is finite.

Illustration of CLT



- The first plot is the distribution of a variable X which is clearly not Gaussian.
- The second, third, and fourth plots are the distribution of $X + X$, $X + X + X$, and $X + X + X + X$.
- Note that the distribution of the sum is approaching a Gaussian.
- `http://en.wikipedia.org/wiki/Image:Central_limit_thm.png`

Gaussian Conclusions

- The Gaussian distribution is unarguably fundamental, and the central limit theorem is profound.
- The CLT is the foundation of much of statistics and inference.
- However, not everything is distributed according to a Gaussian.
- In particular, a variable may fail to be Gaussian if it is the result of a bunch of influences which
  - are multiplicative
  - have infinite variance
  - aren't independent

Discrete Exponential Distribution

- Suppose you are tossing a coin that comes up heads with probability p. Let $x =$ how many tosses before you see a heads.
- The random variable x is distributed according to: $P(x) = (1-p)^{x-1}p$.
- This is an example of a discrete exponential distribution:

$$P(x) = Ae^{-kx} \; . \tag{9}$$

- Exponential distributions arrive in waiting problems of this sort.
- Note that this is a discrete distribution, but that x is unbounded.

Continuous Exponential Distribution

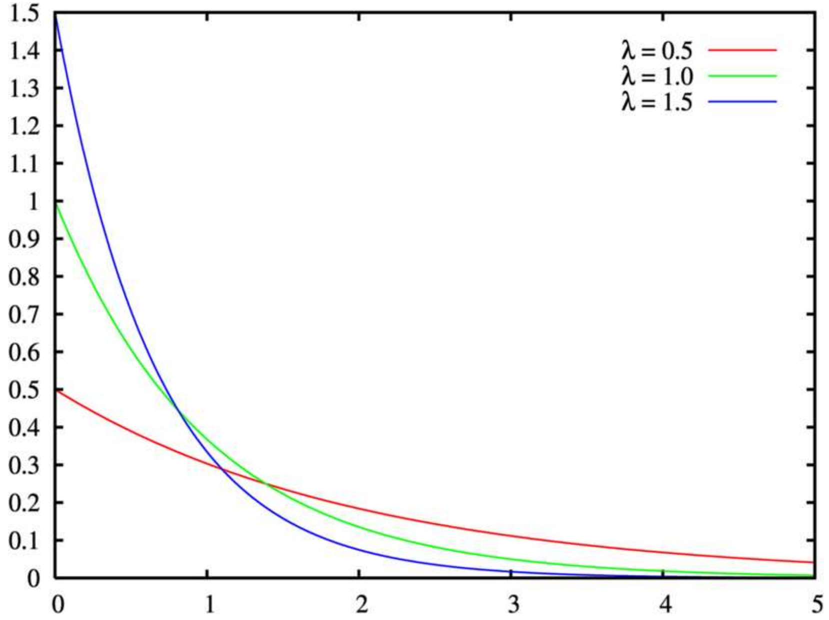
- Consider a Poisson processes in which events occur independently at a constant rate.
- The number of events in a fixed time is a discrete variable and is Poisson distributed.
- The waiting time T between events is a continuous variable and is exponentially distributed:

$$P(T) = Ae^{-kT} , \tag{10}$$

where $T > 0$.

- The mean of this distribution is $1/k$; the variance is $1/k^2$.

Continuous Exponential Distribution



- Many phenomena are distributed exponentially: survival time of many diseases, time between phone calls, distance between mutations on DNA, distance between roadkill on a road, atmospheric density as a function of height.
- Observing an exponential distribution is a clue that the events are independent, or memoryless.

(Central Limit Theorem and Statistics)

- Suppose you have two groups of cancer patients. You give an experimental treatment to one group but not the other.
- You wish to know if the average survival time is different between the two groups.
- So, you calculate the average survival time for each group and then compare them.
- Is the difference significant? In order to answer this question, you need to know how your average estimates are distributed.
- The central limit theorem says that they'll be normally distributed (in the large N limit).
- You can then use this fact to determine statistical significance.

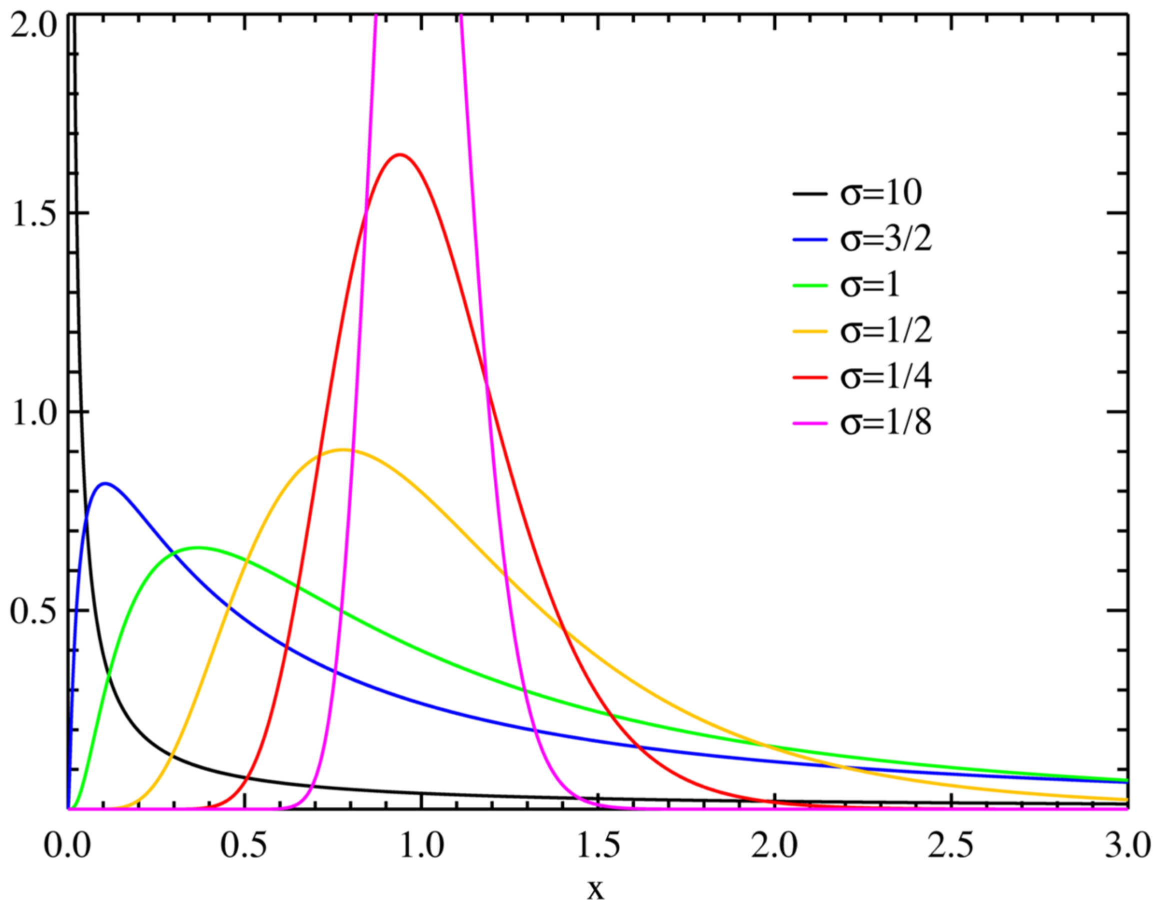
Log-normal distributions

- Gaussian distributions occur when we add up a bunch of random variable.
- If the random variables are somehow multiplied together, then we very often get a log-normal distribution:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} . \quad (11)$$

- In other words, $\ln(x)$ is normally distributed.
- The mean is $e^{\mu+\sigma^2/2}$ and the variance is $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$.
- The log-normal distribution arises from a product of independent, identically distributed random variables.
- The log-normal distribution is skewed and has a heavy tail.

Log-normal Distribution



- http://upload.wikimedia.org/wikipedia/commons/4/46/Lognormal_distribution_PDF.png

Log-normal Distribution

- Log-normals arise as the result of random multiplicative processes.
- Very many phenomena are well described by log-normals:
  - Rates of returns of stocks
  - Number of entries in people's email address books
  - Sizes of oil drops in mayonnaise
  - Latency periods of many diseases
  - The abundance of many species
  - Concentration of elements within a rock
  - Number of letters per sentence
  - Permeability of plant leaves
  - Size distribution of aerosols
- It is possible to mistake a log-normals for a power laws, and vice versa.

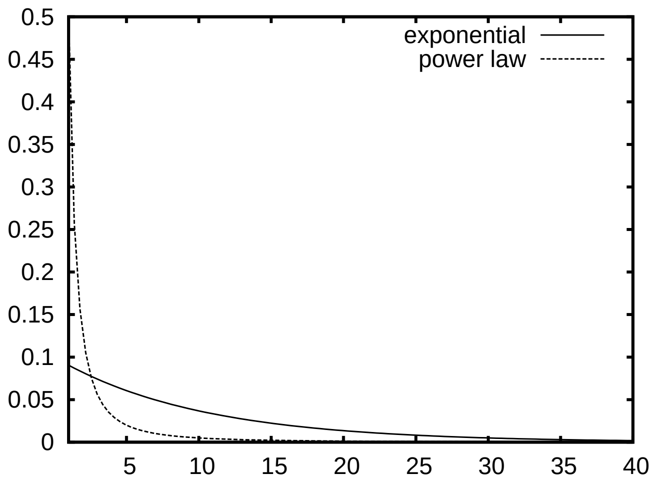
Power Laws

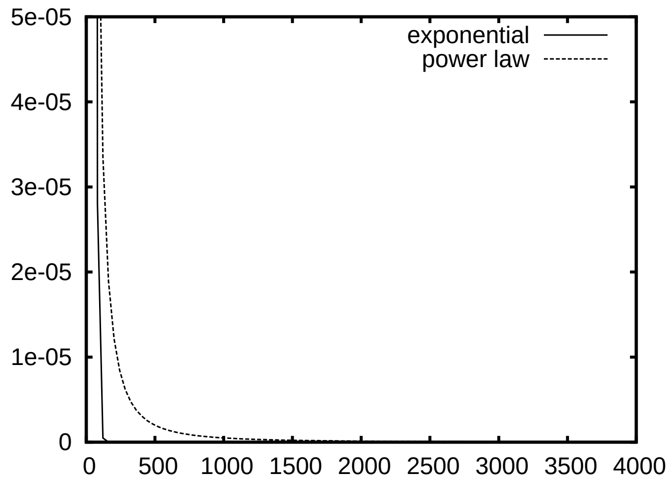
- A power law for a variable x is a distribution of the following form:

$$f(x) = Ax^{-\alpha} . \tag{12}$$

- The variable x could be discrete or continuous.
- Power-laws often arise in frequency plots, where x is the *rank* of a variable and $f(x)$ is the frequency.
- For example, $f(4)$ might be the size of the 4^{th} largest city.
- Power-law distributions are long-tailed.
- If $\alpha \leq 2$, the variance is infinite.
- If $\alpha \leq 1$, the mean is infinite.

The Long Tail





- $\alpha = 2$, mean of exponential $= 0.1$.
- The power law distribution has infinite variance.
- More about power laws in a bit.

Probability Distributions

- Two views of these distributions: something to fit to empirically and the limiting distribution of a stochastic process
- The distributions reviewed above are just the beginning. There are many other useful probability distributions.
- However, the ones I've presented are, in my experience, the ones you are most likely to encounter.
- (However, I probably should have mentioned the Weibull, or stretched exponential distribution, as this is also pretty important.)
- Also, there is no reason why the real world has to adhere to simple stochastic processes and distribution.

Power Laws: What's all the fuss about?

Why are power laws seen as very interesting objects of study. Why?

- They are long-tailed, and thus qualitatively different than Gaussians or exponentials or Poisson distributions. The probability of extreme events is much higher under a power law distribution.
- They are a particularly simple distribution, and so it is noteworthy when a complicated system is well approximated by a power law distribution.
- Power law distributions are scale free, or fractal.
- There are some interesting mechanisms that produce power laws.

Warning: Finding Power laws in Data.

- Finding power laws in empirical data can be tricky, and it is often done wrong.
- Estimating a slope from a log-log plot is *not* the best way to estimate the exponent.
- Instead, use a maximum likelihood estimator.
- Everything looks straighter on a log-log plot.
- Also, it is essential to also try fitting to other candidate distributions, e.g., the log-normal.
- Just because a power-law has a high R^2, meaning it describes a lot of the variance, doesn't mean that the distribution really is a power-law, since other distributions might give higher R^2's.

Empirical Power-law Warnings, continued

- See the excellent paper: Clauset, Shalizi, Newman: Power-law distributions in empirical data, `http://arxiv.org/abs/0706.1062`, 2007.
- See also Shalizi, `http://cscs.umich.edu/~crshalizi/weblog/491.html` and `http://cscs.umich.edu/~crshalizi/weblog/232.html`.
- The blog entries and the paper make excellent reading.
- Their comments about statistics and inference apply to much more than just power laws.

Why are Power Laws Scale-Free?

- The following closely follows Newman (2005).
- A power law distribution is the same, no matter what scale we look at it.
- E.g., we might find that blogs with 200 hits are day are 1/6 as common as blogs with 100 hits a day, and that blogs with 20,000 hits a day are 1/6 as common as blogs with 10,000 hits a day.
- A distribution $p(x)$ is scale free if, for all b and x:

$$p(bx) \;=\; g(b)p(x)\;. \tag{13}$$

- One can show that the *only* $p(x)$ that satisfies this is a power law: $p(x) = cx^{-\alpha}$.
- Thus, power law $\Longleftrightarrow$ scale free.

Generative Mechanisms for Power Laws

- Often one wants to use an observed distribution to infer something about the mechanism that generated that distribution.
- It turns out that there are many interesting ways that power laws can be generated.
- The following papers are excellent reviews:
  - M.E.J. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics 46, 323-351 (2005). `arXiv.org/cond-mat/0412004`
  - D. Sornette, Probability Distributions in Complex Systems, `arxiv.org/abs/0707.2194`.
  - M. Mitzenmacher, A Brief History of Generative Models for Power Law and Lognormal Distributions, Internet Mathematics, vol 1, No. 2, pp. 226-251, 2004.
  - Reed and Hughes, From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. Physical Review E 66:067103. 2002.

Mechanism 1: Exponentially Observing Exponentials

- Suppose a quantity is growing exponentially: $X(t) = e^{\mu t}$.
- Suppose we measure the quantity at a random time T , obtaining the value $\bar{X} = e^{\mu T}$.
- Let T also be exponentially distributed: $P(T > t) = e^{-\nu T}$.
- Then the probability density for $\bar{X}$ is given by $f_{\bar{X}}(x) = kx^{-\mu/\nu - 1}$.
- Like magic, a power law has appeared.
- In general, there are lots of ways to make power laws by combining exponential distributions in different ways.
- See Reed and Hughes, 2002.

Mechanism 2: The Yule Process

- First proposed by Yule in 1922 as an explanation for the distribution of the number of species in a genus.
- Subsequently rediscovered and extended.
- Goes by many names: cumulative advantage, the Matthew Effect, the Gibrat principle, rich-get-richer, and preferential attachment.
- Basic idea. An entity gets more (links, money, species) in proportion to the the number that it currently has.

The Yule Process, continued

- General formulation of the Yule process:
  - System consists of a collection of object with some property (e.g., population, number of links, etc.) measured by k
  - A new object appears with an initial value k_0.
  - There are m new connections/species/people added after each new object appears.
  - The quantity measured by k increases stochastically, in proportion to $c + k_0$.
- This model has three parameters: m, c, and k_0.
- The distribution of k can be solved for exactly. In the large k limit, it has a power law tail with an exponent of:

$$\alpha = 2 + \frac{k_0 + c}{m} . \tag{14}$$

Non-linear Preferential Attachment

- What happens if the probability of a new connection is proportional to k^{γ} instead of being exactly proportional to k.
- Krapivsky, Redner, and Leyvraz (PRL 85(21):4629, 2000) show that:
  - if $\gamma < 1$ the distribution of k is a stretched exponential and not a power law.
  - if $\gamma = 1$ the distribution of k is a power law, as before.
  - if $\gamma > 1$ the distribution of k is a “Winner-Take-All” in which one node ends up with an extremely large number of connections, while the rest are are distributed exponentially.
- Models which require fine tuning of the parameters to produce the desired result are often suspect.
- However, a stretched exponential is not always readily distinguishable from a power law.

Method 3: Continuous Phase Transitions

- At the critical point of many phase transitions, physical quantities such as the specific heat or magnetic susceptibility diverge.
- The origin of this divergence is the fact that at the critical point, there are “long-tailed” correlations across the system which decay as a power law.
- At the critical point, the system is scale-free.
- Away from the critical point, correlations decay exponentially.
- The specific heat and other quantities also obey a power law at the critical point.
- The exponents describing these power laws are *universal*, in the sense that there are only a few possible sets of exponents, and a particular set of exponents describe a very wide range of different systems.

Phase Transitions, continued

- The universality of continuous phase transitions is an amazing fact which is well understood theoretically and demonstrated experimentally.
- Thus, it is sometimes argued that power laws are evidence for a system in a critical state, poised between two different phases.
- Moreover, there are also efforts to find universal power laws: a wide range of systems described by the same exponent.
- Often, the underlying assumption to these attempts is the belief that power laws must arise as the result of long-range correlations or organization of some sort.
- But we've seen that this needn't be the case.
- Also, is there any argument for why things should be tuned to a transition point?

Method 4: Self-Organized Criticality

- There are a number of models which “tune themselves” to a critical state, and hence produce power-law distributed phenomena.
- In dynamical systems parlance, the critical point is an attractor.
- The original model can be thought of as a sandpile, in which grains of sand are added one-by-one. The subsequent distribution of avalanches follows a power law.
- An appealing aspect of SOC is that it does not require parameter fine tuning.
- However, others have argued that this is not the case.
- SOC has generated quite a bit of interest/hype.
- Self-organized critical models are interesting, but in my view are unlikely to explain more than a tiny fraction of the myriad power laws observed empirically.

Method 5: Optimization

- Mandelbrot in the late 1950's argued that power laws could arise from an optimization process for transmitting words.
- The following discussion follows Mitzenmacher (2004).
- Let p_j be the frequency of the $j\mathrm{th}$ most used word.
- Let C_j be the cost of transmitting the j^{th} word.
- Use $C_j \sim \log j$.
- What set of p_j's minimizes the average value of C_j?
- The answer to this question is a power law: $p_j \sim j^{alpha}$.
- There are other optimization processes that lead to power laws, e.g., highly optimized tolerance (J.M. Carlson and J. Doyle, PRL 84, 25292532. 2000.)

Method 6: Many More...

- There are a lots of other ways that one can form power law distributions.
- Many of these involve multiplicative processes of some sort.
- See the review articles by Newman, Mitzenmacher, and Sornette.

Power Laws Summary

- There are many simple, non-complex ways to make power laws.
- These mechanisms are very different from each other.
- They are not necessarily an indicator of complexity or correlation or organization.
- They are not necessarily an indicator of criticality–of a system on the edge of a phase transition.
- Many of the claims in the literature for the existence of power laws may be based on faulty data analysis.
- Power laws and other long-tail distributions are very common and important. Their properties are very different from Gaussian distributions, upon which much of our intuition is based.
- It seems to me that the rich-get-richer models are a good description/explanation of a wide range of phenomena.