Inference, Models and Simulation for Complex Systems — CSCI 70001003
Lectures 2 — 30 August 2011
Prof. Aaron Clauset

1 Power-law distributions

A power-law distribution is a special kind of probability distribution. There are several ways to define them mathematically. Here's one way, for a continuous random variable:

$$p(x) = Cx^{-\alpha} \qquad \text{for } x \geq x_{\min} \ , \tag{1}$$

where the normalization constant $C = (\alpha - 1)x_{\min}^{\alpha-1}$ is derived in the usual way. Note that this expression only makes sense for $\alpha > 1$, which is indeed a requirement for a power-law form to normalize.[1] As a more compact form, we can rewrite Eq. (1) as

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \qquad \text{for } x \geq x_{\min} \ , \tag{2}$$

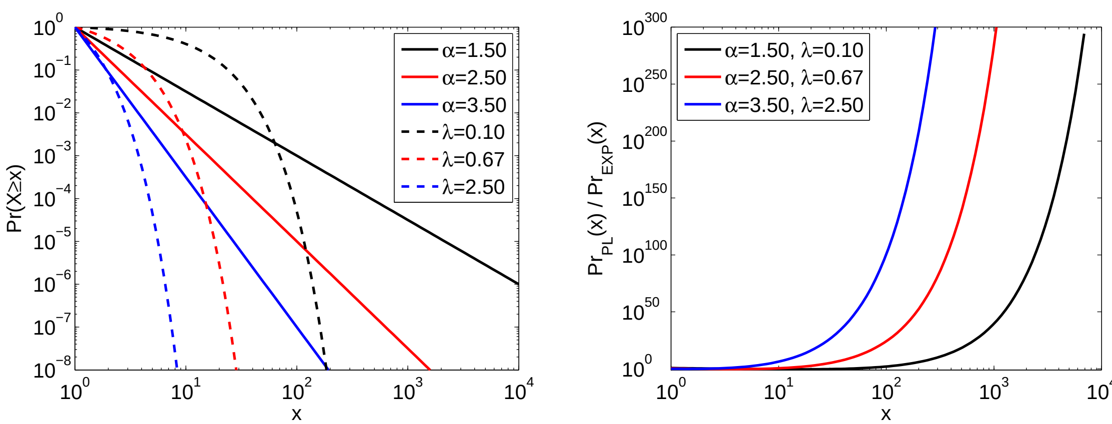

Figure 1: (a) Power-law and exponential distributions, for several choices of parameters. In both cases, $x \geq x_{\min} = 1$. (b) The ratio of power-law and exponential distributions, illustrating that events that are effectively "impossible" (negligible probability under an exponential distribution) become practically commonplace under a power-law distribution.

[1] Mathematically, the only way to have something that behaves like a power-law distribution but with a heavier tail than $\alpha \gtrsim 1$ is to effectively truncate its upper range, e.g., by adding an exponential cutoff in the upper tail like this: $\Pr(x) \propto x^{-\alpha} e^{-\lambda x}$.

1.1 What's weird about power laws

Many empirical quantities cluster around a typical value. The speeds of cars on a highway, the weights of apples in a store, air pressure, sea level, the temperature in New York at noon on Midsummer's Day. All of these things vary somewhat, but their distributions place a negligible amount of probability far from the *typical* value, making the typical value representative of *most* observations. For instance, it is an entirely useful statement to say that an adult male American is about 170cm (about 5 feet 7 inches) tall[2] because not one of the 200 million-odd members of this group deviate very far from this size. Even the largest deviations, which are exceptionally rare, are still only about a factor of two from the mean in either direction and hence the distribution can be well-characterized by quoting just its mean and standard deviation. In short, the underlying processes that generate these distributions fall into the general class well-described by the central limit theorem (recall Lecture 0).

Not all distributions fit this pattern, however, and in some cases the deviation is not a defect or problem, but rather an indication of interesting underlying complexity in the generating process. In particular, the past 15 years have produced countless examples of "non-normal" distributions from complex social, biological and technological systems. As we'll see in the next lecture, there are a large (and increasing) number of ways to produce "heavy-tailed" distributions. The most work has focused on power-law distributions because these have special mathematical properties and can be produced by interesting endogenous processes like feedback loops, self-organization, network effects, etc.

Power-law distributed quantities are not uncommon, and many characterize the distribution of familiar quantities. For instance, consider the populations of the 600 largest cities in the United States (from the 2000 Census).[3] Among these, the average population is only $\langle x \rangle = 165{,}719$, and metropolises like New York City and Los Angles seem to be "outliers" relative to this size. One clue that city sizes are not well explained by a Normal distribution is that the sample standard deviation $\sigma = 410{,}730$ is significantly larger than the sample mean. Indeed, if we modeled the data in this way, we would expect to see 1.8 times fewer cities at least as large as Albuquerque (population 448,607) than we actually do. Further, because it is more than a dozen standard deviations above the mean, we would never expect to see a city as large as New York City (population 8,008,278), and largest we expect to see in a sample of $n = 600$ cities would be Indianapolis (population 781,870).[4]

As a more whimsical second example, consider a world where the heights of Americans were distributed as a power law, with approximately the same average as the true distribution (which is

[2]See `http://en.wikipedia.org/wiki/Human_height#Average_height_around_the_world`

[3]See `http://www.demographia.com/db-uscity98.htm`

[4]The expected maximum size for a sample of n iid random variables drawn can be calculated by solving the following equation for $x_{\max}$: $\frac{1}{n} = \int_{x_{\max}}^{\infty} \Pr(x)\mathrm{d}x$. Do you see why this makes sense?

convincingly Normal when certain exogenous factors are controlled). In this case, we would expect nearly 60,000 individuals to be as tall as the tallest adult male on record, at 2.72 meters. Further, we would expect ridiculous facts such as 10,000 individuals being as tall as an adult male giraffe, one individual as tall as the Empire State Building (381 meters), and 180 million diminutive individuals standing a mere 17 cm tall.

In fact, this same analogy was used in 2006 to describe the counter-intuitive nature of the extreme inequality in the wealth distribution in the United States, whose upper tail is often said to follow a power law.[5]

1.2 Moments

In addition to cropping up as descriptions of many interesting quantities in social, biological and technological systems, power-law distributions have many interesting mathematical properties. Many of these come from the extreme right-skewness of the distributions and the fact that only the first $\lfloor \alpha - 1 \rfloor$ moments of a power-law distribution exist; all the rest are infinite. In general, the kth moment is defined as

$$\begin{aligned} \langle x^k \rangle &= \int_{x_{\min}}^{\infty} x^k \, p(x) \mathrm{d}x \\ &= (\alpha - 1)/x_{\min}^{\alpha-1} \int_{x_{\min}}^{\infty} x^{-\alpha+k} \mathrm{d}x \\ &= x_{\min}^k \left(\frac{\alpha - 1}{\alpha - 1 - k} \right) \qquad \text{for } \alpha > k + 1 \ . \end{aligned} \tag{3}$$

Thus, when $1 < \alpha < 2$, the first moment (the mean or average)is infinite, along with all the higher moments. When $2 < \alpha < 3$, the first moment is finite, but the second (the variance) and higher moments are infinite! In contrast, all the moments of the vast majority of other pdfs are finite.

A consequence of these infinite moments is that empirical estimates of those or nearby moments can converge very slowly due to the regular appearance of extremely large values. Figure 2 shows this numerically using synthetic data. When a moment doesn't exist, the sample estimate grows with sample size n. But, even when the appropriate moment does exist, the sample estimates vary a lot (remember, these data are shown on logarithmic scales), especially for small values of n, and converge very slowly on the true value.

[5] See `http://www.theatlantic.com/magazine/archive/2006/09/the-height-of-inequality/5089/` . The upper tail of the wealth distribution does not in fact follow a perfect power law because there are statistically significant deviations in it, which appear because the wealth of individuals are not iid random variables.

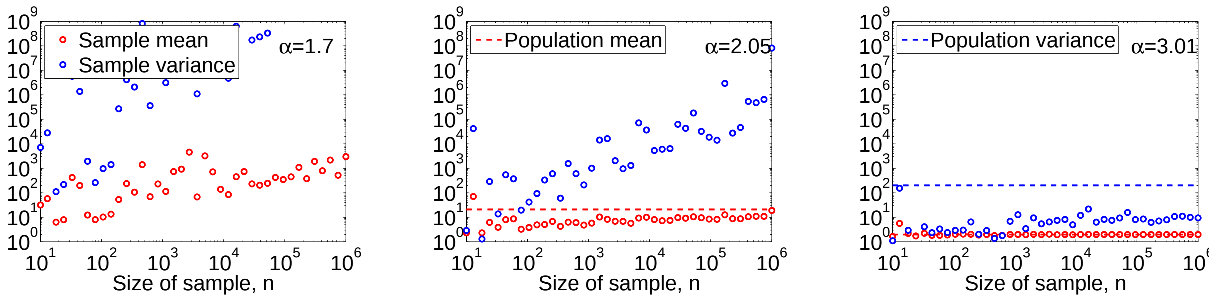

Figure 2: The sample mean and variance for power-law distributions with $\alpha = \{1.7, 2.05, 3.01\}$, for a wide range of sample sizes n. For each value of n, the mean and variance estimates are for the same set of synthetic observations. See Section 2 for Matlab code for these figures.

1.3 Scale invariance

Another interesting property of power-law distributions is "scale invariance." If we compare the densities at $p(x)$ and at some $p(c\,x)$, where c is some constant, they're always proportional. That is, $p(c\,x) \propto p(x)$. This behavior shows that the relative likelihood between small and large events is the same, no matter what choice of "small" we make. That is, the density "scales." Mathematically:

$$\begin{aligned} p(c\,x) &= (\alpha - 1)x_{\min}^{\alpha-1}(c\,x)^{-\alpha} \\ &= c^{-\alpha}\left[(\alpha - 1)x_{\min}^{\alpha-1}x^{-\alpha}\right] \\ &\propto p(x) \ . \end{aligned}$$

Further, it can be shown[6] that a power law form is the *only* function that has this property.

Here's another way of seeing this behavior. If we take the logarithm of both sides of Eq. (1), we get an expression for $\ln p(x)$ that's linear in $\ln x$. That is,

$$\begin{aligned} \ln p(x) &= \ln\left[(\alpha - 1)x_{\min}^{\alpha-1}(x)^{-\alpha}\right] \\ &= \ln C - \alpha \ln x \ . \end{aligned}$$

That is, rescaling $x \to c\,x$ simply shifts the power law up or down on a logarithmic scale. This shows another of the more well-known properties of a power-law distribution: it's a straight line on a log-log plot. This is in contrast to the strongly curved behavior of, say, an exponential distribution, as in Fig. 1.

[6] An exercise left to the reader.

1.4 Top-heavy distributions and the 80–20 "rule"

The extreme right-skewness of power-law distributions also implies some other interesting behaviors. For instance, assume that the distribution of wealth is power-law distributed with some parameter α (which, it turns out, is not a terrible assumption). What fraction W of the total wealth is held by the richest fraction P of the population?

The fraction P of the population whose wealth is at least x is given by the complementary cdf:

$$P(x) = \int_x^\infty C\, y^{-\alpha} \mathrm{d}y = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1} \quad , \tag{4}$$

where $C = (\alpha - 1)x_{\min}^{\alpha-1}$, as above. And the fraction wealth held by those people is given by:

$$W(x) = \frac{\int_x^\infty y\, p(y)\mathrm{d}y}{\int_{x_{\min}}^\infty y\, p(y)\mathrm{d}y} = \left(\frac{x}{x_{\min}}\right)^{-\alpha+2} \quad , \tag{5}$$

where $\alpha > 2$. Solving Eq. (4) for $x/x_{\min}$, and substituting the result into Eq. (5) produces an expression that does not depend on x

$$W = P^{(\alpha-2)/(\alpha-1)} \quad , \tag{6}$$

Fig. 3 shows how skewed or "top heavy" the distribution of wealth can be for several different choices of α, along with similar Lorenz curves for an exponential distribution, for comparison.[7]

This extreme top-heaviness is sometimes called the "80–20 rule," meaning that 80% of the wealth is in the hands of the richest 20% of people. However, as $\alpha \to 2$ this asymmetry gets progressively more extreme, with a smaller and smaller fraction of the population holding a greater and greater proportion of the total wealth. When $\alpha < 2$, the integrals in our calculation above diverge and the total wealth is almost completely held by a single person, i.e., the sum of all the wealth is largely equal to the largest value in the sum.[8]

Gini coefficients are a common way to measure just how skewed the distribution is, and are often quoted by people talking about wealth inequality, and are derived from Lorenz curves. The Gini

[7] If we assume that "wealth" is instead an exponentially distributed random variable, and if repeat the steps to derive the corresponding Lorenz curve, we find $W = P\left(1 - [1 + \lambda x_{\min}]^{-1} \ln P\right)$.

[8] There is an entire branch of theoretical statistics called *extreme value theory* that is devoted to studying the asymptotics of such situations. Within extreme value theory, power-law distributions have special significance because of their infinite moments. Extreme value theory is frequently used in the (re-)insurance industry, where rare but catastrophic losses can lead to huge fluctuations in the profits (and reserves) of insurance companies. It's worth pointing out that most results from this field hold asyptotically, and thus it is not always clear that they hold for the finite-sized samples we will typically encounter in studying complex systems.

coefficient G is defined as twice the area between the observed $W(P)$ function and the "ideal" $W = P$ function (everyone having an equal portion of the total wealth). Since the maximum area between the two curves is $1/2$ (when one individual holds all the wealth), $G \in [0, 1]$ with larger values indicating more skewed distributions. The Wikipedia page for Gini coefficients[9] has a nice map, derived from the CIA *World Fact Book 2009*, showing Gini coefficients for most countries worldwide.

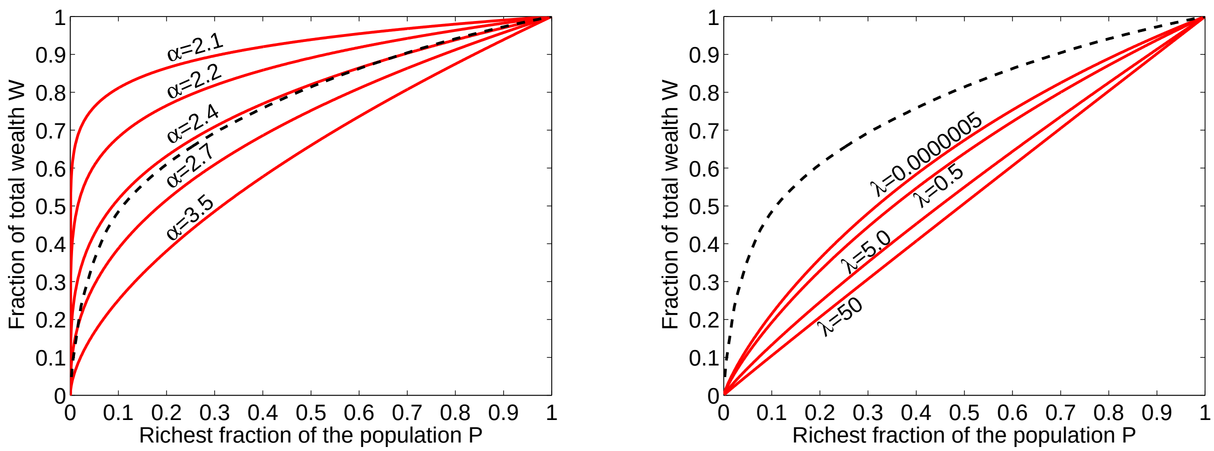

Figure 3: Lorenz curves (after Max Otto Lorenz, 1880–1962, an American economist) for several different power-law (left) and exponential (right) wealth distributions. The dashed line shows the empirical Lorenz curve for the wealthiest individuals in the United States (data from the Forbes 400, 2003). For the exponential distribution, these curves are for $x_{\min} = 1$; setting $x_{\min} = 600{,}000{,}000$, which is the smallest value in the Forbes data, yields a flat line $W = P$; see footnote 2.

1.5 Power-law tails

Equation (2) describes a pdf that follows a power law over its entire range. But some distributions may only exhibit a power law in their *tail*, i.e., when x is sufficiently large. Generally, such distributions can be expressed in the form $\Pr(x) = L(x)\, x^{-\alpha}$, where $L(x)$ represents a "slowly varying function," i.e., as $x \to \infty$, $L(x) \to c$, where c is some constant, and $p(x) \to x^{-\alpha}$.

For instance, consider the *shifted power-law* distribution, which has a form

$$\Pr(x) = \frac{\alpha - 1}{k + x_{\min}} \left(\frac{k + x}{k + x_{\min}} \right)^{-\alpha} \qquad \text{for } x \geq x_{\min} \ , \tag{7}$$

[9] See `http://en.wikipedia.org/wiki/Gini_coefficient`

where k is some constant. (Note that when $k = 0$, we recover Eq. (2) exactly.)

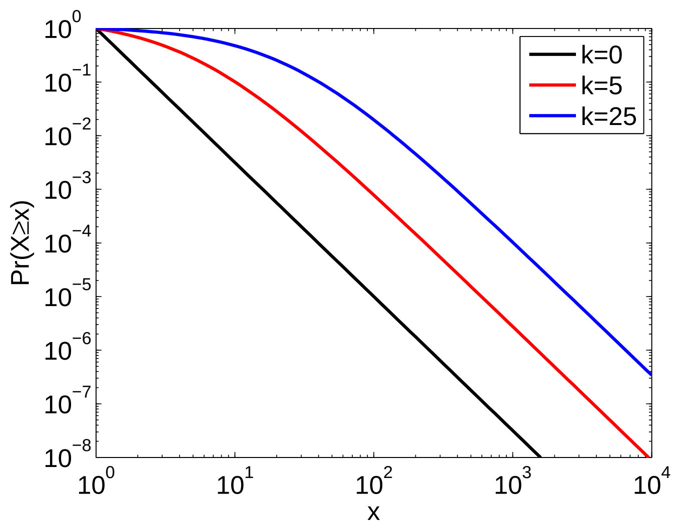

Figure 4: The ccdf of the shifted power-law distribution, for several choices of shift parameter k. Note that the tail does show the power-law form, but the "body" or "head" shows significant curvature.

With a little algebra, we can rewrite Eq. (7):

$$\begin{aligned}\Pr(x) &= C(x+k)^{-\alpha} \qquad\qquad \text{for } x \geq x_{\min} \\ &= C(x+k)^{-\alpha}\left(\frac{x^{-\alpha}}{x^{-\alpha}}\right) \\ &= C\left(1+\frac{k}{x}\right)^{-\alpha} x^{-\alpha} \\ &= L(x)\, x^{-\alpha} \;\; ,\end{aligned}$$

where $L(x) = C\left(1+\frac{k}{x}\right)^{-\alpha} \to 1$ as $x \to \infty$, and thus a shifted power-law distribution has a power-law tail. The function $L(x)$ describes exactly how the deviations from the power-law form decay as we move further out into the tail. When $x \lesssim k$, the "body" term $L(x)$ is large compared to the tail term $x^{-\alpha}$, leading to curvature on the log-log plot. Figure 4 shows some examples of this distribution.

2 Matlab code for Figure 2

Here is the Matlab code for the in-class simulation of the convergence of the sample mean and variance for iid random variables drawn from a power-law distribution. Note: you won't be able to run this code as-is because it calls a function `randht` that you don't have. Once you're solved question 3a from problem set 1, replace that line with the function you derive.

```
% run the experiment 10 times, showing the results of each iteration for 1 second
for iterations=1:10
    alpha=1.7;
    nr = unique(round(logspace(1,6,40)));   % sizes of samples to simulate
    mn = zeros(length(nr),1);               % storage for sample means
    vr = zeros(length(nr),1);               % storage for sample variances
    for i=1:length(nr)
        n = nr(i);                          % set the size of the iid sample
        for j=1:size(mn,2)
            x = randht(n,'powerlaw',alpha); % generate iid PL r.v.
            mn(i,j) = mean(x);              % compute the sample mean
            vr(i,j) = var(x);               % compute the sample variance
        end;
    end;

    % make a pretty figure
    figure(2);
    loglog(nr,mn,'ro','LineWidth',2,'MarkerFaceColor',[1 1 1]); hold on;
    loglog(nr,vr,'bo','LineWidth',2,'MarkerFaceColor',[1 1 1]); hold off;
    set(gca,'FontSize',fs,'XLim',10.^[1 6],'YLim',10.^[0 9]);
    set(gca,'XTick',10.^(1:6),'YTick',10.^(0:9));
    xlabel('Size of sample, n','FontSize',16);
    ylabel('Value','FontSize',16);
    h=legend('Sample mean','Sample variance',2); set(h,'FontSize',16);
    h=text(8*10^4,10^8,strcat('\alpha=',num2str(alpha,'%3.1f'))); set(h,'FontSize',16);
    drawnow;  % force Matlab to draw the figure now (flush the graphics queue)
    pause(1); % pause for 1 second before continuing the loop
end;
```