Stat 720 Notes

Alan Mishler

September 28, 2016

Agenda

- Power law distributions and the "scale-free" property
- Straight lines on log-log plots
- Proper inference for power laws
  - MLE
  - Goodness of Fit
  - Model comparison tests (Vuong)
- So what?

Power law distributions

Remember that degree distributions are typically heavy-tailed and right-skewed. For instance, a typical histogram of the degree of a graph might look something like panel (a) in Figure 1, while a typical complementary CDF (cCDF) might look like the solid line in panel (b). Rarely would you expect to see something like the dotted line in (b).

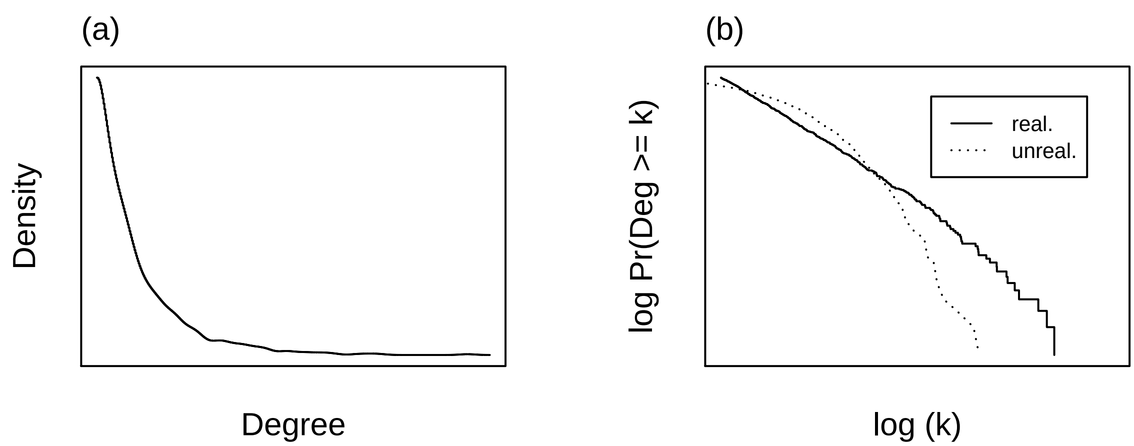

Figure 1: (a) Realistic simulated degree distribution. (b) Realistic simulated complementary CDF (cCDF), and an unrealistic (unlikely) cCDF.

How can these patterns be explained? Answer: power-law distributions. A power law distribution is a distribution of the form

$$p(x) \propto x^{-\alpha}, \text{ for } x \geq x_0$$

where $p(x)$ is either a pdf or a pmf, and x_0 is some minimum possible value. Pareto distributions are continuous power-law distributions, while Zipf distributions (aka Zeta distributions) are discrete power-law distributions.

Properties of power-law distributions

Power-law distributions have two important properties.

1. They're heavy-tailed
2. They're similar (or "scale-free")

1. The heavy-tailed property

From the definition above, notice that

$$\begin{aligned} p(x) &= kx^{-\alpha} \\ \implies \log p(x) &= c - \alpha \log(x) \\ \implies \log Pr(X \geq x) &= c' - (\alpha - 1) \log x \end{aligned}$$

where k, c, and c' are constants. In other words, $\log Pr(X \geq x)$ is a linear function of $\log x$, with slope $= (\alpha - 1)$.

Notice that $E[X^d] = \int_{x_0}^{\infty} x^d x^{-\alpha} c dx = \infty$ if $d > \alpha$. (Again, c is a constant.) For example, if $\alpha = -2$, then the variance is undefined, as are all moments of order > 2.

2. Self-similarity

Saying that power-law distributions are self-similar is equivalent to saying that they are "scale-free." This means that the shape of the density curve is the same regardless of the scale at which you're observing the curve, i.e. how large a portion of the curve you are observing. Mathematically, this means that, for $a \geq x_0$, $Pr(X \geq x | X \geq a) \propto x^{-\alpha+1}$:

$$\begin{aligned} Pr(X \geq x | X \geq a) &= \frac{Pr(X \geq x, X \geq a)}{P(X \geq a)} \\ &= \frac{Pr(X \geq x)}{Pr(X \geq a)} \\ &= \frac{c' x^{-\alpha+1}}{c^* a^{-\alpha+1}} \\ &\propto x^{-\alpha+1} \end{aligned}$$

Power-law distributions are basically the only distributions which have this scale-free property. (Note that "scale-free" $\neq$ "memoryless." The exponential distribution is a memoryless distribution, but it is not scale-free.)

Scale-free distributions are important in physical phase transitions, like between states of matter, and they are long familiar to physicists. At some point, they made the leap from the physical to the social sciences, computational biology, etc. For instance, they are used to model the point at which diseases begin to spread through a network–the so-called *epidemic threshold*.

Inference for power-law distributions, old school (1890s-2000s)

Inferential methods for power-law distributions date to Pareto's work in the 1890s. The traditional procedure involves drawing straight lines on log-log plots, like this:

1. Bin the data, setting bin widths so that all the bins are small but have non-zero counts. Using the model $\log p(x) = c - \alpha \log x$, where $p(x)$ is the proportion of the data in bin x, regress $\log p(x)$ on $\log x$ to get an estimate for α.
2. Alternatively, start from the cCDF. Using that $\log Pr(X \geq x) = c' - (\alpha - 1) \log x$, where $\log Pr(X \geq x)$ is the proportion of the data greater than or equal to bin x, regress $\log Pr(X \geq x)$ on $\log x$ to get an estimate for α.
3. Either way, report the usual standard errors and confidence intervals for $\hat{\alpha}$.
4. Claim goodness of fit by R^2.

Problems with the old school method

1. R^2 is bad, in the usual ways. Among other issues, it has no power to discriminate between power laws and other heavy-tailed distributions. For instance, if X is a log-normal variable, i.e. $logX \sim N(\mu, \sigma^2)$, then it is possible to get R^2 values arbitrarily close to 1.
2. For inference on $\hat{\alpha}$, none of the usual regression assumptions hold (normality, homoscedasticity, etc.).
3. The estimated distribution isn't normalized. For example, the best fit line using the empirical doesn't necessarily go through $(x_0, 1)$, which it should, since x_0 is the minimum possible value of X.
4. The estimate $\hat{\alpha}$ is consistent, but it's not very efficient.

Inference for power-law distributions, new school

The modern, correct way to perform inference for power-law distributions, is as follows:

1. Estimate α by MLE (assuming x_0 is known). The log likelihood is:

$$\log L(\alpha) = \prod_{i=1}^{n} c(\alpha, x_0) x^{-\alpha}$$

where $c(\alpha, x_0)$ is a normalizing constant that depends on α and x_0. Differentiating $\log L(\alpha)$ and setting it equal to 0 yields, for the continuous case:

$$\hat{\alpha}_{MLE} = 1 + \frac{n}{\sum_{i=1}^{n} \log x_i}$$

The MLE is computationally easy to find, it's consistent, it's efficient, and it's asymptotically normal.

2. To test goodness of fit, use a goodness-of-fit test, like the Kolmogorov-Smirnov (KS) test. The test statistic d_{ks} for a KS test is defined as $d_{ks} = \max_x |\hat{F}_n(x) - F(x)|$, where $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1(x_i \leq x)$. This quantity converges in probability to 0, if $X \sim F$, and to $\max_x |G(x) - F(x)|$ if $X \sim G \neq F$.

The KS statistic is tabulated for fixed F. Bootstrapping can be used when the particular power-law distribution F is unknown.

3. Estimate x_0 by minimizing d_{ks}, i.e. seeing how good a fit you get at each candidate value of x_0. This yields an estimator

$$\hat{x}_0 = \text{argmin}_{x_0}\, d_{ks}(\hat{F}_n, PL(al\hat{p}ha(x_0), x_0))$$

where $\hat{F}_n$ is the empirical cumulative distribution and PL is a power law distribution with parameters $\hat{\alpha}$ and x_0.

4. Compare your model to other models (other heavy-tailed distributions) with model comparison metrics like BIC, or using the Vuong likelihood ratio test.

The Vuong test

The Vuong test requires two models with given likelihood functions and MLE estimators:

$$\begin{aligned} &\text{Model 1: log likelihood } \ell_n(\theta), \text{MLE } \hat{\theta} \\ &\text{Model 2: log likelihood } m_n(\psi), \text{MLE } \hat{\psi} \end{aligned}$$

Assumptions

1. There is no overlap; that is, there are no pairs (θ_0, ψ_0) that index the same models. For example, this test would work for an exponential distribution and a normal distribution, but not for an exponential and a Gamma distribution, since one is a special case of the other.
2. Both parameter estimates are assumed to be consistent.

If the assumptions are met, then $\frac{1}{n}(\ell_n(\hat{\theta}) - m_n(\hat{\psi})) \to E[\ell(\theta^*) - m(\psi^*)]$, the expectation under the true distribution. With a few regularity conditions, a Central Limit Theorem holds for $\frac{1}{n}(\ell_n(\hat{\theta}) - m_n(\hat{\psi}))$. If one model has higher likelihood than the other, then the test statistics will tend to $\pm\infty$.

Looking back at the literature

An examination of a bunch of papers that used power-law models yielded the following observations:

1. There are lots of heavy tails.
2. Usually the tails aren't so heavy, under the best power-law model, that the variance is undefined.
3. Very few of the models pass goodness-of-fit tests even at the 10% level. None of the network datasets that were examined passed.
4. In most cases, the power-law models don't provide better fits than other heavy-tailed distributions. Log-normal distributions are particularly strong competitors, as are negative gamma distributions.

Conclusions

1. Power-law distributions are a rough approximation, sometimes.
2. When doing inference with power-law distributions, use proper statistics (new school, not old school).
3. Ask yourself whether it really matters that something is exactly is a power-law. It might be that the only important factor is that the distribution is right-skewed and heavy-tailed, in which case other models might work just as well or better.
4. Look for other sources of information about the degree distribution in order to distinguish among models.