Power Law Derivations

Supplementary to Clauset 2009

Unchitta Kan

January 2021

This short document contains brief derivations of some of the equations in [1], an excellent paper on power law distributions.

Table of Contents

Normalizing constant for continuous power law distribution. A quantity x follows a power law if its distribution is described by a probability density function $p(x)$ such that

$$p(x)dx \; = \; Pr(x \leq X < x + dx) \; = \; Cx^{-\alpha}dx. \tag{1}$$

The above equation is Eq. (2.1) in [1]. Clauset et al then proceeds to give the explicit form of $p(x)$ in Eq. (2.2), which reads

$$p(x) \; = \; \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{\alpha}. \tag{2}$$

To go from Eq. (1) to Eq. (2) here, we solve for C, the normalizing constant, such that $p(x)$ is a valid probability density function, which is to say that it has a non-negative value everywhere and its integral over the entire event space is 1. Since we assume that there is a lower bound $x_{min} > 0$ to the power-law behavior, the first condition is already satisfied and we are only concerned with the area under the curve of $p(x)$. We want to satisfy

$$\int p(x)dx \; = \; 1$$

or, equivalently,

$$\int_{x_{min}}^{\infty} Cx^{-\alpha}dx \;=\; 1 \qquad \Longrightarrow \qquad \int_{x_{min}}^{\infty} x^{-\alpha}dx \;=\; \frac{1}{C}.$$

Now we evaluate the integral:

$$\int_{x_{min}}^{\infty} x^{-\alpha}dx \;=\; \left[\frac{x^{(-\alpha+1)}}{-\alpha+1}\right]_{x_{min}}^{\infty} \;=\; 0 - \frac{x_{min}^{-\alpha+1}}{(-\alpha+1)} \;=\; \frac{x_{min}^{-\alpha+1}}{\alpha-1},$$

where the first equality comes from taking the anti-derivative of $x^{-\alpha}$ and the third inequality follows from the fact that $x^{-\alpha} \to 0$ as $x \to \infty$ for $\alpha > 1$.

So,

$$\frac{1}{C} \;=\; \left(\frac{x_{min}}{\alpha-1}\right) x_{min}^{-\alpha} \qquad \Longrightarrow \qquad C \;=\; \frac{\alpha-1}{x_{min}}\left(\frac{1}{x_{min}^{-\alpha}}\right)$$

and

$$p(x) \;=\; \frac{\alpha-1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\alpha}.$$

Normalizing constant for discrete power law distribution. Next we look at the discrete case, which is the same idea as the continuous case but instead of taking integrals we use an infinite sum instead. We want

$$\sum_{x=x_{min}}^{\infty} p(x) \;=\; \sum_{x=x_{min}}^{\infty} Pr(X=x) \;=\; \sum_{x=x_{min}}^{\infty} Cx^{-\alpha} \;=\; 1.$$

This implies

$$\sum_{x=x_{min}}^{\infty} x^{-\alpha} \;=\; \frac{1}{C}.$$

Expanding the infinite sum, we see

$$\begin{aligned}\frac{1}{C} \;&=\; x_{min}^{-\alpha} + (x_{min}+1)^{-\alpha} + (x_{min}+2)^{-\alpha} + \dots \\ &=\; \sum_{n=0}^{\infty}(x_{min}+n)^{-\alpha} \\ &=\; \zeta(\alpha, x_{min}),\end{aligned}$$

where the last equality is the Hurwitz zeta function as described in [1] (Eq. 2.5 there).

It then follows that

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})}, \qquad (3)$$

the discrete power-law distribution given in [1] (Eq. 2.4 there).

Complementary CDF. The complementary CDF of a power-law distributed variable is

$$\begin{aligned}
P(x) &= \int_x^\infty p(x')dx' \\
&= \int_x^\infty \frac{\alpha-1}{x_{min}}\left(\frac{x'}{x_{min}}\right)^{-\alpha} dx' \\
&= \left[\frac{\alpha-1}{x_{min}^{-\alpha+1}}\frac{x'^{-\alpha+1}}{-\alpha+1}\right]_x^\infty \\
&= 0-\frac{\alpha-1}{x_{min}^{-\alpha+1}}\frac{x^{-\alpha+1}}{-\alpha+1} \;=\; \frac{\alpha-1}{x_{min}^{-\alpha+1}}\frac{x^{-\alpha+1}}{\alpha-1} \\
&= \left(\frac{x}{x_{min}}\right)^{-\alpha+1},
\end{aligned}$$

which is Eq. 2.6 in [1].

The discrete case can be derived in a similar way; we sum $Pr(X=x)$ starting from x to ∞, which gives us the Hurwitz zeta function again in the numerator:

$$\begin{aligned}
P(x) &= \frac{1}{\zeta(\alpha, x_{min})}\sum_x^\infty x^{-\alpha} \\
&= \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})}.
\end{aligned}$$

Maximum Likelihood Estimators. To fully desribe the power-law distribution given empirical data, we need to estimate the value of α, which is the only parameter of this distribution. To do this, Clauset introduces the maximum likelihood estimation method [1]. The idea of this method is as follows: given data points $x_1, \ldots, x_n$, we ask ourselves: how likely is it that we would see these data given α? If we just look at any particular data point x_i, the likelihood of seeing it given α (which we do not know yet), in the continuous case, is

$$Pr(x_i|\alpha) = \frac{\alpha-1}{x_{min}}\left(\frac{x_i}{x_{min}}\right)^{-\alpha}.$$

Assuming that the data points are independently and identically distributed, the probability of seeing all the data points is the product of the probabilities of seeing the individual points, namely

$$Pr(x_1,\ldots,x_n|\alpha) \;=\; \Pi_{i=1}^n Pr(x_i|\alpha) \;=\; \Pi_{i=1}^n \left[\frac{\alpha-1}{x_{min}}\left(\frac{x_i}{x_{min}}\right)^{-\alpha}\right] \;:=\; L(\alpha;\mathbf{x}). \tag{4}$$

Eq. 4 is called the likelihood function L, as that tells us the likelihood of observing these data points. Now, we want to find an estimate α which will maximize this likelihood, $\hat{\alpha} = \arg\max_\alpha Pr(x_1,\ldots,x_n|\alpha)$. To do this, we find where the function is at its maximum, and we learned from calculus that we can do so by finding its derivative and setting it equal to 0. However, Eq. 4 is rather nasty (so are most likelihood functions), so instead of dealing with a product, we deal with a finite sum of the natural log of the factors instead. The equation about to follow (Eq. 5) is referred to as the log-likelihood for this reason.

$$\begin{aligned}\ell(\alpha;\mathbf{x}) &= \ln\left(L(\alpha;\mathbf{x})\right) \qquad (5)\\ &= \sum_{x=1}^{n}\ln\left[\frac{\alpha-1}{x_{min}}\left(\frac{x_i}{x_{min}}\right)^{-\alpha}\right]\\ &= n\ln\left(\frac{\alpha-1}{x_{min}}\right)-\alpha\sum_{x=1}^{n}\ln\left(\frac{x_i}{x_{min}}\right)\\ &= n\ln\left(\alpha-1\right)-n\ln\left(x_{min}\right)-\alpha\sum_{x=1}^{n}\ln\left(\frac{x_i}{x_{min}}\right), \qquad (6)\end{aligned}$$

where we use the properties of natural logs to manipulate the expressions in the last two equalities. We also note that because the natural log function is monotonically increasing, the value of α that maximizes the log-likelihood also maximizes the likelihood itself. Another reason why we would want to use the log-likelihood has to do with concavity. Concavity ensures that there exists a global maximum. Many probability density functions are not concave, but their logs are.

Next, we take the derivative of (Eq. 6) *with respect to* α and set it equal to 0, and solve for $\hat{\alpha}$, the maximum likelihood estimator for α:

$$\begin{aligned}0 &= \frac{\partial\ell}{\partial\alpha}(\alpha;\mathbf{x})\\ &= \frac{n}{\alpha-1}-\sum_{x=1}^{n}\ln\left(\frac{x_i}{x_{min}}\right)\\ \Longrightarrow\ \hat{\alpha} &= 1+n\left[\sum_{x=1}^{n}\ln\left(\frac{x_i}{x_{min}}\right)\right]^{-1}, \qquad (7)\end{aligned}$$

which is indeed as given in [1] (Eq. 3.1 there).

References

[1] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data". In: *SIAM review* 51.4 (2009), pp. 661–703.