Power law distributions

Statistics for Business

Josemari Sarasola - Gizapedia

Power laws

We say X and Y variables are related according to a power law when one of them is proportional to a given power of the other:

$$Y = KX^{\alpha}$$

If we take logarithms, the relation becomes lineal:

$$\ln Y = \ln K + \alpha \ln X$$

In statistics, power law distributions link the x random variable and $p(x)$ probability, with a negative exponent:

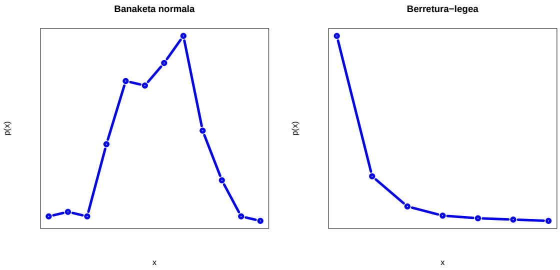
$$p(x) = Kx^{-\alpha}$$

Taking logarithms,

$$\ln p(x) = \ln K - \alpha \ln x$$

That is, probability decreases when x increases.
α parameter gives the rate of decrease: if α is big the rate of decreases bigger, and thus when x is big, the probability or frequency will be samller.

Power laws in statistics



In a power law, there are many small, and few big.

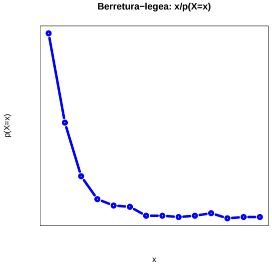
Power laws in statistics

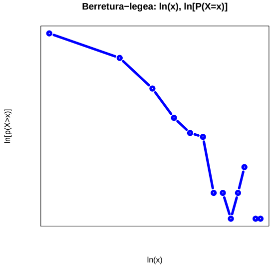
Power laws are usual in distrbutions about sizes and magnitudes:

- Family incomes: many with low incomes; few with high incomes.
- Number of workers: many companies with few workers, few companies with many workers.
- The number of sold books for a given title: many books with few sales, few with big sales.

How to seek for power law distributions

After the normal distribution, power laws are the most usual data. Data with a bell shape and symmetry follow a normal distribution. How to seek for power law distributed data?
Instead of plotting x and $p(x)$, we should plot both's logarithms. In order to have a power law distribution, we should have a (nearly) straight line:



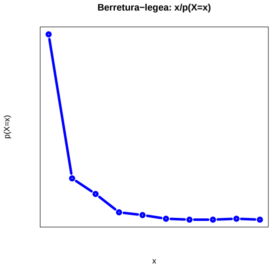


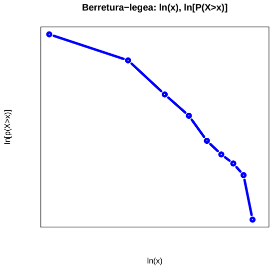
Nevertheless, usually the is "noise" at big values, because in the upper side frequencies are very small (and thus variable). So,

How to seek for power law distributions

Do we have a straight line in the previous plot? In order to see more clearly:

- instead of taking intervals with constant lenght, we take logaritmic intervals: (e.g.: 10-100-1000 or 2-4-8-16)
- or (this works better) plot $\ln x$ and $\ln p(X > x)$, and we should have a straight line:





Minimum in power law distributions

- Power laws usually don't apply along all the support (set of possible values).
- E.g, if the take family incomes there is always a minimum income.
- Because of this, power laws are applied for values bigger than a x_{min} value.
- So, usually we don't have power law distributions but power law tail distributions.

Continuous power law distributions

Densuty functions

We may proof than power law distributions have to follow this density function:

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

We must remark than α must be bigger than 1. In practice this is not a hard condition, because data with $\alpha < 1$ are very rare.

Continuous power law distributions

Distribution function

Probabilities are more esasily calculated with distribution functions:

$$F(x) = P[X < x] = 1 - \left(\frac{x}{x_{min}}\right)^{1-\alpha}$$

$$\overline{F}(x) = P[X > x] = \left(\frac{x}{x_{min}}\right)^{1-\alpha}$$

Expected value

$$\forall \alpha > 2, E[X] = \frac{\alpha - 1}{\alpha - 2} x_{min}$$

For $\alpha \leq 2$ values, expected values doesn't exist, more exactly it goes to infinity. What does this mean? If we would estimate the expected value taking some data and calculating its arithmetic mean, we would a bigger and bigger mean, as n increases, with any limit, till infinity.

Continuous power law distributions

Expected value of the maximum in a sample

For a n sample size,

$$E[X_{max}] = n^{\frac{1}{\alpha - 1}}$$

Thus, as we have $\alpha > 1$, the expected value of the maximum always increases, as n increases. But when we have $\alpha \leq 2$, the expected value of the maximum increases in an explosive way, as we may see in the formula. There we have another argument for the need of having $\alpha > 2$.

Continuous power law distributions

Concentration: Lorenz curve

For a distribution of incomes, whaty is the A percentage of the total wealth owned by the richest P percentage of families?

$$\forall \alpha > 2, A = P^{\frac{\alpha-2}{\alpha-1}}$$

E.g., taking the richest 20%:

α	A
2.1	0.86
2.3	0.69
2.5	0.58
2.7	0.51
2.9	0.46

So, the smaller is α, the bigger is the concentration.

Continuous power law distributions

Conditional distributions

Given a powe law eith a x_{min} minimum value, known that $X > x_0$, the new conditional distribution will also be a power law distribution, with the same α parameter, taking x_0 instead of x_{min} :

$$P[X > x/X > x_0] = \frac{P[X > x]}{P[X > x_0]} = \frac{\left(\dfrac{x}{x_{min}}\right)^{1-\alpha}}{\left(\dfrac{x_0}{x_{min}}\right)^{1-\alpha}} = \left(\frac{x}{x_0}\right)^{1-\alpha}$$

As we see, the last expression is the complementary of the distribution function of a power law distribution.

Continuous power law distributions

Pareto distribution

Pareto distribution is another name for the power law distribution, but with another parametrization. It's named after Vilfredo Pareto (1848-1923), an Italina economist.
This is its distribution function:

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

Thus, in the Pareto distribution α parameter is the $\alpha - 1$ parameter for the power law disribution, as we see with a sinple calculation.
It's recommended to state previously which type of parametrization is used (power law or Pareto).

Continuous power law distributions

Estimation of alfa parameter

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \right]$$

Discrete power laws: Zipf's law

The former distributions are applied for continuous random variables. In other cases, variable is discrete (few different values) or qualitative, (e.g., size of cities, frequency or words, article visits in Gizapedia number of families with 2 or more children), but we find the same power law:

- The biggest city is really very big, and the others have lesser and lesser population, and finally very similar.
- Some words appear very much times, most of them not.
- Some articles in the encyclopedia have many visits, most of them very few.
- Many families with 2 children, but the number of families decreases with 3, 4, 5, 6, ... children.

Discrete power laws: Zipf's law

George Kingsley Zipfek (1902-1950), a linguist in Harvard University, found that the frequency of words in English language follow (approximately) this formula

$$f = \frac{0.1}{k}$$

k is the rank of the word (1, most common; 2, second most common, and so on)) and f the frequency. Thus, the most frequent word ("the") apperas 0.1/1=0.1=10% of the times, the second most frequent 0.1/2=0.05=5%5 of the times, ...
Zipf's law always link rank with frequency or size.

This is a more exact expression of the Zipf's law:

$$f = \frac{1}{k^a}$$

having f the frequency, magnitude or size of the k ranked element, respectig to the other elements and a a parameter, usually bigger than 1
When the number of elements is fixed, in order to have the sum of frequencies 1, the frequescies must be divided by the sum of all the frequencies, and so we will get the f_z normalized frequencies:

$$f_z = \frac{\frac{1}{k^a}}{\sum_{k=1}^{N} \frac{1}{k^a}}$$

But for the exercises we will always take the simplest expression:

$$f = \frac{C}{k}$$

$$f_z = \frac{\frac{C}{k}}{\sum_{k=1}^{N} \frac{C}{k}}$$

Discrete power laws: Zipf's law

How to find Zipf's laws

Plot the logarithms of both the k ranges and f sizes or frequencies. If the result is approximately a straight line, Zipf's law is suitable.

Discrete power laws: Zipf's law

We have a region with 6 cities, and have set $f = \dfrac{3}{k}$ as a Zipf's law for the relative size among them:

Rank	Relative size (f)	Normalized frequency (f_z)
1	3	0.40
2	1.5	0.20
3	1	0.13
4	0.75	0.10
5	0.6	0.08
6	0.5	0.07
	7.35	1

Thus, the most populated city has 40% of the population.

Parameter estimation

I suggest a very simple method of estimation (I accept it is not rigorous, but on the other hand it's very easy): we equal the empirical or "real" frequency with the frequency given by the Zipf's law, C unknown. Coming back to the last example, known that the population in the second most populated city is 25 of the total population:

$$\frac{C}{2} = 0.25 \to \hat{C} = 0.5$$

These estimates are different for different k ranks. We may seek for the best estimate calculating the sum of squared differences between theoretical and empirical frequencies.