# Power laws in economics and/or finance

Thesis presented by
**Valérian Even**

Supervisor
**Zhengyuan Gao (UCL)**

Reader
**François Maniquet (UCL)**

Academic year 2017-2018

**Master 60 en Sciences économiques, Orientation générale**

Ecole d'économie de Louvain/UCL • Place Montesquieu 3 • 1348 Louvain-la-Neuve
Département des Sciences économiques/UNamur • Rempart de la Vierge 8 • 5000 Namur

# Acknowledgements

I would like to express my gratitude to:

- My supervisor Professor Zhengyuan Gao for pointing me towards such an interesting topic, for giving me a lot of freedom in the conception of this work and for his (infinite) patience during the making of this work;
- Professor François Maniquet, first for reading and reviewing this work, but also for his course Advanced Microeconomics I that I found very enlightening and particularly well teached;
- Manu, Hector, Alexandre, Justin and Valentin for their inspirations;
- my parents for their trust;
- and finally Céline for her infinite support.

# Contents

# Introduction

Many empirical quantities congregate around a characteristic value. The speeds of cars on a highway, the heights of adult human beings, the weights of apples in a store, air pressure, sea level, ... All of these quantities display variations, but their distribution gives a negligible amount of probability to the event of being far from the typical value. Consequently, the typical value is an ideal representative candidate for most observations. To a certain degree, it is an entirely useful statement to say that an adult male in Belgium is about 182cm tall because not one of the 10 millions members of this group deviate very far from this size. Even when considering the most extreme deviations, which happen exceptionally rarely, we do not notice deviations greater than a factor two from the mean in either directions and so the distribution can be well-approached by using just its mean and standard deviation.

However, not all distributions follow this pattern. The recent years have produced countless examples of "non-normal" distributions from complex social, biological and technological systems. Among these distributions, the *power law* distribution has received a lot of attention because of its mathematical properties and the fact it can be produced by interesting processes such as network effects or self-organization.

A power law is a relation of the type $y = \frac{c}{x^\alpha}$ between two variables of interest $x$ and $y$. In particular, power laws appear widely in physics, biology, earth sciences, computer science, demography, social sciences, economics and finance. Phenomena that appear to follow a power law include the sizes of craters on the moon and of solar flares, the foraging pattern of various species, the sizes of activity patterns of neuronal populations, the frequencies of words in most languages, the frequencies of family names, the sizes of towns and cities, the distribution of wealth, the sizes of firm and the stock market movements.

It is the aim of the present work to introduce power laws and their properties as well as some of the ways to generate such distributions. The work is organized as follows:

- in Chapter 1, we introduce the notion of power law and study its properties. We also discuss various statistical methods used to determine the presence of power laws and we illustrate those techniques using the data related to Belgian city and town population in 2000. Finally, we end this first chapter by providing additional examples of power law appearing in the world of economics and finance;

- in Chapter 2, we look at procedures from which power laws may arise. In particular, we focus on two methods:

  - the *preferential attachment model* that refers to a process in which a discrete number of objects is randomly added to a set of containers, in particular additional objects are added continuously to the system and are distributed among the containers as an increasing function of the number of objects the containers already have;

- the *percolation process* that is a mechanism revolving around finding a critical value at which a dynamical system suddenly shifts from a convergent state to an exploding state.

# Chapter 1

# Introduction to power laws and a few statistical tools

This first chapter aims at introducing the concept of power laws to the reader. After going through the basic properties of this type of distribution, we will present the basic statistical techniques used to identify and assess the power law behavior of the distribution we are studying. We will use one example to illustrate these tools and to provide a better insight of the techniques used in this section. For more information on that topic, we recommend reading [CSN09, Gab09, Mit04, New05, Sor06, Zaj97].

## 1.1 Motivation and definition

Many observations a scientist can measure vary around a typical value. The most simple example would be measuring the heights of human beings. Adults usually culminate around 180cm but it is extremely unlikely to find an adult person that is 20cm tall or even 400cm tall (the tallest man being 272cm tall and the smallest one being 55cm tall according to the Guinness World Records). But typical other measurements could also be the speed of cars on a highway, the sea level or even the temperature in Louvain-la-Neuve at noon on the first of May.

All these examples capture the fact that the measurements do not deviate too much from a particular value, but not all distributions fit this pattern. Taking for example the population of US cities we would find, using the 2000 US Census,[1] that the average population of a city in the US is 9002.051. Obviously there are cities like New York, Los Angeles or Chicago that are way bigger than this average size. The same remark holds for towns like McMullen that hardly achieve a population of 10. It seems obvious that the measure of the population of a town or city is not characterized by a mean value like the previous examples were.

Another way to put this is by asking the following question: what is the probability that someone has twice my height? For a adult human being, the answer to this question will be very close to zero. Now what if we ask ourselves: what is the probability that someone has twice my wealth? This, of course, will depend on your wealth but there is also a big chunk of the population that are even one hundred times as wealthy as you are.

[1] Data available at `http://tuvalu.santafe.edu/~aaronc/powerlaws/data.htm`

This observation, that the distribution of the number of people with an income or wealth $X$ greater than a large value $x$ is actually proportional to $\frac{1}{x^\alpha}$, was due to Pareto at the end of the $19^{th}$ century, therefore the name of *Pareto law* to describe this particular distribution.

Among the distributions that are not characterized by such central values, the *power law* managed to get a lot of attention, not only for its mathematical properties, but for its appearance in a diverse range of phenomena. Examples of such are sizes of earthquakes, moon craters, solar flares, computer files and wars, frequencies of use of words, frequencies of occurrence of personal names, numbers of papers scientists write, numbers of citations they receive, numbers of hits on web pages, sales of books or music recordings, numbers of species in biological taxa, people's annual income.

Let us now give the mathematical definition of the object that we will study in the subsequent parts of this work:

**Definition.** A nonnegative random variable $X$ is said to have a power law distribution if

$$P(X \geq x) \sim cx^{-\alpha}$$

for some constants $c > 0$ and $\alpha > 0$.

In the next section we will interest ourselves more particularly with the properties exhibited by such a distribution in order to get a better understanding of its nature.

## 1.2 Properties

The aim of this section is to go through the important properties exhibited by power laws. Here we will mainly be concerned with the study of the probability density that will be written $p(x)$.

### 1.2.1 Lower bound and tail

A first remark is that the probability density $p(x) = cx^{-\alpha}$ diverges when $x$ tends to 0: there must a lower bound $x_m$ after which the power law applies. This is something that we see also in practice, few phenomena display a power law for all values of $x$ but rather follow a power law in the *tail* of the distribution. It is additionally very common to find that the parameter $\alpha$ lies in the interval $]2, 3[$, except for a few exceptions.

### 1.2.2 Normalizing constant

The constant $c$ can be found through the normalization requirement that (taking the case where $\alpha > 1$)

$$\begin{aligned}
1 &= \int_{x_m}^{\infty} p(x)dx \\
&= \int_{x_m}^{\infty} cx^{-\alpha}dx \\
&= c \lim_{y\to\infty} \left[\frac{x^{-\alpha+1}}{1-\alpha}\right]_{x_m}^{y} \\
&= c \lim_{y\to\infty} \left(\frac{y^{-\alpha+1}}{1-\alpha} - \frac{x_m^{-\alpha+1}}{1-\alpha}\right) \\
&= c\frac{x_m^{-\alpha+1}}{\alpha-1}
\end{aligned}$$

from which we find

$$c = (\alpha-1)x_m^{\alpha-1}.$$

Note that the normalization only makes sense if $\alpha > 1$ otherwise the limiting term $\lim_{y\to\infty} \frac{y^{-\alpha+1}}{1-\alpha}$ would have diverged and this, as mentioned before, is not something that usually occurs. Thus the normalized expression for the power law is

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

Also notice that if the distribution follows a power law only between some values $x_m$ and $x_M$ then the normalization can always be done (whatever the value of $\alpha$) and yields the following density (calculations are done as before):

$$p(x) = \frac{1-\alpha}{x_M^{1-\alpha} - x_m^{1-\alpha}}x^{-\alpha}.$$

Except when mentioned otherwise, we will consider unbounded power laws.

### 1.2.3 Moments

We can calculate the moments of the distribution. The mean value for example, when it exists, is calculated as

$$\begin{aligned}
E[X] &= \int_{x_m}^{\infty} xp(x)dx \\
&= \int_{x_m}^{\infty} x(\alpha-1)x_m^{\alpha-1}x^{-\alpha}dx \\
&= (\alpha-1)x_m^{\alpha-1}\int_{x_m}^{\infty} x^{-\alpha+1}dx \\
&= (\alpha-1)x_m^{\alpha-1}\lim_{y\to\infty}\left[\frac{x^{-\alpha+2}}{2-\alpha}\right]_{x_m}^{y} \\
&= \frac{\alpha-1}{\alpha-2}x_m^{\alpha-1}x_m^{2-\alpha} \\
&= \frac{\alpha-1}{\alpha-2}x_m.
\end{aligned}$$

To get this expression however, we obviously need that $\alpha > 2$. For a lower value of $\alpha$ the mean diverges and becomes infinite, this result in particular should be interpreted carefully: it is obvious that if we take a finite number of data (that are themselves finite) we will get a finite mean, but the previous result indicates that repeating the experiment multiple times (by taking new measurements) would yield large fluctuations in the mean. In other words, the mean is not a well defined quantity in the case where $\alpha \leq 2$. We will see later that the population of towns or cities falls into this particular case.

More generally, one can find the following expression for the $m^{th}$ moment:

$$E\left[X^r\right] = \frac{\alpha - 1}{\alpha - 1 - r} x_r^m.$$

The general rule is thus that for a power law with exponent $\alpha$, all moments $m > \lfloor \alpha - 1 \rfloor$, where $\lfloor x \rfloor$ is the greatest integer smaller or equal to $x$, exist while the others don't.

It is common in the literature to work with the complementary cumulative distribution function of a variable that follows a power law. We write

$$\begin{aligned} P(x) = P(X \geq x) &= \int_x^\infty p(x')dx' \\ &= \int_x^\infty \frac{\alpha - 1}{x_m} \left(\frac{x'}{x_m}\right)^{-\alpha} dx' \\ &= \frac{\alpha - 1}{x_m^{1-\alpha}} \lim_{y \to \infty} \left[\frac{x'^{1-\alpha}}{1-\alpha}\right]_x^y \\ &= \left(\frac{x}{x_m}\right)^{1-\alpha}. \end{aligned}$$

### 1.2.4 Discrete case

So far we only focused ourselves on the continuous case. Before giving some illustrations of the power law distribution, let us make a quick note on the discrete case. Let us consider the case of integer values with

$$p(x) = P(X = x) = cx^{-\alpha}.$$

We again see that the distribution diverges when $x = 0$ so there must still be a lower bound $x_m > 0$. Calculating the constant via the normalization requirement we find

$$c = \frac{1}{\sum_{x=x_m}^\infty x^{-\alpha}} = \frac{1}{\zeta(\alpha, x_m)}$$

where $\zeta(\alpha, x_m) = \sum_{x=x_m}^\infty \frac{1}{x^\alpha} = \sum_{x=0}^\infty \frac{1}{(x+x_m)^\alpha}$ is the Hurwitz zeta function. In particular this implies

$$P(x) = P(X \geq x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_m)}.$$

Note that everything we did can still be achieved in the discrete but it usually implies more involved calculation techniques compared to the integrals used in the continuous case.

### 1.2.5 Scale-free

Following with the properties of the power law distribution, it is worth noting that it is a *scale-free distribution*. This is one of the most important property displayed by a power law. By that we mean that

$$p(bx) = c(bx)^{-\alpha} = cb^{-\alpha}x^{-\alpha} = b^{-\alpha}p(x)$$

and thus, we see that if we increase the scale of our measurements by a factor $b$, the shape of the distribution remains unchanged (the exponent $\alpha$ is the same), it is only multiplied by some constant $b^{-\alpha}$. It is for example not the case when considering the normal distribution or the exponential distribution.

Actually, we have even more: the power law distribution is in fact the only function satisfying the scale-free criterion. Suppose some probability distribution $p(x)$ has the property

$$p(bx) = g(b)p(x), \tag{1.1}$$

for all $b > 0$. First set $x = x^*$ (note that $x^* > 0$) so that we find $p(bx^*) = g(b)p(x^*)$ and thus $g(b) = \frac{p(bx^*)}{p(x^*)}$. Inserting this last expression in equation 1.1 we find

$$p(bx) = \frac{p(bx^*)}{p(x^*)}p(x).$$

This last expression is true for any $b$ so taking the derivative with respect to $b$ on both sides yield

$$p'(bx)x = \frac{p'(bx^*)x^*}{p(x^*)}p(x)$$

which in the particular case when $b = 1$ gives us

$$p'(x)x = \frac{p'(x^*)x^*}{p(x^*)}p(x)$$

or

$$\frac{p'(x)}{p(x)} = \frac{p'(x^*)x^*}{p(x^*)}\frac{1}{x}$$

which is a first order differential equation easily solved by taking the primitive:

$$\int \frac{p'(x)}{p(x)}dx = \int \frac{p'(x^*)x^*}{p(x^*)}\frac{1}{x}dx = \frac{p'(x^*)x^*}{p(x^*)}\int \frac{1}{x}dx$$

from where we find

$$\ln(p(x)) = \frac{p'(x^*)x^*}{p(x^*)}\ln(x) + C = \ln(x^{\frac{p'(x^*)x^*}{p(x^*)}}) + C.$$

Now the constant $C$ can be simply found by setting $x = x^*$

$$C = \ln(p(x^*)) - \ln(x^{*\frac{p'(x^*)x^*}{p(x^*)}}) = \ln\left(\frac{p(x^*)}{x^{*\frac{p'(x^*)x^*}{p(x^*)}}}\right)$$

Applying the exponential on both sides of the last equation yields the result we were looking for:

$$p(x) = e^C e^{\ln(x^{\frac{p'(x^*)x^*}{p(x^*)}})} = \left( \frac{p(x^*)}{x^{*\frac{p'(x^*)x^*}{p(x^*)}}} \right) x^{\frac{p'(x^*)x^*}{p(x^*)}}$$

thus $p(x)$ is a power law distribution with $\alpha = -\frac{p'(x^*)x^*}{p(x^*)}$.

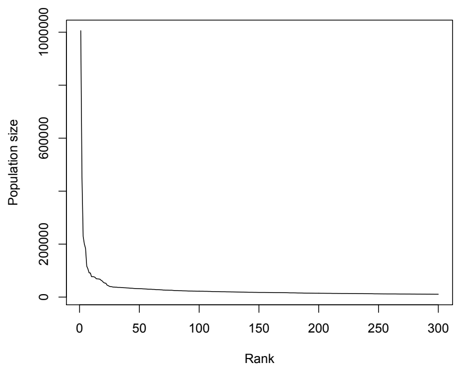
The power law distributions possess other interesting properties like that of the 80/20 rule that shows where the majority of the distribution lies or that of the largest value which shows that the expected largest value among $n$ measurements drawn from a power law distribution grows positively with the number of measurements $n$ (as $n^{1/(\alpha-1)}$ actually). We will not enter the details of the derivations of those properties but rather now present some tools in order to test the power law distribution hypothesis when handling data.

## 1.3 How to recognize a power law?

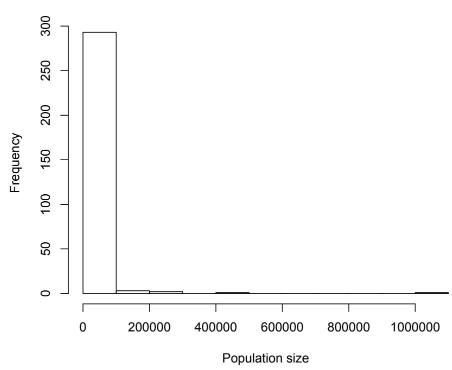
In this section, we will focus on ways to determine if our data exhibit some power law behavior. To illustrate the various techniques, we will apply those on the data related to Belgian city and town population in 2000.[2] Without going too deep into the details, we will also introduce some statistical techniques that allow one to reach conclusions about the hypothesis of facing a power law. These will be completed by adding methods for calculating the parameters of power laws. All calculations and simulations in the present work are done using R.[3] This discussion will closely follow [CSN09, New05, GMY04] and [Gab09].

### 1.3.1 Plot on logarithm scales

The first natural step consists of plotting the data at hand.




(a) Belgian city sizes over rank




(b) Histogram of Belgian city sizes

Figure 1.1

[2]Data available at `http://www.tageo.com/index-e-be-cities-BE.htm`

[3]Although the code will not be displayed here, the author would be glad to share its code with any person interested.

Figure 1.1a shows a simple plot of the population size over the rank of the cities (Bruxelles would be rank 1, Anvers rank 2, and so on) while Figure 1.1b displays the histogram of the distribution of Belgian city sizes. From the latter Figure, we already see that the largest volume of the distribution occurs for fairly small sizes but there is a small number of cities with much higher population, the histogram is thus extremely *right-skewed*. The fact that the histogram is right-skewed typically shows a very different behavior than the one of a normal distribution, this feature and the shape of the curve could be seen as a first indication that the distribution of Belgian city sizes follows a power law.

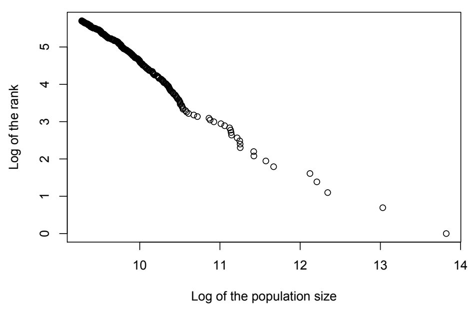
The next test is represented in Figure 1.2a where we see the same graphs as in Figure 1.1 but plotted on logarithmic scales, thus replotted with logarithmic horizontal and vertical axes. As we can see, the histogram plotted this way follows a rather straight line, which means that

$$\log p(x) = -\alpha \log x + C$$

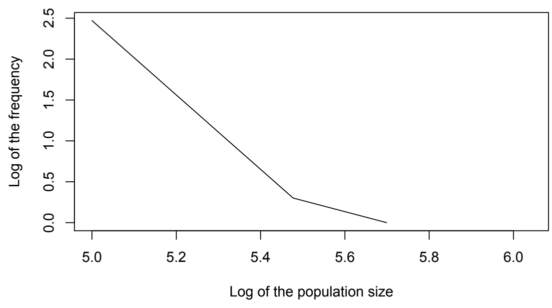
where $\alpha$ and $c$ are constants. Taking exponentials leads to the distribution of power laws

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

The natural way to proceed from here would then be to perform a linear regression in order to extract the slope (parameter $\alpha$). Applying this method yields a pretty accurate result as one can see in Figure 1.3, the details of the statistics are shown in Table 1.1. The scaling parameter is estimated at $\alpha = 1.56704$ while the constant is shown to be $c = 8.80004$; judging by the $p$-values associated to the coefficients, we have to reject the null hypothesis that these coefficients are zero, thus both are significant. The other statistical measures all join hands together to reinforce the idea of the good fitting of the linear regression.




(a) Log-log graph




(b) Histogram on logarithmic scales

Figure 1.2

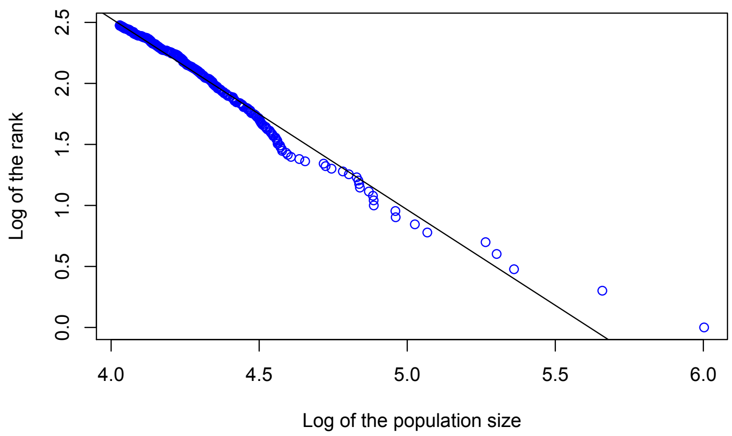



Figure 1.3: Data with regression line

Table 1.1: Linear regression summary

| | Estimate | Std. Error | $p$-value |
|---|---|---|---|
| $c$ | 8.80004 | 0.05923 | $< 2e-16$ |
| $\alpha$ | 1.56704 | 0.01372 | $< 2e-16$ |
| | | Value | |
| Residual Std. Error | | 0.06278 | |
| R-squared | | 0.9777 | |
| Adjusted R-squared | | 0.9776 | |
| F-statistic | | $1.304e+04$ | |
| $p$-value | | $< 2.2e-16$ | |
| AIC | | -805.5532 | |
| BIC | | -794.4418 | |

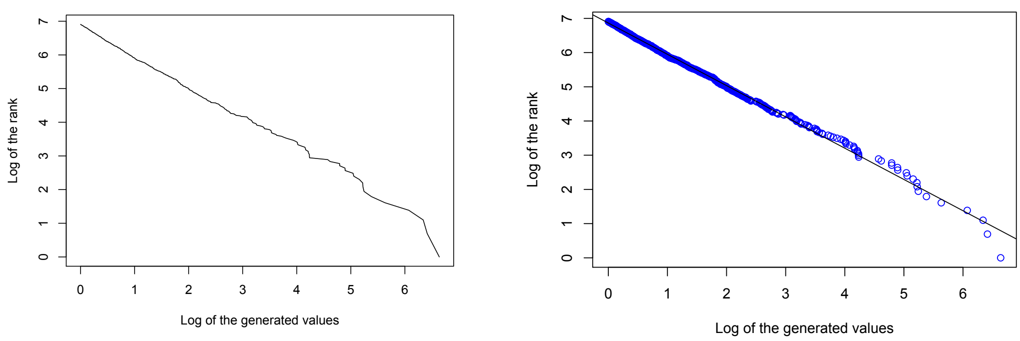
It is however bad practice to stop there and claim that the distribution at hand follows a power law. Although this tool is often used to fit power laws to empirical data, there are several problems with it.

- A first problem is the estimation of the errors: the formula for the computation of the standard error on the slope of a regression line is only appropriate when the assumptions of linear regression hold. Meaning it must include independent, Gaussian noise in the dependent variable at each value of the independent variable. When we do the fitting to the logarithm of a histogram, however, the noise, yet independent, is not Gaussian. The noise in the frequency estimates $p(x)$ themselves is Gaussian, but the noise in their logarithms is not.

- A second problem resides in the fact that a fit to a power law distribution can justify a large fraction of the variance even when the fitted data do not follow a power law, and hence high values of $r^2$ cannot be considered as evidence in favor of the power law form. Distributions like the log-normal, for example, can approximate a power law over many orders of magnitude, yielding high values of $r^2$. It is worth noting, however, that we can use a low value of $r^2$ in order to reject the power law hypothesis.

- A last problem is that we did not account for the fact that $\int_{x_m}^{\infty} p(x)dx$ should be equal to 1 when we did the regression. In general, the constant calculated via regression do not allow us to have a proper probability density function.

To illustrate the previous discussion, we generated a thousand random real numbers drawn from a power law distribution with exponent 2 (see Figure 1.4a) via the following method: for a random real number $r$ uniformly distributed in the range $0 \leq r < 1$, then $x = x_m(1-r)^{-\frac{1}{\alpha-1}}$ is a random power law distributed real number that lies in the range $x_m \leq x < \infty$ with exponent $\alpha$. Calculating the regression line in Figure 1.4b yields an estimated value of $1.9147 \pm 0.0015$ for $\alpha$ which differs from the real value 2.




(a) Log-log graph

(b) Regression line

Figure 1.4

### 1.3.2 Estimation via maximum likelihood

As mentioned in the previous subsection, using the plot on logarithm scales method is not adequate in order to fit power laws to empirical data. A better method for fitting parametrized models to observed data is the method of maximum likelihood.

From the assumption that the data at hand are drawn from a distribution that follows a power law for $x \geq x_m$, we can determine maximum likelihood estimators of the scaling parameter $\alpha$. The estimator in the continuous case (when $x$ is not a discrete integer variable) is given by

$$\hat{\alpha} = 1 + n \left[ \sum_{i=1}^{n} \ln \frac{x_i}{x_m} \right]^{-1} \tag{1.2}$$

where $x_i$ for $i = 1, \ldots, n$ are the observed values of $x$ such that $x_i \geq x_m$. The standard error is derived from the width of the likelihood maximum

$$\sigma = \frac{\hat{\alpha} - 1}{\sqrt{n}} + \mathcal{O}(\frac{1}{n}).$$

The discrete case is more complicated as it involves solving the following transcendental equation

$$\frac{\zeta'(\hat{\alpha})}{\zeta(\hat{\alpha})} = -\frac{\sum_{i=1}^{n} \ln x_i}{n}.$$

A transcendental equation is an equation that includes a function that basically cannot be written in terms of a finite sequence of the algebraic operations of addition, multiplication, and root extraction. Examples of such functions are $\sin(x), c^x, \log_c(x), \ldots$ with $c > 0$ and $c \neq 1$. Such equations often do not have closed-form solutions. It is however possible to approximate the solution by the following expression:

$$\hat{\alpha} = 1 + n \left[ \sum_{i=1}^{n} \ln \frac{x_i}{x_m - \frac{1}{2}} \right]^{-1}.$$

Returning to our previously generated data, we find, using Equation 1.2, $\hat{\alpha} = 1.96$ with standard error 0.03 which is a closer value from the correct one $\alpha = 2$ than what we found previously.

Now remark that we don't usually know what the value of $x_m$ is when treating empirical data. Therefore, before calculating the estimate of the scaling parameter $\alpha$, all samples below $x_m$ need first to be discarded so that we are left with only those for which the power law model holds. Thus, in order for the estimate of $\alpha$ to be conclusive, we also need an accurate method for estimating $x_m$. Choosing a value for $x_m$ that is too low will lead to a biased estimate of the scaling parameter as we will be attempting to fit a power law model to non-power law data. On the other hand, choosing a value for $x_m$ that is too high leads us to throw away legitimate data points. For example, taking $x_m = 10$ in our previous example leads to an estimate of the scaling parameter $\hat{\alpha} = 1.86$ with standard error $\sigma = 0.08$. Actually, even taking a smaller value like $x_m = 2$ already leads to a worst estimate $\hat{\alpha} = 1.92$ with standard error $\sigma = 0.04$, showing how quickly the estimate of the scaling parameter depreciates when considering a wrong value of $x_m$. A method for approximating $x_m$ is to minimize the distance between the power law model and the empirical data, in particular an estimate of $x_m$ could be the value of $x_m$ that *minimize* the Kolmogorov-Smirnov statistics

$$D = \max_{x \geq x_m} |F_n(x) - F(x)|$$

where

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{x_i \leq x}$$

is the empirical cumulative distribution function for the observations with value at least $x_m$ (the term $1_A$ is the indicator function yielding 1 when $A$ is true and 0 otherwise), and $F(x)$ is the cumulative distribution function for the power law model that best fits the data in the region $x \geq x_m$.

Let us now go back to our analysis on the data related to Belgian city population and let us apply on them the estimation techniques we just saw. Results are shown in Table 1.2 and show that the Kolmogorov-Smirnov statistics is minimized when the lower bound is $x_m = 10700$ and take the value 0.0722. We also find $\hat{\alpha} = 1.56$ with error $\sigma = 0.09$ which corresponds to the findings we had when doing the linear regression. Figure 1.5 shows the line resulting from these estimates.

Table 1.2: Parameters estimation summary

| | Value |
|---|---|
| Goodness of fit | 0.0722 |
| $x_m$ | 10700 |
| $\alpha$ | 1.56 |
| $\sigma$ | 0.09 |

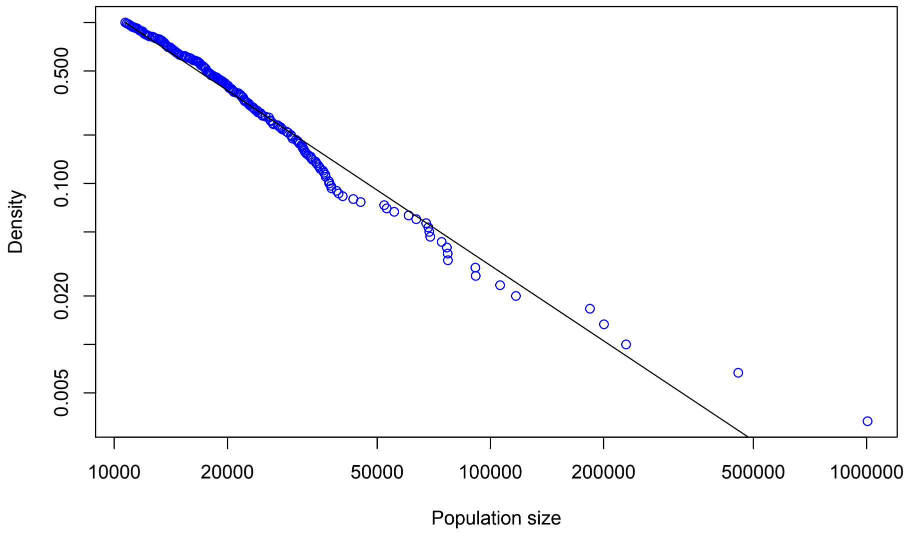



Figure 1.5: Data with fitted distribution

### 1.3.3 Inferring on the model

Note that the techniques presented here only allow us to fit a power law distribution to a given data set and provide estimates of the parameters $\alpha$ and $x_m$. But they do not indicate about whether the power law is a credible fit to the data. We need some way to tell whether the fit is a good match to the data. Without going into details, all those can be found in [CSN09], this hypothesis whether the observed data set actually follows a power law is tested using a goodness-of-fit test, via a bootstrapping procedure. Essentially, hypothesis tests are performed by generating multiple data sets (with parameters $x_m$ and $\alpha$) and then "reinferring" the model parameters. These generate a $p$-value that quantifies the plausibility of the hypothesis.

In order to avoid confusion, please remember that the $p$-value can be interpreted as the probability of being wrong if we reject the hypothesis. Here, the latter is that the observed data set actually follows a power law, so a high result for the $p$-value (say $p > 0.05$) means that there is a significant chance of being wrong by rejecting the hypothesis that our data set indeed follows a power law, in which case we would not discard the hypothesis. Returning to our data and applying the method for testing the power law hypothesis gives 0.11. It means that we have roughly less than 1 chance out of 10 of being wrong to discard the power law hypothesis. In our present concern, we shall not reject the power law hypothesis.

But there is a last point to take into account. Although we could not reject the power law hypothesis, we may still be able to find other distributions, such as an exponential or a log-normal, that might give a better fit. The way to proceed would thus be to reiterate the previous points to find the best fit for another distribution and the corresponding Kolmogorov-Smirnov statistic and finally calculate a $p$-value. If the $p$-value is sufficiently small, we could then rule out the distribution as a model for our data.

In many practical circumstances, however, we are only interested in knowing which distribution, from a list of possible candidates, is the better fit. In our case, we would like to confront the power law distribution with some alternatives distributions. Fortunately, there exist methods that can compare two distributions against each other and which are easier to implement than the KS test. One such method is the *(log)-likelihood ratio test*.

Basically, the idea of the likelihood ratio test is to compute the ratio of the likelihoods of the two competing distributions, or equivalently the logarithm $\mathcal{R}$ of the ratio. The ratio may thus be either positive or negative depending on which distribution is better, or zero in case of a tie between those two.

Basing the result of our analysis on the sign of the log-likelihood ratio alone, however, is not recommended as it will not definitively indicate which model is the better fit. Indeed, this ratio, being a statistical quantity, is subject to statistical fluctuation. In order to decide between the distributions, we should take into account the "size" of the log-likelihood ratio meaning that we want it to be sufficiently positive or negative to avoid the possibility of it being the result of a fluctuation from a true result that is close to zero.

In order to be able to say something significant about whether the observed value of $\mathcal{R}$ is sufficiently far from zero, we need to calculate the standard deviation $\sigma$ on $\mathcal{R}$. This can be done using a method suggested by Vuong [Vuo89]. It gives a $p$-value that can be seen as the probability of being wrong to reject the hypothesis that the observed sign of $\mathcal{R}$ is not statistically significant. To sum up, a "low" $p$-value indicates that we can take the sign as a reliable indicator.

For illustration purpose, we confronted the power law distribution with an exponential distribution, a log-normal distribution and a Poisson distribution. Results are shown in table 1.3. From the tests, we can see that we can discard the exponential and Poisson distributions in favor of the power law distribution. Although it seems at first that the log-normal distribution is a better fit for our data set due to the negative value of the log-likelihood ratio, the $p$-value prevents us from totally rejecting the power law distribution. We thus end up our analysis with two possible candidates, the power law distribution and the log-normal distribution, that appear to be two valid choices to fit the data.[4]

[4]It is worth noting that applying the same technique to the same data set but cut to only include cities with more than 30000 inhabitants shows the power law to be the more descriptive distirbution.

Table 1.3: Parameters estimation summary

| | Exponential | Log-normal | Poisson |
|---|---|---|---|
| Log-likelihood ratio | 2.04 | -0.40 | 1.99 |
| $p$-value | 0.04 | 0.69 | 0.05 |

## 1.4 Additional examples in economics and finance

In addition to the distribution of city sizes in a country or the distribution of income and wealth, we would like to finish this chapter by presenting two other examples of empirical power laws that arise in the context of economics and finance.

### 1.4.1 Firm sizes

Using US Census Bureau data, Axtell [Axt01] analyzes the distribution of firm sizes (measured by number of employees). After putting firms in bins according to their size, he plots the logarithm of the number of firms within a bin. Figure 1.6 shows the result where the line is the linear regression where parameters are calculated by ordinary least square. The result suggests a power law with $\alpha = 2.059 \pm 0.054$.

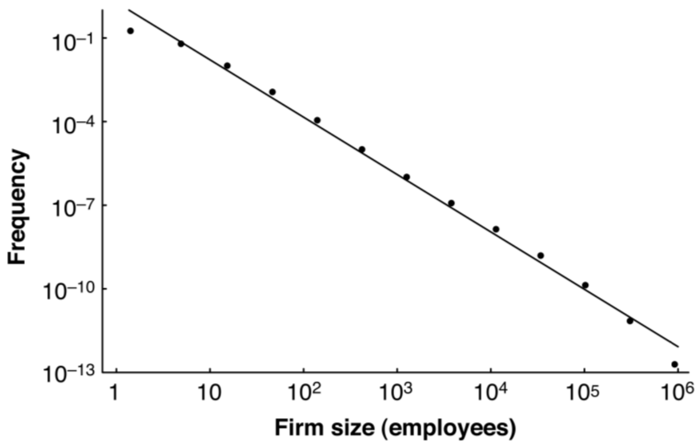



Figure 1.6: Histogram of U.S. firm sizes.
*Source*: Axtell [Axt01].

### 1.4.2 Stock market returns

Known as the *cubic law of stock market returns*, this power law is consistent with

$$P(|r_t| > x) = \frac{a}{x^\beta}$$

with $\beta = 3$. It corresponds to the finding that the probability of discovering extreme values is larger than for a Gaussian distribution of the same mean and standard deviation. Figure 1.7 shows on the left subgraph the distribution for four different sizes of stocks while on the right subgraph is shown the distribution of normalized stock returns, which is calculated as the stock returns divided by their standard deviation. The interesting behavior is here that, after this normalization, the four different distributions appears to follow the same curve. It shows that after some rescaling is done, different systems behave in the same way, allowing for some universal representation.

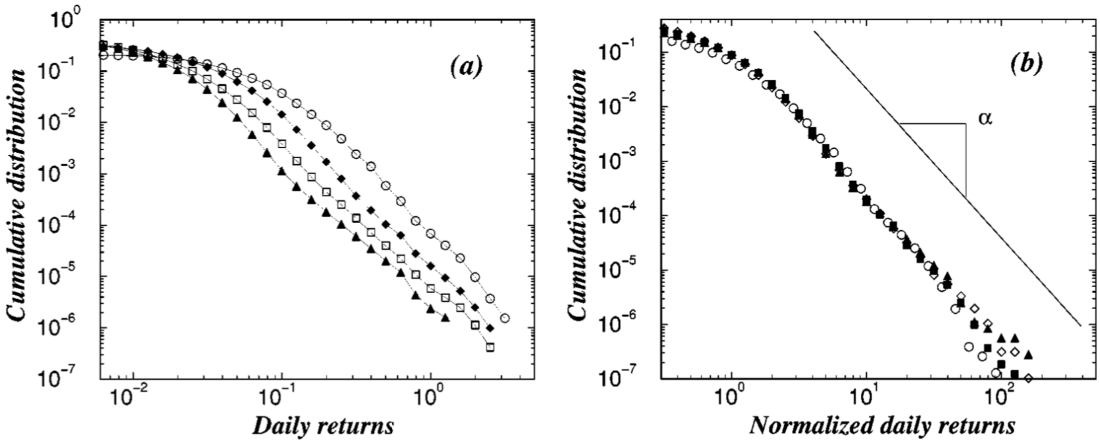



Figure 1.7: Cumulative distribution of daily stock market returns.
*Source*: Plerou [PGA+99].

# Chapter 2

# Generating power laws

In this chapter, we will present various mechanisms yielding power law distributions. In particular, we will focus on two specific mechanisms, the first one being the *proportional random growth* while the second is the *critical phenomena process*. Other mechanisms like the *highly optimized tolerance* mechanism [CD99] received a lot of attention but will unfortunately not be presented in the current work.

The main literature reviewed for these subjects is the following : information about the proportional random growth process can be found in [Sim55, Gab09, Gab16, Mit04] while the reader may find more material about critical phenomena in [BCD05]. Finally we suggest [New05, Sor06, FG08, SR11] which tackle both processes.

## 2.1 Proportional random growth and Yule-Simon process

The *Yule-Simon process* is a model first studied by Yule for modeling the distribution of the sizes of biological taxa (see [WY22, $\mathrm{Y}^{+}25$]). This process was then generalized by the economist Simon [Sim55] in order to analyze a class of distribution functions arising in a wide spectrum of empirical data. These distributions include the distributions of words in prose samples by their frequency of occurrence, the distributions of cities by population in particular, the distributions of incomes by size or the distributions of biological genera by number of species, . . . . This mechanism was rediscovered later in the study of the world wide web growth dynamic [BA99] under the name of *preferential attachment mechanism*. Other names can be found such as the *Matthew effect* [Mer68] or *cumulative advantage* [Pri76].

In the present work, we present the model as described by Simon in [Sim55] for word frequencies.

Assume a book is being written and contains currently $k$ words and write, for $k \geq 1$ and $i \geq 1$, $f(i,k)$ the number of different words that appears exactly $i$ times in the first $k$ words. To illustrate this, let us take the first line of the book The Hobbit by J. R. R. Tolkien:

"In a hole in the ground there lived a hobbit."

In this example, there are exactly 10 words where we find in particular $f(2,10) = 2$ as there are two different words "in" and "a" that are written exactly two times. Also, we find $f(x,10) = 0$ for $x > 2$ and $f(1,10) = 7$.

Now let us assume that the probability that the $(k+1)$-*th* word is a word that has already appeared exactly $i$ times before is proportional to $if(i,k)$. Note that $if(i,k)$ is simply the total number of instances of all words that have appeared exactly $i$ times. The probability is thus given by $K(k)if(i,k)$ where $K(k)$ is the factor of proportionality that we assume depends on the number $k$ of words.

Further assume that there is a constant probability $\alpha \in ]0,1[$ that the $(k+1)$-*th* word is in fact a new word that never occurred before.

The first assumption we made allows us to write the following for $i = 2, \ldots, k+1$:

$$E[f(i,k+1)] = f(i,k) + K(k)(i-1)f(i-1,k) - K(k)if(i,k).$$

Indeed, there is exactly three cases that might appear by adding the $(k+1)$-*th*word: the $(k+1)$-*th* word might be a word that previously appeared exactly $(i-1)$ times, that happens with probability $K(k)(i-1)f(i-1,k)$, and in that case $f(i,k+1) = f(i,k)+1$; or the $(k+1)$-*th* word is a word that previously occurred exactly $i$ times, for which the probability is $K(k)if(i,k)$, and thus $f(i,k+1) = f(i,k)-1$; or it can finally be a whole different case than the two latter cases, that happens thus with probability $1-K(k)(i-1)f(i-1,k)-K(k)if(i,k)$, in which case $f(i,k+1) = f(i,k)$. So adding everything together yields

$$\begin{aligned}
E[f(i,k+1)] &= (f(i,k)+1)K(k)(i-1)f(i-1,k) \\
&\quad + (f(i,k)-1)K(k)if(i,k) \\
&\qquad + f(i,k)\left[1-K(k)(i-1)f(i-1,k)-K(k)if(i,k)\right] \\
&= f(i,k)K(k)(i-1)f(i-1,k) + K(k)(i-1)f(i-1,k) \\
&\quad + f(i,k)K(k)if(i,k) - K(k)if(i,k) \\
&\qquad + f(i,k) - f(i,k)K(k)(i-1)f(i-1,k) - f(i,k)K(k)if(i,k) \\
&= f(i,k) + K(k)(i-1)f(i-1,k) - K(k)if(i,k).
\end{aligned}$$

Note that in the case where $i = k+1$, $f(i,k) = 0$ but the same argument remains applicable as $f(x,x)$ can either be 1 or 0 (also remark that $f(1,1)$ is always 1).

Now according to both of the assumptions, we have

$$E[f(1,k+1)] = f(1,k) + \alpha - K(k)f(1,k).$$

Again, the $(k+1)$-*th* word can be a new word with probability $\alpha$ so that we have the equality $f(1,k+1) = f(1,k)+1$ or it could be a word that only occurred once before, which happens with probability $K(k)f(1,k)$, in which case $f(1,k+1) = f(1,k)-1$ while in the other scenarios, $f(1,k+1) = f(1,k)$.

From now on, we will write $f(i,k+1) = E[f(i,k+1)]$. This allows us to write the *master equation* in the following form:

$$f(i,k+1) = f(i,k) + K(k)(i-1)f(i-1,k) - K(k)if(i,k), \tag{2.1}$$

which holds for all $2 \leq i \leq k+1$, the exception being when $i = 1$ where it takes the particular shape:

$$f(1,k+1) = f(1,k) + \alpha - K(k)f(1,k). \tag{2.2}$$

Let us now try to find the factor $K(k)$. In order to do this, let us consider the sum $\sum_{i=1}^{k} K(k)if(i,k)$. Each member $K(k)if(i,k)$ of this sum is the probability that the $(k+1)$-*th* word is a word that has already appeared exactly $i$ times so the sum of these probabilities turns out to be the probability that the $(k+1)$-*th* word is not an entirely new word, thus

$$\sum_{i=1}^{k} K(k)if(i,k) = 1-\alpha.$$

But $\sum_{i=1}^{k} K(k)if(i,k)$ can also be written as $K(k)\sum_{i=1}^{k} if(i,k)$ where the latter sum is just the total number of occurrences of each different word in a text of $k$ words which is simply $k$, thus

$$\sum_{i=1}^{k} K(k)if(i,k) = K(k)k.$$

So we get the following relationship

$$K(k)k = 1-\alpha$$

which lets us find

$$K(k) = \frac{1-\alpha}{k}. \tag{2.3}$$

### 2.1.1 Master equation and power law

In order to solve Equations 2.1 and 2.2, we make use of a simpler but non-rigorous approach. We recommend reading Appendix A that contains the details of the derivations present in this part.

The main idea is to assume

$$\frac{f(i,k+1)}{f(i,k)} = \frac{k+1}{k} \tag{2.4}$$

for all $k > 1$ and $1 \leq i \leq k$ to help us find a solution. This last assumption induces that

$$\frac{f(i,k+1)}{f(i,k)} = \frac{k+1}{k} = \frac{f(i-1,k+1)}{f(i-1,k)}$$

so that in particular, for $k > 1$ and $2 \leq i \leq k$.

$$\frac{f(i,k+1)}{f(i-1,k+1)} = \frac{f(i,k)}{f(i-1,k)}$$

and thus this ratio does not depend on $k$. We write

$$\beta(i) = \frac{f(i,k+1)}{f(i-1,k+1)}. \tag{2.5}$$

Using the relations 2.3 and 2.4 in Equation 2.2 yields

$$f(1,k) = \frac{\alpha k}{2-\alpha}. \tag{2.6}$$

Let us now turn to Equation 2.1. Using the relations 2.3, 2.4 and 2.5, we can derive

$$\frac{k+1}{k}f(i,k) = f(i,k) + \frac{1-\alpha}{k}(i-1)\frac{f(i,k)}{\beta(i)} - \frac{1-\alpha}{k}if(i,k).$$

This allows us to find a working expression of the factor $\beta(i)$:

$$\beta(i) = \frac{(1-\alpha)(i-1)}{1+(1-\alpha)i}.$$

Introducing

$$\rho = \frac{1}{1-\alpha},$$

the last expression can be simplified to

$$\beta(i) = \frac{i-1}{\rho+i}. \tag{2.7}$$

Let us go back to 2.6, the expression can also be written as

$$f(1,k) = \frac{\alpha k \rho}{\rho+1},$$

and by definition of $\alpha$, $\sum_{i=1}^{k} f(i,k) = \alpha k = n_k$ is the total number of different words in the $k$ first words so we can write

$$f(1,k) = \frac{\rho n_k}{\rho+1}. \tag{2.8}$$

Developing the relation 2.5

$$f(i,k) = \beta(i)f(i-1,k) = \beta(i)\beta(i-1)f(i-2,k) = \cdots = \beta(i)\beta(i-1)\ldots\beta(2)f(1,k),$$

and using 2.7 and 2.8 gives

$$f(i,k) = \frac{\Gamma(i)\Gamma(\rho+1)}{\Gamma(i+\rho+1)}\rho n_k,$$

where $\Gamma(n+1) = n!$ is the gamma function. The gamma function itself can be written using the beta function $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ to give

$$f(i,k) = B(i,\rho+1)\rho n_k. \tag{2.9}$$

We need now to verify that Equation 2.9 is indeed a solution of 2.1. Substituting 2.9 in 2.1 we find on the right hand side

$$B(i,\rho+1)\rho n_k + \frac{1-\alpha}{k}\left[(i-1)B(i-1,\rho+1)\rho n_k - iB(i,\rho+1)\rho n_k\right]. \tag{2.10}$$

Now note that it is possible to rewrite

$$(i-1)B(i-1,\rho+1) = (i+\rho)B(i,\rho+1),$$

so expression 2.10 can be written as

$$B(i,\rho+1)\rho\left(\alpha k + (1-\alpha)\rho\alpha\right) = B(i,\rho+1)\rho\alpha(k+1). \tag{2.11}$$

The last expression is exactly the left hand side of 2.1 since

$$f(i,k+1) = B(i,\rho+1)\rho n_{k+1} = B(i,\rho+1)\rho\alpha(k+1).$$

This shows that Equation 2.9 is indeed a solution of 2.1.

In fact, it also shows that the solution can be approximated by a power law. Since the beta distribution $B(i, \rho + 1)$ is in particular a power law distribution (it can be shown that it behaves as $i^{-(\rho+1)}$) and $f(i, k)$ is proportional to $B(i, \rho + 1)$, we can conclude that $f(i, k)$ follows a power law with exponent $\rho + 1$.

It is worth noting that an interesting possible extension of the Yule-Simon process includes considering the case where $\alpha$ is a function of $k$. In [Sim55], two cases are developed: the first assumes that $\alpha(k) = 0$ as soon as $k$ is larger than a certain number $k_0$ so that the flow of new words swiftly ceases after a certain number $k_0$ of words being introduced, while the second case assumes that $\alpha$ decreases with $k$ which is saying that the flow of new words decreases with the number of words already introduced.

### 2.1.2 General form

Let us wrap up this section by giving the general form of the mechanism which could thus be applied to various other topics such as genera sizes, cities sizes, paper citations, links to web pages and others.

Consider a system composed of a collection of objects, such as genera, cities, papers, web pages and so forth. The number of these objects will grow every so often as cities grow up or people publish new papers. Each object also has some property $k$ associated with it, such as the number of species in a genus, people in a city, or citations to a paper, that is believed to obey a power law, the latter we wish to explain. Newly appearing objects have some initial value of $k$ which we will denote $k_0$. For example, new genera initially have only a single species $k_0 = 1$, but new towns or cities might have quite a large initial population before being considered as a city or town in its own right, $k_0 = 100$ people might do so. Note that the value of $k_0$ could even be zero in some instances: newly published papers do not have any citation at first.

In between the appearance of one object and the next, a number $m$ of new species, people, citations, links ... are added to the entire system. Some cities or papers will get new people or citations, but not necessarily all will gain new properties. And in the simplest case these are added to objects in proportion to the number of properties the object already has. Thus the probability of a city gaining a new member is proportional to the number already there; the probability of a paper getting a new citation is proportional to the number it already has. Note that in this last case, since $k_0 = 0$, we have a problem that is easily countered by assigning new citations in proportion of $k + c$ for some constant $c$ and not simply $k$. This *rich-get-richer* mechanism seems in many cases like a natural process. For example, a paper that already has many citations is more likely to be discovered during a literature search and hence more likely to be cited again, same goes for a website that already has a lot of links pointing to it, this website will be ranked higher in search engines.

Thus this mechanism revolves around three key parameters that are $k_0$, $c$ and $m$. The process could be summarized as follow:

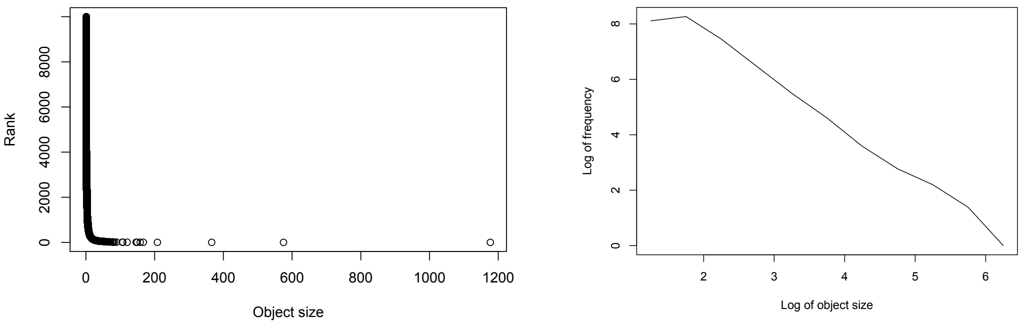
- Start with one object having initial value $k_0$ (but viewed as having $k_0+c$ existing properties),
- Repeat the following process:

[*] with probability $p = \frac{1}{1+m}$, a new object with initial value $k_0$ is created;

[*] with probability $1 - p = \frac{m}{1+m}$, the property of object $i$ is increased by one with probability proportional to the number of already existing properties that belong to object $i$.

We propose to do some simulations of this process for different values of $k_0$, $c$ and $m$ in order to visualize the impact of each variable on the mechanism.

First we simulate the mechanism for the creation of 10000 objects with the following settings: $k_0 = 1$, $c = 0$ and $m = 2$. Results are shown in Figure 2.1 where on the $x$ axis of graph 2.1a is displayed the size of the objects that is measured by the number of properties that ends up in the object. This setting will be our basis to understand the impact of changes in the parameters.

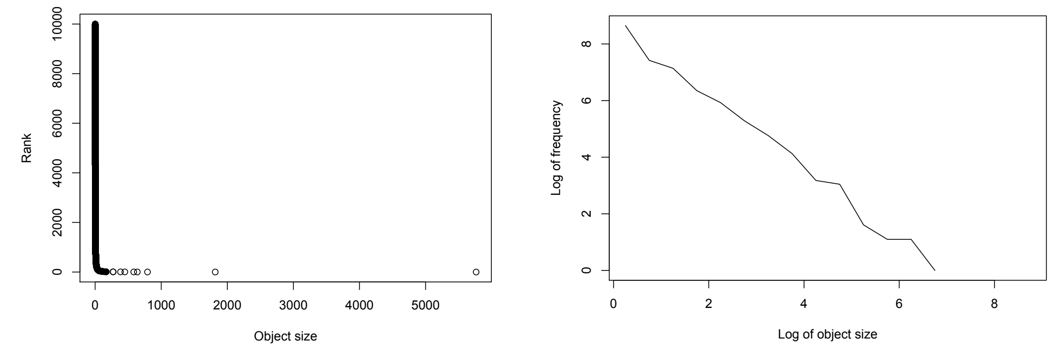



(a) Object size distribution

(b) Histogram in log form

Figure 2.1: Simulation for $k_0 = 1$, $c = 0$ and $m = 2$

In Figure 2.2 is displayed the same experiment with $k_0 = 1$, $c = 0$ and $m = 3.5$. As we can see, as the probability of creating a new object is smaller due to the increase in $m$, the larger object has now a bigger size than before.

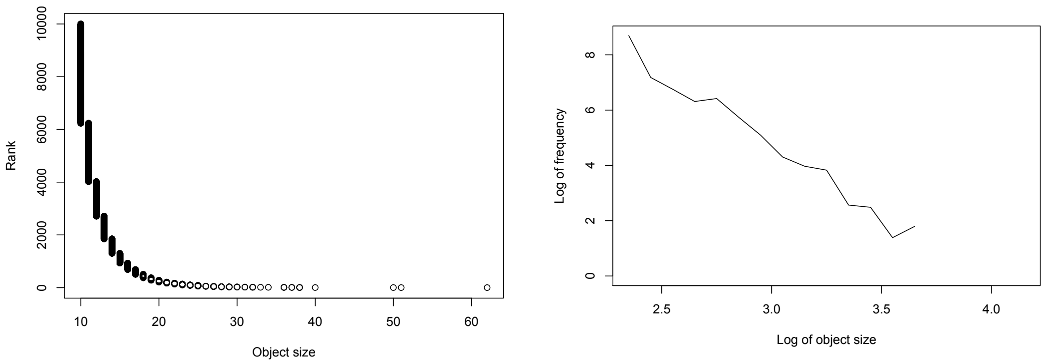



(a) Object size distribution

(b) Histogram in log form

Figure 2.2: Simulation for $k_0 = 1$, $c = 0$ and $m = 3.5$

The effect of increasing the value of $k_0$ is shown in Figure 2.3. As we can see, the size of the objects becomes way smaller but, at the same time, the reversed J shape of the distribution dies slowly compared to what can be seen in Figure 2.1. The idea behind this phenomenon is simply that the new objects start with a greater weight compared to already existing ones and thus the preferential attachment mechanism distributes the new properties more equally between the existing objects.

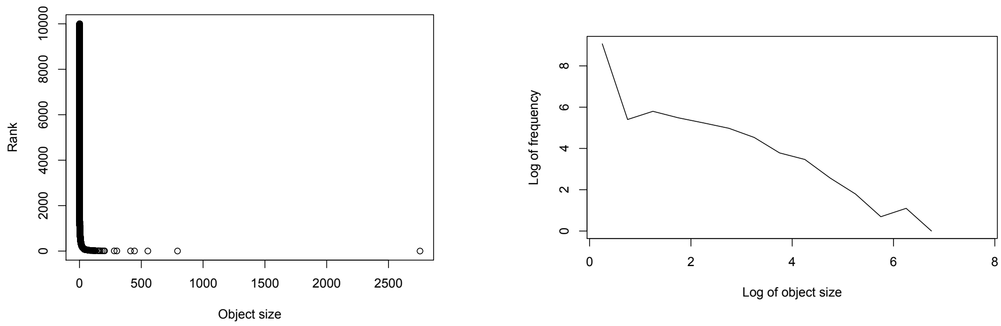



(a) Object size distribution

(b) Histogram in log form

Figure 2.3: Simulation for $k_0 = 10$, $c = 0$ and $m = 2$

Finally, by increasing $c$ to 5 and letting $k_0 = 1$ and $m = 2$ we recover the fast decreasing distribution due to the existence of an object that is way bigger than the others. This is in accord with the theory as the increase of $c$ is just an increase of how we view the already existing objects (they have more attracting power) which leads to greater attachment to these objects in particular.




(a) Object size distribution

(b) Histogram in log form

Figure 2.4: Simulation for $k_0 = 1$, $c = 5$ and $m = 2$

## 2.2 Critical points and percolation process

In this section we will present a totally different mechanism that generates power laws: the *percolation process*, which was founded in a paper of Broadbent and Hammersley [BH57] in 1957. This mechanism then got a lot of attention from the physics community, particularly in geophysics and transport of fluids in porous media.

To illustrate the percolation model with an intuitive example, consider a piece of rock which is subject to alterations due to the application of stress in the presence of water (corrosion) and possibly other processes. At first, we will see single isolated microcracks appear, and then with the increase of the power or time of stress applied to the piece of rock, we will see these microcracks grow in size and multiply. This will lead to an increase of the density of cracks per unit volume. The outcome is that microcracks begin to come together until a *critical density* of cracks is reached at which a main fracture is produced.

The basic idea is that the formation of microfractures prior to the major failure, the divergence of the system, plays a crucial role in the fracture mechanism. The precise point at which the major failure is produced is called the *critical point* or *percolation point*.

Let us further explain what we mean by "major failure". For this, consider a square lattice as in Figure 2.5. Imagine at first that the whole lattice is colored in grey, indicating that all cells are occupied. Suppose now that we go through all the cells of the lattice and for each one of those, we keep the cell intact with probability $p$ and we empty the cell with probability $1-p$, leaving the cell blank so that we are now able to go through this space.

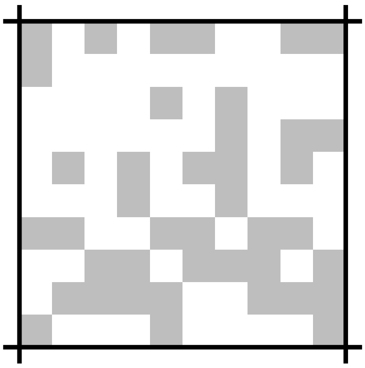


Figure 2.5: Square lattice $10 \times 10$

The question that we ask ourselves now is whether there exists a path, starting from the top of the lattice, across the lattice that ends up at the bottom of the lattice? In our previous example, we see that there is no such path. This is illustrated by Figure 2.6 where we can see that the blue cells are blocked at some level and unable to go lower in the lattice.

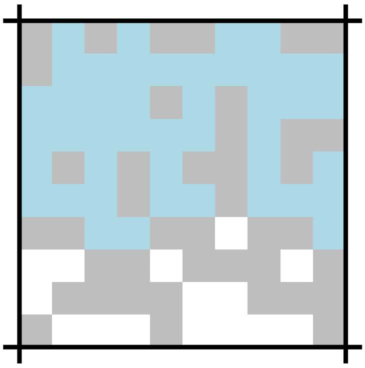


Figure 2.6: Square lattice $10 \times 10$

A high value of $p$ would not introduce enough space to be able to go across the lattice while a small value of $p$ would enable us to go through it very easily as we can see in Figure 2.7 where the blue cells are able to go from top to bottom.

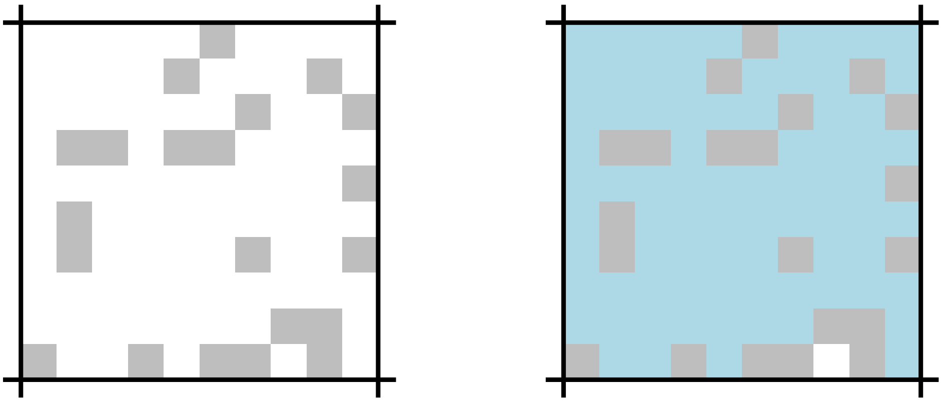


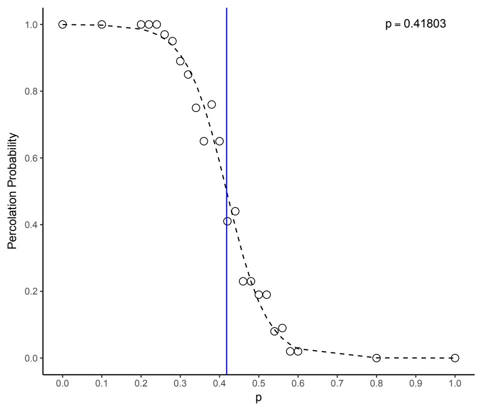
Figure 2.7: Square lattice with $p = 0.2$

If we denote $P_N(p)$ the probability that a path across a lattice of size $N \times N$ exists when a fraction $p$ of the cells are left filled in, it could be interesting to know what is the graph of such probability as a function of $p$. To do that, we create the following simulation: for each value of $p$ (we obviously take a discretize version of the interval $[0, 1]$) we replicate 100 times the experiment of trying to find a path across the lattice and store the results, we then calculate the estimated probability of being able to find such a path for this particular value of $p$ and we finally plot these results in a graph.

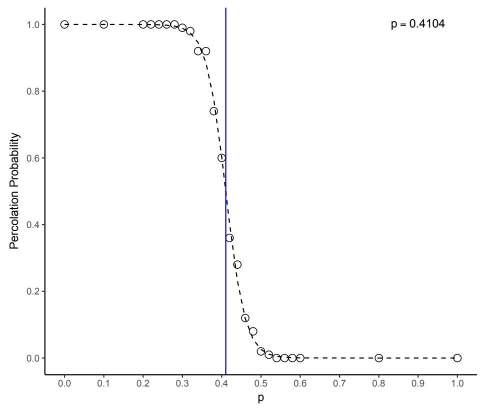
We show in Figure 2.8 the different simulations done by varying the size $N$ of the lattice. An interesting property of the graphs at hand is the shape of the curve when $N$ increases: as we can see, increasing $N$ leads to a sharper turn from always being able to find a crossing path in the lattice to not being able to find such a path. In the limiting case, when $N \to \infty$, it would give a value from which we would instantly shift from the case where there is always a crossing path to the case where there is never a crossing path. The value of $p$ for which the shift occurs is called the *critical value*. The shift is often called *phase transition* as it describes the transfer from one state to another. From the simulations, we find different approximations for the value $p$ given in Table 2.1. The critical value given in [New05] is $0.5927462\ldots$ but corresponds to the probability $1 - p$ in the present context so we find that the critical value should be $p = 0.4072537\ldots$ which we can see our estimated values tending towards to. Figure 2.9 displays two examples in a lattice of size $100 \times 100$ when $p = 0.4072538$.

Table 2.1: Estimated critical value

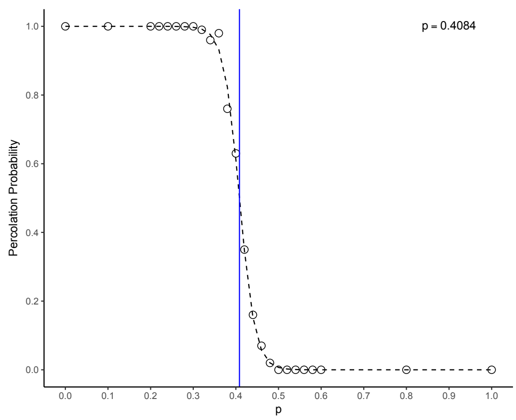
| | Critical value |
|---|---|
| $N = 10$ | 0.4180307 |
| $N = 25$ | 0.4104043 |
| $N = 40$ | 0.4084002 |




(a) Lattice of size $10 \times 10$

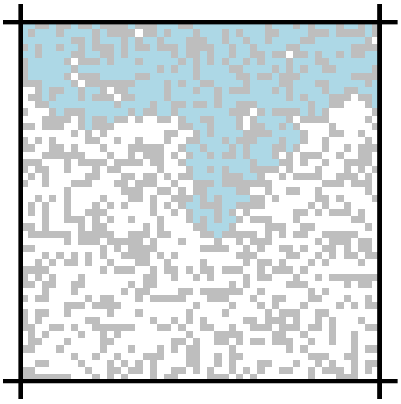



(b) Lattice of size $25 \times 25$

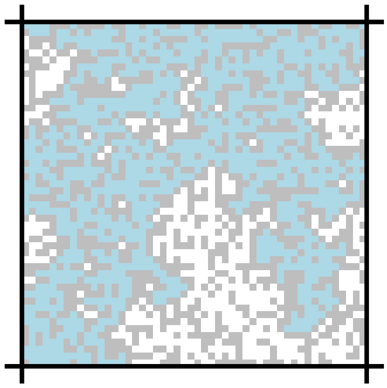



(c) Lattice of size $40 \times 40$

Figure 2.8



(a) No crossing path



(b) Crossing path

Figure 2.9: Lattice $100 \times 100$ with $p = 0.4072538$

### 2.2.1 Rescaling to find a power law

So far we still don't see any power law appearing. It remains thus to answer the question of how this mechanism could generate power laws. Let us go back to the arguments presented in [New05].

To answer that question, let us first change the perspective of the previous problem by now considering *clusters* of *non-colored cells*. More precisely, we are interested in the mean area $\overline{s}$ of the cluster to which a randomly taken cell belongs. When the lattice is fully occupied, that mean area is 0 while it is $aN^2$ when the lattice is fully emptied provided that $a$ is the area of a single cell in the lattice.

The previous question of finding the critical value for which finding a crossing path in the lattice would shift could be restated in terms of the mean area $\overline{s}$: it is equivalent to find the value $p$ for which there is a shift in how the mean area $\overline{s}$ depends on the size of the lattice. Let us explain further, when $p$ is very small, the lattice is almost fully occupied and the mean area $\overline{s}$ will be close to 0. But when $p$ is sufficiently high so that we can find a crossing path that goes through the whole system, the mean area $\overline{s}$, that is influenced by this path, will now depend on the size of the system (see Figure 2.9b), meaning that a larger system will yield a larger value of $\overline{s}$.

Let us now focus on the distribution of cluster sizes and write $p(s)$ the probability that a randomly chosen cell belongs to a cluster of area $s$. Since $p(s)$ is a probability, it has no units, *i.e.* it is a measure without dimensions. However $s$ has a dimension, say square meter for example. Thus in order to create a dimensionless quantity from one that has a unit, we should cancel the unit by making a ratio with a quantity that has the same unit. In the problem at hand, we have 2 others quantities that have the same unit as $s$: the area $a$ of a single cell and the mean area $\overline{s}$ of the cluster to which a randomly chosen square belongs. Note that the area of the whole lattice is not to be taken into account as we are considering the case when $N \to \infty$. We can therefore see $p(s)$ as a function of the following ratios: $\frac{s}{a}$, $\frac{a}{\overline{s}}$ and $\frac{s}{\overline{s}}$. Since the latter is already a function of the first two, we can write

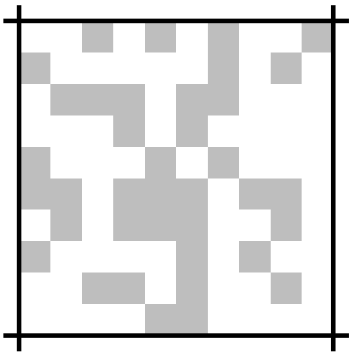
$$p(s) = Cf\left(\frac{s}{a}, \frac{a}{\overline{s}}\right) \tag{2.12}$$

where $C$ is a normalizing constant in order to have $\sum_s p(s) = 1$.

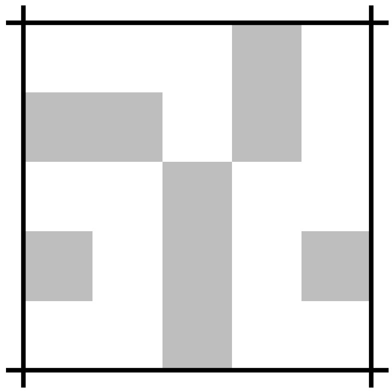
We apply the following distortion: we *simplify* or *rescale* the lattice in order to change the fundamental unit of the lattice. We could for example double the size of the unit cell $a$. Figure 2.10 shows the idea behind the rescaling when $a$ is multiplied by 4. The argument is an asymptotic one: it will therefore only be really valid for larger clusters and, in consequence, the power law will only be exhibited in the tail of the distribution. What is important is that after rescaling, the areas should still be approximatively unchanged, thus the probability $p(s)$ should be unchanged as well as the mean area $\overline{s}$. The only variable that has changed is $a$ that has become $\frac{a}{b}$ and thus we must have

$$p(s) = C'f\left(\frac{s}{\frac{a}{b}}, \frac{\frac{a}{b}}{\overline{s}}\right) = C'f\left(\frac{bs}{a}, \frac{a}{b\overline{s}}\right)$$

for some other normalizing constant $C'$. If we compare this equation with Equation 2.12, we see that this equation behaves (up to a certain constant) like $p(bs)$ but in a system where the mean area is given by $b\overline{s}$.



(a) Before rescaling



(b) After rescaling

Figure 2.10: Rescaling by quadrupling the area $a$

But as we said before, we are interested in a particular point: the critical point. By its definition, this is exactly the point where $\overline{s} \to \infty$ and thus the point at which we can link $p(s)$ and $p(bs)$ the following way:

$$p(s) = C'f\left(\frac{bs}{a}, 0\right) = \frac{C'}{C}Cf\left(\frac{bs}{a}, 0\right) = \frac{C'}{C}p(bs). \tag{2.13}$$

But this expression is exactly like Equation 1.1

$$p(bs) = g(b)p(s)$$

with $g(b) = \frac{C}{C'}$. As we previously saw, this in particular implies that the distribution of $s$ follows a power law.

We would like to finish this part with a few notes. First a note on the machinery used in this section to derive the power law: although the arguments were imprecise, the idea was only to give a hint of the method used to derive the power law; the reader interested in a more accurate discussion is referred to the method of the *renormalization group* [Wil75]. Second is the fact that this model produces power law in very specific circumstances, which is a disadvantage as it is thus unlikely to explain generic power law distributions seen in Chapter 1. It seems indeed inconceivable that if we could find some model for those distributions, the parameters of the real world would fall precisely at the point where the shift occurs. Nonetheless, some authors [BTW87] have studied dynamical systems that actually arrange themselves to fall at the critical point no matter what state we start off. In this case, one says that the system *self-organize* to the critical point.

### 2.2.2 Percolation in network theory

In this subsection, we will see how the previously mentioned phase transition can be recovered in the context of network theory. Indeed, one can think of percolation as the physical mechanics of contagion, therefore perfectly suited to try to answer questions like: ”How infectious does a strain of flu have to be to create a pandemic?”. The main idea being that the full spread of a rumor or information among a set of agents or traders depends on the connectivity that exists between the agents, exactly like in the previous case of the lattice.

In particular, we will give an example of its appearance in the theory of *complex network* which is an emerging field of research that is now extending to many disciplines such as physics, engineering, biology, sociology and economics. A common feature shared by many systems encountered in these different fields is that they can be represented by a graph with the nodes representing a set of individual entities and the links symbolizing the interactions between these entities. Here we will interest ourselves with properties of *typical graphs*, meaning that we are interested in answering questions that are somewhat universal in the fact that they do not hold for a specific graph. The objects used to study such questions are *random graphs*. In the present work, we will focus on one particular model used to generate random graphs: the *Erdös–Rényi random graph model* [ER59].

We propose to go through the Erdös–Rényi random graph model and see where the percolation appears in such a model. First let us recall that a *graph* $G$ is an ordered pair $(V, E)$ where $V$ is a set of vertices or nodes and $E$ is a set of edges, each edge being associated with two vertices. Now there are two versions of the Erdös–Rényi random graph model:

- The first version, due to Erdös and Rényi [ER59] is the $G(N, M)$ model, it assigns equal probability to all graphs that have $N$ nodes and $M$ edges. In the case where $N = 3$ and $M = 2$, there are exactly 3 possible graphs displayed in Figure 2.11, and to each one is thus assigned the probability 1/3.

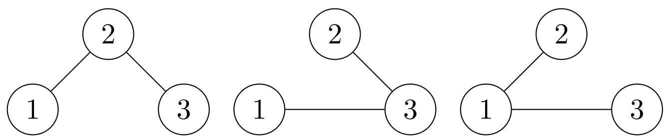
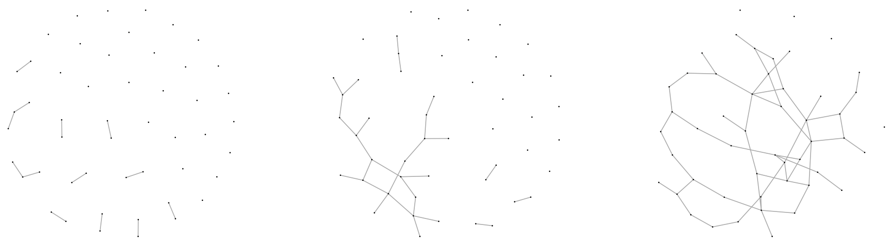



Figure 2.11: Graphs with 3 nodes and 2 edges

- The second version, due to Gilbert [Gil59] is the $G(N, p)$ model. Here we construct a graph by connecting nodes at random. For each pair of nodes, we include an edge between those nodes with probability $p$. Note that the probability is the same for each pair of nodes, it does not decrease nor increase with the number of edge that were already included. We will nevertheless assume that $p$ decreases as $N$ increases. We display examples of such graphs for $N = 50$ and different values of $p$ in Figure 2.12.

These models are closely related, yet the random graph $G(N, p)$ appears to be somewhat more easier to manipulate due to the independence of edges in the graph. Here we will thus focus on the second variation, namely the $G(N, p)$ model in order to give a brief idea on how percolation can arise in this setting.



(a) $p = 0.01$ (b) $p = 0.03$ (c) $p = 0.05$

Figure 2.12: Examples of $G(50, p)$ graphs

Further assume that we are interested in the limiting case $N \to \infty$, capturing the fact that real world complex networks are growing more and more everyday. Although we are not aiming at a fully rigorous discussion, let us explain an idea that could easily be missed when working with random graphs: in fact, a random graph, let's say $G(N, p)$, is a probability space $(\Omega, \mathcal{F}, P)$ where the sample space $\Omega$, set of all possible graphs with $n$ vertices, is of size $|\Omega| = 2^{\binom{N}{2}}$ with $\binom{N}{2}$ being the total number of possible edges that can be created in a graph with $N$ vertices; $\mathcal{F}$ is a collection of all the events we are interested in and $P$ is a probability measuring returning for each graph $G$ with $M$ edges

$$P(G) = p^M (1-p)^{\binom{N}{2}-M}.$$

Thus, by convention, a property of a random graph is always viewed as an average across all collections of outcomes, meaning that we are dealing with expected values. For example, if we are interested to find the mean number of edges in $G(N, p)$, one could consider the event $\{e \in G\}$ of edge $e$ being part of the graph $G \in G(N, p)$ and remark that the number of edges in the graph $G$ is given by

$$X = \sum_{e \in K_N} 1_{e \in G}$$

where $K_N$ is just the full graph with $\binom{N}{2}$ edges and $1_A$ is the indicator function returning 1 when $A$ is true and 0 otherwise. Now we can simply take expectations to find

$$E(X) = E(\sum_{e \in K_N} 1_{e \in G}) = \sum_{e \in K_N} E(1_{e \in G})$$

and since $E(1_{e \in G}) = P(e \in G) = p$ and $|K_N| = \binom{N}{2}$ we find

$$E(X) = \binom{N}{2} p.$$

Other properties of $G(N, p)$ include the degree distribution: for a given vertex $i$, we ask for the probability that its degree $d_i$, the number of edges incident to the vertex, is of size $k$. This follows a binomial with parameters $N - 1$, the number of available other vertices, and $p$:

$$P(d_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

Now since we are interested in large values of $N$, we can approximate the binomial distribution with a Poisson distribution with parameter $\lambda = (N-1)p$ or, since $N$ is large, $\lambda \simeq Np$, thus

$$P(d_i = k) \simeq \frac{\lambda^k}{k!} e^{-\lambda}.$$

One could also show that the expected mean degree of a vertex is given $\lambda = Np$.

Now let's consider $G(N, p)$ as a function of $p$. Trivially, the graph will be empty, meaning that there are no edges, when $p = 0$ while it will be full when $p = 1$. Looking at Figure 2.12, we get a glimpse at the percolation process that occurs in $G(N, p)$: when $p$ is under a certain threshold, the graph have multiple components that are all of very small sizes (they actually have size of order $\mathcal{O}(\log N)$) while a gigantic connected component, that is a connected component of size $\mathcal{O}(N)$, emerges exactly when $p$ reaches that threshold value, giving a hint at the percolation occurring in $G(N, p)$.

We end up this section with a non-rigorous argument to find the critical value $p$ at which the phase transition appears. Let $u$ be the fraction of vertices that do not belong to the giant component. This probability can be seen from another perspective: starting from a vertex, pick another one, there are exactly 4 distinct situations that are depicted in Figure 2.13. First the starting vertex $A$ can be linked to the other vertex $B$ or not and that other vertex can either belong to the giant connected component (noted by GC) or not (denoted ~~GC~~).

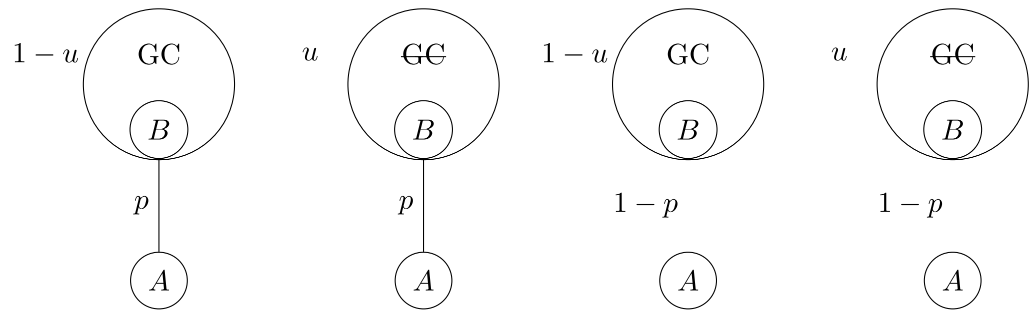



Figure 2.13: The four situations

In order for the first vertex not to be in the giant component we thus either need the two vertices not to be connected, which happens with probability $1 - p$ or to be connected with a node that is not in the giant component which happens with probability $pu$ so that the total probability is $1-p+pu$. Another way to put this is simply saying that the probability for the vertex $A$ not to belong in the giant connected component is the probability $pu$ of being linked to a vertex that does not belong to the giant component, to which we add the probability $(1-p)(1-u)$ of not being linked to a vertex that belongs to the giant component and we finally add the probability $(1-p)u$ of not being linked to a vertex that does not belong to the giant component which gives

$$(1-p)u + (1-p)(1-u) + pu = 1 - p + pu.$$

Now this should hold true for all $N-1$ other vertices in order for the vertex $A$ to be outside of the giant component, thus we have the connection

$$u = (1-p+pu)^{N-1}$$

where we can plug the relation $p = \frac{\lambda}{N}$ to get

$$u = \left(1 + \frac{\lambda}{N}(u-1)\right)^{N-1}.$$

Applying the logarithm on both sides

$$\log(u) = (N-1)\log(1 + \frac{\lambda}{N}(u-1)),$$

then using the famous first order Taylor approximation $\log(1+x) \simeq x$ when $x$ is small to simplify further

$$\log(u) = (N-1)\frac{\lambda}{N}(u-1)$$

and finally remarking that when $N \to \infty$, the ratio $\frac{N-1}{N} \to 1$ yields

$$\log(u) = \lambda(u-1) = -\lambda(1-u)$$

that we rewrite as

$$u = e^{-\lambda(1-u)}. \qquad (2.14)$$

Now note that $u = 1$ is always a solution of this equation for any value of $\lambda$. But the number of solutions depends on the value of $\lambda$ as we can see in Figure 2.14. The highest blue line is the plot of $e^{-\lambda(1-u)}$ for $\lambda = 0.5$, the middle blue line corresponds to $\lambda = 1$ and the lowest $\lambda = 1.5$. What appears from this graph (we could also make a more analytical argument by differentiating $e^{-\lambda(1-u)}$) is that there is only one solution, which is $u = 1$, for $\lambda \leq 1$, this solution corresponds to the fact that there is no gigantic connected component when $\lambda \leq 1$, while we find two solutions as soon as $\lambda > 1$.

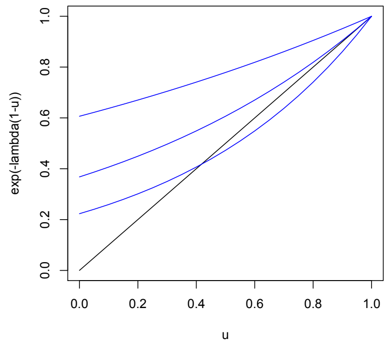



Figure 2.14: Plot of $e^{-\lambda(1-u)}$ for different values of $\lambda$

Now the previous argument do not say that a giant component exists when $\lambda > 1$. In the latter case, we could use the following reasoning: since $\lambda$ gives the expected mean degree of a vertex, starting from any vertex, it will be linked to $\lambda$ other vertices on average. Now since each of these vertices will be themselves connected to $\lambda$ other vertices on average, all of them will have $\lambda^2$ neighbors (see Figure 2.15) and we can continue like that to find that at step $t$ we would have $\lambda^t$ vertices that are at distance less or equal to $s$ from our starting vertex. Now if $\lambda > 1$, this number $\lambda^t$ would keep growing exponentially and thus most of the nodes would be connected together in a giant component. This corresponds exactly to the nonzero solution of Equation 2.14.

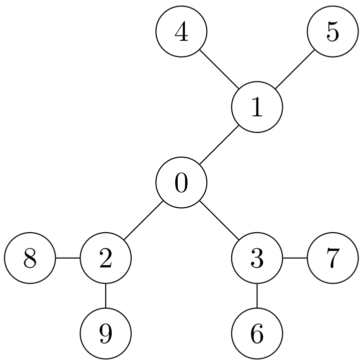
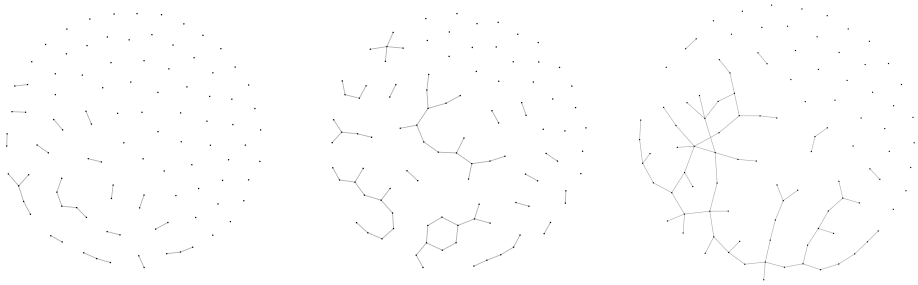



Figure 2.15: Giant component argument

The critical value $\lambda = 1$ corresponds to $p = \frac{1}{N}$. We finish this section with Figure 2.16 showing how the giant component appears when exceeding the threshold value $\frac{1}{100}$ in the $G(100, p)$ model.



(a) $p = 0.005$ (b) $p = 0.01$ (c) $p = 0.015$

Figure 2.16: Percolation in $G(100, p)$ graphs

# Conclusion

In this work, we studied power law distributions and hopefully managed to give a good insight of their strengths and weaknesses and illustrate two different ways that are able to generate such distributions. From a theoretical point of view, the mathematical object is deceptively simple and possesses very interesting properties like that of being a scale-free distribution and even being the only one verifying this property. Also, we saw that power laws emerge very naturally as real contenders to explain data sets that contains fat tails, like in the example of the distribution of Belgian towns and cities population. This makes the power law a good perfomer for a lot of economic data. It is in general quite difficult, however, to rule out other candidates like the log-normal or the exponential distribution, making its legitimacy very unstable. Moreover, although it could not be seen as a true weakness, the power law fails in most cases to explain all the data distribution and only applies to the tail of the latter, making it only able to give information in a certain subset of the distribution.

The two physical mechanism we presented for generating power laws includes the rich-get-richer mechanism in which the most crowded cities win more inhabitants over time in proportion to the number they already have and the critical phenomena in which some scale-factor of the system diverges. There is however still work to be done both experimentally and theoretically before we can claim that we truly understand the physical mechanisms leading to power laws.

# Bibliography

[Axt01] Robert L Axtell. Zipf distribution of us firm sizes. *science*, 293(5536):1818–1820, 2001.

[BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[BCD05] Ted Brookings, JM Carlson, and John Doyle. Three mechanisms for power laws on the cayley tree. *Physical Review E*, 72(5):056120, 2005.

[BH57] Simon R Broadbent and John M Hammersley. Percolation processes: I. crystals and mazes. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 53, pages 629–641. Cambridge University Press, 1957.

[BTW87] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, 59(4):381, 1987.

[CD99] Jean M Carlson and John Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E*, 60(2):1412, 1999.

[CSN09] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[ER59] Paul Erdös and Alfréd Rényi. On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[FG08] J Doyne Farmer and John Geanakoplos. Power laws in economics and elsewhere. In *Santa Fe Institute*, 2008.

[Gab09] Xavier Gabaix. Power laws in economics and finance. *Annu. Rev. Econ.*, 1(1):255–294, 2009.

[Gab16] Xavier Gabaix. Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1):185–206, 2016.

[Gil59] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[GMY04] Michel L Goldstein, Steven A Morris, and Gary G Yen. Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 41(2):255–258, 2004.

[Mer68] Robert K Merton. The matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810):56–63, 1968.

[Mit04] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

[New05] Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[PGA+99] Vasiliki Plerou, Parameswaran Gopikrishnan, Luis A Nunes Amaral, Martin Meyer, and H Eugene Stanley. Scaling of the distribution of price fluctuations of individual companies. *Physical review e*, 60(6):6519, 1999.

[Pri76] Derek de Solla Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology*, 27(5):292–306, 1976.

[Sim55] Herbert A Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[Sor06] Didier Sornette. *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media, 2006.

[SR11] Mikhail V Simkin and Vwani P Roychowdhury. Re-inventing willis. *Physics Reports*, 502(1):1–35, 2011.

[Vuo89] Quang H Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

[Wil75] Kenneth G Wilson. The renormalization group: Critical phenomena and the kondo problem. *Reviews of modern physics*, 47(4):773, 1975.

[WY22] John C Willis and G Udny Yule. Some statistics of evolution and geographical distribution in plants and animals, and their significance, 1922.

[Y+25] G Udny Yule et al. II.—A mathematical theory of evolution, based on the conclusions of dr. jc willis, fr s. *Phil. Trans. R. Soc. Lond. B*, 213(402-410):21–87, 1925.

[Zaj97] Daniel Zajdenweber. Scale invariance in economics and in finance. In *Scale invariance and beyond*, pages 185–194. Springer, 1997.