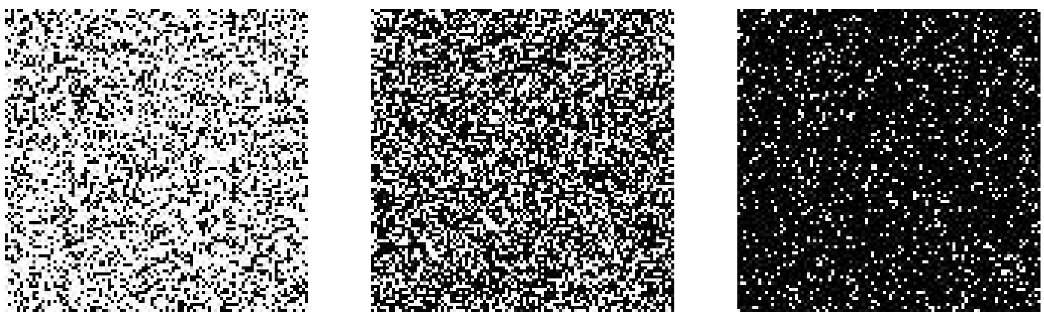
Phase transitions and power laws

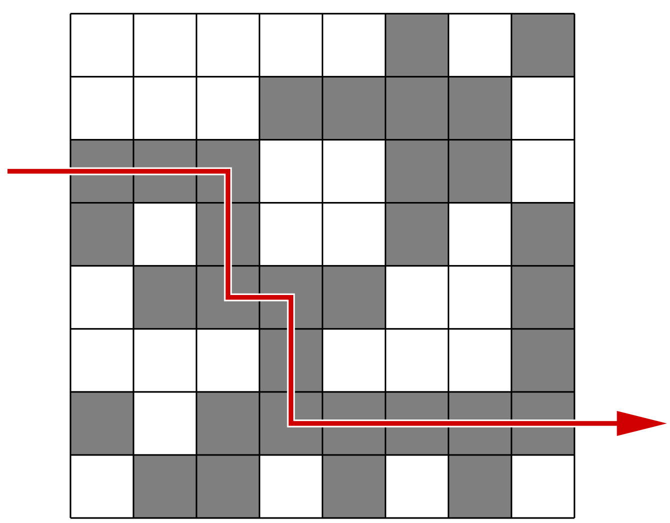
1 Phase transitions

Consider the percolation model:

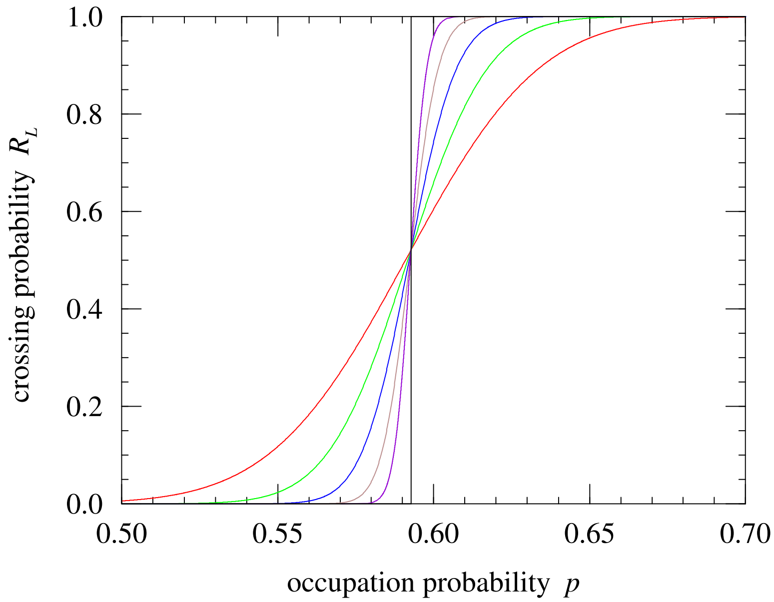


$p = 0.3$ $p = p_c$ $p = 0.9$

Is there a path across the lattice from one side to the other? Clearly if p is very low there is not. If it's high there is. Somewhere in between a path appears.

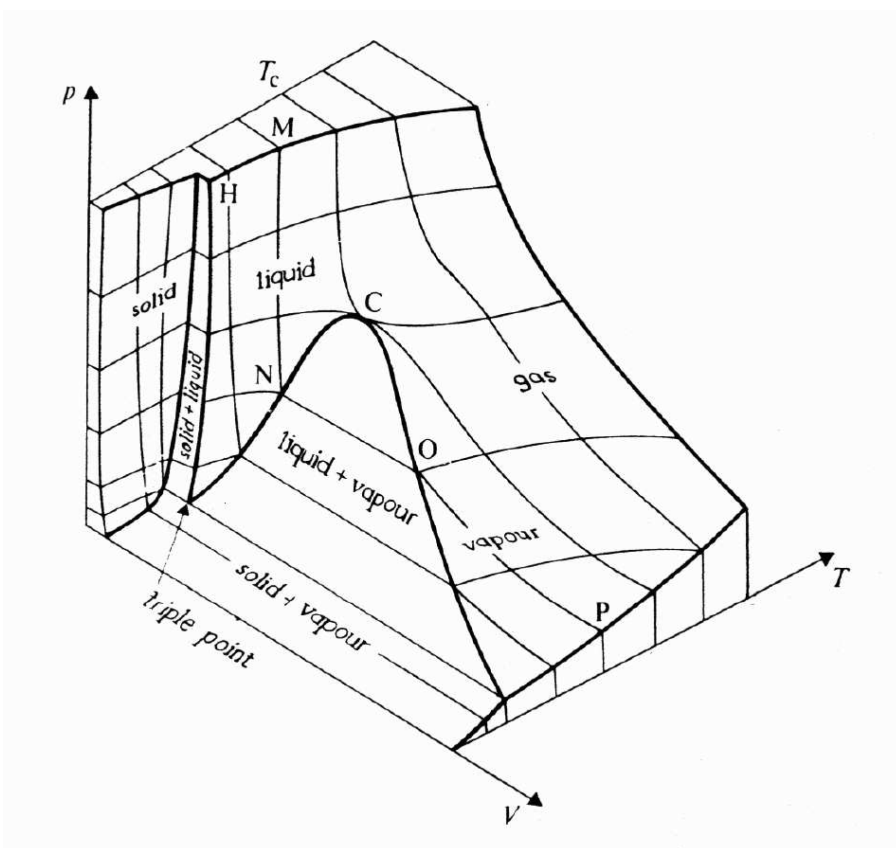


Let $R_L(p)$ be the probability that a path exists across a system of $L \times L$ sites when a fraction p of the sites are filled in. Here's what $R_L(p)$ looks like:



As L becomes large, the change from no-path to path becomes sharper and sharper. When $L \to \infty$, it is a step—an instantaneous transition. This is an example of a **phase transition**.

Another common example of a phase transition is the change from liquid water to steam or water vapor:



In fact there are two different transitions taking place here:

1. When liquid water changes to water vapor at constant temperature, we have a **first order phase transition**, which is characterized by a **coexistence region** in which water and vapor are seen simultaneously, and by a **latent heat**—a finite amount of work needs to be done in order to drive the system through the transition.

2. As temperature increases the size of the coexistence region diminishes and finally vanishes. The point at which it vanishes is a **continuous phase transition**, like the transition in percolation.

These appear as the line NO (first order transition) and the point C (continuous transition) in the figure.

Other continous phase transitions include:

- Curie point T_c of a magnet.
- The superconducting transition (either high- or low-T_c superconductors).
- The epidemic transition of a disease $R_0 = 1$.
- The error catastrophe of population genetics.
- The formation of a giant component in a random graph (the "small-world" transition).
- The solvability transition in computation complexity theory (e.g., in satisfiability).
- *Maybe* the frozen/chaotic transition in cellular automata (but probably not: the Game of Life, which is capable of universal computation, has a large but not infinite correlation length).
- *Maybe* the frozen/chaotic transition in Kauffman's NKC model of coevolution.

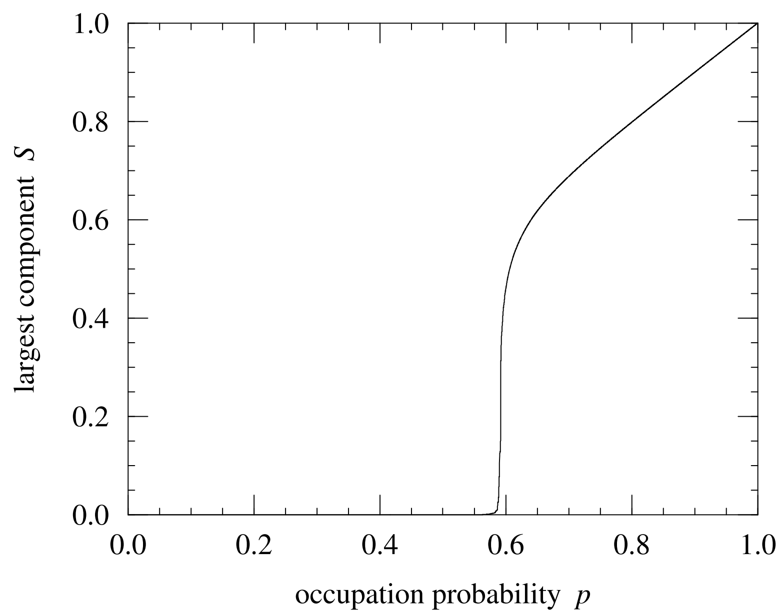
And many more, in all different subjects.

2 Scaling theory

All phase transitions have an **order parameter**, which is a quantity which is zero on one side of the transition and non-zero on the other:

system	independent variable	order parameter
percolation	site occupation probability	fractional size of spanning cluster
magnet	temperature	magnetization
superconductor	temperature	fraction of electrons in Bose condensate
disease	reproductive ratio	fraction of population affected by average outbreak
evolution	mutation rate	fraction of population at fitness optimum
random graph	mean degree	fractional size of giant component
satisfiability	ratio of variables to clauses	fraction of problems satisfiable

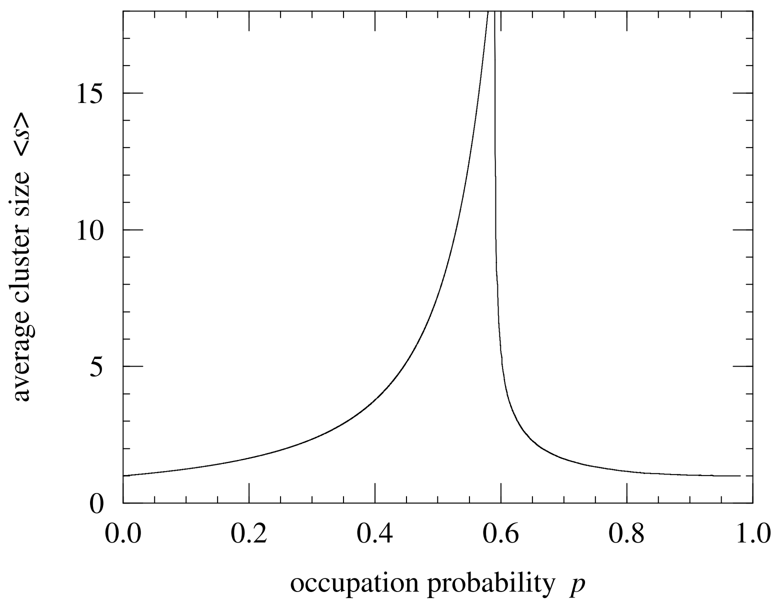
A continuous transition in one in which the order parameter varies continuously as we go through the transition point. Example, percolation:



It can have infinite gradient at the transition, and often does, but it cannot be discontinuous.

> At finite system sizes the transition is not instantaneous. Only in the limit of large system size do we get a sharp step at the transition.

If one calculates the average cluster size, this must diverge at the transition. This divergence is a classic example of a critical phenomenon:



We define $n_s(p)$ to be the density of clusters of size s at occupation fraction p, excluding the spanning clusters. It must have the scaling form

$$n_s = f(s/\langle s\rangle)g(s). \tag{1}$$

Important stuff: Consider the following scaling argument. If we change the *scale* on which we measure areas on our lattice by a factor of b, then all clusters change size according to $s \to bs$. Of course, the physics of the system hasn't changed, only how we measure it, so this change of variables cannot change the distribution n_s, except by a numerical factor to keep the normalization correct.

The argument of $f(x)$ doesn't change anyway, because s and $\langle s\rangle$ both change by the same factor b. But the argument of $g(x)$ does change. Thus $g(x)$ must satisfy

$$g(bx) = k(b)g(x), \tag{2}$$

where $k(b)$ is the numerical factor, which can depend on b but not x. Let us choose the normalization of g so that $g(1) = 1$. Then, setting $x = 1$ above we have

$$g(b) = k(b) \tag{3}$$

for all b and hence $k(x)$ and $g(x)$ are the same function. Thus

$$g(xy) = g(x)g(y). \tag{4}$$

To solve this equation, we take the derivative with respect to y:

$$\frac{\partial}{\partial y}g(xy) = xg'(xy) = g(x)g'(y), \tag{5}$$

then set $y = 1$ to get

$$xg'(x) = g(x)g'(1), \tag{6}$$

whose solution is

$$\log g(x) = g'(1) \log x + c, \tag{7}$$

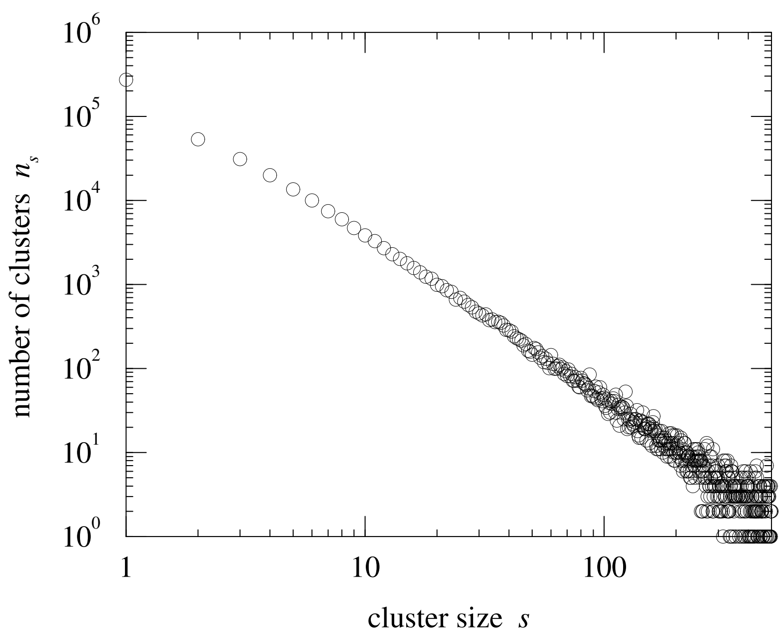
where c is an integration constant. Given $g(1) = 1$, we must have $c = 0$, and hence

$$g(x) = x^{-\tau}, \tag{8}$$

where $\tau = -g'(1)$. This functional form is called a **power law**. The quantity τ is a **critical exponent**.

> The distribution of cluster sizes becomes a power law exactly at the critical point. Indeed, the same arguments imply that all distributions will become power laws at the critical point. This is one of the characteristic features of phase transitions.

One of the nice things about power laws is that if $g(x) \sim x^{-\tau}$, then $\log g(x) \sim -\tau \log x +$ constant, so power laws give straight lines on log–log graphs:



3 Renormalization group

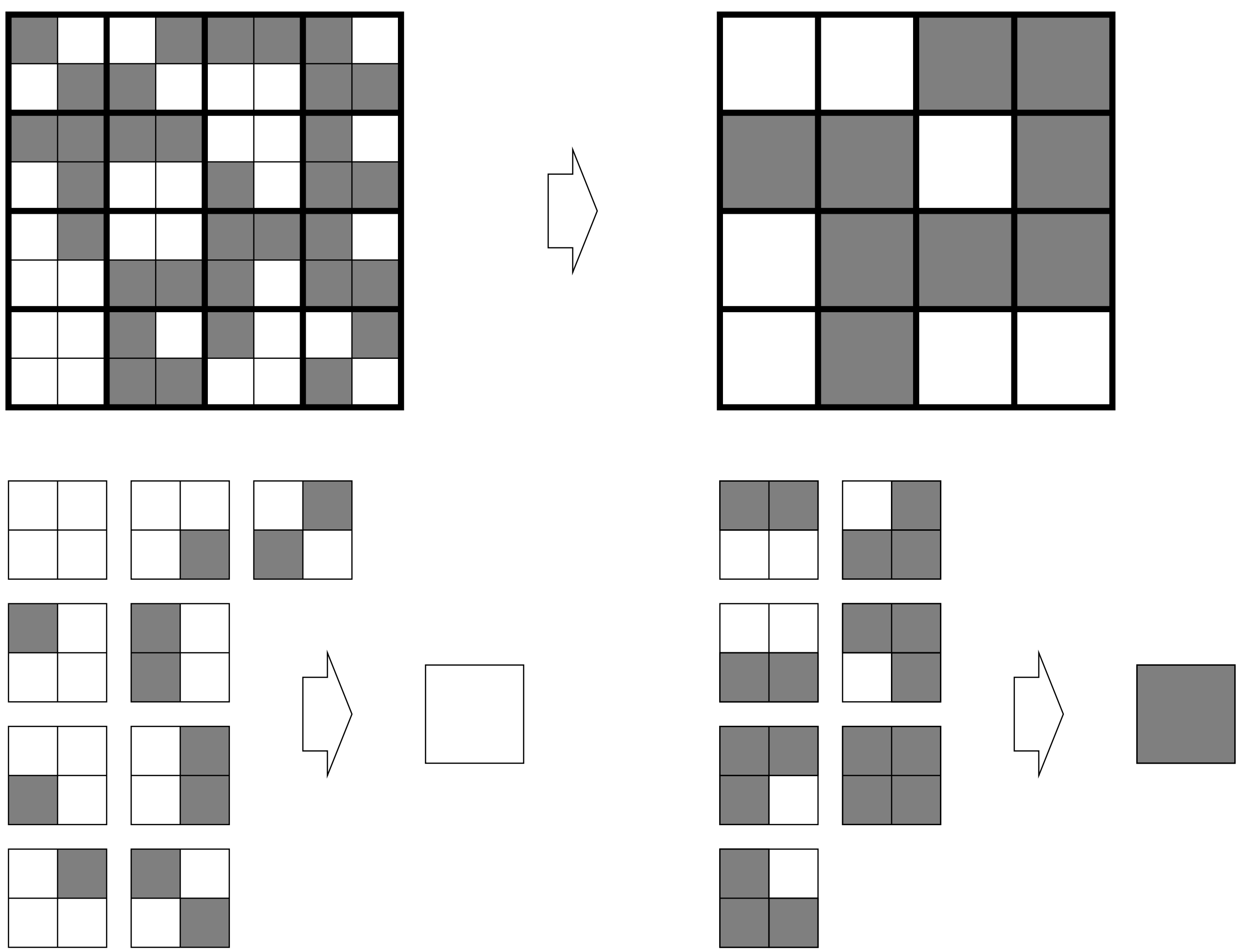
Calculating the properties of systems at or near the critical point was one of the abiding problems of twentieth century physics, until it was solved beautifully by Ken Wilson and Michael Fisher in the 1970s with their invention of the **renormalization group** (RG). Here's how you would use the RG to calculate the position of the phase transition in percolation.

3.1 RG for percolation

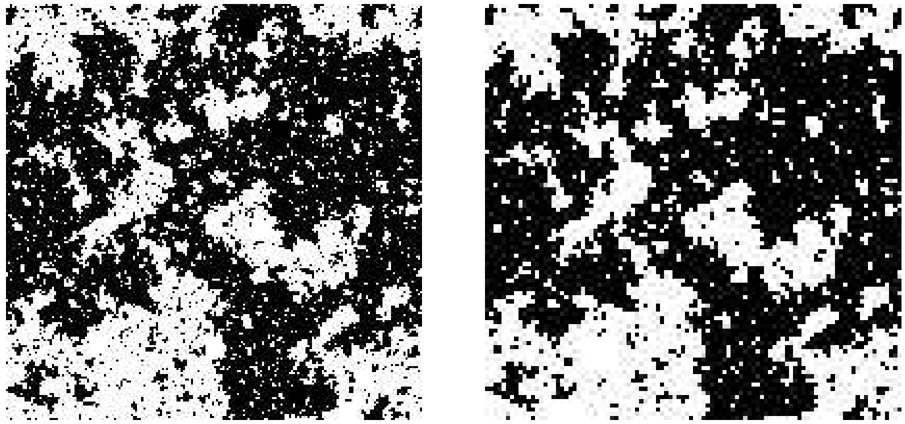
Here is a simple example of the (real-space) renormalization group for our percolation problem.

- Close to the critical point, the cluster size distribution becomes power-law, and averages over the distribution are dominated by the large-s contributions, i.e., by large clusters.
- The large clusters are invariant when we rescale the system.

Simple rescaling transformation:



On a large system:



As we can see, the transformation preserves most of the large-scale structure of the configuration, although a lot of the small detail is lost.

What is the occupation probability of the new state? The probabilities of the 4×4 blocks which map to an occupied site sum to

$$p' = 2p^2(1-p)^2 + 4p^3(1-p) + p^4 = 2p^2 - p^4. \tag{9}$$

If we are precisely at the transition point, then the distribution of cluster sizes doesn't change—it is a power law before and a power law after we rescale. This means that p_c is the point at which $p' = p$, or

$$p_c^4 - 2p_c^2 + p_c = 0, \tag{10}$$

which has solutions 0 and 1 (not likely), or

$$p_c = \frac{-1+\sqrt{5}}{2} = 0.618\ldots \tag{11}$$

The result from numerical simulations is $p_c = 0.593\ldots$, so we're within a few percent. This is typical of RG methods—the answers are pretty good with little effort, but the errors are rather uncontrolled.

4 Power laws

As we saw, a power law is a function of the form

$$f(x) = f(1)x^{-\tau}, \tag{12}$$

where τ is a constant. Power laws occur at critical points, but as we will see they occur elsewhere also. A straight-line form on logarithmic scales indicates a power law.

The power law is, in some respects, a rather surprising functional form. We expect to see exponential distributions, as arise from the maximization of the Gibbs entropy. Exponentials also arise in many other contexts, particularly in the probabilities of things happening many times and in solutions of first order differential equations. We also expect to see the Gaussian distribution, or its close relatives the binomial and Poisson distributions, which arise naturally through additive random processes and the central limit theorem.

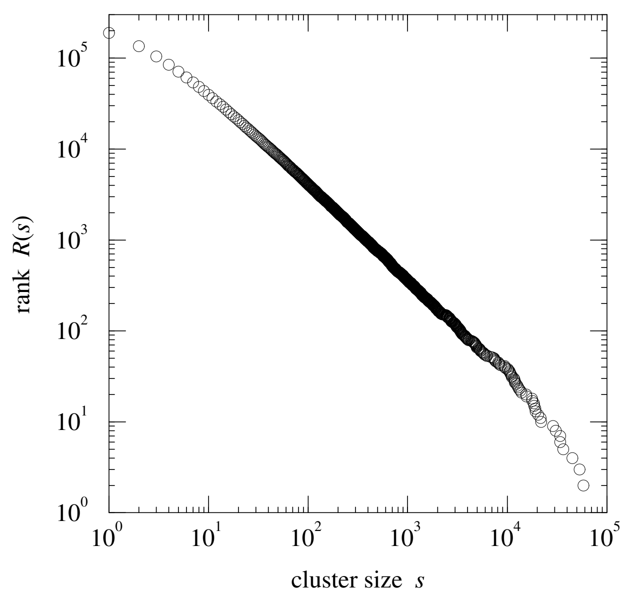
But power laws are a considerably rarer phenomenon in the real world. There are however quite a number of situations where they do crop up, and this leads to speculation about what their cause might be. Many physicists have suggested that critical phenomena may be an explanation for power laws in real-world data. In recent years the investigation of power-law forms in various systems has been one of the principal preoccupations of the complex systems community.

Rank/frequency plots: Suppose a certain quantity x has a power-law distribution, as in Eq. (12) above. Then the integral under the distribution from x to ∞ is

$$R(x) = f(1)\int_x^\infty y^{-\tau}\,\mathrm{d}y = \frac{f(1)}{1-\tau}x^{-\tau+1}. \tag{13}$$

This quantity is called the **rank**. If $f(x)$ is a histogram of x, then $R(x)$ is the number of measurements which had a value greater than or equal to x. To put it another way, if we number n measurements from 1 (greatest) to n (smallest), then the number given to a measurement x is $R(x)$. Above we see that if $f(x)$ is a power law, then so is $R(x)$. Often one plots a so-called **rank/frequency plot**, which is $R(x)$ plotted against x. On log-log scales this should be a straight line with slope $-\tau + 1$. This is better than making a histogram, because it doesn't require us to bin the data—each data point gets counted separately.

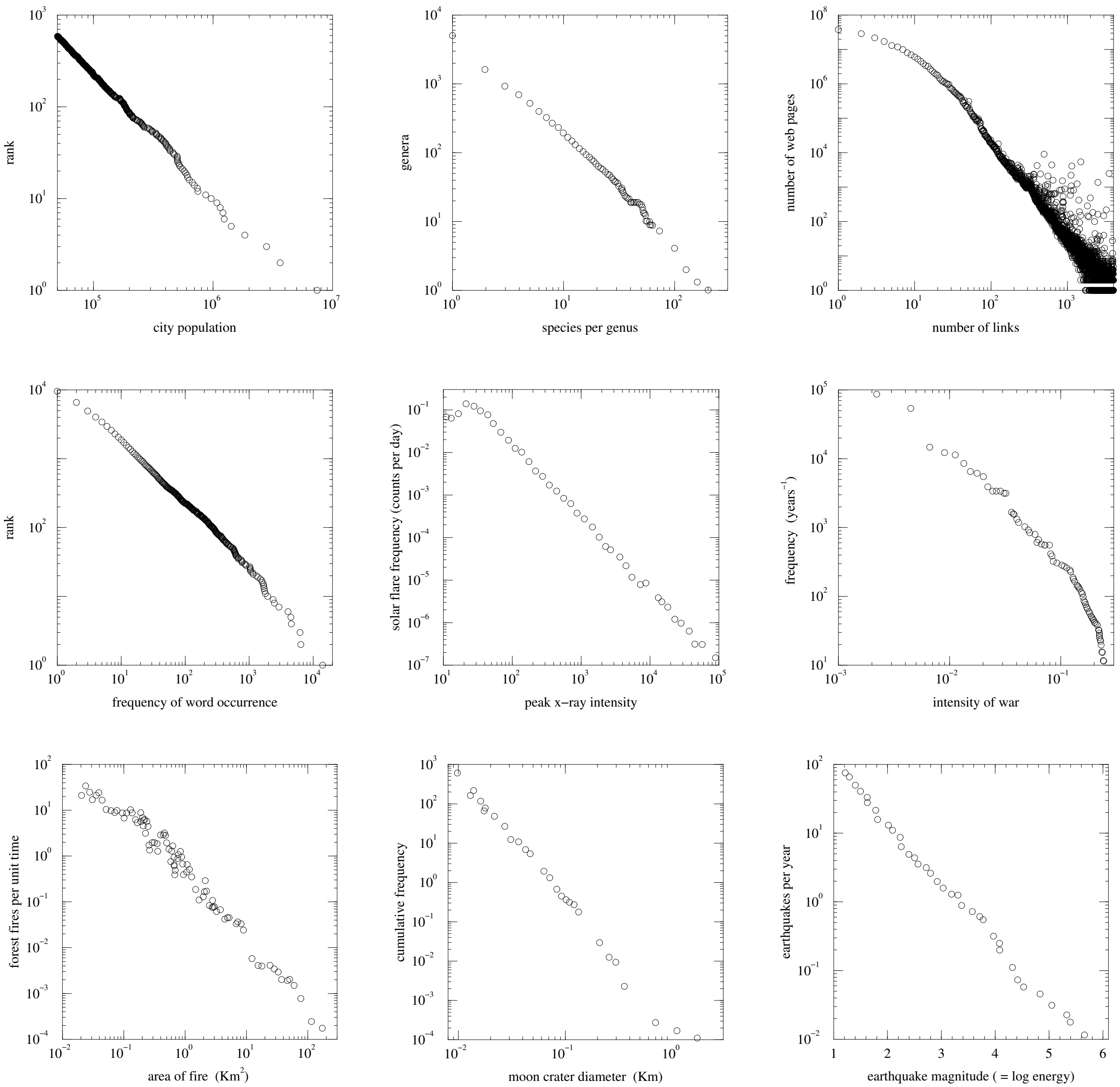
Here is a rank/frequency plot for our cluster size data:



Note that the thing that rank is plotted against is not necessarily a frequency. In the first such plots, the measured data were frequencies, and the name has stuck, but in most cases the independent variable is something other than frequency, such as here, where it is cluster size. Rank/frequency plots are also sometimes called cumulative distribution functions or cumulative histograms.

4.1 Examples of power laws

Here are some examples of data from various situations that show power laws. Some are normal histograms, some are rank/frequency plots, depending on how the original data were published.



Top left to bottom right, these show:

1. the populations of US cites with more than 50 000 inhabitants, from the US Census;
2. the number of species per genus of angiosperms (flowering plants), from Willis (1922);
3. number of links in pages on the world-wide web, from Broder *et al.* (1999);

4. frequency of occurrence of words in the English language, from *Moby Dick*;
5. peak intensity of solar flares, from Lu and Hamilton (1991);
6. normalized intensities of wars 1495–1973, from Roberts and Turcotte (1998).
7. areas burned by forest fires in the Australian Capital Territory, 1926–1991, from Malamud *et al.* (1998);
8. sizes of craters on the Moon, from Gehrels (1994);
9. magnitudes of earthquakes in the Southeastern United States, 1974–1983, from Johnston and Nava (1985);

Many other examples exist, including sizes of avalanches, sizes of meteors, wealth and income distributions, and fluctuations in economic indices.

So are all these power laws examples of critical phenomena? Some physicists would like to claim that they are. The most cursory inspection, however, reveals that this is an idiotic claim. There are many ways in which power laws can be produced, and only a few real power laws come from critical phenomena. What's interesting though is that the number of known ways of producing power laws is not very large. If you find a power law, like the ones above, the number of mechanisms that could be behind it is rather small, and so the mere existence of a power law gives you a lot of help in working out the physical mechanism behind an observed phenomenon.

4.2 Mechanisms for creating power laws

Here are the main known mechanisms by which power laws are created.

4.2.1 Critical phenomena

We saw this one earlier. This is the classic power-law-producing mechanism of statistical physics, but it has a big disadvantage as an explanation for natural phenomena: to see the power law you have to be exactly at the critical point. If you are not at the critical point then, as we saw, you get some more general form such as

$$n_s = s^{-\tau} f(s/\langle s\rangle), \tag{14}$$

which is usually not a power law. This makes ordinary critical phenomena a rather unlikely explanation of an observed power law. (But see "self-organized criticality" below.)

4.2.2 Combination of exponentials

Suppose that some quantity y has an exponential distribution:

$$p(y) \propto \mathrm{e}^{-\alpha y}. \tag{15}$$

But suppose that the real quantity we are interested in is x, which is given by

$$x = \mathrm{e}^{\beta y}. \tag{16}$$

Then the probability distribution of x is

$$p(x) = p(y)\frac{\mathrm{d}y}{\mathrm{d}x} \propto \frac{\mathrm{e}^{-\alpha y}}{\beta \mathrm{e}^{\beta y}} = \frac{x^{-\alpha/\beta-1}}{\beta}, \tag{17}$$

which is a power law with exponent $\tau = 1 + \alpha/\beta$.

Example: (Prof. Moore already mentioned this one.) Suppose that the frequency with which words are used in a language goes down exponentially with their length ℓ (on average):

$$f(\ell) \propto \mathrm{e}^{-\alpha\ell}. \tag{18}$$

But the number of possible words of length ℓ clearly goes up exponentially with length

$$n(\ell) \propto \mathrm{e}^{\beta\ell}. \tag{19}$$

So the number of words used with frequency f is

$$n \propto \mathrm{e}^{\beta\ell} = [\exp(-\alpha\ell)]^{-\beta/\alpha} = f^{-\tau}, \tag{20}$$

where $\tau = \beta/\alpha$. This is known as Zipf's law of word frequency.

4.2.3 Reciprocals of things

Suppose we are interested in a quantity x, which is proportional to the reciprocal of some other quantity y: $x = c/y$. And suppose y can take both positive and negative values, having some distribution $p(y)$ which passes through zero (and is smooth there). Then the distribution of x is

$$p(x) = p(y)\frac{\mathrm{d}y}{\mathrm{d}x} \simeq -\frac{c}{y^2} = \frac{-1}{c}\,\frac{1}{x^2}, \tag{21}$$

where the approximate equality pertains close to the origin. In fact, if x is *any* inverse power of y, then we get a power-law distribution in x.

A particular example of this is measurements of relative changes in quantities. Suppose we are interested in the fractional change

$$x = \frac{\Delta y}{y}, \tag{22}$$

in the quantity y. If Δy is distributed according to any smooth distribution, then the large values of x are dominated by cases where y is close to zero, and have the same power-law distribution $p(x) \propto x^{-2}$.

4.2.4 Random multiplicative processes

We know that random additive processes—ones in which a bunch of random numbers are added together—give results which are distributed according to a normal distribution. What happens if we *multiply* a bunch of random numbers?

$$x = \prod_{i=1}^{N} r_i \tag{23}$$

gives

$$\log x = \sum_i \log r_i, \tag{24}$$

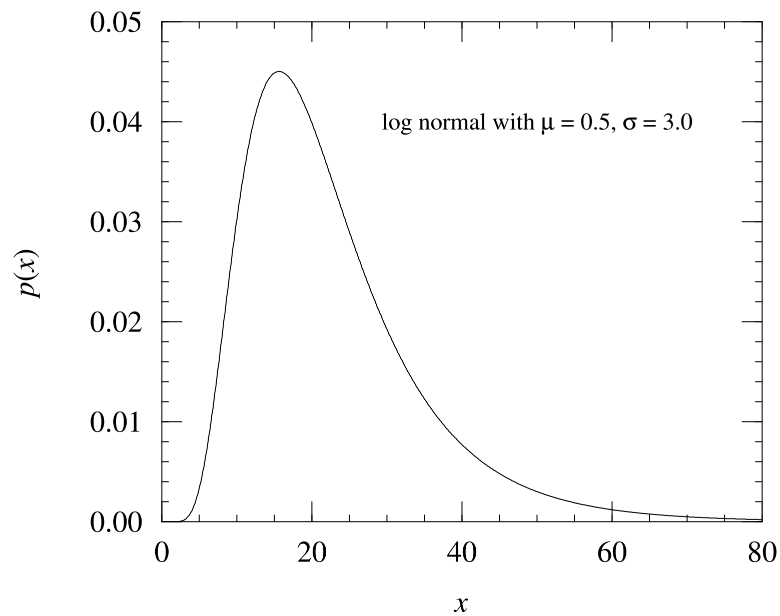
so $\log x$ is the sum of random numbers, and hence is normally distributed. In other words

$$p(\log x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(\log x - \mu)^2}{2\sigma^2}\right]. \tag{25}$$

Thus the distribution of x is

$$p(x) = p(\log x)\frac{\mathrm{d}\log x}{\mathrm{d}x} = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[\frac{(\log x - \mu)^2}{2\sigma^2}\right]. \tag{26}$$

This distribution is called the **log-normal** distribution.



(Note that formally there is also a divergence in the distribution at the origin. Does this matter?)

If $\log x$ is close to its mean μ, then the exponential is roughly constant, and the variation in $p(x)$ comes primarily from the leading factor of $1/x$. But the regime in which $\log x$ is close to its mean can correspond to a very large range of x. Suppose for example that we are multiplying together 100 numbers of typical logarithm 1. Then $\mu \simeq 100$, and $\sigma \simeq 10$. This means that the exponential will be roughly constant in the range $90 < \log x < 110$, which corresponds to a range of x from 1×10^{39} to 3×10^{43}, which is more than four orders of magnitude. We would expect to see a good power-law with slope -1 over this range.

Example: A classic example of a multiplicative random process is wealth accumulation. Rich people make their money by investing the money they already have, and we assume that the rate at which they make money in this fashion is in proportion to their current wealth, with some fluctuations which are related to the wisdom of their investment strategy and the current state of the economy. In other words, in each investment period, one's wealth is multiplied by some number which fluctuates randomly with some distribution. So it is a multiplicative random process. If all the rich people start off with roughly the same amount of money ("the first million"), then after some time, their wealths should be distributed according to a power law. In fact, this is just what one sees. This explanation was first proposed by Herb Simon in 1955.

Another example: It is proposed that web pages acquire links in proportion to the number they already have, so that the distribution of numbers of links should be a power law. Indeed, it is found to be a power law, but it doesn't have slope -1. Instead the slope is about -2. Why is this?

Fragmentation: Suppose we break a stick of unit length into two parts at a position which is a random fraction of the way along the stick's length. Then we break the resulting pieces again, and again, and so on. After many breaks, the length of one of the remaining pieces will be $\prod_i x_i$, where x_i is the position of the ith break. Again this is a product of random numbers and the resulting distribution of lengths will be a power law. This is thought to be the explanation for the power-law distribution of the sizes of meteors (and also meteor craters).

4.2.5 Random extremal processes

Some processes show *temporal* randomness, i.e., events which happen at times which are distributed according to a power law. One such example is the Omori law for earthquake aftershocks (time t between aftershocks is found to be distributed according to t^{-1}). One possible explanation is the **random extremal process**, or **record dynamics**. Suppose we generate a stream of uncorrelated random numbers x_i, and keep a record of the largest one we have seen so far. What is the average spacing between the record-breaking events?

Suppose it takes time t_1 to get a certain record breaking event. Then on average it will take as long again before we get another event of the same magnitude (or greater). So $t_2 = 2t_1$, and repeating the same argument,

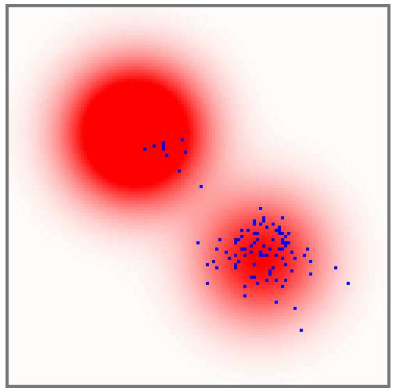
$$t_n = 2t_{n-1} = 2^n t_0. \tag{27}$$

Then the number δn of events in an interval δt is on average

$$\delta n = \delta t \Big/ \frac{\mathrm{d}t}{\mathrm{d}n} = \frac{\delta t}{t_0 2^n \log 2} = \frac{\delta t}{t \log 2}. \tag{28}$$

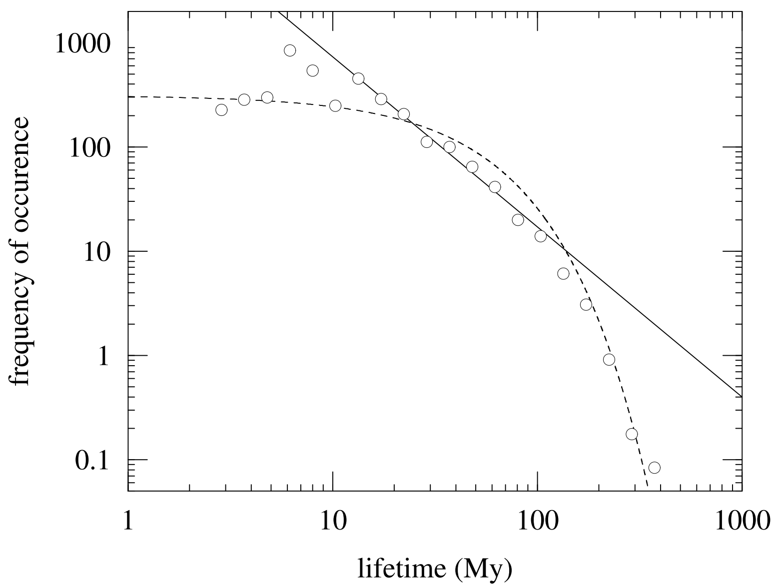
So the distribution of times between records is a power-law with slope -1, just like the Omori law.

Example: Suppose a population is evolving on a fitness landscape. Most of the time most of the population is localized around a fitness peak (it forms a species). Occasionally however, an offshoot population makes it to an adjacent peak, like this:



If the new peak is higher than the old one, the whole population will move there and we get a "punctuation." How often will this happen? Well, if the landscape is very high-dimensional, and mutation is random, then the heights of the peaks sampled will be independent random variables, and the dynamics will obey the rules described above, with punctuations happening with a power-law distribution of times separating them. Also the times between events, which are the lifetimes of species, will get longer.

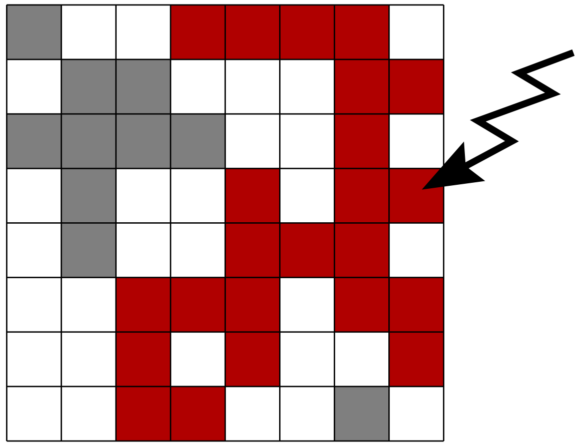
Interestingly, the lifetimes of species *do* get longer, and maybe they have a power-law distribution:



4.2.6 Self-organized criticality

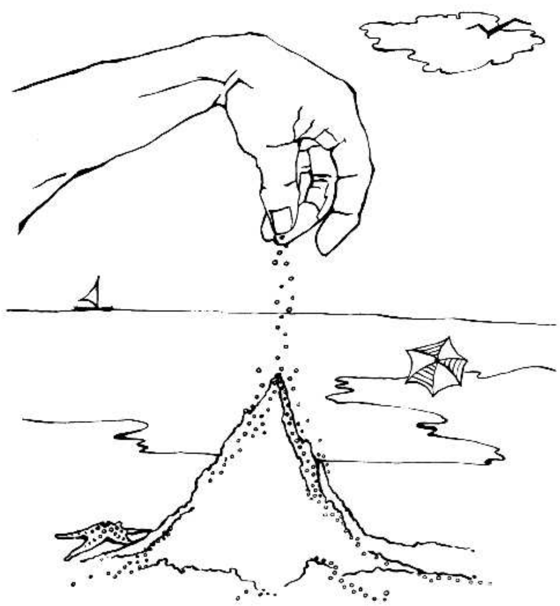
We said in Section 4.2.1 that criticality was not a good way of generating power laws, because it required you to be exactly at the critical point, which is unlikely. However, there is way around this. Some systems drive themselves to their own critical points and so produce power laws. This is called **self-organized criticality**. Here's a classic example: self-organized percolation, also called the **forest fire model**.

"Trees" appear on a square grid at a rate of one per unit time. At a much lower rate, "fires" start at random points on the grid. If there is a tree at the point where a fire starts, the fire destroys that tree and then spreads to any adjacent trees and destroys them, and so on until no trees are left for it to spread to. In other words, the fire destroys the percolation cluster of trees at the point where it strikes.



If there is a spanning cluster, then there is a finite chance of hitting it and burning it all, which means that within a finite time, it will be gone. So as soon as the system passes the percolation threshold, it gets knocked down below it again. Thus it always stays right at the threshold. The result is a system which always has a power-law distribution of fire sizes regardless of what state you start it off in.

The sandpile: Another famous example of a self-organized critical system is the sandpile:



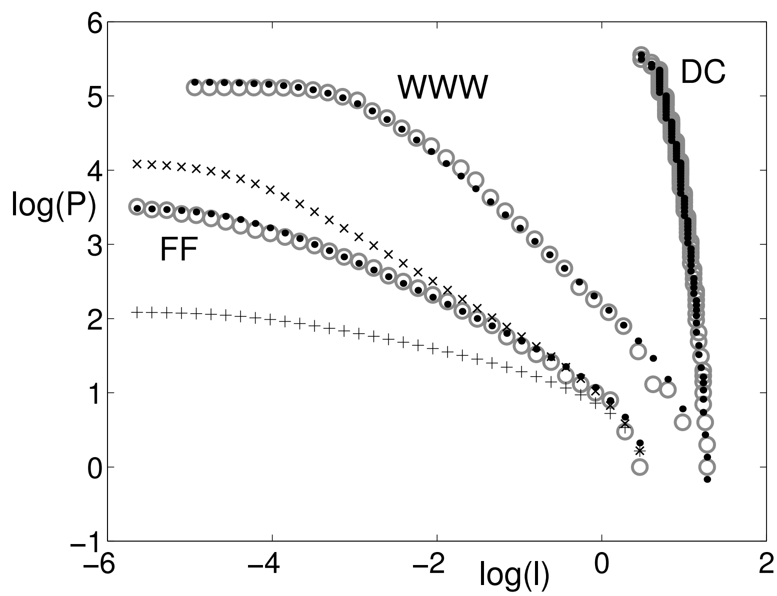
Sand is dropped on the top of the pile. The slope of the edges thus builds up. Avalanches start to happen, and they get bigger as the slope increases. At some point—the critical point—their size diverges, and we get mass transport of sand down the pile. This reduces the slope again and so we move back below the critical point. Overall therefore, we hover around criticality and get power-law distributions in the sizes of avalanches.

Both real avalanches and real forest fires are found to have power-law distributions. Perhaps this is the explanation? Maybe, but here's an alternative explanation.

4.2.7 Highly optimized tolerance

Suppose instead of allowing trees to grow at random in the forest, we place them specifically. How should we place them to minimize the average amount of damage done by the fires? Another way of looking at this is that we should place fire-breaks between forest stands and optimize the positions of these breaks. If fires are started by sparks which land uniformly at random everywhere in the forest, then the solution to this optimization problem is simple—cut the forest into equally sized chunks. However, if there are more sparks in some areas than others, it turns out that the average damage done by a fire is minimized by cutting the forest into chunks whose size varies in inverse proportion to the rate at which sparks land in that area. You can show that if you take this result and use it to work out what the distribution of the sizes of fires is, you get a distribution which follows a power law for a wide variety of choices of the distribution of sparks. Thus a power law is generated by the actions of an external agent (the forester) aiming the optimize the behavior of a system (the forest).

This theory has been applied to (amongst other things) real forest fires and the distribution of file sizes on the world-wide web:



Pretty impressive, huh? (Well, maybe, but the jury's still out on this one.)