

# Entropy, Power Laws, and Economics

Tom Carter

Complex Systems Summer School  
SFI, 2007

<http://astarte.csustan.edu/~tom/>

Santa Fe

June, 2007

# Contents

<b>Mathematics of Information</b>	<b>6</b>
<b>Some entropy theory</b>	<b>13</b>
<b>A Maximum Entropy Principle</b>	<b>17</b>
<b>Application: Economics I</b>	<b>20</b>
<b>Fit to Real World™</b>	<b>26</b>
<b>A bit about Power Laws</b>	<b>31</b>
<b>Application: Economics II</b>	<b>41</b>
<b>References</b>	<b>48</b>

# The quotes



- ⊙ Science, wisdom, and counting
- ⊙ Surprise, information, and miracles
- ⊙ Information (and hope)
- ⊙ H (or S) for Entropy

To topics ←

# Science, wisdom, and counting ←

“Science is organized knowledge. Wisdom is organized life.”

- Immanuel Kant

“My own suspicion is that the universe is not only stranger than we suppose, but stranger than we can suppose.”

- John Haldane

“Not everything that can be counted counts, and not everything that counts can be counted.”

- Albert Einstein (1879-1955)

“The laws of probability, so true in general, so fallacious in particular .”

- Edward Gibbon

# Surprise, information, and miracles



“The opposite of a correct statement is a false statement. The opposite of a profound truth may well be another profound truth.”

- Niels Bohr (1885-1962)

“I heard someone tried the monkeys-on-typewriters bit trying for the plays of W. Shakespeare, but all they got was the collected works of Francis Bacon.”

- Bill Hirst

“There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle.”

- Albert Einstein (1879-1955)

# Mathematics of Information



- We would like to develop a usable measure of the *information* we get from observing the occurrence of an event having probability  $p$ . Our first reduction will be to ignore any particular features of the event, and only observe whether or not it happened. Thus we will think of an event as the observance of a symbol whose probability of occurring is  $p$ . We will thus be defining the *information* in terms of the probability  $p$ .

The approach we will be taking here is axiomatic: on the next page is a list of the four fundamental axioms we will use. Note that we can apply this axiomatic system in any context in which we have available a set of non-negative real numbers. A specific special case of interest is *probabilities* (i.e., real numbers between 0 and 1), which motivated the selection of axioms ...

- We will want our *information* measure  $I(p)$  to have several properties:
  1. Information is a non-negative quantity:  
 $I(p) \geq 0$ .
  2. If an event has probability 1, we get no information from the occurrence of the event:  $I(1) = 0$ .
  3. If two independent events occur (whose joint probability is the product of their individual probabilities), then the information we get from observing the events is the sum of the two informations:  $I(p_1 * p_2) = I(p_1) + I(p_2)$ . (This is the critical property . . . )
  4. We will want our *information* measure to be a continuous (and, in fact, monotonic) function of the probability (slight changes in probability should result in slight changes in *information*).

- We can therefore derive the following:

1.  $I(p^2) = I(p * p) = I(p) + I(p) = 2 * I(p)$

2. Thus, further,  $I(p^n) = n * I(p)$   
(by induction . . .)

3.  $I(p) = I((p^{1/m})^m) = m * I(p^{1/m})$ , so  
 $I(p^{1/m}) = \frac{1}{m} * I(p)$  and thus in general

$$I(p^{n/m}) = \frac{n}{m} * I(p)$$

4. And thus, by continuity, we get, for  
 $0 < p \leq 1$ , and  $a > 0$  a real number:

$$I(p^a) = a * I(p)$$

- From this, we can derive the nice property:

$$I(p) = -\log_b(p) = \log_b(1/p)$$

for some base  $b$ .



- Summarizing: from the four properties,
  1.  $I(p) \geq 0$
  2.  $I(p_1 * p_2) = I(p_1) + I(p_2)$
  3.  $I(p)$  is monotonic and continuous in  $p$
  4.  $I(1) = 0$

we can derive that

$$I(p) = \log_b(1/p) = -\log_b(p),$$

for some positive constant  $b$ . The base  $b$  determines the units we are using.

We can change the units by changing the base, using the formulas, for  $b_1, b_2, x > 0$ ,

$$x = b_1^{\log_{b_1}(x)}$$

and therefore

$$\log_{b_2}(x) = \log_{b_2}(b_1^{\log_{b_1}(x)}) = (\log_{b_2}(b_1))(\log_{b_1}(x)).$$

- Thus, using different bases for the logarithm results in *information* measures which are just constant multiples of each other, corresponding with measurements in different units:
  1.  $\log_2$  units are *bits* (from 'binary')
  2.  $\log_3$  units are *trits*(from 'trinary')
  3.  $\log_e$  units are *nats* (from 'natural logarithm') (We'll use  $\ln(x)$  for  $\log_e(x)$ )
  4.  $\log_{10}$  units are *Hartleys*, after an early worker in the field.
- Unless we want to emphasize the units, we need not bother to specify the base for the logarithm, and will write  $\log(p)$ . Typically, we will think in terms of  $\log_2(p)$ .

- For example, flipping a fair coin once will give us events  $h$  and  $t$  each with probability  $1/2$ , and thus a single flip of a coin gives us  $-\log_2(1/2) = 1$  bit of information (whether it comes up  $h$  or  $t$ ).

Flipping a fair coin  $n$  times (or, equivalently, flipping  $n$  fair coins) gives us  $-\log_2((1/2)^n) = \log_2(2^n) = n * \log_2(2) = n$  bits of information.

We could enumerate a sequence of 25 flips as, for example:

*hthhththhhthttththhhthtt*

or, using 1 for  $h$  and 0 for  $t$ , the 25 bits

1011001011101000101110100.

We thus get the nice fact that  $n$  flips of a fair coin gives us  $n$  bits of information, and takes  $n$  binary digits to specify. That these two are the same reassures us that we have done a good job in our definition of our *information* measure . . .

# Information (and hope) ←

“In Cyberspace, the First Amendment is a local ordinance.”

- John Perry Barlow

“Groundless hope, like unconditional love, is the only kind worth having.”

- John Perry Barlow

“The most interesting facts are those which can be used several times, those which have a chance of recurring. . . . Which, then, are the facts that have a chance of recurring? In the first place, simple facts.”

H. Poincare, 1908

# Some entropy theory ←

- One question we might ask here is, what is the average amount of information we will get (per observation) from observing events from a probability distribution  $P$ ? In particular, what is the expected value of the information?
- Suppose we have a discrete probability distribution  $P = \{p_1, p_2, \dots, p_n\}$ , with  $p_i \geq 0$  and  $\sum_{i=1}^n p_i = 1$ , or a continuous distribution  $p(x)$  with  $p(x) \geq 0$  and  $\int p(x)dx = 1$ , we can define the *expected value* of an associated discrete set  $F = \{f_1, f_2, \dots, f_n\}$  or function  $F(x)$  by:

$$\langle F \rangle = \sum_{i=1}^n f_i p_i$$

or

$$\langle F(x) \rangle = \int F(x) p(x) dx.$$

With these ideas in mind, we can define the *entropy of a distribution* by:

$$H(P) = \langle I(p) \rangle .$$

In other words, we can define the *entropy* of a probability distribution as the *expected value* of the *information* of the distribution.

In particular, for a discrete distribution  $P = \{p_1, p_2, \dots, p_n\}$ , we have the entropy:

$$H(P) = \sum_{i=1}^n p_i \log \left( \frac{1}{p_i} \right) .$$

Several questions probably come to mind at this point:

- What properties does the function  $H(P)$  have? For example, does it have a maximum, and if so where?
- Is *entropy* a reasonable name for this? In particular, the name *entropy* is already in use in thermodynamics. How are these uses of the term related to each other?
- What can we do with this new tool?
- Let me start with an easy one. Why use the letter  $H$  for entropy? What follows is a slight variation of a footnote, p. 105, in the book *Spikes* by Rieke, et al. :-)

# H (or S) for Entropy ←

“The enthalpy is [often] written  $U$ .  $V$  is the volume, and  $Z$  is the partition function.  $P$  and  $Q$  are the position and momentum of a particle.  $R$  is the gas constant, and of course  $T$  is temperature.  $W$  is the number of ways of configuring our system (the number of states), and we have to keep  $X$  and  $Y$  in case we need more variables. Going back to the first half of the alphabet,  $A$ ,  $F$ , and  $G$  are all different kinds of free energies (the last named for Gibbs).  $B$  is a virial coefficient or a magnetic field.  $I$  will be used as a symbol for information;  $J$  and  $L$  are angular momenta.  $K$  is Kelvin, which is the proper unit of  $T$ .  $M$  is magnetization, and  $N$  is a number, possibly Avogadro’s, and  $O$  is too easily confused with 0. This leaves  $S$  . . .” and  $H$ . In *Spikes* they also eliminate  $H$  (e.g., as the Hamiltonian). I, on the other hand, along with Shannon and others, prefer to honor Hartley. Thus,  $H$  for entropy . . .



# A Maximum Entropy Principle ←

- Suppose we have a system for which we can measure certain macroscopic characteristics. Suppose further that the system is made up of many microscopic elements, and that the system is free to vary among various states. Then (a generic version of) the Second Law of Thermodynamics says that with probability essentially equal to 1, the system will be observed in states with maximum entropy.

We will then sometimes be able to gain understanding of the system by applying a *maximum information entropy* principle (MEP), and, using Lagrange multipliers, derive formulae for aspects of the system.

- Suppose we have a set of macroscopic measurable characteristics  $f_k$ ,  $k = 1, 2, \dots, M$  (which we can think of as constraints on the system), which we assume are related to microscopic characteristics via:

$$\sum_i p_i * f_i^{(k)} = f_k.$$

Of course, we also have the constraints:

$$p_i \geq 0, \text{ and}$$

$$\sum_i p_i = 1.$$

We want to maximize the entropy,  $\sum_i p_i \log(1/p_i)$ , subject to these constraints. Using Lagrange multipliers  $\lambda_k$  (one for each constraint), we have the general solution:

$$p_i = \exp \left( -\lambda - \sum_k \lambda_k f_i^{(k)} \right).$$

If we define  $Z$ , called the partition function, by

$$Z(\lambda_1, \dots, \lambda_M) = \sum_i \exp \left( - \sum_k \lambda_k f_i^{(k)} \right),$$

then we have  $e^\lambda = Z$ , or  $\lambda = \ln(Z)$ .

# Application: Economics I (a Boltzmann Economy) ←

- Our first example here is a very simple economy. Suppose there is a fixed amount of money ( $M$  dollars), and a fixed number of agents ( $N$ ) in the economy. Suppose that during each time step, each agent randomly selects another agent and transfers one dollar to the selected agent. An agent having no money doesn't go in debt. What will the long term (stable) distribution of money be?

This is not a very realistic economy – there is no growth, only a redistribution of money (by a random process). For the sake of argument, we can imagine that every agent starts with approximately the same amount of money, although in the long run, the starting distribution shouldn't matter.

- For this example, we are interested in looking at the distribution of money in the economy, so we are looking at the probabilities  $\{p_i\}$  that an agent has the amount of money  $i$ . We are hoping to develop a model for the collection  $\{p_i\}$ .

If we let  $n_i$  be the number of agents who have  $i$  dollars, we have two constraints:

$$\sum_i n_i * i = M$$

and

$$\sum_i n_i = N.$$

Phrased differently (using  $p_i = \frac{n_i}{N}$ ), this says

$$\sum_i p_i * i = \frac{M}{N}$$

and

$$\sum_i p_i = 1.$$

- We now apply Lagrange multipliers:

$$L = \sum_i p_i \ln(1/p_i) - \lambda \left[ \sum_i p_i * i - \frac{M}{N} \right] - \mu \left[ \sum_i p_i - 1 \right],$$

from which we get

$$\frac{\partial L}{\partial p_i} = -[1 + \ln(p_i)] - \lambda i - \mu = 0.$$

We can solve this for  $p_i$ :

$$\ln(p_i) = -\lambda i - (1 + \mu)$$

and so

$$p_i = e^{-\lambda_0} e^{-\lambda i}$$

(where we have set  $1 + \mu \equiv \lambda_0$ ).

- Putting in constraints, we have

$$\begin{aligned}
 1 &= \sum_i p_i \\
 &= \sum_i e^{-\lambda_0} e^{-\lambda i} \\
 &= e^{-\lambda_0} \sum_{i=0}^M e^{-\lambda i},
 \end{aligned}$$

and

$$\begin{aligned}
 \frac{M}{N} &= \sum_i p_i * i \\
 &= \sum_i e^{-\lambda_0} e^{-\lambda i} * i \\
 &= e^{-\lambda_0} \sum_{i=0}^M e^{-\lambda i} * i.
 \end{aligned}$$

We can approximate (for large  $M$ )

$$\sum_{i=0}^M e^{-\lambda i} \approx \int_0^M e^{-\lambda x} dx \approx \frac{1}{\lambda},$$

and

$$\sum_{i=0}^M e^{-\lambda i} * i \approx \int_0^M x e^{-\lambda x} dx \approx \frac{1}{\lambda^2}.$$

From these we have (approximately)

$$e^{\lambda_0} = \frac{1}{\lambda}$$

and

$$e^{\lambda_0} \frac{M}{N} = \frac{1}{\lambda^2}.$$

From this, we get

$$\lambda = \frac{N}{M} = e^{-\lambda_0},$$

and thus (letting  $T = \frac{M}{N}$ ) we have:

$$\begin{aligned} p_i &= e^{-\lambda_0} e^{-\lambda i} \\ &= \frac{1}{T} e^{-\frac{i}{T}}. \end{aligned}$$

This is a Boltzmann-Gibbs distribution, where we can think of  $T$  (the average amount of money per agent) as the “temperature,” and thus we have a “Boltzmann economy” . . .

Note: this distribution also solves the functional equation

$$p(m_1)p(m_2) = p(m_1 + m_2).$$



- This example, and related topics, are discussed in

*Statistical mechanics of money*

by Adrian Drăgulescu and Victor M. Yakovenko,

<http://arxiv.org/abs/cond-mat/0001432>

and

*Statistical mechanics of money: How saving propensity affects its distribution*

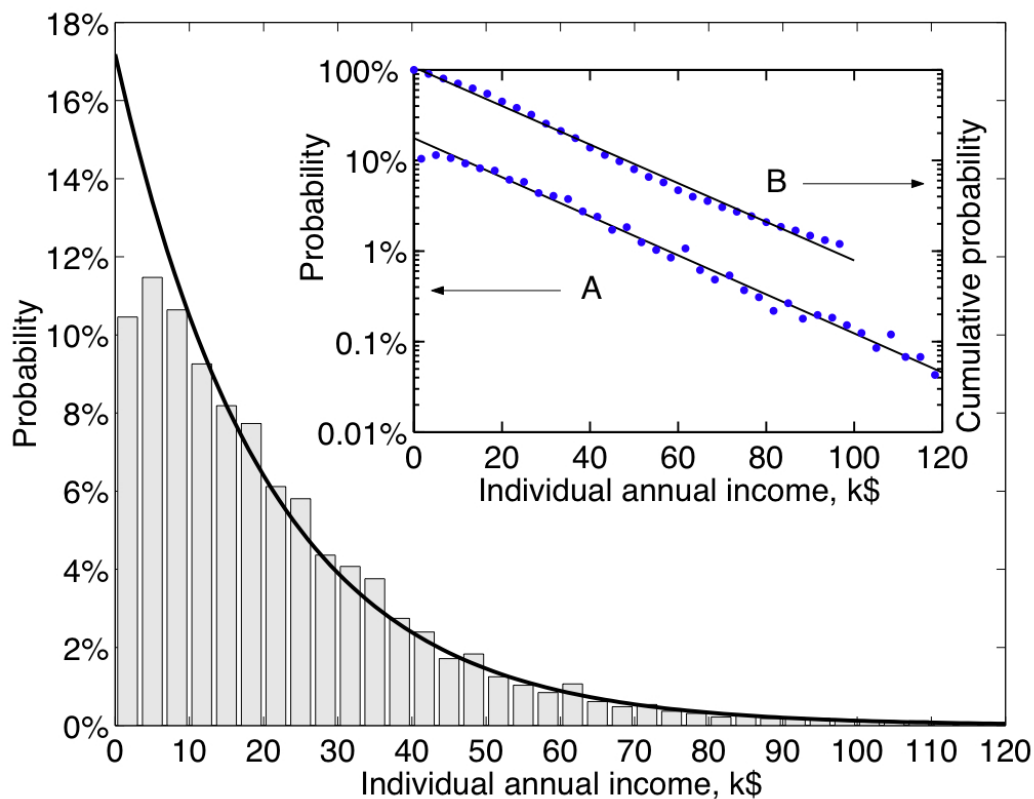
by Anirban Chakraborti and Bikas K. Chakrabarti

<http://arxiv.org/abs/cond-mat/0004256>

# Fit of this model to the Real World™ ←

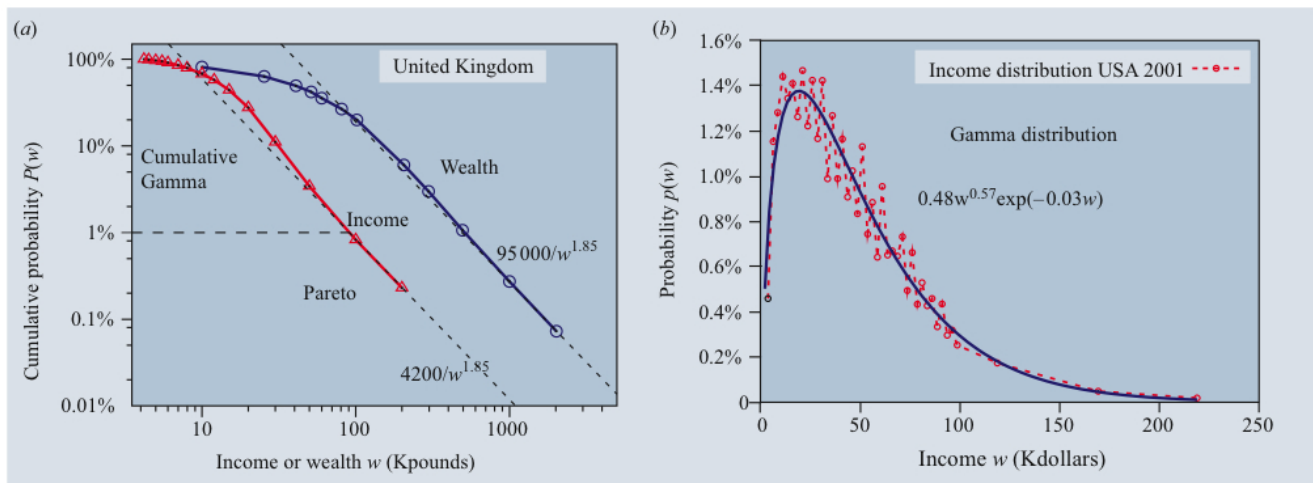
- How well does this model seem to fit to the Real World?

For a fairly large range of individuals, it actually does a decent job. Here is a graphical representation of U.S. census data for 1996 (see [33]):



The black line is  $p(x) = \frac{c}{R} e^{-\frac{x}{R}} \approx \frac{3.4}{20} e^{-\frac{x}{20}}$ .

- However, for the wealthy it doesn't do such a good job. Here are some graphical representations of U.K. and U.S. data for 1996-2001 (see [8]):



**Figure 1.** (a) Cumulative wealth (1996) and income (1998–9) distributions in the United Kingdom; for the upper class (almost 1% of the population) the distribution is well fitted with a Pareto's law. (b) Income distribution (2001) in the USA; for the low–middle classes (almost 99% of the population), the distribution is well fitted with a gamma distribution, equation (15).

As can be seen on the left graph, the wealth distribution for the U.K. wealthy in 1996 is close to a linear fit in log – log coordinates.

Can we modify the model somewhat to capture other characteristics of the data?

- There are a wide variety of important distributions that are observed in data sets. For example:

- Normal (gaussian) distribution:

$$p(x) \sim \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Natural explanation: Central limit theorem; sum of random variables (with finite second moment):

$$X_n = \sum_{i=1}^n x_i$$

Many applications:

- \* Maxwell: distribution of velocities of gas particles
- \* IQ
- \* heights of individuals

Distribution is *thin tailed* – no one is 20 feet tall . . .

– Exponential distribution:

$$p(x) \sim \exp(-x/x_0)$$

Natural explanation 1: Survival time for constant probability decay.

Natural explanation 2: Equilibrium statistical mechanics (see above – maximum entropy subject to constraint on mean).

Many applications:

\* Radioactive decay.

\* Equilibrium statistical mechanics (Boltzmann-Gibbs distribution)

Characteristic scale is  $x_0$ ; distribution is thin tailed.

– Power law (see below):

$$p(x) \sim x^{-\alpha}$$

# A bit about Power Laws ←

- Various researchers in various fields at various times have observed that many datasets seem to reflect a relationship of the form

$$p(x) \sim x^{-\alpha}$$

for a fairly broad range of values of  $x$ . These sorts of data relations are often called *power laws*, and have been the subject of fairly intensive interest and study.

An early researcher, Vilfredo Pareto, observed in the late 1800s that pretty uniformly across geographical locations, wealth was distributed through the population according to a power law, and hence such distributions are often called *Pareto distributions*.

A variety of other names have been applied to these distributions:

- Power law distribution
- Pareto's law
- Zipf's law
- Lotka's law
- Bradford's law
- Zeta distribution
- Scale free distribution
- Rank-size rule

My general rule of thumb is that if something has lots of names, it is likely to be important . . .

- These distributions have been observed in many places (as noted, for example, in Wikipedia):
  - Frequencies of words in longer texts
  - The size of human settlements (few cities, many hamlets/villages)
  - File size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones)
  - Clusters of Bose-Einstein condensate near absolute zero
  - The value of oil reserves in oil fields (a few large fields, many small fields)
  - The length distribution in jobs assigned supercomputers (a few large ones, many small ones)
  - The standardized price returns on individual stocks



- Size of sand particles
- Number of species per genus (please note the subjectivity involved: The tendency to divide a genus into two or more increases with the number of species in it)
- Areas burnt in forest fires
- There are a variety of important properties of power laws:
  - Distribution has fat / heavy tails (extreme events are more likely than one might expect . . . ). Stock market volatility; sizes of storms / floods, etc.
  - A power law is a linear relation between logarithms:

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

– Power laws are scale invariant:

Sufficient:

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

$$x \rightarrow cx$$

$$p(x) \rightarrow Kc^{-\alpha}x^{-\alpha} = c^{-\alpha}p(x)$$

Necessary: Scale invariant is defined as

$$p(cx) = K(c)p(x)$$

Power law is the only solution (0 and 1 are trivial solutions).

- Power laws are actually asymptotic relations. We can't define a power law on  $[0, \infty]$ :

If  $\alpha > 1$ , not integrable at 0.

If  $\alpha \leq 1$ , not integrable at  $\infty$ .

Thus, when we say something is a power law, we mean either within a range, or as  $x \rightarrow 0$  or as  $x \rightarrow \infty$ .

- Moments: power laws have a threshold above which moments don't exist. For  $p(x) \sim x^{-(\alpha+1)}$ , when  $\alpha > m$ ,

$$\begin{aligned}\gamma(m) &= \int_a^\infty x^m p(x) dx \\ &= \int_a^\infty x^m x^{-(\alpha+1)} dx \\ &= \infty\end{aligned}$$

- The lack of moments is conserved under aggregation . . . If  $\alpha(x)$  is the *tail exponent* of the random variable  $x$  (the value above which moments don't exist), then

$$\begin{aligned}\alpha(x + y) &= \min(\alpha(x), \alpha(y)) \\ \alpha(xy) &= \min(\alpha(x), \alpha(y)) \\ \alpha(x^k) &= \alpha(x)/k.\end{aligned}$$

- Power laws are generic for heavy / fat tailed distributions. In other words, any “reasonable” distribution with fat tails (i.e., with moments that don’t exist) is a power law:

$$\begin{aligned}P(X > x) &= 1 - \Phi_\alpha(x) \\ &= 1 - \exp(-x^{-\alpha}) \\ &\approx 1 - (1 - x^{-\alpha}) \\ &= x^{-\alpha}\end{aligned}$$

(there is some discussion of *extreme value distributions* that goes here, with discussion of Fréchet, Weibull, and Gumbel distributions – specifically Fréchet distributions (with fat tails) . . . perhaps another place or time).

- Some mechanism for generating power laws:
  - Critical points and deterministic dynamics
  - Non-equilibrium statistical mechanics
  - Random processes
  - Mixtures
  - Maximization principles
  - Preferential attachment
  - Dimensional constraints

- Multiplicative (random) processes generate log-normal distributions, which can look like power law distributions across various ranges of the variable. If  $a(t)$  is a random variable:

$$x(t + 1) = a(t)x(t)$$

$$x(t) = \prod_{i=0}^{t-1} a(i)x(0)$$

$$\log(x(t)) = \sum_{i=0}^{t-1} \log(a(i)) + \log(x(0))$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2 / 2\sigma^2}$$

$$\log(f(x)) = -\frac{(\log(x))^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1\right) \log(x) + \text{const}$$

In particular, if  $\sigma$  is large in comparison with  $\log(x)$ , then it will look like

$$\log(f(x)) \approx \log(x^{-1}),$$

which is a one-over-x power law distribution . . .

- Other distributions that have power-law appearing regions:
  - Mixed multiplicative / additive processes (*Kesten processes*):

$$x(t + 1) = a(t)x(t) + b(t)$$

- Stable multiplicative random walk with reflecting barrier.

Both of these will look log-normal in their bodies, and like power laws in their tails.

(Various pieces of this section draw from lectures / notes by Doyne Farmer on power laws in financial markets – my thanks to him . . . )

# Application: Economics II (a power law) ←

- Suppose that a (simple) economy is made up of many agents  $a$ , each with wealth at time  $t$  in the amount of  $w(a, t)$ . (I'll leave it to you to come up with a reasonable definition of “wealth” – of course we will want to make sure that the definition of “wealth” is applied consistently across all the agents.) We can also look at the total wealth in the economy  $W(t) = \sum_a w(a, t)$ .

For this example, we are interested in looking at the distribution of wealth in the economy, so we will assume there is some collection  $\{w_i\}$  of possible values for the wealth an agent can have, and associated probabilities  $\{p_i\}$  that an agent has wealth  $w_i$ . We are hoping to develop a model for the collection  $\{p_i\}$ .



- In order to apply the maximum entropy principle, we want to look at global (aggregate/macro) observables of the system that reflect (or are made up of) characteristics of (micro) elements of the system.

For this example, we can look at the growth rate of the economy. A reasonable way to think about this is to let

$R_i = w_i(t_1)/w_i(t_0)$  and  $R = W(t_1)/W(t_0)$  (where  $t_0$  and  $t_1$  represent time steps of the economy). The growth rate will then be  $\ln(R)$ . We then have the two constraints on the  $p_i$ :

$$\sum_i p_i * \ln(R_i) = \ln(R)$$

and

$$\sum_i p_i = 1.$$

- We now apply Lagrange multipliers:

$$L = \sum_i p_i \ln(1/p_i) - \lambda \left[ \sum_i p_i \ln(R_i) - \ln(R) \right] - \mu \left[ \sum_i p_i - 1 \right],$$

from which we get

$$\frac{\partial L}{\partial p_i} = -[1 + \ln(p_i)] - \lambda \ln(R_i) - \mu = 0.$$

We can solve this for  $p_i$ :

$$p_i = e^{-\lambda_0} e^{-\lambda \ln(R_i)} = e^{-\lambda_0} R_i^{-\lambda}$$

(where we have set  $1 + \mu \equiv \lambda_0$ ).

Solving, we get  $\lambda_0 = \ln(Z(\lambda))$ , where  $Z(\lambda) \equiv \sum_i R_i^{-\lambda}$  (the partition function) normalizes the probability distribution to sum to 1. From this we see the power law (for  $\lambda > 1$ ):

$$p_i = \frac{R_i^{-\lambda}}{Z(\lambda)}.$$

- We might actually like to calculate specific values of  $\lambda$ , so we will do the process again in a continuous version. In this version, we will let  $R = w(T)/w(0)$  be the relative wealth at time  $T$ . We want to find the probability density function  $f(R)$ , that is:

$$\max_{\{f\}} H(f) = - \int_1^{\infty} f(R) \ln(f(R)) dR,$$

subject to

$$\int_1^{\infty} f(R) dR = 1,$$

$$\int_1^{\infty} f(R) \ln(R) dR = C \ln(R),$$

where  $C$  is the average number of transactions per time step.

We need to apply the calculus of variations to maximize over a class of functions.

When we are solving an extremal problem of the form

$$\int F[x, f(x), f'(x)]dx,$$

we work to solve

$$\frac{\partial F}{\partial f(x)} - \frac{d}{dx} \left( \frac{\partial F}{\partial f'(x)} \right) = 0.$$

Our Lagrangian is of the form

$$L \equiv - \int_1^{\infty} f(R) \ln(f(R)) dr - \mu \left( \int_1^{\infty} f(R) dR - 1 \right) - \lambda \left( \int_1^{\infty} f(R) \ln(R) dR - C * \ln(R) \right).$$

Since this does not depend on  $f'(x)$ , we look at:

$$\frac{\partial[-f(R) \ln f(R) - \mu(f(R) - 1) - \lambda(f(R) \ln R - R)]}{\partial f(R)} = 0$$

from which we get

$$f(R) = e^{-(\lambda_0 - \lambda \ln(R))} = R^{-\lambda} e^{-\lambda_0},$$

where again  $\lambda_0 \equiv 1 + \mu$ .

We can use the first constraint to solve for  $e^{\lambda_0}$ :

$$e^{\lambda_0} = \int_1^{\infty} R^{-\lambda} dR = \left[ \frac{R^{-\lambda+1}}{1-\lambda} \right]_1^{\infty} = \frac{1}{\lambda-1},$$

assuming  $\lambda > 1$ . We therefore have a power law distribution for wealth of the form:

$$f(R) = (\lambda - 1)R^{-\lambda}.$$

To solve for  $\lambda$ , we use:

$$C * \ln(R) = (\lambda - 1) \int_1^{\infty} R^{-\lambda} \ln(R) dR.$$

Using integration by parts, we get

$$\begin{aligned} C * \ln(R) &= (\lambda - 1) \left[ \ln(R) \frac{R^{1-\lambda}}{1-\lambda} \right]_1^{\infty} \\ &\quad - (\lambda - 1) \int_1^{\infty} \frac{R^{-\lambda}}{1-\lambda} dR \\ &= (\lambda - 1) \left[ \ln(R) \frac{R^{1-\lambda}}{1-\lambda} \right]_1^{\infty} + \left[ \frac{R^{1-\lambda}}{1-\lambda} \right]_1^{\infty}. \end{aligned}$$

By L'Hôpital's rule, the first term goes to zero as  $R \rightarrow \infty$ , so we are left with

$$C * \ln(R) = \left[ \frac{R^{1-\lambda}}{1-\lambda} \right]_1^\infty = \frac{1}{\lambda-1},$$

or, in other terms,

$$\lambda - 1 = C * \ln(R^{-1}).$$

For much more discussion of this example, see the paper *A Statistical Equilibrium Model of Wealth Distribution* by Mishael Milakovic, February, 2001, available on the web at:

<http://astarte.csustan.edu/~tom/SFI-CSSS/Wealth/wealth-Milakovic.pdf>



# References

- [1] Bar-Yam, Yaneer, *Dynamics of Complex Systems (Studies in Nonlinearity)* , Westview Press, Boulder, 1997.
- [2] Brillouin, L., *Science and information theory* Academic Press, New York, 1956.
- [3] Brooks, Daniel R., and Wiley, E. O., *Evolution as Entropy, Toward a Unified Theory of Biology*, Second Edition, University of Chicago Press, Chicago, 1988.
- [4] Campbell, Jeremy, *Grammatical Man, Information, Entropy, Language, and Life*, Simon and Schuster, New York, 1982.
- [5] Cover, T. M., and Thomas J. A., *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
- [6] DeLillo, Don, *White Noise*, Viking/Penguin, New York, 1984.
- [7] Drăgulescu, A. and Yakovenko, V.M., *Statistical mechanics of money*, Eur. Phys. J. B 17, 723-729 (2000). <http://arxiv.org/abs/cond-mat/0001432>

- [8] Drăgulescu, A. and Yakovenko, V.M., *Evidence for the exponential distribution of income in the USA*, Eur. Phys. J. B 20, 585-589 (2001).  
<http://arxiv.org/pdf/cond-mat/0008305.pdf>
- [9] Feller, W., *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1957.
- [10] Feynman, Richard, *Feynman lectures on computation*, Addison-Wesley, Reading, 1996.
- [11] Gatlin, L. L., *Information Theory and the Living System*, Columbia University Press, New York, 1972.
- [12] Greven, A., Keller, G., Warnecke, G., *Entropy*, Princeton Univ. Press, Princeton, 2003.
- [13] Haken, Hermann, *Information and Self-Organization, a Macroscopic Approach to Complex Systems*, Springer-Verlag, Berlin/New York, 1988.
- [14] Hamming, R. W., Error detecting and error correcting codes, *Bell Syst. Tech. J.* **29** 147, 1950.
- [15] Hamming, R. W., *Coding and information theory*, 2nd ed, Prentice-Hall, Englewood Cliffs, 1986.



- [16] Hill, R., *A first course in coding theory* Clarendon Press, Oxford, 1986.
- [17] Hodges, A., *Alan Turing: the enigma* Vintage, London, 1983.
- [18] Hofstadter, Douglas R., *Metamagical Themas: Questing for the Essence of Mind and Pattern*, Basic Books, New York, 1985
- [19] Jones, D. S., *Elementary information theory* Clarendon Press, Oxford, 1979.
- [20] Knuth, Eldon L., *Introduction to Statistical Thermodynamics*, McGraw-Hill, New York, 1966.
- [21] Landauer, R., Information is physical, *Phys. Today*, May 1991 23-29.
- [22] Landauer, R., The physical nature of information, *Phys. Lett. A*, **217** 188, 1996.
- [23] van Lint, J. H., *Coding Theory*, Springer-Verlag, New York/Berlin, 1982.
- [24] Lipton, R. J., Using DNA to solve NP-complete problems, *Science*, **268** 542–545, Apr. 28, 1995.
- [25] MacWilliams, F. J., and Sloane, N. J. A., *The theory of error correcting codes*, Elsevier Science, Amsterdam, 1977.

- [26] Martin, N. F. G., and England, J. W., *Mathematical Theory of Entropy*, Addison-Wesley, Reading, 1981.
- [27] Maxwell, J. C., *Theory of heat* Longmans, Green and Co, London, 1871.
- [28] von Neumann, John, Probabilistic logic and the synthesis of reliable organisms from unreliable components, in *automata studies*(*Shanon, McCarthy eds*), 1956 .
- [29] Papadimitriou, C. H., *Computational Complexity*, Addison-Wesley, Reading, 1994.
- [30] Pierce, John R., *An Introduction to Information Theory – Symbols, Signals and Noise*, (second revised edition), Dover Publications, New York, 1980.
- [31] Roman, Steven, *Introduction to Coding and Information Theory*, Springer-Verlag, Berlin/New York, 1997.
- [32] Sampson, Jeffrey R., *Adaptive Information Processing, an Introductory Survey*, Springer-Verlag, Berlin/New York, 1976.

- [33] Scafetta, N., Picozzi, S. and West, B. J., *An out-of-equilibrium model of the distributions of wealth*, QUANTITATIVE FINANCE VOLUME 4 (2004)353-364.  
DOI: 10.1088/1469-7688/4/3/010
- [34] Schroeder, Manfred, *Fractals, Chaos, Power Laws, Minutes from an Infinite Paradise*, W. H. Freeman, New York, 1991.
- [35] Shannon, C. E., A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379; also p. 623, 1948.
- [36] Slepian, D., ed., *Key papers in the development of information theory* IEEE Press, New York, 1974.
- [37] Turing, A. M., On computable numbers, with an application to the Entscheidungsproblem, *Proc. Lond. Math. Soc. Ser. 2* **42**, 230 ; see also *Proc. Lond. Math. Soc. Ser. 2* **43**, 544, 1936.
- [38] Zurek, W. H., Thermodynamic cost of computation, algorithmic complexity and the information metric, *Nature* **341** 119-124, 1989.

[To top ←](#)