Lotka’s Law (and Other Power Laws)

T.C. Scott

presented at Near India Pvt Ltd. No. 71/72, Jyoti Nivas College Road, Koramangala, Bangalore 560095, India

May 27, 2017
Outline of Presentation: Explanation through Examples

1. Definitions
2. Power Law: Underlying mechanism behind Pareto Principle
3. Two examples of publications.
4. Artwork
5. Other Power laws. E.g. Zipf’s law.
   5.1 word frequency
   5.2 company size
   5.3 Investment
   5.4 Gutenberg-Richter law: seismology
   5.5 Moore’s Law: exponential growth.
   5.6 death statistics
   5.7 Population vs. e.g. city ranking
   5.8 general cases
6. Artifacts/Comments
7. Conclusions
Lotka’s Law

- Author publication frequency in *any* given field.
- Number of authors writing $X$ papers is $\propto 1/X^n$.
- $n \approx 2$: Inverse Square Law but not always. Determined by fitting.
- **Gist:** As number of articles published increases, authors producing that many publications become less frequent.
- ”*Success breeds success*” ☺.
- ”*Disaster compounds disaster*” ☻.
Applications of Power Law

Publications (of course) but also:

**Art Productivity.** E.g.: number of $X$ statues of Venus vs $Y$ of artists who depicted Venus.

**Geological:** Sizes of earthquakes, craters on the moon and of solar flares.

**Nature:** Foraging pattern of various species.

**Military:** fighter pilot kills 🎃.

**Languages:** Frequencies of words in docs; of family name.

**”Human affairs”:** Sizes of power outages and wars.

**Business rankings:** Company size based on e.g. employee size, VC capital investment.

e tc... etc...
Lotka’s law formula:
\[ C = X^nY \quad \Leftrightarrow \quad Y = \frac{C}{X^n} \]

where
- \( X \): number of publications.
- \( Y \): Number of authors writing \( X \) publications.
- \( n \approx 2 \) but not always.
- applies to \textit{any} field (if model holds).
- applies well for relatively short time windows.
Get a Feel for it

Assume $n = 2$.

For example, we have:

1. Say 60% will have just one publication. ($C = 0.6$)
2. 15% will have 2 publications ($\frac{1}{2^2} \times 0.60$).
3. 7% of authors will have 3 publications ($\frac{1}{3^2} \times 0.60$).
4. etc...
Figure: On the right is the long tail; on the left is called the Head (also known as the 80-20 rule or Pareto Principle).
Power Law Curve Features: More “commercial” version

Figure: From Chris Anderson, editor-in-chief of Wired Magazine.
The Law of the Vital few:

or principle of factor sparsity

- 20% of people dominate 80% of wealth.
- 20% of inventory occupies 80% of warehouse.
- 20% of sales staff make 80% of revenue.
- 20% of products make 80% of sales.
- 20% of criminals cause 80% of crimes.
- 20% of hazards cause 80% of injuries.
- 20% of patients use 80% of health care resources.

Not exact. Idea works but percentages vary.
E.g.: 30% vs. 70%.
Manipulation: put data on a Log-Log scale

Take logs on both sides of $C = X^NY$:

$$\log(C') = n \log(X) + \log(Y)$$

$$\Rightarrow \log(Y) = -n \log(X) + \log(C')$$

- Get linear $F = mx + b$ relation on a log scale.
- Verify with Pearson correlation.
- Least-Squares curve fit.
- (Non-linear) curve fit.

Tony C. Scott  Lotka’s Law
Useful Tool: Pearson Correlation

Think of angle $\theta$ between 2 vectors $\mathbf{u}$ and $\mathbf{v}$:

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| \ ||\mathbf{v}||}$$

Now subtract out mean (or average) from each vector:

$$\mathbf{u} \leftarrow \mathbf{u} - \bar{\mathbf{u}} \quad \mathbf{v} \leftarrow \mathbf{v} - \bar{\mathbf{v}}$$

Obtain:

$$\rho(\mathbf{u}, \mathbf{v}) = \frac{\sum_{i=1}^{N} (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^{N} (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^{N} (v_i - \bar{v})^2}}$$

$\Rightarrow$ Calculate $\rho(\log(X), \log(Y))$
Log-Log Plot

\[ \log(F_1/F_2) \]

\[ \log(x_2/x_1) \]

\[ m \]

\[ 1 \]

1. Dickson H. Leavens in 1953 based his study of econometricians on the work of 721 authors who presented papers at meetings of the Econometric Society or had articles published in the first twenty volumes of Econometrica (1933-52) - Table 3 of Potter’s paper.

2. University of Illinois, Library of Urbana-Champaign, Study of Personal Authors in the Card Catalog. The Illinois catalog contains records for about 2.5 million titles. A random sample of 2345 personal authors was drawn - Table 5 of Potter’s paper.
Test Case 1

**Table:** Data from Table 3 of W.G. Potter Reference

<table>
<thead>
<tr>
<th>No. Contributions</th>
<th>No. Contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>436</td>
</tr>
<tr>
<td>2</td>
<td>107</td>
</tr>
<tr>
<td>3</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
</tr>
</tbody>
</table>
Test Case 1 - Log-Log Scale

Correlation of Log Data = -0.94    C = 275.99    n = -1.7
Test Case 1 - Linear Scale

Linear Data vs $Y = C \times X^n = 275.99 \times X^{-1.7}$
Test Case 2

Table: Data from Table 5 of W.G. Potter Reference

<table>
<thead>
<tr>
<th>No. Works</th>
<th>No. Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1489</td>
</tr>
<tr>
<td>2</td>
<td>343</td>
</tr>
<tr>
<td>3</td>
<td>160</td>
</tr>
<tr>
<td>4</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>1</td>
</tr>
</tbody>
</table>
Test Case 2 - Log-Log Scale

Correlation of Log Data = -0.95  C = 731.61  n = -1.8

Tony C. Scott  Lotka’s Law
Test Case 2 - Linear Scale

Linear Data vs $Y = C X^n = 731.61 X^{-1.8}$
Lotka’s Law of Productivity in Artwork

Sources:


- Adjacent paper: “Quantitative Comparison between the Italian and French Iconography of Venus from the Middle Ages to Modern Times” by K. Bender.

- Fig. 11 of Adjacent paper generated from the two Topical Catalogs “The French Venus” and “The Venus of the Low Countries”.
Figure: Fig. 11 of Adjacent paper generated from the two Topical Catalogs “The French Venus” and “The Venus of the Low Countries”.
Lotka’s Law Example - “Italian Job”

Productivity (Italian)

Log (proportional number of artists) vs Log (number of works)

$y = -1.6826x - 0.4036$

$R^2 = 0.9286$
Related metrics-Generalizations

**Zipf’s Law:** frequency of any word is inversely proportional to its rank in the frequency table (Lotka’s law is a special application).

**Bradford’s Law:** estimates exponentially diminishing returns of extending a search for references in science journals.

**Benford’s Law:** also called first-digit law of data:
1. 30% chance lead digit is 1,
2. 18% chance lead digit is 2,
3. larger digits occur with diminishing frequency according to an inverse power law.

**Gutenberg-Richter law:** for Seismology

**Moore’s Law:** Not a power law: exponential law.
Zipf’s Law


- \( N \) be the number of elements and \( k \) be their rank,
- \( s \) be the value of the exponent characterizing the distribution.

Zipf’s law predicts that out of a population of \( N \) elements, the frequency of elements of rank \( k \), \( f(k; s, N) \), is:

\[
f(k; s, N) = \frac{1/k^s}{\sum_{n=1}^{N}(1/n^s)}
\]  

(1)

Zipf’s law holds if the number of elements with a given frequency is a random variable with power law distribution \( p(f) = \alpha f^{-1-1/s} \).

Often written in terms of Riemann \( \zeta \) function or distribution.

**Note:** Lotka’s law is a special case of Zipf’s law. \( (n = -1 - 1/s) \)
Zip’s law Example: Word frequency of Moby Dick

Figure: From Whim Online Magazine
Another Zipf’s Law example

From “CONVERSABLE ECONOMIST” by Timothy Taylor

An example looking at the distribution of the size of US firms:

- Size measured by the number of employees on the horizontal axis,
- The number of firms of this size, measured on the vertical axis.
- Both axes are measured in powers of 10.
- Again, the end result is very closely obtained by Zipf’s Law.
Zipf’s Law Example: Firm Size

Figure 3
Log Frequency versus log Size of US firms (by Number of Employees) for 1997

Notes: Ordinary least squares (OLS) fit gives a slope of 2.06 (s.e. = 0.054; $R^2 = 0.99$). This corresponds to a frequency $f(S) \sim S^{-2.059}$, which is a power law distribution with exponent 1.059. This is very close to an ideal Zipf’s law, which would have an exponent $\zeta = 1$. 
Question: What does the distribution of returns in venture fund look like?

Naive Answer: rank companies from best to worst according to their return in multiple of dollars invested:

- People tend to group investments into three buckets.
- The bad companies go to zero.
- The mediocre ones do maybe $1x$ (stock with multiple of 1 in "liquidation preference"), so you don’t lose much or gain much.
- And then the great companies do maybe $3 - 10x$.

This model misses a key insight that actual returns are incredibly skewed.
Look at Founders Fund’s 2005 fund,
The best investment ended up being worth about as much as all the rest combined.
The investment in the second best company was about as valuable as number three through the rest.
This same dynamic generally held true throughout the fund: This is a power law distribution.

To a first approximation, a VC portfolio will only make money if your best company investment ends up being worth more than your whole fund.

In practice, it’s quite hard to be profitable as a VC if you don’t get to those numbers.
Power Law Example: Peter Thiel’s CS183

![Graph showing investment return vs company rank]

- **investment return**
- **company rank**

- Actual
- Perceived
The Gutenberg-Richter law (GR law) expresses the relationship between the magnitude and total number of earthquakes in any given region and time period of at least that magnitude.

\[ \log_{10} N = a - b \, m \]  \hspace{1cm} (2)

or

\[ N = 10^{a-bm} \]  \hspace{1cm} (3)

where

- \( N \) is the number of events having a magnitude \( \geq m \).
- \( a \) and \( b \) are constants.
GR law: Example - Data over 6 decades

From Kim Christensen, Proceedings of the National Academy of Sciences of the USA, vol. 99 no. suppl 1, 2509-2513.

Figure: Gutenberg-Richter exponent $b = 1$ (dashed red line). Roll-off for $m < 2$ is due to difficulties with detecting very small earthquakes.
Moore’s Law

- Observation that the number of transistors in a dense integrated circuit *doubles* approximately every two years.
- Moore’s prediction proved accurate for several decades, and has been used in the semiconductor industry to guide long-term planning and to set targets for research and development.
- Main reason is the invention of the integrated circuit.
- Not a Power Law but an *exponential law*: See patterns on a Semi-Log scale.
Moore's Law: Example

Microprocessor Transistor Counts 1971-2011 & Moore's Law

Figure: Plot of CPU transistor counts vs. dates of introduction.
**Figure:** Updated version of Moore’s Law over 120 Years (by Ray Kurzweil). The 7 most recent data points are all NVIDIA GPUs.
Overall Growth of Data

- Complementary relationship to Moore’s law: exponential growth of (Big) Data.
- IDC says that in 2005 there were 130 exabytes of data created, about 2,837 exabytes in 2012, and a forecast of 40,000 exabytes by 2020...
- ICD says Digital Universe will be 35 Zettabytes by 2020. 1 Zettabyte = 1000 exabytes = 1 billion terabytes.
- Also an exponential law.
Figure: 1 Exabyte = $10^{18}$ bytes = 1000 petabytes = 1 million terabytes = 1 billion gigabytes.
How much Data can be analysed?

1. The rate of data growth is *outstripping* Moore’s Law;
2. The amount of data we want to store would be too expensive with current technology;
3. The speed at which read-heads in large disk drives can retrieve data is not keeping up with data growth meaning that retrieval times are simply too long, and
4. Newly available semi-structured and unstructured data can yield new insights with the new technology.

IDC estimates that only $\frac{1}{2}$% of data today is being analyzed; that 3% of data today is tagged and could be analyzed, and that of the 2,837 exabytes created in 2012, 23% would be useful if tagged and analyzed. Clearly a great deal of potential value is being lost.
Figure: “Data is growing faster than Moore’s Law”. Source: Capgemini Consulting Technology Outsourcing, June 23, 2015
Examples from “The Challenge Network”

Death Statistics: Imagine terrorist incidents and their consequences. If you plot the number of terrorist incidents of a certain lethality against the frequency with which such incidents occur in time, you get a scatter that forms a nearly straight line when plotted on a log scale.

Other Examples: How probable is it that a meteor of a given size will hit the moon over such-and-so range of time? How likely is a financial crash at some arbitrary period in time?

Again Zipf distributions apply.
Power Law Example: Death Statistics

Relative probability of an incident occurring at this scale

Number of casualties per incident

Deaths
Injuries
All casualties
Rank-Size distribution

Rank-size distribution is the distribution of size by rank, in decreasing order of size.

▶ E.g. if a data set consists of items of sizes 5, 100, 5, and 8, the rank-size distribution is 100, 8, 5, 5 (ranks 1 through 4).

▶ Also known as the rank-frequency distribution, when the source data are from a frequency distribution.

▶ These distributions frequently follow a power law distribution, or less well-known ones such as a stretched exponential function \( f_\beta(t) = \exp(-t^\beta) \) or parabolic fractal distribution, at least approximately for certain ranges of ranks.
Figure: Rank-size distribution of the population of countries follows a stretched exponential distribution except for China and India.
Zipf’s Law and World city populations

world city populations for 8 countries
log-size vs log-rank
General Power Law Examples

Lotka’s Law
Get the idea?

We get power laws or nearly so for:

- Laws of Physics ($k/r^2$): (Kepler, Newton, Electrosatic Coulomb Law)
- Publications
- Bug frequencies
- Frequencies governing city and company sizes
- markers for trends, distributions, predictions
- markers for success, victories, accomplishments
- markers for droughts, diseases, disasters!
Why power law relationships show up so often?

▶ Say you start of with a random distribution of something,
▶ The different points in your data all tend to experience proportional growth (positive or negative).
▶ However, this particular data (like city size) can’t turn negative.
▶ In addition, the total size of the system can’t grow in an unbounded way.
▶ with a few additional assumptions resulting from the actual dynamics, this process will result in the kind of straight-line power laws shown here.

Of course, this kind of explanation then needs to be adapted and applied to each specific situation.
Note those points for large $X$:

- Lotka formula underestimates number of works of more prolific authors but applies fairly well for the less prolific.
- Known as the **long tail** of frequency distributions.
- Martize-Mekler (2009), Egghe (2010) alternative:

$$Y = c \frac{(X_{max} + 1 - X)^b}{X^a}$$

where $a$, $b$ and $c$ are parameters and...

- $X_{max}$ is maximum number of works produced by $Y$ authors and this for $X = 1$, $C = cX_{max}^b$. 
Estimating exponent $n$

Non-linear Fitting: instead of linear fit of log data.

Two point fitting method: Use cumulative frequency i.e. 'bundling' of frequencies w.r.t. reference value. Smoother functions.

Maximum likelihood: gives the observed data the greatest probability.

Kolmogorov-Smirnov (K-S): Recommended: does not assume independent and identically distributed data.
Improvements: Kolmogorov-Smirnov

Minimize Kolmogorov-Smirnov (K-S) statistic

\[ \hat{n} = \arg \min_n D_n \]

i.e. find \( n \) such that \( D_n \) is smallest and:

\[ D_n = \max_X |P_{\text{data}}(X) - P_{\text{Lotka}}(X)| \]

- **cmds**: cumulative distribution of exponent \( n \).
- \( P_{\text{data}} \) cmds of Data.
- \( P_{\text{Lotka}} \) cmds of Lotka’s model.
Power laws like Lotka’s law appear to be *ubiquitous* to nature and even subjective man-made data, more than many people realize.

The potential for applications of these power laws is underestimated and far from exhausted.

When in doubt: put the frequency statistics on a log-log scale.

Accuracy can be estimated in advance using *Pearson correlation* on log data.

Threshold at e.g. $> 75\%$.

If Threshold is met: produce the Lotka (Power) plot.

Deal with refinements of metric *later*．

Keep *Pareto principle* in mind for *all* data.