# WIKIPEDIA

Zipf's law

Zipf's law (/zɪf/, not /tsɪpf/ as in German) is an empirical law formulated using mathematical statistics that refers to the fact that for many types of data studied in the physical and social sciences, the rank-frequency distribution is an inverse relation. The **Zipfian distribution** is one of a family of related discrete power law probability distributions. It is related to the zeta distribution, but is not identical.

Zipf's law was originally formulated in terms of quantitative linguistics, stating that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table. Thus the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc. For example, in the Brown Corpus of American English text, the word "*the*" is the most frequently occurring word, and by itself accounts for nearly 7% of all word occurrences (69,971 out of slightly over 1 million). True to Zipf's Law, the second-place word "*of*" accounts for slightly over 3.5% of words (36,411 occurrences), followed by "*and*" (28,852). Only 135 vocabulary items are needed to account for half the Brown Corpus.[1]

The law is named after the American

Zipf's law
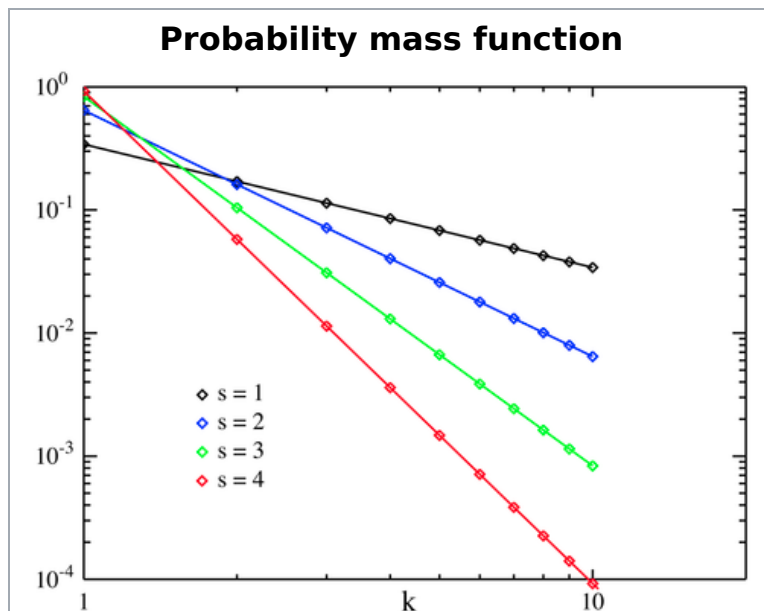
Zipf PMF for *N* = 10 on a log–log scale. The horizontal axis is the index *k* . (Note that the function is only defined at integer values of *k*. The connecting lines do not indicate continuity.)
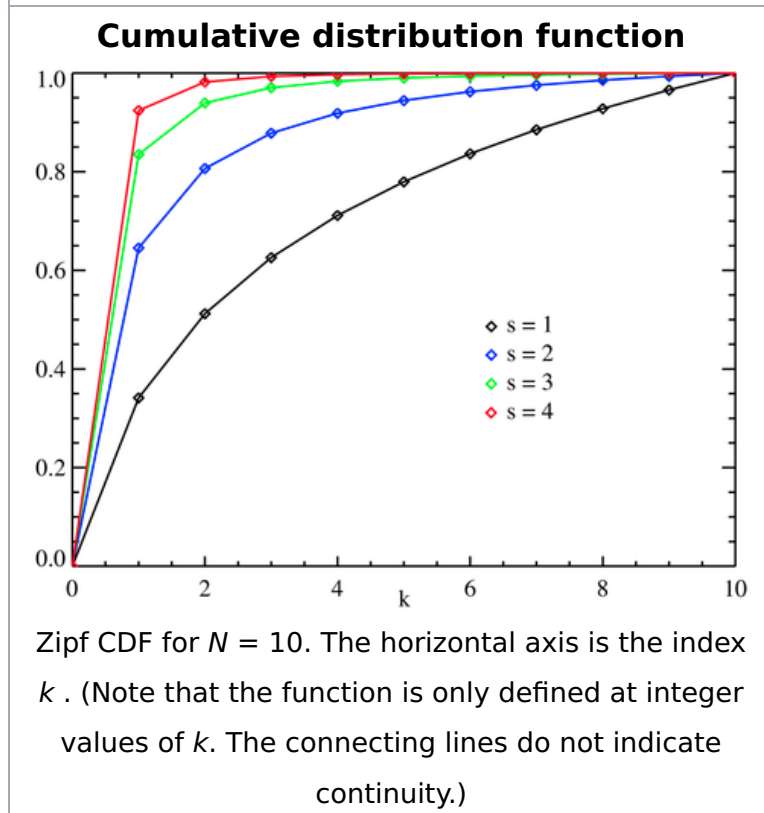
Zipf CDF for *N* = 10. The horizontal axis is the index *k* . (Note that the function is only defined at integer values of *k*. The connecting lines do not indicate continuity.)

linguist George Kingsley Zipf (1902–1950), who popularized it and sought to explain it (Zipf 1935, 1949), though he did not claim to have originated it.[2] The French stenographer Jean-Baptiste Estoup (1868–1950) appears to have noticed the regularity before Zipf.[3] It was also noted in 1913 by German physicist Felix Auerbach (1856–1933).[4]

The law is similar in concept, though not identical in distribution, to Benford's law.

Parameters	$s \geq 0$ (real) $N \in \{1, 2, 3 \ldots\}$ (integer)
Support	$k \in \{1, 2, \ldots, N\}$
PMF	$\dfrac{1/k^s}{H_{N,s}}$ where $H_{N,s}$ is the Nth generalized harmonic number
CDF	$\dfrac{H_{k,s}}{H_{N,s}}$
Mean	$\dfrac{H_{N,s-1}}{H_{N,s}}$
Mode	1
Variance	$\dfrac{H_{N,s-2}}{H_{N,s}} - \dfrac{H_{N,s-1}^2}{H_{N,s}^2}$
Entropy	$\dfrac{s}{H_{N,s}} \sum_{k=1}^{N} \dfrac{\ln(k)}{k^s} + \ln(H_{N,s})$
MGF	$\dfrac{1}{H_{N,s}} \sum_{n=1}^{N} \dfrac{e^{nt}}{n^s}$
CF	$\dfrac{1}{H_{N,s}} \sum_{n=1}^{N} \dfrac{e^{int}}{n^s}$

Contents

Other data sets

The same relationship occurs in many other rankings of human-created systems,[5] such as the ranks of mathematical expressions or ranks of notes in music[6] and even in uncontrolled environments, such as corporation sizes, income rankings, ranks of number of people watching the same TV channel,[7] cells' transcriptomes[8] and so on. The appearance of the distribution in rankings of cities by population was first noticed by Felix Auerbach in 1913,[4] leading to a broad literature of Zipf's law for cities.[9] However, more recent empirical[10][11] and theoretical[12] studies have challenged the relevance of Zipf's law for cities.

Empirically, a data set can be tested to see whether Zipf's law applies by checking the goodness of fit of an empirical distribution to the hypothesized power law distribution with a Kolmogorov–Smirnov test, and then comparing the (log) likelihood ratio of the power law distribution to alternative distributions like an exponential distribution or lognormal distribution.[13]

Theoretical review

Zipf's law is most easily observed by plotting the data on a log-log graph, with the axes being log (rank order) and log (frequency). For example, the word *"the"* (as described above) would appear at $x = \log(1)$, $y = \log(69971)$. It is also possible to plot reciprocal rank against frequency or reciprocal frequency or interword interval against rank.[2] The data conform to Zipf's law to the extent that the plot is linear.

Formally, let:

- *N* be the number of elements;
- *k* be their rank;
- *s* be the value of the exponent characterizing the distribution.

Zipf's law then predicts that out of a population of *N* elements, the normalized frequency of the element of rank *k*, $f(k;s,N)$, is:

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

Zipf's law holds if the number of elements with a given frequency is a random variable with power law distribution $p(f) = \alpha f^{-1-1/s}$.[14]

It has been claimed that this representation of Zipf's law is more suitable for statistical testing, and in this way it has been analyzed in more than 30,000 English texts. The goodness-of-fit tests yield that only about 15% of the texts are statistically compatible with this form of Zipf's law. Slight variations in the definition of Zipf's law can increase this percentage up to close to 50%.[15]

In the example of the frequency of words in the English language, *N* is the number of words in the English language and, if we use the classic version of Zipf's law, the exponent *s* is 1. $f(k; s,N)$ will then be the fraction of the time the *k*th most common word occurs.

The law may also be written:

$$f(k; s, N) = \frac{1}{k^s H_{N,s}}$$

where $H_{N,s}$ is the Nth generalized harmonic number.

The simplest case of Zipf's law is a "$\frac{1}{f}$" function. Given a set of Zipfian distributed frequencies, sorted from most common to least common, the second most common frequency will occur half as often as the first, the third most common frequency will occur $\frac{1}{3}$ as often as the first, and the nth most common frequency will occur $\frac{1}{n}$ as often as the first. However, this cannot hold exactly, because items must occur an integer number of times; there cannot be 2.5 occurrences of a word. Nevertheless, over fairly wide ranges, and to a fairly good approximation, many natural phenomena obey Zipf's law.

In human languages, word frequencies have a very heavy-tailed distribution, and can therefore be modeled reasonably well by a Zipf distribution with an s close to 1.
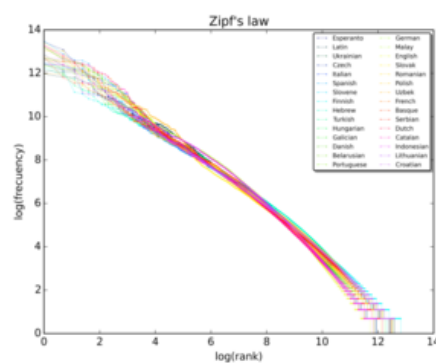
As long as the exponent s exceeds 1, it is possible for such a law to hold with infinitely many words, since if $s > 1$ then

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} < \infty.$$

where ζ is Riemann's zeta function.

Statistical explanation

Although Zipf's Law holds for all languages, even non-natural ones like Esperanto,[16] the reason is still not well understood.[17] However, it may be partially explained by the statistical analysis of randomly generated texts. Wentian Li has shown that in a document in which each character has been chosen randomly from a uniform distribution of all letters (plus a space character), the "words" with different lengths follow the macro-trend of the Zipf's law (the more probable words are the shortest with equal probability).[18] Vitold Belevitch, in a paper entitled *On the Statistical Laws of Linguistic Distribution*, offers a mathematical derivation. He took a large class of well-behaved statistical distributions (not only the normal

A plot of the rank versus frequency for the first 10 million words in 30 Wikipedias (dumps from October 2015) in a log-log scale.

distribution) and expressed them in terms of rank. He then expanded each expression into a Taylor series. In every case Belevitch obtained the remarkable result that a first-order truncation of the series resulted in Zipf's law. Further, a second-order truncation of the Taylor series resulted in Mandelbrot's law.[19][20]

The principle of least effort is another possible explanation: Zipf himself proposed that neither speakers nor hearers using a given language want to work any harder than

necessary to reach understanding, and the process that results in approximately equal distribution of effort leads to the observed Zipf distribution.[21][22]

Similarly, preferential attachment (intuitively, "the rich get richer" or "success breeds success") that results in the Yule–Simon distribution has been shown to fit word frequency versus rank in language[23] and population versus city rank[24] better than Zipf's law. It was originally derived to explain population versus rank in species by Yule, and applied to cities by Simon.

Mathematical explanation

Atlas models are systems of exchangeable positive-valued *diffusion processes* with drift and variance parameters that depend only on the rank of the process. It has been shown mathematically that Zipf's law holds for Atlas models that satisfy certain natural regularity conditions.[25] Atlas models can be used to represent empirical systems of time-dependent multivariate data, including, e.g., the frequency of words in a written language or the size of companies. An Atlas model that represents an empirical system will have the same stationary distribution as the empirical system, so if the Atlas model follows Zipf's law, the system will also follow Zipf's law. Since Atlas models that satisfy natural regularity conditions follow Zipf's law, this accounts for its universality.[26]

In the figure above of the 10 million Wikipedia words, the log-log plots are not precisely straight lines but rather slightly concave curves with a tangent of slope -1 at some point along the curve. Such distributions are usually referred to as *quasi-Zipfian* distributions, and most systems of time-dependent empirical data that are said to follow Zipf's law are actually quasi-Zipfian. Quasi-Zipfian systems can be represented by *quasi-Atlas models*, and quasi-Atlas models are amenable to mathematical treatment similar to that for Zipf's law.

Related laws

Zipf's law in fact refers more generally to frequency distributions of "rank data", in which the relative frequency of the *n*th-ranked item is given by the zeta distribution, $1/(n^s \zeta(s))$, where the parameter $s > 1$ indexes the members of this family of probability distributions. Indeed, *Zipf's law* is sometimes synonymous with "zeta distribution", since probability distributions are sometimes called "laws". This distribution is sometimes called the **Zipfian** distribution.

A generalization of Zipf's law is the Zipf–Mandelbrot law, proposed by Benoit Mandelbrot, whose frequencies are:

$$f(k; N, q, s) = \frac{[\text{constant}]}{(k + q)^s}.$$

The "constant" is the reciprocal of the Hurwitz zeta function evaluated at *s*. In practice, as

easily observable in distribution plots for large corpora, the observed distribution can be modelled more accurately as a sum of separate distributions for different subsets or subtypes of words that follow different parameterizations of the Zipf–Mandelbrot distribution, in particular the closed class of functional words exhibit s lower than 1, while open-ended vocabulary growth with document size and corpus size require s greater than 1 for convergence of the Generalized Harmonic Series.[2]

Zipfian distributions can be obtained from Pareto distributions by an exchange of variables.[14]

The Zipf distribution is sometimes called the **discrete Pareto distribution**[27] because it is analogous to the continuous Pareto distribution in the same way that the discrete uniform distribution is analogous to the continuous uniform distribution.

The tail frequencies of the Yule–Simon distribution are approximately
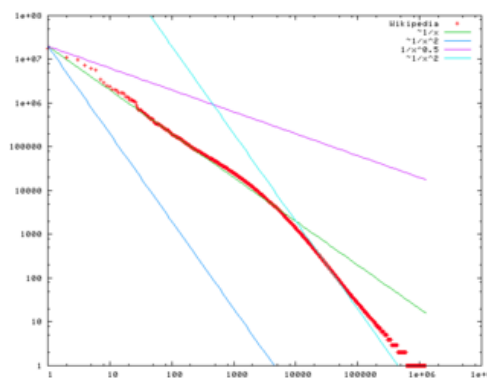
$$f(k; \rho) \approx \frac{[\text{constant}]}{k^{\rho+1}}$$

for any choice of $\rho > 0$.

In the parabolic fractal distribution, the logarithm of the frequency is a quadratic polynomial of the logarithm of the rank. This can markedly improve the fit over a simple power-law relationship.[28] Like fractal dimension, it is possible to calculate Zipf dimension, which is a useful parameter in the analysis of texts.[29]

A plot of word frequency in Wikipedia (November 27, 2006). The plot is in log-log coordinates. x is rank of a word in the frequency table; y is the total number of the word's occurrences. Most popular words are "the", "of" and "and", as expected. Zipf's law corresponds to the middle linear portion of the curve, roughly following the green ($1/x$) line, while the early part is closer to the magenta ($1/x^{0.5}$) line while the later part is closer to the cyan ($1/(k + x)^{2.0}$) line. These lines correspond to three distinct parameterizations of the Zipf–Mandelbrot distribution, overall a broken power law with three segments: a head, middle, and tail.

It has been argued that Benford's law is a special bounded case of Zipf's law,[28] with the connection between these two laws being explained by their both originating from scale invariant functional relations from statistical physics and critical phenomena.[30] The ratios of probabilities in Benford's law are not constant. The leading digits of data satisfying Zipf's law with s = 1 satisfy Benford's law.

n	Benford's law: $P(n) =$ $\log_{10}(n+1) - \log_{10}(n)$	$\frac{\log(P(n)/P(n-1))}{\log(n/(n-1))}$
1	0.30103000	
2	0.17609126	−0.7735840
3	0.12493874	−0.8463832
4	0.09691001	−0.8830605
5	0.07918125	−0.9054412
6	0.06694679	−0.9205788
7	0.05799195	−0.9315169
8	0.05115252	−0.9397966
9	0.04575749	−0.9462848

Applications

In information theory, a symbol (event, signal) of probability p contains $-\log_2(1/p)$ bits of information. Hence, Zipf's law for natural numbers: $\Pr(x) \approx 1/x$ is equivalent with number x containing $\log_2(x)$ bits of information. To add information from a symbol of probability p into information already stored in a natural number x, we should go to x' such that $\log_2(x') \approx \log_2(x) + \log_2(1/p)$, or equivalently $x' \approx x/p$. For instance, in standard binary system we would have $x' = 2x + s$, what is optimal for $\Pr(s=0) = \Pr(s=1) = 1/2$ probability distribution. Using $x' \approx x/p$ rule for a general probability distribution is the base of Asymmetric Numeral Systems family of entropy coding methods used in data compression, whose state distribution is also governed by Zipf's law.

Zipf's law has been used for extraction of parallel fragments of texts out of comparable corpora.[31] Zipf's law has also been used by Laurance Doyle and others at the SETI Institute as part of the search for extraterrestrial intelligence.[32][33]

The Voynich Manuscript, which is a 15th-century codex, also falls in line with Zipf's law, indicating that text is most likely not a hoax but rather written in an obscure language or cipher.[34][35]

See also

- 1% rule (Internet culture)
- Benford's law
- Bradford's law
- Brevity law
- Demographic gravitation

- Frequency list
- Gibrat's law
- Hapax legomenon
- Heaps' law
- King effect
- Lorenz curve
- Lotka's law
- Menzerath's law
- Pareto distribution
- Pareto principle, a.k.a. the "80–20 rule"
- Price's law
- Principle of least effort
- Rank-size distribution
- Stigler's law of eponymy
- Long tail

References

1. Fagan, Stephen; Gençay, Ramazan (2010), "An introduction to textual econometrics", in Ullah, Aman; Giles, David E. A. (eds.), *Handbook of Empirical Economics and Finance*, CRC Press, pp. 133–153, ISBN 9781420070361. P. 139 (https://books.google.com/books?hl=en &lr=&id=QAUv9R6bJzwC&oi=fnd&pg=PA139): "For example, in the Brown Corpus, consisting of over one million words, half of the word volume consists of repeated uses of only 135 words."

2. Powers, David M W (1998). *Applications and explanations of Zipf's law* (http://aclweb.org/anthology/W98-1218). Joint conference on new methods in language processing and computational natural language learning. Association for Computational Linguistics. pp. 151–160.

3. Christopher D. Manning, Hinrich Schütze *Foundations of Statistical Natural Language Processing*, MIT Press (1999), ISBN 978-0-262-13360-9, p. 24

4. Auerbach F. (1913) Das Gesetz der Bevölkerungskonzentration. Petermann's Geographische Mitteilungen 59, 74–76

5. Piantadosi, Steven (March 25, 2014). "Zipf's word frequency law in natural language: A critical review and future directions" (https://www. ncbi.nlm.nih.gov/pmc/articles/PMC4176592). *Psychon Bull Rev*. **21** (5): 1112–1130. doi:10.3758/s13423-014-0585-6 (https://doi.org/10.375 8%2Fs13423-014-0585-6). PMC 4176592 (https://www.ncbi.nlm.nih.go v/pmc/articles/PMC4176592). PMID 24664880 (https://pubmed.ncbi.nl m.nih.gov/24664880).

6. Zanette, Damián H. (June 7, 2004). "Zipf's law and the creation of musical context". arXiv:cs/0406015 (https://arxiv.org/abs/cs/0406015).

7. M. Eriksson, S.M. Hasibur Rahman, F. Fraille, M. Sjöström, Efficient Interactive Multicast over DVB-T2 - Utilizing Dynamic SFNs and PARPS (http://apachepersonal.miun.se/~mageri/myresearch/bmsb2013-Eriks son.pdf) Archived (https://web.archive.org/web/20140502183246/htt p://apachepersonal.miun.se/~mageri/myresearch/bmsb2013-Eriksson. pdf) 2014-05-02 at the Wayback Machine, 2013 IEEE International Conference on Computer and Information Technology (BMSB'13), London, UK, June 2013. Suggests a heterogeneous Zipf-law TV channel-selection model

8. Lazzardi, Silvia; Valle, Filippo; Mazzolini, Andrea; Scialdone, Antonio; Caselle, Michele; Osella, Matteo (2021-06-17). "Emergent Statistical Laws in Single-Cell Transcriptomic Data" (https://www.biorxiv.org/cont ent/10.1101/2021.06.16.448706v1). *bioRxiv*: 2021–06.16.448706. doi:10.1101/2021.06.16.448706 (https://doi.org/10.1101%2F2021.06. 16.448706). S2CID 235482777 (https://api.semanticscholar.org/Corpu sID:235482777). Retrieved 2021-06-18.

9. Gabaix, Xavier (1999). "Zipf's Law for Cities: An Explanation" (https:// www.jstor.org/stable/2586883). *The Quarterly Journal of Economics*. **114** (3): 739–767. doi:10.1162/003355399556133 (https://doi.org/10. 1162%2F003355399556133). ISSN 0033-5533 (https://www.worldcat. org/issn/0033-5533). JSTOR 2586883 (https://www.jstor.org/stable/258 6883).

10. Arshad, Sidra; Hu, Shougeng; Ashraf, Badar Nadeem (2018-02-15). "Zipf's law and city size distribution: A survey of the literature and future research agenda" (https://www.sciencedirect.com/science/articl e/pii/S0378437117310130). *Physica A: Statistical Mechanics and Its Applications*. **492**: 75–92. Bibcode:2018PhyA..492...75A (https://ui.ads abs.harvard.edu/abs/2018PhyA..492...75A). doi:10.1016/j.physa.2017.10.005 (https://doi.org/10.1016%2Fj.physa. 2017.10.005). ISSN 0378-4371 (https://www.worldcat.org/issn/0378-4 371).

11. Gan, Li; Li, Dong; Song, Shunfeng (2006-08-01). "Is the Zipf law spurious in explaining city-size distributions?" (https://www.sciencedir ect.com/science/article/pii/S0165176506000772). *Economics Letters*. **92** (2): 256–262. doi:10.1016/j.econlet.2006.03.004 (https://doi.org/1 0.1016%2Fj.econlet.2006.03.004). ISSN 0165-1765 (https://www.worl dcat.org/issn/0165-1765).

12. Verbavatz, Vincent; Barthelemy, Marc (November 2020). "The growth equation of cities" (https://www.nature.com/articles/s41586-020-2900- x). *Nature*. **587** (7834): 397–401. arXiv:2011.09403 (https://arxiv.org/ abs/2011.09403). Bibcode:2020Natur.587..397V (https://ui.adsabs.har vard.edu/abs/2020Natur.587..397V). doi:10.1038/s41586-020-2900-x (https://doi.org/10.1038%2Fs41586-020-2900-x). ISSN 1476-4687 (htt ps://www.worldcat.org/issn/1476-4687). PMID 33208958 (https://pubm ed.ncbi.nlm.nih.gov/33208958). S2CID 227012701 (https://api.semant icscholar.org/CorpusID:227012701).

13. Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. SIAM Review, 51(4), 661–703. doi:10.1137/070710111 (https://doi.org/10.1137%2F070710111)

14. Adamic, Lada A. (2000) "Zipf, Power-laws, and Pareto - a ranking tutorial", originally published at (http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html) .parc.xerox.com (http://www.parc.xerox.com/istl/groups/iea/papers/ranking/ranking.html) Archived (https://web.archive.org/web/20071026062626/http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html) 2007-10-26 at the Wayback Machine

15. Moreno-Sánchez, I; Font-Clos, F; Corral, A (2016). "Large-Scale Analysis of Zipf's Law in English Texts" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4723055). *PLOS ONE*. **11** (1): e0147073. arXiv:1509.04486 (https://arxiv.org/abs/1509.04486). Bibcode:2016PLoSO..1147073M (https://ui.adsabs.harvard.edu/abs/2016PLoSO..1147073M). doi:10.1371/journal.pone.0147073 (https://doi.org/10.1371%2Fjournal.pone.0147073). PMC 4723055 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4723055). PMID 26800025 (https://pubmed.ncbi.nlm.nih.gov/26800025).

16. Bill Manaris; Luca Pellicoro; George Pothering; Harland Hodges (13 February 2006). *Investigating Esperanto's statistical proportions relative to other languages using neural networks and Zipf's law* (https://web.archive.org/web/20160305040450/http://www.cs.cofc.edu/~manaris/uploads/Main/IASTED2006.pdf) (PDF). *Artificial Intelligence and Applications*. Innsbruck, Austria. pp. 102–108. Archived from the original (http://www.cs.cofc.edu/~manaris/uploads/Main/IASTED2006.pdf) (PDF) on 5 March 2016.

17. Léon Brillouin, *La science et la théorie de l'information*, 1959, réédité en 1988, traduction anglaise rééditée en 2004

18. Wentian Li (1992). "Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution". *IEEE Transactions on Information Theory*. **38** (6): 1842–1845. CiteSeerX 10.1.1.164.8422 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.164.8422). doi:10.1109/18.165464 (https://doi.org/10.1109%2F18.165464).

19. Neumann, Peter G. "Statistical metalinguistics and Zipf/Pareto/Mandelbrot" (http://www.csl.sri.com/users/neumann/#12a), *SRI International Computer Science Laboratory*, accessed and archived (https://web.archive.org/web/20110605012951/http://www.csl.sri.com/users/neumann/) 29 May 2011.

20. Belevitch V (18 December 1959). "On the statistical laws of linguistic distributions" (http://www.csl.sri.com/users/neumann/belevitch.pdf) (PDF). *Annales de la Société Scientifique de Bruxelles*. I. **73**: 310–326.

21. Zipf GK (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, Massachusetts: Addison-Wesley. p. 1.

22. Ramon Ferrer i Cancho & Ricard V. Sole (2003). "Least effort and the origins of scaling in human language" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC298679). *Proceedings of the National Academy of Sciences of the United States of America*. **100** (3): 788–791. Bibcode:2003PNAS..100..788C (https://ui.adsabs.harvard.edu/abs/2003PNAS..100..788C). doi:10.1073/pnas.0335980100 (https://doi.org/10.1073%2Fpnas.0335980100). PMC 298679 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC298679). PMID 12540826 (https://pubmed.ncbi.nlm.nih.gov/12540826).

23. Lin, Ruokuang; Ma, Qianli D. Y.; Bian, Chunhua (2014). "Scaling laws in human speech, decreasing emergence of new words and a generalized model". arXiv:1412.4846 (https://arxiv.org/abs/1412.4846) [cs.CL (https://arxiv.org/archive/cs.CL)].

24. Vitanov, Nikolay K.; Ausloos, Marcel; Bian, Chunhua (2015). "Test of two hypotheses explaining the size of populations in a system of cities". *Journal of Applied Statistics*. **42** (12): 2686–2693. arXiv:1506.08535 (https://arxiv.org/abs/1506.08535). Bibcode:2015arXiv150608535V (https://ui.adsabs.harvard.edu/abs/2015arXiv150608535V). doi:10.1080/02664763.2015.1047744 (https://doi.org/10.1080%2F02664763.2015.1047744). S2CID 10599428 (https://api.semanticscholar.org/CorpusID:10599428).

25. Ricardo T. Fernholz; Robert Fernholz (December 2020). "Zipf's law for atlas models" (https://www.cambridge.org/core/journals/journal-of-applied-probability/article/abs/zipfs-law-for-atlas-models/5D6B730DDEE4C05CF494213FDA57B064). *Journal of Applied Probability*. **57** (4): 1276–1297. doi:10.1017/jpr.2020.64 (https://doi.org/10.1017%2Fjpr.2020.64). S2CID 146808080 (https://api.semanticscholar.org/CorpusID:146808080).

26. Terence Tao (2012). "E Pluribus Unum: From Complexity, Universality" (https://direct.mit.edu/daed/article/141/3/23/27037/E-pluribus-unum-From-Complexity-Universality). *Daedalus*. **141** (3): 23–34. doi:10.1162/DAED_a_00158 (https://doi.org/10.1162%2FDAED_a_00158). S2CID 14535989 (https://api.semanticscholar.org/CorpusID:14535989).

27. N. L. Johnson; S. Kotz & A. W. Kemp (1992). *Univariate Discrete Distributions* (second ed.). New York: John Wiley & Sons, Inc. ISBN 978-0-471-54897-3., p. 466.

28. Johan Gerard van der Galien (2003-11-08). "Factorial randomness: the Laws of Benford and Zipf with respect to the first digit distribution of the factor sequence from the natural numbers" (https://web.archive.org/web/20070305150334/http://home.zonnet.nl/galien8/factor/factor.html). Archived from the original (http://home.zonnet.nl/galien8/factor/factor.html) on 2007-03-05. Retrieved 8 July 2016.

29. Eftekhari, Ali (2006). "Fractal geometry of texts: An initial application to the works of Shakespeare". *Journal of Quantitative Linguistic*. **13** (2–3): 177–193. doi:10.1080/09296170600850106 (https://doi.org/10.1080%2F09296170600850106). S2CID 17657731 (https://api.semanticscholar.org/CorpusID:17657731).

30. Pietronero, L.; Tosatti, E.; Tosatti, V.; Vespignani, A. (2001). "Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf". *Physica A*. **293** (1–2): 297–304. Bibcode:2001PhyA..293..297P (https://ui.adsabs.harvard.edu/abs/2001PhyA..293..297P). doi:10.1016/S0378-4371(00)00633-6 (https://doi.org/10.1016%2FS0378-4371%2800%2900633-6).

31. Mohammadi, Mehdi (2016). "Parallel Document Identification using Zipf's Law" (https://comparable.limsi.fr/bucc2016/pdf/BUCC04.pdf) (PDF). *Proceedings of the Ninth Workshop on Building and Using Comparable Corpora*. LREC 2016. Portorož, Slovenia. pp. 21–25. Archived (https://web.archive.org/web/20180323154706/https://comparable.limsi.fr/bucc2016/pdf/BUCC04.pdf) (PDF) from the original on 2018-03-23.

32. Doyle, Laurance R.; Mao, Tianhua (2016-11-18). "Why Alien Language Would Stand Out Among All the Noise of the Universe" (http://cosmos.nautil.us/feature/54/listening-for-extraterrestrial-blah-blah). *Nautilus Quarterly*.

33. Kershenbaum, Arik (2021-03-16). *The Zoologist's Guide to the Galaxy: What Animals on Earth Reveal About Aliens--and Ourselves*. Penguin. pp. 251–256. ISBN 978-1-9848-8197-7. OCLC 1242873084 (https://www.worldcat.org/oclc/1242873084).

34. #author.fullName}. "Mystery text's language-like patterns may be an elaborate hoax" (https://www.newscientist.com/article/2106915-mystery-texts-language-like-patterns-may-be-an-elaborate-hoax/). *New Scientist*. Retrieved 2022-02-25. {{cite web}}: |last= has generic name (help)

35. Montemurro, Marcelo A.; Zanette, Damián H. (2013-06-21). "Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3689824/). *PLoS ONE*. **8** (6): e66344. doi:10.1371/journal.pone.0066344 (https://doi.org/10.1371%2Fjournal.pone.0066344). ISSN 1932-6203 (https://www.worldcat.org/issn/1932-6203). PMC 3689824 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3689824). PMID 23805215 (https://pubmed.ncbi.nlm.nih.gov/23805215).

Further reading

Primary:

- George K. Zipf (1949) *Human Behavior and the Principle of Least Effort*. Addison-Wesley. "Online text [1] (https://archive.org/details/in.e rnet.dli.2015.90211)"
- George K. Zipf (1935) *The Psychobiology of Language*. Houghton-Mifflin.

Secondary:

- Alexander Gelbukh and Grigori Sidorov (2001) "Zipf and Heaps Laws' Coefficients Depend on Language" (http://www.gelbukh.com/CV/Public ations/2001/CICLing-2001-Zipf.htm). Proc. CICLing-2001, *Conference on Intelligent Text Processing and Computational Linguistics*, February 18–24, 2001, Mexico City. Lecture Notes in Computer Science N 2004, ISSN 0302-9743 (https://www.worldcat.org/search?fq=x0:jrnl&q=n2:0 302-9743), ISBN 3-540-41687-0, Springer-Verlag: 332–335.
- Damián H. Zanette (2006) "Zipf's law and the creation of musical context (http://xxx.arxiv.org/abs/cs.CL/0406015)," *Musicae Scientiae* *10*: 3–18.
- Frans J. Van Droogenbroeck (2016), Handling the Zipf distribution in computerized authorship attribution (https://www.academia.edu/2414 7736/)
- Frans J. Van Droogenbroeck (2019), An essential rephrasing of the Zipf-Mandelbrot law to solve authorship attribution applications by Gaussian statistics (https://www.academia.edu/40029629)
- Kali R. (2003) "The city as a giant component: a random graph approach to Zipf's law," *Applied Economics Letters 10*: 717–720(4)
- Gabaix, Xavier (August 1999). "Zipf's Law for Cities: An Explanation" (http://pages.stern.nyu.edu/~xgabaix/papers/zipf.pdf) (PDF). *Quarterly Journal of Economics*. **114** (3): 739–67. CiteSeerX 10.1.1.180.4097 (ht tps://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.180.4097). doi:10.1162/003355399556133 (https://doi.org/10.1162%2F00335539 9556133). ISSN 0033-5533 (https://www.worldcat.org/issn/0033-5533).
- Axtell, Robert L; Zipf distribution of US firm sizes (http://www.science mag.org/content/293/5536/1818.short), Science, 293, 5536, 1818, 2001, American Association for the Advancement of Science
- Ramu Chenna, Toby Gibson; Evaluation of the Suitability of a Zipfian Gap Model for Pairwise Sequence Alignment (http://www.worldcomp-p roceedings.com/proc/p2011/BIC4329.pdf), International Conference on Bioinformatics Computational Biology: 2011.
- Shyklo A. (2017); Simple Explanation of Zipf's Mystery via New Rank-Share Distribution, Derived from Combinatorics of the Ranking Process (https://ssrn.com/abstract=2918642), Available at SSRN: https://ssrn.com/abstract=2918642.

External links

- Strogatz, Steven (2009-05-29). "Guest Column: Math and the City" (http://judson.blogs.nytimes.com/2009/05/19/math-and-the-city/). *The New York Times*. Retrieved 2009-05-29.—An article on Zipf's law applied to city populations
- Seeing Around Corners (Artificial societies turn up Zipf's law) (https://www.theatlantic.com/issues/2002/04/rauch.htm)
- PlanetMath article on Zipf's law (https://web.archive.org/web/20021018011011/http://planetmath.org/encyclopedia/ZipfsLaw.html)
- Distributions de type "fractal parabolique" dans la Nature (French, with English summary) (http://www.hubbertpeak.com/laherrere/fractal.htm)
- An analysis of income distribution (https://www.newscientist.com/article.ns?id=mg18524904.300)
- Zipf List of French words (http://www.lexique.org/listes/liste_mots.txt)
- Zipf list for English, French, Spanish, Italian, Swedish, Icelandic, Latin, Portuguese and Finnish from Gutenberg Project and online calculator to rank words in texts (http://1.1o1.in/en/webtools/semantic-depth)
- Citations and the Zipf–Mandelbrot's law (https://arxiv.org/abs/physics/9901035)
- Zipf's Law examples and modelling (1985) (http://www.geoffkirby.co.uk/ZIPFSLAW.pdf)
- Complex systems: Unzipping Zipf's law (2011) (http://www.nature.com/nature/journal/v474/n7350/full/474164a.html)
- Benford's law, Zipf's law, and the Pareto distribution (http://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/) by Terence Tao.
- "Zipf law" (https://www.encyclopediaofmath.org/index.php?title=Zipf_law), *Encyclopedia of Mathematics*, EMS Press, 2001 [1994]

This page was last edited on 25 February 2022, at 11:07.