Zipf–Mandelbrot law

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **Zipf-Mandelbrot law** is a discrete probability distribution. Also known as the Pareto-Zipf law, it is a power-law distribution on ranked data, named after the linguist George Kingsley Zipf who suggested a simpler distribution called Zipf's law, and the mathematician Benoît Mandelbrot, who subsequently generalized it.

Zipf–Mandelbrot

parameters:	$N \in \{1, 2, 3 \ldots\}$ (integer) $q \in [0; \infty)$ (real) $s > 0$ (real)
support:	$k \in \{1, 2, \ldots, N\}$
pmf:	$\frac{1/(k+q)^s}{H_{N,q,s}}$
cdf:	$\frac{H_{k,q,s}}{H_{N,q,s}}$
mean:	$\frac{H_{N,q,s-1}}{H_{N,q,s}} - q$
median:	
mode:	1
variance:	
skewness:	
ex.kurtosis:	
entropy:	
mgf:	
cf:	

The probability mass function is given by:

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

where $H_{N,q,s}$ is given by:

$$H_{N,q,s} = \sum_{i=1}^{N} \frac{1}{(i+q)^s}$$

which may be thought of as a generalization of a harmonic number. In the limit as N approaches infinity, this becomes the Hurwitz zeta function $\zeta(q,s)$. For finite N and $q = 0$ the Zipf–Mandelbrot law becomes Zipf's law. For infinite N and $q = 0$ it becomes a Zeta distribution.

Contents

- 1 Applications
- 2 Notes
- 3 References
- 4 External links

Applications

The distribution of words ranked by their frequency in a random text corpus is generally a power-law distribution, known as Zipf's law.

If one plots the frequency rank of words contained in a large corpus of text data versus the number of occurrences or actual frequencies, one obtains a power-law distribution, with exponent close to one (but see Gelbukh & Sidorov, 2001).

In ecological field studies, the relative abundance distribution (i.e. the graph of the number of species observed as a function of their abundance) is often found to conform to a Zipf–Mandelbrot law.[1]

Within music, many metrics of measuring "pleasing" music conform to Zipf–Mandlebrot distributions.[2]

Notes

1. ^ Mouillot, D; Lepretre, A (2000). "Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity" (http://cat.inist.fr/?aModele=afficheN&cpsidt=1411186) . *Environmental Monitoring and Assessment* (Springer) **63** (2): 279–295. doi:10.1023/A:1006297211561 (http://dx.doi.org/10.1023%2FA%3A1006297211561) . http://cat.inist.fr/?aModele=afficheN&cpsidt=1411186. Retrieved 24 Dec 2008.
2. ^ Manris, B; Vaughan, D, Wagner, CS, Romero, J, Davis, RB. "Evolutionary Music and the Zipf-Mandelbrot Law: Developing Fitness Functions for Pleasant Music" (http://shaunwagner.com/writings_computer_evomus.html) . *Proceedings of 1st European Workshop on Evolutionary Music and Art (EvoMUSART2003)* **611**. http://shaunwagner.com/writings_computer_evomus.html.

References

- Mandelbrot, Benoît (1965). "Information Theory and Psycholinguistics". In B.B. Wolman and E. Nagel. *Scientific psychology*. Basic Books. Reprinted as
  - Mandelbrot, Benoît (1968) [1965]. "Information Theory and Psycholinguistics". In R.C. Oldfield and J.C. Marchall. *Language*. Penguin Books.
- Zipf, George Kingsley (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press.

External links

- Z. K. Silagadze: Citations and the Zipf-Mandelbrot's law (http://arxiv.org

/abs/physics/9901035)
- NIST: Zipf's law (http://www.nist.gov/dads/HTML/zipfslaw.html)
- W. Li's References on Zipf's law (http://www.nslij-genetics.org/wli/zipf/index.html)
- Gelbukh & Sidorov, 2001: Zipf and Heaps Laws' Coefficients Depend on Language (http://www.gelbukh.com/CV/Publications/2001/CICLing-2001-Zipf.htm)

Retrieved from "http://en.wikipedia.org/wiki/Zipf%E2%80%93Mandelbrot_law"
Categories: Discrete distributions | Power laws | Computational linguistics | Quantitative linguistics