

Why Zipf's Law: A Symmetry-Based Explanation

Daniel Cervantes¹, Olga Kosheleva², and
Vladik Kreinovich¹

Departments of ¹Computer Science and ²Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
drcervantes@miners.utep.edu, olgak@utep.edu,
vladik@utep.edu

Abstract

In many practical situations, we have probability distributions for which, for large values of the corresponding quantity x , the probability density has the form $\rho(x) \sim x^{-\alpha}$ for some $\alpha > 0$. While, in principle, we have laws corresponding to different α , most frequently, we encounter situations – first described by Zipf for linguistics – when $\alpha \approx 1$. The fact that Zipf's has appeared frequently in many different situations seems to indicate that there must be some fundamental reason behind this law. In this paper, we provide a possible explanation.

Mathematics Subject Classification: 60K10 58D19 58J70

Keywords: Zipf's law; power distributions; symmetries

1 Power Laws are Ubiquitous

In many practical situations, we have probability distributions for which:

- for large values of the corresponding quantity x ,
- the probability density has the form $\rho(x) \sim x^{-\alpha}$ for some $\alpha > 0$.

In principle, we have laws corresponding to different α . However, most frequently, we encounter situations – first described by Zipf for linguistics – when

$$\alpha \approx 1.$$

In linguistics, this law means that:

- if we sort words by frequency,
- then the k -th word in this ordering has frequency

$$f_k \approx \frac{c}{k};$$

see, e.g., [1] are references therein.

2 Why Zipf's Law?

The fact that Zipf's law appears frequently in many different situations seems to indicate that there must be some fundamental reason behind this law.

In this paper, we provide a possible explanation.

3 First Explanation

In many real-life cases, the corresponding phenomenon do not have a preferred value of the quantity x . As a result, the corresponding equations do not change if we simply change the measuring unit.

It is reasonable to require that corresponding probability distribution should also not change. How can we describe this requirement in precise terms?

- If we replace the measuring unit by a λ time smaller one,
- then all numerical values of a quantity x are multiplied by λ :

$$x \rightarrow x' = \lambda \cdot x.$$

For example, 2 m becomes 200 cm.

Similarly, the probability density – number of events per unit of x :

- becomes λ times smaller
- when this unit decreases by a factor of λ :

$$\rho'(x') = \rho'(\lambda \cdot x) = \frac{\rho(x)}{\lambda}.$$

We require that the formula for the probability density remains the same in both units, i.e., that

$$\rho'(x) = \rho(x).$$

Then we conclude that $\rho(\lambda \cdot x) = \frac{\rho(x)}{\lambda}$. For $x = 1$ and $\lambda = z$, if we denote $c \stackrel{\text{def}}{=} \rho(1)$, we get

$$\rho(z) = \frac{c}{z}.$$

This exactly Zipf's law.

4 Alternative Explanation

Instead of probabilities, we may have general densities with

$$\int \rho(x) dx < +\infty.$$

In this case, scale-invariance means that

$$\rho(\lambda \cdot x) = c(\lambda) \cdot \rho(x)$$

for some function $c(\lambda)$.

If we differentiate both sides by λ , we get:

$$x \cdot \rho'(\lambda \cdot x) = c'(\lambda) \cdot \rho(x).$$

In particular, for $\lambda = 1$, we get:

$$x \cdot \rho'(x) = c \cdot \rho(x), \text{ where } c \stackrel{\text{def}}{=} c'(1).$$

In other words,

$$x \cdot \frac{d\rho}{dx} = c \cdot \rho.$$

If we move all the terms containing ρ to one side and all the terms containing x to another side, we get

$$\frac{d\rho}{\rho} = c \cdot \frac{dx}{x}.$$

Integrating both sides, we get

$$\ln(\rho) = c \cdot \ln(x) + C.$$

Applying exp to both sides, we get

$$\rho(x) = a \cdot x^{-\alpha}, \text{ where } a \stackrel{\text{def}}{=} e^C \text{ and } \alpha \stackrel{\text{def}}{=} -c.$$

The finiteness requirement $\int \rho(x) dx < +\infty$ implies that $\alpha > 1$. We may have many different contributing processes with different α :

$$\rho(x) = \sum_i a_i \cdot x^{-\alpha_i}.$$

Asymptotically, the terms with the smallest α_i prevail:

$$\rho(x) = \sum_i a_i \cdot x^{-\alpha_i} \sim x^{-\min_i \alpha_i}.$$

When we have many different processes, with high probability, some of them will be close to 1. So, we have $\rho(x) \sim x^{-\alpha}$ with $\alpha \approx 1$.

This is exactly Zipf's law.

Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122.

References

- [1] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, California, 1983.

Received: March 31, 2018