Why Zipf's Law: A Symmetry-Based Explanation

Daniel Cervantes, Olga Kosheleva, and Vladik Kreinovich
University of Texas at El Paso, El Paso, TX 79968, USA
drcervantes@miners.utep.edu, olgak@utep.edu, vladik@utep.edu

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 $\alpha > 0$.

Zipf's law is the most frequent. While, in principle, we have laws corresponding to different α, most frequently, we encounter situations – first described by Zipf for linguistics – when $\alpha \approx 1$.

Why Zipf's law? 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 talk, we provide a possible explanation.

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 does not change if we simply change the measuring unit. It is reasonable to require that the formulas describing the 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: $x \to x' = \lambda \cdot x$ (e.g., 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: $\rho'(x') = \rho'(\lambda \cdot x) = \dfrac{\rho(x)}{\lambda}$. If we require that the formula remains the same in both units, i.e., that $\rho'(x) = \rho(x)$, then we conclude that $\rho(\lambda \cdot x) = \dfrac{\rho(x)}{\lambda}$. For $x = 1$ and $\lambda = z$, if we denote $c \stackrel{\text{def}}{=} \rho(1)$, we get $\rho(z) = \dfrac{c}{z}$ – exactly Zipf's law.

Alternative explanation. Instead of probabilities, we may have general densities with $\int \rho(x)\,dx < +\infty$. In this case, scale-invariance implies $\rho(x) \sim x^{-\alpha}$ for some α. The finiteness requirement $\int \rho(x)\,dx < +\infty$ implies that $\alpha > 1$. We may have many different contributing processes with different α: $\rho(x) = \sum\limits_i c_i \cdot x^{-\alpha_i}$. Asymptotically, the terms with the smallest α_i prevail. 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 $\alpha \approx 1$.