Why Zipf's Law: A Symmetry-Based Expalnation

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 . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .

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

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

Power Laws are . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .

2. Why Zipf's Law?

- *Fact:* fact that Zipf's las appears frequently in many different situations.
- This seems to indicate that there must be some fundamental reason behind this law.
- In this talk, we provide a possible explanation.

Power Laws are . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .

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 doe 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: $x \to x' = \lambda \cdot x$.
- Example: 2 m becomes 200 cm.

Power Laws are . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .

4. First Explanation (cont-d)

- 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 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

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

- This exactly Zipf's law.

Power Laws are . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .

5. 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,
$$x \cdot \frac{d\rho}{dx} = c \cdot \rho.$$

Power Laws are . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .

6. Alternative Explanation (cont-d)

- 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:

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

Power Laws are . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .

7. Alternative Explanation (cont-d)

- Asymptotically, the terms with the smallest α_i prevail:

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

- 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.

Power Laws are . . .
Why Zipf's Law?
First Explanation
First Explanation (cont-d)
Alternative Explanation
Alternative . . .
Alternative . . .