Set-5: Modelling data with power laws (Pareto's law and Zipf's law)

Shail Patel (202201297)* and Divyakumar Tandel (202201469)†
Dhirubhai Ambani Institute of Information & Communication Technology, Gandhinagar, Gujarat 382007, India
CS302, Modeling and Simulation

In this lab, we modeled the wealth distribution in India in the year 2021 and the dependency network of Debian using the Pareto's law and Zipf's law.

I. PARETO DISTRIBUTION OF WEALTH IN INDIA

A. Model

To model the power law that the data follow,apply the function

$$N(x) = A + Bx^{-\alpha} \tag{1}$$

Given that x is the amount of wealth and N(x) is the frequency distribution of wealth holders.

B. Results

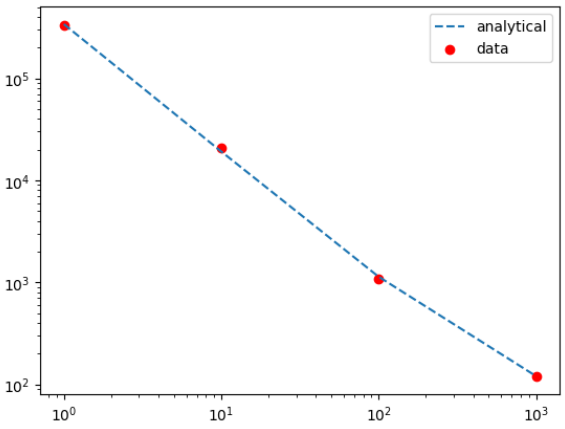

FIG. 1: The dotted graph is the analytical solution of the power law. The red dots are the data points given in the dataset.This shows Pareto distribution of wealth in India in the year 2021.

Here, A = 60, B = 340000 and $\alpha = 5/4$.

*Electronic address: 202201297@daiict.ac.in
†Electronic address: 202201469@daiict.ac.in

II. ZIPF'S LAW IN THE DEPENDENCY NETWORK OF DEBIAN

A. Model

The global distribution of power law is given by

$$\phi(x) = [\eta + (\frac{x+\lambda}{c})^{-\mu\alpha}]^{-1/\mu} \tag{2}$$

in which α is a power law exponent, μ is a non-linear saturation exponent, η is a 'tuning' parameter for non-linearity, and λ is another parameter that is instrumental in setting a limiting scale for poorly connected nodes.

With $\mu = -1$ (which implies a power law in the distribution) and with $\alpha = -2$ (which implies that the power law is specifically Zipf's law), the saturation properties of the network (for any value of λ and η) can be abstracted from equation (2) as

$$\phi(x) = \eta + (\frac{c}{x+\lambda})^2 \tag{3}$$

B. Results

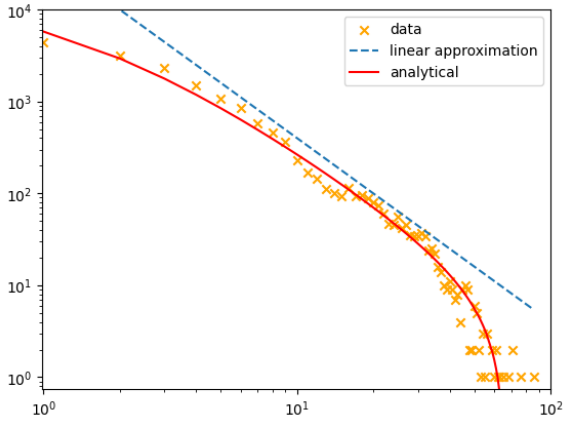

FIG. 2: This graph shows the network of incoming links in the Etch release.

Here, $\alpha = -2$, $\mu = -1$, $\eta = -8$, $\lambda = 1.5$ and c = 190.

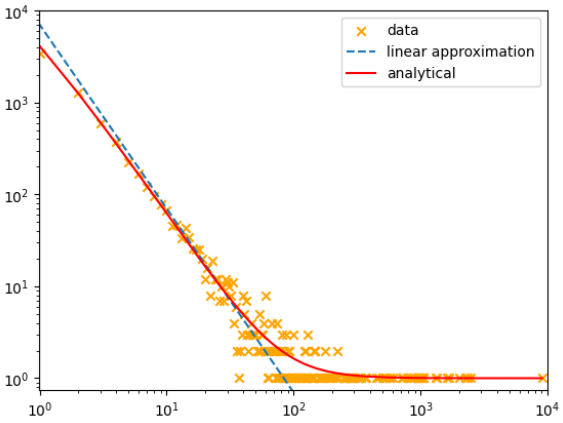

FIG. 3: This graph shows the network of outgoing links in the Etch release.

Here, $\alpha = -2$, $\mu = -1$, $\eta = 1$, $\lambda = 0.25$ and the data are adapted for c = 80. A single top node is to be seen for x = 9025.

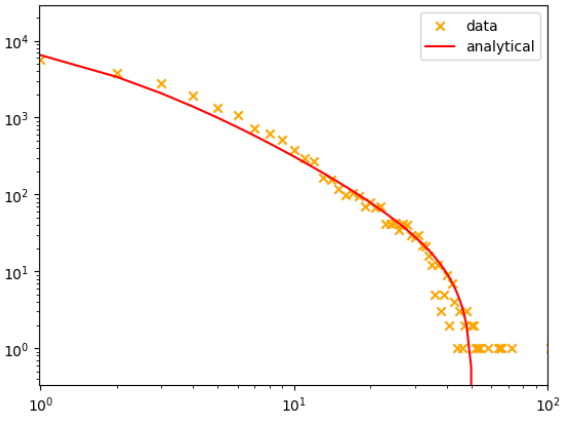

FIG. 4: This graph shows the network of incoming links in the Lenny release.

Here, $\alpha = -2$, $\mu = -1$, $\eta = -15$, $\lambda = 1.6$ and c = 210.

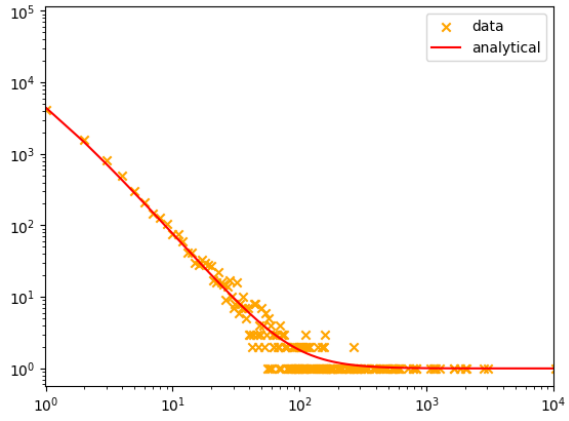

FIG. 5: This graph shows the network of outgoing links in the Lenny release.

Here, $\alpha = -2$, $\mu = -1$, $\eta = 1$, $\lambda = 0.35$ and the data is fitted for c = 90. A solitary top node is to be seen for x = 10446.

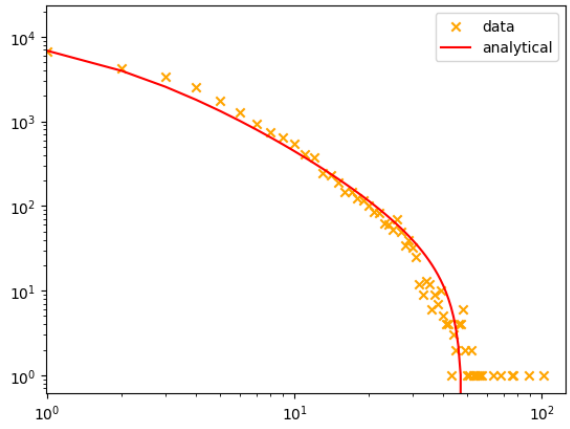

FIG. 6: This graph shows the network of incoming links in the Squeeze release.

Here, $\eta = -28$, $\lambda = 2.2$ and c = 265.

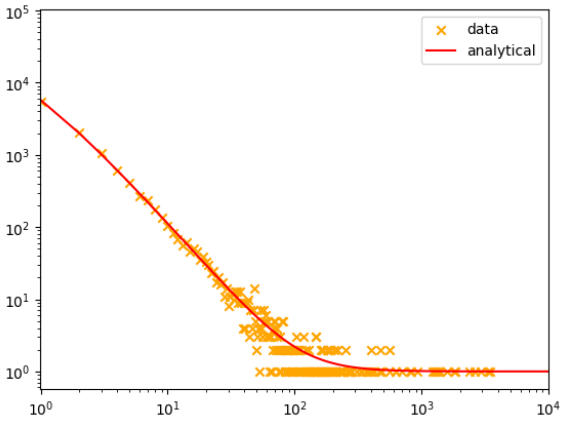

FIG. 7: This graph shows the network of outgoing links in the Squeeze release.

Here, $\eta = 1$, $\lambda = 0.45$ and the data is fitted for c = 110. The richest node in this distribution has 12470 links.

III. SALIENT FEATURES:

1. A mathematical pattern that is closely similar to Zipf's law is the power law distribution, which describes the frequency of packages in the Debian dependency network.

2. The value η in Debian data Modelling for Outgoing data, however, models the saturation behaviour towards a limiting scale of ϕ for large values of x.

3. According to the Pareto distribution, a very small percent of the population controls majority of the country's wealth.

[1] Rajiv Nair, G. Nagarjuna, and Arnab K. Ray, *'Finite-Size Effects in the Dependency Networks of Free and Open-Source Software'*.