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

Natansh Shah (202201445)* and Tirth Modi (202201513)†
Dhirubhai Ambani Institute of Information & Communication Technology, Gandhinagar, Gujarat 382007, India
CS302, Modelling and Simulation

This study models power law distributions in two distinct domains: wealth distribution in India using Pareto's law and dependency networks in Debian using Zipf's law. By analyzing real-world data, it reproduces key plots and derives insights into inequality and structural dependencies.

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 shows pareto distribution of wealth in India.

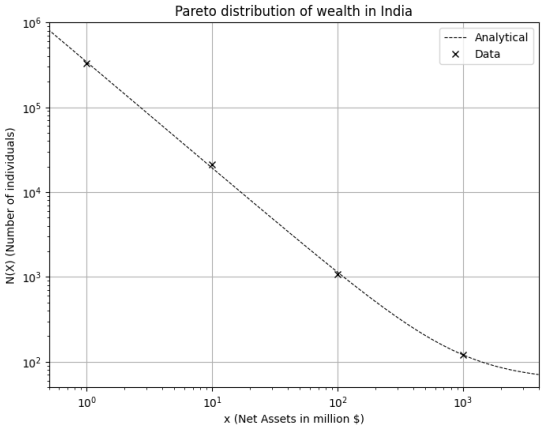

FIG. 1: Here A=60, B=340000 and $\alpha = 5/4$.

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

A. Model

The global power law distribution 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 nonlinear saturation exponent, η is a "tuning" parameter for nonlinearity, and λ is another parameter that is instrumental in setting a limiting scale for the poorly connected nodes.

With $\mu = -1$ (implying a power-law in the distribution) and with $\alpha = -2$ (implying 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 shows the network of incoming links in the Etch release.

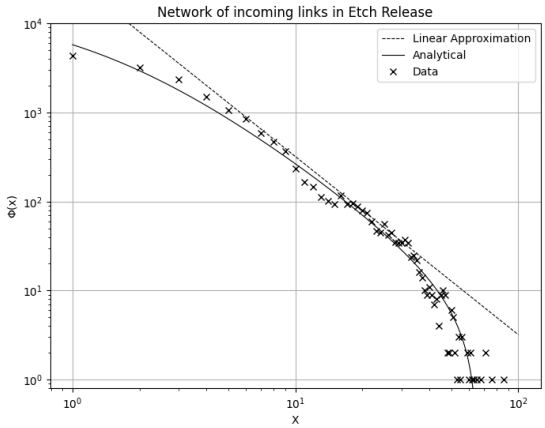

FIG. 2: Here $\alpha = -2$, $\mu = -1$, $\eta = -8$, $\lambda = 1.5$ and c = 190.

*Electronic address: 202201445@daiict.ac.in
†Electronic address: 202201513@daiict.ac.in

Fig. 3 shows the network of outgoing links in the Etch release.

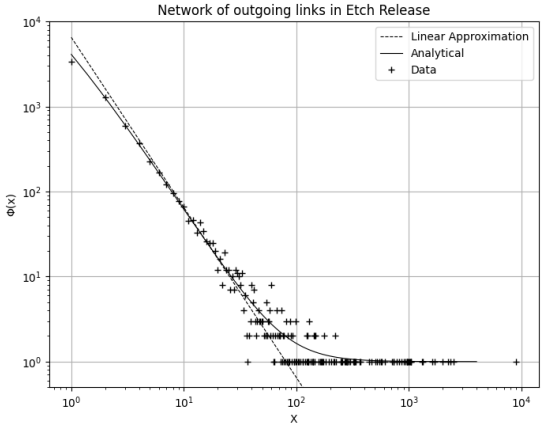

FIG. 3: Here $\alpha = -2$, $\mu = -1$, $\eta = 1$, $\lambda = 0.25$ and the data is fitted for c = 80. A solitary top node is to be seen for x = 9025.

Fig. 4 shows the network of incoming links in the Lenny release.

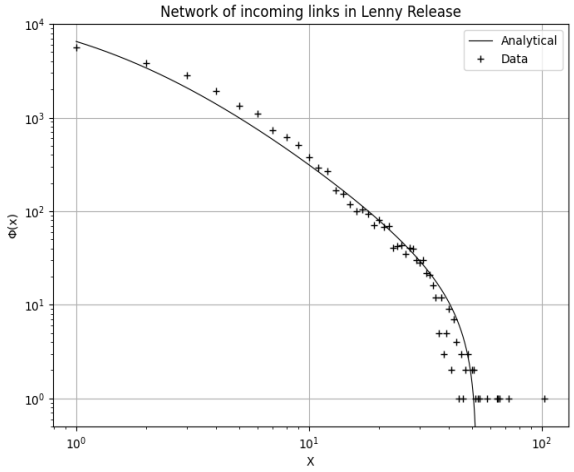

FIG. 4: Here $\alpha = -2$, $\mu = -1$, $\eta = -15$, $\lambda = 1.6$ and c = 210.

Fig. 5 shows the network of outgoing links in the Lenny release.

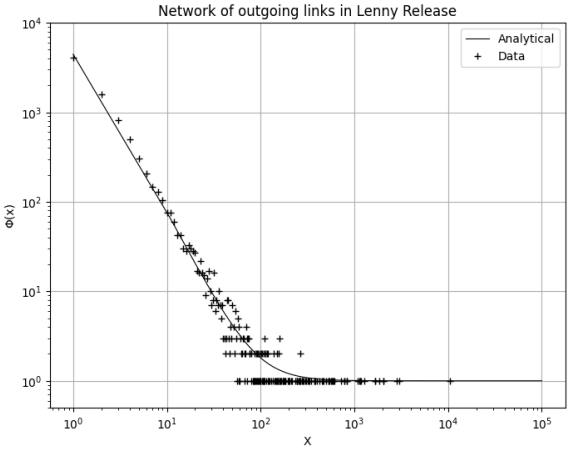

FIG. 5: Here $\alpha = -2$, $\mu = -1$, $\eta = 1$, $\lambda = 0.35$ and the data is fitted for c = 90. A solitary top node is to be seen for x = 10446.

Fig. 6 shows the network of incoming links in the Squeeze release.

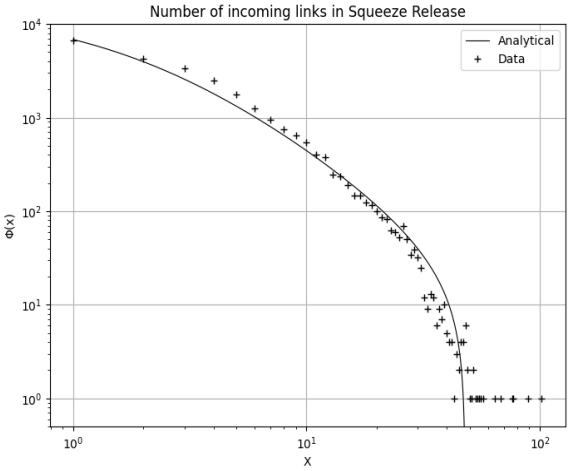

FIG. 6: Here $\eta = -28$, $\lambda = 2.2$ and c = 265.

Fig. 7 shows the network of outgoing links in the Squeeze release.

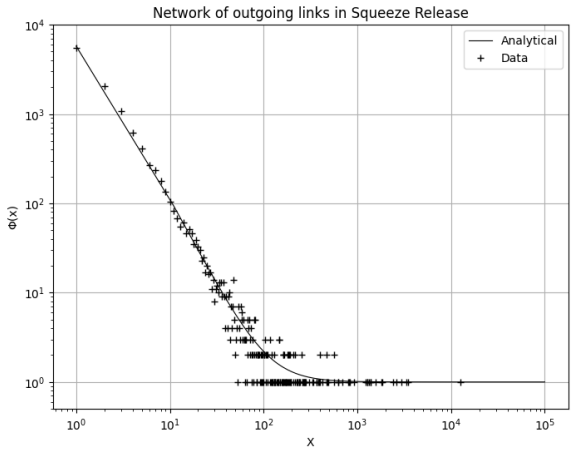

FIG. 7: Here $\eta = 1$, $\lambda = 0.45$ and the data is fitted for c = 110. The richest node in this distribution has 12470 links.

III. CONCLUSIONS

In conclusion, our exploration involved a basic mathematical model, the logistic equation, with respect to pareto distribution and ziff's law.

⇒ 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.

⇒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.

⇒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'*.