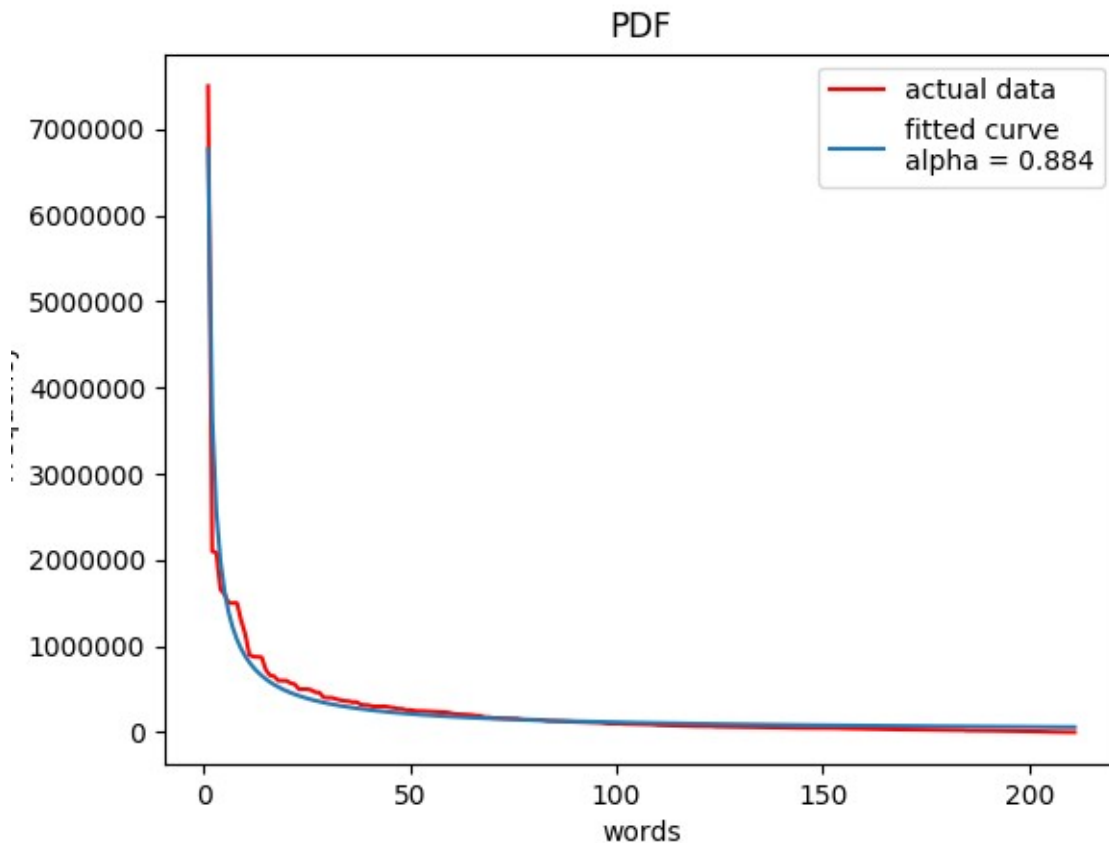# **Project report**

Data set used : The numbers of customers affected in electrical blackouts in the United States between 1984 and 2002.

This kind of distribution in frequency is known as Zipf's law. It is an emperical law formulated using mathematical statistics. Normally Zipf's Law holds for all languages(the frequency of any word is inversely proportional to its rank in the frequency table), even non-natural ones like Esperanto. But here we can observe it on the blackout data as well. Though the actual reason for occuring this pattern in several physical and social phenomena is still not well understood, but two theories ,
1. principle of least effort (in case of language usage, neither speakers nor hearers using a given language want to work any harder than necessary to reach understanding) and
2. preferential attachment (intuitively, "the rich get richer" or "success breeds success")
can be possible explanation in some particular cases.

Mridul Gain (17MCMC40)

Log-log representation
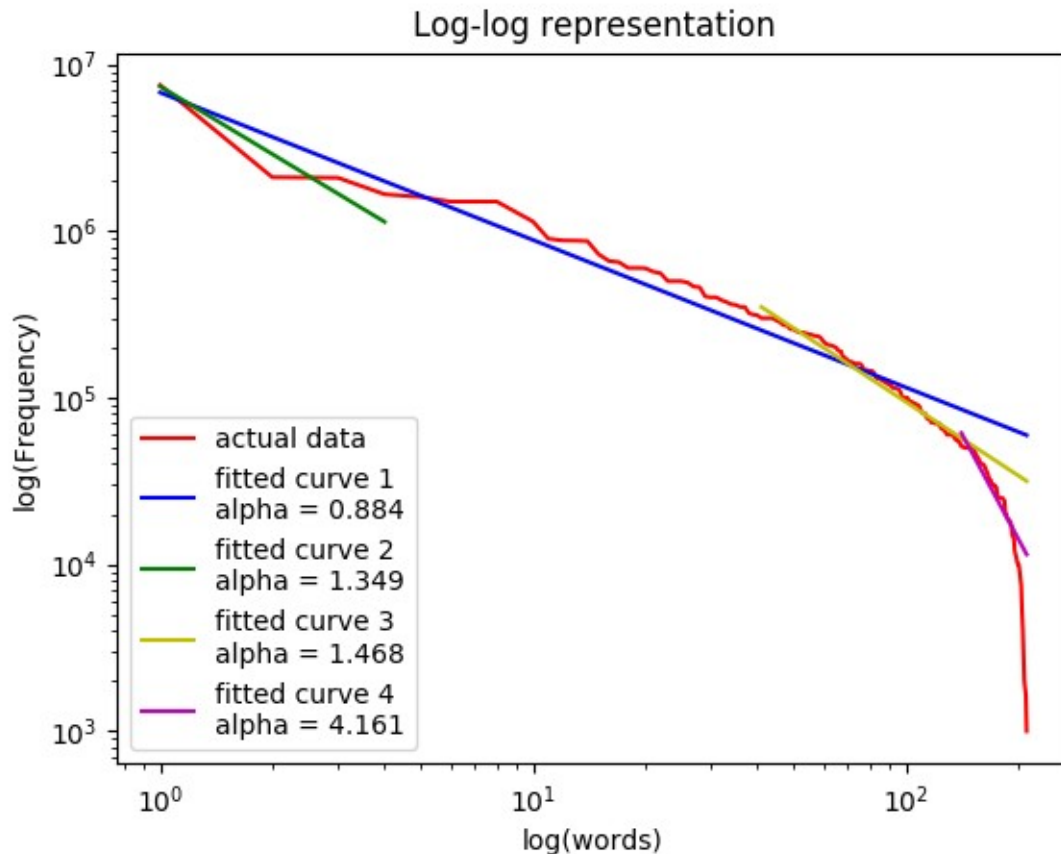
An interesting outcome of the experiment is observed when log-log plot is used. A broken power law in the frequency distribution is noticed which shows there are large ammount (more than 50%) of data which are not lying on the primary power law curve. Zipf's law corresponds to the middle linear portion of the curve, roughly following the blue ($\alpha = 0.884$) line, while the early part is closer to the green ($\alpha = 1.349$) line while the later part is closer to the yellow ($\alpha = 1.468$) and magenta ($\alpha = 4.161$) line. These lines correspond to three distinct parameterizations of the Zipf–Mandelbrot distribution, overall a broken power law with three segments: a head, middle, and tail.

Resources:

https://en.wikipedia.org/wiki/Zipf%27s_law

https://en.wikipedia.org/wiki/Power_law

https://www.statisticshowto.datasciencecentral.com/power-law/