HW 1 Part 1

Samanvitha Basole Riti Gupta Anand

September 10, 2018

1 1. MLE

(i) Maximum Likelihood Estimate for:

(a) neither - $\frac{198}{884421} = \mathbf{0.0002239}$

(b) a - $\frac{14945}{884421} = \mathbf{0.01690}$

(c) borrower - $\frac{2}{884421} = \mathbf{0.000002261}$

(d) nor - $\frac{1028}{884421} = \mathbf{0.001162}$

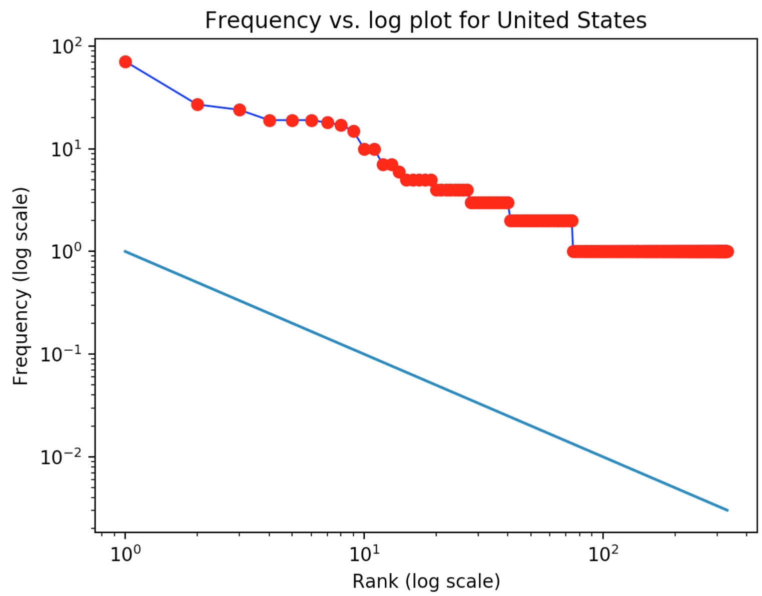
(e) lender - $\frac{2}{884421} = \mathbf{0.000002261}$

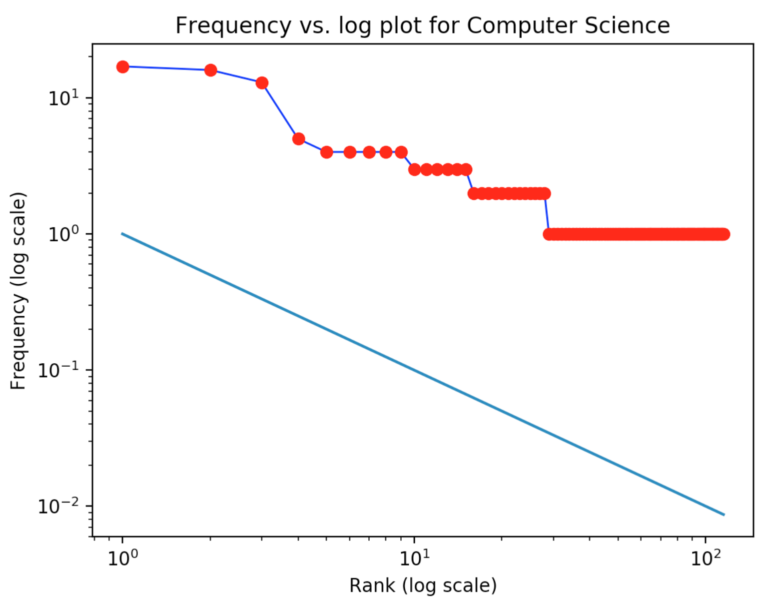
(f) be - $\frac{7166}{884421} = \mathbf{0.008102}$

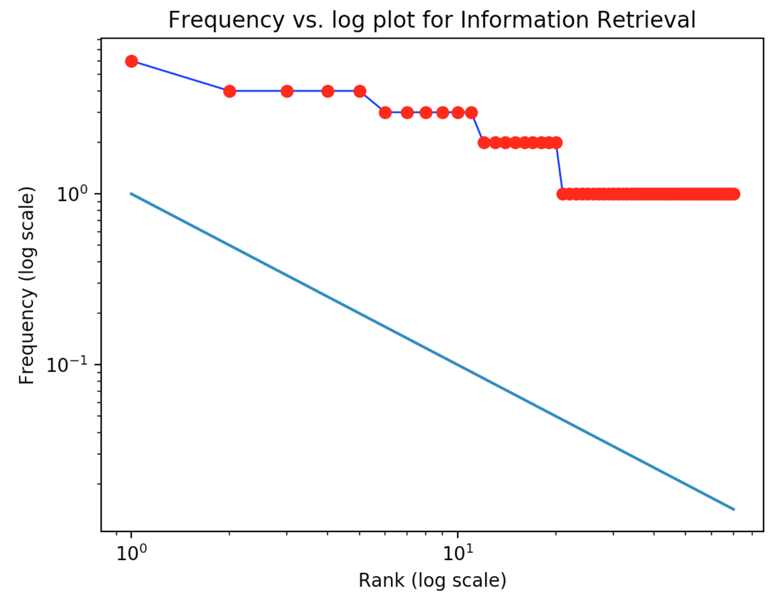
(ii) Neither a borrower nor a lender be - $0.0002239 * 0.01690 * 0.000002261 * 0.001162 * 0.01690 * 0.000002261 * 0.008102 = 3.269e^{-24}$

2 2. Frequency-Rank log scale plot

The following plots were created for three small wikipedia texts with search terms: United States, Computer Science, and Information Retrieval.


Frequency vs. log plot for United States
Frequency (log scale)
Rank (log scale)
10^2
10^1
10^0
10^{-1}
10^{-2}
10^0
10^1
10^2


Frequency vs. log plot for Computer Science
Frequency (log scale)
Rank (log scale)
10^1
10^0
10^{-1}
10^{-2}
10^0
10^1
10^2



The below screenshot lists the data for 'Computer Science':

Rank	Frequency	Zipf	Word
1	17	1	the
2	16	0.5	of
3	13	0.333333	and
4	5	0.25	computer
5	4	0.2	study
6	4	0.166667	theory
7	4	0.142857	computation
8	4	0.125	to
9	4	0.111111	fields
10	3	0.1	that
11	3	0.0909091	in
12	3	0.0833333	programming
13	3	0.0769231	science
14	3	0.0714286	is
15	3	0.0666667	computational
16	2	0.0625	considers
17	2	0.0588235	design
18	2	0.0555556	as
19	2	0.0526316	such
20	2	0.05	use
21	2	0.047619	computers
22	2	0.0454545	various
23	2	0.0434783	systems
24	2	0.0416667	challenges

In all of these graphs, the frequency of a word is inversely proportional to its rank. Even for this small data, Zipf's Law is still followed. It can also be seen that the slope of the regression line for the data is very similar to the slope of the Zipf's Law line. As the search term got more specific (United States → Computer Science → Information Retreival), the slope got more horizontal. The line for the 'United States' term and the line for the Zipf's Law are almost parallel.

3 3. Markov Model

$$M = \begin{bmatrix} 0.0 & 0.4 & 0.0 & 0.6 \\ 0.7 & 0.0 & 0.3 & 0.0 \\ 0.0 & 0.8 & 0.0 & 0.2 \\ 0.1 & 0.9 & 0.0 & 0.0 \end{bmatrix}$$

The results for the calculations using the transition matrix are as follows:

```
[1, 0, 0, 0]M^5:
[ 0.36612  0.36304  0.1458   0.12504]

[1, 0, 0, 0]M^10:
[ 0.3012797   0.38989305  0.12005916  0.1887681 ]

[1, 0, 0, 0]M^15:
[ 0.29392532  0.39073658  0.11734021  0.19799789]

[1, 0, 0, 0]M^20:
[ 0.29323018  0.39046775  0.11711497  0.1991871 ]

[1, 0, 0, 0]M^25:
[ 0.29318647  0.39038465  0.11710684  0.19932203]

[1, 0, 0, 0]M^30:
[ 0.2931879   0.39036847  0.11710883  0.1993348 ]

[1, 0, 0, 0]M^35:
[ 0.29318905  0.39036587  0.11710948  0.1993356 ]

[1, 0, 0, 0]M^40:
[ 0.29318931  0.3903655   0.11710961  0.19933557]

[1, 0, 0, 0]M^45:
[ 0.29318936  0.39036545  0.11710963  0.19933555]

[1, 0, 0, 0]M^50:
[ 0.29318937  0.39036545  0.11710963  0.19933555]
```

The results converge since the values remain almost the same after many it-

erations. According to the Perron-Frobenius theorem, if by whatever means, $\lim_{k\to\infty} M^k$ is found , then the stationary distribution of the Markov chain can be easily determined for any starting distribution.
If the Markov chain is time-homogeneous, then the transition matrix M is the same after each step, so the k-step transition probability can be computed as the k-th power of the transition matrix, M^k.
If the Markov chain is irreducible and aperiodic, then there is a unique stationary distribution π. Additionally, in this case M^k converges to a rank-one matrix in which each row is the stationary distribution π, that is, $\lim_{k\to\infty} M^k = 1\pi$ where 1 is the column vector with all entries equal to 1.