PRINCIPLES OF COMPLEX SYSTEMS

ASSIGNMENT 4

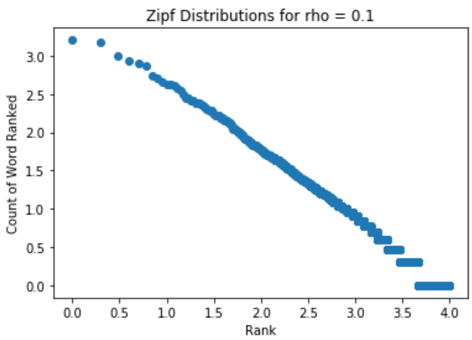
January 6, 2023

Maxfield Green
University of Vermont
Comp. Systems and Data Sci.

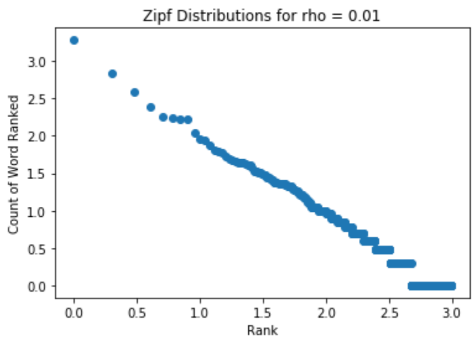
0.1 Problem 1

Rich-Gets-Richer Model
Show Zips distributions for ρ= 0.10, 0.01, and 0.001. and perform regression to test $\alpha = 1 - \rho$ Run the simulation for long enough to produce decent scaling laws.

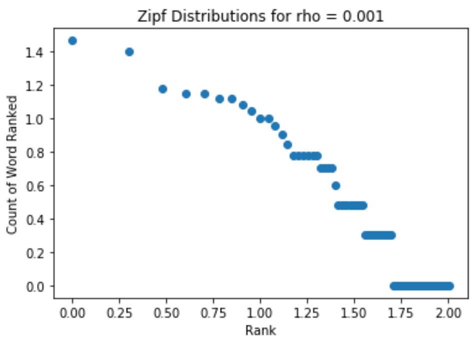
The first mover advantage is very clear. We can see that the first elephant model attracts the largest share of future elephants on the market. We will see this to be true in the following analysis of Simon's Model and the Rich Get Richer story.



(a)



(b)



(c)

Figure 1: Simon's Model Power Law Distributions for different values of ρ

0.2 Problem 2

For Herbert Simon's model of what we've called Random Competitive Replication, we found in class that the normalized number of groups in the long time limit, n_k, satisfies the following difference equation:

$$\frac{n_k}{n_{k-1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} \tag{1}$$

where $k \leq 2$. The model parameter ρ is the probability that a newly arriving node forms a group of its own(or is a novel word, starts a new city, has a unique flavor, ect). For $k = 1$, we have instead

$$n_1 = \rho - (1-\rho)n_1 \tag{2}$$

which directly give us n_1 in terms of ρ.
(a) Derive the exact solution for n_k in terms of gamma functions and ultimately a beta function.

$$n_k = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} * (n_k - 1) \tag{3}$$

$$\tag{4}$$

We can expand this to the following form:

$$n_k = [\frac{(k-1)(1-\rho)}{1+(1-\rho)k}][\frac{(k-2)(1-\rho)}{1+(1-\rho)(k-1)}][\frac{(k-3)(1-\rho)}{1+(1-\rho)(k-2)}]...[\frac{\zeta(1-\rho)}{1+(1-\rho)(\eta)}] \tag{5}$$

$$\tag{6}$$

$$\tag{7}$$

The numerator can be re-written as : $[(1-p)^k(\Gamma(k)])$
The denominator can be re-written as follows if $z = (1-\rho)$:

$$[1+zk][1+z(k-1)]...[1+z]... \tag{8}$$

$$= k^k(\frac{1}{z}+k)(\frac{1}{z}+k-1)(\frac{1}{z}+k-2)...(\frac{1}{z})(\frac{1}{z}-1).. \tag{9}$$

$$= \frac{k^k(\frac{1}{z}+k)(\frac{1}{z})..}{(\frac{1}{z})(\frac{1}{z}-1)...} \tag{10}$$

$$= Z^K\frac{\Gamma(\frac{1}{z}+k+1)}{\Gamma(\frac{1}{z}+1)} \tag{11}$$

$$\tag{12}$$

Stitching our original fraction back together results in:

$$\frac{\Gamma(k)z^k\Gamma(\frac{1}{z}+1)}{z^k\Gamma(\frac{1}{z}+k+1)} \tag{13}$$

$$= \frac{\Gamma(k)\Gamma(\frac{1}{z}+1)}{\Gamma(\frac{1}{z}+k+1)} \tag{14}$$

This particular form of Gamma functions allows us to construct a beta function:

$$B(k,\frac{1}{z}+1) = \frac{\Gamma(k)\Gamma(\frac{1}{z}+1)}{\Gamma(\frac{1}{z}+k+1)} \tag{15}$$

With another use of Stirlings approximation, for large x and fixed y.

$$B(x,y)\ x^{-y} \tag{16}$$

$$n_k = k^{\frac{1}{\rho-1}+1} \tag{17}$$

This solution converges to 0 for all values of ρ

0.3 Problem 3

What happens to γ in the limits ρ -> 0 and ρ->1? Explain in a sentence or two whats going on in these cases and how the specific limiting value of γ makes sense.

$$\lim_{\rho \to 0} \gamma = 2 \tag{18}$$

$$\lim_{\rho \to 1} \gamma = \infty \tag{19}$$

$$\tag{20}$$

When ρ -> 0, almost all of all actions are "mutations", therefor the size of the first member group is going to skyrocket. There will be very few other groups that exist in the distribution. This will lead way to a power law distribution with γ -> $\infty. In the alternate case, when \rho$ -> 1. γ = 2. In this case, the distribution is more balanced and has a heavier tale. There will be more groups after a given time step of running Simon's model.

0.4 Problem 4

Part A : In Simon's Original model, the expected total number of distinct groups at time t is pt. Recall the each group is made up of elements of a particular flavor.
In class, we figured out that the number of groups containing only one element is:

$$n_1^{(g)} = \frac{N_1(t)}{pt} = \frac{1}{2-p} \tag{21}$$

Part A : Find the form of n_2 and n_3, the fraction of groups that are of size 2 and size 3. The Stochastic Difference Equation states:

$$< N_{k,t+1} - N_{k,t} = (1-\rho)((k-1)\frac{N_{k-1,t}}{t} - k\frac{N_{k,t}}{t}) > \tag{22}$$

We solve directly for n_k for any $k > 1$ assuming that the distribution stabilizes. Thus, (23)

$$n_k = \frac{(1-\rho)(k-1)(n_{k-1})}{1+(1-\rho)k} \tag{24}$$

Given n_1, after some basic substitution, we get:

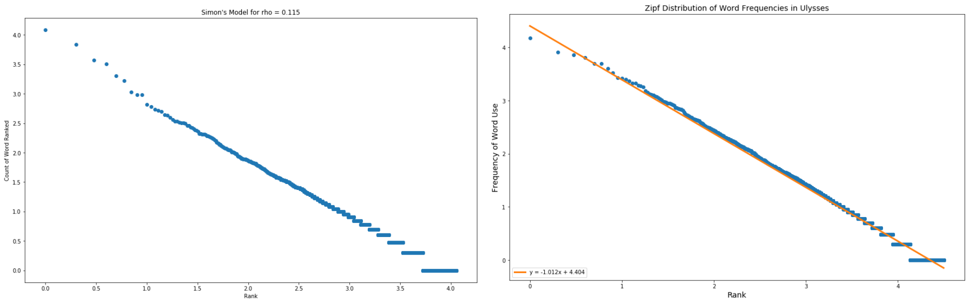
$$n_1 = \frac{1}{2-\rho} \tag{25}$$

$$n_2 = \frac{1-\rho}{3-2(\rho))(2-\rho)} \tag{26}$$

$$n_3 = \frac{2(1-\rho)^2}{(4-3(\rho))(3-2(\rho))(2-\rho)} \tag{27}$$

Part B : How Does the power law distribution of word rank frequencies in the text corpus compared to a distribution from Simon's Model with $\rho = 0.115$?

We see that the estimation for the innovation rate $\rho_{est} = 0.115$ is quite good, the model generated data follows the same Power Law distribution as the frequency groups from the Ulysses text corpus.



(a) **(b)**

Figure 2: Zipfs Law in Ulysses Text Corpus and Power Law Size Distribution generated from Simon s Model with ρ = 0.115.

Part C : Compare the computed and calculated composition of group size 1,2 and 3. The theoretical values of n(1-3) hold up remarkably well against the empirical values found in the Ulysses Corpus.

$$n_1 = \frac{1}{2-\rho} = 0.5305 \tag{28}$$

$$\text{Computed Value} = 0.5293 \tag{29}$$

$$n_2 = \frac{1-\rho}{3-2(\rho))(2-\rho)} = 0.1556 \tag{30}$$

$$\text{Computed Value} = 0.1739 \tag{31}$$

$$n_3 = \frac{2(1-\rho)^2}{(4-3(\rho))(3-2(\rho))(2-\rho) = 0.08208} \tag{32}$$

$$\text{computed value} = 0.0815 \tag{33}$$

0.5 Problem 5

Consider a set of N samples, randomly chosen according to the probability distribution $P_k = ck^{-\gamma}$ where $2 \le \gamma \le 3$. K is discrete.
(a) Estimate mink_{max}, the approximate minimum of the largest sample in the network, finding how it depends on N.

$$\sum_{k=minKmax}^{\infty} cK^{-\gamma} = \frac{1}{N} \tag{34}$$
$$\sum_{k=minKmax}^{\infty} K^{-\gamma} = \frac{1}{NC} \tag{35}$$
$$Q_1^{-\gamma} + Q_2^{-\gamma} + Q_3^{-\gamma} + ... + Q_n^{-\gamma} = \frac{1}{CN} \tag{36}$$
$$N(Q^{-\gamma} = \frac{1}{NC} \tag{37}$$
$$minK_{max} = \frac{N^{\frac{-1}{\gamma}}}{C} \tag{38}$$

(b) Find <k_{max}>

$$E[x] = xF(x) \tag{39}$$
$$= N * \frac{N^{\frac{-1}{\gamma}}}{C} \tag{40}$$
$$= \frac{N^{\frac{1-\gamma}{\gamma}}}{C} \tag{41}$$

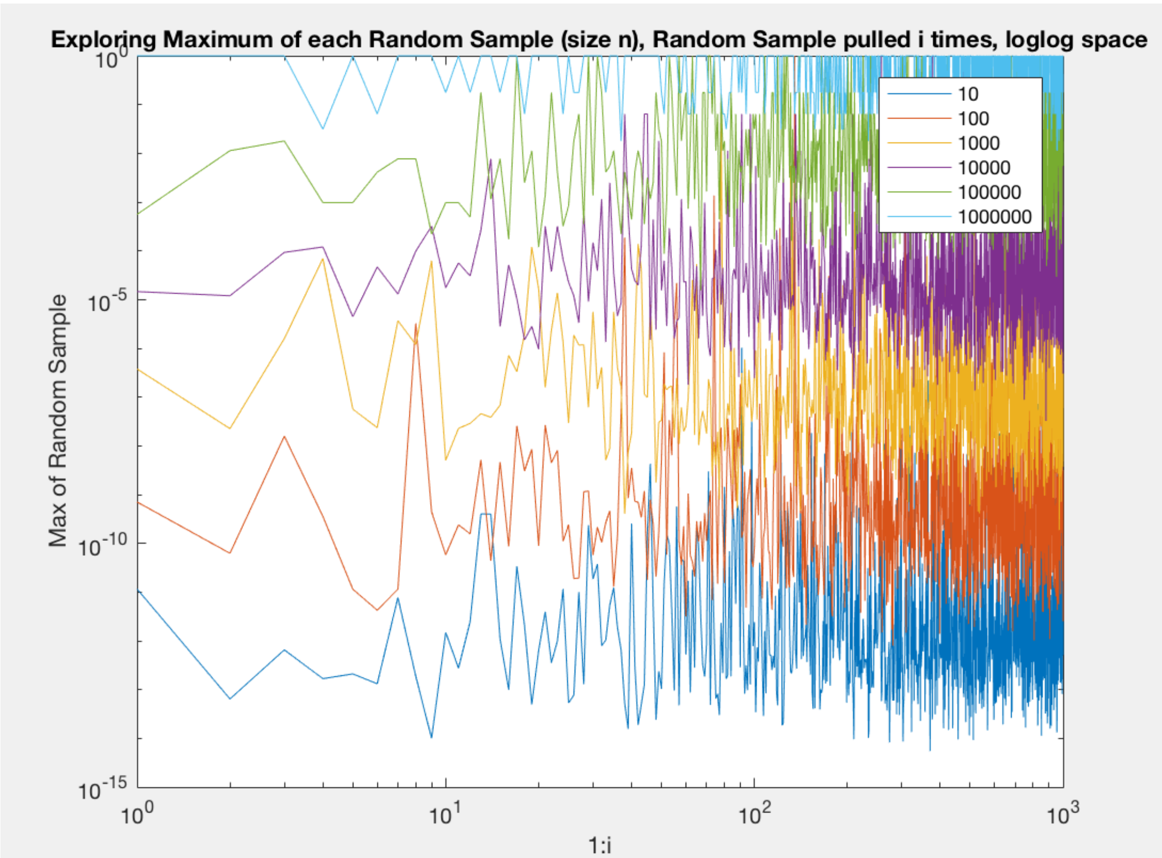
0.6 Problem 6

Sampling probability distributions.

(a) Plot k_{max} for i = 1,2,..,n where i is the sample number. The plot below compares the distribution of kMax produced from different sample sizes.

To save computation time, Sophie and I plotted this once on one computer after ensuring our methods were correct. This is the beautiful product.

The graph shows us that each sample size produces a different maximum K regime. The smaller the sample size, the higher the probability.



kMax for different sample sizes

(b) For each set, find the max value, then find the average max value for each N. Plot <Kmax> as a function of N and calculate the scaling using least squares.

Does your scaling match up?

The calculated scaling exponent is 1.7. This is surprising as the lower cutoff for γ is 2. It is possible that the data sampled is not completely representative of the total generated distribution.