CSYS 300 PoCS Assignment 4

Thomas Waters

September 28th 2018

Code is located at

https://github.com/Evelios/PoCS_Assignment_04

1 Problem 1

This problem goal was to simulate Simon's rich-gets-richer model and to see how the simulations match up to the theoretical expectations for the model. In this case I simulated the models for probabilities $p = [0.1, 0.01, 0.001]$ I performed these sumulations with 20,000 samples and repeated the simulation 100 times, combining the output results. From those output results I performed linear regrssion on them to determine the α values to compare them against the expected results. I first had to perform filtering on the tail of the distribution because the range of the distribution is highly dependent on the length of the simulation.

To correlate the results of the simulations and the predictions of the models, the following are used to correlate the values of the slope γ to the probability of mutation ρ in terms of the value of α.

$$\alpha = \frac{1}{\gamma - 1} \tag{1}$$

$$\alpha = 1 - \rho \tag{2}$$

$$1 - \rho = \frac{1}{\gamma - 1} \tag{3}$$

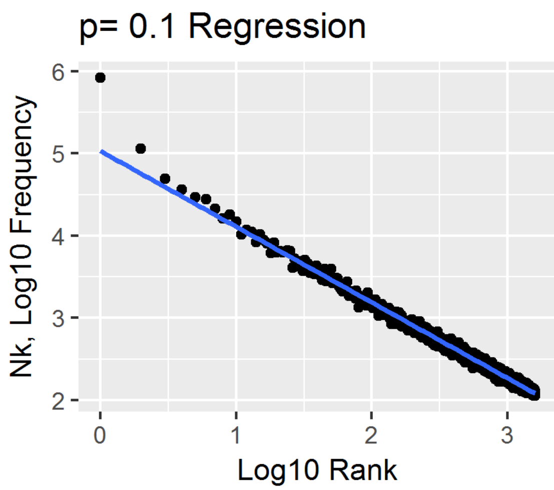
$$\gamma - 1 = \frac{1}{1 - \rho} \tag{4}$$

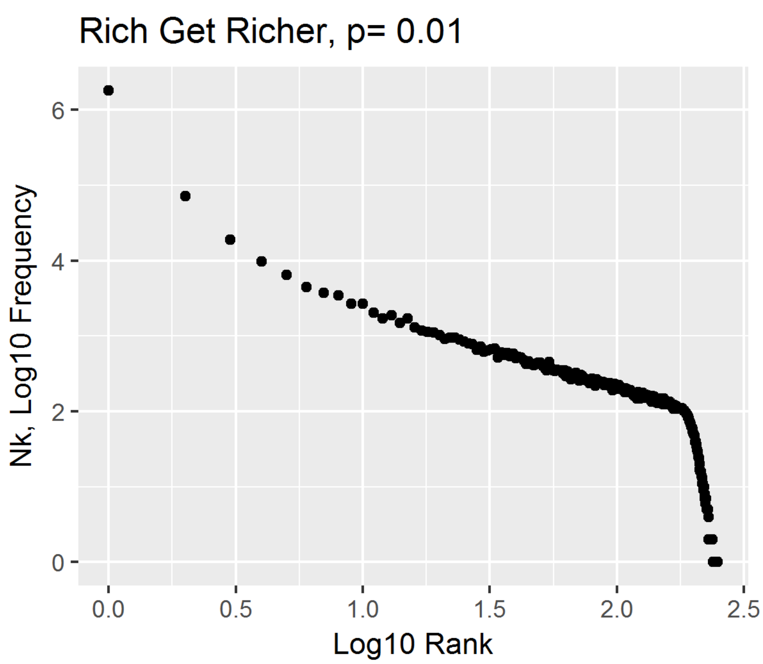
$$\gamma = \frac{1}{1-\rho} + 1 \tag{5}$$

Using these equations I can compare the values of the expected and measured values for γ for the different mutation rates.

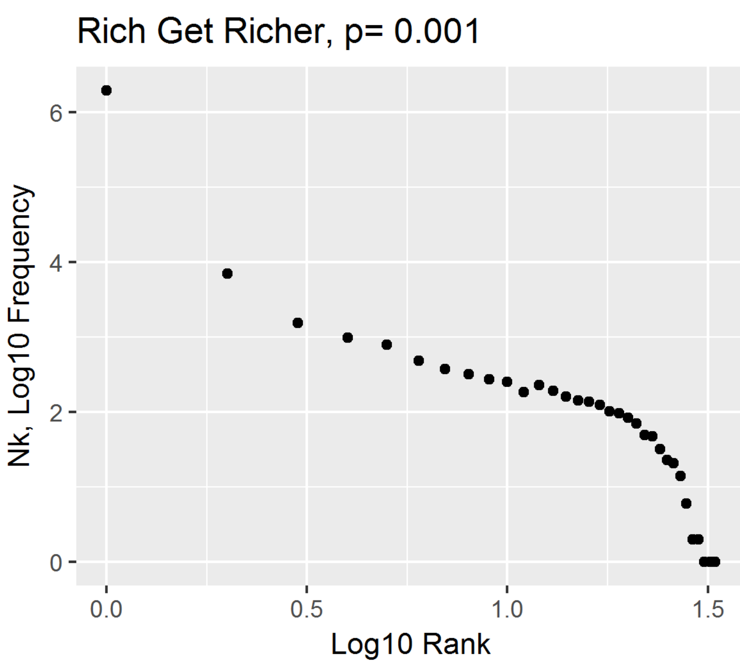
ρ	γ (expected)	γ (measured)
0.1	2.111	0.9277
0.01	2.010	1.227
0.001	2.001	2.559

My data shows no clear correlaction of any of the expected results from the model. I don't have a good answer for that, but it shows no clear trend for $\gamma \to 2$ as $\rho \to 0$.

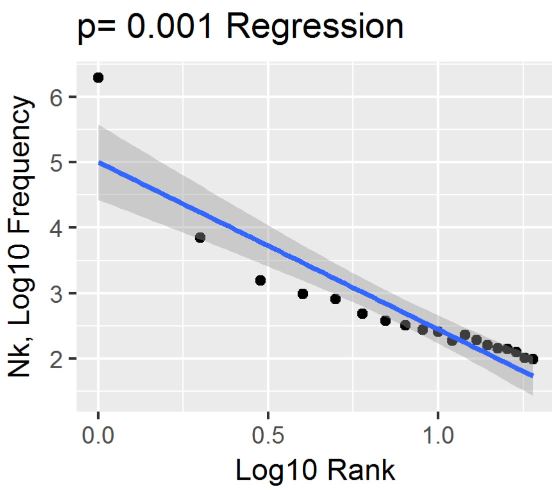

p= 0.1 Regression
Nk, Log10 Frequency
6
5
4
3
2
0
1
2
3
Log10 Rank


Rich Get Richer, p= 0.01
Nk, Log10 Frequency
6
4
2
0
0.0
0.5
1.0
1.5
2.0
2.5
Log10 Rank


p= 0.01 Regression
Nk, Log10 Frequency
Log10 Rank
6
5
4
3
2
0.0
0.5
1.0
1.5
2.0


Rich Get Richer, p= 0.001
Nk, Log10 Frequency
Log10 Rank
6
4
2
0
0.0
0.5
1.0
1.5

4



2 Problem 2

Normalized number of groups inthe long time limit, n_k satisfies the following difference equation with $k \geq 2$,

$$\frac{n_k}{n_{k-1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} \tag{6}$$

For $k = 1$,

$$n_1 = \rho - (1-\rho)n_1 \tag{7}$$

The following steps are to derive the exact solution for n_k. The following derivation will use both the gamma function and the beta function with the following definitions,

$$\Gamma(x) = (k-1)! \tag{8}$$

$$\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{9}$$

So now simplifying for n_k

$$\frac{n_k}{n_{k-1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} \tag{10}$$

$$n_k = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} n_{k-1} \tag{11}$$

$$= \left[\frac{(k-1)(1-\rho)}{1+(1-\rho)k}\right] \left[\frac{(k-2)(1-\rho)}{1+(1-\rho)(k-2)}\right] n_{k-2} \tag{12}$$

$$= \left[\frac{(k-1)(1-\rho)}{1+(1-\rho)k}\right] \left[\frac{(k-2)(1-\rho)}{1+(1-\rho)(k-2)}\right] ...n_2 n_1 \tag{13}$$

$$= \left[\frac{(k-1)(1-\rho)}{1+(1-\rho)k}\right] \left[\frac{(k-2)(1-\rho)}{1+(1-\rho)(k-1)}\right] ... \left[\frac{(2-1)(1-\rho)}{1+(1-\rho)(2)}\right] n_1 \tag{14}$$

$$= \frac{(1-\rho)^{k-1}}{(1-\rho)^{k-1}} \left[\frac{(k-1)}{\frac{1}{1-\rho}+k}\right] \left[\frac{(k-2)}{\frac{1}{1-\rho}+(k-1)}\right] ... \left[\frac{(2-1)}{\frac{1}{1-\rho}+(2)}\right] n_1 \tag{15}$$

$$= \Gamma(k) \left[\frac{1}{\frac{1}{1-\rho}+k}\right] \left[\frac{1}{\frac{1}{1-\rho}+(k-1)}\right] ... \left[\frac{1}{\frac{1}{1-\rho}+(2)}\right] n_1 \tag{16}$$

$$n_k = \frac{\Gamma(k)\,\Gamma(1-\rho+1)}{\Gamma(1-\rho+1+k)} n_1 \tag{17}$$

$$n_k = \frac{\Gamma(k)\,\Gamma(\rho)}{\Gamma(\rho+k)} n_1 \tag{18}$$

$$n_k = \beta(k,\rho)\, n_1 \tag{19}$$

3 Problem 3

From the lectures we arrived at the following,

$$\gamma = 1 + \frac{1}{1-\rho} \tag{20}$$

For $\rho \to 0$,

$$\gamma = \lim_{\rho \to 0} 1 + \frac{1}{1-\rho} = 2 \tag{21}$$

For $\rho \to 1$,

$$\gamma = \lim_{\rho \to 1} 1 + \frac{1}{1-\rho} = \infty \tag{22}$$

For the case of $\gamma \to 2$, there is no mutation rate and the entire population is in the same groupn, this would cause the slope of the distribution to become infinite because there would be nothing to scale against. In the case where $\gamma \to \infty$, we are guarenteed to make a new group every time. This would cause the frequency of all the groups to be one, causing the power law slope to go to zero because each group occurs as often as every other one.

4 Problem 4

$$n_1 = \frac{\rho}{2-\rho} \tag{23}$$

$$n_k^{(g)} = \frac{1}{\rho t} N_{kt} = \frac{1}{\rho} n_k \tag{24}$$

Solving for $n_2^{(g)}$,

$$\frac{n_k}{n_{k-1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} \tag{25}$$

$$n_k = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} n_{k-1} \tag{26}$$

$$n_2 = \frac{(2-1)(1-\rho)}{1+(1-\rho)2} n_1 \tag{27}$$

$$n_2 = \frac{(1-\rho)}{1+2(1-\rho)} \left[\frac{\rho}{2-\rho}\right] \tag{28}$$

$$n_2^{(g)} = \frac{1}{\rho} n_2 = \frac{1}{\rho} \left[\frac{1-\rho}{3-2\rho}\right] \left[\frac{\rho}{2-\rho}\right] \tag{29}$$

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

Now for $n_3^{(g)}$,

$$n_k = \frac{(k-1)(1-\rho)}{1+(1-\rho)k} n_{k-1} \tag{31}$$

$$n_3 = \frac{(3-1)(1-\rho)}{1+(1-\rho)3} n_2 \tag{32}$$

$$n_3 = \frac{2(1-\rho)}{4-3\rho}\left[\frac{(1-\rho)}{3-2\rho}\right]\left[\frac{\rho}{2-\rho}\right] \tag{33}$$

$$n_3^{(g)} = \frac{1}{\rho t}N_{3t} = \frac{1}{\rho}n_3 \tag{34}$$

$$n_3^{(g)} = \frac{1}{\rho}\left[\frac{2(1-\rho)}{4-3\rho}\right]\left[\frac{(1-\rho)}{3-2\rho}\right]\left[\frac{\rho}{2-\rho}\right] \tag{35}$$

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

5 Problem 5

Here I am using the Hurwitz Zeta Function,

$$\zeta\left(s,q\right) = \sum_{n=0}^{\infty}\frac{1}{(q+n)^s}$$

Trying to estimate the minimum of the largest sample in the network.

$$P_k = ck^{-\gamma} \tag{36}$$

$$\sum_{k=min\,k_{max}}^{\infty} P_k = \frac{1}{N} \tag{37}$$

$$\sum_{k=min\,k_{max}}^{\infty} ck^{-\gamma} = \frac{1}{N} \tag{38}$$

$$c\sum_{k=min\,k_{max}}^{\infty}\frac{1}{k^{\gamma}} = \frac{1}{N} \tag{39}$$

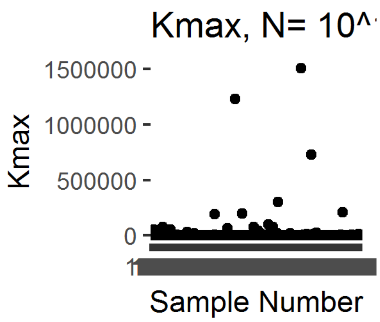
$$c\sum_{k=0}^{\infty}\frac{1}{((min\,k_{max})+k)^{\gamma}} = \frac{1}{N} \tag{40}$$

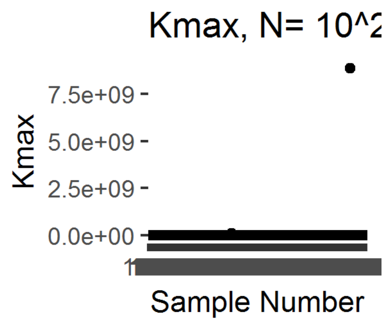
$$c\,\zeta\left(\gamma, min\,k_{max}\right) = \frac{1}{N} \tag{41}$$

6 Problem 6

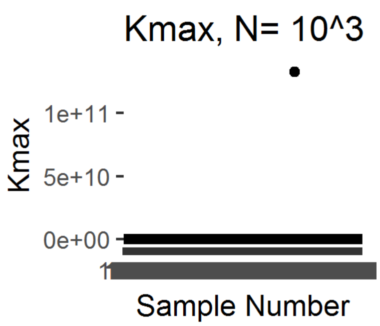
Unfortunately on this one, the plots out of R on such small scales caused some problems for both the titles and the labels.

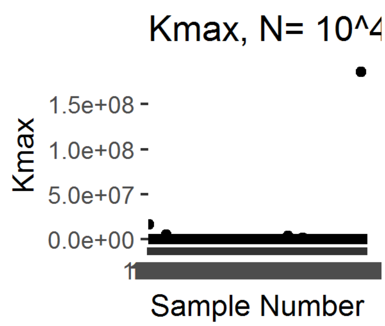
Showing k_max for $N = 10^1$ (left) and for $N = 10^2$ (right).



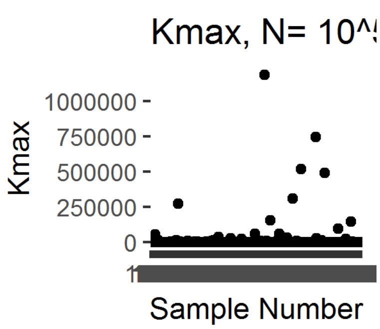


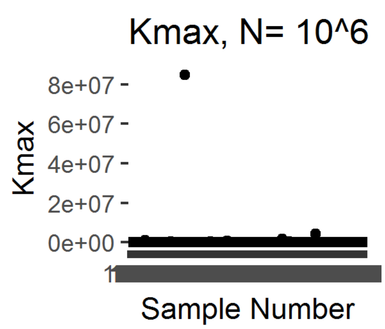
Showing k_max for $N = 10^3$ (left) and for $N = 10^4$ (right).





Showing k_max for $N = 10^5$ (left) and for $N = 10^6$ (right).





Unfortunately for this homework it seems that again, something is off. Here, unfortunately because of the nature of power law distributions I can't tell if it is the nature of the distribution but I assume it is probably a result of my algorithm. Here, I would expect from the averaging from 1000 simulation that any of the erratic behavior from the output of the power law distributions would be mitigated. Here, unfortunately, from my graph of the average for k_max, I don't see any relationship in the average value of k_max in terms of the sample size.