

# PoCS Assignment 4

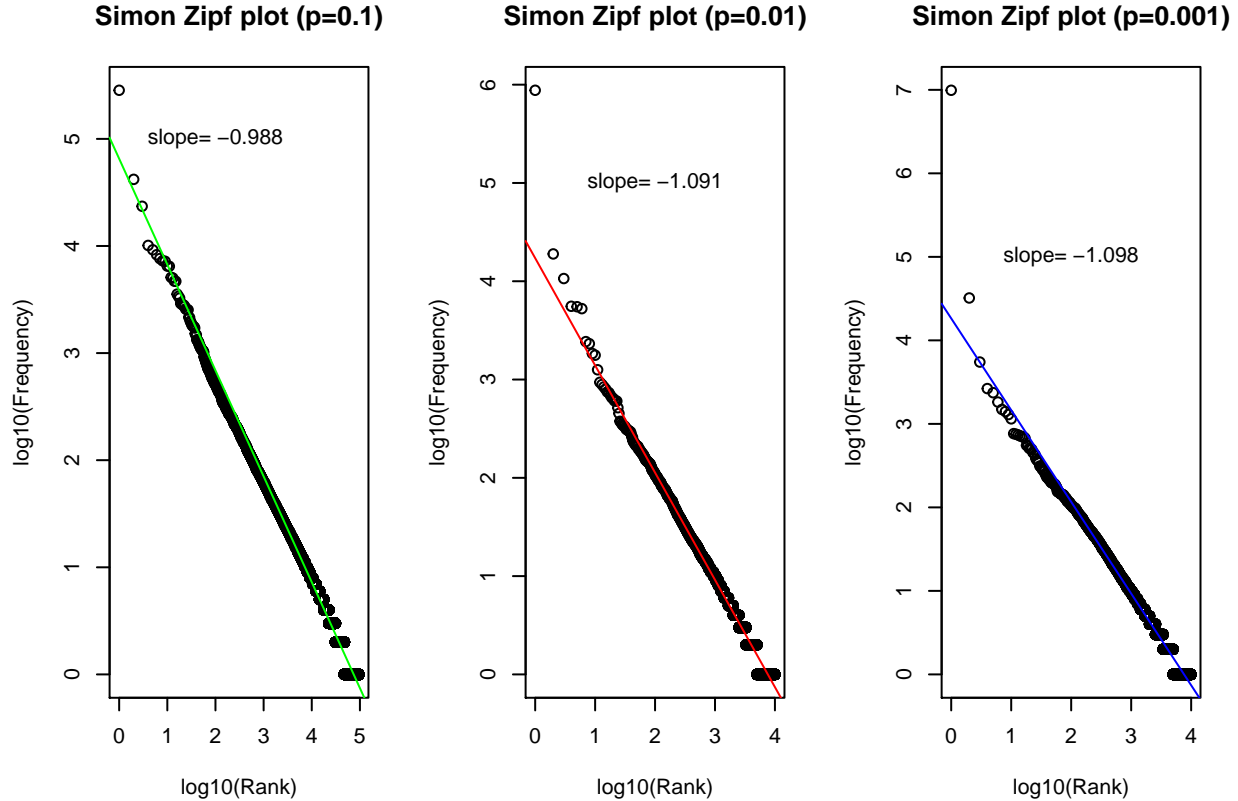
*P. Alexander Burnham*

*September 25, 2018*

PoCS Assignment github repo: <https://github.com/alexburn17/BurnhamPoCS>

Worked with: Yu Han, Edison & Kewang

## Problem 1:



The slopes would be closer to the expected values of 0.9, 0.99 and 0.999 if they were repeated more times and the distributions were trimmed before the regressions were conducted, however, they are very close and we see an increase in the value as  $\rho$  decreases.

## Problem 2:

a)

Given:  $\frac{n_k}{n_{k+1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k}$ ,  $z = (1-\rho)$ ,  $n_1 = \frac{\rho}{2-\rho}$

Gamma Function says:  $\Gamma(k) = (k-1)!$

From Video:

Step 1: write difference equation for an idea of what is going on

$$n_k = \left[ \frac{(k-1)(1-\rho)}{1+(1-\rho)k} \right]$$

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

$$n_{k-2} = \left[ \frac{(k-3)(1-\rho)}{1+(1-\rho)(k-2)} \right]$$

etc.

Step 2: Denominator sub in  $z$  for  $1-\rho$

$$[(1+zk)(1+zk-1)(1+z)(k-2)\dots(1+2z)(1+z)] \frac{1}{1+z}$$

simplify...

$$z^k \left(\frac{1}{z} + k\right) \left(\frac{1}{z} + k - 1\right) \dots \left(\frac{1}{z} + 1\right) \left(\frac{1}{1+z}\right)$$

$$Denominator = \frac{z^k \Gamma(\frac{1}{z} + k - 1)}{\Gamma(\frac{1}{z} + 1)} * \frac{1}{1+z}$$

Step 3: Numerator sub in  $z$  for  $1-\rho$  and  $\Gamma(k)$  for  $(k-1)!$

$$= (k-1)(1-\rho)(k-2)(1-\rho)(k-3)(1-\rho)\dots(1-\rho)$$

simplify...

$$= (1-\rho)^{(k-1)}(k-1)!$$

$$= z^{(k-1)}\Gamma(k)$$

Step 4: put numerator of denominator, multiply by  $n_1$  and simplify

$$= \frac{z^{(k-1)}\Gamma(k)}{\frac{z^k \Gamma(\frac{1}{z} + k - 1)}{\Gamma(\frac{1}{z} + 1)} * \frac{1}{1+z}} * \frac{\rho}{2-\rho}$$

cross out  $z^k$  and move  $z^{-1}$  to denominator sub in:  $\rho = 1-z$

$$= \frac{\Gamma(k)}{z \frac{\Gamma(\frac{1}{z} + k - 1)}{\Gamma(\frac{1}{z} + 1)} * \frac{1}{1+z}} * \frac{(1-z)}{2-(1-z)}$$

$$= \frac{\Gamma(k)}{z} (1+z) \frac{\Gamma(\frac{1}{z} + 1)}{\Gamma(\frac{1}{z} + k - 1)} * \frac{(1-z)}{2-(1-z)}$$

simplify and sub  $\rho$  back in:

$$= \frac{\Gamma(k)\Gamma(\frac{1}{z} + 1)}{\Gamma(\frac{1}{z} + k - 1)} * \frac{\rho}{1-\rho}$$

Use beta function to derive:

$$\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \frac{(x-1)!(y-1)!}{(x+y-1)!}$$

$$n_k = \frac{\rho}{1-\rho} * \beta(k, \frac{1}{1-\rho} + 1)$$

b)

$$n_k = \frac{\rho}{1-\rho} * \beta(k, \frac{1}{1-\rho} + 1)$$

use the fact that  $\beta(k, \gamma) = k^{-\gamma}$

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

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

$$\gamma = \frac{2-\rho}{1-\rho}$$

**Problem 3:**

$$\lim_{\gamma \rightarrow 0} \gamma = \frac{2-\rho}{1-\rho}$$

As  $\rho$  approaches 0,  $\gamma$  goes to 2.

$$\lim_{\gamma \rightarrow 1} \gamma = \frac{2-\rho}{1-\rho}$$

As  $\rho$  approaches 1,  $\gamma$  goes to  $\infty$ .

**Problem 4:**

a)

Given:

$$n_1^{(g)} = \frac{N_1 t}{\rho t} = \frac{1}{2-\rho} \text{ \& \ } \frac{n_k}{n_{k+1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k}$$

Step 1: Set up to find  $n_1^{(g)}$  and plug in  $k = 2$  and multiply by  $n_1^{(g)}$

$$\frac{n_2^{(g)}}{n_1^{(g)}} = \frac{\frac{N_2}{\rho}}{\frac{N_1}{\rho}} = \frac{N_2}{N_1} = \frac{(2-1)(1-\rho)}{1+(1-\rho)2} * n_1^{(g)}$$

$$n_2^{(g)} = \frac{(2-1)(1-\rho)}{1+(1-\rho)2} * \frac{1}{2-\rho}$$

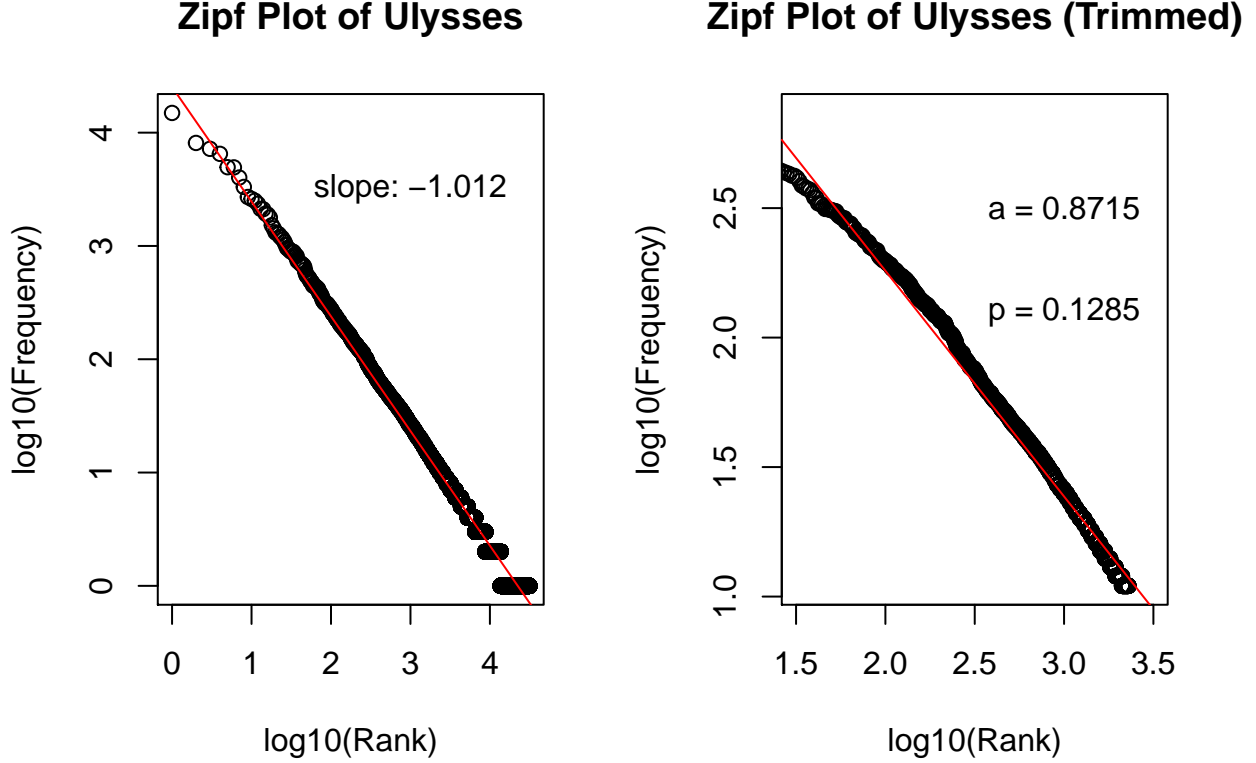
$$n_2^{(g)} = \frac{(1-\rho)}{3-2\rho} * \frac{1}{2-\rho}$$

Step 2: find  $n_3^{(g)}$  in the same way

$$\frac{n_3^{(g)}}{n_2^{(g)}} = \frac{\frac{N_3}{\rho}}{\frac{N_2}{\rho}} = \frac{N_3}{N_2} = \frac{(3-1)(1-\rho)}{1+(1-\rho)3} * n_2^{(g)}$$

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

b)



The estimate of  $\rho$  was 0.1285 based on a trimmed distribution to increase accuracy.

Plugging this value in for  $n_1^{(g)}, n_2^{(g)}, n_3^{(g)}$  gives the theoretical estimates. The actual values are found by taking the number of groups with frequencies of 1, 2 and 3 in turn over the total number of groups. This gives the following values:

Actual (measured from data set):

- $n_1^{(g)} = 0.565$
- $n_2^{(g)} = 0.156$
- $n_3^{(g)} = 0.071$

Theoretical (using estimate for  $\rho$ ):

- $n_1^{(g)} = 0.534$
- $n_2^{(g)} = 0.1697$
- $n_3^{(g)} = 0.082$

These values are all very similar meaning that our estimate matches our theoretical values well.

c)

**Problem 5:**

a)

Given:  $P_k = ck^{-\gamma}$

Step 1: find c

$$\int_{k_{min}}^{\infty} ck^{-\gamma} = 1$$

$$c \frac{1}{-\gamma + 1} k^{(-\gamma+1)} \Big|_{k_{min}}^{\infty} = 1$$

$$c \frac{1}{\gamma - 1} k_{min}^{(\gamma-1)} = 1$$

find c:

$$c = (\gamma - 1) k_{min}^{(\gamma-1)}$$

Step 2: find  $k_{max}$

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

$$\int_{k_{max}}^{\infty} ck^{-\gamma} = \frac{1}{N}$$

$$\int_{k_{max}}^{\infty} k^{-\gamma} = \frac{1}{cN}$$

Find antiderivative

$$\frac{1}{\gamma - 1} k_{max}^{-\gamma+1} = \frac{1}{cN}$$

multiply both sides by  $(\gamma - 1)$

$$k_{max}^{-\gamma+1} = \frac{\gamma - 1}{cN}$$

$$k_{max} = \left( \frac{\gamma - 1}{c} \right)^{\left( \frac{1}{1-\gamma} \right)} \frac{1}{N^{\left( \frac{1}{1-\gamma} \right)}}$$

Step 3: sub in C and simplify:

$$k_{max} = \left( \frac{\gamma - 1}{(\gamma - 1) k_{min}^{(\gamma-1)}} \right)^{\left( \frac{1}{1-\gamma} \right)} \frac{1}{N^{\left( \frac{1}{1-\gamma} \right)}}$$

cancel terms:

$$k_{max} = k_{min} \frac{1}{N^{(\frac{1}{1-\gamma})}}$$

Answer:

$$k_{max} = k_{min} N^{(\frac{1}{\gamma-1})}$$

b)

Step 1:

$$= \int_{k_{max}}^{\infty} c k^{-\gamma+1} dk$$

antiderivative:

$$= c \frac{1}{2-\gamma} k^{2-\gamma} \Big|_{k_{max}}^{\infty}$$

$$= c \frac{1}{\gamma-2} k_{max}^{2-\gamma}$$

plug in  $c$  and  $k_{max}$

$$= (\gamma-1) k_{min}^{(\gamma-1)} \frac{1}{\gamma-2} \left( k_{min} \frac{1}{N^{(\frac{1}{1-\gamma})}} \right)^{2-\gamma}$$

Step 2: simplify

$$\frac{\gamma-1}{\gamma-2} k_{min}^{(\gamma-1)} k_{min}^{(2-\gamma)} \frac{1}{N^{(\frac{2-\gamma}{1-\gamma})}}$$

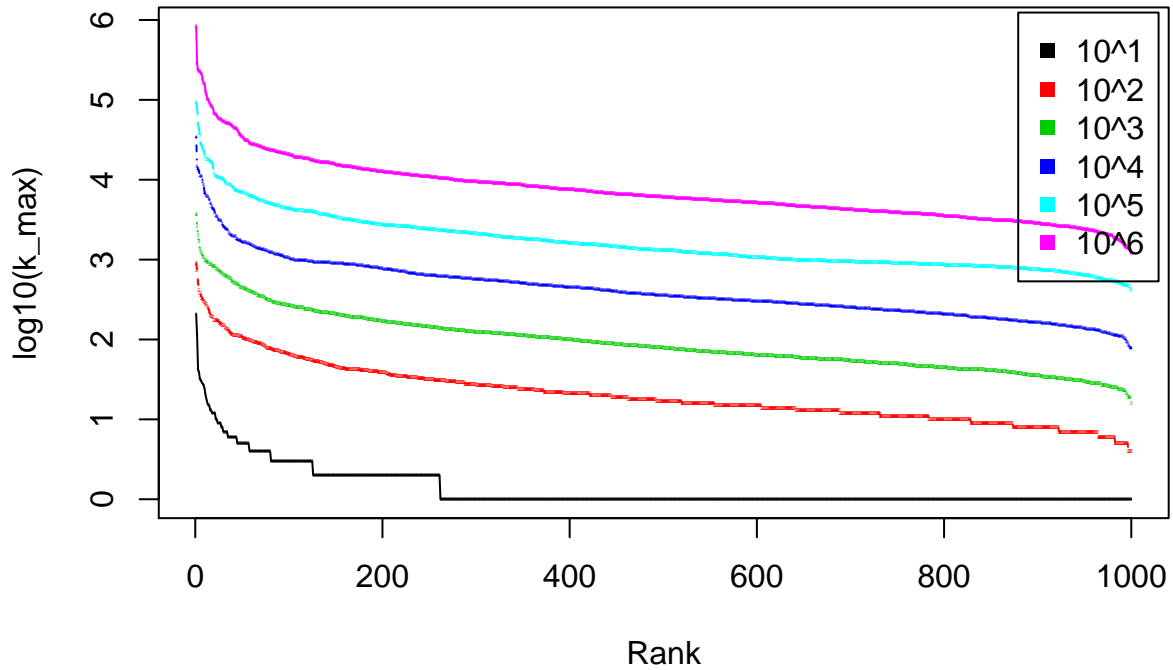
$$k_{min} \frac{\gamma-1}{\gamma-2} * \frac{1}{N^{(\frac{2-\gamma}{1-\gamma})}}$$

$$k_{min} \frac{\gamma-1}{\gamma-2} * N^{(\frac{2-\gamma}{\gamma-1})}$$

**Problem 6:**

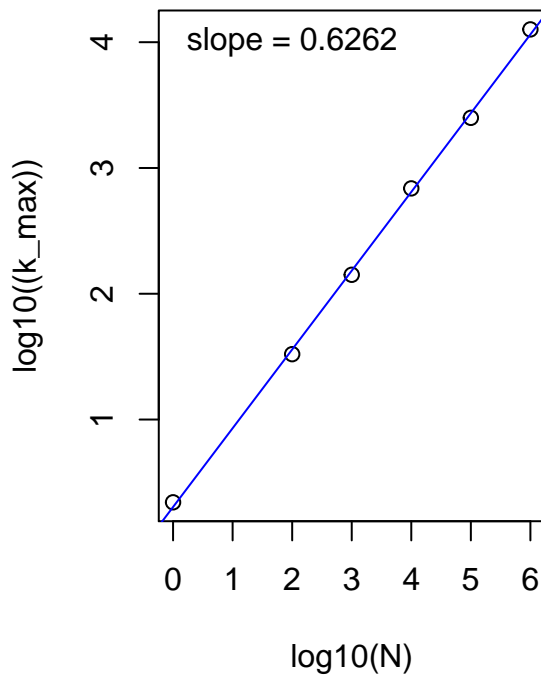
a)

## K\_max by N

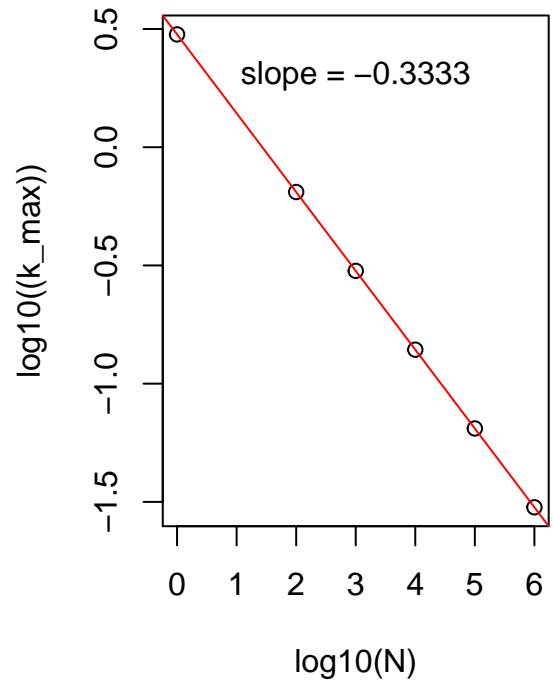


b)

## K\_max estimate by N



## K\_max theoretical by N



My theoretical value and empirical values do not match up. The one I calculated by sampling a powerlaw distribution has a positive slope of around 0.6. This makes sense as the max  $k$  should increase as the sample size increases. However, my derivation of the theoretical max  $k$  has a negative slope of -0.3. This does not make sense and leads me to believe that my model is correct and the theoretical values are incorrect.  $K_{max}$

should increase as  $N$  increases as this increase in  $N$  gives higher probability of capturing a large  $K$ . The theoretical value looks like it is flipped about the  $y$  axis as the spacing of the points is smaller.