PoCS What's The Story?

**Principles of Complex Systems, Vols. 1, 2, & 3D**
**CSYS/MATH 300, 303, & 394**
**University of Vermont, Fall 2022**
**Solutions to Assignment 07**
**Emergency B-Vord**

**Name: Krishna Kannan Srinivasan**

**Conspirators:**

---

Begin to think about projects.

See assignment 9 for instructions including details for the first presentation.

1. (3 + 3)

   You've earlier determined the theoretical scaling of the large sample of a power-law size distribution as a function of sample number.

   Let's see how well things match up with simulations.

   For $\gamma = 5/2$, generate $n = 1000$ sets each of $N = 10$, $10^2$, $10^3$, $10^4$, $10^5$, and $10^6$ samples, using $P_k = ck^{-5/2}$ with $k = 1, 2, 3, \ldots$

   How do we computationally sample from a discrete probability distribution?

   Note: We examined some of these in class. See slides on power-law size distributions.

   Perishing Monk Hint: You can use a continuum approximation to speed things up. See below.

   (a) For each value of sample size $N$, sequentially create $n$ sets of $N$ samples. For each set, determine and record the maximum value of the set's $N$ samples. (You can discard each set once you have found the maximum sample.)

   You should have $k_{\max,i}$ for $i = 1, 2, \ldots, n$ where $i$ is the set number. For each $N$, plot the $n$ values of $k_{\max,i}$ as a function of $i$.

   If you think of $n$ as time $t$, you will be plotting a kind of time series.

   These plots should give a sense of the unevenness of the maximum value of $k$, a feature of power-law size distributions.

   (b) Now find the average maximum value $\langle\rangle i k_{\max,i}$ for each $N$.

   The steps again here are:
   1. Sample $N$ times from $P_k$;
   2. Determine the maximum of the sample, $k_{\max}$;

3. Repeat steps 1 and 2 a total $n$ times and take the average of the $n$ values of $k_{\max}$ you have obtained.

   Plot $\langle k_{\max} \rangle$ as a function of $N$ on double logarithmic axes, and calculate the scaling using least squares. Report error estimates.

   Does your scaling match up with your theoretical estimate for $\gamma = 5/2$?

How to sample from your power law distribution (and similarly upsetting things):

We now turn our problem of randomly selecting from this distribution into randomly selecting from the uniform distribution. After playing around a little, $k = 10^6$ seems like a good upper limit for the number of samples we're talking about.

Using Matlab (or some ghastly alternative), we create a cdf for $P_k$ for $k = 1, 2, \ldots, 10^6$ and one final entry $k > 10^6$ (for which the cdf will be 1).

We generate a random number $x$ and find the value of $k$ for which the cdf is the first to meet or exceed $x$. This gives us our sample $k$ according to $P_k$ and we repeat as needed. We would use the exactly normalized $P_k = \frac{1}{\zeta(5/2)} k^{-5/2}$ where $\zeta$ is the Riemann zeta function.

Now, we can use a quick and dirty method by approximating $P_k$ with a continuous function $P(z) = (\gamma - 1) z^{-\gamma}$ for $z \geq 1$ (we have used the normalization coefficient found in assignment 1 for $a = 1$ and $b = \infty$). Writing $F(z)$ as the cdf for $P(z)$, we have $F(z) = 1 - z^{-(\gamma-1)} = 1 - z^{-3/2}$. Inverting, we obtain $z = [1 - F(z)]^{-2/3}$. We replace $F(z)$ with our random number $x$ and round the value of $z$ to finally get an estimate of $k$.

**Solution:**

(a) Plots of $k_{max}$ vs experiment for each N

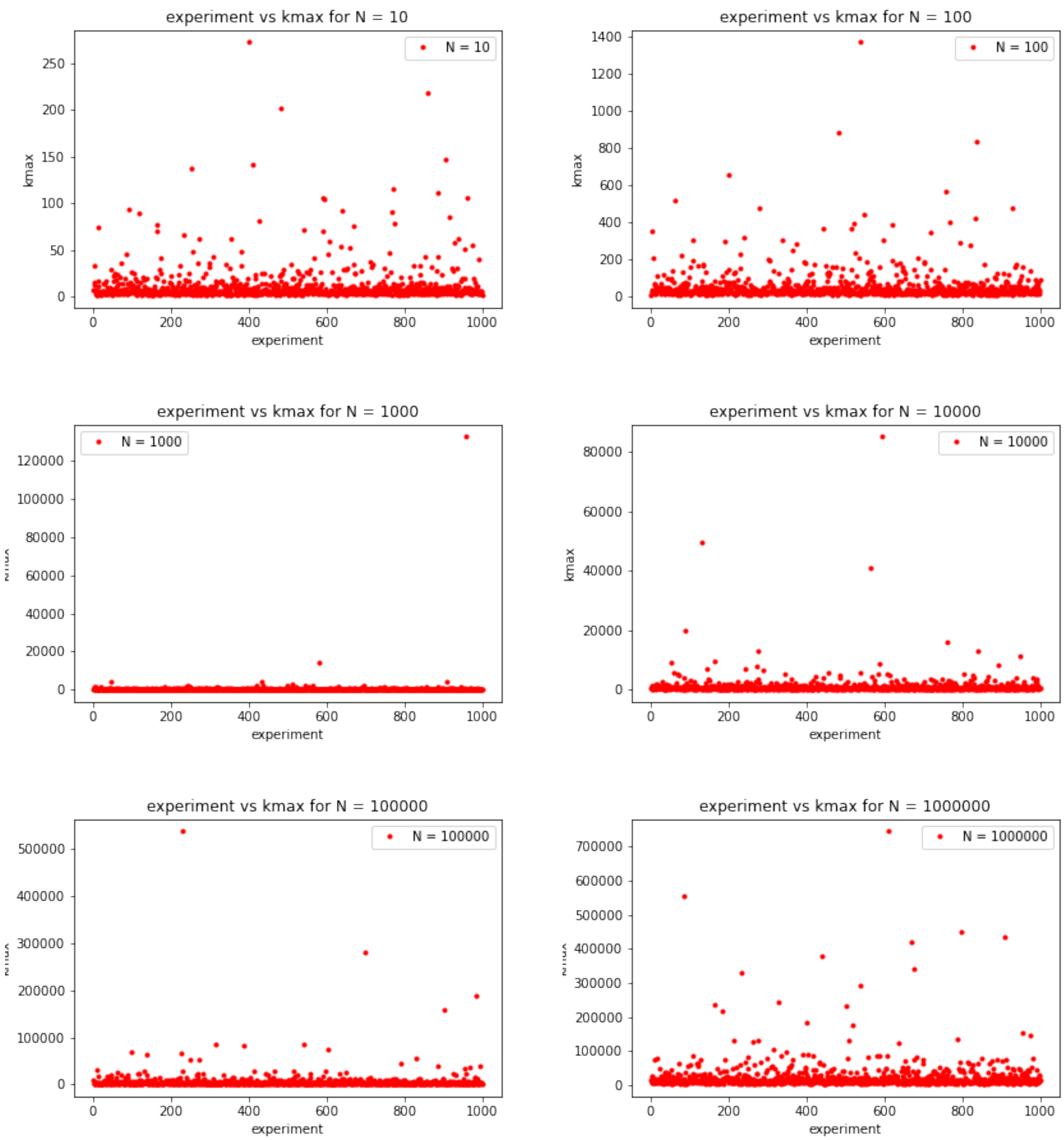



Figure 1: Kmax vs experiments(1 to 1000) for various N

(b) Plots of Average of $k_{max}$ from all experiments for each N

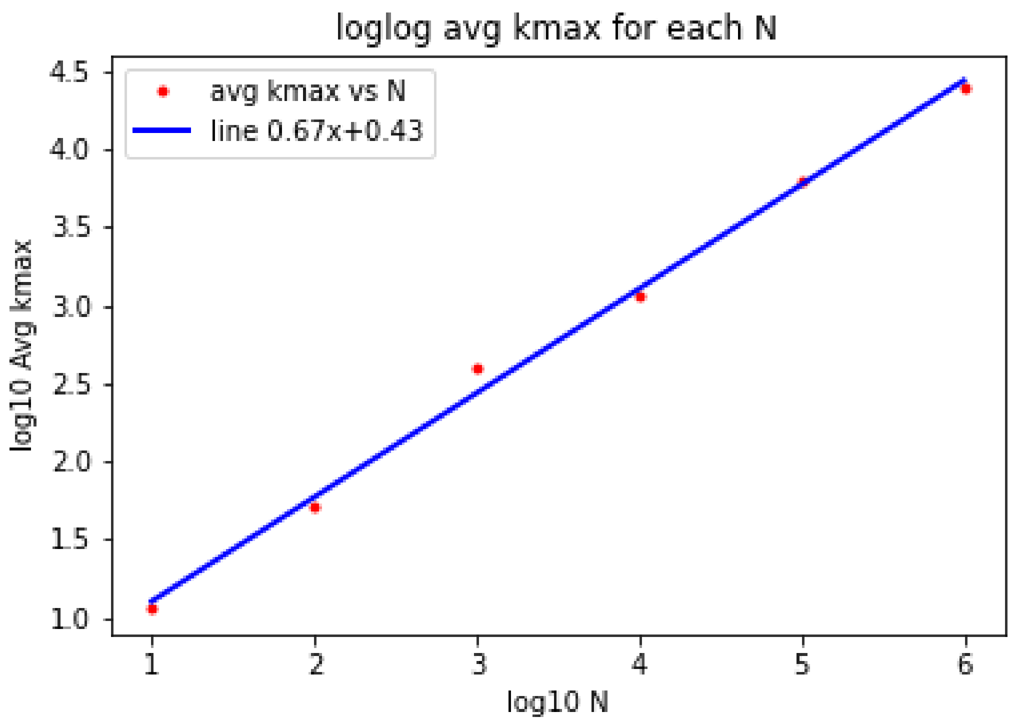



Figure 2: Average $k_{max}$ for each N

Theoretical estimate of Average of $k_{\max,i}$ in terms N:

$$Average k_{\max,i} \propto N^{\frac{-1}{\gamma-1}}$$

For $\gamma = 5/2$, the exponent we find is $2/3$. Here comparing with theoretical estimate of $\approxeq 0.67$, we get a slope of 0.67 which matches up.

Sum of least squares errors of fit: 0.035 which is very less. □

2. (3 + 3 points) *Zipfarama via Optimization:*

Complete the Mandelbrotian derivation of Zipf's law by minimizing the function

$$\Psi(p_1, p_2, \ldots, p_n) = F(p_1, p_2, \ldots, p_n) + \lambda G(p_1, p_2, \ldots, p_n)$$

where the 'cost over information' function is

$$F(p_1, p_2, \ldots, p_n) = \frac{C}{H} = \frac{\sum_{i=1}^{n} p_i \ln(i+a)}{-g \sum_{i=1}^{n} p_i \ln p_i}$$

and the constraint function is

$$G(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i - 1 \quad (= 0)$$

to find

$$p_j = e^{-1-\lambda H^2/gC}(j+a)^{-H/gC}.$$

Then use the constraint equation, $\sum_{j=1}^{n} p_j = 1$ to show that

$$p_j = (j+a)^{-\alpha}.$$

where $\alpha = H/gC$.

3 points: When finding $\lambda$, find an expression connecting $\lambda$, $g$, $C$, and $H$.

The Perishing Monks who have returned say the way is sneaky. Before collapsing, one monk mumbled something about substituting the form you find for $\ln p_i$ into $H$'s definition (but do not replace $p_i$).

Note: We have now allowed the cost factor to be $(j+a)$ rather than $(j+1)$.

**Solution:**

We have

$$\Psi(p_1, p_2, \ldots, p_n) = F(p_1, p_2, \ldots, p_n) + \lambda G(p_1, p_2, \ldots, p_n)$$

Finding the minimum of $\Psi$ entails,

$$\frac{\partial \Psi(p_1, p_2, \ldots, p_n)}{\partial p_j} = 0$$

for some fixed $p_j$.

We have

$$F = \frac{C}{H}$$

$$\frac{\partial F}{\partial p_j} = \frac{\partial C}{\partial p_j}\frac{1}{H} + \frac{\partial H}{\partial p_j}C\frac{-1}{H^2}$$

Also,

$$G(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i - 1$$

Then,

$$\frac{\partial G}{\partial p_j} = \frac{\partial(p_1 + p_2 + \ldots + p_j + \ldots + p_n)}{\partial p_j} - 0$$

$$\frac{\partial G}{\partial p_j} = 1$$

Then,

$$\frac{\partial \Psi}{\partial p_j} = \lambda + \frac{1}{H}\frac{\partial C}{\partial p_j} - \frac{C}{H^2}\frac{\partial H}{\partial p_j} \tag{1}$$

$$\frac{\partial C}{\partial p_j} = \frac{\partial}{\partial p_j}\left(\sum_{i=1}^{n} p_i \ln(i+a)\right) \tag{2}$$

$$\frac{\partial H}{\partial p_j} = \frac{\partial}{\partial p_j}\left(-g\sum_{i=1}^{n} p_i \ln(p_i)\right) \tag{3}$$

Equation 2 $\Rightarrow$

$$\frac{\partial C}{\partial p_j} = \frac{\partial}{\partial p_j}\left(p_1 \ln(1+a) + \ldots + p_j \ln(j+a) + \ldots + p_n \ln(n+a)\right)$$

$$\frac{\partial C}{\partial p_j} = \ln(j+a)$$

Equation 3 $\Rightarrow$

$$\frac{\partial H}{\partial p_j} = -g\frac{\partial}{\partial p_j}\left(p_j \ln(p_j)\right)$$

$$\frac{\partial H}{\partial p_j} = -g\left(1+\ln(p_j)\right)$$

Then, Equation 1 becomes,

$$\frac{\partial \Psi}{\partial p_j} = \lambda + \frac{1}{H}\ln(j+a) + \frac{Cg}{H^2}\left[1+\ln(p_j)\right]$$

$$\frac{Cg}{H^2}\left[1+\ln(p_j)\right] + \frac{1}{H}\ln(j+a) + \lambda = 0$$

$$\frac{Cg\left(1+\ln(p_j)\right) + H\ln(j+a) + \lambda H^2}{H^2} = 0$$

$$Cg\left(1+\ln(p_j)\right) + H\ln(j+a) + \lambda H^2 = 0$$

with $H \neq 0$,

$$Cg\left(1+\ln(p_j)\right) + H\ln(j+a) + \lambda H^2 = 0$$

Rearranging,

$$\ln(p_j) = \frac{-\lambda H^2 - H\ln(j+a)}{Cg} - 1 \tag{4}$$

Solving for $p_j$,

$$p_j = e^{\frac{-\lambda H^2}{Cg} - \frac{H\ln(j+a)}{Cg} - 1}$$

$$p_j = e^{\frac{-\lambda H^2}{Cg} - \ln(j+a)^{\frac{-H}{Cg}} - 1}$$

$$p_j = e^{\frac{-\lambda H^2}{Cg}-1}(j+a)^{\frac{-H}{Cg}} \tag{5}$$

Comparing with $(j+a)^{-\alpha}$,

$$p_j \propto (j+a)^{\frac{-H}{Cg}}$$

with $\alpha = \frac{H}{gC}$.

Now to find $\lambda$ in terms of g, C and H:

We have,

$$H = -g\sum_{i=1}^{n} p_i \ln p_i$$

Equation 4 gives us,

$$\ln(p_i) = \frac{-\lambda H^2 - H\ln(i+a)}{Cg} - 1$$

Substituting in expression for H,

$$H = -g\sum_{i=1}^{n} p_i \left(\frac{-\lambda H^2 - H\ln(i+a)}{Cg} - 1\right)$$

$$H = -g\left(\frac{-\lambda H^2}{Cg} - 1\right)\sum_{i=1}^{n} p_i + g\left(\frac{H}{Cg}\right)\sum_{i=1}^{n} p_i \ln(i+a)$$

Using constraint equation, $\sum_{i=1}^{n} p_i = 1$, and $C = \sum_{i=1}^{n} p_i \ln(i+a)$,

$$H = -g\left(\frac{-\lambda H^2}{Cg} - 1\right)1 + g\left(\frac{H}{Cg}\right)C$$

Rearranging and simplifying we have

$$\frac{\lambda H^2}{C} = -g$$

$$\lambda = \frac{-gC}{H^2}$$

Substituting in equation 5,

$$p_j = e^{\frac{-\frac{-gC}{H^2}H^2}{Cg}-1}(j+a)^{\frac{-H}{Cg}}$$

$$p_j = e^{1-1}(j+a)^{\frac{-H}{Cg}}$$

$$p_j = (j + a)^{\frac{-H}{Cg}}$$

Hence shown.

$\square$

3. (3 + 3) Carrying on from the previous problem:

   (a) For $n \to \infty$, use some computation tool (e.g., Matlab, an abacus, but not a clever friend who's really into computers) to determine that $\alpha \simeq 1.73$ for $a = 1$. (Recall: we expect $\alpha < 1$ for $\gamma > 2$)

   (b) For finite $n$, find an approximate estimate of $a$ in terms of $n$ that yields $\alpha = 1$.

   (Hint: use an integral approximation for the relevant sum.)

   What happens to $a$ as $n \to \infty$?

**Solution:**
(a) We have,

$$p_j = (j + a)^{-\alpha}$$

with $\alpha = \frac{-H}{Cg}$

For

(a) a $= 1$

(b) $n \to \infty$

What is $\alpha$?

We know that

$$\sum_{i=1}^{n} p_i = 1$$

$$\sum_{i=1}^{\infty} (j + 1)^{-\alpha} = 1$$

if $k = j + 1$,

$$\sum_{i=1}^{\infty} (j + 1)^{-\alpha} = \sum_{k=1}^{\infty} k^{-\alpha} - 1$$

Then,

$$\sum_{k=1}^{\infty} k^{-\alpha} - 1 = 1$$

$$\sum_{k=1}^{\infty} k^{-\alpha} = 2$$

We know that, $\zeta(\alpha) = \sum_{x=1}^{\infty} x^{-\alpha}$
Then,

$$\zeta(\alpha) = 2$$

Finding the roots of the equation $\zeta(\alpha) - 2 = 0$,

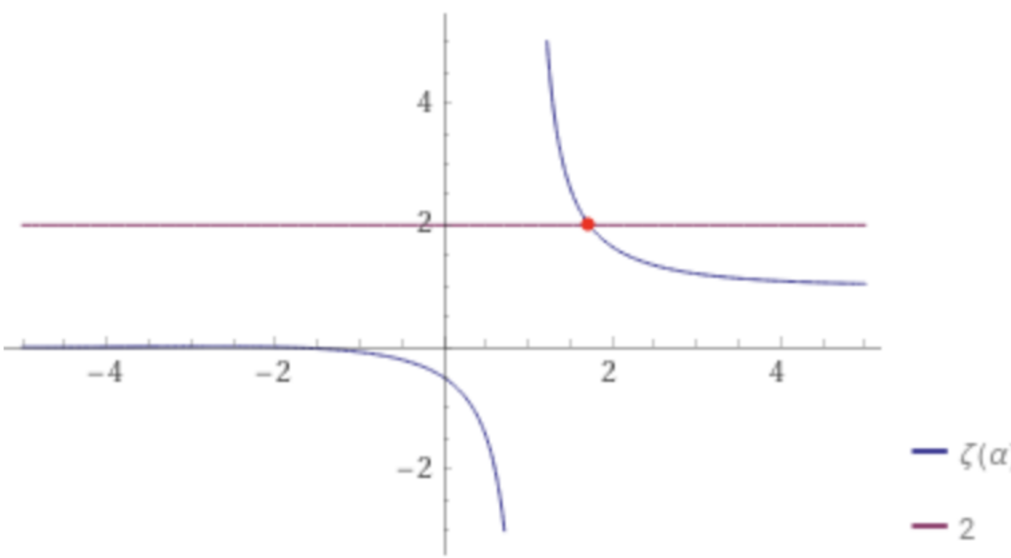
We find $\alpha \simeq 1.73$

From wolfram alpha, we see

Input

$\zeta(\alpha) = 2$

Plot




Numerical solution

$\alpha \approx 1.72864723899818\ldots$

(b) We have

$$\sum_{i=1}^{n} p_i = 1$$

$$\sum_{i=1}^{n} (j+a)^{-\alpha} = 1$$

Approximating the discrete sum as an integral,

$$\int_{1}^{n} (x+a)^{-\alpha}\, dx = 1$$

Assuming $\alpha = 1$, what is a in terms of n? where n is finite.

$$\int_{1}^{n} (x+a)^{-1}\, dx = 1$$

$$\ln(x+a)\Big|_{1}^{n} = 1, \text{for } a > 0 \text{ and } x > 0$$

$$\ln(n+a) - \ln(1+a) = 1$$

$$\ln(\frac{n+a}{1+a}) = 1$$

Taking exponents on both sides,

$$\frac{n+a}{1+a} = e$$

$$n + a = e + ea$$

$$a = \frac{e-n}{1-e}$$

$$\lim_{n\to\infty} a = \lim_{n\to\infty} n \lim_{n\to\infty} \frac{e/n - 1}{1-e} \to \infty$$

$\square$

$\alpha = 1$ setting does not hold good when we have a very large data set.