P o C S What's The Story?

Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 6701, 6713, & a pretend number
University of Vermont, Fall 2024
"He's going to be all right"
Assignment 07

Buster Bluth, Arrested Development, Hand to God, S2E12.
Episode links: Wikipedia, IMDB, Fandom, TV Tropes.

Due: Monday, October 14, by 11:59 pm
https://pdodds.w3.uvm.edu/teaching/courses/2024-2025pocsverse/assignments/07/
Some useful reminders:
Deliverator: Prof. Peter Sheridan Dodds (contact through Teams)
Office: The Ether and/or Innovation, fourth floor
Office hours: See Teams calendar
Course website: https://pdodds.w3.uvm.edu/teaching/courses/2024-2025pocsverse
Overleaf: LaTeX templates and settings for all assignments are available at https://www.overleaf.com/read/tsxfwwmwdgxj.

Some guidelines:

1. Each student should submit their own assignment.
2. All parts are worth 3 points unless marked otherwise.
3. Please show all your work/workings/workingses clearly and list the names of others with whom you ~~conspired~~ collaborated.
4. We recommend that you write up your assignments in LaTeX (using the Overleaf template). However, if you are new to LaTeX or it is all proving too much, you may submit handwritten versions. Whatever you do, please only submit single PDFs.
5. For coding, we recommend you improve your skills with Python, R, and/or Julia. **Please do not use any kind of AI thing.** The (evil) Deliverator uses (evil) Matlab.
6. There is no need to include your code but you can if you are feeling especially proud.

Assignment submission:

Via Brightspace (which is not to be confused with the death vortex of the same name).

Again: One PDF document per assignment only.

Please submit your project's current draft in pdf format via Brightspace four days after the due date for this assignment (normally a Friday). For teams, please list all team member names clearly at the start.

---

1. (3 + 3 = 6 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 λ, find an expression connecting λ, 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 $(j+1)$.

2. (3 points) Carrying on from the previous problem:

   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 $\gamma > 2$)

3. (3 points) 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 $n \to \infty$?

4. (3 + 3 = 6 points)

   Repeat the preceding assignment's question on the largest sample for a power-law size distribution, now with $\gamma = 3/2$.

   As $1 < \gamma < 2$, we should see a very different behavior.

   Here's the question reprinted with γ switched to 3/2.

   The key change in the question is in the form of $F(z)$ (last paragraph).

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

   How do we computationally sample from a discrete probability distribution?

   Hint: You can use a continuum approximation to speed things up. In fact, taking the exact continuum version from the first two assignments will work.

   (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 $\langle k_{\max} \rangle$ 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 of 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 $\gamma = 3/2$?

How to sample from your power law distribution (and similar kinds of beasts):

We now turn our problem of randomly selecting from this distribution into randomly selecting from the uniform distribution.

Because the tail of power-law size distributions can be so long, trying to sample from a discrete distribution can be either painfully slow or even computationally impossible. Brute force often works but not here.

We use a continuous approximation for P_k to make sampling both possible and fast.

We first approximate P_k with $P(z) = (\gamma - 1)z^{-\gamma}$ 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 $F(z) = 1 - z^{-(\gamma-1)} = 1 - z^{-1/2}$ when γ=3/2.

Inverting, we obtain $z = [1 - F(z)]^{-1/(\gamma-1)} = [1 - F(z)]^{-2}$ when γ=3/2.

We now replace $F(z)$ with our random number x and round the value of z to finally get an estimate of k.

In sum, given x is distributed uniformly on $[0, 1]$, then

$$k = \left[(1 - x)^{-2}\right]$$

is approximately distributed according to a power-law size distribution $P_k = ck^{-3/2}$ where $[\ \cdot\ ]$ indicates rounding to the nearest integer.

5. (3 + 3 + 3 + 3 + 3 = 15 points)

   We take a look at the 80/20 rule, 1 per centers, and similar concepts.

   Take x to be the wealth held by an individual in a population of n people, and the number of individuals with wealth between x and $x + \mathrm{d}x$ to be approximately $N(x)\mathrm{d}x$.

   Given a power-law size frequency distribution $N(x) = cx^{-\gamma}$ where $x_{\min} \ll x \ll \infty$, determine the value of γ for which the so-called 80/20 rule holds.

   In other words, find γ for which the bottom 4/5 of the population holds 1/5 of the overall wealth, and the top 1/5 holds the remaining 4/5.

   Note that inherent in our construction of the wealth frequency distribution is that the population is ordered by increasing wealth.

   Assume the mean is finite, i.e., $\gamma > 2$.

   (a) Determine the total wealth W in the system given $\int_{x_{\min}}^{\infty} \mathrm{d}x N(x) = n$.

   (b) Imagine that the bottom $100\,\theta_{\mathrm{pop}}$ percent of the population holds $100\,\theta_{\mathrm{wealth}}$ percent of the wealth.

Show γ depends on θ_{pop} and θ_{wealth} as

$$\gamma = 1 + \frac{\ln \frac{1}{(1-\theta_{\text{pop}})}}{\ln \frac{1}{(1-\theta_{\text{pop}})} - \ln \frac{1}{(1-\theta_{\text{wealth}})}}. \tag{1}$$

(c) Given the above, is every pairing of θ_{pop} and θ_{wealth} possible?

(d) Find γ for the 80/20 requirement ($\theta_{\text{pop}} = 4/5$ and $\theta_{\text{wealth}} = 1/5$).

(e) For general γ, determine the fraction of wealth θ_{wealth} that the bottom fraction θ_{pop} of the population possesses as a function of θ_{pop}. Call this function f: $\theta_{\text{wealth}} = f(\theta_{\text{pop}})$.

For the "80/20" γ you found in (d), plot the curve $f(\theta_{\text{pop}})$ and indicate the 80/20 rule.