Principles of Complex Systems, Vols. 1 and 2
CSYS/MATH 6701, 6713
University of Vermont, Fall 2025
"I wanna ... I wanna change my answer."
Assignment 03

P o C S | What's The Story?

It's Always Sunny in Philadelphia ☑: The Gang Goes on Family Fight, S10E08 ☑
Episode links: IMDB ☑, Fandom ☑, TV Tropes ☑.

Due: Wednesday, September 17, by 11:59 pm

https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/assignments/03/

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/2025-2026pocsverse

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. And it's going to be a no for the catachrestic Excel. Please do not use any kind of AI thing unless directed. 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, just a weird coincidence). Again: One PDF document per assignment only.

All about power law size distributions (basic computations and some real life data from Google).

Note 1: Please do not use Mathematica, etc. for any symbolic work—you can do all of these calculations by hand. Yes you can!

Note 2: Otherwise, use whatever tools you like for the data analysis.

30 points total

1. (3 points)

   As in assignment 1, consider a random variable X with a probability distribution given by
   $$P(x) = cx^{-\gamma},$$
   where c is a normalization constant you determined in the first assignment, and $0 < a \leq x \leq b$. (a and b are the lower and upper cutoffs respectively.) Assume that $\gamma > 1$.

   Note: For all answers you obtain for the questions below, replace c by the expression you obtained in the first assignment, and simplify expressions as much as possible.

   Compute the nth moment of X which is in general defined as:
   $$\langle x^n \rangle = \int_a^b x^n P(x) \, dx.$$

2. (3 points)

   In the limit $b \to \infty$, how does the nth moment behave as a function of γ?

3. (6 points total, 3 + 3)

   (a) Find σ, the standard deviation of X for finite a and b, then obtain the limiting form of σ as $b \to \infty$, noting any constraints we must place on γ for the mean and the standard deviation to remain finite as $b \to \infty$.
   Some help: the form of σ^2 as $b \to \infty$ should reduce to
   $$= \frac{(\gamma - c_1)}{(\gamma - c_2)(\gamma - c_3)^2} a^2$$
   where c_1, c_2, and c_3 are simple, meaningful constants to be determined (by you).

2

(b) For the case of $b \to \infty$, how does σ behave as a function of γ, given the constraints you have already placed on γ? More specifically, how does σ behave as γ reaches the ends of its allowable range?

4. (6 points total, $3 + 3$) Drawing on a Google vocabulary data set (see below for links)

   (a) Plot the frequency distribution N_k representing how many distinct words appear k times in this particular corpus as a function of k.

   (b) Repeat the same plot in log-log space (using base 10, i.e., plot $\log_{10} N_k$ as a function of $\log_{10} k$).

5. (3 points)

   Using your eyeballs, indicate over what range power-law scaling appears to hold and, estimate, using least squares regression over this range, the exponent in the fit $N_k \sim k^{-\gamma}$ (we'll return to this estimate later).

6. (3 points)

   Compute the mean and standard deviation for the entire sample (not just for the restricted range you used in the preceding question). Based on your answers to Questions 2 and 3 as well as material from the lectures, do these values for the mean and standard deviation make sense given your estimate of γ?

   Hint: note that we calculate the mean and variance from the distribution N_k; a common mistake is to treat the distribution as the set of samples. Another routine misstep is to average numbers in log space (oops!) and to average only over the range of k values you used to estimate γ.

   The data for N_k and k (links are clickable):

   - Compressed text file (first column $= k$, second column $= N_k$):
     https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/docs/vocab_cs_mod.txt.gz

   - Uncompressed text file (first column $= k$, second column $= N_k$):
     https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/docs/vocab_cs_mod.txt

   - Matlab file (`wordfreqs` $= k$, `counts` $= N_k$):
     https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/docs/google_vocab_freqs.mat

   The raw frequencies of individual words:

- https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/docs/google_vocab_rawwordfreqs.txt.gz

- https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/docs/google_vocab_rawwordfreqs.txt

- https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/docs/google_vocab_rawwordfreqs.mat

Note: 'words' here include any separate textual object including numbers, websites, html markup, etc.

Note: To keep the file to a reasonable size, the minimum number of appearances is $k_{\min} = 200$ corresponding to $N_{200} = 48030$ distinct words that each appear 200 times.

7. (6 points total: $3 + 3$)

   (a) (3 points) A parent has two children, not twins, and one is a girl born on a Tuesday. What's the probability that both children are girls?

   Assume 50/50 birth probabilities.

   (b) (3 points) Same as the previous question but we now know that one is a girl born on December 31. Again, what's the probability that both are girls?

   (c) Optional, ungraded, to think about: Once you have a calculation of probabilities, can you also add a visual explanation?

   (d) Optional, ungraded, to think about: Once you have the answer, can you improve our intuition here?

   Why does adding the more detailed piece of information of the Tuesday birth change the probability that both are girls?