**Principles of Complex Systems, Vols. 1, 2, & 3D**
**CSYS/MATH 6701, 6713, & a pretend number**
**University of Vermont, Fall 2024**
**"Attacked by a loose seal"**
**Assignment 05**

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

---

**Due:** Monday, September 30, by 11:59 pm
https://pdodds.w3.uvm.edu/teaching/courses/2024-2025pocsverse/assignments/05/
*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.

---

Project start up details.

- Use this Overleaf LaTeX template:

  https://github.com/petersheridandodds/universal-paper-template

Notes for Baby Name analysis:

- You will have the baby name data sets on hand from the previous assignment.
- Recommended: Use the Allotaxonometer for All created by Jonathan St.-Onge and team.

  Here's the app (which should be the easiest to use):

  https://allotaxp.vercel.app/

  Git repo:

  https://github.com/jstonge/allotaxp

  Observable:

  https://observablehq.com/@jstonge/allotaxonometer-4-all

- Alternatively, if you 'love' Matlab:

  https://gitlab.com/compstorylab/allotaxonometer.

  Note that you will need to separately install the command epstopdf (which should be part of a well stocked unix system with LaTeX).

1. (3 + 3 + 3 + 3 + 3 + 3)

   Generate allotaxonographs comparing the following pairs with the stated values of $\alpha$ (6 plots):

   See alternative below if you cannot get any allotaxonometer code to work on your system.

   (a) Baby girl names in 2022 versus baby girl names in 2023 ($\alpha = 1/6$)

(b) Baby boy names in 2022 versus baby boy names in 2023 ($\alpha = 1/6$)

(c) Baby girl names in 1973 versus baby girl names in 2023 ($\alpha = \infty$)

(d) Baby boy names in 1973 versus baby boy names in 2023 ($\alpha = \infty$)

(e) Baby girl names in 1973 versus baby boy names in 1973 ($\alpha = \infty$)

(f) Baby girl names in 2023 versus baby boy names in 2023 ($\alpha = \infty$)

Note that the javascript version does not yet have contour lines which are needed to guide the choice of $\alpha$. Comparing baby name distributions in adjacent years (e.g., baby girl names in 2022 and 2023) will require an $\alpha$ close to 0. Here, we're going with 1/6. But in general, if you can see contour lines, you can choose an $\alpha$ that fits best.

Online appendices for the main allotaxonometry papers is here: http://compstorylab.org/allotaxonometry/.

More information for the Matlab version:

The gitlab repository:

https://gitlab.com/compstorylab/allotaxonometer/

For example baby names code, look through the main script here:

https://gitlab.com/compstorylab/allotaxonometer/figures/babynames/figures/

See if you can get this script to run as is.

Contains overview, examples, links to papers, figure-making code, etc.

**Alternative:**

Using rank-turbulence divergence with $\alpha = 0$ and $\infty$, list the top 30 contributing baby names for the four comparisons listed above.

Indicate which year each contributing baby name comes from in parentheses.

For ordering, you do not need to compute RTD in full but rather just the core structure:

$$\left| \frac{1}{[r_{\tau,1}]^{\alpha}} - \frac{1}{[r_{\tau,2}]^{\alpha}} \right| . \tag{1}$$

Recall that for $\alpha = 0$ and $\alpha = \infty$, the essential core structure becomes:

$$\left| \ln \frac{r_{\tau,1}}{r_{\tau,2}} \right| \text{ and } \max_{\tau} \left\{ \frac{1}{r_{\tau,1}}, \frac{1}{r_{\tau,2}} \right\} . \tag{2}$$

2. (3 points)

Everyday random walks and the Central Limit Theorem:

Show that the observation that the number of discrete random walks of duration $t = 2n$ starting at $x_0 = 0$ and ending at displacement $x_{2n} = 2k$ where $k \in \{0, \pm 1, \pm 2, \ldots, \pm n\}$ is

$$N(0, 2k, 2n) = \binom{2n}{n+k} = \binom{2n}{n-k}$$

leads to a Gaussian distribution for large $t = 2n$:

$$\mathbf{Pr}(x_t \equiv x) \simeq \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

Please note that $k \ll n$.

Stirling's sterling approximation will prove most helpful.

**Hint:** You should be able to reach this form:

$$\frac{\text{Some stuff not involving spotted quokkas}}{\text{Some other quokka-free stuff} \times (1 - k^2/n^2)^{n+1/2}(1 + k/n)^k (1 - k/n)^{-k}}.$$

Lots of sneakiness here. You'll want to examine the natural log of the piece shown above, and see how it behaves for large $n$.

You may very well need to use the Taylor expansion $\ln(1 + z) \simeq z$.

Exponentiate and carry on.

**Tip:** If at any point quokkas appear in your expression, you're in real trouble. Get some fresh air and start again.

3. (3 points)

   From lectures, show that the number of distinct 1-$d$ random walk that start at $x = i$ and end at $x = j$ after $t$ time steps is

   $$N(i, j, t) = \binom{t}{(t + j - i)/2}.$$

   Assume that $j$ is reachable from $i$ after $t$ time steps.

   **Hint—Counting random walks**:
   http://www.youtube.com/watch?v=daSIYz-0U3E

4. (3 + 3)

   *Discrete random walks:*

   In class, we argued that the number of random walks returning to the origin for the first time after $2n$ time steps is given by

   $$N_{\mathrm{fr}}(2n) = 2 \times N^{+}_{\mathrm{fr}}(2n) = 2\left(N(1, 1, 2n - 2) - N(-1, 1, 2n - 2)\right)$$

where

$$N(i, j, t) = \binom{t}{(t + j - i)/2}.$$

Find the leading order term for $N_{\mathrm{fr}}(2n)$ as $n \to \infty$.

Two-step approach:

(a) Combine the terms to form a single fraction,

(b) and then again use Stirling's approximation.

If you enjoy this sort of thing, you may like to explore the same problem for random walks in higher dimensions. Seek out George Pólya.

And we are connecting to much other good stuff in combinatorics; more to come in the solutions.