PoCS What's The Story?

Principles of Complex Systems, Vols. 1 & 2, CSYS/MATH 300 and 303
University of Vermont, Fall 2021
Assignment 05 • code name: Bored On Board Onboard

Due: Friday, October 1, by 11:59 pm, 2021.
Relevant clips, episodes, and slides are listed on the assignment's page:
https://pdodds.w3.uvm.edu//teaching/courses/2021-2022principles-of-complex-systems//assignments/05/
Some useful reminders:
Deliverator: Prof. Peter Sheridan Dodds (contact through Teams)
Assistant Deliverator: Michael Arnold (contact through Teams)
Office: The Ether
Office hours: TBD
Course website:
https://pdodds.w3.uvm.edu//teaching/courses/2021-2022principles-of-complex-systems

---

All parts are worth 3 points unless marked otherwise. Please show all your workingses clearly and list the names of others with whom you collaborated.

For coding, we recommend you improve your skills with Python, R, and/or Julia. The Deliverator uses Matlab.

Graduate students are requested to use LaTeX (or related TeX variant). If you are new to LaTeX, please endeavor to submit at least n questions per assignment in LaTeX, where n is the assignment number.

Assignment submission: Via Blackboard.

Please submit your project's current draft in pdf format via Blackboard by the same time specified for this assignment. For teams, please list all team member names clearly at the start.

---

1. Generate allotaxonographs comparing the following four pairs:

   (a) Baby girl names in 1952 versus baby girl names in 2002.
   (b) Baby boy names in 1952 versus baby boy names in 2002.
   (c) Baby girl names in 1952 versus baby boy names in 1952.
   (d) Baby girl names in 2002 versus baby boy names in 2002.

Use rank-turbulence divergence with $\alpha = \infty$.

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

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

Notes:

- You will the data sets on hand from the previous assignment.
- Matlab. Yes. You will need to install it. Please connect with Assistant to the Regional Deliverator.
- Unix systems will work (Linux, the Apple things, etc.).
- As is, you will need the command epstopdf. Please install if not on deck already.

2. 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:

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

   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 penguins appear in your expression, you're in real trouble. Get some fresh air and start again.

3. 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. *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

$$N_{\text{first return}}(2n) = N_{\text{fr}}(2n) = N(1, 1, 2n - 2) - N(-1, 1, 2n - 2)$$

where

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

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

Two-step approach:

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

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

5. (3 + 3)

More on the peculiar nature of distributions of power law tails:

Consider a set of N samples, randomly chosen according to the probability distribution $P_k = ck^{-\gamma}$ where $k = 1, 2, 3, \ldots$.

(a) Estimate $\min k_{\max}$, the approximate minimum of the largest sample in the system, finding how it depends on N.

(Hint: we expect on the order of 1 of the N samples to have a value of $\min k_{\max}$ or greater.)

Hint—Some visual help on setting this problem up:
http://www.youtube.com/watch?v=4tqlEuXA7QQ

(b) Determine the average value of samples with value $k \geq \min k_{\max}$ to find how the expected value of $k_{\max}$ (i.e., $\langle k_{\max} \rangle$) scales with N.

Notes:

- For language, this scaling is known as Heap's law.
- In a later assignment, we will test this scaling by (thoughtfully) sampling from power-law size distributions.