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 05
SQUEAK

Name: Krishna Kannan Srinivasan

Conspirators:

Notes for Baby Name analysis:

- You will have the data sets on hand from the previous assignment.
- Unix systems will work (Linux, the Apple things, etc.). Windows may not (alternative given below).
- As is, you will need the command epstopdf. Please install if not on deck already.

1. Plot time series for the rank of these baby names in the US over all years in the census data:

   - Shirley.
   - Desmond.
   - Madison.
   - Aiden.
   - A name of your choice.

   Note that if you plotted relative frequency rather than rank, you would need to know (or estimate) the overall number of babies born. Ranks are both easy simple to work with and easy to understand.

Solution:

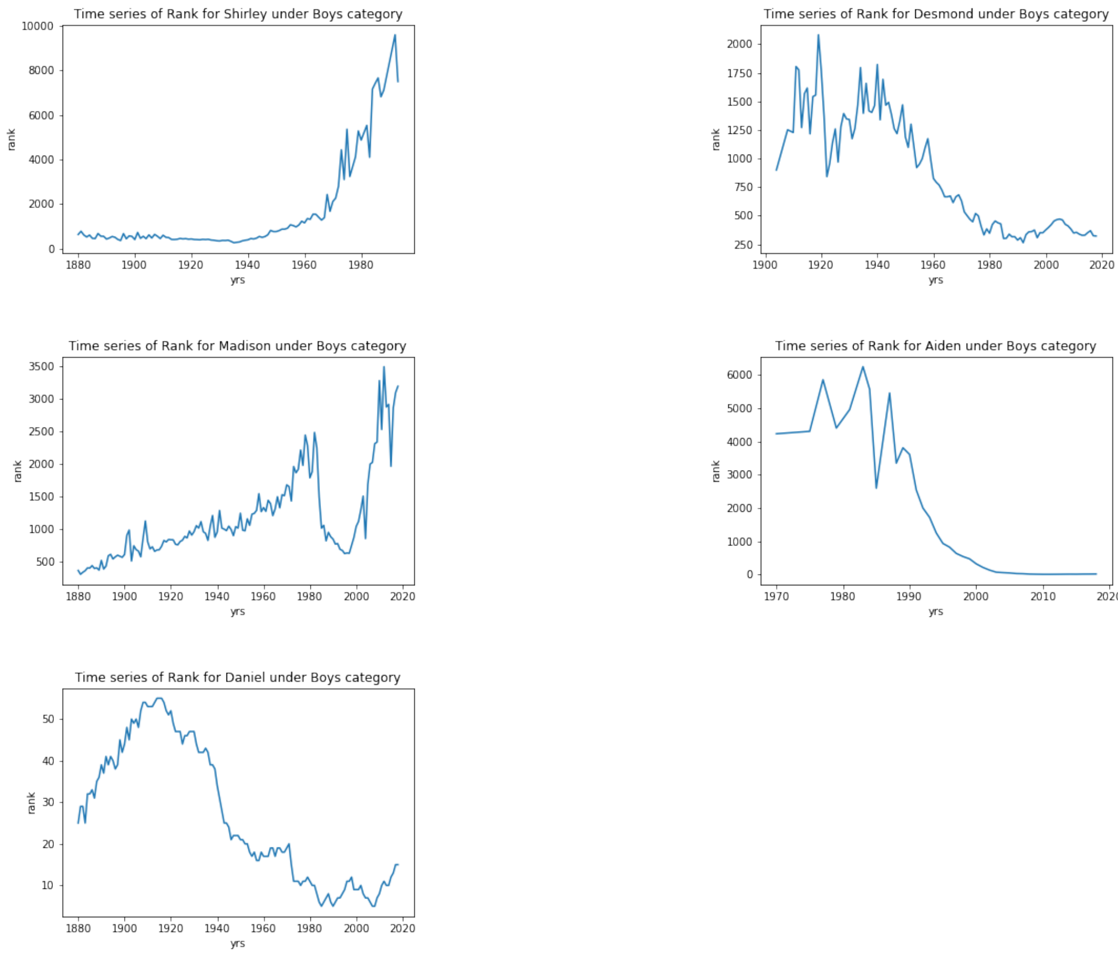

Figure 1: Boys Category

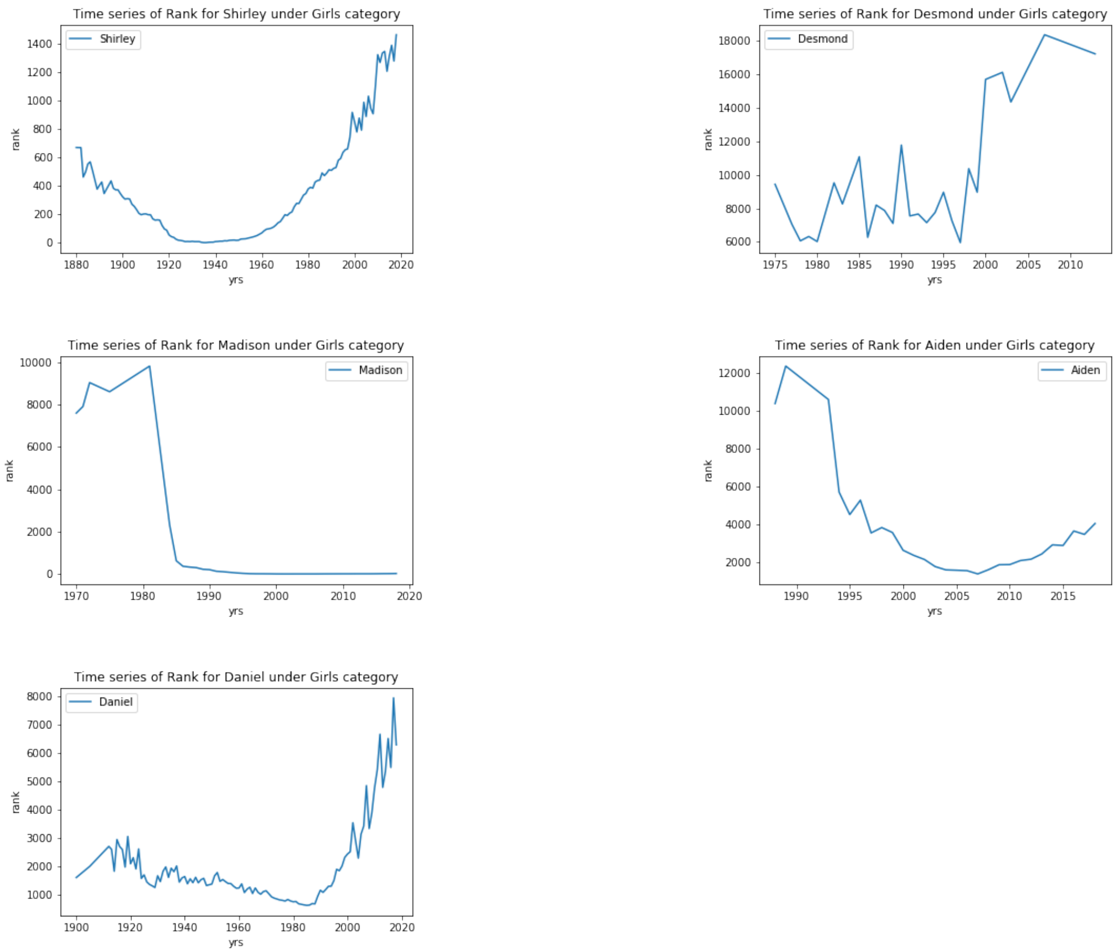

Figure 2: Girls Category

2. (3 + 3 + 3 + 3)

   Only this question requires The Laboratory of the Matrix (Matlab). See alternative below if you cannot get the allotaxonometer to work on your system.

   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 $\alpha = 0$ and $\alpha = \infty$.

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

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.

Solution:

Alternative: Cases $\alpha = 0$ and $\alpha = \infty$ are computed for system 1 names: rank and system 2names: rank For name in system1 if not in system 2, compute using lowest rank for system2.
Top 30 names and their Rank Turbulence scores for each case of α

(a) Baby girl names in 1952 versus baby girl names in 2002.

```
Baby girl names in 1952 versus baby girl names in 2002

Alpha = 0 [('Vola', 8.76), ('Vonne', 8.76), ('Wallie', 8.76), ('Wally', 8.76), ('Walterine', 8.76), ('Weida', 8.76),
('Wileen', 8.76), ('Wilene', 8.76), ('Willean', 8.76), ('Williette', 8.76), ('Willine', 8.76), ('Willistine', 8.76),
('Willo', 8.76), ('Willye', 8.76), ('Winfred', 8.76), ('Woodie', 8.76), ('Yevette', 8.76), ('Yola', 8.76), ('Yvone',
8.76), ('Vija', 8.76), ('Vilda', 8.76), ('Vilia', 8.76), ('Vincetta', 8.76), ('Viona', 8.76), ('Victorine', 8.76),
('Vessie', 8.76), ('Vickiann', 8.76), ('Unice', 8.75), ('Vail', 8.75), ('Valita', 8.75)]

Alpha = inf [('Linda', 1.0), ('Emily', 1.0), ('Willie', 1.0), ('Gale', 1.0), ('Gay', 1.0), ('Laverne', 1.0), ('Dian
n', 1.0), ('Earnestine', 1.0), ('Bettye', 1.0), ('Bette', 1.0), ('Myrtle', 1.0), ('Pat', 1.0), ('Delois', 1.0), ('Ess
ie', 1.0), ('Kathi', 1.0), ('Earlene', 1.0), ('Pearlie', 1.0), ('Cathie', 1.0), ('Earline', 1.0), ('Eula', 1.0), ('Be
verley', 1.0), ('Alfreda', 1.0), ('Melva', 1.0), ('Vickey', 1.0), ('Pamala', 1.0), ('Gaye', 1.0), ('Debrah', 1.0),
('Pam', 1.0), ('Harriett', 1.0), ('Peggie', 1.0)]
```

(b) Baby boy names in 1952 versus baby boy names in 2002.

```
Baby boy names in 1952 versus baby boy names in 2002.

Alpha = 0 [('Zebbie', 8.36), ('Willia', 8.35), ('Wilmot', 8.35), ('Woodroe', 8.35), ('Wydell', 8.35), ('Wilgus', 8.3
5), ('Webber', 8.35), ('Wendolyn', 8.35), ('Valiant', 8.35), ('Varnell', 8.35), ('Vasco', 8.35), ('Veldon', 8.35),
('Vernel', 8.35), ('Vilas', 8.35), ('Virgin', 8.35), ('Virlyn', 8.35), ('Vollie', 8.35), ('Vondell', 8.35), ('Vytas',
8.35), ('Wadie', 8.35), ('Wandell', 8.35), ('Warden', 8.35), ('Turhan', 8.35), ('Udell', 8.35), ('Ulyess', 8.35), ('U
ral', 8.35), ('Thurl', 8.35), ('Thurlow', 8.35), ('Tiburcio', 8.35), ('Tee', 8.34)]

Alpha = inf [('James', 1.0), ('Jacob', 1.0), ('Gail', 1.0), ('Merrill', 1.0), ('Linda', 1.0), ('Dickie', 1.0), ('Conn
ie', 1.0), ('Burt', 1.0), ('Buford', 1.0), ('Gaylord', 1.0), ('Gearld', 1.0), ('Carol', 1.0), ('Patricia', 1.0), ('Ga
yle', 1.0), ('Lindsay', 1.0), ('Bud', 1.0), ('Alva', 1.0), ('Lacy', 1.0), ('Deborah', 1.0), ('Len', 1.0), ('Sherwoo
d', 1.0), ('Beverly', 1.0), ('Burl', 1.0), ('Lynwood', 1.0), ('Milford', 1.0), ('Normand', 1.0), ('Meredith', 1.0),
('Barbara', 1.0), ('Verne', 1.0), ('Cleve', 1.0)]
```

(c) Baby girl names in 1952 versus baby boy names in 1952.

```
Baby girl names in 1952 versus baby boy names in 1952.

Alpha = 0 [('Yevette', 8.76), ('Yola', 8.76), ('Yvone', 8.76), ('Walterine', 8.76), ('Weida', 8.76), ('Wileen', 8.7
6), ('Wilene', 8.76), ('Willean', 8.76), ('Williette', 8.76), ('Willine', 8.76), ('Willistine', 8.76), ('Willo', 8.7
6), ('Willye', 8.76), ('Veronia', 8.76), ('Veronique', 8.76), ('Vessie', 8.76), ('Vickiann', 8.76), ('Vicktoria', 8.7
6), ('Victoriana', 8.76), ('Victorine', 8.76), ('Victory', 8.76), ('Vija', 8.76), ('Vilda', 8.76), ('Vilia', 8.76),
('Vincetta', 8.76), ('Viona', 8.76), ('Viviana', 8.76), ('Vola', 8.76), ('Vonne', 8.76), ('Unice', 8.75)]

Alpha = inf [('Linda', 1.0), ('Colleen', 1.0), ('Claudia', 1.0), ('Sue', 1.0), ('Marlene', 1.0), ('Lynda', 1.0), ('Ro
semary', 1.0), ('Jeanette', 1.0), ('Cindy', 1.0), ('Beth', 1.0), ('Christina', 1.0), ('Terri', 1.0), ('Sara', 1.0),
('Paulette', 1.0), ('Carole', 1.0), ('Candace', 1.0), ('Belinda', 1.0), ('Amy', 1.0), ('Edna', 1.0), ('Lillian', 1.
0), ('Emma', 1.0), ('Esther', 1.0), ('Jacquelyn', 1.0), ('Roxanne', 1.0), ('Veronica', 1.0), ('Marianne', 1.0), ('Bec
ky', 1.0), ('Grace', 1.0), ('Melinda', 1.0), ('Clara', 1.0)]
```

(d) Baby girl names in 2002 versus baby boy names in 2002.

```
Baby girl names in 2002 versus baby boy names in 2002 .

Alpha = 0 [('Zayonna', 9.8), ('Zeenat', 9.8), ('Zeidy', 9.8), ('Zelie', 9.8), ('Zelma', 9.8), ('Zeneida', 9.8), ('Zen
y', 9.8), ('Zera', 9.8), ('Ziann', 9.8), ('Zillah', 9.8), ('Zinia', 9.8), ('Zitlalli', 9.8), ('Zniya', 9.8), ('Zoa',
9.8), ('Zona', 9.8), ('Zuha', 9.8), ('Zula', 9.8), ('Zulay', 9.8), ('Zuleica', 9.8), ('Zyara', 9.8), ('Zylah', 9.8),
('Yzabel', 9.8), ('Zacaria', 9.8), ('Zakaiya', 9.8), ('Zanea', 9.8), ('Zaret', 9.8), ('Zareth', 9.8), ('Zareyah', 9.
8), ('Zarianna', 9.8), ('Zariel', 9.8)]

Alpha = inf [('Emily', 1.0), ('Katelyn', 1.0), ('Caroline', 1.0), ('Angelina', 1.0), ('Lily', 1.0), ('Ava', 1.0), ('E
lla', 1.0), ('Ariana', 1.0), ('Claire', 1.0), ('Lillian', 1.0), ('Marissa', 1.0), ('Katie', 1.0), ('Breanna', 1.0),
('Alexandria', 1.0), ('Laura', 1.0), ('Molly', 1.0), ('Isabelle', 1.0), ('Sofia', 1.0), ('Arianna', 1.0), ('Alicia',
1.0), ('Caitlin', 1.0), ('Mikayla', 1.0), ('Jillian', 1.0), ('Margaret', 1.0), ('Sabrina', 1.0), ('Lydia', 1.0), ('Am
elia', 1.0), ('Tiffany', 1.0), ('Erica', 1.0), ('Miranda', 1.0)]
```

□

Alternative:

Using rank-turbulence divergence with $\alpha = 0$ and ∞, 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:

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

3. 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.

Solution: We have,

$$\begin{aligned} N(0,2k,2n) &= \binom{2n}{n+k} \\ &= \frac{(2n)!}{(n+k)!(2n-n-k)!} \\ &= \frac{(2n)!}{(n+k)!(n-k)!} \end{aligned} \tag{3}$$

Using Sterling's approximation,

$$n! = \sqrt{(2\pi)} n^{n+\frac{1}{2}} e^{-n}$$

3 becomes,

$$\begin{aligned} N(0,2k,2n) &= \frac{\sqrt{(2\pi)}(2n)^{2n+\frac{1}{2}} e^{-2n}}{\sqrt{(2\pi)}(n+k)^{n+k+\frac{1}{2}} e^{-(n+k)} \sqrt{(2\pi)}(n-k)^{n-k+\frac{1}{2}} e^{-(n-k)}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{2n^{2n+\frac{1}{2}} e^{-2n}}{(n+k)^{n+k+\frac{1}{2}}(n-k)^{n-k+\frac{1}{2}} e^{-2n}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{2^{2n+\frac{1}{2}} n^{2n+\frac{1}{2}}}{(1+\frac{k}{n})^{n+k+\frac{1}{2}} n^{n+k+\frac{1}{2}} (1-\frac{k}{n})^{n-k+\frac{1}{2}} n^{n-k+\frac{1}{2}}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{2^{2n+\frac{1}{2}} n^{2n+\frac{1}{2}}}{n^{2n+\frac{1}{2}} n^{\frac{1}{2}} (1+\frac{k}{n})^{n+k+\frac{1}{2}} (1-\frac{k}{n})^{n-k+\frac{1}{2}}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{2^{2n+\frac{1}{2}}}{n^{\frac{1}{2}} (1+\frac{k}{n})^{n+\frac{1}{2}} (1+\frac{k}{n})^{k} (1-\frac{k}{n})^{n+\frac{1}{2}} (1-\frac{k}{n})^{-k}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{2^{2n+\frac{1}{2}}}{n^{\frac{1}{2}} (1-\frac{k^2}{n^2})^{n+\frac{1}{2}} (1+\frac{k}{n})^{k} (1-\frac{k}{n})^{-k}} \end{aligned} \tag{4}$$

Let $\delta = n^{\frac{1}{2}}(1-\frac{k^2}{n^2})^{n+\frac{1}{2}}(1+\frac{k}{n})^k(1-\frac{k}{n})^{-k}$

$$\ln\delta = (n+\frac{1}{2})\ln(1-\frac{k^2}{n^2}) + k\ln(1+\frac{k}{n}) - k\ln(1-\frac{k}{n})$$

Using Taylor's expansion, $\ln(1+x) \ = x, x \ll 1$
Since $k \ll n$, $\frac{k}{n} \ll 1$ Then,

$$\ln \delta \simeq (n+\frac{1}{2})(-\frac{k^2}{n^2}) + k(\frac{k}{n}) - k(-\frac{k}{n})$$

$$\ln \delta \simeq -\frac{k^2}{n} - \frac{k}{2n^2} + \frac{k^2}{n} + \frac{k^2}{n}$$

$$\ln \delta \simeq \frac{k^2}{n} - \frac{k}{2n^2}$$

Since $k \ll n, k^2 \ll n^2, \lim_{n\to\infty} \frac{k^2}{2n} = 0$

$$\ln \delta \simeq \frac{k^2}{n}$$

Taking exponential on both sides,

$$\delta \simeq = e^{\frac{k^2}{n}}$$

Then,

$$N(0, 2k, 2n) \simeq \frac{1}{\sqrt{2\pi}} \frac{2^{2n+\frac{1}{2}}}{n^{\frac{1}{2}} e^{\frac{k^2}{n}}}$$

$$N(0, 2k, 2n) \simeq \frac{1}{\sqrt{2\pi}} \frac{2^{2n+\frac{1}{2}} 2^{\frac{1}{2}} e^{-\frac{k^2}{n}}}{2^{\frac{1}{2}} n^{\frac{1}{2}}}$$

$$N(0, 2k, 2n) \simeq \frac{1}{\sqrt{2\pi}} \frac{2^{2n+1} e^{-\frac{(2k)^2}{2^2 n}}}{(2n)^{\frac{1}{2}}}$$

For large n, $2^{2n+1} \simeq 2^{2n}$.

$$N(0, 2k, 2n) \simeq \frac{1}{\sqrt{2\pi}} \frac{2^{2n} e^{-\frac{(2k)^2}{2^2 n}}}{(2n)^{\frac{1}{2}}}$$

We know that,

$$P(0, 2k, 2n) = \frac{1}{2^{2n}} N(0, 2k, 2n)$$

$$P(0, 2k, 2n) = \frac{1}{2^{2n}} \frac{1}{\sqrt{2\pi}} \frac{2^{2n} e^{-\frac{(2k)^2}{2^2 n}}}{(2n)^{\frac{1}{2}}}$$

$$P(0, 2k, 2n) = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{(2k)^2}{2^2 n}}}{(2n)^{\frac{1}{2}}}$$

Let $x = 2k, t = 2n$,

$$P(0, 2k, 2n) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-x^2}{2t}}$$

Which is a normal distribution.
Hence shown.

$\square$

4. 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

Solution:

If we start a random walk from i and end at j after t steps, the number of ways this is possible can be calculated as follows,
Each step is either in positive direction from x^{th} step or a negative direction from x^{th} step.
Let P be the number of positive steps,
N be the number of negative steps.
Then,

(a) P + N = t
(b) P - N = j - i , which is the displacement from i to j

Solving for P,

$$P = \frac{t + j - 1}{2}$$

$$N = \frac{2t + j + 1}{2}$$

The number of possible random walks is the number of ways we can chose P steps from t total steps. Since $\binom{n}{k} = \binom{n}{(n-k)}$,

$$\binom{t}{P} = \binom{t}{t - P} = \binom{t}{N}$$

$$N(i, j, k) = \binom{t}{P}$$

$$N(i,j,k) = \binom{t}{\frac{t+j-1}{2}}$$

Hence shown.

$\square$

5. *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.

   Solution:

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

   $$\begin{aligned} N_{fr}(2n) &= N(1,1,2n-2) - N(-1,1,2n-2) \\ &= \binom{2n-2}{n-1} - \binom{2n-2}{n} \\ &= \frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{(2n-2)!}{n!(n-2)!} \end{aligned} \tag{5}$$

   Since $n! = n(n-1)!$, and $(n-2)! = \frac{(n-1)!}{(n-1)}$

   $$N_{fr}(2n) = \frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{(2n-2)!}{n(n-1)!(n-1)!/(n-1)} = \frac{(2n-2)!}{(n-1)!(n-1)!}\left[1 - \frac{(n-1)}{n}\right]$$

$$N_{fr}(2n) = \frac{(2n-2)!}{n(n-1)!(n-1)!}$$

$$\lim_{n\to\infty} N_{fr}(2n) = \lim_{n\to\infty} \frac{(2n-2)!}{n(n-1)!(n-1)!}$$

Using Sterling's approximation,

$$n! = \sqrt{(2\pi)} n^{n+\frac{1}{2}} e^{-n}$$

$$\lim_{n\to\infty} N_{fr}(2n) \simeq \frac{1}{n} \frac{\sqrt{2\pi}(2n-2)^{2n-2+1/2} e^{-(2n-2)}}{\sqrt{2\pi}(n-1)^{n-1+1/2} e^{-(n-1)} \sqrt{2\pi}(n-1)^{n-1+1/2} e^{-(n-1)}}$$

$$\lim_{n\to\infty} N_{fr}(2n) \simeq \frac{1}{n} \frac{(2n-2)^{2n-2+1/2} e^{-(2n-2)}}{\sqrt{2\pi}(n-1)^{2n-1} e^{-(2n-2)}}$$

$$\lim_{n\to\infty} N_{fr}(2n) \simeq \frac{1}{n} \frac{2^{2n-2+1/2}(n-1)^{2n-2+1/2}}{\sqrt{2\pi}(n-1)^{2n-1}}$$

$$\lim_{n\to\infty} N_{fr}(2n) \simeq \frac{1}{n} \frac{2^{2n-3/2}}{\sqrt{2\pi}(n-1)^{1/2}}$$

$$\lim_{n\to\infty} N_{fr}(2n) \simeq \frac{1}{n} \frac{2^{2n-3/2}}{\sqrt{2\pi} n^{1/2}(1-\frac{1}{n})^{1/2}}$$

$$\lim_{n\to\infty} (1-\frac{1}{n})^{1/2} = 1$$

Then,

$$\lim_{n\to\infty} N_{fr}(2n) \simeq \frac{1}{n} \frac{2^{2n}}{\sqrt{2\pi} 2^{3/2} n^{3/2}}$$

$$\lim_{n\to\infty} N_{fr}(2n) \simeq \frac{1}{n} \frac{2^{2n}}{\sqrt{2\pi}(2n)^{3/2}}$$

Here 2^{2n} is the leading order term that dominates the behavior(blows up) as $n \to \infty$

□