CSYS 300 Assignment 4

Missy Westland (mwestlan@uvm.edu)

October 8, 2021

Link to Github for Assignment Code: https://github.com/b0ws3r/PoCS

Question 1.

Allotaxonometry.
Rank-turbulence divergence (RTD) is defined as:

$$
\begin{aligned}
D^{\mathrm{R}}_{\alpha}(R_1 \,\|\, R_2) &= \sum_{\tau \in R_{1,2;\alpha}} \delta D^{\mathrm{R}}_{\alpha,\tau}(R_1 \,\|\, R_2) \\
&= \frac{1}{\mathcal{N}_{1,2;\alpha}} \frac{\alpha+1}{\alpha} \sum_{\tau \in R_{1,2;\alpha}} \left| \frac{1}{[r_{\tau,1}]^{\alpha}} - \frac{1}{[r_{\tau,2}]^{\alpha}} \right|^{1/(\alpha+1)}
\end{aligned}
$$

Find the limits of RTD for:
(a) $\alpha \to 0$
(b) $\alpha \to \infty$

Leave $\frac{1}{\mathcal{N}_{1,2;\alpha}}$ as a constant.

Responses:

(a)

$$
\begin{aligned}
&\lim_{\alpha\to 0} \frac{\alpha+1}{\alpha} \sum \left| \frac{1}{(r_{\tau,1})^{\alpha}} - \frac{1}{(r_{\tau,2})^{\alpha}} \right|^{\frac{1}{\alpha+1}} \\
&= \lim_{\alpha\to 0} \frac{\alpha+1}{\alpha} \sum \left| e^{-\alpha \ln r_{\tau,1}} - e^{-\alpha \ln r_{\tau,2}} \right|^{\frac{1}{\alpha+1}} \\
&= \lim_{\alpha\to 0} \frac{\alpha+1}{\alpha} \sum \left| (1 - \alpha \ln r_{\tau,1}) - (1 - \alpha \ln r_{\tau,2}) \right|^{\frac{1}{\alpha+1}} \\
&= \lim_{\alpha\to 0} \frac{\alpha+1}{\alpha} \sum \left| \alpha \ln r_{\tau,2} - \alpha \ln r_{\tau,1}) \right|^{\frac{1}{\alpha+1}} \\
&= \lim_{\alpha\to 0} \frac{\alpha+1}{\alpha} \sum \left| \alpha \ln \frac{r_{\tau,2}}{r_{\tau,1}} \right|^{\frac{1}{\alpha+1}} \\
&= \lim_{\alpha\to 0} \alpha+1 \sum \left| \ln \frac{r_{\tau,2}}{r_{\tau,1}} \right|^{\frac{1}{\alpha+1}} \\
&= \sum \left| \ln \frac{r_{\tau,2}}{r_{\tau,1}} \right|
\end{aligned}
$$

(b)

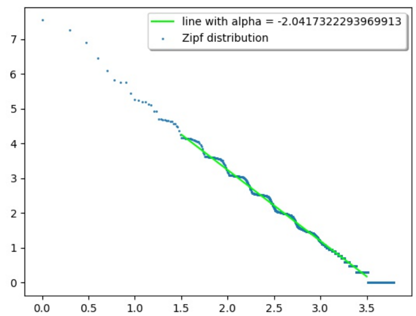
$$
\begin{aligned}
&\lim_{\alpha\to\infty}\frac{\alpha+1}{\alpha}\sum\left|\frac{1}{(r_{\tau,1})^{\alpha}}-\frac{1}{(r_{\tau,2})^{\alpha}}\right|^{\frac{1}{\alpha+1}}\\
&=\lim_{\alpha\to\infty}\sum\left|\frac{1}{(r_{\tau,1})^{\alpha}}-\frac{1}{(r_{\tau,2})^{\alpha}}\right|^{\frac{1}{\alpha+1}}\\
&=\lim_{\alpha\to\infty}\sum\left|\frac{1}{\min(r_{\tau,1},r_{\tau,1})}^{\frac{\alpha}{\alpha+1}}\left(1-\frac{\min(r_{\tau,1},r_{\tau,1})}{\max(r_{\tau,1},r_{\tau,1})}^{\alpha}\right)\right|^{\frac{1}{\alpha+1}}\\
&=\lim_{\alpha\to\infty}\sum\left|\frac{1}{\min(r_{\tau,1},r_{\tau,1})}^{\frac{\alpha}{\alpha+1}}\right|\\
&=\sum\left|\frac{1}{\min(r_{\tau,1},r_{\tau,1})}\right|
\end{aligned}
$$

Question 2.

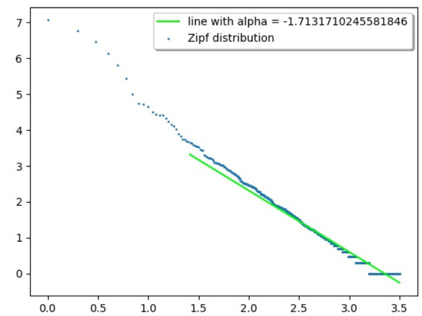
Code up Simon's rich-gets-richer model. Show Zipf distributions for $\rho = .1, .01,$ and $.001.$ and perform regressions to test $\alpha = 1 - \rho$.
Run the simulation for long enough to produce decent scaling laws (recall: three orders of magnitude is good).
Averaging over simulations will produce cleaner results so try 10 and then, if possible, 100.

Note the first mover advantage.

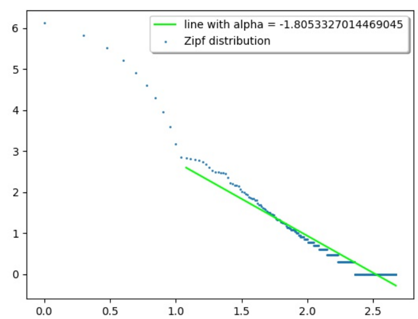
Responses:
I'm not sure my approach to combining runs was entirely correct, as this resulted in aggegating results rather than averaging anything. The expected result was that the ρ values should have approached 1, but this was not observed (probably human error). Code here



(a) Here we see the worst fit - The first mover's advantage is the largest here as a consequence of a low innovation rate



(b)



(c)

Question 3.

For Herbert Simon's model of what we've called Random Competitive Replication, we found in class that the normalized number of groups in the long time limit, n_k, satisfies the following difference equation:

$$\frac{n_k}{n_{k-1}} = \frac{(k-1)(1-\rho)}{1+(1-\rho)k}$$

where k $\geq$ 2. The model parameter ρ is the probability that a newly arriving node forms a group of its own (or is a novel word, starts a new city, has a unique flavor, etc.). For $k = 1$, we have instead

$$n_1 = \rho - (1-\rho)n_1$$

which directly gives us n_1 in terms of ρ.

(a) Derive the exact solution for n_k in terms of gamma functions and ultimately the beta function.

Note: Simon's own calculation is slightly awry. The end result is good however.

Responses:

$$
\begin{aligned}
n_k &= \frac{(k-1)(1-\rho)}{1+k(1-\rho)} n_{k-1} \\
&= \frac{(k-1)(1-\rho)}{1+k(1-\rho)} * \frac{(k-2)(1-\rho)}{1+(k-2)(1-\rho)} n_{k-2}
\end{aligned}
$$

Note in the numerator, $(1-\rho)$ will go to $(1-\rho)^{k-1}$ by the time we reach term n_1, and the other term will become $(k-1)!$

$$
\begin{aligned}
&= \frac{(1-\rho)^{k-1}(k-1)!}{(1+k(1-\rho))(1+(k-1)(1-\rho))\dots(1+(k-(k-1))(1-\rho))} * n_1 \\
&= \frac{(1-\rho)^{k-1}(k-1)!}{(1-\rho)^k(\frac{1}{1-\rho}+k)(\frac{1}{1-\rho}+(k-1))\dots(\frac{1}{1-\rho}+(k-(k-1)))} * n_1
\end{aligned}
$$

We can then reduce the $(1-\rho)^{k-1}$ in the numerator with the term $(1-\rho)^k$ in the denominator.

$$
\begin{aligned}
&= \frac{(k-1)!}{(1-\rho)(\frac{1}{1-\rho}+k)(\frac{1}{1-\rho}+(k-1))\dots(\frac{1}{1-\rho}+(k-(k-1)))} * n_1 \\
&= \frac{\Gamma(k)}{\Gamma(k+\frac{1}{1-\rho}+1)} * \frac{1}{1-\rho} * n_1 \\
&= \Gamma(k)\frac{\Gamma(\frac{1}{1-\rho})}{\Gamma(k+1+\frac{1}{1-\rho})} * \frac{1}{1-\rho} * n_1 \\
&= \Gamma(k)\frac{\Gamma(\frac{1}{1-\rho})}{\Gamma(k+\frac{1}{1-\rho})} * \frac{1}{(1-\rho)(k+\frac{1}{(1-\rho)})} * n_1 \\
&= \beta(k, \frac{1}{1-\rho}) * \frac{1}{(1-\rho)k+1} * n_1 \\
&\text{Now, substitute } n_1 = \frac{\rho}{2-\rho} \\
&= \beta(k, \frac{1}{1-\rho}) * \frac{1}{k(1-\rho)+1} * \frac{\rho}{2-\rho}
\end{aligned}
$$

(b) From this exact form, determine the large k behavior for $n_k(\sim k^{-\gamma}$ and identify the exponent γ in terms of ρ. You are welcome to use the fact that $B(x,y) \sim x^{-y}$ for large x and fixed y (use Stirling's approximation or possibly Wikipedia)

Take $\gamma = 1 + \frac{1}{1-\rho}$, and use the fact that $\beta(k, \frac{1}{1-\rho}) \approx k^{-\frac{1}{1-\rho}}$

$$
\begin{aligned}
&\lim_{k\to\infty} \frac{1}{k(1-\rho)+1} * \frac{\rho}{2-\rho} * k^{-(\gamma-1)} \\
&= \lim_{k\to\infty} \frac{\rho}{2-\rho} \frac{1}{k(1-\rho+\frac{1}{k})k^{\gamma-1}} \\
&= \lim_{k\to\infty} \frac{\rho}{2-\rho} \frac{1}{(1-\rho+\frac{1}{k})k^{\gamma}} \\
&= 0
\end{aligned}
$$

Question 4.

What happens to γ in the limits $\rho \to 0$ and $\rho \to 1$? Explain in a sentence or two what is going on in these cases and how the specific limiting value of γ makes sense.

Responses:

- $\gamma \to 0$
  As $\rho \to 0$, we see that $\gamma = 1 + \frac{1}{1-\rho} \to 1 + 1 = 2$. This means that there will be many of the same type in the system, since the innovation rate is low.
- $\gamma \to 1$
  As $\rho \to 1$, we see that $\gamma = 1 + \frac{1}{1-\rho} \to 1 + \infty = \infty$. This means that there will be infinite variance, due to the high innovation rate.

Question 5.

In Simon's original model, the expected total number of distinct groups at time t is pt. Recall that each group is made up of elements of a particular flavor. In class, we derived the fraction of groups containing only 1 element, finding $n_1^{(g)} = \frac{N_1(t)}{pt} = \frac{1}{2-\rho}$
(a) Find the form of $n_2^{(g)}$ and $n_3^{(g)}$, the fraction of groups that are of size 2 and size 3.

(b) Using data for James Joyce's Ulysses (see below), first show that Simon's estimate for the innovation rate $\rho_{est} \simeq 0.115$ is reasonably accurate for the version of the text's word counts given below.
Hint: You should find a slightly higher number than Simon did.
Hint: Do not compute ρ_{est} from an estimate of γ.

(c) Now compare the theoretical estimates for $n_1^{(g)}$, $n_2^{(g)}$, $n_3^{(g)}$, with empirical values you obtain for Ulysses. The data:

Responses:
Code here

(a) Using the fact that $n_k = \frac{(k-1)(1-\rho)}{1+k(1-\rho)} n_{k-1}$, and $n_1 = \frac{1}{2-\rho}$

$$\begin{aligned} n_2 &= \frac{(2-1)(1-\rho)}{1+2(1-\rho)} * n_1 \\ &= \frac{1-\rho}{1+2(1-\rho)} * \frac{1}{2-\rho} \end{aligned}$$

$$\begin{aligned} n_3 &= \frac{(3-1)(1-\rho)}{1+3(1-\rho)} * n_2 \\ &= \frac{2(1-\rho)}{1+3(1-\rho)} * \frac{1-\rho}{1+2(1-\rho)} * \frac{1}{2-\rho} \\ &= \frac{2(1-\rho)^2}{(1+3(1-\rho))(1+2(1-\rho))} * \frac{1}{2-\rho} \end{aligned}$$

(b) Using Simon's estimate that

$$\rho_{est} = \frac{\#\text{ of groups of unique words}}{\text{total }\#\text{ of words}}$$

We find that $\rho_{est} = 31398/264706 \approx 0.11861$

(c)

k	**Theoretical**	**Empirical**
1	0.53152	0.56494
2	0.16957	0.15565
3	0.08202	0.07137

Question 6.

Repeat the preceding analyses for Ulysses for Jane Austen's "Pride and Prejudice" and Alexandre Dumas' "Le comte de Monte-Cristo" (in the original French), working this time from the original texts.

Responses: Code here

Pride and Prejudice:

$\rho_{est} = 0.097997$

k	theory	empirical
1	0.525762	0.529626
2	0.169129	0.149258
3	0.082328	0.076959

Comte de Monte Cristo:

$\rho_{est} = 0.163118$

k	theory	empirical
1	0.544401	0.616629
2	0.170396	0.151442
3	0.081239	0.063460