Principles of Complex Systems, CSYS/MATH 300
University of Vermont, Spring 2013
Assignment 4 • code name: Dumbledore

Dispersed: Monday, February 11, 2013.
Due: By start of lecture, 11:30 am, Thursday, February 21, 2013.
Some useful reminders:
Instructor: Peter Dodds
Office: Farrell Hall, second floor, Trinity Campus
E-mail: peter.dodds@uvm.edu
Office hours: 1:00 pm to 4:00 pm, Wednesday
Course website: http://www.uvm.edu/∼pdodds/teaching/courses/2013-01UVM-300

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

Graduate students are requested to use LaTeX (or related TeX variant).

---

1. (3 + 3 points) *Zipfarama via Optimization:*

   Complete the Mandelbrotian derivation of Zipf's law by minimizing the function

   $$\Psi(p_1, p_2, \ldots, p_n) = F(p_1, p_2, \ldots, p_n) + \lambda G(p_1, p_2, \ldots, p_n)$$

   where the 'cost over information' function is

   $$F(p_1, p_2, \ldots, p_n) = \frac{C}{H} = \frac{\sum_{i=1}^{n} p_i \ln(i+a)}{-g \sum_{i=1}^{n} p_i \ln p_i}$$

   and the constraint function is

   $$G(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i - 1 = 0$$

   to find

   $$p_j = (j+a)^{-\alpha}$$

   where $\alpha = H/gC$.

   3 points: When finding λ, find an expression connecting λ, g, C, and H.

   Hint: one way may be to substitute the form you find for $\ln p_i$ into H's definition (but do not replace p_i).

   Note: We have now allowed the cost factor to be $(j+a)$ rather than $(j+1)$.

2. (3 + 3)

   (a) For $n \to \infty$, use some computation tool (e.g., Matlab, an abacus, but not a clever friend who's really into computers) to determine that $\alpha \simeq 1.73$ for $a = 1$. (Recall: we expect $\alpha < 1$ for $\gamma > 2$)

   (b) For finite n, find an approximate estimate of a in terms of n that yields $\alpha = 1$.

   (Hint: use an integral approximation for the relevant sum.)

   What happens to a as $n \to \infty$?

3. (3 + 3)

   Consider a set of N samples, randomly chosen according to the probability distribution $P_k = ck^{-\gamma}$ where $k \geq 1$ and $2 < \gamma < 3$. (Note that k is discrete rather than continuous.)

   (a) Estimate $\min k_{\max}$, the approximate minimum of the largest sample in the network, 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.)

   (Extra hint: `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.

   For language, this scaling is known as Heap's law.

4. (3 + 3)

   Let's see how well your answer for the previous question works.

   For $\gamma = 5/2$, generate $n = 1000$ sets each of $N = 10$, 10^2, 10^3, 10^4, 10^5, and 10^6 samples, using $P_k = ck^{-5/2}$ with $k = 1, 2, 3, \ldots$

   Question: how do we computationally sample from a discrete probability distribution?

   (a) For each value of N, plot the maximum value of the $n = 1000$ samples as a function of sample number. These plots should give a sense of the unevenness of the maximum value of k, a feature of power-law size distributions.

   (b) For each set, find the maximum value. Then find the average maximum value for each N. Plot $\langle k_{\max} \rangle$ as a function of N and calculate the scaling using least squares.