PoCS What's The Story?

**Principles of Complex Systems, Vols. 1, 2, & 3D**
**CSYS/MATH 6701, 6713, & a pretend number**
**University of Vermont, Fall 2023**
**Assignment 09**
**"Add Eat, Pray, Love soundtrack to workout mix"**

**Due:** Sunday, November 12, by 11:59 pm
https://pdodds.w3.uvm.edu/teaching/courses/2023-2024pocsverse/assignments/09/
*Some useful reminders:*
**Deliverator:** Prof. Peter Sheridan Dodds (contact through Teams)
**Assistant Deliverator:** Chris O'Neil (contact through Teams)
**Office:** The Ether
**Office hours:** See Teams calendar
**Course website:** https://pdodds.w3.uvm.edu/teaching/courses/2023-2024pocsverse
**Overleaf:** LaTeX templates and settings for all assignments are available at https://www.overleaf.com/read/tsxfwwmwdgxj.

---

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

For coding, we recommend you improve your skills with Python, R, and/or Julia. The (evil) Deliverator uses (evil) 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 Brightspace or other preferred death vortex.

---

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

---

1. (3 + 3)

   You've earlier determined the theoretical scaling of the large sample of a power-law size distribution as a function of sample number.

   Let's see how well things match up with simulations.

   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 $k = 1, 2, 3, \ldots$

How do we computationally sample from a discrete probability distribution?

Note: We examined some of these in class. See slides on power-law size distributions.

Perishing Monk Hint: You can use a continuum approximation to speed things up. See below.

(a) For each value of sample size $N$, sequentially create $n$ sets of $N$ samples. For each set, determine and record the maximum value of the set's $N$ samples. (You can discard each set once you have found the maximum sample.)

You should have $k_{\mathrm{max},i}$ for $i = 1, 2, \ldots, n$ where $i$ is the set number. For each $N$, plot the $n$ values of $k_{\mathrm{max},i}$ as a function of $i$.

If you think of $n$ as time $t$, you will be plotting a kind of time series.

These plots should give a sense of the unevenness of the maximum value of $k$, a feature of power-law size distributions.

(b) Now find the average maximum value $\langle k_{\mathrm{max}} \rangle$ for each $N$.

The steps again here are:

1. Sample $N$ times from $P_k$;
2. Determine the maximum of the sample, $k_{\mathrm{max}}$;
3. Repeat steps 1 and 2 a total of $n$ times and take the average of the $n$ values of $k_{\mathrm{max}}$ you have obtained.

Plot $\langle k_{\mathrm{max}} \rangle$ as a function of $N$ on double logarithmic axes, and calculate the scaling using least squares. Report error estimates.

Does your scaling match up with your theoretical estimate for $\gamma = 5/2$?

How to sample from your power law distribution (and similarly upsetting things):

We now turn our problem of randomly selecting from this distribution into randomly selecting from the uniform distribution. After playing around a little, $k = 10^6$ seems like a good upper limit for the number of samples we're talking about.

Using Matlab (or some ghastly alternative), we create a cdf for $P_k$ for $k = 1, 2, \ldots, 10^6$ and one final entry $k > 10^6$ (for which the cdf will be 1).

We generate a random number $x$ and find the value of $k$ for which the cdf is the first to meet or exceed $x$. This gives us our sample $k$ according to $P_k$ and we repeat as needed. We would use the exactly normalized $P_k = \frac{1}{\zeta(5/2)} k^{-5/2}$ where $\zeta$ is the Riemann zeta function.

Now, we can use a rough method by approximating $P_k$ with a continuous function $P(z) = (\gamma - 1) z^{-\gamma}$ for $z \geq 1$ (we have used the normalization coefficient found in assignment 1 for $a = 1$ and $b = \infty$). Writing $F(z)$ as the cdf for $P(z)$, we have

$F(z) = 1 - z^{-(\gamma-1)} = 1 - z^{-3/2}$. Inverting, we obtain $z = [1 - F(z)]^{-1/(\gamma-1)} = [1 - F(z)]^{-2/3}$. We replace $F(z)$ with our random number $x$ and round the value of $z$ to finally get an estimate of $k$.

In sum, given $x$ is distributed uniformly on $[0, 1]$, then

$$k = \left[(1 - x)^{-2/3}\right]$$

is distributed according to a power-law size distribution $P_k = ck^{-5/2}$ where $[\,\cdot\,]$ indicates rounding to the nearest integer.

2. (3 + 3)

   Repeat the preceding question for $\gamma = 3/2$.

   As $1 < \gamma < 2$, we should see a very different behavior.

   Here's the question reprinted with $\gamma$ switched to 3/2.

   The key change in the question is in the form of $F(z)$ (last paragraph).

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

   How do we computationally sample from a discrete probability distribution?

   Hint: You can use a continuum approximation to speed things up. In fact, taking the exact continuum version from the first two assignments will work.

   (a) For each value of sample size $N$, sequentially create $n$ sets of $N$ samples. For each set, determine and record the maximum value of the set's $N$ samples. (You can discard each set once you have found the maximum sample.)

   You should have $k_{\max,i}$ for $i = 1, 2, \ldots, n$ where $i$ is the set number. For each $N$, plot the $n$ values of $k_{\max,i}$ as a function of $i$.

   If you think of $n$ as time $t$, you will be plotting a kind of time series.

   These plots should give a sense of the unevenness of the maximum value of $k$, a feature of power-law size distributions.

   (b) Now find the average maximum value $\langle k_{\max} \rangle$ for each $N$.

   The steps again here are:
   1. Sample $N$ times from $P_k$;
   2. Determine the maximum of the sample, $k_{\max}$;
   3. Repeat steps 1 and 2 a total of $n$ times and take the average of the $n$ values of $k_{\max}$ you have obtained.

   Plot $\langle k_{\max} \rangle$ as a function of $N$ on double logarithmic axes, and calculate the scaling using least squares. Report error estimates.

   Does your scaling match up with your theoretical estimate for $\gamma = 3/2$?

How to sample from your power law distribution (and similar kinds of beasts):

We now turn our problem of randomly selecting from this distribution into randomly selecting from the uniform distribution. After playing around a little, $k = 10^6$ seems like a good upper limit for the number of samples we're talking about.

Using Matlab (or some ghastly alternative), we create a cdf for $P_k$ for $k = 1, 2, \ldots, 10^6$ and one final entry $k > 10^6$ (for which the cdf will be 1).

We generate a random number $x$ and find the value of $k$ for which the cdf is the first to meet or exceed $x$. This gives us our sample $k$ according to $P_k$ and we repeat as needed. We would use the exactly normalized $P_k = \frac{1}{\zeta(3/2)} k^{-3/2}$ where $\zeta$ is the Riemann zeta function.

Now, we can use a quick and dirty method by approximating $P_k$ with a continuous function $P(z) = (\gamma - 1)z^{-\gamma}$ for $z \geq 1$ (we have used the normalization coefficient found in assignment 1 for $a = 1$ and $b = \infty$). Writing $F(z)$ as the cdf for $P(z)$, we have $F(z) = 1 - z^{-(\gamma-1)} = 1 - z^{-1/2}$. Inverting, we obtain $z = [1 - F(z)]^{-1/(\gamma-1)} = [1 - F(z)]^{-2}$.

We now replace $F(z)$ with our random number $x$ and round the value of $z$ to finally get an estimate of $k$.

In sum, given $x$ is distributed uniformly on $[0, 1]$, then

$$k = \left[(1 - x)^{-2}\right]$$

is distributed according to a power-law size distribution $P_k = ck^{-2}$ where $[\ \cdot\ ]$ indicates rounding to the nearest integer.