CELLULAR AND SYSTEMS MODELING - HW3

Student: Shu-Ting Cho (shc167@pitt.edu)

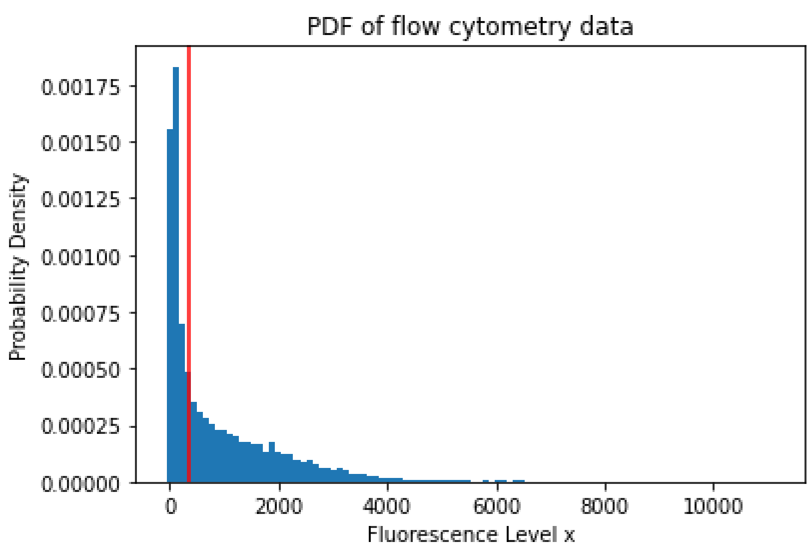
1. (10 points) Problem 5.6. (You don't need to write up problem 5.5 but do consider potential effects of log transforming the data as you are doing.)

5.6 *Flow cytometry data*

First work Problem 5.5; then obtain Dataset 8.

A. Erez and coauthors activated many individual immune system cells (mouse T lymphocytes), treated them with a drug of interest, and then exposed them to other cells that would normally stimulate them (antigen presenting cells). One actor in the response mechanism was fluorescently labeled in the cells (ppERK), and the fluorescence levels were found for each cell individually (**flow cytometry**). Each row in the dataset gives the fluorescence intensity x in arbitrary units, and hence is proportional to the population of the response molecule in a single cell.

a. Make a histogram of x. Rescale it to obtain an estimate of the PDF. How many maxima does it have? Where are they located? Also indicate on your graph the location of the median (Problem 5.1).



One maximum

At the bin with its center = 41.44811577

Median = 353.0818785

CELLULAR AND SYSTEMS MODELING - HW3

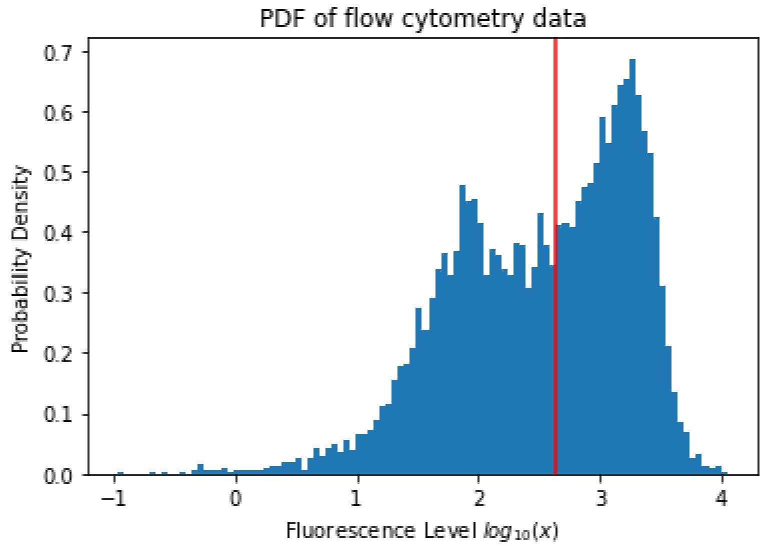
Student: Shu-Ting Cho (shc167@pitt.edu)

b. Data of this sort are very often presented logarithmically. Make a histogram of $\log_{10}(x)$. Rescale it to obtain an estimate of the PDF. How many maxima does it have? Where are they located? Are there any data points that cannot be represented in this way? Also indicate on your graph the location of the median and comment.

Two maxima

3.2447199838175247 and 1.8421976

There are 569 negative values in the input data that cannot take log value.



Median = 2.6376633355495307

The new median is located between two local maxima, and the $10^{2.6377}$ ~= 434.1735 is higher than the median(x) = 353.0819. This is because the negative values were removed when we presented the data logarithmically.

c. If your graphs appear qualitatively different, what could be the reason? Do you think there were two distinct subpopulations of cells?

The log transformation reduces the skewness of the original data (high frequency for x values lower than the median). It becomes more “normal” distributed.

No, the second bump with the peak at ~3.24 appears in the log-transformed plot is because of the long tail of the original data. The second bump has x from ~353 to ~10^4, which are the values higher than the median in the part a graph, thus is not a subpopulation.

CELLULAR AND SYSTEMS MODELING - HW3

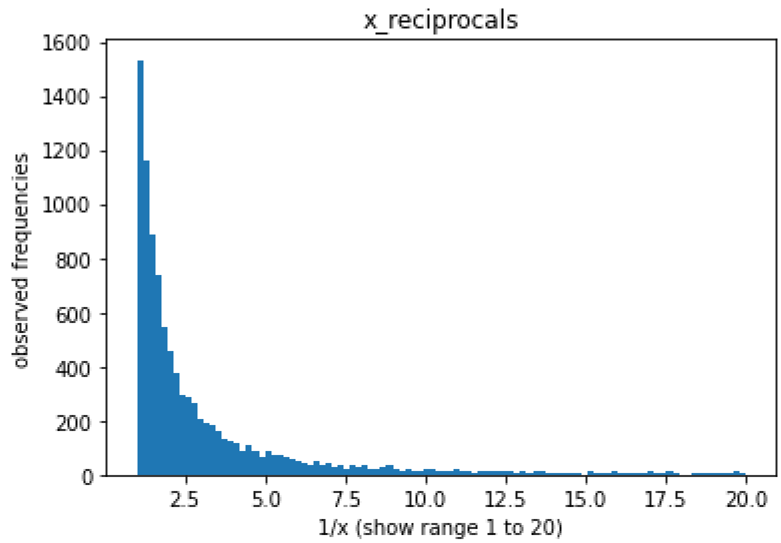
Student: Shu-Ting Cho (shc167@pitt.edu)

2. (10 points) Problem 5.7

5.7 *Simulation via transformation*

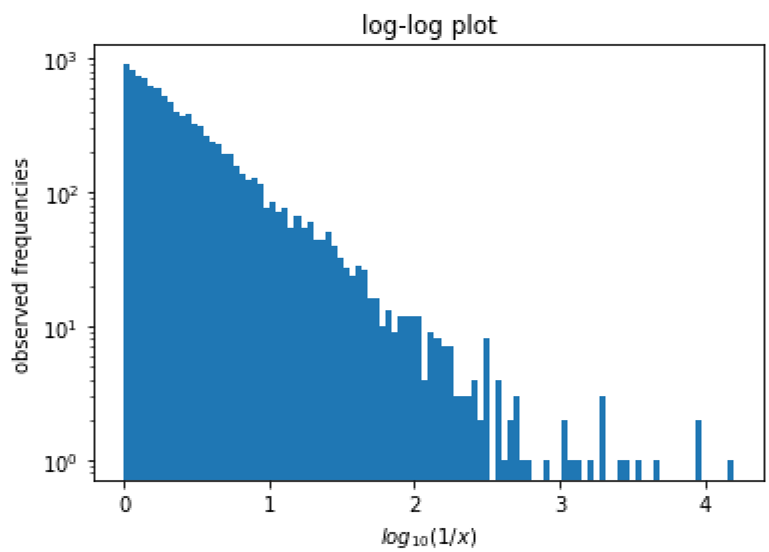
Section 5.2.7 explained how to take Uniformly distributed random numbers supplied by a computer and convert them into a modified series with a more interesting distribution.

a. As an example of this procedure, generate 10 000 Uniformly distributed real numbers x between 0 and 1, find their reciprocals $1/x$, and make a histogram of the results. (You'll need to make some decisions about what range of values to histogram and how many bins to use in that range.)



Bins=100, range=1~20

b. The character of this distribution is revealed most clearly if the counts in each bin (observed frequencies) are presented as a log-log plot, so make such a graph.



CELLULAR AND SYSTEMS MODELING - HW3

Student: Shu-Ting Cho (shc167@pitt.edu)

c. Draw a conclusion about this distribution by inspecting your graph. Explain mathematically why you got this result.

The y values of the log-log plot are linearly descent from the lowest x, with a slope ~ -1.

The distribution of observe frequency of 1/x is a power-law distribution:
$y = 1840.7x^{-2.023}$
Transforming a power-law distribution to log-log plot makes the distribution become linear.

The original function: $y = x^{-1}$ (power relation)
Taking $\log_{10}$: $log_{10}y = -1log_{10}x$ (linear function of $\log_{10}x$)

The original distribution: $y = 1840.7x^{-2.023}$ (power relation)
Taking $\log_{10}$ on both axes: $log_{10}y = 3.265 - 2.023log_{10}x$ (linear function of $\log_{10}x$)
The log-log graph is therefore a straight line.

d. Comment on the high-x (lower right) end of your graph. Does it look messy? What happens if you replace 10 000 by, say, 50 000 samples?

Yes, there are several values observed at $\log_{10}(1/x) > 3$, which mess up the linear descendence of the y values.

It will make the log-log histogram more smooth on the high-x end of the graph.

CELLULAR AND SYSTEMS MODELING - HW3

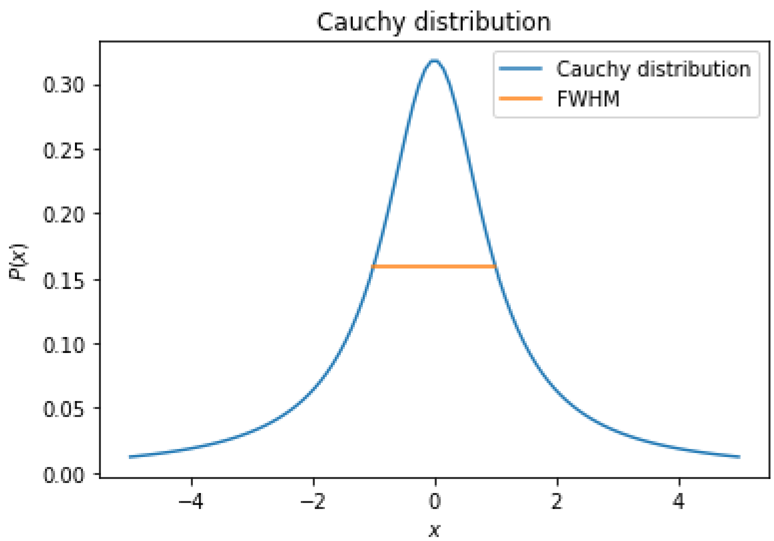
Student: Shu-Ting Cho (shc167@pitt.edu)

3. (5 points) Problem 5.17 (a-e). (You don't need to write up problem 5.3

5.17 *Tail probabilities*

In this problem, you will explore the probability of obtaining extreme ("tail") values from a Gaussian versus a Power-law distribution. "Typical" draws from a Gaussian distribution produce values within about one standard deviation of the expectation, whereas Power-law distributions are more likely to generate large deviations. Your job is to make this intuition more precise. First work Problem 5.3 if you haven't already done so.

a. Make a graph of the Cauchy distribution with $\mu_x = 0$ and $\eta = 1$. Its variance is infinite, but we can still quantify the width of its central peak by the full width at half maximum (FWHM, Figure 5.2, page 108), defined as twice the value of x at which $\wp$ equals $\frac{1}{2}\wp(0)$. Find this value.



$$\frac{A}{1+x^2} = \frac{1}{2}P(0) = \frac{A}{2}$$

$$\Rightarrow x^2 = 1 \Rightarrow x = \pm 1$$

$$\Rightarrow FWHM = 2 * 1 = 2$$

Student: Shu-Ting Cho (shc167@pitt.edu)

b. Calculate the FWHM for a Gaussian distribution with standard deviation σ. What value of σ gives the same FWHM as the Cauchy distribution in (a)? Add a graph of the Gaussian with this σ (and expectation equal to zero) to your graph, and comment.

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = \frac{1}{2}P(0) = \frac{1}{\sigma\sqrt{2\pi}} \times \frac{1}{2}$$

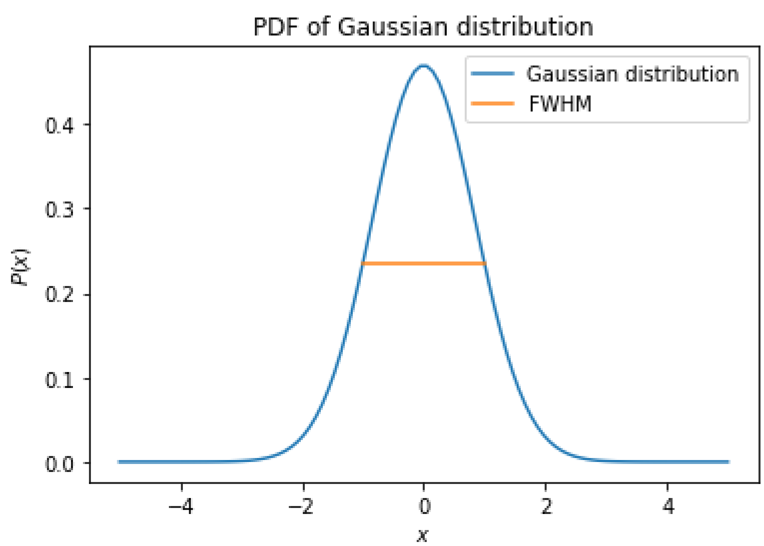
$$\Rightarrow e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = \frac{1}{2}$$

$$same\ FWHM \Rightarrow let\ x = \pm\ 1$$

$$\Rightarrow e^{-\frac{1}{2\sigma^2}} = \frac{1}{2}$$

$$\Rightarrow \frac{1}{\sigma^2} = -\ 2 \times ln\frac{1}{2}$$

$$\Rightarrow \sigma = \frac{1}{\sqrt{2 \times ln2}} = 0.8493218003$$



With the same FWHM (shown as orange line in the graph), Gaussian has a higher P(μ) with narrower bump.

Student: Shu-Ting Cho (shc167@pitt.edu)

c. For the Cauchy distribution, calculate $\mathcal{P}(|x| > \mathrm{FWHM}/2)$.

$$FWHM/2 = 1$$

$$P(|x| > 1) = 2 \times P(x > 1)$$

$$P(|x| < 1) = \int_{-1}^{1} \frac{A}{1+x^2} = \int_{-1}^{1} \frac{1}{\pi(1+x^2)} = 0.5$$

$$P(|x| > FWHM/2) = 0.5$$

d. Repeat (c) for the Gaussian distribution you found in (b). You will need to do this calculation numerically, either by integrating the Gaussian distribution or by learning about, and then computing, the "error function."

$$P(x < 1) = \int_{-1}^{1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = 0.760968108543862$$

$$P(|x| > FWHM/2) = 0.239031891456138$$

e. Repeat (c) and (d) for more extreme events, with $|x| > \frac{3}{2}\mathrm{FWHM}$.

$$\frac{3}{2} FWHM = 3$$

 # Cauchy distribution:

$$P(|x| < 3) = \int_{-3}^{3} \frac{1}{\pi(1+x^2)} = 0.7951672353008665$$

$$P(|x| > \frac{3}{2} FWHM) = 0.2048327646991335$$

 # Gaussian distribution:

$$P(|x| < 3) = \int_{-3}^{3} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = 0.9995879293201379$$

$$P(|x| > \frac{3}{2} FWHM) = 0.0004120706798621$$

Student: Shu-Ting Cho (shc167@pitt.edu)

4. (10 points) Problem 6.2

6.2 *Motion in a trap*

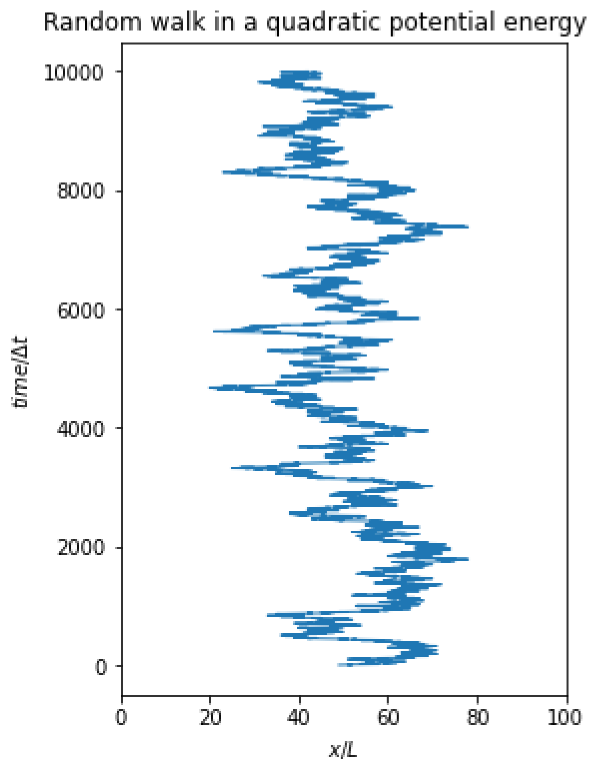
Write a computer code to generate data like that shown in Figure 6.2a.

a. Specifically, use the potential energy function in Equation 6.3. But forbid the particle from leaving the range $0 \le x \le 100L$, as follows: If $x = 0$, set $\mathcal{P}_- = 0$; and if $x = 99L$, set $\mathcal{P}_+ = 0$. Physically, we can imagine "hard walls" ($U = \infty$) at these locations. As in Problem 6.1, release every walker from the center position.

$$U(x) = 0.0025\zeta D\left(\frac{x}{L} - 50\right)^2. \quad (6.3)$$

$$\Delta U = U(x+1) - U(x-1) = 0.0025\zeta D\left[(x+1-50)^2 - (x-1-50)^2\right]$$

$$\mathcal{P}_+ = \frac{1}{2}\left(1 - \frac{\Delta U}{2\zeta D}\right). \quad (6.2)$$



b. $\boxed{T_2}$ Make an animated graphic of a representative trajectory.

Please see the attached "Q6.2b.mp4" file for the animation.

CELLULAR AND SYSTEMS MODELING - HW3

Student: Shu-Ting Cho (shc167@pitt.edu)

5. (15 points) Problem 6.3

6.3 *Equilibrium distribution in a trap*

Write a computer code to generate data like that shown in Figures 6.3a;b. Specifically, model 10 000 trajectories, each with 50 000 steps and each starting from the center position. As in the preceding problem, use the potential energy function in Equation 6.3 and implement hard walls at the ends. It may start to get computationally intensive to simulate half a billion steps! But here is a time-saving trick.

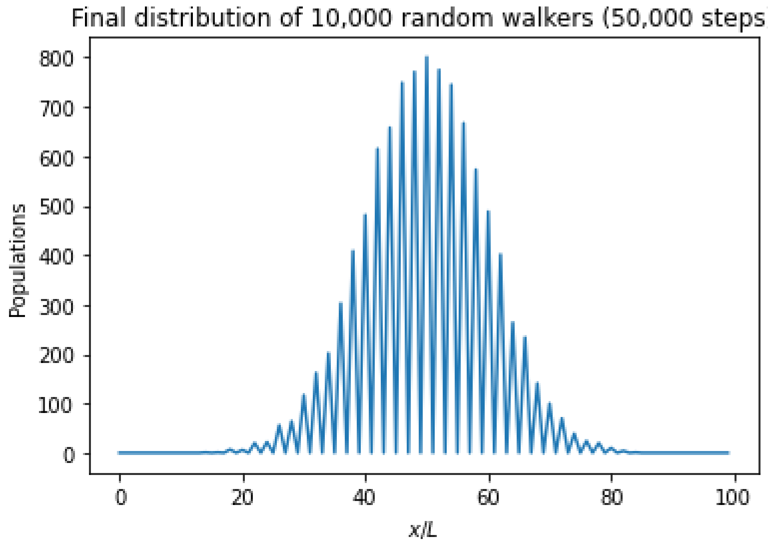
For this problem, don't attempt to follow the individual trajectories. All we need are the *populations* at each spatial position, for each time. Thus, your code need only retain an array containing those populations. Also, the problem has the Markov property that each walker's next step depends only on its current position, not on its past history. So you can proceed as follows:

- For each time step, first make a new array to hold the populations at the next time step.
- Next, consider each location ($k = x/L = 0, \ldots, 99$) in turn. If the population at position k is not zero, then it will get partitioned into a subpopulation stepping right, with probability $\mathcal{P}_+(k)$, and those stepping left, with probability $1 - \mathcal{P}_+$. The partitioning is random, but Section 4.2.2 (page 73) argued that it follows a Binomial distribution. Thus, a *single* draw from the appropriate Binomial will establish the fates of *all* walkers currently at k.[19]
- Use the preceding result to update the new populations at $k-1$ and $k+1$, respectively. Step through all k values.
- Copy the updated populations into the main population counter array and repeat for the desired number of time steps.

CELLULAR AND SYSTEMS MODELING - HW3

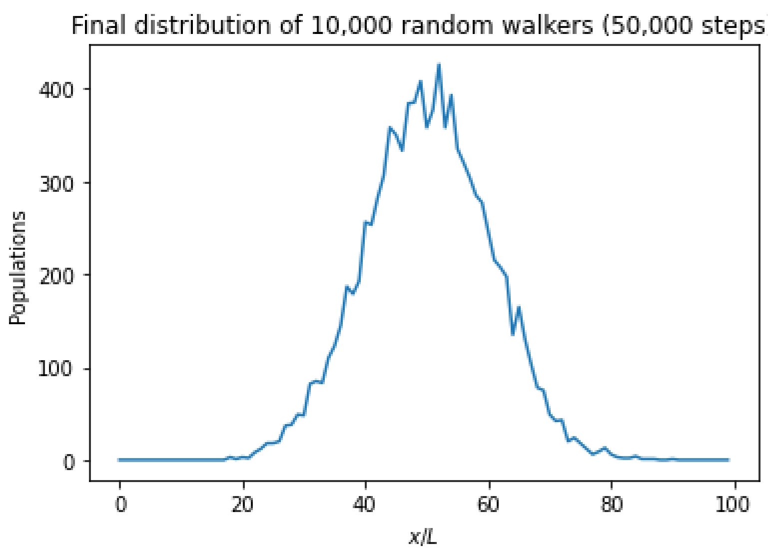
Student: Shu-Ting Cho (shc167@pitt.edu)

a. Carry out the above steps and show the final distribution.



b. If you released all the walkers exactly at $x = 50L$, then your graph will have an unpleasant jagged character. Why did that happen? Try releasing just half of the walkers at 50 and the the other half at 49. Why does that help?

At each time step, all walkers will leave their current position x and all walkers start from the same position, so the distribution array will always be [...., N_{x1}, 0, N_{x3}, 0, N_{x5}, 0, …].



Because half of the walkers start from x=50 and the other half walkers start from the adjacent position x=49, they will be like two subgroups of walkers with similar distribution. Thus, they complement each other's "0" and make the distribution more smooth.

CELLULAR AND SYSTEMS MODELING - HW3

Student: Shu-Ting Cho (shc167@pitt.edu)

c. $\boxed{T_2}$ Make an animated graphic of the time development of the probability distribution, estimated from your finite sample. How quickly does it reach nearly equilibrium form? How much does it then jitter around that form?

It reaches nearly equilibrium form after around 150 steps.
The probabilities remains a normal distribution with very little variation (< 1%) to the the end of simulation.

Please see the attached “Q6.3c.y55.mp4” and “Q6.3c.y10.mp4” files for the animations.
“Q6.3c.y55.mp4” and “Q6.3c.y10.mp4” set ylim to 55 and 10, respectively.

Because the distribution reaches its equilibrium around 150 steps, and showing 50,000 steps with fps=24 will make the animation longer than half an hour, which make this file too long and too big. Therefore, I only output the first 300 steps(frames) as a representative animated graphic.