 HW 4 – Due start of class 10/14

Homework is due at the start of class on 10/14. Please turn in a hardcopy. It can be handwritten. Please show all computations.

Reading from your textbook: Chapter 20. Remember that when reading you should cover all the "Answers" and derivations and try them yourself first.

Exercises:

1. Exercise 20.2 from your textbook. You can use an integration package for this exercise, or else do it by hand (it's not hard).
   In this exercise, you are asked to compute the fraction of load, q, made up by just the top (largest) 1% of all jobs. Note: For non-negative job size distribution, X, with density function $f_X(\cdot)$, we define q as follows:
   $$q = \frac{\int_{[t \text{ in top } 1\%]} t f_X(t) dt}{\int_0^\infty t f_X(t) dt} .$$

2. This problem asks you to study "jobs" in your research. Because this is not just an academic exercise, you may find it hard to do some parts of this problem. Please get help as needed. Start early because it will take time to collect the data.
   (a) Describe the setting that you're working in. Explain what is meant by a "job" for you. For example, is it a "read," or a "write" or a "compute" or something else?
   (b) Your overall goal is to understand S, which is a random variable representing a job's size (service time). To do this, start by measuring the size of a *large* number of jobs. A job's size is how long it would take if it were running "alone" in the system (not queueing behind any other jobs). Ideally you should measure the sizes of 10^6 jobs or more. If you can't do this, just measure as many as you can.
   (c) What is $\mathbf{E}[S]$ based on your data? What is $\mathbf{Var}(S)$ based on your data? What is C_S^2 based on your data?
   (d) Make a plot like we did in class, where the y-axis shows $\mathbf{P}\{S > x\}$ and the x-axis shows x. Your plot can be on a linear scale or a log-log scale – whatever is clearest. Show your plot. You might also want to show your job sizes using a histogram if that helps.
   (e) Provide some description of your size distribution in words. Does it look like an Exponential? Like a Pareto? Does it look like a Uniform? Like a Normal? Is it a combination of several distributions, where say half the jobs are Normally-distributed around some small mean and the other half are Normally-distributed around some larger mean, but there are no jobs inbetween? Can you provide some explanation for why your distribution looks the way it does?

(f) What is the 50%-tile of your job size distribution (this is the median of those job sizes that you collected). What is the 90%-tile of your job size distribution (this is the point x where 10% of your jobs are bigger than x)? How about the 99%-tile?

(g) Determine the load made up by the top 1% of jobs. To do this, determine quantity q based on your samples:

$$q = \frac{\text{Sum of sizes of all jobs larger or equal to the 99\%-tile}}{\text{Sum of sizes of all jobs}} .$$

Clearly it should be the case that $q > 1\%$. However is it a lot higher than that? Does your distribution have a heavy tail?

(h) Explain any other interesting things that you have observed about your job sizes.