Problem Set 3

Collaborated with: Steven Brotz, David Kawashima, Roshan Bal, Jessica Choi, Abraham Hussain, Tristan Nee

Problem 1

a. The probability that a particular types a particular c- letter word is $(\frac{1-q}{n})^c q$. This is because we need c letters which each happen with probability $\frac{1-q}{n}$ followed by a space which occurs with probability q.

b. We first calculate the bounds for the ranks given that the word is length c. The lower and upper bounds are precisely $r_{low} = \sum_{i=0}^{c-1} n^i = \frac{n^c-1}{n-1}$ and $r_{high} = \sum_{i=0}^{c} n^i = \frac{n^{c+1}-1}{n-1}$. From the above lower and upper bounds, we can see that c can actually be bounded asymptotically. Specifically, $\log_n r - 1 \leq c \leq \log_n r$. Now, we calculate $\lim_{r\to\infty} \frac{\log P_r}{\log r}$ for both $c = r_{high}$ and $c = r_{low}$. We first start with $c = r_{high}$:

$$\lim_{r\to\infty} \frac{\log P_r}{\log r}$$

$$= \lim_{r\to\infty} \frac{\log((\frac{1-q}{n})^c q)}{\log r}$$

$$= \lim_{r\to\infty} \frac{\log((\frac{1-q}{n})^{\log_n r} q)}{\log r}$$

$$= \lim_{r\to\infty} \frac{\log q + \log((\frac{1-q}{n})^{\log_n r})}{\log r}$$

$$= \lim_{r\to\infty} \frac{\log q}{\log r} + \frac{\log((\frac{1-q}{n})^{\log_n r})}{\log r}$$

$$= \lim_{r\to\infty} \frac{\log((\frac{1-q}{n})^{\log_n r})}{\log r}$$

$$= \lim_{r\to\infty} \frac{\log_n r \log((\frac{1-q}{n})^)}{\log r}$$

$$= \lim_{r\to\infty} \frac{\frac{\log r}{\log n} \log((\frac{1-q}{n})^)}{\log r}$$

$$= \lim_{r\to\infty} \frac{\log(\frac{1-q}{n})}{\log n}$$

$$= \lim_{r\to\infty} \log_n \frac{1-q}{n}$$

$$= \log_n(\frac{1-q}{n})$$

Now, we repeat this process for $c = r_{low}$.

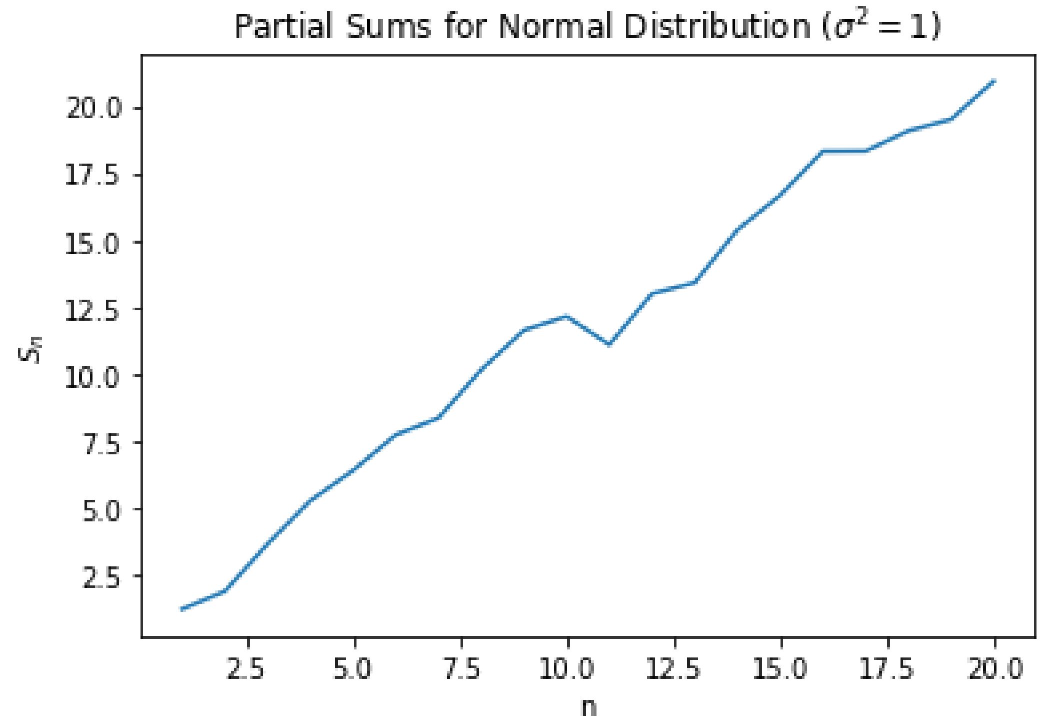
$$\begin{aligned}
&\lim_{r\to\infty} \frac{\log P_r}{\log r} \\
&= \lim_{r\to\infty} \frac{\log((\frac{1-q}{n})^c q)}{\log r} \\
&= \lim_{r\to\infty} \frac{\log((\frac{1-q}{n})^{\log_n r-1} q)}{\log r} \\
&= \lim_{r\to\infty} \frac{\log q + \log((\frac{1-q}{n})^{\log_n r-1})}{\log r} \\
&= \lim_{r\to\infty} \frac{\log q}{\log r} + \frac{\log((\frac{1-q}{n})^{\log_n r-1})}{\log r} \\
&= \lim_{r\to\infty} \frac{\log((\frac{1-q}{n})^{\log_n r-1})}{\log r} \\
&= \lim_{r\to\infty} \frac{(\log_n r-1)\log((\frac{1-q}{n})^)}{\log r} \\
&= \lim_{r\to\infty} \frac{(\log_n r)\log((\frac{1-q}{n})) - \log((\frac{1-q}{n}))}{\log r} \\
&= \lim_{r\to\infty} \frac{\frac{\log r}{\log n}\log((\frac{1-q}{n})^)}{\log r} - \frac{(\log(\frac{1-q}{n}))}{\log r} \\
&= \lim_{r\to\infty} \frac{\log(\frac{1-q}{n})}{\log n} \\
&= \lim_{r\to\infty} \log_n \frac{1-q}{n} \\
&= \log_n(\frac{1-q}{n})
\end{aligned}$$

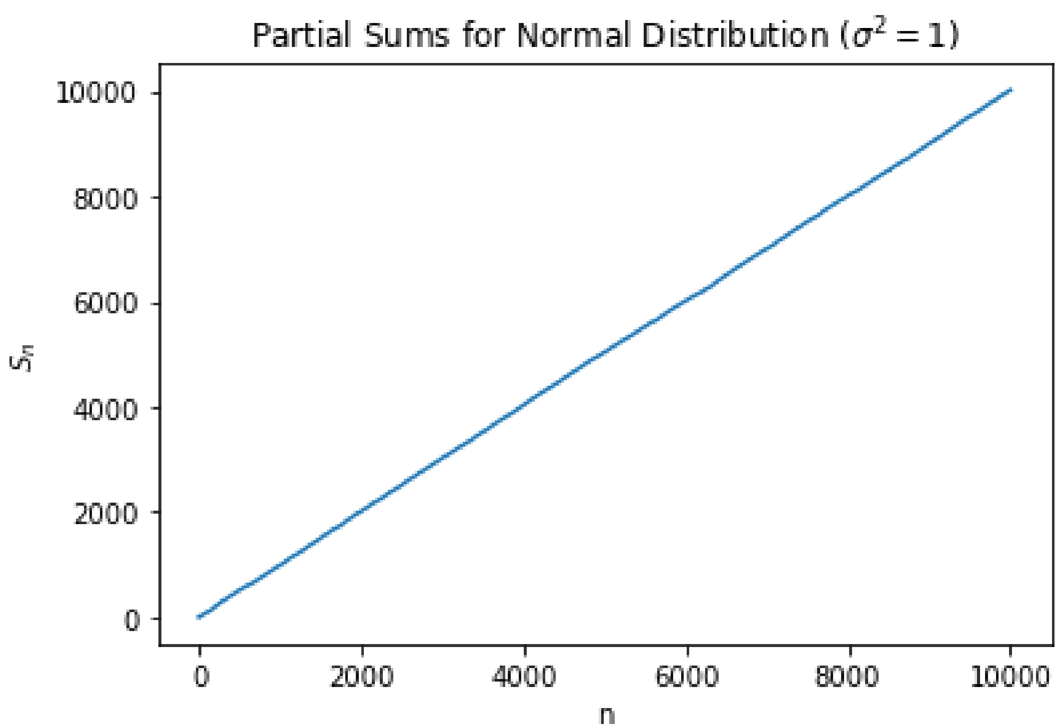
Thus, by the squeeze theorem, because the limits for both the upper and lower bounds converge to $\log_n(\frac{1-q}{n})$, then we know that $\lim_{r\to\infty} \frac{\log P_r}{\log r} = \log_n(\frac{1-q}{n})$, as desired.

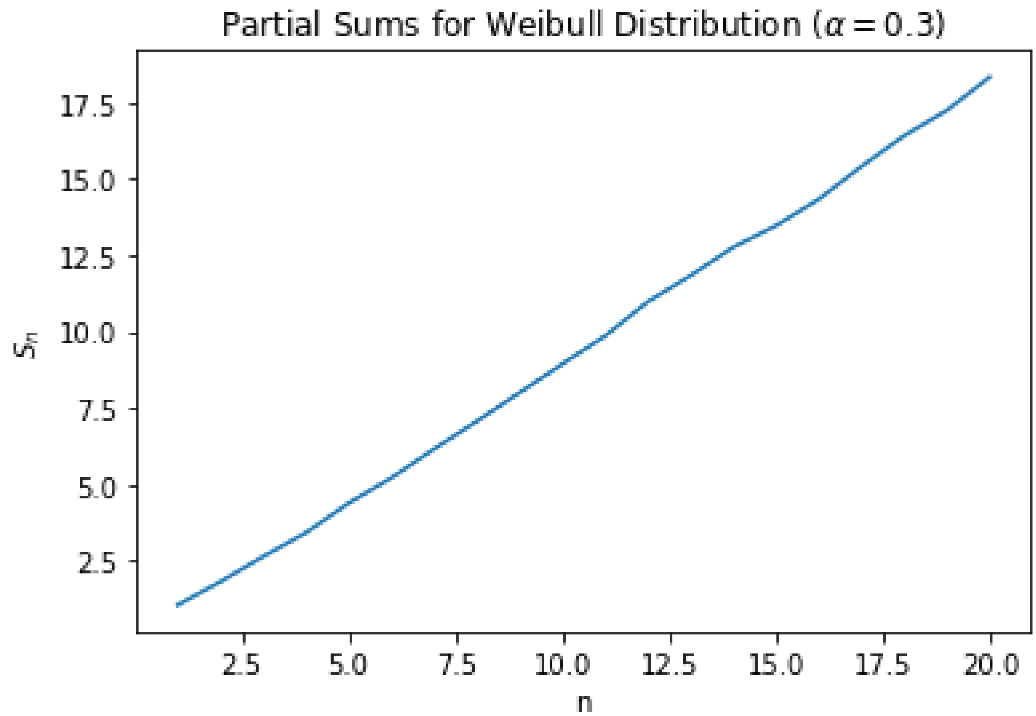
c. Looking at part b, we see that on a log log scale and fixing q and n, that P_r is linear. This suggests that the probability distribution is heavy tailed. If we look at the distribution from part a however, we see that it is exponential. Thus, we have a heavier tail for the distribution of part b. This suggests that the difference in probability between typing a $c-$letter word and a $(c+k)$ letter word $(k > 0)$ is greater than the difference in probability between typing a word with rank r and a word with rank $(r+h), h > 0$ for large h and k.

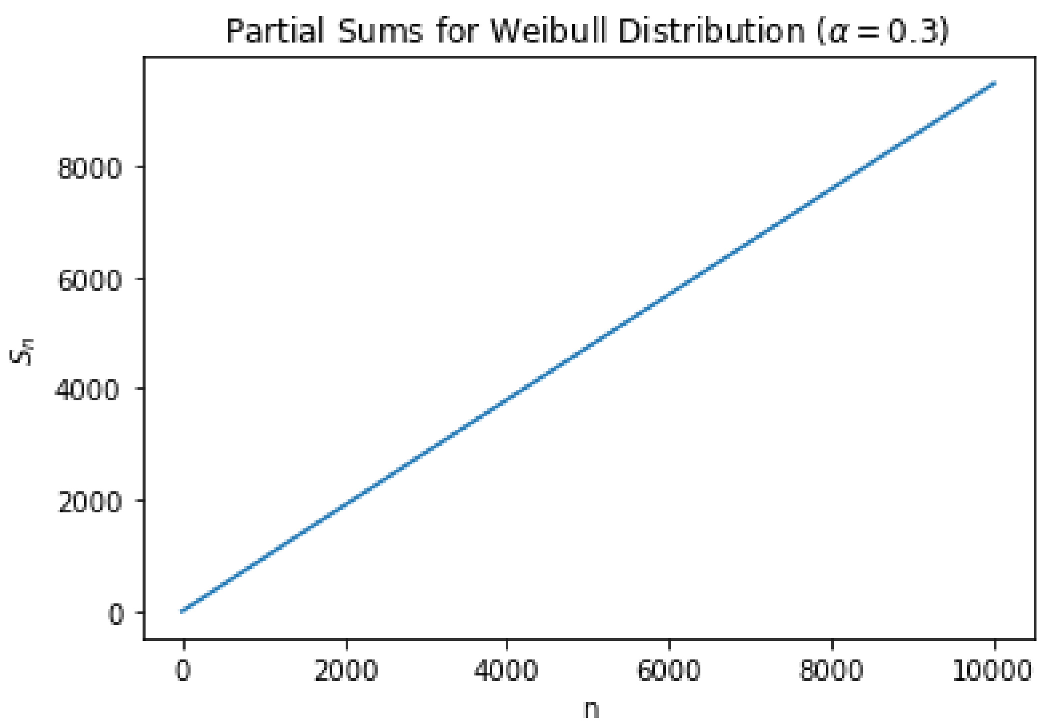
Problem 2

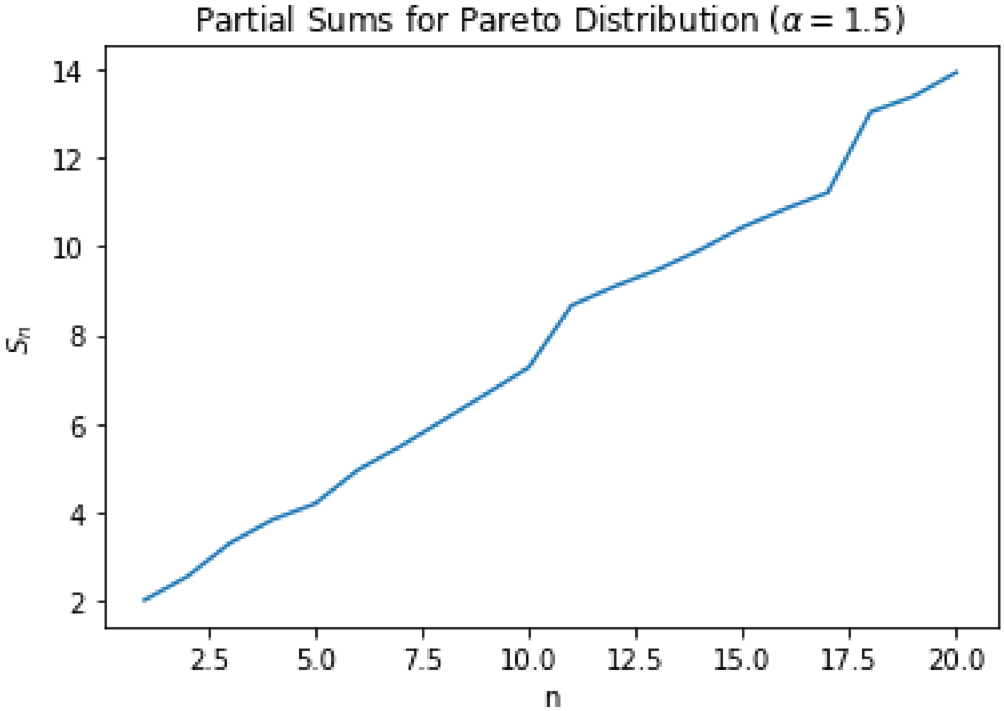
a.


Partial Sums for Normal Distribution ($\sigma^2 = 1$)
S_n
n
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5

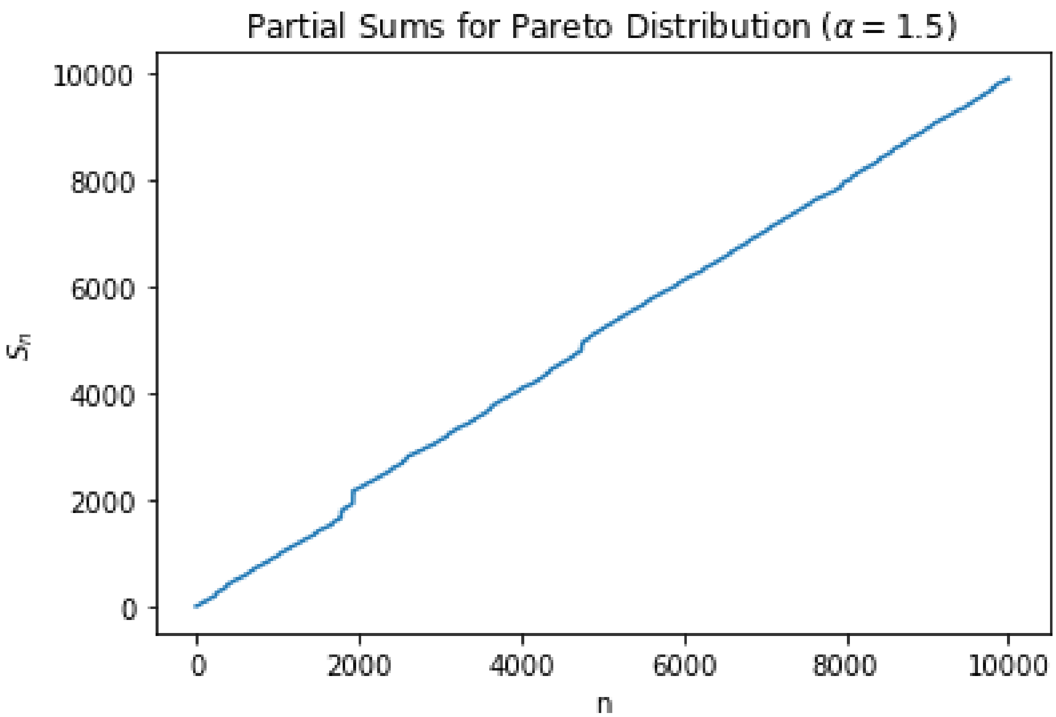

Partial Sums for Normal Distribution ($\sigma^2 = 1$)
S_n
n
10000
8000
6000
4000
2000
0


Partial Sums for Weibull Distribution ($\alpha = 0.3$)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
S_n
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
n

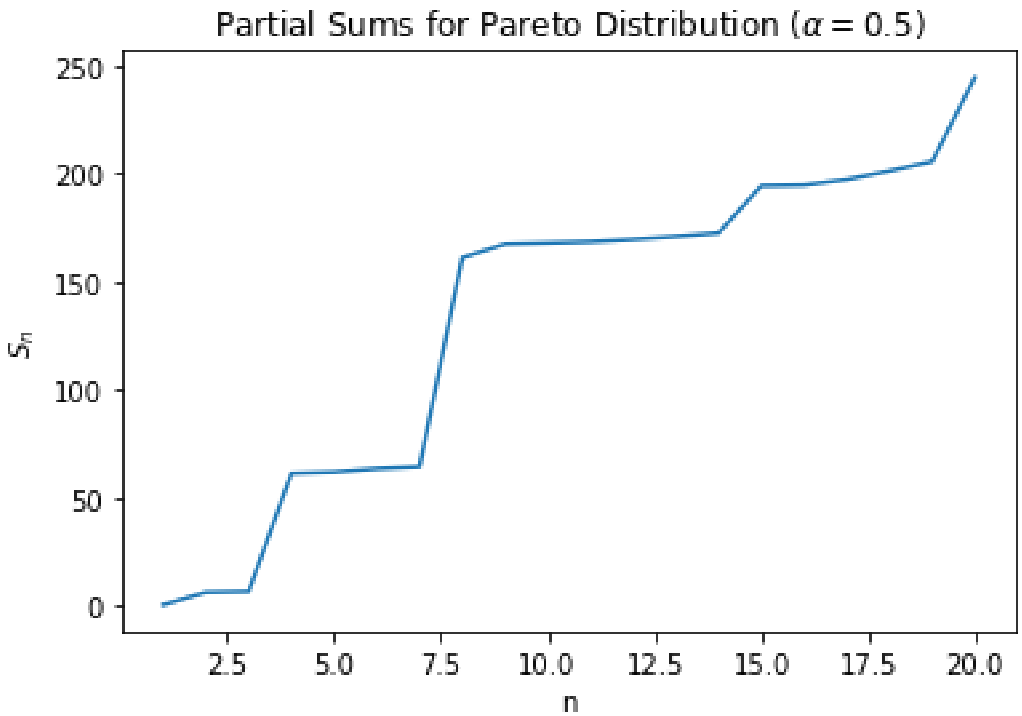

Partial Sums for Weibull Distribution ($\alpha = 0.3$)
8000
6000
4000
2000
0
S_n
0
2000
4000
6000
8000
10000
n

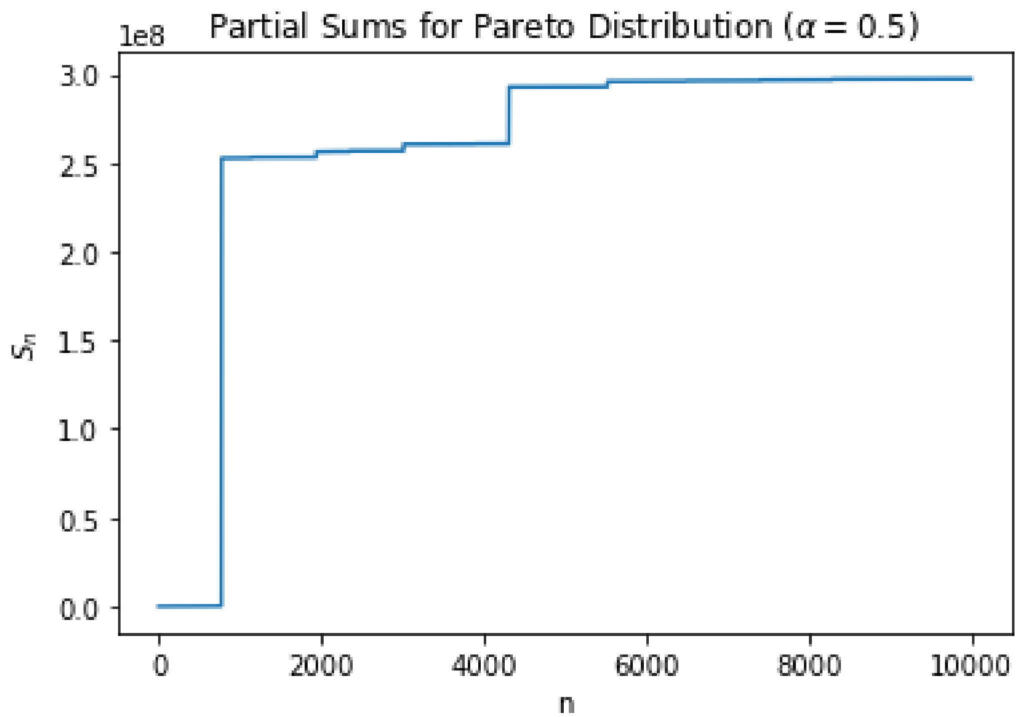


Partial Sums for Pareto Distribution ($\alpha = 1.5$)
S_n
n
2
4
6
8
10
12
14
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0



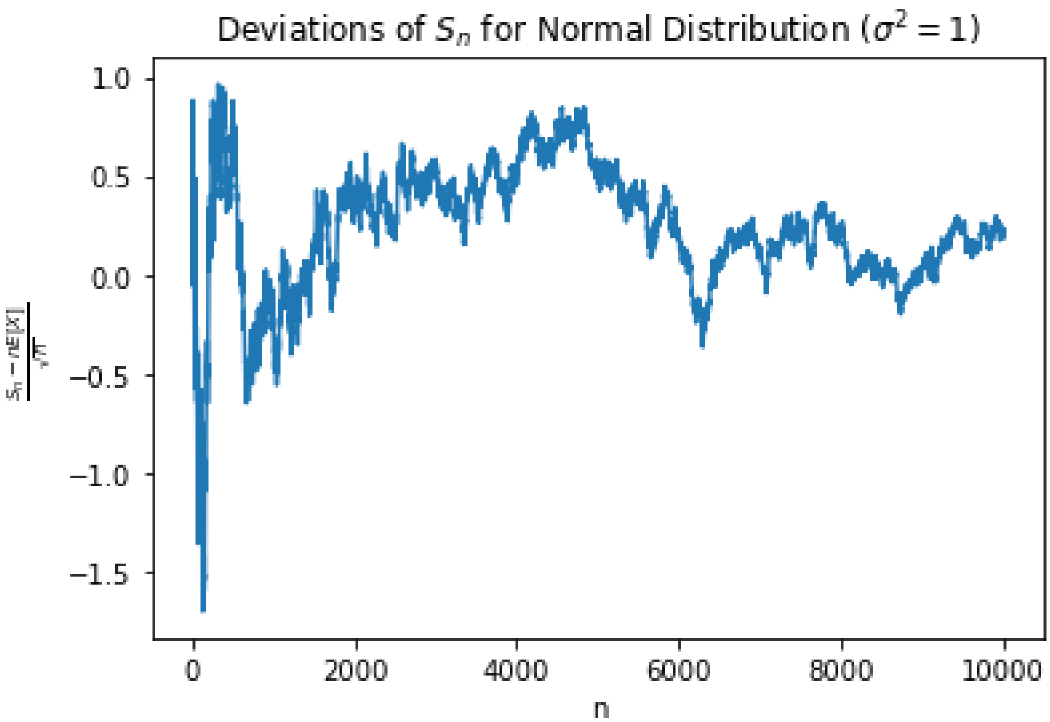
Partial Sums for Pareto Distribution ($\alpha = 1.5$)
S_n
n
0
2000
4000
6000
8000
10000

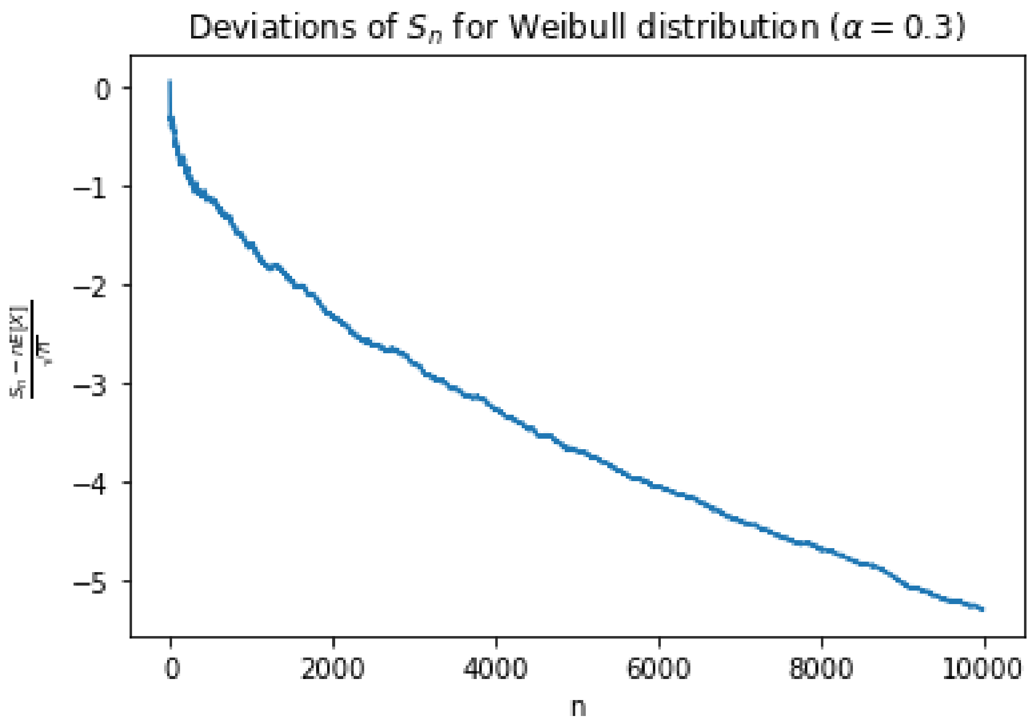


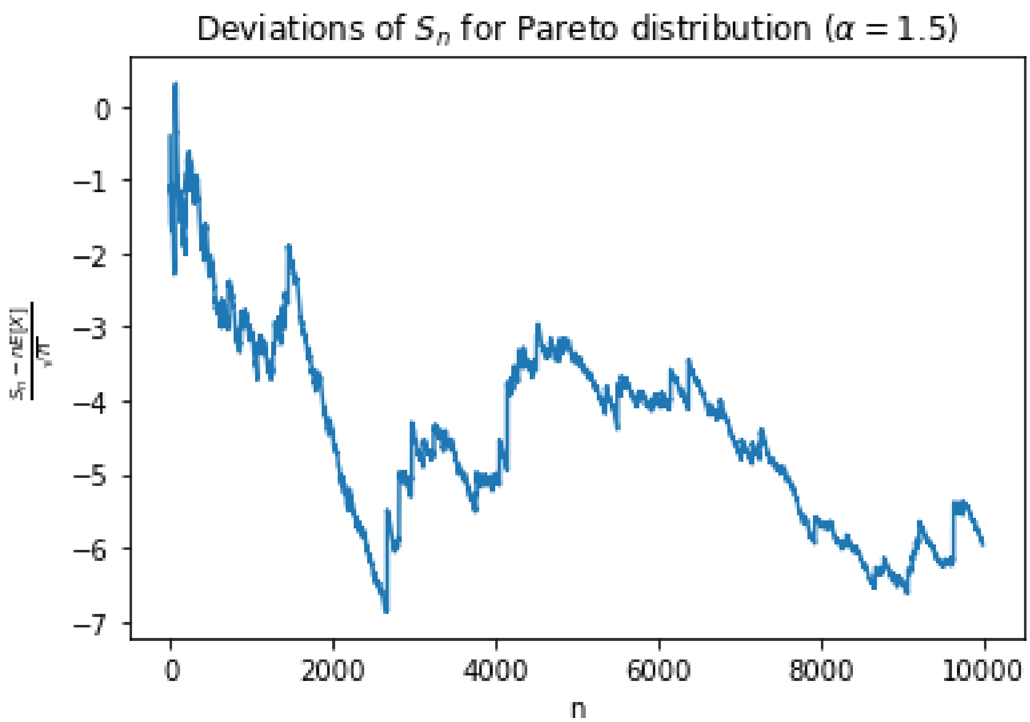


What sticks out here is the catastrophe vs. conspiracy of the different graphs. We see that for our light tailed distributions (normal and weibull), that we essentially have smooth, straight lines when we look at partial sums S_n vs n for our large n (10,000 data points). This reflects the conspiracy principle for light-tailed distributions, where certain samples may be slightly higher or lower than the mean, but it is highly unlikely to see a sample that is extremely high or low compared to the mean. However, if we look at heavy tailed distributions like our pareto, the same graph for large n is much more rugged, with certain jerks here and there. For our first pareto distribution ($\alpha = 1.5$), it may resemble a line at first glance, but notice at around n = 2000 and n = 5000, we have certain spikes in our S_n. This is especially pronounced in our second pareto distribution ($\alpha = 0.5$) where we have drastic jumps in partial sums. These traits reflect the catastrophe principle common to heavy-tailed distributions like the pareto. In such heavy-tailed distributions, it is not terribly uncommon to see extremely high sample values which, when calculating the partial sums, will cause a sudden spike in partial sum after coming across one of these catastrophe values. Within the context of law of large numbers, we see that law of large numbers holds for distributions with finite variance (weibull, normal) as the partial sum plots for large n are essentially straight but do not hold for the pareto distributions which have infinite variance. On a similar note, comparing $n = 20$ to $n = 10000$ for the weibull and normal, we see that we have a lot more variation in S_n for $n = 20$ than compared to $n = 10000$ which supports the law of large numbers (as n gets bigger and bigger, the sample mean approaches the population mean).

b.

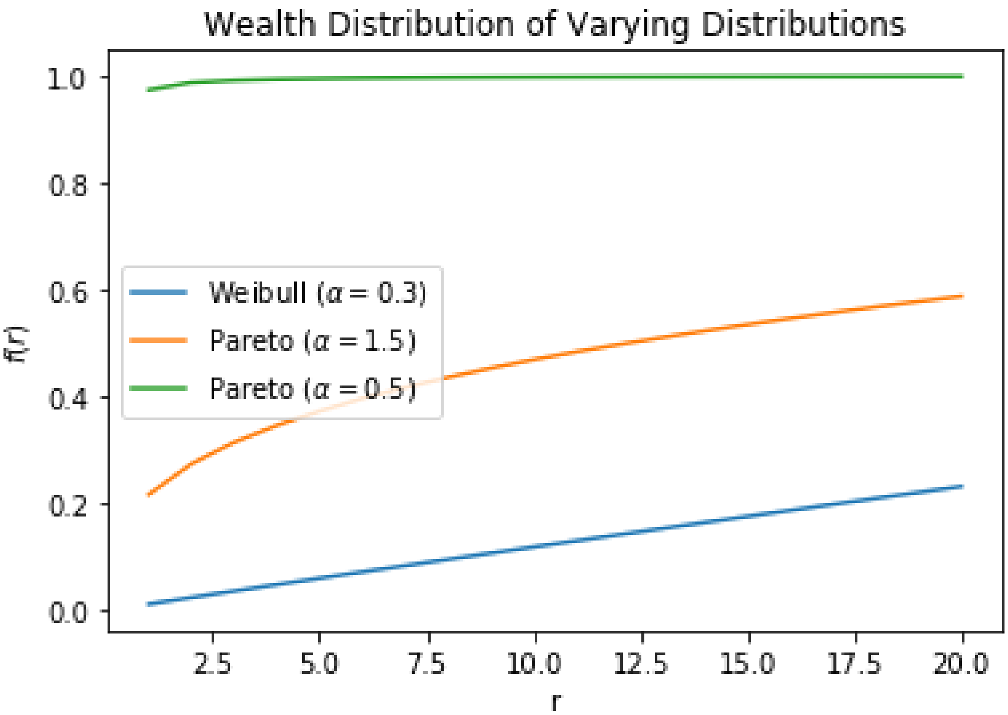






The central limit theorem appears to hold for all the above cases. For each graph, we see that deviations of S_n are consistently on the order of $\sqrt{n}$.

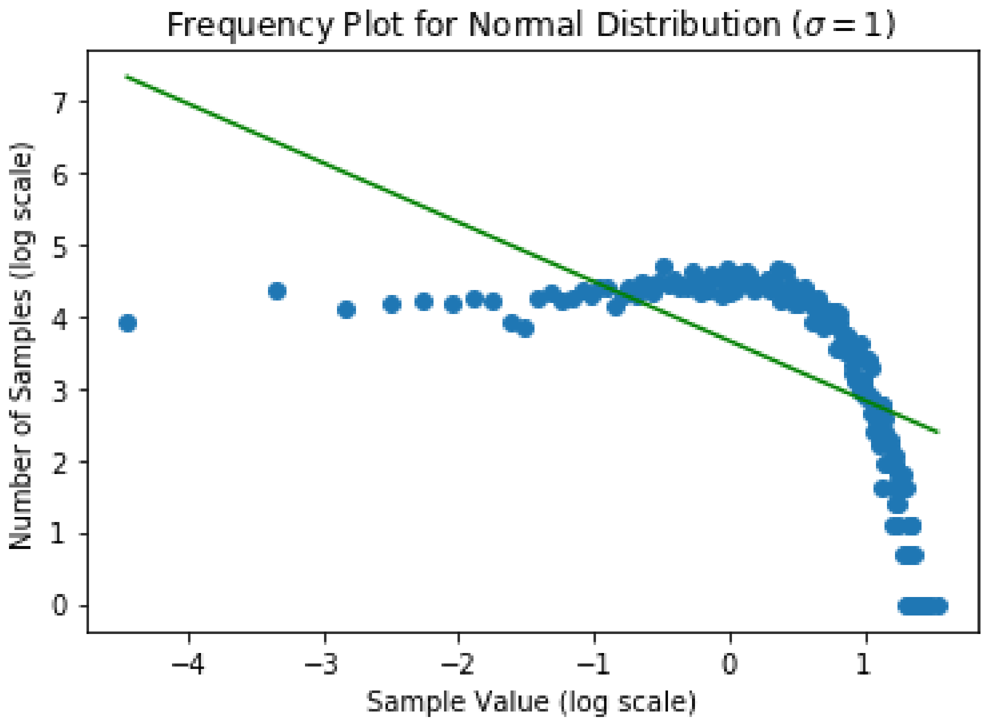
c.



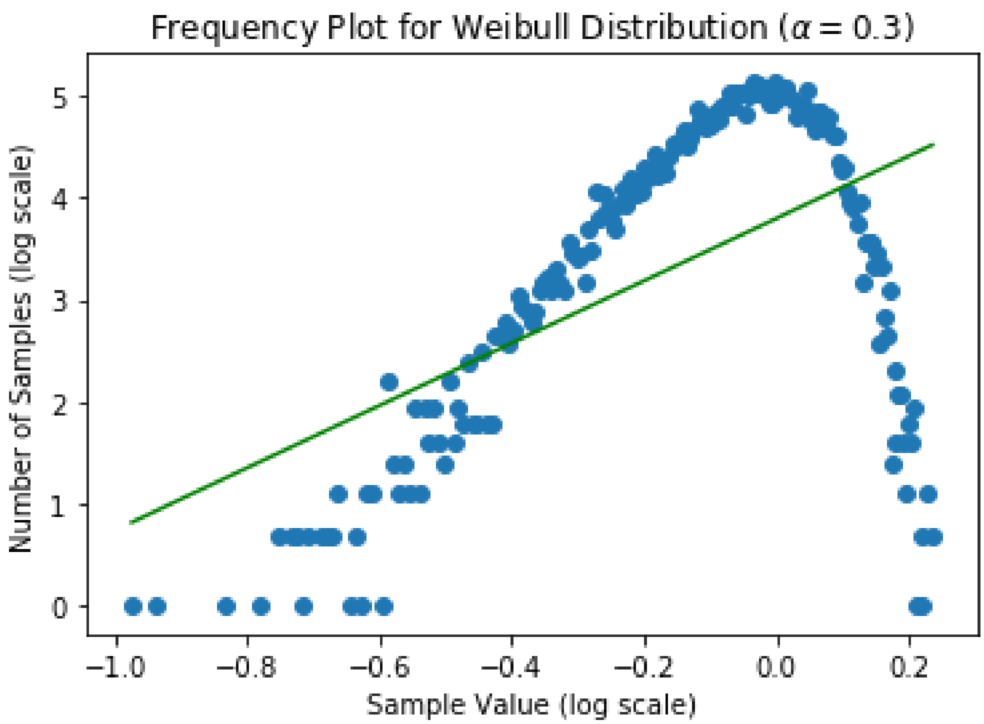
We see the wealth disparity in our heavy-tailed pareto distributions. If we just look at the pareto with $\alpha = 1.5$, we see that the top 20% have about 60% of the wealth, and in the pareto with $\alpha = 0.5$, we see that the top 2-3% of wealthiest people own close to all the wealth. However, in the weibull distribution, the wealth distribution seems to be a lot more even; the top 20% richest people do end up having about 20% of total wealth. The 80-20 rule is indeed a good marker for heavy-tailed distributions.

d.

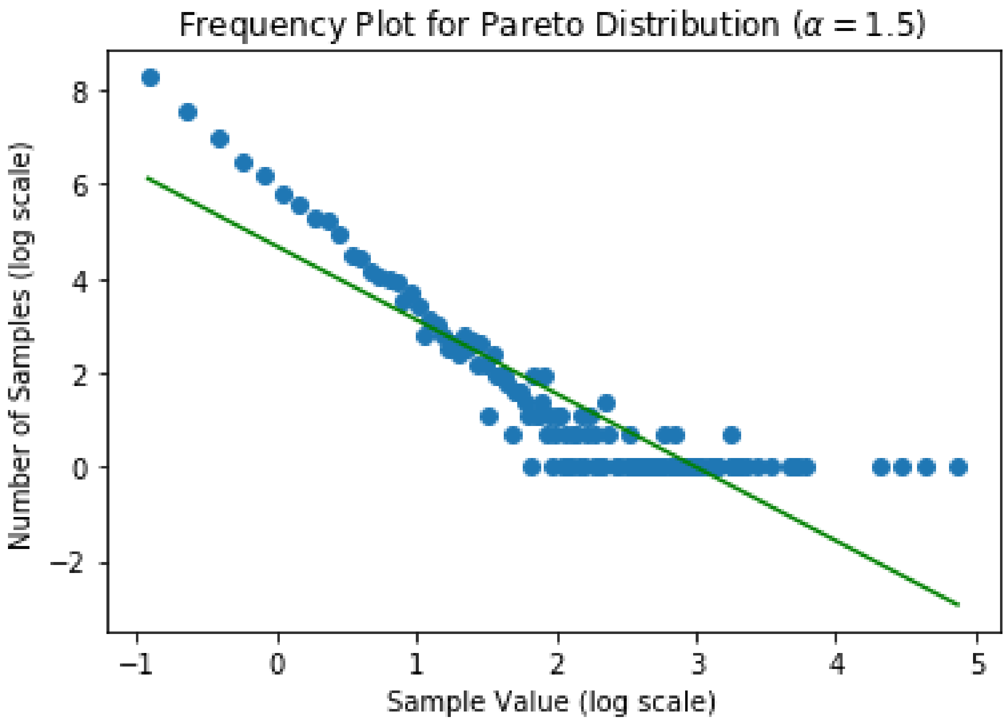
Frequency plots (log/log scale)



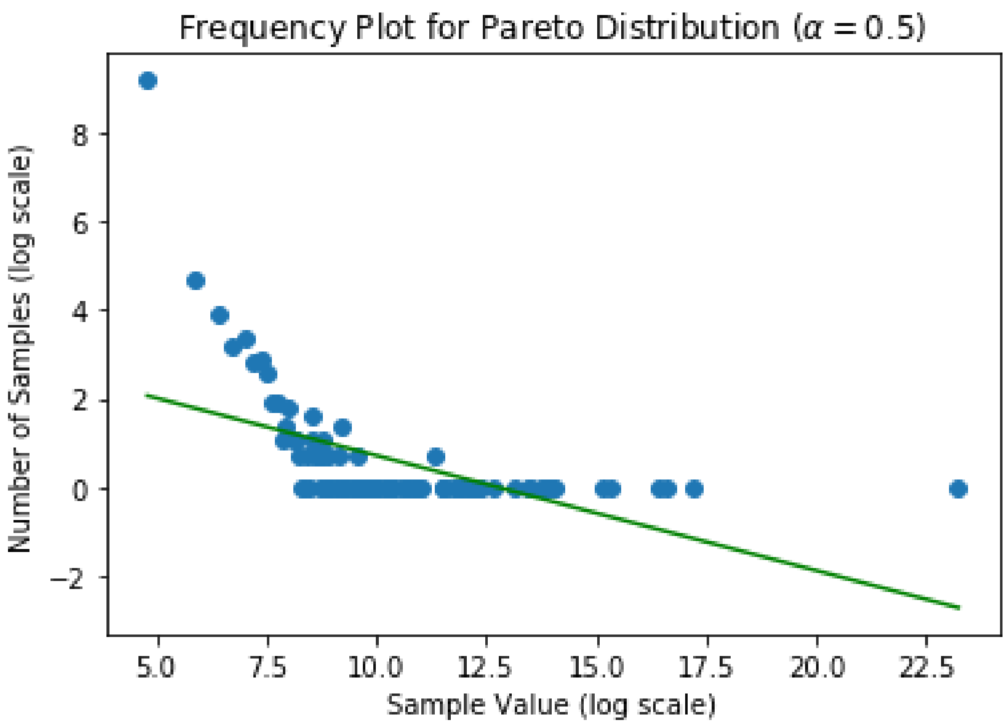
$R^2 = 0.359$



$R^2 = 0.312$

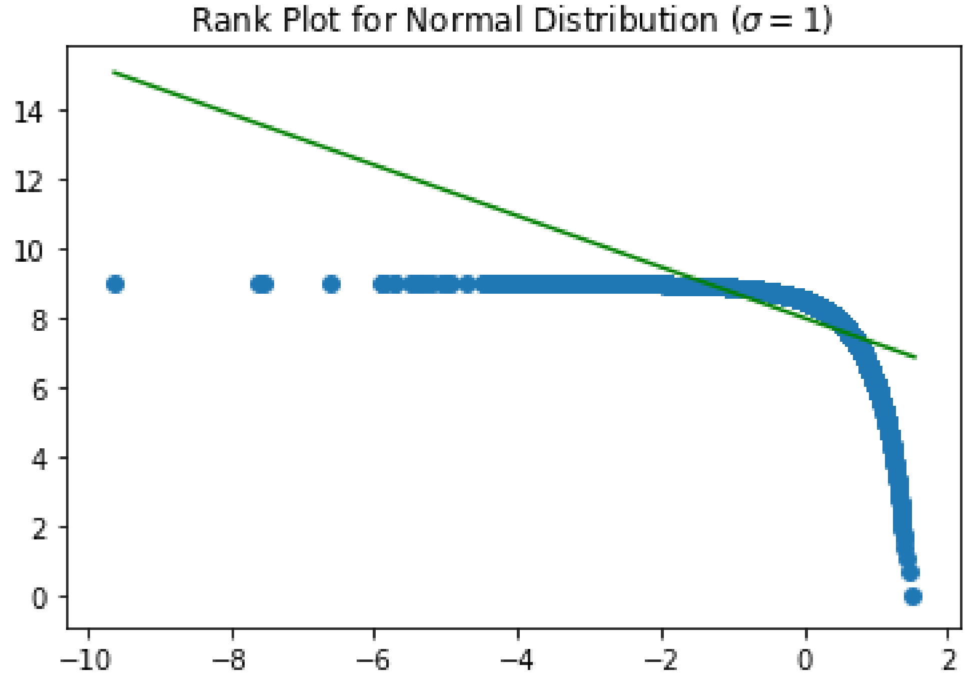


$R^2 = 0.785$

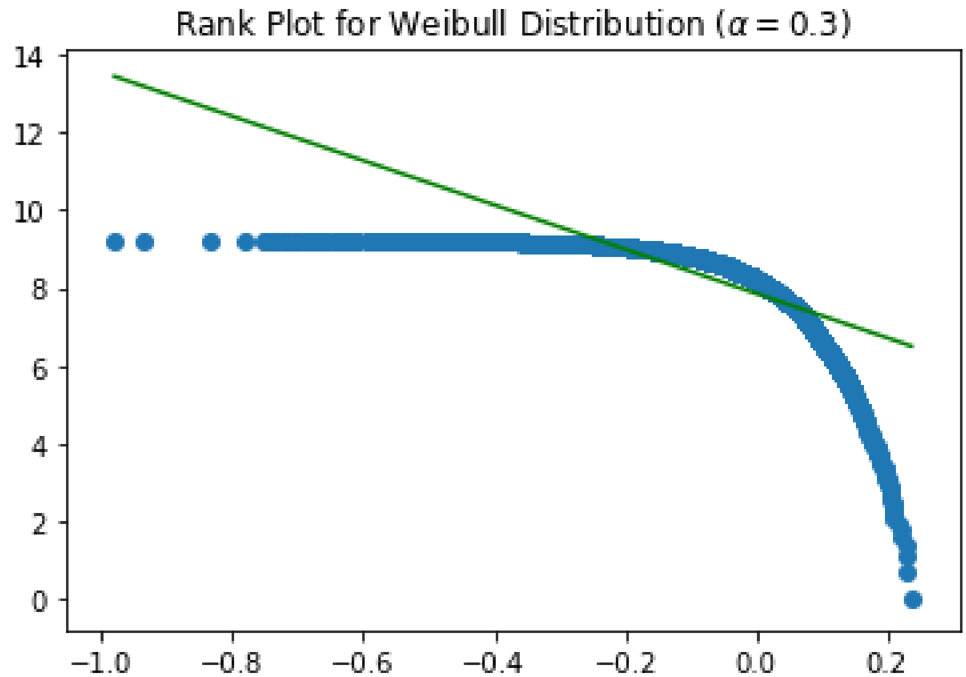


$R^2 = 0.295$

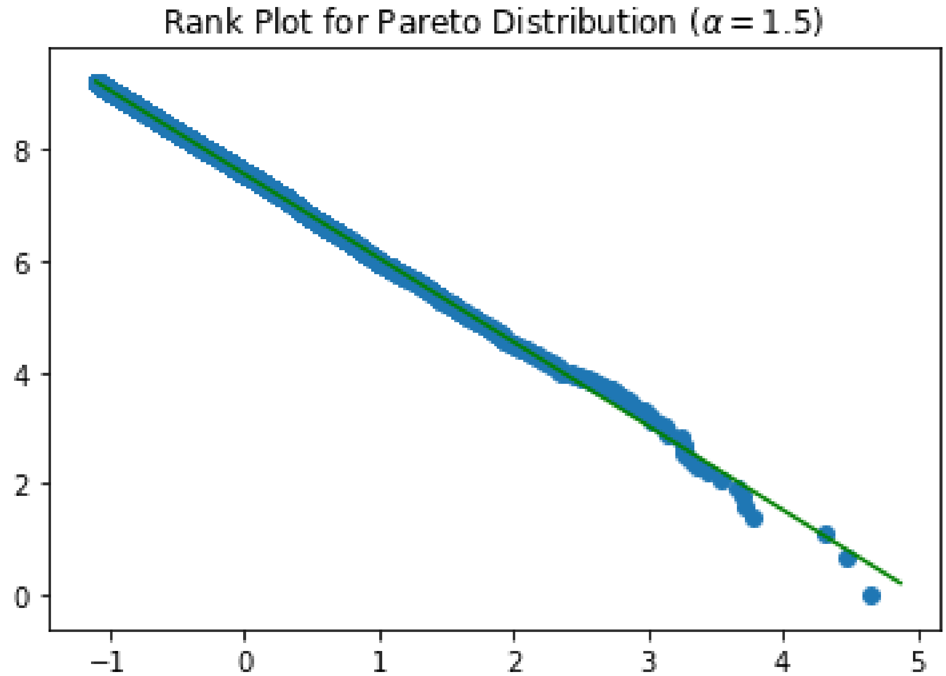
Rank plots (log/log scale)



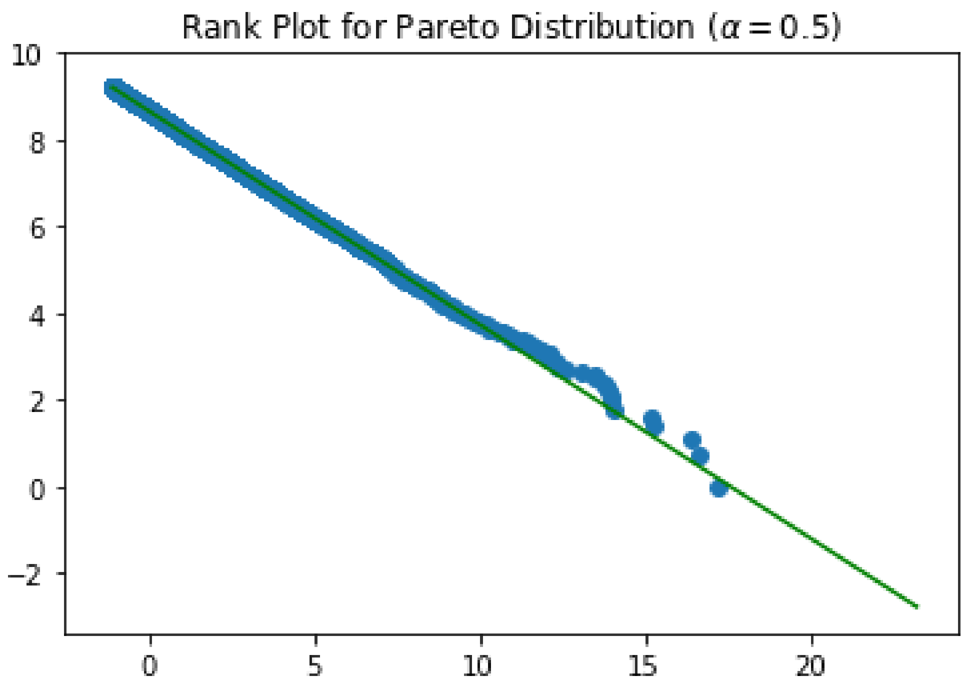
$R^2 = 0.487$



$R^2 = 0.616$



$R^2 = 0.999$



$R^2 = 0.999$

From the shapes of both the frequency plots and the rank plots, we see that on log-log scales, the heavy-tailed pareto distributions look a lot more linear than the weibull and normal. Looking at the frequency plots specifically, we also see that the pareto distributions have a lot more sample values that are far greater than the mean than compared to the that of other distributions. Looking at the rank plots, what stands out is the staggering linearity of the pareto vs. the others. This is expressed in the ridiculously high $R^2 = 0.999$ for pareto while the $R^2 = 0.487$ for the normal and $R^2 = 0.616$ for the weibull.

Problem 3

i. From our assumptions, we know that the number of nodes in the graph, n, is on the order of $\Theta(c)$ and that the number of edges in the graph, is on the order of $\Theta(c^{2/\alpha})$. If we divide the number of edges by the number of nodes, we will get a number on the order of the average degree of a node in the graph. This is precisely $\Theta(c^{2/\alpha} - 1)$ which is also $O(c^{2/\alpha} - 1)$. Now, to bound this, we use the weak law of large numbers (presented in problem set 1). From the weak law of large numbers, we know that as n grows, that the average degree approaches the expected degree and thus, with constant probability, a node selected uniformly at random will have "small" degree, specifically $O(c^{2/\alpha} - 1)$.

ii. We first find the probability that a node selected uniformly at random will be a node of "large" degree. We use Markov's inequality to do this. Specifically, we let d be degree, and so we want to find $P(d \geq c^{(1/\alpha)} - \epsilon)$. Appying Markov's inequality and the $E[d]$ or expected degree from part i, we have:

$$P(d \geq c^{(1/\alpha)-\epsilon}) \leq \frac{\Theta(c^{2/\alpha-1})}{c^{1/\alpha-\epsilon}}$$

$$P(d \geq c^{(1/\alpha)-\epsilon}) \leq \Theta(c^{1/\alpha-1+\epsilon})$$

Now, multiplying this probability by the total number of nodes which is $\Theta(c)$ gives us expected number of nodes with "large" degree. We call this $E[l]$ which is:

$$E[l] = \Theta(c^{1/\alpha-1+\epsilon})\Theta(c)$$
$$E[l] = \Theta(c^{1/\alpha+\epsilon})$$

Now, to get the total number of stubs adjacent to these large nodes, we have to multiply $E[l]$ by the degree expected of a "large" node as defined in the problem (having degree $\geq c^{1/\alpha-\epsilon}$ or equivalently, degree of $\Omega(c^{1/\alpha-\epsilon})$). Thus, our total number of stubs adjacent to nodes of "large" degree is:

$$\Theta(c^{1/\alpha+\epsilon})\Omega(c^{1/\alpha-\epsilon}) = \Theta(c^{2/\alpha}) = \Theta(m)$$

Thus, we have shown the total number of stubs adjacent to nodes of "large" degree is $\Theta(m)$.

iii. We first provide some intuition. From part i, we know that a node selected uniformly at random will have a degree that is $O(c^{2/\alpha-1})$. Thus, v_1 has degree $O(c^{2/\alpha-1})$. Now, from part ii, we know that total number of stubs adjacent to nodes of "large" degree is $\Theta(m)$, so thus, our v_1 will have a neighbor v_2 which has "large" degree. Now, to get the ratio, we divide expected degree of v_1 from the "large" degree of v_2, so we get:

$$\frac{\Omega(c^{1/\alpha-\epsilon})}{O(c^{2/\alpha-1})}$$

Recognizing that n is on the order of $\Theta(c)$, we can simplify the above expression to get:

$$\frac{\Omega(n^{1/\alpha-\epsilon})}{O(n^{2/\alpha-1})} = \Omega(n^{1-(1/\alpha)-\epsilon})$$

as desired.

Problem 4

a. As mentioned in the problem, we can split this up into two cases: whether $i = t$ or $i < t$ which relates them to whether they are eligible for purchases by individualistic or social behavior. If $i = t$, then we are guaranteed that the sales volume will increase by 1 due to the individual purchase. If $i < t$, then the sales volume will increase by 1 if product i is chosen to be the social product (which happens with probability of $\frac{m_i(t)}{2t-1}$, where $(2t-1)$ is the total number of purchases made up to time t. Thus, for product i, we get that the expected increase in market size $(m_i(t))$:

$$\begin{cases} 1 & i = t \\ \frac{m_i(t)}{2t-1} & i < t \end{cases}$$

b. This is pretty intuitive to see. For $t > i$, we are concerned with the social purchase aspect. We know that the likelihood that a consumer chooses a given product for his social purchase is proportional to the current volume, $m_i(t)$ of the product. Thus, that gives us our numerator. The denominator is simply the total number of purchases so far. For each t, we know that two purchases are made except for the first customer who does not make a social purchase. Thus, the denominator is $(2t-1)$, and so the rate of change of volume of product i at time t is simply $\frac{m_i(t)}{2t-1}$ for $t > i$.

To show this using mean value and continuous time approximation, we have that:

$$\frac{m_i(t) + m_i(t-1)}{2(t + (t-1))} \approx \frac{m_i(t)}{2t-1}$$

and

$$\frac{m_i(t) - m_i(t-1)}{t - (t-1)} = \frac{dm_i(t)}{dt}$$

Thus, we get $\frac{dm_i(t)}{dt} = \frac{m_i(t)}{2t-1}$ for $t > i$.

c.

$$\frac{dm_i(t)}{dt} = \frac{m_i(t)}{2t-1}$$

$$\frac{dm_i(t)}{m_i(t)} = \frac{dt}{2t-1}$$

Let's take the integral of both sides:

$$\int \frac{dm_i(t)}{m_i(t)} = \int \frac{dt}{2t-1}$$

$$\ln(m_i(t)) = \frac{\ln(2t-1)}{2} + C$$

Using the approximation given in the problem that $\ln(2t-1) \approx \ln(2t)$, we get:

$$\ln(m_i(t)) = \frac{\ln(2t)}{2} + C$$

Let's take the exponent of both sides and solve for our constant.

$$m_i(t) = A\sqrt{2t}, A = e^C$$

We use the fact that $m_i(i) = 1$ to solve for A.

$$1 = A\sqrt{2i} \longrightarrow A = \frac{1}{\sqrt{2i}} \to C = \ln \frac{1}{\sqrt{2i}}$$

Thus, our final equation for $m_i(t)$ is :

$$m_i(t) = \frac{1}{\sqrt{2i}}\sqrt{2t} = \sqrt{\frac{t}{i}}$$

d. The market share is:

$$\frac{\sqrt{\frac{t}{i}}}{2t-1}$$

To understand how this relates to Amazon's success, we look at the market share of the most popular product as time goes on.

$$\lim_{t\to\infty} \frac{\sqrt{\frac{t}{i}}}{2t-1} = 0$$

This shows that Amazon's business model accounts for the heavy-tailed world we live in. Instead of focusing on a few popular items (in which case, we see market share dying as time moves on), Amazon chooses to maintain a huge variety of items to profit.

Problem 5

a. We first give the probability $P(l,k)$ that the shortest path from node A_i to node A_j has length l given that node A_j is k hops away from A_i along the ring. To build some intuition, we do a few examples. For $l = 1$, we are required to have an edge from A_i to the center and an edge from A_j to the center, which happens altogether at a probability p^2. For $l = 2$, we realize that we have two paths for this to happen (an edge from A_i to center and an edge from A_{j-1} to center or an edge from A_{i+1} to center and an edge from A_j to center). This happens with probability $2p^2(1-p)$. The $(1-p)$ comes from not having an edge so that the minimum length would be 1 instead of 2. With this, we get our general formula for $P(l,k)$ for $l < k$.

$$P(l,k) = lp^2(1-p)^{l-1}, l < k$$

Now, for the special case for $l = k$, we not only have the term above, but we also need to account for the scenario where there are no edges and the scenario where we have a single pair of adjacent edges to the middle. Thus, for $P(l,k)$ when $l = k$, we have:

$$P(l,k) = lp^2(1-p)^{l-1} + (1-p)^{k+1} + (k+1)(p)(1-p)^k, l = k$$

Now, putting this altogether, we get:

$$P(l,k) = \begin{cases} lp^2(1-p)^{l-1} & l < k \\ lp^2(1-p)^{l-1} + (1-p)^{k+1} + (k+1)(p)(1-p)^k & l = k \\ 0 & l > k \end{cases}$$

Now, to get the expected value of the shortest path length from A_i to A_j, we simply sum over the lengths with their respective probabilities above and get:

$$E[k] = (p^2 \sum_{l=1}^{k} l^2(1-p)^{l-1}) + k(1-p)^{k+1} + k(k+1)(1-p)^k p$$

Plugging this into mathematica, we get:

$$E[k] = -\frac{k^2p^2(1-p)^k + 2kp(1-p)^k - p(1-p)^k + 2(1-p)^k + p - 2}{p} + k(k+1)p(1-p)^k + k(1-p)^{k+1}$$

b. To get the average shortest path length, we have a few key observations. First off, given n nodes, we need to count the number of nodes that are 1 hop from each other, number of nodes that are two hops from each other, ... , number of nodes that are $(n-1)$ hops from each other (we take $A_i \to A_j$ to be different than $A_j \to A_i$). Due to the ring structure, given n nodes, we will have n pairs of nodes that are 1 hop from each other, n pairs that are 2 hops from each other, ..., n pairs of nodes that are $(n-1)$ hops from one another. Thus, to get the total sum of all shortest path lengths, for each k, we multiply the numbers of pairs of nodes that are k hops from each other by $E[k]$ (calculated in part a). Overall, we have $2\binom{n}{2} = n(n-1)$ total pairs of nodes we are summing over. Thus, our expected average shortest path length is:

$$\frac{n\sum_{k=1}^{n-1} E[k]}{n(n-1)}$$

and substituting $E[k]$, we get:

$$\frac{\sum_{k=1}^{n-1}(p^2\sum_{l=1}^{k} l^2(1-p)^{l-1}) + k(1-p)^{k+1} + k(k+1)(1-p)^k p}{n-1}$$

Now, plugging the above expression into mathematica, we get:

$$-\frac{np^2 - np(1-p)^n + 2p(1-p)^n - 3(1-p)^n - 2np - 2p + 3}{(n-1)p^2}$$

The average shortest path length between nodes on the ring of the graph if there was no central node would precisely be $\frac{n\sum_{k=1}^{n-1} k}{n(n-1)} = \frac{\sum_{k=1}^{n-1} k}{n-1} = \frac{n}{2}$. This can be seen because without a central node, we will have n pairs (as before, $A_i \to A_j$ is different than $A_j \to A_i$)of nodes with distance 1 apart form each other, n pairs with distance 2, ... n pairs with distance $(n-1)$ apart from one another.

Now, comparing these two average shortest path lengths, we see that for large n (as $n \to \infty$), that our expected average shortest path length with the central node converges to some positive constant C.

$$\lim_{n\to\infty} -\frac{np^2 - np(1-p)^n + 2p(1-p)^n - 3(1-p)^n - 2np - 2p + 3}{(n-1)p^2} = C, C > 0$$

.
However, in the case where there is no central node, we have that as n grows large, the expected average shortest path length becomes infinitely large.

$$\lim_{n\to\infty} \frac{n}{2} = \infty$$

Thus, we see that having this central node drastically cuts down on average shortest path length, especially as n gets bigger and bigger.

c. No, it does not always find the shortest path. The one edge case where it does not found the shortest path is if along the route, there is a single edge from that particular node to the middle node. In this case, the algorithm will have the packet be sent off to the middle node, but then realizing that there is no other edge for the packet to travel along from the middle, will go back along the ring using the edge it came from. This would add 1 to the path length, and so by this algorithm, we are guaranteed to find a path with length, at most, 1 longer than the length of the shortest path.