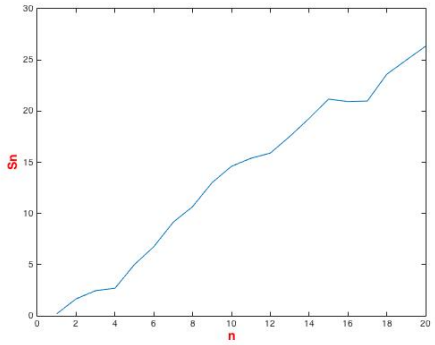
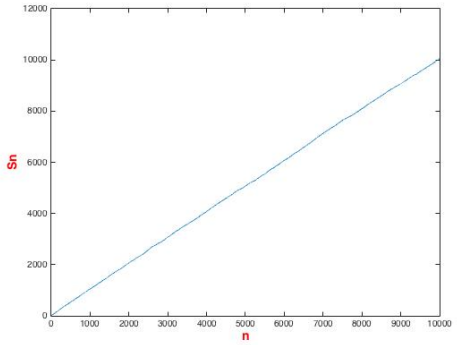
2. Heavy vs. light

a) Law of Large Number:

i. Normal distribution:

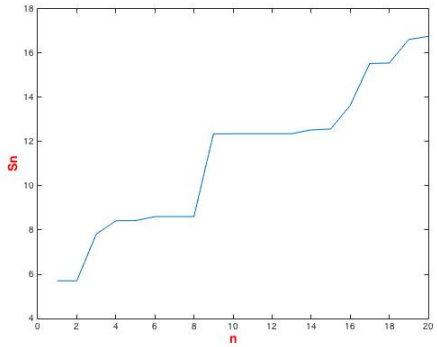


n=20

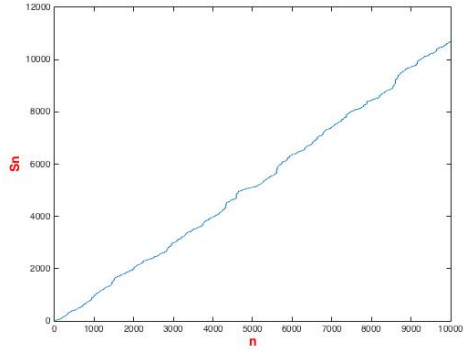


n = 10000

ii. Weibull distribution.

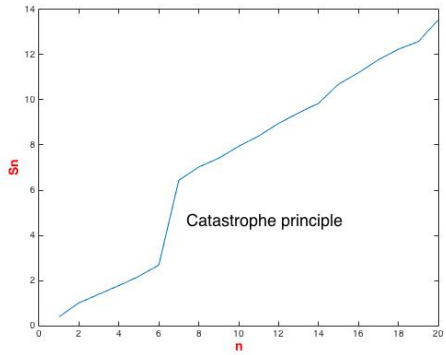


n=20

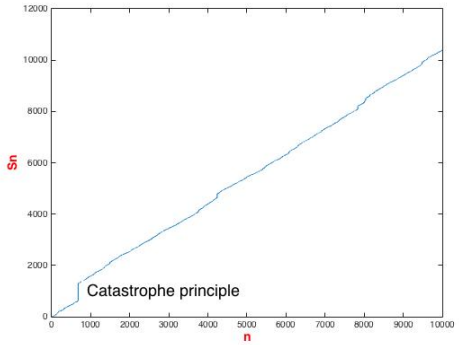


n = 10000

iii. Pareto distribution($\alpha = 1.5$)

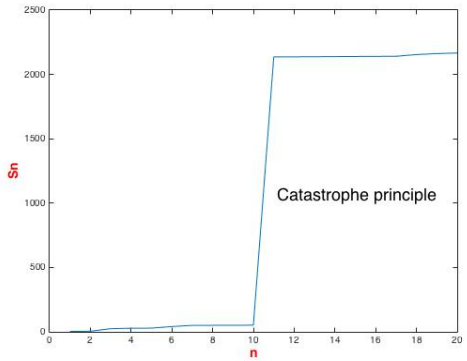


n=20

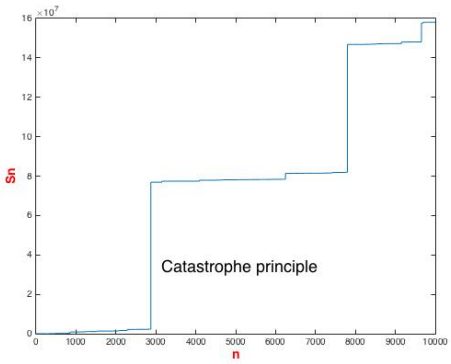


n = 10000

iv. Pareto distribution ($\alpha = 0.5$)



n=20



n = 10000

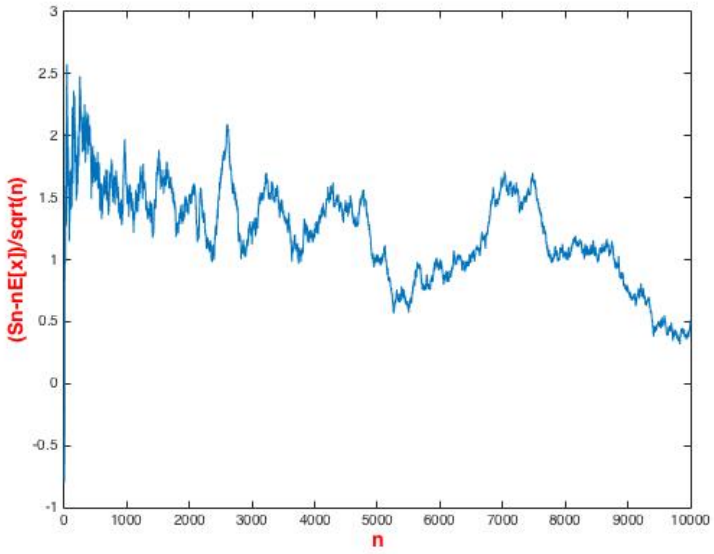
Interpretation:

In probability theory, the **law of large numbers** (**LLN**) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed. Here, in this problem, we can see when n is 10000, which is large enough, S_n is linear in Normal distribution and Weibull distribution.

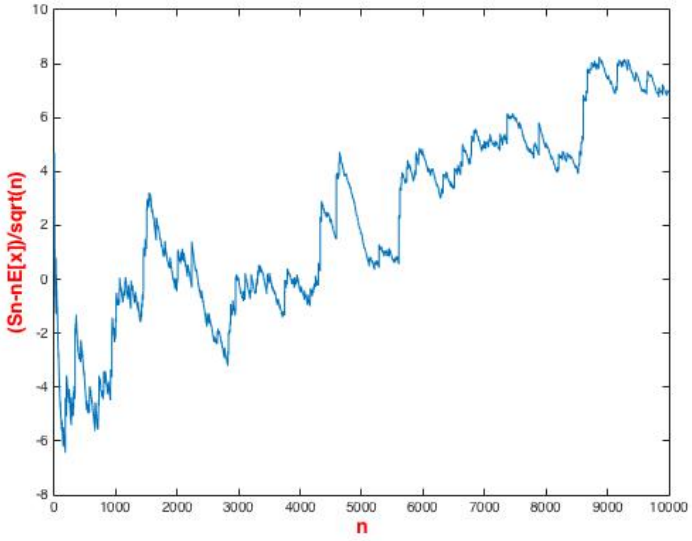
However, in Pareto distribution, there are big jumps which prevent the plot being purely linear, this is the Catastrophe principle of the Heavy-tail distribution, where individual point may contributes large in S_n.

b) Central Limit Theorem:

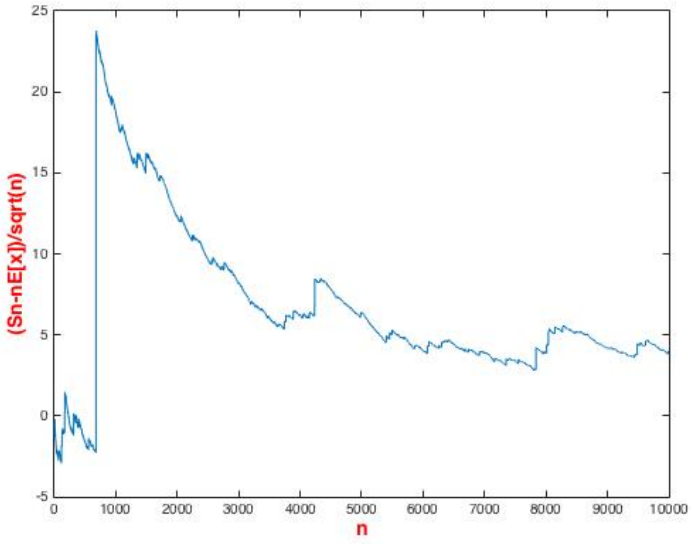
I. Normal distribution



II. Weibull distribution



iii. Pareto distribution($\alpha = 1.5$)

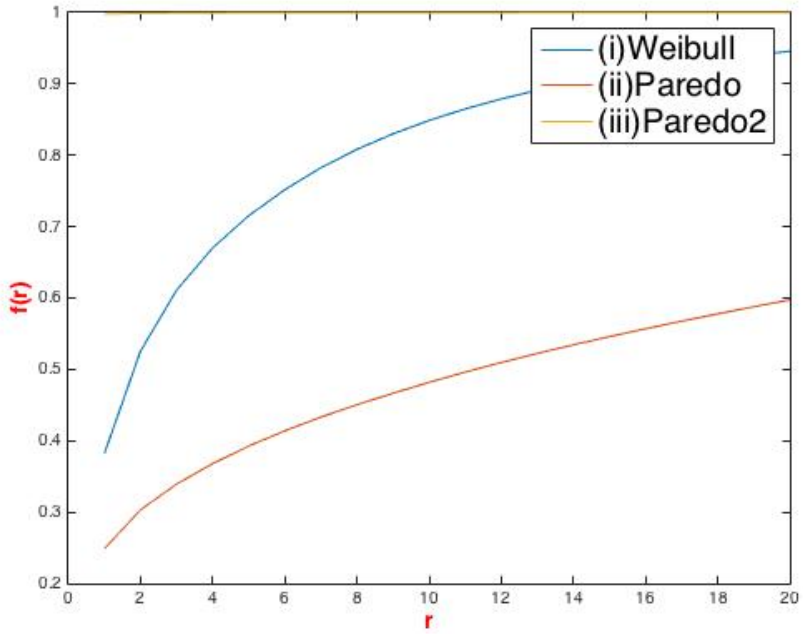


Interpretation:

Central Limit Theorem states that when n is large enough, $\frac{S_n - nE[X]}{\sqrt{n}} \sim N(0,\ 1)$,

Where the pareto distribution seems not follow this theorem in the plot, it has large jump here which can also be explained by the catastrophy principles stated in problem(a).

c) The 80-20 Rule:

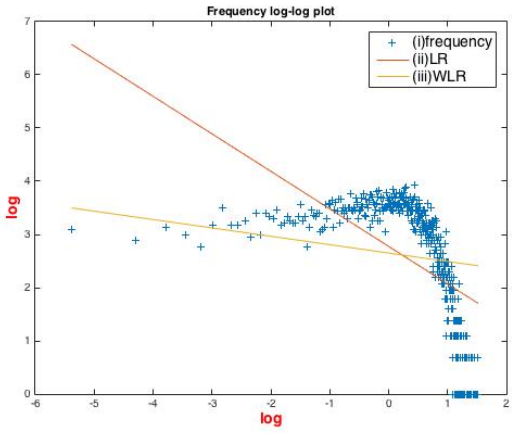


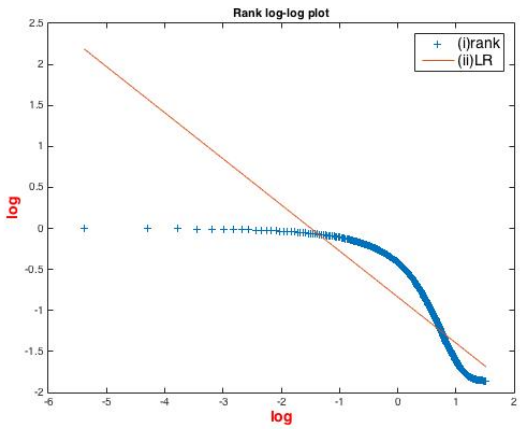
Interpretation:

This example is extremely illustrated by the Pareto 2 which in on the top of the plot. It means that the very few percent of people hold a large amount of wealth that at the very beginning of the plot, it already have large f(x).

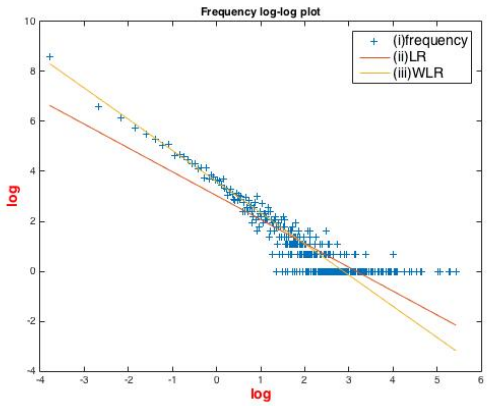
d) Identifying heavy tails:

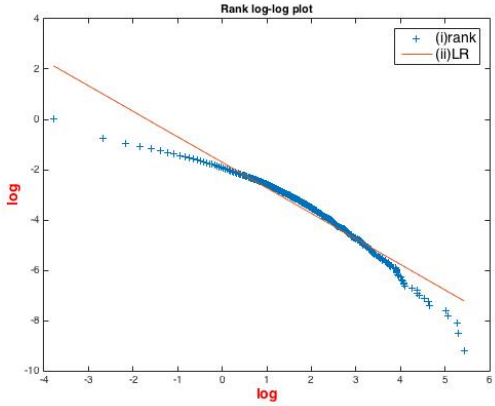
i. Normal distribution



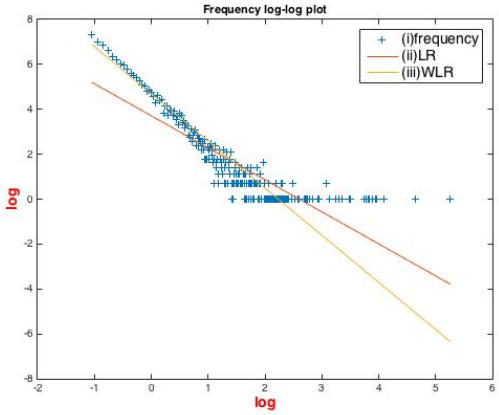


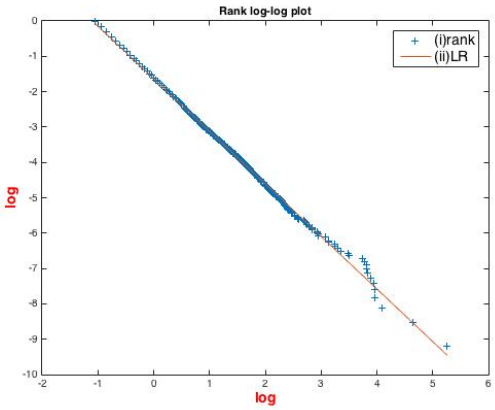
ii. Weibull distribution



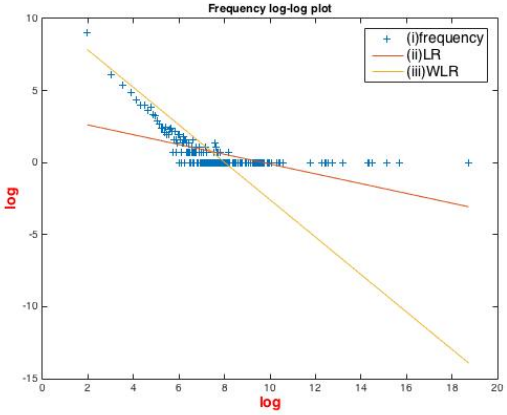


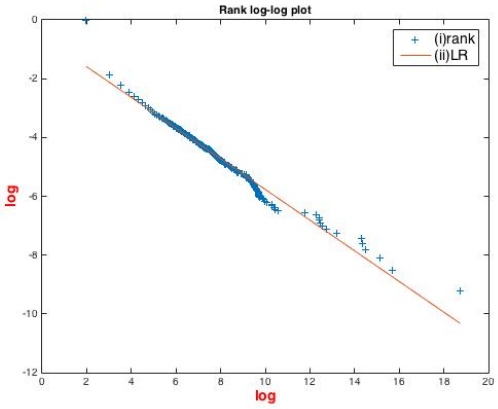
iii. Pareto distribution($\alpha = 1.5$)





iv. Pareto distribution ($\alpha = 0.5$)





Interpretation:

Choose of number of bars:

Normal: 500

Weibull: 5000

Pareto1: 5000
Pareto2:10000000

LR refers to linear regression
WLR refers to weighted linear regression

I can tell from the plot that normal distribution is not linear on log-log plot, while Pareto seems to be linear with some noise.