Project 1

inż. Sebastian Janus, inż. Wojciech Strycharczyk

April 5, 2023

In this project we will use the Law of Large Numbers to build a quick procedure to visually check existance of the moments in probabulity distributions. For our analysis, we have selected

1. Exponential distribution,
2. Gamma distribtion,
3. Lognormal distribution
4. Chi-squared (χ^2) distribution,
5. Pareto distribution.

For each distribution we chose different sets of parameters.

1 Law of Large Numbers

Law of Large Numbers (LLN) is a theorem that says the average of the results obtained from a large number of trials should be close to the expected value as more trials are performed.

We can use the LLN to verify if the moments exist for the given probability distribution visually. Let $X_1, X_2, \ldots, X_n$ be the i.i.d. random variables from the given probability distribution. Then, if the n-th moment is well-defined, the following series

$$X_1^n, \frac{X_1^n + X_2^n}{2}, \frac{X_1^n + X_2^n + X_3^n}{3}, \ldots, \frac{X_1^n + \ldots + X_n^n}{n}$$

should converge to a constant.

In case of the first moment, we expect this series to converge to the theoretical expected value of the analysed probability distribution. When checking higher moments, we will just be looking at the convergence to the constant.

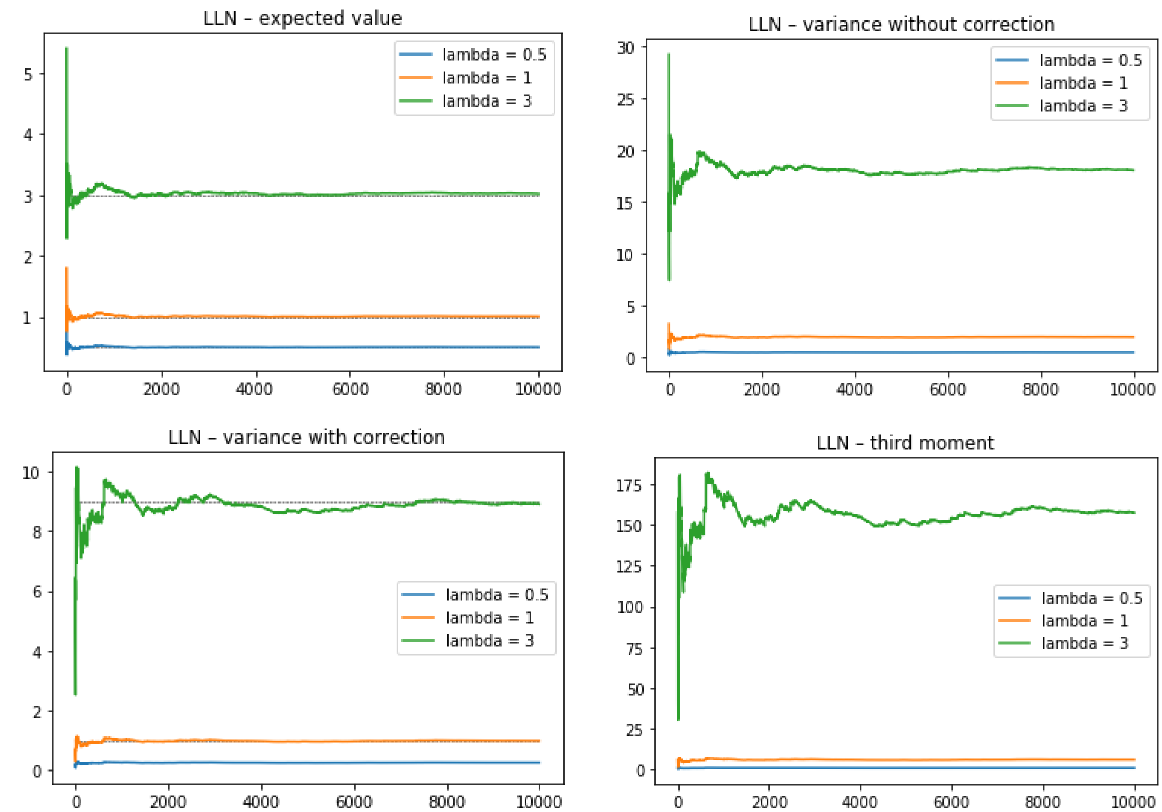
Of course one can apply additional 'correction' to check if the series converge to the theoretical value of the higher moments. In this project, we applied the correction to the second moment calculations.

2 Exponential distribution

The exponential distribution is a probability distribution defined by the probability density function

$$f_X(x) = \lambda e^{-\lambda x} \, \mathbf{1}_{x \geq 0}.$$

λ is the rate parameter defined on $(0, \infty)$. We can define the following theoretical properties: 1. expected value: $\mathbb{E}X = \lambda^{-1}$, 1. variance: $\text{Var}X = \lambda^{-2}$.



We selected three lambdas, i.e. 0.5, 1 and 3. We conducted the simulation for $N = 10^4$ to check the convergence of the above-mentioned series. In all of the cases, we observe that expected value, variance and third moment exist for the exponential distribution. Moreover, for the expected value and for the 'corrected' variance the series converge to the theoretical values, denoted by the dashed lines on the charts.

3 Gamma distribution

The gamma distribution is a probability distribution defined by the probability density function

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

α is the shape parameter and β is the rate parameter, both defined on $(0, \infty)$. We can define:

1. expected value: $\mathbb{E}X = \frac{\alpha}{\beta}$,
2. variance: $\text{Var}X = \frac{\alpha}{\beta^2}$.

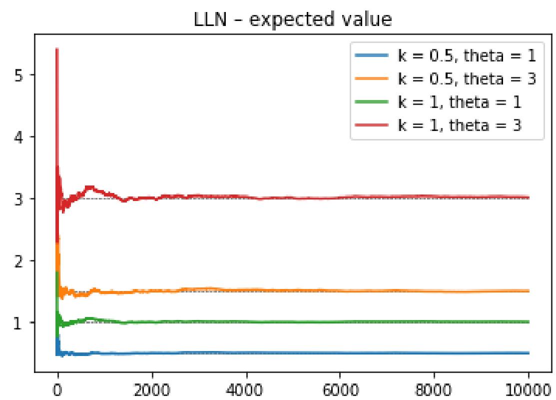
We can also define gamma distribution using the scale parameter:

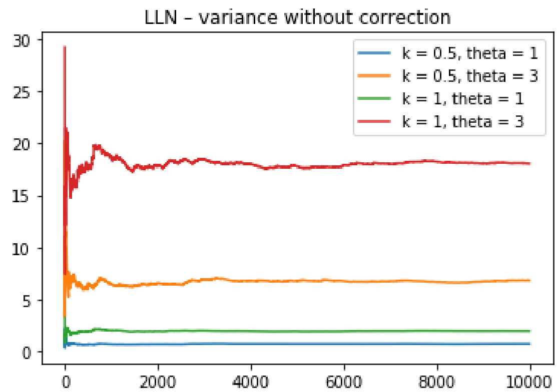
$$f(x) = \frac{1}{\Gamma(k)\,\theta^k} x^{k-1} e^{-\frac{x}{\theta}},$$

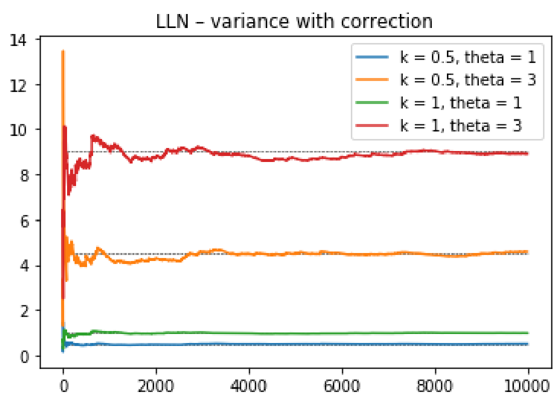
where k is the shape parameter and θ is the scale parameter, both defined on $(0, \infty)$. Then we define:

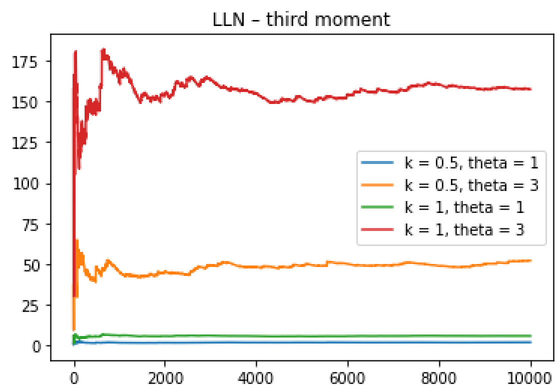
1. expected value: $\mathbb{E}X = k\theta$,
2. variance: $\mathrm{Var}X = k\theta^2$.

We will stick to the second formula where the scale parameter is defined.









We cross-joined two sets of parameters, i.e. $k = \{0.5, 1\}$ and $\theta = \{1, 3\}$. We conducted the simulation for $N = 10^4$ to check the convergence of the above-mentioned series. In all of the cases, we observe that expected value, variance and third moment exist for the gamma distribution. Moreover, for the expected value and for the 'corrected' variance the series converge to the theoretical values, denoted by the dashed lines on the charts.

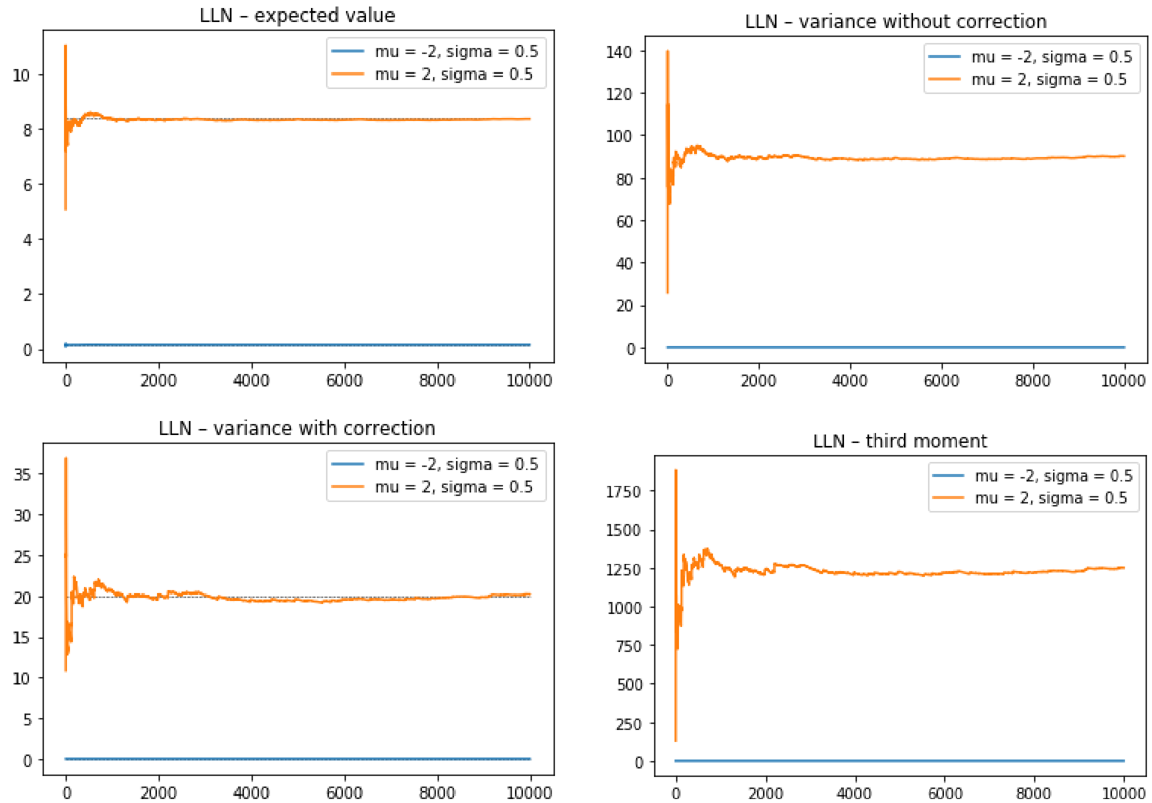
4 Log-normal distribution

The log-normal distribution, that is the transformation of normal distribution, is defined by the probability density function

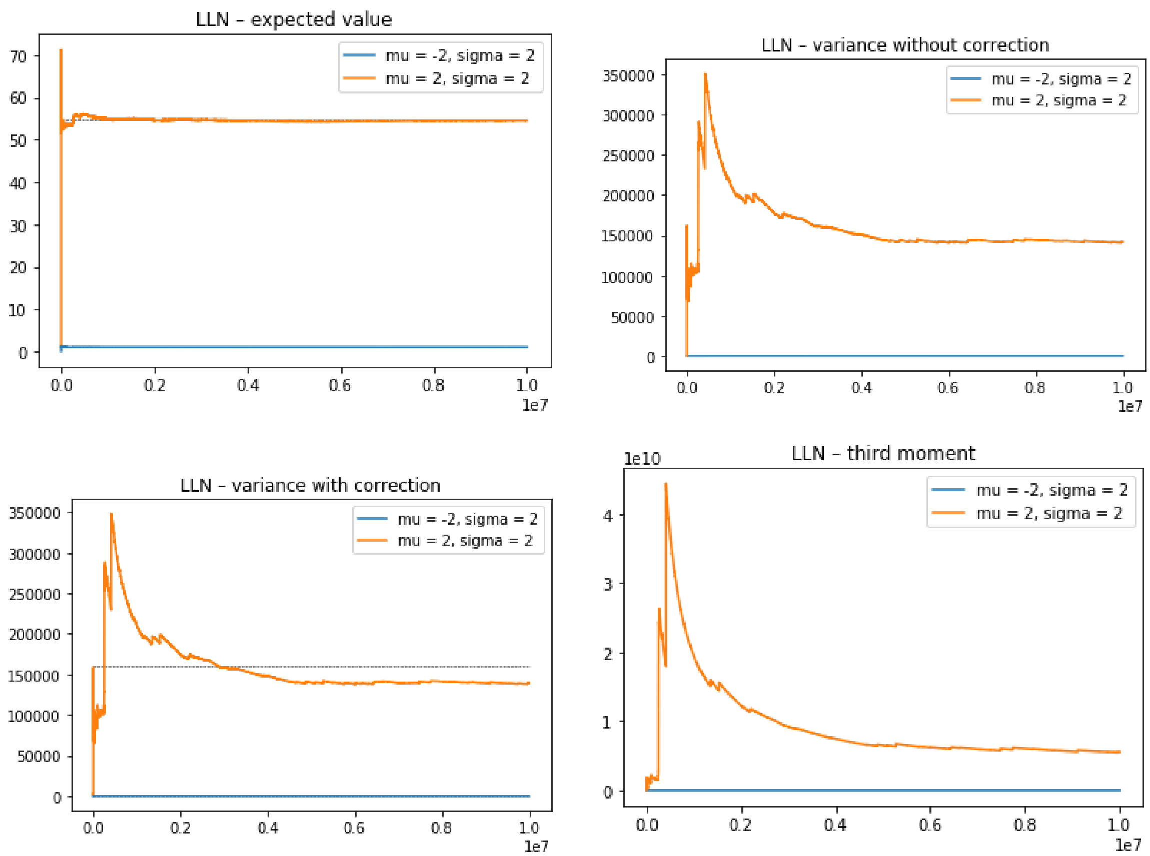
$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right).$$

$\mu \in \mathbb{R}$ and $\sigma > 0$ are the expected value and standard deviation of the variable's natural logarithm. We can define:

1. expected value: $\mathbb{E}X = \exp\left(\mu + \frac{\sigma^2}{2}\right)$,
2. variance: $\mathrm{Var}X = \left[\exp\left(\sigma^2\right) - 1\right]\exp\left(2\mu + \sigma^2\right)$.



Firstly, we selected two values of μ, i.e. -2 and 2 and one σ equaled 0.5. We conducted the simulation for $N = 10^4$ to check the convergence of the above-mentioned series. In all of the cases, we observe that expected value, variance and third moment exist for the log-normal distribution. Moreover, for the expected value and for the 'corrected' variance the series converge to the theoretical values, denoted by the dashed lines on the charts.



Next, we selected the same values of μ and this time σ equaled 2. We noticed that the speed of convergence is very slow for higher values of σ and that's why we have decided to increase N to 10^7. We observe that eventualy, the series converges to the constant, same as in the case of smaller σ.

5 Pareto distribution

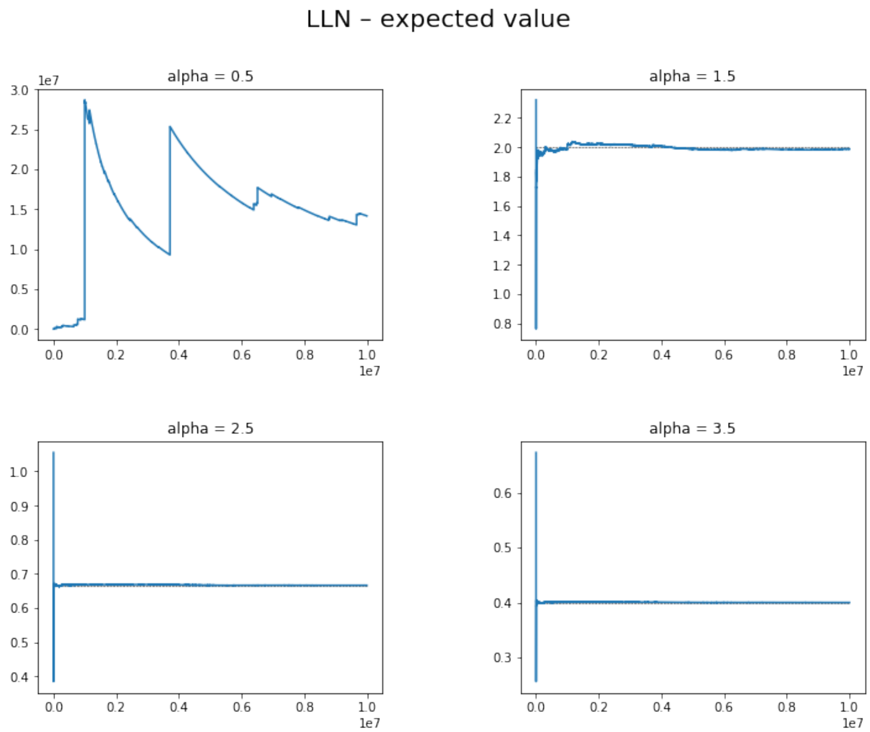
The Pareto distribution is a probability distribution defined by the probability density function

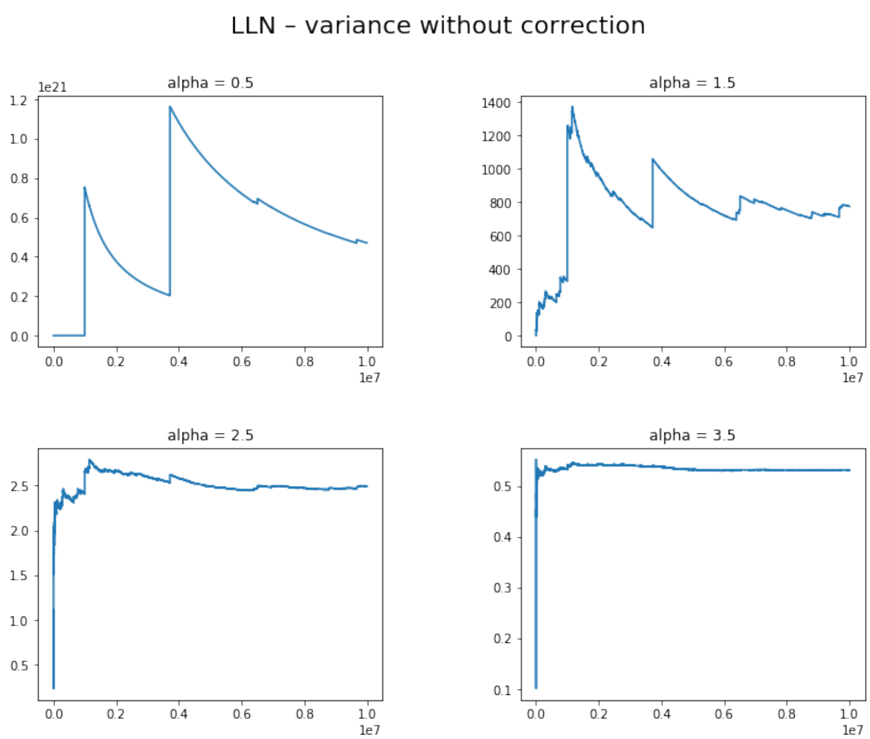
$$f(x) = \frac{\alpha m^\alpha}{x^{\alpha+1}} \mathbf{1}_{x \geq m}.$$

$m > 0$ is the scale parameter and $\alpha > 0$ is the shape parameter. Also $x \in [m, \infty)$. We can define:

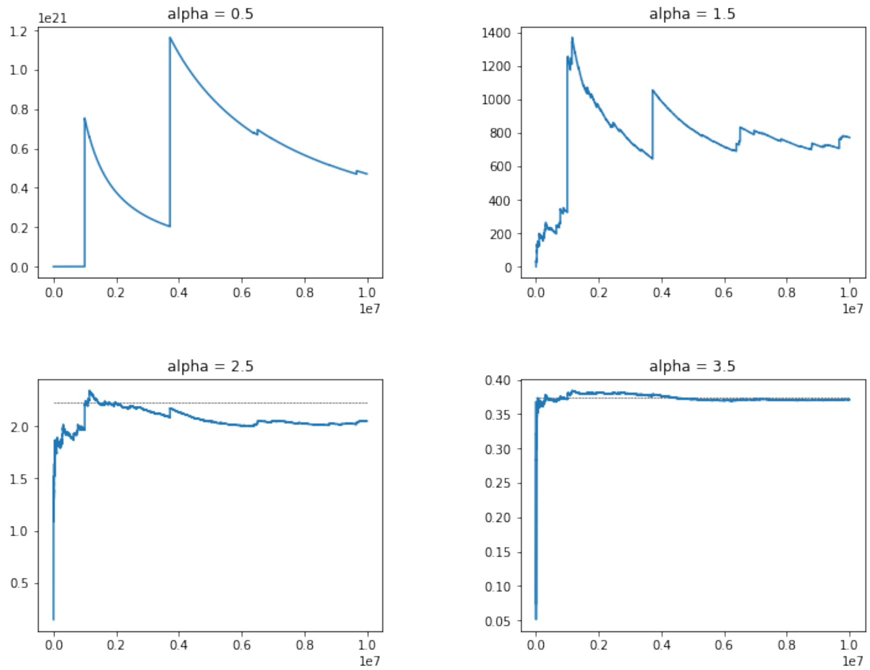
1. expected value: $\mathbb{E}X = \frac{\lambda}{\alpha-1}$ when $\alpha > 1$,
2. variance: $\mathrm{Var}X = \frac{\lambda^2\alpha}{(\alpha-1)^2(\alpha-2)}$ when $\alpha > 2$.

Generally, k-th moment exists for $\alpha > k$.

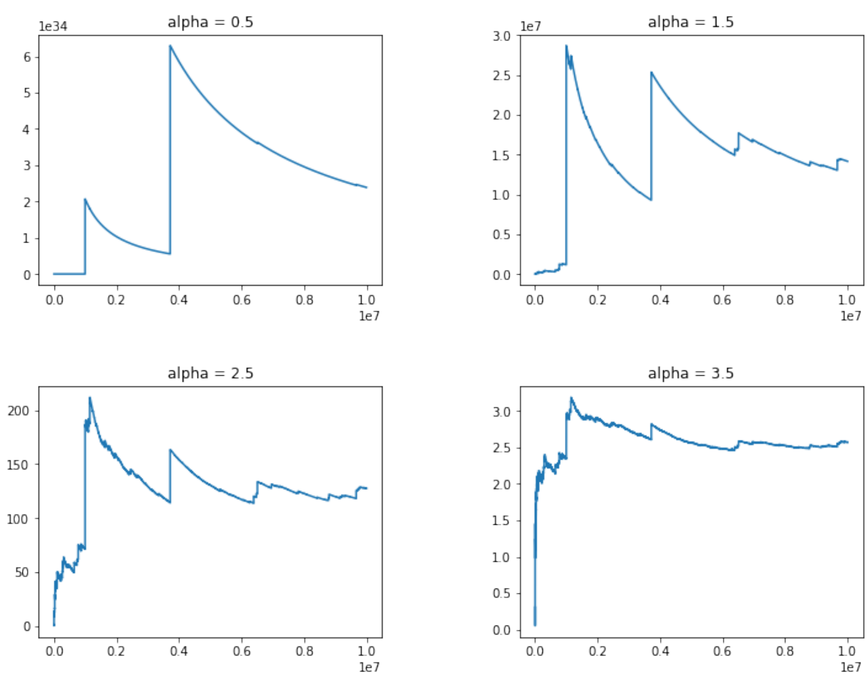

LLN – expected value
alpha = 0.5
alpha = 1.5
alpha = 2.5
alpha = 3.5
1e7


LLN – variance without correction
alpha = 0.5
alpha = 1.5
alpha = 2.5
alpha = 3.5
1e21
1e7

LLN – variance with correction



LLN – third moment



In case of the Pareto distribution, we have selected four α values to check the existance of the first, second and third moment, i.e. $\{0.5, 1.5, 2.5, 3.5\}$. As expected, we observed that none of the abovementioned moments exist for the $\alpha = 0.5 < 1$. For $\alpha = 1.5 \in [1, 2)$ we can see that only the expected value is well-defined. For $\alpha = 2.5 \in [2, 3)$ we observe that the mean and expected value exist (but not the third moment) and for $\alpha = 3.5 > 3$ all three moments exist.