Report 1

Aleksander Jakóbczyk i Bogdan Banasiak
Index number: 255939 i 256456

1 Information and formulas

Stable random variable

There are two parameterizations of a random variable from an alpha stable distribution $S(\alpha,\beta,\gamma,\delta;0)$ and $S(\alpha,\beta,\gamma,\delta;1)$. They are uniquely determined by the characteristic function.

Definition 1. *A random variable X is stable if and only if $X =^d aZ+b$, with $\alpha \in (0,2],\ \beta \in [-1,1],\ a \neq 1,\ b \in \mathbb{R}$ and Z is a random variable with characteristic function*

$$\varphi_Z(u) = \exp(iuZ) = \begin{cases} \exp\left(-|u|^\alpha (1 - i\beta \tan(\frac{\pi\alpha}{2})(\operatorname{sign} u)\right) & \alpha \neq 1, \\ \exp\left(-|u|^\alpha (1 + i\beta \frac{2}{\pi}(\operatorname{sign} u) \ln|u|\right) & \alpha = 1. \end{cases}$$

Definition 2. *Let $X \sim S(\alpha,\beta,\gamma,\delta;0)$ with $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\gamma \geq 0$, $\delta \in \mathbb{R}$ then*

$$X \stackrel{d}{=} \begin{cases} \gamma(Z - \beta\tan(\frac{\pi\alpha}{2}) + \delta) & \alpha \neq 1, \\ \gamma Z + \delta & \alpha = 1, \end{cases}$$

where $Z = Z(\alpha,\beta)$ is given by 1. X has characteristic function

$$E\exp(iuX) = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left[1 + i\beta\left(\tan\frac{\pi\alpha}{2}\right)(\operatorname{sign} u)\left(|\gamma u|^{1-\alpha} - 1\right)\right] + i\delta u\right) & \alpha \neq 1 \\ \exp\left(-\gamma|u|\left[1 + i\beta\frac{2}{\pi}(\operatorname{sign} u)\log(\gamma|u|)\right] + i\delta u\right) & \alpha = 1 \end{cases}$$

Definition 3. *Let $X \sim S(\alpha,\beta,\gamma,\delta;1)$ with $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\gamma \geq 0$, $\delta \in \mathbb{R}$ then*

$$X \stackrel{d}{=} \begin{cases} \gamma Z + \delta & \alpha \neq 1, \\ \gamma Z + (\delta + \beta\frac{2}{\pi}\ln\gamma) & \alpha = 1, \end{cases}$$

where $Z = Z(\alpha,\beta)$ is given by 1. X has characteristic function

$$E\exp(iuX) = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left[1 - i\beta\left(\tan\frac{\pi\alpha}{2}\right)(\operatorname{sign} u)\right] + i\delta u\right) & \alpha \neq 1 \\ \exp\left(-\gamma|u|\left[1 + i\beta\frac{2}{\pi}(\operatorname{sign} u)\log|u|\right] + i\delta u\right) & \alpha = 1 \end{cases}$$

Above we defined the general stable law in the 0-parameterization and 1-parameterization. Alternatively, we can swap between the parameterizations using the following theorem;

Theorem 1. *Let $Z \sim S(\alpha,\beta,1,0;0)$ with $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\gamma \geq 0$, $\delta \in \mathbb{R}$ then*

$$\begin{cases} \gamma Z + \delta + \beta\gamma\tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1, \\ \gamma Z + \delta + \beta\frac{2}{\pi}\ln\gamma & \alpha = 1 \end{cases} \sim S(\alpha,\beta,\gamma,\delta;1),$$

The tail exponent estimation method gives us the information about the index of stability. The tails of stable random variable are asymptotically power laws.

Theorem 2 (Tail approximation)**.** *Let $X \sim S(\alpha,\beta,\gamma,\delta;k)$ with $\alpha \in (0,2)$, $\beta \in (-1,1]$, $k = 0,1$ then as $x \to \infty$:*

$$\begin{aligned} 1 - F_X(x) &\sim \gamma^\alpha c_a(1+\beta)x^{-\alpha}, \\ f_X(x) &\sim \alpha\gamma^\alpha c_a(1+\beta)x^{-(\alpha+1)} \end{aligned}$$

where $c_a = \sin(\frac{\pi\alpha}{2})\Gamma(\alpha)/\pi$ and $f(x) \sim g(x)$ as $x \to a$ means $\lim_{x\to a} h(x)/f(x) = 1$. Using the reflection property, the lower tail properties are similar: for $\beta \in [-1,1)]$ as $x \to \infty$:

$$\begin{aligned} F_X(-x) &\sim \gamma^\alpha c_a(1-\beta)x^{-\alpha}, \\ f_X(-x) &\sim \alpha\gamma^\alpha(1-\beta)c_a x^{-(\alpha+1)} \end{aligned}$$

It follows from the above theorem that for $x \to \infty$

$$1 - F(x) \sim Cx^{-\alpha} \implies \ln(1-F(x)) \sim \ln(C) - \alpha\ln(x) \tag{1}$$

Theorem 3. *Let $X \sim S(\alpha,\beta,\gamma,\delta;k)$ wwith $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\gamma \geq 0$, $\delta \in \mathbb{R}$, $k = 0,1$ then*

$$\begin{aligned} |\varphi_X(u)| = e^{-C|u|^\alpha} &\implies \\ \ln|\varphi_X(u)| = -C|u|^\alpha &\implies \\ \ln(-\ln|\varphi_X(u)|) = \ln(C) + \alpha\ln|u| & \end{aligned}$$

Theorems 2 and 3 allows us to determine the stability index using linear regression.

Definition 4. *Let $X_1,\dots,X_n$ be i.i.d. random variables with the common cumulative distribution function $F(t)$. Then the empirical distribution function is defined as*

$$\hat{F_n}(x) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{X_k \leq x\}}$$

Definition 5. *Let $X_1,\dots,X_n$ be i.i.d. random variables with the characteristic function $\varphi(t)$. Then the empirical characteristic function is defined as*

$$\hat{\varphi_n}(u) = \frac{1}{n}\sum_{k=1}^{n}e^{-iuX_k}$$

2 Compare two estimators of α parameter we introduced in the laboratories

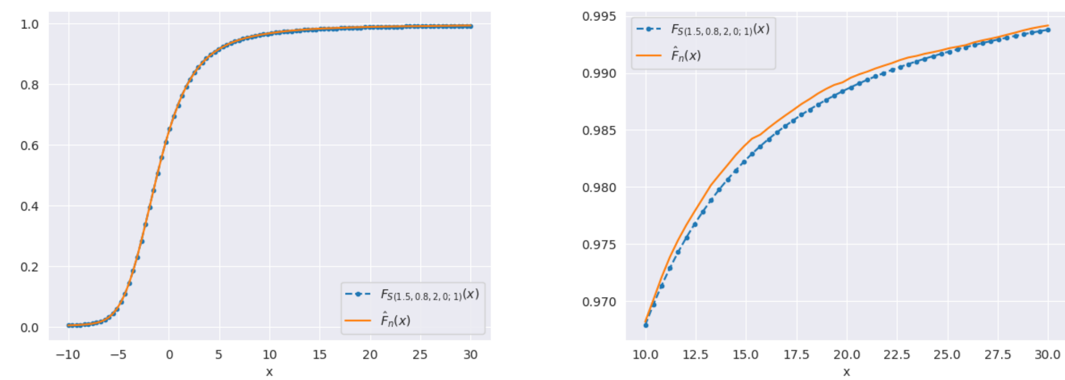
We will be conducting our simulations on consider one set of parameters $(\alpha,\beta,\gamma,\delta) = (1.5, 0.8, 2, 0)$.

2.1 Based on the ECDF

2.1. i)

We gernerated $n = 10^6$ samples of alfa-stable distribution. A figure 1 depicts cumulative distribution function of theoretical variable (blue) compared to generated one (orange). Based on this graph and the KS test p-value $= 0.14623$ we do not reject the hypothesis of different distributions and assume that the program that generates the variables is working correctly.

Graph at the right side shows CDF for higher values of x, thanks to which we can use the first method of estimation parameter α based on theorem 2 and properties 1.



(a) CDF simulated for $x \in (-10, 30)$.

(b) CDF simulated of tail for $x \in (10, 30)$.

Figure 1: CDF of a random variable with an α-stable distribution based on $n = 10^6$ realizations of the random variable.

Following the first method, using logarithm of distribution's tail, we successfully fitted parameters to our model. Results are shown at figure 2, where blue line is a teoretical distribution, orange line represents empirical one and green is a fitted model.

2.1. ii)

After checking the correctness of the method, we estimated parameter α using a Monte Carlo simulation with 1000 steps and 20000 samples on each step. Distribution of estimated parameters can be observed at figure 3, where we have placed density histogram and boxplot. We can see, that the mean of estimation is correct, but there are also a lot of outliers.

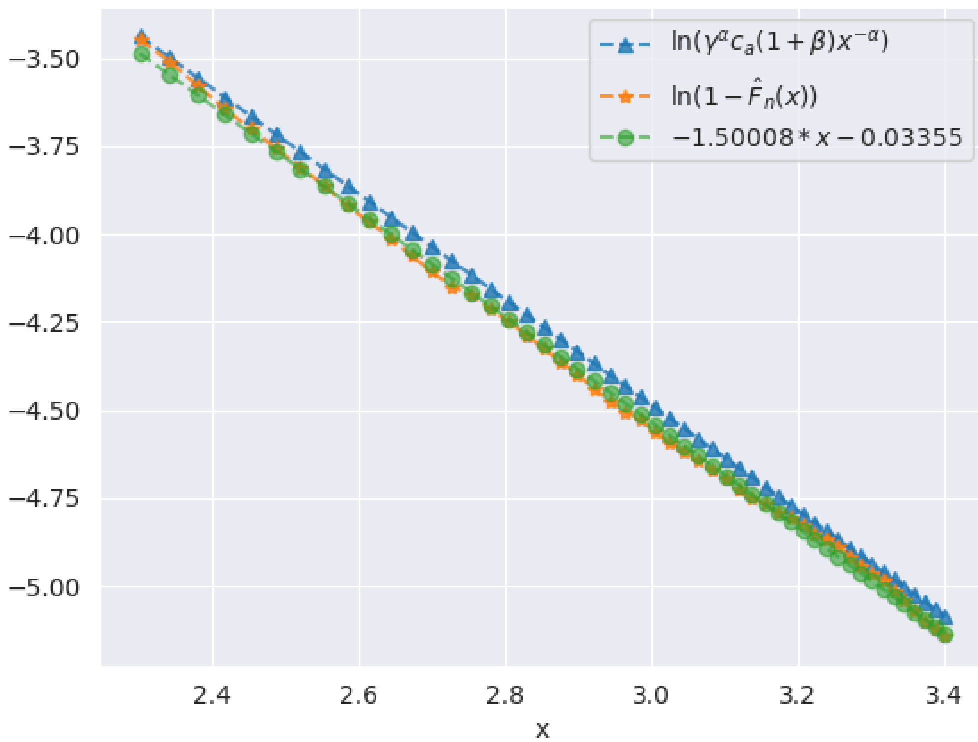
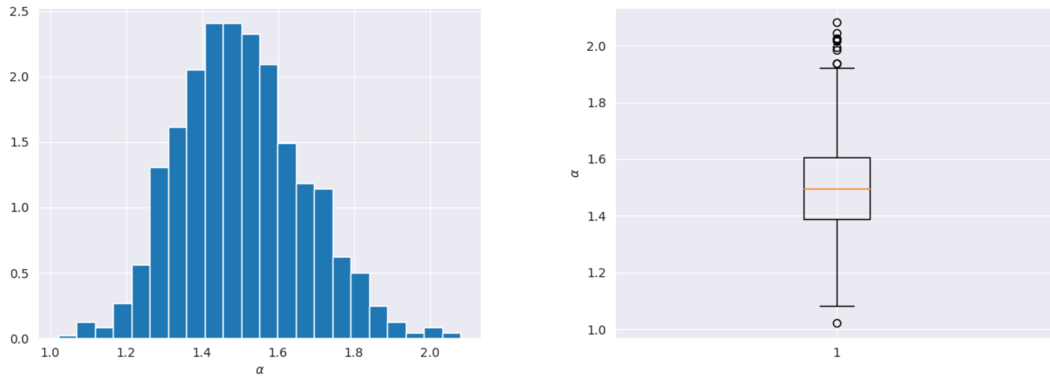


Figure 2: Plot of tail approximation method fitted line.



(a) Histogram of the $\hat{\alpha}$ estimator.

(b) Boxplot

Figure 3: Distribution of the estimator of the α parameter based on Monte Carlo simulations for tail approximation method.

In addition, we have included basic statistics in the table 1. We can see, that the distribution of estimator has high std. Maximal and minimal values are absolutly too far from true value of $\hat{\alpha}$. Skewness is equal to 0.28487 and with the analysis of the graph we assess, that distribution is right-skewed. The kurtosis is close to zero, so we can assume that it is leptokurtic distribution, so we can expect havy tails and this is correct with values of quantiles.

count	mean	std	min	25%	50%	75%	max	skewness	kurtosis
1000	1.5081	0.1697	0.994552	1.3893	1.5051	1.6112	2.0974	0.2849	0.0931

Table 1: Table of basic statistics of the $\hat{\alpha}$ estimator for tail approximation method.

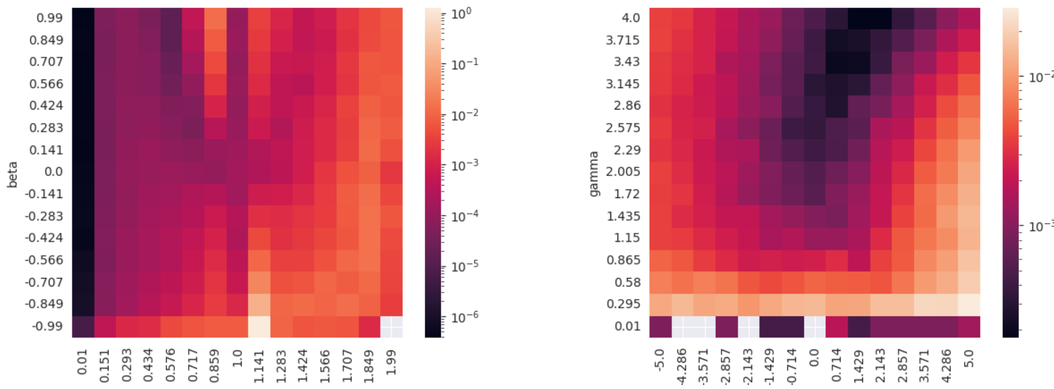
2.1. iii)

We checked, how selected parameters affect MSE and MAE. We made simulation of 100 steps of Monte Carlo and 5000 samples of generated $\hat{\alpha}$, to create adequate heatmaps.

On figure 4 we placed results for Mean Squared Error. At the left side we inserted dependencies of α and β using $\gamma = 2$ and $\delta = 0$ and at the right side we placed dependencies of γ and δ using $\alpha = 1.5$ and $\beta = 0.8$

The main conclusion is, that the smaller β we take, the worse results we obtain. The same result we get for α.

We get smaller errors within increasing γ, but for δ we obtain the best results where it is equal to 0 and farther we go from zero, then worse. Delta should not have influance on the highest of error becouse this method uses aproximation of infinitive values, but our domain is the range from 10 to 30. We did it becouse we decided to optimalise the process of generating $\hat{\alpha}$, so in this case, selection of δ has big influance on the generated tails. Anyway, the differences of results depended on γ and δ are smaller than dependent on α and β.

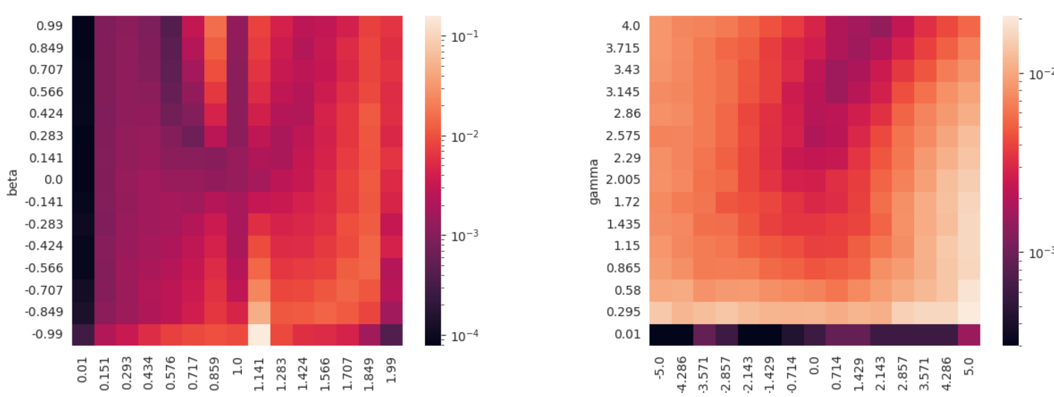


(a) Depending on α, β parameters and $\gamma = 2, \delta = 0$

(b) Depending on γ, δ parameters and $\alpha = 1.5, \beta = 0.8$

Figure 4: Heatmaps of Mean Squared Error (MSE) based on Monte Carlo simulations for tail approximation method.

We created also a heatmaps of MAE (figure 5). We used the same parameters as in the case with MSE. In this example we obtain the same dependences, but the only one difference is with the scale of errors, what is meaningful, if we would check the formula of MAE.



(a) Depending on α, β parameters and $\gamma = 2, \delta = 0$

(b) Depending on γ, δ parameters and $\alpha = 1.5, \beta = 0.8$

Figure 5: Heatmaps of Mean Absolute Error (MAE) based on Monte Carlo simulations for tail approximation method.

2.2 Based on the CF

2.2. i)

Now we will consider second method of estimation parameter α [3]. First we crated an empirical characteristic function by generating 10000 samples from alpha-stabil distribution. At figure 6 we presented comparison of theoretical (orange line) and empirical (blue line) CF. We can see, that empirical plot is similar to theoretical one.

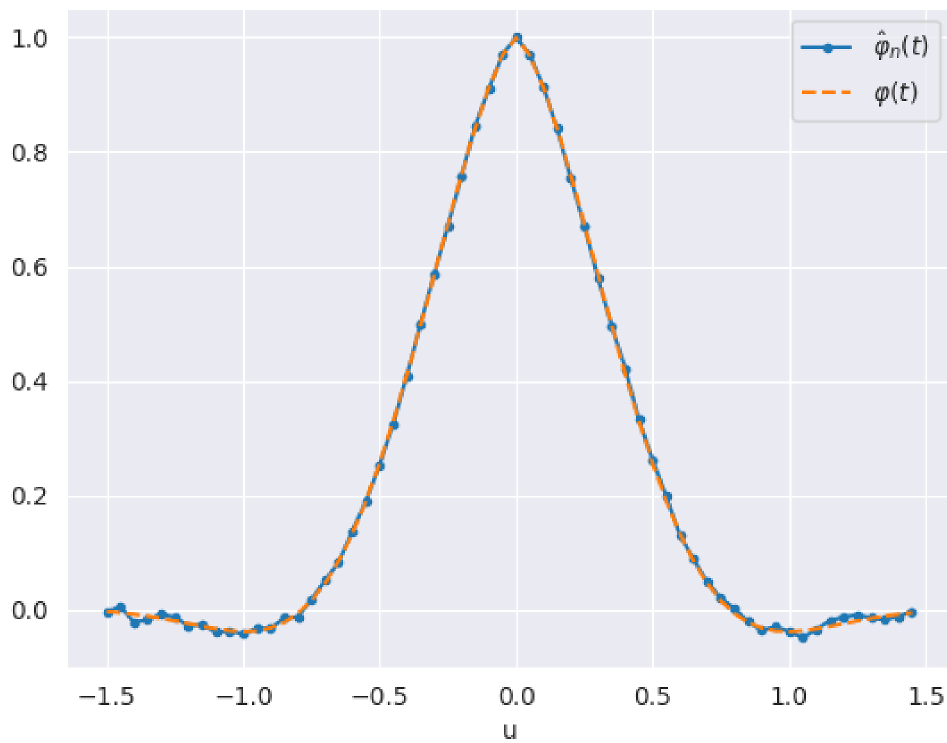


Figure 6: CF of a random variable with an α-stable distribution based on $n = 10000$ realizations of the random variable for each t.

Following this method, using double logarithm on CF, we successfully fitted parameters to our model. Correctness of the fitting is shown in figure 8a, where blue line is theoretical function, orange in estimated one and green is a line represanting suitable model.

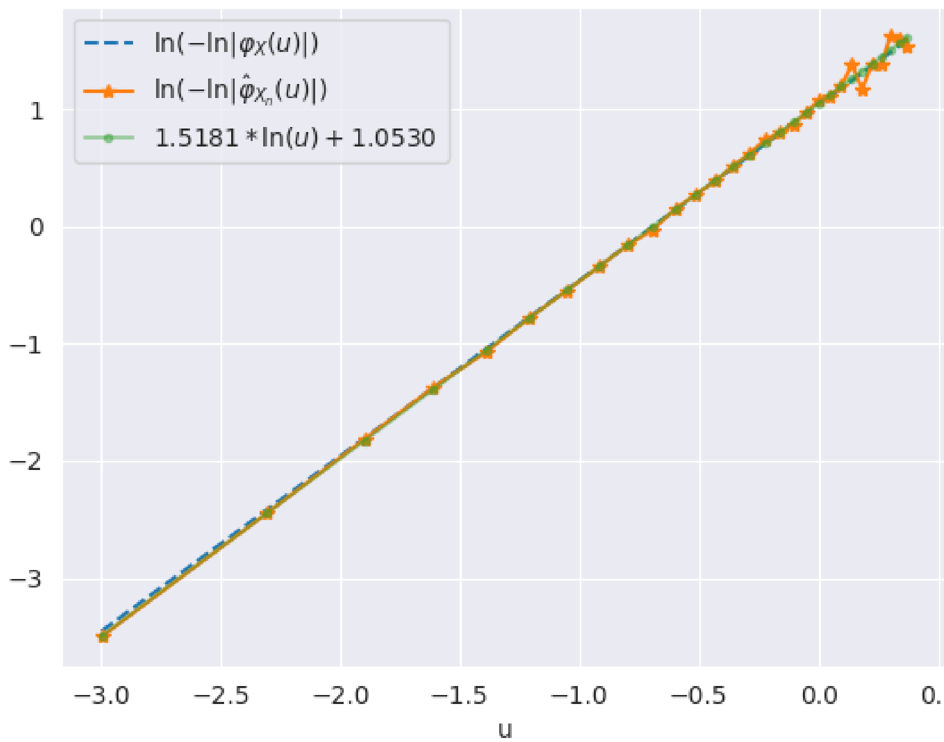
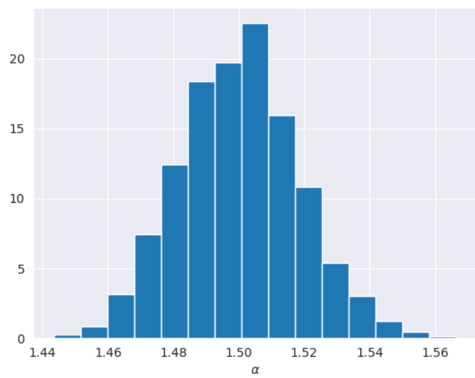


Figure 7: Plot of characteristic function method fitted line.

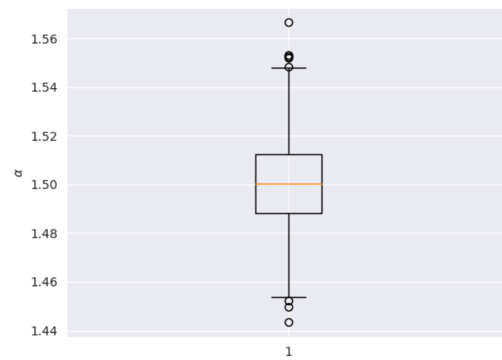
2.2. ii)

We checked the correctness of this method by estimating parameter α using a Monte Carlo simulation with 1000 steps and 5000 samples on each step.

Distribution of estimated parameters can be observed in figure 8, where we have placed density histogram and boxplot and we have included basic statistics in the table 2. We can see, that the mean of estimation is correct. We obtain better std then using tail approximation method. Whatsmore, in this case, the skewness is low, close to 0. The kurtosis is equal to -0.1926, so the distribution is platokurtic.



(a) Histogram of the $\hat{\alpha}$ estimator.



(b) Boxplot of the $\hat{\alpha}$ estimator.

Figure 8: Distribution of the estimator of the α parameter based on Monte Carlo simulations for characteristic function method.

count	mean	std	min	25%	50%	75%	max	skewness	kurtosis
1000.0	1.4989	0.0185	1.4468	1.4868	1.498	1.5118	1.5543	0.0294	-0.1926

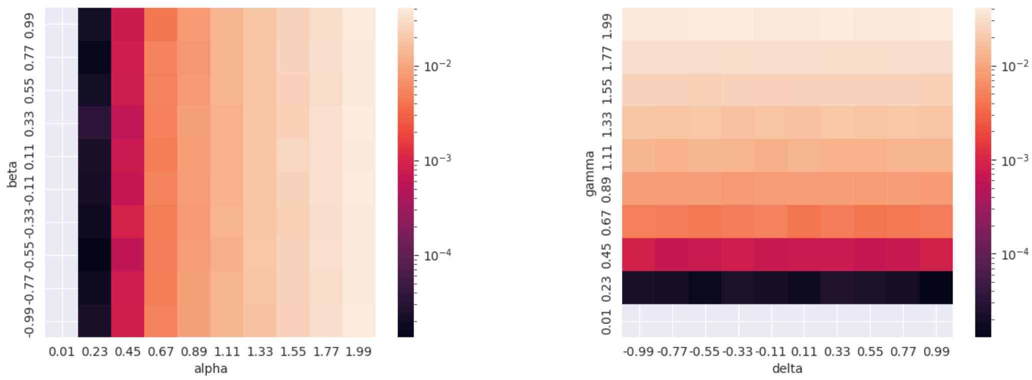
Table 2: Table of basic statistics of the $\hat{\alpha}$ estimator for tail approximation method.

2.2. iii)

We checked, how selected parameters affect MSE and MAE. We made simulation of 100 steps of Monte Carlo and 5000 samples of generated $\hat{\alpha}$, to create adequate heatmaps.

On figure 9 we placed results for Mean Squared Error. At the left side we inserted dependencies of α and β using $\gamma = 2$ and $\delta = 0$ and at the right side we placed dependencies of γ and δ using $\alpha = 1.5$ and $\beta = 0.8$

In this case, the dependencies are really visuable. We get bigger errors within increasing α and γ. Changing β and δ has no effect on the size of the error.



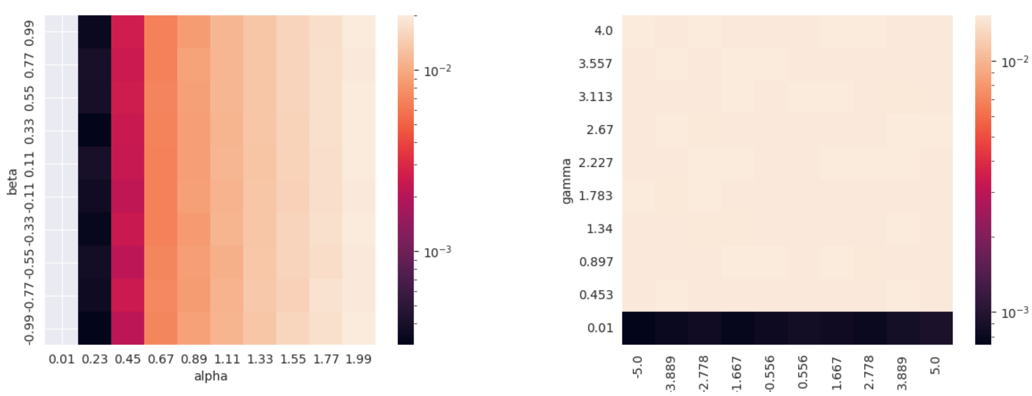
(a) Depending on α, β parameters and $\gamma = 2, \delta = 0$

(b) Depending on γ, δ parameters and $\alpha = 1.5, \beta = 0.8$

Figure 9: Heatmaps of Mean Squared Error (MSE) based on Monte Carlo simulations for characteristic function method.

On figure 10 we placed results for Mean Absolute Error. At the left side we inserted dependencies of α and β using $\gamma = 2$ and $\delta = 0$ and at the right side we placed dependencies of γ and δ using $\alpha = 1.5$ and $\beta = 0.8$

The conclussion is that the size of error depend only on α. The bigger α, the higher error we get. High error at the bottom of right graph is coused by taking γ to closed to 0, which is out of domain of alpha-stable distribution.



(a) Depending on α, β parameters and $\gamma = 2, \delta = 0$

(b) Depending on γ, δ parameters and $\alpha = 1.5, \beta = 0.8$

Figure 10: Heatmaps of Mean Absolute Error (MAE) based on Monte Carlo simulations for characteristic function method.

2.3 Summary

We presented the operation of both methods, which estimate parameter α. We proved, that both of them works correctly. Errors, we obtianed by using the method which bases on characteristic function, depend strongly on parameter α.