Generation of 0.01 Quantile Distribution for 10-Day Returns

Introduction

This report examines the process of generating the 0.01 quantile distribution for 10-day returns obtained from a time series of 1-day returns generated from a stable distribution. We will use the Monte Carlo method for numerical estimation and verify whether the chosen number of simulations is sufficient to achieve the required accuracy.

Data Generation

About Generating the Original Data

For the creation of the base time series, random noise was generated using a stable Lévy distribution. This distribution is characterized by heavy tails and asymmetry, making it suitable for modeling financial time series where extreme events and anomalous fluctuations are often observed.

The noise was generated using the `levy_stable.rvs` function from the SciPy library, which allows for the specification of parameters of the stable distribution:

$$\text{Noise}_t = \text{Stable}(\alpha, \beta, \gamma, \delta), \quad t = 1, \ldots, n_{\text{days}}$$

Where:

- $\alpha = 1.7$ — the "tail thickness" parameter, which indicates the presence of heavy tails, important for modeling rare but large events;
- $\beta = 0.0$ — the symmetry parameter, indicating that the distribution is symmetric (no trend);
- $\gamma = 1.0$ — the scale parameter, which defines the variance;
- $\delta = 1.0$ — the shift parameter, allowing for the specification of the average return.

The generated noise represents random fluctuations that can serve as the basis for constructing more complex time series.

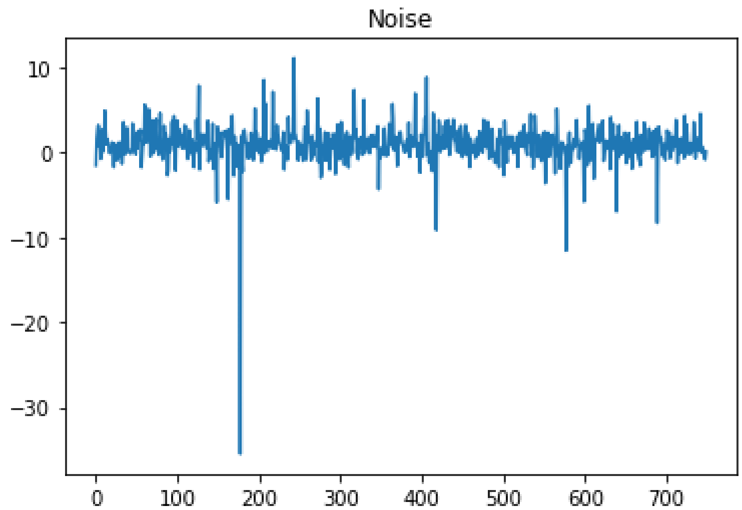

Fig. 1. Generated 1-Day Returns

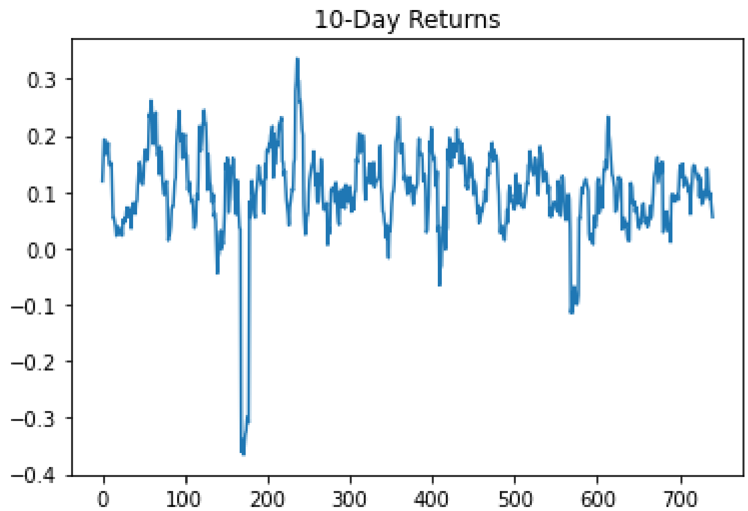

Fig. 2. Generated 10-Day Returns

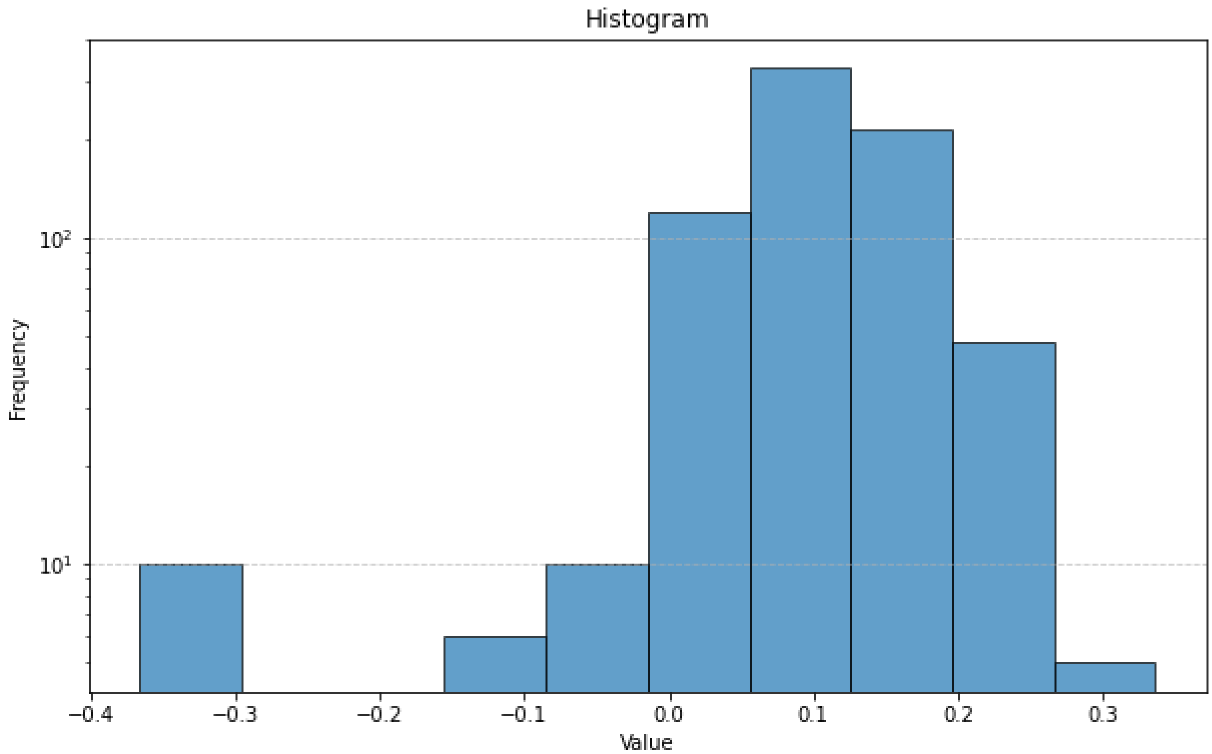

Fig. 3. Distribution Histogram

About Generating Other Time Series

About Generating the Autoregressive Time Series

For generating a time series with autocorrelation, a first-order autoregressive process (AR(1)) was used. This process models the dependence of the current value on the previous one, which is typical for many financial time series where "memory" effects are observed.

Mathematically, the AR(1) process is described by the recurrence formula:

$$r_t = \phi r_{t-1} + \epsilon_t, \quad t = 1, \ldots, n_{\text{days}}$$

Where:

- r_t — the current value of the time series,
- $\phi = 0.8$ — the autocorrelation coefficient, which determines the degree of dependence of the current value on the previous one,
- $\varepsilon_t \sim \text{Stable}(\alpha, \beta, \gamma, \delta)$ — random noise generated by the Lévy distribution.

This process takes into account the autocorrelation effect, where current changes in the time series partially depend on previous values. Such a process is often used to model financial time series with long-term dependencies.

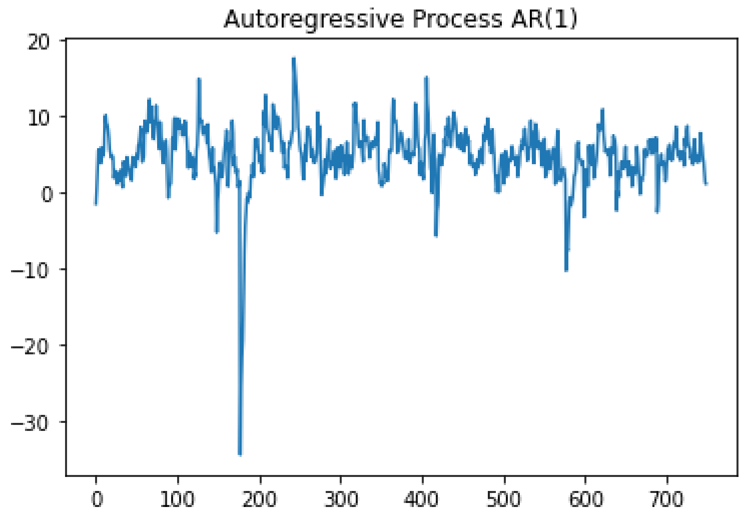

Fig. 4. Generated 1-Day Returns

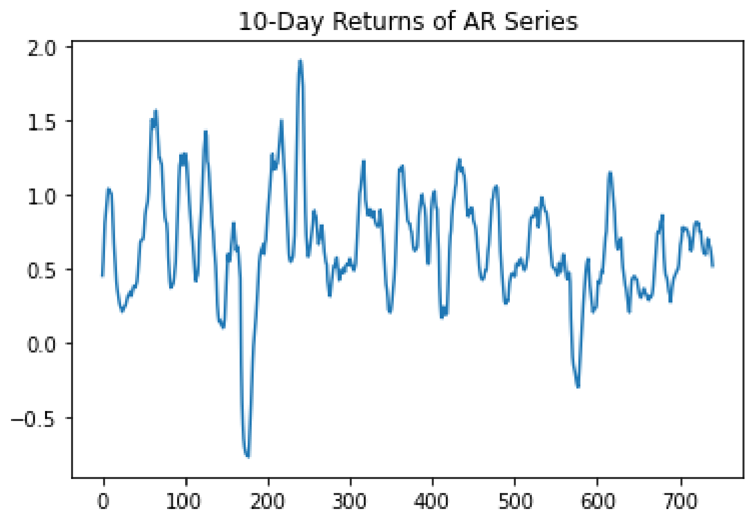

Fig. 5. Generated 10-Day Returns

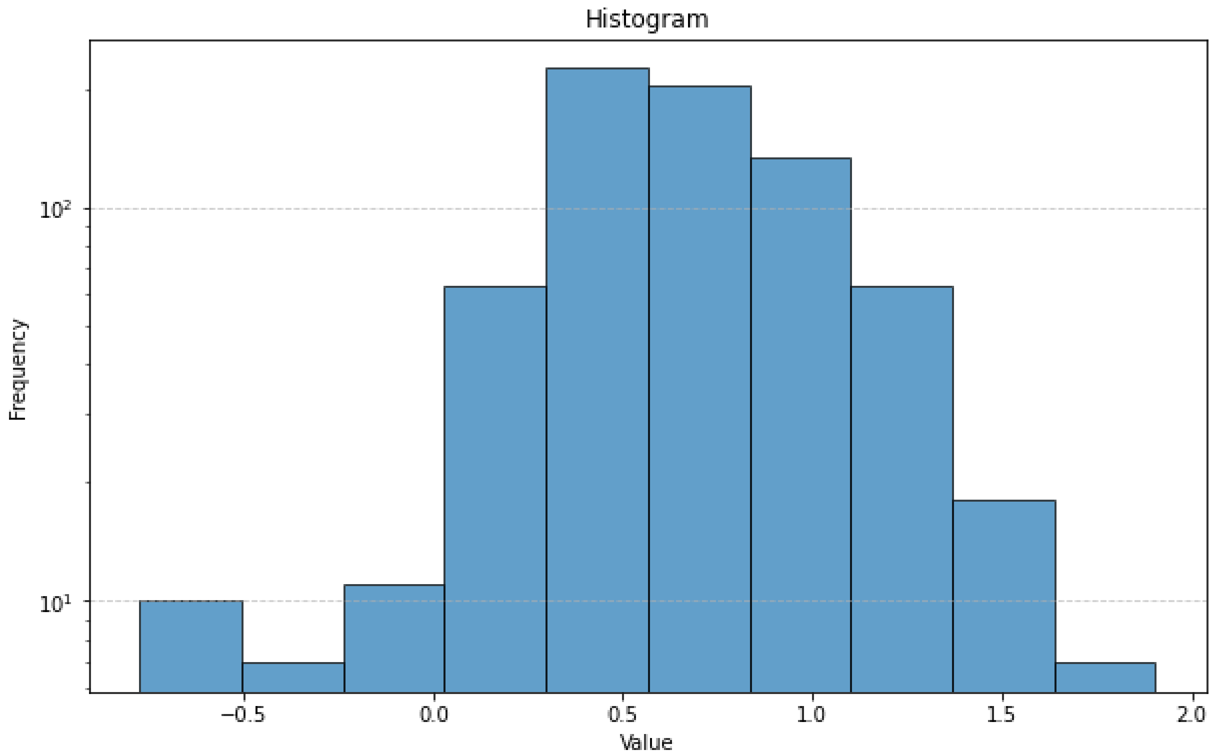

Fig. 6. Distribution Histogram

About Generating a Complex Time Series

For more complex modeling, a combination of trend, seasonality, and noise was used. In this case, the time series includes:

- A linear trend, described by the formula $\beta_0 + \beta_1 t$, where $\beta_1 = 0.03$ — the trend coefficient determining the rate of change of the time series over time;
- Seasonality, modeled using a sinusoidal function $\sin\left(\frac{2\pi t}{T}\right)$, where $T = 500$ — the period of seasonal fluctuations, and $A = 5$ — the amplitude of the seasonal fluctuation.

The complete model for generating the complex time series is:

$$r_t = \beta_0 + \beta_1 t + A \sin\left(\frac{2\pi t}{T}\right) + \epsilon_t, \quad t = 1, \ldots, n_{\text{days}}$$

Where:

- β_0 — initial return,
- β_1 — rate of growth or decline,
- A — amplitude of seasonal fluctuations,
- ε_t — random noise.

Thus, we obtain a time series with long-term trends, seasonal fluctuations, and random fluctuations, which is a more accurate representation of real data than just using noise.

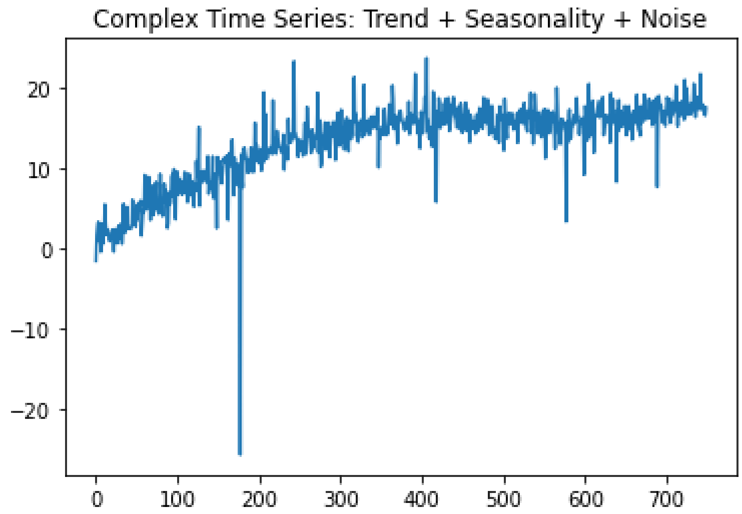

Fig. 7. Generated 1-Day Returns

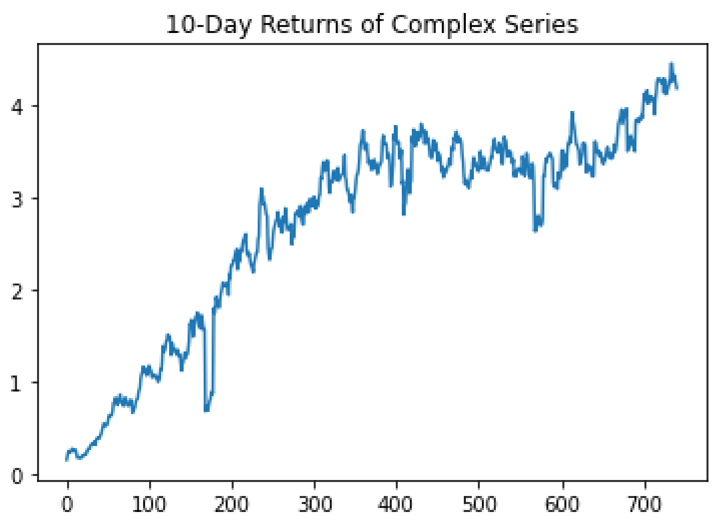

Fig. 8. Generated 10-Day Returns

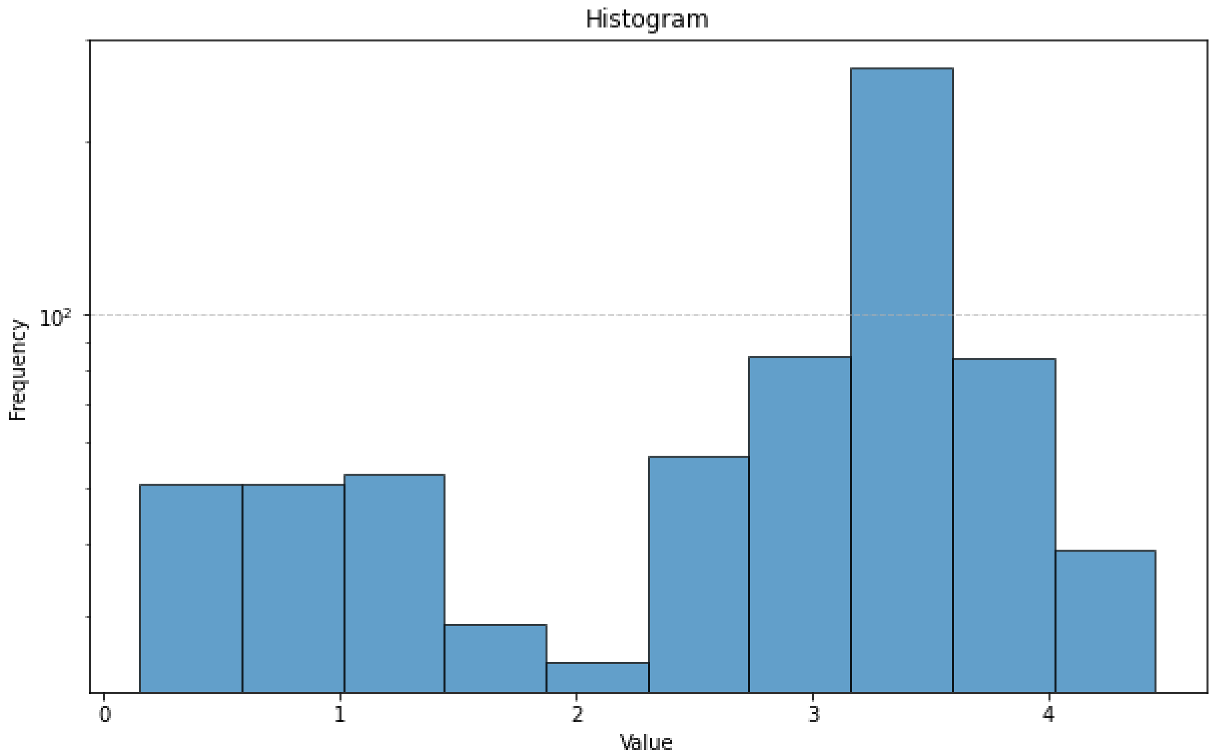

Fig. 9. Distribution Histogram

About Generating 10-Day Returns

After generating time series for all data types (noise, AR(1), complex series), 10-day returns were calculated using the overlapping window method. This method allows for a denser sample, as each new interval includes information from previous periods.

The 10-day return for each interval is calculated using the following formula:

$$R_t^{(10)} = \prod_{i=0}^{9}(1 + r_{t+i}) - 1, \quad t = 1, \ldots, 741,$$

Where:

- r_t — the return value on day t,
- $R_t^{(10)}$ — the 10-day return for the interval from t to $t+9$.

This approach allows for overlapping periods and provides a more accurate estimate of the 10-day return, which is especially important for evaluating rare events such as extreme losses or gains, which can be useful for risk analysis.

After obtaining the 10-day returns for each series, their quantile values were calculated, including the 0.01 quantile, allowing the study of extreme events in the data.

Simulation

About the Monte Carlo Method

The Monte Carlo method is a numerical technique that uses random samples for approximate solutions to problems that are difficult to solve analytically. This method is used for estimating statistical characteristics of random processes, such as quantiles (e.g., the 0.01 quantile), by repeatedly simulating random data.

In this case, we use the Monte Carlo method to estimate the 0.01 quantile for different types of time series, including random noise, AR(1) process, and complex time series with trend and seasonality. The main process involves generating random samples from a Lévy stable distribution and calculating the quantile for each sample. As the number of simulations increases, the quantile estimate becomes more accurate.

Formally, the quantile estimation task can be expressed as:

$$Q_\alpha = \text{Quantile}(X_1, X_2, \ldots, X_N, \alpha)$$

Where $X_1, X_2, \ldots, X_N$ are N independent random variables generated according to the given distribution, and α is the quantile of interest (e.g., the 0.01 quantile).

About Convergence Checking

To ensure that the simulation results are reliable and accurate, it is important to check the convergence of quantile estimates. This means that as the number of simulations increases, the quantile estimates should stabilize and converge to the true value.

Several methods are used for convergence checking:

- **Stopping Criterion**: The simulation process continues until the difference between the mean quantile values at the current and previous simulation steps becomes smaller than a given threshold ϵ:

$$|\text{mean}(\text{quantiles}) - \text{mean}(\text{quantiles}[:-1])| < \epsilon.$$

This criterion checks whether the quantile values have stabilized, and if so, the simulation process can be stopped.

- **Statistical Tests**: To further verify the stability of results, various statistical tests can be applied. One such test is the **Kruskal-Wallis test**, which checks for statistically significant differences between multiple data groups. This helps ensure that the quantile distribution does not change as the number of simulations increases.
- **Standard Error**: To evaluate the accuracy of the results, the standard error is calculated. For simulations, it is determined as:

$$\text{SE} = \frac{\sigma}{\sqrt{N}},$$

Where σ is the standard deviation of the quantiles, and N is the number of simulations. The smaller the standard error, the more accurate the quantile estimate.

About the Kruskal-Wallis Test

The Kruskal-Wallis test is a non-parametric statistical test used to check the hypothesis of equality of medians across several independent samples. In the context of quantile simulation, it is used to assess whether the quantile results have stabilized with increasing simulations, and whether the quantile distributions change across different simulations.

The test is based on computing the statistic H, defined as:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{n_i(\bar{R}_i - \bar{R})^2}{s^2}$$

Where:

- N — total number of observations,
- k — number of groups (in this case, the number of simulations),
- n_i — number of observations in the i-th group,
- $\bar{R}_i$ — average rank in group i,
- $\bar{R}$ — overall average rank across all groups,
- s^2 — overall variance term.

If the computed H statistic exceeds the critical value for a given significance level (e.g., the p-value is less than 0.05), the null hypothesis is rejected, indicating statistically significant differences between the groups. This suggests that the simulation results have not stabilized yet.

In the context of quantile estimation, the Kruskal-Wallis test helps verify whether the quantile distribution is changing as the number of simulations increases. If significant differences are found, it may indicate that more simulations are needed for convergence.

About the Simulation Results

As a result of running Monte Carlo simulations for various types of processes (noise, AR(1), complex time series), we obtain the following data:

1. **Number of Simulations**: After the stopping criterion is met, the number of simulations is recorded. This number depends on the convergence of quantile estimates — the process continues until the quantile values stabilize.

2. **Average Quantile Values**: For each simulation, the 0.01 quantile is calculated, and then the average value for all simulations is computed. This helps assess the central tendency of the quantiles.
3. **Standard Deviation**: The standard deviation of the quantiles helps understand the variability of the estimates. The smaller the deviation, the more accurate the estimate.
4. **Histograms of Quantile Distributions**: Histograms visually show how the quantile values distribute across each simulation iteration. This helps identify the stability or instability of the results.
5. **Kruskal-Wallis Test**: By applying this test, we check if there are statistically significant differences between the quantile groups. If the differences are significant, this suggests that the simulations have not yet stabilized.
6. **Standard Error Estimation**: The standard error helps evaluate the accuracy of the results, taking into account the number of simulations.
7. **Theoretical Error**: The theoretical error is evaluated as the ratio of the standard deviation to the square root of the number of simulations, providing an estimate of how accurately the quantile was estimated.

Ultimately, the results of the simulation and statistical tests allow us to assess how accurate and stable the quantile estimates are for different processes, and to check their statistical properties for convergence and accuracy.

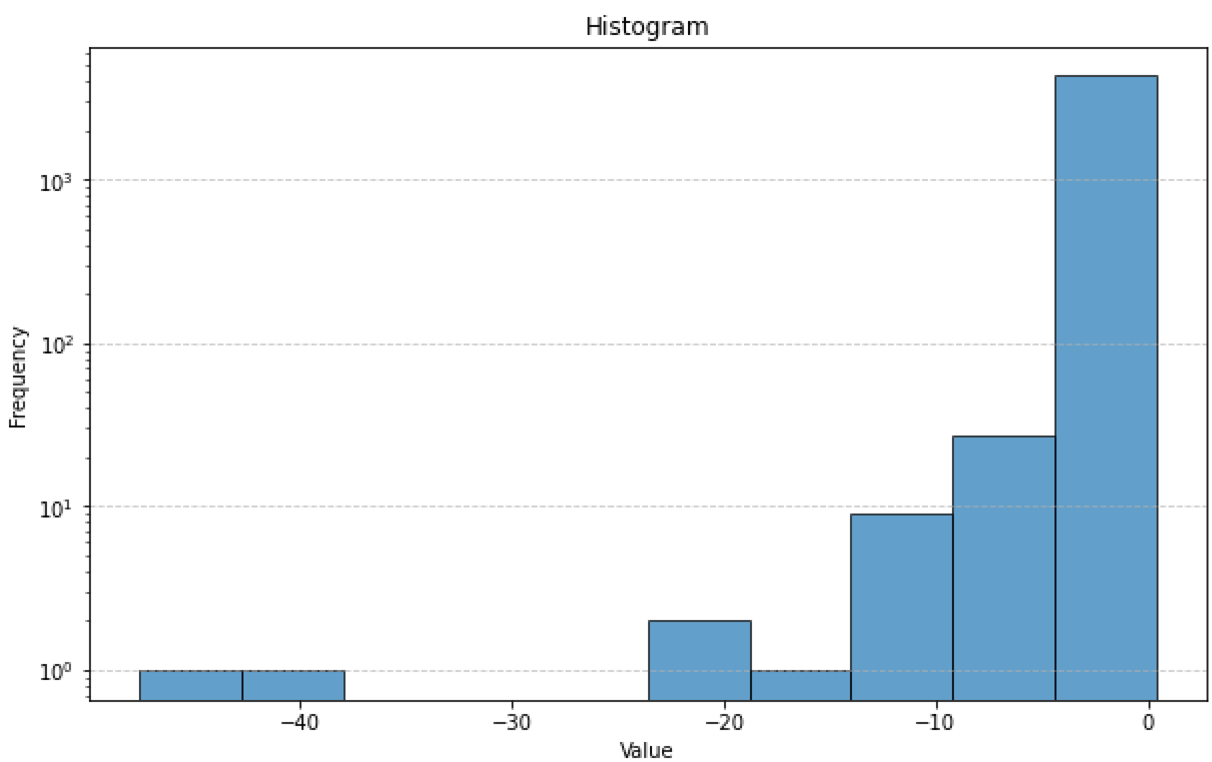

Fig. 10. Quantile Distribution

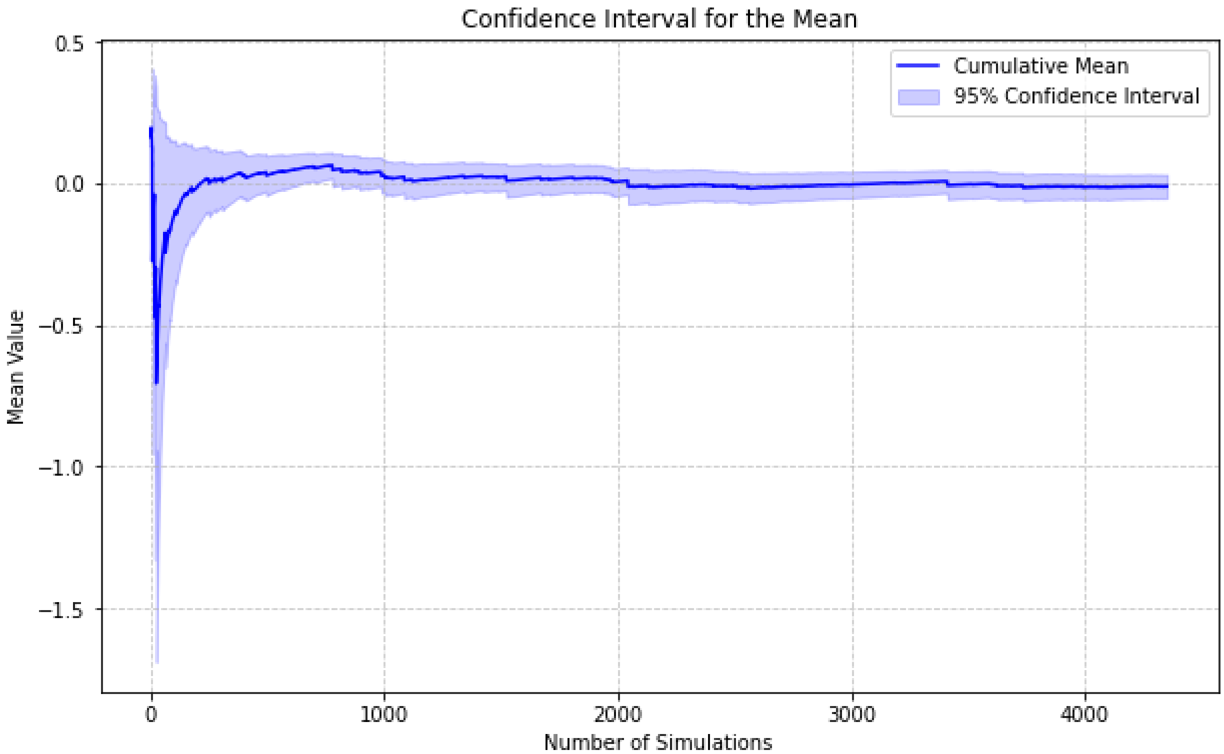

Fig. 11. Cumulative Mean and Confidence Intervals

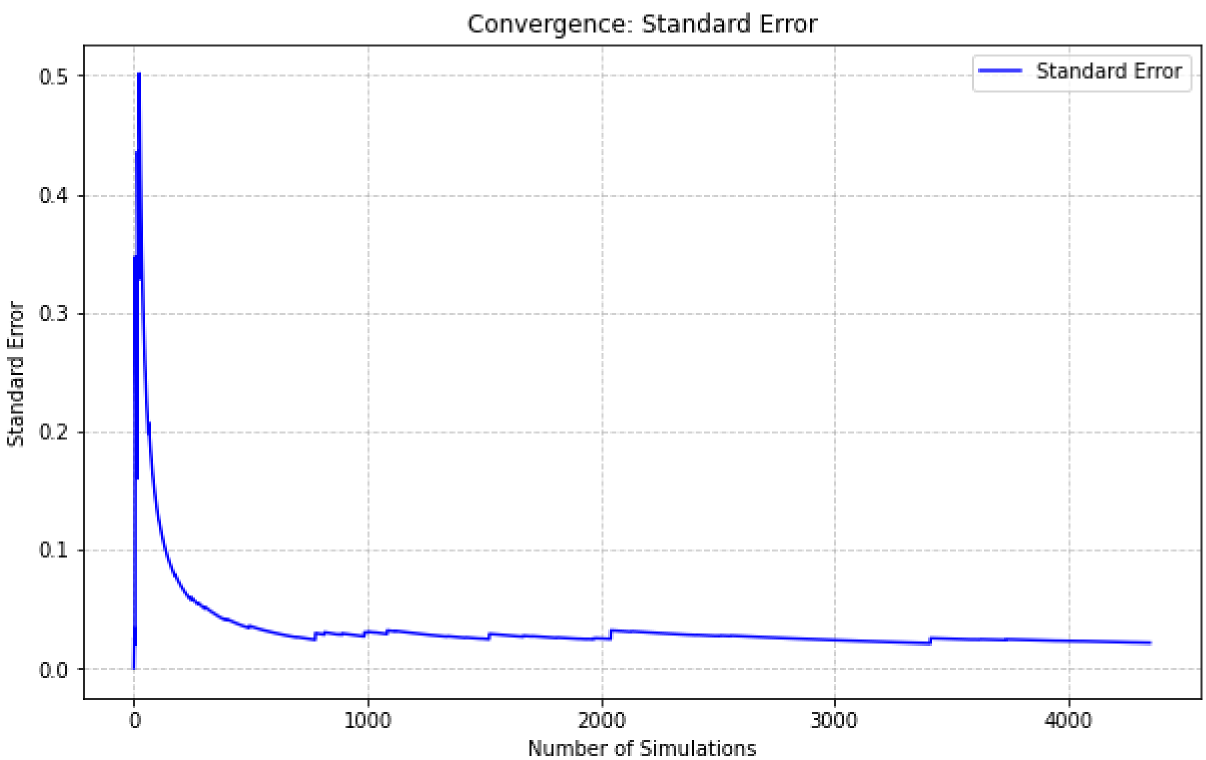

Fig. 12. Standard Error Graph