Генерация распределения 0.01-квантили 10-дневных доходностей

Введение

В этом отчете рассмотрен процесс генерации распределения 0.01-квантили для 10-дневных доходностей, полученных из временного ряда 1-дневных доходностей, сгенерированных из устойчивого распределения. Мы будем использовать метод Монте-Карло для численной оценки и проверим достаточность выбранного числа симуляций для достижения требуемой точности.

Генерация данных

О генерации изначальных данных

Для создания базового временного ряда был использован случайный шум, сгенерированный с помощью устойчивого распределения Леви. Это распределение характеризуется тяжёлыми хвостами и асимметрией, что делает его подходящим для моделирования финансовых временных рядов, где часто наблюдаются экстремальные события и аномальные колебания.

Генерация шума была выполнена с использованием функции `levy_stable.rvs` из библиотеки SciPy, которая позволяет задать параметры устойчивого распределения:

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

где:

- $\alpha = 1.7$ — параметр "толщины хвостов" распределения, который указывает на наличие тяжёлых хвостов, что важно для моделирования редких, но крупных событий;
- $\beta = 0.0$ — параметр симметрии, указывающий на симметричность распределения (отсутствие тренда);
- $\gamma = 1.0$ — масштабный параметр, определяющий дисперсию;
- $\delta = 1.0$ — сдвиг, который позволяет задать среднее значение доходности.

Генерируемый шум представляет собой случайные колебания, которые могут служить основой для построения более сложных временных рядов.

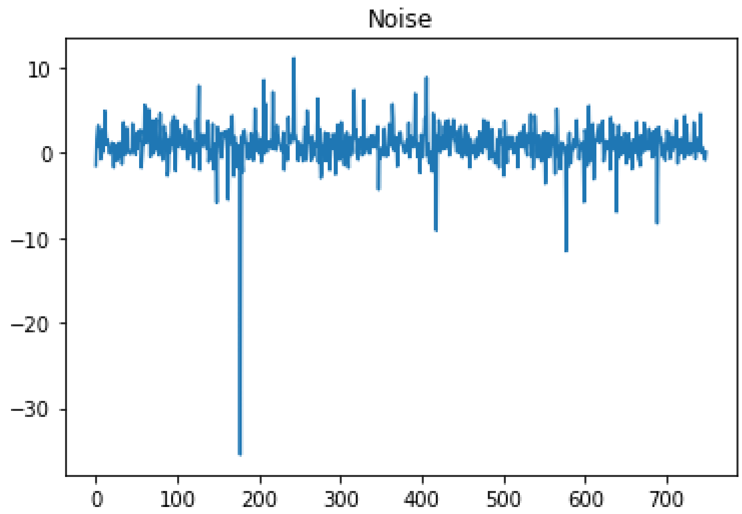

Рис.1. Сгенерированные однодневные доходности

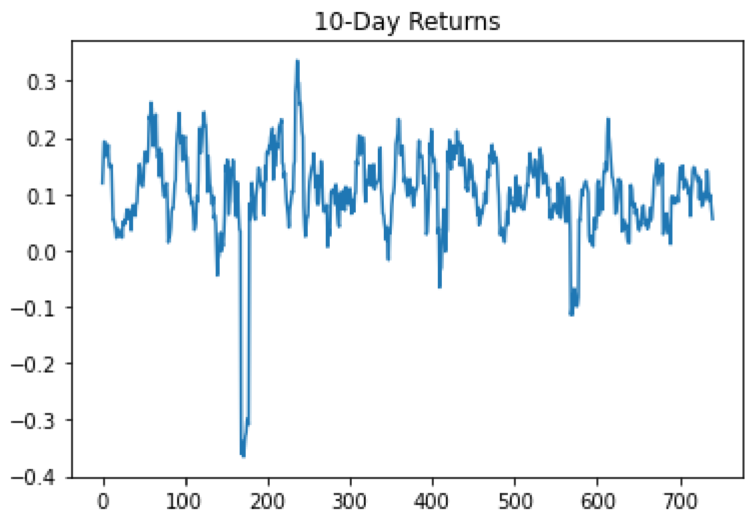

Рис.2. Сгенерированные десятидневные доходности

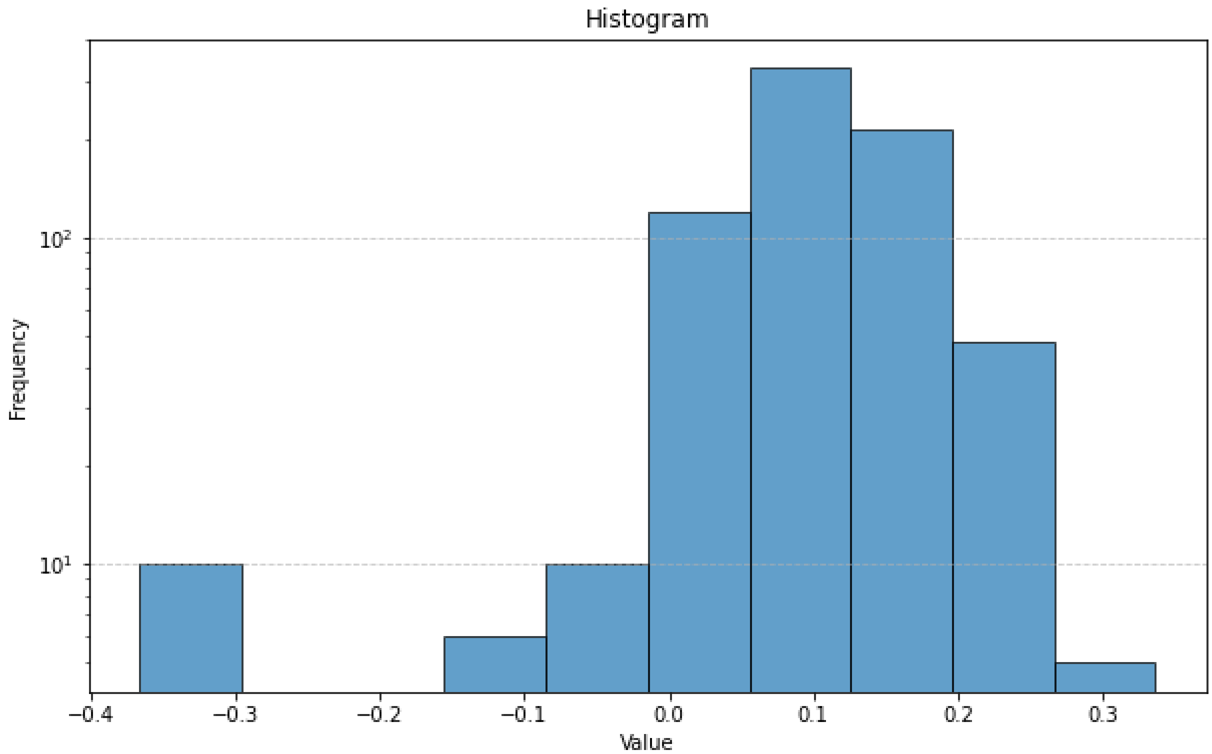

Рис.3. Гистограмма распределения

О генерации остальных временных рядов

О генерации авторегрессионного временного ряда

Для создания временного ряда с автокорреляцией был использован процесс авторегрессии первого порядка (AR(1)). Этот процесс моделирует зависимость текущего значения от предыдущего, что характерно для многих финансовых временных рядов, где наблюдаются эффекты "памяти".

Математически процесс AR(1) описывается рекуррентной формулой:

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

где:

- r_t — текущее значение временного ряда,
- $\phi = 0.8$ — коэффициент автокорреляции, который определяет степень зависимости текущего значения от предыдущего,
- $\varepsilon_t \sim \text{Stable}(\alpha, \beta, \gamma, \delta)$ — случайный шум, сгенерированный по распределению Леви.

Этот процесс позволяет учесть эффект автокорреляции, где текущие изменения во временном ряде частично зависят от предыдущих значений. Такой процесс часто используется для моделирования финансовых временных рядов с долговременной зависимостью.

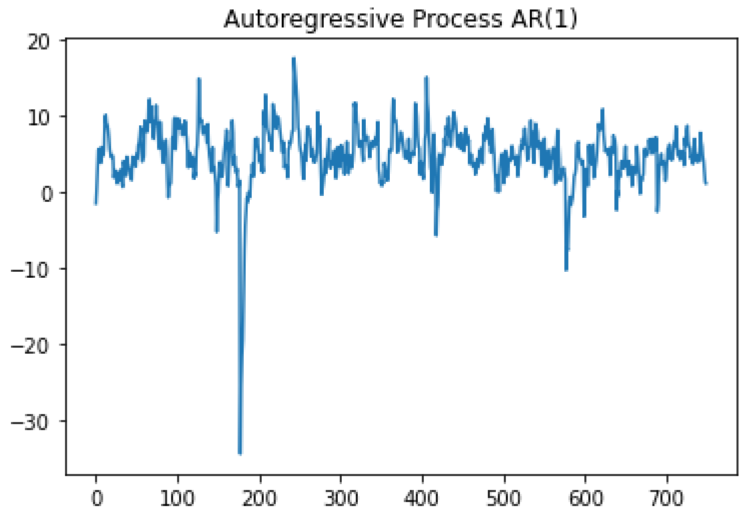

Рис.4. Сгенерированные однодневные доходности

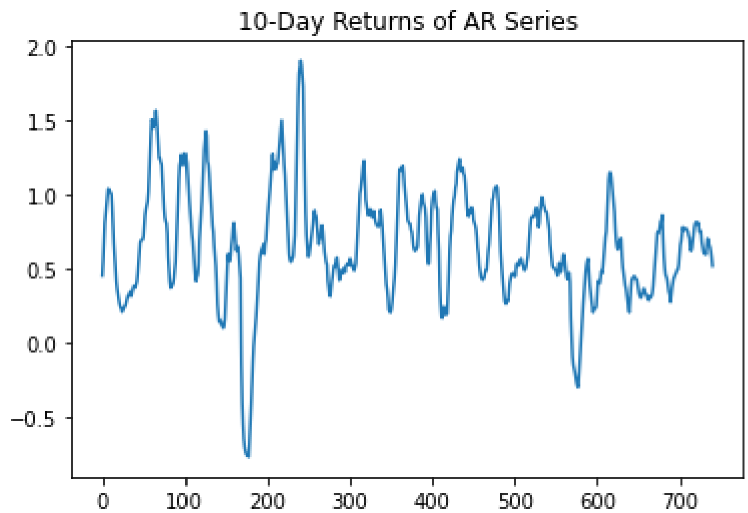

Рис.5. Сгенерированные десятидневные доходности

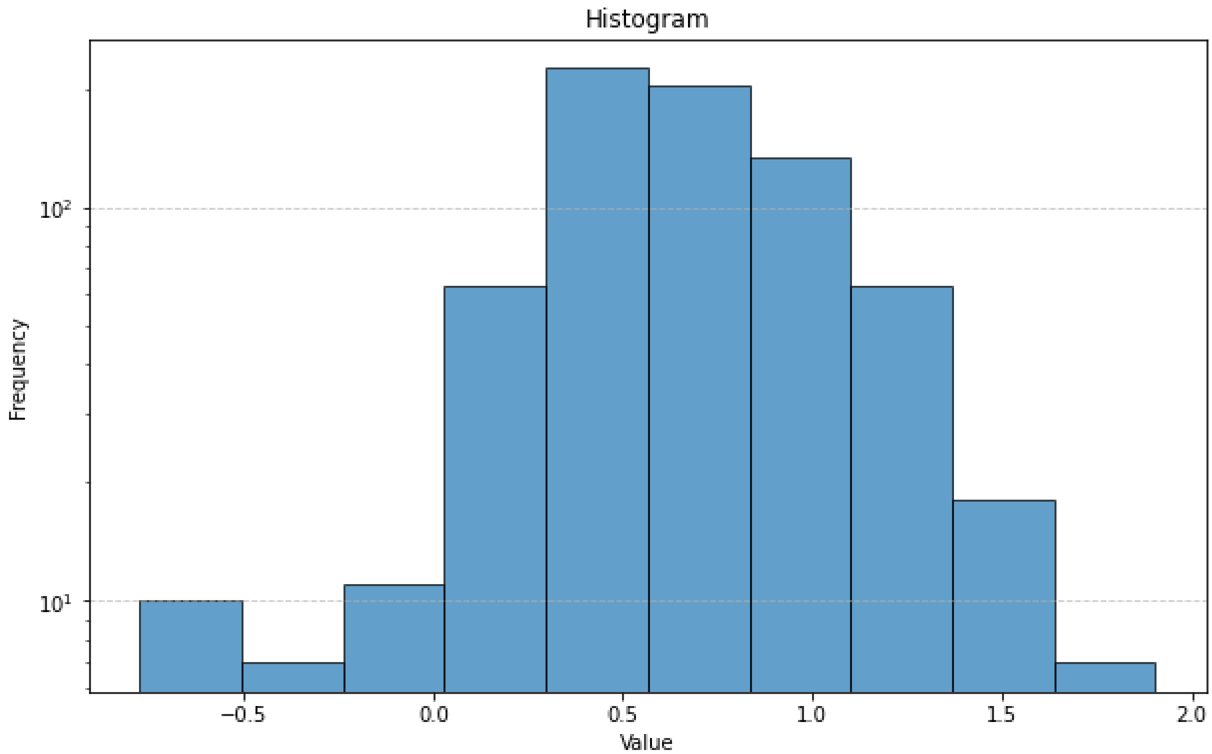

Рис.6. Гистограмма распределения

О генерации сложного временного ряда

Для более сложного моделирования использовалась комбинация тренда, сезонности и шума. В этом случае временной ряд включает в себя:

- Линейный тренд, который описывается формулой $\beta_0 + \beta_1 t$, где $\beta_1 = 0.03$ — коэффициент тренда, определяющий скорость изменения временного ряда со временем;
- Сезонность, которая моделируется с помощью синусоидальной функции $\sin\left(\frac{2\pi t}{T}\right)$, где $T = 500$ — период сезонных колебаний, а $A = 5$ — амплитуда сезонного колебания.

Полная модель для генерирования сложного временного ряда:

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

где:

- β_0 — начальная доходность,
- β_1 — скорость роста или снижения,
- A — амплитуда сезонных колебаний,
- ε_t — случайный шум.

Таким образом, получаем временной ряд с долговременной тенденцией, сезонными колебаниями и случайными флуктуациями, что является более точным представлением реальных данных, чем просто использование шума.

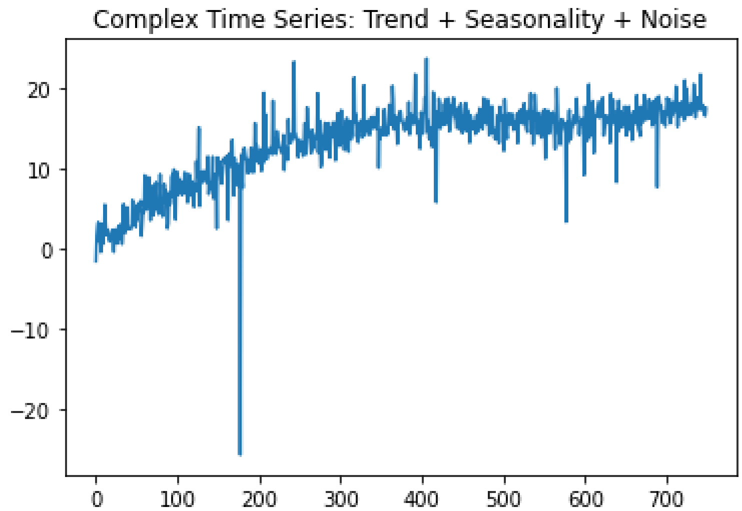

Рис.7. Сгенерированные однодневные доходности

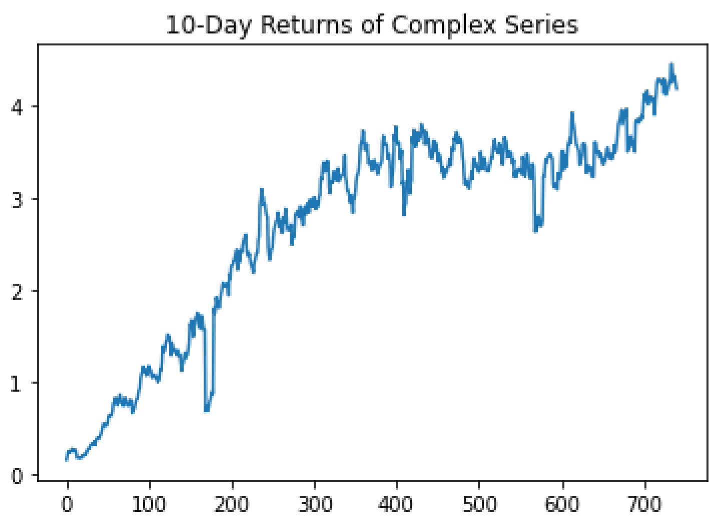

Рис.8. Сгенерированные десятидневные доходности

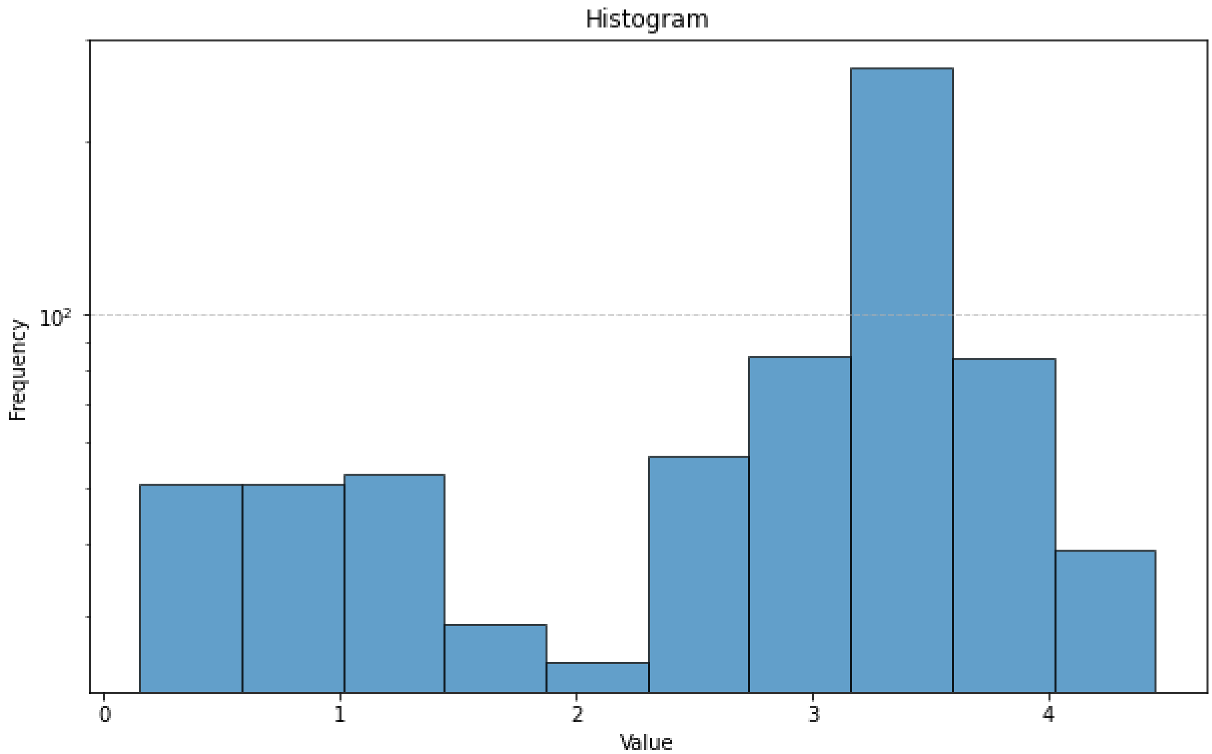

Рис. 9. Гистограмма распределения

О генерации 10-дневных доходностей

После генерации временных рядов для всех типов данных (шум, AR(1), сложный ряд) были рассчитаны 10-дневные доходности с использованием метода перекрывающихся окон. Этот метод позволяет получить более плотную выборку, так как каждый новый интервал включает информацию о предыдущих периодах.

10-дневная доходность для каждого интервала рассчитывается по следующей формуле:

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

где:

- r_t — значение доходности в день tt,
- $R_t^{(10)}$ — 10-дневная доходность на интервале от t до $t+9$.

Этот подход позволяет учесть перекрывающиеся периоды и более точно оценить доходность за 10 дней, что особенно важно для оценки редких событий, таких как экстремальные потери или прибыли, которые могут быть полезны для анализа рисков.

После получения 10-дневных доходностей для каждого ряда были рассчитаны их квантильные значения, в том числе 0.01-квантиль, что позволяет изучить экстремальные события в данных.

Симуляция

О методе Монте-Карло

Метод Монте-Карло представляет собой численный метод, использующий случайные выборки для приближённого решения задач, которые трудно решить аналитически. Этот метод применяется для оценки статистических характеристик случайных процессов, таких как квантиль (например, 0.01-квантиль), путём многократного моделирования случайных данных.

В данном случае мы используем метод Монте-Карло для оценки 0.01-квантили для различных типов временных рядов, включая случайный шум, процесс AR(1) и сложные временные ряды с трендом и сезонностью. Основной процесс заключается в генерации случайных выборок из устойчивого распределения Леви и вычислении квантиля для каждой выборки. С увеличением числа симуляций оценка квантиля становится более точной.

Формально, задача оценки квантиля может быть выражена как:

$$Q_\alpha = \mathrm{Quantile}(X_1, X_2, \dots, X_N, \alpha)$$

где $X_1, X_2, \dots, X_N$ — это N независимых случайных величин, генерируемых по заданному распределению, а α — интересующий нас квантиль (например, 0.01-квантили).

О проверке сходимости

Для того чтобы результаты симуляций были надёжными и точными, важно проверить сходимость оценок квантилей. Это означает, что с увеличением числа симуляций оценки квантилей должны стабилизироваться и сходиться к истинному значению.

Для проверки сходимости используются несколько методов:

- **Критерий остановки**: Процесс симуляции продолжается до тех пор, пока разница между средними значениями квантилей на текущем шаге и на предыдущем шаге симуляции не станет меньше заданного порогового значения ϵ\epsilon:

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

Этот критерий проверяет, стабилизировались ли значения квантилей, и, если да, процесс симуляции может быть завершён.

- **Статистические тесты**: Для дополнительной проверки стабильности результатов могут использоваться различные статистические тесты. Одним из таких тестов является **критерий Крускала-Уоллиса**, который проверяет, есть ли статистически значимые различия между несколькими группами данных. Это позволяет убедиться в том, что распределение квантилей не изменяется при увеличении числа симуляций.

- **Стандартная ошибка**: Для оценки точности результатов вычисляется стандартная ошибка, которая для симуляций определяется как:

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

где σ — стандартное отклонение квантилей, а N — количество симуляций. Чем меньше стандартная ошибка, тем точнее оценка квантиля.

О критерии Крускала-Уоллиса

Критерий Крускала-Уоллиса — это непараметрический статистический тест, который используется для проверки гипотезы о равенстве медиан нескольких независимых выборок. В контексте симуляции квантилей он применяется для оценки, стабилизировались ли результаты квантилей при увеличении числа симуляций и не изменяются ли распределения квантилей в различных симуляциях.

Тест Крускала-Уоллиса проверяет гипотезу о том, что несколько выборок имеют одинаковые медианы. Формально, нулевая гипотеза H_0 утверждает, что выборки имеют одинаковые медианы, а альтернативная гипотеза H_1 — что хотя бы одна из выборок имеет медиану, отличающуюся от других.

Тест основан на вычислении статистики H, которая определяется как:

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

где:

- N — общее количество наблюдений,
- k — количество групп (в нашем случае это количество симуляций),
- n_i — количество наблюдений в i-й группе,
- $\bar{R}_i$ — средний ранг в группе i,
- $\bar{R}$ — общий средний ранг по всем группам,
- s^2 — общий дисперсионный элемент.

Если вычисленная статистика H превышает критическое значение для заданного уровня значимости (например, p-уровень меньше 0.05), то нулевая гипотеза отвергается, что указывает на наличие статистически значимых различий между группами. Это может свидетельствовать о том, что результаты симуляции ещё не стабилизировались.

В контексте задачи, критерий Крускала-Уоллиса позволяет проверить, не изменяется ли распределение квантилей с увеличением числа симуляций. Если результат теста указывает на значительные различия, это может означать, что симуляция ещё не стабилизировалась и необходимо продолжить её выполнение.

О результатах симуляции

В результате выполнения симуляций с использованием метода Монте-Карло для различных типов процессов (шум, AR(1), сложные временные ряды) мы получаем следующие данные:

1. **Количество симуляций**: После того как критерий остановки выполнен, количество симуляций фиксируется. Это количество зависит от сходимости оценок квантилей — процесс продолжается до тех пор, пока значения квантилей не стабилизируются.
2. **Среднее значение квантилей**: Для каждой симуляции вычисляется значение 0.01-квантили, и затем рассчитывается среднее значение для всех симуляций. Это позволяет оценить центральную тенденцию квантилей.
3. **Стандартное отклонение**: Стандартное отклонение квантилей помогает понять вариативность оценок. Чем меньше это отклонение, тем точнее оценка.
4. **Гистограммы распределений квантилей**: Гистограммы позволяют визуально оценить, как распределяются значения квантилей по каждой итерации симуляции. Это помогает выявить стабильность или нестабильность результатов.
5. **Тест Крускала-Уоллиса**: Применяя этот тест, мы проверяем, есть ли статистически значимые различия между группами квантилей. Если различия значимы, это может означать, что сходимость ещё не достигнута.
6. **Оценка стандартной ошибки**: Стандартная ошибка помогает оценить точность полученных результатов, учитывая количество симуляций.
7. **Теоретическая ошибка**: Теоретическая ошибка оценивается как отношение стандартного отклонения к квадратному корню из количества симуляций, что даёт представление о том, насколько точно оценён квантиль.

В итоге, результаты симуляции и статистических тестов позволяют оценить, насколько точно и стабильно получены квантильные оценки для различных типов процессов, а также проверяются их статистические характеристики на сходимость и точность.

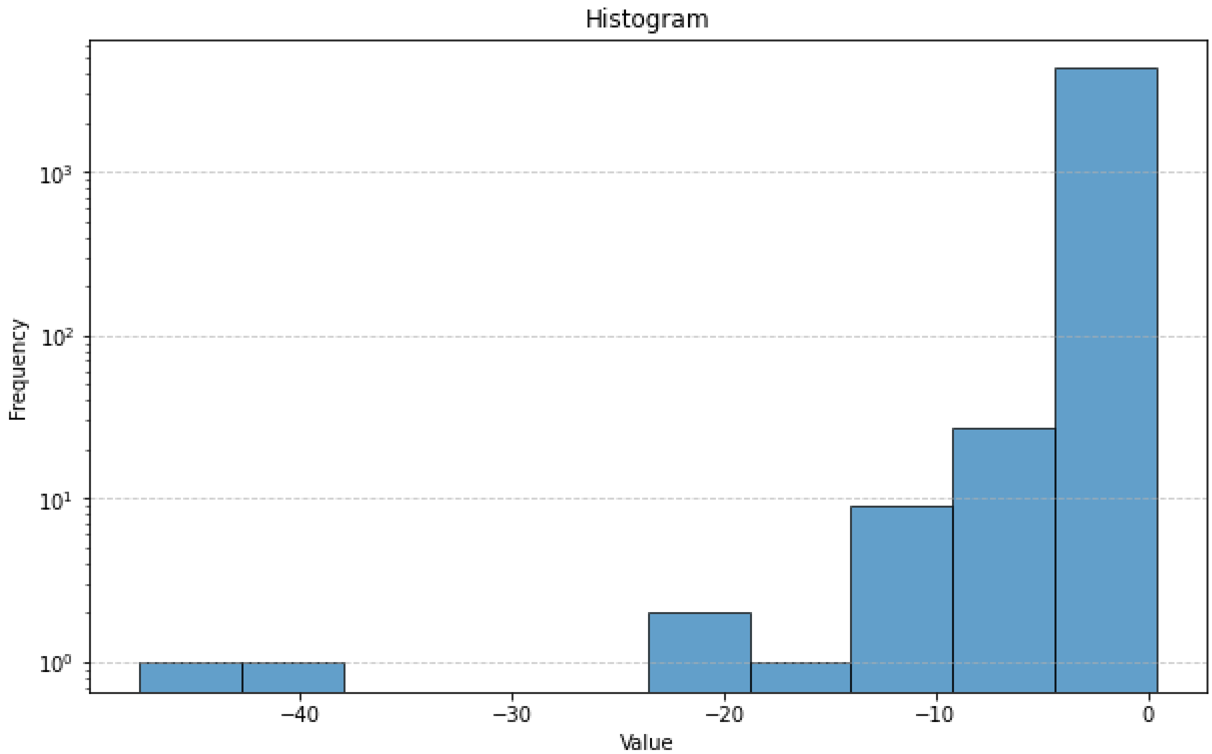

Рис. 10. Распределение квантилей

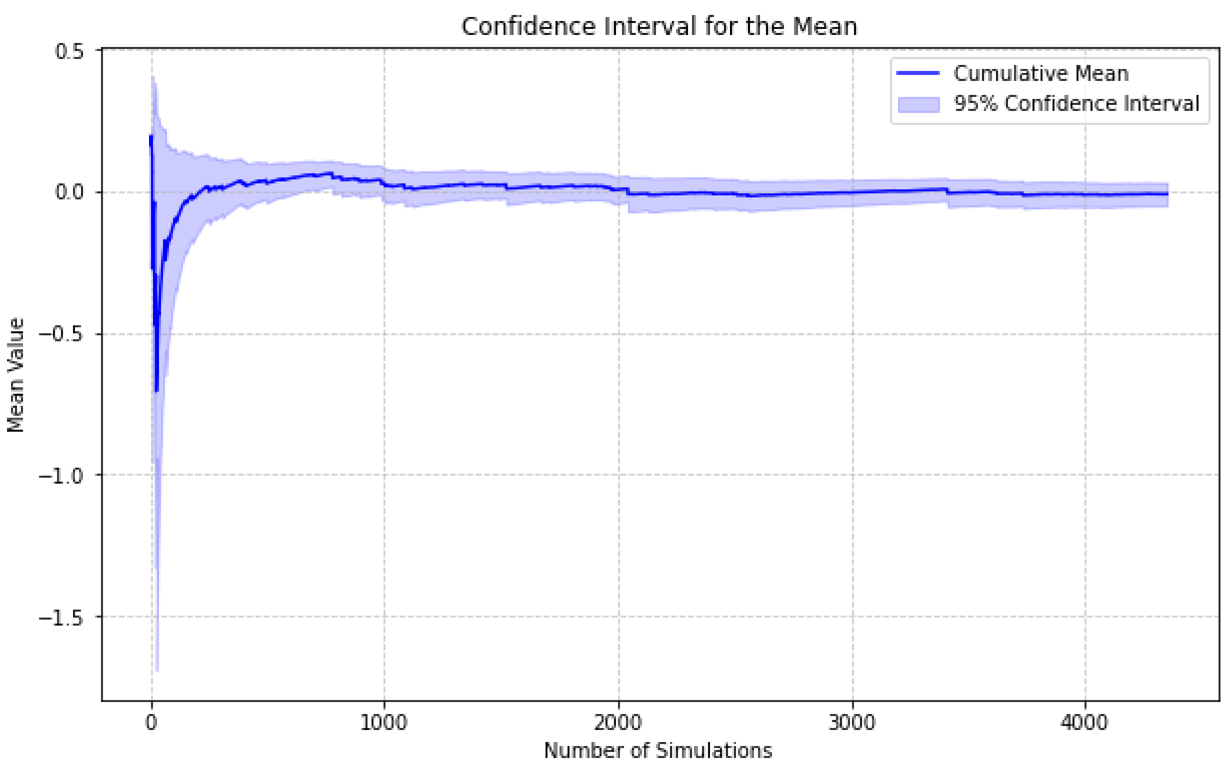

Рис. 11. График кумулятивного среднего и доверительных интервалов

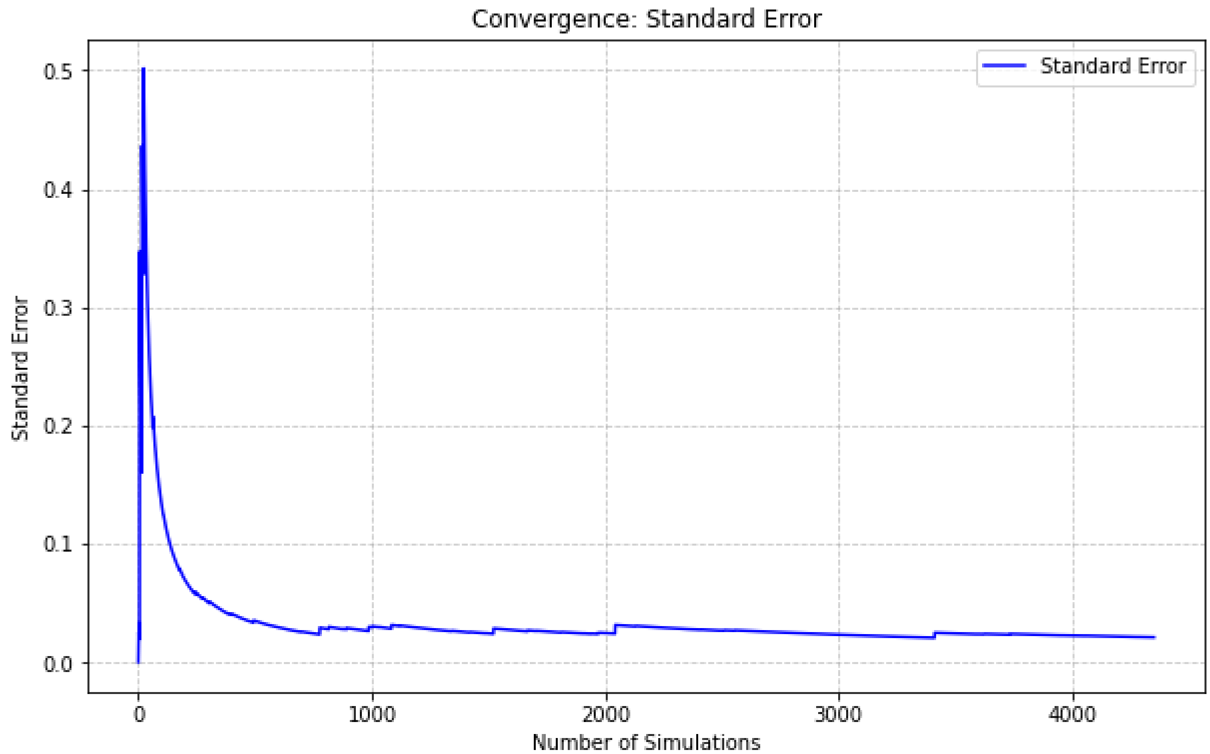

Рис. 12. График стандартных ошибок