

# Отчёт по проекту "Библиотека для проверки статистических гипотез"

Ляпшиева Анастасия, Балакаева Мария, Козлов Илья

Декабрь 2024

Данный документ представляет собой отчет по проекту в рамках курса "Программная инженерия и C++ для количественного анализа и алгоритмической торговли" и посвящен реализации методов по проверке гипотез о совпадении параметров двух коррелированных нормальных распределений, описанных в статье [2]. Результатом работы является библиотека *statisticslib*, написанная на языке Python.

## 1 Функциональные и системные требования

### Функциональные требования:

- Моделирование динамических процессов таких как арифметическое броуновское движение и процесс Кокса-Ингерсолла-Росса;
- Анализ временных рядов;
- Статистический анализ и реализация статистических методов, изложенных в статье;
- Легкая интеграция дополнительных статистических методов и моделей.

### Системные требования:

- Python 3.8+;
- Компилятор C++ (для сборки .cpp файлов);
- Необходимые библиотеки:
  - NumPy, SciPy (для статистических расчетов);
  - Matplotlib (для визуализации);
  - SWIG (для создания Python-обертки).

## 2 Программная архитектура

Для максимальной гибкости к добавлению новых моделей и статистических тестов было решено использовать **модульную архитектуру**. Каждый модуль выполняет определенную функцию и имеет свою логику и интерфейс для взаимодействия с другими модулями, более подробное описание есть в [4].

Наш проект разделен на **три основных модуля**:

- *models*: отвечает за математические и стохастические модели;
- *stats*: предоставляет инструменты для статистического анализа;
- *timeseries*: предназначен для работы с временными рядами.

Каждый модуль имеет четко определенную область ответственности и может использоваться независимо или совместно с другими модулями.

Для визуализации дизайна проекта была выбрана UML-диаграмма, потому что она позволяет четко показать классы, их связи, зависимости и иерархию. Она является более информативной наряду с другими методами визуализации, потому что имеет строгий шаблон, соответствующий профессиональным стандартам.

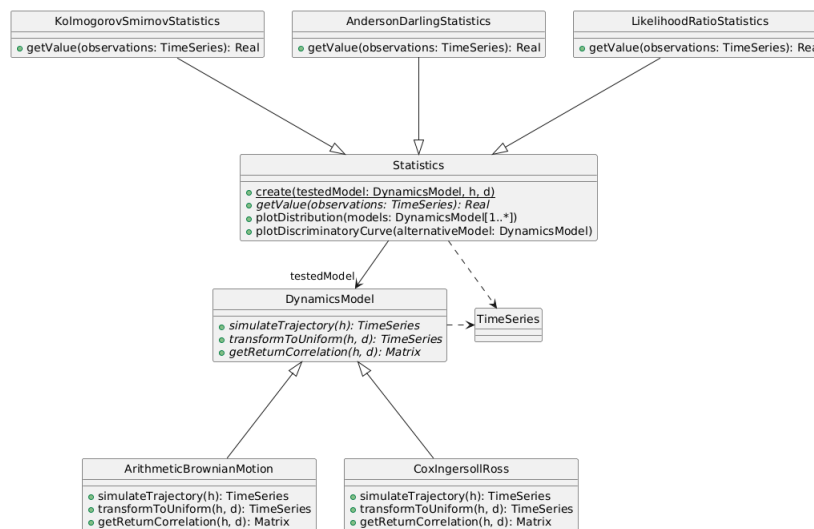


Рис. 1: Дизайн проекта.

## 3 Пайплайн проекта

В ходе работы над проектом были последовательно выполнены следующие задачи:

- Выбор темы проекта;
- Анализ и выбор архитектуры;
- Определение конечного результата;
- Работа над статьей и написание ноутбука *ClaytonImplementation.ipynb*;
- Расширение статьи и применение освоенного инструментария к модели Кокса-Ингерсолла-Росса;
- Написание библиотеки *statisticslib*;
- Проверена работоспособность библиотеки в файлах *launchBM.ipynb* и *launchCIR.ipynb* для броуновского движения и процесса CIR соответственно;
- Была попытка ускорить программу с помощью переноса кода симулятора траекторий броуновского движения. Был переписан код файла *ArithmeticBrownianMotion.py* на язык C++, но не удалось применить swig для портирования на python;
- Настройка GitHub Actions для использования Doxygen;
- Оформление документации проекта.

В ноутбуке *ClaytonImplementation.ipynb* были сохранены все оригинальные названия и обозначения из статьи, для удобства понимания. Однако при написании библиотеки *statisticslib* мы пользовались принципами, изложенными в книге [3], и обдумывали название каждой функции и переменной для удобного использования этой библиотеки.

## 4 Реализация

Имплементация статьи сделана в ноутбуке *ClaytonImplementation.ipynb*, в котором шаг за шагом с **подробными комментариями** проделано то, что описано в [2].

Основные шаги имплементации:

- Анализ содержания статьи: изучение подходов к бэкстестингу моделей прогнозирования волатильности;
- Реализация статистических гипотез: проверка гипотезы на принятие или отвержение истинности гипотезы. Подробнее об этом можно найти в [6].
- Использование корреляционной структуры: модификация статистических тестов для учета корреляций в пересекающихся наблюдениях;
- Генерация графиков: построение визуализаций для демонстрации мощности статистических тестов;

- Расширение моделей: применение методов статьи к более сложным стохастическим моделям, выбранным и изученным по [5].

## 5 Заключение

В рамках проекта был реализован подход, описанный в статье [2], посвященный реализации методов по проверке гипотез о совпадении параметров двух коррелированных нормальных распределений. Итоговым результатом стала гибкая и готовая к дальнейшему использованию библиотека, разработанная на основе тщательно спроектированного дизайна.

Проект проиллюстрировал актуальность темы: использование скорректированных статистик для учета корреляционной структуры перекрывающихся наблюдений улучшает дискриминационную мощность тестов. Анализ открытых источников показал, что доступных реализаций этой статьи ранее не существовало.

Актуальность выбранной темы подчеркивается так же в недавней работе [7], в которой были предложены новые методы исследований. Таким образом наш проект может быть расширен путем реализации описанных подходов, что повысит точность результатов.

## Список литературы

- [1] Лекции и семинары по курсу "*Программирование на C++ для алгоритмической торговли и количественного анализа*", 2024.
- [2] Michael A. Clayton, *Backtesting Volatility Assumptions using Overlapping Observations*, September 25, 2019.
- [3] Роберт Мартин, *Чистый код*, 2019.
- [4] Роберт Мартин, *Чистая архитектура*, 2021.
- [5] А.В. Булинский и А.Н.Ширяев, *Теория случайных процессов*, М.:ФИЗМАТЛИТ, 2005.
- [6] Я.Р. Магнус, П.К. Катышев и А.А. Пересецкий, *Эконометрика*, 2004.
- [7] Nikolai Nowaczyk, Vladimir Piterbarg *Backtesting Correlated Quantities*, September 13, 2023