**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ**

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ**

**Кафедра теории вероятностей и математической статистики**

**АППРОКСИМАЦИЯ ДОХОДНОСТЕЙ АКЦИЙ С ПОМОЩЬЮ МАРКОВСКИХ ПРОЦЕССОВ ВЫСШИХ ПОРЯДКОВ**

Курсовой проект

Олихвера Дмитрия Юрьевича

студента 3 курса,  
специальность  
«актуарная математика»

Научный руководитель:

кандидат физико-математических наук,

доцент П.М. Лаппо

Минск, 2022

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра теории вероятностей и математической статистики

**ЗАДАНИЕ НА КУРСОВОЙ ПРОЕКТ**

Студент 3 курса **Олихвер Д.Ю**

1. Тема **Аппоксимация доходностей акций с помощью процессов Маркова высших порядков**

2. Срок представления курсового проекта к защите 11 декабря 2022 года

3. Исходные данные для научного исследования

1. [Вероятностно-статистический анализ цепей Маркова высокого порядка](https://elib.bsu.by/handle/123456789/17034) [Харин, Ю. С.](https://elib.bsu.by/browse?type=author&value=%D0%A5%D0%B0%D1%80%D0%B8%D0%BD%2C+%D0%AE.+%D0%A1.).
2. Харин, Ю.С. Практикум на ЭВМ по математической статистике. / М.Д. Степанова. – Мн.: изд-во «Университетское», 1987.
3. Цепи Маркова: https://habr.com/ru/post/455762/
4. [Теория вероятностей, математическая и прикладная статистика](https://elib.bsu.by/handle/123456789/110455) [Харин, Ю. С.](https://elib.bsu.by/browse?type=author&value=%D0%A5%D0%B0%D1%80%D0%B8%D0%BD%2C+%D0%AE.+%D0%A1.)*;*[Зуев, Н. М.](https://elib.bsu.by/browse?type=author&value=%D0%97%D1%83%D0%B5%D0%B2%2C+%D0%9D.+%D0%9C.)*;*[Жук, Е. Е.](https://elib.bsu.by/browse?type=author&value=%D0%96%D1%83%D0%BA%2C+%D0%95.+%D0%95.)

4. Содержание курсового проекта

1. Изучить процессы Маркова.

4.2 Выяснить, можно ли прогнозировать изменение доходностей акций, используя процессы Маркова высших порядков.

4.3 Разработать соответствующее программное обеспечение.

Руководитель курсового проекта \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ П. М. Лаппо

Задание принял к исполнению \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д. Ю. Олихвер

**Оглавление**

[Введение 4](#_Toc1643657541)

[Глава 1. Изучение процессов Маркова 5](#_Toc998250178)

[Глава 2. Использование Марковских процесов высших порядков для аппроксимации доходностей акций 8](#_Toc478788239)

[Глава 3. Разработка программного обеспечения 10](#_Toc1848782814)

[Заключение 13](#_Toc1187113105)

[Список использованных источников 14](#_Toc1250276365)

[ПРИЛОЖЕНИЕ А 15](#_Toc1951499379)

# **Введение**

Роль рынка цепных бумаг проявляется в результатах его функционирования для экономики, населения, государства. Основная роль рынка ценных бумаг проявляется через воздействие на социально-экономическое развитие страны путем: обеспечения непрерывности процесса кругооборота капитала за счет привлечения денежных средств, которые находятся в свободном обращении в форме инвестиций, с целью развития компаний; свободного перелива капитала между секторами рыночной экономики (из убыточных отраслей в развивающиеся), что создает наиболее благоприятные условия для расширения производства (увеличения запасов, обновления и создания дополнительных основных фондов); предоставления множества инструментов инвестирования с разными условиями и соглашениями (включая продажу акций, облигаций, евронот, еврооблигаций, варрантов и т.д.); содействия платежному и денежному обороту (вексель, чек), перераспределению и экономному использованию материальных и денежных ресурсов, повышению платежеспособного спроса населения, покрытию дефицита федерального бюджета, развитию внешнеэкономических связей.

Давая общую оценку значения ценных бумаг в экономике можно выделить следующие важнейшие моменты. Во-первых, ценные бумаги выступают гибким инструментом инвестирования свободны денежных средств юридических и физических лиц. Во-вторых, размещение ценных бумаг - эффективный способ мобилизации ресурсов для развития производства и удовлетворения других общественных потребностей. В-третьих, ценные бумага активно участвуют в обслуживании товарного и денежного обращения.

Целью работы является аппроксимация доходностей акций рассматривая динамику изменения цен на акции при помощи процессов Маркова высших порядков. Для достижения цели мною сформулированы задачи работы: рассмотреть понятие и сущность процессов Маркова и процессов Маркова высших порядков; охарактеризовать возможность использовать цепи Маркова для прогноза доходностей акций; разработать программное обеспечение для аппроксимации доходностей акций.

# **Глава 1. Изучение процессов Маркова**

Существуют хорошо известные семейства случайных процессов: гауссовы процессы, пуассоновские процессы, авторегрессивные модели, модели скользящего среднего, цепи Маркова и другие. Каждое из этих отдельных случаев имеет определённые свойства, позволяющие нам лучше исследовать и понимать их.  
  
Одно из свойств, сильно упрощающее исследование случайного процесса — это «марковское свойство». Если объяснять очень неформальным языком, то марковское свойство сообщает нам, что если мы знаем значение, полученное каким-то случайным процессом в заданный момент времени, то не получим никакой дополнительной информации о будущем поведении процесса, собирая другие сведения о его прошлом. Более математическим языком: в любой момент времени условное распределение будущих состояний процесса с заданными текущим и прошлыми состояниями зависит только от текущего состояния, но не от прошлых состояний (свойство отсутствия памяти). Случайный процесс с марковским свойством называется марковским процессом.

Марковское свойство обозначает, что если мы знаем текущее состояние в заданный момент времени, то нам не нужна никакая дополнительная информация о будущем, собираемая из прошлого.  
  
На основании этого определения мы можем сформулировать определение «однородных цепей Маркова с дискретным временем» (в дальнейшем для простоты мы их будем называть «цепями Маркова»). **Цепь Маркова** — это марковский процесс с дискретным временем и дискретным пространством состояний. Итак, цепь Маркова — это дискретная последовательность состояний, каждое из которых берётся из дискретного пространства состояний (конечного или бесконечного), удовлетворяющее марковскому свойству.

Математически мы можем обозначить цепь Маркова так:

https://habrastorage.org/r/w1560/getpro/habr/post_images/114/3b7/fc3/1143b7fc371f3142534c2b886bf3e69c.png

где в каждый момент времени процесс берёт свои значения из дискретного множества E, такого, что

https://habrastorage.org/r/w1560/getpro/habr/post_images/886/a22/d76/886a22d7671798102ee3d94fe9868b81.png

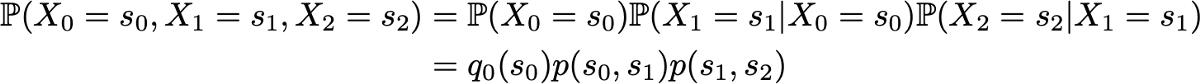
Тогда марковское свойство подразумевает, что у нас есть

https://habrastorage.org/r/w1560/getpro/habr/post_images/edc/8bf/384/edc8bf38422705e72c9dd7d094b249db.png

И затем проявляется упрощение, определяемое марковским допущением. И в самом деле, в случае длинных цепей мы получим для последних состояний сильно условные вероятности. Однако в случае цепей Маркова мы можем упростить это выражение, воспользовавшись тем, что

https://habrastorage.org/r/w1560/getpro/habr/post_images/0ab/7f6/568/0ab7f6568e28bb562ebb287252422d51.png

получив таким образом



Так как они полностью характеризуют вероятностную динамику процесса, многие сложные события можно вычислить только на основании исходного распределения вероятностей q0 и матрицы переходной вероятности p. Стоит также привести ещё одну базовую связь: выражение распределения вероятностей во время n+1, выраженное относительно распределения

вероятностей во время n  






**Марковские процессы высших порядков** являются логическим продолжением Марковских процессов первого порядка. В отличие от марковских процессов первого порядка, здесь рассматривают не только состояние исследуемого объекта на данный момент, но и N – 1 предыдущих состояний, где N – порядкок процесса. Так, если рассматривать степень 5, то помимо текущего состояния нужно обратиться к 4 предыдущим состояниям.

# **Глава 2. Использование Марковских процесов высших порядков для аппроксимации доходностей акций**

Для анализа был выбран медийный конгломерат Kering SA. Данные для анализа были взяты за период 2005-12-12 ­– 2022-12-9 с периодичность в одну неделю и представлены в следующем виде:

Date,Open,High,Low,Close,Adj Close,Volume

Пример: 2005-12-12,86.577,86.949493,83.928795,84.718826,49.856682,2440378

Использование Марковских процессов высшего порядка для аппроксимации доходностей используется следующим образом:

1. На основе данных динамики ценн на акции считается доходность для времени t
2. Доходности разбиваются на зоны
3. Для каждой посчитанной доходности, а конкретно ее зоны, считается переход из зоны доходности в момент времени t – 1 в зону доходности в момент t
4. На основе полученных переходов формируется вероятные последовательности

Конечная аппроксимация алгоритма выглядит следующим образом:

Probability of next sequence to be (3, 4, 4, 1) is 0.027777777777777776. Profitability for next week: [0.06760930595444994, 0.10595405724911736]

Probability of next sequence to be (3, 4, 4, 2) is 0.1111111111111111. Profitability for next week: [0.029264554659782532, 0.06760930595444994]

Probability of next sequence to be (3, 4, 4, 3) is 0.3333333333333333. Profitability for next week: [-0.00908019663488488, 0.029264554659782532]

Probability of next sequence to be (3, 4, 4, 4) is 0.5. Profitability for next week: [-0.04742494792955229, -0.00908019663488488]

Probability of next sequence to be (3, 4, 4, 5) is 0.027777777777777776. Profitability for next week: [-0.0857696992242197, -0.04742494792955229],

где (3, 4, 4, 1), (3, 4, 4, 2), (3, 4, 4, 3), (3, 4, 4, 4) – вероятная последовательность, где первые три цифры – текущее и два предыдущих состояния, а последняя цифра – зона доходности, в которую с определенной вероятностью доходность попадет при следующем моменте времени t.

Таким образом можно трактовать ответ, где – зона доходности, , t – текущий момент времени,  *–* минимальная возможная доходность*, –* максимальная возможная доходность:

*Вероятность того, что после последовательности (, , …, ,) доходность перейдет в определенное состояние является . Доходность будет находится в отрезке [,].*

Расчет доходностей происходит по формуле: , где – доходность в момент *i*, цена в момент времени *i, .*

Вероятность следующей последовательности высчитывается по формуле , где – количество последовательностей , –количество последовательностей , где соответственно.

Сумма вероятностей всех исходов должна равняться единице.

# **Глава 3. Разработка программного обеспечения**

Результаты аппроксимации очень сильно зависит от количества выбранного количества зон для разбиения доходностей (далее Z) и порядка Марковского процесса (далее D).

В примере выше Z = 7 и D = 3. При аппроксимации с параметрами, больше указанных в примере, несмотря на более узкий диапазон возможной доходности, возникает вероятность отсутствия прогноза вовсе. Так при той же выборке с параметрами Z = 9 и D = 6 вовсе нет аппроксимации, из за отсутствия последней последовательности в предыдущих данных. Также при высоких значениях Z и D возникают вероятные последовательности со 100% исходом, что будет являться неверным:

(3, 4, 2, 3, 4, 4, 2, 2) - 1.0

(4, 2, 3, 4, 4, 2, 2, 0) - 1.0

(2, 3, 4, 4, 2, 2, 0, 2) - 1.0

(3, 4, 4, 2, 2, 0, 2, 3) - 1.0

(4, 4, 2, 2, 0, 2, 3, 3) - 1.0

(4, 2, 2, 0, 2, 3, 3, 2) - 1.0

(2, 2, 0, 2, 3, 3, 2, 3) - 1.0

Выбор Z и D для алгоритма аппроксимации являются очень важными для точности алгоритма. При использовании алгоритма с выборкой доходностей по неделям за 15 лет возникают грубые неточности.

Для проверки работы алгоритма были удалены последние строки из данных цен на акции, таким образом получив возможность проверить точность аппроксимации алгоритма, имея реальную доходность акций за следующий период.

1. Данные 2005-12-12 – 2022-11-28, Z = 7, D = 3

Probability of next sequence to be (2, 2, 2, 1) is 0.05357142857142857. Profitability for next week: [0.04569801950035428, 0.09499841402206952]

Probability of next sequence to be (2, 2, 2, 2) is 0.6488095238095238. Profitability for next week: [-0.0036023750213609596, 0.04569801950035428]

Probability of next sequence to be (2, 2, 2, 3) is 0.2976190476190476. Profitability for next week: [-0.05290276954307621, -0.0036023750213609596]. **Реальная доходность составила**: ~0.023. **Аппроксимация верна**

2. Данные 2005-12-12 – 2022-12-05, Z = 7, D = 3 Probability of next sequence to be (2, 2, 3, 1) is 0.046511627906976744. Profitability for next week: [0.04569801950035428, 0.09499841402206952]

Probability of next sequence to be (2, 2, 3, 2) is 0.37209302325581395. Profitability for next week: [-0.0036023750213609596, 0.04569801950035428]

Probability of next sequence to be (2, 2, 3, 3) is 0.4883720930232558. Profitability for next week: [-0.05290276954307621, -0.0036023750213609596]

Probability of next sequence to be (2, 2, 3, 4) is 0.08139534883720931. Profitability for next week: [-0.10220316406479146, -0.05290276954307621]

Probability of next sequence to be (2, 2, 3, 5) is 0.011627906976744186. Profitability for next week: [-0.1515035585865067, -0.10220316406479146] **Реальная доходность составила**: ~0.00427. **Алгоритм попал во вторую по вероятности область**

3. Данные 2005-12-12 – 2022-11-21, Z = 7, D = 3

Probability of next sequence to be (1, 2, 2, 1) is 0.125. Profitability for next week: [0.04569801950035428, 0.09499841402206952]

Probability of next sequence to be (1, 2, 2, 2) is 0.625. Profitability for next week: [-0.0036023750213609596, 0.04569801950035428]

Probability of next sequence to be (1, 2, 2, 3) is 0.1875. Profitability for next week: [-0.05290276954307621, -0.0036023750213609596]

Probability of next sequence to be (1, 2, 2, 5) is 0.0625. Profitability for next week: [-0.1515035585865067, -0.10220316406479146] **Реальная доходность составила**: ~0.03481. **Аппроксимация верна**

4. Данные 2005-12-12 – 2022-11-14, Z = 7, D = 3

Probability of next sequence to be (1, 2, 2, 1) is 0.125. Profitability for next week: [0.04569801950035428, 0.09499841402206952]

Probability of next sequence to be (1, 2, 2, 2) is 0.625. Profitability for next week: [-0.0036023750213609596, 0.04569801950035428]

Probability of next sequence to be (1, 2, 2, 3) is 0.1875. Profitability for next week: [-0.05290276954307621, -0.0036023750213609596]

Probability of next sequence to be (1, 2, 2, 5) is 0.0625. Profitability for next week: [-0.1515035585865067, -0.10220316406479146] **Реальная доходность составила**: ~0.04222. **Аппроксимация верна**

# **Заключение**

В данной работе выявлено можно ли использовать подход с процессами Маркова высших порядков для аппроксимации доходностей акций.

Во время ознакомления с материалами были изучены другие примениения процессов Маркова высших порядков в других сферах, например прогнозирование погоды.

В результате работы был спроектирован алгоритм, который позволяет с определенной погрешностью спрогнозировать доходность акций на следующую неделю. Тем не менее нельзя постоянно основываться на прогнозы алгоритма, потому, что цены зависят от множества факторов. В современном мире это стоит особенно остро, ведь даже одно интервью, одна запись или война могут как значительно повысить, так понизить цену акций.

**Список использованных источников**

1. [Вероятностно-статистический анализ цепей Маркова высокого порядка](https://elib.bsu.by/handle/123456789/17034) [Харин, Ю. С.](https://elib.bsu.by/browse?type=author&value=%D0%A5%D0%B0%D1%80%D0%B8%D0%BD%2C+%D0%AE.+%D0%A1.).
2. Харин, Ю.С. Практикум на ЭВМ по математической статистике. / М.Д. Степанова. – Мн.: изд-во «Университетское», 1987.
3. Цепи Маркова: https://habr.com/ru/post/455762/
4. [Теория вероятностей, математическая и прикладная статистика](https://elib.bsu.by/handle/123456789/110455) [Харин, Ю. С.](https://elib.bsu.by/browse?type=author&value=%D0%A5%D0%B0%D1%80%D0%B8%D0%BD%2C+%D0%AE.+%D0%A1.)*;*[Зуев, Н. М.](https://elib.bsu.by/browse?type=author&value=%D0%97%D1%83%D0%B5%D0%B2%2C+%D0%9D.+%D0%9C.)*;*[Жук, Е. Е.](https://elib.bsu.by/browse?type=author&value=%D0%96%D1%83%D0%BA%2C+%D0%95.+%D0%95.)
5. Цепи Маркова: <https://thecode.media/markov-chain/>
6. Higher Order Markov Chains: <https://link.springer.com/chapter/10.1007/0-387-29337-X_6>

# ПРИЛОЖЕНИЕ А

Листинг программы на языке Python

f = open("KER.PA.csv", 'r')  
  
f.readline()  
  
max\_ = 0  
min\_ = 2147483647  
Zones\_amount = 9  
Degree = 3  
duration = 0  
zones = []  
profitabilities = []  
avgs = []  
  
s = f.readline().split(',')  
prev\_price = (float(s[3]) + float(s[2])) / 2  
  
  
def find\_profitability\_zone(p):  
 counter = 0  
  
 m = max\_ - duration  
 while m >= min\_:  
 if p <= m:  
 counter += 1  
 m -= duration  
 else:  
 return counter  
 return counter - 1  
  
  
def get\_zone\_limits(z):  
 return [max\_ - ((z + 1) \* duration), max\_ - (z \* duration)]  
  
  
def get\_average\_zone\_profitability(zone):  
 return (max\_ - (zone \* duration) + max\_ - ((zone + 1) \* duration)) / 2  
  
  
for line in f:  
 high = float(line.split(',')[2])  
 low = float(line.split(',')[3])  
  
 avg = (high + low) / 2  
 avgs.append(avg)  
 profitability = (avg - prev\_price) / prev\_price  
 profitabilities.append(profitability)  
 prev\_price = avg  
  
 if profitability > max\_:  
 max\_ = profitability  
  
 if profitability < min\_:  
 min\_ = profitability  
  
duration = (max\_ - min\_) / Zones\_amount  
  
print("Profitabilities: {}".format(profitabilities))  
  
zones = [find\_profitability\_zone(p) for p in profitabilities]  
print("Zones: {}".format(zones))  
  
occurences = {}  
sequences = {}  
probabilities = {}  
  
for i in range(len(zones) - Degree):  
 sequence = tuple(zones[j] for j in range(i, i + Degree + 1))  
 z = tuple(sequence[:-1])  
 if sequence in occurences.keys():  
 occurences[sequence] = occurences[sequence] + 1  
 else:  
 occurences[sequence] = 1  
  
 if z in sequences.keys():  
 sequences[z] += 1  
 else:  
 sequences[z] = 1  
  
print(occurences)  
print(sequences)  
  
for o in occurences:  
 sequence = tuple(o[:-1])  
 probabilities[o] = occurences[o] / sequences[sequence]  
 print("{} - {}".format(o, probabilities[o]))  
  
last\_sequence = [i for i in zones[-Degree:]]  
print(last\_sequence)  
  
printed = 0  
for i in range(Zones\_amount + 1):  
 last\_sequence.append(i)  
 prob\_sequence = tuple(j for j in last\_sequence)  
 if prob\_sequence in probabilities.keys():  
 print("Probability of next sequence to be {} is {}. Profitability for next week: {}".format(prob\_sequence,  
 probabilities[  
 prob\_sequence],  
 get\_zone\_limits(i)))  
 printed += 1  
 last\_sequence.pop()  
if printed == 0:  
 print("No probability for future week due to absence of previous records for {} sequence".format(last\_sequence))