**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ**
**РОССИЙСКОЙ ФЕДЕРАЦИИ**
**ФГБОУ ВО «Тувинский Государственный Университет»**
**ФИЗИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ**
**КАФЕДРА МАТЕМАТИКИ И МЕТОДИКИ ПРЕПОДАВАНИЯ МАТЕМАТИКИ**

Выпускная квалификационная работа
(бакалаврская работа)

# Применение цепей Маркова для построения прогнозов

Работа допущена к защите
Зав. кафедрой ____________
**Танзы М.В., к.п.н., доцент**____
(фамилия, и.о., уч.степень, звание)

Студентки 4_курса_5_группы
направления подготовки 01.03.01
профиль «Математика»
очной формы обучения
Самбала Айраны Витальевны
(Ф.И.О.)
______________________________
(подпись)
«__»_____________________20___г.

Работа защищена «__»________20__г.
С оценкой ______________________
Председатель ГЭК ______________
(подпись)
**Сенашов В. И. д.ф-м.н., профессор**
**ведущий научный сотрудник Института**
**вычислительного моделирования**
**СО РАН, Красноярск**

Члены комиссии _________________
(подпись)
________________________________
(фамилия, и. о., должность, уч.степень, звание)
________________________________
(подписи)

Научный руководитель: _____________
__Хурума Анна Кыс-ооловна, ст. преп.__

**Кызыл – 2020г.**

# СОДЕРЖАНИЕ

# ВВЕДЕНИЕ.

Теория случайных процессов - это наука, изучающая закономерности случайных явлений в динамике их развития. Понятие случайного процесса появилось в начале прошлого века и связано с именами А. Н. Колмогорова, А.Я. Хинчина, Е.Е. Слуцкого, Н. Винера и других.

Это понятие в наши дни является одним из центральных не только в теории вероятностей, но также в естествознании, инженерном деле, экономике, организации производства, теории связи. Теория случайных процессов принадлежит к категории наиболее быстро развивающихся математических дисциплин. Несомненно, это обстоятельство в значительной мере определяется ее глубокими связями с практикой.

Процессы, протекающие в различных областях практической деятельности, в том числе в системе образования, требуют постоянной объективной оценки, корректировки и управления. Однако без прогнозирования управление невозможно. Поэтому возникает необходимость прогнозирования показателей качества как на завершающих этапах обучения, так и в ходе учебного процесса.

Прогнозирование успеваемости студентов по той или иной дисциплине позволяет сформировать индивидуальную траекторию работы обучаемого в семестре и тем самым повысить уровень их профессиональной подготовки.

**Актуальность исследования** – построение прогнозов, базирующихся на современных математических подходах с применение цепей Маркова, позволяет прогнозировать динамику развития основных показателей исследуемого процесса.

*В работе показана возможность применения аппарата марковских цепей для прогнозирования успеваемости студентов. На основании данных об успеваемости на начальных курсах строится прогноз об успеваемости на старших курсах. Расчетные данные сравниваются с экспериментальными. В работе представлена методология оценки и прогнозирования качества образования.*

Марковские случайные процессы названы по имени выдающегося русского математика А.А.Маркова (1856-1922), впервые начавшего изучение вероятностной связи случайных величин и создавшего теорию, которую можно назвать "динамикой вероятностей". В дальнейшем основы этой теории явились исходной базой общей теории случайных процессов, а также таких важных прикладных наук, как теория диффузионных процессов, теория надежности, теория массового обслуживания и т.д. В настоящее время теория марковских процессов и ее приложения широко применяются в самых различных областях.

В последние десятилетия стала очевидна необходимость использования широкого класса вероятностных математических моделей для практического анализа и изучения вопросов, связанных с социально-экономической жизнью современного общества. Изучение массовых и социальных процессов, вопросов воспроизводства населения в целом, демографических прогнозов и политики сегодня остается первостепенной и актуальной задачей. Только современный аппарат математической статистики и теории случайных процессов дает общие приемы, разработанные частные схемы для решения огромного спектра практических задач, возникающих при изучении различных явлений. Как и много лет назад эти реальные задачи иллюстрируют дальнейшее совершенствование и развитие современной науки.

Цепи Маркова – одна из основных и актуальных тем в нынешнее время в современной математике. Цепи Маркова являются обобщением схемы Бернулли, которая была написана в XVII, а марковские цепи получили сравнительно недавно свое признание. Очень много процессов в нынешнее время решаются с помощью схем Бернулли или цепей Маркова. Вся поисковая система Интернета основана на этих процессах.

**Объект исследования:** процесс обучения студентов ФМФ.

**Предмет исследования:** прогнозирование учебной деятельности студентов за весь период учебы с первого по девятый семестр ФМФ.

**Цель исследования** заключается в том, чтобы изучить практическое применение цепей Маркова для прогнозирования в разных областях и построить пример применения цепи Маркова для прогноза результатов обучения студентов.

На основании объекта и предмета исследования для достижения поставленных целей нами были сформулированы следующие **задачи:**

1. Определить подходы к прогнозированию успеваемости студентов

2. Выявить особенности использования цепей Маркова как метода прогнозирования.

3. Показать практическое применение цепей Маркова к прогнозированию успеваемости студентов.

Практическая значимость исследования. Работа представляет методологию оценки и прогнозирования успеваемости студентов. Эта методология основана на аппарате марковских цепей. Результаты исследования могут быть использованы преподавателями высших учебных заведений при прогнозировании успеваемости студентов.

**Структура работы.** Работа состоит из введения, двух глав, заключения и списка литературы.

# ГЛАВА 1. ТЕОРЕТИЧЕСКИЙ АНАЛИЗ ИСПОЛЬЗОВАНИЯ ЦЕПЕЙ МАРКОВА ПРИ ПРОГНОЗИРОВАНИИ УСПЕВАЕМОСТИ СТУДЕНТОВ.

## 1.1. Подходы к прогнозированию успеваемости студентов.

В последние годы смешанное обучение (комбинация традиционного и online обучения) стало одной из самых востребованных образовательных технологией в связи с растущим внедрением систем управления обучением (LMS) в образовательный процесс [5]. В России самой популярной системой с открытым исходным кодом является Moodle, ее используют более 65 % вузов. В ряде университетов внедрена Blackboad, самая распространенная в мире система с закрытым исходным кодом. LMS поддерживает вовлечение студентов в учебный процесс, создает условия для активного взаимодействия студентов и преподавателей, осуществляет сбор данных о поведении обучающихся на online платформе [5]. Наличие системы организации обучения, достижения в области статистики, быстрое развитие программных и аналитических методов, а также возможность применения принципов бизнес-аналитики к процессу обучения, привели к появлению учебной аналитики как одного из наиболее перспективных направлений исследований в области компьютерной поддержки образования [6].

Учебная аналитика - это измерение, сбор, анализ и представление данных об обучающихся и их действиях с целью понимания и оптимизации учебного процесса и той среды, где это этот процесс происходит [4]. Анализ таких данных позволяет контролировать регулярность занятий слушателя, осуществлять мониторинг его успеваемости, следить за ходом выполнения контрольных заданий [4]. В последние годы учебная аналитика послужила концептуальной основой для анализа образовательных данных online курсов, прогнозирования успешности обучающихся [4], сбора данных об их поведении и своевременного вмешательства преподавателя в процесс обучения.

Приведем краткий обзор существующих подходов к прогнозированию успеваемости.

Моисеев В.Б., Зубков А.Ф., Деркаченко В.Н. в своей работе [17] описывают подход на базе регрессионных моделей, позволяющий выявить связь между уровнем знаний и умений студентов по общепрофессиональным и специальным дисциплинам в зависимости от обеспечивающих курсов с помощью многофакторной линейной регрессионной модели.

В работе [10] Гоголева Н.Г., Тарасова О.Ю. показывают возможность применения аппарата Марковских цепей для прогнозирования успеваемости студентов. На основании данных об успеваемости на начальных курсах строится прогноз об успеваемости на старших курсах. Расчетные данные сравниваются с экспериментальными посредством составления уравнения парной линейной регрессии y=a+b*x.

Статья «Классификация студентов по уровню успеваемости с помощью аппарата дискриминантного анализа» [11] посвящена разработке алгоритма выявления факторов, влияющих на успеваемость студента. В основе этого подхода лежит математический инструментарий дискриминантного анализа, который реализует классификацию с «обучением». Для проведения классификации студентов по уровню успеваемости используются следующие дискриминантные переменные: средний балл аттестата, пол, тип населенного пункта, количество детей в семье, возраст отца, возраст матери, количество больничных листов в течение года, проводимое в интернете время; наличие родителей, количество друзей, количество прочитанных в месяц книг, средний балл в колледже. Корректировка некоторых из перечисленных факторов (напр., количество прочитанных в месяц книг и др.) позволит повысить качество подготовки учащихся. Дискриминантная модель оптимально разделяет множество объектов на подмножества и проводит классификацию новых объектов в тех случаях, когда неизвестно заранее, к какому из существующих классов они принадлежат.

В статье «Анализ и прогнозирование успеваемости студентов на основе радиальной базисной нейронной сети» [20] рассматриваются аспекты применения радиальных базисных нейронных сетей для решения задачи прогнозирования успеваемости студентов. Прогнозируется успеваемость студентов по отдельным дисциплинам текущего семестра. На вход сети поступает следующая информация: средняя оценка студента по предшествующим смежным дисциплинам, освоение которых в полном объеме является необходимым условием; наличие свободного посещения; оценка за тестирование остаточных знаний, проводимое перед началом изучения данной дисциплины; наличие задолженностей по другим дисциплинам. На выходе сети получена прогнозируемая по дисциплине оценка.

Подход, описанный Ясинским И.Ф., Семеновым М.Б., позволяет получить оценку перспективности абитуриентов на основе использования трехслойной нейронной сети. Цель подхода в выявлении как наиболее одаренных студентов, так и тех, кому могут потребоваться дополнительные занятия. На входе сети используются следующие характеристики: место жительства; номер школы; выпускные оценки по физике и математике при окончании школы; профессия отца и матери; сведения о семейном доходе. На выходе сети получают оценку перспективности абитуриента[25].

Шевченко В.А. предложена методика прогнозирования успеваемости студентов на основе методов кластерного анализа. В основе подхода лежит модификация метода к-средних Мак-Кина. Количество кластеров задается явно и равно 4, каждому кластеру ставится в соответствие типологическая группа, характеризующая успеваемость студентов: «отлично», «хорошо», «удовлетворительно» и «плохо». Прогнозирование выполняется по каждой отдельной дисциплине, в качестве оценочных параметров выступают факторы, значения которых можно оценить в начальный момент изучения дисциплины: уровень начальных знаний студентов (определяется в начале первого занятия по дисциплине); уровень компетенций, сформированных у студентов по первой теме дисциплины (определяется на первой лабораторной работе); количество пропусков занятий студентами на момент составления прогноза. Для прогнозирования задаются эталонные объекты, соответствующие центрам четырех заданных групп. Прогнозирование будущей типологической группы студента осуществляется по расстоянию до ближайшего эталона [25].

В литературе также описывается подход к прогнозированию успеваемости студентов в зависимости от мотивационных предпочтений учения на основе нечеткой логики [13]. В результате с учетом степени мотивации учения прогнозируется возможность перехода студентов между типологическими группами в зависимости от уровня успеваемости по выбранному предмету.

Анализ литературных источников позволяет выделить следующие основные группы методов прогнозирования успеваемости студентов:

- методы, основанные на регрессионных моделях,
- методы, основанные на дискриминантных моделях,
- нейросетевые технологии,
- методы кластерного анализа и таксономии.

Как видно из приведенного выше обзора, подходы к прогнозированию успеваемости весьма разнообразны и, как правило, отличаются:

- конечной целью прогнозирования,
- набором оценочных факторов,
- периодом прогноза.

Одни - просто прогнозируют будущую типологическую группу студентов (по уровню успеваемости, по степени мотивации учения), другие - конечную оценку по отдельной дисциплине, третьи определяют факторы, влияющие на успеваемость, цель четвертых - оценка перспективности абитуриентов и т.д.

Ниже приводится описание подхода на базе методов таксономии и эталонов. Методы таксономии позволяют рассматривать множество исходных данных практически произвольной природы, а также производить разбиение объектов не по одному параметру, а по целому набору признаков, что дает возможность анализа результатов учебной деятельности студентов за весь период обучения.

В данной работе проверяется гипотеза о том, что при смешанной модели данные учебной аналитики LMS наряду с экспертным мнением преподавателя позволяют прогнозировать успешность студентов. Понятие «успешность обучения» имеет разную трактовку, некоторые ученые связывают его с успеваемостью студентов, Б. Г. Ананьев определяет его как «оптимальное сочетание темпа, напряженности, индивидуального своеобразия (стиля) учебной работы, степени прилежания и усилий, которые прилагает обучаемый для достижения определенных результатов» [7].

На основе данных учебной аналитики LMS можно получить большой массив эмпирических данных о поведении студента на online платформе, определяющих вовлеченность, непрерывность, результативность, самостоятельность и другие аспекты учебной деятельности студента [20]. Модели прогнозирования только на основе этих данных позволяют соотнести текущую деятельность обучающегося с его будущими перспективами, например, «забросить учебу» или получить высокий итоговый результат. Такую диагностику необходимо проводить как можно раньше, чтобы у преподавателей было достаточно времени для проведения образовательных мероприятий, способствующих повышению успешности обучения студента.

Одной из возможностей учебной аналитики является накопление данных о поведении студентов, обучавшихся в предыдущие временные периоды. Это позволяет выделить группы студентов со сходным поведением и при прогнозировании учитывать накопленный «опыт», а не только текущую информацию о деятельности студента.

В работе [23] отмечается, что для прогнозирования чаще всего используются линейные модели, такие как линейная регрессия и логистическая регрессия из-за их простоты и возможности интерпретировать линейные закономерности между данными учебной аналитики и достижениями студентов. В [19] авторы использовали online данные о поведении более 220 студентов инженерных специальностей для изучения факторов, влияющих на успешность их обучения. В [5] описывается методика прогнозирования прогресса студентов средствами регрессионного анализа на основе более 20 эмпирических показателей. В работе [2] на основе k-medoids кластеризации студенты были разделены на группы успешности выполнения самостоятельной работы. Затем на основе регрессионных моделей сформулированы стратегии поддержки студентов каждой группы.

Метод опорных векторов (SupportedVectorMachine), относящийся к алгоритмам линейного машинного обучения с учителем, в контексте прогнозирования использовался для нескольких целей.

Другим методом прогнозирования является использование деревьев решений (DecisionTrees). В [2] с помощью интеллектуального анализа данных и деревьев решений выделяется несколько моделей поведения студентов при online обучении, которые определяют способ взаимодействия обучающегося с содержимым, его длительность и перспективу завершения online курса. В [10] предлагается классификационная модель для выявления студентов, которые могут «забросить» учебу в течение одной недели. Для моделирования используется байесовский классификатор и алгоритм C 4.5 построения дерева решений. Результаты исследования показали, что коэффициент удержания был бы выше, если бы преподаватели своевременно вмешивались в процесс обучения на основе результатов прогнозирования.

В работах [10;12] для прогнозирования использовался алгоритм «случайный лес» (RandomForest), суть которого заключается в использовании большого количества деревьев решения, каждое из которых само по себе дает очень невысокое качество классификации, но за счет их большого количества получается хороший результат. Пример использования алгоритма для прогноза отсева обучающихся показан в работе [12]. Авторы разработали разные модели деревьев решений, в том числе RandomForest, оценили точность прогноза и показали, что использование последней улучшает точность прогноза до 91 %. Также были отмечены проблемы с производительностью и масштабированием модели.

В работе [14] прогнозирование выполняется на основе вероятностной марковской модели. Авторами предлагается отслеживать индивидуальные траектории прохождения слушателями курса и на их основе определять вероятности переходов между группами слушателей по успеваемости в процессе обучения. Это позволяет предсказывать результаты выполнения заданий курса разными категориями обучающихся.

Определенная часть элементов учебного процесса при смешанном обучении может быть полностью реализована на online платформе. Прежде всего, это самостоятельная работа студентов, а также выполнение различных контрольно-измерительных мероприятий текущего и промежуточного контроля. Для оценки качества самостоятельной работы применяется совокупность показателей, которые при смешанном обучении могут быть рассчитаны на основе данных учебной аналитики. Учебную аналитику используют и для анализа выполнения студентами контрольно-измерительных заданий. Кроме того, при смешанном обучении преподаватель может непосредственно наблюдать за студентами на аудиторном занятии, его экспертные оценки могут дополнить данные для прогнозирования успешности обучения, что позволит улучшить прогнозирующие характеристики моделей.

Отличительные особенности предлагаемого подхода состоят в следующем:

– прогнозируется успеваемость за весь оставшийся период обучения студента;

– в качестве оценочных факторов используются текущие результаты обучения студентов по всем ранее изученным дисциплинам;

– прогноз основывается на прецедентах и состоит в поиске группы студентов - аналога, обучающихся на старших курсах или закончивших обучение.

## 1.2. Цепи Маркова как метод прогнозирования.

Особое место в теории случайных процессов занимают марковские процессы. Случайный процесс, протекающий в системе I с дискретными состояниями $i_1,i_2,\ldots, i_i,\ldots$, называется марковским, или случайным процессом без последействия, если для любого момента времени $t_0$ вероятность каждого из состояний системы в будущем (при $t>t_0$) зависит только от ее состояния в настоящем ($t=t_0$) и не зависит от того, когда и как она пришла в это состояние, то есть не зависит от ее поведения в прошлом (при $t<t_0$). (Будущее зависит от прошлого через настоящее).

Марковские процессы делятся на процессы с дискретным и с непрерывным временем. В некоторых источниках под цепями Маркова понимают только марковские процессы с дискретным временем, однако есть и авторы (Кельберт, Сухов), которые все Марковские процессы называют цепями Маркова. В данной работе под цепями Маркова будем понимать Марковские процессы с дискретным временем.

Итак, цепью Маркова называется Марковский случайный процесс с дискретным временем, в котором его возможные состояния $i_1,i_2,\ldots, i_i$ можно заранее перечислить, а переход из состояния в состояние происходит мгновенно (скачком), но только в определенные моменты времени ($t_0$, $t_1$...), называемые шагами процесса [9].

Случайный механизм, вызывающий изменение состояния, описывается матрицей перехода P с элементами $p_{ij}$, где $i,j\in I$. Элемент $p_{ij}$ равен вероятности, с которой система перейдет из состояния i в состояние j за единицу времени. Таким образом $p_{ij}$ – это условная вероятность того, что система будет находиться в состоянии j в следующий момент, при условии, что в данный момент она находится в состоянии i. значит все элементы P неотрицательны, но не превышают 1, и сумма элементов в любой строке равна 1:

$$0 \leq p_{ij} \leq 1 \; \forall i,j \in I$$

$$\sum_{j\in I} p_{ij} = 1 \ \forall i \in I.$$

Матрица P, обладающая такими свойствами называется стохастической, т. е. вероятностной.

Простейший случай имеет вид 2x2 (пространство из 2 состояний). Можно считать, что состояниями являются 0 и 1 [8].

Тогда элементы матрицы имеют вид $p_{ij}$, I, j=0,1, а стохастическую матрицу можно представить в виде:

$$\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}, \text{где } 0 \le \alpha, \beta \le 1.$$

В частности при $\alpha = \beta$=0 получаем единичную, а при $\alpha = \beta$=1 антидиагональную матрицу

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Система с единичной матрицей остается в начальном состоянии навсегда, а в антидиагональном случае она меняет состояние в каждый момент времени, переходя из 0 в 1 и обратно.

С другой стороны при $\alpha = \beta = \frac{1}{2}$ - мы получаем матрицу

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

В этом случае система может либо остаться в том же состоянии, либо поменять его с вероятностью $\frac{1}{2}$.

Пусть $X_n$- состояние системы в момент n. Правила задающие марковскую цепь с начальным распределением λ и матрицей перехода P таковы:

1. $X_0$ имеет распределение λ:$P(X_0 = i) = \lambda_i \ \forall i \in I$
2. Более общим образом,$\forall n \forall$ и$i_0, \dots, i_n \in I$ вероятность $P(X_0 = i_0,\ X_1 = i_1, \dots ,\ X_n = i_n)$ того, что система находится в состоянии $i_0, i_1, \dots, i_n$ в моменты времени 0, 1, ... , $n$ записывается как произведение $P(X_0 = i_0,\ X_1 = i_1, \dots ,\ X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} 1$ - это частный случай 2 при n=0.

Для условной вероятности $P(X_{n+1} = j|\, X_{0=}i_0,\, X_1 = i_1, \dots ,\ X_n = i_n)$ того, что состояние в момент n+1 есть при условии что заданы состояния $i_0, \dots, i_{n-1}$ *и* $i_n$=i *в моменты времени 0,..., n — 1*,n [8]:

$$P(X_{n+1} = j|\, X_{0=}i_0,\, X_1 = i_1, \dots , \qquad X_n = i_n)$$

$$= \frac{P(X_0 = i_0, \dots, X_{n-1} = i_n - 1,\ X_n = i, \dots ,\ X_{n+1} = j)}{P(X_0 = i_0,\, X_{n-1} = i_{n-1}, \dots ,\ X_n = i)} =$$

$$= \frac{\lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i} p_{ij}}{\lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i}} = p_{ij}.$$

Таким образом, при условии, что $X_{0=}\, i_0,\, X_{n-1} = i_{n-1}, \dots ,\ X_n = i,\ X_{n+1}$ имеет распределение $p_{ij}, j \in I$. В частности, условное распределение $X_{n+1}$ не зависит от $i_0, i_1, \dots, i_{n-1}$ .т.е. зависит только от состояния $i$ в последний предшествующий момент n.

Эта формула иллюстрирует свойство ограниченной памяти цепи Маркова.

Теперь нас интересует вероятность $P(X_n = j)$ того, что в момент $n$ наша система находится в состоянии j. Для n=1:

$$P(X_n = j) = \sum\nolimits_{i \in I} P(\ X_0 = i,\ X_1 = j)$$

где $i$- все возможные начальные состояния.

Тогда, $\sum_{i \in I} P(\ X_0 = i,\ X_1 = j) = \sum_{i \in I} \lambda_i p_{i,j} = (\lambda P)_j$ для $n = 1$.

И для общих значений n:

$$P(X_n = j) = \sum_{i_0, i_1, \dots, i_{n-1}} P(X_0 = i_0, \dots, X_{n-1} = i_n - 1,\ X_n = j) =$$

$$= \sum_{i_0, i_1, \dots, i_{n-1}} \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1}} j = (\lambda P^n)_j$$

где $P^n$ - n-я степень матрицы P. Таким образом стохастический вектор, описывающий распределение случайной величины $X_n$, можно получить помножив матрицу $P^n$ к начальному стохастическому вектору λ.

Теперь аналогично:

$$P(X_n = i, X_{n+1} = j) = \sum_{i_0, i_1, \dots, i_{n-1}} P(X_0 = i_0, \dots, X_n = i,\ X_{n+1} = j) =$$

$$= \sum_{i_0, i_1, \dots, i_{n-1}} \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i} p_{ij} = (\lambda P^n)_j p_{ij}$$

и отсюда следует, что

$$P(X_{n+1} = j | X_n = i) = \frac{P(X_n = i, X_{n+1} = j)}{P(X_n = i)} = \frac{(\lambda P^n)_j p_{ij}}{(\lambda P^n)_j} = p_{ij.}$$

То есть элемент $p_{ij}$ равен условной вероятности того, что в следующий момент состояние будет у, если в данный момент оно есть i [14].

$$P(X_0 = i, X_n = j) = \sum_{i_1, \dots, i_{n-1}} P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1},\ X_n = j) =$$

$$= \sum_{i_0, i_1, \dots, i_{n-1}} \lambda_i p_{i i_1} \dots p_{i_{n-1} j} p_{ij} = \lambda_i (P^n)_{ij}$$

$$P(X_n = j | X_0 = i) = \frac{P(X_0 = i, X_n = j)}{P(X_0 = i)} = \frac{\lambda_i (P^n)_{ij}}{\lambda_j} = p_{ij.}$$

Значит, элемент $\lambda_i (P^n)_{ij}$ матрицы $P^n$ дает вероятность перехода за $n$ шагов из состояния $i$ в состояние j.

В общем случае $P(X_k = i, X_{n+k} = j) = (\lambda P^k)_j (P^n)_{ij}$ и

$$P(X_{k+n} = j | k = i) = \frac{P(X_k = i, X_{k+n} = j)}{P(X_k = i)} = \frac{(\lambda P^k)_j (P^n)_{ij}}{(\lambda P^k)_j} = (P^n)_{ij}.$$

Теперь можно дать точное определение цепи Маркова с дискретным временем:

Говорят, что последовательность случайных величин $X_n$ со значениями в конечном или счетном множестве I образует цепь Маркова с дискретным временем с начальным распределением λ и матрицей перехода P, если $\forall i_0, i_1, \dots, i_{n-1} \in I$ совместное распределение $P(X_0 = i_0,\ X_1 = i_1, \dots ,\ X_n = i_n)$ равно $\lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$.

Далее приведем примеры применения цепей Маркова для прогнозирования различных процессов.

**<u>Пример 1</u>**

**Условие.**

Погода классифицируется в прогнозах как ясная, умеренно пасмурная и пасмурная.

– Если погода ясная, то вероятность, что она будет ясной на следующий день, составляет 0.5; вероятность, что она будет умеренно пасмурной, равна 0.4; а вероятность пасмурной погоды на следующий день составляет 0.1.

– Если погода умеренно пасмурная, то вероятность, что на следующий день она будет ясной, равна 0.3; вероятность, что погода останется умеренно пасмурной, равна 0.5; а вероятность пасмурной погоды на следующий день составляет 0.2.

– Если же погода пасмурная, то вероятность, что она будет ясной на следующий день составляет 0.2; вероятность что она станет умеренно пасмурной, равна 0.4; вероятность что на следующий день она останется пасмурной, равна 0.4.

– Вопрос 1: Если вероятность ясной погоды в воскресенье равна 0.6, а вероятность умеренно пасмурной — 0.4, то какова вероятность, что погода в понедельник будет ясной?

– Вопрос 2: Какова вероятность, что во вторник погода будет умеренно пасмурной?

**Решение:**

Если порядок, в котором перечисляются погодные условия, таков: ясно, умеренно пасмурно и пасмурно, то:

$$p^{(0)} = (0{,}6; 0{,}4; 0)$$

$$T = \begin{vmatrix} 0{,}5 & 0{,}4 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{vmatrix}.$$

Следовательно, $p^{(1)} = (0{,}6; 0{,}4; 0) * \begin{vmatrix} 0{,}5 & 0{,}4 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{vmatrix} = (0{,}42; 0{,}44; 0{,}14)$ и вероятность, что в понедельник будет ясная погода, равна $0{,}42$.

Компоненты матрицы С вычисляются следующим образом [16]:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}=0.6\cdot 0.5+0.4\cdot 0.3+0\cdot 0.2=0.3+0.12+0=0.42$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}=0.6\cdot 0.4+0.4\cdot 0.5+0\cdot 0.4=0.24+0.2+0=0.44$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}=0.6\cdot 0.1+0.4\cdot 0.2+0\cdot 0.4=0.06+0.08+0=0.14$

Пусть $p_1^{(2)}$ – вероятность того, что во вторник будет ясная погода, $p_2^{(2)}$– вероятность того, что во вторник будет умеренно пасмурно и $p_3^{(2)}$– вероятность того, что во вторник будет пасмурно [19].

Пусть $p^{(2)} = (p_1^{(2)}, p_2^{(2)}, p_3^{(2)})$

Тогда $p^{(2)} = (0{,}42; 0{,}44; 0{,}14) * \begin{vmatrix} 0{,}5 & 0{,}4 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{vmatrix} = (0{,}37; 0{,}444; 0{,}186).$

Компоненты матрицы С вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}=0.42\cdot 0.5+0.44\cdot 0.3+0.14\cdot 0.2=0.21+0.132+0.028=0.37$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}=0.42\cdot 0.4+0.44\cdot 0.5+0.14\cdot 0.4=0.168+0.22+0.056=0.444$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}=0.42\cdot 0.1+0.44\cdot 0.2+0.14\cdot 0.4=0.042+0.088+0.056=0.186$

Следовательно, вероятность того, что во вторник будет умеренно пасмурная погода равна 0.444.

**Пример 2.** Оценка будущих продаж.

Цепи Маркова также применяются при оценке будущих продаж. Например, сделав опрос среди покупателей той или иной марки автомобиля об их следующем выборе, можно составить матрицу *T*.

**Условие.**

В процессе опроса владельцев автомобилей трех американских марок: марки *A*, марки *B*, марки *C*, им был задан вопрос о том, какую торговую марку они бы выбрали для следующей покупки.

1. Среди владельцев автомобилей марки *A*

20% сказали что выберут опять эту же марку, 50% сказали, что они бы перешли на марку *B*%, а 30% заявили, что предпочли бы марку *C*.

Среди владельцев автомобилей марки $B$ 20% сказали, что перейдут на марку $A$, в то время как 70% заявили, что приобрели бы опять автомобиль марки $B$, а 10% заявили, что в следующий раз предпочли бы марку $C$.

Среди владельцев автомобилей $C$ 30% ответили, что перешли бы на марку $A$, 30% сказали, что перешли бы на марку $B$, а 40% заявили, что остались бы верны той же марке $C$

Вопрос 1: Если некто приобрел автомобиль марки $A$, то какова вероятность, что его второй машиной будет автомобиль марки $C$?

Вопрос 2: Если при покупке первой машины покупатель подбросил монету, выбирая между автомобилями марки $B$ и $C$, то какова вероятность, что его третьей машиной станет автомобиль марки $B$?

**Решение.**

Матрица перехода для этого события имеет вид:

$$\begin{vmatrix} 0,2 & 0,5 & 0,3 \\ 0,2 & 0,7 & 0,1 \\ 0,3 & 0,3 & 0,4 \end{vmatrix}.$$

Для ответа на первый вопрос имеем:

$$p^{(1)} = (1; 0; 0) * \begin{vmatrix} 0,2 & 0,5 & 0,3 \\ 0,2 & 0,7 & 0,1 \\ 0,3 & 0,3 & 0,4 \end{vmatrix} = (0,2; 0,5; 0,3).$$

Компоненты матрицы С вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}=1\cdot 0.2+0\cdot 0.2+0\cdot 0.3=0.2+0+0=0.2$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}=1\cdot 0.5+0\cdot 0.7+0\cdot 0.3=0.5+0+0=0.5$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}=1\cdot 0.3+0\cdot 0.1+0\cdot 0.4=0.3+0+0=0.3$

Вероятность того, что вторая машина будет марки С равна 0,3. Для ответа на второй вопрос требуется найти [19]

$$\mathrm{T}^{(2)} = \begin{vmatrix} 0,23 & 0,54 & 0,23 \\ 0,21 & 0,62 & 0,17 \\ 0,24 & 0,48 & 0,28 \end{vmatrix}.$$

Для нахождения ответа на второй вопрос имеем:

$$p^{(2)} = (0; 0{,}5; 0{,}5) * \begin{vmatrix} 0{,}23 & 0{,}54 & 0{,}23 \\ 0{,}21 & 0{,}62 & 0{,}17 \\ 0{,}24 & 0{,}48 & 0{,}28 \end{vmatrix} = (0{,}225; 0{,}55; 0{,}225).$$

Компоненты матрицы С вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}=0\cdot 0.23+0.5\cdot 0.21+0.5\cdot 0.24= 0+0.105+0.12=0.225$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}=0\cdot 0.54+0.5\cdot 0.62+0.5\cdot 0.48=0+0.31+ 0.24=0.55$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}=0\cdot 0.23+0.5\cdot 0.17+0.5\cdot 0.28=0+0.085+0.14=0.225$

Поэтому вероятность того, что второй автомобиль будет марки А равна 0,225.

**Пример 3.** Данные, полученные при исследовании рынка ценных бумаг показали, что рыночная цена одной акции акционерного общества А открытого типа может колебаться в пределах от 1 руб. до 10 руб. включительно.

Рассматривая в качестве системы S одну акцию этого акционерного общества, будем интересоваться следующими четырьмя состояниями этой системы, характеризующимися рыночной ценой акции:

$s_1$=от 1 руб. до 4 руб.;

$s_2$=от 4 руб. до 7 руб.;

$s_3$=от 7 руб. до 9 руб.;

$s_4$=от 9 руб. до 10 руб. включительно.

Замечено, что рыночная цена акции в будущем зависит (существенно) ее цены в текущий момент времени. При этом в силу случайных воздействий рынка изменение рыночной цены акции может произойти в любой случайный момент времени. Переходы системы Sиз состояния в состояние происходит со следующими плотностями вероятностей переходов, почти не изменяющимися с течением времени:

$$\mathrm{P} = \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 3 & 2 & 0 & 1 \\ 0 & 0 & 4 & 0 \end{pmatrix}.$$

Попытаемся составить (приближенный) долгосрочный прогноз рыночной цены акции и ответить на вопрос: стоит ли приобретать акции акционерного общества А по цене 6 руб. за акцию?

Из условий примера следует, что в системе S протекает дискретный однородный марковский процесс с непрерывным временем. Следовательно, все потоки событий, порождающие переходы системы S из состояния в состояние, - простейшие.

Размеченный граф состояний системы S выглядит следующим образом (см. рис. 1.1.).

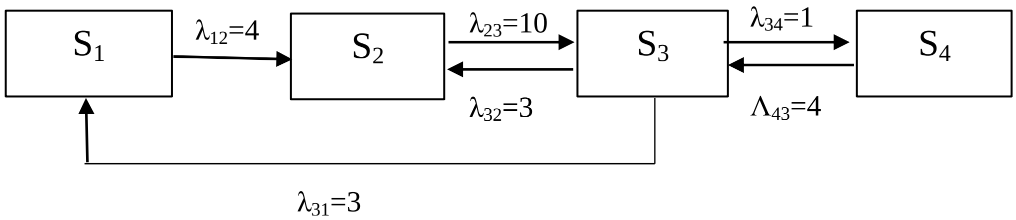
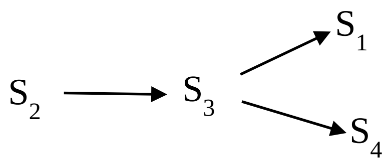



**Рис. 1.1. Размеченный граф состояний системы S.**

По графу видно, что система S эргодична, то есть из любого своего состояния может перейти (за конечное число шагов) в любое другое состояние).

Например, из состояния $S_2$ система S может перейти в любое другое состояние по следующему пути:




Итак, выполняются все условия теоремы, по которой существуют предельные вероятности состояний $p_1$,$p_2$, $p_3$, $p_4$, не зависящие от времени и от состояний системы S в начальный момент времени. Эти предельные вероятности как раз и дают нам информацию о долгосрочном прогнозе рыночной цены акции.

Составим по одному из трех выше данных правил систему четырех линейных алгебраических уравнений с четырьмя неизвестными: $p_1$,$p_2$, $p_3$, $p_4$.

$$\begin{cases} -4p_1 + 3p_3 = 0, \\ 4p_1 - 10p_2 + 2p_3 = 0, \\ 10p_2 - 6p_3 + 4p_4 = 0, \\ p_3 - 4p_4 = 0. \end{cases}$$

Из последнего уравнения системы $p_3 = 4p_4$. Подставим это в 1-е уравнение системы, найдем $p_1 = 3p_4$. Если найденные значения $p_3$ и $p_1$, выраженные через $p_4$, подставить во 2-е уравнение системы, то получим: $p_2 = 2p_4$. Таким образом, мы нашли общее решение системы:

($p_1 = 3p_4$; $p_2 = 2p_4$; $p_3 = 4p_4$; $p_4$), зависящее от одного свободного параметра $p_4 \in [0,1]$и представляющее собой множество всех частных решений. Из этих частных решений найдем то, которое удовлетворяет нормировочному условию $p_1 + p_2 + p_3 + p_4 = 1$. Подставим в это равенство найденные значения $p_1, p_2, p_3$и получим что $p_4 = 0{,}1$.

Тогда $p_1 = 0{,}3$; $p_2 = 0{,}2$; $p_3 = 0{,}4$; $p_4 = 0{,}1$.

Таким образом, долгосрочный прогноз рыночной цены акции состоит в том, что по истечении достаточного времени вероятнее всего ($p_3 =$ 0,4>$p_1, p_2, p_4$) цена акции будет колебаться в пределах от 7 до 9 руб. Поэтому стоит приобрести акции по цене 6 руб.

# ГЛАВА 2. ПРИМЕНЕНИЕ ЦЕПЕЙ МАРКОВА ПРИ ПРОГНОЗИРОВАНИИ УСПЕВАЕМОСТИ СТУДЕНТОВ.

## 2.1. Математическая модель прогнозирования успеваемости студентов с применением цепей Маркова.

Процесс называется марковским, если в любой момент времени вероятность любого состояния системы в будущем зависит только от состояния системы в текущий момент и не зависит от того, каким образом система пришла в это состояние. [1] Переходной вероятностью $p_{ij}$ называется условная вероятность перехода системы на к-ом шаге в состояние $S_j$ при условии, что на (к-1) шаге система находилась в $S_j$. Матрицей перехода системы называют матрицу, которая содержит все переходные вероятности этой системы. Равенство Маркова связывает матрицу перехода $P_n$ за п шагов с матрицей перехода $P_1$ за 1 шаг.

$$P_n = P_1^n.$$

Под состоянием системы мы будем понимать определенное распределение студентов по группам, в зависимости от успеваемости. Считаем, что состояние системы в текущем году зависит только от её состояния в предыдущий год, то есть мы определяем данный процесс изменения системы как цепь Маркова. Система имеет 4 состояния: $S_0$, $S_1$, $S_2$, $S_3$ - где $S_0$-состояние системы после первого года обучения, $S_1$ -состояние системы после второго года, $S_2$ -состояние системы после третьего года, $S_3$ - состояние системы после четвертого года (соответствует окончанию бакалавриата). Используя данные результатов экзаменов после первого курса, вычислим средний балл для каждого студента и распределим всех студентов на п групп. Далее смотрим, как меняется количество студентов в каждой группе в процессе обучения. Состояние системы описывается матрицей-столбцом, каждый элемент которой представляет собой наличие студентов в группе после соответствующего года.

$$S_k = \begin{bmatrix} S_1^k \\ S_2^k \\ . \\ S_n^k \end{bmatrix}.$$

Здесь $S_i^k$-число студентов, оказавшихся в i-ой группе после к+1 курса. Здесь и далее матрицей перехода студентов $A_{0k}$ будем называть матрицу, каждый элемент которой $a_{ij}$ представляет число студентов, перешедших из i-ой группы в у-ую после (к+1)-го курса. Элементы матрицы перехода системы $P_{0k}$ определяются по формуле статистической вероятности. Применительно к данной задаче формула статистической вероятности имеет вид

$$P_{ij} = \frac{a_{ij}}{S_i^k}.$$

Где $a_{ij}$ - число студентов, перешедших из i-ой группы в j-ую после (к+1)-го курса, $S_i^k$число студентов, оказавшихся в i-ой группе после к-го курса.

Используя собранные данные, получим начальное состояние системы Б0 и матрицу переходов студентов $A_{01}$[14].

$$S_k = \begin{bmatrix} S_1^0 \\ S_2^0 \\ . \\ S_n^0 \end{bmatrix} \qquad A_{01} = \begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ a_{31} & a_{32} & . & a_{3n} \\ a_{41} & a_{42} & . & a_{4n} \end{bmatrix}.$$

Здесь $S_i^0$ -число студентов, оказавшихся в i-ой группе после первого курса, $a_{ij}$ - число студентов, перешедших из i-ой группы в j-ую после второго курса.

Следующим шагом является составление матрицы вероятностей перехода $P_{01}$, элемент матрицы которой определяется по формуле.

$$P_{ij} = \frac{a_{ij}}{S_i^0}.$$

Далее, используя равенство Маркова, определим матрицу вероятностей переходов из начального состояния So (после первого курса) в состояние $S_2$ (после третьего курса) за два шага. Используя равенство Маркова (1), получим:

$$P_{02} = P_{01}^2.$$

Чтобы получить матрицу перехода студентов из состояния Б0 (после первого курса) в состояние $S_2$ (после третьего курса) А02, необходимо матрицу перехода системы $P_{02}$ поэлементно умножить на соответствующие элементы матрицы начального состояния системы, при этом округляя полученные результаты до целого числа, тогда

$$a_{ij}^{02} = p_{ij}^{02} \cdot S_i^0.$$

Теперь можно определить число студентов в каждой группе после окончания 3 курса.

$$s_j^2 = \sum_{i=1}^{n} a_{ij}^{02}.$$

Полученные значения запишем в матрицу-столбец $S_2$, каждый элемент которой представляет собой расчетное число студентов в соответствующей группе после окончания 3-го курса. Далее составляем матрицу-столбец $S_2'$, основываясь на данных эксперимента. Повторяем эту же процедуру для 4-го курса [22].

Таким образом, у нас есть реальные данные, и спрогнозированные, что позволяет построить уравнение регрессии, где зависимой переменной будут реальные данные, а независимой -спрогнозированные.

## 2.2. Анализ исходных данных

Была использована база данных из дипломной работы «Сравнительный статистический анализ успеваемости студентов ФМФ 2005 -2008 годов поступления в условиях обязательного ЕГЭ» студентов Гончаровой Ю.А., Ооржак О.М. 2005 года поступления.

| | Вступ-ые (мат-ка: ЕГЭ,уст.) | 1 семестр (мат.ан1, алг1, геом1) | 2 семестр (мат.ан2, алг2, геом2) | 3 семестр (мат.ан3, алг3) | 4 семестр (мат.ан4, геом3,прогр.) | 5 семестр (мат.лог., ТФДП, теор.чис.) | 6 семестр (теор.вер., диф.ур., чис.мет.) | 7 семестр (ТиМОМ, ТиМОИ, ТФКП) | 8 семестр (чис. сист., функ. ан.) | 9 семестр (ТОИ) |
|---|---|---|---|---|---|---|---|---|---|---|
| №1 | **3** | **3,3** | 3,3 | 3 | 4,3 | 4 | 3,3 | 4 | 4 | 4 |
| №2 | 3,5 | 5 | 5 | 5 | 4,3 | 5 | 4,7 | 5 | 5 | 5 |
| №3 | 4,5 | 4,3 | 4,3 | 4,5 | 4,7 | 4,3 | 4,3 | 5 | 4,5 | 5 |
| №4 | **3** | **3** | 3 | 3 | 3 | 3 | 3,3 | 4,3 | 3 | 5 |
| №5 | 3,5 | 3,3 | 3 | 4 | 3 | 4,3 | 3 | 4,3 | 4 | 3 |
| №6 | **3** | **3** | 3 | 3 | 3 | 3 | 3 | 3,7 | 3 | 4 |
| №7 | **3** | **4** | 3,7 | 3 | 4,3 | 4 | 4 | 4,7 | 5 | 4 |
| №8 | 3,5 | 3 | 3 | 3 | 3,3 | 3 | 3 | 4,3 | 3 | 3 |
| №9 | 3,5 | 3 | 4 | 4 | 3 | 3,3 | 4 | 4 | 4,5 | 3 |
| №10 | 3,5 | 3,3 | 4 | 4 | 3,3 | 3,3 | 3 | 4,3 | 3,5 | 3 |
| №11 | 3,5 | 3,7 | 3,3 | 4 | 3,3 | 4 | 4 | 4,7 | 4 | 4 |
| №12 | 4 | 3,3 | 3 | 3 | 3 | 3 | 3,3 | 4 | 4 | 3 |
| №13 | 3,5 | 3,3 | 4,7 | 4 | 4 | 4 | 5 | 5 | 4,5 | 4 |
| Ср. балл | 3,5 | 3,5 | 3,6 | 3,7 | 3,6 | 3,7 | 3,7 | 4,4 | 4 | 3,8 |

Для наглядности представим также график - динамика экзаменационных оценок по профильным дисциплинам за 1-9 семестры студентов специальности «математика-информатика» 2005 года поступления.

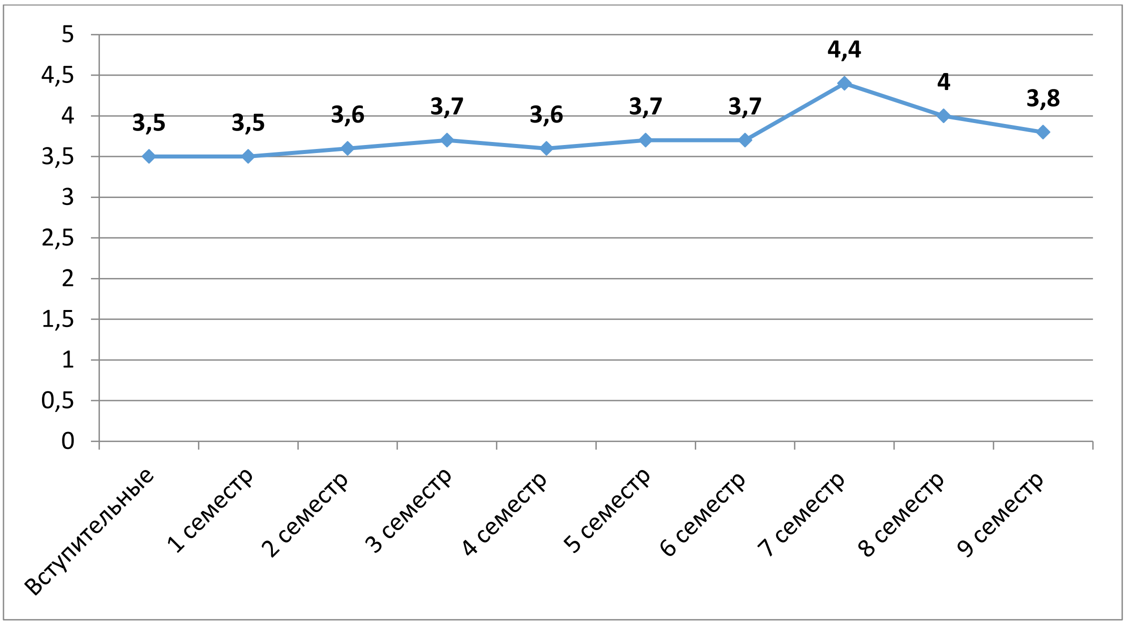



**Рис. 2.1. Динамика экзаменационных оценок за 1-9 семестры. Средние оценки по профильным дисциплинам группы (13 студентов) специальности «математика-информатика» 2005 года поступления.**

## 2.3. Практическое применение цепей Маркова для прогнозирования успеваемости студентов.

**Задача №1.** Исследуется успеваемость в группе студентов 5 курса 1 группы (13 студентов) специальности «математика-информатика» 2005 года поступления. Исходя из среднего балла оценок, полученных в сессию, сформированы четыре возможных состояния: «двоечник» (студент отчислен за неуспеваемость), «троечник», «хорошист» и «отличник».

В соответствии с вкладываемым в указанное понятие смыслом данные состояния удобно обозначить как $E_2$, $E_3$, $E_4$, $E_5$.

Согласно статистическим данным стохастическая матрица P имеет вид (запишем матрицу вместе с состояниями). Отчисленные после 1 семестра студенты не учитываются, все остальные 13 человек закончили высшее учебное заведение.

$$P = \begin{array}{ccccc} & E_2 & E_3 & E_4 & E_5 \\ E_2 & 1 & 0 & 0 & 0 \\ E_3 & 0 & 0{,}75 & 0{,}25 & 0 \\ E_4 & 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ E_5 & 0 & 0 & 1 & 0 \end{array}$$

Во второй строке: вступительные экзамены четыре студента сдали на «удовлетворительно», после первого семестра, один студент сдал сессию на «хорошо» (вероятность 0,25), три студента – остались троечниками (0,75).

В третьей строке: всего по итогам вступительных экзаменов восемь «хорошистов», шесть стали «троечниками» (0,75), один – остался «хорошистом» (0,125), один – «отличником» (0,125).

В четвертой строке, вступительные экзамены на отлично сдал один студент, который после 1 семестра стал «хорошистом».

Шаг процесса – один семестр, все время исследования – девять семестров. Пусть в начальном состоянии, после сессии первого семестра, студент находиться в состоянии $E_4$. Найдем векторы безусловных вероятностей после первого и второго шагов, то есть осуществим вероятностный прогноз на два временных интервала: после 2-го и 3-го семестров. После первого семестра (начальный вектор): $(p_1(0))=(0{:}0;\ 1;0)$. После второго семестра:

$$\{p_1(1)\} = \{p_1(0)\} \cdot P = (0{:}\,0;\ 1;0) \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix}$$

$$= (0;\ 0{,}75;\ 0{,}125;\ 0{,}125).$$

Компоненты матрицы C вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0\cdot 0+1\cdot 0+0\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0\cdot 0.75+1\cdot 0.75+0\cdot 0=0+0+0.75+0=0.75$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0\cdot 0.25+1\cdot 0.125+0\cdot 1=0+0+0.125+0=0.125$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44=}0\cdot 0+0\cdot 0+1\cdot 0.125+0\cdot 0=0+0+0.125+0=0.125$

После третьего семестра:

$$\{p_1(2)\} = \{p_1(1)\} \cdot P = (0;\ 0{,}75;\ 0{,}125;\ 0{,}125) \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix} =$$

*=(0;0,6563;0,3281;0,0156).*

Компоненты матрицы C вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0.75\cdot 0+0.125\cdot 0+0.125\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0.75\cdot 0.75+0.125\cdot 0.75+0.125\cdot 0=0+0.5625+0.0938+0=0.6563$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0.75\cdot 0.25+0.125\cdot 0.125+0.125\cdot 1=0+0.1875+0.0156+0.125=0.3281$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44}=0\cdot 0+0.75\cdot 0+0.125\cdot 0.125+0.125\cdot 0=0+0+0.0156+0=0.0156$

Таким образом, «хорошист» в первом семестре после сессии третьего семестра с вероятностью 0 окажется отчисленным, с вероятностью 0,6563 будет «троечником», с вероятностью 0,3281 – «хорошистом» и с вероятностью 0,0156 – «отличником».

Выполним расчеты после последующих семестров.

После четвертого семестра:

$$\{p_1(3)\} = \{p_1(2)\} \cdot P =$$

$$= (0; 0{,}6563; 0{,}3281; 0{,}0156) \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix} =$$

*=(0;0,7383;0,2207;0,0410).*

Компоненты матрицы C вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0.6563\cdot 0+0.3281\cdot 0+0.0156\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0.6563\cdot 0.75+0.3281\cdot 0.75+0.0156\cdot 0=0+0.492225+0.246075+0=0.7383$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0.6563\cdot 0.25+0.3281\cdot 0.125+0.0156\cdot 1=0+0.1641+0.0410+0.0156=0.2207$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44}=0\cdot 0+0.6563\cdot 0+0.3281\cdot 0.125+0.0156\cdot 0=0+0+0.0410+0=0.0410$

После пятого семестра:

$$\{p_1(4)\} = \{p_1(3)\} \cdot P =$$

$$= (0; 0{,}7383; 0{,}2207; 0{,}0410) \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix} =$$

*=(0;0,7192;0,2532;0,0276).*

Компоненты матрицы C вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0.7383\cdot 0+0.2207\cdot 0+0.041\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0.7383\cdot 0.75+0.2207\cdot 0.75+0.041\cdot 0=0++0.5537+0.1655+0=0.7192$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0.7383\cdot 0.25+0.2207\cdot 0.125+0.041\cdot 1=0+0.1846+0.0276+0.041=0.2532$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44}=0\cdot 0+0.7383\cdot 0+0.2207\cdot 0.125+0.041\cdot 0=0+0+0.0276+0=0.0276$

После шестого семестра:

$$\{p_1(5)\} = \{p_1(4)\} \cdot P = (0;0{,}7192;0{,}2532;0{,}0276) \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix} =$$

$=(0;0{,}7293;0{,}2391;0{,}0316)$.

Компоненты матрицы C вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0.7192\cdot 0+0.2532\cdot 0+0.0276\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0.7192\cdot 0.75+0.2532\cdot 0.75+0.0276\cdot 0=0+ 0.5394+0.1899+0=0.7293$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0.7192\cdot 0.25+0.2532\cdot 0.125+0.0276\cdot 1=0+ 0.1798+0.0317+0.0276=0.2391$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44}=0\cdot 0+0.7192\cdot 0+0.2532\cdot 0.125+0.0276\cdot 0=0+0+ 0.0316+0=0.0316$

После седьмого семестра:

$$\{p_1(6)\} = \{p_1(5)\} \cdot P = (0;0{,}7293;0{,}2391;0{,}0316) \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix} =$$

$=(0;0{,}7263;0{,}2438;0{,}0299)$.

Компоненты матрицы C вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0.7293\cdot 0+0.2391\cdot 0+0.0316\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0.7293\cdot 0.75+0.2391\cdot 0.75+0.0316\cdot 0=0+ 0.546975+0.179325+0=0.7263$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0.7293\cdot 0.25+0.2391\cdot 0.125+0.0316\cdot 1=0+ 0.1823+0.0299+0.0316=0.2438$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44}=0\cdot 0+0.7293\cdot 0+0.2391\cdot 0.125+0.0316\cdot 0=0+0+ 0.0299+0=0.0299$

После восьмого семестра:

$$\{p_1(7)\} = \{p_1(6)\} \cdot P = (0;0{,}7263;0{,}2438;0{,}0299) \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix} =$$

$=(0;0{,}7276;0{,}2420;0{,}0304)$.

Компоненты матрицы С вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0.7263\cdot 0+0.2438\cdot 0+0.0299\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0.7263\cdot 0.75+0.2438\cdot 0.75+0.0299\cdot 0=0+0.5447+0.1829+0=0.7276$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0.7263\cdot 0.25+0.2438\cdot 0.125+0.0299\cdot 1=0+0.1816+0.0305+0.0299=0.2420$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44}=0\cdot 0+0.7263\cdot 0+0.2438\cdot 0.125+0.0299\cdot 0=0+0+0.0304+0=0.0304$

После девятого семестра:

$$\{p_1(8)\} = \{p_1(7)\} \cdot P = (0;0{,}7276;0{,}2420;0{,}0304)\cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0{,}75 & 0{,}25 & 0 \\ 0 & 0{,}75 & 0{,}125 & 0{,}125 \\ 0 & 0 & 1 & 0 \end{vmatrix} =$$

$$=(0;0{,}7272;0{,}2426;0{,}0302).$$

Компоненты матрицы С вычисляются следующим образом:

$c_{11}=a_{11}\cdot b_{11}+a_{12}\cdot b_{21}+a_{13}\cdot b_{31}+a_{14}\cdot b_{41}=0\cdot 1+0.7276\cdot 0+0.242\cdot 0+0.0304\cdot 0=0+0+0+0=0$

$c_{12}=a_{11}\cdot b_{12}+a_{12}\cdot b_{22}+a_{13}\cdot b_{32}+a_{14}\cdot b_{42}=0\cdot 0+0.7276\cdot 0.75+0.242\cdot 0.75+0.0304\cdot 0=0+0.5457+0.1815+0=0.7272$

$c_{13}=a_{11}\cdot b_{13}+a_{12}\cdot b_{23}+a_{13}\cdot b_{33}+a_{14}\cdot b_{43}=0\cdot 0+0.7276\cdot 0.25+0.242\cdot 0.125+0.0304\cdot 1=0+0.1819+0.0303+0.0304=0.2426$

$c_{14}=a_{11}\cdot b_{14}+a_{12}\cdot b_{24}+a_{13}\cdot b_{34}+a_{14}\cdot b_{44}=0\cdot 0+0.7276\cdot 0+0.242\cdot 0.125+0.0304\cdot 0=0+0+0.0302+0=0.0302$

Таким образом, «хорошист» в первом семестре после сессии девятого семестра с вероятностью 0 окажется отчисленным, с вероятностью 0.7272 будет «троечником», с вероятностью 0.2426 – «хорошистом» и с вероятностью 0.0302– «отличником».

**Пример 2. Прогнозирование средней успеваемости студентов.**

Начальными данными для модели является таблица успеваемости студентов 5 курса 1 группы (13 студентов) специальности «математика-информатика» 2005 года поступления.

Исходя из данных успеваемости студентов после первого курса, вычислим средний балл группы и распределим всех студентов на 9 семестров: Далее расчет проводился по формулам, указанным в п.2.1.

| | 1 семестр (мат.ан1, алг1,геом1) | 2 семестр (мат.ан2, алг2,геом2) | 3 семестр (мат.ан3,алг3) | 4 семестр (мат.ан4,геом3, прогр.) | 5 семестр (мат.лог.,ТФДП, теор.чис.) | 6 семестр (теор.вер., диф.ур.,чис.мет.) | 7 семестр (ТиМОМ, ТиМОИ,ТФКП) | 8 семестр (чис.сист.,функ.ан.) | 9 семестр (ТОИ) |
|---|---|---|---|---|---|---|---|---|---|
| Ср. балл | 3,5 | 3,6 | 3,7 | 3,6 | 3,7 | 3,7 | 4,4 | 4 | 3,8 |
| Расчетные данные | 3,5 | 3,2 | 3,5 | 3,9 | 4,0 | 3,5 | 4,5 | 4,1 | 4,0 |

Сравнение экспериментальных и расчетных данных распределения студентов после 1-9 семестров представлено на рис.2.2 .По горизонтальной оси отложен номер семестра, по вертикальной оси – средний балл студентов. Расчетные данные получены с использованием уравнения Маркова.

Далее были построена регрессионная модель успеваемости студентов после 9-го курса. Составлялись уравнения парной линейной регрессии y=a+bx, где переменной x обозначены реальные данные, а переменной y - спрогнозированные. Результаты сведены в таблицу. Видно, что коэффициент корреляции r(xy) больше 0.7 (коэффициент детерминации $R^2$ больше 0.5), что говорит о сильной связи между исследуемыми параметрами (один признак определяет другой больше, чем на половину).

**Результаты регрессионного анализа**

| A | b | r(xy) | $R^{2}$" |
|---|---|---|---|
| 1.468326 | 0.799508 | 0.756728 | 0.623316 |

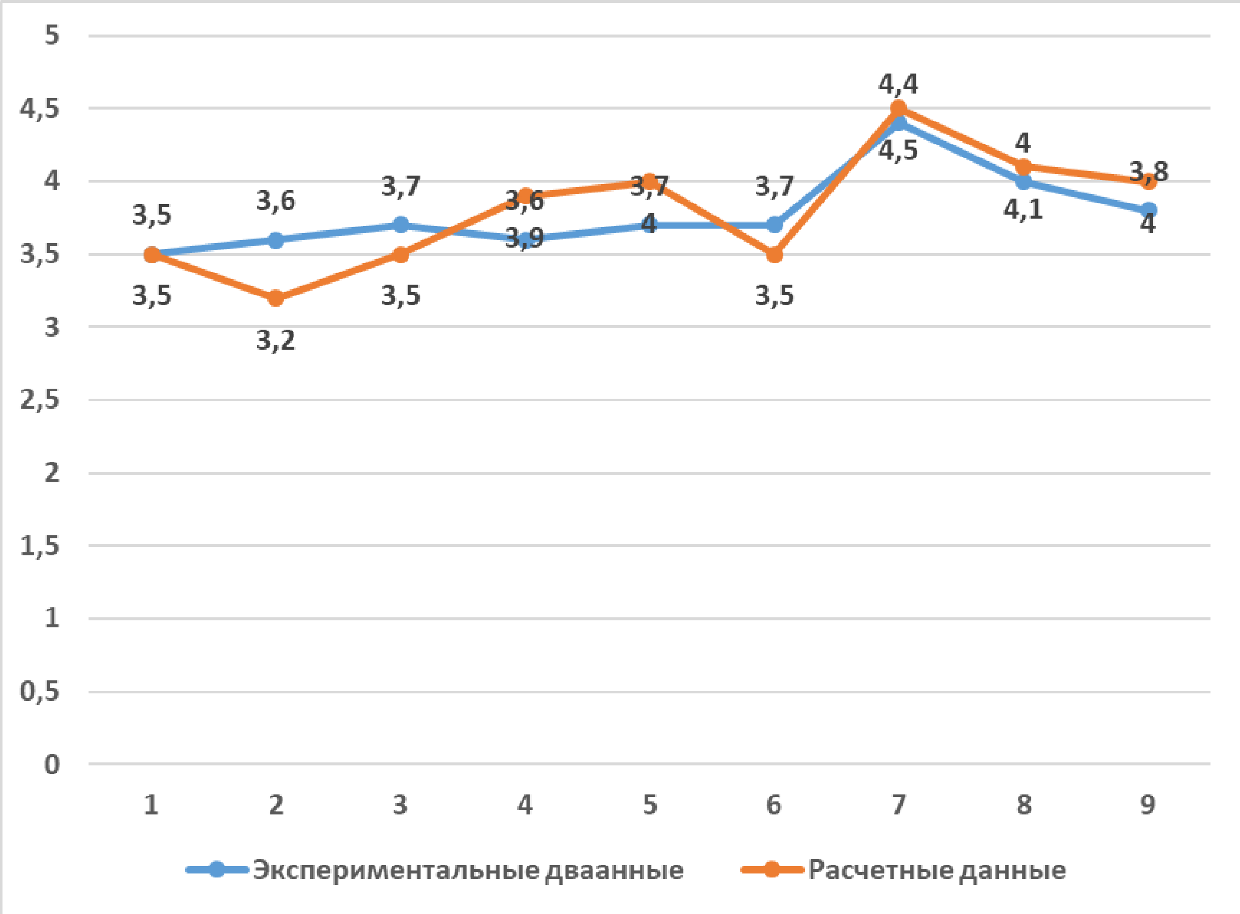



**Рис. 2.2. Расчетные и экспериментальные результаты успеваемости студентов после 1-9 семестров.**

Эта методология основана на аппарате Марковских цепей. Моделирование показало удовлетворительное совпадение результатов расчета с экспериментальными данными.

**Пример 3. Прогнозирование успешности студентов при переходе с одной группы студентов в другую.**

Для раннего прогнозирования успешности обучения классификация студентов осуществлялась после выполнения каждого контрольно-измерительного инструмента. В таблице приведены результаты классификации в соотнесении с группой успеваемости по каждому из семестров.

| | 1 семестр (мат.ан1, алг1,гео м1) | 2 семестр (мат.ан2, алг2,гео м2) | 3 семестр (мат.ан3,а лг3) | 4 семестр (мат.ан4,ге ом3, прогр.) | 5 семестр (мат.лог.,ТФ ДП, теор.чис.) | 6 семестр (теор.вер., диф.ур.,чис. мет.) | 7 семестр (ТиМОМ, ТиМОИ,ТФ КП) | 8 семестр (чис.сист.,функ. ан.) | 9 семес тр (ТОИ ) |
|---|---|---|---|---|---|---|---|---|---|
| Уд. | 9 | 7 | 6 | 8 | 6 | 7 | 0 | 3 | 5 |
| Хо р. | 3 | 4 | 5 | 4 | 6 | 4 | 8 | 5 | 5 |
| От л. | 1 | 2 | 2 | 1 | 1 | 2 | 5 | 5 | 3 |

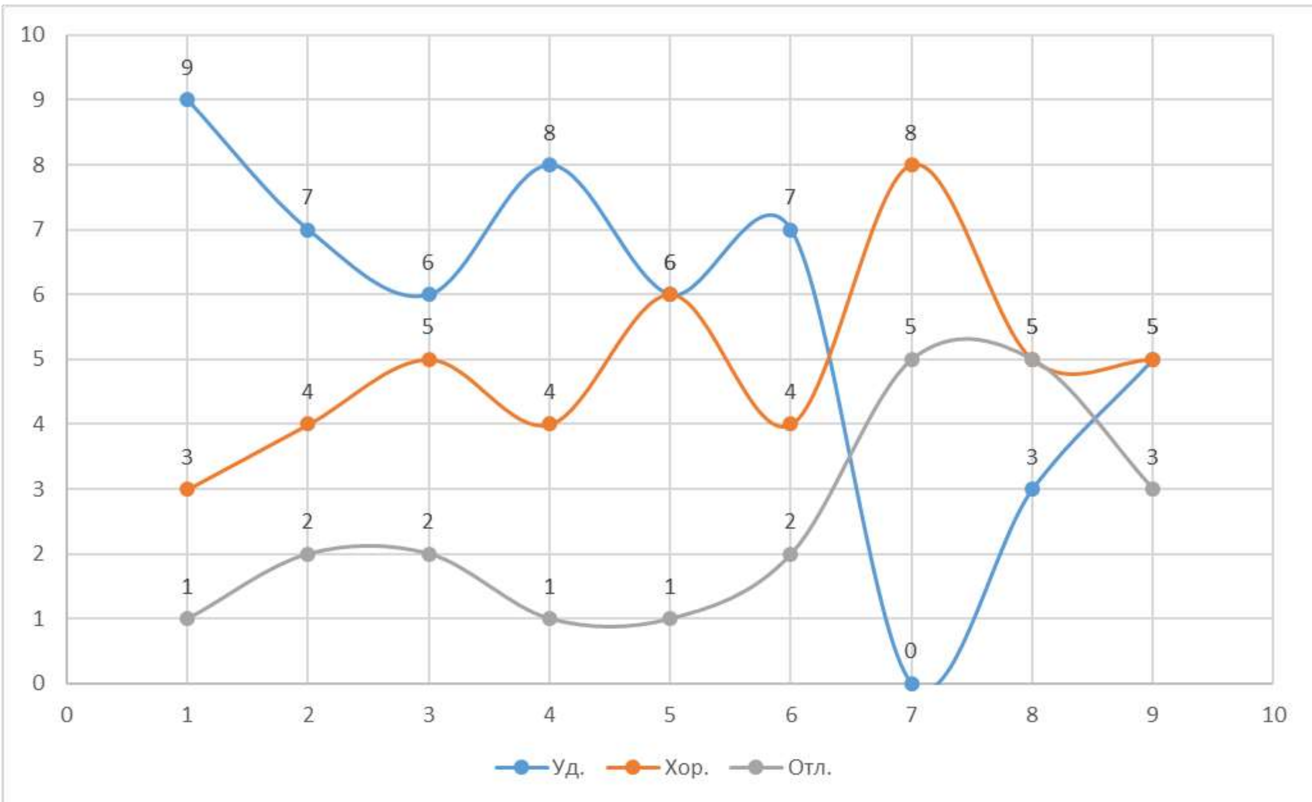



**Рис.2.3. Классификация выборки.**

Мониторинг состава групп успешности со стороны преподавателя позволяет выявить потенциально «выпадающих» студентов за несколько месяцев до итоговой аттестации, когда ситуацию еще можно исправить. Все студенты, попавшие в 3-ю группу, входят в группу «риска, о каждом из них формируется полный отчет об их деятельности на курсе и предоставляется преподавателю.

В таблице приведен процент студентов, отнесенных классификатором к группе 3 с низким уровнем успешности в течение семестра определенное число раз, в соотнесении с полученной оценкой по дисциплине.

| | 1 семестр (мат.ан1, алг1,гео м1) | 2 семестр (мат.ан2, алг2,гео м2) | 3 семестр (мат.ан3,а лг3) | 4 семестр (мат.ан4,ге ом3, прогр.) | 5 семестр (мат.лог.,ТФ ДП, теор.чис.) | 6 семестр (теор.вер., диф.ур.,чис. мет.) | 7 семестр (ТиМОМ, ТиМОИ,ТФ КП) | 8 семестр (чис.сист.,функ. ан.) | 9 семес тр (ТОИ) |
|---|---|---|---|---|---|---|---|---|---|
| Отл. | 7,7% | 15,4% | 15,4% | 7,7% | 7,2% | 15,4% | 38,5% | 38,5% | 23,1 % |
| Хор. | 23,1% | 30,8% | 38,5% | 30,8% | 46,1% | 30,8% | 61,5% | 38,5% | 38,5 % |
| Уд. | 69,2% | 53,8% | 46,1% | 61,5% | 46,1% | 53,8% | 0% | 23% | 38,5 % |

По окончанию семестра вновь проводится классификация студентов на группы успешности для каждой контрольной точки, чтобы учесть результаты тех попыток, которые студенты выполнили не в заданный преподавателем интервал времени. На основе этих данных, были построены матрицы переходов студентов по группам успешности. На рисунке 2.4 показан граф вероятности переходов после первого семестра.

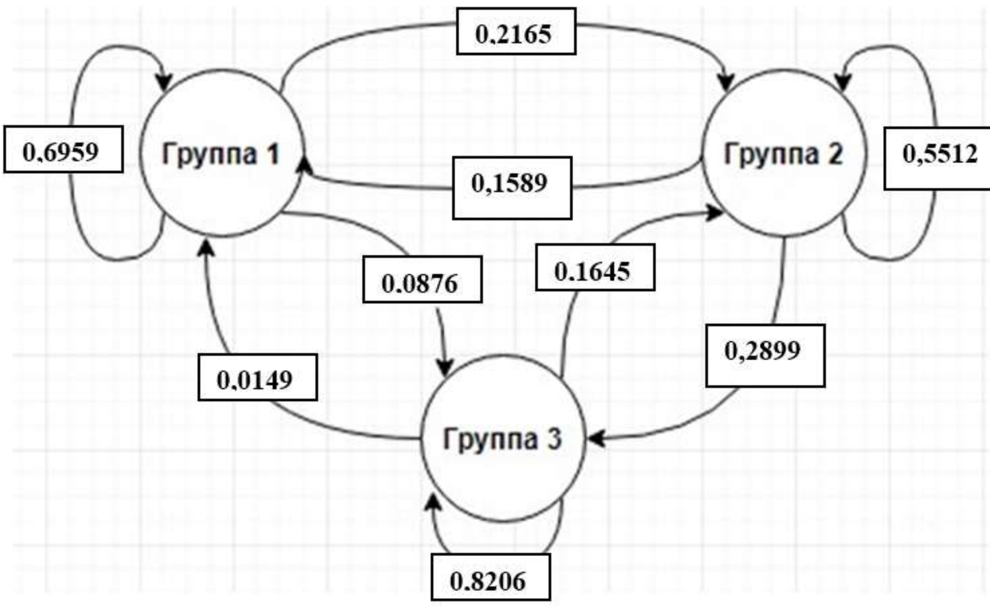



**Рис. 2.4. Марковская цепь, для описания переходов между группами успешности после второго семестра.**

На рисунке видно, что студенты первой группы с вероятностью в 69,6 % в ней и остались, это хороший показатель. А вот тот факт, что 82 % студентов с низкой успешностью остались в своей группе, говорит о том, что с 3-ей группой студентов не было проведено необходимой работы со стороны преподавателя. Более 24 % студентов от общей численности понизили свою группу успешности (31 % студентов первой группы и 29 % второй) и только 10 % ее повысили (16 % студентов второй группы и 18 % третьей).

# ЗАКЛЮЧЕНИЕ

Марковские случайные процессы названы по имени выдающегося русского математика А.А.Маркова (1856-1922), впервые начавшего изучение вероятностной связи случайных величин и создавшего теорию, которую можно назвать "динамикой вероятностей". В дальнейшем основы этой теории явились исходной базой общей теории случайных процессов, а также таких важных прикладных наук, как теория диффузионных процессов, теория надежности, теория массового обслуживания и т.д. В настоящее время теория марковских процессов и ее приложения широко применяются в самых различных областях.

В работе показана возможность применения аппарата Марковских цепей для прогнозирования успеваемости студентов. На основании данных об успеваемости на начальных курсах строится прогноз об успеваемости на старших курсах. Расчетные данные сравниваются с экспериментальными. В работе представлена методология оценки и прогнозирования качества образования.

Основными методами исследования являются: теоретический анализ и обобщение научно-исследовательских работ, теоретические и практические методы педагогического исследования, методы статистической обработки эмпирических данных, методы машинного обучения и методы моделирования случайных событий.

Проведенное исследование выявило, что прогнозирование должно осуществляться на основе критериев, определяющих успешность обучения, метрики для которых можно получить на основе данных учебной аналитики. Классификацию студентов на группы успешности по выбранным критериям необходимо проводить для каждого контрольно-измерительного инструмента непосредственно после его выполнения студентами, чтобы своевременно выявить обучающихся, нуждающихся в особом внимании со стороны преподавателя. Для прогнозирования успешности обучения других потоков студентов целесообразно накапливать информацию о динамике переходов обучающихся между группами успешности, используя дискретные цепи Маркова.

Прогнозирование успешности студентов на основе данных учебной аналитики позволяет выделить обучающихся «группы риска», предсказывать распределение студентов по группам успешности и при необходимости корректировать учебно-методические материалы.

Проведенное исследование выявило следующие особенности прогнозирования:

–прогнозирование должно осуществляться на основе критериев, определяющих успешность обучения, таких как своевременность, результативность, непрерывность, самостоятельность; для всех факторов существуют метрики, построенные на основе данных учебной аналитики;

–классификацию студентов на группы успешности следует проводить для каждого контрольно-измерительного инструмента непосредственно после его выполнения студентами, чтобы своевременно выявить обучающихся «группы риска», нуждающихся в особом внимании со стороны преподавателя;

–необходимо накапливать информацию о динамике переходов обучающихся между группами успешности, используя дискретные цепи Маркова, что позволит прогнозировать успешность обучения других потоков студентов и в следующем учебном году.

Таким образом, в работе было изучено практическое применение цепей Маркова для прогнозирования в разных областях и построены примеры применения цепей Маркова для прогноза результатов обучения студентов, цель исследования достигнута и все поставленные во введении задачи выполнены.

## СПИСОК ЛИТЕРАТУРЫ

1. Андронов, А. М.Теория вероятностей и математическая статистика: учебник для вузов / А. М. Андронов, Е. А. Копытов, Л. Я. Гринглаз. - СПб. : Питер, 2004. - 461 с. - (Учебник для вузов).

2. Апатова Н.В., Гапонов А.И., Майорова А.Н. Прогнозирование успеваемости студентов на основе нечеткой логики. // Современные наукоемкие технологии, № 4, 2017, С. 7-11.

3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.

4. Будаева А.А. Об одном подходе к прогнозированию успеваемости студентов// В сборнике: ГГ-технологии: Теория и Практика. Материалы семинара. 2018. С. 28-33.

5. Будаева А.А. Прогнозирование успеваемости студентов вуза методами таксономии. // Вестник Института цивилизации. 2007. № 7. С. 124-127.

6. Быстрова Т. Ю., Ларионова В. А., Синицын Е. В., Толмачев А. В. Учебная аналитика МООК как инструмент прогнозирования успешности обучающихся // Вопросы образования. - 2018. - № 4. - С. 139-166.

7. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL. Учебное пособие. 2-е изд. — М.: Форум, 2008. — 464 с.

8. Гмурман, В. Е.Теория вероятностей и математическая статистика: учеб.пособие для студентов вузов / В. Е. Гмурман. - М. :Высш.шк., 2000. - 479с.

9. Гнеденко Б.В. Курс теории вероятностей: Учебник – Изд. бе, перераб. и доп. – М.: Наука. Гл. ред. физ.-мат. лит., 1988. — 448 с.

10. Гоголева Н.Г., Тарасова О.Ю. Прогнозирование успеваемости студентов вуза на основе цепей Маркова (на примере СПБ ГЭТУ «ЛЭТИ»). //Научные труды SWorld, 2016. Т. 5. № 44. С. 104-109.

11. Гончарова Ю.А., Ооржак О.М. Статистический анализ и проверка гипотез в успеваемости студентов ФМФ 2005-2008 годов поступления методами пакета STATISTICA. Выпускная квалификационная работа студенток 5 курса ФМФ (научный руководитель А.И.Жданок), Кызыл, ТывГУ, каф.мат. анализа и МПМ, 2008.

12. Гроппен В.О., Будаева А.А. Эталоны как уникальный инструмент постановки и решения задач теории принятия решений. // В сборнике: ГГ-технологии: Теория и Практика. Материалы семинара. 2017. С. 3-20.

13. Жданок А. И. Электронный учебник «Конечные цепи Маркова» на платформе кафедры информатики ТувГУ в системе дистанционного обучения ТувГУ (ЦДО ТувГУ).

14. Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. -М.: МЦНМО, 2017. Т. 2. -560 с.

15. Королюк В.С. и др. Справочник по теории вероятностей и математической статистике/В.С. Королюк, Н.И. Портенко, А.В. Скороход, А.Ф. Турбин. — М.: Наука. Главная редакция физико-математической литературы, 1985. — 640 с.

16. Медик В. А.. Токмачев М. С. Математическая статистика в медицине и биологии - Нов-город. 1998.-243 с.

17. Медик В. А.. Токмачев М. С. Фишман Б. Б. Статистика в медицине и биологии: Руковод-ство. В 2-х томах / Под ред. Ю. М. Комарова. Т. 1. Теоретическая статистика.- М.: Ме-дицина. 2000.-412 с.

18. Моисеев В.Б., Зубков А.Ф., Деркаченко В.Н. Прогнозирование успеваемости студентов по общепрофессиональным и специальным дисциплинам на основе регрессионных моделей. // Научно-технические ведомости СПбГПУ № 6, 2010, С. 169-173.

19. Панова Н.Ф., Денисова Н.В. Классификация студентов по уровню успеваемости с помощью аппарата дискриминантного анализа. // Вестник Оренбургского государственного университета. 2014. № 8 (169). С. 33-36.

20. Патаракин Е. Д. Совместная сетевая деятельность и поддерживающая ее учебная аналитика // Высшее образование в России. - 2015. - № 5. - С. 145-154.

21. Прошкина Е.Н., Балышова И.Ю. Анализ и прогнозирование успеваемости студентов на основе радиальной базисной нейронной сети. //Технические науки: традиции и инновации: материалы III Междунар.науч.конф. (г.Самара, март 2018 г.). - Казань: Молодой ученый, 2018. - К, С. 24-28.

22. Свешников А.А. и др. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. – М.: Наука, 1970. – 632с.

23. Хурума А.К. Дистанционный курс «Теория вероятностей и математическая статистика» на платформе Moodle в системе дистанционного обучения ТувГУ (ЦДО ТувГУ).

24. Шевченко В.А. Прогнозирование успеваемости студентов на основе методов кластерного анализа. //Вестник ХНАДУ, вып. 68, 2015, С. 15-18.

25. Шестакова Л. Г., Рихтер Т. В. Показатели оценки и самооценки готовности студентов к самоорганизации // ScienceforEducationToday. - 2019. - Т. 9, № 3. - С. 138-150.

26. Ясинский И.Ф., Семенова М.Б. Опыт прогнозирования успеваемости студентов при помощи нейросетевой технологии. // Вестник ИГЭУ, вып. 5, 2007, С.29-31.

27. Kemeny J. G., Snell J. L., Finite Markov chains. — The University Series in Undergraduate Mathematics. — Princeton: Van Nostrand, 1960 (перевод: КемениДж. Дж.,СнеллДж. Л. КонечныецепиМаркова. — М.: Наука. 1970. — 272 с.)