

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГУ»)

ИНСТИТУТ ЛИНГВИСТИКИ
Учебно-научный центр компьютерной лингвистики

Основы теории вероятностей

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.04.03 Фундаментальная и прикладная лингвистика

Код и наименование направления подготовки/специальности

Фундаментальная и компьютерная лингвистика

Наименование направленности (профиля)/ специализации

Уровень высшего образования: магистратура

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2023

Основы теории вероятностей

Рабочая программа дисциплины

Составитель(и):

Старший преподаватель И.П.Рыгаев

Ответственный редактор:

К.ф.н, доцент Н.А.Коротаев

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной лингвистики

№ 6 от 12 апреля 2023 г.

ОГЛАВЛЕНИЕ

1.	Пояснительная записка	
1.1.	Цель и задачи дисциплины	
1.2.	Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	
1.3.	Место дисциплины в структуре образовательной программы	
2.	Структура дисциплины	
3.	Содержание дисциплины	
4.	Образовательные технологии	
5.	Оценка планируемых результатов обучения	
5.1	Система оценивания	
5.2	Критерии выставления оценки по дисциплине	
5.3	Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	
6.	Учебно-методическое и информационное обеспечение дисциплины	
6.1	Список источников и литературы	
6.2	Перечень ресурсов информационно-телекоммуникационной сети «Интернет».	
6.3	Профессиональные базы данных и информационно-справочные системы	
7.	Материально-техническое обеспечение дисциплины	
8.	Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	
9.	Методические материалы	
9.1	Планы семинарских/ практических/ лабораторных занятий	
9.2	Методические рекомендации по подготовке письменных работ	
9.3	Иные материалы	

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Предметом дисциплины является ознакомление слушателей с современными статистическими моделями и основами теории вероятностей, необходимыми для дальнейшего изучения и использования методов математической статистики. Курс предваряет более практическую дисциплину «Анализ лингвистических данных».

Курс направлен на решение следующих задач:

- освоение магистрантами базовых знаний по теории вероятностей;
- изучение классических статистических моделей;
- приобретение навыков по применению теории вероятностей и статистических моделей для решения практических задач в области компьютерной лингвистики;
- выработка у магистрантов умения пользоваться средствами программы Microsoft Excel для решения задач по теории вероятностей и комбинаторике.

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.1 Знает принципы сбора, отбора и обобщения информации	<p>Знать:</p> <ul style="list-style-type: none"> – принципы сбора, отбора и обобщения информации; <p>Уметь:</p> <ul style="list-style-type: none"> – учитывать приоритетность требований к проекту, подготавливать функциональные, конструктивные и технологические обоснования; – разрабатывать, обосновывать, согласовывать и реализовывать разделы проекта; <p>Владеть:</p> <ul style="list-style-type: none"> – навыками разработки, управления и оценки эффективности реализации проекта на всех этапах жизненного цикла.
ПК-2 Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий;	ПК-2.1 Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые	<p>Знать:</p> <ul style="list-style-type: none"> – основные принципы обработки информации; – базовые принципы корпусной лингвистики, лексикографии, математической статистики; – базовые представления о

фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных	языковом разнообразии; – наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных; Уметь: – пользоваться лингвистическими корпусами, электронными словарями и базами данных; – применять основные принципы корпусной лингвистики, лексикографии, математической статистики; Владеть: – принципами создания лингвистических корпусов, электронных словарей и баз данных.
---	--	--

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Основы теории вероятностей» является элективной дисциплиной и относится к части, формируемой участниками образовательных отношений.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: Основания математики. В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: Основы математической статистики.

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических часов.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
2	Лекции	4
2	Практические занятия	26
Всего:		30

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 78 академических часа.

3. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание
1.	Введение. Применение теории вероятности и статистики в теоретической и прикладной лингвистике	<ul style="list-style-type: none"> • Лингвистика как эмпирическая наука; • Поиск закономерностей в лингвистических данных; • Вероятностная интерпретация задач компьютерной лингвистики.
	Базовые понятия теории вероятностей.	<ul style="list-style-type: none"> • Вероятность, свойства вероятности; • Пространство событий. Независимые и зависимые события; • Совместная вероятность. Условная вероятность, формула Байеса и формула полной вероятности.
3.	Напоминание: основные понятия математического анализа, операции дифференцирования и интегрирования.	<ul style="list-style-type: none"> • Функции. Предел функции в точке, производная функции в точке, дифференцируемые функции. Экстремум функции; • Первообразная, неопределённый интеграл. Определённый интеграл, формула Ньютона-Лейбница.
4.	Дискретные случайные величины. Математическое ожидание, дисперсия	<ul style="list-style-type: none"> • Понятие случайной величины; • Основные дискретные распределения: бернуллиево, биномиальное, геометрическое; • Математическое ожидание, дисперсия, ковариация, их свойства. Пуассоновское распределение.
5.	Непрерывные случайные величины. Нормальное распределение. Закон больших чисел и центральная предельная теорема	<ul style="list-style-type: none"> • Функция распределения и плотность вероятности; • Математическое ожидание для непрерывной случайной величины; • Основные непрерывные распределения: равномерное, показательное, бета и гамма-распределения, нормальное распределение; • Неравенства Чебышева и Маркова. Закон больших чисел. Центральная предельная теорема.
6.	Основные понятия математической статистики. Оценки параметров случайных величин	<ul style="list-style-type: none"> • Оценки параметров случайных величин; • Эмпирическое среднее и эмпирическая дисперсия; • Несмещённые и состоятельные оценки.
7.	Метод максимального правдоподобия и метод моментов.	<ul style="list-style-type: none"> • Оценки параметров распределения; • Метод максимального правдоподобия и метод моментов; • Априорное распределение на множестве параметров; • Оценки максимального правдоподобия для основных распределений.
8.	Критерии значимости для оценок случайных величин. Ошибки первого и второго рода	<ul style="list-style-type: none"> • Проверка статистических гипотез; • Нулевая гипотеза, ошибки первого и второго рода;

		<ul style="list-style-type: none"> Уровень значимости критерия, критерий Стьюдента.
9.	Таблицы сопряженности. Применение критериев значимости к таблицам сопряженности	<ul style="list-style-type: none"> Таблицы сопряженности. Критерий хи-квадрат для таблиц сопряженности, его применимость
10.	Применение статистических критериев к реальным лингвистическим данным.	<ul style="list-style-type: none"> Исследовательский проект

4. Образовательные технологии

Для проведения учебных занятий по дисциплине используются различные образовательные технологии. Для организации учебного процесса может быть использовано электронное обучение и (или) дистанционные образовательные технологии.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- домашние задания	5 баллов	30 баллов
- выполнение заданий на семинаре	5 баллов	10 баллов
- участие в соревновании	20 баллов	20 баллов
Промежуточная аттестация – зачет		40 баллов
Итого за семестр		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	отлично/ зачтено	Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	хорошо/ зачтено	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	удовлетво- рительно/ зачтено	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	неудовлет- ворительно/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

В качестве домашних заданий предлагаются задания следующих типов

- Д31. Базовые понятия теории вероятностей.
- Д32. Вычисление производных и интегралов.
- Д33. Характеристики дискретных случайных величин.
- Д34. Характеристики непрерывных случайных величин.
- Д35. Оценки и их значимость.
- Д36: Оценки значимости для таблиц сопряжённости

Исследовательский проект (часть 3): Применение статистических критериев к реальным лингвистическим данным

Зачет ориентирован на следующие контрольные вопросы

Вычисление вероятности события при условии другого события, формулы Байеса и полной вероятности.

Нахождение математического ожидания, дисперсии и других характеристик для заданных случайных величин и распределений.

Проверка несмещённости оценки для параметра распределения.

Проверка статистической гипотезы на основе выборки с помощью данного критерия.

Анализ таблицы сопряжённости с помощью критерия хи-квадрат и других критериев.

Нахождение оптимальных параметров распределения с помощью метода максимального правдоподобия.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Основная литература

1. *А.В. Антонов, М.С. Никулин, А.М. Никулин, В.А. Чепурко.* Теория надежности. Статистические модели. Учеб.пособие — М.: ИНФРА-М, 2018. — 576 с.+ Доп. Материалы
2. *Белько И.В.* Теория вероятностей, математическая статистика, математическое программирование: Учебное пособие. - 1. - Москва; Минск: ООО "Научно издательский центр ИНФРА-М»: ООО "Новое знание", 2016. - 299 с.
3. *Глинский В.В.* Статистика: Учебник. - 4; перераб. и доп. - Москва: ООО "Научно-издательский центр ИНФРА-М", 2017. - 355 с.
4. *Кочетков Е.С.* Теория вероятностей и математическая статистика: Учебник. - 2; испр. и перераб. - Москва; Москва: Издательство "ФОРУМ»: ООО "Научно-издательский центр ИНФРА-М", 2018. - 240 с.
5. *М.Б. Лагутин.* Наглядная математическая статистика. М.: БИНОМ, Лаборатория знаний, 2007.

Рекомендованная литература

1. *А. А. Боровков* Математическая статистика. М.: Наука, 1984.
2. *Ю. А. Розанов.* Теория вероятностей, случайные процессы и математическая статистика.
3. *Г. В. Ивченко, В. И. Медведев.* Математическая статистика. М.: Высшая школа, 1984.
4. *М.Б. Лагутин.* Наглядная математическая статистика. М.: БИНОМ, Лаборатория знаний, 2007.
5. *В. Н. Тутубалин.* Теория вероятностей и случайных процессов. М.: Издательство МГУ, 1992.

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

№ п/п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках

	национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

7. Материально-техническое обеспечение дисциплины

№п /п	Наименование ПО	Производитель	Способ распространения
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.

- для глухих и слабослышащих: в печатной форме, в форме электронного документа.

- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA SE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;

- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;

- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

9.1 Планы семинарских/ практических/ лабораторных занятий

1. Введение. Применение теории вероятности и статистики в теоретической и прикладной лингвистике • Лингвистика как эмпирическая наука;

- Поиск закономерностей в лингвистических данных;

- Вероятностная интерпретация задач компьютерной лингвистики.

2. Базовые понятия теории вероятностей. • Вероятность, свойства вероятности;

- Пространство событий. Независимые и зависимые события;

- Совместная вероятность. Условная вероятность, формула Байеса и формула полной вероятности.

3. Напоминание: основные понятия математического анализа, операции дифференцирования и интегрирования.
 - Функции. Предел функции в точке, производная функции в точке, дифференцируемые функции. Экстремум функции;
 - Первообразная, неопределённый интеграл. Определённый интеграл, формула Ньютона-Лейбница.
4. Дискретные случайные величины. Математическое ожидание, дисперсия
 - Понятие случайной величины;
 - Основные дискретные распределения: бернуллиево, биномиальное, геометрическое;
 - Математическое ожидание, дисперсия, ковариация, их свойства. Пуассоновское распределение.
5. Непрерывные случайные величины. Нормальное распределение. Закон больших чисел и центральная предельная теорема
 - Функция распределения и плотность вероятности;
 - Математическое ожидание для непрерывной случайной величины;
 - Основные непрерывные распределения: равномерное, показательное, бета и гамма-распределения, нормальное распределение;
 - Неравенства Чебышева и Маркова. Закон больших чисел. Центральная предельная теорема.
6. Основные понятия математической статистики. Оценки параметров случайных величин
 - Оценки параметров случайных величин;
 - Эмпирическое среднее и эмпирическая дисперсия;
 - Несмещённые и состоятельные оценки.
7. Метод максимального правдоподобия и метод моментов.
 - Оценки параметров распределения;
 - Метод максимального правдоподобия и метод моментов;
 - Априорное распределение на множестве параметров;
 - Оценки максимального правдоподобия для основных распределений.
8. Критерии значимости для оценок случайных величин. Ошибки первого и второго рода
 - Проверка статистических гипотез;
 - Нулевая гипотеза, ошибки первого и второго рода;
 - Уровень значимости критерия, критерий Стьюдента.
9. Таблицы сопряжённости. Применение критериев значимости к таблицам сопряжённости
 - Таблицы сопряжённости. Критерий хи-квадрат для таблиц сопряжённости, его применимость
10. Применение статистических критериев к реальным лингвистическим данным.
 - Исследовательский проект

9.2 Иные материалы

Все необходимые для обучения материалы даются на лекциях и практических занятиях.