Министерство образования и науки РФ
Федеральное государственное автономное образовательное учреждение высшего образования
«Новосибирский национальный исследовательский государственный университет» (Новосибирский государственный университет, НГУ)
Гуманитарный факультет

Программа рассмотрена
на заседании кафедры
фундаментальной и прикладной
лингвистики
29.08.2014
_____
Зав. кафедрой, проф. М.К. Тимофеева

Утверждаю

_____
декан гуманитарного
факультета, профессор
1.09.2014
Л.Г. Панин

Основная образовательная программа
высшего образования

Направление подготовки
035800 – Фундаментальная и прикладная лингвистика

Квалификация (степень) выпускника –
бакалавр

ПРОГРАММА УЧЕБНОГО КУРСА «ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

(128 часов, 4 з.е.)

1. **Наименование дисциплины**

ПРОГРАММА УЧЕБНОГО КУРСА
«ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Программа дисциплины «Теория вероятностей и математическая статистика» составлена в соответствии с требованиями к обязательному минимуму содержания и уровню подготовки дипломированного бакалавра по направлению 035800 «Фундаментальная и прикладная лингвистика» в целях обеспечения реализации учебного процесса в НГУ.

Автор Савельев Лев Яковлевич, к. ф. - м. н., профессор

Цели освоения дисциплины

Дисциплина «Теория вероятностей и математическая статистика» является теоретической базой всех статистических дисциплин. Она необходима для изучения ряда дисциплин учебного плана. Курс позволяет овладеть основными вероятностными и статистическими методами, применяемыми в современных лингвистических теориях, моделях и прикладных лингвистических ресурсах.

Теория вероятностей и математическая статистика имеют важное методологическое значение в познавательном процессе. Содержание курса “Вероятность и статистика” является важной составляющей теоретико-методологической базы ряда последующих основных и специальных курсов.

Курс соответствует двум приоритетным направлениям Программы развития НГУ: математика, гуманитарные науки.

2. В результате освоения дисциплины обучающийся должен:

- **Знать:** основные понятия теории вероятностей и математической статистики; основные распределения дискретных и непрерывных случайных величин; методы вычисления вероятностей случайных событий, значений функций распределения, средних значений и других числовых характеристик случайных величин; методы статистических оценок параметров и статистического выбора гипотез; по окончании курса студенты должны обладать набором знаний по сбору, обработке и анализу статистических данных.
- **Уметь:** составлять и решать различные прикладные вероятностные и статистические задачи, связанные с лингвистикой, используя изученные теоретические и эмпирические распределения; проводить грамотный статистический анализ имеющегося материала; по окончании курса студенты должны обладать набором навыков по сбору, обработке и анализу статистических данных, получению на их основе содержательных выводов..
- **Владеть:** методами вычисления вероятностей случайных событий , значений функций распределения, средних значений и других числовых характеристик случайных величин; методами статистических оценок параметров и статистического

выбора гипотез; по окончании курса студенты должны владеть методами решения прикладных и теоретических задач, а также выполнять компьютерные вычисления с реальными данными; иметь практические навыки статистического анализа лингвистических материалов.

Перечисленные результаты образования являются основой для формирования следующих общекультурных и общепрофессиональных компетенций:

***а) общекультурными* (*ОК*)**

- владением культурой мышления, способностью к обобщению, анализу, восприятию информации, постановке цели и выбору путей её достижения — ОК-1
- умением логически верно, аргументировано и ясно строить устную и письменную речь — ОК-2
- стремлением к саморазвитию, повышению своей квалификации и мастерства — ОК-6
- способностью применять методы математического анализа и моделирования в профессиональной деятельности — ОК-10

***б) профессиональными (ПК)*:**

общепрофессиональными:

- знанием основных понятий и категорий современной лингвистики — ПК-1
- знанием основ математических дисциплин, которые используются при формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур: теории множеств, математического анализа, теории вероятностей и математической статистики, теории информации и кодирования, математической логики, математической теории грамматик — ПК-2

дополнительными:

- способностью работать в междисциплинарной команде — ПК-26
- способностью общаться с экспертами в других областях знаний — ПК-27
- умением видеть междисциплинарные связи изучаемых дисциплин и пониманием их значения для будущей профессиональной деятельности — ПК-28

4. Место дисциплины в структуре образовательной программы

Важное место среди разнообразного инструментария, используемого в гуманитарном образовании и лингвистических исследованиях, занимают методы и модели специальных выборочных разработок, анализ текстов и стилей, теория лингвистических переменных, статистический анализ информации, вероятностные модели языков. Понимание и осмысленное использование этого инструментария в решении исследовательских и практических задач невозможны без определенного запаса знаний и навыков в области теории вероятностей и математической статистики. Именно на овладение необходимым минимумом этих знаний и навыков и нацелен курс “Теория вероятностей и математическая статистика”. Построение методов и организационных форм обучения учитывает основную направленность курса: он обращен к пользователю (а не к разработчику) описываемых в курсе методов и моделей. Поэтому курс “Теория

вероятностей и математическая статистика" нацелен в первую очередь на разъяснение их прикладных возможностей и на изложение рекомендаций по их использованию.

Для усвоения материала курса теории вероятностей и математической статистики необходимы достаточно полные курсы

1. Алгебры
2. Математического анализа

В свою очередь материал курса теории вероятностей и математической статистики используется при изложении целого ряда дисциплин, связанных с математической лингвистикой:

1. Прикладной лингвистики
2. Структурной лингвистики
3. Криптологии
4. Стилеметрии
5. Корпусной лингвистики
6. Компьютерной лингвистики
7. Нейролингвистики и психолингвистики (методы проведения лингвистических экспериментов)

Естественно, наиболее тесные связи с теорией вероятностей и математической статистикой имеет статистическая лингвистика**,** дисциплина, изучающая количественные закономерности естественного языка, проявляющиеся в текстах. В основе статистической лингвистики лежит предположение, что некоторые численные характеристики и функциональные зависимости между ними, полученные для ограниченной совокупности текстов, характеризуют язык в целом или его функциональные стили (публицистический, научный, художественный и т.п.). Практически важной и наиболее изученной числовой характеристикой является относительная частота употребления различных лингвистических единиц (букв, фонем, слогов, слов, синтаксических конструкций), их классов (например, гласных, согласных, частей речи) и сочетаний (например, последовательностей из n букв). Данные о частоте слов (иногда словосочетаний) отражаются в частотных словарях. Важную роль в статистической лингвистика играет функциональная зависимость, приближенно описывающая связь между частотой слова и его номером (рангом) в последовательности по убыванию частот. Статистическая лингвистика изучает также зависимости между частотой и длиной слова (в числе слогов), числом его значений и возрастом. Накопленные данные используются для выявления особенностей стиля отдельных авторов, атрибуции текстов, дешифровки исторических письменностей, для решения задач стенографии, теории связи, а также информатики. Статистическая лингвистика при получении численных характеристик использует методы математической статистики и некоторые методы теории информации для определения энтропии и избыточности языка, а для установления связи между наблюдаемыми характеристиками и выбора наиболее существенных из них — метод математических моделей, базирующихся на понятиях теории вероятностей и математической лингвистики. Возможно, и более широкое понимание статистической лингвистики как науки, эффективно применяющей методы теории вероятностей и математической статистики для проверки лингвистических гипотез, которые могут носить и качественный характер.

4. Объем дисциплины в зачетных единицах с указанием количества академических, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся.

Общая трудоемкость дисциплины составляет 4 зачетных единицы, 128 часа. Из них на контактную работу с преподавателем 64 часа (32 часа – лекции, 32 часа – практические занятия), на самостоятельную работу студентов – 64 часа. Занятий в интерактивной форме – 16 часов.

5. Содержание дисциплины "Теория вероятностей и математическая статистика", структурированное по темам с указанием отведенного на них количества астрономических часов и видов учебных занятий

№ п/п	**Раздел дисциплины**	**Семестр**	**Неделя семестра**	**Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)**				***Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)***
1	Теория вероятностей и математическая статистика	2	16	Семестровые задания 32	Самостоятельные письменные работы 8	Консультации 16		*Коллоквиумы с выставлением оценок, 7-я, 14-я недели*

Программа курса

Лекции	Раздел, тема, содержание занятий	Количество часов			Литература
		лекций	упражнений	Самост. занятий	(пункты учебников)
1	2	3	4	5	6
	Р а з д е л 1 Теория вероятностей				
1	*Тема 1.1.* — Введение в курс: теории вероятностей и математической статистики, связь с математической лингвистикой, роль в лингвистических исследованиях Правила действий со случайными событиями и вероятностями	2	2	2	Пиотровский, Введение Савельев, ЭТВ 1-2 Коваленко, гл 1
2	*Тема 1.2.* — Случайные величины, их распределения и основные числовые характеристики **.** — Важнейшие распределения. Наиболее распространенные в математической лингвистике распределения	2	2	2	Савельев, ЭТВ 1 Коваленко, гл 2 Коваленко, гл 5 Пиотровский, гл 6
3	*Тема 1.3.* — Неравенство Чебышева, законы больших чисел, центральная предельная теорема.	2	2	2	Савельев, ЭТВ 1 Коваленко, гл 7

					Пиотровский, гл 6
4	*Тема 1.4.* — Информация и энтропия. Информационные измерения в текстах. Вычисление энтропии текста.	2	2	2	Савельев, ЭТВ 1 Пиотровский, гл 2, 5 Шеннон, с 669-686
5	*Тема 1.5.* — Стохастическая зависимость, коэффициенты корреляции, регрессии и информации. Вероятностное моделирование текста и составляющих его единиц.	2	2	2	Савельев, ЭТВ 1 Коваленко, гл 9 Пиотровский, гл 6 Шеннон, с 243-332
6	*Тема 1.6.* — Цепи Маркова и их реализации. Элементы теории случайных процессов. Марковские модели в лингвистике.	2	2	2	Савельев, ЭТВ 1 Коваленко, гл 8 Кемени, гл 2 Шеннон, с 243-332
7	*Тема 1.7.* — Элементы криптографии	2	2	2	Ященко, гл 1-7 Приложение
8	*Тема 1.8.* —Информационные технологии в стохастических исследованиях	2	2	2	Воробьев, гл 2 **.**
		16	16	16	
	Р а з д е л 2 . Математическая статистика				
9	*Тема 2.1.* — Основы статистического описания: генеральная совокупность, выборка, основные выборочные характеристики и анализ их поведения. Комбинаторика лингвистических единиц. Вероятность и информация лингвистических событий.	2	2	2	Пиотровский, гл 5
10	*Тема 2.2.* —Вариационный ряд и порядковые статистики. Первичная статистическая обработка текста.	2	2	2	Пиотровский, гл 7
11	*Тема 2.3.* —Методы статистической оценки параметров. Построение интервальных оценок. Статистическая модель текста и вероятностные характеристики нормы языка.	2	2	2	Коваленко, гл 10 Пиотровский, гл 8
12	*Тема 2.4.* — Статистическая проверка гипотез: основные типы статистических критериев, их общая логическая схема, критерии согласия и однородности.	2	2	2	Коваленко, гл 11 Пиотровский, гл 9
13	*Тема 2.5.* —Теории статистических игр. Выбор из нескольких гипотез. Исследование свойств языка.	2	2	2	Пиотровский, гл 9
14	*Тема 2.6.* — Метод наименьших квадратов и обработка наблюдений. Элементы регрессионного анализа. Модель	2	2	2	Линник, введение Дрейпер,

	дисперсионного анализа.				гл 1-2, 9
15	*Тема 2.7.* —Методы классификации и распознавания. Дискримииантный анализ, кластерный анализ, факторный анализ	2	2	2	Айвазян, гл 1-2, 11,14
16	*Тема 2.8.* —Информационные технологии в статистических исследованиях	2	2	2	Воробьев, гл 2
		16	16	16	

Оставшиеся часы: дополнительные занятия (консультации, сдача задолженностей)

6. Рекомендации для самостоятельной работы обучающихся по дисциплине

Темы для самостоятельной подготовки

Раздел 1. Теория вероятностей

Тема 1.1. Теория вероятности - раздел математики, изучающий случайные события. Она находит зависимости между их появлениями, вычисляя вероятности их появлений. Математическая статистика - раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов. Прикладная и математическая лингвистика существенно использует методы теории вероятностей и математической статистики.

Дискретные, непрерывные и общие вероятностные пространства. Правила сложения и умножения вероятностей. Зависимость и условные вероятности. Формула полной вероятности. Формула Байеса. Примеры применения в прикладной и математической лингвистике.

Тема 1.2. Случайные величины и переменные. Индикаторы случайных событий. Распределение и функции распределения. Свойства функции распределения. Плотность распределения. Характеристические функции и их свойства. Стохастическая независимость и зависимость. Среднее значение и его свойства. Дисперсия и стандартное отклонение. Последовательности независимых и зависимых случайных величин.

Важнейшие распределения:

- Биномиальное распределение
- Полиномиальное распределение
- Распределение Пуассона
- Экспоненциальные распределения
- Нормальное распределение
- Хи-квадрат распределение
- Распределение Стьюдента
- Распределение Коши

Тема 1.3. Практика изучения случайных явлений показывает, что хотя результаты отдельных наблюдений, даже проведенных в одинаковых условиях, могут сильно отличаться, в то же время средние результаты для достаточно большого числа наблюдений устойчивы и слабо зависят от результатов отдельных наблюдений. Теоретическим обоснованием этого замечательного свойства случайных явлений является закон больших чисел. Этим названием объединена группа теорем, устанавливающих устойчивость средних результатов большого количества случайных явлений и объясняющих причину этой устойчивости. Неравенство Чебышева и его следствия. Закон

больших чисел в форме Бернулли. Закон больших чисел в форме Чебышева. Центральная предельная теорема объясняет широкое распространение нормального закона распределения. Теорема утверждает, что всегда, когда случайная величина образуется в результате сложения большого числа независимых случайных величин с конечными дисперсиями, закон распределения этой случайной величины оказывается практически нормальным законом.

Тема 1.4. Информация и энтропия являются универсальными естественно научными понятиями. Энтропия используется как мера неопределенности. Например, в последовательности букв, составляющих какое-либо предложение на русском языке, разные буквы появляются с разной частотой, поэтому неопределённость появления для некоторых букв меньше, чем для других. Если же учесть, что некоторые сочетания букв встречаются очень редко, то неопределённость ещё более уменьшается. Впервые понятия энтропия и информация связал К.Шеннон. В теории вероятностей информация и энтропия точно определяются. В дискретном случае определения формулируются достаточно просто. В математической лингвистике широко используются информационные модели текстов. Предсказание и энтропия английского текста. Комбинаторика лингвистических единиц. Вероятность и информация лингвистических событий

Тема 1.5 или парный р в теории вероятностей и статистике — это показатель характера взаимного стохастического влияния изменения двух случайных величин. Случайные величины могут быть связаны даже функциональной зависимостью (каждому значению одной случайной величины соответствует единственное значение другой случайной величины), но коэффициент их корреляции будет равен нулю. Для метрических величин применяется коэффициент корреляции Пирсона, точная формула которого была введена Фрэнсисом Гальтоном: Коэффициенты ранговой корреляции Кенделла и Спирмена применяются для выявления взаимосвязи между количественными или качественными показателями, если их можно ранжировать. В коэффициенте корреляции знаков Фехнера подсчитывается количество совпадений и несовпадений знаков отклонений значений показателей от их среднего значения. Используя индикаторы можно коэффициенты корреляции и регрессии для случайных величин использовать для измерения зависимости между случайными событиями.

Коэффициент информации и его свойства. Для измерения зависимости между случайными переменными кроме коэффициента корреляции, оценивающего только линейную зависимость, применяется коэффициент информации, оценивающий и нелинейную стохастическую зависимость.

Тема 1.6. В простейшем случае условное распределение последующего состояния цепи Маркова первого порядка зависит только от текущего состояния и не зависит от всех предыдущих состояний (в отличие от сложных цепей Маркова высших порядков). Рассматриваются простые однородные марковские цепи с конечным множеством состояний. Свойства матриц переходных вероятностей. Классификация состояний. Предельные вероятности. Распределения и характеристики различных типов серий в марковских последовательностях. Марковские модели прогноза. Примеры применения в прикладной и математической лингвистике.

Общее описание случайного процесса. Реализации случайного процесса. Конечномерные распределения случайного процесса. Гауссовские и пуассоновские случайные процессы, примеры их использования.

Пример Маркова с текстом из «Евгения Онегина». Марковское модели текстов Шеннона. Проблема моделирования словообразования. Словообразовательные гнезда и цепи. Распределения длин слов и предложений в различных языках.

Тема 1.7. Основные понятия криптографии. Примеры простых шифров: шифр Цезаря, шифр простой подстановки, шифр Вижинера, матричная система, шифр Плейфер. Криптография и теория сложности. Система шифрования RSA с открытым ключом. Компьютерная криптография. Теория связи в секретные системы. Алгебра секретных систем.

Тема 1.8. Информационные технологии в стохастических исследованиях. Примеры применения компьютерных программ в вычислениях характеристик стохастических распределений и случайных переменных. Компьютерное моделирование распределений.

Раздел 2. Математическая статистика

Тема 2.1. — Статистическое описание. Генеральная совокупность лингвистических объектов. Выборка, основные выборочные характеристики и анализ их поведения. Методы организации статистического наблюдения над текстом: случайный выбор, механический выбор, серийный выбор, типический выбор.

Тема 22. Вариационный ряд и порядковые статистики. Первичная статистическая обработка текста. Способы представления синтаксической структуры предложения в системах обработки данных. Статистическая совокупность лингвистических объектов и ее организация. Вариационные ряды лингвистических признаков. Статистические характеристики лингвистических вариационных рядов. Исследование лингвистических вариационных рядов с помощью эмпирических моментов.

Тема 2.3. Классификация оценок, несмещенность, эффективность и состоятельность. Методы статистической оценки параметров, метод максимального правдоподобия и метод минимума хи-квадрата. Построение интервальных оценок. Доверительные интервалы для среднего и дисперсии нормального распределения. Оценка коэффициентов корреляции и регрессии по выборке. Точечная оценка параметров генеральной лингвистической совокупности. Оценка среднего с помощью доверительных интервалов и статистическая параметризация стилей. Оценка функции генерального распределения по данным лингво-статистического наблюдения. Статистическая модель текста и вероятностные характеристики нормы языка.

Тема 2.4. Статистическая проверка гипотез: основные типы статистических критериев, их общая логическая схема, критерии согласия и однородности. Критерий хи-квадрат для простой гипотезы. Критерий хи-квадрат для оценки параметров по выборке. Критерий согласия Колмогорова. Наиболее мощные критерии. Гипотеза о лексической нормативности текста и ее проверка с помощью порядковых критериев. Проверка гипотез о расхождениях статистических характеристик языков, функциональных стилей и подъязыков с помощью параметрических критериев. Проверка статистических гипотез о тождестве двух лингвистических распределений. Распределение средних длин словоформ в языках мира

Тема 2.5. Теория статистических игр. Проблема статистического решения. Допустимые распределения. Решения и решающие правила. Средний риск. Байесовские и минимаксные стратегии. Выбор из нескольких гипотез. Доминантные смысловые единицы и элементы заполнения текста.

Тема 2.6. Элементы регрессионного анализа. Метод наименьших квадратов. Модель дисперсионного анализа. Регрессионный анализ — статистический метод исследования зависимости между зависимой переменной и одной или несколькими независимыми

переменными. Терминология отражает лишь стохастическую зависимость переменных, а не причинно-следственные отношения. Рассматривается простейшая линейная модель. Уравнения регрессии. Подбор прямой методом наименьших квадратов. Матричный подход к линейной регрессии. Исследование остатков. Случай двух предикторных переменных. Множественная регрессия и построение математической модели. Приложение множественной регрессии к задачам дисперсионного анализа.

Тема 2.7. Методы классификации и распознавания. Дискриминантный анализ, кластерный анализ, факторный анализ. Сущность задач классификации и снижения размерности пространства признаков. Основные идеи многомерного статистического анализа. Основные этапы в решении задач классификации и снижения размерности. Теоретические результаты классификации при известных распределениях и использовании обучающих выборок (дискриминантный анализ). Практические рекомендации. Применения дискриминантного анализа. Методы автоматической классификации (кластерный анализ). Метод главных компонент. Модели и методы факторного анализа. Многомерное шкалирование. Программное обеспечение для задач классификации и снижения размерности.

Тема 2.8. Информационные технологии в статистических исследованиях. Программная система «Математика». Символьные вычисления. Встроенная графика. Работа со списками. Вычисление выражений. Специализированный пакет «Статистика».

7. Фонд оценочных средств для промежуточной аттестации обучающихся по дисциплине.

Перечень тем для курсовых работ и рефератов по теории вероятностей

- Алгебра вероятностей событий. Стохастическая независимость и зависимость событий. Условные вероятности. Формула полной вероятности и формула Байеса. Примеры вероятностей лингвистических событий.
- Алгебра случайных величин. Распределения. Стохастическая независимость и зависимость случайных величин, ковариация. Среднее значение и дисперсия, их свойства. Условные средние. Формула полной среднего. Примеры лингвистических случайных переменных.
- Коэффициент корреляции и его свойства. Коэффициенты регрессии случайных величин и событий. Ранговые коэффициенты корреляции. Стохастическая близость событий. Примеры применения в математической лингвистике.
- Биномиальное и полиномиальное распределения. Примеры применения в математической лингвистике.
- Экспоненциальные распределения и распределение Пуассона. Примеры применения в математической лингвистике.
- Нормальное распределение и его свойства. Хи-квадрат распределение и распределение Стьюдента. Примеры применения в математической лингвистике.
- Центральная предельная теорема для последовательностей независимых одинаково распределенных случайных величин. Примеры применения в математической лингвистике.
- Неравенство Чебышева и его следствия. Закон больших чисел в форме Бернулли. Закон больших чисел в форме Чебышева. Примеры применения в математической лингвистике.

- Информация и энтропия. Их основные свойства. Коэффициент информации и его свойства. Вероятность и информация лингвистических событий. Предсказание и энтропия английского текста.
- Коэффициент информации и его свойства. Измерение стохастической зависимости случайных величин. Примеры применения в математической лингвистике.
- Простые однородные марковские цепи с конечным множеством состояний. Свойства матриц переходных вероятностей. Классификация состояний марковской цепи. Предельные вероятности. Марковские модели в математической лингвистике.
- Проблема моделирования словообразования. Пример, аналогичный примеру Маркова с текстом из «Евгения Онегина». Распределения длин слов и предложений в различных языках.
- Основные понятия криптографии. Примеры простых шифров: шифр Цезаря, шифр простой подстановки, шифр Вижинера, матричная система, шифр Плейфер.

Перечень тем для курсовых работ и рефератов по математической статистике

- Статистическое описание. Генеральная совокупность лингвистических объектов. Выборка, основные выборочные характеристики и анализ их поведения. Методы организации статистического наблюдения над текстом.
- Вариационный ряд и порядковые статистики. Первичная статистическая обработка текста. Вариационные ряды лингвистических признаков. Статистические характеристики лингвистических вариационных рядов.
- Классификация оценок, несмещенность, эффективность и состоятельность. Методы статистической оценки параметров, метод максимального правдоподобия и метод минимума хи-квадрата. Примеры применения в математической лингвистике.
- Построение интервальных оценок. Доверительные интервалы для среднего и дисперсии нормального распределения. Оценка среднего с помощью доверительных интервалов и статистическая параметризация стилей.
- Оценка коэффициентов корреляции и регрессии по выборке. Точечная оценка параметров генеральной лингвистической совокупности. Оценка функции генерального распределения по данным лингво-статистического наблюдения.
- Статистическая проверка гипотез: основные типы статистических критериев, их общая логическая схема, критерии согласия и однородности. Примеры применения в математической лингвистике.
- Критерий хи-квадрат для простой гипотезы. Критерий хи-квадрат для оценки параметров по выборке. Критерий согласия Колмогорова. Примеры применения в математической лингвистике.
- Гипотеза о лексической нормативности текста и ее проверка с помощью порядковых критериев. Проверка гипотез о расхождениях статистических характеристик языков, функциональных стилей и подъязыков с помощью параметрических критериев. Проверка статистических гипотез о тождестве двух лингвистических распределений.
- Проблема статистического решения. Байесовские и минимаксные стратегии. Выбор из нескольких гипотез. Примеры применения в математической лингвистике.
- Элементы регрессионного анализа. Метод наименьших квадратов. Модель дисперсионного анализа. Примеры применения в математической лингвистике.
- Методы классификации и распознавания. Дискриминантный анализ, кластерный анализ, факторный анализ. Примеры применения в математической лингвистике.
- Информационные технологии в статистических исследованиях. Программная система «Математика». Специализированный пакет «Статистика».

Список экзаменационных вопросов по теории вероятностей

- Алгебра вероятностей событий.
- Стохастическая независимость и зависимость событий.
- Условные вероятности.
- Формула полной вероятности и формула Байеса.
- Примеры вероятностей лингвистических событий.
- Алгебра случайных величин.
- Основные распределения.
- Стохастическая независимость и зависимость случайных величин.
- Среднее значение и дисперсия, их свойства.
- Условные средние.
- Формула полного среднего.
- Примеры лингвистических случайных переменных.
- Коэффициент корреляции и его свойства.
- Коэффициенты регрессии случайных величин и событий. Примеры применения в математической лингвистике.
- Биномиальное и полиномиальное распределения. Примеры применения в математической лингвистике.
- Распределение Пуассона. Примеры применения в математической лингвистике.
- Нормальное распределение и его свойства. Примеры применения в математической лингвистике.
- Центральная предельная теорема для последовательностей независимых одинаково распределенных случайных величин. Примеры применения в математической лингвистике.
- Закон больших чисел в форме Бернулли. Примеры применения в математической лингвистике.
- Информация и энтропия.
- Вероятность и информация лингвистических событий.

Перечень вопросов по математической статистике

- Что такое случайная переменная?
- Каковы принципы группировки данных?
- Что такое теоретическое распределение?
- Что такое эмпирическое распределение?
- Каковы возможные причины нескольких вершин вариационных кривых?
- Что такое вариационный размах и лимиты?
- Какие две группы показателей позволяют характеризовать вариационные ряды?
- Свойства среднего арифметического.
- Свойства среднего стохастического.
- Степени свободы критерия хи-квадрат.
- Что такое доверительные интервалы?
- Отличаются ли друг от друга по закономерностям случайной вариации выборочная и генеральная совокупности?
- В какой степени среднее арифметическое выборочной совокупности характеризует среднее арифметическое генеральной совокупности?
- Объясните, в чем заключается закон больших чисел.
- Кратко охарактеризуйте основные предпосылки выборочного метода.
- Объясните сущность нулевой гипотезы и дайте примеры.
- Проиллюстрируйте вычисление энтропии на примере.
- Что такое корреляция?

- Какая разница между корреляционной и функциональной зависимостью?
- Какая разница между положительной и отрицательной корреляциями?
- Что такое информация?
- Является ли наличие корреляции доказательством причинной зависимости между изучаемыми варьирующими признаками?

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.

а) основная литература:

1. **Колемаев В.А.** Теория вероятностей и математическая статистика : учебник для студентов высших учебных заведений, обучающихся по экономическим специальностям / В.А. Колемаев, В.Н. Калинина .— 3-е изд., перераб. и доп .— Москва : КНОРУС, 2013 .— 375,

2. **Учебно-**методическая литература на сайте НГУ (http://mmf.nsu.ru/education/materials#algebra):

Лотов В.И. Лекции по теории вероятностей: Учебное пособие / Новосиб. гос. ун-т. Новосибирск, 2011. 113 с.(link is external)
Коршунов Д.А., Фосс С.Г. Сборник задач и упражнений по теории вероятностей: Учебное пособие. - 2-е изд., испр. - Новосибирск: Новосиб. гос. ун-т, 2003, 119 с.
Коршунов Д.А., Чернова Н.И. Сборник задач и упражнений по математической статистике: Учебное пособие. - 2-е изд., испр. - Новосибирск: Изд-во Института математики, 2004. — 128 с.

б) дополнительная литература:

1. **Коваленко И.Н., Филиппова А.А.** Теория вероятностей и математическая статистика. Москва, Высшая школа, 1982.
2. **Савельев Л.Я.** Элементарная теория вероятностей, 1 - 2. Новосибирск, НГУ, 2005.
3. **Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А.** Математическая лингвистика. Москва, Высшая школа, 1977.
4. **Шеннон К.** Работы по теории информации и кибернетике. Москва, ИЛ, 1963.
5. **Кемени Дж., Снелл Дж.** Конечные цепи Маркова. Москва, Наука, 1970.
6. **Введение в криптографию. Под ред. В.В. Ященко.** Серия «Новые математические дисциплины». Москва, МЦНМО – ЧеРо, 2000.
7. **Айвазян С. А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.** Классификация и снижение размерности. Москва, Финансы и Статистика, 1989.
8. **Линник Ю.В.** Метод наименьших квадратов и основы теории наблюдений. Москва, ФМ 1962.
9. **Дрейпер Н., Смит Г**. Прикладной регрессионный анализ, 1 - 2. Москва, Финансы и Статистика, 1986 - 1987.
10. **Воробьев Е.М.** Введение в программу «Математика». Москва, Финансы и Статистика, 1998.
11. **Феллер В.** Введение в теорию вероятностей и ее приложения, том 1. Москва.: Мир, 1984.
12. **Феллер В.** Введение в теорию вероятностей и ее приложения, том 2. Москва.: Мир, 1984.
13. **Айвазян С. А., Енюков И. С., Мешалкин Л. Д.** Прикладная статистика. Основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.

14. **Айвазян С. А., Енюков И. С., Мешалкин Л. Д.** Прикладная статистика. Статистическое оценивание зависимостей.— М.: Финансы и статистика, 1985.
15. **Баранов А.Н.** Введение в прикладную лингвистику.

10. Методические указания для обучающихся по дисциплине.

В ходе проведения лекций и практических занятий предусмотрено выполнение студентами двух самостоятельных работ по каждому из разделов, а также экзамены по теории вероятностей и математической статистике. В итоговой оценке по каждому экзамену учитывается результат выполнения самостоятельных семестровых, лабораторных и контрольных работ. Все они вместе с активностью на практических занятиях и посещаемостью оцениваются по 10-ти бальной шкале. Суммарный рейтинг определяет поправку к оценке на экзамене и итоговую оценку за семестр, выставляемую в ведомость.

Критерии оценки ответов на вопросы билета

10 (десять) баллов:
студент демонстрирует системность и глубину знаний, точно и полно использует терминологию, умеет объяснить происхождение термина, дать исчерпывающее определение; использует в своем ответе знания, полученные при изучении соответствующих курсов; стилистически грамотно, логически правильно излагает ответы на вопросы; дает исчерпывающие ответы на дополнительные вопросы преподавателя по темам, предусмотренным учебной программой.

9 (девять) баллов:
студент демонстрирует системность и глубину знаний, в том числе полученных при изучении основной и дополнительной литературы; точно использует лингвистическую терминологию; умеет стилистически правильно излагать материал, логично строить ответ; полно и правильно отвечает на дополнительные вопросы преподавателя по темам, предусмотренным учебной программой, смежным с вопросами билета.

8 (восемь) баллов:
студент демонстрирует системность и глубину знаний в объеме учебной программы; владеет необходимой для ответа терминологией; логически правильно строит ответ на вопросы, делает обоснованные выводы; полно и правильно отвечает на дополнительные вопросы преподавателя по теме вопросов экзаменационного билета.

7 (семь) баллов:
студент демонстрирует глубину знаний; использует необходимую для ответа терминологию; логически правильно излагает ответы на вопросы, делает обоснованные выводы; полно раскрывает вопросы билета. Допускаются незначительные неточности в ответах на вопросы билета и на дополнительные вопросы.

6 (шесть) баллов:
студент демонстрирует достаточную полноту знаний в объеме учебной программы, владеет необходимой терминологией; в целом, раскрывает вопросы билета; ответ строит логически правильно, однако допустил незначительные ошибки, содержательные погрешности, которые были исправлены при ответе на дополнительные вопросы экзаменатора.

5 (пять) баллов:
студент демонстрирует достаточные знания по вопросам билета; в ответе допускает отдельные несущественные ошибки и неточности, в том числе в употреблении терминологии, которые затем исправляет, отвечая на вопросы экзаменаторов.

4 (четыре) балла:
студент демонстрирует неполные знания по вопросам билета, но способен дополнить свой ответ, отвечая на наводящие вопросы экзаменаторов; не совсем точно использует терминологию; в ответе встречаются ошибки.

3 (три) балла:
студент демонстрирует поверхностные знания по вопросам билета и дополнительным вопросам; усвоил только часть терминологии; при ответе допускает многочисленные ошибки.
2 (два) балла:
студент демонстрирует фрагментарные знания в рамках учебной программы; не владеет минимально необходимой терминологией; в ответе допускает грубые ошибки.
1 (один) балл:
студент демонстрирует отсутствие знаний; не ответил или отказался отвечать на вопросы билета.

Образовательные технологии

В курсе используются следующие методы и формы работы:
- лекции (2 часа в неделю)
- семинары (2 часа в неделю)
- консультации преподавателя
- самостоятельная работа с литературой.

Выполняются еженедельные письменные домашние задания. В ходе обучения слушатели выполняют индивидуальные научно-исследовательские задания в рамках проблематики курса. На семинарах анализируются отдельные теоретические и практические вопросы, выполняются практические задания, обсуждаются доклады и рефераты студентов. Даются месячные задания для самостоятельной работы. Итоговая аттестация проводится в виде зачета и экзамена в каждом семестре. Групповые и индивидуальные консультации предполагается проводить каждую неделю. В каждом семестре студенты готовят реферат и устный доклад на заданную тему. К общему списку вопросов и задач добавляются индивидуальные задания научно-исследовательского характера. При проведении практических занятий, особенно по математической статистике, предполагается использование компьютерных технологий.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

Программа «Математика 7.01», специализированный пакет программ «Статистика»

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Ноутбук, медиапроектор, мультемидийные презентации