**Программа учебной дисциплины Теория вероятностей и математическая статистика**

Утверждена

Академическим советом ООП

Протокол № от «\_\_»\_\_\_\_\_20\_\_ г.

|  |  |
| --- | --- |
| Авторы | д. ф.-м. н., проф. Ульянов В.В. (vulyanov@hse.ru),д.ф.-м.н. Шабанов Д.А. (dashabanov@hse.ru),д.ф.-м.н. Шапошников С.В. (sshaposhnikov@hse.ru) |
| Число кредитов  | ??? |
| Контактная работа (час.)  | 144 |
| Самостоятельная работа (час.)  | 160 |
| Курс  | 2 |
| Формат изучения дисциплины | Без использования онлайн курса |

1. **ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ**

Цель освоения дисциплины «Теория вероятностей и математическая статистика» — познакомить слушателей с основными понятиями, фактами и методами теории вероятностей и математической статистики, а также с их возможными приложениями для статистической обработки реальных данных.

В результате освоения дисциплины студент должен:

* Знать основные понятия теории вероятностей и математической статистики, их основные результаты и математические методы анализа.
* Уметь применять математические методы и модели к анализу случайных явлений для их адекватного описания и понимания.
* Владеть навыками решения стандартных задач теории вероятностей и математической статистики, а также применением основных аналитических инструментов для анализа вероятностных и статистических задач.

«Теория вероятностей и математическая статистика» является самостоятельной учебной дисциплиной, относится к математическому и естественнонаучному циклу дисциплин. Для специализации 01.03.02 «Прикладная математика и информатика» настоящая дисциплина является базовой.

Для освоения учебной дисциплины студенты должны владеть знаниями и навыками в объеме программы средней школы по математике и освоить учебные курсы:

* Дискретная математика,
* Математический анализ-1,-2,
* Линейная алгебра и геометрия,

Основные положения дисциплины могут быть использованы в дальнейшем при изучении следующих дисциплин:

* Машинное обучение 1;
* Машинное обучение 2;
* Теория информации;
* Прикладной статистический анализ данных

# Содержание УЧЕБНОЙ ДИСЦИПЛИНЫ

**Тема 1. Дискретные вероятностные пространства.**

Теория вероятностей как наука о случайных явлениях. Принцип устойчивости частот в природе. Вероятностное пространство как математическая модель эксперимента со случайными исходами. Дискретное вероятностное пространство (Ω,P). Простейшие свойства вероятности. Классическая модель вероятностного пространства, основные примеры. Условные вероятности. Формула полной вероятности и формула Байеса. Пример применения: задача о последнем пассажире в задаче о сумаcшедшей старушке. Независимость событий на вероятностном пространстве. Попарная независимость и независимость в совокупности. Пример Бернштейна. Независимость событий, связанных с последним и предпоследним пассажиром, в задаче о сумаcшедшей старушке.

**Тема 2. Случайные величины в дискретных вероятностных пространствах**

Случайные величины в дискретных вероятностных пространствах. Распределение случайной величины, основные примеры дискретных распределений случайных величин. Независимость случайных величин. Математическое ожидание случайной величины и его основные свойства. Дисперсия случайной величины, ковариация двух случайных величин. Их основные свойства. Дисперсия суммы независимых случайных величин.

**Тема 3. Закон больших чисел**

Неравенства Маркова и Чебышева. Закон больших чисел и его смысл. Сходимость по вероятности и сходимость с вероятностью 1. Их эквивалентность для дискретных вероятностных пространств. Схема испытаний Бернулли. Аппроксимация биномиального распределения: теорема Пуассона и теорема Муавра-Лапласа (б/д). Интерпретация теоремы Муавра-Лапласа, как оценки скорости сходимости в законе больших чисел для схемы Бернулли. Неравенство Чернова для вероятности уклонения от среднего в схеме Бернулли. Сравнение оценок скорости убывания вероятности уклонения от среднего в схеме Бернулли по неравенству Чернова и по неравенству Чебышева.

**Тема 4. Общее понятие вероятностного пространства.**

Общее понятие вероятностного пространства. Тройка Колмогорова (Ω,F,P). Вероятностные меры на прямой. Борелевская сигма-алгебра, доказательства существования. Функция распределения вероятностной меры на прямой, лемма о ее трех основных свойствах. Примеры функций распределения. Теорема Каратеодори о продолжении вероятностной меры (б/д). Взаимная однозначность функций распределения на прямой и вероятностных мер (только идея доказательства). Классификация вероятностных мер на прямой по функциям распределения: дискретные, абсолютно непрерывные, сингулярные. Основные примеры распределений. Теорема Лебега о представлении произвольной функции распределения (без док-ва).

**Тема 5. Непрерывные случайные величины.**

Случайные величины и векторы в общих вероятностных пространствах. Понятие измеримости и его смысл. Борелевская сигма-алгебра в Rn. Эквивалентные определения случайной величины и случайного вектора. Индикаторы и константы как простейшие случайные величины. Действия над случайными величинами и векторами. Борелевские функции. Непрерывные функции как борелевские. Арифметические операции над случайными величинами, взятие пределов и точных верхних/нижних граней у последовательностей случайных величин. Построение математического ожидания в общем случае. Простые случайные величины, определение математического ожидания для них и доказательство его основных свойств. Неотрицательные случайные величины, приближение простыми (явный вид последовательности), доказательство корректности определения математического ожидания. Определение математического ожидания в произвольном случае. Независимость случайных величин в общем случае. Критерий независимости (без док-ва). Независимость случайных векторов. Лемма о независимости функций от независимых случайных величин или векторов. Теорема о математическом ожидании произведения независимых случайных величин. Функция распределения и распределение случайной величины. Формулы подсчета математических ожиданий. Подсчет с помощью рядов в дискретном случае. Функция распределения вероятностной меры в Rn, n>1, ее основные свойства. Теорема о взаимной однозначности мер и функций распределения (б/д). Понятие плотности многомерного распределения. Совместное распределение конечного набора случайных величин (распределение случайного вектора). Совместная функция распределения, совместная плотность, их вычисление в случае независимости. Формула вычисления математического ожидания функций от случайных величин в случае наличия совместной плотности. Формула свертки для плотности суммы независимых случайных величин. Примеры вычисления. Дисперсия, ковариация и коэффициент корреляции. Лемма об их основных свойствах. Дисперсия суммы независимых случайных величин. Математическое ожидание и матрица ковариаций случайного вектора. Симметричность и неотрицательная определенность матрицы ковариаций.

**Тема 6. Сходимости случайных величин.**

Виды сходимостей случайных величин: с вероятностью 1, по вероятности, в среднем порядка p>0, по распределению. Критерий сходимости с вероятностью 1. Теорема о взаимоотношении различных видов сходимостей. Достаточное условие сходимости почти наверное для последовательности случайных величин. Усиленный закон больших чисел для случайных величин с конечным четвертым моментом. Усиленный закон больших чисел в форме Колмогорова (б/д). Смысл усиленного закона больших чисел. Предельный переход под знаком математического ожидания. Теорема о монотонной сходимости, лемма Фату и теорема Лебега о мажорируемой сходимости.

**Тема 7. Характеристические функции.**

Характеристические функции случайных величин. Их основные свойства. Примеры вычислений характеристических функций: биномиальное и экспоненциальное распределения. Пример вычисления распределения суммы независимых пуассоновских случайных величин с помощью характеристических функций. Теорема единственности для характеристических функций случайных величин. Вычисление характеристической функции для стандартной случайной величины. Следствие: распределение суммы независимых нормальных случайных величин. Формула обращения для нахождения плотности (б/д). Теорема о производных характеристической функции. Характеристические функции случайных векторов (совместная характеристическая функция). Критерий независимости набора случайных величин для характеристических функций.

**Тема 8. Предельные теоремы.**

Теорема непрерывности для характеристических функций (б/д). Центральная предельная теорема для независимых одинаково распределенных случайных величин. Пример применения. Скорость сходимости в ЦПТ: теорема Берри - Эссеена (б/д). Сходимости случайных векторов: с вероятностью 1, по вероятности, по распределению. Эквивалентное определение сходимости по распределению для случайных величин (б/д). Взаимосвязь многомерной сходимости и одномерных сходимостей координат. Лемма о существовании подпоследовательности, сходящейся п.н., если последовательность сходится по вероятности. Теорема о наследовании сходимости. Лемма Слуцкого. Примеры применения леммы Слуцкого и теоремы о наследования сходимости.

**Тема 9. Многомерное нормальное распределение.**

Гауссовские случайные векторы (многомерное нормальное распределение). Теорема о трех эквивалентных определениях. Следствия: смысл параметров, корректность определения, линейные преобразования. Критерий независимости компонент гауссовского вектора. Теорема о плотности гауссовского случайного вектора. Многомерная центральная предельная теорема (б/д).

**Тема 10. Условное математическое ожидание.**

Условное математическое ожидание: определение и явная формула для вычисления в случае, если случайная величина условие имеет дискретное распределение. Условное математическое ожидание E(X|Y=y), связь с E(X|Y). Условное распределение и условная плотность. Вычисление условного математического ожидания с помощью условной плотности (б/д). Теорема о достаточном условии существования условной плотности. Основные свойства условного математического ожидания, свойства условного математического ожидания E(X|Y=y).

**Тема** **11. Основные понятия математической статистики.**

Основные задачи математической статистики. Статистическая структура. Параметрические семейства, примеры. Выборка. Выборочное пространство. Статистика. Эмпирическая функция распределения. Гистограмма. Теорема Гливенко-Кантелли. Выборочные моменты, их асимптотическая нормальность. Вариационный ряд. Порядковые статистики и их распределения. Точечные оценки. Несмещенные оценки, их свойства, примеры. Состоятельные оценки, их свойства, примеры. Оптимальные оценки. Теорема о единственности оптимальной оценки. Функции правдоподобия для дискретного и абсолютно непрерывного случаев. Информация Фишера. Неравенство Рао-Крамера. Относительная частота события как оптимальная оценка неизвестной вероятности. Эффективность точечных оценок. Эффективные оценки. Критерий эффективности.

**Тема** **12. Оценивание параметров.**

Основные методы нахождения точечных оценок параметров: метод максимального правдоподобия (ММП) и метод моментов (ММ). Теорема о сходимости по вероятности для непрерывных функций от случайных величин, сходящихся по вероятности. Оценки максимального правдоподобия. Уравнения правдоподобия. Утверждения об эффективных оценках, достаточных статистиках и оценках максимального правдоподобия. Принцип инвариантности для оценок максимального правдоподобия. Свойства ОМП. Достаточные статистики. Критерий факторизации. Теорема Рао-Блекуэлла-Колмогорова. Полные статистики. Оптимальность оценок, являющихся функцией полной достаточной статистики. Распределения, связанные с гауссовским (распределение хи-квадрат, распределение Стьюдента, распределение Фишера). Доверительные интервалы. Трактовка коэффициента доверия. Методы построения интервальных оценок: использование точечных оценок, метод центральной статистики, использование центральной предельной теоремы. Построение доверительных интервалов параметров в одновыборочных и двувыборочных гауссовских моделях. Асимптотические доверительные интервалы. Преобразование, стабилизирующее дисперсию.

**Раздел 13. Проверка статистических гипотез.**

Проверка гипотез. Задача о различении двух гипотез о доле шаров в урне. Оценка снизу для числа наблюдений, необходимых для различения гипотез с заданной точностью. Примеры гипотез: о виде распределений, об однородности выборок, о независимости. Простые и сложные гипотезы. Статистический критерий. Нерандомизированные S-критерии. Ошибки первого и второго рода. Уровень значимости, мощность критерия. Рандомизированные критерии. Функция мощности критерия. Состоятельные критерии. Несмещённые критерии. Понятие равномерно наиболее мощного (РНМ) и локально наиболее мощного (ЛНМ) критерия. Алгоритм проверки статистической гипотезы. Лемма Неймана-Пирсона. Построение РНМ критериев в одновыборочных гауссовских моделях. Критерий Стьюдента и критерий Фишера для проверки параметрических гипотез в двухвыборочных гауссовских моделях. Критерий Колмогорова для проверки простой гипотезы о виде распределения случайной величины. Критерии согласия Пирсона хи-квадрат для дискретных и абсолютно непрерывных распределений для проверки простых и сложных гипотез о виде распределения случайной величины. Асимптотика для критерия Пирсона. Критерий однородности хи-квадрат. Критерий, основанный на выборочном коэффициенте корреляции. Таблица сопряжённости признаков. Критерии проверки независимости двух случайных величин. Связь между задачами проверки гипотез и доверительным оцениванием.

**Раздел 14. Регрессионный анализ.**

Модель линейной регрессии. Методы оценивания параметров в линейной регрессионной модели (МНК, ВМНК, МНМ, ранговый). МНК-оценка параметров и её свойства. Критерии проверки адекватности гауссовской линейной регрессионной модели.

# ОЦЕНИВАНИЕ

|  |  |  |  |
| --- | --- | --- | --- |
| Тип контроля | Форма контроля | 1 год | Параметры  |
| 1 | 2 | 3 | 4 |
| Текущий(неделя) | Контрольная работа |  | 8 |  | 9 | Письменная работа 80 минут |
|  |  |  |  |  |
| Коллоквиум |  | 7 |  | 8 | Устное собеседование |
| Домашнее задание |  | 8 |  | 9 |  |
| Промежу­точный | Экзамен |  | э |  |  | Письменная работа на 120 минут |
| Итоговый | Экзамен  |  |  |  | э | Письменная работа на 120 минут |

## Критерии оценки знаний, навыков

Для прохождения контроля студент должен, как минимум, продемонстрировать знания основных определений, формулировок теорем и доказательства базовых теоретических утверждений; умение решать типовые задачи, разобранные на семинарских занятиях. Контрольные работы и домашние задания заключаются в решении задач. Проверяется ответ и ход решения. Коллоквиум заключается в устных ответах на вопросы по знанию теоретического материала. На экзамене необходимо решить определенное число задач.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

**Порядок формирования оценок по дисциплине**

Работа будет оценена в ноль баллов, если в ней получено значение вероятности, не принадлежащее отрезку [0;1].

Оценки за промежуточный контроль формируются следующим образом.

***Промежуточный контроль (экзамен) во втором модуле.***

Оценка промежуточного контроля 1-го этапа рассчитывается по формуле

О *промежуточная 1 этапа* = 0.7·О *накопленная 1 этапа* + 0.3·О *промежуточный экзамен №1* ,

где

 О *накопленная 1 этапа* = (3/7)\* О *КЛ№1* +(3/7)\* О *КР№1* +(1/7)\* О *ДЗ№1*

Оценка за экзаменационную письменную работу (О *промежуточный экзамен №1*) – не блокирующая.

***Промежуточный контроль (экзамен) в четвертом модуле.***

В диплом выставляется результирующая оценка по учебной дисциплине, которая рассчитывается по следующей формуле:

О *результ* = 0.8\*О*накопленная итоговая* + 0.2\*О *итоговый экзамен*

где О*накопленная итоговая* формируется следующим образом:

0.8\*О*накопленная итоговая*=0.4\*О*промежуточная 1 этапа* + 0.17\*О*КР№2* + 0.17\*О*КЛ№2*+ 0.06\*О*ДЗ№2*.

Правило округления оценок – арифметическое.

Оценка за итоговый контроль – не блокирующая.

# ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

## Примеры вопросов для оценки качества освоения дисциплины

**Тема 1**.

1. Дайте определение дискретного вероятностного пространства.

2. Дайте определение независимости в совокупности конечного набора событий.

3. Что такое условная вероятность события?

4. Приведите пример двух зависимых событий

**Тема 2.**

1. Дайте определение случайной величины.

2. Что такое математическое ожидание случайной величины в дискретном вероятностном пространстве?

3. Сформулируйте основные свойства математического ожидания.

4. Приведите пример зависимых и некоррелированных случайных величин.

**Тема 3**.

1. Какой из видов сходимостей сильнее в дискретных вероятностных пространствах: по распределению или по вероятности?

2. Сформулируйте закон больших чисел в форме Чебышева.

3. Сравните скорости сходимости к нулю вероятности уклонения от среднего значения в схеме Бернулли, получающиеся по неравенствам Чебышева и Чернова.

**Тема 4**.

1. Дайте определение вероятностного пространства в аксиоматике Колмогорова.

2. Сформулируйте основные свойства функции распределения на прямой.

3. Дайте определение нормального распределения.

**Тема 5.**

1. Дайте определение случайной величины в общем вероятностном пространстве.

2. Приведете формулу вычисления дисперсии случайной величины с заданном плотностью p(x).

3. Что такое формула свертки?

4. Приведите критерий независимости случайных величин.

5. Сформулируйте основные свойства математического ожидания.

6. Дайте определение матрицы ковариаций случайного вектора.

**Тема 6.**

1. Дайте определения основных видов сходимостей случайных величин.

2. Сформулируйте усиленный закон больших чисел.

3. Поясните связь усиленного закона больших чисел и принципа устойчивости частот.

**Тема 7.**

1. Что такое характеристическая функция случайной величины?

2. Вычислите характеристическую функцию бернуллиевской случайной величины.

3. Каковы основные свойства характеристических функций?

4. Сформулируйте критерия независимости случайных величин для характеристических функций.

**Тема 8.**

1. Дайте два эквивалентных определения сходимости по распределению случайных величин.

2. Сформулируйте центральную предельную теорему.

3. Какова скорость сходимости в центральной предельной теореме?

4. В чем состоит метод характеристических функций для доказательства предельных теорем?

**Тема 9.**

1. Дайте три эквивалентных определения гауссовского случайного вектора.

2. Каковы основные свойства гауссовских случайных векторов.

3. При каких условиях гауссовский случайный вектор имеет плотность?

**Тема 10.**

1. Дайте определение условного математического ожидания случайной величины относительно другой случайной величины.

2. Каковы основные свойства условного математического ожидания?

3. Что такое условная плотность одной случайной величины относительно другой?

4. Каков способ вычисления условного математического ожидания с помощью условной плотности?

**Тема 11.**

1. Что такое выборка?

2. Приведите пример состоятельной, но не несмещенной оценки.

3. Сформулируйте неравенство Рао-Крамера.

4. Каков критерий эффективности оценки?

5. Сравните понятия оптимальности и эффективности оценок.

**Тема 12.**

1. В чем состоит метод максимального правдоподобия для нахождения оценок?

2. Каковы основные свойства оценки максимального правдоподобия?

3. Дайте определение достаточной статистики.

4. Каков способ нахождения оптимальных оценок с помощью полных достаточных статистик?

5. Приведите пример построения доверительного интервала для параметра бернуллиевского распределения.

**Тема 13.**

1. В чем состоят ошибки первого и второго родов при проверке гипотез?

2. Что такое равномерно наиболее мощный критерий?

3. Сформулируйте лемму Неймана-Пирсона.

4. В чем состоит критерий согласия хи-квадрат?

5. Приведите пример состоятельного критерия для проверки однородности двух выборок.

**Тема 14.**

1. В чем состоит задача линейной регрессии?

2. Какова формула вычисления оценки наименьших квадратов?

3. Каковы основные свойства оценки наименьших квадратов в случае гауссовской линейной модели?

1. **РЕСУРСЫ**
	1. **Основная литература**

1. Ширяев А.Н. Вероятность. В 2-х кн., 3-е изд., М.: МЦНМО, 2004

2. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. М.: Изд-во ЛКИ, 2010.

3. Гнеденко Б.В. Курс теории вероятностей. 7-е изд. М.: Едиториал УРСС, 2001.

**2. Дополнительная литература**

1. Ульянов В.В., Ушаков В.Г., Байрамов Н.Р., Нагапетян Т.А. Задачи по математической статистике с решениями. МАКС Пресс Москва, 2008.
2. Севастьянов Б. А. Курс теории вероятностей и математической статистики. - 2-е изд. - М.-Ижевск: Институт компьютерных исследований, 2004.
3. Чистяков В.П. Курс теории вероятностей - М.: Наука, 1987.
4. Леман Э. Проверка статистических гипотез. – М.: Наука, 1964.
5. Горяинова Е.Р., Наумов А.В., Сиротин А.Н. Решение задач по теории вероятностей.- М.: МАИ, 2001.
6. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. М.: Наука, 1989.
7. Боровков А. А. Теория вероятностей. - 4-е изд. - М.: Едиториал УРСС, 2003.
8. Боровков А.А., Математическая статистика. СПб: Изд-во «Лань», 2010.
9. Секей Г., Парадоксы в теории вероятностей и математической статистике. Москва-Ижевск: Институт компьютерных исследований, 2003.
10. Spokoiny V., Dickhaus T. Basics of Modern Mathematical Statistics. Springer-Verlag, Berlin Heidelberg, 2015
11. Haerdle W.K., Spokoiny V., Panov V., Wang W., Basics of Modern Mathematical Statistics, Exercises and Solutions. Springer-Verlag, Berlin Heidelberg, 2014.
12. Чибисов Д.М., Пагурова В.И. Задачи по математической. статистике. М.,«МГУ», 1990.
13. Леман Э. Теория точечного оценивания. М., «Наука», 1991.

**3. Программное обеспечение**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Не требуется**.**

**4. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |
|  |  |  |
|  |  |  |
|  |  |
|  |  |  |

Не требуется.

**5. Материально-техническое обеспечение дисциплины**

Учебные аудитории для лекционных, семинарских и самостоятельных занятий по дисциплине не требуют специального технического оснащения.