**Правительство Российской Федерации**

**Федеральное государственное автономное образовательное учреждение высшего профессионального образования
«Национальный исследовательский университет
«Высшая школа экономики»**

**Факультет Компьютерных наук**

 **Департамент больших данных и информационного поиска**

УТВЕРЖДАЮ

Академический руководитель

образовательной программы

по направлению 01.03.02

 «Прикладная математика и информатика»

А.С. Конушин

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 2015 г.

**Программа дисциплины**

**Теория вероятностей и математическая статистика**

Для направления 01.03.02 «Прикладная математика и информатика» подготовки бакалавров

Авторы программы:

**д. ф.-м. н., проф. Ульянов В.В.** (vulyanov@hse.ru), **д.ф.-м.н. Шабанов Д.А.** ()

Одобрена на заседании Департамента больших данных и информационного поиска

 «\_\_\_» \_\_\_\_\_\_ 2015  г.

Руководитель департамента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.П. Ветров

Рекомендована Академическим советом

образовательной программы

«Прикладная математика и информатика» «\_\_\_»\_\_\_\_\_\_\_\_\_\_ 2015  г.

Менеджер Департамента больших данных и информационного поиска

 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Москва, 2015

*Настоящая программа не может быть использована другими подразделениями
университета и другими вузами без разрешения кафедры-разработчика программы*

# Область применения и нормативные ссылки

Настоящая программа учебной дисциплины «Теория вероятностей и математическая статистика» (2-ый год обучения) устанавливает минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов направления подготовки 01.03.02 «Прикладная математика и информатика», обучающихся по бакалаврской программе «Прикладная математика и информатика» изучающих дисциплину «Теория вероятностей и математическая статистика».

Программа разработана в соответствии с:

* образовательным стандартом федерального государственного автономного образовательного учреждения высшего профессионального образования НИУ ВШЭ подготовки бакалавров по направлению 010400.62 «Прикладная математика и информатика»;
* Образовательной программой подготовки бакалавров по направлению 01.03.02 «Прикладная математика и информатика»,
* Рабочим учебным планом университета подготовки бакалавров по направлению 01.03.02 «Прикладная математика и информатика», утвержденным в 2015г.

# Цели освоения дисциплины

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.

# Компетенции обучающегося, формируемые в результате освоения дисциплины

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.

В результате освоения дисциплины студент осваивает следующие компетенции:

| Компетенция | Код по ФГОС/ НИУ | Дескрипторы – основные признаки освоения (показатели достижения результата) | Формы и методы обучения, способствующие формированию и развитию компетенции |
| --- | --- | --- | --- |
| Общенаучная | ОНК-1 | Способность к анализу и синтезу на основе системного подхода | Стандартные (лекционно-семинарские) |
| Общенаучная | ОНК-2 | Способность перейти от проблемной ситуации к проблемам, задачам и лежащим в их основе противоречиям | Стандартные (лекционно-семинарские) |
| Общенаучная | ОНК-3 | Способность использовать методы критического анализа, развития научных теорий, опровержения и фальсификации, оценить качество исследований в некоторой предметной области | Стандартные (лекционно-семинарские) |
| Общенаучная | ОНК-4 | Готовность использовать основные законы естественнонаучных дисциплин в профессиональной деятельности, применять методы математического анализа и моделирования, теоретического и экспериментального исследования при работе в какой-либо предметной области | Стандартные (лекционно-семинарские) |
| Общенаучная | ОНК-5 | Готовность выявить естественнонаучную сущность проблем, возникающих в ходе профессиональной деятельности, привлечь их для решения соответствующий аппарат дисциплины | Стандартные (лекционно-семинарские) |
| Общенаучная | ОНК-6 | Способность приобретать новые знания с использованием научной методологии и современных образовательных и информационных технологий | Стандартные (лекционно-семинарские) |
| Общенаучная | ОНК-7 | Способность порождать новые идеи (креативность) | Стандартные (лекционно-семинарские) |
| Профессиональные | ПК-1 | Способность демонстрации общенаучных базовых знаний естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой | Стандартные (лекционно-семинарские) |
| Профессиональные | ПК-2 | способность понимать и применять в исследовательской и прикладной деятельности современный математический аппарат | Стандартные (лекционно-семинарские) |
| Профессиональные | ПК-3 | способность в составе научно-исследовательского и производственного коллектива решать задачи профессиональной деятельности в соответствии с профилем подготовки, общаться с экспертами в других предметных областях | Стандартные (лекционно-семинарские) |
| Профессиональные | ПК-4 | способность критически оценивать собственную квалификацию и её востребованность, переосмысливать накопленный практический опыт, изменять при необходимости вид и характер своей профессиональной деятельности | Стандартные (лекционно-семинарские) |
| Профессиональные | ПК-8 | способность решать задачи производственной и технологической деятельности на профессиональном уровне, включая разработку математических моделей, алгоритмических и программных решений | Стандартные (лекционно-семинарские) |

# Место дисциплины в структуре образовательной программы

«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;
* Эконометрика;
* Анализ данных.

# Тематический план учебной дисциплины

#

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название раздела | Всего часов  | Аудиторные часы | Самостоя­тельная работа |
| Лекции | Семинары | Практические занятия |
| 1 | Вероятностные пространства | 32 | 8 | 8 |  | 16 |
| 2 | Случайные величины | 36 | 8 | 8 |  | 20 |
| 3 | Сходимости случайных величин | 28 | 6 | 6 |  | 16 |
| 4 | Характеристические функции | 28 | 6 | 6 |  | 16 |
| 5 | Предельные теоремы | 20 | 4 | 4 |  | 12 |
| 6 | Условное математическое ожидание | 20 | 4 | 4 |  | 12 |
| 7 | Основные понятия математической статистики | 28 | 6 | 6 |  | 16 |
| 8 | Оценивание параметров | 42 | 14 | 14 |  | 24 |
| 9 | Проверка статистических гипотез | 42 | 12 | 12 |  | 18 |
| 10 | Регрессионный анализ | 20 | 4 | 4 |  | 12 |
|  |  |  |  |  |  |  |
|  | Итого: | 304 | 72 | 72 |  | 160 |

# Формы контроля знаний студентов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Тип контроля | Форма контроля | 2 год | Кафедра | Параметры  |
| 1 | 2 | 3 | 4 |
| Текущий | Контрольная работа | 1 | 2 | 1 | 2 |  | Письменная работа 40 минут |
| Домашнее задание |  | 1 |  | 1 |  | Всего 5-10 задач |
| Промежуточный | Экзамен |  | 1 |  |  |  | Письменный экзамен на 90 минут |
| Промежуточный | Экзамен |  |  |  | 1 |  | Письменный экзамен на 90 минут |

## Критерии оценки знаний, навыков

В рамках курса студенты должны выполнить две домашние письменные работы, написать контрольные работы и сдать два экзамена (после 2 и 4 модулей). Каждое задание и экзамен оцениваются по 10-балльной шкале. Контрольные работы и домашние задания заключаются в решении задач. Проверяется ответ и ход решения. На экзамене также необходимо изложить теоретический лекционный материал.

## Порядок формирования оценок по дисциплине

**Промежуточный контроль во втором модуле**

Накопленная оценка (1-й и 2-й модули) за текущий контроль формируется из оценок за домашние задания, работу на семинарах и контрольные работы следующим образом:

$$О\_{накопленная1}=0.1\*O\_{д/з1}+0.8\*O\_{к/р1}+0.1\*O\_{сем1}$$

Оценка промежуточного контроля за первый этап рассчитывается следующим образом:

$О\_{итог1}=$ $0.4\*O\_{накопленная1}+0.6\*O\_{экз1}$

**Промежуточный контроль в четвертом модуле**

Накопленная оценка (3-й и 4-й модули) за текущий контроль формируется из оценок за домашние задания, работу на семинарах и контрольные работы следующим образом:

$$О\_{накопленная2}=0.1\*O\_{д/з2}+0.8\*O\_{к/р2}+0.1\*O\_{сем2}$$

Оценка промежуточного контроля за первый этап рассчитывается следующим образом:

$О\_{итог2}=$ $0.4\*O\_{накопленная2}+0.6\*O\_{экз2}$

**Общая итоговая оценка**

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

$$О\_{итог}=0.5\*O\_{итог1}+0.5\*O\_{итог2}$$

Все оценки выставляются по 10-балльной системе, способ округления всегда арифметический.

# Содержание дисциплины

**Раздел 1. Вероятностные пространства.**

Теория вероятностей как наука о случайных явлениях. Принцип устойчивости частот в природе. Вероятностное пространство как математическая модель эксперимента со случайными исходами. Вероятностное пространство (Ω,F,P). Аксиомы Колмогорова. Простейшие свойства вероятностной меры. Теорема о непрерывности вероятностной меры. Дискретные вероятностные пространства. Классическое определение вероятности. Примеры. Геометрические вероятности. Условные вероятности. Формула полной вероятности. Формула Байеса. Системы множеств (алгебры, σ-алгебры). Наименьшая алгебра (σ-алгебра), порожденная системой множеств. Борелевские σ-алгебры в R, Rn. Независимость событий и систем событий. Пример Бернштейна. Лемма о достаточном условии независимости σ-алгебр. Теорема Каратеодори о продолжении вероятностной меры. Вероятностные меры на (R, B(R)). Функция распределения вероятностной меры, ее свойства. Классификация функций распределения, теорема Лебега (б/д), примеры. Теорема о построении вероятностной меры на (R, B(R)) по функции распределения (идея док-ва). Вероятностные меры на (Rn,B(Rn)). Многомерная функция распределения, ее свойства. Примеры. Теорема о построении вероятностной меры на (Rn, B(Rn)) по функции распределения (б/д).

**Раздел 2. Случайные величины.**

Случайные величины в дискретных вероятностных пространствах. Независимость случайных величин. Математическое ожидание случайной величины, его основные свойства. Дисперсия, ковариация и их свойства. Случайные элементы, случайные величины и векторы в произвольных пространствах. Действия над случайными величинами. Характеристики случайной величины (вектора): распределение вероятностей, функция распределения, σ-алгебра, порожденная случайной величиной. Независимость произвольного набора случайных величин. Критерий независимости случайных величины в терминах функций распределения. Математическое ожидание случайной величины (интеграл Лебега по вероятностной мере). Основные свойства математического ожидания. Дисперсия, ковариация и их свойства. Матрица ковариаций случайного вектора, ее неотрицательная определенность. Формула пересчета математических ожиданий. Теорема о замене переменных в интеграле Лебега. Прямое произведение вероятностных пространств. Теорема Фубини (б/д). Совместное распределение конечного набора случайных величин. Свертка распределений.

**Раздел 3. Сходимости случайных величин.**

Виды сходимости случайных величин (почти наверное, по вероятности, в пространстве Lp, по распределению), их взаимосвязи. Критерий Коши сходимости с вероятностью 1. Неравенство Маркова. Неравенство Чебышева. Закон больших чисел в форме Чебышева. Неравенство Колмогорова. Теорема о сходимости почти наверное ряда из случайных величин. Усиленный закон больших чисел для независимых случайных величин с ограниченными дисперсиями.

Предельный переход под знаком интеграла Лебега. Теорема о монотонной сходимости, лемма Фату, теорема Лебега о мажорированной сходимости. Лемма Бореля - Кантелли. Усиленный закон больших чисел в форме Колмогорова.

**Раздел 4. Характеристические функции.**

Характеристические функции вероятностных мер, функций распределения, случайных величин и векторов. Примеры. Основные свойства характеристических функций случайных величин. Теорема единственности для характеристических функций. Независимость компонент случайного вектора в терминах характеристических функций. Гауссовские случайные векторы (многомерное нормальное распределение), три эквивалентных определения. Основные свойства гауссовских случайных векторов.

**Раздел 5. Предельные теоремы.**

Слабая сходимость и сходимость в основном вероятностных мер. Теорема Александрова (б/д). Теорема об эквивалентности сходимости по распределению случайных величин и сходимости функций распределения во всех точках непрерывности предельной функции. Схема испытаний Бернулли и полиномиальная схема. Предельные теоремы для схемы Бернулли: теорема Пуассона и теорема Муавра - Лапласа (б/д). Их интерпретация в терминах сходимостей случайных величин. Теорема непрерывности для характеристических функций (б/д). Центральная предельная теорема для независимых одинаково распределенных случайных величин. Теорема Берри - Эссеена о скорости сходимости в центральной предельной теореме (б/д). Многомерная центральная предельная теорема.

**Раздел 6. Условное математическое ожидание.**

Условное математическое ожидание случайной величины относительно сигма-алгебры. Обоснование существования условного математического ожидания. Явный вид условного математического ожидания в случае, если сигма-алгебра порождена конечным или счетным разбиением. Свойства условного математического ожидания. Условные распределения, условные плотности. Теорема о достаточном условии существования условной плотности. Вычисление условных математических ожиданий с помощью условной плотности. Примеры.

**Раздел 7. 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.

**Раздел 8. 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 9. 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 10. 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.

# Образовательные технологии

В учебном процессе используются следующие образовательные технологии:

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Вид занятия | Форма проведения занятий | Цель |
| 1 | Лекция | Изложение теоретического материала. | Получение теоретических знаний по предмету. |
| 2 | Семинар | Решение задач. | Получение практических навыков в решении задач. |
| 3 | Самостоятельная работа студента | Решение задач. | Усовершенствование практических навыков.  |

# Оценочные средства для текущего контроля и аттестации студента

## Вопросы для оценки качества освоения дисциплины

**Раздел 1.**

1. Дайте определение вероятностного пространства.

2. Дайте определение независимости в совокупности конечного набора событий.

3. Что такое условная вероятность события?

4. Сформулируйте основные свойства функции распределения на прямой.

5. Какова классификация функций распределения на прямой.

**Раздел 2.**

1. Дайте определение случайной величины и случайного вектора.

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

3. Что такое нормальная случайная величина?

4. Сформулируйте основные свойства математического ожидания.

5. Что такое формула свертки?

**Раздел 3.**

1. Дайте определение сходимости с вероятностью 1 последовательности случайных величин?

2. Какой из видов сходимостей сильнее: по распределению или по вероятности?

3. Сформулируйте закон больших чисел в форме Чебышева.

4. Дайте формулировку усиленного закона больших чисел в форму Колмогорова.

5. Приведите пример последовательности случайных величин, которая сходится по вероятности, но не сходится почти наверное.

**Раздел 4.**

1. Что такое характеристическая функция случайной величины?

2. Найдите характеристическую функцию бернуллиевской случайной величины?

3. Каковы основные свойства характеристических функций?

4. Сформулируйте критерия независимости случайных величин для характеристических функций.

5. Дайте три эквивалентных определения гауссовских случайных векторов.

**Раздел 5.**

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

2. Сформулируйте центральную предельную теорему.

3. Какова скорость сходимости в центральной предельной теореме?

4. Дайте интерпретацию теоремы Пуассона в терминах сходимостей случайных величин.

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

**Раздел 6.**

1. Дайте определение условного математического ожидания случайной величины относительно сигма-алгебры?

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?

**Раздел 7.**

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.

**Раздел 8.**

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.

**Раздел 9.**

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.

**Раздел 10.**

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?

# Учебно-методическое и информационное обеспечение дисциплины

## Базовые учебники

1. Ширяев А.Н. Вероятность. В 2-х кн., 4-е изд., М.: МЦНМО, 2007

2. Ивченко Г.И., Медведев Ю.И., Математическая статистика. М.: Книжный дом «ЛИБРОКОМ», 2014.

## Основная 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. Гнеденко Б.В. Курс теории вероятностей. М.: Едиториал УРСС, 2005.

7. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. М.: Наука, 1989.

## Дополнительная литература

1. Боровков А. А. Теория вероятностей. - 4-е изд. - М.: Едиториал УРСС, 2003.
2. Боровков А.А., Математическая статистика. СПб: Изд-во «Лань», 2010.
3. Секей Г., Парадоксы в теории вероятностей и математической статистике. Москва-Ижевск: Институт компьютерных исследований, 2003.
4. Spokoiny V., Dickhaus T. Basics of Modern Mathematical Statistics. Springer-Verlag, Berlin Heidelberg, 2015
5. Haerdle W.K., Spokoiny V., Panov V., Wang W., Basics of Modern Mathematical Statistics, Exercises and Solutions. Springer-Verlag, Berlin Heidelberg, 2014.
6. Д.М. Чибисов, В.И. Пагурова. Задачи по математической. статистике. М.,«МГУ», 1990.
7. Э. Леман. Теория точечного оценивания. М., «Наука», 1991.