

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО
Директор физтех-школы
прикладной математики и
информатики
А.М. Райгородский

Рабочая программа дисциплины (модуля)

по дисциплине: Математическая статистика
по направлению: Прикладная математика и информатика
профиль подготовки: Искусственный интеллект и большие данные
Сетевое обучение
кафедра дискретной математики
курс: 2
квалификация: бакалавр

Семестр, формы промежуточной аттестации: 4 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 84 час.

Подготовка к экзамену: 0 час.

Всего часов: 144, всего зач. ед.: 4

Программу составил: И.В. Родионов, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры дискретной математики 30.08.2022

Аннотация

Курс математической статистики изучается студентами ФПМИ в пятом семестре и является обязательной дисциплиной базовой части. В рамках данной дисциплины студенты проходят следующие темы: сходимости случайных векторов, статистики и оценки, методы нахождения оценок, эффективные оценки, достаточные статистики и оптимальные оценки, байесовские оценки, доверительное оценивание, линейная регрессионная модель, проверка гипотез и равномерно наиболее мощные критерии, критерии, основанные на нормальности, критерии согласия, критерии проверки независимости.

1. Цели и задачи

Цель дисциплины

изучение математических и теоретических основ современного статистического анализа, а также подготовка слушателей к дальнейшей самостоятельной работе в области анализа статистических задач прикладной математики, физики и экономики.

Задачи дисциплины

изучение математических основ математической статистики;
приобретение слушателями теоретических знаний в области современного статистического анализа.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения	ОПК-2.1. Использует знания основных положений и концепций в области программирования, архитектуру языков программирования, основную терминологию и базовые алгоритмы, основные требования информационной безопасности для практического применения
	ОПК-2.2. Анализирует типовые языки программирования, составляет программы
	ОПК-2.3 Применяет на практике опыт решения задач с использованием базовых алгоритмов, анализа типов коммуникаций и интеграции различных типов программного обеспечения

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные понятия математической статистики;
- основные подходы к сравнению оценок параметров неизвестного распределения;
- асимптотические и неасимптотические свойства оценок параметров неизвестного распределения;
- основные методы построения оценок с хорошими асимптотическими свойствами: метод моментов, метод максимального правдоподобия, метод выборочных квантилей;
- понятие эффективных оценок и неравенство информации Рао-Крамера;
- определение и главные свойства условного математического ожидания случайной величины относительно сигма-алгебры или другой случайной величины;
- определение общей линейной регрессионной модели и метод наименьших квадратов;
- многомерное нормальное распределение и его основные свойства;

- базовые понятия теории проверки статистических гипотез;
- лемму Неймана – Пирсона и теорему о монотонном отношении правдоподобия;
- критерий хи-квадрат Пирсона для проверки простых гипотез в схеме Бернулли.

уметь:

- обосновывать асимптотические свойства оценок с помощью применения предельных теорем теории вероятностей;
- строить оценки с хорошими асимптотическими свойствами для параметров неизвестного распределения по заданной выборке из него;
- находить байесовские оценки по заданному априорному распределению;
- вычислять условные математические ожидания с помощью условных распределений;
- находить оптимальные оценки с помощью полных достаточных статистик;
- строить точные и асимптотические доверительные интервалы, и области для параметров неизвестного распределения;
- находить оптимальные оценки и доверительные области в гауссовской линейной модели;
- строить равномерно наиболее мощные критерии в случае параметрического семейства с монотонным отношением правдоподобия;
- строить F-критерий для проверки линейных гипотез в линейной гауссовской модели.

владеть:

- основными методами математической статистики построения точечных и доверительных оценок: методом моментов, выборочных квантилей, максимального правдоподобия, методом наименьших квадратов, методом центральной статистики.
- навыками асимптотического анализа статистических критериев;
- навыками применения теорем математической статистики в прикладных задачах физики и экономики.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	F-критерий для проверки линейных гипотез в гауссовской линейной модели.	4	4		14
2	Вероятностно-статистическая модель.	4	4		14
3	Основная задача математической статистики.	4	4		15
4	Различные виды сходимостей случайных векторов.	6	6		15
5	Статистики и оценки.	6	6		15
6	Эмпирическое распределение и эмпирическая функция распределения.	6	6		11

Итого часов	30	30		84
Подготовка к экзамену	0 час.			
Общая трудоёмкость	144 час., 4 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 4 (Весенний)

1. F-критерий для проверки линейных гипотез в гауссовской линейной модели.
F-критерий для проверки линейных гипотез в гауссовской линейной модели. Пример с двумя гауссовскими выборками, отличающимися сдвигом: проверка гипотезы об их однородности.

2. Вероятностно-статистическая модель.

Вероятностно-статистическая модель. Понятия наблюдения и выборки. Параметрическая статистическая модель. Моделирование выборки из неизвестного распределения, принадлежащему параметрическому семейству.

3. Основная задача математической статистики.

Основная задача математической статистики. Примеры: выборка и линейная модель.

4. Различные виды сходимостей случайных векторов.

Различные виды сходимостей случайных векторов: с вероятностью 1, по вероятности, по распределению. Три знаменитых теоремы: закон больших чисел, усиленный закон больших чисел, центральная предельная теорема. Теорема о наследовании сходимости и лемма Слуцкого. Пример применения леммы Слуцкого.

5. Статистики и оценки.

Статистики и оценки. Примеры статистик: выборочные характеристики, порядковые статистики. Основные свойства оценок: несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность. Примеры. Наследование состоятельности и сильной состоятельности при взятии непрерывной функции. Лемма о наследовании асимптотической нормальности.

6. Эмпирическое распределение и эмпирическая функция распределения.

Эмпирическое распределение и эмпирическая функция распределения. Обоснованность основной задачи математической статистики и теорема Гливенко-Кантелли.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Стандартная учебная аудитория, оборудованная проектором и интерактивной доской.

6. Перечень рекомендуемой литературы

Основная литература

1. Математическая статистика [Текст] : [учебник для вузов] / А. А. Боровков .— [3-е изд., испр.] .— М. : Физматлит, 2007 .— 704 с.
2. Введение в математическую статистику [Текст] : [учебник для вузов] / Г. И. Ивченко, Ю. И. Медведев .— М. : ЛКИ, 2010, 2014, 2015 .— 600 с.
3. Теория вероятностей и математическая статистика [Текст] : учеб. пособие для вузов / П. П. Бочаров, А. В. Печинкин .— М. : Физматлит, 2005 .— 295 с. : ил. + pdf-версия. - Библиогр.: с. 292. - ISBN 5-9221-0633-3. — Полный текст (Доступ из сети МФТИ / Удаленный доступ).

Дополнительная литература

1. Наглядная математическая статистика [Текст] : учеб. пособие для вузов / М. Б. Лагутин .— 2-е изд., испр. — М. : Бином. Лаб. знаний, 2009 .— 472 с.

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Не используются

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

На лекционных занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

В процессе самостоятельной работы обучающихся возможно использование таких программных средств, как Mathcad, MATLAB, Maple и др.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

1. Рекомендуется успешно сдавать контрольные работы, так как это упрощает итоговую аттестацию по предмету.

2. Для подготовки к итоговой аттестации по предмету лучше всего пользоваться материалами лекций.

ПРИЛОЖЕНИЕ

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Прикладная математика и информатика
профиль подготовки: Искусственный интеллект и большие данные
Сетевое обучение
кафедра дискретной математики
курс: 2
квалификация: бакалавр

Семестр, формы промежуточной аттестации: 4 (весенний) - Дифференцированный зачет

Разработчик: И.В. Родионов, канд. физ.-мат. наук, доцент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения	ОПК-2.1. Использует знания основных положений и концепций в области программирования, архитектуру языков программирования, основную терминологию и базовые алгоритмы, основные требования информационной безопасности для практического применения
	ОПК-2.2. Анализирует типовые языки программирования, составляет программы
	ОПК-2.3 Применяет на практике опыт решения задач с использованием базовых алгоритмов, анализа типов коммуникаций и интеграции различных типов программного обеспечения

2. Показатели оценивания компетенций

В результате изучения дисциплины «Математическая статистика» обучающийся должен:

знать:

- основные понятия математической статистики;
- основные подходы к сравнению оценок параметров неизвестного распределения;
- асимптотические и неасимптотические свойства оценок параметров неизвестного распределения;
- основные методы построения оценок с хорошими асимптотическими свойствами: метод моментов, метод максимального правдоподобия, метод выборочных квантилей;
- понятие эффективных оценок и неравенство информации Рао-Крамера;
- определение и главные свойства условного математического ожидания случайной величины относительно сигма-алгебры или другой случайной величины;
- определение общей линейной регрессионной модели и метод наименьших квадратов;
- многомерное нормальное распределение и его основные свойства;
- базовые понятия теории проверки статистических гипотез;
- лемму Неймана – Пирсона и теорему о монотонном отношении правдоподобия;
- критерий хи-квадрат Пирсона для проверки простых гипотез в схеме Бернулли.

уметь:

- обосновывать асимптотические свойства оценок с помощью применения предельных теорем теории вероятностей;
- строить оценки с хорошими асимптотическими свойствами для параметров неизвестного распределения по заданной выборке из него;
- находить байесовские оценки по заданному априорному распределению;
- вычислять условные математические ожидания с помощью условных распределений;
- находить оптимальные оценки с помощью полных достаточных статистик;
- строить точные и асимптотические доверительные интервалы, и области для параметров неизвестного распределения;
- находить оптимальные оценки и доверительные области в гауссовской линейной модели;
- строить равномерно наиболее мощные критерии в случае параметрического семейства с монотонным отношением правдоподобия;
- строить F-критерий для проверки линейных гипотез в линейной гауссовской модели.

владеть:

- основными методами математической статистики построения точечных и доверительных оценок: методом моментов, выборочных квантилей, максимального правдоподобия, методом наименьших квадратов, методом центральной статистики.
- навыками асимптотического анализа статистических критериев;
- навыками применения теорем математической статистики в прикладных задачах физики и экономики.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Примеры домашних заданий:

1. Найдите оптимальную оценку параметра $\theta > 0$ по выборке из распределения: а) $N(\theta, 1)$, б) $R(0, \theta)$, в) $Pois(\theta)$, г) $Bin(1, \theta)$ (здесь $(0,1)$). 6. Пусть X_1, \dots, X_n выборка из равномерного распределения на отрезке $[0, \theta]$, $\theta > 0$. Постройте доверительный интервал для уровня доверия γ , используя статистику а) X , б) $X(1)$, в) $X(n)$.
2. Пусть X_1, \dots, X_n выборка из нормального распределения с параметрами $(\theta, 1)$. Найдите байесовскую оценку параметра θ , если априорное распределение есть $Bin(1, p)$. Будет ли полученная оценка состоятельной оценкой параметра θ ?
3. Имеется 2 объекта с весами a и b . Мы взвесили с ошибками первый, второй и оба объекта вместе, причём дисперсия ошибки в последнем случае была в 4 раза больше. Свести задачу к линейной регрессионной модели и найти оценки наименьших квадратов для a и b .
4. X_1, \dots, X_n выборка из экспоненциального распределения с параметром θ . Постройте равномерно наиболее мощный критерий уровня значимости проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы а) $H_1: \theta > \theta_0$, б) $H_1: \theta < \theta_0$.

Контрольная No.1

1. $X_1 \dots X_n$ -- выборка из $U[0, \theta]$. Для какой величины $\tau(\theta)$ статистика $\sum_{i=1}^n X_i$ является асимптотически нормальной оценкой? Вычислите асимптотическую дисперсию данной оценки.
2. $X_1 \dots X_n$ -- выборка из дискретного распределения $P(X_1 = -1) = (1 - \theta)/(1 + \theta)$; $P(X_1 = 2) = 2\theta/(1 + \theta)$. Найдите оценки параметра θ по методу моментов и по методу максимального правдоподобия.
3. $X_1 \dots X_n$ --- выборка из распределения с плотностью $f(x) = \theta^2 e^{-\theta x} I(x \in [0, +\infty))$. Для каких функций $\tau(\theta)$ существует эффективная оценка? Вычислите информацию Фишера $i(\theta)$ одного элемента выборки.

Контрольная No.2

1. Пусть $X_1 \dots X_n$ -- выборка из экспоненциального распределения со сдвигом, т.е. имеющего плотность $p\theta(x) = \exp\{-(x - \theta)\} I\{x > \theta\}$. Найдите полную достаточную статистику в данной модели.
2. Пусть $X_1 \dots X_n$ -- выборка из распределения с плотностью $p\theta(x) = (4\theta^4)/(x^5) I\{x \geq \theta\}$. С помощью статистики $X(n)$ постройте точный доверительный интервал уровня доверия γ параметра θ .
3. Взвешивание двух грузов массами a и b на одних и тех же весах производится следующим образом: n раз взвешивается первый груз, m раз взвешивается второй груз и k раз взвешиваются эти два груза вместе. В предположении, что все ошибки измерения имеют распределение $N(0; \sigma^2)$, постройте доверительные интервалы уровня доверия для a и для b при неизвестном параметре σ^2 .
4. На участке сажают липы, березы, дубы и осины. На каком уровне значимости можно принять гипотезу о том, что дубы и осины сажали с одинаковой вероятностью, липы и березы -- с одинаковой вероятностью, а березы -- с вероятностью в 3 раза большей, чем дубы? Было посажено 50 лип, 65 берез, 17 дубов, 28 осин. Квантили распределения χ^2_3 : $z_{\{0,8\}} = 4,6416$; $z_{\{0,9\}} = 6,2514$; $z_{\{0,95\}} = 7,8147$; $z_{\{0,975\}} = 9,3484$.

Темы рефератов

- Статистический критерий.
- Критерий Неймана-Пирсона.
- Построение оптимального критерия Неймана-Пирсона для параметра (математического ожидания) нормального закона распределения с известной дисперсией.
- Определение минимального объема выборки.
- Сложные параметрические гипотезы. Построение критерия для проверки сложных параметрических гипотез.
- Проверка гипотез о математическом ожидании.
- Проверка гипотез о равенстве двух выборочных средних (о равенстве математических ожиданий).
- Проверка гипотезы о равенстве дисперсий двух совокупностей.
- Критерий согласия (критерий Пирсона).
- Некоторые распределения случайных величин.

Правила оценивания курса следующие:

1. Предварительное количество задач, с которым студент выходит на экзамен, рассчитывается по формуле $Q = 5 \min(K_1 + K_2 + S, 5)$, где $K_1, K_2, S \in \{0, 1, 2\}$ количество баллов, полученных за первую контрольную, вторую контрольную и работу на семинарах соответственно.
2. На экзамене студенту выдается Q задач, на решение каждой из которых отводится 20 минут. По истечении этого времени студенту выдается билет. Оценка за курс выставляется по формуле $\max(T/2R, 2)$, где T оценка за ответ билета по 10-балльной шкале, R , $0 < R \leq Q$ количество задач, не решенных студентом. Семинаристам рекомендуется ставить студенту 2 балла за работу на семинарах, если он решит в течение семестра не менее 40 задач из домашних заданий, 1 балл если решено не менее 20, но менее 40 задач. Семинаристу рекомендуется ставить 2 балла за решение контрольной, если студент решит не менее 2 задач, 1 балл если студент решит не менее 1 задачи, но менее 2 задач.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Вопросы:

1. Виды сходимостей случайных векторов: с вероятностью 1, по вероятности, по распределению. Взаимоотношения между различными видами сходимостей. Усиленный закон больших чисел для случайных векторов. Многомерная центральная предельная теорема.
2. Теорема о наследовании сходимости и лемма Слуцкого. Пример их применения.
3. Гауссовские случайные векторы (многомерное нормальное распределение). Теорема о трех эквивалентных определениях. Смысл параметров гауссовского вектора.

4. Основные свойства гауссовских векторов: линейные преобразования и критерий независимости компонент. Теорема об ортогональном разложении гауссовского вектора.
5. Вероятностно-статистическая модель, параметрическая модель. Выборка, эмпирическое распределение. Теорема Гливленко-Кантелли.
6. Статистики и оценки. Несмещенность, состоятельность, сильная состоятельность и асимптотическая нормальность. Лемма о наследовании асимптотической нормальности.
7. Оценка параметра по методу подстановки. Оценка параметра по методу моментов. Теорема о состоятельности оценки метода моментов.
8. Квантили распределения, выборочные квантили, выборочная медиана. Теорема об асимптотической нормальности выборочной p -квантили.
9. Функция потерь и функция риска. Равномерный подход к сравнению оценок. Байесовский, минимаксный и асимптотический подходы к сравнению оценок.
10. Информация Фишера и вклад наблюдения. Неравенство Рао-Крамера. Эффективные оценки и критерий эффективности.
11. Понятие плотности для дискретной случайной величины. Доминируемое семейство распределений. Функция правдоподобия и оценка максимального правдоподобия. Теорема об экстремальном свойстве правдоподобия.
12. Теорема о существовании состоятельного решения уравнения правдоподобия. Состоятельность оценки максимального правдоподобия. Теорема об асимптотической нормальности решения уравнения правдоподобия.
13. Теорема Бахадура. Асимптотически эффективные оценки. Асимптотическая эффективность и эффективность оценки максимального правдоподобия.
14. Условное математическое ожидание случайной величины относительно сигма-алгебры. Заряд на вероятностном пространстве. Теорема Радона-Никодима.
15. Свойства условного математического ожидания (9 штук).
16. Условное математическое ожидание. Условное распределение и условная плотность одной случайной величины относительно другой. Теорема о вычислении условного математического ожидания с помощью условной плотности. Теорема о достаточном условии существования условной плотности.
17. Достаточные статистики. Критерий факторизации Неймана-Фишера. Теорема Колмогорова-Блекуэлла-Рао и следствие из нее.
18. Полные статистики. Теорема об оптимальной оценке. Экспоненциальное семейство распределений. Теорема о полной достаточной статистике в экспоненциальном семействе.
19. Доверительные интервалы и доверительные области. Понятие центральной статистики и метод построения доверительных областей с ее помощью. Асимптотические доверительные интервалы.
20. Линейная регрессионная модель. Оценка наименьших квадратов, формула ее вычисления. Несмещенные оценки для параметров линейной регрессионной модели.
21. Проверка статистических гипотез: гипотеза и альтернатива, критерий проверки гипотезы, ошибки первого и второго родов, уровень значимости и размер критерия, функция мощности. Несмещенность и состоятельность статистического критерия.
22. Сравнения критериев, равномерно наиболее мощные критерии. Лемма Неймана-Пирсона о проверке простых гипотез. Следствие из нее. Параметрические семейства распределений с монотонным отношением правдоподобия. Теорема о семействе с монотонным отношением правдоподобия.
23. Линейная гауссовская модель, достаточная статистика в линейной гауссовской модели. Линейные гипотезы, F -критерий для проверки линейной гипотезы в гауссовской линейной модели.
24. Статистика хи-квадрат Пирсона в схеме Бернулли с m исходами. Теорема Пирсона. Критерий согласия Пирсона.

Примеры билетов:

Билет №1

1. Виды сходимостей случайных векторов: с вероятностью 1, по вероятности, по распределению.

2. Основные свойства гауссовских векторов: линейные преобразования и критерий независимости компонент. Теорема об ортогональном разложении гауссовского вектора.

Билет №2

1. Вероятностно-статистическая модель, параметрическая модель.

2. Оценка параметра по методу подстановки. Оценка параметра по методу моментов. Теорема о состоятельности оценки метода моментов.

Критерии оценивания

- оценка «отлично (10)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений
- оценка «хорошо (7)» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (6)» выставляется студенту, если он знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (5)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «удовлетворительно (4)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины.