

**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:	Биотехнология
профиль подготовки:	Системная и синтетическая биология Физтех-школа Биологической и Медицинской Физики кафедра дискретной математики
курс:	3
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час.

Всего часов: 45, всего зач. ед.: 1

Программу составил: А.М. Райгородский, д-р физ.-мат. наук, заведующий кафедрой

Программа обсуждена на заседании кафедры дискретной математики 04.06.2020

Аннотация

Курс математической статистики 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-5 Способен участвовать в проведении фундаментальных и прикладных исследований и разработок, самостоятельно осваивать новые теоретические, в том числе, математические методы исследований	ОПК-5.1 Способен решать поставленные задачи в области теоретических и экспериментальных исследований и разработок

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	Вероятностно-статистическая модель.	4			
2	Основная задача математической статистики.	4			3
3	Различные виды сходимостей случайных векторов.	6			6
4	Статистики и оценки.	6			
5	Эмпирическое распределение и эмпирическая функция распределения.	6			6
6	F-критерий для проверки линейных гипотез в гауссовской линейной модели.	4			
Итого часов		30			15
Подготовка к экзамену		0 час.			
Общая трудоёмкость		45 час., 1 зач.ед.			

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

Семестр: 6 (Весенний)

1. Вероятностно-статистическая модель.

Примеры несмещенных и состоятельных оценок (моменты, дисперсия); смещенных, но состоятельных оценок; несостоятельных, но несмещенных оценок. Оценки функций от параметров. Пример ситуации, в которой не существует несмещенной оценки некоторой функции от параметра.

2. Основная задача математической статистики.

Байесовская и минимаксная стратегии. Минимаксность байесовской стратегии с постоянным риском.

3. Различные виды сходимостей случайных векторов.

Теоремы об асимптотической нормальности выборочного среднего и медианы в модели симметричного распределения с неизвестным параметром сдвига.

4. Статистики и оценки.

Напоминание правила трех сигм и пояснения в терминах этого правила. Пример со «смешанным» нормальным распределением (медиана vs. выборочное среднее).

5. Эмпирическое распределение и эмпирическая функция распределения.

Оценки максимального правдоподобия (о.м.п.) и их свойства (состоятельность, асимптотическая нормальность и эффективность). О.м.п. для параметра сдвига в распределении Лапласа как пример асимптотически нормальной о.м.п. в нерегулярной модели.

6. F-критерий для проверки линейных гипотез в гауссовской линейной модели.

Линейная гауссовская модель и метод наименьших квадратов. Основные свойства многомерного гауссовского закона. Распределение (центральное) хи-квадрат Пирсона и лемма о круговом гауссовском распределении. Линейная гауссовская модель, полные достаточные статистики

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

Стандартная учебная аудитория.

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:	Биотехнология
профиль подготовки:	Системная и синтетическая биология Физтех-школа Биологической и Медицинской Физики кафедра дискретной математики
курс:	3
квалификация:	бакалавр
Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет	
Разработчик:	А.М. Райгородский, д-р физ.-мат. наук, заведующий кафедрой

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-5 Способен участвовать в проведении фундаментальных и прикладных исследований и разработок, самостоятельно осваивать новые теоретические, в том числе, математические методы исследований	ОПК-5.1 Способен решать поставленные задачи в области теоретических и экспериментальных исследований и разработок

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 (μ, σ^2) . Найдите байесовскую оценку параметра μ , если априорное распределение есть $\text{Bin}(1, p)$. Будет ли полученная оценка состоятельной оценкой параметра?
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 λ . Постройте равномерно наиболее мощный критерий уровня значимости проверки гипотезы $H_0: \lambda = 0$ против альтернативы а) $H_1: \lambda > 0$, б) $H_1: \lambda < 0$.

Контрольная No.1

1. $X_1 \dots X_n$ -- выборка из $U[0, \theta]$. Для какой величины $\tau(\theta)$ статистика $X_e - X$ 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 $f(x) = \theta^2 e^{-\theta x} I(\in [0, 2; +\infty])$. 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 $\chi^2_3: z_{\{0,8\}} = 4,6416; z_{\{0,9\}} = 6,2514; z_{\{0,95\}} = 7,8147; z_{\{0,975\}} = 9,3484$.

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

1. Основная задача математической статистики. Параметрическая статистика. Оценки («статистики»). Вариационный ряд и порядковые статистики. Напоминание о распределении порядковых статистик.
2. Несмещенность и состоятельность оценок. Примеры несмещенных и состоятельных оценок (моменты, дисперсия); смещенных, но состоятельных оценок; несостоятельных, но несмещенных оценок. Оценки функций от параметров. Пример ситуации, в которой не существует несмещенной оценки некоторой функции от параметра.
3. Эмпирическая функция распределения. Ее несмещенность и состоятельность по отношению к теоретической функции распределения.
4. Функция штрафа (потерь) и функция риска. Байесовская и минимаксная стратегии. Минимаксность байесовской стратегии с постоянным риском.
5. Байесовская оценка в схеме Бернулли. Минимаксная оценка в схеме Бернулли (статистика Ходжеса – Лемана). Сравнение статистики Ходжеса – Лемана с выборочным средним.
6. Асимптотическая нормальность оценки. Теоремы об асимптотической нормальности выборочного среднего и медианы в модели симметричного распределения с неизвестным параметром сдвига.

7. Относительная асимптотическая эффективность. Напоминание правила трех сигм и пояснения в терминах этого правила. Пример со «смешанным» нормальным распределением (медиана vs. выборочное среднее).
8. Условия регулярности и неравенство Рао – Крамера (случай несмещенной оценки). Информация Фишера. Эффективные и сверхэффективные оценки.
9. Метод моментов. Состоятельность оценки.
10. Метод максимального правдоподобия. Оценки максимального правдоподобия (о.м.п.) и их свойства (состоятельность, асимптотическая нормальность и эффективность). О.м.п. для параметра сдвига в распределении Лапласа как пример асимптотически нормальной о.м.п. в нерегулярной модели.
11. О.м.п. в стандартных моделях (биномиальная, нормальная, Коши и др.).
12. Достаточные статистики в дискретном случае. Критерий факторизации. Примеры.
13. Условное математическое ожидание относительно разбиения. Условные плотности. Свойства условного математического ожидания (б/д в «непрерывном» случае, но с доказательством в случае условного математического ожидания относительно разбиения).
14. Достаточные статистики в общем случае: общий критерий факторизации как определение. Примеры.
15. Экспоненциальное семейство. Пессимистическая теорема (б/д).
16. Теорема Колмогорова – Блекуэлла – Рао об улучшении несмещенной оценки.
17. Полные статистики. Критерий полноты и достаточности для экспоненциального семейства (б/д).
18. Теорема Гливленко – Кантелли (формулировка). Общая задача о равномерной сходимости в законах больших чисел. Задача о треугольниках на плоскости. Размерность Вапника – Червоненкиса. Пример с пространством и открытыми полупространствами. Две теоремы Вапника – Червоненкиса (б/д). Теорема Гливленко – Кантелли как частный случай теоремы Вапника – Червоненкиса. Теорема о треугольниках как частный случай теоремы Вапника – Червоненкиса.
19. Статистика Колмогорова и ее независимость от вида непрерывной функции распределения (только строго монотонный случай). Асимптотическая теорема Колмогорова (б/д).
20. Точные и асимптотические доверительные интервалы. Односторонние и двусторонние интервалы.
21. Квантили распределений. Напоминание распределения хи-квадрат. Распределение Стьюдента.
22. Построение доверительных интервалов для параметров нормального распределения: для среднего при известной дисперсии, для дисперсии при известном среднем и для одновременно неизвестных среднего и дисперсии.
23. Доверительные интервалы для квантилей.
24. Критерий согласия Колмогорова.
25. Критерий согласия хи-квадрат Пирсона.
26. Сравнение двух простых гипотез. Ошибки первого и второго рода.
27. Случай абсолютно непрерывных распределений: отношение правдоподобия и теорема Неймана – Пирсона.
28. Критерий однородности Смирнова для двух независимых выборок.
29. Критерий однородности хи-квадрат для нескольких независимых выборок.
30. Проверка равенства средних у двух нормальных выборок с известными дисперсиями; с неизвестными дисперсиями. Проверка равенства дисперсий у двух нормальных выборок с неизвестными средними. Распределение Фишера.

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 всесторонне 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 зачета обучающиеся могут пользоваться программой дисциплины.