

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ:

Директор



А.В. Замятин

20 22 г.

Рабочая программа дисциплины

Математическая статистика

по направлению подготовки

02.03.02 Фундаментальная информатика и информационные технологии

Направленность (профиль) подготовки:

Искусственный интеллект и разработка программных продуктов

Форма обучения

Очная

Квалификация

Бакалавр

Год приема

2022

Код дисциплины в учебном плане: Б1.О.02.11

СОГЛАСОВАНО:

Руководитель ОП

А.В. Замятин

Председатель УМК

С.П. Сущенко

Томск – 2022

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

– ОПК-1 – Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.3. Обладает необходимыми знаниями для исследования информационных систем и их компонент.

ИОПК-1.2. Использует фундаментальные знания, полученные в области математических и (или) естественных наук в профессиональной деятельности.

ИОПК-1.1. Применяет фундаментальные знания, полученные в области математических и (или) естественных наук.

2. Задачи освоения дисциплины

– Освоить аппарат математической статистики, ее методы, основные понятия и задачи.

– Научиться применять методы математической статистики для решения практических задач профессиональной деятельности.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к обязательной части образовательной программы. Дисциплина входит в модуль "Математика".

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Пятый семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Математический анализ», «Линейная алгебра и аналитическая геометрия», «Теория вероятностей».

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 32 ч.

-практические занятия: 16 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Элементы выборочной теории

Случайная выборка. Статистика. Порядковые статистики, вариационный ряд. Эмпирическая функция распределения, ее статистические свойства. Теорема Гливленко и теорема Колмогорова для эмпирической функции распределения. Эмпирическая плотность распределения: гистограмма, полигон частот, ядерная оценка плотности.

Эмпирическая функция распределения и ядерная оценка плотности в случае многомерной выборки.

Тема 2. Выборочные характеристики

Примеры выборочных характеристик (выборочное среднее, выборочная дисперсия, выборочные квантили, выборочный коэффициент корреляции и др.). Выборочные характеристики как функционалы от эмпирической функции распределения. Два типа статистик. Теоремы непрерывности для функций от выборочных моментов. Асимптотическая нормальность выборочных моментов.

Тема 3. Точечное оценивание параметров распределения

Точечное оценивание. Статистические оценки и общие требования к ним. Состоятельность, асимптотическая нормальность. Несмещенные оценки с минимальной дисперсией. Оптимальные оценки. Понятие функции правдоподобия, вклада выборки, функции информации Фишера. Неравенство Рао-Крамера и эффективные оценки. Экспоненциальная модель. Неравенство Рао-Крамера в случае векторного параметра. Достаточные статистики и оптимальные оценки. Критерий факторизации.

Тема 4. Методы точечной оценки параметров распределений

Метод подстановки. Метод моментов. Метод максимального правдоподобия (ММП). Принцип инвариантности для ММП, асимптотические свойства оценок ММП. Мультиномиальные оценки максимального правдоподобия. Байесовский и минимаксный подходы к оцениванию параметров.

Тема 5. Интервальное оценивание

Понятие доверительного интервала. Построение доверительных интервалов для математического ожидания случайной величины в случаях известной и неизвестной дисперсии. Доверительные интервалы для неизвестной дисперсии при неизвестных других параметрах. Доверительные интервалы для среднего и дисперсии в случае нормальной модели. Доверительные области для многомерного параметра.

Тема 6. Проверка статистических гипотез

Понятие статистической гипотезы и статистического критерия. Простые и сложные гипотезы. Общий принцип построения статистических критериев и их характеристики. Статистика критерия, критическая область критерия. Уровень значимости, функция мощности и мощность критерия. Несмещенные и состоятельные критерии. Критерии Проверка гипотез о виде распределения. Критерии согласия Колмогорова и хи-квадрат К. Пирсона для простых гипотез, критерий согласия хи-квадрат для сложной гипотезы. Гипотеза однородности. Критерии однородности Смирнова, хи-квадрат, Манна-Уитни. Гипотеза независимости. Критерий независимости хи-квадрат. Критерий Спирмена.

Тема 7. Параметрические гипотезы

Понятие параметрической гипотезы. Общий принцип выбора критической области. Вероятности ошибок первого и второго родов. Равномерно наиболее мощные критерии.

Критерий Неймана-Пирсона для проверки двух простых гипотез. Сложные гипотезы. Критерий отношения правдоподобия проверки общих гипотез. Байесовское решающее правило.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, проведения контрольных работ, тестов по лекционному материалу, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в пятом семестре проводится в письменной форме по билетам. Экзаменационный билет состоит из пяти вопросов. Продолжительность экзамена 1,5 часа.

Первый вопрос представляет собой теоретическую часть курса, излагается в развернутой форме, проверяет ИОПК-1.1.

Ответы на вопросы со второго по пятый предполагают решение задач в развернутой форме, краткую интерпретацию полученных результатов и проверяют ИОПК-1.2 и ИОПК-1.3.

Перечень теоретических вопросов

1. В чем отличие теории вероятностей от математической статистики. Задачи математической статистики;
2. Случайная выборка, статистика, порядковая статистика;
3. Эмпирическая функция распределения (э.ф.р.) для одномерной случайной величины;
4. Статистические свойства э.ф.р. при конечном объеме выборки;
5. Предельные теоремы для э.ф.р.
6. Гистограмма и полигон частот;
7. Ядерная оценка плотности для одномерной случайной величины;
8. Асимптотические свойства ядерной оценки плотности;
9. Э.ф.р. и ядерная оценка плотности для многомерной выборки
10. Выборочное среднее, статистические свойства;
11. Выборочная дисперсия, подправленная выборочная дисперсия;
12. Выборочная квантиль порядка p ;
13. Выборочный коэффициент корреляции;
14. Асимптотическая нормальность выборочных моментов;
15. Теорема непрерывности для функций от выборочных моментов;
16. Точечная оценка, примеры;
17. Состоятельные оценки, несмещенные оценки;
18. Оптимальные оценки;
19. Функция правдоподобия;
20. Функция вклада выборки;
21. Функция информации Фишера;
22. Что показывает неравенство Рао-Крамера;
23. Эффективные оценки, критерий эффективности;
24. Достаточные статистики;
25. Критерий факторизации;
26. Метод подстановки оценивания параметров;
27. Метод моментов в оценке параметров распределения;
28. Метод максимального правдоподобия оценивания параметров;
29. Метод мультиномиального правдоподобия оценивания параметров;
30. Байесовское оценивание параметров;
31. Минимаксное оценивание параметров;
32. Доверительный интервал;
33. Доверительные области;
34. Статистическая гипотеза и статистический критерий;
35. Критическая область критерия;
36. Уровень значимости критерия;
37. Мощность критерия;
38. Несмещенность критерия;
39. Состоятельность критерия;
40. Гипотеза о виде распределения, критерии согласия;
41. Гипотеза однородности, критерии однородности;

42. Гипотеза независимости, критерии независимости;
43. Параметрические гипотезы. Общий принцип выбора критической области.
44. Вероятности ошибок первого и второго родов.
45. Равномерно наиболее мощные критерии.
46. Критерий Неймана-Пирсона для проверки двух простых гипотез.
47. Сложные гипотезы. Критерий отношения правдоподобия проверки общих гипотез.
48. Байесовское решающее правило.

Примеры задач:

Статистическое оценивание.

Задача 1. По реализации выборки X_1, \dots, X_n построить оценку методом максимального правдоподобия для параметра экспоненциального распределения.

Задача 2. По реализации выборки X_1, \dots, X_n построить оценку методом максимального правдоподобия для параметра закона Пуассона.

Задача 3. По реализации выборки X_1, \dots, X_n построить оценки методом максимального правдоподобия для параметров нормального распределения.

Задача 4. Построить доверительный интервал для математического ожидания по реализации выборки $-1.25, 0.11, 2.37, 3.45$ из нормального закона с дисперсией, равной $0,49$. Принять $\gamma=0.96$.

Задача 5. Построить доверительный интервал для математического ожидания по реализации выборки $-1.62, .54, 2.12, 3.72$ из нормального закона с неизвестной дисперсией. Принять $\gamma=0.98$.

Задача 6. Построить доверительный интервал для математического ожидания случайной величины X с дисперсией, равной 4 , при выборке объема $n=100$ и выборочному среднему равному 10 . Принять $\gamma=0.97$.

Задача 7. Подсчитайте ранговый коэффициент корреляции Спирмена между двумя случайными величинами X и Y по следующим данным:

$$\begin{aligned} X_1=1,5, \quad X_2=2, \quad X_3=4, \quad X_4=1, \quad X_5=3 \\ Y_1=3, \quad Y_2=2, \quad Y_3=2,1, \quad Y_4=1, \quad Y_5=4 \end{aligned}$$

Задача 8. Найти достаточную статистику для: параметра распределения Пуассона, для параметров равномерного в $[a,b]$ распределения, для параметров нормального распределения.

Проверка статистических гипотез

Задача 9. При 65 подбрасываниях монеты герб появился 25 раз. Можно ли считать монету симметричной? Принять уровень значимости $\alpha=0.10$.

Задача 10. При 160 подбрасываниях игральной кости шестерка выпала 25 раз. Можно ли считать кость правильной? Принять $\alpha=0.05$.

Задача 11. При 120 подбрасываниях игральной кости пятерка выпала 25 раз, а шестерка 15 раз. Можно ли считать кость правильной? Принять $\alpha=0.01$.

Задача 12. Можно ли считать два потока абитуриентов однородными, если итоги экзамена по математике на каждом потоке оказались следующими:

1-й поток: баллы «2», «3», «4» и «5» получили соответственно $45, 40, 70$ и 35 человек;
2-й поток: баллы «2», «3», «4» и «5» получили соответственно $40, 35, 65$ и 30 человек.
Уровень значимости $\alpha=0,05$.

Задача 13. Комплектующие изделия одного наименования поступают с трех предприятий A, B , и C . Результаты проверки изделий следующие. Предприятие A : годные – 30 , негодные – 2 , предприятие B : годные – 38 , негодные – 3 , предприятие C : годные – 54 , негодные – 7 . Можно ли считать, что качество изделий не зависит от поставщика? Принять уровень значимости $\alpha=0,1$.

Задача 14. По реализации выборки $-1.56, 0.22, 2.34, 3.75$ из нормального закона с дисперсией, равной $0,49$, и неизвестным математическим ожиданием a проверить гипотезы $H_0: a=1.2$ и $H_1: a=2$. Принять уровень значимости $\alpha=0,01$.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Экзаменационный билет состоит из пяти заданий (вопросов).

Если студент ответил на пять вопросов, то ставится оценка «отлично», на четыре вопроса – оценка «хорошо», на три вопроса – оценка «удовлетворительно», на два вопроса – оценка «неудовлетворительно».

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=00000>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине (Приложение 1).

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Боровков А. А. Математическая статистика: учебник / А. А. Боровков. - Изд. 4-е, стер. - Санкт-Петербург [и др.]: Лань, 2010. - 703 с.: табл.- URL: http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=3810.

– Ивченко Г. И. Введение в математическую статистику / Г. И. Ивченко, Ю. И. Медведев. - Москва: Изд-во ЛКИ, сор. 2010. - 599 с.

б) дополнительная литература:

– Бородин А. Н. Элементарный курс теории вероятностей и математической статистики: [учебное пособие для вузов по нематематическим специальностям] / А. Н. Бородин. - Изд. 8-е, стер. - Санкт-Петербург [и др.]: Лань, 2016. - 254 с.

– Шуленин В. П. Математическая статистика: [учебник] / В. П. Шуленин ; Том. гос. ун-т. – Томск : Изд-во НТЛ, 2012. – Ч. 1. – 539 с. – URL:

<http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000434962> ; Ч. 2. – 387 с. – URL:

<http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000434963> ; Ч. 3. – 518 с. – URL:

<http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000440075>

- URL: http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=2026.

– Ватугин В. А. Теория вероятностей и математическая статистика в задачах: учеб. пособие для вузов / В. А. Ватугин, Г. И. Ивченко, Ю. И. Медведев, В. П. Чистяков. - М.: Дрофа, 2005. - 315 с.

– Ивченко Г.И., Медведев Ю.И., Чистяков А.В. Задачи с решениями по математической статистике – М.: Дрофа, 2007.

в) ресурсы сети Интернет:

– открытые онлайн-курсы

– Журнал «Эксперт» - <http://www.expert.ru>

– Официальный сайт Федеральной службы государственной статистики РФ - www.gsk.ru

– Официальный сайт Всемирного банка - www.worldbank.org

– Общероссийская Сеть КонсультантПлюс Справочная правовая система. <http://www.consultant.ru>

в) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

– ЭБС Лань – <http://e.lanbook.com/>

- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

- Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных:

- Университетская информационная система РОССИЯ – <https://uisrussia.msu.ru/>
- Единая межведомственная информационно-статистическая система (ЕМИСС) – <https://www.fedstat.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Дмитриев Юрий Глебович, д-р.физ.-мат.наук, доцент, профессор кафедры системного анализа и математического моделирования ИПМКН ТГУ.