

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

Институт прикладной математики и компьютерных наук



Рабочая программа дисциплины

Непараметрические методы анализа данных

по направлению подготовки

01.03.02 Прикладная математика и информатика

Направленность (профиль) подготовки:

Математическое моделирование и информационные системы

Форма обучения

Очная

Квалификация

Бакалавр

Год приема

2023

Код дисциплины в учебном плане: Б1.В.ДВ.01.01.02

СОГЛАСОВАНО:

Руководитель ОП

 Л.А. Нежелская

Председатель УМК

 С.П. Сущенко

Томск – 2023

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.

– ОПК-3 – Способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности.

– ОПК-4. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.

– ПК-2. Способен формализовать и алгоритмизировать поставленную задачу, написать программный код, а также верифицировать работоспособность программного обеспечения и исправить дефекты

– ПК-3. Способен формализовывать, согласовывать и документировать требования к системе и подсистеме, обрабатывать запросы на изменение требований к системе и подсистеме, выявлять и формализовывать риски, анализировать проблемные ситуации.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.1. Демонстрирует навыки работы с учебной литературой по основным естественнонаучным и математическим дисциплинам.

ИОПК-1.2. Демонстрирует навыки выполнения стандартных действий, решения типовых задач с учетом основных понятий и общих закономерностей, формулируемых в рамках базовых математических и естественнонаучных дисциплин.

ИОПК-1.3. Демонстрирует навыки использования основных понятий, фактов, концепций, принципов математики, информатики и естественных наук для решения практических задач, связанных с прикладной математикой и информатикой.

ИОПК-3.1. Демонстрирует навыки применения современного математического аппарата для построения адекватных математических моделей реальных процессов, объектов и систем в своей предметной области.

ИОПК-3.2. Демонстрирует умение собирать и обрабатывать статистические, экспериментальные, теоретические и т.п. данные для построения математических моделей, расчетов и конкретных практических выводов.

ИОПК-3.3. Демонстрирует способность критически переосмысливать накопленный опыт, модифицировать при необходимости вид и характер разрабатываемой математической модели.

ИОПК-4.3. Использует современные информационные технологии на всех этапах решения задач профессиональной деятельности.

ИПК-2.3. Осуществляет работу с системой контроля версий, рефакторинг и оптимизацию программного кода.

ИПК-3.1. Реализовывает построение формализованной математической модели системы (подсистемы), введение целевой функции системы, подсистемы и ограничений, соответствующих требованиям к системе (подсистеме).

ИПК-3.2. Адаптирует формализованную математическую модель системы (подсистемы) к изменению требований (ограничений к целевой функции) к системе (подсистеме).

ИПК-3.3. Выявляет и формализовывает в виде математической модели возникающие при функционировании системы (подсистемы) риски; выявляет и анализирует проблемные ситуации.

2. Задачи освоения дисциплины

– Освоить современные методы обработки статистических данных в условиях непараметрической статистической модели наблюдений при неизвестном распределении вероятности изучаемого случайного явления.

– Научиться применять эффективные методы решения задач синтеза и анализа непараметрических ядерных оценок различных характеристик систем, зависящих от неизвестных распределений наблюдаемых последовательностей с различными типами стохастической зависимости. Особое внимание уделяется исследованию производственных функций и связанных с ними характеристик, проблемам управления производственными процессами, прогнозированию цен акций, паев фондов и т.п.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений. Дисциплина входит в модуль «Прикладная математика».

4. Семестр освоения и форма промежуточной аттестации по дисциплине

Восьмой 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», «Математическая статистика», «Эконометрика», «Макроэкономика», «Микроэкономика», «Экономико-математическое моделирование I».

6. Язык реализации

Русский

7. Объем дисциплины

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

– лекции: 16 ч.

– лабораторные работы: 16 ч.

в том числе практическая подготовка: 16 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение. Общие вопросы

Актуальность развития методов условного непараметрического оценивания для стохастических систем. Типы функционалов: интегральные, условные, характеристические, дополненные функционалы.

Тема 2. Непараметрическое оценивание функционалов распределений

Параметрический и непараметрический подходы к оцениванию статистических характеристик. Эмпирическая функция распределения, её свойства. Статистики как функционалы от эмпирических распределений. Непараметрические ядерные оценки плотности распределения.

Тема 3. Общие теоремы сходимости функций от статистик

Введение. Постановка задачи. Сходимость по распределению первых моментов функций статистик. Среднеквадратическая сходимость оценки подстановки. Моменты отклонений оценки подстановки, ее смещение, дисперсия и СКО.

Тема 4. Ядерные оценки базовых функционалов по независимым выборкам

Построение оценок базовых функционалов и их производных. Сходимость в среднеквадратическом оценок базовых функционалов и их производных. Улучшение

скорости сходимости СКО. Улучшенные оценки полиномиальной схемы и локальной аппроксимации. Теоретическое сравнение оценок: асимптотическая относительная эффективность. Моделирование улучшенных и обычных оценок базовых функционалов. Сходимость по распределению оценок базовых функционалов и их производных.

Тема 5. Идентификация производственных функций

Непараметрическая идентификация двухфакторной производственной функции. Параметрический и непараметрический подходы в регрессионном анализе показателей инвестиционной деятельности. Оценивание предельной продуктивности ресурсов и предельной нормы технического замещения факторов производства. Рекуррентное оценивание функций от функционалов многомерной плотности и их производных. Рекуррентная идентификация двухфакторной производственной функции и ее характеристик. Сравнение рекуррентных и обычных оценок.

Тема 6. Идентификация статистических характеристик по зависимым выборкам

Слабая зависимость. Соотношения между различными коэффициентами слабой зависимости. Примеры последовательностей с перемешиванием. Слабая зависимость и ядерные оценки функционалов. Оценивание функций от функционалов по наблюдениям сильного перемешивания. Непараметрическая идентификация нелинейной авторегрессии и прогнозирование цен акций.

Тема 7. Непараметрические методы идентификации стохастических систем

Применение оценок регрессии, функции чувствительности, остаточной дисперсии при идентификации стохастических систем. Регрессионная модель. Идентификация в широком смысле. Рекуррентный алгоритм обучения в задачах классификации. Обнаружение детерминированного тренда на фоне аддитивной помехи.

Тема 8. Непараметрические методы управления для стохастических систем

Управление выходными параметрами объекта. Синтез алгоритмов управления. Управление технологией конвертерного производства стали. Моделирование алгоритма управления. Непараметрические алгоритмы идентификации и управления для случайных процессов. Дискретно-аналоговые алгоритмы.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости лекций и лабораторных работ, проведения контрольных работ и фиксируется в форме контрольной точки не менее одного раза в семестр. Контрольные работы оцениваются по пятибалльной шкале. К контрольной точке все необходимые лабораторные работы должны быть выполнены и сданы преподавателю.

| № п\п | Темы лабораторных занятий | Объем, час. |
|-------|---|-------------|
| 1. | Моделирование непараметрических оценок ядерного типа. Моделирование непараметрических оценок плотности и регрессии с использованием ядер Валле-Пуссена, Чебышева-Эрмита и тригонометрического ядра. | 4 |
| 2. | Моделирование непараметрических алгоритмов идентификации двухфакторной производственной функции. | 2 |
| 3. | Параметрический и непараметрический подходы в регрессионном анализе показателей инвестиционной деятельности. | 2 |
| 4. | Оценивание предельной продуктивности ресурсов и предельной нормы технического замещения факторов производства. | 2 |
| 5. | Непараметрическая идентификация модели изменения реальных, взятых из открытого доступа Интернет, цен акций. Построение прогнозов цен акций на основе полученной модели, исследование точностных свойств | 4 |

| | | |
|-------------|--|----|
| | алгоритмов. | |
| 6. | Моделирование алгоритмов управления технологией конвертерного производства стали | 2 |
| Всего часов | | 16 |

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в восьмом семестре проводится в письменной форме по билетам. Продолжительность экзамена 1,5 часа. Студент допускается к экзамену, если им выполнены все лабораторные работы.

Типовые экзаменационные билеты имеют следующий вид:

Экзаменационный билет № 1

1. Параметрический и непараметрический подходы к оцениванию статистических характеристик.
2. Непараметрическая идентификация нелинейной авторегрессии и прогнозирование цен акций.

Экзаменационный билет № 2

1. Статистики как функционалы от эмпирических распределений.
2. Рекуррентный алгоритм обучения в задачах классификации.

Примерный перечень теоретических вопросов

1. Актуальность развития методов условного непараметрического оценивания для стохастических систем.
2. Типы функционалов: интегральные, условные, характеризационные, дополненные функционалы.
3. Непараметрическое оценивание функционалов распределений.
4. Параметрический и непараметрический подходы к оцениванию статистических характеристик.
5. Эмпирическая функция распределения, её свойства.
6. Статистики как функционалы от эмпирических распределений.
7. Непараметрические ядерные оценки плотности распределения.
8. Сходимость по распределению первых моментов функций статистик.
9. Среднеквадратическая сходимость оценки подстановки.
10. Моменты отклонений оценки подстановки, её смещение, дисперсия и СКО.
11. Построение оценок базовых функционалов и их производных.
12. Сходимость в среднеквадратическом оценок базовых функционалов и их производных.
13. Улучшение скорости сходимости СКО. Улучшенные оценки полиномиальной схемы и локальной аппроксимации.
14. Теоретическое сравнение оценок: асимптотическая относительная эффективность.
15. Сходимость по распределению оценок базовых функционалов и их производных.
16. Непараметрическая идентификация двухфакторной производственной функции.

17. Параметрический и непараметрический подходы в регрессионном анализе показателей инвестиционной деятельности.
18. Рекуррентное оценивание функций от функционалов многомерной плотности.
19. Рекуррентная идентификация двухфакторной производственной функции и ее характеристик. Сравнение рекуррентных и обычных оценок.
20. Слабая зависимость. Соотношения между различными коэффициентами слабой зависимости. Примеры последовательностей с перемешиванием.
21. Слабая зависимость и ядерные оценки функционалов. Оценивание функций от функционалов по наблюдениям сильного перемешивания.
22. Непараметрическая идентификация нелинейной авторегрессии и прогнозирование цен акций.
23. Применение оценок регрессии, функции чувствительности, остаточной дисперсии при идентификации стохастических систем.
24. Регрессионная модель. Идентификация в широком смысле.
25. Рекуррентный алгоритм обучения в задачах классификации.
26. Обнаружение детерминированного тренда на фоне аддитивной помехи.
27. Управление выходными параметрами объекта. Синтез алгоритмов управления.
28. Управление технологией конвертерного производства стали.
29. Непараметрические алгоритмы идентификации и управления для случайных процессов. Дискретно-аналоговые алгоритмы.

Результаты ответов на вопросы экзамена оцениваются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

| Неудовлетворительно | Удовлетворительно | Хорошо | Отлично |
|---|---|---|--|
| Не ответил ни на один из основных вопросов. | Ответил на один из основных вопросов и на один - два из трех дополнительных вопросов. | Ответил на оба вопроса, содержащихся в экзаменационном билете, и на дополнительные вопросы, но с замечаниями. | Уверенно и правильно ответил на все основные и дополнительные вопросы. |

11. Учебно-методическое обеспечение

а) Кошкин Г.М. Непараметрическая идентификация экономических систем / Учебное пособие. – Томск: Изд-во НТЛ, 2007. – 308 с.

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине предоставляются преподавателями в течение обучения.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

Добровидов А.В., Кошкин Г.М. Непараметрическое оценивание сигналов. – М.: Наука. Физматлит. 1997. – 336 с.

– Васильев В.А., Добровидов А.В., Кошкин Г.М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей. – М.: Наука. Физматлит. 1997. – 336 с.

– Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник – М.: Издательский дом «Дело» РАНХиГС, 2021. – 504 с.

– Кореева Е.Б., Ростова Е.П. Методы социально-экономического прогнозирования. Электронный курс – Самара: Самарский государственный аэрокосмический университет, 2013. – 90 с.

- Кошкин Г.М. Непараметрическая идентификация экономических систем / Учебное пособие. – Томск: Изд-во НТЛ, 2007. – 308 с.
- Дмитриев Ю.Г., Кошкин Г.М., Симахин В.А., Тарасенко Ф.П., Шуленин В.П. Непараметрическое оценивание функционалов по стационарным выборкам. – Томск: Изд-во Том. ун-та, 1974. – 93 с.
- Кошкин Г.М., Пивен И.Г. Непараметрическая идентификация стохастических систем. Научное издание. –Хабаровск: Российская академия наук, Дальневосточное отделение. 2009. – 336 с.

б) дополнительная литература:

- Alexander V. Dobrovidov, Gennady M. Koshkin, Vyacheslav A. Vasiliev. Non-parametric State Space Models. – Heber, UT 84032, USA. Kendrick Press, Inc., 2012. – 501 p.
- А.В. Китаева, Г.М. Кошкин, Н.В. Степанова. Ядерное оценивание функций от условных распределений. – Барнаул: Изд-во Алт. ун-та, 2013. – 122 с.
- Valery Smagin, Gennady Koshkin. Kalman Filtering Discrete Systems with Use of Nonparametric Algorithms for Unknown Input and Parameters. Journal of Information, Control and Management Systems. Vol.13. No. 1. 2015. – P. 67-74.
- Дмитриев Ю.Г., Кошкин Г.М. Использование дополнительной информации при непараметрическом оценивании функционалов плотности. Автоматика и телемеханика. 1987. N 10. С. 47-59.
- Yury G. Dmitriev, Gennady M. Koshkin, Vadim Yu. Lukov. Combined Identification and Prediction Algorithms \ \ Advances in Computer Science Research (ACSR), volume 72, IV International Research Conference "Information Technologies in Science, Management, Social Sphere and Medicine" (ITSMSSM 2017), Atlantis Press, 2017, pp. 244-247.
- Алексеев Е.Р., Чеснокова О.В., Рудченко Е.А. Scilab: Решение инженерных и математических задач. – М.: ALT Linux; БИНОМ, 2008. – 260 с.
- Домбровский В. В. Эконометрика [Электронный ресурс]. – Томск: [б. и.], 2016. – <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000550882>

в) ресурсы сети Интернет:

- открытые онлайн-курсы.

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 Web Apps (Word Excel MS PowerPoint Outlook);
- R-Studio, Mathcad, Matlab;
- публично доступные облачные технологии (Google Docs, Яндекс диск).

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 Орайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

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

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешенном формате («Актру»).

15. Информация о разработчиках

Кошкин Геннадий Михайлович, д-р физ.-мат. наук, профессор, кафедра системного анализа и математического моделирования, профессор;

Андреевская Анна Андреевна, ассистент кафедры системного анализа и математического моделирования.