**Программа утверждена на заседании кафедры теории вероятностей**

**Протокол № от 2017 г.**

**Рабочая программа дисциплины (модуля)**

1. Код и наименование дисциплины (модуля): Избранные главы теории вероятностей.

2. Уровень высшего образования – специалитет.

3. Направление подготовки: 01.05.01 Фундаментальные математика и механика. Специализация: Фундаментальная математика.

4. Место дисциплины (модуля) в структуре ООП: вариативная часть ООП. Является специальной дисциплиной (спецкурсом) для студентов 3-6 годов обучения и аспирантов, специализирующихся в данной научной области или смежной научной области, спецкурсом по выбору студента.

Освоение дисциплины необходимо для последующего изучения дисциплин образовательной программы: курсовая работа, научно-исследовательская практика, преддипломная практика, выпускная квалификационная работа, научно-исследовательская работа.

5. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

Получить знания в современной области теории вероятностей, имеющей разнообразные приложения. Научиться применять изложенную теорию для решения теоретических и прикладных задач.

6. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических или астрономических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся:

Объем дисциплины (модуля) составляет 3 зачетных единицы, всего 108 часа, из которых 44 (46\*) часа составляет контактная работа студента с преподавателем (34 (36\*) часа занятия лекционного типа, 12 часов мероприятия текущего контроля успеваемости и промежуточной аттестации), 64 (62\*) часа составляет самостоятельная работа студента.

*\* - если специальный курс читается в нечетном семестре (продолжительность нечетного семестра 18 недель, четного семестра 17 недель).*

7. Входные требования для освоения дисциплины (модуля), предварительные условия.

Для того чтобы изучение дисциплины было возможно, обучающийся должен

1. освоить следующие дисциплины образовательной программы: математический анализ, линейную алгебру и геометрию, действительный анализ, теорию вероятностей.
2. обладать следующими компетенциями:

Знать: основные результаты изученных разделов математики и основные идеи доказательств этих результатов.

Уметь: решать стандартные задачи математического анализа, линейной алгебры и геометрии, действительного анализа, теории вероятностей и применять идеи, использованные в их решениях, для решения аналогичных и более сложных задач.

Владеть: основными понятиями, результатами и методами из упомянутых разделов математики.

8. Формат обучения.

Очная форма обучения, лекционные занятия.

9. Содержание дисциплины (модуля), структурированное по темам (Перечень тем см. Приложения).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Наименование и краткое содержание разделов и тем дисциплины (модуля),**  **форма промежуточной аттестации по дисциплине (модулю)** | **Всего**  **(часы**) | В том числе | | | | | | | | |
| **Контактная работа (работа во взаимодействии с преподавателем), часы**  из них | | | | | | **Самостоятельная работа обучающегося, часы**  из них | | |
| Занятия лекционного типа | Занятия семинарского типа | Групповые консультации | Индивидуальные консультации | Учебные занятия, направленные на проведение текущего контроля успеваемости, промежуточной аттестации | **Всего** | Выполнение домашних заданий | Подготовка рефератовит.п.. | **Всего** |
| Тема 1 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 2 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 3 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 4 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 5 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 6 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 7 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 8 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Текущий контроль успеваемости | 6 |  |  |  |  | 2 | 2 | 4 |  | 4 |
| Тема 9 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 10 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 11 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 12 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 13 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 14 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 15 | 6 | 2 |  |  |  |  | 2 | 4 |  | 4 |
| Тема 16 | 4 |  |  |  |  |  | 0 | 4 |  | 4 |
| Тема 17 | 2\* |  |  |  |  |  |  | 2\* |  | 2\* |
| Промежуточная аттестация  *экзамен*  *зачет* | 8 (6\*) |  |  |  |  | 2 | 2 | 6(4\*) |  | 6 (4\*) |
| **Итого** | 108 | 30 |  |  |  | 4 | 34 | 74 |  | 74 |

10. Перечень учебно-методического обеспечения для самостоятельной работы студентов по дисциплине (модулю):

Конспекты лекций, списки задач к лекциям, основная и дополнительная учебная литература.

11. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю).

* Перечень компетенций:
* Описание шкал оценивания*:*

*экзамен с оценкой по пятибалльной шкале*

*зачет («зачтено» или «не зачтено»)*

* Критерии и процедуры оценивания результатов обучения по дисциплине (модулю), характеризующих этапы формирования компетенций.
* Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения, характеризующих этапы формирования компетенций. См. Приложения.

12. Ресурсное обеспечение:

Перечень основной учебной литературы: см. Приложение

Перечень дополнительной учебной литературы: см. Приложения

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»: см. Приложения.

Описание материально-технической базы: аудитории для проведения лекционных занятий.

13. Язык преподавания: русский (при необходимости – английский).

ПРИЛОЖЕНИЕ

1. ИЗБРАННЫЕ ГЛАВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. «Введение в биоинформатику»
2. Преподаватель – проф. А.В.Булинский
3. Аннотация курса: специальный курс для студентов 3-6 курсов и аспирантов. Рассматривается ряд стохастических моделей, широко используемых в современной биоинформатике (например, модели скрытых марковских цепей и их обобщения). Излагаются новые статистические методы анализа медико-биологических данных. Большое внимание уделено идентификации значимых факторов, влияющих на изучаемый отклик.
4. Тематическое содержание курса:

|  |  |
| --- | --- |
| Тема 1 | Строение клетки. Структура ДНК |
| Тема 2 | Скрытые марковские цепи. Алгоритм Витерби |
| Тема 3 | Алгоритм Баума – Велча. Скрытые полумарковские цепи. Применение к анализу ДНК (идентификация кодируюшего экзона) |
| Тема 4 | Обратимые марковские цепи. Случайные блуждания на графах. Процессы рождения и гибели с дискретным временем |
| Тема 5 | Метод Монте-Карло по марковским цепям. Алгоритм Метрополиса - Хастингса |
| Тема 6 | Проблемы компьютерного моделирования заданных распределений. Алгоритм Проппа - Вилсона |
| Тема 7 | Моделирование гиббсовского распределения (метод “sandwiching”) |
| Тема 8 | Байесовский подход в статистике. |
| Тема 9 | Применение метода Монте – Карло в байесовской статистике |
| Тема 10 | Проверка гипотез. |
| Тема 11 | Проблемы выбора вероятностно-статистической модели |
| Тема 12 | Методы современного регрессионного анализа (LASSO и др.). |
| Тема 13 | Эпистазис. Статистические методы выявления значимых факторов, влияющих на изучаемый отклик |
| Тема 14 | Кросс-валидация. MDR-EFE метод. Критерий сильной состоятельности оценок дисперсии функционала ошибки предсказания отклика. |
| Тема 15 | Регуляризация используемых оценок. Центральная предельная теорема для регуляризованных статистик |
| Тема 16 | Анализ стратифицированных выборок. |
| Тема 17 | Ансамблевые методы (усреднение результатов различных выборок) |
| Тема 18 | Проблемы устойчивости статистических выводов |

*\* - если специальный курс читается в нечетном семестре (продолжительность нечетного семестра 18 недель, четного семестра 17 недель).*

1. Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения, характеризующих этапы формирования компетенций.

*Программа экзамена (или вопросы к зачету)*

1. Строение клетки. Структура ДНК. Гены. Центральная догма молекулярной биологии.
2. Скрытые марковские цепи. Алгоритм Витерби.
3. Алгоритм Баума – Велча. Скрытые полумарковские цепи. Применение к анализу ДНК (идентификация кодируюшего экзона).
4. Обратимые марковские цепи. Случайные блуждания на графах. Процессы рождения и гибели с дискретным временем
5. Метод Монте-Карло по марковским цепям. Алгоритм Метрополиса – Хастингса
6. Проблемы компьютерного моделирования заданных распределений. Алгоритм Проппа – Вилсона (теорема сходимости).
7. Моделирование гиббсовского распределения (метод “sandwiching”).
8. Байесовский подход к статистике. Пример с моделями биномиального и отрицательного биномиального распределений. Сопряженные распределения.
9. Применение метода Монте – Карло в байесовской статистике.
10. Проверка гипотез. Теория Неймана – Пирсона. Виды дивергенций. Информация Кульбака – Лейблера. Лемма Стейна о вероятностях ошибок первого и второго рода.
11. Проблемы выбора вероятностно-статистической модели. Новая XOR-модель.
12. Методы современного регрессионного анализа (LASSO, логистическая регрессия, логическая регрессия и др.). Оценки Надарая – Ватсона.
13. Эпистазис. Статистические методы выявления значимых факторов, влияющих на изучаемый отклик.
14. Кросс-валидация. MDR-EFE метод. Критерий сильной состоятельности оценок дисперсии функционала ошибки предсказания отклика.
15. Регуляризация используемых оценок. Центральная предельная теорема для регуляризованных статистик.
16. Анализ стратифицированных выборок.
17. Ансамблевые методы (усреднение результатов различных выборок)
18. Проблемы устойчивости статистических выводов

*Экзаменационные билеты (билеты к устному зачету) формируются в виде двух вопросов (А и Б) из указанного списка и одной задачи (В), примеры задач см. далее.*

Образцы билетов.

Билет №1.

А. Скрытые марковские цепи. Алгоритм Витерби.

Б. Виды дивергенций. Информация Кульбака – Лейблера. Лемма Стейна об ошибках первого и второго рода.

В. Получить формулу для байесовского классификатора.

Билет №2.

А. Эпистазис. Статистические методы выявления значимых факторов, влияющих на изучаемый отклик.

Б. Применение метода Монте – Карло в байесовской статистике.

В. Показать на примерах чувствительность алгоритма Проппа – Вилсона к выбору обновляющей функции.

Билет №3

А. Кросс-валидация. MDR-EFE метод. Критерий сильной состоятельности оценок дисперсии функционала ошибки предсказания отклика.

Б. Анализ стратифицированных выборок.

В. Пусть X,X\_1,X\_2,… - независимые одинаково распределенные случайные векторы со значениями в [0,1]^d. Обозначим R(d,n) математическое ожидание расстояния (используя sup-норму в R^d) от X до ближайшего соседа среди X\_1,…,X\_n. Доказать, что R(d,n) > d/(2(d+1)n^{1/d}).

1. Перечень основной и дополнительной учебной литературы, ресурсов информационно-телекоммуникационной сети «Интернет»:

Основная литература:

1. V.Bolon-Canedo, N.Sanchez-Marono, A.Alonso-Betanzos. Feature Selection for High Dimensional Data. Springer, Cham, 2015.
2. A.Bulinski. Some statistical methods in genetics. In: V.Schmidt (Ed.) Stochastic Geometry, Spatial Statistics and Random Fields. LNM 2120, Springer, Cham, 2015, p. 293 - 320.
3. R. Jiang, X. Zhang, M.Q. Zhang (Eds). Basics of Bioinformatics. Springer. Heidelberg, 2013.
4. R.Mitra, P.Muller (Eds). Nonparametric Bayesian Inference in Statistics. Springer, Cham, 2015.
5. G.B.Singh. Fundamentals of Bioinformatics and Computational Biology. Springer, Cham, 2015.

Дополнительная литература:

1. P.Buhlmann, S.van de Geer. Statistics for High-Dimensional Data. Springer, Heidelberg, 2011.
2. K-A.Do, Z.S.Qin, M.Vannucci. Advances in Statistical Bioinformatics. Cambridge University Press, 2013.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

http://lib.mexmat.ru/

<http://elibrary.ru/>

<http://www.mathnet.ru/>

<http://www.sciencedirect.com/>

<http://www.ams.org/mathscinet/>

http://new.math.msu.su/department/probab/index-k.html