

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

**УТВЕРЖДЕНО**

**Директор физтех-школы  
биологической и медицинской  
физики**

**Д.В. Кузьмин**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Машинное обучение
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Алгоритмическая биология
	Физтех-школа Биологической и Медицинской Физики
	центр образовательных программ Физтех-школы биологической и медицинской физики
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестры, формы промежуточной аттестации:

2 (весенний) - Зачет

3 (осенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: Л.А. Урошлев, канд. физ.-мат. наук

Программа обсуждена на заседании центра образовательных программ Физтех-школы биологической и медицинской физики 10.06.2021

## Аннотация

Дисциплина направлена на освоение основных понятий и методов теории вероятностей. После изучения курса студенты изучат методы математической статистики, математический аппарат, применяемый в теории математической статистики, основы машинного обучения и теории вероятностей, а также смогут развить логическое мышление, строить вероятностные модели и решать на их основе задачи различного уровня сложности.

### 1. Цели и задачи

#### Цель дисциплины

- освоение основных понятий и методов теории вероятностей, развитие способностей к логическому мышлению, получение навыков построения вероятностных моделей и решения на их основе задач различного уровня сложности. ознакомление с теорией и методами математической статистики. В результате изучения дисциплины студенты должны освоить математический аппарат, применяемый в теории математической статистики для машинного обучения с тем, чтобы уметь грамотно формулировать задачу в терминах теории, адекватно ее формализовать, обоснованно выбирать методы для решения поставленной задачи и правильно интерпретировать полученные результаты.

#### Задачи дисциплины

- освоение основных понятий и методов теории вероятностей;
- изучение основных методов решения вероятностных задач;
- ознакомление с наиболее важными для приложений законами распределения вероятностей;
- приобретение фундаментальных знаний по теории вероятностей для обеспечения освоения дисциплин, базирующихся на понятиях и методах теории вероятностей.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности
	УК-6.2 Оценивает свою деятельность, соотносит цели, способы и средства выполнения деятельности с её результатами
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основы машинного обучения и теории вероятностей в рамках программы курса.

уметь:

- грамотно формулировать задачу в терминах теории вероятности, адекватно ее формализовать, обоснованно выбирать методы для решения поставленной задачи, грамотно выбирать вероятностные модели для описания биоинформатических объектов и моделей, правильно интерпретировать полученные результаты.

владеть:

- навыками решения практических задач, начиная с постановки задачи, выбора вероятностной модели и до трактовки результата в области биоинформатики.

### 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	Классическая модель вероятностного пространства	3	3		3
2	Аксиоматическое построение теории вероятностей	3	3		3
3	Случайные величины и случайные векторы	3	3		3
4	Числовые характеристики распределений случайных величин	3	3		3
5	Производящие и характеристические функции случайных величин и векторов	3	3		3
6	Предельные теоремы	2	2		4
7	Основные понятия математической статистики	2	1		2
8	Точечное оценивание	1	2		2
9	Сравнение оценок	2	2		4
10	Эффективные оценки	1	1		4
11	Интервальное оценивание	1	1		2
12	Распределения, связанные с нормальным	2	2		4
13	Проверка гипотез	1	1		2
14	Критерии согласия	1	1		2
15	Исследование статистической зависимости	2	2		4
Итого часов		30	30		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

## Семестр: 2 (Весенний)

### 1. Классическая модель вероятностного пространства

Случайные события и соотношения между ними. Классическое определение вероятности. Геометрические вероятности. Простейшие комбинаторные теоремы. Условная вероятность. Теоремы умножения и сложения вероятностей. Независимость событий. Формулы полной вероятности и Байеса. Схема испытаний Бернулли. Биномиальный закон распределения вероятностей. Полиномиальная схема испытаний. Полиномиальный закон распределения вероятностей. Производящие функции распределений.

### 2. Аксиоматическое построение теории вероятностей

Аксиома непрерывности и ее эквивалентность аксиоме счетной аддитивности. Свойства вероятности. Борелевская сигма-алгебра множеств вещественной прямой. Функция распределения на прямой. Борелевская сигма-алгебра множеств и функция распределения в пространстве. Способы задания вероятностных мер на построенных измеримых пространствах. Типы вероятностных мер.

### 3. Случайные величины и случайные векторы

Случайная величина, ее распределение вероятностей и функция распределения. Типы случайных величин. Борелевские функции. Случайный вектор, его распределение вероятностей и функция распределения. Независимость случайных величин. Законы распределения функций случайных величин. Композиция (свертка) распределений.

### 4. Числовые характеристики распределений случайных величин

Математическое ожидание, его свойства и теорема о его вычислении. Дисперсия и ее свойства. Неравенство Чебышева. Математическое ожидание и дисперсия независимых случайных величин. Моменты высших порядков. Неравенства Гельдера, Йенсена и Ляпунова. Ковариационная матрица случайного вектора. Коэффициент корреляции. Условные распределения и условные математические ожидания.

### 5. Производящие и характеристические функции случайных величин и векторов

Производящие функции и факториальные моменты целочисленных неотрицательных случайных величин. Производящие функции случайных векторов. Характеристические функции случайных величин. Формула обращения. Теорема единственности. Теорема непрерывности. Семиинварианты случайных величин. Характеристические функции случайных векторов.

## Семестр: 3 (Осенний)

### 6. Предельные теоремы

Типы сходимости последовательности случайных величин. Закон больших чисел. Теорема Хинчина. Теорема Бернулли. Теорема Чебышева. Центральная предельная теорема. Теорема Леви. Локальная и интегральная теоремы Муавра-Лапласа. Теорема Ляпунова. Теорема Линдеберга. Следствия теоремы Линдеберга.

### 7. Основные понятия математической статистики

Вариационный ряд выборки и эмпирическая функция распределения. Выборочные характеристики. Асимптотическое поведение выборочных моментов. Порядковые статистики и их распределение.

## 8. Точечное оценивание

Точечные оценки и их свойства. Метод моментов. Свойства оценок метода моментов. Метод максимального правдоподобия.

## 9. Сравнение оценок

Среднеквадратический подход к сравнению оценок. Асимптотический подход к сравнению оценок.

## 10. Эффективные оценки

Регулярность семейства распределений. Неравенство Рао—Крамера. Эффективность оценок. Построение эффективных оценок. Байесовский подход к оцениванию параметров. Полные и достаточные статистики.

## 11. Интервальное оценивание

Доверительные интервалы. Принципы построения доверительных интервалов.

## 12. Распределения, связанные с нормальным

Основные статистические распределения. Преобразования нормальных выборок. Многомерное нормальное распределение. Свойства нормальных векторов.

## 13. Проверка гипотез

Гипотезы и критерии. Подходы к сравнению критериев. Построение оптимальных критериев. Критерий Неймана-Пирсона. Сложные гипотезы. Критерий отношения правдоподобия.

## 14. Критерии согласия

Общий вид критериев согласия. Понятие статистической значимости. Критерии для проверки гипотезы о распределении. Критерии для проверки однородности. Критерий хи-квадрат для проверки независимости. Проверка простых гипотез о параметрах нормального распределения. Статистическая значимость в анализе биологических последовательностей.

## 15. Исследование статистической зависимости

Математическая модель регрессии. Общая модель линейной регрессии.

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

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

## 6. Перечень рекомендуемой литературы

### Основная литература

Предоставляется на базовой кафедре:

1. Теория вероятностей / А. А. Боровков .— М. : Едиториал УРСС, 2003 .— 472 с.
2. Вероятность : в 2 т. : учебник для вузов / А. Н. Ширяев .— 4-е перераб. и доп. — М. : МЦНМО, 2007, 2011 .— Т. 2 : Суммы и последовательности случайных величин - стационарные, мартингалы, марковские цепи. - 2007, 2011. - 416 с.

### Дополнительная литература

Предоставляется на базовой кафедре:

1. Курс теории вероятностей : учебник для вузов / Б. В. Гнеденко ; Моск. гос. ун-т им. М. В. Ломоносова. — 10-е изд. доп. — М. : ЛИБРОКОМ, 2011. — 485 с.

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

Научно-библиографические и патентные базы данных в области физико-химической биологии, доступные по сети Интернет в бесплатном режиме - Science Citation Index (Web of Science), Medline (PubMed), Научная электронная библиотека (НЭБ), Российская патентная БД ФГУ ФИПС и американская патентная БД USPAFULL; электронные адреса крупных научных издательств, предоставляющих доступ к полным текстам текущих и архивным выпускам этих журналов.

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

Доступ в Интернет, UNIX сервер с отдельным аккаунтом для каждого студента. Для части занятий потребуется Zoom. Google Drive для доступа к материалам курса. Приветствуется наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

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

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

## ПРИЛОЖЕНИЕ

### ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Алгоритмическая биология Физтех-школа Биологической и Медицинской Физики центр образовательных программ Физтех-школы биологической и медицинской физики
<b>курс:</b>	<u>1</u>
<b>квалификация:</b>	магистр
Семестры, формы промежуточной аттестации:	
2 (весенний) - Зачет	
3 (осенний) - Экзамен	
<b>Разработчик:</b>	Л.А. Урошлев, канд. физ.-мат. наук

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности
	УК-6.2 Оценивает свою деятельность, соотносит цели, способы и средства выполнения деятельности с её результатами
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение» обучающийся должен:

### знать:

- основы машинного обучения и теории вероятностей в рамках программы курса.

### уметь:

- грамотно формулировать задачу в терминах теории вероятности, адекватно ее формализовать, обоснованно выбирать методы для решения поставленной задачи, грамотно выбирать вероятностные модели для описания биоинформатических объектов и моделей, правильно интерпретировать полученные результаты.

### владеть:

- навыками решения практических задач, начиная с постановки задачи, выбора вероятностной модели и до трактовки результата в области биоинформатики.

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

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

1. Операции над случайными событиями
2. Комбинаторика
3. Схема Бернулли
4. Геометрическая схема
5. Условная вероятность. Формулы полной вероятности и Байеса
6. Распределения случайных величин



7. Математическое ожидание и дисперсия случайных величин
8. Ковариация и корреляция
9. Условные распределения и математические ожидания
10. Нормальное распределение
11. Сходимость почти наверное и сходимость по вероятности
12. Характеристические функции
13. Законы больших чисел
14. Центральная предельная теорема
15. Цепи Маркова

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Возможно выполнение патентного поиска в качестве самостоятельной задачи. Успешное выполнение всех заданий по курсу и выполнение контрольных срезов знаний дает преимущество на экзамене.

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

Вопросы для зачета:

1. Случайные события и соотношения между ними. Классическое определение вероятности.
2. Геометрические вероятности. Простейшие комбинаторные теоремы.
3. Условная вероятность. Теоремы умножения и сложения вероятностей.
4. Независимость событий. Формулы полной вероятности и Байеса.
5. Схема испытаний Бернулли. Биномиальный закон распределения вероятностей.
6. Полиномиальная схема испытаний. Полиномиальный закон распределения вероятностей.
7. Производящие функции распределений.
8. Аксиома непрерывности и ее эквивалентность аксиоме счетной аддитивности.
9. Свойства вероятности. Борелевская сигма-алгебра множеств вещественной прямой.
10. Функция распределения на прямой. Борелевская сигма-алгебра множеств и функция распределения в пространстве.
11. Способы задания вероятностных мер на построенных измеримых пространствах. Типы вероятностных мер.

Вопросы для экзамена:

1. Типы сходимости последовательности случайных величин.
2. Закон больших чисел.
3. Теорема Хинчина.
4. Теорема Бернулли.
5. Теорема Чебышева.
6. Центральная предельная теорема.
7. Теорема Леви.
8. Локальная и интегральная теоремы Муавра-Лапласа.
9. Теорема Ляпунова.
10. Теорема Линдеберга. Следствия теоремы Линдеберга.
11. Регулярность семейства распределений. Неравенство Рао—Крамера.
12. Эффективность оценок. Построение эффективных оценок.
13. Байесовский подход к оцениванию параметров. Полные и достаточные статистики.

Пример билета на зачете:

Билет 1.

1. Случайные события и соотношения между ними. Классическое определение вероятности.
2. Геометрические вероятности. Простейшие комбинаторные теоремы.

Пример экзаменационного билета:

Билет 1.

1.Регулярность семейства распределений. Неравенство Рао—Крамера.

2.Эффективность оценок. Построение эффективных оценок.

### Критерии оценивания

Оценка отлично (10 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9 баллов) - 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 (8 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6 баллов) - 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 (5 баллов) - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2 балла) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1 балл) - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

Оценка "зачтено" - ставится при правильном ответе на вопросы к зачету,

Оценка "не зачтено" - ставится при неправильном ответе.

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

Аттестация по дисциплине осуществляется в форме зачета и экзамена. При проведении устного зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету при устном ответе не должен превышать одного астрономического часа. При проведении устного экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету при устном ответе не должен превышать одного астрономического часа.