

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
"Казанский (Приволжский) федеральный университет"  
Институт математики и механики им. Н.И. Лобачевского



*подписано электронно-цифровой подписью*

## Программа дисциплины

Вероятностные модели генетики

Направление подготовки: 02.03.01 - Математика и компьютерные науки

Профиль подготовки: Наука о данных

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2022

## Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
  - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
  - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): профессор, д.н. (профессор) Володин И.Н. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), igornvolodin@gmail.com ; доцент, к.н. Салимов Р.Ф. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), Rustem.Salimov@kpfu.ru ; доцент, к.н. (доцент) Симушкин С.В. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), Sergey.Simushkin@kpfu.ru

**1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО**

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-2	Способен демонстрировать базовые знания математических и естественных наук, основ программирования и информационных технологий

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

методы построения некоторых вероятностных и статистических моделей в генетике;  
основные алгоритмы для реализации построенных моделей;

Должен уметь:

применять полученные знания в построении новых моделей;  
реализовывать построенные модели в виде алгоритмов и программ;  
проверять полученные алгоритмы и программы с помощью методов статистического моделирования;

Должен владеть:

навыками выбора методов построения вероятностных моделей;  
навыками применения основных классических алгоритмов разработки;

**2. Место дисциплины (модуля) в структуре ОПОП ВО**

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.08.02 Дисциплины (модули)" основной профессиональной образовательной программы 02.03.01 "Математика и компьютерные науки (Наука о данных)" и относится к дисциплинам по выбору части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 4 курсе в 8 семестре.

**3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся**

Общая трудоемкость дисциплины составляет 4 зачетных(ые) единиц(ы) на 144 часа(ов).

Контактная работа - 41 часа(ов), в том числе лекции - 20 часа(ов), практические занятия - 20 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 1 часа(ов).

Самостоятельная работа - 103 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 8 семестре.

**4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий**

**4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)**

N	Разделы дисциплины / модуля	Се-местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само-стоя-тельная ра-бота
			Лекции, всего	Лекции в эл. форме	Практи-ческие занятия, всего	Практи-ческие в эл. форме	Лабора-торные работы, всего	Лабора-торные в эл. форме	
1	Тема 1. Анализ одной последовательности ДНК	8	5	0	5	0	0	0	26

N	Разделы дисциплины / модуля	Се- местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само- стоя- тель- ная рабо- та
			Лекции, всего	Лекции в эл. форме	Практи- ческие занятия, всего	Практи- ческие в эл. форме	Лабора- торные работы, всего	Лабора- торные в эл. форме	
2.	Тема 2. Анализ нескольких последовательностей ДНК или последовательностей протеинов	8	5	0	5	0	0	0	26
3.	Тема 3. Базовый метод поиска локальных выравниваний	8	5	0	5	0	0	0	26
4.	Тема 4. Экспрессия генов, микрочипы и многомерный анализ	8	5	0	5	0	0	0	25
	Итого		20	0	20	0	0	0	103

#### 4.2 Содержание дисциплины (модуля)

##### Тема 1. Анализ одной последовательности ДНК

Краткое введение в генетику. Понятие нуклеотида, дезоксирибонуклеиновой кислоты, рибонуклеиновой кислоты, генов, генотипа., белков, аминокислот. Анализ одной последовательности ДНК. Задача секвенирования, шотган-секвенирования: контиги и якорные контиги. Моделирование ДНК, моделирование сигналов ДНК. Задача длинных повторов. R-сканы. Анализ шаблонов: с повторения и без повторений. Мтoивы.

##### Тема 2. Анализ нескольких последовательностей ДНК или последовательностей протеинов

Анализ нескольких последовательностей ДНК или протеиновых последовательностей:

Частотное сравнение двух последовательностей. Простой тест схожести по выравниванию.

Алгоритмы выравнивания двух последовательностей: глобальное сравнение с гэпами и алгоритмы динамического программирования. Линейная модель с гэпами, локальное выравнивание. Ограничения алгоритмов динамического программирования для выравниваний. Последовательности протеинов и матрицы переходов: BLOSUM матрицы, PAM матрицы. Простая симметричная эволюционная матрица.

##### Тема 3. Базовый метод поиска локальных выравниваний

Введение в BLAST.

Сравнение двух выровненных последовательностей. Случайное блуждание(BLAST). Вычисление параметров. Подсчёт вклада. Ограничения и аппроксимации для р-значения. Нормализованные и битовые вклады. Количество высоко оцененных отклонений. Сумма по Карлину - Альтшулу

Сравнение двух невыровненных последовательностей: теоретическая и эмпирическая постановки. Краевые эффекты. Множественное тестирование.

Минимальнозначимая длина.

Связь с последовательным анализом.

##### Тема 4. Экспрессия генов, микрочипы и многомерный анализ

Введение в микрочипы. Данные для отклонений и вариации.

Статистически анализ данных с микрочипов для одного гена. Определение выражен ил ген. Тестирование для дифференциальных экспрессий.

Дифференциальная экспрессия - множественные гены. Ранговые списки. Выбор статистики. Доврительное оценивание. FWER, FDR, ANOVA.

Главные компоненты и микрочипы.

Методы кластеризации.

#### 5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

## **6. Фонд оценочных средств по дисциплине (модулю)**

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемыми результатами обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

## **7. Перечень литературы, необходимой для освоения дисциплины (модуля)**

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

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

Genetics - <http://learn.genetics.utah.edu/>

Genetics for Probability - <http://www.stat.washington.edu/thompson/S394/genetics.html>

Probabilities in genetics - <https://www.khanacademy.org/science/ap-biology/heredity/mendelian-genetics-ap/a/probabilities-in-genetics>

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

Вид работ	Методические рекомендации
лекции	Лекционный материал содержит большой объём теоретического материала, для освоения которого необходимо производить чтение уже начитанного материала с выделением плохо освоенных частей и решением возникающих задач. Кроме того, необходимо по каждой лекции выделять основные части представленных теорем и утверждений, вокруг которых формировать весь остов теории. Студентам следует стремиться к активизации знаний на занятиях и по другим дисциплинам, которые используются при построении лекции.
практические занятия	Для решения практических задач используются стандартные компьютерные средства обработки статистических данных. Поэтому первый шаг в решении таких задач состоит в доскональном освоении возможностей этих средств обработки. Главное здесь - правильная организация данных, представление их в виде удобной таблицы - фрейма. Интерпретация полученных результатов подразумевает подробное ознакомление с соответствующим теоретическим материалом
самостоятельная работа	В программу курса не входит метод Монте-Карло, поэтому основы этого метода рекомендуется изучить самостоятельно. При выполнении самостоятельных заданий, а также при изучении новых методов, описанных на занятиях, рекомендуется проверять вероятностные утверждения с помощью метода Монте-Карло. Это позволит студенту лучше освоить сам метод, и лучше понять смысл проверяемых утверждений.
зачет	Следует понимать, что некоторые вопросы могут иметь множество ответов (например, вопрос "На основе чего сравнивать различные модели регрессии?"), однако это множество ответов может выходить за рамки курса. Знание методов, выходящих за рамки курса, а также способность сравнивать различные методы, приветствуется на зачете.

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

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

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

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Компьютерный класс.

### 12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;

- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;

- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 02.03.01 "Математика и компьютерные науки" и профилю подготовки "Наука о данных".

### Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 02.03.01 - Математика и компьютерные науки

Профиль подготовки: Наука о данных

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2022

#### Основная литература:

1. Сазанов, А. А. Генетика : учебное пособие / А. А. Сазанов. - Санкт-Петербург : ЛГУ им. А. С. Пушкина, 2011. - 264 с. - Текст : электронный. - URL: <https://znanium.com/catalog/product/445036> (дата обращения: 28.06.2023). - Режим доступа: по подписке.
2. Бородин, А. Н. Элементарный курс теории вероятностей и математической статистики / А. Н. Бородин. - 9-е изд., стер. - Санкт-Петербург : Лань, 2023. - 256 с. - ISBN 978-5-507-47132-4. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/330488> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.
3. Боровков, А. А. Математическая статистика : учебник для вузов / А. А. Боровков. - 5-е изд., стер. - Санкт-Петербург : Лань, 2021. - 704 с. - ISBN 978-5-8114-7677-0. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/164711> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.
4. Лагутин, М. Б. Наглядная математическая статистика : учебное пособие / М. Б. Лагутин. - 9-е изд. (эл.). - Москва : Лаборатория знаний, 2023. - 475 с. - ISBN 978-5-93208-651-3. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/319217> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.

#### Дополнительная литература:

1. Сазанов, А. А. Основы генетики : учебное пособие / А. А. Сазанов. - Санкт-Петербург : ЛГУ им. А. С. Пушкина, 2012. - 240 с. - ISBN 978-5-8290-1132-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/445015> (дата обращения: 28.01.2022). - Режим доступа: по подписке.
2. Володин И.Н, Тихонов О. Е., Турилова Е. А. Математические основы вероятности [Текст: электронный ресурс] : [учебное пособие] / Володин И. Н., Тихонов О. Е., Турилова Е. А. ; Казан. гос. ун-т, Каф. мат. статистики .- Электронные данные (1 файл: 0,73 Мб) .- (Казань : Казанский федеральный университет, 2013) .- Загл. с экрана .- Для 5-го семестра .- Документ является электронной копией оригинала: Математические основы вероятности: [учебное пособие] / Володин И. Н., Тихонов О. Е., Турилова Е. А. -- Казань: [Казанский государственный университет], 2006. -- Фондодержатель Научная библиотека Казанского федерального университета .- Режим доступа: открытый. Оригинал копии: Математические основы вероятности : [учебное пособие] / Володин И. Н., Тихонов О. Е., Турилова Е. А. ; Казан. гос. ун-т, Каф. мат. статистики .- Казань : [Казанский государственный университет], 2006 .- 163 с. URL: [http://libweb.kpfu.ru/ebooks/09\\_66%20\\_ds005.pdf](http://libweb.kpfu.ru/ebooks/09_66%20_ds005.pdf)
3. Ширяев, А.Н. Вероятность-1 [Электронный ресурс] : учебное пособие / А.Н. Ширяев. - Электрон. дан. - Москва : МЦНМО, 2007. - 552 с. - Режим доступа: <https://e.lanbook.com/book/9448>



*Приложение 3  
к рабочей программе дисциплины (модуля)  
Б1.В.ДВ.08.02 Вероятностные модели генетики*

**Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем**

Направление подготовки: 02.03.01 - Математика и компьютерные науки

Профиль подготовки: Наука о данных

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2022

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows