УТВЕРЖДЕНО
Проректор по учебной работе
и довузовской подготовке
А. А. Воронов
25 июня 2019 г.

ПРОГРАММА

по дисциплине: **Математическая статистика**
по направлению подготовки: 01.03.02 «Прикладная математика и информатика»
физтех-школа: **ФАКТ**
кафедра: **высшей математики**
курс: 3
семестр: 5

Трудоёмкость:
Базовая часть — 3 зач. ед.;
лекции — 30 часов
практические занятия — 30 часов
лабораторные занятия — нет

Экзамен — 5 семестр

ВСЕГО АУДИТОРНЫХ ЧАСОВ — 60

Самостоятельная работа:
45 часов

Программу составил
к. ф.-м. н., доцент С. Д. Животов

Программа принята на заседании кафедры
высшей математики 22 мая 2019 г.

Заведующий кафедрой
д. ф.-м. н., профессор Г. Е. Иванов

1. Основные распределения: нормальное, χ^2, Стьюдента, Фишера. Квантили.
2. Леммы о распределении линейных и квадратичных форм нормально распределенного вектора. Теорема Фишера.
3. Выборка и генеральная совокупность (вероятностная модель). Обработка выборки -– размах, медиана, мода, коэффициент асимметрии, вариационный ряд. Законы распределения элементов вариационного ряда. Bootstrap.
4. Требования к оценкам. Достаточное условие состоятельности. Единственность эффективной оценки.
5. Неравенство Крамера–Рао и следствия из него. Многомерный аналог неравенства.
6. Функция распределения и гистограмма. Ядерное оценивание плотности.
7. Частота появления события как несмещенная, состоятельная и эффективная оценка вероятности. Оценка начальных моментов. Оценки математического ожидания, дисперсии, ковариации и коэффициента корреляции.
8. Методы получения оценок: метод моментов, метод максимального правдоподобия. Асимптотические свойства оценок, полученных методом максимального правдоподобия. Байесовский метод получения оценок.
9. Доверительные интервалы и методы их построения (точный, асимптотический, численный). Асимптотические доверительные интервалы для функции распределения и моментов. Асимптотический доверительный интервал для вероятности появления события.
10. Точные доверительные интервалы для параметров нормального распределения.
11. Проверка статистических гипотез. Ошибки первого и второго рода. Мощность критерия.
12. Критерий χ^2 для проверки гипотезы о законе распределения.
13. Критерий χ^2 для проверки гипотез однородности и независимости.
14. Критерий Колмогорова для проверки гипотезы о законе распределения, критерий Смирнова для проверки гипотезы однородности, критерий проверки гипотезы случайности.
15. Проверка параметрических гипотез. Теорема Неймана–Пирсона для проверки простой гипотезы против простой альтернативы.
16. Критерий отношения правдоподобия для проверки сложных гипотез. Асимптотический метод проверки сложных гипотез. Проверка гипотез о вероятности появления события.

17. Проверка гипотез о параметрах нормального распределения. Проверка гипотез о коэффициенте корреляции. Проверка гипотез о равенстве средних и равенстве дисперсий.
18. Дисперсионный анализ (ANOVA). Корреляционный анализ (ANCOVA).
19. Регрессионный анализ. Несмещенность, состоятельность и эффективность оценки коэффициентов регрессии. Оценка остаточной дисперсии. Закон распределения коэффициентов регрессии. Закон распределения остаточной суммы квадратов.
20. Статистический анализ коэффициентов регрессии: проверка значимости коэффициентов, проверка простейших линейных гипотез. Проверка значимости всей регрессии, сравнение двух уравнений регрессии. Коэффициент детерминации.
21. Проверка адекватности модели. Кросс–проверка. Проверка основных предположений регрессионного анализа. Доверительный интервал для прогноза.

Литература

1. *Ивченко Г. И. Медведев Ю. И.* Введение в математическую статистику. — Москва : Изд-во, ЛКИ, 2010.
2. *Лагутин М. Б.* Наглядная математическая статистика. — Москва : БИНОМ. Лаборатория знаний, 2007.
3. *Боровков А. А.* Математическая статистика. — 3-е изд. — Москва : Физматлит, 2007.
4. *Кобзарь А. И.* Прикладная математическая статистика: для инженеров и научных работников. — Москва : Физматлит, 2006.
5. *Животов С. Д.* Элементы корреляционно-регрессионного анализа: учебно-методическое пособие. — Москва : МФТИ, 2017.

ПЕРВОЕ ЗАДАНИЕ

(срок сдачи 21–26 октября)

1. Случайная величина имеет экспоненциальный закон распределения

$$p(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

Сгенерируйте выборку объема $n = 25$.

a) Определить моду, медиану, размах, коэффициент асимметрии;
b) Построить эмпирическую функцию распределения и гистограмму;
c) Найти ядерную оценку плотности распределения;
d) Определить плотность распределения среднего арифметического элементов выборки. Сравнить с бутстраповской оценкой плотности;

e) Найти бутстраповскую оценку плотности распределения коэффициента асимметрии;
f) Найти плотность совместного распределения i-го и j-го элементов вариационного ряда.

2. Случайная величина распределена равномерно на отрезке $[0, \theta]$. По выборке объема n найдены оценки параметра θ : $\tilde{\theta_1} = 2\bar{x}$, $\tilde{\theta_2} = x_{\min}$, $\tilde{\theta_3} = x_{\max}$, $\tilde{\theta_4} = x_{\min} + x_{\max}$, $\tilde{\theta_5} = \left(x_1 + \frac{\sum\limits_{k=2}^{n} x_k}{(n-1)} \right)$.

a) Проверить оценки на несмещенность и состоятельность. Исправить эти оценки, если необходимо. Какая из исправленных оценок более эффективна?
b) Сгенерируйте выборку объема $N = 100$ для некоторого значения параметра θ. Для всех значений $1 \le n \le N$ вычислите исправленные оценки и постройте графики $|\tilde{\theta_k} - \theta|$ в зависимости от n.

3. Случайная величина имеет экспоненциальное распределение $p(x) = \begin{cases} e^{-x/\theta}/\theta, & x \ge 0 \\ 0, & x < 0 \end{cases}$, $\theta > 0$. По выборке объема $n = 3$ найдены оценки параметра θ : $\tilde{\theta_1} = \bar{x}$, $\tilde{\theta_2} = \frac{x_{\min} + x_{\max}}{2}$, $\tilde{\theta_3} = x_{(2)}$ (второй член вариационного ряда).

a) Проверить оценки на несмещенность. Исправить эти оценки, если необходимо. Какая из исправленных оценок более эффективна?
b) Исследовать эти оценки на эффективность с помощью неравенства Крамера–Рао.

4. Случайная величина распределена равномерно на отрезке $[0, \theta]$.

a) По выборке объема n найти оценки параметра θ методом моментов и методом максимального правдоподобия.
b) Построить точный доверительный интервал для параметра θ.
c) Построить асимптотический доверительный интервал для параметра θ.
d) Сгенерируйте выборку объема $n = 100$ для некоторого значения параметра θ. Вычислите указанные выше доверительные интервалы для доверительной вероятности 0.95.
e) Численно постройте бутстраповский доверительный интервал.
f) Сравнить все интервалы.

5. Случайная величина имеет распределение Парето:

$$p(x) = \begin{cases} \dfrac{\theta - 1}{x^{\theta}}, & x \geq 1 \\ 0, & x < 1 \end{cases}, \quad \theta > 1.$$

a) По выборке объема n найти оценку параметра θ методом максимального правдоподобия.
b) Построить доверительный интервал для медианы.
c) Построить точный доверительный интервал для параметра θ.
d) Построить асимптотический доверительный интервал для параметра θ.
e) Сгенерируйте выборку объема $n = 100$ для некоторого значения параметра θ. Вычислите указанные выше доверительные интервалы для доверительной вероятности 0.95.
f) Численно постройте бутстраповский доверительный интервал.
g) Сравнить все интервалы.

6. При эпидемии гриппа из 200 контролируемых людей однократное заболевание наблюдалось у 181 человека, а дважды болели гриппом 9 человек. Правдоподобна ли гипотеза о том, что в течение эпидемии гриппа число заболеваний отдельного человека суть случайная величина, подчиняющаяся биномиальному распределению с количеством испытаний $n = 2$?

7. Произведено измерение размеров деталей в двух партиях по 100 деталей в каждой партии. В первой партии оказалось 25 деталей с заниженным размером, 50 деталей с точным размером, 25 деталей с завышенным размером, а во второй партии аналогичные числа оказались равны 52, 41, 7 соответственно. Проверить гипотезу о независимости номера партии деталей и размера детали.

8. Участники олимпиады по математике разбиты на два потока по 300 человек в каждом. Итоги олимпиады оказались следующими: в первом потоке оценки 2, 3, 4, 5 получили соответственно 33, 43, 80 и 144 человека, соответствующие данные для второго потока 39, 35, 72 и 154. Можно ли считать оба потока однородными?

9. При снятии показаний измерительного прибора десятые доли деления шкалы прибора оцениваются «на глаз» наблюдателем. Количества цифр 0, 1, 2, ..., 9, записанных наблюдателем в качестве десятых долей при 100 независимых измерениях, равны 5, 8, 6, 12, 14, 18, 11, 6, 13, 7 соответственно.

a) Проверить гипотезу о согласии данных с законом равномерного распределения с помощью критерия χ^2.
b) Проверить гипотезу о согласии данных с законом нормального распределения с помощью критерия χ^2.
c) Проверить гипотезу о согласии данных с законом нормального распределения с помощью критерия Колмогорова (распределение критерия определить бутстрапом).

10. Основная гипотеза H_0 состоит в том, что случайная величина имеет распределение с плотностью $p_0(x) = \begin{cases} 1, & x \in (0, 1) \\ 0, & x \notin (0, 1) \end{cases}$. Альтернатива H_1 состоит в том, что случайная величина имеет распределение с плотностью $p_1(x) = \begin{cases} \dfrac{e^{1-x}}{e-1}, & x \in (0, 1) \\ 0, & x \notin (0, 1) \end{cases}$.

a) Построить наиболее мощный критерий проверки этих гипотез по выборке объема $n = 1$ с уровнем значимости α, найти ошибки первого и второго рода, мощность критерия.
b) Построить наиболее мощный критерий проверки этих гипотез по выборке объема $n = 2$ с уровнем значимости α, найти ошибки первого и второго рода, мощность критерия.
c) Построить наиболее мощный асимптотический критерий проверки этих гипотез по выборке объема n с уровнем значимости α, найти ошибки первого и второго рода, мощность критерия.
d) Построить критерий по выборке объема n с критической областью $x_{\min} < c$ и уровнем значимости α, найти ошибки первого и второго рода, мощность критерия.

11. У игрока, наблюдавшего за игрой в четырехгранные кости, создалось впечатление, что четверка выпадает в 8 случаях из 24, тройка в 4, а остальные две грани выпадают равновероятно. Получив приглашение принять участие в игре, игрок попросил разрешения предварительно проверить свою гипотезу на двух производимых подряд бросаниях кости. Единственная рассматриваемая им альтернатива состоит в том, что игральная кость сделана «честно». Найти наиболее мощный критерий с уровнем значимости 0,2. Найти мощность критерия.

12. Дана выборка объемы $n = 3$ из нормального распределения $N(a, \sigma^2)$: $-1.11; -6.10; 2.42$. Проверьте гипотезу $a = 0$, против альтернативы $a > 0$.

13. Пусть z_n и y_m — независимые случайные выборки из нормального распределения с параметрами $a, \sigma_x^2 = 2$ и $b, \sigma_x^2 = 1$ соответственно. Используя реализации случайных выборок: $x = \{-1.11, -6.10, 2.42\}$, $y = \{-2.29, -2.91\}$, проверить гипотезу о равенстве средних против альтернативы $a < b$.

ВТОРОЕ ЗАДАНИЕ

(срок сдачи 16–21 декабря)

1. Случайный вектор $(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \eta)$ имеет компоненты, распределенные по следующему закону: $\xi_1 \sim R(-1, 1)$, $\eta \sim N(2 + 3x_1 - 2x_2 + x_3 + x_4 - x_5, 1.5^2)$, где x_i это значения, которые принимает случайная величина ξ_i. Сгенерировать выборку объема $n = 50$.

a) Проверить переменные ξ_i на мультиколлинеарность.

b) Определить уравнение линейной регрессии $\eta = \beta_0 + \sum\limits_{i=1}^{5} \beta_i \xi_i$ и проверить значимость коэффициентов.

c) Определить коэффициент детерминации и проверить его значимость.

d) Найти значение в точке $x_i = 0$ и построить 95% доверительный интервал.

e) Проверить предположение о независимости ошибок измерения.

f) Проверить предположение о нормальности распределения ошибок.

g) Провести кросс-проверку регрессии.

h) Проверить адекватность регрессии, сделав 5 повторных измерений в одной точке.

i) Удалить переменную, соответствующую наименее значимому коэффициенту и повторить пункты b и c, сравнить уравнения регрессии.

j) Сравнить уравнения регрессии бутстрапом.

2. В таблице приведены данные о содержании иммуноглобулина $Ig\ A$ в сыворотке крови у больных пяти возрастных групп:

№	Содержание $Ig\ A (\%)$
1	83, 85
2	84, 85, 85, 86, 86, 87
3	86, 87, 87, 87, 88, 88, 88, 88, 88, 89, 90
4	89, 90, 90, 91
5	90, 92

a) Предполагая, что выборки получены из нормального закона распределения, определить влияние возраста на содержание иммуноглобулина в крови с помощью дисперсионного анализа (ANOVA).

b) Проверить гипотезу о равенстве дисперсий.

c) Определить влияние возраста на содержание иммуноглобулина в крови с помощью регрессионного анализа.

d) Провести попарное сравнение средних в рамках регрессионной модели.

Составитель задания к. ф.-м. н., доцент С. Д. Животов