

Программа курса „Математическая статистика“

лектор — к.ф.-м.н. М. Е. Жуковский

Весна 2014

1. Различные виды сходимостей случайных векторов: с вероятностью 1, по вероятности, по распределению. Законы больших чисел для сходимостей случайных векторов. Теорема о наследовании сходимости и лемма Слуцкого. Пример применения леммы Слуцкого. Многомерная центральная предельная теорема (б/д).
2. Эмпирическое распределение и эмпирическая функция распределения. Теорема Гливенко–Кантелли.
3. Вероятностно–статистическая модель. Понятия наблюдения и выборки. Параметрическая статистическая модель. Моделирование выборки из неизвестного распределения, принадлежащему параметрическому семейству.
4. Статистики и оценки. Примеры статистик: выборочные характеристики, порядковые статистики, выборочные квантили. Основные свойства оценок: несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность. Примеры.
5. Наследование состоятельности и сильной состоятельности при взятии непрерывной функции. Лемма о наследовании асимптотической нормальности. Примеры. Асимптотическая нормальность выборочной квантили и выборочной медианы. Примеры.
6. Метод моментов, состоятельность оценки метода моментов. Примеры.
7. Понятие плотности в дискретном случае. Метод максимального правдоподобия. Примеры. Экстремальное свойство функции правдоподобия. Состоятельность оценки максимального правдоподобия. Примеры.
8. Сравнение оценок, функция потерь и функция риска. Подходы к сравнению оценок: равномерный, байесовский, минимаксный, асимптотический. Примеры.
9. Неравенство Рао–Крамера и эффективные оценки. Критерий эффективности оценки.
10. Асимптотическая нормальность оценки максимального правдоподобия в регулярном случае для одномерного параметра. Эффективность и асимптотическая эффективность оценки максимального правдоподобия.

11. Условное математическое ожидание случайной величины относительно σ -алгебры. Теорема Радона–Никодима (б/д) и обоснование существования условного математического ожидания. Явный вид условного математического ожидания в случае, если σ -алгебра порождена счетным разбиением. Основные свойства условного математического ожидания.
12. Условные распределения и условные плотности. Достаточное условие существования условной плотности. Вычисление условного математического ожидания с помощью условной плотности.
13. Достаточные статистики и σ -алгебры. Критерий факторизации Неймана–Фишера (док-во для дискретного и абсолютно непрерывного случаев). Примеры. Теорема Колмогорова–Блекуэлла–Рао об улучшении несмещенных оценок.
14. Полные достаточные статистики. Единственность наилучшей несмещенной оценки. Экспоненциальное семейство распределений. Теорема о полной достаточной статистике в экспоненциальном семействе (б/д). Нахождение оптимальных оценок с помощью полных достаточных статистик. Примеры.
15. Доверительные интервалы. Метод центральной статистики. Асимптотические доверительные интервалы. Построение асимптотических доверительных интервалов с помощью асимптотически нормальных оценок. Примеры.
16. Линейная регрессионная модель. Оценка наименьших квадратов, ее основные свойства. Несмещенная оценка для дисперсии ошибки измерений σ^2 .
17. Линейная гауссовская модель. Достаточные статистики в линейной гауссовской модели. Наилучшие несмещенные оценки параметров в линейной гауссовской модели, их распределения.
18. Распределения хи-квадрат, Стьюдента и Фишера. Теорема об ортогональных разложениях гауссовского вектора. Доверительные интервалы для параметров гауссовской линейной модели. Примеры.
19. Проверка статистических гипотез: общие принципы и основные понятия (критическое множество, уровень значимости, альтернативы, ошибки первого и второго родов, функция мощности). Сравнения критериев: наиболее мощные и равномерно наиболее мощные критерии. Байесовский и минимаксный подходы. Состоятельность статистического критерия.
20. Лемма Неймана–Пирсона. Построение с ее помощью наиболее мощных критериев. Примеры. Теорема о монотонном отношении правдоподобия (б/д). Построение равномерно наиболее мощных критериев для односторонних альтернатив. Пример построения равномерно наиболее мощного критерия в случае отсутствия монотонного отношения правдоподобия.
21. F -критерий для проверки линейных гипотез в гауссовской линейной модели. Пример с двумя гауссовскими выборками, отличающимися сдвигом: проверка гипотезы об их однородности.

22. Проверка непараметрических гипотез. Теорема Пирсона. Критерий согласия Пирсона для проверки простой гипотезы в схеме испытаний Бернулли с m исходами. Состоятельность критерия Пирсона. Расширение области применения критерия Пирсона.
23. Критерий Колмогорова–Смирнова. Теорема Колмогорова (б/д). Явный вид статистики Колмогорова. Критерий Мизеса–Смирнова. Теорема Мизеса–Смирнова (б/д). Явный вид статистики Мизеса–Смирнова.

СПИСОК ЛИТЕРАТУРЫ

1. *Боровков А. А.* Математическая статистика. — 3-е изд. — М.: Физматлит, 2007.
2. *Севастьянов Б. А.* Курс теории вероятностей и математической статистики. — 2-е изд. — М.-Ижевск: Институт компьютерных исследований, 2004.
3. *Тюрин Ю. Н.* Математическая статистика. Записки лекций. — М.: изд-во ЦПИ механико-математического факультета МГУ, 2003.
4. *Ширяев А. Н.* Вероятность. В 2-х кн. — 3-е изд. — М.: МЦНМО, 2004.

Задачи для самостоятельного решения

1. Сходимости случайных векторов, гауссовские векторы

- 1** Пусть последовательность случайных векторов $\xi_1, \dots, \xi_n, \dots$ сходится по распределению к константе C . Докажите, что тогда $\xi_n \xrightarrow{P} C$.
- 2** Задана выборка X_1, \dots, X_n из распределения $\mathcal{N}(0, \sigma^2)$. Рассмотрим статистики $Y = \frac{1}{n} \sum_{i=1}^n |X_i|$, $Z = \frac{1}{n} \sum_{i=1}^n X_i^2$ и $T = \sqrt{\frac{2}{\pi}} Z/Y$. Найдите предел сходимости по распределению выражения

$$\sqrt{n}(T - \sigma).$$

- 3** Пусть $\xi_n \xrightarrow{d} \xi$ — случайные векторы размерности m , а $h(x_1, \dots, x_m)$ — функция m переменных, дифференцируемая в точке $a \in \mathbb{R}^m$. Найдите предел сходимости по распределению для выражения

$$\frac{h(a + b_n \xi_n) - h(a)}{b_n},$$

где $b_n \rightarrow 0$ — произвольная последовательность положительных чисел.

- 4** Пусть $\{\xi_n\}_{n=1}^{\infty}$ и $\{\eta_n\}_{n=1}^{\infty}$ — две последовательности случайных величин, причем для каждого $n \geq 1$ величины ξ_n и η_n независимы. Пусть $\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$. Используя метод характеристических функций, докажите, что ξ и η — тоже независимы.
- 5** Пусть $X = (\xi, \eta)$ — гауссовский вектор. Подберите такие числа x_1, x_2 , что случайные величины $\eta + x_1 \xi$, $\eta + x_2 \xi$ являются независимыми.
- 6** Случайные величины X и Y — независимые нормальные с параметрами $(0, 1)$. Докажите, что распределение случайной величины $Z = (X + a)^2 + (Y + b)^2$ зависит только лишь от величины $r = \sqrt{a^2 + b^2}$.
- 7** Приведите пример таких двух случайных величин X и Y , что обе они — стандартные нормальные, $EXY = 0$, но вектор (X, Y) не является гауссовским.
- 8** Пусть X_1, \dots, X_n — выборка из распределения Лапласа с параметром σ , имеющего плотность

$$p(x) = \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}.$$

Рассмотрим $Y = \frac{1}{n} \sum_{i=1}^n |X_i|$, $Z = \frac{1}{n} \sum_{i=1}^n X_i^2$. Используя многомерную центральную предельную теорему, найдите предел по распределению для выражения

$$\sqrt{n}(T - \sigma),$$

где $T = Z^2/(4Y^3)$.

2. Свойства оценок

- 1 Пусть X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, \theta]$. Проверьте на несмещенность, состоятельность и сильную состоятельность следующие оценки параметра θ : $2\bar{X}$, $\bar{X} + X_{(n)}/2$, $(n+1)X_{(1)}$, $X_{(1)} + X_{(n)}$, $\frac{n+1}{n}X_{(n)}$.
- 2 Пусть X_1, \dots, X_n — выборка из распределения $Bin(1, \theta)$. Для каких функций $\tau(\theta)$ существуют несмещенные оценки?
- 3 Пусть $\hat{\theta}_n(X)$ — асимптотически нормальная оценка параметра θ с асимптотической дисперсией $\sigma^2(\theta)$. Докажите, что тогда $\hat{\theta}_n(X)$ является состоятельной оценкой θ .
- 4 Пусть X_1, \dots, X_n — выборка из распределения с параметром σ^2 . Пусть, кроме того, $DX_1 = \sigma^2$. Докажите, что статистика $s^2 = 1/n \sum_{i=1}^n (X_i - \bar{X})^2$ равна $\overline{X^2} - (\bar{X})^2$ и является состоятельной оценкой σ^2 . Является ли она несмещенной оценкой того же параметра?
- 5 Пусть X_1, \dots, X_n — выборка из экспоненциального распределения с параметром θ . Покажите, что для любого $k \in \mathbb{N}$ статистика $\sqrt[k]{k!/\bar{X}^k}$ является асимптотически нормальной оценкой параметра θ . Найдите ее асимптотическую дисперсию.

3. Основные методы нахождения оценок

- 1 Найдите оценки по методу моментов со стандартными пробными функциями для следующих распределений: а) $\mathcal{N}(a, \sigma^2)$, б) $\Gamma(\alpha, \lambda)$, в) $R(a, b)$, г) $Pois(\lambda)$, д) $Bin(m, p)$, е) $Geom(p)$, ж) $Beta(\lambda_1, \lambda_2)$.
- 2 Найдите оценки по методу максимального правдоподобия для следующих распределений:
 - а) $\mathcal{N}(a, \sigma^2)$ в трех случаях: когда неизвестен только один из параметров и когда неизвестны оба параметра; б) $\Gamma(\alpha, \lambda)$, если параметр λ известен; в) $R(a, b)$; г) $Pois(\lambda)$; д) $Bin(m, p)$, если параметр m известен; е) $Geom(p)$.
- 3 X_1, \dots, X_n — выборка из распределения с плотностью

$$p_{\alpha, \beta}(x) = \frac{1}{\alpha} e^{(\beta-x)/\alpha} I_{[\beta, +\infty)}(x).$$

где $\theta = (\alpha, \beta)$ — двумерный параметр. Найдите для θ оценку максимального правдоподобия. Докажите, что полученная для α оценка $\hat{\alpha}_n$ является асимптотически нормальной, и найдите ее асимптотическую дисперсию.

- 4 Найдите оценку максимального правдоподобия для параметра сдвига в модели распределения Коши,

$$p_{\theta}(x) = \frac{1}{\pi(1 + (x - \theta)^2)},$$

если выборка состоит из а) одного наблюдения, б) двух наблюдений (т.е. $n = 1, 2$).

4. Сравнение оценок. Эффективные оценки

- 1 Пусть X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, \theta]$. Сравните следующие оценки параметра θ в равномерном подходе с квадратичной функцией потерь: $2\bar{X}$, $(n+1)X_{(1)}$, $\frac{n+1}{n}X_{(n)}$.
- 2 Пусть $\theta_1^*(X)$ и $\theta_2^*(X)$ — две наилучшие оценки параметра θ с одинаковыми математическими ожиданиями. Докажите, что тогда для любого θ они совпадают почти наверное, т.е. $\theta_1^*(X) = \theta_2^*(X)$ P_θ -п.н.
- 3 Пусть X_1, \dots, X_n — выборка из биномиального распределения с параметрами (m, p) , причем m известно. Найдите информацию Фишера $i(p)$ в данной модели, а также эффективную оценку параметра p .
- 4 Пусть X_1, \dots, X_n — выборка из экспоненциального распределения с параметром θ . Для какой функции $\tau(\theta)$ существует эффективная оценка? Вычислите информацию Фишера $i(\theta)$ одного наблюдения в данной модели.
- 5 Пусть X_1, \dots, X_n — выборка из нормального распределения с параметрами (a, σ^2) . Найдите эффективную оценку
 - а) параметра a , если σ известно;
 - б) параметра σ^2 , если a известно.Вычислите информацию Фишера одного наблюдения в обоих случаях.

5. Условные математические ожидания и условные распределения

- 1 Из урны, содержащей 2 черных, 4 красных и 2 белых шара вытаскиваются с возвращением по одному n раз шары. Пусть X — количество черных, а Y — количество красных шаров, появившихся в этих опытах. Найдите $E(X|X+Y)$.
- 2 Пусть X_1, \dots, X_n — выборка из распределения с конечным математическим ожиданием, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ — симметричная (значение не меняет при любой перестановке аргументов) борелевская функция. Докажите, что $\forall i, j \in \{1, \dots, n\}$ справедливо равенство $E(X_i|g(X_1, \dots, X_n)) = E(X_j|g(X_1, \dots, X_n))$. Найдите $E(X_1|\sum_{i=1}^n X_i)$.
- 3 Пусть (X, Y) — гауссовский вектор, $(X, Y) \sim \mathcal{N}(a, \Sigma)$. Найдите $E(X|Y)$ и $E(X|X+Y)$.
- 4 Случайная величина X имеет равномерное распределение на отрезке $[-1, 1]$. Вычислите $E(X|X^2)$.
- 5 Пусть X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, 1]$. Найдите а) $E(X_1|X_{(1)})$, б) $E(X_1|X_{(n)})$.
- 6 Пусть X и Y — независимые случайные величины, равномерно распределенные на отрезке $[0, 2]$. Найдите $E(Y^2|X/Y)$.
- 7 Пусть X и Y — независимые случайные величины, X имеет равномерное распределение на отрезке $[0, 1]$, а Y — экспоненциальное распределение с параметром 1. Найдите $E(Y|X/Y)$.
- 8 Пусть X и Y — независимые случайные величины, X имеет равномерное распределение на отрезке $[0, 1]$, а Y — экспоненциальное с параметром 1. Найдите $E(Y^3/X^2|X/Y)$.

6. Достаточные статистики и оптимальные оценки

- 1 Найдите достаточные статистики для следующих параметрических распределений:
а) $\mathcal{N}(a, \sigma^2)$, б) $\Gamma(\alpha, \lambda)$, в) $R(a, b)$, г) $Pois(\lambda)$, д) $Bin(1, p)$, е) $Geom(p)$.
- 2 Найдите оптимальную оценку параметра $\theta > 0$ по выборке из распределения: а) $\mathcal{N}(\theta, 1)$, б) $R(0, \theta)$, в) $Pois(\theta)$, г) $Bin(1, \theta)$ (здесь $\theta \in (0, 1)$).
- 3 Пусть X_1, \dots, X_n — выборка из нормального распределения с параметрами (a, σ^2) , $a \in \mathbb{R}$, $\sigma > 0$. Найдите оптимальную оценку параметра $\theta = (a, \sigma^2)$.
- 4 Пусть X_1, \dots, X_n — выборка из экспоненциального распределения с параметром $\theta > 0$. Найдите оптимальные оценки для θ и $\tau(\theta) = \theta^{1/2}$.
- 5 Пусть X_1, \dots, X_n — выборка из нормального распределения с параметрами $(0, \theta^2)$. Найдите оптимальную оценку для θ .
- 6 Пусть X_1, \dots, X_n — выборка из пуассоновского распределения с параметром $\theta > 0$. Найдите $E\left(X_1^2 \mid \sum_{i=1}^n X_i\right)$.

7. Доверительные интервалы

- 1 Пусть X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, \theta]$, $\theta > 0$. Постройте доверительный интервал для θ уровня доверия α , используя статистику
а) \bar{X} , б) $X_{(1)}$, в) $X_{(n)}$.
- 2 Пусть X_1, \dots, X_n — выборка из распределения Коши со сдвигом, т.е.

$$p_\theta(x) = \frac{1}{\pi(1 + (x - \theta)^2)}.$$

Постройте асимптотический доверительный интервал для θ уровня доверия α .

- 3 Пусть X_1, \dots, X_n — выборка из пуассоновского распределения с параметром θ . Постройте асимптотический доверительный интервал для θ уровня доверия α .
- 4 Пусть X_1, \dots, X_n — выборка из гамма-распределения с параметрами (θ, λ) . Постройте асимптотический доверительный интервал для θ уровня доверия α , если а) λ известно, б) λ неизвестно.

8. Оценки наименьших квадратов. Гауссовская линейная модель

- 1** В четырехугольнике $ABCD$ независимые равноточные измерения углов ABD , DBC , ABC , BCD , CDB , BDA , CDA , DAB (в градусах) дали результаты 50.78, 30.25, 78.29, 99.57, 50.42, 40.59, 88.87, 89.86 соответственно. Считая, что ошибки измерений распределены нормально по закону $\mathcal{N}(0, \sigma^2)$, найдите оптимальные оценки углов $\beta_1 = ABD$, $\beta_2 = DBC$, $\beta_3 = CDB$, $\beta_4 = BDA$ и неизвестной дисперсии σ^2 .
- 2** Пусть $X_i = \beta_1 + i\beta_2 + \varepsilon_0 + \dots + \varepsilon_i$, $i = 0, 1, \dots, n$, где β_1, β_2 — неизвестные параметры, а $\varepsilon_0, \dots, \varepsilon_n$ — независимые, распределенные по закону $\mathcal{N}(0, \sigma^2)$ случайные величины. Сведите задачу к линейной модели и найдите оценки наименьших квадратов для β_1 и β_2 , а также несмещенную оценку для σ^2 .
- 3** Пусть X_1, \dots, X_n — выборка из нормального распределения с параметрами (a, σ^2) . Докажите, что статистики \bar{X} и $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ независимы и вычислите распределение статистики nS^2 .
- 4** Пусть $X_i, i \in \{1, \dots, n\}$ — независимые случайные величины, распределенные по нормальному закону с параметрами $(a + bi, \sigma^2)$. Постройте точные доверительные интервалы для параметров a, b, σ^2 .

9. Проверка статистических гипотез

- 1 Имеется X_1 — выборка объема 1. Основная гипотеза H_0 состоит в том, что X_1 имеет равномерное распределение на отрезке $[0, 1]$, альтернатива — в том, что X_1 имеет показательное распределение с параметром 1. Постройте наиболее мощный критерий уровня значимости α для различения этих гипотез и вычислите его мощность.
- 2 X_1, \dots, X_n — выборка из экспоненциального распределения с параметром θ . Постройте равномерно наиболее мощный критерий уровня значимости α проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы
а) $H_1 : \theta > \theta_0$, б) $H_1 : \theta < \theta_0$.
- 3 X_1, \dots, X_n — выборка из нормального распределения с параметрами $(\theta, 1)$. Постройте равномерно наиболее мощный критерий уровня значимости α проверки
а) гипотезы $H_0 : \theta \geq \theta_0$ против альтернативы $H_1 : \theta < \theta_0$,
б) гипотезы $H_0 : \theta \leq \theta_0$ против альтернативы $H_1 : \theta > \theta_0$.
- 4 X_1, \dots, X_n — выборка из биномиального распределения $Bin(1, \theta)$. Докажите, что не существует равномерного наиболее мощного критерия произвольного уровня значимости α для проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы $H_1 : \theta \neq \theta_0$.
- 5 X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, \theta]$, $\theta > 0$. Постройте равномерно наиболее мощный критерий уровня значимости α для проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы $H_1 : \theta \neq \theta_0$ в виде

$$S(X) = \{X_{(n)} > \theta_0\} \cup \{X_{(n)} \leq c\theta_0\}.$$

- 6 По выборке объёма n из нормального распределения с неизвестным средним a и неизвестной дисперсией построить байесовский критерий для различения двух простых гипотез о параметре a , если априорные вероятности гипотез равны.
- 7 Пусть X_1, \dots, X_n — выборка из геометрического распределения с параметром p , где p может принимать лишь значения $1/2$ и $1/4$ с априорными вероятностями $1/3$ и $2/3$ соответственно. Построить байесовский критерий.
- 8 По выборке объёма 1 из нормального распределения с неизвестным средним a и неизвестной дисперсией построить минимаксный критерий для различения двух простых гипотез о параметре a .

10. Критерии согласия: критерий Колмогорова и критерий хи-квадрат

- 1 Имеется выборка X_1, X_2, X_3 объема 3. Для проверки гипотезы о том, что выборка взята из равномерного на отрезке $[0,1]$ распределения, используется критерий Колмогорова: гипотеза о равномерности отвергается, если $\sup_{y \in [0,1]} |F_3^*(y) - y| > 1/3$. Сформулировать этот критерий в явном виде в терминах порядковых статистик. Чему равен размер этого критерия?
- 2 Доказать, что при условии $0 \leq X_{(1)} \leq X_{(n)} \leq 1$ справедливо равенство $\int_0^1 (F_n^*(y) - y)^2 dy = \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^n (X_{(k)} - (2k-1)/n)^2$ (с помощью этого представления часто вычисляется значение статистика ω^2).
- 3 Доказать состоятельность критерия Колмогорова.
- 4 Цифры $0, 1, 2, \dots, 9$ среди 800 первых десятичных знаков числа π появились 74, 92, 83, 79, 80, 73, 77, 75, 76, 91 раз соответственно. Проверьте гипотезу о согласии этих данных с законом равномерного распределения на множестве $\{0, 1, \dots, 9\}$ на уровне значимости
а) 0.05, б) 0.5, в) 0.8.
- 5 Пользуясь интегральной теоремой Муавра–Лапласа, доказать теорему Пирсона при $k = 2$.

11. F-критерий

- 1 Постройте F -критерий уровня значимости α для проверки гипотезы $H_0 : \beta_2 = \beta_1$ в задаче 9.2.
- 2 X_1, \dots, X_n — выборка из распределения $\mathcal{N}(a_1, \sigma^2)$, Y_1, \dots, Y_m — выборка из распределения $\mathcal{N}(a_2, \sigma^2)$, Z_1, \dots, Z_k — выборка из распределения $\mathcal{N}(a_3, \sigma^2)$. Постройте F -критерий размера α для проверки гипотезы $H_0 : a_1 = a_2$ и $a_1 + a_2 = a_3$.