Программа курса „Математическая статистика“

лектор — к.ф.-м.н. Д. А. Шабанов

Весна 2011

1. Основная задача математической статистики. Понятие вероятностно-статистической модели. Примеры: выборка и линейная модель.

2. Различные виды сходимостей случайных величин и векторов: с вероятностью 1, по вероятности, по распределению. Три знаменитых теоремы: закон больших чисел, усиленный закон больших чисел, центральная предельная теорема. Теорема о наследовании сходимости и лемма Слуцкого. Лемма о наследовании „асимптотической нормальности“.

3. Эмпирическое распределение и эмпирическая функция распределения. Обоснованность основной задачи математической статистики и теорема Гливенко–Кантелли.

4. Параметрическая статистическая модель. Запас параметрических моделей, напоминание о распределении сумм независимых одинаково распределенных случайных величин.

5. Статистики и оценки. Примеры статистик: выборочные характеристики, порядковые статистики. Лемма о совместной плотности элементов вариационного ряда.

6. Основные свойства оценок: несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность. Примеры.

7. Методы нахождения оценок, общий принцип подстановки. Метод моментов, состоятельность оценки метода моментов. Выборочные квантили и выборочная медиана, использование выборочных квантилей для нахождения состоятельной оценки в „неинтегрируемых случаях“. Примеры.

8. Сравнение оценок, функция потерь и функция риска. Подходы к сравнению оценок: равномерный, байесовский, минимаксный, асимптотический. Допустимые оценки.

9. Понятие плотности в дискретном случае и условия регулярности семейства распределений. Неравенство Рао–Крамера и эффективные оценки. Критерий эффективности оценки (б/д). Многомерное неравенство Рао-Крамера.

10. Метод максимального правдоподобия. Примеры. Экстремальное свойство функции правдоподобия. Состоятельность и 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. Теорема о наилучшем квадратичном прогнозе.

13. Байесовские оценки. Оптимальность байесовской оценки при байесовском подходе к сравнению оценок.

14. Достаточные оценки и σ-алгебры. Критерий факторизации Неймана–Фишера (док-во для дискретного и абсолютно непрерывного случаев). Примеры. Теорема Колмогорова–Блекуэлла—Рао об улучшении несмещенных оценок. Многомерный вариант теоремы Колмогорова–Блекуэлла—Рао.

15. Полные достаточные статистики. Единственность наилучшей несмещенной оценки. Примеры. Экспоненциальное семейство распределений. Теорема о полной достаточной статистике в экспоненциальном семействе (б/д). Нахождение оптимальных оценок с помощью полных достаточных статистик.

16. Доверительные интервалы и доверительные области. Метод центральной статистики. Асимптотические доверительные интервалы. Построение асимптотических доверительных интервалов с помощью асимптотически нормальных оценок. Примеры.

17. Линейная регрессионная модель. Оценки наименьших квадратов, их основные свойства. Теорема о наилучшей оценке в классе линейных оценок (б/д). Несмещенная оценка для дисперсии ошибки измерений σ^2.

18. Гауссовские случайные векторы (многомерное нормальное распределение): три эквивалентных определениях, основные свойства, критерий независимости компонент гауссовского вектора. Теорема об ортогональных разложениях гауссовского вектора. Многомерная центральная предельная теорема (б/д).

19. Линейная гауссовская модель. Достаточные статистики в линейной гауссовской модели. Наилучшие несмещенные оценки параметров в линейной гауссовской модели, их распределения.

20. Распределения хи-квадрат, Стьюдента и Фишера. Доверительные интервалы и эллипсоиды для параметров гауссовской линейной модели. Примеры.

21. 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. Несмещенность и состоятельность статистического критерия.

22. 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.

23. Двойственность доверительного оценивания и проверки гипотез.

24. F-критерий для проверки линейных гипотез в гауссовской линейной модели. Пример с двумя гауссовскими выборками, отличающимися сдвигом: проверка гипотезы об их однородности.

25. Проверка непараметрических гипотез. Теорема Пирсона. Критерий согласия Пирсона для проверки простой гипотезы в схеме испытаний Бернулли с m исходами.

СПИСОК ЛИТЕРАТУРЫ

1. *Боровков А. А.* Математическая статистика. — 3-е изд. — М.: Физматлит, 2007.

2. *Ивченко Г. И. и Медведев Ю. И.* Математическая статистика. — М.: Высшая школа, 1984.

3. *Лагутин М. Б.* Наглядная математическая статистика. — М.: БИНОМ. Лаборатория знаний, 2007.

4. *Леман Э.* Теория точечного оценивания. — Пер. с англ. — М.: Наука, 1991.

5. *Севастьянов Б. А.* Курс теории вероятностей и математической статистики. — 2-е изд. — М.-Ижевск: Институт компьютерных исследований, 2004.

6. *Тюрин Ю. Н.* Математическая статистика. Записки лекций. — М.: изд-во ЦПИ механико-математического факультета МГУ, 2003.

7. *Ширяев А. Н.* Вероятность. В 2-х кн. — 3-е изд. — М.: МЦНМО, 2004.

Задачи для самостоятельного решения

Сходимости случайных векторов

1 Приведите пример такой последовательности случайных векторов $\{(\xi_n, \eta_n)\}_{n\in\mathbb{N}}$ и таких случайных величин ξ и η, что $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} \eta$, но $(\xi_n, \eta_n) \overset{d}{\nrightarrow} (\xi, \eta)$.

2 Пусть последовательность случайных векторов $\xi_1, \ldots, \xi_n, \ldots$ сходится по распределению к константе C. Докажите, что тогда $\xi_n \xrightarrow{\mathsf{P}} C$.

3 Задана выборка $X_1, \ldots, X_n$ 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

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

4 Пусть $\xi_n \xrightarrow{d} \xi$ — случайные векторы размерности m, а $h(x_1, \ldots, x_m)$ — 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 $b_n \to 0$ — произвольная последовательность положительных чисел.

5* Пусть $\{\xi_n, n \in \mathbb{N}\}$ — независимые одинаково распределенные невырожденные случайные величины с конечным вторым моментом. Пусть $\mathsf{E}\xi_i = a$, $S_n = \xi_1 + \ldots + \xi_n$. Докажите, что у выражения

$$\sqrt{n}\left(\frac{S_n}{n} - a\right)$$

есть предел сходимости по распределению, но нет предела сходимости по вероятности.

Свойства оценок

1 Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Проверьте на несмещенность, состоятельность и сильную состоятельность следующие оценки параметра θ: $2\overline{X}$, $\overline{X} + X_{(n)}/2$, $(n+1)X_{(1)}$, $X_{(1)} + X_{(n)}$, $\frac{n+1}{n}X_{(n)}$.

2 Пусть $X_1, \ldots, X_n$ — выборка из распределения $Bin(1, \theta)$. Для каких функций $\tau(\theta)$ существуют несмещенные оценки?

3 Пусть X_1 — выборка из распределения $Bin(n, \theta)$. Найдите несмещенную оценку для $\tau(\theta) = \theta^k(1-\theta)^s$, $k + s < n$.

4 Пусть $\widehat{\theta}_n(X)$ — асимптотически нормальная оценка параметра θ с асимптотической дисперсией $\sigma^2(\theta)$. Докажите, что тогда $\widehat{\theta}_n(X)$ является состоятельной оценкой θ.

5 Пусть $X_1, \dots, X_n$ — выборка из экспоненциального распределения с параметром θ. Покажите, что для любого $k \in \mathbb{N}$ статистика $\sqrt[k]{k!/\overline{X^k}}$ является асимптотически нормальной оценкой параметра θ. Найдите ее асимптотическую дисперсию.

6* Проверьте, являются ли оценки из задачи №1 также и асимптотически нормальными оценками параметра θ.

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 $\widehat{\alpha}_n$ является асимптотически нормальной, и найдите ее асимптотическую дисперсию.

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

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

5* 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. $n = 1, 2$).

Сравнение оценок. Байесовские 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: $2\overline{X}$, $(n+1)X_{(1)}$, $\frac{n+1}{n}X_{(n)}$.

2 $X_1, \dots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Найдите наилучшую оценку параметра θ в равномерном подходе с квадратичной функцией потерь в классе оценок $\{\mu X_{(1)} :\ \mu \in \mathbb{R}\}$.

3 Пусть $X_1, \dots, X_n$ — выборка из распределения Лапласа со сдвигом,

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

Сравните в асимптотическом подходе оценки, полученные по методу моментов со стандартной пробной функцией и по методу максимального правдоподобия.

4 Пусть $X_1, \dots, X_n$ — выборка из нормального распределения с параметрами $(\theta, 1)$. Найдите байесовскую оценку параметра θ, если априорное распределение θ есть $Bin(1, p)$. Будет ли полученная оценка состоятельной оценкой параметра θ?

5 Пусть $X_1, \dots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Найдите байесовскую оценку параметра θ, если θ имеет априорное распределение

а) равномерное на отрезке $[0, 1]$.

б) с плотностью $q(t) = 1/t^2$ при $t \geq 1$.

Проверьте полученные оценки на состоятельность.

6* Пусть $X_1, \dots, X_n$ — выборка из нормального распределения с параметрами $(\theta, 1)$. Найдите байесовскую оценку параметра θ, если априорное распределение θ есть $\mathcal{N}(b, \sigma^2)$.

Эффективные оценки и информация Фишера

1 Пусть $X_1, \dots, X_n$ — выборка из биномиального распределения с параметрами (m, p), причем m известно. Найдите информацию Фишера в данной модели, а также эффективную оценку параметра p.

2 Пусть $X_1, \dots, X_n$ — выборка из экспоненциального распределения с параметром θ^{-1}. Найдите эффективную оценку параметра θ. Существует ли эффективная оценка параметра θ^{-1}?

3 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.

4* Пусть $X_1, \dots, X_n$ — выборка из нормального распределения с параметрами (a, σ^2). Найдите матрицу информации Фишера в данной модели. Существует ли эффективная оценка двумерного параметра $\theta = (a, \sigma^2)$?

Условные математические ожидания и условные распределения

1 Пусть $X_1, \dots, X_n$ — выборка из распределения с конечным математическим ожиданием. Найдите $\mathsf{E}(X_1 | \sum_{i=1}^n X_i)$.

2 Пусть $X_1, \dots, X_n$ — выборка из равномерного распределения на отрезке $[0, 1]$. Найдите а) $\mathsf{E}(X_1 | X_{(1)})$, б) $\mathsf{E}(X_1 | X_{(n)})$.

3 Пусть X и Y — независимые случайные величины, равномерно распределенные на отрезке $[0,2]$. Найдите $\mathsf{E}(Y|X/Y)$.

4 Пусть X и Y — независимые случайные величины, имеющие экспоненциальное распределение с параметром 1. Найдите $\mathsf{E}(Y|XY)$.

5* Пусть X и Y — независимые случайные величины, X имеет равномерное распределение на отрезке $[0,1]$, а Y — экспоненциальное с параметром 1. Найдите $\mathsf{E}(Y|X/Y)$.

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) $Geom(p)$, ж) $Beta(\lambda_1,\lambda_2)$.

2 Найдите оптимальную оценку параметра θ по выборке из распределения: а) $\mathcal{N}(\theta,1)$, б) $R(0,\theta)$, в) $Pois(\theta)$, г) $Bin(1,\theta)$.

3 Пусть $X_1,\dots,X_n$ — выборка из нормального распределения с параметрами (a,σ^2). Найдите оптимальную оценку параметра $\theta=(a,\sigma^2)$.

4 Пусть $X_1,\dots,X_n$ — выборка из экспоненциального распределения с параметром θ. Найдите оптимальные оценки θ и $\tau(\theta)=\theta^{1/2}$.

5 Пусть $X_1,\dots,X_n$ — выборка из нормального распределения с параметрами $(0,\theta^2)$. Найдите оптимальные оценки θ и $\tau(\theta)=\theta^2$.

6 Пусть $X_1,\dots,X_n$ — выборка из пуассоновского распределения с параметром $\theta>0$. Найдите $\mathsf{E}\left(X_1^2|\sum\limits_{i=1}^n X_i\right)$.

7* Пусть $X=(X_1,\dots,X_n)$ — выборка из показательного распределения с параметром $\theta>0$. Предположим, что наблюдаемы лишь r первых порядковых статистик: $X_{(1)},\dots,X_{(r)}$. Укажите в этих условиях достаточную статистику для θ и оптимальную оценку $\tau(\theta)=\theta^{-1}$.

8* Пусть $\theta_1^*(X)$ и $\theta_2^*(X)$ — две оптимальные оценки параметра θ. Докажите, что тогда для любого θ они совпадают почти наверное, т.е. $\theta_1^*(X)=\theta_2^*(X)$ P_θ-п.н.

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) $\overline{X}$, 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.

5* Пусть $X_1, \dots, X_n$ — выборка из нормального распределения с параметрами (a, σ^2). Постройте точную доверительную область уровня доверия α для двумерного параметра $\theta = (a, \sigma^2)$.

Гауссовские векторы

1 Пусть (X, Y) — гауссовский вектор, $(X, Y) \sim \mathcal{N}(a, \Sigma)$. Найдите $\mathsf{E}(X|Y)$.

2 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 $r = \sqrt{a^2 + b^2}$.

3 Приведите пример таких двух случайных величин X и Y, что обе они — стандартные нормальные, $\mathsf{E}XY = 0$, но вектор (X, Y) не является гауссовским.

4 Пусть случайные величины $\xi_1, \dots, \xi_n$ независимы и имеют стандартное нормальное распределение. Докажите, что случайная величина $\eta = \xi_1^2 + \ldots + \xi_n^2$ имеет распределение χ_n^2.

5 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 статистики $Y = \frac{1}{n}\sum_{i=1}^n |X_i|$, $Z = \frac{1}{n}\sum_{i=1}^n X_i^2$. Докажите, что оценка $T = Z^2/(4Y^3)$ является асимптотически нормальной оценкой σ, и найдите ее асимптотическую дисперсию.

6* Пусть $X_1, \dots, X_n$ — выборка из некоторого распределения с конечными первыми четырьмя моментами: a_1, a_2, a_3, a_4. Докажите, что статистика $S^2 = \frac{1}{n}\sum\limits_{i=1}^n (X_i - \overline{X})^2$ является асимптотически нормальной оценкой $\sigma^2 = \mathsf{D}X_i$, и найдите ее асимптотическую дисперсию.

7* Вычислите плотности для распределений Фишера и Стьюдента.

Оценки наименьших квадратов. F-критерий

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 $\beta_1 = ABD$, $\beta_2 = DBC$, $\beta_3 = BD$, $\beta_4 = BDA$ и неизвестной дисперсии σ^2.

2 Пусть $X_i = \beta_1 + i\beta_2 + \varepsilon_0 + \ldots + \varepsilon_i$, $i = 0, 1, \ldots, n$, где β_1, β_2 — неизвестные параметры, а $\varepsilon_0, \ldots, \varepsilon_n$ — 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 Постройте F-критерий уровня значимости α для проверки гипотезы $H_0:\ \beta_2 = \beta_1$ в предыдущей задаче.

4 $X_1, \ldots, X_n$ — выборка из распределения $\mathcal{N}(a_1, \sigma^2)$, $Y_1, \ldots, Y_m$ — выборка из распределения $\mathcal{N}(a_2, \sigma^2)$, $Z_1, \ldots, Z_k$ — выборка из распределения $\mathcal{N}(a_3, \sigma^2)$. Постройте F-критерий размера α для проверки гипотезы $H_0:\ a_1 = a_2$ и $a_1 + a_2 = a_3$.

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 Игральный автомат по закону не может быть запрограммирован выигрывать больше трети раз. С игральным автоматом сыграно 10 раз. Постройте критерий уровня значимости 0.95 для проверки гипотезы о честности автомата против альтернативы о его нечестности. При каком уровне значимости можно утверждать, что автомат нечестный, если игрок проиграл 10 раз? 9 раз?

3 $X_1,\ \ldots,\ X_n$ — выборка из пуассоновского 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) $H_1:\ \theta < \theta_0$.

4 $X_1,\ \ldots,\ X_n$ — 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 $H_1:\ \theta > \theta_0$,.

5 $X_1,\ \ldots,\ X_n$ — выборка из показательного распределения с параметром $\theta > 0$. Постройте равномерно наиболее мощный критерий уровня значимости α проверки гипотезы $H_0:\ \theta = \theta_0$ против альтернативы а) $H_1:\ \theta > \theta_0$, б) $H_1:\ \theta < \theta_0$.

6 Цифры $0, 1, 2, \ldots, 9$ 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 $\{0, 1, \ldots, 9\}$ на уровне значимости
а) 0.05, б) 0.5, в) 0.8.

7* $X_1,\ \ldots,\ X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$, $\theta > 0$. Постройте равномерно наиболее мощный критерий уровня значимости α проверки гипотезы $H_0:\ \theta = \theta_0$ против альтернативы $H_1:\ \theta \neq \theta_0$.