Задачи по курсу „Математическая статистика“

лектор — к.ф.-м.н. И.В. Родионов

Весна 2016

1. Сходимости случайных векторов

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

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

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

3 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.

4 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

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

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

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

3 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.

4 Пусть $X_1, ..., X_n$ выборка из распределения с параметром σ^2. Пусть, кроме того, $\mathsf{D}X_1 = \sigma^2$. Докажите, что статистика $s^2 = 1/n \sum_{i=1}^n (X_i - \overline{X})^2$ равна $\overline{X^2} - (\overline{X})^2$ и является состоятельной оценкой σ^2. Является ли она несмещенной оценкой того же параметра?

5 Пусть $X_1, \ldots, X_n$ — 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.

3. Основные методы нахождения оценок

1 Найдите оценки по методу моментов 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) $Beta(\lambda_1, \lambda_2)$, з) распределения Парето с плотностью $p(x) = \gamma x^{-\gamma-1} I\{x > 1\}$, и) $Cauchy(\theta)$.

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) $Geom(p)$, ж) для распределения с плотностью $p(x) = \frac{2\theta^2}{x^3} I\{x > \theta\}$.

3 $X_1, \ldots, X_n$ — 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 Коши,

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

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

5 Предложите асимптотически нормальную оценку параметра θ^2 в модели распределения Коши со сдвигом (см. задача 3.4), а также найдите её асимптотическую дисперсию.

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

2 Пусть $\theta_1^*(X)$ и $\theta_2^*(X)$ — две наилучшие в среднеквадратическом подходе 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 Пусть $X_1, \dots, X_n$ — выборка из распределения с конечным математическим ожиданием, $g : \mathbb{R}^n \to \mathbb{R}$ — симметричная (значение не меняет при любой перестановке аргументов) борелевская функция. Докажите, что $\forall i, j \in \{1, ..., n\}$ справедливо равенство $\mathsf{E}(X_i | g(X_1, ..., X_n)) = \mathsf{E}(X_j | g(X_1, ..., X_n))$. Найдите $\mathsf{E}(X_1 | \sum_{i=1}^n X_i)$.

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

3 Случайная величина X имеет равномерное распределение на отрезке $[-1, 1]$. Вычислите $\mathsf{E}(X|X^2)$.

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

5 Пусть X и Y — независимые случайные величины, X имеет равномерное распределение на отрезке $[0, 1]$, а Y — экспоненциальное с параметром 1. Найдите $\mathsf{E}(Y^3/X^2 | X/Y)$.

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 $X_1, \ldots, X_n$ — 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 $X_1, \ldots, X_n$ — выборка из экспоненциального распределения с параметром $\theta > 0$. Найдите оптимальные оценки для θ и $\tau(\theta) = \theta^{1/2}$.

5 Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с параметрами $(0, \theta^2)$. Найдите оптимальную оценку для θ.

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

7. Доверительные интервалы

1 Пусть $X_1, \ldots, X_n$ — 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 $X_1, \ldots, X_n$ — выборка из распределения плотностью $p_\theta(x) = \frac{3x^2}{8\theta^3} I\{x \in [0, 2\theta]\}$. С помощью статистики $X_{(1)}$ постройте точный доверительный интервал уровня доверия γ для параметра θ.

3 Пусть $X_1, \ldots, X_n$ — 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.

4 Пусть $X_1, \ldots, X_n$ — выборка из пуассоновского распределения с параметром θ. Постройте асимптотический доверительный интервал для θ уровня доверия α.

5 Пусть $X_1, \ldots, X_n$ — 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. Байесовские оценки.

1 Пусть $X_1, \ldots, X_n$ – 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?

2 Пусть $X_1, \ldots, X_n$ – выборка из равномерного распределения на отрезке $[0, \theta]$. Найдите байесовскую оценку параметра θ, если θ имеет априорное распределение
(a) равномерное на отрезке $[0, 1]$;
(b) с плотностью $q(t) = 1/t^2$ при $t \geq 1$.
Проверьте полученные оценки на состоятельность.

3 Пусть $X_1, \ldots, X_n$ – выборка из нормального распределения с параметрами $(\theta, 1)$. Найдите байесовскую оценку параметра θ, если априорное распределение θ есть $N(a, \sigma^2)$.

4 Пусть $X_1, \ldots, X_n$ – выборка из распределения
(a) $N(\theta, 1)$,
(b) $N(0, \theta)$, где $\theta > 0$,
(c) $Bin(n, p)$, где $p \in (0, 1)$.
Подберите сопряженное распределение и найдите байесовскую оценку. Сравните ее с оценкой максимального правдоподобия.

9. 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 $X_i = \beta_1 + i\beta_2 + \varepsilon_0 + \ldots + \varepsilon_i$, $i = 0, 1, \ldots, n$, 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 Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с параметрами (a, σ^2). Докажите, что статистики $\overline{X}$ и $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ независимы и вычислите распределение статистики nS^2.

4 Пусть $X_i, i \in \{1, \ldots, n\}$ — независимые случайные величины, распределенные по нормальному закону с параметрами $(a + bi, \sigma^2)$. Постройте точные доверительные интервалы для параметров a, b, σ^2.

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

4 $X_1, \ \ldots, \ X_n$ — выборка из распределения $Bern(\theta)$. 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 $X_1, \ \ldots, \ X_n$ — 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

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

11. Критерий Колмогорова, критерий хи-квадрат, F-критерий

1 Пусть $X_1, \ldots, X_n$ — выборка из распределения $\mathcal{N}(a_1, \sigma_1^2)$, $Y_1, \ldots, Y_m$ — выборка из распределения $\mathcal{N}(a_2, \sigma_2^2)$. Предложить критерий для проверки гипотезы $H_0: \ \sigma_1^2 = \sigma_2^2$.

2 Доказать состоятельность критерия Колмогорова.

3 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 0.05 и 0.1. (квантили распределения хи-квадрат можно в интернете найти).

4 Постройте F-критерий уровня значимости α для проверки гипотезы $H_0: \ \beta_2 = \beta_1$ в задаче 9.2.

5 $X_1, \ldots, X_n$ — 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 $a_1 + a_2 = a_3$.

12. Коэффициенты корреляции.

1 Получить коэффициенты корреляции Пирсона, Спирмэна и Кэндалла с помощью обобщённого коэффициента корреляции.

2 Доказать, что в случае верности гипотезы о независимости выборок $D\rho_S = \frac{1}{n-1}$, где ρ_S — коэффициент корреляции Спирмэна.

3 Темпы роста ВВП России за 2006-2012 в процентах (всего и на душу населения, 2005 - 100%): 108,2 117,4 123,5 113,9 119,0 124,1 128,4 108,5 118,0 124,2 114,5 119,6 124,6 128,7. Те же цифры для Украины: 107 116 118 101 105 111 111 108 117 121 103 108 114 115. Проверить гипотезу о независимости данных выборок.

4 Переставим использованные при введении коэффициента корреляции Спирмэна пары рангов (R_i, S_i), $i = 1, \dots, n$, в порядке возрастания первой компоненты — получается набор $(1, T_1), \dots, (n, T_n)$. Доказать, что $\rho_S = 1 - \frac{12}{n^3-n} \sum\limits_{i<j} (j-i) I\{T_i > T_j\}$ и $\tau = 1 - \frac{4}{n^2-n} \sum\limits_{i<j} I\{T_i > T_j\}$, где τ — коэффициент корреляции Кэндалла.