

6 сентября 2026 г.

Задача 1.

Пусть $\{X_i\}_{i=1}^n$ — *н.о.р.* выборка из распределения с функцией распределения $F_X(x)$. Обозначим за $\hat{F}(x)$ э.ф.р., построенную по этой выборке.

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq x\}.$$

- а) Возьмём случайную величину с плотностью $f_X(x) = 3x^2, x \in [0, 1]$. Известна реализация выборки с $n = 4 : x_1 = 0.3, x_2 = 0.65, x_3 = 0.8, x_4 = 1$. Постройте график получившейся э.ф.р. $\hat{F}(x)$ и сравните его с графиком $F_X(x)$.
- б) Пользуясь тем, что при фиксированном x индикаторы $\mathbb{1}\{X_i < x\}$ являются *н.о.р.* бернуллиевскими случайными величинами, найдите $\mathbb{E}\hat{F}(x)$ и $\mathbb{V}\hat{F}(x)$.
- в) Рассмотрим случайные функции вида

$$Z_x^{(n)} = \sqrt{n}(\hat{F}(x) - F_X(x)), \quad x \in [0, 1].$$

Для произвольных $0 < x_1 < x_2 < 1$ найдите $\lim_{n \rightarrow \infty} \text{cov}(Z_{x_1}^{(n)}, Z_{x_2}^{(n)})$.

Задача 2.

Пусть $X_1, \dots, X_n$ — случайная выборка из распределения, в котором $\mathbb{E}X_1 = m, \mathbb{V}X_1 = \sigma^2$.

- а) Докажите, что статистика $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ равна $\bar{X}^2 - (\bar{X})^2$ и является состоятельной оценкой σ^2 . *Подсказка:* используйте закон больших чисел и теорему Манна-Вальда.
- б) Является ли она несмещённой оценкой того же параметра?
- в) Аргументируйте, что $\sqrt{n}(\bar{X} - m) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.
- г) Докажите, что $\sqrt{\frac{\sigma^2}{S^2}} \frac{\sqrt{n}(\bar{X} - m)}{\sqrt{\sigma^2}} \xrightarrow{d} \mathcal{N}(0, 1)$, где S^2 — несмещённая оценка дисперсии. Изменится ли этот результат, если S^2 заменить на $\hat{\sigma}^2$?
- д) $C(\alpha)$ статистики, предложенные Нейманом, имеют общий вид

$$Z_n(b) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{f(X_i) - \mathbb{E}_b(f(X_i))}{\sigma_b}, \quad (1)$$

где b — некоторый параметр распределения, $f(\cdot)$ — произвольная измеримая функция, такая, что $f(X)$ имеет конечную дисперсию σ_b^2 ; также $\mathbb{E}_b(f(X_i))$ — математическое ожидание $f(X_i)$ при настоящем значении b .

Нейман показал, что условие ортогональности $\frac{\partial}{\partial b} \mathbb{E}_b(f(X_i)) = 0$ является необходимым и достаточным для $Z_n(\hat{b}) \xrightarrow{d} \mathcal{N}(0, 1)$, где $\hat{b}$ — любая асимптотически нормальная оценка b .

Рассмотрите выражение $\frac{\sqrt{n}(\bar{X}-m)}{\sqrt{\sigma^2}}$, как пример $C(\alpha)$ статистики Неймана и проверьте условие ортогональности для неё.

Задача 3.

Пусть $\{X_i\}_{i=1}^n$ – случайная выборка из распределения $\Gamma(a, b)$ с неизвестными параметрами. Плотность распределения имеет вид

$$p(x; a, b) = \frac{b^{-a} x^{a-1}}{\Gamma(a)} e^{-\frac{x}{b}} \mathbb{1}(x \geq 0),$$

где $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$.

- Найдите выражения для $\mathbb{E}X_1$ и $\mathbb{V}X_1$. Предположим, что мы хотим тестировать гипотезу $H_0 : a = a_0$ против альтернативы $H_1 : a \neq a_0$; найдите $\hat{b}$ – оценку для b , состоятельную при H_0 .
- Рассмотрите выражение $Z_n(b) = \frac{\sqrt{n}(\bar{X} - \mathbb{E}X_1)}{\sqrt{\mathbb{V}X_1}}$ как пример $C(\alpha)$ статистики Неймана (см. определение в задаче #2 п. е). Проверьте выполнение условия ортогональности для используемой функции. Будет ли $Z_n(\hat{b})$ иметь предел по распределению $\mathcal{N}(0, 1)$?
- Постройте ортогонализированную статистику для этого же теста, имеющую вид

$$B_n(b) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\partial}{\partial a} \ell_i(a, b) - \frac{I_{ab}}{I_{bb}} \frac{\partial}{\partial b} \ell_i(a, b) \right],$$

где I_{ab}, I_{bb} – элементы информационной матрицы Фишера, $\ell_i(a, b)$ – логарифмическая функция правдоподобия для i -го наблюдения.

- Проведите Монте-Карло эксперимент, заключающийся в генерации выборки $\{X_i\}_{i=1}^n$ и вычислении статистик $Z_n(\hat{b})$ и $B_n(\hat{b})$ (в предположении, что нулевая гипотеза верна, т.е. при a_0 соответствующем популяционному параметру). Постройте гистограммы значений полученных статистик с наложенной плотностью $\mathcal{N}(0, 1)$.

Задача 4.

Целью этой задачи является проверка доверительных интервалов для параметра равномерного распределения.

Сгенерируйте случайную выборку $\{X_i\}_{i=1}^n$ размером $n = 20$ из равномерного распределения $U[0, \theta]$, где θ = число букв в вашей фамилии. Используя полученную выборку, постройте статистики $X_{(n)}$, а также $Z = \overline{X^3}$ и $Y = \overline{X^2}$, аналогичные задаче #3 из семинара #1, – и вычислите следующие доверительные интервалы для параметра θ с $\alpha = 0.05$:

- Асимптотический доверительный интервал на основе $\frac{4Z}{3Y}$;
- Доверительный интервал (точный или асимптотический) на основе оценки θ методом максимального правдоподобия;
- Непараметрический бутстрап-интервал методом Холла на основе $\frac{4Z}{3Y}$;
- Параметрический бутстрап-интервал на основе $\frac{4Z}{3Y}$. В этом случае бутстрап-выборки – случайные выборки нужного размера из распределения $U[0, \hat{\theta}]$.

- д) Непараметрический бутстрап-интервал методом Холла на основе оценки максимального правдоподобия.
- е) Сравните ширину доверительных интервалов и для каждого варианта бутстрапа и оцените его фактический уровень доверия – то есть, сделайте Монте-Карло эксперимент, в котором генерируется 10^4 выборок, аналогичных исходной, строятся доверительные интервалы и вычисляется доля случаев, когда построенный интервал накрыл истинное значение параметра. Отличаются ли полученные значения от расчётного уровня?

Задача 5.

Используя теоретические соображения, основанные на разложении Эджворта, предложите (обоснованно) асимптотически стандартные нормальные статистики со следующими асимптотиками приближения по ЦПТ:

- а) Приближение имеет скорость порядка $n^{-1/2}$.
- б) Приближение имеет скорость порядка n^{-1} .
- с) Проведите Монте-Карло эксперимент, подтверждающий ваши предположения. Для этого постройте э.ф.р. статистики при разных размерах выборки n и оцените расстояния от этих функции до ф.р. $\mathcal{N}(0, 1)$, используя равномерную и L^2 метрику. После этого постройте зависимость метрик расстояния от n (удобнее использовать логарифмическую шкалу).
- д) Вычислите 3й и 4й кумулянты выбранных статистик из п. а) и б) и постройте примеры графиков их э.ф.р. с наложенными приближениями по Эджворту до 3го и 4го порядка;
- е) Проведите аналогичный Монте-Карло эксперимент со статистикой из п. б), но посчитанной для биномиального распределения $B(15, 0.6)$. Отличается ли для неё скорость сходимости? Если да, то почему?