

# Теория вероятностей и математическая статистика, Коллоквиум IV

Версия от 13.06.2021 17:40

## Содержание

|      |   |    |
|------|---|----|
| 1.   | Выборка, оценка, статистика. Несмещенность, состоятельность, асимптотическая нормальность и эффективность оценок. Пример отсутствия несмещенной оценки. Отсутствие эффективной оценки в классе всех оценок. Единственность эффективной оценки. Состоятельность асимптотической нормальной оценки. | 2  |
| 1.1. | Выборка, оценка, статистика.  | 2  |
| 1.2. | Несмещенность, состоятельность, асимптотическая нормальность и эффективность оценок.  | 2  |
| 1.3. | Пример отсутствия несмещенной оценки.   | 3  |
| 1.4. | Отсутствие эффективной оценки в классе всех оценок.   | 3  |
| 1.5. | Состоятельность асимптотической нормальной оценки.  | 4  |
| 2.   | Метод моментов и его состоятельность. Метод максимального правдоподобия. Энтропия и состоятельность оценки максимального правдоподобия.   | 5  |
| 2.1. | Метод моментов.   | 5  |
| 2.2. | Метод максимального правдоподобия.  | 5  |
| 3.   | Информация Фишера и неравенство Рао-Крамера. Критерий равенства в неравенстве Рао-Крамера.  | 8  |
| 4.   | Доверительные интервалы. Различные методы построения доверительных интервалов (с помощью неравенств на вероятность больших отклонений, с помощью центральной статистики, с помощью асимптотически нормальной оценки). Примеры.  | 8  |
| 4.1. | Доверительные интервалы   | 8  |
| 4.2. | Различные методы построения доверительных интервалов (с помощью неравенств на вероятность больших отклонений, с помощью центральной статистики, с помощью асимптотически нормальной оценки)   | 8  |
| 5.   | Построение точных доверительных интервалов для параметров нормального распределения.  | 10 |
| 6.   | Проверка гипотез. Ошибки 1-го и 2-го рода. Уровень значимости и мощность статистического критерия. Пример построения критерия с помощью доверительного интервала. Нижняя оценка суммы вероятностей ошибок 1-го и 2-го рода.   | 12 |
| 6.1. | Проверка гипотез.   | 12 |
| 6.2. | Ошибки 1-го и 2-го рода.  | 12 |
| 6.3. | Уровень значимости и мощность статистического критерия.   | 12 |
| 6.4. | Пример построения критерия с помощью доверительного интервала.  | 12 |
| 6.5. | Нижняя оценка суммы вероятностей ошибок 1-го и 2-го рода.   | 13 |
| 7.   | Теорема Неймана-Пирсона и пример её применения.   | 14 |
| 7.1. | Теорема Неймана-Пирсона.  | 14 |
| 7.2. | Пример применения теоремы Неймана-Пирсона.  | 14 |
| 8.   | Эмпирическая функция распределения. Теорема Гливенко-Кантелли.  | 15 |

# 1. Выборка, оценка, статистика. Несмещенность, состоятельность, асимптотическая нормальность и эффективность оценок. Пример отсутствия несмещенной оценки. Отсутствие эффективной оценки в классе всех оценок. Единственность эффективной оценки. Состоятельность асимптотической нормальной оценки.

Предположим, нам известно, что неизвестное распределение принадлежит какому-то конкретному семейству распределений с функциями распределения  $F_\theta$ , где  $\theta \in \Theta \subseteq \mathbb{R}^k$ . Тогда задачей статистики является оценка неизвестного параметра  $\theta_0 \in \Theta$ , соответствующего нашему неизвестному распределению.

Например, пусть  $X$  есть случайная величина и мы знаем распределение этой случайной величины  $F_\theta(t)$  с точностью до  $\theta$  (например,  $\mathcal{N}(0, \theta)$ ). Задача статистики заключается в том, чтобы оценить параметр  $\theta$ .

Рассмотрим пример, показывающий, что теория вероятностей и математическая статистика изучают разные вещи:

**Пример.** Пусть в ящике  $N$  шаров,  $M$  из них чёрные. Мы достали из ящика  $n$  шаров. Теория вероятностей задаётся вопросом, с какой вероятностью среди вытянутых шаров есть  $m$  чёрных. Математическая статистика задаётся вопросом, сколько всего в ящике чёрных шаров (какое  $M$ ), если мы достали  $n$  шаров и  $m$  из них чёрные.

## 1.1. Выборка, оценка, статистика.

**Определение.** Вектор  $X = (X_1, \dots, X_n)$  с независимыми компонентами, где каждая случайная величина имеет одно и то же распределение, называется **выборкой**.

**Определение.** Произвольная функция  $T_n(X)$ , принимающая выборку как аргумент, называется **статистикой**.

Важно то, что статистика зависит от случайных величин  $X_1, \dots, X_n$ , и не зависит от  $\theta$ .

**Пример.** Выборочное среднее  $\overline{X_n}$  является статистикой

$$T(X) = \overline{X_n} = \frac{X_1 + \dots + X_n}{n}.$$

Когда проводится серия независимых экспериментов с функцией распределения  $F_\theta$  ( $\theta$  неизвестно), мы получаем выборку  $X = (X_1, \dots, X_n)$ . По выборке хочется определить значение  $\hat{\theta}$ , которое в каком-либо смысле близко к реальному  $\theta$ .

**Определение.** Статистика  $\hat{\theta}(X_1, \dots, X_n)$  со значением из множества параметров  $\Theta$  называется **оценкой** неизвестного параметра.

## 1.2. Несмещенность, состоятельность, асимптотическая нормальность и эффективность оценок.

Хотим, чтобы оценка  $\hat{\theta}(X)$  была в каком-то смысле близка к реальному  $\theta$ . Далее описаны вещи, под которыми можно понимать близость.

**Определение.** Оценка  $\hat{\theta}_n(X)$  является **несмещенной**, если  $E_\theta [\hat{\theta}_n(X)] = \theta$ .

Напомним, что последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  *по вероятности*, если для любого  $\varepsilon > 0$  выполняется

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \varepsilon] = 0.$$

**Определение.** Оценка  $\hat{\theta}_n(X)$  является **состоятельной**, если  $\hat{\theta}_n(X) \xrightarrow[n \rightarrow \infty]{P_\theta} \theta$  для любого  $\theta \in \Theta$ .

Обычно состоятельность оценки является следствием закона больших чисел.

**Пример.** Пусть  $\hat{\theta}_n = \overline{X_n} = \frac{X_1 + \dots + X_n}{n}$  и  $E[X_1] = \theta$ . Тогда  $\hat{\theta}_n = \overline{X_n} \xrightarrow[n \rightarrow \infty]{\text{ЗБЧ}} E[X_1] = \theta$ , откуда следует, что оценка  $\hat{\theta}_n$  состоятельная.

Напомним, что последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  *почти наверное*, если выполняется  $P\left[\lim_{n \rightarrow \infty} X_n = X\right] = 1$ .

**Определение.** Оценка  $\hat{\theta}_n(X)$  является **сильно состоятельной**, если  $\hat{\theta}_n(X) \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \theta$  для любого  $\theta \in \Theta$ .

Напомним, что последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  *по распределению*, если  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  в каждой точке  $x$ , где непрерывна  $F_X$ .

**Определение.** Оценка  $\hat{\theta}_n(X)$  является **асимптотически нормальной** оценкой параметра  $\theta$  с коэффициентом  $\sigma^2(\theta)$ , если

$$\sqrt{n} \left( \hat{\theta}_n(X) - \theta \right) \xrightarrow[n \rightarrow \infty]{d_\theta} \mathcal{N}(0, \sigma^2(\theta)),$$

что эквивалентно  $\frac{\sqrt{n} \left( \hat{\theta}_n(X) - \theta \right)}{\sigma(\theta)} \xrightarrow[n \rightarrow \infty]{d_\theta} \mathcal{N}(0, 1)$ .

Коэффициент  $\sigma^2(\theta)$  называется асимптотической дисперсией.

**Определение.** Пусть  $K$  это некоторое множество (класс) оценок (например,  $K$  — несмещенные оценки).

Оценка  $\hat{\theta}_n(X) \in K$  является **эффективной в классе  $K$** , если для каждого  $\theta \in \Theta$  и для каждой оценки  $\theta_n^* \in K$  выполняется

$$\mathbb{E}_\theta \left[ \hat{\theta}_n(X) - \theta \right]^2 \leq \mathbb{E}_\theta \left[ \theta_n^*(X) - \theta \right]^2.$$

Если  $K$  это класс несмещенных оценок, то условие можно переписать следующим образом:

$$\mathbb{D}_\theta \left[ \hat{\theta}_n(X) \right] \leq \mathbb{D}_\theta \left[ \theta_n^*(X) \right]$$

**Определение.** Несмещенную оценку  $\hat{\theta}_n(X)$  в классе всех несмещенных оценок будем называть просто **эффективной**.

### 1.3. Пример отсутствия несмещенной оценки.

**Пример.** Пусть  $X_i$  это случайная величина Бернулли, то есть  $X_i$  принимает значение 1 с вероятностью  $p \in (0; 1)$  и 0 с вероятностью  $1 - p$ , и  $\theta = \sin(p)$ .

Запишем математическое ожидание оценки по определению:

$$\mathbb{E} \left[ \hat{\theta}_n(X) \right] = \sum_{(x_1, \dots, x_n) \in \{0, 1\}^n} \hat{\theta}_n(x_1, \dots, x_n) \cdot p^{x_1 + \dots + x_n} \cdot (1 - p)^{n - x_1 - \dots - x_n}.$$

То есть математическое ожидание оценки это некий полином от  $p$ , а полином не может быть равняться  $\sin p$ , поэтому оценка не может быть несмещенной. Вместо синуса можно было рассмотреть что-то другое.

### 1.4. Отсутствие эффективной оценки в классе всех оценок.

**Утверждение.** Не существует эффективной оценки в классе всех оценок.

*Доказательство.* Пусть  $\hat{\theta}_n(X)$  эффективна в классе всех оценок. Тогда должно выполняться

$$\mathbb{E}_\theta \left[ \hat{\theta}_n(X) - \theta \right]^2 \leq \mathbb{E}_\theta \left[ \theta_n^*(X) - \theta \right]^2$$

для любой  $\theta_n^*(X)$ . В том числе, это должно выполняться для любой  $\theta_n^*(X) = C = \text{const}$ :

$$\mathbb{E}_\theta \left[ \hat{\theta}_n(X) - \theta \right]^2 \leq \mathbb{E}_\theta \left[ C - \theta \right]^2.$$

Это неравенство также должно выполняться для любого  $\theta$ . В частности, для  $\theta = C$ . Тогда

$$\mathbb{E}_\theta \left[ \hat{\theta}_n(X) - \theta \right]^2 \leq 0.$$

Это означает, что  $\hat{\theta}_n(X) = C$  почти наверное. Но мы ведь могли взять и другое  $C' \neq C$  и ровно по тем же соображениям получить

$$\hat{\theta}_n(X) = C \neq C' = \hat{\theta}_n(X).$$

Пришли к противоречию. ■

Следующего утверждения нет в программе коллоквиума, но тем не менее оно полезное.

**Утверждение.** Если эффективная оценка существует, то она единственная.

*Доказательство.* Пусть  $\hat{\theta}_n(X)$  и  $\theta_n^*(X)$  — эффективные оценки.

Тогда по определению эффективности:

- $E_\theta [\hat{\theta}_n(X)] = \theta = E_\theta [\theta_n^*]$  для любого  $\theta$ .
- $D_\theta [\hat{\theta}_n(X)] \leq D_\theta [\tilde{\theta}(X)]$  и  $D_\theta [\theta_n^*(X)] \leq D_\theta [\tilde{\theta}(X)]$  для любой несмещенной  $\tilde{\theta}_n(X)$ .

Рассмотрим полусумму  $T_n = \frac{\hat{\theta}_n(X) + \theta_n^*(X)}{2}$ .  $T_n$  является несмещенной оценкой (по свойству для суммы математического ожидания).

Кратко напомним, что  $\text{cov}(\xi, \nu)$  это билинейная форма и по неравенству параллелограмма

$$\text{cov}(\xi + \nu, \xi + \nu) + \text{cov}(\xi - \nu, \xi - \nu) = 2 \cdot \text{cov}(\xi, \xi) + 2 \cdot \text{cov}(\nu, \nu).$$

Делим обе части на 4, вспоминаем, что  $D[\xi] = \text{cov}(\xi, \xi)$ , и получаем

$$D \left[ \frac{\xi + \nu}{2} \right] + D \left[ \frac{\xi - \nu}{2} \right] = \frac{1}{2} D[\xi] + \frac{1}{2} D[\nu].$$

Тогда

$$\begin{aligned} D_\theta \left[ \frac{\hat{\theta}_n(X) + \theta_n^*(X)}{2} \right] &= \frac{D[\hat{\theta}_n(X)] + D[\theta_n^*(X)]}{2} - D_\theta \left[ \frac{\hat{\theta}_n(X) - \theta_n^*(X)}{2} \right] \\ &[\text{вычитается неотрицательное значение (дисперсия)}] \leq \frac{D[\hat{\theta}_n(X)] + D[\theta_n^*(X)]}{2} \\ &[\text{так как оценки эффективные}] \leq D_\theta \left[ \frac{\hat{\theta}_n(X) + \theta_n^*(X)}{2} \right]. \end{aligned}$$

Последнее равенство может показаться неочевидным. Так как  $\hat{\theta}_n(X)$  это эффективная оценка, выполняется неравенство

$$D_\theta [\hat{\theta}_n(X)] \leq D_\theta \left[ \frac{\hat{\theta}_n(X) + \theta_n^*(X)}{2} \right] \iff \frac{1}{2} D_\theta [\hat{\theta}_n(X)] \leq \frac{1}{2} D_\theta \left[ \frac{\hat{\theta}_n(X) + \theta_n^*(X)}{2} \right].$$

Аналогичное неравенство получается и для оценки  $\theta_n^*(X)$ . Суммируем эти два неравенства и получаем

$$\frac{D_\theta [\hat{\theta}_n(X)] + D_\theta [\theta_n^*(X)]}{2} \leq D_\theta \left[ \frac{\hat{\theta}_n(X) + \theta_n^*(X)}{2} \right].$$

Мы закончили тем же, с чего и начали. Тогда все неравенства это равенство. Значит,  $D_\theta \left[ \frac{\hat{\theta}_n(X) - \theta_n^*(X)}{2} \right] = 0$ .

Из этого следует, что

$$\hat{\theta}_n(X) - \theta_n^*(X) = E_\theta [\hat{\theta}_n(X) - \theta_n^*(X)] = E_\theta [\hat{\theta}_n(X)] - E_\theta [\theta_n^*(X)] = \theta - \theta = 0.$$

Получается,  $\hat{\theta}_n(X) = \theta_n^*(X)$  почти наверное. ■

### 1.5. Состоятельность асимптотической нормальной оценки.

**Утверждение.** Если оценка асимптотически нормальная, то она состоятельная.

*Доказательство.* Пусть  $\hat{\theta}_n(X)$  — асимптотически нормальная оценка параметра  $\theta$ .

По определению асимптотической нормальности  $\sqrt{n} (\hat{\theta}_n(X) - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2(\theta))$ .

Про сходимость по распределению произведения мы знаем, что если один из пределов сходится к константе, то верна арифметика:

$$\frac{1}{\sqrt{n}} \cdot Z \xrightarrow[n \rightarrow \infty]{d} 0 \cdot Z = 0.$$

Тогда

$$\hat{\theta}_n(X) - \theta = \frac{1}{\sqrt{n}} \cdot \sqrt{n} (\hat{\theta}_n(X) - \theta) \xrightarrow[n \rightarrow \infty]{d_\theta} 0 \cdot \mathcal{N}(0, \sigma^2(\theta)) = 0.$$

Так как есть сходимость по распределению к константе, то есть сходимость по вероятности к константе (факт с предыдущего коллоквиума):

$$\hat{\theta}_n(X) - \theta \xrightarrow[n \rightarrow \infty]{d_\theta} 0 \iff \hat{\theta}_n(X) \xrightarrow[n \rightarrow \infty]{P_\theta} \theta,$$

а это и есть определение состоятельности. ■

## 2. Метод моментов и его состоятельность. Метод максимального правдоподобия. Энтропия и состоятельность оценки максимального правдоподобия.

### 2.1. Метод моментов.

Пусть  $(X_1, \dots, X_n)$  — выборка, где  $X_i$  задано распределением  $F_\theta$ . Хотим найти состоятельную оценку параметра  $\theta$ . Пусть  $g$  — непрерывная функция, причём  $\mathbb{E}_\theta (|g(X_1)|) < \infty$ . Посчитаем матожидание  $\mathbb{E}_\theta (g(X_1)) = f(\theta)$ . Предположим, что  $\exists f^{-1}$ , и она непрерывна. Т.к. на практике мы не знаем параметр  $\theta$ , то мы не можем посчитать такое матожидание. Но мы можем приближённо посчитать  $f(\theta)$  воспользовавшись ЗБЧ.

По ЗБЧ

$$\frac{g(X_1) + \dots + g(X_n)}{n} \xrightarrow{P} \mathbb{E}_\theta (g(X_1)) = f(\theta)$$

Теперь в силу обратимости  $f$  можно получить сходимость к  $\theta$ :

$$f^{-1} \left( \frac{g(X_1) + \dots + g(X_n)}{n} \right) \xrightarrow{P} \theta$$

Оценкой параметра  $\theta$  назовём функцию  $\hat{\theta}(X_1, \dots, X_n) = f^{-1} \left( \frac{g(X_1) + \dots + g(X_n)}{n} \right)$ . Состоятельность оценки очевидна. Вместо ЗБЧ можно применять УЗБЧ, и получить сильную состоятельность.

### 2.2. Метод максимального правдоподобия.

**Определение 1.** Обобщённой плотностью  $\rho_X$  случайной величины  $X$  назовём функцию плотности  $X$ , если случайная величина является непрерывно, или функцию  $\rho_X(t) = P(X = t)$  в случае, если  $X$  имеет дискретное распределение.

**Определение 2.** Пусть  $X = (X_1, \dots, X_n)$  — выборка из распределения с обобщённой плотностью  $\rho_\theta$ . Обобщённая плотность вектора  $X$  называется функцией правдоподобия, и имеет вид

$$p(X, \theta) = \rho_\theta(X_1) \cdot \dots \cdot \rho_\theta(X_n)$$

Функцию  $\ln p(X, \theta)$  называют логарифмической функцией правдоподобия и обозначают  $L(X, \theta)$ .

**Определение 3.** Пусть  $\rho_0, \rho_1$  — положительные вероятностные плотности. Выражение

$$\int \ln \frac{\rho_1(x)}{\rho_0(x)} \cdot \rho_1(x) dx$$

называется энтропией распределения с плотностью  $\rho_1$  относительно распределения с плотностью  $\rho_0$ .

**Замечание.** Здесь и далее интегралы без пределов интегрирования обозначают интегрирование по множеству, на котором задано распределение. Они вовсе не означают неопределённый интеграл.

Следующее утверждение показывает, что энтропия в некотором смысле оценивает расстояние между распределениями:

**Лемма.** (Информационное неравенство)

Пусть  $\rho_0, \rho_1$  — положительные вероятностные плотности. Тогда

$$\int \ln \frac{\rho_1(x)}{\rho_0(x)} \cdot \rho_1(x) dx \geq 0$$

Равенство достигается тогда и только тогда, когда  $\rho_0 = \rho_1$ .

*Доказательство.* Домножим обе части неравенства на  $(-1)$  и будем выводить оценку сверху:

$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x) dx \stackrel{?}{\leq} 0$$

Воспользуемся неравенством  $\ln x \leq x - 1$  (очевидно, если, например, посмотреть на графики этих функций: у них есть единственное пересечение в точке  $x = 1$ ):

$$\ln \frac{\rho_0(x)}{\rho_1(x)} \rho_1(x) \leq \left( \frac{\rho_0}{\rho_1} - 1 \right) \rho_1 = \rho_0 - \rho_1$$

Проинтегрируем обе части:

$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x) dx \leq \int \rho_0(x) dx - \int \rho_1(x) dx$$

Оба интеграла слева равны 1, в силу того, что под интегралами стоят плотности. Таким образом оценку сверху мы доказали, найдём теперь, когда достигается равенство.

Пусть в неравенстве достиглось равенство, т.е. известно, что

$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x) dx = 0 \quad \int \rho_0(x) dx - \int \rho_1(x) dx = 0 \iff \int \left( \frac{\rho_0}{\rho_1} - 1 \right) dx = 0$$

Тогда

$$\int \left( \left( \frac{\rho_0}{\rho_1} - 1 \right) - \ln \frac{\rho_0}{\rho_1} \right) \rho_1 dx = 0$$

Так как  $\ln x \leq x - 1$ , то  $0 \leq x - 1 - \ln x$ , и функция в скобках неотрицательна. Теперь очевидно, что 0 достигается только в случае  $\rho_0 = \rho_1$ :

$$\left( \frac{\rho_0}{\rho_1} - 1 \right) - \ln \frac{\rho_0}{\rho_1} = 0 \iff \left( \frac{\rho_0}{\rho_1} - 1 \right) = \ln \frac{\rho_0}{\rho_1} \iff \rho_0 = \rho_1$$

■

Вывели утверждение, которое показывает, что энтропия, в некотором смысле оценивает расстояние между плотностями, т.е. расстояние между распределениями. Теперь будем применять это утверждение для построения оценки.

Пусть есть выборка  $X = (X_1, \dots, X_n)$  из распределения с обобщённой плотностью  $\rho_\theta$ . Пусть реальное значение параметра  $\theta$  равно  $\theta_1$ . Рассмотрим функцию следующего вида:

$$W(\theta) = \mathbb{E}_{\theta_1} \ln \rho_\theta(X_1) = \int \ln \rho_\theta(x) \rho_{\theta_1}(x) dx$$

Можно показать, что  $W(\theta) \leq W(\theta_1) \forall \theta$ , действительно:

$$W(\theta) - W(\theta_1) = \int \ln \rho_\theta(x) \rho_{\theta_1}(x) dx - \int \ln \rho_{\theta_1}(x) \rho_{\theta_1}(x) dx = \int \ln \frac{\rho_\theta(x)}{\rho_{\theta_1}(x)} \rho_{\theta_1}(x) dx \leq 0$$

Причём наибольшее значение  $W(\theta)$  достигается при  $\theta = \theta_1$ . Таким образом можно естественно оценить реальный параметр, если найти точку максимума функции  $W(\theta)$ . В чём проблема: мы не знаем  $\rho_{\theta_1}$ , и потому функция  $W(\theta)$  нам так же не известна. Решение проблемы:  $W(\theta)$  это некоторое матожидание. По ЗБЧ известно, что выборочное среднее по вероятности сходится к матожиданию. Т.е.

$$\frac{\ln \rho_\theta(X_1) + \dots + \ln \rho_\theta(X_n)}{n} \xrightarrow{P} \mathbb{E}_{\theta_1} \ln \rho_\theta(X_1) = W(\theta)$$

Немного преобразуем левую часть:

$$\frac{\ln \rho_\theta(X_1) + \dots + \ln \rho_\theta(X_n)}{n} = \frac{1}{n} \sum_{i=1}^n \ln \rho_\theta(X_i) = \frac{1}{n} L(X, \theta)$$

Таким образом, вместо того, чтобы искать максимум неизвестной функции, мы будем искать максимум того, что к ней приближается, и найденное значение и будем называть **оценкой максимального правдоподобия**.

**Определение 4.** Оценкой максимального правдоподобия параметра  $\theta$  называется максимум функции  $L(X, \theta)$ .

**Предложение.** (Состоятельность оценки максимального правдоподобия.)

Пусть  $\theta \in (a, b)$ , и на этом отрезке функция  $\theta \rightarrow L(X, \theta)$  имеет единственную точку локального максимума  $\hat{\theta}$ . Тогда  $\hat{\theta} \xrightarrow{P} \theta_0$ .

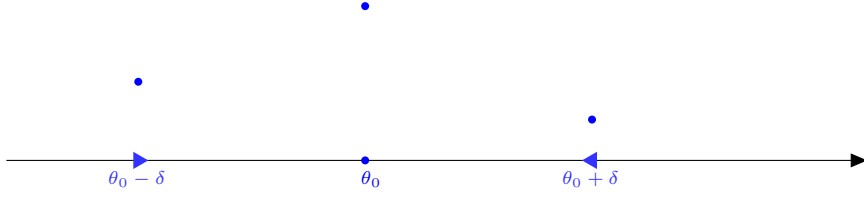
*Доказательство.* Будем пользоваться тем, что  $\frac{1}{n}L(X, \theta) \xrightarrow{P} W(\theta)$ . Хотим доказать, что  $P(|\hat{\theta} - \theta_0| \geq \delta) \rightarrow 0 \forall \delta > 0$  (просто определение сходимости по вероятности). Рассмотрим точки  $\theta_0 - \delta, \theta_0 + \delta$ . Про эти точки известно следующее:

$$\begin{cases} W(\theta_0) > W(\theta_0 - \delta) \xleftarrow{P} \frac{1}{n}L(X, \theta_0 - \delta) \\ W(\theta_0) > W(\theta_0 + \delta) \xleftarrow{P} \frac{1}{n}L(X, \theta_0 + \delta) \end{cases}$$

Можно ожидать, что при достаточно большом  $n$ , с вероятностью, близкой к 1 будут выполнены неравенства

$$\begin{cases} \frac{1}{n}L(X, \theta_0) > \frac{1}{n}L(X, \theta_0 - \delta) \\ \frac{1}{n}L(X, \theta_0) > \frac{1}{n}L(X, \theta_0 + \delta) \end{cases}$$

Посмотрим теперь на функцию  $\theta \rightarrow L(X, \theta)$ :



Ясно, что на интервале  $(\theta_0 - \delta, \theta_0 + \delta)$  существует точка, значение в которой строго больше, чем на концах, а значит, функция имеет на этом отрезке точку локального максимума. Т.е. для точки локального максимума  $\hat{\theta}$  выполнено  $|\hat{\theta} - \theta_0| < \delta$ . Чтобы завершить доказательство, нужно обосновать фразу “при достаточно большом  $n$ , с вероятностью, близкой к 1...”. Другими словами, хотим доказать, что

$$P\left(\frac{1}{n}L(X, \theta_0 - \delta) < \frac{1}{n}L(X, \theta_0)\right) \xrightarrow{n \rightarrow \infty} 1$$

Положим  $W(\theta_0) - W(\theta_0 - \delta) = \varepsilon > 0$ . Из ЗБЧ следует, что

$$\begin{cases} P\left(\left|\frac{1}{n}L(X, \theta_0 - \delta) - W(\theta_0 - \delta)\right| \geq \frac{\varepsilon}{4}\right) \rightarrow 0 \\ P\left(\left|\frac{1}{n}L(X, \theta_0) - W(\theta_0)\right| \geq \frac{\varepsilon}{4}\right) \rightarrow 0 \end{cases}$$

Поймём, почему из этого следует, что  $P\left(\frac{1}{n}L(X, \theta_0 - \delta) \geq \frac{1}{n}L(X, \theta_0)\right) \rightarrow 0 (*)$ . Пусть величины  $\frac{1}{n}L(X, \theta_0 - \delta)$  и  $W(\theta_0 - \delta)$  отличаются менее, чем на  $\frac{\varepsilon}{4}$ . Аналогично для  $\frac{1}{n}L(X, \theta_0)$  и  $W(\theta_0)$ . Тогда верна следующая цепочка равенств:

$$W(\theta_0) \leq \frac{1}{n}L(X, \theta_0) + \frac{\varepsilon}{4} \stackrel{(1)}{\leq} \frac{1}{n}L(X, \theta_0 - \delta) + \frac{\varepsilon}{4} \leq W(\theta_0 - \delta) + \frac{\varepsilon}{2}$$

Переход (1) следует из неравенства (\*). Тогда мы получаем, что  $W(\theta_0) - W(\theta_0 - \delta) \leq \frac{\varepsilon}{2}$ . Но  $W(\theta_0) - W(\theta_0 - \delta) = \varepsilon$ , и получается противоречие. Значит, или  $\frac{1}{n}L(X, \theta_0 - \delta)$  и  $W(\theta_0 - \delta)$  отличаются более, чем на  $\frac{\varepsilon}{4}$ , или же величины  $\frac{1}{n}L(X, \theta_0)$  и  $W(\theta_0)$ . Но тогда мы получаем, что исход из события  $\{\frac{1}{n}L(X, \theta_0 - \delta) \geq \frac{1}{n}L(X, \theta_0)\}$  лежит в объединении

$$\left\{\left|\frac{1}{n}L(X, \theta_0 - \delta) - W(\theta_0 - \delta)\right| \geq \frac{\varepsilon}{4}\right\} \cup \left\{\left|\frac{1}{n}L(X, \theta_0) - W(\theta_0)\right| \geq \frac{\varepsilon}{4}\right\}$$

А вероятность таких событий стремится к нулю. Теперь методом пристального взгляда можно заметить, что мы всё доказали. ■

3. **Информация Фишера и неравенство Рао-Крамера. Критерий равенства в неравенстве Рао-Крамера.**

4. **Доверительные интервалы. Различные методы построения доверительных интервалов (с помощью неравенств на вероятность больших уклонений, с помощью центральной статистики, с помощью асимптотически нормальной оценки). Примеры.**

#### 4.1. Доверительные интервалы

Знать, что оценка  $\hat{\theta}_n(X)$  состоятельна (сходится по вероятности к  $\theta$ ) это, конечно, круто, но особо много информации о ней нам не даёт. Нам хотелось бы знать как быстро она куда-то там сходится – хотим для фиксированного  $\alpha \in (0, 1)$  и фиксированного  $\varepsilon > 0$  знать такой номер  $n$ , что  $P_\theta(|\hat{\theta}_n(X) - \theta| < \varepsilon) > 1 - \alpha$ .

**Определение.**  $(\hat{\theta}_1(X), \hat{\theta}_2(X))$  – доверительный интервал уровня доверия  $1 - \alpha$ , если

$$P_\theta(\theta \in (\hat{\theta}_1(X), \hat{\theta}_2(X))) \geq 1 - \alpha$$

$$P_\theta(\hat{\theta}_1(X) \leq \theta \leq \hat{\theta}_2(X)) \geq 1 - \alpha$$

**Определение.** Последовательность оценок  $\hat{\theta}_1^n(X), \hat{\theta}_2^n(X)$  образует асимптотический доверительный интервал, если  $\liminf_{n \rightarrow \infty} P_\theta(\hat{\theta}_1^n(X) \leq \theta \leq \hat{\theta}_2^n(X)) \geq 1 - \alpha$

**Пример.** Пусть есть выборка из случайных величин с нормальным распределением  $X_j \sim \mathcal{N}(\theta, 1)$ .

Знаем, что  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \xrightarrow{P_\theta} \theta$  (ЗБЧ) – среднее хорошо приближает  $\theta$ .

Посмотрим на разность эмпирического среднего и реальной  $\theta$ :  $\frac{X_1 + \dots + X_n}{n} - \theta = \frac{\overbrace{(X_1 - \theta)}^{\sim \mathcal{N}(0,1)} + \dots + \overbrace{(X_n - \theta)}^{\sim \mathcal{N}(0,1)}}{n} \sim \mathcal{N}(0, \frac{1}{n})$   
 $\Rightarrow \sqrt{n}(\frac{X_1 + \dots + X_n}{n} - \theta) \sim \mathcal{N}(0, 1)$

Теперь по таблице значений функции распределения нормального закона найдём квантили  $z_{\frac{\alpha}{2}}$  и  $z_{1-\frac{\alpha}{2}}$ :  $\Phi(z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}, \Phi(z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ .  
 $\Rightarrow$

$$\begin{aligned} P_\theta(z_{\frac{\alpha}{2}} \leq \sqrt{n}(\frac{X_1 + \dots + X_n}{n} - \theta) \leq z_{1-\frac{\alpha}{2}}) &= \Phi(z_{1-\frac{\alpha}{2}}) - \Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha \\ P(\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} - \bar{X}_n \leq -\theta \leq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} - \bar{X}_n) &= 1 - \alpha \\ P(\bar{X}_n - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \theta \leq \bar{X}_n - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}) &= 1 - \alpha \end{aligned}$$

Заметим, что мы взяли симметричный интервал:  $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$ . В таком случае наш интервал принимает вид:

$(\bar{X}_n - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}})$ . В таком случае длина этого интервала равна  $O(\frac{1}{\sqrt{n}})$

Но зачем мы решили взять симметричный интервал? Вспомним, что мы от него хотим: минимальной длины. А какой интервал на графике нормального распределения будет захватывать нужную площадь и при этом быть самым коротким среди всех? Правильно, симметричный с центром в пике колокола нормального распределения.

4.2. **Различные методы построения доверительных интервалов (с помощью неравенств на вероятность больших уклонений, с помощью центральной статистики, с помощью асимптотически нормальной оценки)**

##### 1. Неравенства Чебышёва или Чернова

$$X_1, \dots, X_n \sim \text{Bern}(\theta), P(X_j = 1) = \theta$$

Чебышёв:

$$P_\theta(|\bar{X}_n - \theta| \geq \varepsilon) \leq \frac{\mathbb{D}X_1}{n\varepsilon^2} = \frac{\theta(1-\theta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} = \alpha \Rightarrow \varepsilon = \frac{1}{\sqrt{2n\alpha}} \Rightarrow P_\theta(\bar{X}_n - \frac{1}{\sqrt{2n\alpha}} < \theta < \bar{X}_n + \frac{1}{\sqrt{2n\alpha}}) \geq 1 - \alpha.$$



$$\begin{aligned}
& \text{Чернов: } P_\theta(|\bar{X}_n - \theta| \geq \varepsilon) \leq 2e^{-\frac{n\varepsilon^2}{4}} = \alpha \\
& -\frac{n\varepsilon^2}{4} = \ln \frac{\alpha}{2} \\
& \varepsilon = 2\sqrt{-\frac{\ln \frac{\alpha}{2}}{n}} \\
& \implies P_\theta(\bar{X}_n - 2\sqrt{-\frac{\ln \frac{\alpha}{2}}{n}} < \theta < \bar{X}_n + 2\sqrt{-\frac{\ln \frac{\alpha}{2}}{n}}) \geq 1 - \alpha
\end{aligned}$$

Заметим, что в обеих оценках мы получили, что длина интервала равна  $O(\frac{1}{\sqrt{n}})$ , но несложно заметить, что константа Чернова значительно лучше, чем у Чебышёва.

## 2. Метод центральной статистики

**Определение.**  $V(X, \theta)$  называется центральной статистикой, если:

- (a) её распределение не зависит от  $\theta$ :  $P_\theta(V(X, \theta) \leq t) = F(t)$
- (b)  $\forall X: \theta \mapsto V(X, \theta)$  — монотонная

Пусть у нас есть такая статистика. Вопрос: как с её помощью строить доверительные интервалы? Предельно просто: подберём числа  $t_1$  и  $t_2$  таким образом, чтобы  $P_\theta(t_1 \leq V(X, \theta) \leq t_2) \geq 1 - \alpha$ . Мы можем так сделать, потому что распределение  $V$  не зависит от  $\theta$ . Теперь поскольку при любом  $X$  наша функция монотонна, то данная оценка равносильна тому, что  $P_\theta(\hat{\theta}_1(X) \leq \theta \leq \hat{\theta}_2(X)) \geq 1 - \alpha$  — чисто из-за монотонности по  $\theta$ .

**Пример.**  $X_j \sim \mathcal{U}(0, \theta) \implies \theta^{-1}X_j \sim \mathcal{U}(0, 1)$ . Это уже центральная статистика, однако она зависит всего от одного элемента выборки. Рассмотрим  $X_{(n)} = \max_{1 \leq j \leq n} X_j$ :  $P_\theta(\theta^{-1}X_{(n)} \leq t) = P_\theta(\max_{1 \leq j \leq n} \theta^{-1}X_j \leq t) = \prod_{j=1}^n P_\theta(\underbrace{\theta^{-1}X_j}_{\sim \mathcal{U}(0,1)} \leq t) = t^n$

Теперь грубо попробуем оценить, куда там наша статистика попадает:

$$P_\theta(\underbrace{t}_{t_1} \leq \theta^{-1}X_{(n)} \leq \underbrace{1}_{t_2}) = 1 - t^n = 1 - \alpha \implies t = \alpha^{\frac{1}{n}}$$

Теперь попробуем вытащить отсюда  $\theta$ :

$$P_\theta(\alpha^{\frac{1}{n}} \leq \theta^{-1}X_{(n)} \leq 1) = 1 - \alpha$$

$$P_\theta(\frac{\alpha^{\frac{1}{n}}}{X_{(n)}} \leq \theta^{-1} \leq \frac{1}{X_{(n)}}) = 1 - \alpha$$

$$P_\theta(\underbrace{X_{(n)}}_{\hat{\theta}_1(X)} \leq \theta \leq \underbrace{\frac{X_{(n)}}{\alpha^{\frac{1}{n}}}}_{\hat{\theta}_2(X)}) = 1 - \alpha$$

Теперь посмотрим на длину полученного доверительного интервала:

$$(\alpha^{-\frac{1}{n}} - 1)X_{(n)}$$

Что мы можем сказать про  $\alpha^{-\frac{1}{n}} - 1$ ? Разложим это дело по Тейлору:

$$\alpha^{-\frac{1}{n}} - 1 \sim e^{-\frac{\ln \alpha}{n}} - 1 \sim \frac{-\ln \alpha}{n} = O(\frac{1}{n}) \rightarrow 0$$

Получается длина доверительного интервала с ростом количества элементов выборки стремится к нулю. Получается мы построили что-то более менее разумное.

Часто в роли центральной статистики можно взять следующую лабуду:  $V(X, \theta) = -\sum_{j=1}^n \ln F_\theta(X_j)$  — это сумма независимых распределений, поэтому достаточно показать что одно не зависит от  $\theta$  — тогда в силу независимости сумма тоже будет не зависеть от  $\theta$ :

$P_\theta(-\ln F_\theta(X_j) \leq t) = P_\theta(F_\theta(X_j) \geq e^{-t}) = P_\theta(X_j \geq F_\theta^{-1}(e^{-t})) = 1 - F_\theta(F_\theta^{-1}(e^{-t})) = 1 - e^{-t}$ , а это экспоненциальное распределение. Сумма экспоненциальных распределений это Гамма распределение  $\implies V(X, \theta) = \Gamma(n, 1)$

## 3. Построение асимптотических доверительных интервалов

Пусть у нас есть  $\hat{\theta}_n(X)$  — асимптотически нормальная оценка  $\theta$  с асимптотической дисперсией  $\sigma^2(\theta)$ . Это значит, что

$$\frac{\sqrt{n}(\hat{\theta}_n(X) - \theta)}{\sigma(\theta)} \xrightarrow{d_\theta} Z \sim \mathcal{N}(0, 1)$$

Теперь мы хотим получить доверительный интервал. Если бы у нас  $\sigma(\theta)$  была константой, то мы могли бы уже привычно взять там квантили нормального распределения, туды сюды и получить интервал:

$$P_\theta(t_1 \leq \frac{\sqrt{n}(\hat{\theta}_n(X) - \theta)}{\sigma(\theta)} \leq t_2) \rightarrow \Phi(t_2) - \Phi(t_1) = 1 - \alpha.$$

Тогда мы могли бы просто взять такие  $\Phi(t_2) = 1 - \frac{\alpha}{2}$  и  $\Phi(t_1) = \frac{\alpha}{2}$  и получить, что

$$P_\theta(\hat{\theta}_n(X) - \frac{t_{1-\frac{\alpha}{2}}\sigma(\theta)}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n(X) + \frac{t_{1-\frac{\alpha}{2}}\sigma(\theta)}{\sqrt{n}}) \rightarrow 1 - \alpha$$

Но тут есть проблема — у нас слева и справа есть  $\sigma(\theta)$  в числителе, что совершенно ломает корректность статистики, мы ведь хотим чтобы штуки слева и справа от  $\theta$  в неравенстве не зависели от  $\theta$ . Как это решать? Очень просто, перейти от  $\sigma(\theta)$  к состоятельной оценке  $\sigma(\theta)$ . Возможны следующие случаи:

(a)  $\sigma$  — непрерывная функция

Тогда  $\sigma(\hat{\theta}_n(X)) \xrightarrow{P_\theta} \sigma(\theta)$  и мы можем везде в наших рассуждениях заменить  $\sigma(\theta)$  на  $\sigma(\hat{\theta}_n(X))$  и сходимость сохранится:

$$P_\theta(\hat{\theta}_n(X) - \frac{t_{1-\frac{\alpha}{2}}\sigma(\hat{\theta}_n(X))}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n(X) + \frac{t_{1-\frac{\alpha}{2}}\sigma(\hat{\theta}_n(X))}{\sqrt{n}}) \rightarrow 1 - \alpha$$

(b) Изначально было ЦПТ

$$\hat{\theta}_n(X) = \overline{X_n}, \sigma^2(\theta) = \mathbb{D}_\theta X$$

В таком случае мы можем использовать выборочную дисперсию в качестве состоятельной оценки дисперсии:

$$s^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \overline{X_n})^2 \text{ — выборочная дисперсия}$$

(c) Можно поправить нашу асимптотическую дисперсию: подобрать такую функцию  $\varphi$ , что

$$\sqrt{n}(\varphi(\hat{\theta}_n(X)) - \varphi(\theta)) \rightarrow \underbrace{\mathcal{N}(0, 1)}_{=\varphi'(\theta) \cdot \mathcal{N}(0, \sigma^2(\theta))} \implies \varphi'(\theta)\sigma^2(\theta) = 1$$

## 5. Построение точных доверительных интервалов для параметров нормального распределения.

Оцениваем параметры случайной величины  $\sim \mathcal{N}(a, \sigma^2)$  по выборке  $X_1, \dots, X_n$ :

1.  $\sigma$  — известно. Оцениваем матожидание  $a$ :

Центрируем и нормируем случайную величину разности оценки и параметра:

$$\sqrt{n} \frac{\overline{X_n} - a}{\sigma} \sim \mathcal{N}(0, 1) \text{ — центральная статистика}$$

Пусть для  $t_{1-\frac{\alpha}{2}}$  верно, что  $\Phi(t_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ , тогда

$$P_\theta \left( -t_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\overline{X_n} - a)}{\sigma} \leq t_{1-\frac{\alpha}{2}} \right) = 1 - \alpha \Leftrightarrow P_\theta \left( \underbrace{\frac{\overline{X_n} - \frac{t_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}}{\hat{\Theta}_1(x)}}_{\hat{\Theta}_1(x)} \leq a \leq \underbrace{\frac{\overline{X_n} + \frac{t_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}}{\hat{\Theta}_2(x)}}_{\hat{\Theta}_2(x)} \right) = 1 - \alpha$$

2.  $\sigma$  не известно.

**Лемма.** Пусть  $X = (X_1, \dots, X_n) \sim \mathcal{N}(a, \text{diag}(\sigma^2))$ ,  $\text{diag}(\sigma^2) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$ ,  $\{X_j\}$  независимы,  $\mathbb{D}X_j = \sigma^2$ ,  $\mathbb{E}X_j = a$ .

Тогда  $\overline{X_n}$  и  $\sum_{j=1}^n (X_j - \overline{X_n})^2$  независимы.

*Доказательство.* Не ограничивая общности, считаем, что  $\mathbb{E}X_j = a = 0$ .

Рассмотрим  $U = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \star & \cdots & \star \\ \star & \cdots & \star \end{pmatrix}$  — ортогональную, и случайную величину  $u = UX \sim \mathcal{N}(0, \text{diag}(\sigma^2))$ ,

$$(U^* C_x U = U^* \text{diag}(\sigma^2) U = \sigma^2 U^* I U = \sigma^2 U^* U = \sigma^2 I = \text{diag}(\sigma^2))$$

$\{u_\sigma\}$  — незав.,  $\mathbb{E}u_j = 0$ ,  $\mathbb{D}u_j = \sigma^2$ .

$$\sum_{j=2}^n u_j^2 = |u|^2 - u_1^2 = |u|^2 - \left( \frac{\sum X_j}{\sqrt{n}} \right)^2 = |X|^2 - nX_{(n)}^2 \quad (\text{т.к. } U \text{ — орт.})$$

$$\sum_{j=1}^n (X_j - \overline{X}_n)^2 = \sum_{j=1}^n X_j^2 - 2n\overline{X}_n \sum_{j=1}^n X_j + n\overline{X}_n^2 = \sum_{j=1}^n X_j^2 - n\overline{X}_n^2$$

$$\sum_{j=2}^n u_j^2 = \sum_{j=1}^n (X_j - \overline{X}_n)^2 \text{ — независимо с } u_1 = \sqrt{n}\overline{X}_n$$

При этом  $\sum_{j=2}^n u_j^2 = \sigma^2 \sum_{j=2}^n (\sigma^{-1} u_j)^2$ , и  $\sigma^{-1} u_j \sim \mathcal{N}(0, 1)$ , то есть

$$\chi_n^2 = \sum_{j=1}^n (X_j - \overline{X}_n)^2 = \sum_{j=1}^n \xi_k^2, \quad \{\xi_k\} \text{ — незав., } \xi_k \sim \mathcal{N}(0, 1),$$

где  $\chi_n^2$  — распределение  $\chi$ -квадрат с  $n$  степенями свободы, распределение величины  $\sum_{j=1}^n \eta_j^2$ ,  $\eta_j$  — независимые нормально распределенные с параметрами 0 и 1 величины.

В итоге  $\overline{X}_n$  и  $s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \overline{X}_n)^2$  независимы и  $\sigma^{-2} s^2 \sim \frac{1}{n-1} \chi_{n-1}^2$ . ■

Так как  $\sigma$  неизвестна, заменим ее на  $\sqrt{s^2}$  и получим статистику:

$$T_{n-1}(X) = \frac{\sqrt{n}(X_n - a)}{\sqrt{s^2}} = \frac{\frac{\sqrt{n}(X_n - a)}{\sigma}}{\frac{\sigma}{\sqrt{\sigma^2 s^2}}} \sim \frac{\xi}{\sqrt{\frac{1}{n-1} \chi_{n-1}^2}},$$

$\xi$  и  $\chi$  независимы.

$T_n = \frac{\xi}{\sqrt{\frac{1}{n} \chi_n^2}}$  — распределение Стьюдента с  $(n-1)$  степенями свободы.

Его плотность:

$$\varrho(x) = C_n \left( 1 + \frac{x^2}{n-1} \right)^{-n/2}$$

Т.к. плотность симметрична, можем выбрать 1 квантиль:

$$F_{T_{n-1}}(t_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}, \quad F_{T_{n-1}}(-t_{1-\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

Попадаем в случай 1 с известной дисперсией:

$$P_\theta \left( -t_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\overline{X}_n - a)}{\sqrt{s^2}} \leq t_{1-\frac{\alpha}{2}} \right) = 1 - \alpha \Leftrightarrow P_\theta \left( \underbrace{\overline{X}_n - \frac{t_{1-\frac{\alpha}{2}} \sqrt{s^2}}{\sqrt{n}}}_{\Theta_1(x)} \leq a \leq \underbrace{\overline{X}_n + \frac{t_{1-\frac{\alpha}{2}} \sqrt{s^2}}{\sqrt{n}}}_{\Theta_2(x)} \right) = 1 - \alpha$$

## 6. Проверка гипотез. Ошибки 1-го и 2-го рода. Уровень значимости и мощность статистического критерия. Пример построения критерия с помощью доверительного интервала. Нижняя оценка суммы вероятностей ошибок 1-го и 2-го рода.

### 6.1. Проверка гипотез.

Пусть есть выборка  $X_1, \dots, X_n$  с распределением  $P_\theta$ .

**Определение.** Предположения о значениях  $\theta$  и называются статистическими гипотезами.

**Пример.**  $H_0: \theta \in \Theta_0$  — статистическая гипотеза

**Определение.** Простая гипотеза (одноточечная гипотеза) — гипотеза вида  $H_0: \theta = \theta_0$ , где  $\Theta = \{\theta_0\}$

**Определение.** Гипотеза  $H_1: \theta \in \Theta_1$  — альтернативная гипотеза

**Пример.**  $H_1: \theta = \bar{\theta}_0$  — альтернативная гипотеза

Для проверки гипотез, строят критерий на основе критического множества как правило  $\subset \mathbb{R}^n$ , то есть действуют по такому принципу:

Выделяют в области значения параметров критического множества  $K$ , так, что  $\forall \theta \in \Theta_0, P_\theta((X_1 \dots X_n) \in K)$  — «маленькая», тогда  $X_1 \dots X_n \in K$  свидетельствует против гипотезы  $H_0$ , то есть если  $X = (X_1 \dots X_n) \in K \Rightarrow H_0$  отклоняется, иначе принимается.

Далее  $X = (X_1 \dots X_n)$

### 6.2. Ошибки 1-го и 2-го рода.

Пусть у нас есть критическое множество  $K$ . При проверке гипотез мы могли совершить две ошибки:

**Определение.** Ошибка первого рода: отклонение верной гипотезы  $H_0$ , то есть это  $P_{\theta \in \Theta_0}(X \in K)$ . В случае простой гипотезы  $P_{\theta_0}(X \in K)$

**Определение.** Ошибка второго рода: принятие ложной гипотезы  $H_0$ , то есть это  $P_{\theta \in \Theta_1}(X \notin K)$ . В случае простой гипотезы  $P_{\theta_1}(X \notin K)$

### 6.3. Уровень значимости и мощность статистического критерия.

**Определение.** Критерий  $K$  имеет уровень значимости  $\alpha$ , если вероятность ошибки первого рода меньше либо равна  $\alpha$ , то есть  $P_{\theta \in \Theta_0}(X \in K) \leq \alpha$ .

**Определение.** Мощность критерия  $K$  это величина, равная  $1 -$  вероятность ошибки второго рода, то есть  $1 - P_{\theta \in \Theta_1}(X \notin K) = P_{\theta \in \Theta_1}(X \in K)$ . В случае простой гипотезы величина  $\beta = P_{\theta_1}(X \in K)$  — мощность

Если имеются два критерия  $K, S$  уровня значимости  $\alpha$ , то  $K$  более мощный, чем  $S$  если

$$\forall \theta \in \Theta_1: P_\theta(X \in K) \geq P_\theta(X \in S)$$

### 6.4. Пример построения критерия с помощью доверительного интервала.

**Пример.** Пусть  $X_1, \dots, X_n \sim \mathcal{N}(\theta, 1)$

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1$

Ранее при данных условиях мы получили следующий доверительный интервал:

$$P_{\theta_0} \left( \bar{X}_n - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \theta_0 \leq \bar{X}_n + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right) = P_{\theta_0} \left( \theta_0 - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \bar{X}_n \leq \theta_0 + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right) = 1 - \alpha$$

Выберем критическое множество  $K: \left\{ X: \overline{X_n} > \theta_0 + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right\} \cup \left\{ X: \overline{X_n} < \theta_0 - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right\}$

Тогда  $P_{\theta_0}(X \in K) = 1 - (1 - \alpha) = \alpha$

Найдем ошибку второго рода:

$$\begin{aligned} P_{\theta_1}(X \notin K) &= P_{\theta_1} \left( \theta_0 - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \overline{X_n} \leq \theta_0 + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right) = \\ &= P_{\theta_1} \left( \sqrt{n}(\theta_0 - \theta_1) - Z_{1-\frac{\alpha}{2}} \leq \underbrace{\sqrt{n}(\overline{X_n} - \theta_1)}_{\sim \mathcal{N}(0,1)} \leq \sqrt{n}(\theta_0 - \theta_1) + Z_{1-\frac{\alpha}{2}} \right) = \\ &= \Phi(\sqrt{n}(\theta_0 - \theta_1) + Z_{1-\frac{\alpha}{2}}) - \Phi(\sqrt{n}(\theta_0 - \theta_1) - Z_{1-\frac{\alpha}{2}}) \end{aligned}$$

Посмотрим, что происходит при  $n \rightarrow \infty$ :

$$1. \theta_0 > \theta_1 \implies P_{\theta_1}(X \notin K) \rightarrow 0$$

$$2. \theta_0 < \theta_1 \implies P_{\theta_1}(X \notin K) \rightarrow 0$$

Таким образом, мы получили состоятельный критерий.

### 6.5. Нижняя оценка суммы вероятностей ошибок 1-го и 2-го рода.

**Теорема.** Пусть у нас есть две гипотезы:

$$1. H_0: \rho = f_0$$

$$2. H_1: \rho = f_1$$

Сумма ошибки первого рода и ошибки второго рода больше либо равна  $1 - \frac{1}{2} \int_{\mathbb{R}^n} |f_0 - f_1| dx$

*Доказательство.* Найдем сумму ошибок первого и второго рода:

$$P_0(X \in K) + P_1(X \notin K) = \int_K f_0 dx + \underbrace{\int_{\mathbb{R}^n \setminus K} f_1 dx}_{1 - \int_K f_1 dx} = 1 + \int_K (f_0 - f_1) dx \geq$$

$$\begin{aligned} &\text{Введем множество } S = \{f_0 \leq f_1\} \\ &\geq 1 + \int_{K \cap S} (f_0 - f_1) dx \geq 1 + \int_S (f_0 - f_1) dx \end{aligned}$$

Рассмотрим отдельно интеграл  $\int_S (f_0 - f_1) dx$ :

$$\int_S (f_0 - f_1) dx = \int_S f_0 dx - \int_S f_1 dx = 1 - \int_{\mathbb{R}^n \setminus S} f_0 dx - 1 + \int_{\mathbb{R}^n \setminus S} f_1 dx = \int_{\mathbb{R}^n \setminus S} (f_1 - f_0) dx$$

В силу того, как мы выбрали множество  $S$ , можно увидеть, что

$$1. \int_S (f_0 - f_1) dx = - \int_S |f_0 - f_1| dx$$

$$2. \int_{\mathbb{R}^n \setminus S} (f_1 - f_0) dx = - \int_{\mathbb{R}^n \setminus S} |f_0 - f_1| dx$$

Тогда мы получаем, что

$$- \int_{\mathbb{R}^n \setminus S} |f_0 - f_1| dx = - \int_S |f_0 - f_1| dx = - \frac{1}{2} \int_{\mathbb{R}^n} |f_0 - f_1| dx \implies P_0(X \in K) + P_1(X \notin K) \geq 1 - \frac{1}{2} \int_{\mathbb{R}^n} |f_0 - f_1| dx$$

■

## 7. Теорема Неймана-Пирсона и пример её применения.

### 7.1. Теорема Неймана-Пирсона.

Пусть гипотеза  $H_0$  утверждает, что плотность выборки – это  $f_0$ , а альтернативная гипотеза  $H_1$  утверждает, что плотность выборки – это  $f_1$ .

Предположим, что  $\forall \alpha \in [0, 1] \exists t := t(\alpha) : P_0(f_1(x) \geq t f_0(x)) = \alpha$ .

**Теорема (Неймана-Пирсона).** В такой постановке наиболее мощный критерий уровня значимости  $\alpha$  имеет вид  $K_{t(\alpha)} := \{f_1(x) \geq t(\alpha)f_0(x)\}$ .

*Доказательство.* Пусть  $S$  – тоже критерий уровня значимости  $\alpha$ :  $P_0(X \in S) \leq \alpha = P_0(X \in K_{t(\alpha)})$ . Хотим сравнить  $P_1(X \in K_{t(\alpha)}) - P_1(X \in S)$ . Хотим, чтобы это было больше либо равно нулю. Это и будет означать, что у нас критерий наиболее мощный.

$$P_1(X \in K_{t(\alpha)}) - P_1(X \in S) = \int_{K_{t(\alpha)}} f_1 dx - \int_S f_1 dx = [\text{можем выкинуть пересечение, так как на пересечении эти интегралы просто сократятся}] = \int_{K_{t(\alpha)} \setminus S} f_1 dx - \int_{S \setminus K_{t(\alpha)}} f_1 dx.$$

Заметим, что на  $S \setminus K_{t(\alpha)}$  выполнено  $f_1 < t(\alpha)f_0$ , так как это взято из дополнения к  $K_{t(\alpha)}$ , где по условию выполняется  $f_1(x) \geq t(\alpha)f_0(x)$ . Поэтому имеем:  $\int_{K_{t(\alpha)} \setminus S} f_1 dx - \int_{S \setminus K_{t(\alpha)}} f_1 dx \geq t(\alpha) \int_{K_{t(\alpha)} \setminus S} f_0 dx - t(\alpha) \int_{S \setminus K_{t(\alpha)}} f_0 dx = [\text{снова добавим пересечение и вынесем } t(\alpha)] = t(\alpha) \cdot \left( \int_{K_{t(\alpha)}} f_0 dx - \int_S f_0 dx \right) = t(\alpha) \cdot (P_0(X \in K_{t(\alpha)}) - P_0(X \in S)) \geq 0$  из построения критерия  $S$  ( $P_0(X \in S) \leq \alpha = P_0(X \in K_{t(\alpha)})$ ).

Получили:  $P_1(X \in K_{t(\alpha)}) - P_1(X \in S) \geq 0$ , что и требовалось доказать. ■

### 7.2. Пример применения теоремы Неймана-Пирсона.

**Пример.** Пусть у нас выборка из нормального закона  $N(\theta, 1)$ . Пусть наша гипотеза  $H_0$  говорит, что  $\theta = \theta_0$ , а альтернативная гипотеза  $H_1$  говорит, что  $\theta = \theta_1 > \theta_0$ .

$$f_1(X) = \frac{1}{\sqrt{2\pi}^n} \cdot \exp\left(-\frac{1}{2} \sum_{j=1}^n (X_j - \theta_1)^2\right)$$

$$f_0(X) = \frac{1}{\sqrt{2\pi}^n} \cdot \exp\left(-\frac{1}{2} \sum_{j=1}^n (X_j - \theta_0)^2\right)$$

Зададим критерий  $K_t$  из теоремы Неймана-Пирсона (ничего в 0 не обращается – сразу можем поделить):

$$K_t = \left\{ \frac{f_1}{f_0} \geq t \right\} = \left\{ \exp\left(\frac{1}{2} \sum_{j=1}^n [(X_j - \theta_0)^2 - (X_j - \theta_1)^2] \right) \geq t \right\} = [\text{логарифмируем, раскрываем скобки, умножаем на два}] = \left\{ \sum_{j=1}^n [2X_j(\theta_1 - \theta_0) + n(\theta_0^2 - \theta_1^2)] \geq 2 \ln t \right\} = \left\{ (\theta_1 - \theta_0) \overline{X}_n \geq \frac{\ln t}{n} - \frac{(\theta_0^2 - \theta_1^2)}{2} \right\} = [\text{по условию } \theta_1 > \theta_0 \Rightarrow \text{поделим}]$$

$$= \left\{ \overline{X}_n \geq \frac{\frac{\ln t}{n} - \frac{(\theta_0^2 - \theta_1^2)}{2}}{\theta_1 - \theta_0} \right\}$$

Таким образом пришли к тому, что  $K_t = \left\{ \frac{f_1}{f_0} \geq t \right\}$  равносильно множеству  $\tilde{K}_s = \{\overline{X}_n \geq s\}$ . Равносильно в том смысле, что для каждого  $t$  мы можем подобрать  $s(t)$ , что множество  $K_t$  совпадает с  $\tilde{K}_{s(t)}$ . Теперь будем искать критические множества именно в таком виде (для удобства).

Должно выполняться:  $P_0(X \in K_t) = \alpha \Leftrightarrow P_0(X \in \tilde{K}_{s(t)}) = \alpha$ . А что это за вероятности? Это вероятность  $P_{\theta_0}(\overline{X}_n \geq s) = \alpha$

То есть,  $P_{\theta_0}(\sqrt{n}(\overline{X}_n - \theta_0) \geq \sqrt{n}(s - \theta_0)) = \alpha$ , где  $\sqrt{n}(\overline{X}_n - \theta_0) \sim N(0, 1)$ , поэтому тут просто написано, что  $1 - \Phi(\sqrt{n}(s - \theta_0)) = \alpha$ .

Значит, выбираем квантиль нормального закона уровня  $1 - \alpha$ :  $Z_{1-\alpha} = \sqrt{n}(s - \theta_0) \Rightarrow s = \theta_0 + \frac{Z_{1-\alpha}}{\sqrt{n}}$ . Выразили  $s$ .

Таким образом, наше критическое множество  $\left\{ \overline{X}_n \geq \theta_0 + \frac{Z_{1-\alpha}}{\sqrt{n}} \right\}$ . Это критерий уровня значимости  $\alpha$ .

Теперь посчитаем мощность (это же самый мощный критерий):

$$P_{\theta_1} \left( \overline{X}_n \geq \theta_0 + \frac{Z_{1-\alpha}}{\sqrt{n}} \right) = P_{\theta_1} \left( \sqrt{n}(\overline{X}_n - \theta_1) \geq \sqrt{n}(\theta_0 - \theta_1) + Z_{1-\alpha} \right) = 1 - \Phi(\sqrt{n}(\theta_0 - \theta_1) + Z_{1-\alpha}).$$

Заметим, что если объём выборки  $n$  устремить к бесконечности, то точка, в которой мы берём  $\Phi$  стремится к минус бесконечности (так как  $(\theta_0 - \theta_1) < 0$  по условию), поэтому мощность стремится к 1.

По теореме Неймана-Пирсона выписанная мощность максимальна.

## 8. Эмпирическая функция распределения. Теорема Гливенко-Кантелли.