

Математическая статистика.

Андрей Тищенко @AndrewTGk

2024/2025

Лекция 10 января

Преамбула

Статистика. Мнения о появлении этого слова:

1. Статистиками в Германии назывались люди, собирающие данные о населении и передающие их государству.
2. В определённый день в Венеции народ выстаивался для выплаты налогов (строго фиксированных, в зависимости от рода действий). Государство собирало данные обо всём населении. Это происходило до появления статистиков в Германии, поэтому мы будем считать, что статистика пошла из Венеции.

Задача статистики — по результатам наблюдений построить вероятностную модель наблюдаемой случайной величины.

Основные определения

Определение

Однородной выборкой объёма n называется случайный вектор $X = (X_1, \dots, X_n)$, компоненты которого являются независимыми и одинаково распределёнными. Элементы вектора X называются элементами выборки

Определение

Если элементы выборки имеют распределение $F_\xi(x)$, то говорят, что выборка соответствует распределению $F_\xi(x)$ или порождена случайной величиной ξ с распределением $F_\xi(x)$.

Определение

Детерминированный вектор $x = (x_1, \dots, x_n)$, компоненты которого x_i являются реализациями соответствующих случайных величин X_i , ($i = \overline{1, n}$), называется реализацией выборки.
То есть реализацией однородной выборки X является вектор x .

Определение

Выборочным пространством S называется множество всех возможных реализаций выборки $x = (x_1, \dots, x_n)$. Например, для вектора $X = (x_1, \dots, x_{10})$ каждый элемент x_i которой порождён случайной величиной $\xi \sim U(0, 1)$, выборочным пространством является \mathbb{R}^{10}

Определение

Пусть реализация выборки упорядочена по возрастанию:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Где $x_{(i)}$ — i по возрастанию элемент.

Обозначим $X_{(k)}$ случайную величину, реализация которой при каждой реализации x выборки X принимает значение $x_{(k)}$. Тогда последовательность $X_{(1)}, \dots, X_{(n)}$ называется вариационным рядом выборки.

Определение

Случайная величина $X_{(k)}$ называется k -ой порядковой статистикой выборки.

Определение

Случайные величины $X_{(1)}, X_{(n)}$ называются эстремальными порядковыми статистиками.

Определение

Порядковая статистика $X_{([n \cdot p])}$ называется выборочной квантилью уровня p , где $p \in [0, 1]$

Определение

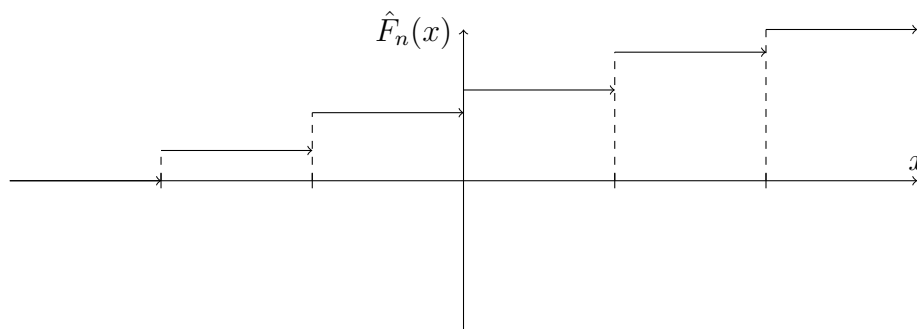
Эмпирическая функция распределения. Пусть выборка X_1, \dots, X_n имеет распределение $F_\xi(x)$. Тогда функция

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n I(X_k \leq x)$$

I — индикаторная функция. $I = \begin{cases} 1, & \text{если аргумент верен} \\ 0, & \text{иначе} \end{cases}$

Называется эмпирической функцией распределения выборки X

Пусть x_1, \dots, x_n — реализация выборки X_1, \dots, X_n



Свойства $\hat{F}_n(x)$

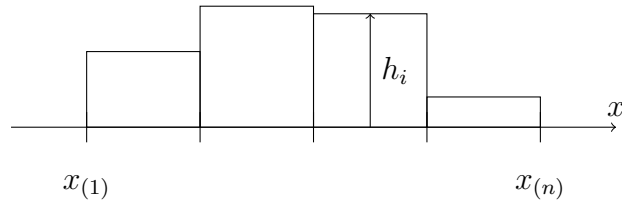
$$1. \forall x \in \mathbb{R} \quad E\hat{F}_n(x) = E\left(\frac{1}{n} \sum_{k=1}^n I(X_k \leq x)\right) = \frac{1}{n} \sum_{k=1}^n EI(X_k \leq x) = P(X_1 \leq x) = F_\xi(x)$$

2. По усиленному закону больших чисел (УЗБЧ)

$$\forall x \in \mathbb{R} \quad \hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n I(X_k \leq x) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} EI(X_k \leq x) = F_\xi(x)$$

Гистограмма

Разбить \mathbb{R} на $(m + 2)$ непересекающихся интервала. Рассматриваются $x_{(1)}, \dots, x_{(n)}$



Размах выборки $r = x_{(n)} - x_{(1)}$

$\Delta = \frac{r}{m}$ — ширина интервала.

$h_k = \frac{\nu_k}{\Delta}$, $k = \overline{1, k}$, где ν_k — количество попаданий на интервал.