Лекция 1

Ilya Yaroshevskiy

6 сентября

Содержание

1 Введение

Определение. Генеральная совокупность — множество всех исходов определенного всех эксперимента

Определение. Выборочная совокупность — множество исходов наблюдаемых экспериментов

Примечание. Выборка *репрезентативная* если ее распределение совпадает с распределением генеральной совокупности

Определение. 1 Пусть проведено n наблюдаемых независимых экспериментов в которых наблюдаемые величины приняли значения: $X_1, X_2, \dots, X_n$. Набор этих данных называется **выборкой объема** n

Определение. 2 **Выборкой объема** n называется набор из n независимых олинакого распределенных случайных величин

2 Выборочные характеристики

Пусть имеется выборка в смысле 1. Ее можно понимать как следующую дискретную случайную величину:

X_1	X_2	X_3	$\dots$	X_n	$\sum$
p_i	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$	1

$$X = \sum_{i=1}^{n} \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i$$

— **матожидание**?

$$S^2 = \sum_{i=1}^{n} (X_i - X)^2 \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} (X_i - X)^2$$

— **дисперсия**

$$F_n^*(z) = \frac{1}{n} \sum_{i=1}^{n} I(X_i < z)$$

— **функция распределения**, где $I(X < z) = \begin{cases} 1 & , X < z \\ 0 & , X_i \geq z \end{cases}$ — индикатор

Теорема 2.1. $\forall z \in \mathbb{R}$

$$F_n^*(z) \xrightarrow[n \to \infty]{p} F(z)$$

Доказательство.

$$EI(X_1 < z) = 1 \cdot p(X < z) + 0 \cdot p(X_1 \geq z) = p(X_1 < z) = F(z)$$

— функция распределения X_1
По ЗБЧ Хнчина:

$$F_n^*(z) = \frac{\sum I(X_i < z)}{n} \xrightarrow{p} EI(X_1 < z) = F(z)$$

□

Примечание. На самом деле имеется даже равномерная сходимость по вероятности: *теорема Гливенко-Кантеля*:

$$\sup_{z \in \mathbb{R}} |F_n^*(z) - F(z)| \xrightarrow{p} [n \to \infty] 0$$

3 Первоначальная обработка

Определение. Если упорядочить выборку по возрастанию(ранжирование) то получим **вариационный ряд**:

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

Если учесть повторяющиеся экземпляры, получим **частотный вариационный ряд**

$X_{(1)}$	$X_{(2)}$	$\dots$	$X_{(k)}$	$\sum$
n_1	n_2	$\dots$	n_k	n
p_1	$\frac{n_1}{n}$	$\dots$	$\frac{n_k}{n}$	1

- $h = X_{\max - X_{\min}}$ — **размах выборки**

Примечание. Допустим разбили интервал $(X_{\min, X_{\max})}$ на k интервалов, чаще всего одинаковой длинны. Тогда $h = \frac{h}{k}$. Тогда вариационный ряд можно заменить интервальным вариационным рядом.

Пример.

i	l_1	l_2	$\dots$	l_k	$\sum$
m_i	m_2	m_2	$\dots$	m_k	n
$\frac{m_i}{n}$	$\frac{m_1}{m}$	$\frac{m_2}{m}$	$\dots$	$\frac{m_k}{m}$	1

На координатной плоскости построим прямоугольники:

- l_i — основание прямоугольника соотвествующего интервала
- $\frac{m_i}{nl_i}$ — высота прямоугольника

Получаем ступенчатую фигуру площади 1, которая называется **гистограмма**.

Теорема 3.1. При

- $n \to \infty$
- $k(n) \to \infty$
- $\frac{k(n)}{n} \to 0$

Гистограмма по вероятности будет стремиться к плотности распределения:

$$\frac{m_i}{n} \xrightarrow{p} p(X_i \in l_i) = \int_{l_i} f(x)dx$$

Примечание. Чаще всего число интервалов берется по формулу Стержеса.

$$k \approx 1 + \log_k n$$

$$k \approx \sqrt[3]{n}$$

Примечание. Иногда выборка изображаеся в виде полигона

- l_i^m — середина i-го интервала