Математическая Статистика

Лекции

Автор конспектов: Чубий Савва Андреевич

Преподаватель: Горяинова Елена Рудольфовна

2024–2025

2025-01-10

2025-01-10

Введение

- Теория вероятностей: по известной модели ищем параметры
- Математическая статистика: по наблюдаемой величине строим модель

Определения

Опр. Однородной выборкой объема N называется случайный вектор $X = (X_1, ..., X_n)$, компоненты которого — независимые и одинокого распределенные.

Примечание: в ближайшее время будем обсуждать только **однородные** выборки.

Опр. Компоненты однородной выборки называются **элементами выборки**.

Опр. Если все элементы выборки $X_1, ..., X_n$ выборки имеют распределение $F_\xi(x)$, то говорят, что **выборка соответствует распределению $F_\xi(x)$ или выборка порождена случайной величиной ξ**.

Примечание: обычно $F_\xi(x)$ мы не знаем. Часто известно только семейство распределения: например, «гауссово» или «непрерывное».

Опр. Детерминированный вектор $x = (x_1, ..., x_n)$, где X_i есть реализация СВ $X_i, i = \overline{1...n}$ называется **реализацией выборки X**.

Опр. Выборочным пространством S называется множество всех возможных реализаций выборки.

Опр. Пара $(S, \mathcal{F})$, где $\mathcal{F}$ – семейство распределений, порождающих X, называется **моделью**.

Опр. Упорядочим реализацию $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$.

Пусть СВ X_k есть такой элемент выборки, реализация которого, который при любой реализации $X_1, ..., X_n$, принимает значение $X_{(k)}$.

Тогда пос-ть $X_{(1)}, ..., X_{(n)}$ называется **вариационным рядом** выборки, а $X_{(k)}$ — k-ой порядковой статистикой.

Опр. Порядковые статистики $X_{(1)}$ и $X_{(n)}$ называются **экстремальными**.

Когда мы зафиксируем конкретную реализацию, $X_{(i)}$ станет числом, до этого $X_{(i)}$ — СВ.

Эмпирическая функция распределения

Опр. Пусть $X_1, ..., X_n$ соответствует распределению $F_\xi(x)$, тогда функция:

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

называется **эмпирической функцией распределения**.

<u>Свойства</u>

- $E\hat{F}_n(x) = P(x_k \leq x) = F_\xi(x)$
- По УЗБЧ:

$$\frac{1}{n} \sum_{k=1}^{n} I(x_k \leq x) \overset{\text{п. н.}}{\underset{n\to\infty}{\to}} F_\xi(x)$$

Гистограмма

$x_1, ..., x_n$ — реализация

Для построения гистограммы:
1. Разбить R^1 на $m + 2$ не пересекающихся интервала
2. Обычно рассматривают **размах** выборки $r = x_{(n)} - x_{(1)}$
3. Первый и последний интервалы ($\left(-\infty, x_{(1)}\right)$ и $\left(x_{(n)}, +\infty\right)$) пустые, остальные — длинны равной $\Delta = \frac{r}{m}$
4. Для каждого интервала вычисляем частоту попаданию в него: ν_k
5. На каждом интервале строим прямоугольник высотой $h_k = \frac{\nu_k}{\Delta}$

При увеличении m гистограмма стремится к плотности.

Гистограмма ростов

$$x_{(1)} = 163$$
$$x_{(39)} = 203$$
$$m = 5$$
$$r = 203 - 163 = 40$$
$$\Delta = 8$$

Интервал	Частота	Высота
$[163, 171)$	$\frac{8}{39}$	...
$[171, 179)$	$\frac{11}{39}$	...
$[179, 187)$	$\frac{18}{39}$	...
$[187, 195)$	$\frac{1}{39}$	...
$[195, 203)$	$\frac{1}{39}$	...

100 наблюдений = 7 интервалов