Математическая статистика

Домашняя работа № 2

Основные понятия математической статистики

Попов Юрий, СКБ-172

ОГЛАВЛЕНИЕ

Предисловие

Все графики, которые в дальнейшем будут вставлены в эту работу, были сконструированы с помощью библиотеки matplotlib в Jupyter Notebook, который будет приложен вместе с работой (Mathematical statistics $DZ2$.ipynb)

Большая часть определений, которые представлены в этой работы взять с лекций нашего курса.

Также некоторые определения взяты из источника Г.И. Ивченко, Ю.И. Медведев "Ведение в математическую статистику"

Задание 2.1 *Моделирование выбранных случайных величин*

Реализация выборки

Определение 1. *Реализация выборки* - это набор из n наблюдений $\hat{x} = (x_1, x_2, \ldots, x_n)$

Задание 2.2 *Построение эмпирической функции распределения*

Эмпирическая функция распределения

Определение 2. Для произвольного числа $x \in R$ рассмотрим случайную величину

$$\mu_n(x) = \sum_{i=1}^{n} Ind(X_i \leq x)$$

равную числу элементов выборки меньших или равных x. Тогда функция $\hat{F}(x) = \frac{\mu_n(x)}{n}$ называется *эмпирической функцией распределения(э.ф.р)*

Эмпирическая функция распределения принимает значения $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$

$$P(\hat{F}(x) = \frac{k}{n}) = C_k^m F^k(x)(1 - F(x))^{n-k}$$

Задание 2.3 *Построение вариационного ряда выборки*

Вариационный ряд выборки

Определение 3. Пусть есть

$$\vec{X} = (X_1, \ldots, X_n),$$

где $X_i, i = \overline{1, n}$—независимые одинаково распределенные случайные величины из распределения ξ. И $\vec{x} = (x_1, \cdots, x_n)$ является реализацией имеющейся выборки $\vec{X}$. Отсортируем вектор $\vec{x}$ по возрастанию:

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

Тогда $x_{(1)} = \min(x_1, x_2, \ldots x_n)$, а $x_{(n)} = \max(x_1, x_2, \ldots, x_n)$. Через $X_{(i)}$ обозначают случайную величину, которая для каждой реализации выборки принимает значение $X_{(i)}$. Вектор $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ называют вариационным рядом выборки.

Квантиль

Определение 4. *Квантилью уровня* $\alpha \in (0, 1)$ функции распределения $F(x)$ называется величина $\zeta_\alpha = \sup\{x : F(x) \leq p\} = F^{-1}(p)$.

Выборочный квантиль

Определение 5. *Выборочными квантилями* называют квантили выборочного распределения.

Для каждой выборки построим вариационный ряд. Вот, примеры вариационных рядов для выборок объема n = 5 и для n = 10:

Вариационные ряды для 5 реализаций выборки, объема 10:

X:	0	1	2	5	7	14	16	19	39
f:	1	1	1	2	1	1	1	1	1

X:	0	1	2	3	6	8	23
f:	1	4	1	1	1	1	1

X:	0	1	2	3	6	9	12	13	15	33
f:	1	1	1	1	1	1	1	1	1	1

X:	0	1	3	5	16	28
f:	4	2	1	1	1	1

X:	0	1	3	4	5	6	8	12	17
f:	1	1	1	2	1	1	1	1	1

Вариационные ряды для 5 реализаций выборки, объема 5:

X:	0	1	5	15
f:	1	2	1	1

X:	0	2	6	9	25
f:	1	1	1	1	1

X:	1	4	5	7	8
f:	1	1	1	1	1

X:	0	8	13	17	21
f:	1	1	1	1	1

X:	0	1	3	7	18
f:	1	1	1	1	1

Рис. 1: n = 5

Перейдем к нахождению выборочной квантили.

Поясним простыми словами на примере уровня квантиля 0.1, что такое выборочная квантиль:

Выборочная квантиль уровня 0,1 - это точка, левее которой (включая её саму) лежит не менее 10% всей выборки, и правее которой (снова включая её саму) - не менее $100\% - 10\% = 90\%$ выборки.

Будем искать выборочную квантиль графическим способом: проведем прямую $y =$(уровень квантиля) до пересечения с графиком. И определим x, при котором прямая пересекает график. Если точка пересечения одна, то именно это значения x и будет выборочной квантиль. А если пересечение проходит по отрезку, то квантилей будет много. Например, если отрезок от 3 до 4 является прямой y, то квантилью будет любое число от 3 до 4. ля определённости в практической статистике в таких случаях выбирают по какому-то правилу одно из чисел "отрезка квантилей". Например, середину отрезка - в данном случае 3,5.

Задание 2.4 *Построение гистограммы и полигон частот*

Гистограмма

Определение 4.Для непрерывной случайной величины ξ, обладающей непрерывной плотностью $f(x)$, также можно построить по соответствующей выборке $X = (X_1, \dots, X_n)$ статистический аналог $\hat{f}_n(x)$ для плотности $f(x)$, который называется *гистограммой*

Полигон частот

Определение 5. Наряду с гистограммой, в качестве приближения для неизвестной теоретической плотности $f(x)$ можно использовать кусочно-линейный график, называемый *полигоном частот*, и который строится так: если построена гистограмма $\hat{f}_n(x)$, то ординаты, соответствующие серединам интервалов группировки, последовательно соединяют отрезками прямых.

Литература

[1]

[2] ссылка1

[3] ссылка2

[4] // ссылка3

[5] // ссылка4