Математическая статистика

Математическая статистика

В теории вероятностей изучались математические модели случайных явлений. Предполагается, что известны законы распределения случайных величин. В реальности эти законы неизвестны, есть только выборка наблюдений.

Математическая статистика

В теории вероятностей изучались математические модели случайных явлений. Предполагается, что известны законы распределения случайных величин. В реальности эти законы неизвестны, есть только выборка наблюдений.

Как восстановить эти законы?

Математическая статистика

В теории вероятностей изучались математические модели случайных явлений. Предполагается, что известны законы распределения случайных величин. В реальности эти законы неизвестны, есть только выборка наблюдений.

Как восстановить эти законы?

Это - задача математической статистики.
Информация берётся из наблюдений над случайными величинами.

Математическая статистика

В теории вероятностей изучались математические модели случайных явлений. Предполагается, что известны законы распределения случайных величин. В реальности эти законы неизвестны, есть только выборка наблюдений.

Как восстановить эти законы?

Это - задача математической статистики.
Информация берётся из наблюдений над случайными величинами.

Цель - разработка методов сбора, описания и анализа статистических данных с целью выявления в них закономерностей.

Математическая статистика

В теории вероятностей изучались математические модели случайных явлений. Предполагается, что известны законы распределения случайных величин. В реальности эти законы неизвестны, есть только выборка наблюдений.

Как восстановить эти законы?

Это - задача математической статистики.
Информация берётся из наблюдений над случайными величинами.

Цель - разработка методов сбора, описания и анализа статистических данных с целью выявления в них закономерностей.

Связь между ТВ и МС: теоретические модели используются для обоснования правильности результатов анализа статистических данных.

Предположим, что существует статистическая совокупность, состоящая из большого числа единиц.

Предположим, что существует статистическая совокупность, состоящая из большого числа единиц.

Нас интересуют некоторые свойства совокупности.

Идея выборочного метода: выбрать из этой совокупности сравнительно небольшое число объектов, и по свойствам этих объектов судить о свойствах всей совокупности.

Предположим, что существует статистическая совокупность, состоящая из большого числа единиц.

Нас интересуют некоторые свойства совокупности.

Идея выборочного метода: выбрать из этой совокупности сравнительно небольшое число объектов, и по свойствам этих объектов судить о свойствах всей совокупности.

Основное требование: выборка должна быть репрезентативной, т.е. точно выражать свойства, которые имеет генеральная совокупность.

Элементы выборки должны быть независимыми и одинаково распределенными.

Предположим, что существует статистическая совокупность, состоящая из большого числа единиц.

Нас интересуют некоторые свойства совокупности.

Идея выборочного метода: выбрать из этой совокупности сравнительно небольшое число объектов, и по свойствам этих объектов судить о свойствах всей совокупности.

Основное требование: выборка должна быть репрезентативной, т.е. точно выражать свойства, которые имеет генеральная совокупность.

Элементы выборки должны быть независимыми и одинаково распределенными.

Альтернатива – сплошное исследование

Основные задачи математической статистики.

Основные задачи математической статистики.

1. ***Построение закона распределения по опытным данным***.

Основные задачи математической статистики.

1. ***Построение закона распределения по опытным данным***.
2. ***Статистическое оценивание***.

Основные задачи математической статистики.

1. ***Построение закона распределения по опытным данным***.
2. ***Статистическое оценивание***.
3. ***Проверка статистических гипотез***.

Основные задачи математической статистики.

1. ***Построение закона распределения по опытным данным***.
2. ***Статистическое оценивание***.
3. ***Проверка статистических гипотез***.
4. ***Определение статистической взаимосвязи между случайными величинами***.

Основные задачи математической статистики.

1. ***Построение закона распределения по опытным данным***.
2. ***Статистическое оценивание***.
3. ***Проверка статистических гипотез***.
4. ***Определение статистической взаимосвязи между случайными величинами***.
5. ***Построение прогнозных моделей по экспериментальным данным***

Основные понятия математической статистики

Основные понятия математической статистики

Генеральная совокупность объектов - те объекты, явления, события и т.п., которые входят в круг интересов исследователя

Основные понятия математической статистики

Генеральная совокупность объектов - те объекты, явления, события и т.п., которые входят в круг интересов исследователя

Выборка объектов $o_1,...,o_n$ - часть статистической совокупности; формируется в результате случайного отбора некоторых представителей совокупности.

n - *объем выборки*.

Основные понятия математической статистики

Генеральная совокупность объектов - те объекты, явления, события и т.п., которые входят в круг интересов исследователя

Выборка объектов $o_1,...,o_n$ - часть статистической совокупности; формируется в результате случайного отбора некоторых представителей совокупности.

n - *объем выборки*.

Переменная – характеризует какое-либо свойство объекта (синонимы: *показатель, признак*).

Два типа переменных: дискретные и непрерывные.

Вариационный ряд и порядковые статистики

Вариационный ряд и порядковые статистики

Рассмотрим случай одной переменной X. Пусть имеется выборка наблюдений

$$X_1, X_2, \ldots, X_n.$$

Вариационный ряд и порядковые статистики

Рассмотрим случай одной переменной X. Пусть имеется выборка наблюдений

$$X_1, X_2, \ldots, X_n.$$

Расположим элементы выборки в порядке возрастания (неубывания):

$$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}.$$

Вариационный ряд и порядковые статистики

Рассмотрим случай одной переменной X. Пусть имеется выборка наблюдений

$$X_1, X_2, \ldots, X_n.$$

Расположим элементы выборки в порядке возрастания (неубывания):

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}.$$

Упорядоченная выборка называется вариационным рядом, $X_{(i)}$ называется i-й порядковой статистикой, различные элементы ряда – вариантами.

Число вариант обозначим через m, где $m \leq n$.

Вариационный ряд и порядковые статистики

Рассмотрим случай одной переменной X. Пусть имеется выборка наблюдений

$$X_1, X_2, \ldots, X_n.$$

Расположим элементы выборки в порядке возрастания (неубывания):

$$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}.$$

Упорядоченная выборка называется вариационным рядом, $X_{(i)}$ называется i-й порядковой статистикой, различные элементы ряда – вариантами.

Число вариант обозначим через m, где $m \le n$.

Частота варианты n_i - число, показывающее, сколько раз встречается в выборке i-я варианта a_i.

Относительная частота варианты:

$$\mathrm{w}_i = \frac{n_i}{n} = \frac{n_i}{\sum_{i=1}^{m} n_i}.$$

Относительная частота варианты:

$$\mathrm{w}_i = \frac{n_i}{n} = \frac{n_i}{\sum_{i=1}^{m} n_i}.$$

Пример.

Имеется выборка

$$1,3,5,2,1,4,1,5,5,4,2,3,3,3,1,$$

n = 15.

Относительная частота варианты:

$$\mathrm{w}_i = \frac{n_i}{n} = \frac{n_i}{\sum\limits_{i=1}^{m} n_i}.$$

Пример.

Имеется выборка

1,3,5,2,1,4,1,5,5,4,2,3,3,3,1,

n = 15.

Вариационный ряд:

1,1,1,1,2,2,3,3,3,3,4,4,5,5,5

m=5

Относительная частота варианты:

$$w_i = \frac{n_i}{n} = \frac{n_i}{\sum_{i=1}^{m} n_i}.$$

Пример.

Имеется выборка

1,3,5,2,1,4,1,5,5,4,2,3,3,3,1,

n = 15.

Вариационный ряд:

1,1,1,1,2,2,3,3,3,3,4,4,5,5,5

m=5

Таблица вариант и их частот:

a_i	1	2	3	4	5
n_i	4	2	4	2	3
w_i	4/15	2/15	4/15	2/15	3/15

Распределение порядковой статистики

Имеет ли случайная величина $X_{(i)}$ то же распределение, что и X_i ?

Распределение порядковой статистики

Имеет ли случайная величина $X_{(i)}$ то же распределение, что и X_i ?

Введем вспомогательную случайную функцию:

$\mu_n\left(x\right)$- количество наблюдений $X_i < x$.

Распределение порядковой статистики

Имеет ли случайная величина $X_{(i)}$ то же распределение, что и X_i ?

Введем вспомогательную случайную функцию: $\mu_n(x)$- количество наблюдений $X_i < x$.

Найдем $P\left\{\mu_n(x) = k\right\}$.

Распределение порядковой статистики

Имеет ли случайная величина $X_{(i)}$ то же распределение, что и X_i ?

Введем вспомогательную случайную функцию: $\mu_n\left(x\right)$- количество наблюдений $X_i < x$.

Найдем $P\left\{\mu_n\left(x\right) = k\right\}$.

Событие $\mu_n\left(x\right) = k$ означает, что в интервал $(-\infty, x)$ попало ровно k наблюдений, а в интервал $[x, +\infty)$ ровно $\left(n-k\right)$ наблюдений.

Распределение порядковой статистики

Имеет ли случайная величина $X_{(i)}$ то же распределение, что и X_i ?

Введем вспомогательную случайную функцию:

$\mu_n\left(x\right)$- количество наблюдений $X_i < x$.

Найдем $P\left\{\mu_n\left(x\right) = k\right\}$.

Событие $\mu_n\left(x\right) = k$ означает, что в интервал $(-\infty, x)$ попало ровно k наблюдений, а в интервал $[x, +\infty)$ ровно $\left(n - k\right)$ наблюдений.

Число способов, которыми можно выбрать k элементов из n равно C_n^k, в результате получаем:

Распределение порядковой статистики

Имеет ли случайная величина $X_{(i)}$ то же распределение, что и X_i ?

Введем вспомогательную случайную функцию:

$\mu_n\left(x\right)$- количество наблюдений $X_i < x$.

Найдем $P\left\{\mu_n\left(x\right) = k\right\}$.

Событие $\mu_n\left(x\right) = k$ означает, что в интервал $(-\infty, x)$ попало ровно k наблюдений, а в интервал $[x, +\infty)$ ровно $\left(n-k\right)$ наблюдений.

Число способов, которыми можно выбрать k элементов из n равно C_n^k, в результате получаем:

$$P\left\{\mu_n\left(x\right) = k\right\} = C_n^k F^k\left(x\right)\left(1 - F\left(x\right)\right)^{n-k}$$

Найдем теперь функцию распределения:

Найдем теперь функцию распределения:

$$F_{X_{(i)}}(x) = P\left\{X_{(i)} < x\right\} = P\left\{\mu_n\left(x\right) \geq i\right\} =$$

Найдем теперь функцию распределения:

$$F_{X_{(i)}}(x) = P\left\{X_{(i)} < x\right\} = P\left\{\mu_n\left(x\right) \geq i\right\} =$$

$$= P\left\{\mu_n\left(x\right) = i \vee \mu_n\left(x\right) = i + 1 \vee \ldots \vee \mu_n\left(x\right) = n\right\} =$$

Найдем теперь функцию распределения:

$$F_{X_{(i)}}(x) = P\left\{X_{(i)} < x\right\} = P\left\{\mu_n\left(x\right) \geq i\right\} =$$

$$= P\left\{\mu_n\left(x\right) = i \vee \mu_n\left(x\right) = i+1 \vee ... \vee \mu_n\left(x\right) = n\right\} =$$

$$= \sum_{k=i}^{n} P\left\{\mu_n\left(x\right) = k\right\} = \sum_{k=i}^{n} C_n^k F\left(x\right)^k \left(1 - F\left(x\right)\right)^{n-k}$$

Найдем теперь функцию распределения:

$$F_{X_{(i)}}(x) = P\left\{X_{(i)} < x\right\} = P\left\{\mu_n\left(x\right) \geq i\right\} =$$

$$= P\left\{\mu_n\left(x\right) = i \vee \mu_n\left(x\right) = i+1 \vee \ldots \vee \mu_n\left(x\right) = n\right\} =$$

$$= \sum_{k=i}^{n} P\left\{\mu_n\left(x\right) = k\right\} = \sum_{k=i}^{n} C_n^k F\left(x\right)^k \left(1 - F\left(x\right)\right)^{n-k}$$

Таким образом, распределение $X_{(i)}$ отличается от распределения X_i

Найдем теперь функцию распределения:

$$F_{X_{(i)}}(x) = P\left\{X_{(i)} < x\right\} = P\left\{\mu_n\left(x\right) \geq i\right\} =$$

$$= P\left\{\mu_n\left(x\right) = i \vee \mu_n\left(x\right) = i + 1 \vee \ldots \vee \mu_n\left(x\right) = n\right\} =$$

$$= \sum_{k=i}^{n} P\left\{\mu_n\left(x\right) = k\right\} = \sum_{k=i}^{n} C_n^k F\left(x\right)^k \left(1 - F\left(x\right)\right)^{n-k}$$

Таким образом, распределение $X_{(i)}$ отличается от распределения X_i

Накопленная (кумулятивная) частота $\mu_n(x)$ - число элементов выборки, меньших, чем x.

Зависимость $\mu_n(x)$ называют кумулятивной кривой.

Относительная кумулятивная частота $w_n(x) = \dfrac{\mu_n(x)}{n}$.

Относительная кумулятивная частота $w_n(x) = \dfrac{\mu_n(x)}{n}$.

Относительная кумулятивная частота $w_n(x) = \dfrac{\mu_n(x)}{n}$.

Зависимость $w_n(x)$ называют эмпирической функцией распределения:

$$\boxed{F_n(x) = \frac{\mu_n(x)}{n}}.$$

Относительная кумулятивная частота $w_n(x) = \dfrac{\mu_n(x)}{n}$.

Зависимость $w_n(x)$ называют эмпирической функцией распределения:

$$\boxed{F_n(x) = \frac{\mu_n(x)}{n}}.$$

$\forall x \in R,\quad F_n(x)$ – дискретная случайная величина,

принимающая значения

Относительная кумулятивная частота $w_n(x) = \dfrac{\mu_n(x)}{n}$.

Зависимость $w_n(x)$ называют эмпирической функцией распределения:

$$\boxed{F_n(x) = \frac{\mu_n(x)}{n}}.$$

$\forall x \in R$, $F_n(x)$ – дискретная случайная величина,

принимающая значения

$$0 = \frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, \frac{n}{n} = 1$$

Относительная кумулятивная частота $w_n(x) = \dfrac{\mu_n(x)}{n}$.

Зависимость $w_n(x)$ называют эмпирической функцией распределения:

$$\boxed{F_n(x) = \frac{\mu_n(x)}{n}}.$$

$\forall x \in R,\quad F_n(x)$ – дискретная случайная величина,

принимающая значения

$$0 = \frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, \frac{n}{n} = 1$$

при этом

$$P\left\{F_n(x) = \frac{k}{n}\right\} = P\left\{\mu_n(x) = k\right\} = C_n^k F^k(x)\left(1 - F(x)\right)^{n-k}$$

Используя варианты можно записать

$$F_n(x) = \begin{cases} 0, x \le a_1 \\ \dfrac{1}{n}\sum\limits_{i=1}^{k} n_i, a_k < x \le a_{k+1}, k = 1, \dots, m-1 \\ 1, x > a_m \end{cases}$$

Пример. Построим эмпирическую функцию распределения для данных вариант:

a_i	2	6	10
n_i	12	18	30
w_i	0,2	0,3	0,5

<u>Пример</u>. Построим эмпирическую функцию распределения для данных вариант:

a_i	2	6	10
n_i	12	18	30
w_i	0,2	0,3	0,5

$$n=\sum_{i=1}^{3} n_i = 60;\ \sum_{i=1}^{3} w_i = 1,$$

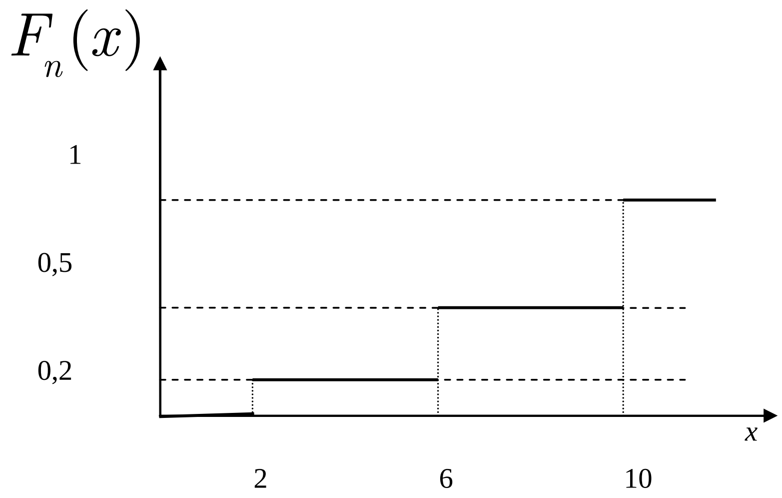
$$F_n(x)=\begin{cases} 0; & x \le 2 \\ 0,2; & 2 < x \le 6 \\ 0,5; & 6 < x \le 10 \\ 1; & x > 10 \end{cases};$$

Пример. Построим эмпирическую функцию распределения для данных вариант:

a_i	2	6	10
n_i	12	18	30
w_i	0,2	0,3	0,5

$$n = \sum_{i=1}^{3} n_i = 60; \sum_{i=1}^{3} w_i = 1,$$

$$F_n(x) = \begin{cases} 0; & x \le 2 \\ 0,2; & 2 < x \le 6 \\ 0,5; & 6 < x \le 10 \\ 1; & x > 10 \end{cases};$$



Оценивание функции распределения по выборке

Оценивание функции распределения по выборке

Очевидно, что свойства эмпирической функции распределения аналогичны свойствам теоретической функции распределения:

Оценивание функции распределения по выборке

Очевидно, что свойства эмпирической функции распределения аналогичны свойствам теоретической функции распределения:

$$1)\forall x \in (-\infty, \infty), \quad 0 \le F_n(x) \le 1;$$

Оценивание функции распределения по выборке

Очевидно, что свойства эмпирической функции распределения аналогичны свойствам теоретической функции распределения:

$$1)\ \forall x \in (-\infty, \infty),\quad 0 \le F_n(x) \le 1;$$

$$2)\ \forall (x_1 \le x_2) \Rightarrow F_n(x_1) \le F_n(x_2);$$

Оценивание функции распределения по выборке

Очевидно, что свойства эмпирической функции распределения аналогичны свойствам теоретической функции распределения:

1) $\forall x \in (-\infty, \infty), \quad 0 \le F_n(x) \le 1;$

2) $\forall (x_1 \le x_2) \Rightarrow F_n(x_1) \le F_n(x_2);$

3) $F_n(x)$ ступенчатая функция, непрерывная слева;

Оценивание функции распределения по выборке

Очевидно, что свойства эмпирической функции распределения аналогичны свойствам теоретической функции распределения:

1) $\forall x \in (-\infty, \infty),\ \ 0 \le F_n(x) \le 1;$

2) $\forall (x_1 \le x_2) \Rightarrow F_n(x_1) \le F_n(x_2);$

3) $F_n(x)$ ступенчатая функция, непрерывная слева;

4) если $x < x_{\min}$, то $F_n(x) = 0;$

Оценивание функции распределения по выборке

Очевидно, что свойства эмпирической функции распределения аналогичны свойствам теоретической функции распределения:

1) $\forall x \in (-\infty, \infty), \quad 0 \le F_n(x) \le 1;$

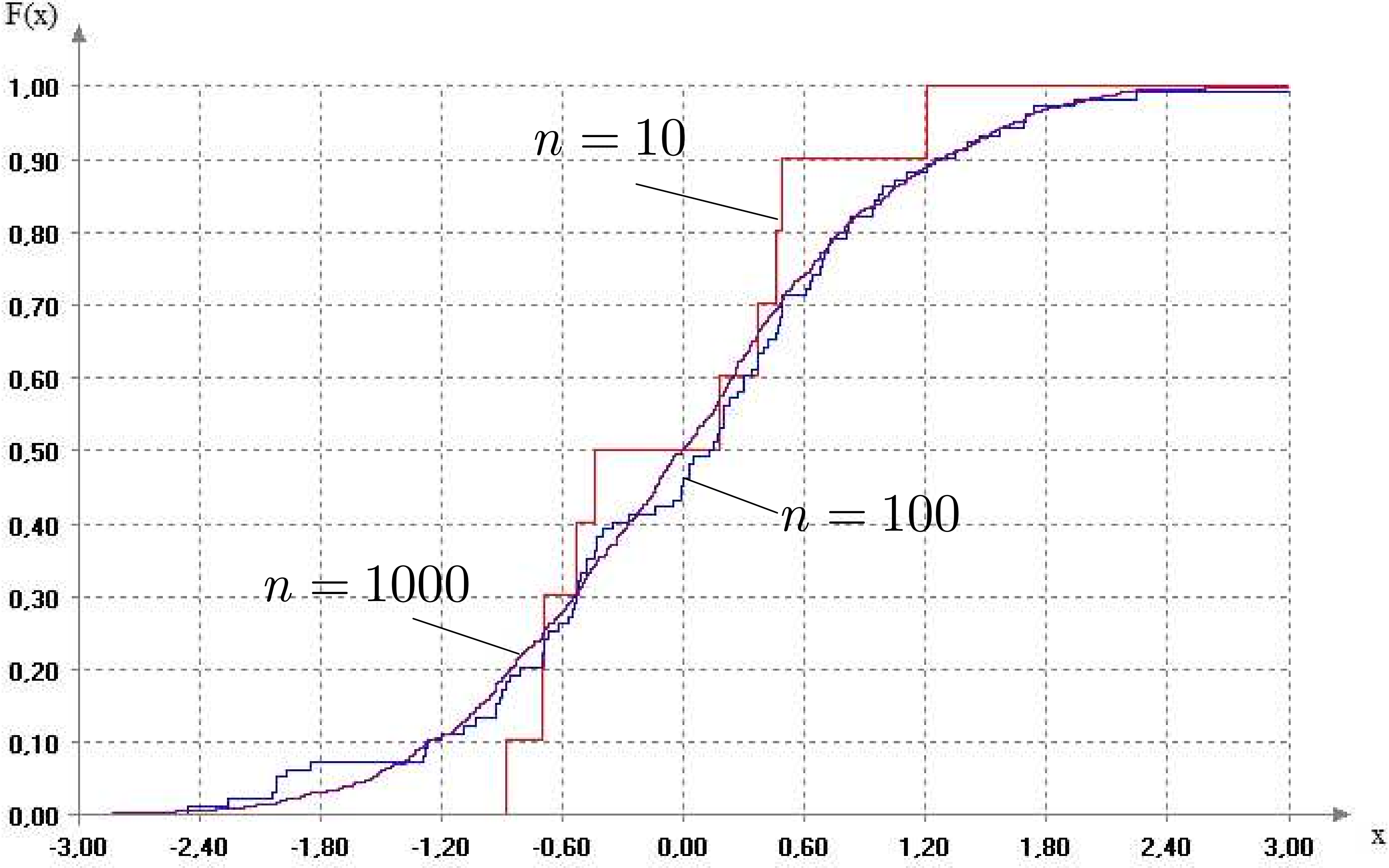
2) $\forall (x_1 \le x_2) \Rightarrow F_n(x_1) \le F_n(x_2);$

3) $F_n(x)$ ступенчатая функция, непрерывная слева;

4) если $x < x_{\min}$, то $F_n(x) = 0;$

5) если $x > x_{\max}$, то $F_n(x) = 1.$

Как меняется эмпирическая функция распределения при увеличении объема выборки?



Сходимость эмпирической функции распределения к теоретической

Теорема

Пусть $F_n(x)$ – эмпирическая функция распределения по выборке случайной величины X, имеющей функцию распределения $F(x)$.

Сходимость эмпирической функции распределения к теоретической

Теорема

Пусть $F_n(x)$ – эмпирическая функция распределения по выборке случайной величины X, имеющей функцию распределения $F(x)$.

Тогда $F_n(x) \xrightarrow{p} F(x)$:

Сходимость эмпирической функции распределения к теоретической

Теорема

Пусть $F_n\left(x\right)$ – эмпирическая функция распределения по выборке случайной величины X, имеющей функцию распределения $F(x)$.

Тогда $F_n(x) \xrightarrow{p} F(x)$:

$$\forall \left|x\right| < \infty, \forall \varepsilon > 0 \;\; \lim_{n\to\infty} P\left\{\left|F_n\left(x\right) - F\left(x\right)\right| < \varepsilon\right\} = 1$$

Доказательство.

Пусть

$$Z_i = \begin{cases} 1, & X_i < x \\ 0, & \text{иначе} \end{cases}, \quad i = 1, \ldots, n.$$

Доказательство.

Пусть

$$Z_i = \begin{cases} 1, & X_i < x \\ 0, & \textit{иначе} \end{cases}, \; i = 1, \ldots, n.$$

Доказательство.

Пусть

$$Z_i = \begin{cases} 1, & X_i < x \\ 0, & \textit{иначе} \end{cases}, \; i = 1, \ldots, n.$$

Все Z_i независимы, одинаково распределены по закону Бернулли:

Доказательство.

Пусть

$$Z_i = \begin{cases} 1, & X_i < x \\ 0, & иначе \end{cases}, \; i = 1,\dots,n.$$

Все Z_i независимы, одинаково распределены по закону Бернулли:

$$P(Z_i = 1) = P(X_i < x) = F(x),$$

Доказательство.

Пусть

$$Z_i = \begin{cases} 1, & X_i < x \\ 0, & \textit{иначе} \end{cases}, \; i = 1, \ldots, n.$$

Все Z_i независимы, одинаково распределены по закону Бернулли:

$$P(Z_i = 1) = P(X_i < x) = F(x),$$
$$P(Z_i = 0) = 1 - F(x).$$

Доказательство.

Пусть

$$Z_i = \begin{cases} 1, & X_i < x \\ 0, & \textit{иначе} \end{cases}, \ i = 1, \ldots, n.$$

Все Z_i независимы, одинаково распределены по закону Бернулли:

$$P(Z_i = 1) = P(X_i < x) = F(x),$$

$$P(Z_i = 0) = 1 - F(x).$$

Сумма $Z_1 + \ldots + Z_n = \mu_n(x)$.

Доказательство.

Пусть

$$Z_i = \begin{cases} 1, & X_i < x \\ 0, & \textit{иначе} \end{cases}, \ i = 1, \ldots, n.$$

Все Z_i независимы, одинаково распределены по закону Бернулли:

$$P(Z_i = 1) = P(X_i < x) = F(x),$$
$$P(Z_i = 0) = 1 - F(x).$$

Сумма $Z_1 + \ldots + Z_n = \mu_n(x)$.

По закону больших чисел

$$F_n(x) = \frac{\mu_n(x)}{n} = \frac{Z_1 + \ldots + Z_n}{n} \xrightarrow{p} EZ_i = F(x).$$

Теорема (Гливенко-Кантелли).

$$P\left\{\lim_{n\to\infty}\sup_{x} |F_n(x) - F(x)| = 0\right\} = 1,$$

Теорема (Гливенко-Кантелли).

$$P\left\{\lim_{n\to\infty}\sup_{x} | F_n(x) - F(x) | = 0\right\} = 1,$$

т.е. эмпирическая функция распределения равномерно сходится к истинной функции распределения.

Теорема (Гливенко-Кантелли).

$$P\left\{\lim_{n\to\infty}\sup_{x} | F_n (x) - F(x) | = 0\right\} = 1,$$

т.е. эмпирическая функция распределения равномерно сходится к истинной функции распределения.

Замечание 1. Последовательность случайных величин Y_n сходится к Y *почти наверное*, если

$$P\left\{\lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right\} = 1$$

Теорема (Гливенко-Кантелли).

$$P\left\{\lim_{n\to\infty}\sup_{x} | F_n(x) - F(x) | = 0\right\} = 1,$$

т.е. эмпирическая функция распределения равномерно сходится к истинной функции распределения.

Замечание 1. Последовательность случайных величин Y_n сходится к Y *почти наверное*, если

$$P\left\{\lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right\} = 1$$

Замечание 2. Из сходимости почти наверное следует сходимость по вероятности.

$$Y_n \xrightarrow{P} Y$$

Теорема (Гливенко-Кантелли).

$$P\left\{\lim_{n\to\infty}\sup_x |F_n(x)-F(x)|=0\right\}=1,$$

т.е. эмпирическая функция распределения равномерно сходится к истинной функции распределения.

Замечание 1. Последовательность случайных величин Y_n сходится к Y *почти наверное*, если

$$P\left\{\lim_{n\to\infty} Y_n(\omega)=Y(\omega)\right\}=1$$

Замечание 2. Из сходимости почти наверное следует сходимость по вероятности.

$$Y_n \xrightarrow{P} Y$$

Замечание 3. Из сходимости по вероятности следует сходимость по распределению.

$$F_{Y_n}(x)\to F_Y(x)$$

Теорема (Колмогоров).

Если функция $F(x)$ непрерывна, то

$$P\left\{\sup_x \left|F_n(x) - F(x)\right| < \frac{z}{\sqrt{n}}\right\} \to K(z) \text{ при } n \to \infty,$$

Теорема (Колмогоров).

Если функция $F(x)$ непрерывна, то

$$P\left\{\sup_x \left|F_n(x) - F(x)\right| < \frac{z}{\sqrt{n}}\right\} \to K(z) \text{ при } n \to \infty,$$

где $$K(z) = \begin{cases} 0 \ \text{при} \quad z \le 0, \\ \sum_{\kappa=-\infty}^{+\infty} (-1)^{\kappa} \exp(-2\kappa^2 z^2), z > 0. \end{cases}$$

Теорема (Колмогоров).

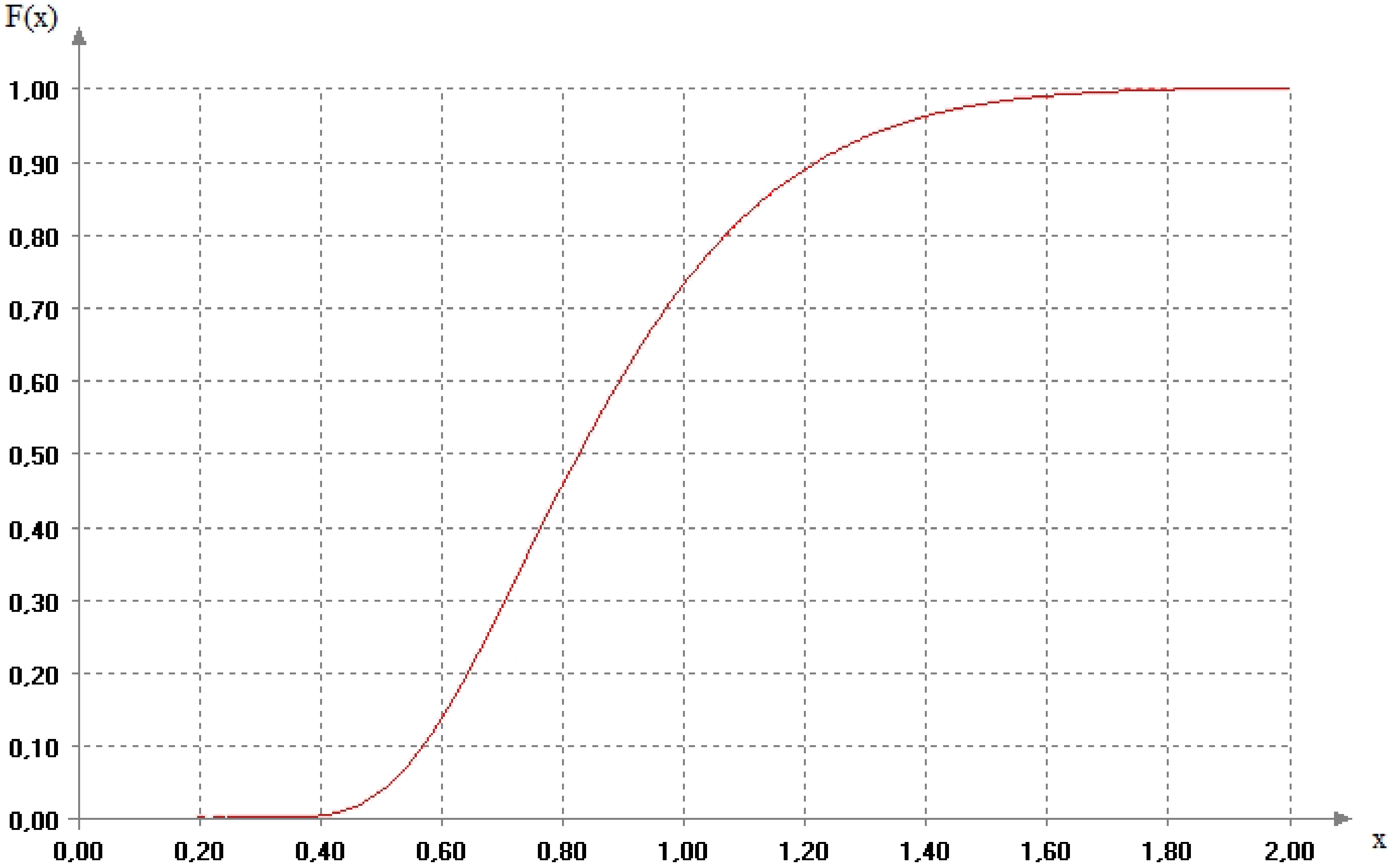
Если функция $F(x)$ непрерывна, то

$$P\left\{\sup_x \left|F_n(x) - F(x)\right| < \frac{z}{\sqrt{n}}\right\} \to K(z) \text{ при } n \to \infty,$$

где $$K(z) = \begin{cases} 0 \text{ при } \quad z \le 0, \\ \sum\limits_{\kappa=-\infty}^{+\infty} (-1)^{\kappa} \exp(-2\kappa^2 z^2), z > 0. \end{cases}$$

Функция $K(z)$ называется функцией распределения Колмогорова. Её значения табулированы.

Функция распределения Колмогорова



Оценивание плотности (ряда) распределения

Пусть переменная X - дискретная, тогда вариационный ряд называют дискретным.

Оценивание плотности (ряда) распределения

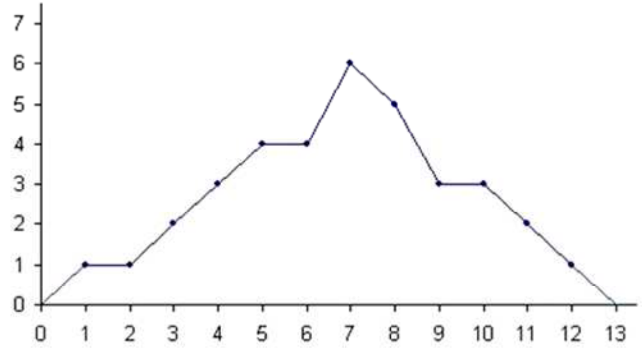
Пусть переменная X - дискретная, тогда вариационный ряд называют дискретным.

Для дискретного ряда можно построить полигон – график зависимости n_i от i (w_i от i) или от вариант a_i.

Оценивание плотности (ряда) распределения

Пусть переменная X - дискретная, тогда вариационный ряд называют дискретным.

Для дискретного ряда можно построить полигон – график зависимости n_i от i (w_i от i) или от вариант a_i.

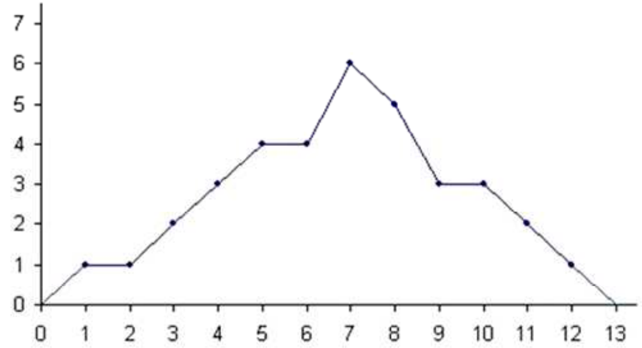


или

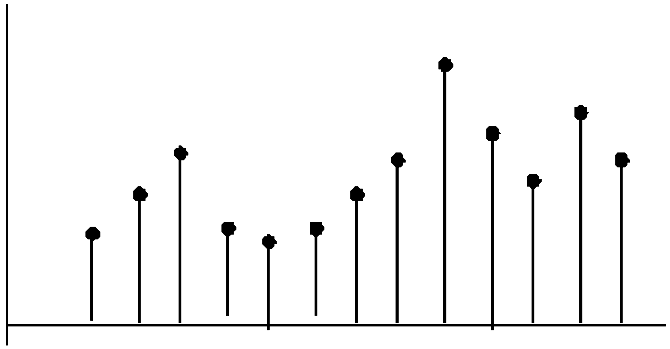
Оценивание плотности (ряда) распределения

Пусть переменная X - дискретная, тогда вариационный ряд называют дискретным.

Для дискретного ряда можно построить полигон – график зависимости n_i от i (w_i от i) или от вариант a_i.



или



Пусть переменная X непрерывная, тогда ряд также называют непрерывным.

Для непрерывного ряда введем интервал группирования $h_i,\ i=1,...,m$, где m - число интервалов.

Пусть переменная X непрерывная, тогда ряд также называют непрерывным**.**

Для непрерывного ряда введем интервал группирования $h_i,\ i=1,...,m$, где m - число интервалов.

Пусть $n_i,\ w_i$ - частоты попадания вариант ряда в i-й интервал.

Пусть переменная X непрерывная, тогда ряд также называют непрерывным.

Для непрерывного ряда введем интервал группирования $h_i,\ i=1,...,m$, где m - число интервалов.

Пусть $n_i,\ w_i$ - частоты попадания вариант ряда в i-й интервал.

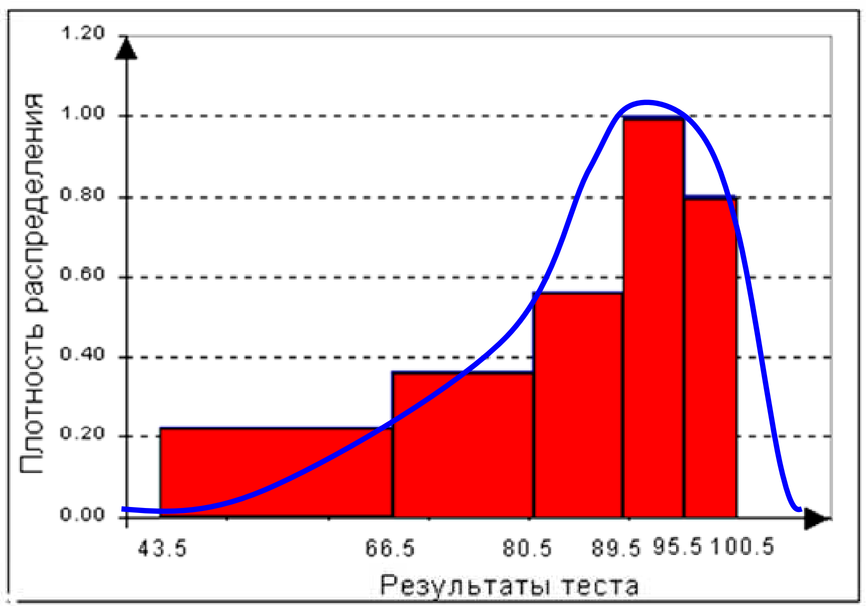
Гистограмма – график, состоящий из прямоугольников шириной, равной величине интервала h_i и высотой, равной $\dfrac{n_i}{h_i}$

Пусть переменная X непрерывная, тогда ряд также называют непрерывным.

Для непрерывного ряда введем интервал группирования h_i, $i = 1, ..., m$, где m - число интервалов.

Пусть n_i, w_i - частоты попадания вариант ряда в i-й интервал.

Пример. Г.Р. результатов теста по иностранному языку.



Гистограмма – график, состоящий из прямоугольников шириной, равной величине интервала h_i и высотой, равной $\frac{n_i}{h_i}$

(или $\frac{w_i}{h_i}$ - аналог функции плотности распределения).

Пример. Дан сгруппированный ряд. Требуется построить гистограмму относительных частот.

границы интервалов	10-20	20-30	30-40	40-50	50-60
частоты	1	2	7	18	12

Пример. Дан сгруппированный ряд. Требуется построить гистограмму относительных частот.

границы интервалов	10-20	20-30	30-40	40-50	50-60
частоты	1	2	7	18	12

Относительные частоты $w_i = \dfrac{n_i}{n}$, высоты прямоугольников $\dfrac{w_i}{h}$, где $n = \sum_{i=1}^{m} n_i = 1+2+7+18+12 = 40$, $h = 10$. Таблица:

Пример. Дан сгруппированный ряд. Требуется построить гистограмму относительных частот.

границы интервалов	10-20	20-30	30-40	40-50	50-60
частоты	1	2	7	18	12

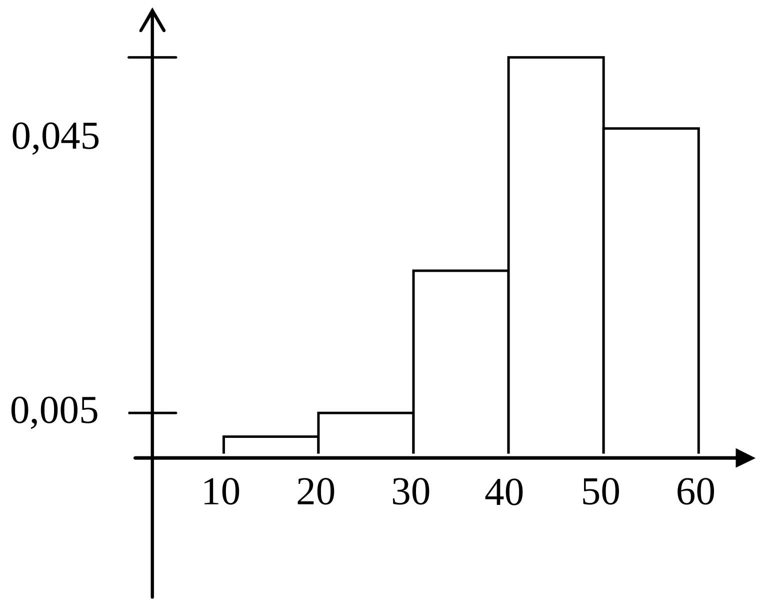
Относительные частоты $w_i = \dfrac{n_i}{n}$, высоты

прямоугольников $\dfrac{w_i}{h}$, где $n = \sum_{i=1}^{m} n_i = 1 + 2 + 7 + 18 + 12 = 40$,

$h = 10$. Таблица:

границы	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
w_i	1/40 = 0,025	2/40 = 0,05	7/40 = 0,175	18/40 = 0,45	12/40 = 0,3
$\dfrac{w_i}{h}$	0,025/10 = 0,0025	0,05/10 = 0,005	0,175/10 = 0,0175	0,45/10 = 0,045	0,3/10 = 0,03

Гистограмма:



Ядерное оценивание плотности распределения

Ядерное оценивание плотности распределения

Пусть $g(t)$ – неотрицательная функция ("ядро"), удовлетворяющая условиям:

Ядерное оценивание плотности распределения

Пусть $g(t)$ – неотрицательная функция ("ядро"), удовлетворяющая условиям:

1. $g(t) = g(-t)$,

Ядерное оценивание плотности распределения

Пусть $g(t)$ – неотрицательная функция ("ядро"), удовлетворяющая условиям:

1. $g(t) = g(-t)$,
2. $\int\limits_{-\infty}^{+\infty} g(t)dt = 1$,

Ядерное оценивание плотности распределения

Пусть $g(t)$ – неотрицательная функция ("ядро"), удовлетворяющая условиям:

1. $g(t) = g(-t)$,
2. $\int\limits_{-\infty}^{+\infty} g(t)dt = 1$,
3. $\int\limits_{-\infty}^{+\infty} t^2 g(t)dt = 1$

Ядерное оценивание плотности распределения

Пусть $g(t)$ – неотрицательная функция ("ядро"), удовлетворяющая условиям:

1. $g(t) = g(-t)$,
2. $\int\limits_{-\infty}^{+\infty} g(t)dt = 1$,
3. $\int\limits_{-\infty}^{+\infty} t^2 g(t)dt = 1$
4. $\int\limits_{-\infty}^{+\infty} t^m g(t)dt < \infty; 0 \leq m < \infty$,

тогда функцию плотности можно оценить:

$$\hat{f}_n(x) = \frac{1}{n\lambda_n} \sum_{i=1}^{n} g\left(\frac{x - X_i}{\lambda_n}\right)$$

тогда функцию плотности можно оценить:

$$\hat{f}_n(x) = \frac{1}{n\lambda_n} \sum_{i=1}^{n} g\left(\frac{x - X_i}{\lambda_n} \right)$$

где λ_n - параметр размытости:

$$\lim_{n \to \infty} \lambda_n = 0 \text{ и } \lim_{n \to \infty} n\lambda_n = \infty,$$

тогда функцию плотности можно оценить:

$$\hat{f}_n(x) = \frac{1}{n\lambda_n} \sum_{i=1}^{n} g\left(\frac{x - X_i}{\lambda_n}\right)$$

где λ_n - параметр размытости:

$$\lim_{n\to\infty} \lambda_n = 0 \text{ и } \lim_{n\to\infty} n\lambda_n = \infty,$$

а также функцию распределения:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} G\left(\frac{x - X_i}{\lambda_n}\right)$$

тогда функцию плотности можно оценить:

$$\hat{f}_n(x) = \frac{1}{n\lambda_n} \sum_{i=1}^{n} g\left(\frac{x - X_i}{\lambda_n}\right)$$

где λ_n - параметр размытости:

$$\lim_{n \to \infty} \lambda_n = 0 \text{ и } \lim_{n \to \infty} n\lambda_n = \infty,$$

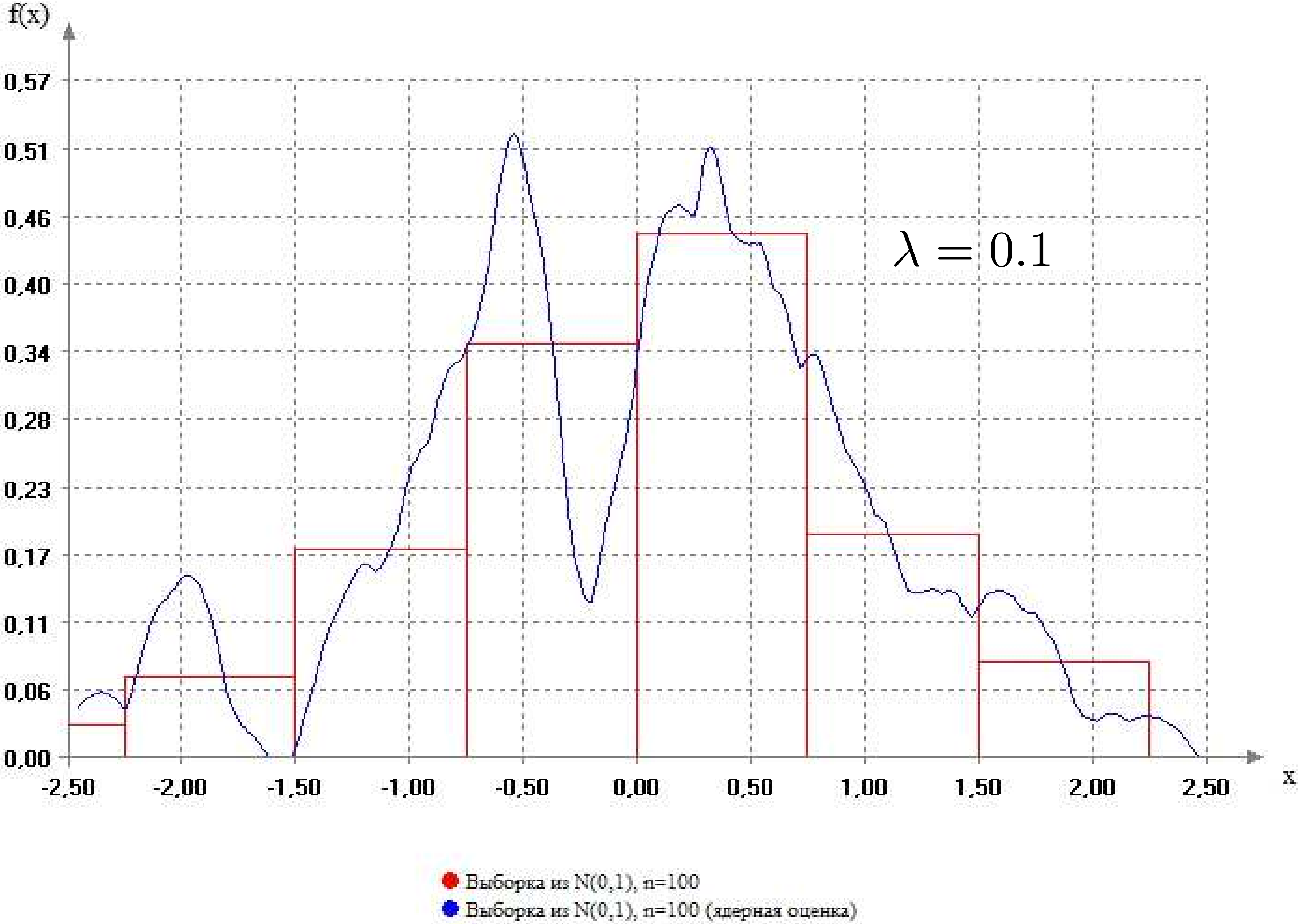
а также функцию распределения:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} G\left(\frac{x - X_i}{\lambda_n}\right)$$

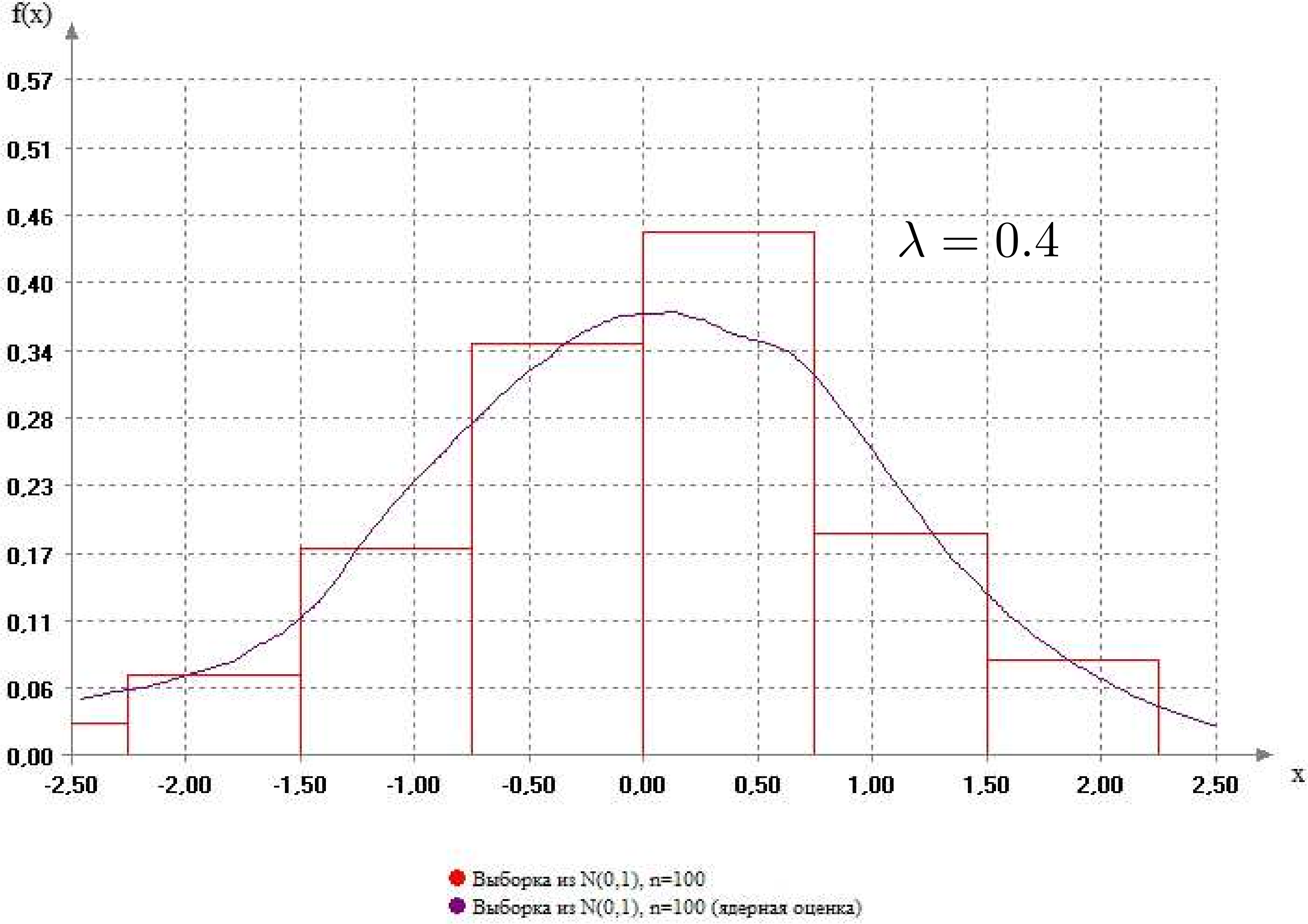
где

$$G(x) = \int_{\infty}^{x} g(x)dt.$$

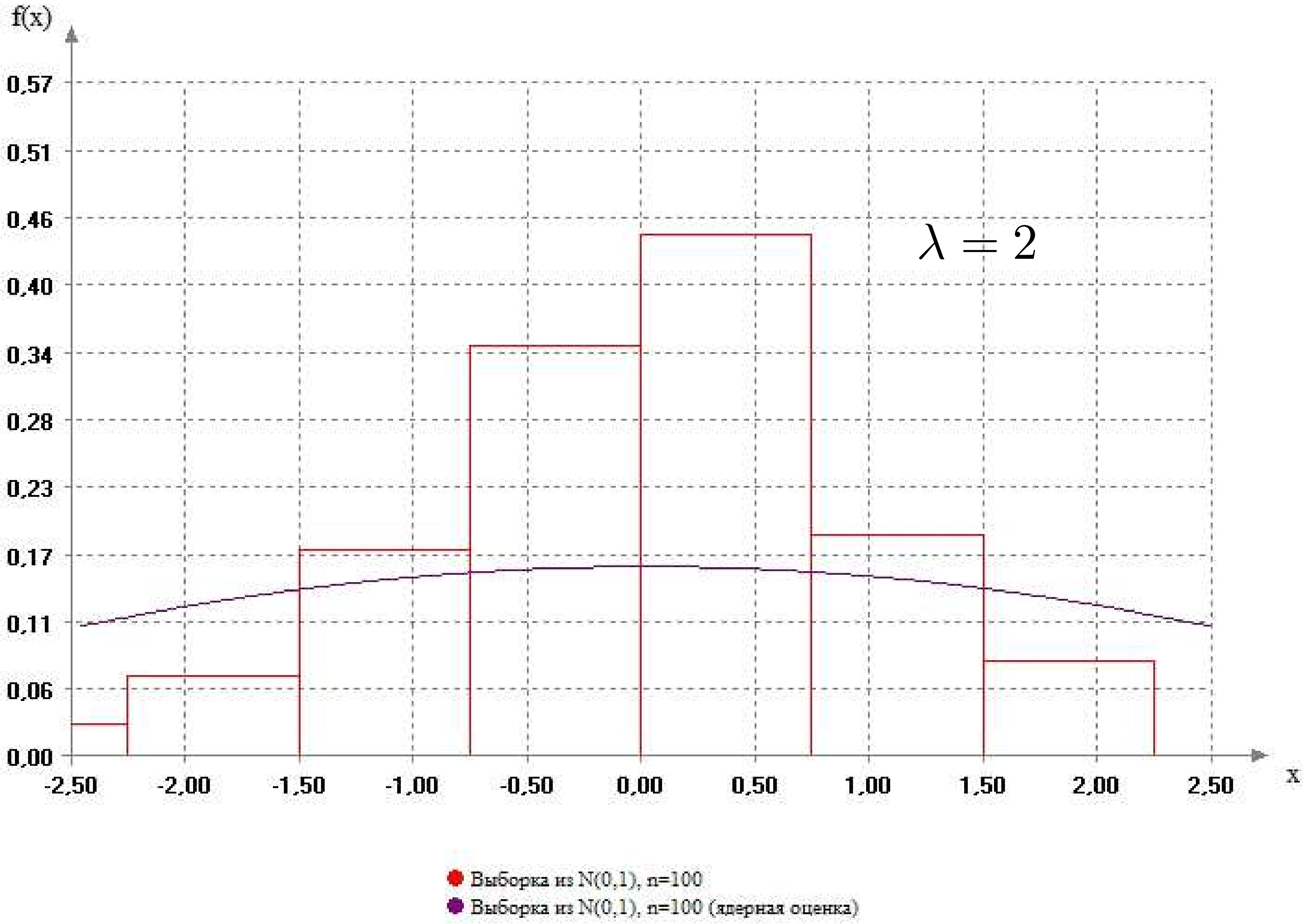
Ядерная оценка плотности



Ядерная оценка плотности



Ядерная оценка плотности



Достоинства "ядерного" оценивания плотности

Достоинства "ядерного" оценивания плотности

- Гладкость оценки

Достоинства "ядерного" оценивания плотности

- Гладкость оценки
- Быстрая сходимость к функции плотности

Достоинства "ядерного" оценивания плотности

- Гладкость оценки
- Быстрая сходимость к функции плотности

Недостатки

- требуется подбирать функцию ядра и параметр размытости

Достоинства "ядерного" оценивания плотности

- Гладкость оценки
- Быстрая сходимость к функции плотности

Недостатки

- требуется подбирать функцию ядра и параметр размытости
- проблема оценивания плотности распределений, определенных на отрезке или полупрямой