

Лекция 2

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Предмет и методы

ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Что общего и чем отличаются ТВ и МС?

▲ ТВ: разработка методов нахождения вероятностей сложных событий, исходя из известных вероятностей более простых событий.

▲ МС:

- прикладная дисциплина, базируется на понятиях и методах теории вероятностей, но решает задачи, обратные теории вероятностей
- восстанавливает по данным измерений или наблюдений неизвестные вероятности событий или неизвестные законы распределения случайных величин.

- разрабатывает методы, позволяющие по статистическим данным делать выбор одного из нескольких, противоречащих друг другу, предположений (гипотез) относительно законов распределения случайных величин или о значениях параметров распределений.
- разрабатывает методы получения, описания и обработки опытных данных для изучения закономерностей случайных массовых явлений

Особенность идей и методов математической статистики — универсальность, возможность использования в различных приложениях.

ОСНОВНЫЕ ПОНЯТИЯ МС

- ▲ Генеральная совокупность
- ▲ Выборка
- ▲ Вариационный ряд
- ▲ Теоретическая функция распределения

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

Пусть исследуется некоторая совокупность объектов, каждому из которых ставится в соответствие некоторая числовая функция — случайная величина X распределенная по некоторому неизвестному закону.

Практически, мы отождествляем наблюдаемые объекты и сопоставляемые им случайные величины, абстрагируясь от физической природы объектов.

Поэтому **генеральной совокупностью** будем считать множество значений, которые может принимать случайная величина **X** .

Выборка

В ходе каждого из испытаний мы случайным образом выбираем один из элементов генеральной совокупности и находим соответствующее ему значение X . Набор чисел

$$X_1, X_2, \dots, X_n$$

будем называть **выборкой** объема n из генеральной совокупности, а числа X_i — элементами выборки.

ВЫБОРКИ ДОЛЖНЫ БЫТЬ РЕПРЕЗЕНТАТИВНЫМИ

Т.Е. Представительными, — должны давать обоснованное представление о генеральной совокупности.

Чтобы обеспечить представительность, выборка должна быть *случайной*.

Теоретическая функция распределения

Рассмотрим выборку единичного объема X_1 .

Поскольку выбор случаен, то X_1 – случайная величина и, как всякая случайная величина, имеет функцию распределения $F(x) = P(X_1 < x)$. Для выборки произвольного объема n каждый элемент будет иметь точно такую же функцию распределения, если

- выборка с возвращением или
- генеральная совокупность бесконечного объема.

С точки зрения теории вероятностей выборку

$$X_1, X_2, \dots, X_n$$

можно трактовать как совокупность **независимых, одинаково распределенных** случайных величин с функцией распределения $F(x) = P(X < x)$.

Функция $F(x)$ называется *теоретической* функцией распределения.

Совместная функция распределения выборки задается формулой:

$$F_{x_1, \dots, x_n}(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n) = F(x_1) \boxtimes F(x_n)$$

Простейшие статистические преобразования

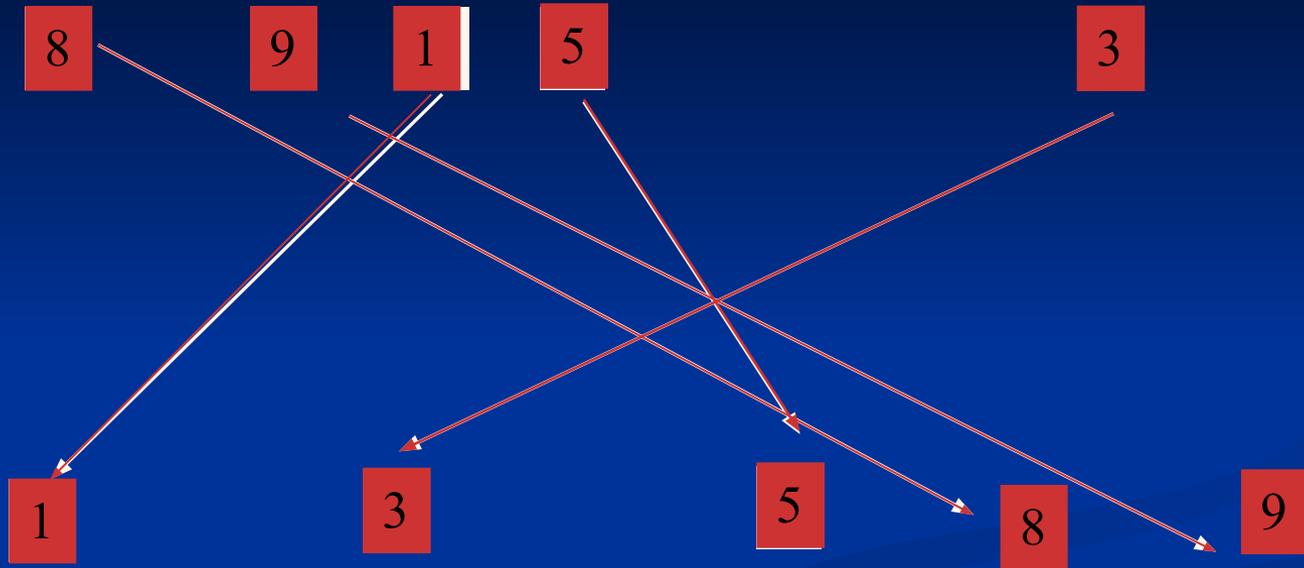
Вариационный и статистический ряды

Вариационный ряд $X_{(1)}, \dots, X_{(n)}$ представляет собой ту же выборку X_1, \dots, X_n , но расположенную в порядке возрастания элементов:

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

Такое преобразование выборки не приводит к потере информации относительно теоретической функции распределения

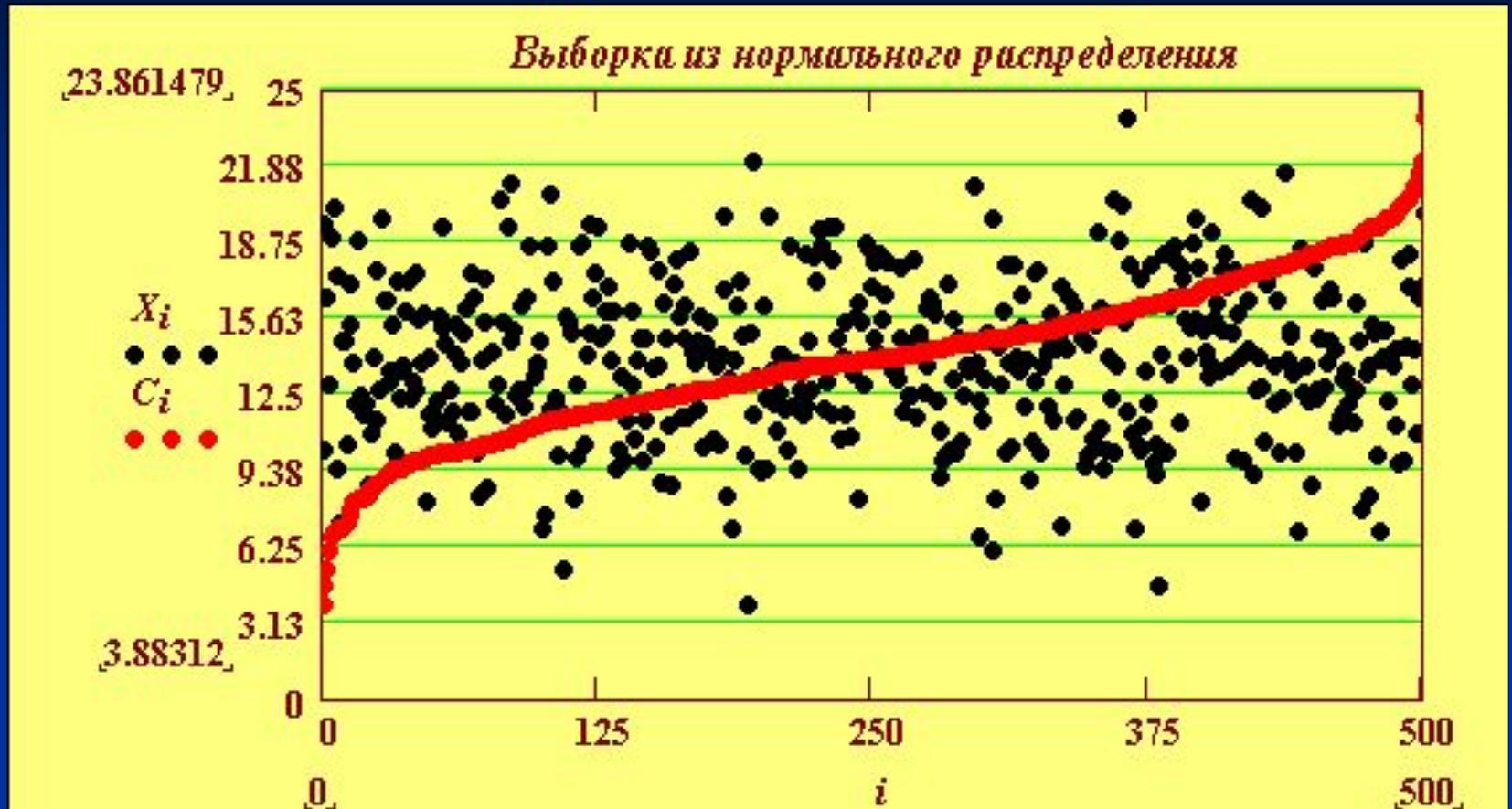
Выборка



Вариационный ряд

РАНГ элемента выборки -- порядковый номер элемента в вариационном ряду

ВЫБОРКА И ВАРИАЦИОННЫЙ РЯД



Если среди элементов выборки (**вариационного ряда**) есть одинаковые, то наряду с **ВАРИАЦИОННЫМ** рядом используется **СТАТИСТИЧЕСКИЙ РЯД** --- таблица, в которой указаны все различные значения вариационного ряда (**ВАРИАНТЫ**) и их количество.

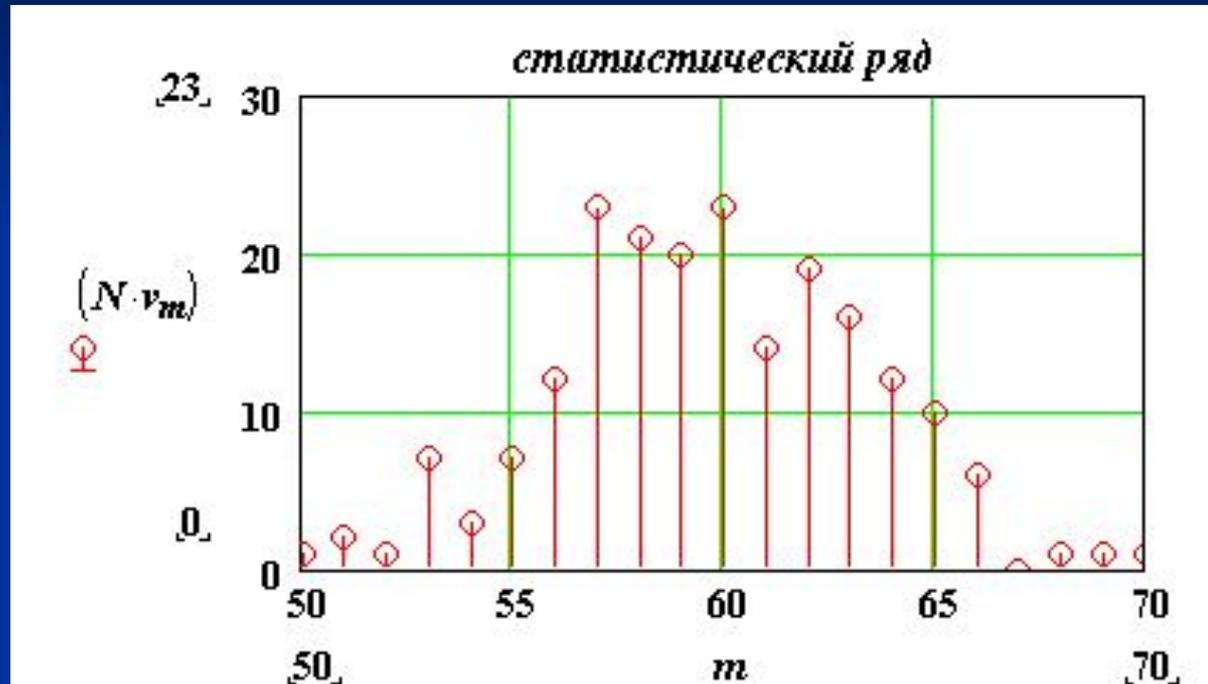
Статистический ряд характерен для выборок из дискретных распределений, а также и для выборок из непрерывных распределений, полученных при измерениях с округлением.

Статистический ряд

$$Z_1 < Z_2 < \dots < Z_k$$

Z_1	...	Z_k
m_1	...	m_k

ВЫБОРКА ИЗ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ



СТАТИСТИКИ

Статистика S --- это произвольная измеримая k -мерная функция от выборки, не содержащая неизвестных параметров распределений.

$$S = (S_1, S_2, \dots, S_k)$$

Достаточные статистики --- такие, которые содержат всю ту информацию о теоретической функции распределения, что и выборка

$$X_1, X_2, \dots, X_n$$

ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ (ЭФР)---

аналог теоретической функции:

$$F_e(x) = m_x / n$$

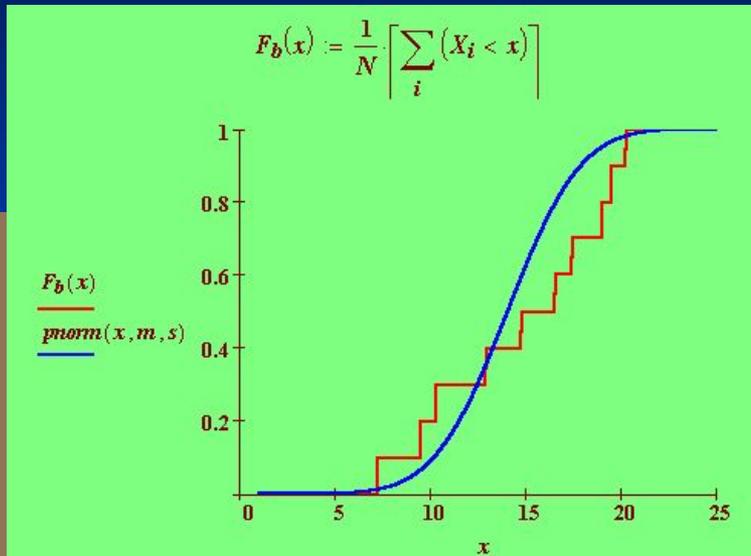
где

m_x --- число элементов выборки, значения которых

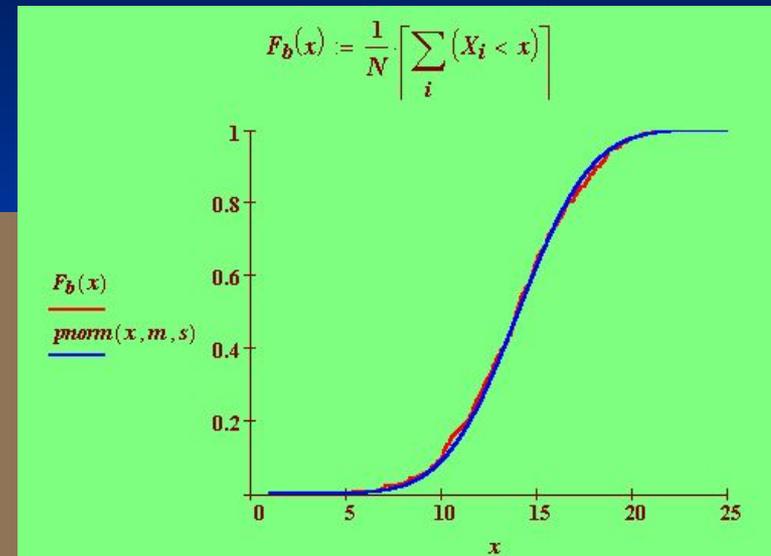
не превышает данное x

n ---объем выборки.

Теоретическая функция распределения и её оценка

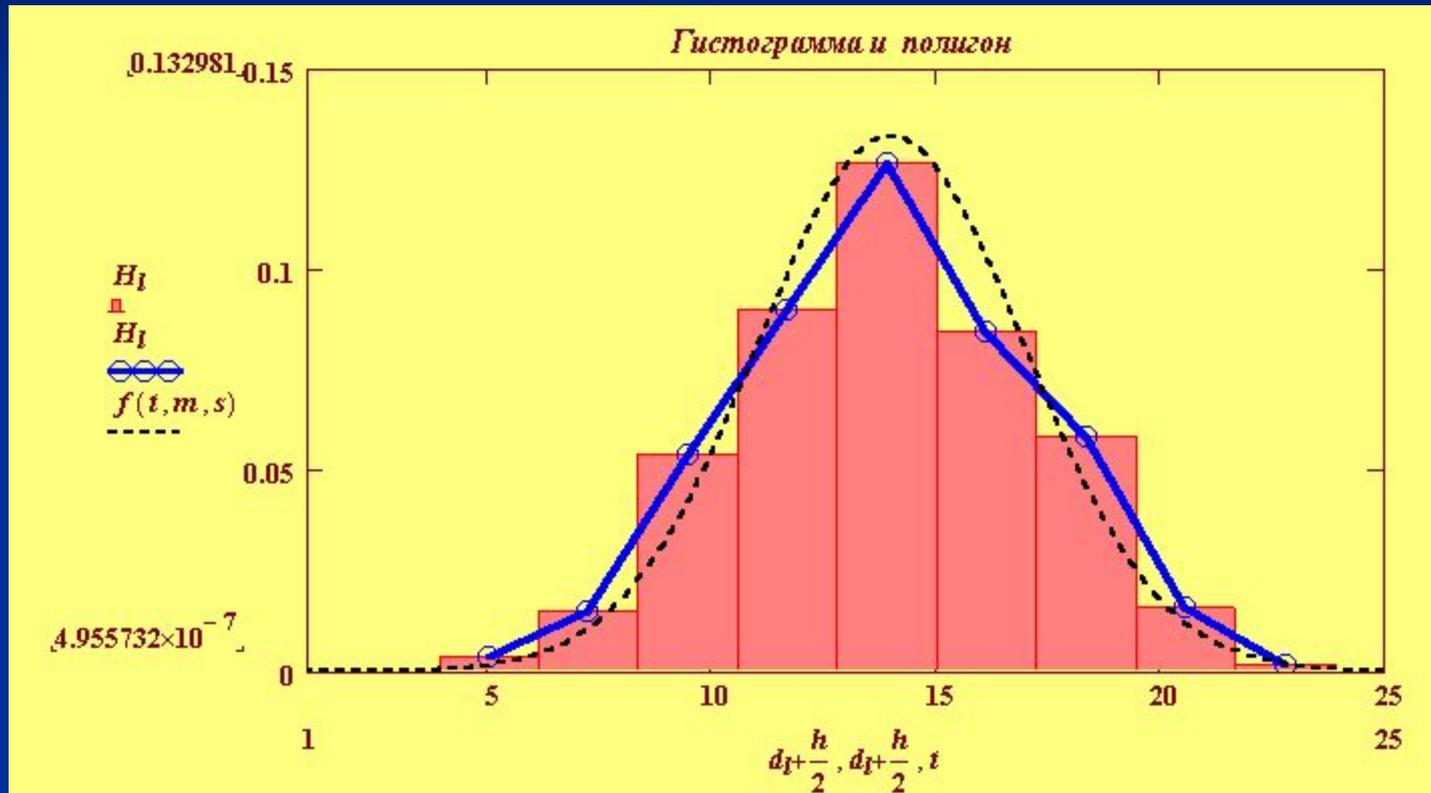


$n = 10$

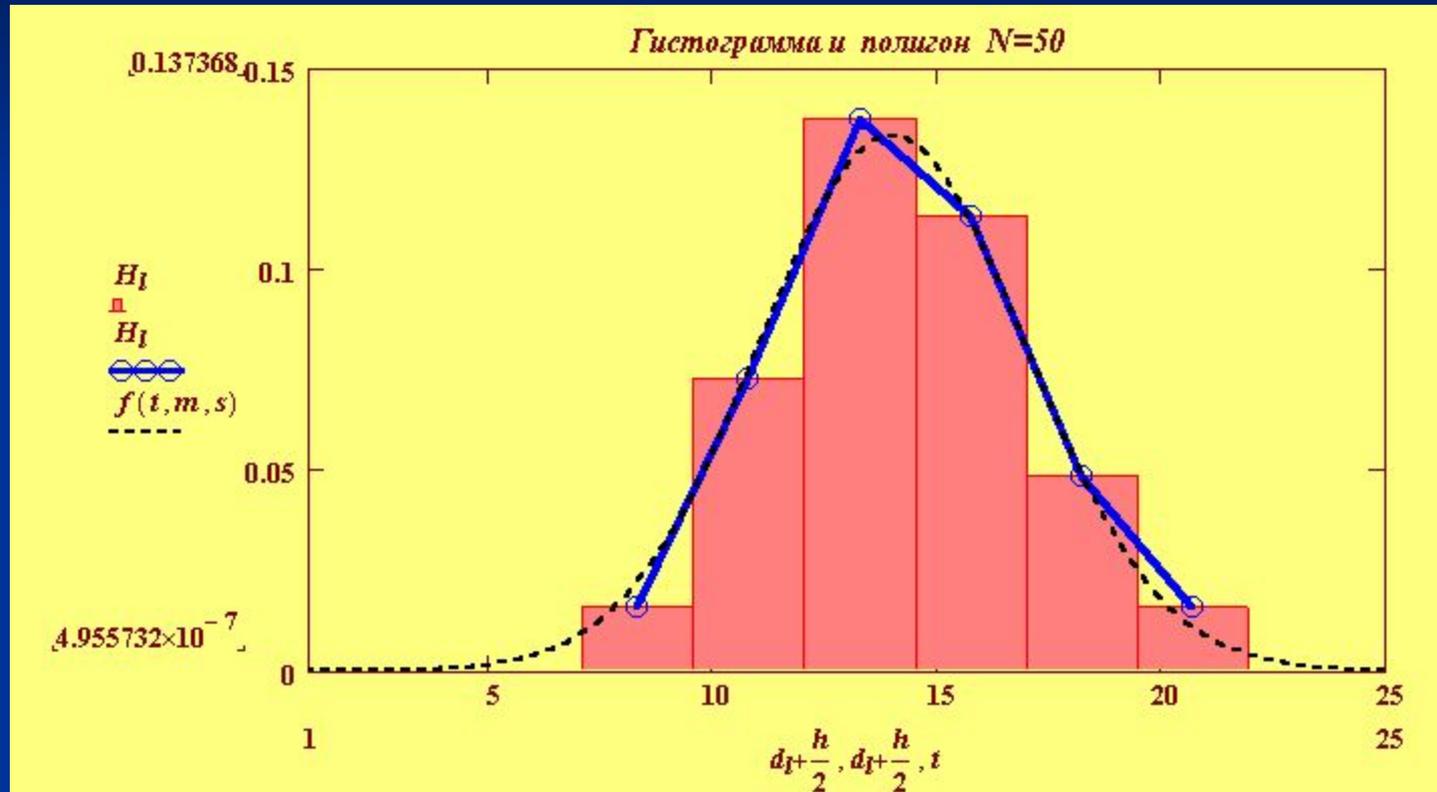


$n = 500$

Гистограмма и полигон



Гистограмма и полигон



Предельное поведение

$$\frac{m_k}{n \cdot h_k} \xrightarrow{p} f(x_k)$$

$$F_e(x) = \frac{m_x}{n} \xrightarrow{p} F(x)$$