Статистические методы обработки медико-биологических данных
Нормальный закон распределения

Тема: Статистические методы обработки медико-биологических данных

Нормальный закон распределения.

План лекции:

- Понятие случайных **дискретных** и **непрерывных** величин.
- Распределения и характеристики случайных величин.
- Нормальный закон **распределения.** **Кривая** Гаусса и ее особенности. Правило «трёх **сигм».**

В медицине необходимо вести учет, анализ и прогноз различных массовых явлений. В целом, массовым явлениям присущи свои особые закономерности. К доктору обращаются пациенты с различными заболеваниями. Болезнь конкретного человека - случайное событие у врача. Но случайные события предсказуемы, например, в период эпидемии гриппа наиболее часто встречаются заболевания гриппом.

Закономерности массовых случайных событий - статистических данных, отражающих эти события, - изучаются с помощью математической статистики.

Типичная задача математической статистики - это приближенная оценка неизвестной вероятности случайного события по результатам наблюдений, экспериментов, когда событие может происходить или не осуществляться.

Случайной величиной

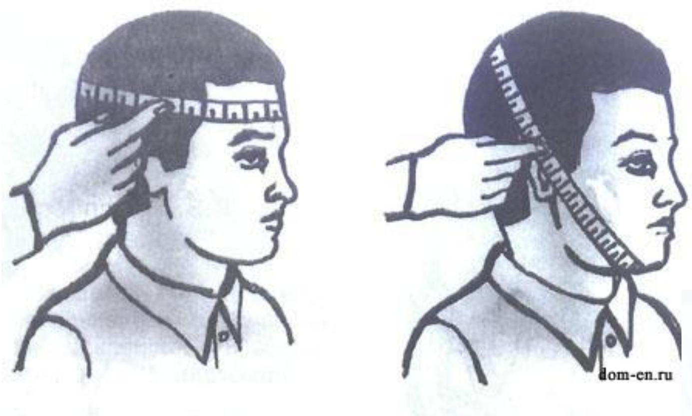
называется переменная **величина, значение** которой зависит от исхода некоторого **испытания.**

Дискретная **Непрерывная**

Дискретной называется случайная величина, которая может принимать значения некоторой конечной или бесконечной числовой последовательности (число слов в **тексте, студентов** в аудитории, больных в клинике...)

Непрерывной называется случайная **величина,** которая может принимать любые значения внутри некоторого интервала (масса, температура, рост...)

Дискретная ? или



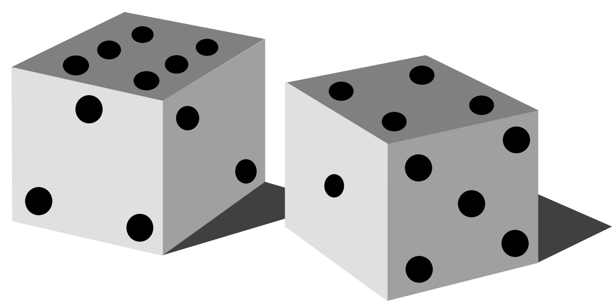
Непрерывная?

Статистический ряд – результаты измерений для статистического исследования, записанные последовательно по порядку их получения. Удобнее представить в таблице.

$X_1, X_2, X_3 \ldots\ldots\ldots X_n$

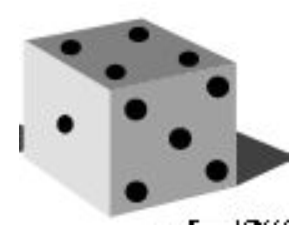
X_1	$X_{2},$	X_3	…	…	…	…	…
…	…	…	…	…	…	…	…
…. .	…	…. .	…	…	…	….	…
….	…	..	…	…	…	….	X_n

Распределение дискретной случайной величины.



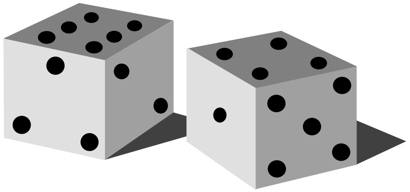
Дискретная случайная величина считается заданной, если указаны ее возможные значения и соответствующие им вероятности

Совокупность X и P называется **распределением дискретной случайной величины**.



Дискретные случайные величины x_i	X_1	X_2	X_3	X_4	X_5	X_6
Вероятность p_i	P_1	P_2	P_3	P_4	P_5	P_6

n – общее число случайных событий



Дискретные случайные величины x_i	X_1	X_2	X_3	…	…	Xn
Вероятность p_i	P_1	P_2	P_3	…	…	Pn

Дискретные случайные величины x_i	X_1 1	X_2 2	X_3 3	X_4 4	X_5 5	X_6 6
Вероятность p_i	P_1 1/6	P_2 1/6	P_3 1/6	P_4 1/6	P_5 1/6	P_6 1/6

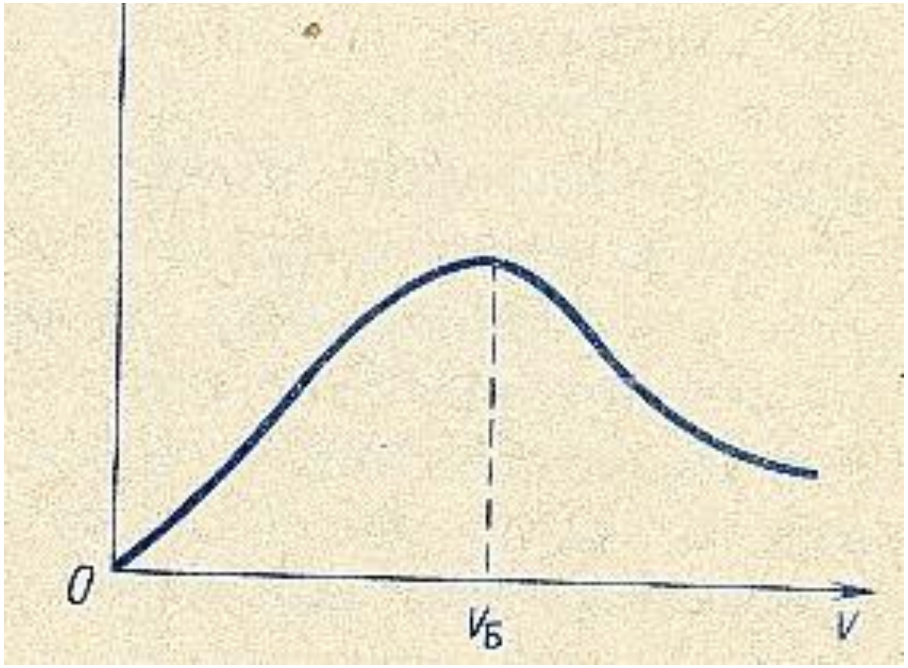
$$\sum_{i=1}^{n} p_i = 1$$

условие нормировки дискретных случайных величин

Различные распределения

1. Биномиальное распределение (позволяет определить вероятность того, что событие *A* произойдет m раз при *n* испытаниях).

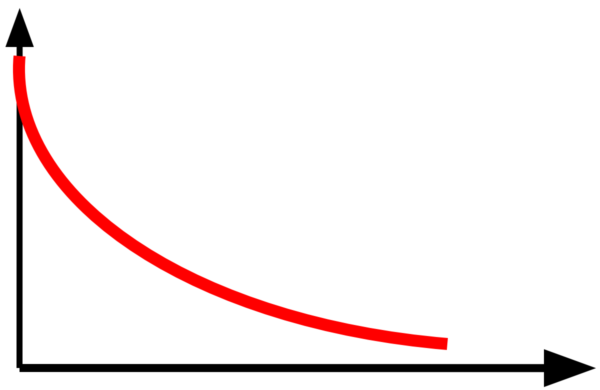
2. Распределение Максвелла (распределение молекул газа по скоростям, кинетическим энергиям). График - кривая ***Максвелла.***

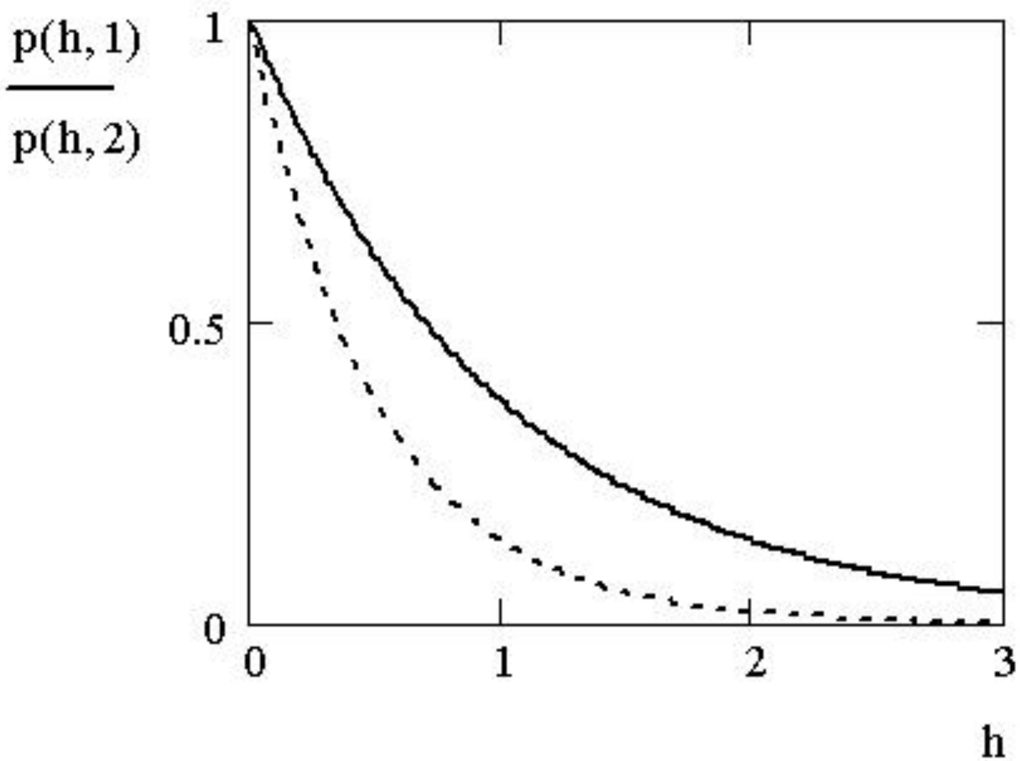


3. Распределение Больцмана

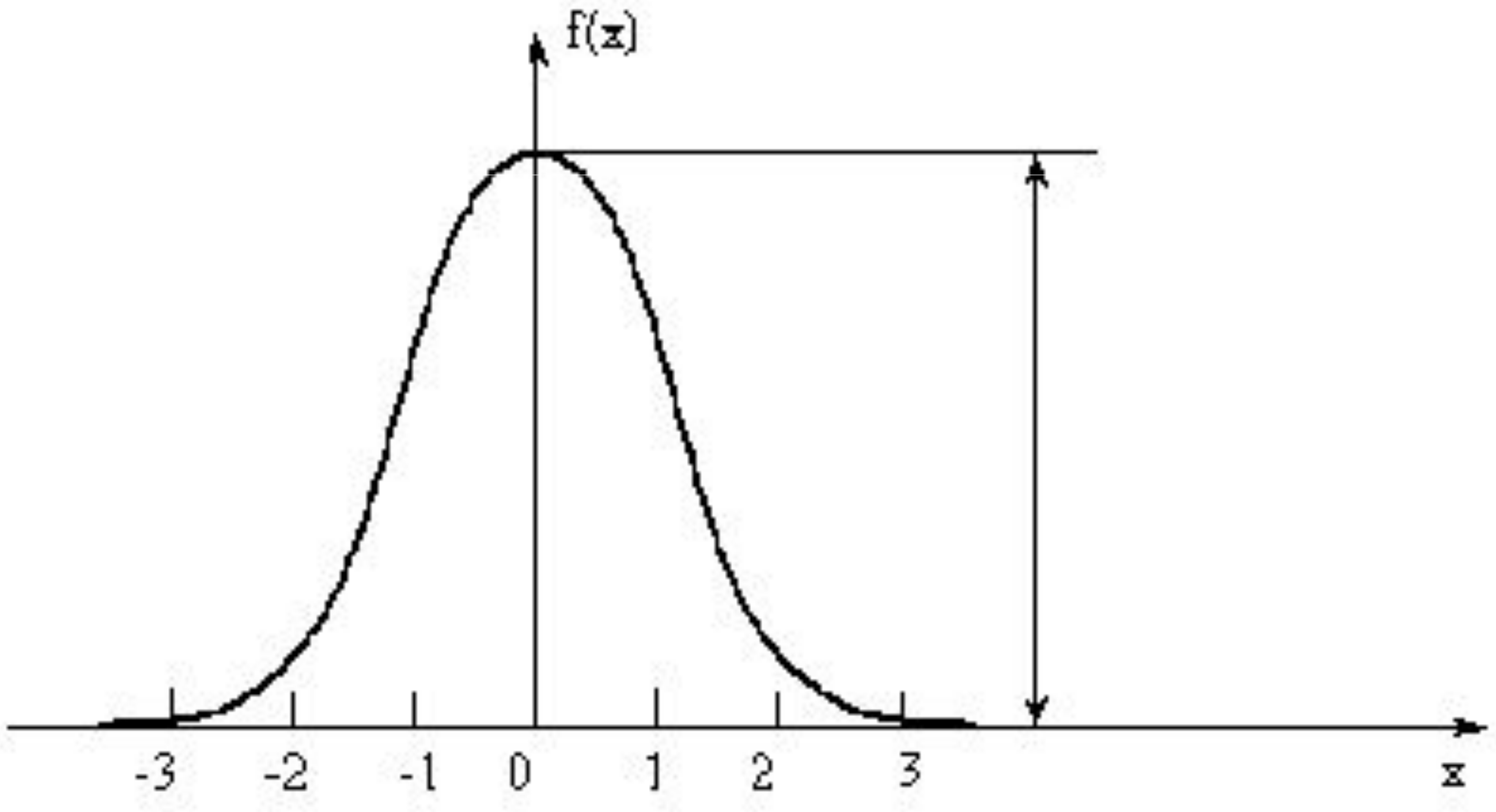
(распределение частиц по потенциальным энергиям в силовых полях - гравитационном, электрическом).

График - экспонента





4. Нормальное распределение (график - кривая Гаусса)



5. Распределение Пуассона и др. ...

Нормальный закон распределения имеет **важное практическое значение в естественных науках.** Оказывается, распределение роста, массы новорожденных и много других случайных событий физической и биологической природы описываются нормальным законом распределения и графически иллюстрируются кривой Гаусса.

Числовые характеристики дискретных случайных величин.

1.**Математическое ожидание** случайной величины есть сумма произведений всех возможных ее значений на вероятности этих значений:

$$M(x) = \sum_{i=1} x_i p_i$$

$$M(x) = x_1 \cdot p_1 + x_2 \cdot p_2 + ... + x_n \cdot p_n$$

2. Среднее арифметическое значение

$$\overline{X} = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3 + m_4 x_4 \ldots + \ldots m_n x_n}{n}$$

$$= \frac{m_1}{n} x_1 + \ldots \frac{m_n}{n} x_n = p_1 x_1 + \ldots + p_n x_n$$

n – число измерений.

$\overline{X}$ <X>

Если n велико , то относительные частоты m/n = p,

а среднее арифметическое значение практически равно математическому ожиданию.

$$\langle X\rangle = M(x)$$

Математическое ожидание часто отождествляют со средним значением

3. **Дисперсия случайной величины** – это математическое ожидание квадрата отклонения случайной величины от ее математического ожидания;

$$D(x) = M(x^2) - [M(x)]^2$$

$$D(x) = M[x - M(x)]^2$$

Дисперсия характеризует **рассеяние** случайных величин относительно математического ожидания.

0 125 130

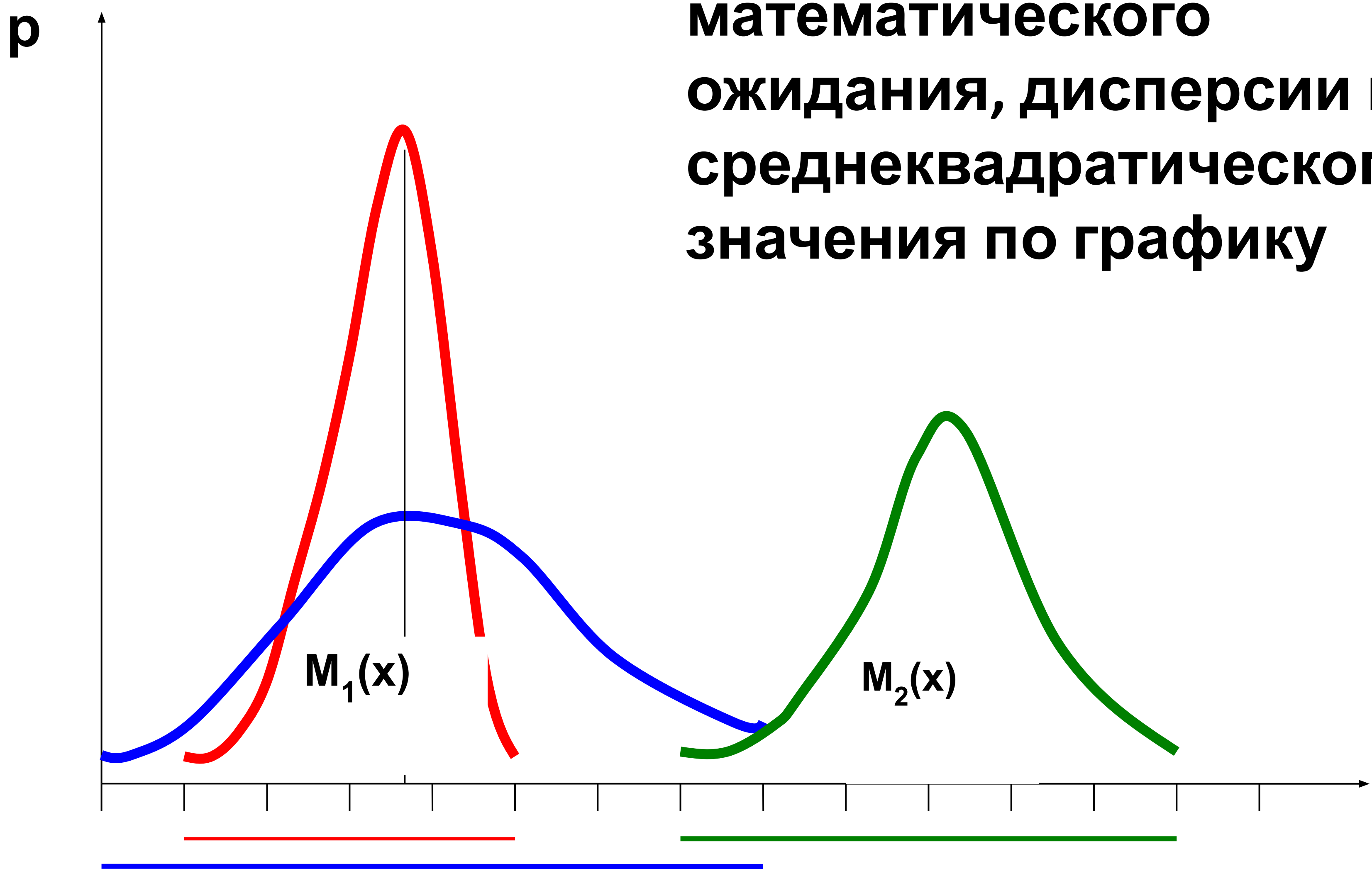
120 150 180

Размерность дисперсии - квадрат размерности случайной величины, поэтому введена величина

$$\sigma = \sqrt{D(x)}$$

4. σ - среднеквадратическое отклонение, которое имеет размерность случайной величины.

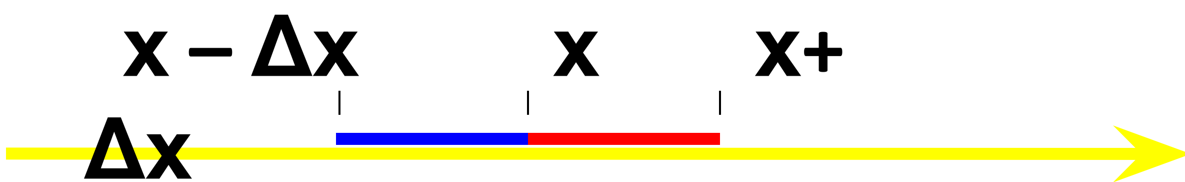
Сравнительный анализ значений математического ожидания, дисперсии и среднеквадратического значения по графику



Числовые характеристики непрерывных случайных величин.

непрерывная случайная величина X принимает значения между x и x± Δx

$$dP = f(x)dx$$



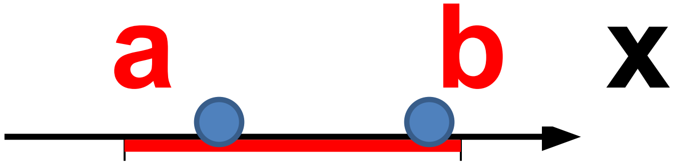
$dP = f(x)dx$, где $f(x)$ - плотность вероятности или функция распределения вероятности.

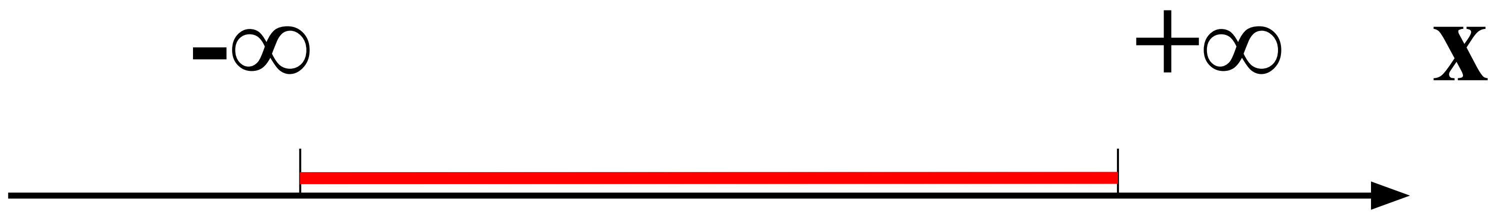
функция распределения вероятности показывает, как изменяется вероятность, отнесенная к интервалу dx случайной величины в зависимости от значения самой этой величины: *f* (x) = dP/dx

$$P_{ab} = \int_a^b f(x)dx$$

$$P_{ab} = \int_a^b f(x)dx$$

- вероятность того, что случайная величина принимает значения в интервале (ab).





Какова вероятность того, что случайная величина находится в данном интервале?

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

условие нормировки для непрерывной случайной величины

1. Математическое ожидание M(x):

$$M(x) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

2. Дисперсия D(x) :

$$D(x) = \int_{-\infty}^{+\infty} [x - M(x)]^2 \cdot f(x) dx$$

3. Среднеквадратическое отклонение,

которое имеет размерность случайной величины.

$$\sigma = \sqrt{D(x)}$$

Нормальный закон распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{[(x - M(x)]^2}{2\sigma^2}\right]$$

exp - экспонента;

$e^{\pm x} = \exp(\pm x)$;

График нормального закона - кривая Гаусса.

Учитывая, что $\sigma = \sqrt{D}$

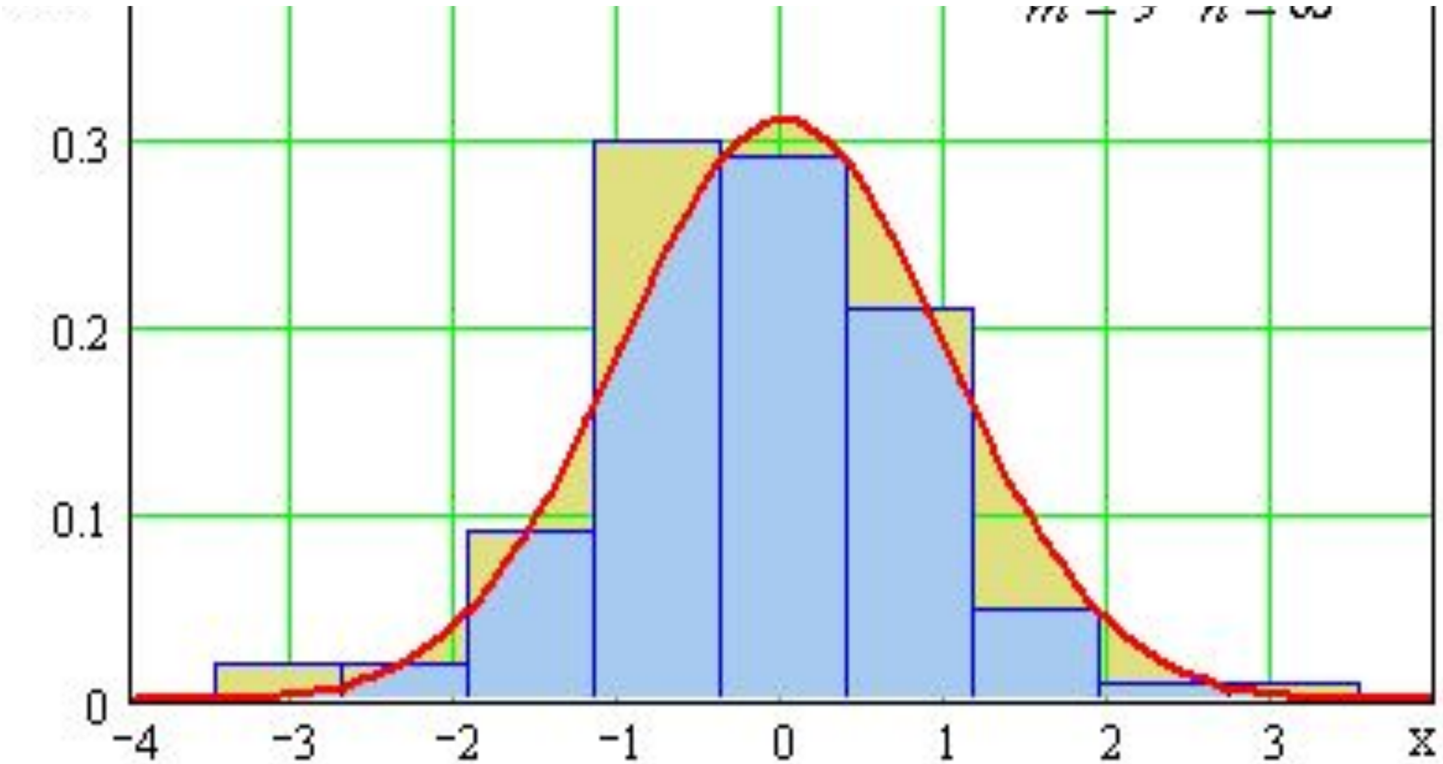
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{[x-M(x)]^{2}}{2\sigma^{2}}\right]}$$

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-[x-M(x)]^2}{2\sigma^2}\right];$$

$$f(x)=\frac{1}{\sqrt{2\pi D}}\exp\left[-\frac{[x-M(x)]^2}{2D}\right]$$

Особенности кривой Гаусса

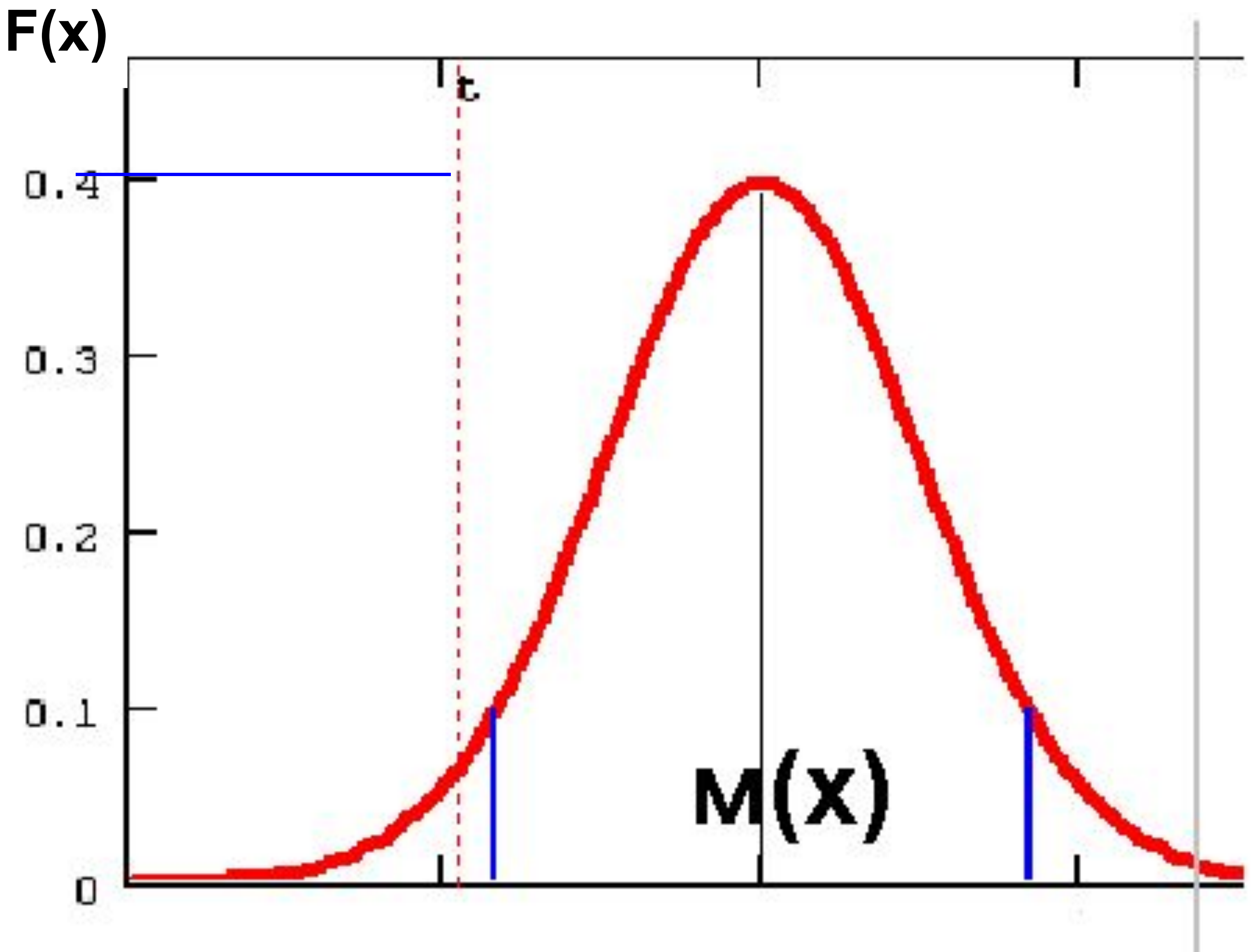
- колоколообразная форма



- ветви – экспоненты (возрастающая и убывающая)

• симметрия относительно M(X)=x.

M(X) - центр рассеивания



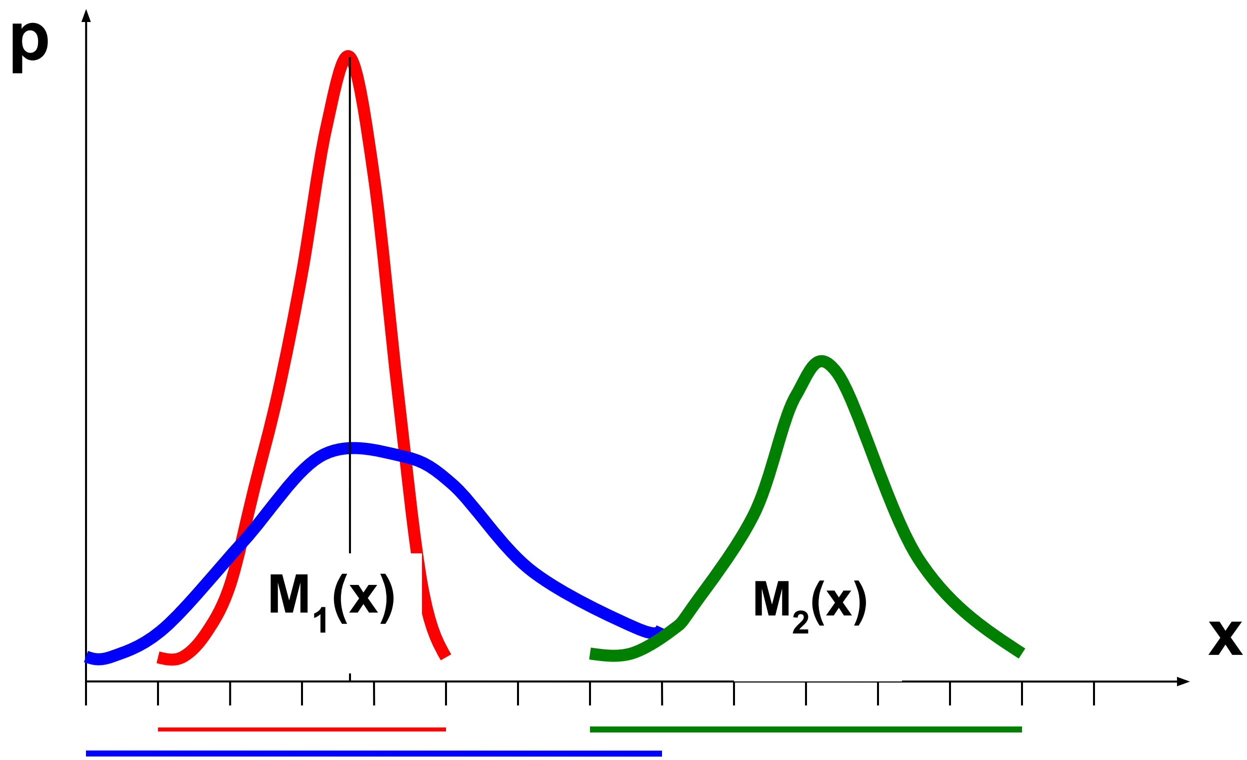
по данной формуле определяем координаты вершины кривой Гаусса, когда x = M(x).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{[x-M(x)]^2}{2\sigma^2}\right]}$$

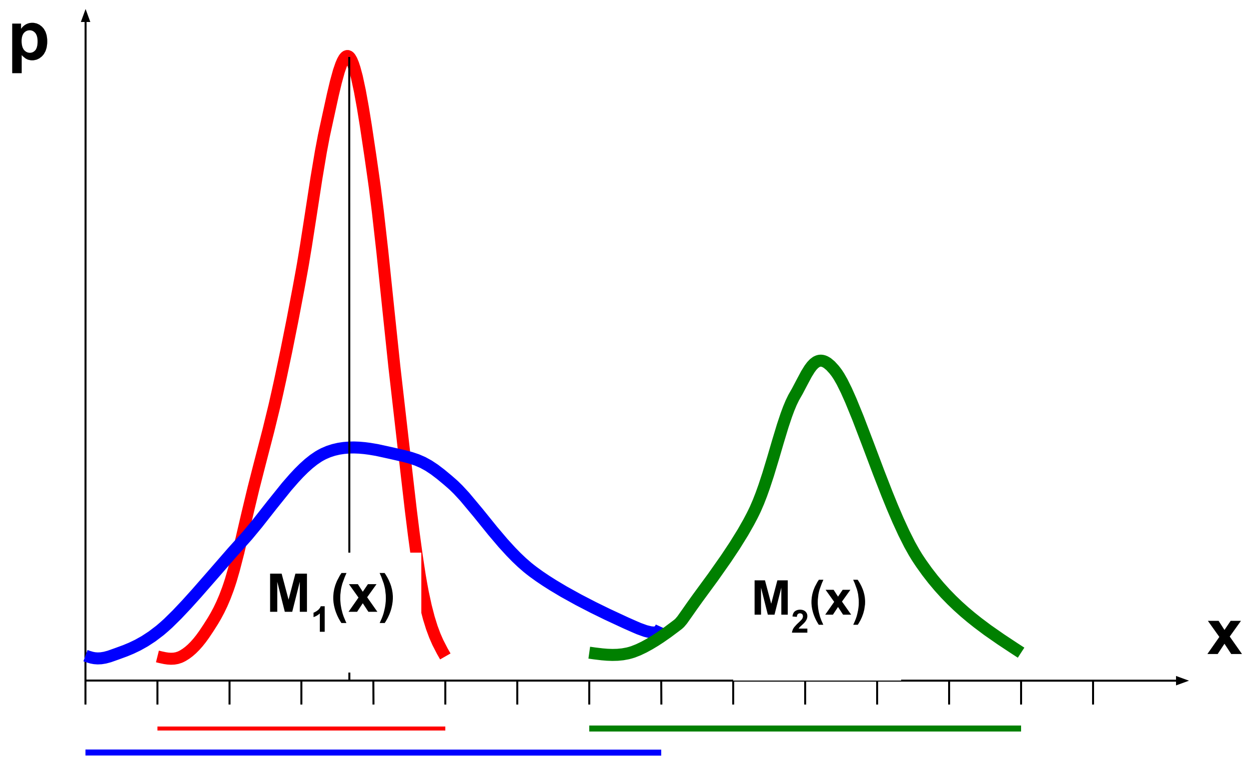
Вершина графика

$$f(x)_{\max} = \frac{1}{\sigma\sqrt{2\pi}}$$

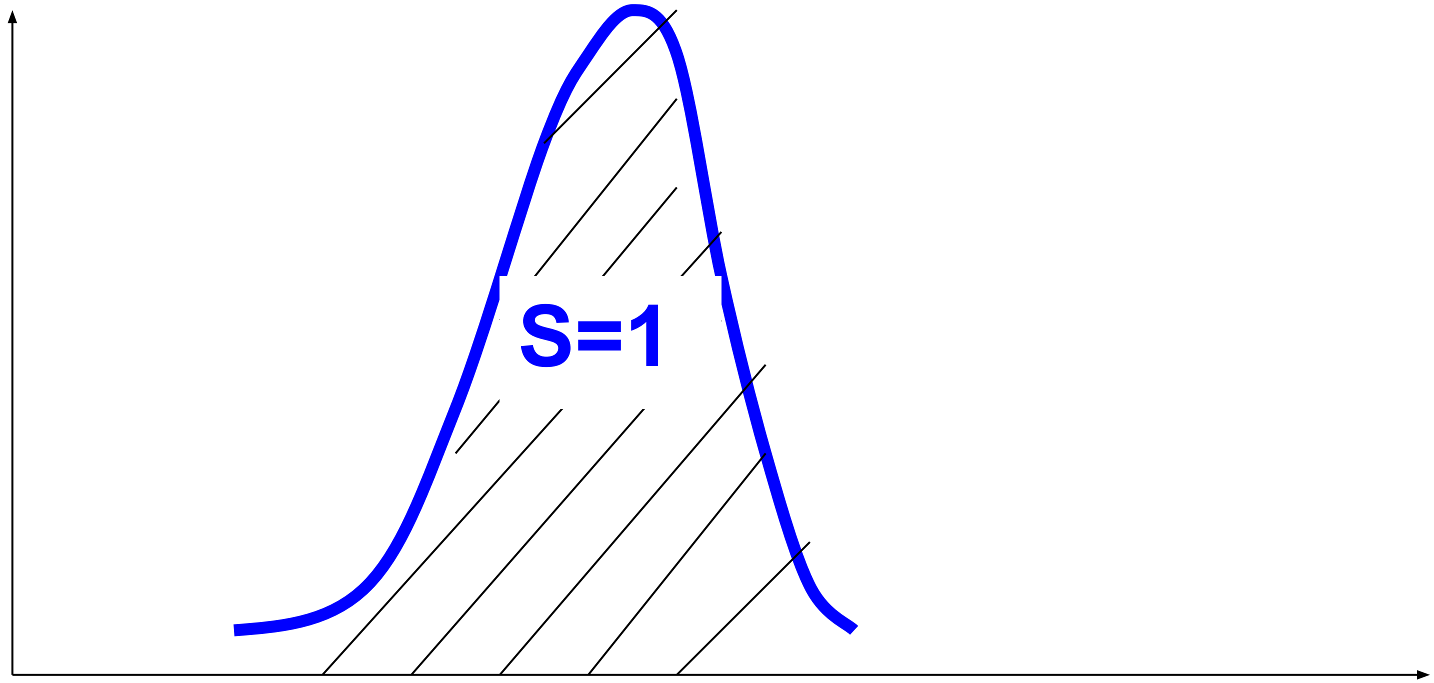
•ветви асимптотически приближаются к оси x. Чем больше σ, тем менее острая вершина.



•изменение математического ожидания M(X) сдвигает влево или вправо вершину кривой Гаусса

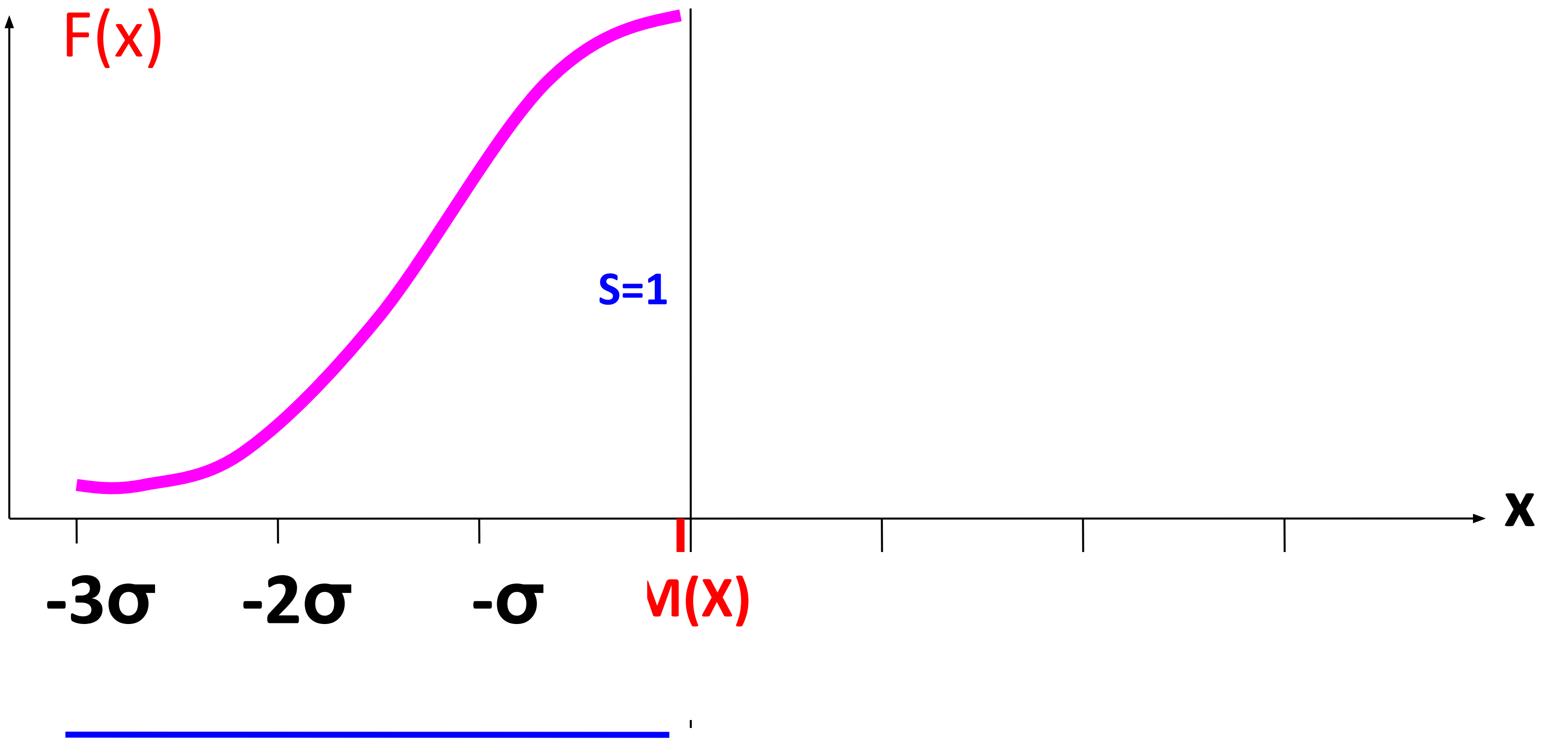


•площадь, заключенная под кривой равна 1 (условие нормировки)

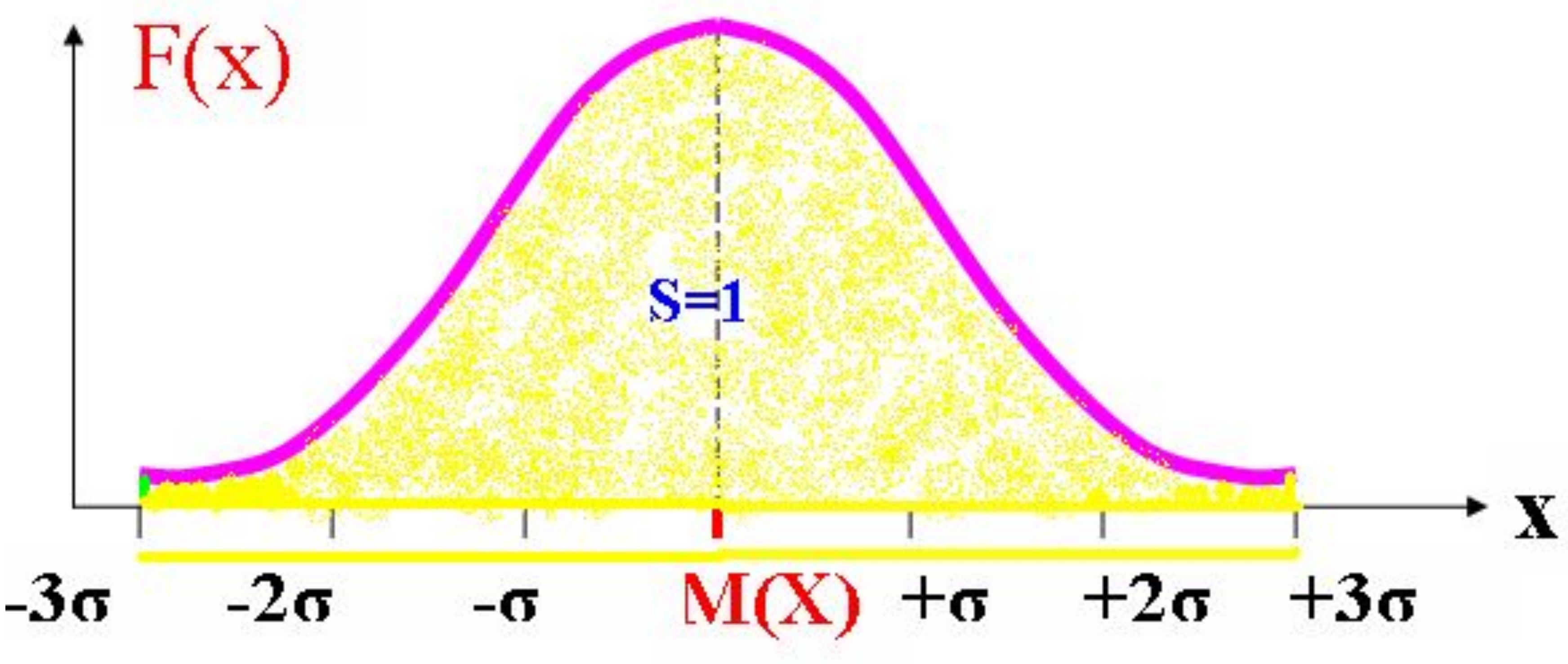


•**выполняется правило "трёх сигм".**

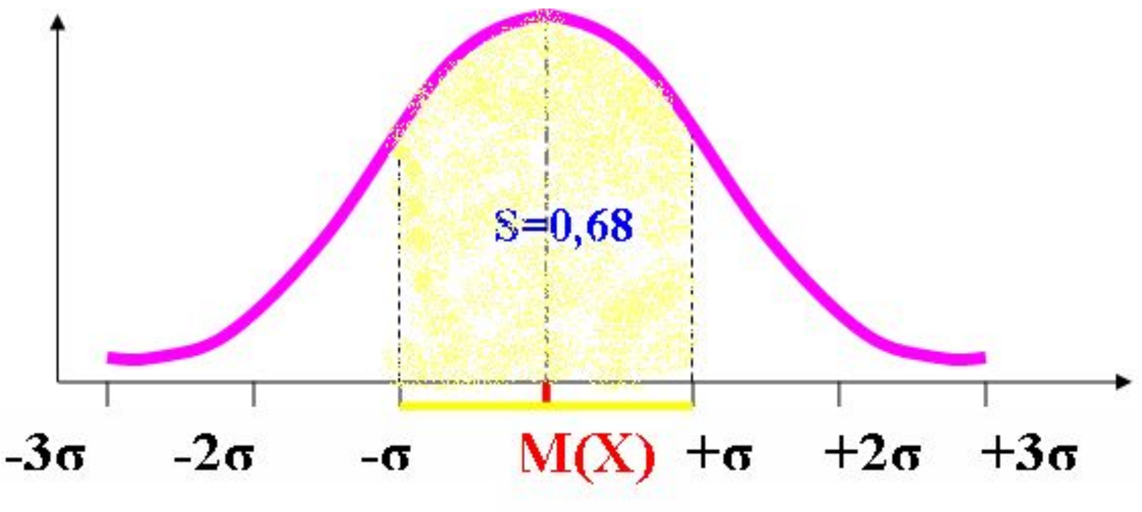
ГРАФИЧЕСКАЯ ИЛЛЮСТРАЦИЯ ПРАВИЛА«**3 σ**".



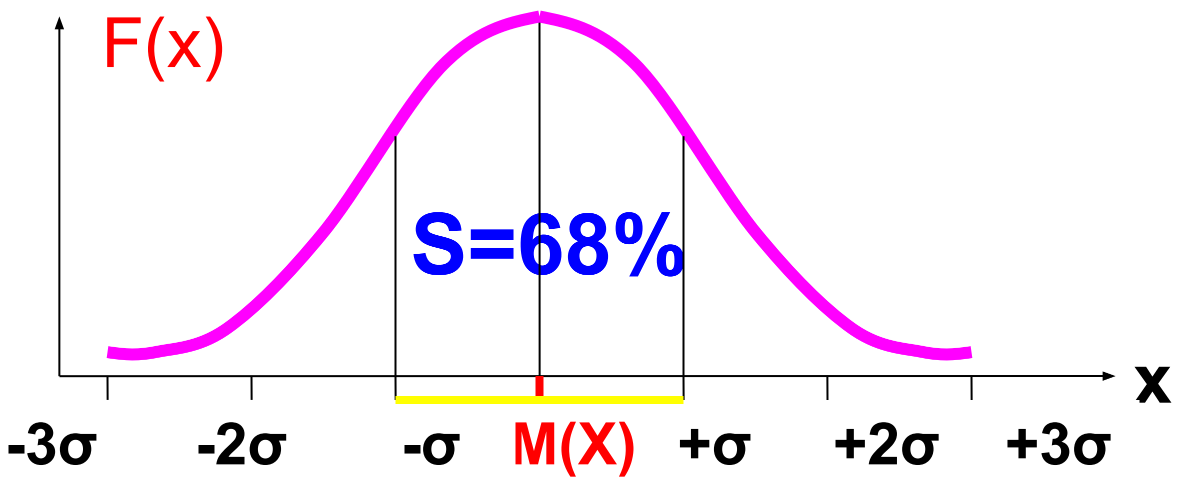
- Вероятность появления случайной величины в интервале значений $M(X) \pm 3\sigma$ равна 99,97% Это соответствует условию нормировки - **площадь под кривой равна 1**, т.е - практически все случайные величины нормального распределения находятся под кривой Гаусса.



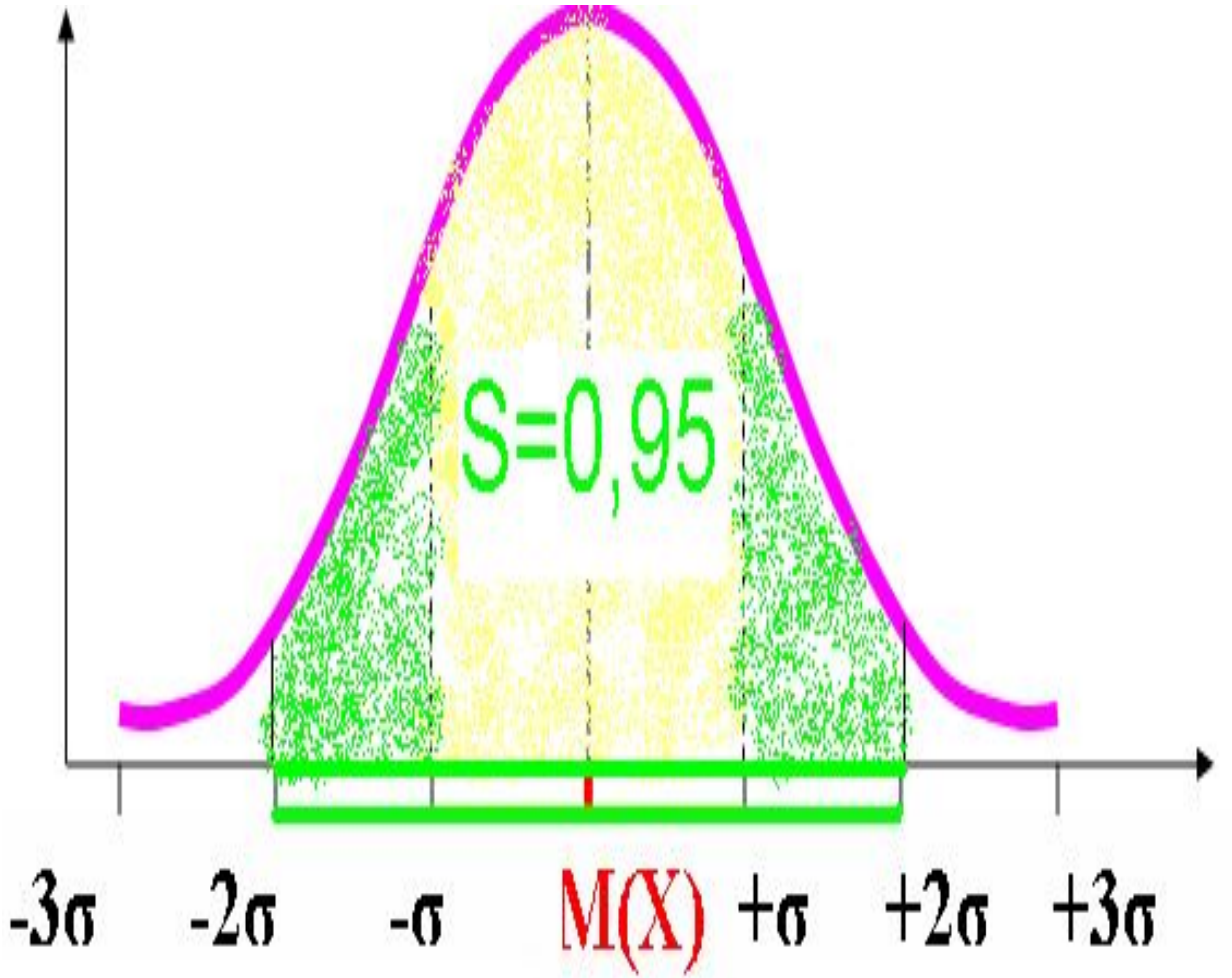
- Вероятность появления случайной величины в интервале значений M(x)±σ равна 68%



M(x)±σ



- Вероятность появления случайной величины в интервале значений M(x)± 2σ равна 95%

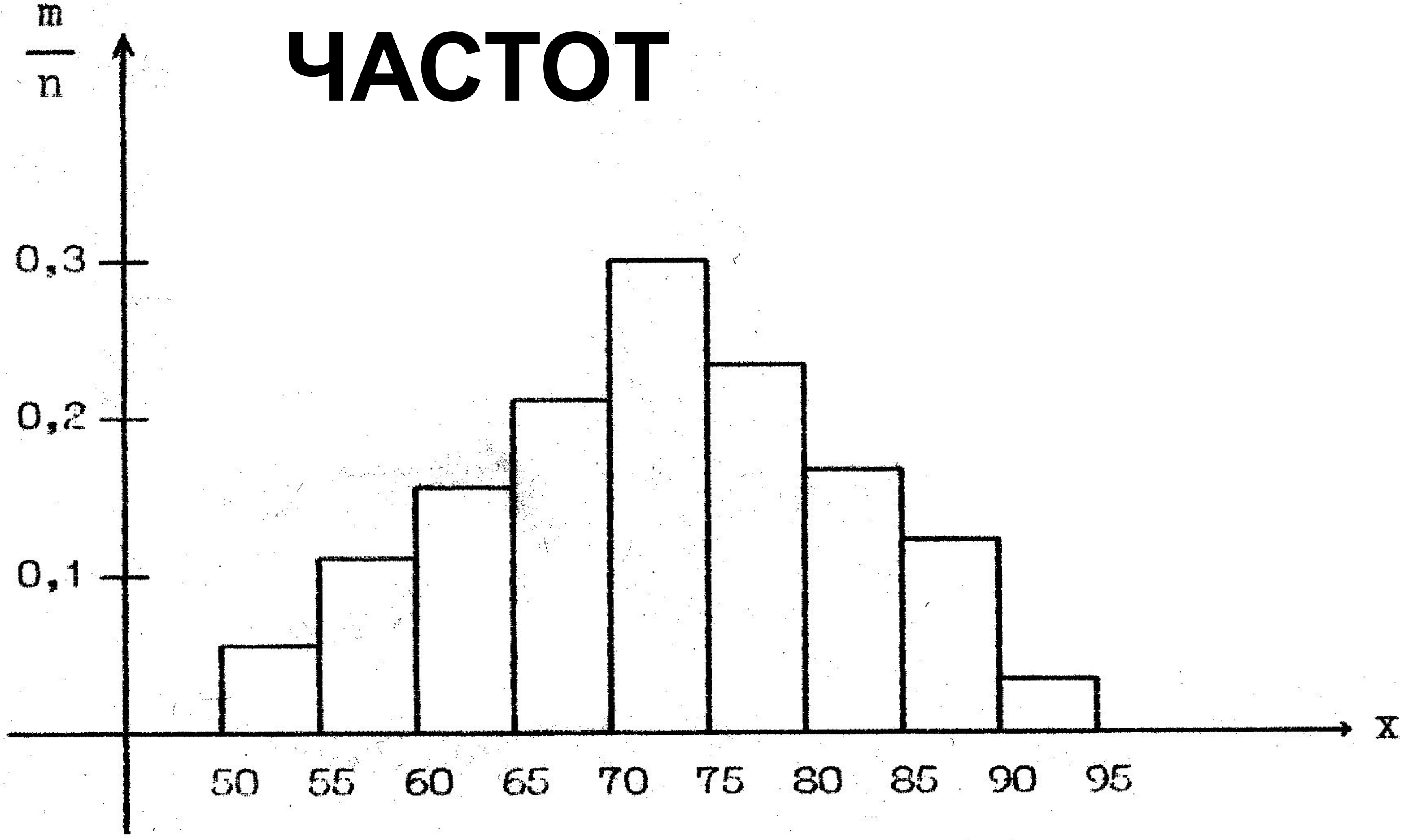


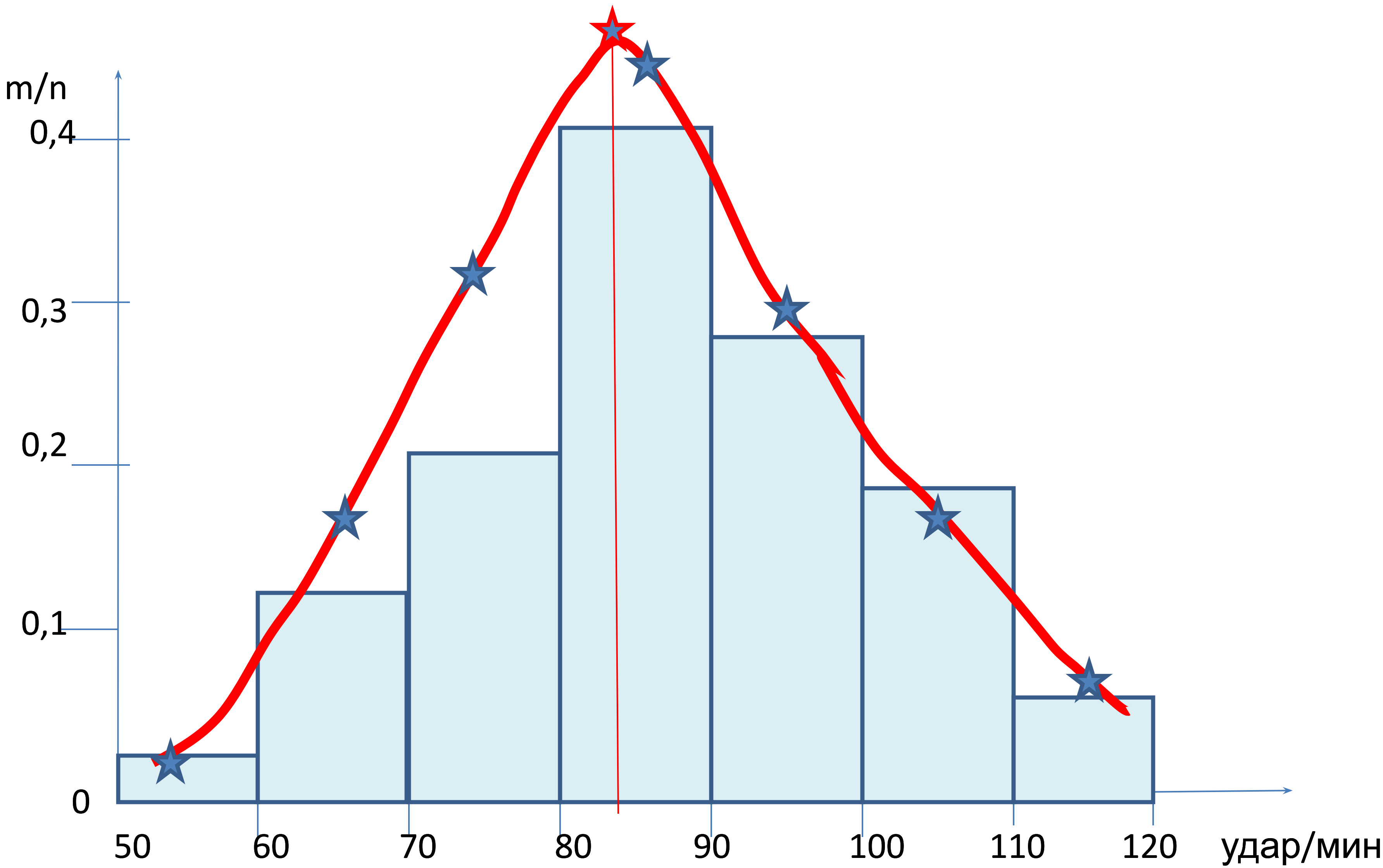
M(x)± 2σ

Для нормального закона **распределения** характерен симметричный вид **гистограммы**

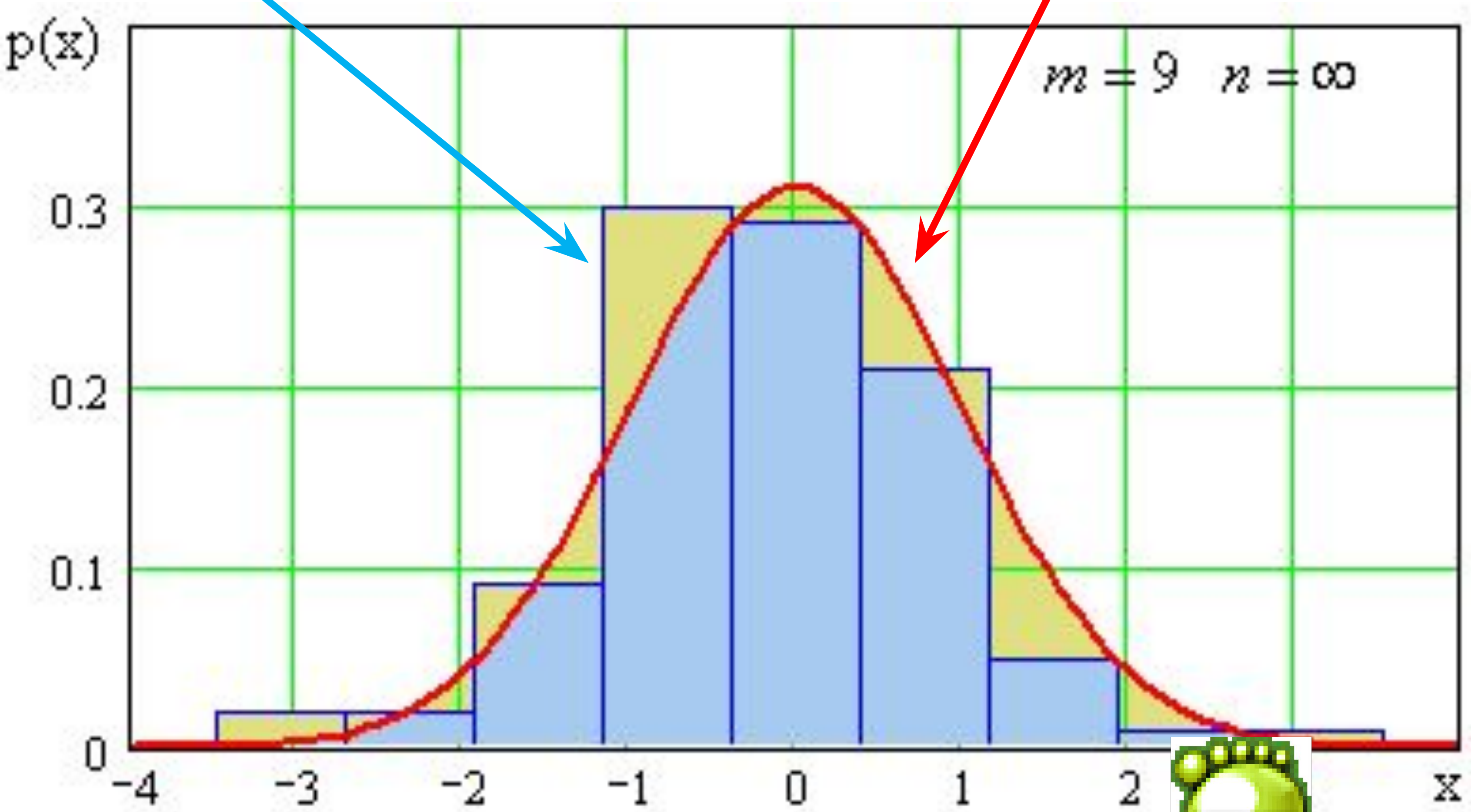
• Гистограмма частот - совокупность смежных прямоугольников, построенных на одной прямой линии. Основания прямоугольников одинаковы. Высоты прямоугольников равны относительной частоте **m/n** (вероятности).

ГИСТОГРАММА ЧАСТОТ




m/n
0,4
0,3
0,2
0,1
0
50
60
70
80
90
100
110
120
удар/мин

Гистограмма частот и кривая Гаусса



ãèñòîãðàììà êðèâàÿ Ãàóññà.mp