

# Случайные величины: законы распределения

# Что было: понятие о случайной величине

**СЛУЧАЙНОЙ ВЕЛИЧИНОЙ** называется величина, которая в результате испытания примет одно и только одно возможное значение, наперед не известное и зависящее от случайных причин, которые заранее не могут быть учтены.

**Функцией распределения случайной величины  $X$**  называется функция  $F(x)$ , выражающая для каждого  $x$  вероятность того, что случайная величина  $X$  примет значение, меньшее  $x$

$$F(x) = P(X < x).$$

# Что было: функция распределения

## Интегральная функция распределения

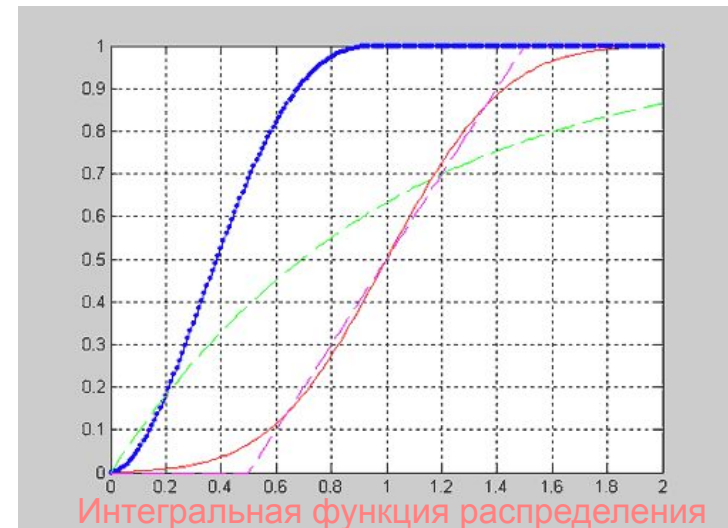
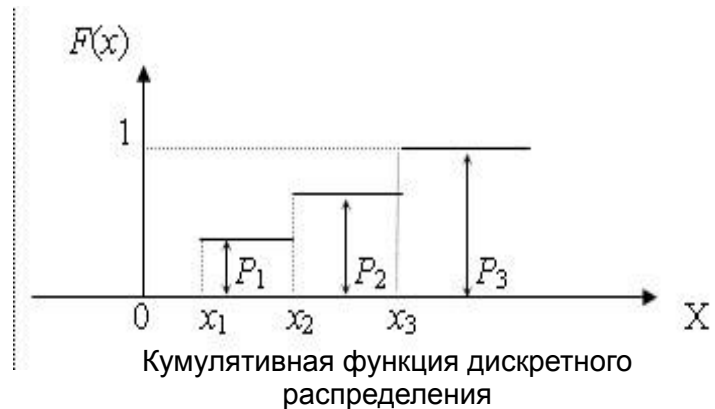
$$P(X \leq x) = F(x)$$

и ее свойства:

- 1)  $0 \leq F(x) \leq 1$ ;
- 2)  $F(-\infty) = 0$ ;
- 3)  $F(+\infty) = 1$ ;
- 4) для  $x_2 > x_1$  всегда  $F_2 > F_1$ ;

$$P(a < X \leq b) = F(b) - F(a)$$

вероятность попадания  $X$  на отрезок  $(a, b)$



# Что было: функция распределения

Дифференциальная функция вероятности:

существует только для непрерывных случайных величин!

$\lim_{\Delta x \rightarrow 0} \Delta F / \Delta x = F'(x) = f(x)$  - плотность вероятности

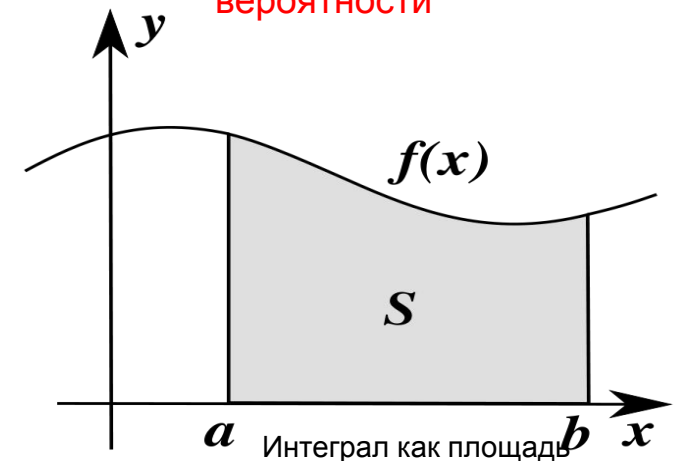
И наоборот:  $\int_{-\infty}^x f(x) dx = F(x)$

Свойства: 1)  $f(x) \geq 0$

2)  $\int f(x) dx = 1$

$$P(a < X \leq b) = \int_a^b f(x) dx = F(b) - F(a) = S$$

вероятность попадания  $X$  на отрезок  $(a, b)$



# Характеристики функции распределения

## Дискретная случайная величина

- Математическое ожидание:

$$M[x] = \sum_{i=1}^n (x_i * p_i)$$

- Дисперсия

$$D[x] = \sum_{i=1}^n ((x_i - M[X])^2 * p_i)$$

- Мода (значение с наибольшей вероятностью)

$$M_o = X_i \mid p(x_i) = p_{\max}$$

- Медиана

$$P(X < Me[X]) = P(X > Me[X]) = \frac{1}{2}, \quad F(Me[X]) = \frac{1}{2}$$

## Непрерывная случайная величина

- Математическое ожидание:

$$M[X] = \int_{-\infty}^{\infty} x * f(x) dx$$

- Дисперсия

$$D[X] = \int_{-\infty}^{\infty} (x_i - M[X])^2 * f(x) dx$$

- Мода (значение с наибольшей плотностью вероятности)

$$M_o = x_i \mid f(x_i) = \max$$

- Медиана

# Знаем:

какие бывают случайные величины;

что такое интегральная (кумулятивная) функция распределения и распределение плотности вероятности;

вероятность попадания  $X$  на отрезок  $(a,b)$ ;

как описать распределение  $F(x)$ .

**Не знаем, какие бывают  $F(x)$**

# Законы распределения случайных величин

# Равномерное распределение №1

**Непрерывная** случайная величина имеет равномерный закон распределения на  $(a, b)$ , если ее **плотность вероятности постоянна** на этом отрезке и равна 0 вне его.

Функция  $P(X < x) = F(x)$  имеет вид

$F(x) = 0$  при  $x \leq a$

$F(x) = (x-a)/(b-a)$  при  $a < x \leq b$

$F(x) = 1$  при  $x > b$

- Математическое ожидание:  $M[x] = (a+b)/2$
- Дисперсия:  $D[x] = (b-a)^2/12$

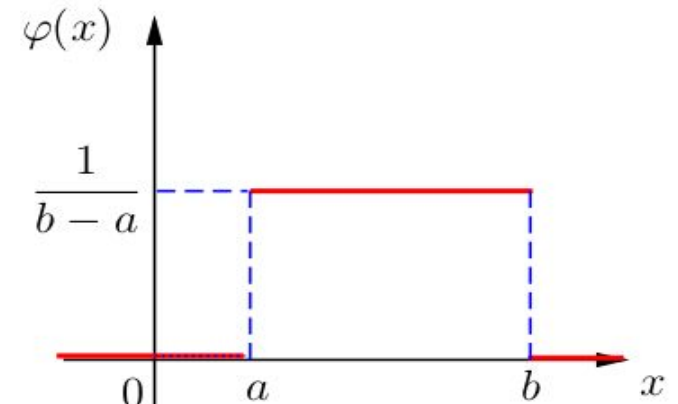


График плотности вероятности

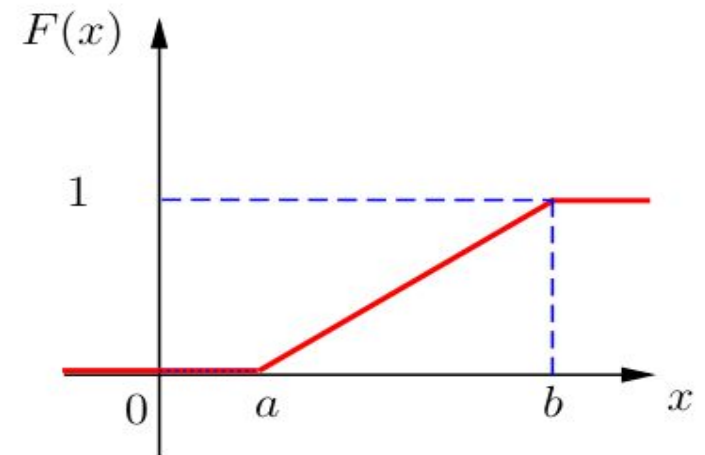


График интегральной функции распределения



# Равномерное распределение №2

Дискретная случайная величина имеет равномерное распределение, если ее **функция вероятности** на всей области определения (a,b) имеет вид

$$P(x) = 1/n,$$

где n — число исходов

- $M[x] = (a+b)/2$  - мат.ожидание
- $D[x] = (n^2-1)/12$  - дисперсия

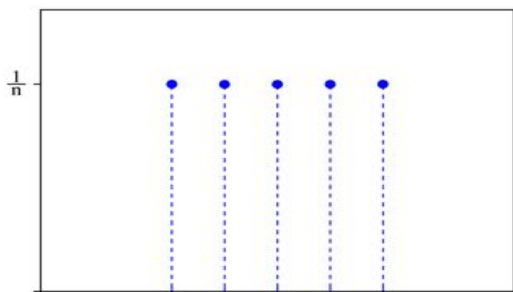


График характеристической функции

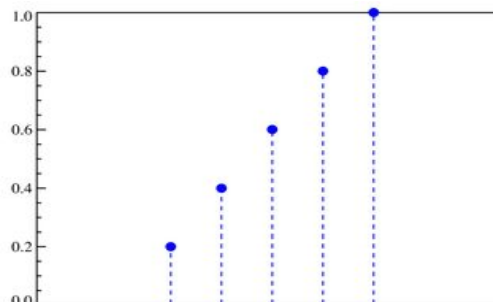
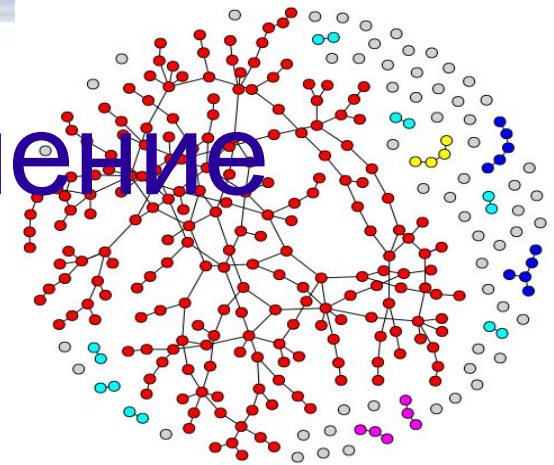


График кумулятивной функции

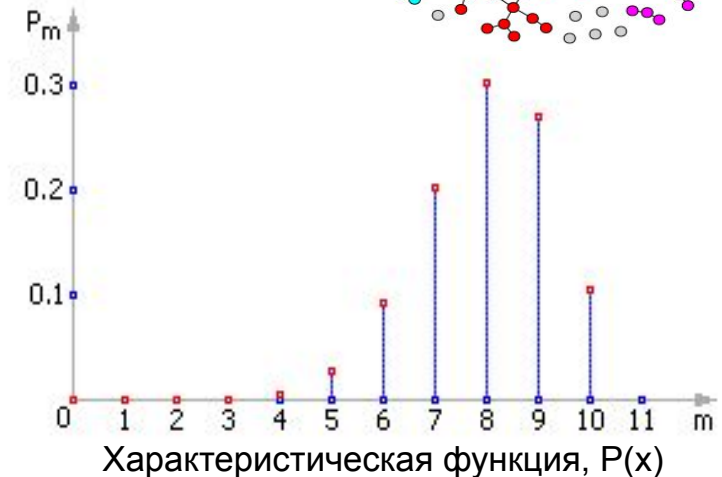


# Биномиальное распределение



Дискретная случайная величина  $X$  распределена по биномиальному закону, если она имеет значения  $\{0 \dots n\}$ , а вероятность  $X=m$

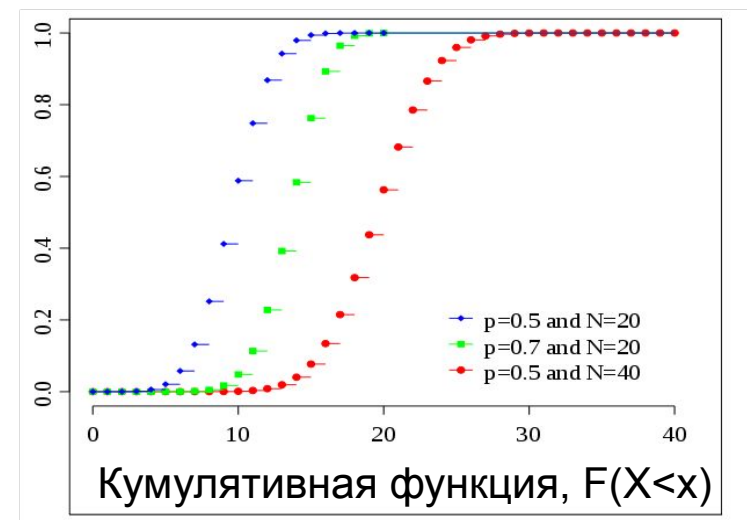
$$P(X=m) = C_n^m * p^m * q^{n-m}$$



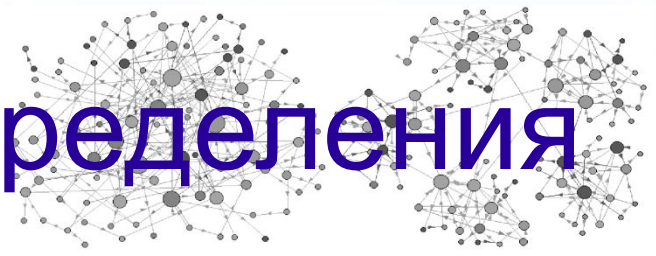
Биномиальное распределение описывает вероятность  $m$  успехов при  $n$  возможных исходов

- $M[X]=n*p$  - мат. ожидание
- $D[X]=n*p*q$  - дисперсия,

где  $p$  - вероятность успеха,  
 $q$  - вероятность неуспеха



# Степенной закон распределения



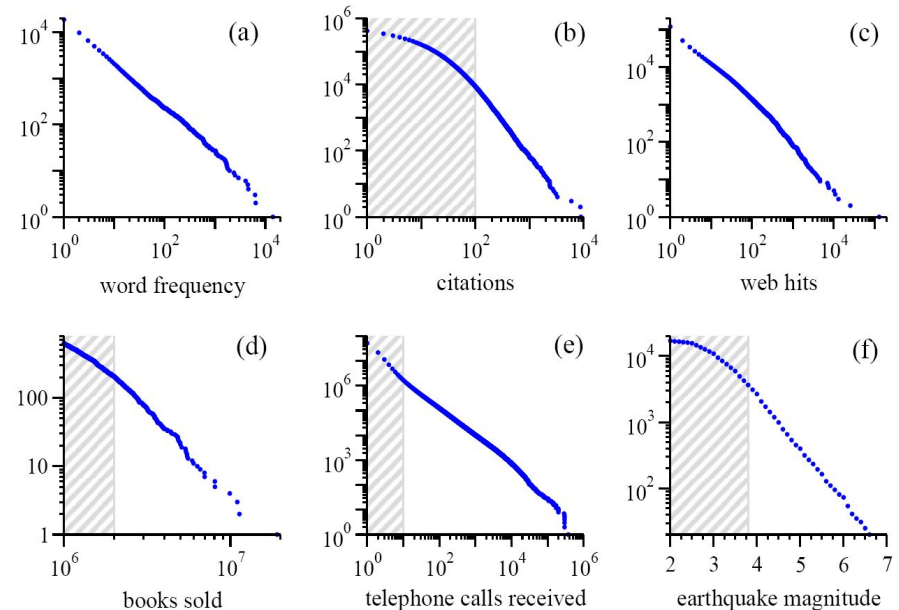
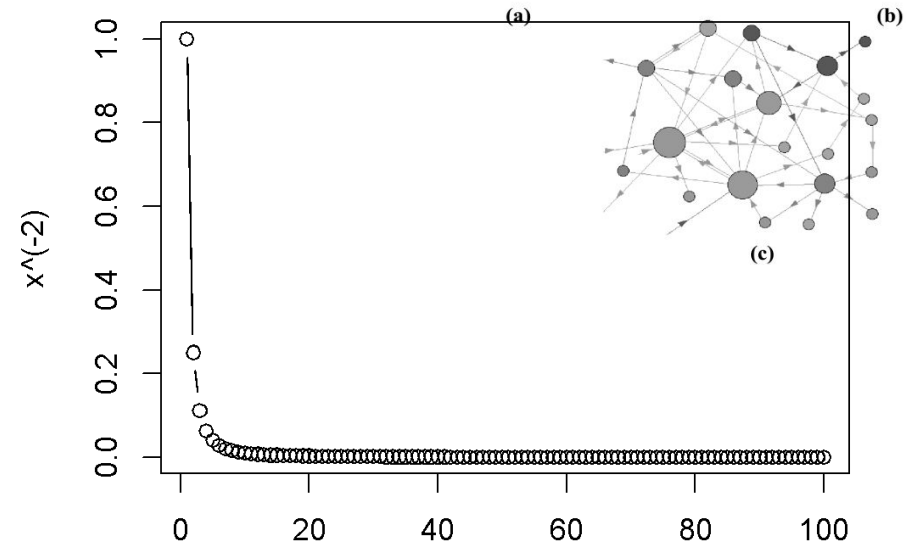
- Случайная величина имеет степенной закон распределения, если ее **плотность вероятности** имеет вид:

$$f(x) = Cx^{-\alpha},$$

при  $\alpha = [2, 3]$

## Свойства:

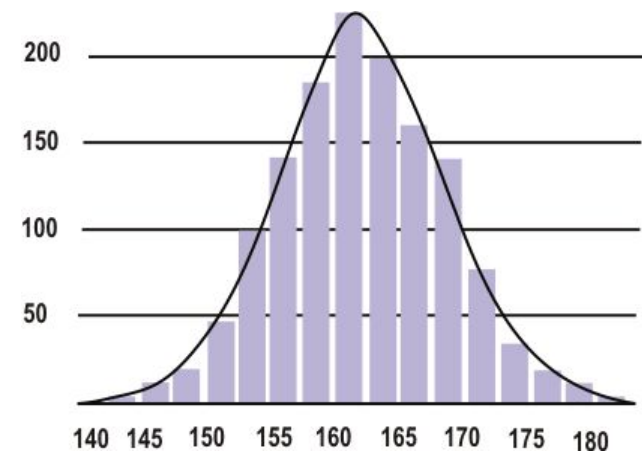
- асимметричное распределение с «тяжелым» хвостом
- прямая линия на log-log шкале;
- Вид графика не зависит от масштаба (scale invariance)



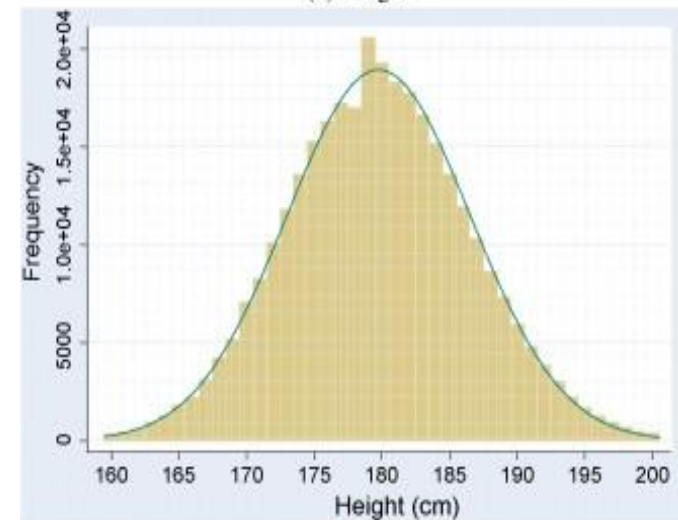
# Нормальное распределение

## Центральная предельная теорема в применении к $\Psi$ :

Если индивидуальная изменчивость некоторого свойства есть следствие действия множества причин, то распределение частот для всего многообразия проявлений этого свойства в генеральной совокупности соответствует кривой нормального распределения



(a) Height



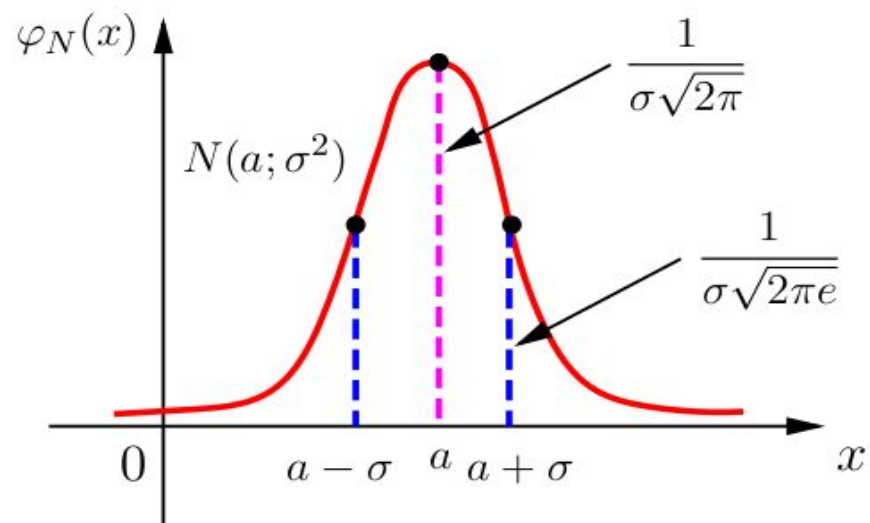
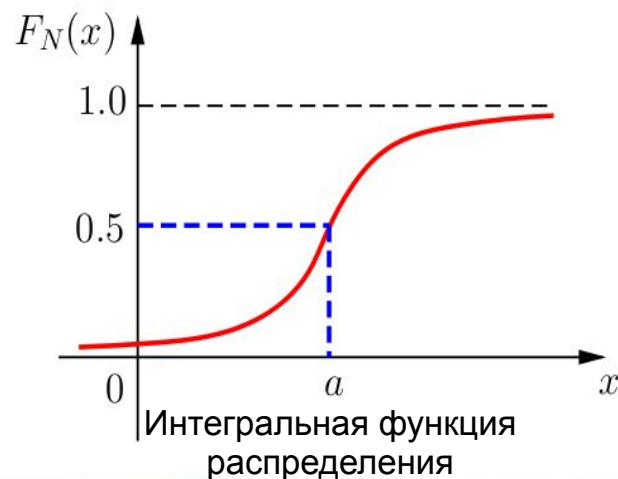
# Закон нормального распределения

Непрерывная случайная величина  $X$  имеет нормальный закон распределения (закон Гаусса) с параметрами  $\alpha$  и  $\beta$ , если ее плотность вероятности имеет вид:

$$f(x) = \frac{1}{\beta\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2\beta^2}}$$

Где:

- $\beta$  — среднеквадратичное отклонение ( $\sigma$ );
- $\alpha$  — среднее ( $M$ );
- $e$ ,  $\pi$  - константы

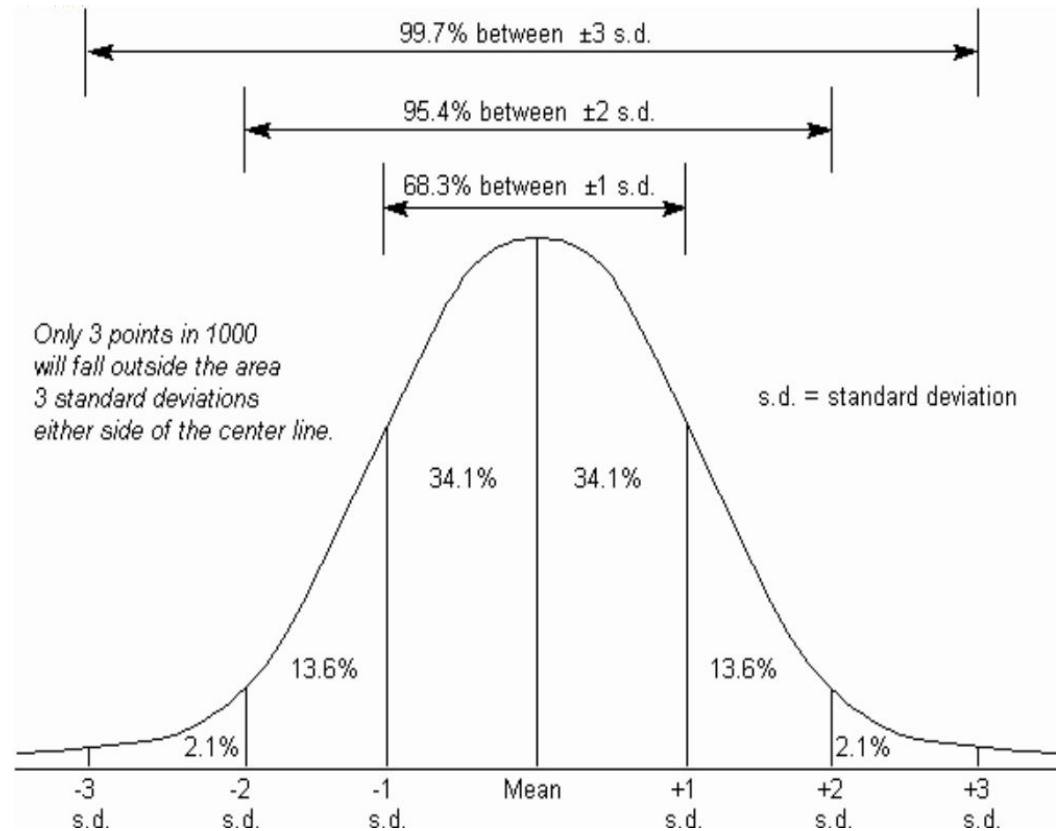


Гауссиана — график нормального распределения

# Правило 3 сигм

При нормальном распределении:

- $M(\pm)\sigma = 68,26\%$
  - $M(\pm)2\sigma = 95,44\%$
  - $M(\pm)3\sigma = 99,72\%$ ,
- $M(\pm)3\sigma$  - интервал всех возможных значений



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2\sigma^2}} dt$$

$$P(c < X < d) = \Phi\left(\frac{d-M}{s}\right) - \Phi\left(\frac{c-M}{s}\right)$$

# Свойства нормального распределения

- Правило 3 сигм (99,72% значений лежат в рамках  $M \pm 3\sigma$ )
- Распределение симметрично ( $A=0$ ), эксцесс, т.е. мера остроты пика или  $E = 0$
- Мода, медиана и среднее совпадают
- Значения, лежащие на равном расстоянии от  $M$  (среднего), будут иметь равную частоту в репрезентативной выборке

# Проверка распределения на «нормальность»

- Графический способ;
- Статистический критерий Колмогорова-Смирнова ( $N > 50$  человек) ;
- $W$ -критерий Шапиро-Уилка ( $N > 8$  человек);
- Критерий асимметрии и эксцесса
- См. ГОСТ Р ИСО 5479—2002



# Критерий асимметрии и эксцесса

1. Определить среднее арифметическое ( $M$ ) и стандартное отклонение ( $\sigma$ ).

2. Рассчитать показатели асимметрии и эксцесса.

$$A = \frac{\sum (X_i - M)^3}{N \cdot \sigma^3} \quad E = \frac{\sum (X_i - M)^4}{N \cdot \sigma^4} - 3$$

3. Рассчитать критические значения  $A$  и  $E$

$$A_{кр} = \sqrt{\frac{6 \cdot (n-1)}{(n+1) \cdot (n+3)}} \quad E_{кр} = \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}}$$

$A$   $E$

4. Если  $A < A_{кр}$  и  $E < E_{кр}$ , распределение нормально

# Закон нормального распределения:

## **следствия**

- Знаем, какой процент испытуемых наберет определенные баллы по тесту;
- Стандартизируем на этой основе баллы по тесту;
- Оцениваем параметры генеральной совокупности по выборочным данным;
- Рассчитываем статистическую значимость наших выводов;
- И задействуем его во всей индуктивной статистике в той или иной степени...