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

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

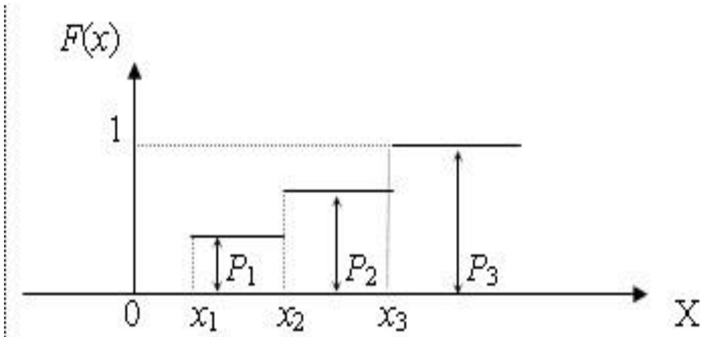
Что было: функция распределения

Интегральная функция распределения

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

и ее свойства:

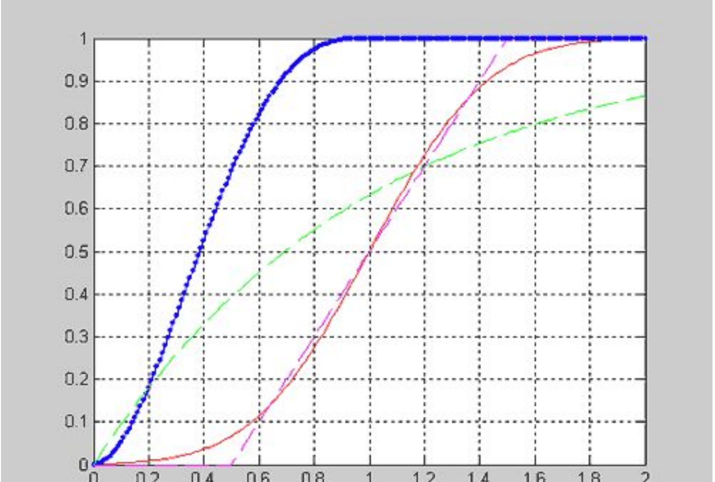
- 1) $0 \le F(x) \le 1$;
- 2) $F(-\infty) = 0$;
- 3) $F(+\infty) = 1$;
- 4) для $x_2 > x_1$ всегда $F_2 > F_1$;



Кумулятивная функция дискретного распределения

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

вероятность попадания X на отрезок (a,b)



Интегральная функция распределения

Что было: функция распределения

Дифференциальная функция вероятности:

существует только для непрерывных случайных величин!

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

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

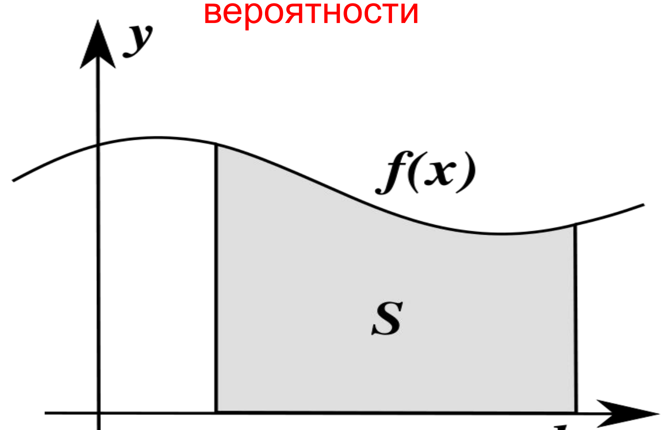
Свойства: 1) $f(x) \geq 0$

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

Функция плотности вероятности

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

вероятность попадания X на отрезок (a,b)



Интеграл как площадь

Характеристики функции распределения

Дискретная случайная величина

- Математическое ожидание:

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

- Дисперсия

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

- Мода (значение с наибольшей вероятностью)

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

- Медиана

Непрерывная случайная величина

Математическое ожидание:

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

Дисперсия

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

Мода (значение с наибольшей плотностью вероятности)

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

Медиана

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

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.

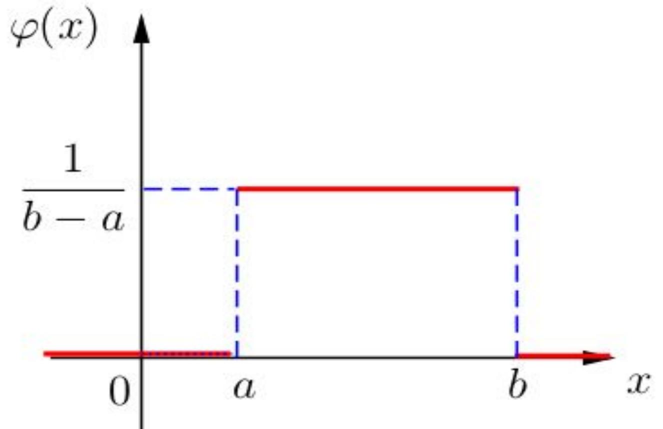


График плотности вероятности

Функция P(X<x)=F(X) имеет вид

F(x)=0 при x≤a

F(x)=(x-a)/(b-a) при a<x≤b

F(x)=1 при x>b

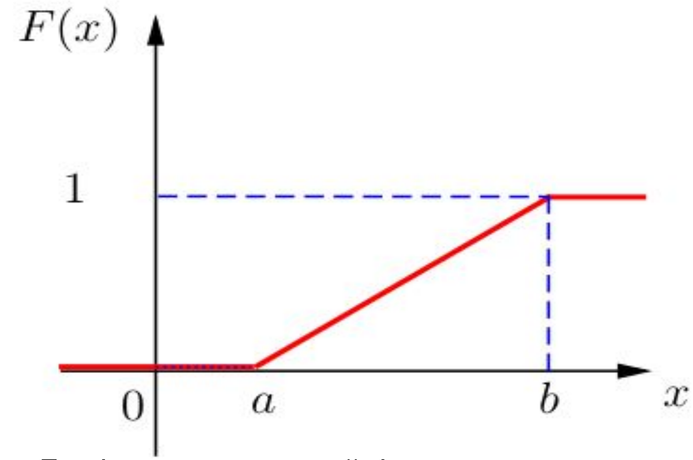


График интегральной функции распределения

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

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

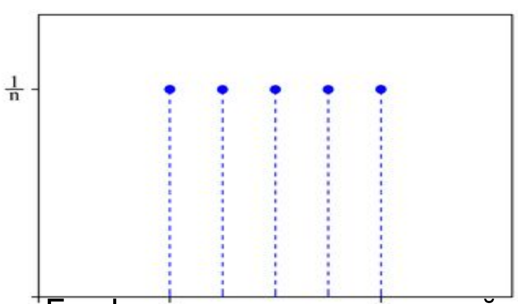


График характеристической функции

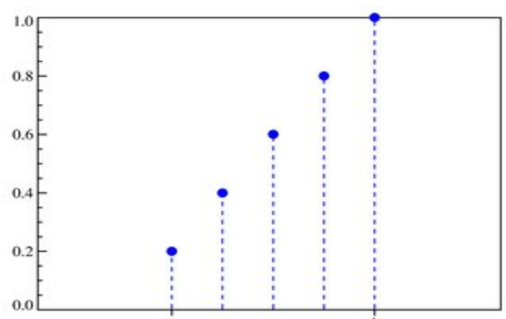
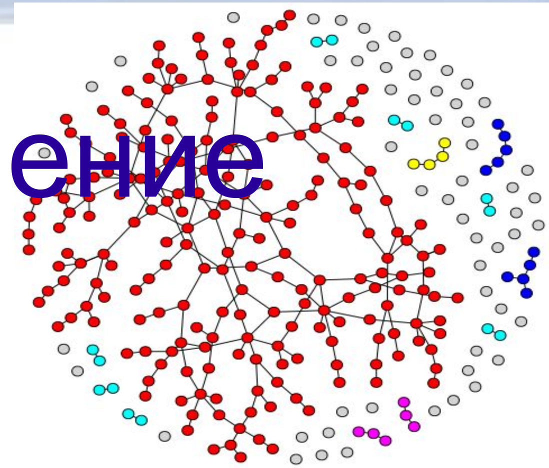


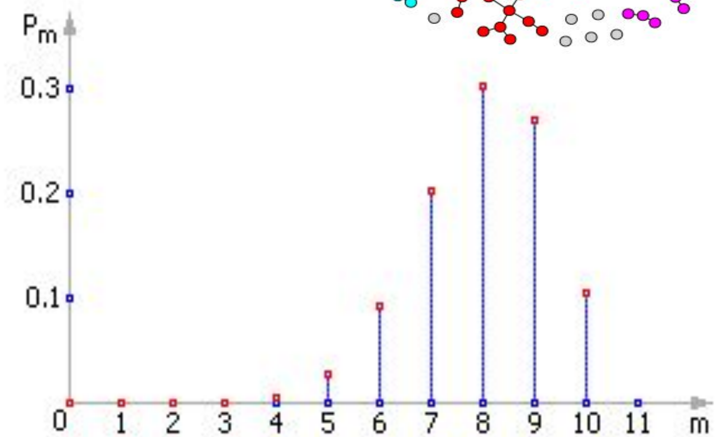
График кумулятивной функции

Биномиальное распределение



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

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

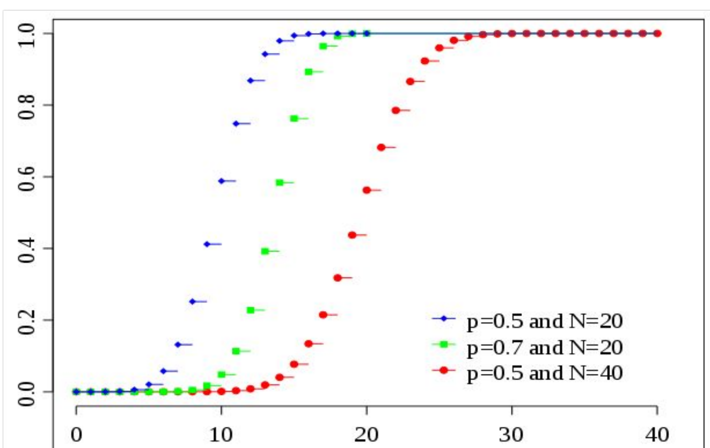


Характеристическая функция, P(x)

Биномиальное распределение описывает вероятность m успехов при n возможных исходов

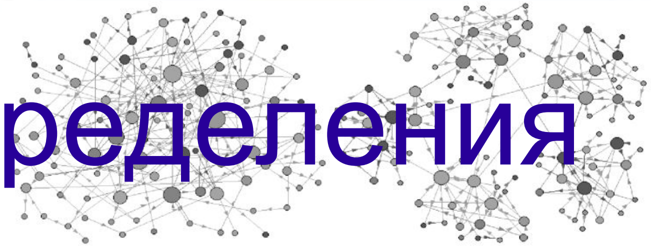
- M[X]=n*p - мат. ожидание
- D[X]=n*p*q - дисперсия,

где p - вероятность успеха,
q - вероятность неуспеха



Кумулятивная функция, F(X<x)

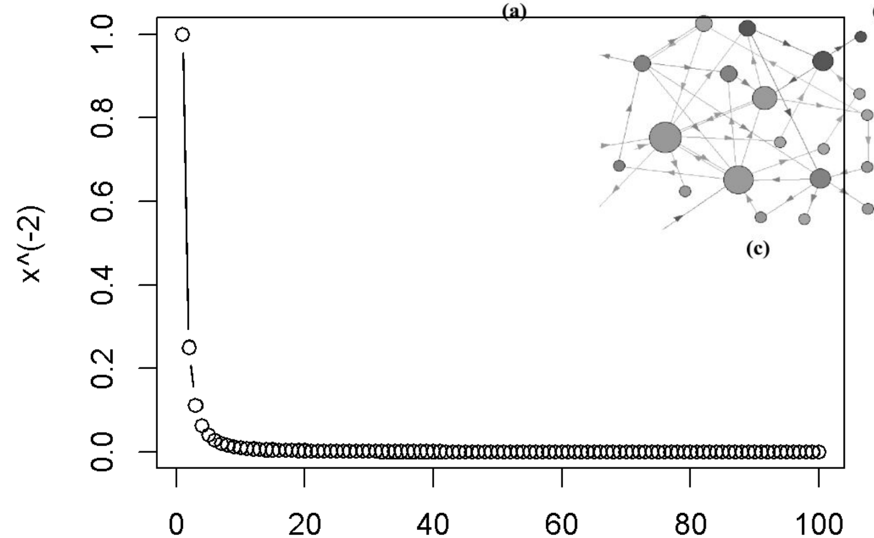
Степенной закон распределения



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

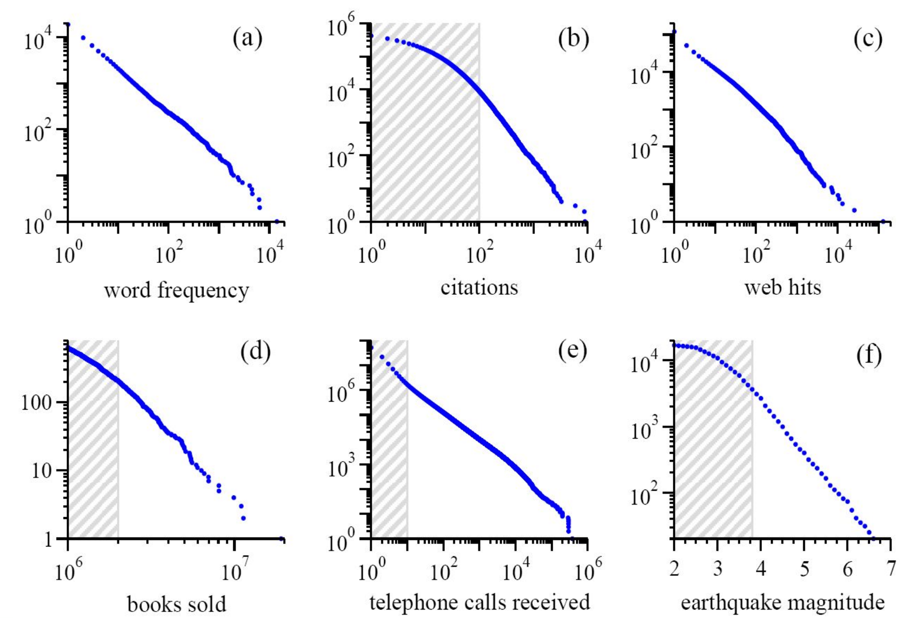
$$f(x)=Cx^{-\alpha} ,$$

при $\alpha=[2,3]$



Свойства:

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



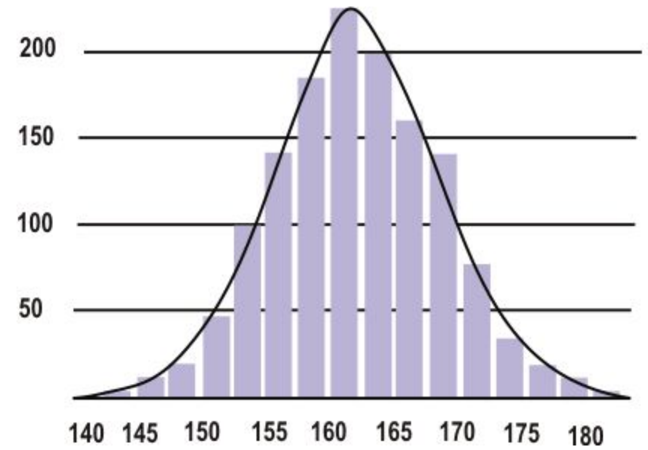
Принцип Парето: 80/20

M.E.J. Newman. Power laws, Pareto distribution and Zipf's law/ arXiv:cond-mat/0412004

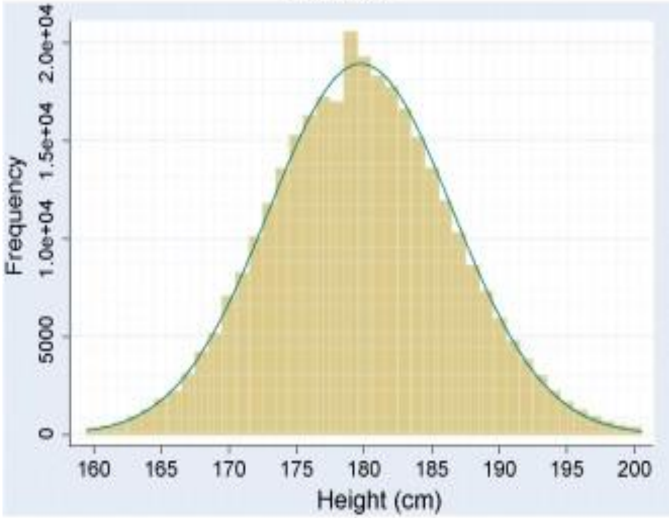
Нормальное распределение

Центральная предельная теорема в применении к Ψ:

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



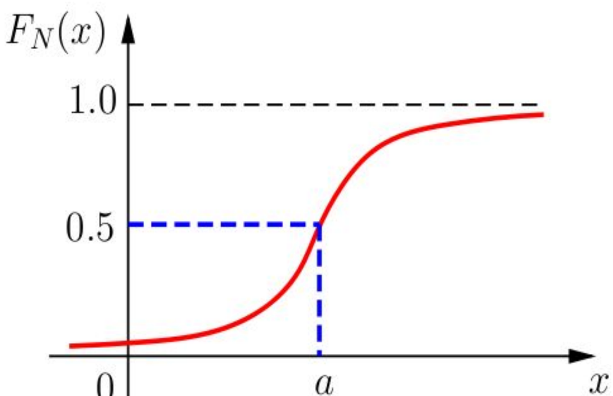
Закон нормального распределения

Непрерывная случайная величина X имеет нормальный закон распределения (закон Гаусса) с параметрами α и β, если ее плотность вероятности имеет вид:

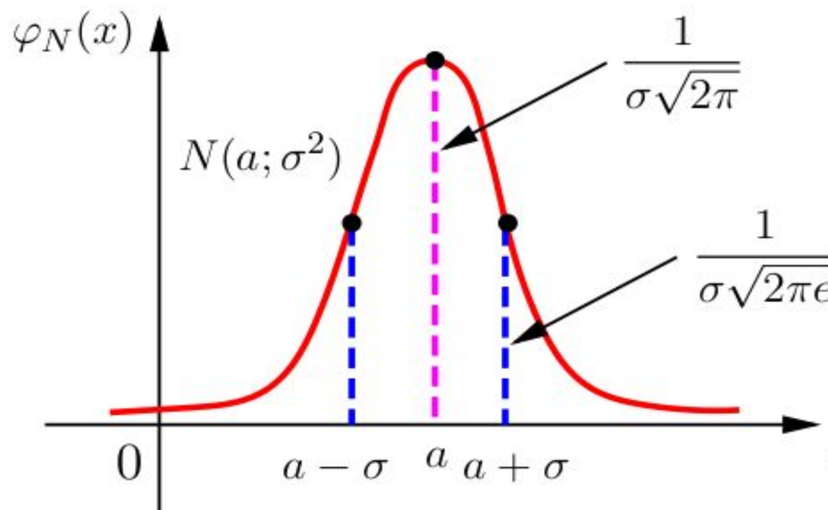
$$f(x) = \frac{1}{\beta\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\beta^2}}$$

Где:

- β — среднеквадратичное отклонение (σ);
- α — среднее (M);
- e, π - константы



Интегральная функция распределения



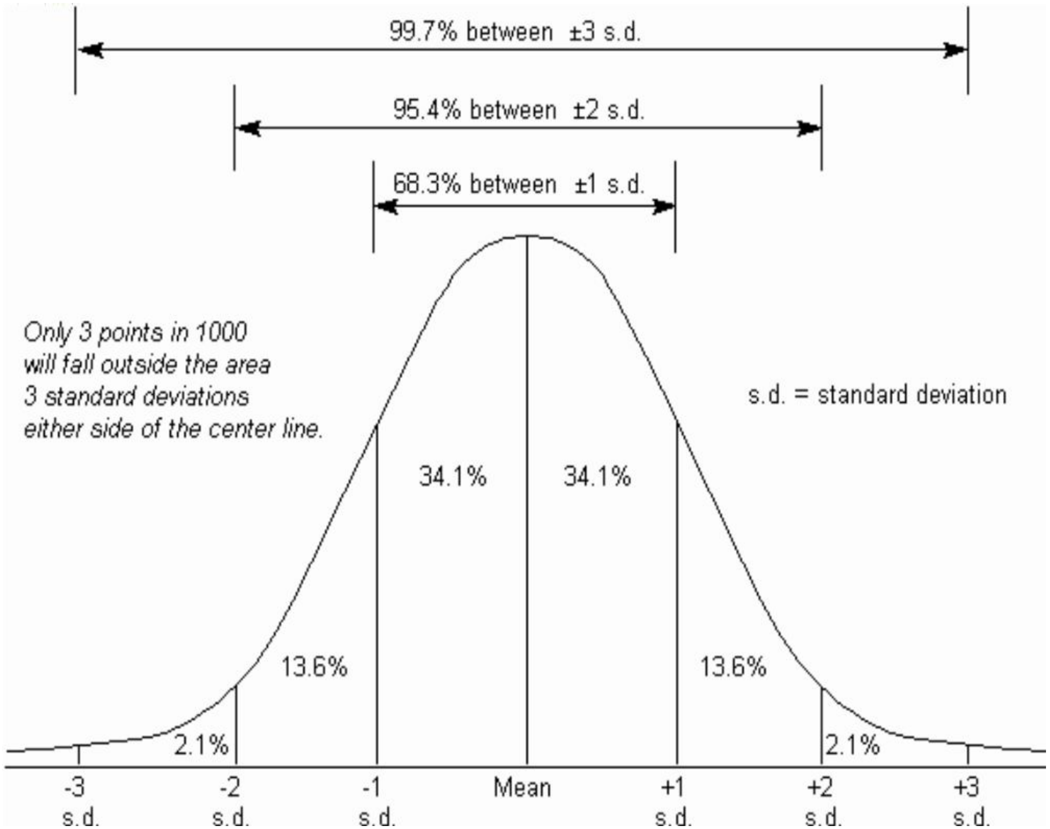
Гауссиана — график нормального распределения

Правило 3 сигм

При нормальном распределении:

- M(+/-)σ=68,26%
- M(+/-)2σ=95,44%
- M(+/-)3σ=99,72%,

M(+/-)3σ - интервал всех возможных значений



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

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

Свойства нормального распределения

- Правило 3 сигм (99,72% значений лежат в рамках $M+/-3\sigma$)
- 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 ассиметрии и эксцесса
- См. ГОСТ Р ИСО 5479—2002

Критерий асимметрии и эксцесса

1. Определить среднее арифметическое (М) и стандартное отклонение (σ).

2. Рассчитать показатели асимметрии и эксцесса.

- A= $\dfrac{\sum\left(X_i - M\right)^3}{N \cdot \sigma^3}$ E= $\dfrac{\sum\left(X_i - M^4\right)}{N \cdot \sigma^4}$ -3

3. Рассчитать критические значения A и E

$$A_{кр} = 3 \cdot \sqrt{\frac{6 \cdot (n-1)}{(n+1)\cdot(n+3)}}$$

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

4. Если A<Aкр и E<Eкр, 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...