Базовые понятия теории вероятностей

Событие – это любой исход какого – либо вероятностного эксперимента.

Вероятность события - это отношение числа исходов, благоприятствующих появлению данного события, к общему числу исходов, данного вероятностного эксперимента

$$P(A)=\frac{m}{n}$$

$$0 \le P(A) \le 1$$

Случайная величина – это величина, которая может принимать то или иное значение, из некоторого множества значений.

Соответствие между всевозможными значениями СВ и их вероятностями называется законом распределения СВ. Аналитически закон распределения СВ задается либо функцией распределения, либо плотностью вероятностей.

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

$$f(x)=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x)$$

Числовые характеристики СВ

Математическое ожидание. Характеризует среднее ожидаемое значение СВ.

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

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

Дисперсия. Оценивает разброс возможных значений СВ относительно ее среднего значения (математического ожидания) .

$$D(X)=M\left(\left(X-M(X)\right)^2\right)=M\left(X^2\right)-M^2(X)$$

Среднее квадратическое отклонение.

$$\sigma_x=\sqrt{D(X)}$$

Коэффициент вариации. Для оценки разброса значений СВ в процентах относительно ее среднего значения.

$$V(X)=\frac{\sigma_x}{|M(X)|}100\ \%$$

Закон распределения непрерывной случайной переменной

В случае, когда X непрерывная случайная переменная, ее закон распределения вероятностей выражается с помощью функции плотности вероятностей, который по определению есть:

$$f(t) = \lim \frac{P(t \le x \le t + \Delta t)}{\Delta t} \quad \text{при} \quad \Delta t \Rightarrow 0$$

где: $P(t \le x \le t+\Delta t)$ – вероятность того, что случайная переменная X примет в опыте значение, лежащее в интервале $(t, t+\Delta t)$

Свойства функции плотности вероятностей

1. Функция плотности вероятности неотрицательна $p_x(t) \geq 0$

2. Справедливо равенство:

$$\int_{-\infty}^{\infty} f(t)dt = 1$$

3. Вероятность попадания СВ x на отрезок [a, b] есть:

$$P_x(a \leq x \leq b) = \int_a^b f(t)dt$$

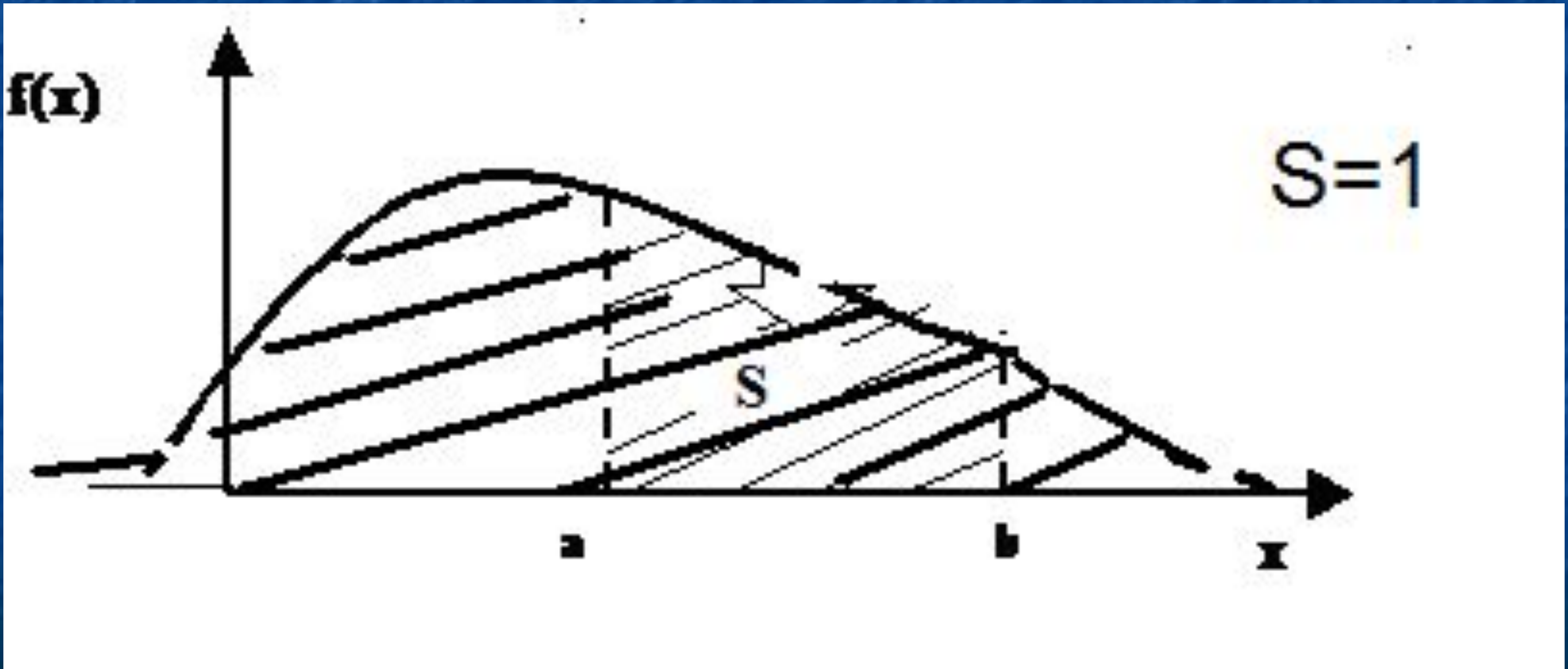
4. Функция распределения вероятностей связана с функцией плотности вероятностей выражением:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Свойства функции плотности вероятностей

2. Справедливо равенство:

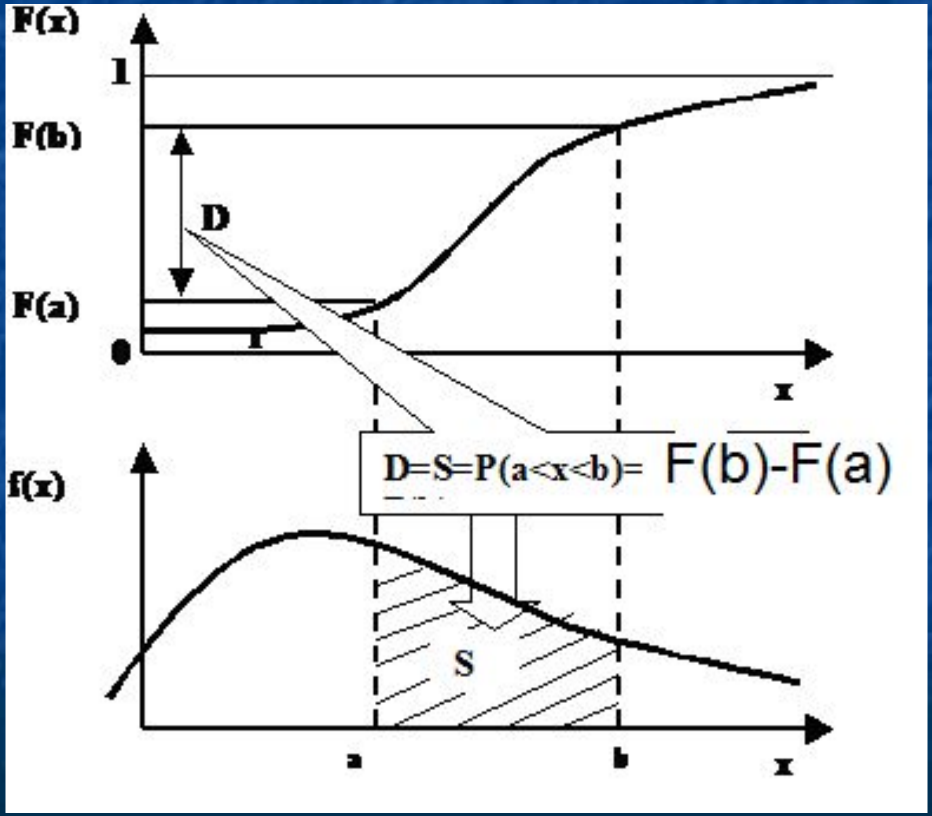
$$\int_{-\infty}^{\infty} f(t)dt = 1$$



Свойства функции плотности вероятностей

Вероятность попадания СВ x на отрезок [a, b] есть:

$$P_x(a \le x \le b) = \int_a^b f(t)dt$$



Квантили

Квантиль **порядка *P* одномерного распределения есть такое значение случайной величины *xP*, для которого выполняется равенство**

$$P(X < x_P) = F\left(x_P\right) = P, \qquad \text{где} \quad 0 < P < 1$$

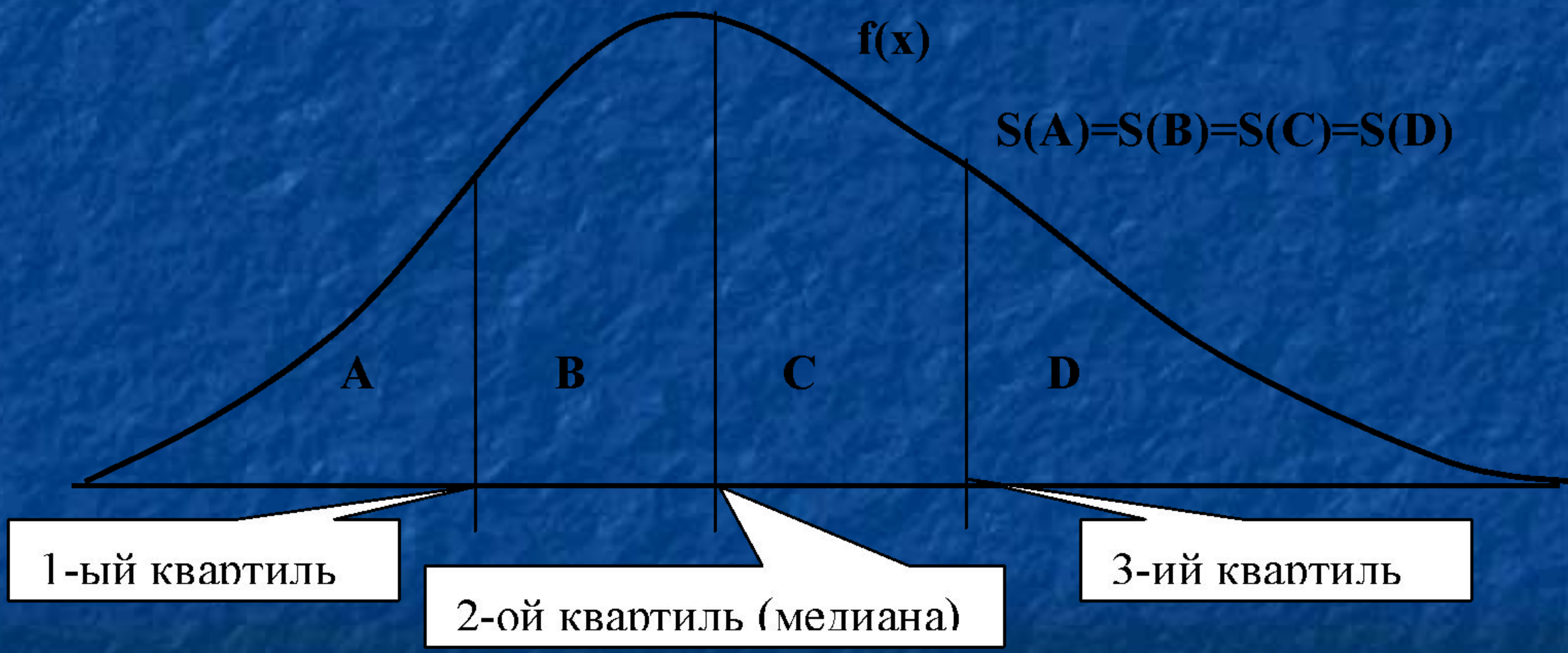
$$\int_{-\infty}^{x_p} f(t)dt = P$$

точка *xP* делит площадь подграфика функции плотности распределения на две части таким образом, что площадь левой части равна *P*

x1/2 - медиана *квартилями* (quartile), децили (decile),

2.5 и 97.5-ые центили, а так же 5-й и 95-ый центили. Первая пара широко используется при построении 95% доверительного интервала, а вторая - для проверки статистических гипотез при уровне значимости, равном 5%.

Квартили для непрерывного распределения



Практически достоверное событие

Определение. Событие V, связанное с некоторым опытом, называется «практически достоверным», если вероятность его появления удовлетворяет условию: $0.95 \le P(V) \le 1$

Любое случайное событие W, связанное с опытом, вероятность которого $0 < P(W) \le 0.05$, называется «практически невозможным».

Установлено, что практически достоверное событие, *как правило*, появляется при первом проведении опыта.

Если этого не происходит, значит нарушены условия опыта.

Числовые характеристики СВ

Математическое ожидание. Характеризует среднее ожидаемое значение СВ.

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

Дисперсия. Оценивает разброс возможных значений СВ относительно ее среднего значения (математического ожидания) .

$$D(X) = M\left((X - M(X))^2\right) = M\left(X^2\right) - M^2(X)$$

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

Среднее квадратическое отклонение.

$$\sigma_x = \sqrt{D(X)}$$

Коэффициент вариации. Для оценки разброса значений СВ в процентах относительно ее среднего значения.

$$V(X) = \frac{\sigma_x}{|M(X)|} 100\ \%$$

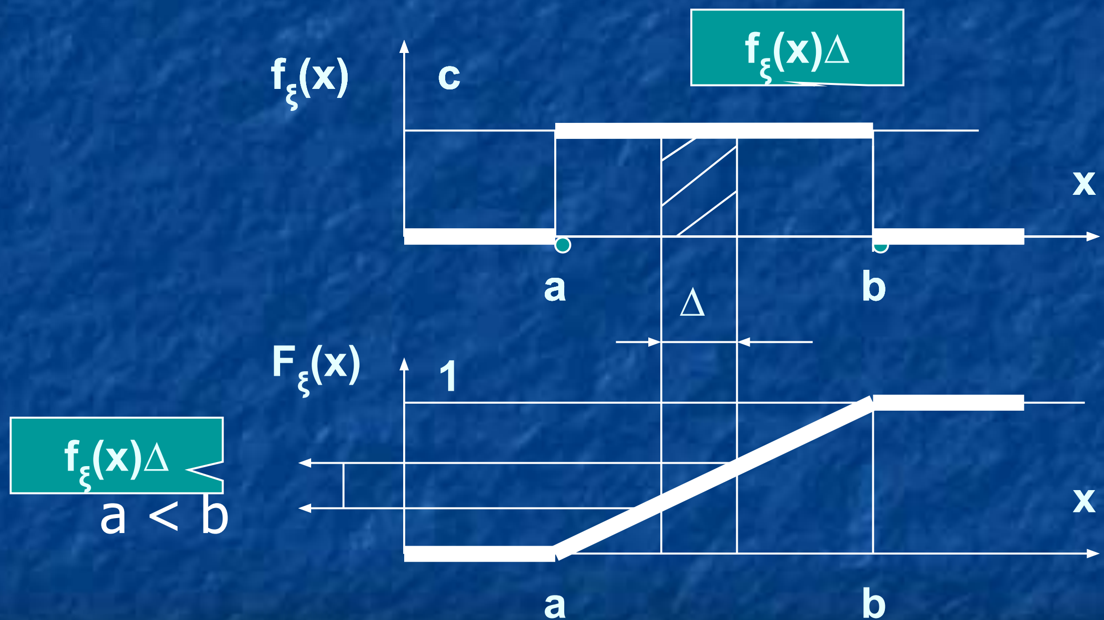
$$F_\xi(x) = P(\text{точка попала в область } D_x) = \frac{\text{площадь } D_x}{\text{площадь } \Omega} = \int_{-\infty}^{x} f_\xi(t)\,dt$$

Пример задания НСВ

<u>Определение.</u> Распределение называется равномерным если имеет функцию плотности распределения вида:

$$\begin{cases} f_\xi(x) = c, & \text{если } a \le x \le b \\ f_\xi(x) = 0, & \text{если } a > x > b \end{cases} \qquad c = \frac{1}{b-a}$$

$$\int_{-\infty}^{\infty} f_\xi(x)dx = \int_a^b f_\xi(x)dx = \int_a^b c\,d(x) = cx\Big|_a^b = c(b-a) = 1$$



$$F_\xi(x) = \begin{cases} 0, & x \notin [a,b] \\ \dfrac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$$

$$P_\xi(x) = P(\text{точка попала в область } D_x) = \frac{\text{площадь } D_x}{\text{площадь } \Omega} = \int_{-\infty}^{x} f_\xi(t)\,dt$$

Законы распределения СВ

Нормальное распределение

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

$$m=M(X),\ \sigma=\sigma_x,\ D(X)=\sigma^2 \qquad X\sim N(m,\sigma)$$

Распределение χ^2 (хи – квадрат)

$$x_i\sim N(m_i,\sigma_i)$$

$$U_i=\frac{(x_i-m_i)}{\sigma_i},\ (i=1,2,\ldots n) \qquad \chi^2=\sum_{i=0}^{n}U_i^2=U_1^2+U_2^2+\ldots+U_n^2$$

$$U_i\sim N(0,1) \qquad M(\chi^2)=n \qquad D(\chi^2)=2n$$

Законы распределения СВ

Распределение Стьюдента

$$U_i \sim N(0,1)$$

$$V \sim \chi^2(n)$$

$$T = \frac{U}{\sqrt{V/n}}$$

$$M(T) = 0$$

$$D(T) = \frac{n}{n-2}$$

Распределение Фишера

$$V \sim \chi^2(n)$$

$$W \sim \chi^2(m)$$

$$F = \frac{V/m}{W/n} = \frac{\frac{1}{m}\chi^2(m)}{\frac{1}{n}\chi^2(n)}$$

$$M(F) = \frac{n}{n-2}, \qquad (n > 2)$$

$$D(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \qquad (n > 4)$$

Многомерные случайные величины
Условные законы распределения

Упорядоченный набор СВ $X=\left(X_1,X_2,...,X_n\right)$ называется многомерной СВ

$$F\left(x_1,x_2,...,x_n\right)=P\left(X_1<x_1,X_2<x_2,...,X_n<x_n\right)$$

В двумерном случае для СВ $\left(X,Y\right)$ $F\left(x,y\right)=P\left(X<x,Y<y\right)$

Условным законом распределения одной из одномерных составляющих двумерной случайной величины , называется ее закон распределения, вычисленный при условии, что другая составляющая приняла определенное значение (или попала в какой-то интервал).

Условное математическое ожидание случайной величины Y при $X=x$, т.е. $M_x\left(Y\right)$, есть функция от x, называемая функцией регрессии или просто регрессией Y по X. График этой функций называются линией регрессии.

Ковариация двух случайных величин – мера линейной зависимости двух СВ

$$Cov\left(X,Y\right)=M\left[\left(X-M\left(X\right)\right)\left(Y-M\left(Y\right)\right)\right]$$

Коэффициент корреляции двух случайных величин

$$\rho=\frac{Cov\left(X,Y\right)}{\sigma_x\sigma_y}$$

Выборочные характеристики

Генеральная совокупность N	Выборка n
$\bar{x}_R = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\bar{x}_b = \frac{1}{n}\sum_{i=1}^{n} x_i$
$D_r = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_r)^2$	$D_b = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_b)^2$
$\sigma_r = \sqrt{D_r}$	$\sigma_b = \sqrt{D_b} = \sqrt{\overline{x^2} - (\bar{x})^2}$
	$V = \frac{\sigma_b}{x_b} \cdot 100\%$

Статистические выводы: оценки и проверка гипотез

Статистические выводы - это заключения о генеральной совокупности на основе выборки, случайно отобранной из генеральной совокупности

Процесс нахождения оценок параметров генеральной совокупности по определенному правилу называется оцениванием

- оценивание ***вида*** распределения

- оценивание ***параметров*** распределения

Точечной оценкой Θ^* параметра Θ называется числовое значение этого параметра, полученное по выборке объема n

Качество оценок характеризуется следующими основными свойствами:

Оценка ***несмещенная***, если математическое ожидание оценки равно оцениваемому параметру

Оценка ***эффективна***, если ее дисперсия меньше дисперсии оценки, полученной по любой другой выборке такого же объема

Оценка ***состоятельна***, если она дает истинное значение параметра при достаточно большом объеме выборки вне зависимости от значений входящих в нее конкретных наблюдений.

Статистическая проверка гипотез

Статистической называют гипотезу о виде закона распределения или о параметрах известного распределения.

Гипотеза $\boldsymbol{H_0}$, подлежащая проверке, называется нулевой (основной). Наряду с ней рассматривают гипотезу $\boldsymbol{H_1}$(альтернативную) которая будет приниматься, если отклоняется $\boldsymbol{H_0}$.

Сущность проверки статистической гипотезы заключается в том, чтобы установить, согласуются или нет данные наблюдений и выдвинутая гипотеза.

*Ошибка **первого рода** состоит в том, что будет отвернута правильная нулевая гипотеза.*

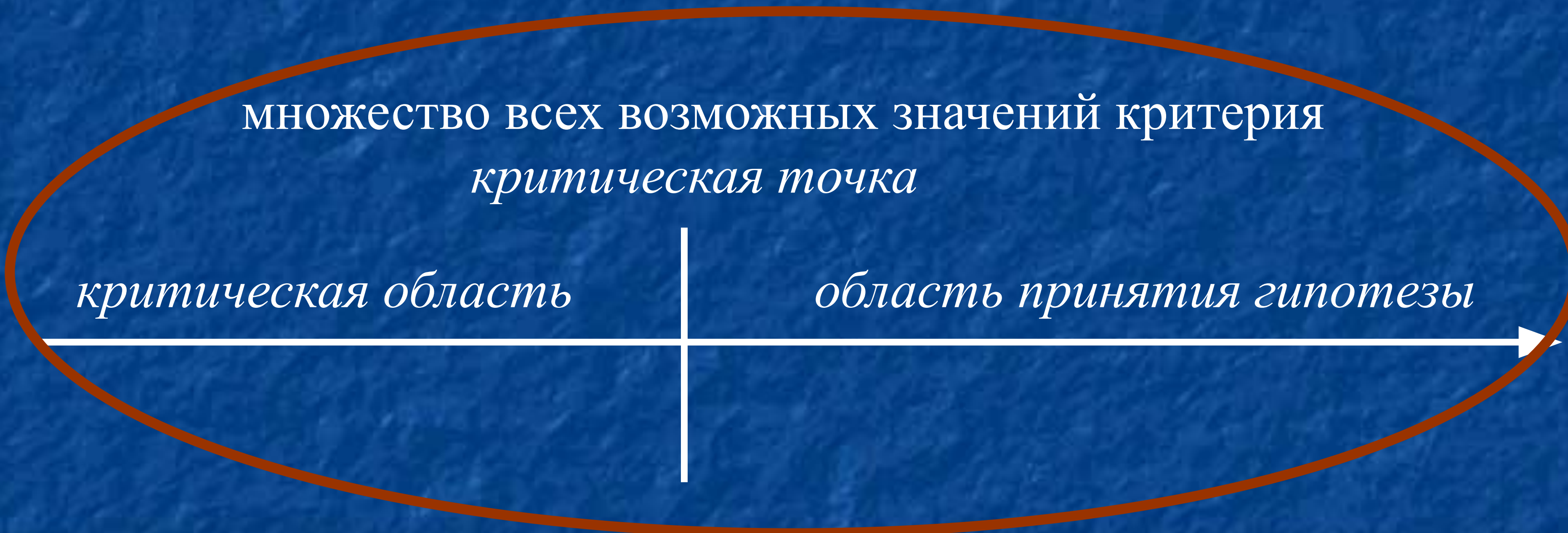
$\boldsymbol{\alpha}$ *- вероятность совершить ошибку первого рода или уровень значимости*

*Ошибка **второго рода** состоит в том, что будет принята нулевая гипотеза, в то время как в действительности верна альтернативная гипотеза.*

$\boldsymbol{\beta}$ *- вероятность совершить ошибку второго рода*

Статистическая проверка гипотез

Статистическим критерием (статистикой) называют специально подобранную **СВ**, закон распределения которой которой известен и которая служит для проверки нулевой гипотезы



U (или **Z**) –стандартизированное нормальное распределение;
T - по закону Стьюдента;
$\boldsymbol{\chi^2}$ - по закону χ^2 ;
F –распределение Фишера.