2. Элементы теории вероятностей и математической статистики

2.1. Случайные величины и их числовые характеристики

Под случайной величиной понимается переменная, которая в результате испытаний в зависимости от случая может принимать любое значение из множества своих возможных значений.

Случайная величина называется дискретной, если множество ее возможных значений дискретно, или непрерывной, если это множество непрерывно.

Пример дискретной случайной величины: число выстрелов до первого попадания в цель.

Пример непрерывной случайной величины: - дальность полета артиллерийского снаряда.

Наиболее полным и исчерпывающим описанием случайной величины является ее закон распределения.

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и вероятностями их реализации.

Для дискретной случайной величины закон распределения может быть задан таблицей, аналитически (в виде формулы) и графически.

При табличном задании закона распределения дискретной случайной величины первая строка таблицы содержит возможные значения, а вторая - их вероятности.

X	X_1	X_2	…	X_n
P	P_1	P_2	…	P_n

Такая таблица называется *рядом распределения* дискретной случайной величины.

Сумма вероятностей во второй строке таблицы всегда равна единице, поскольку в результате испытания какое-то из возможных значений случайной величины обязательно реализуется.

Пример. В денежной лотерее выпущено 100 билетов. Разыгрывается один выигрыш 500 руб. и 10 выигрышей по 100 руб. Найти закон распределения случайной величины X – стоимости выигрыша для владельца одного лотерейного билета.

Решение. Запишем возможные значения X: $X_1 = 500$; $X_2 = 100$; $X_3 = 0$. Вероятности этих возможных значений таковы: $P_1 = 1/100$; $P_2 = 10/100$; $P_3 = 89/100$. Таким образом, получаем следующий закон распределения

X	500	100	0
P	0,01	0,1	0,89

2.2 Математические операции над случайными величинами

Определим понятие независимости случайных величин.

Две случайных величины называются независимыми, если закон распределения одной из них не изменяется от того, какие возможные значения приняла другая величина.

Например. Если имеются различные лотереи, то случайные величины X и Y, выражающие суммы выигрыша по билету разных лотерей, являются независимыми величинами. Если же за X и Y взять выигрыши двух различных билетов в одной лотерее, то эти величины окажутся зависимыми, так как при выигрыше одного билета вероятность выигрыша других билетов уменьшается.

Математические операции над случайными величинами. Рассмотрим случайную величину, имеющую распределение

X_i	X_1	X_2	…	X_n
P_i	P_1	P_2	…	P_n

1. Произведением kX случайной величины X на постоянную величину k называется случайная величина kX_i, принимающая свои значения с теми же вероятностями P_i.

2. m-ной степенью случайной величины X называется случайная величина X_i^m, принимающая свои значения с теми же вероятностями P_i.

Пример. Дана случайная величина X

X_I	-2	1	2
P_i	0,5	0,3	0,2

Найти закон распределения случайной величины $Y=X^2$

Решение. Величина Y принимает значения 1 и 4. Значение 1 она принимает с вероятностью 0,3, а значение 4 с вероятностью 0,2+0,5=0,7. Поэтому распределение величины Y имеет вид:

Y_i	1	4
P_i	0,3	0,7

4

3. Суммой (разностью или произведением) случайных величин X и Y *называется случайная величина* X_i+Y_i (X_i-Y_i, $X_i \cdot Y_i$) с вероятностями P_{ij} того, что величина X_i имеет вероятность P_i, а величина Y_j вероятность P_j. Если случайные величины независимы, то $P_{ij} = P_i \cdot P_j$.

Пример. Даны законы распределения двух независимых случайных величин

X_i	0	2
P_i	0,3	0,7
Y_i	3	1
P_i	0,4	0,6

Найти закон распределения случайной величины $Z=X+Y$.

Решение . Величина Z может принимать следующие значения: 1 с вероятностью P=0,18; 3 с вероятностью P =0,12+0,42; 5 с вероятностью P=0,28. Поэтому закон распределения будет иметь вид:

Z_I	1	3	5
P_i	0,18	0,54	0,28

Легко убедиться, что сумма вероятностей действительно равна 1

2.3. Математическое ожидание дискретной случайной величины

Математическим ожиданием , или средним значением, *M(X)* дискретной случайной величины *X* называется сумма произведений всех ее значений на соответствующие вероятности:

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

Пример. Дана случайная величина X

X_I	-2	1	2
P_i	0,5	0,3	0,2

Требуется найти ее среднее значение.

Решение. Вычислим среднее значение в соответствии с приведенным выше определением

$$M(X) = -2 \cdot 0{,}5 + 1 \cdot 0{,}3 + 2 \cdot 0{,}2 = -0{,}3.$$

Свойства математического ожидания.

1. Математическое ожидание постоянной величины равно самой постоянной. $M(C)=C$.

2. Постоянный множитель можно выносить за знак математического ожидания: $M(kX)=kM(X)$.

3. Математическое ожидание суммы конечного числа случайных величин равно сумме их математических ожиданий, т.е.

$$M(X \pm Y) = M(X) \pm M(Y).$$

Доказательство этого утверждения получить самостоятельно.

4. Математическое ожидание произведения конечного числа независимых случайных величин равно произведению их математических ожиданий.

$$M(X\cdot Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_i p_j=$$

$$\sum_{i=1}^{n}x_i p_i\cdot\sum_{j=1}^{m}y_j p_j=M(X)\cdot M(Y).$$

5. Математическое ожидание отклонения случайной величины от ее математического ожидания равно нулю:

$$M(X-M(X))=M(X)-M(X)=0.$$

6. Если все значения случайной величины увеличить или уменьшить на некоторое постоянное число С, то на эту же постоянную величину С изменится и математическое ожидание.

$$M(X \pm C)=M(X) \pm M(C)=M(X) \pm C.$$

Доказательство этого положения следует из п. 3.

2.4. Дисперсия дискретной случайной величины.

В дальнейшем будем использовать для математического ожидания случайной величины X обозначение

$$M(X) \equiv \overline{X}.$$

Определение Дисперсией *D(X)* случайной величины *X* называется математическое ожидание квадрата ее отклонения от математического ожидания:

$$D(X) = \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot p_i.$$

Наряду с дисперсией вводят величину среднего квадратического отклонения

$$\sigma = \sqrt{D(X)},$$

которая характеризует степень рассеяния индивидуальных значений случайной величины от среднего значения.

Для дисперсии наряду с введенным выше обозначением *D(X)* используется и обозначение

$$\sigma^2 \equiv D(X),$$

которое мы будем широко применять в дальнейшем

Свойства дисперсии:

1. Дисперсия постоянной величины равна нулю.
2. Изменение всех значений признака на одну и ту же величину не изменяет величину дисперсии.
3. Уменьшение или увеличение всех значений признака в k раз приводит к уменьшению или увеличению дисперсии в K^2 раз.
4. Дисперсия алгебраической суммы или разности конечного числа независимых случайных величин равна сумме их дисперсий .

Доказательство утверждений 1– 4 необходимо произвести самостоятельно.

5. Дисперсия относительно любой величины A связана с дисперсией относительно среднего значения следующим соотношением

$$\sigma^2{}_A = \sum_{i=1}^{n}(x_i - A)^2 \cdot p_i =$$

$$= \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - A)^2 \cdot p_i =$$

$$\sum_{i=1}^{n}\left[(x_i - \bar{x})^2 + (x_i - \bar{x})(\bar{x} - A) + (\bar{x} - A)^2\right] \cdot p_i =$$

$$\sigma^2 + (\bar{x} - A)^2.$$

6. Дисперсия случайной величины равна разности между математическим ожиданием квадрата случайной величины и квадратом ее математического ожидания. Действительно

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot p_i =$$

$$= \sum_{i=1}^{n} (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \cdot p_i =$$

$$= \sum_{i=1}^{n} x_i^2 \cdot p_i - 2\bar{x} \sum_{i=1}^{n} x_i \cdot p_i + \bar{x}^2 = \overline{x^2} - \bar{x}^2.$$

Задача. Найти дисперсию случайной величины X, имеющей закон распределения

X_I	-2	1	2
P_i	0,5	0,3	0,2

Решение. Среднее значение было найдено ранее и равно $-0,3$. Поэтому применяя определение дисперсии, имеем

$$\sigma^2 = (-2+0{,}3)^2 \cdot 0{,}5 + (1+0{,}3)^2 \cdot 0{,}3 +$$
$$+ (2+0{,}3)^2 \cdot 0{,}2 = 3{,}01.$$

Важно! Математическое ожидание, дисперсия и среднее квадратическое отклонения, характеризуя случайную величину, сами случайными величинами не являются.

Если случайной величиной является доходность некоторых акций, то средняя величина характеризует прогнозируемую доходность, а среднее квадратическое отклонение – меру колеблемости доходов от среднего значения т.е. риск данного актива.

2.5. Функция распределения непрерывной случайной величины

Задание закона распределения в виде таблицы неприменимо для непрерывных случайных величин.

Возможен другой подход при котором задается вероятность того, что случайная величина *X* примет значение меньшее чем *x*.

Определение. *Функцией распределения* случайной величины *X* называется функция *F(x)* , которая для каждого значения *x* определяет вероятность того, что случайная величина *X* примет значение меньшее нежели *x* . Эту величину называют иногда интегральной функцией распределения.

Рассмотрим общие свойства функции распределения.

1. Функция распределения случайной величины есть положительно определенная неубывающая функция, значения которой заключены между нулем и единицей:

$$0 \le F(x) \le 1; \quad \lim_{x \to -\infty} F(x) = 0; \quad \lim_{x \to \infty} F(x) = 1.$$

2. Вероятность попадания случайной величины в интервал значений $[x_1, x_2]$ (включая x_1) равна приращению функции распределения на этом интервале

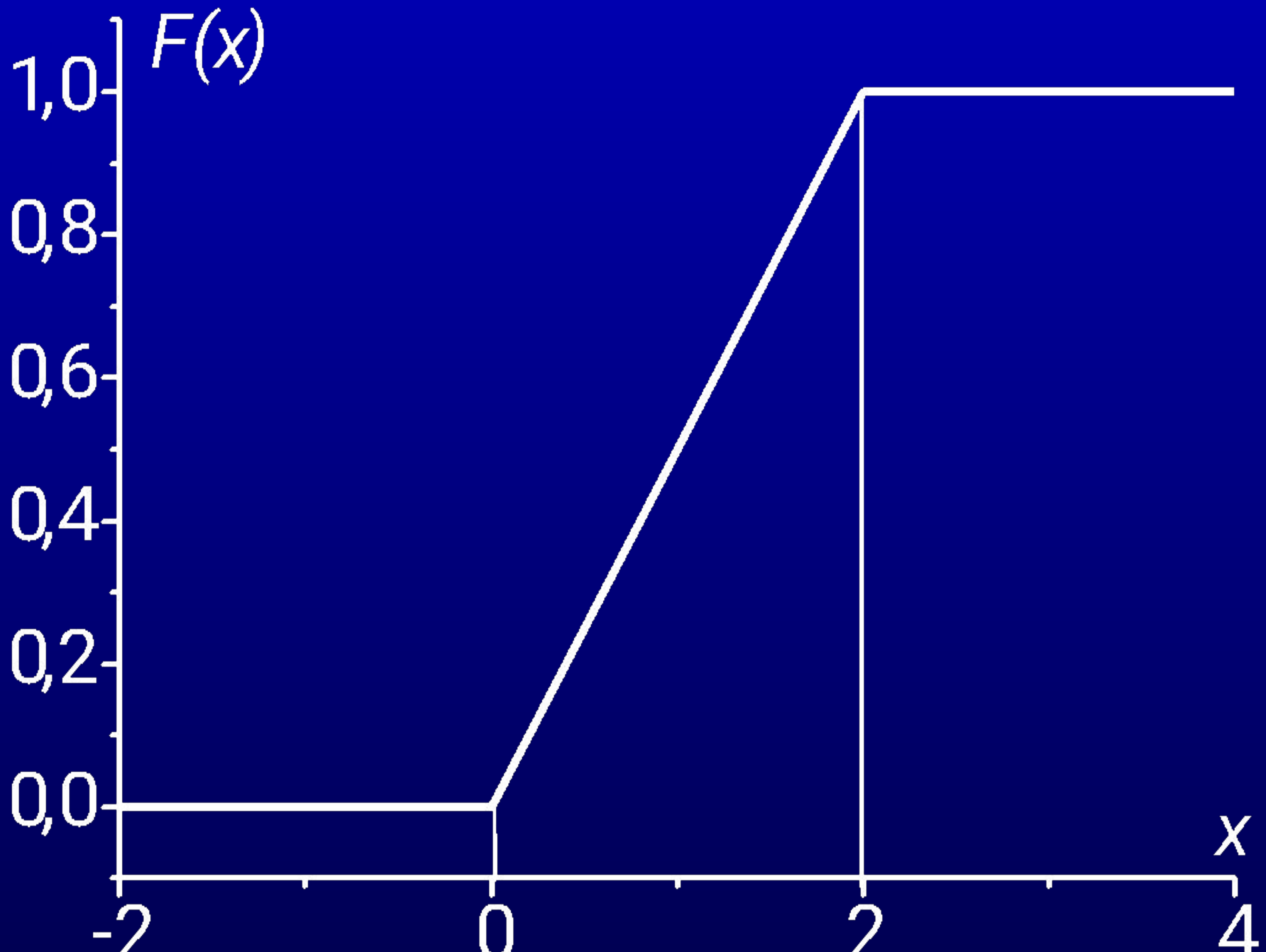
$$P(x_1 \le X < x_2) = F(x_2) - F(x_1).$$

Задача. Функция распределения случайной величины имеет вид

$$F(x) = \begin{cases} 0 & \text{при} \quad x < 0; \\ x/2 & \text{при} \quad 0 < x \le 2; \\ 1 & \text{при} \quad x > 2. \end{cases}$$

Найти вероятность того, что случайная величина примет значение в интервале [1; 3].

График интегральной функции распределения



Решение. Исходя из определения имеем:

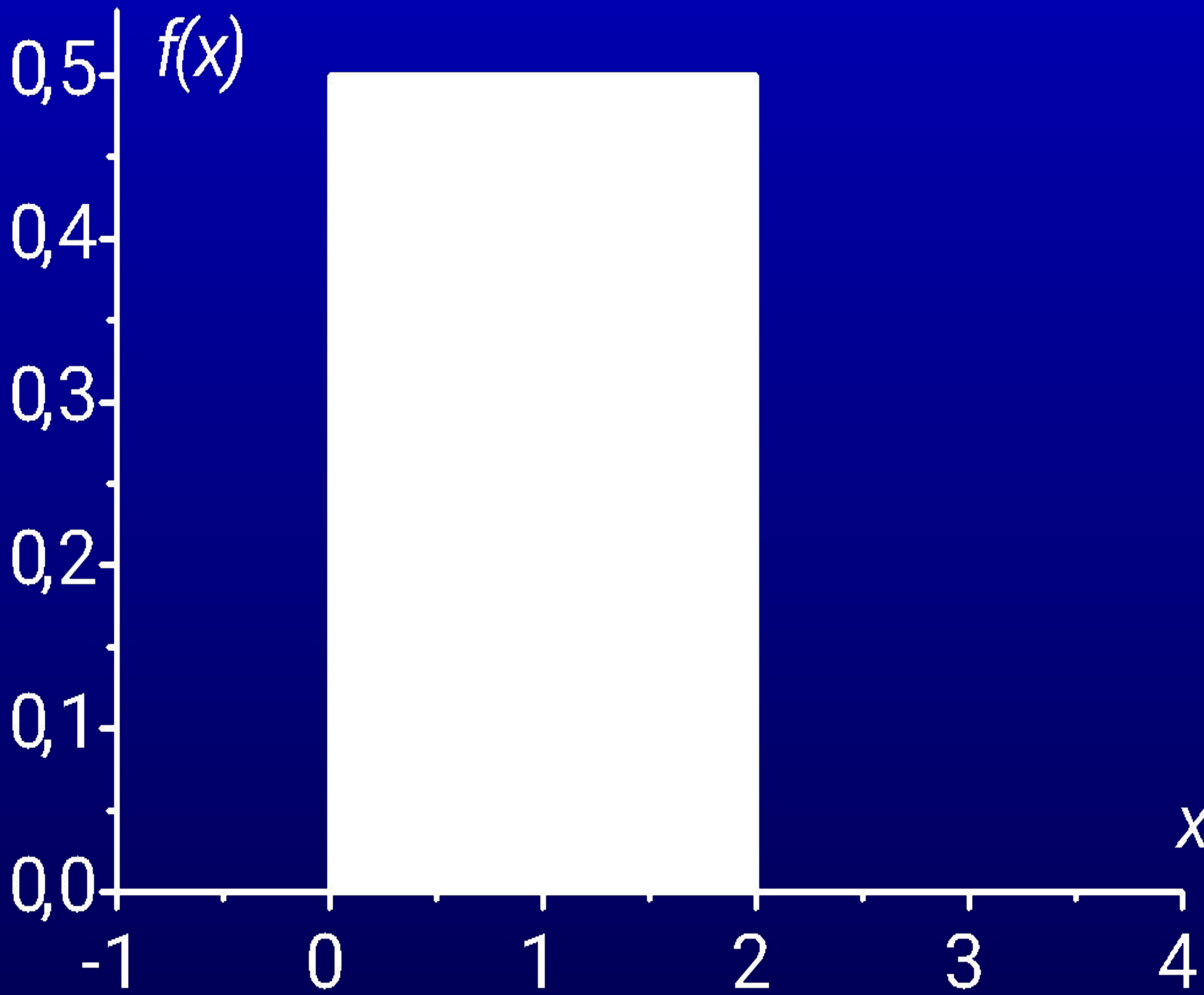
$$P(1 \leq X < 3) = F(3) - F(1) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Для непрерывной случайной величины чаще задается не функция распределения *F(x)* , а другая величина, которая называется плотностью вероятности или плотностью распределения.

Функция плотности вероятности случайной величины X определяется выражением

$$\lim_{\Delta x \to 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x} = \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{\mathrm{dF(x)}}{\mathrm{dx}}.$$

Плотность распределения

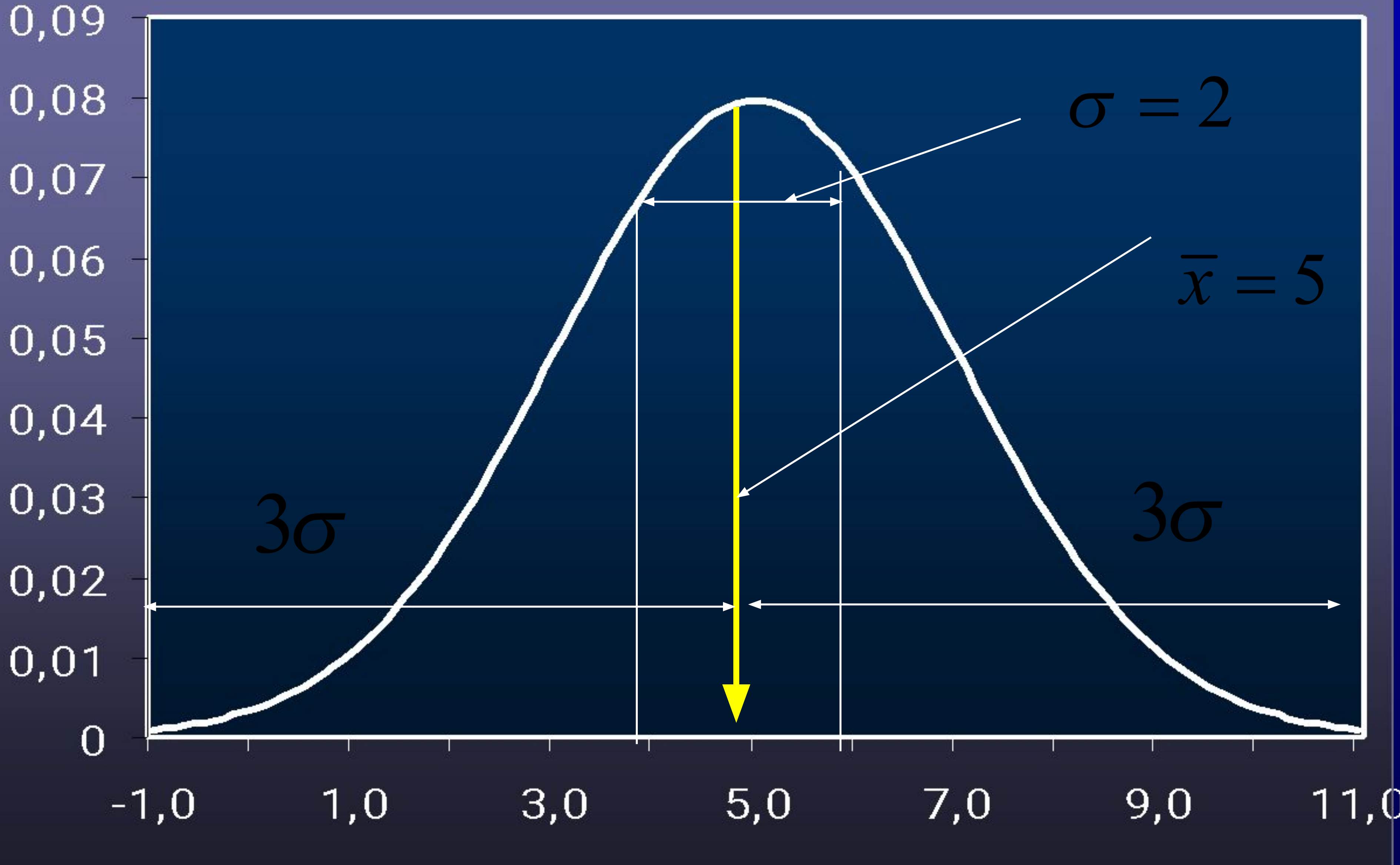


2.6. Нормальное распределение

Нормальное распределение широко используется в математической статистике как предполагаемое теоретическое распределение. Оно зависит от двух параметров σ и $\overline{x}$ и определяется с помощью следующей функции для плотности распределения:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-\overline{x})^2}{2\sigma^2}}.$$

Кривая плотности нормального распределения   2



Свойства нормального распределения

1. Нормальное распределение является симметричным относительно прямой $x = \bar{x}$.

2. Кривая имеет горизонтальную асимптоту – ось абсцисс (при $x \to \infty$ кривая приближается к оси абсцисс).

3. Кривая имеет максимум в точке $x = \bar{x}$, который равен

$$1/\sigma\sqrt{2\pi}$$

4. Площадь между осью абсцисс и кривой нормального распределения равна единице.
5. В промежутке между значениями

$$\bar{x} - 3\sigma \le x \le \bar{x} + 3\sigma$$

содержится 99,73% всей площади кривой, а это означает, что 99,73% всех членов совокупности сосредоточены в этом интервале, если распределение нормальное

Найдем вероятность попадания случайной величины распределенной нормально в интервал значений от x_1 до x_2. В соответствии общим определением функции плотности распределения находим:

$$P(x_1 \le x < x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\bar{x})^2}{2\sigma^2}} dx.$$

Преобразуем это выражение, вводя новую переменную интегрирования

$$t = \frac{x - \bar{x}}{\sigma}; \quad dt = \frac{dx}{\sigma}.$$

В результате получаем:

$$P(x_1 \le x < x_2) = \int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt =$$

$$-\int_{0}^{t_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + \int_{0}^{t_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt,$$

$$t_1 = \frac{x_1 - \bar{x}}{\sigma}; \qquad t_2 = \frac{x_2 - \bar{x}}{\sigma}.$$

Введем в рассмотрение функцию Лапласа, которая определяется выражением

$$\Phi(x)=\int_0^x \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt;\ \Phi(-x)=-\Phi(x).$$

Используя это определение, очевидно, получаем

$$P(x_1 \le x < x_2)=\Phi\left(\frac{x_2-\bar{x}}{\sigma}\right)-\Phi\left(\frac{x_1-\bar{x}}{\sigma}\right).$$

Геометрически искомая вероятность представляет собой площадь между ординатами x_1 и x_2 и кривой.

Геометрический смысл плотности вероятности

2.7 Универсальные распределения

Рассмотрим несколько основных законов распределения, составляющий необходимый математический аппарат для построения в дальнейшем статистических критериев и оценок, применяемых в эконометрике.

Причиной, по которым они играют заметную роль в статистике, является их универсальность. Для их построения не нужно задавать параметры, как для нормального распределения. Они однозначно определяются лишь параметрами, которые обычно известны.

Распределение χ^2 (хи-квадрат)

Распределением χ^2 (хи-квадрат) с ***к*** степенями свободы называется распределение суммы квадратов ***к*** независимых случайных величин, распределенных по стандартному нормальному закону

$$\chi^2 = \sum_{i=1}^{k} Z_i^2,$$

где случайные величины Z_i (i = 1,2,…k) имеют нормальное распределение с математическим ожиданием равным 0 и дисперсией равной 1.

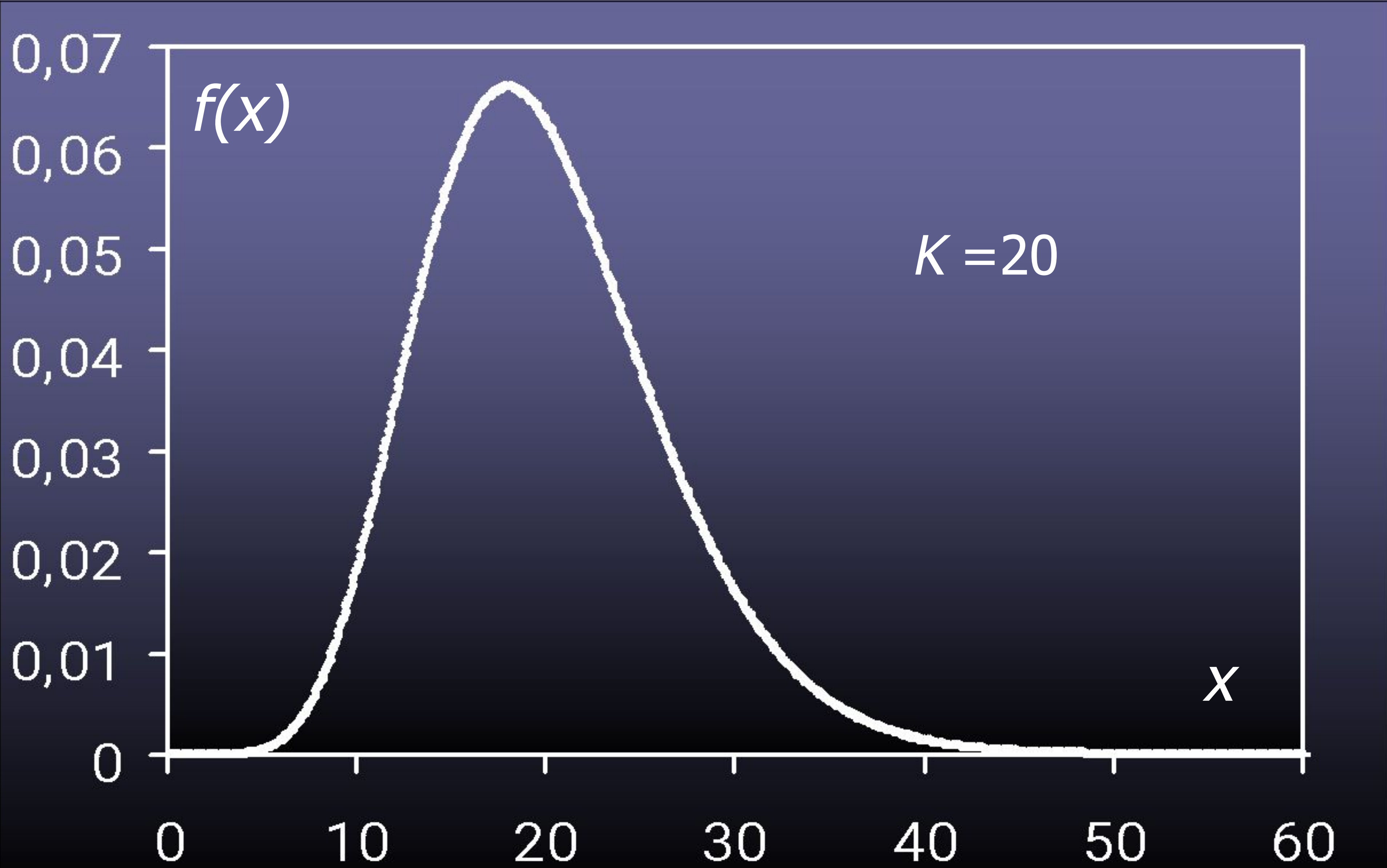
Функция плотности распределения хи-квадрат зависит лишь только от одного параметра – числа степеней свободы.

Числом степеней свободы *k* распределения называется число независимых значений случайной величины. Это число равно числу наблюдений (вариантов) *n* за вычетом числа уравнений связи *L*, которые накладываются на эти наблюдения. Например, если величины X_i связаны линейным соотношением

$$\sum_{i=1}^{n} x_i = const,$$

тогда число степеней свободы *k* будет равным *n-1* .

График функции плотности распределения хи–квадрат



Распределение Стьюдента

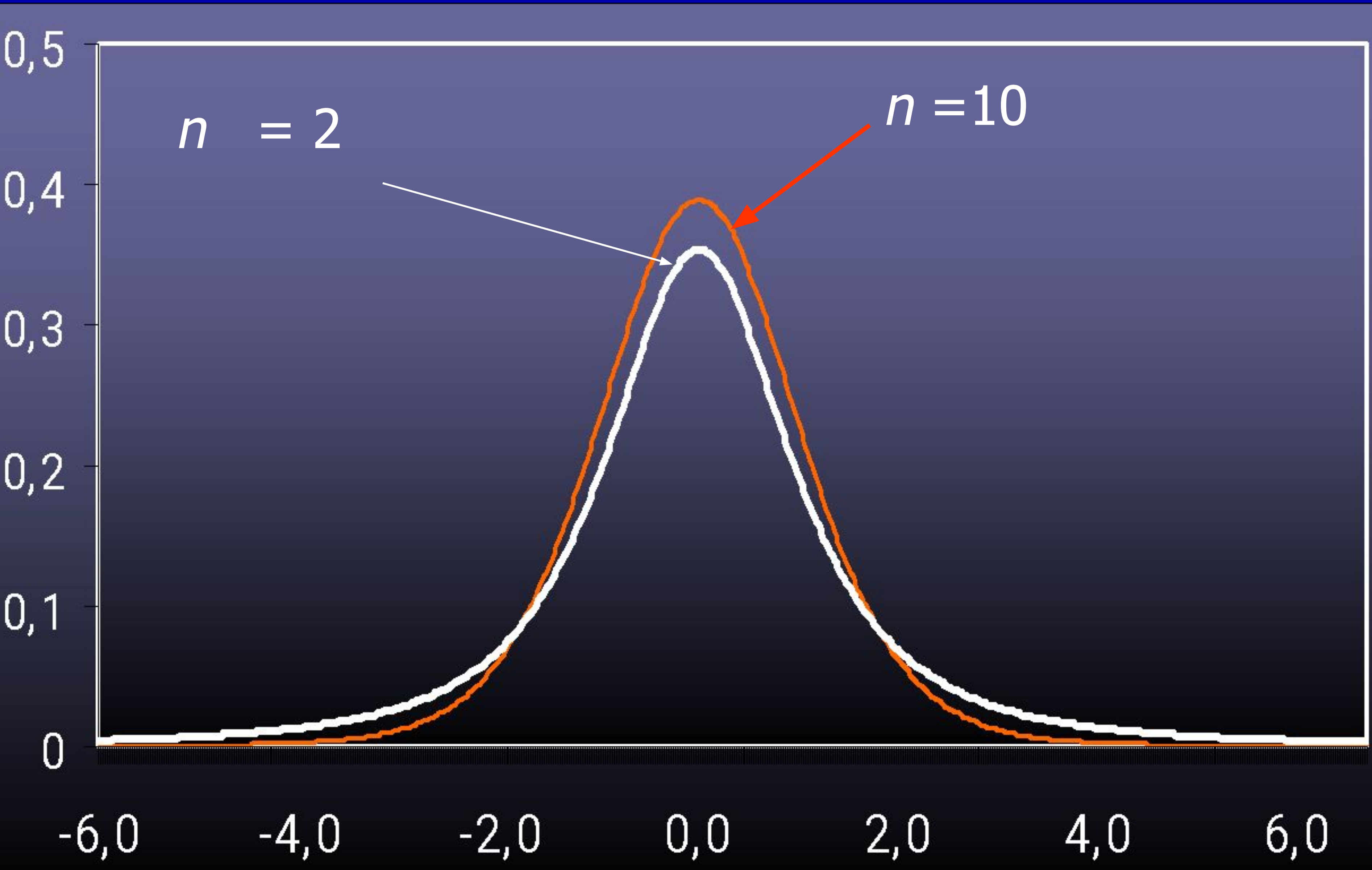
Распределением Стьюдента или t – распределением называется распределение случайной величины

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}},$$

где Z – случайная величина, имеющая стандартное нормальное распределение с математическим ожиданием равным нулю и дисперсией равной единице. χ^2

независимая от Z случайная величина, имеющая распределение хи-квадрат с k степенями свободы.

Графики распределения Стьюдента



Распределение Фишера–Снедекора

Распределением Фишера-Снедекора или F – распределением называется распределение случайной величины

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)},$$ где $\chi^2(k_1)$ и $\chi^2(k_2)$ - случайные величины, имеющие χ^2 распределение.

Графики распределения Фишера-Снедекора