Типовые законы распределения

Для изучения основных законов распределения вероятностей введем понятие **индикатора случайного события А - это дискретная случайная величина Х, которая равна 1 при осуществлении события А и 0 при осуществлении А:**

$$X = \begin{cases} 1, A \\ 0, \overline{A} \end{cases}$$

ряд распределения вероятностей индикатора случайного события:

x_i	0	1
p_i	q	p

где ***p - вероятность осуществления А;***
q = 1 -p - вероятность осуществления . $\overline{A}$
Числовые характеристики индикатора случайного события:

$$m_x = p,\ D_x = qp.$$

Геометрическое распределение

имеет дискретная случайная величина **X, если она принимает значения 0, 1, ..., ∝ с вероятностями:**

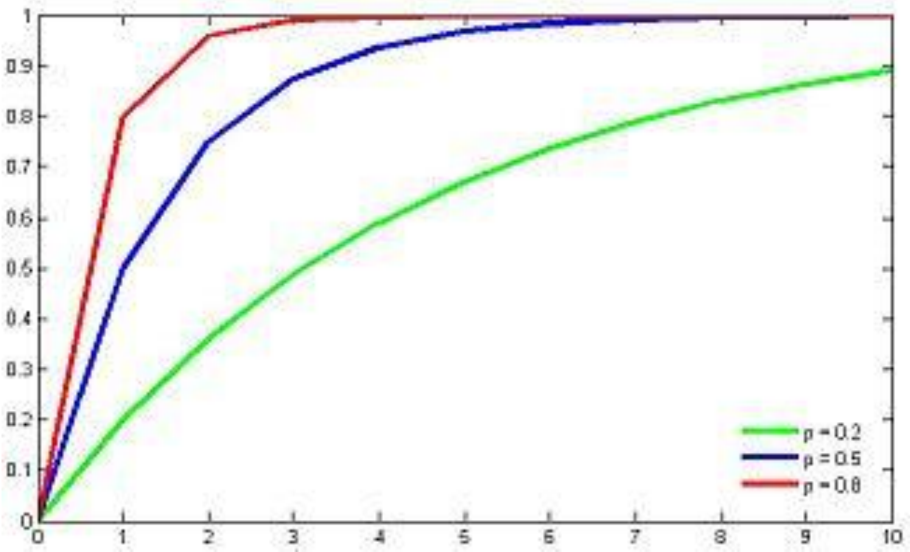
$$P(X = i) = p_i = q\,p,$$

где ***p - параметр распределения (0 < p <1), q=1 - p.***

Числовые характеристики геометрического распределения:

$$m_x = q\,p;\ D_x = q/p^2\ .$$

Условия возникновения. Проводится ряд одинаковых независимых опытов до первого появления некоторого события А. Случайная величина Х - число проведенных безуспешных опытов до первого появления события А



Представьте, вы бросаете монету, вероятность успеха, например, выпадения герба есть p, вероятность неудачи $q = 1 - p$.
Посчитаем, сколько раз вам нужно бросить монету до появления первого успеха (первого герба).
Обозначим эту величину через ν .
Очевидно, это случайная величина.
Зададимся вопросом: какое распределение имеет данная случайная величина?
Ответ прост: случайная величина *n* имеет *геометрическое распределение*, задаваемое формулой:

$$P(\nu = k) = p(1-p)^{k-1}$$

где $k = 1, 2\ldots$

В частности, вероятность того, что герб выпадет на первом шаге равна $P(\nu = 1) = p$
Вероятность того, что герб выпадает впервые на втором шаге (а до того выпадала решетка), равна $P(\nu = 2) = p(1-p)$ и т.д.

Геометрическое распределение является частным случаем отрицательного биномиального распределения.

Геометрическая случайная величина обладает свойством отсутствия последействия: знание о том, что у вас не было успеха в течение n предыдущих бросков, никак не влияет на распределение оставшегося числа бросков до появления герба.

Биномиальное распределение

имеет дискретная случайная величина ***X*, *если она принимает значения 0, 1, ..., n со следующими вероятностями***

$$p(X = i) = p_i = \frac{n!}{i!(n-i)!} p^i q^{n-1},$$

где **n, *p - параметры распределения (0 < p < 1), q = 1 - p.***
Числовые характеристики биномиального распределения:

$$m_x = np; \ D_x = nqp.$$

Условия возникновения. Проводится n одинаковых независимых испытаний, в каждом из которых событие А появляется с вероятностью р. Случайная величина Х - число опытов, в которых произошло событие А.

Если использовать только бумагу и калькулятор, то расчеты по формуле биноминального распределения, несмотря на отсутствие интегралов, даются довольно тяжело.
К примеру значение 100! – имеет более 150 знаков. Вручную рассчитать такое невозможно.
Раньше, да и сейчас тоже, для вычисления подобных величин использовали приближенные формулы.

В настоящий момент целесообразно использовать специальное ПО, типа MS Excel. Таким образом, любой пользователь вполне может вычислить вероятность значения биномиально распределенной случайной величины.

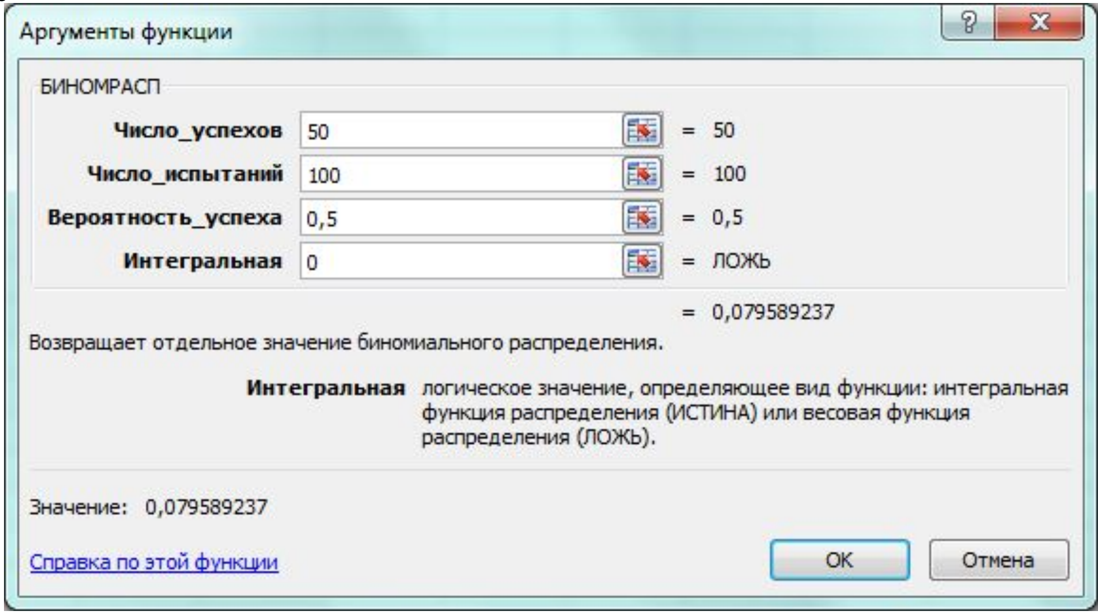
Для закрепления материала задействуем Excel пока в качестве обычного калькулятора, т.е. произведем поэтапное вычисление по формуле биномиального распределения. **Рассчитаем, например, вероятность выпадения 50 орлов**..

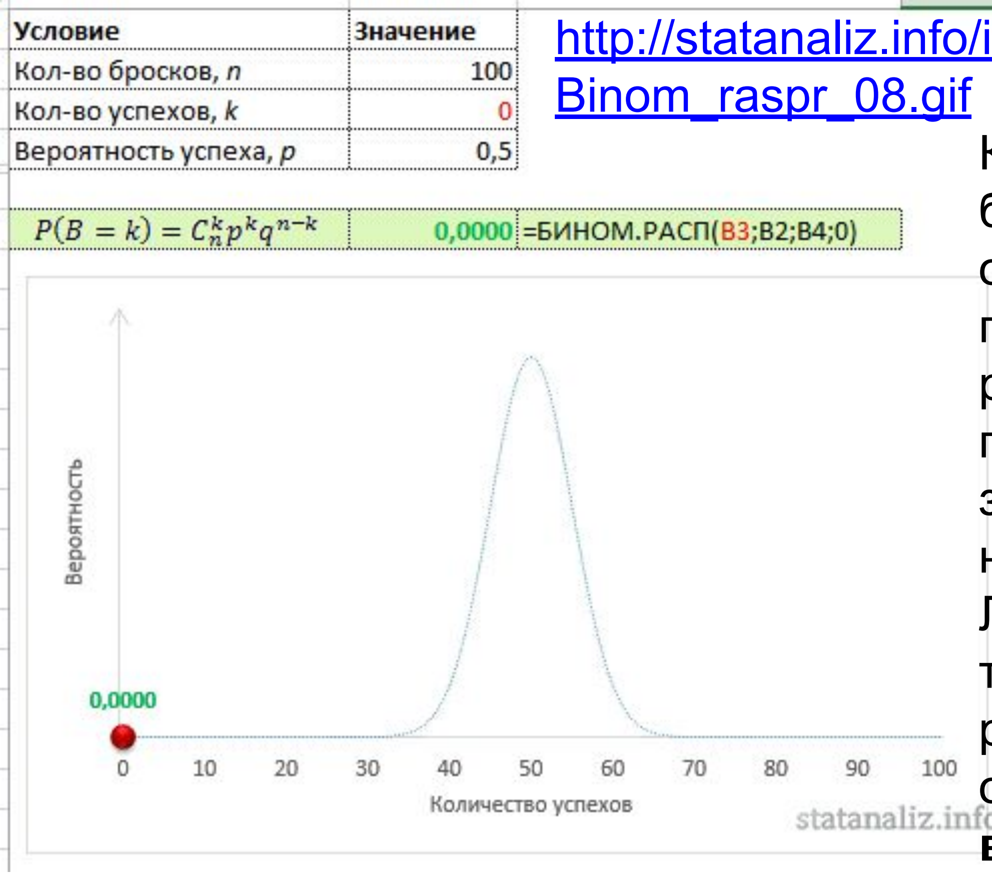
	A	B	C
1	**Условие**	**Значение**	
2	Кол-во бросков, n	100	
3	Кол-во успехов, k	50	
4	Вероятность успеха, p	0,5	
5			
6	**Расчет**	**Значение**	**Формула**
7	n!	9,3326E+157	=ФАКТР(B2)
8	k!	3,04141E+64	=ФАКТР(B3)
9	$C_n^k = \frac{n!}{k!(n-k)!}$	1,00891E+29	=B7/(B8*ФАКТР(B2-B3))
10	p^k	8,88178E-16	=СТЕПЕНЬ(B4;B3)
11	q^{n-k}	8,88178E-16	=СТЕПЕНЬ((1-B4);(B2-B3))
12	$P(B=k) = C_n^k p^k q^{n-k}$	0,0796	=B9*B10*B11
13			

Как видно, промежуточные результаты имеют такой масштаб, что не помещаются в ячейку, хотя везде и используются простые функции типа: ФАКТР (вычисление факториала), СТЕПЕНЬ (возведение числа в степень), а также операторы умножения и деления.

Более того, этот расчет довольно громоздок, во всяком случаен не является компактным, т.к. задействовано много ячеек.

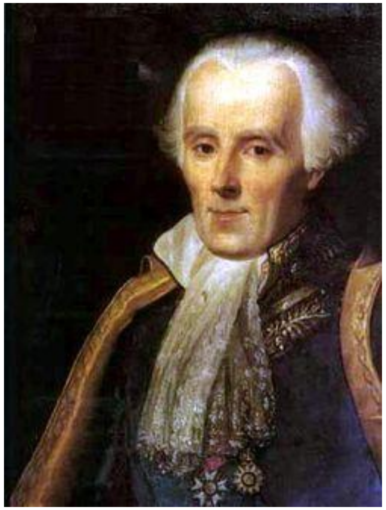
В общем в Excel предусмотрена готовая функция для вычисления вероятностей биномиального распределения. Функция называется БИНОМРАСП.





http://statanaliz.info/images/Metody/terver/Binom_raspr_08.gif

Кто-то может спросить, а не похоже ли биномиальное распределение на... Да, очень похоже. Еще Муавр (в 1733 г.) говорил, что биномиальное распределение при больших выборках приближается к нормальному закону (не знаю, как это тогда называлось), но его никто не слушал. Только Гаусс, а затем и Лаплас через 60-70 лет вновь открыли и тщательно изучили нормальной закон распределения. На графике выше отлично видно, что **максимальная вероятность приходится на математическое ожидание, а по мере отклонения от него, резко снижается**. Также, как и у нормального закона. Биномиальное распределение имеет большое практическое значение, встречается довольно часто. С помощью Excel расчеты проводятся легко и быстро



Распределение Пуассона

имеет дискретная случайная величина X, если она принимает значения 0, 1, …, ∝
со следующими вероятностями:

$$p(X=i)=p_i=\frac{a!}{i!}e^{-ai},$$

Условия возникновения:
Распределение Пуассона является предельным случаем биномиального, когда число опытов ***n неограниченно увеличивается, а вероятность р события А в одном опыте стремится к 0, так что существует предел***

$$\lim_{\substack{n\to\infty \\ p\to 0}} np = a.$$

Случайная величина ***X - число событий пуассоновского потока, поступивших в течение интервала τ, причем параметр а = λτ,***

где λ - интенсивность потока.

Рассмотрим временную ось, на которой будем отмечать моменты возникновения случайных событий (например, отказы компонентов в сложном техническом устройстве, заявки на обслуживание и т.п.).

Последовательность таких моментов называется ***потоком случайных событий.***

Поток случайных событий называется ***стационарным, если число событий, приходящихся на интервал г, в общем случае не зависит от расположения этого участка на временной оси и определяется только его длительностью, т.е. среднее число событий в единице времени λ (интенсивность потока) постоянно.***

Поток случайных событий называется ординарным, если
вероятность попадания в некоторый участок
Δt двух и более случайных событий значительно меньше, чем вероятность
попадания 1-го события.

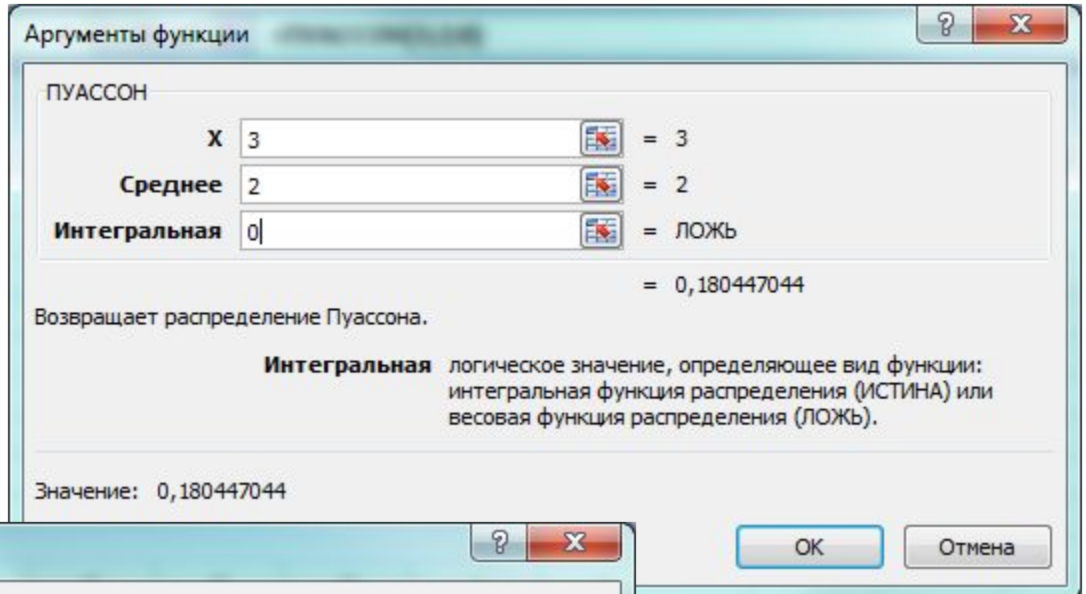
В потоке отсутствует последействие, если вероятность попадания событий на участок τ не зависит от того, сколько событий попало на другие участки, не пересекающиеся с данным.

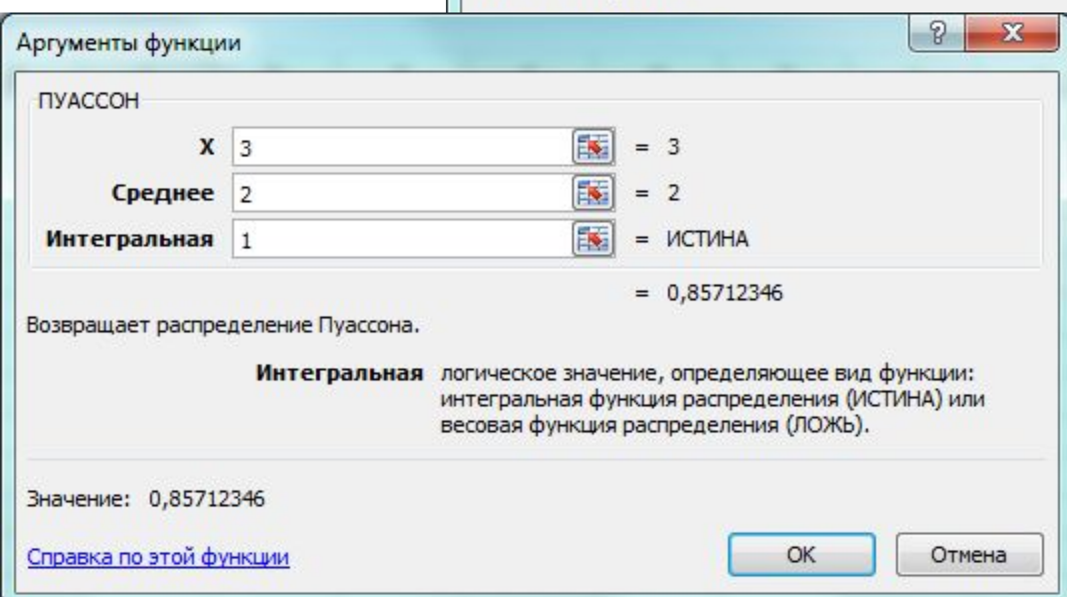
Поток случайных событий называется пуассоновским или простейшим, если он является **стационарным, ординарным и без последействия**.

ПУАССОН

Применяется к: Excel 2016 Excel 2013 Excel 2010 Excel 2007

Возвращает распределение Пуассона. Обычное применение распределения Пуассона состоит в предсказании количества событий, происходящих за определенное время, например количества машин, появляющихся на площади за одну минуту.





x Обязательный. Количество событий.
Среднее Обязательный. Ожидаемое числовое значение.
Интегральная Обязательный.

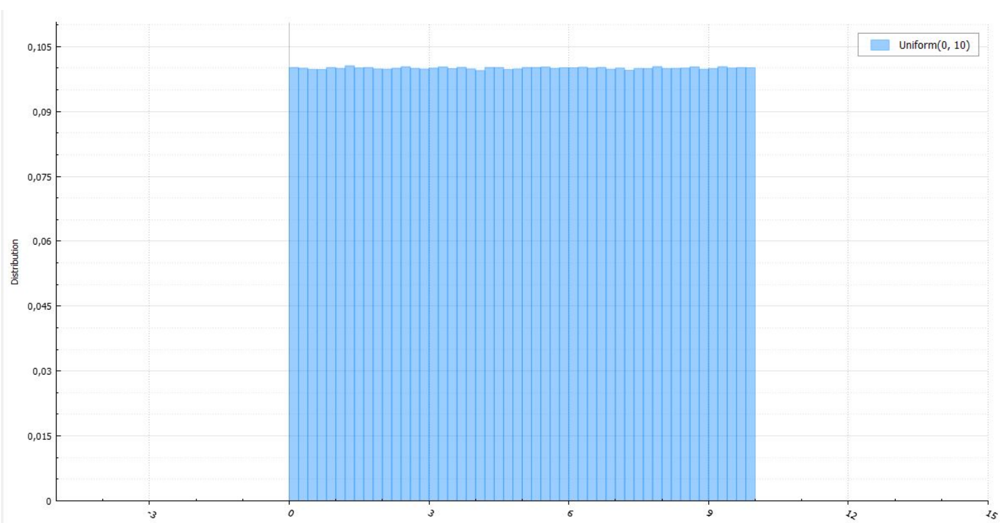
Логическое значение, определяющее форму возвращаемого распределения вероятностей. Если аргумент "интегральная" имеет значение ИСТИНА, то функция ПУАССОН возвращает интегральное распределение Пуассона, то есть вероятность того, что число случайных событий окажется в диапазоне от 0 до x включительно. Если этот аргумент имеет значение ЛОЖЬ, то возвращается весовая функция распределения Пуассона, то есть вероятность точного равенства числа произошедших событий значению x.

Равномерное распределение

имеет непрерывная случайная величина X, если ее плотность вероятности в некотором интервале [a; b] постоянна, т.е. если все значения X в этом интервале равновероятны:

$$f(x) = \begin{cases} 0, x < a; \\ \dfrac{1}{b-a}, a \le x \le b, \end{cases} \qquad F(x) = \begin{cases} 0, x < a; \\ \dfrac{x-a}{b-a}, a \le x \le b, \end{cases}$$

Ниже приведен график плотности равномерного распределения.

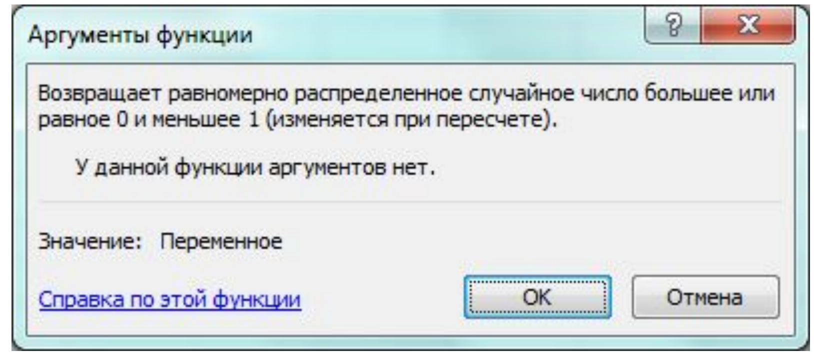


Числовые характеристики равномерно распределенной случайной величины:

$$m_x = \frac{a+b}{2}, \quad D_x = \frac{(b-a)^2}{12}.$$

При необходимости определения параметров a и b по известным m_x, D_x используют следующие формулы:

$$a = m_x + \sigma_x \sqrt{3},$$
$$b = m_x - \sigma_x \sqrt{3}.$$



Условия возникновения:

Случайная величина ***X - ошибки округления при ограниченной разрядной сетке:***

округление до меньшего целого, ***X □ [-1; 0],*** $\boldsymbol{m_x = -0{,}5}$***;***

округление до большего целого, ***X □ [-0; 1],*** $\boldsymbol{m_x = 0{,}5}$***;***

округление до ближайшего целого, ***X □ [- 0,5; 0,5],*** $\boldsymbol{m_x = 0}$***, где 1 - вес младшего разряда.***

Случайная величина ***X - погрешность считывания значений с аналоговой шкалы измерительного прибора, X □ [- 0,5; 0,5],*** $\boldsymbol{m_x = 0}$***, где 1- цена деления шкалы.***

Генераторы псевдослучайных величин, например *RANDOM, встроенные в языки программирования высокого уровня.*

Преобразование равномерно распределенной случайной величины в нормально распределенную

Этот вопрос уже давно подробно изучен, и наиболее широкое распространение получил метод полярных координат, предложенный Джорджем Боксом, Мервином Мюллером и Джорджем Марсальей в 1958 году.

Данный метод позволяет получить пару независимых нормально распределенных случайных величин с математическим ожиданием 0 и дисперсией 1 следующим образом:

$$Z_0 = u \cdot \sqrt{\frac{-2 \ln s}{s}}$$

$$Z_1 = v \cdot \sqrt{\frac{-2 \ln s}{s}}$$

где Z_0 и Z_1 — искомые значения, $s = u^2 + v^2$, а u и v — равномерно распределенные на отрезке (-1, 1) случайные величины, подобранные таким образом, чтобы выполнялось условие

$$0 < s < 1.$$

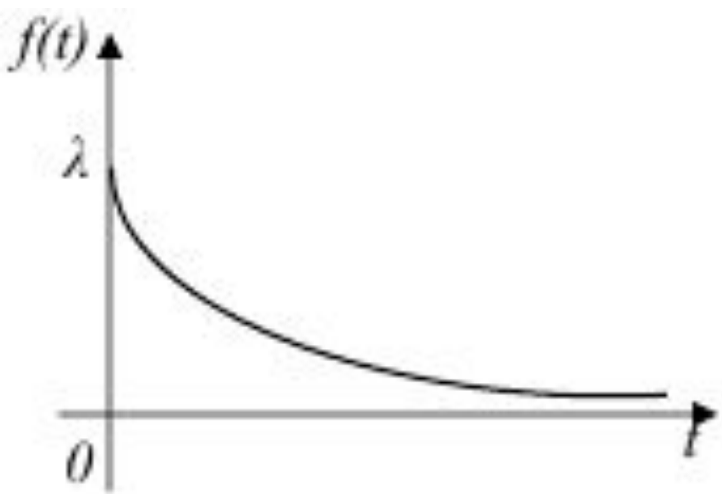
Экспоненциальное распределение

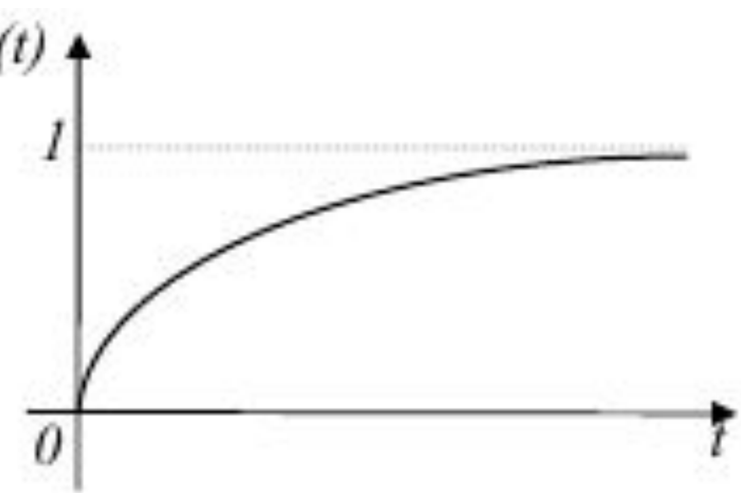
или показательное распределение имеет непрерывная случайная величина Т, принимающая только положительные значения, если ее плотность вероятности и функция распределения равны:

$$f(t) = \begin{cases} 0, t < 0; \\ \lambda e^{-\lambda t}, t \geq 0, \end{cases} \qquad F(t) = \begin{cases} 0, t < 0; \\ 1 - \lambda e^{-\lambda t}, t \leq 0. \end{cases}$$

где λ - *параметр распределения (λ > 0).*

Ниже приведены графики плотности и функции экспоненциального распределения.

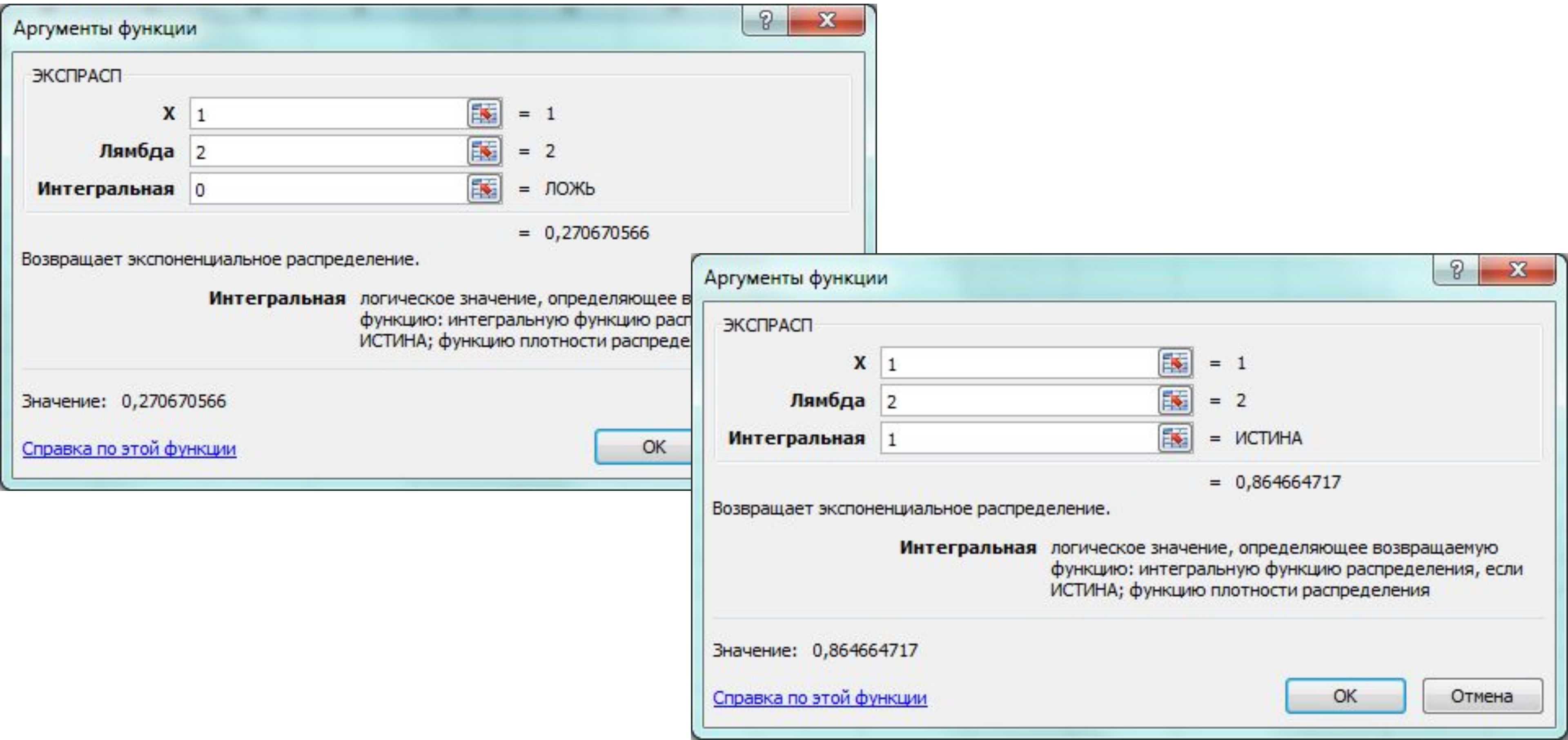




Числовые характеристики экспоненциальной случайной величины:

$$m_T = \frac{1}{\lambda}; D_T = \frac{1}{\lambda^2}$$

Условия возникновения. Случайная величина T – интервал времени между двумя соседними событиями в простейшем или пуассоновском потоке случайных событий, причем параметр распределения λ – интенсивность потока.



Распределение χ (хи - квадрат)

Рассмотрим распределение некоторых случайных величин, представляющих функции нормальных величин, используемых в математической статистике.

Пусть случайная величина **Y, распределена по нормальному** закону ***.***

$$Y \in N(a, \sigma^2).$$

Тогда случайная величина

$$U = \frac{Y - a}{\sigma} = \chi$$

распределена по нормальному закону с параметрами

$$M(U) = 0 \text{ и } \sigma(\mathrm{U}) = 1,\ \ \mathrm{U} \in \mathrm{N}(0,1).$$

Квадрат такой стандартизованной случайной величины

называется **случайной величиной *χ²* *(хи - квадрат) с одной степенью свободы.***

$$U^2 = \left(\frac{Y - a}{\sigma}\right)^2 = \chi^2$$

Рассмотрим n независимых случайных величин Y_1, Y_2, ..., Y_n, распределенных по нормальному закону с $M(Y_i) = a_i$ и средними квадратическими отклонениями σ_i, i=1…n.

Образуем для каждой из этих случайных величин стандартизованную случайную величину

$$U_i = \frac{Y_i - a_i}{\sigma_i}, i = \overline{1, n}.$$

Сумма квадратов стандартизованных переменных

$$\chi^2 = U_1^2 + U_2^2 + \dots + U_n^2 = \left(\frac{Y_1 - a_1}{\sigma_1} \right) + \left(\frac{Y_2 - a_2}{\sigma_2} \right) + \dots + \left(\frac{Y_n - a_n}{\sigma_n} \right)$$

называется случайной величиной χ^2 с ***n степенями свободы. Плотность распределения случайной величины х имеет вид:***

$$f(\chi^2)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}(\chi^2)^{\frac{n}{2}-1}e^{-\frac{\chi^2}{2}}, \text{если } \chi^2>0, \\ 0, \chi^2\le 0\end{cases}$$

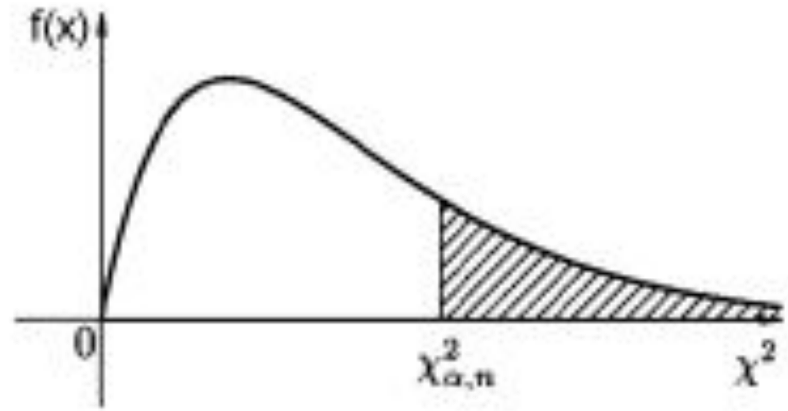
где $$\Gamma(p)=\int_0^\infty t^{p-1}e^{-t}dt$$

гамма-функция Эйлера и является обобщением понятия факториала: *Г**(р)=(р-1)! для целых положительных р.***

Итак, распределение $\boldsymbol{\chi^2}$ зависит от одного параметра n - числа степеней свободы. С возрастанием n распределение $\boldsymbol{\chi^2}$ приближается к нормальному закону распределения (при $n>30$ распределение $\boldsymbol{\chi^2}$ практически не отличается от нормального)

На практике, как правило, используются не ***f*** $\boldsymbol{(\chi^2)}$ ***и F***$\boldsymbol{(\chi^2)}$***, а квантили*** $\boldsymbol{\chi^2}$ ***- распределения*** $\boldsymbol{\chi^2_{\lambda,n}}$.

Квантилем $\chi^2_{\lambda,n}$., отвечающим заданному уровню вероятности λ, называется такое значение $\chi^2 = \chi^2_{\lambda,n}$, при котором площадь заштрихованной криволинейной трапеции (см. рис.5.) была бы равна λ.



Распределение Стьюдента

Распределение Стьюдента (t–распределение) имеет важное значение при статистических вычислениях, связанных с нормальным законом, а именно тогда, когда среднее квадратическое отклонение σ неизвестно и **подлежит определению по опытным данным**.

Пусть Y: Y_1, Y_2, ..., Y_n – независимые случайные величины, имеющие нормальное распределение с параметрами M (Y)=M (Y_i)=0 и $\sigma_Y = \sigma_{Yi} = 1$, i =1,n .

Случайная величина

$$t = \frac{Y}{\sqrt{\frac{1}{n\sum_{i=1}^{n} Y_i^2}}} = \frac{Y}{\sqrt{\frac{1}{n}\chi_n^2}}$$

является функцией нормально распределенных случайных величин и называется безразмерной дробью Стьюдента.

Плотность распределения случайной величины t имеет вид:

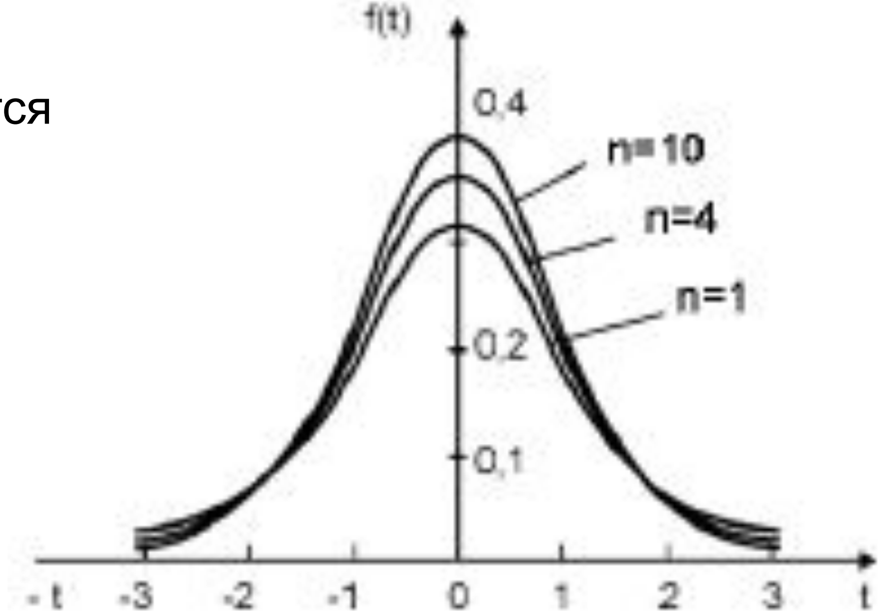
$$f(t) = S(t,n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < t < +\infty,$$

где n - число слагаемых в подкоренном выражении дроби Стьюдента.
Из формулы видно, что распределение случайной величины t зависит только от одного параметра - числа степеней свободы n, равного числу слагаемых в подкоренном выражении дроби Стьюдента ( ).

Известно, что математическое ожидание и дисперсия случайной величины t соответственно равны

$$M(t) = 0; D(t) = \frac{n}{n-2}; (n > 2).$$

На рис.. изображен график плотности распределения Стьюдента при различных степенях свободы. Замечаем, что при увеличении числа степеней свободы n он приближается к кривой Гаусса.

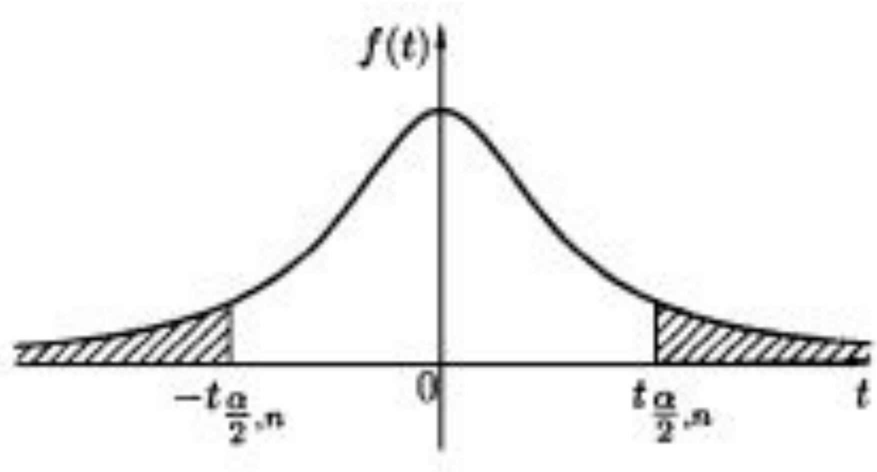


В статистических расчетах используются квантили **t-**распределения *t** .

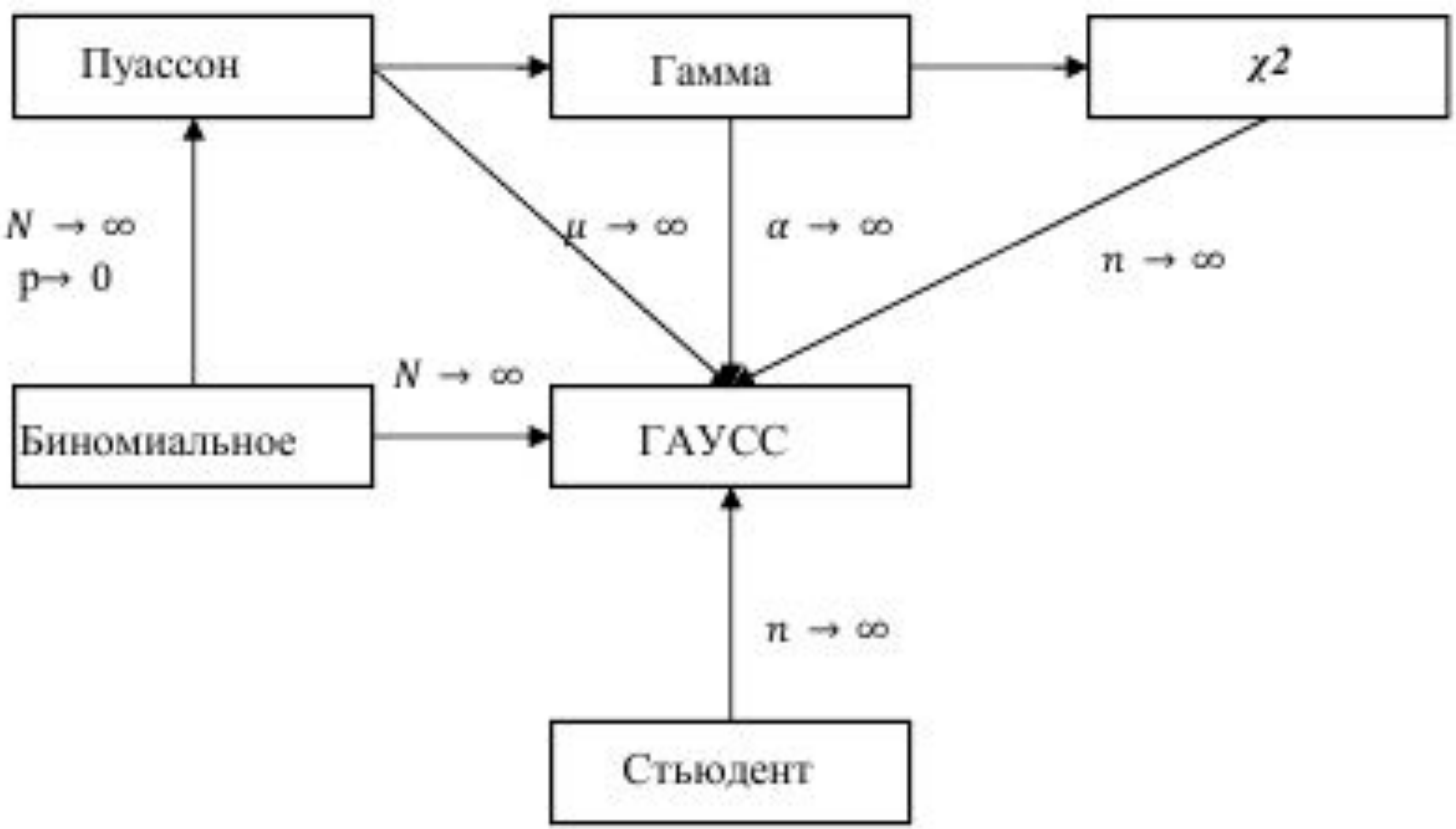
Значения квантилей находятся из решения уравнения:

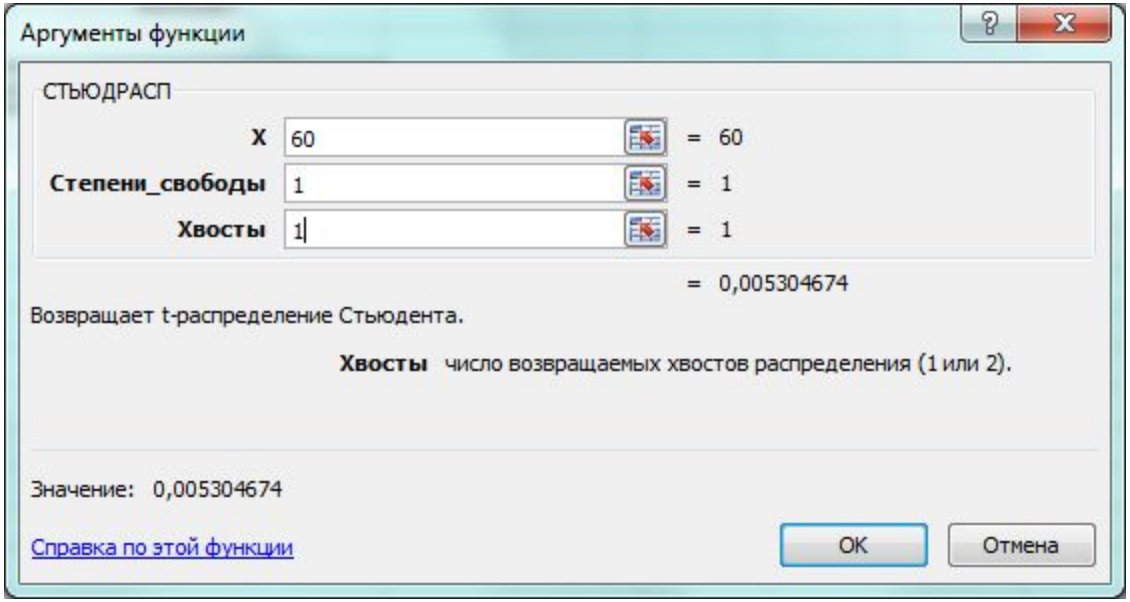
$$P(|t| > t_*) = 2\int_{t_*}^{\infty} f(t)dt = \infty.$$

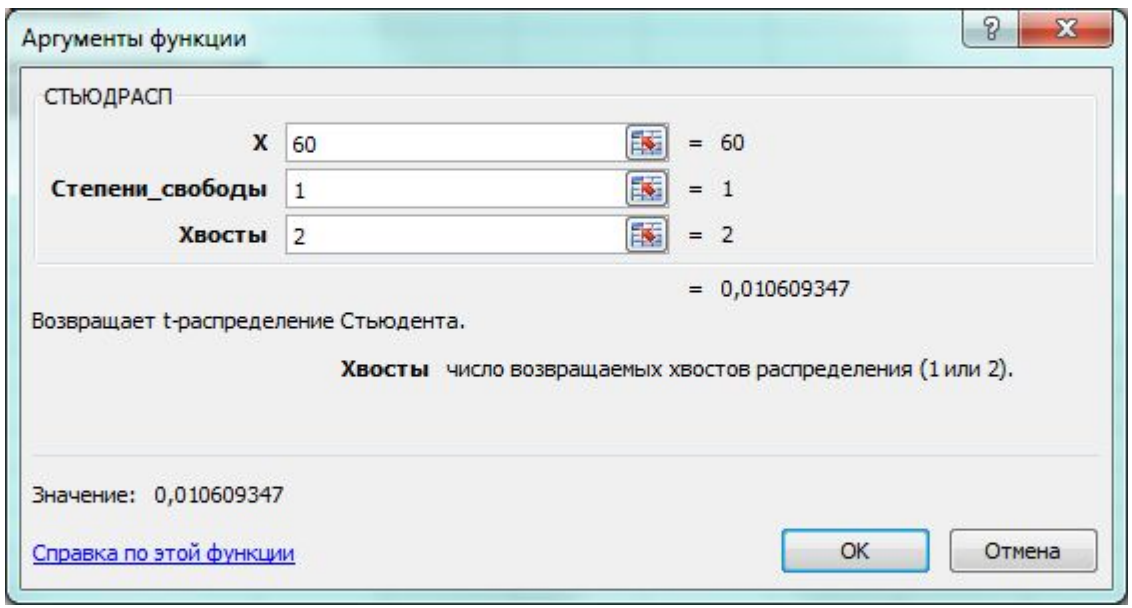
С геометрической точки зрения, нахождение квантилей t_* заключается в том выборе значения $t = t_*$, *при котором* суммарная площадь заштрихованных на рис. криволинейных трапеций была бы равна λ.



На рис. графически представлено соотношение между основными законами распределениями вероятностей.




Аргументы функции
СТЬЮДРАСП
X
60
= 60
Степени_свободы
1
= 1
Хвосты
1
= 1
= 0,005304674
Возвращает t-распределение Стьюдента.
Хвосты число возвращаемых хвостов распределения (1 или 2).
Значение: 0,005304674
Справка по этой функции
ОК
Отмена


Аргументы функции
СТЬЮДРАСП
X
60
= 60
Степени_свободы
1
= 1
Хвосты
2
= 2
= 0,010609347
Возвращает t-распределение Стьюдента.
Хвосты число возвращаемых хвостов распределения (1 или 2).
Значение: 0,010609347
Справка по этой функции
ОК
Отмена