Распределения Пирсона χ^2 (хи – квадрат), Стьюдента и Фишера

Распределение суммы k квадратов независимых случайных величин со стандартным нормальным распределением называется распределением χ^2 (хи-квадрат) с k степенями свободы и обозначается H_k.

Пусть $z_1, \ldots, z_k$ — совместно независимые стандартные нормальные случайные величины, то есть: $z_i \sim N(0,1)$. Тогда случайная величина

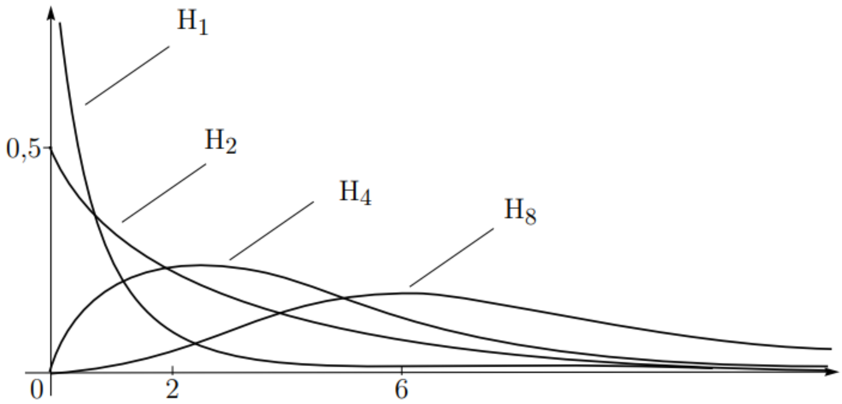
$$x = z_1^2 + \ldots + z_k^2$$

имеет распределение хи-квадрат с k степенями свободы, то есть $x \sim f_{\chi^2(k)}(x)$, или, если записать по-другому:

$$x = \sum_{i=1}^{k} z_i^2 \sim \chi^2(k).$$

Плотности распределений H_k при $k = 1,\ 2,\ 4,\ 8$ показаны на рис.

Максимум функции плотности при $k \geq 2$ распределения H_k достигается в точке k-2 (мода)



Свойство 1. Если случайные величины $\chi^2 \in \mathrm{H}_k$ и $\psi^2 \in \mathrm{H}_m$ независимы, то их сумма $\chi^2 + \psi^2$ имеет распределение H_{k+m}.

Свойство 2. Если величина χ^2 имеет распределение H_k, то

$$\mathsf{E}\chi^2 = k \quad \text{и} \quad \mathsf{D}\chi^2 = 2k.$$

Свойство 3. Пусть $\chi_n^2 \in \mathrm{H}_n$. Тогда при $n \to \infty$

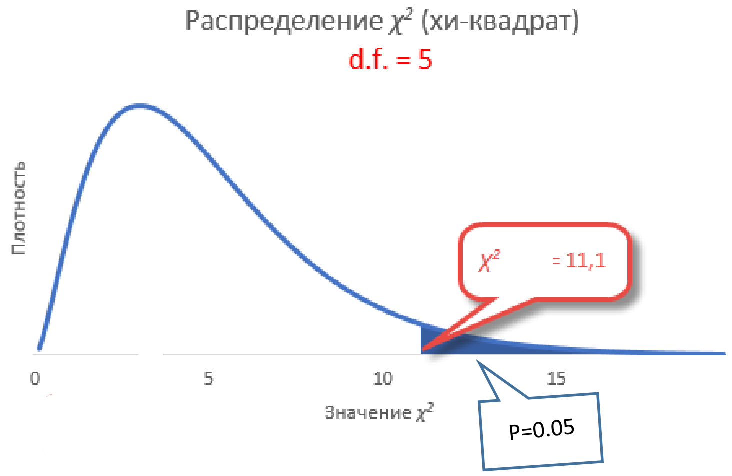
$$\frac{\chi_n^2}{n} \xrightarrow{\text{p}} 1, \qquad \frac{\chi_n^2 - n}{\sqrt{2n}} \Rightarrow \mathrm{N}_{0,1}.$$

Квантиль – это аргумент функции распределения, которой соответствует заданная вероятность.

Степени свободы (df)	Значение χ^2 [3]									
1	0,004	0,02	0,06	0,15	0,46	1,07	1,64	2,71	**3,84**	6,63	10,83
2	0,10	0,21	0,45	0,71	1,39	2,41	3,22	4,61	**5,99**	9,21	13,82
3	0,35	0,58	1,01	1,42	2,37	3,66	4,64	6,25	**7,81**	11,34	16,27
4	0,71	1,06	1,65	2,20	3,36	4,88	5,99	7,78	**9,49**	13,28	18,47
5	1,14	1,61	2,34	3,00	4,35	6,06	7,29	9,24	**11,07**	15,09	20,52
6	1,63	2,20	3,07	3,83	5,35	7,23	8,56	10,64	**12,59**	16,81	22,46
7	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,02	**14,07**	18,48	24,32
8	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	**15,51**	20,09	26,12
9	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	**16,92**	21,67	27,88
10	3,94	4,87	6,18	7,27	9,34	11,78	13,44	15,99	**18,31**	23,21	29,59
p-значение	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	**0,05**	0,01	0,001

.

Например, для χ^2 с 5 степенями свободы квантиль с фиксированной вероятностью 0,95 равен 11,07, т.к. $\boldsymbol{P}(\chi^2 \leq 11.07) = 1 - 0.05 = 0.95$



$\mathbf{P}(\chi^2 > 11.1) = 0.05$.

Задание. Попробуйте Excel, сравнивая с таблицей

ХИ2.РАСП,

ХИ2.РАСП.ПХ

ХИ2.ОБР

ХИ2.ОБР.ПХ

Распределение Стью́дента. Английский статистик Госсет, публиковавший научные труды под псевдонимом Стьюдент, ввёл следующее распределение.

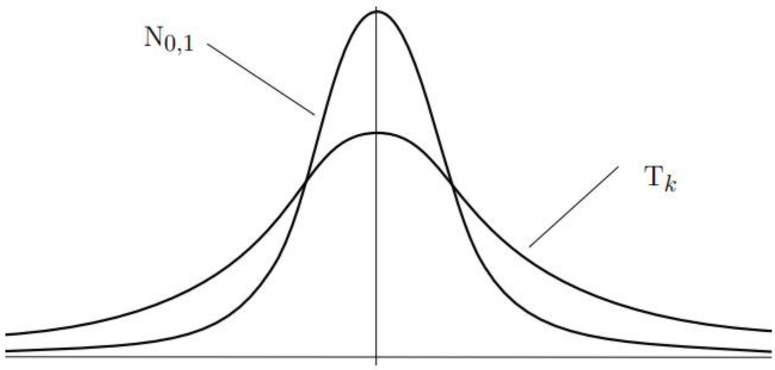
О п р е д е л е н и е . Пусть $\xi_0,\ \xi_1,\ \ldots,\ \xi_k$ независимы и имеют стандартное нормальное распределение. Распределение случайной величины

$$t_k = \frac{\xi_0}{\sqrt{\dfrac{\xi_1^2 + \ldots + \xi_k^2}{k}}}$$

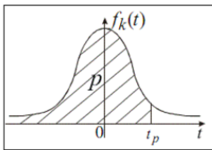
называется *распределением Стью́дента* с k степенями свободы и обозначается T_k.

Распределение Стьюдента совпадает с распределением случайной величины $t_k = \dfrac{\xi}{\sqrt{\chi_k^2 / k}}$, где $\xi \in \mathrm{N}_{0,1}$ и $\chi_k^2 \in \mathrm{H}_k$ независимы.

Графики плотностей стандартного нормального распределения и распределения Стьюдента приведены для сравнения на рис.



$E(T_k)=0$; $D(T_k)=k/(k-2)$, при $k>2$



Число степеней свободы k	Вероятность, p					
	0,9	**0,95**	**0,975**	**0,99**	**0,995**	**0,9995**
1	2	3	4	5	6	7
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
12	1,356	1,782	2,179	2,681	3,055	4,318
14	1,345	1,761	2,145	2,625	2,977	4,140
16	1,337	1,746	2,120	2,584	2,921	4,015
18	1,330	1,734	2,101	2,552	2,878	3,922
20	1,325	1,725	2,086	2,528	2,845	3,850
22	1,321	1,717	2,074	2,508	2,819	3,792
24	1,318	1,711	2,064	2,492	2,797	3,745
26	1,315	1,706	2,056	2,479	2,779	3,707
28	1,313	1,701	2,048	2,467	2,763	3,674
30	1,310	1,697	2,042	2,457	2,750	3,646
60	1,296	1,671	2,000	2,390	2,660	3,460
120	1,289	1,658	1,980	2,358	2,617	3,373

Задание. Попробуйте Excel, сравнивая с таблицей

Распределение Фишера. Следущее распределение тоже тесно связано с нормальным распределением, но понадобится нам не при построении доверительных интервалов, а чуть позже — в задачах проверки гипотез. Там же мы поймём, почему его называют *распределением дисперсионного отношения*.

Определение . Пусть χ_k^2 имеет распределение H_k, а ψ_n^2 — распределение H_n, причём эти случайные величины независимы. Распределение случайной величины

$$f_{k,n} = \frac{\chi_k^2 \,/\, k}{\psi_n^2 \,/\, n} = \frac{n}{k} \cdot \frac{\chi_k^2}{\psi_n^2}$$

называется *распределением Фишера* с k и n степенями свободы и обозначается $\mathrm{F}_{k,n}$.

Свойства распределения Фишера (или *Фишера — Снедекора*):

Свойство . Если случайная величина $f_{k,n}$ имеет распределение Фишера $\mathrm{F}_{k,n}$, то $1/f_{k,n}$ имеет распределение Фишера $\mathrm{F}_{n,k}$.

Заметим, что распределения $\mathrm{F}_{k,n}$ и $\mathrm{F}_{n,k}$ различаются, но связаны соотношением: для любого $x > 0$

$$\mathrm{F}_{k,n}(x) = \mathsf{P}(f_{k,n} < x) = \mathsf{P}\left(\frac{1}{f_{k,n}} > \frac{1}{x}\right) = 1 - \mathrm{F}_{n,k}\left(\frac{1}{x}\right).$$

Свойство . Распределение Фишера $\mathrm{F}_{k,n}$ слабо сходится к вырожденному в точке $c = 1$ распределению при любом стремлении k и n к бесконечности.

Доказательство. Пусть $\xi_1, \xi_2, \ldots$ и $\eta_1, \eta_2, \ldots$ — две независимые последовательности, составленные из независимых случайных величин со стандартным нормальным распределением. Требуемое утверждение вытекает из того, что любая последовательность случайных величин $f_{k,n}$, распределение которой совпадает с распределением отношения двух средних арифметических

$$\frac{\xi_1^2 + \ldots + \xi_k^2}{k} \quad \text{и} \quad \frac{\eta_1^2 + \ldots + \eta_n^2}{n},$$

сходится к единице по вероятности при $k \to \infty$, $n \to \infty$ по ЗБЧ. □

Свойство 11. Пусть $t_k \Subset \mathrm{T}_k$ — случайная величина, имеющая распределение Стьюдента. Тогда $t_k^2 \Subset \mathrm{F}_{1,k}$.