Распределение хи-квадрат (закон Пирсона)

Распределением хи-квадрат с n степенями свободы (обозначается χ_n^2) называется распределение суммы квадратов n независимых случайных величин со стандартным нормальным распределением, т.е. $\chi_n^2 = \sum_{i=1}^{n} \xi_i^2$, если $\xi_i \in N(0, 1)$.

Такое же распределение будет иметь величина

$$\chi_n^2 = \sum_{i=1}^{n} \frac{(\xi_i - a)^2}{\sigma^2}, \text{ если } \xi_i \in N(a, \sigma^2).$$

Плотность распределения случайной величины χ_n^2 с n степенями свободы имеет вид

$$p(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}} \Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Функция распределения:

$$F(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}} \Gamma\left(\dfrac{n}{2}\right)} \displaystyle\int_0^x t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Основные числовые характеристики распределения χ_n^2 следующие:

математическое ожидание — $M\chi_n^2 = n$;

дисперсия — $D\chi_n^2 = 2n$;

асимметрия — $\beta_{\chi^2} = \sqrt{\dfrac{8}{n}}$;

эксцесс — $\nu_{\chi 2} = 12/n$.

На рис. 1 представлен график плотности распределения хи-квадрат с 5 степенями свободы.

Рис. 1

Плотность χ_n^2 распределения зависит от одного параметра n — числа степеней свободы. При $n \le 2$ функция плотности убывает, а при $n > 2$ имеет единственный максимум в точке $x_{mod} = n - 2$. С ростом числа степеней свободы n распределение χ^2_n приближается к нормальному со средним n и дисперсией $2n$ (в смысле асимптотической нормальности). Общий вид графика представлен на рис. 2

Заштрихованная область соответствует вероятности α и определяет *квантиль уровня* α распределения хи-квадрат.

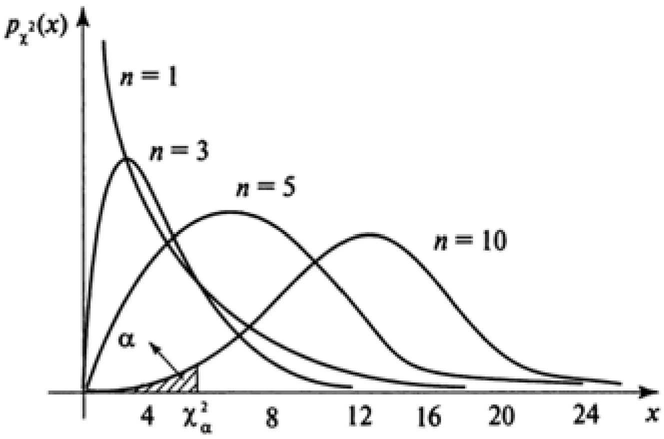

Рис..2

Распределение Стьюдента

Пусть $n + 1$ случайных величин $\xi_0, \xi_1, \ldots, \xi_n$ независимы и имеют стандартное нормальное распределение: $\xi_i \in N(0,1)$. Пусть $\eta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i^2}$. Случайная величина $t_n = \frac{\xi_0}{\eta}$ называется **безразмерной дробью Стьюдента**, а ее распределение — **распределением Стьюдента с n степенями свободы**.

Плотность распределения t_n имеет вид

$$p_t(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad -\infty < x < +\infty.$$

Плотность распределения Стьюдента зависит от одного параметра — числа степеней свободы. С ростом числа степеней свободы распределение Стьюдента сходится к стандартному нормальному.

Основные числовые характеристики:

1) мода и медиана равны математическому ожиданию и равны нулю, т.е. распределение унимодально и симметрично относительно точки $x = 0$;

2) дисперсия $-$ $Dt_n = n/(n - 2)$ и существует только при $n > 2$;

3) асимметрия $-$ $\beta = 0$;

4) эксцесс — $\nu_n = 6/(n - 4)$.

На рис. 3 представлен график плотности распределения Стьюдента с тремя степенями свободы.

Плотность распределения Стьюдента зависит от одного параметра n — числа степеней свободы. Общий вид кривой распределения представлен на рис. 4.

Рис. 3

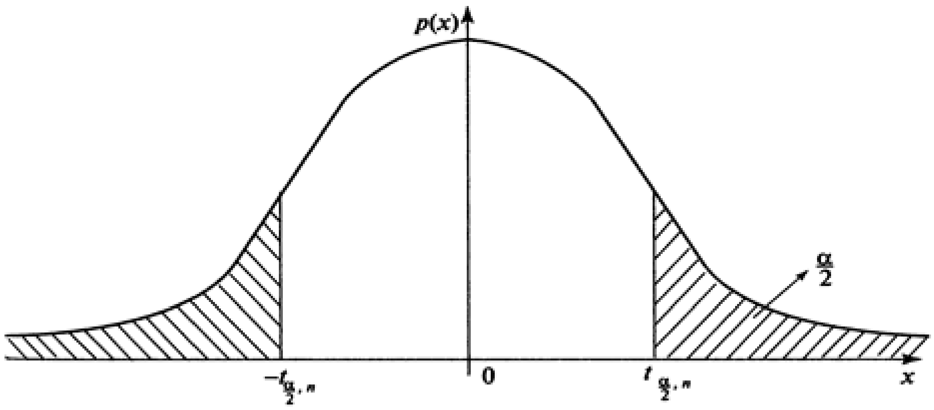

Рис. 4

Заштрихованным областям (слева и справа) соответствуют вероятности по $\alpha/2$, в сумме дающие вероятность α и определяющие *критические точки* распределения Стьюдента для двусторонней области.

Распределение Фишера

Если χ_n^2 и χ_m^2 — независимые случайные величины, распределенные по закону χ^2 с числами степеней свободы n и m соответственно, то случайная величина

$$F(n, m) = \frac{\chi_n^2/n}{\chi_m^2/m}$$

имеет распределение, которое называют **распределением Фишера—Снедекора** (F-распределением) с числами степеней свободы n и m, или *распределением дисперсионного отношения*.

Функция распределения и функция плотности имеют вид

$$F_{n,m}(x)=\frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\int_0^x \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}u^{\frac{n}{2}-1}}{(nu+m)^{\frac{n+m}{2}}}\,du;$$

$$p_{n,m}(x)=\frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\cdot\frac{n^{\frac{n}{2}}m^{\frac{m}{2}}x^{\frac{n}{2}-1}}{(nx+m)^{\frac{n+m}{2}}},\quad x>0.$$

Распределение Фишера определяется двумя параметрами: n и m, называемыми числами степеней свободы.

Основные числовые характеристики:

1) математическое ожидание $M_F=\dfrac{m}{m-2}$ существует только при $m > 2$;

2) дисперсия $D_F=\dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ существует только при $m > 4$;

3) мода $x_{\text{mod}}=\dfrac{m(n-2)}{n(m+2)}$ при $n > 1$;

4) асимметрия $\beta_F=\dfrac{(2n+m-2)\sqrt{8(m-4)}}{(m-6)\sqrt{n(n+m-2)}}$, при $m > 6$;

5) эксцесс $\nu_F=\dfrac{3(m-6)(2+\frac{1}{2}\beta_F^2)}{m-8}-3$, при $m > 8$.

Из этих формул следует, что F-распределение всегда имеет модальное значение, меньшее единицы, и среднее значение, большее единицы, а также положительную асимметрию.

На рис. 5 представлен график плотности распределения Фишера с числами степеней свободы $n = 10$ и $m = 15$.

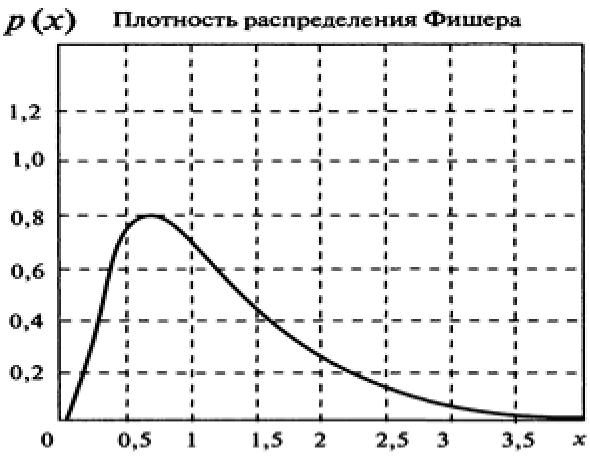

Рис. 5

Общий вид графика плотности распределения Фишера при различных значениях параметров представлен на рис. 6

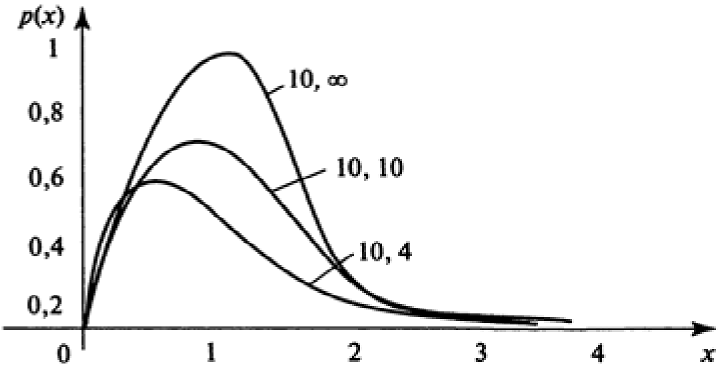

Рис. 6