Лекция 6

6. Нормальное и связанные с ним распределения

6.1. Некоторые факты из теории распределений

1. Распределение линейной функции

Пусть задана случайная величина ξ с функцией распределения $F(x) = \mathsf{P}\{\,\xi < x\,\}$. Рассмотрим линейную функцию $\eta = a\xi + b$, $a \neq 0$. Обозначим через $G(y)$ функцию распределения η.

$$\eta < y \;\Leftrightarrow\; a\xi + b < y \;\Leftrightarrow\; \begin{cases} \xi < (y-b)/a, & a > 0, \\ \xi > (y-b)/a, & a < 0. \end{cases}$$

Следовательно, $G(y) = F\left(\dfrac{y-b}{a}\right)$ при $a > 0$. Если $a < 0$, то

$$G(y) = \mathsf{P}\{\xi > (y-b)/a\} = 1 - \mathsf{P}\{\xi \le (y-b)/a\} = 1 - F\left(\frac{y-b}{a} + 0\right).$$

Таким образом,

$$G(y) = \begin{cases} F\left(\dfrac{y-b}{a}\right), & a > 0, \\ 1 - F\left(\dfrac{y-b}{a} + 0\right), & a < 0. \end{cases} \tag{1.1}$$

Если $F(x)$ абсолютно непрерывна и $f(x)$ её плотность, то

$$g(y) = G'(y) = \frac{1}{|\,a\,|}\, f\left(\frac{y-b}{a}\right). \tag{1.2}$$

2. Распределение квадратичной функции

Рассмотрим функцию $\eta = \xi^2$. Пусть $G(y)$ — её функция распределения. Так как $\eta \ge 0$, то $G(y) = 0$ при $y \le 0$. Если $y > 0$, то

$$\eta < y \;\Leftrightarrow\; -\sqrt{y} < \xi < \sqrt{y}.$$

Следовательно, $G(y) = F(\sqrt{y}) - F(-\sqrt{y} + 0)$. В результате

$$G(y) = \begin{cases} 0, & y \le 0, \\ F(\sqrt{y}) - F(-\sqrt{y} + 0), & y > 0. \end{cases} \tag{1.3}$$

Если $F(x)$ абсолютно непрерывна и $f(x)$ её плотность, то

$$g(y) = G'(y) = \begin{cases} 0, & y \le 0, \\ \dfrac{1}{2\sqrt{y}}\big(f(\sqrt{y}) - f(-\sqrt{y})\big), & y > 0. \end{cases} \tag{1.4}$$

3. Распределение частного случайных величин

Пусть плотность распределения вероятностей величины (ξ, η) равна $p(y,z)$. Найдём функцию распределения частного $\zeta = \xi/\eta$. По определению

$$F_{\xi}(x) = \mathsf{P}\{\,\xi/\eta < x\,\}.$$

Так как

$$\{\,\xi/\eta < x,\ \eta \neq 0\,\} = \{\,\xi < x\eta,\ \eta > 0\,\} = \{\,\xi > x\eta,\ \eta < 0\,\},$$

то

$$F_{\xi}(x) = \int_0^{\infty}\int_{-\infty}^{xz} p(y,z)\,dy\,dz + \int_{-\infty}^{0}\int_{xz}^{\infty} p(y,z)\,dy\,dz. \tag{1.5}$$

Дифференцируя (1.5), получим формулу для плотности:

$$p_{\xi}(x) = \int_0^{\infty} zp(zx,z)\,dz - \int_{-\infty}^{0} zp(zx,z)\,dz. \tag{1.6}$$

Пусть теперь случайные величины ξ, η независимы и $p_1(y)$, $p_1(z)$ — их плотности распределения. Тогда формула (1.5) запишется в виде

$$F_{\xi}(x) = \int_0^{\infty} F_1(xz)\,p_2(z)\,dz - \int_{-\infty}^{0} (1 - F_1(xz))\,p_2(z)\,dz, \tag{1.7}$$

а формула (1.6) в виде

$$p_{\xi}(x) = \int_0^{\infty} zp_1(zx)p_2(z)\,dz - \int_{-\infty}^{0} zp_1(zx)p_2(z)\,dz. \tag{1.8}$$

6.2. Нормальное распределение

Стандартная нормальная функция распределения определяется равенством

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt. \tag{2.1}$$

Соответствующая *стандартная нормальная плотность вероятности* есть

$$\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}. \tag{2.2}$$

С функцией $\Phi(x)$ тесно связаны специальные функции $\operatorname{erf} x$, $\operatorname{erfc} x$:

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}}\int_0^{x} e^{-t^2}dt, \quad \operatorname{erfc} x = 1 - \operatorname{erf} x = \frac{2}{\sqrt{\pi}}\int_x^{+\infty} e^{-t^2}dt, \tag{2.3}$$

называемые *функцией ошибок* и *дополнительной функцией ошибок* соответственно. Функция $\Phi(x)$ выражается через $\operatorname{erf} x$ следующим образом:

$$\Phi(x) = \frac{1}{2}\left\{1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right\}. \tag{2.4}$$

Рассмотрим моменты стандартно нормальной функции распределения

$$\alpha_k = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^k e^{-t^2/2}\,dt.$$

Очевидно, что все нечётные моменты равны нулю. Чётные моменты легко вычисляются с помощью формулы

$$\int_0^\infty x^{\alpha-1} \mathrm{e}^{-Ax^2}\, dx = \frac{1}{2}\, A^{-\alpha/2}\, \Gamma(\alpha/2), \quad \alpha > 0,\ A > 0. \tag{2.5}$$

В результате получаем

$$\alpha_k = \begin{cases} 0, & \text{при } k = 2m+1, \\ (2m-1)!!, & \text{при } k = 2m. \end{cases} \tag{2.6}$$

Случайная величина ξ называется *нормально распределённой с параметрами a и σ^2* или, короче, *нормальной* (a, σ^2), если функция распределения величины ξ есть $\Phi\left(\dfrac{x-a}{\sigma}\right)$, где $\sigma > 0$ и a — постоянные. Плотность вероятности равна

$$\frac{1}{\sigma}\, \Phi'\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-a)^2/2\sigma^2}$$

Найдём математическое ожидание и дисперсию. Имеем

$$\mathsf{M}\xi = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-(x-a)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (a + \sigma y)\, e^{-y^2/2}\, dy = a,$$

$$\mathsf{D}\xi = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-(x-a)^2/2\sigma^2} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2}\, dy = \sigma^2.$$

Из (2.6) следует, что моменты относительно среднего значения величины ξ равны

$$\mu_{2m+1} = 0, \quad \mu_{2m} = (2m-1)!!\, \sigma^{2m}. \tag{2.7}$$

В частности, коэффициенты асимметрии и эксцесса равны

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = 0, \quad \gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = 0.$$

Если случайная величина ξ нормальна (a, σ^2), то из (1.2) следует, что линейная функция $A\xi + B$ нормальна $(Aa + B, A^2\sigma^2)$.

6.3. Распределение χ^2

Пусть ξ_1, ξ_2, …, ξ_n — независимые случайные величины $\mathcal{N}(0, 1)$. Функция распределения величины

$$\chi^2 = \sum_{k=1}^{n} \xi_k^2 \tag{3.1}$$

называется *χ^2-распределением с n степенями свободы*. Обозначим $K_n(x) = \mathsf{P}\{\chi^2 < x\}$.

Вычислим $K_n(x)$. Очевидно, что $K_n(x) = 0$ при $x \le 0$. Пусть $x > 0$. Величина $K_n(x)$ равна вероятности попадания точки $(\xi_1, \dots, \xi_n)$ внутрь сферы

$$\sum_{k=1}^{n} \xi_k^2 = x.$$

Следовательно,

$$K_n(x) = (2\pi)^{-n/2} \underset{\sum x_i^2 < x}{\int \cdots \int} \exp\Big\{-\sum_{i=1}^{n} \frac{x_i^2}{2}\Big\}\, dx_1 \ldots dx_n.$$

Написанный интеграл легко сводится к выражению:

$$K_n(x) = (2\pi)^{-n/2} C_n \int_0^{\sqrt{x}} r^{n-1} \exp(-r^2/2)\, dr.$$

Константа C_n равна площади единичной n-мерной сферы:

$$C_n = \frac{n\,\pi^{n/2}}{\Gamma(n/2+1)}.$$

Так как $\Gamma(n/2+1) = \dfrac{n}{2}\,\Gamma(n/2)$, то в результате

$$K_n(x) = \frac{1}{2^{n/2-1}\Gamma(n/2)} \int_0^{\sqrt{x}} r^{n-1} \exp(-r^2/2)\, dr = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{x} t^{n/2-1} \exp(-t/2)\, dt.$$

Здесь мы сделали замену $r = \sqrt{t}$. Отсюда получаем выражение для плотности распределения $\chi^2(n)$

$$k_n(x) = \begin{cases} \dfrac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)}\, e^{-x/2}, & x > 0, \\ 0, & x \le 0. \end{cases} \tag{3.2}$$

Вычислим моменты χ^2-распределения. Имеем

$$\begin{aligned} \alpha_\nu &= \int_{-\infty}^{\infty} x^\nu k_n(x)\, dx = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{\infty} x^{\nu+n/2-1} e^{-x/2}\, dx = \\ &= \frac{2^{n/2+\nu}}{2^{n/2}\Gamma(n/2)} \int_0^{\infty} y^{\nu+n/2-1} e^{-y}\, dy = \frac{2^\nu}{\Gamma(n/2)}\, \Gamma(n/2+\nu). \end{aligned}$$

Так как

$$\begin{aligned} \Gamma(n/2+\nu) &= (n/2+\nu-1)(n/2+\nu-2)\ldots(n/2)\,\Gamma(n/2) = \\ &= n(n+2)\ldots(n+2\nu-2)\,\Gamma(n/2), \end{aligned}$$

то мы получаем

$$\alpha_\nu = n(n+2)\ldots(n+2\nu-2). \tag{3.3}$$

Отсюда, в частности,

$$\mathsf{M}\chi^2 = \alpha_1 = n, \quad \mathsf{D}\chi^2 = \alpha_2 - \alpha_1^2 = 2n. \tag{3.4}$$

Если каждая из независимых величин $\xi_1, \ldots, \xi_n$ нормальна $(0, \sigma^2)$, то величины $\dfrac{\xi_1}{\sigma}, \ldots, \dfrac{\xi_n}{\sigma}$ независимы и нормальны $(0,1)$. Следовательно, плотность вероятности величины $\displaystyle\sum_{i=1}^{n} \left(\frac{\xi_i}{\sigma}\right)^2$ равна $k_n(x)$. Тогда получаем функцию плотности величины $\displaystyle\sum_{i=1}^{n} \xi_i^2$:

$$\frac{1}{\sigma^2}\, k_n\left(\frac{x}{\sigma^2}\right) = \frac{1}{2^{n/2}\sigma^n\Gamma(n/2)}\, x^{n/2-1} e^{-x/2\sigma^2}, \quad x > 0.$$

Аналогичными рассуждениями находим функции плотности вероятности величин

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i^2, \quad \sqrt{\sum_{i=1}^{n}\xi_i^2}, \quad \sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i^2}.$$

Результаты даны в следующей ниже таблице. Случайные величины ξ_1, ξ_2, $\ldots$, ξ_n предполагаются независимыми и нормальными $(0, \sigma^2)$.

Величина	Плотность вероятности
$\sum_{k=1}^{n}\xi_k^2$	$\frac{1}{\sigma^2}\,k_n\left(\frac{x}{\sigma^2}\right) = \frac{1}{2^{n/2}\sigma^n\Gamma(n/2)}\,x^{n/2-1}\mathrm{e}^{-x/2\sigma^2}$
$\frac{1}{n}\sum_{k=1}^{n}\xi_k^2$	$\frac{n}{\sigma^2}\,k_n\left(\frac{nx}{\sigma^2}\right) = \frac{(n/2)^{n/2}}{\sigma^n\Gamma(n/2)}\,x^{n/2-1}\mathrm{e}^{-nx/2\sigma^2}$
$\sqrt{\sum_{k=1}^{n}\xi_k^2}$	$\frac{2x}{\sigma^2}\,k_n\left(\frac{x^2}{\sigma^2}\right) = \frac{2}{2^{n/2}\sigma^n\Gamma(n/2)}\,x^{n-1}\mathrm{e}^{-x^2/2\sigma^2}$
$\sqrt{\frac{1}{n}\sum_{k=1}^{n}\xi_k^2}$	$\frac{2nx}{\sigma^2}\,k_n\left(\frac{nx^2}{\sigma^2}\right) = \frac{2(n/2)^{n/2}}{\sigma^n\Gamma(n/2)}\,x^{n-1}\mathrm{e}^{-nx^2/2\sigma^2}$

Распределение $\chi^2(n)$ при больших значениях n $(n > 30)$ с достаточной для практических расчётов точностью аппроксимируется нормальным распределением. Это свойство используется для приближённого выражения квантилей $\chi^2_p(n)$ распределения $\chi^2(n)$ через квантили u_p нормального распределения $N(0,1)$. Обычно используют следующие две формулы:

$$\chi^2_p(n) \approx \frac{1}{2}\left(u_p + \sqrt{2n-1}\,\right)^2, \tag{3.5}$$

$$\chi^2_p(n) \approx n\left(1 - \frac{2}{9n} + u_p\sqrt{\frac{2}{9n}}\right)^3. \tag{3.6}$$

Формула (3.5) применяется при $n \geq 30$ и $p \geq 0{,}5$, а формула (3.6) применяется при $n \geq 30$ и малых p.

Пример. Найдём квантили $\chi^2_{0,01}(10)$, $\chi^2_{0,95}(40)$, $\chi^2_{0,01}(40)$.

Из таблицы находим $\chi^2_{0,01}(10) = 2{,}56$.

Так как $u_{0,95} = 1{,}645$, то из формулы (3.5) получаем

$$\chi^2_{0,95}(40) \approx \frac{1}{2}\left(1{,}645 + \sqrt{2\cdot 40 - 1}\,\right)^2 \approx 55{,}47.$$

Точное значение квантиля по таблице: $\chi^2_{0,95}(40) = 55{,}8$.

По формуле (3.6), используя значение $u_{0,01} = -u_{0,99} = -2{,}326$, получим

$$\chi^2_{0,01}(40) \approx 40\left(1 - \frac{2}{9 \cdot 40} - 2{,}326\sqrt{\frac{2}{9 \cdot 40}}\right)^3 \approx 22{,}14.$$

Точное значение квантиля по таблице: $\chi^2_{0,01}(40) = 22{,}2$.

6.4. Распределение Стьюдента

Пусть $n+1$ случайных величин ξ и $\xi_1, \ldots, \xi_n$ независимы и нормальны $(0, \sigma)$. Положим

$$\eta = \sqrt{\frac{1}{n}\sum_{k=1}^{n} \xi_k^2},$$

где берётся положительная ветвь корня, и рассмотрим величину

$$t = \frac{\xi}{\eta} = \frac{\xi}{\sqrt{\frac{1}{n}\sum_{k=1}^{n} \xi_k^2}}. \tag{4.1}$$

Обозначим через S_n функцию распределения величины t. Таким образом

$$S_n(x) = \mathsf{P}\{\, t < x \,\} = \mathsf{P}\left\{\, \xi/\eta < x \,\right\}.$$

Плотности распределений для ξ и η равны $p_\xi(y) = \dfrac{1}{\sigma}\Phi\left(\dfrac{y}{\sigma}\right)$ и $p_\eta(z) = \dfrac{2nz}{\sigma^2}\, k_n\left(\dfrac{nz^2}{\sigma^2}\right)$ (см. таблицу 4 строка). Воспользуемся формулой (1.8). Так как $p_\eta(z) = 0$ при $z \le 0$, то

$$s_n(x) = \int_0^\infty z p_\xi(zx)\, p_\eta(z)\, dz = c_n \int_0^\infty z^n e^{-(x^2+n)z^2/2\sigma^2}\, dz,$$

где

$$c_n = \sqrt{\frac{2}{\pi}}\, \frac{(n/2)^{n/2}}{\sigma^{n+1}\Gamma(n/2)}.$$

Из (2.5) получаем

$$\int_0^\infty z^n e^{-(x^2+n)z^2/2\sigma^2}\, dz = \frac{(2\sigma^2)^{(n+1)/2}}{2(x^2+n)^{(n+1)/2}}\, \Gamma((n+1)/2),$$

В результате

$$s_n(x) = \frac{1}{\sqrt{\pi n}}\, \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}. \tag{4.2}$$

Распределение, задаваемое функцией плотности $s_n(x)$, известно под названием *распределения Стьюдента* или *t-распределения*. Оно было впервые использовано в одной статистической проблеме Госсетом (William Sealy Gosset), писавшим под псевдонимом *Student*. Как и в случае χ^2-распределения, параметр n называют *числом степеней свободы* распределения Стьюдента.

При $\nu < n$ момент ν-го порядка конечен. В частности, математическое ожидание конечно при $n > 1$, а дисперсия — при $n > 2$. Несложное вычисление даёт

$$\mathsf{M}\, t = 0, \quad \mathsf{D}\, t = \frac{n}{n-2}. \tag{4.3}$$

Так как плотность распределения Стьюдента симметрична, то для квантилей $t_p(n)$ справедливо соотношение $t_p(n) = -t_{1-p}(n)$. При больших n ($n > 30$) для квантилей $t_p(n)$ выполнено приближённое равенство $t_p(n) \approx u_p$. Более точная формула имеет вид

$$t_p(n) \approx u_p \left(\left(1 - \frac{1}{4n}\right)^2 - \frac{u_p^2}{2n} \right)^{-1/2}. \tag{4.4}$$

Пример. Найдём квантили $t_{0,05}(8)$ и $t_{0,90}(40)$.

Из таблицы находим $t_{0,95}(8) = 1{,}860$. Используя равенство $t_p(n) = -t_{1-p}(n)$, получим $t_{0,05}(8) = -t_{0,95}(8) = 1{,}860$.

Квантиль $t_{0,90}(40)$ определим, используя формулу (4.4). Так как $u_{0,90} = 1{,}282$, то

$$t_{0,90}(40) \approx 1{,}282 \left(\left(1 - \frac{1}{4 \cdot 40}\right)^2 - \frac{1{,}282^2}{2 \cdot 40} \right)^{-1/2} = 1{,}30369 \approx 1{,}304.$$

Точное значение квантиля по таблице: $t_{0,90}(40) = 1{,}303$.

6.5. Распределение Фишера

Пусть $m+n$ случайных величин $\xi_1, \ldots, \xi_m$ и $\eta_1, \ldots, \eta_n$ независимы и нормальны $(0, \sigma)$. Положим

$$\xi = \frac{1}{m} \sum_{k=1}^{m} \xi_k^2, \quad \eta = \frac{1}{n} \sum_{k=1}^{n} \eta_k^2,$$

и рассмотрим величину

$$F(m,n) = \frac{\xi}{\eta} = \frac{\dfrac{1}{m} \displaystyle\sum_{k=1}^{m} \xi_k^2}{\dfrac{1}{n} \displaystyle\sum_{k=1}^{n} \eta_k^2}. \tag{5.1}$$

Обозначим через F_{nm} функцию распределения величины $F(m,n)$. Таким образом

$$F_{mn}(x) = \mathsf{P}\left\{ F(m,n) < x \right\} = \mathsf{P}\left\{ \xi/\eta < x \right\}.$$

Очевидно, что $F(m,n) \geq 0$ и $F_{nm}(x) = 0$ при $x \leq 0$.

Плотности распределений величин ξ и η равны $p_\xi(y) = \dfrac{m}{\sigma^2}\, k_m\left(\dfrac{my}{\sigma^2}\right)$ и $p_\eta(z) = \dfrac{n}{\sigma^2}\, k_n\left(\dfrac{nz}{\sigma^2}\right)$ (см. таблицу 2 строка). Воспользуемся формулой (1.8). Так как $p_\eta(z) = 0$ при $z \leq 0$, то

$$f_{mn}(x) = \int_0^\infty z p_\xi(zx)\, p_\eta(z)\, dz = a_{mn} \int_0^\infty z^{(m+n)/2-1} e^{-(mx+n)z/2\sigma^2}\, dz,$$

где

$$a_{mn} = \frac{(m/2)^{m/2}(n/2)^{n/2}}{\sigma^{m+n}\Gamma(m/2)\Gamma(n/2)}\, x^{m/2-1}.$$

Из известной формулы

$$\int_0^\infty x^{\beta-1} \mathrm{e}^{-Bx}\, dx = B^{-\beta}\Gamma(\beta), \quad B > 0,\ \beta > 0.$$

получаем

$$\int_0^\infty z^{(m+n)/2-1} e^{-(mx+n)z/2\sigma^2}\, dz = \frac{(2\sigma^2)^{(m+n)/2}}{(mx+n)^{(m+n)/2}}\,\Gamma((m+n)/2),$$

В результате

$$f_{mn}(x) = \frac{\Gamma((m+n)/2)(m/n)^{m/2}}{\Gamma(m/2)\Gamma(n/2)}\,\frac{x^{m/2-1}}{(1+mx/n)^{(m+n)/2}}. \tag{5.2}$$

Распределение с плотностью $f_{mn}(x)$ называется *распределением Фишера* или *F-распределением с n, m степенями свободы*. Математическое ожидание и дисперсия равны

$$\mathsf{M}F(m,n) = \frac{n}{n-2},\ n > 2, \quad \mathsf{D}F(m,n) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)},\ n > 4. \tag{5.3}$$

Квантили распределения Фишера порядка p и $1-p$ связаны следующим соотношением:

$$F_{1-p}(m,n) = \frac{1}{F_p(n,m)}. \tag{5.4}$$

При $m \gg 1$, $n \gg 1$ справедлива приближённая формула:

$$F_{1-p}(m,n) = \frac{n}{n-2}\sqrt{\frac{2(n+m-2)}{m(n-4)}}\,u_p + \frac{n}{n-2}. \tag{5.5}$$

Пример. Найдём квантили $F_{0{,}01}(3,5)$, $F_{0{,}05}(60,120)$,

Используя соотношение (5.4) и таблицу, получаем

$$F_{0{,}01}(3,5) = \frac{1}{F_{0{,}99}(5,3)} = \frac{1}{28{,}24} \approx 0{,}0354.$$

По формуле (5.5), используя значение $u_{0{,}05} = -u_{0{,}95} = -1{,}645$, получим

$$F_{0{,}05}(60,120) = \frac{120}{120-2}\sqrt{\frac{2(60+120-2)}{60(120-4)}}\,(-1{,}645) + \frac{120}{120-2} \approx 0{,}6386.$$

По таблице значение квантиля $F_{0{,}05}(60,120)$ равно

$$F_{0{,}05}(60,120) = \frac{1}{F_{0{,}95}(120,60)} = \frac{1}{1{,}43} \approx 0{,}699.$$