11. Нормальное распределение

Определение. Нормальным (или гауссовским) распределением называется непрерывное распределение с плотностью

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty$$

график этой плотности изображен на Рисунке 1.

С помощью известного интеграла Пуассона

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}\,,$$

проверяем условие нормировки:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = 1\,.$$

Математическое ожидание случайной величины ξ с нормальным распределением равно нулю,

$$E\xi = \int_{-\infty}^{\infty} x p(x)\,dx = 0,$$

поскольку функция x нечетна, а $p(x)$ четна: $p(-x) = p(x)$. Следовательно, дисперсия равна

$$D\xi = E\xi^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\,dx;$$

интеграл вычисляем по частям:

$$u = x, \qquad v = e^{-\frac{x^2}{2}}\,,$$

$$D\xi = \frac{1}{\sqrt{2\pi}} \left(-x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx \right) = 1.$$

Это распределение далее будет называться **стандартное нормальное** и символически обозначаться $\mathcal{N}(0,1)$. Соответствующую функцию распределения обозначим $F_0(x)$:

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du;$$

также ноликом будем отмечать плотность стандартного нормального:

$$p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Так как $p_0(x)$ симметрична относительно 0, $p_0(x) = p_0(-x)$, то $F_0(0) = \frac{1}{2}$ (медиана равна 0), также справедливо соотношение

$$F_0(-x) = 1 - F_0(x), \qquad x \geq 0.$$

Вероятности событий, связанных со случайной величиной $\xi \sim \mathcal{N}(0,1)$ (значок $\sim$ будет обозначать фразу: ”случайная величина … имеет распределение …”), вычисляются с помощью функции $F_0(x)$:

$$P\{c_1 < \xi \le c_2\} = F_0(c_2) - F_0(c_1), \tag{1}$$

$$P\{|\xi| \le \delta\} = P\{-\delta < \xi \le \delta\} = F_0(\delta) - F_0(-\delta) = 2F_0(\delta) - 1. \tag{2}$$

В тех задачах, где необходимо найти, при каком δ вероятность $P\{|\xi| \le \delta\}$ принимает заданное значение,

$$P\{|\xi| \le \delta\} = \alpha \qquad (0< \alpha < 1)$$

применяется обратная функция:

$$2F_0(\delta) - 1 = \alpha \Rightarrow \delta = F_0^{-1}\left(\frac{1+\alpha}{2}\right).$$

Пусть a – произвольное число, и $\sigma > 0$; сделаем линейное преобразование случайной величины $\xi \sim \mathcal{N}(0,1)$,

$$\eta = a + \sigma\xi,$$

тогда

$$p_\eta(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

$$F_\eta(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{(x-a)/\sigma} e^{-\frac{u^2}{2}} du = F_0\left(\frac{x-a}{\sigma}\right). \tag{3}$$

Это распределение называется нормальным распределением общего вида.

Его плотность симметрична относительно значения $x = a$, где она достигает максимума. Разброс значений случайной величины относительно точки a определяется величиной σ. Из свойств математического ожидания и дисперсии следует, $E\eta = a,\ D\eta = \sigma^2$; соответственно, обозначение нормального распределения общего вида будет иметь вид: $\eta \sim \mathcal{N}(a, \sigma^2)$.

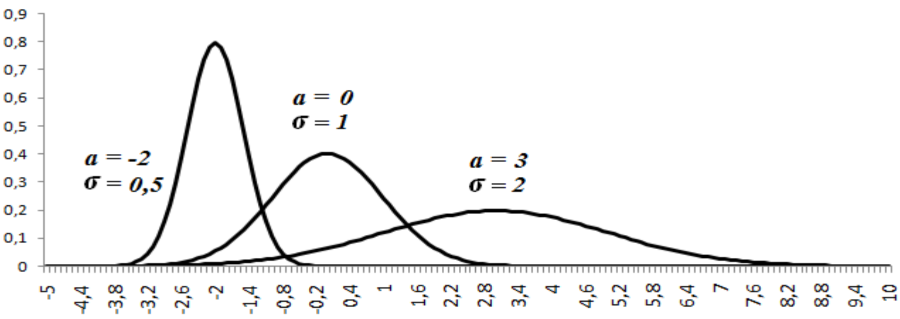

Рисунок 1. Плотность нормального распределения $\mathcal{N}(a, \sigma^2)$

На Рисунке 1показана зависимость плотности нормального распределения от параметров a, σ.

Для вычисления вероятности $P\{c_1 < \eta \le c_2\}$, связанной со случайной величиной η, имеющей распределение $\mathcal{N}(a, \sigma^2)$, сделаем преобразование

$$P\{c_1 < \eta \le c_2\} = P\left\{\frac{c_1 - a}{\sigma} < \frac{\eta - a}{\sigma} \le \frac{c_2 - a}{\sigma}\right\}$$

и аналогичным образом,

$$P\{|\eta - a| \leq \delta\} = P\left\{\left|\frac{\eta - a}{\sigma}\right| \leq \frac{\delta}{\sigma}\right\}.$$

Учитывая, что случайная величина $\frac{\eta - a}{\sigma}$ имеет стандартное нормальное распределение, получаем две формулы, обобщающие (1) и (2):

$$P\{c_1 < \eta \leq c_2\} = F_0\left(\frac{c_2 - a}{\sigma}\right) - F_0\left(\frac{c_1 - a}{\sigma}\right), \quad (4)$$

$$P\{|\eta - a| \leq \delta\} = 2F_0\left(\frac{\delta}{\sigma}\right) - 1. \quad (5)$$

Определение. Двумерным нормальным распределением называется совместное распределение двух случайных величин ξ, η, имеющее плотность

$$p_{\xi,\eta}(x, y) = \frac{exp\left\{-\frac{1}{2(1-r^{2)}}\left[\left(\frac{x - a_1}{\sigma_1}\right)^2 - 2r\frac{x - a_1}{\sigma_1}\frac{y - a_2}{\sigma_2} + \left(\frac{y - a_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} \quad (6)$$

здесь a_1, a_2 - заданные вещественные числа, $\sigma_1 \sigma_2$ положительные и $|r| < 1$.

Теорема 1. Если случайные величины ξ, η имеют двумерное нормальное распределение, то ξ и η каждая в отдельности имеет нормальное распределение, причем $a_1 = E\xi,\ a_2 = E\eta\ , \sigma_1^2 = D\xi,\ \ \sigma_2^2 = D\eta$ и $r = r(\xi, \eta)$ - коэффициент корреляции ξ, η.

Доказательство. Согласно основному свойству совместной плотности,

$$p_\eta(y) = \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y)dx.$$

Подставим сюда плотность двумерного гауссовского распределения, в которой очевидным образом сгруппируем слагаемые и выделим полный квадрат по переменной *x:*

$$\left[\left(\frac{x - a_1}{\sigma_1}\right)^2 - 2r\frac{x - a_1}{\sigma_1}\frac{y - a_2}{\sigma_2} + \left(\frac{y - a_2}{\sigma_2}\right)^2\right] =$$

$$= \left[\left(\frac{x - a_1}{\sigma_1} - r\frac{y - a_2}{\sigma_2}\right)^2 + \left(\frac{y - a_2}{\sigma_2}\right)^2(1 - r^2)\right],$$

а затем преобразуем интеграл:

$$p_\eta(y) = \frac{exp\left(-\frac{(y - a_2)^2}{2\sigma_2^2}\right)}{\sqrt{2\pi}\sigma_2}\int_{-\infty}^{\infty}\frac{exp\left\{-\frac{1}{2\sigma_1^2(1 - r^2)}\left[x - a_1 - r\frac{\sigma_1}{\sigma_2}(y - a_2)\right]^2\right\}}{\sqrt{2\pi}\sigma_1\sqrt{1 - r^2}}dx.$$

Интеграл справа представляет собой интеграл по всей вещественной оси от некоторой гауссовской плотности, поэтому он равен 1 (условие нормировки). Оставшееся выражение и есть плотность распределения $\mathcal{N}(a_2, \sigma_2^2)$. Точно также выводится распределение для ξ, таким образом, первая часть Теоремы, касающаяся математических ожиданий и дисперсий, доказана.

Вычислим коэффициент корреляции,

$$r(\xi,\eta)=\frac{E\xi\eta-E\xi E\eta}{\sqrt{D\xi D\eta}}=\frac{E\xi\eta-a_1a_2}{\sigma_1\sigma_2}; \qquad (4)$$

$$E\xi\eta=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}xyp_{\xi,\eta}(x,y)dxdy=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int\limits_{-\infty}^{\infty}y\cdot exp\left(-\frac{(y-a_2)^2}{2\sigma_2^2}\right)dy*$$

$$*\int\limits_{-\infty}^{\infty}x\cdot exp\left\{-\frac{1}{2\sigma_1^2(1-r^2)}\left[x-a_1-r\frac{\sigma_1}{\sigma_2}(y-a_2)\right]^2\right\}dx,$$

здесь использовано полученное выше представление совместной плотности. Интеграл по переменной x преобразуем следующим образом:

$$\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-r^2}}\int\limits_{-\infty}^{\infty}x\cdot exp\left\{-\frac{1}{2\sigma_1^2(1-r^2)}\left[x-\left(a_1+r\frac{\sigma_1}{\sigma_2}(y-a_2)\right)\right]^2\right\}dx=$$

$$=a_1+r\tfrac{\sigma_1}{\sigma_2}(y-a_2),$$

тогда

$$E\xi\eta=\frac{1}{\sqrt{2\pi}\sigma_2}\int\limits_{-\infty}^{\infty}y\left[a_1+r\frac{\sigma_1}{\sigma_2}(y-a_2)\right]exp\left(-\frac{(y-a_2)^2}{2\sigma_2^2}\right)dy=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2}\int\limits_{-\infty}^{\infty}[(y-a_2)+a_2]\left[a_1+r\frac{\sigma_1}{\sigma_2}(y-a_2)\right]exp\left(-\frac{(y-a_2)^2}{2\sigma_2^2}\right)dy=$$

$$=a_1a_2+\ r\frac{\sigma_1\sigma_2^2}{\sigma_2}=a_1a_2+\ r\sigma_1\sigma_2,$$

и подставляя в (16.2), убеждаемся, что действительно $r(\xi,\eta)=r$. ■

Отметим одно свойство нормального распределения (как будет видно в будущем - очень важное свойство): если ξ,η имеют двумерное нормальное распределение и они некоррелированы, то ξи η независимы. Это прямо следует из формулы (6), поскольку при $r = 0$ совместная плотность превращается в произведение двух плотностей:

$$p_{\xi,\eta}(x,y)=\frac{exp\left\{-\frac{1}{2}\left[\left(\frac{x-a_1}{\sigma_1}\right)^2+\left(\frac{y-a_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2}=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-a_1)^2}{2\sigma_1^2}}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-a_2)^2}{2\sigma_2^2}}=p_\xi(x)p_\eta(x).$$