

Лекция 6

05.03.2012

Пусть  $X$  и  $Y$  – случайные векторы.

**Преобразование векторов**

Пусть  $X \in R^n$  – случайный вектор с плотностью распределения  $f(x)$ ,  $x \in R^n$ ,  $G: R^n \rightarrow R^n$  – обратимое отображение, и пусть  $Y = G(X)$  и  $y = G(x)$ . Тогда если  $dV_x$  – малая окрестность точки  $x$  и  $dV_y = G(dV_x)$ , то  $\Pr(Y \in dV_y) = \Pr(X \in dV_x)$ , т.е.  $f_Y(y) |dV_y| = f_X(x) |dV_x|$ . Но

$|dV_y| = |G'(x)| |dV_x|$ , где  $|G'(x)|$  – якобиан отображения  $G$ . Таким образом,  $f_Y(y) = |G'(x)|^{-1} f_X(x)$ .

В частности, если  $G(x) = Ax + b$ , где  $A$  –  $n \times n$  невырожденная матрица,  $b$  –  $n \times 1$  вектор, то

$$G'(x) = A \text{ и } f_Y(y) = |A^{-1}| f_X(A^{-1}(y - b)). \quad (1)$$

Для  $n = 1$  получаем формулу  $f_Y(y) = f_X(G^{-1}(y)) / |G'(G^{-1}(y))|$

**Многомерное нормальное распределение**

Пусть  $Z_1, \dots, Z_n$  – независимые стандартные нормальные случайные величины.

**Определение.** Вектор  $\varepsilon = [Z_1, \dots, Z_n]'$  называется стандартным нормальным  $n$ -мерным вектором,  $\varepsilon \sim N(0, I)$ .

Нетрудно проверить, что

$$f_\varepsilon(x) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{x'x}{2}\right), \quad x \in R^n,$$
$$E(\varepsilon) = 0_n, \quad V(\varepsilon) = I_n$$

Пусть  $\varepsilon = [Z_1, \dots, Z_n]'$  – стандартный нормальный  $n$ -мерный вектор,  $A$  –  $m \times n$  матрица,  $b$  –  $m \times 1$  вектор.

**Определение.** Вектор  $X = A\varepsilon + b$  называется нормальным  $m$ -мерным вектором.

Нетрудно проверить, что

$$E(X) = b, \quad V(X) = AA' \equiv \Sigma.$$

Обозначение:  $X \sim N(b, \Sigma)$ .

Пусть  $m = n$  и матрица  $A$  является невырожденной, и пусть  $X = A\varepsilon + b \sim N(b, \Sigma)$ .

**Предложение.** Плотность распределения случайного  $n$ -мерного вектора  $X$  есть

$$f(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - b)' \Sigma^{-1} (x - b)\right),$$

**Свойства нормального распределения**

1. Если  $X \sim N(b, \Sigma)$ , то  $E(X) = b$ ,  $V(X) = \Sigma$  – доказано ранее.

2. Пусть  $X \sim N(b, \Sigma_X)$ ,  $X \in R^n$  – нормальный случайный вектор,  $B - m \times n$  матрица,  $d - m \times 1$  вектор. Вектор  $Y = BX + d$  является нормальным. Действительно, по определению  $X = A\varepsilon + b$ , где  $A - n \times n$  матрица,  $b - n \times 1$  вектор. Тогда

$$Y = BX + d = B(A\varepsilon + b) + d = (BA)\varepsilon + (Bb + d)$$

– нормальный вектор. При этом  $E(Y) = Bb + d$ ,  $V(Y) = B\Sigma_X B'$ . В частности, линейная комбинация компонент нормального вектора есть нормальная случайная величина.

3. Любой подвектор нормального вектора тоже нормален. Действительно, пусть, например,

$X \sim N(b, \Sigma)$  и  $X = \begin{bmatrix} U \\ W \end{bmatrix}$ , где  $U - k \times 1$  вектор,  $W - (n - k) \times 1$  вектор. Тогда  $U = BX$ , где

$B = [I_k; O_{n-k}]$ . В силу п. 2  $U$  – нормальный вектор.

4. Если  $X$  и  $Y$  – независимые нормальные векторы, то  $U = [X', Y']'$  – нормальный вектор.

Действительно, пусть для простоты средние равны нулю. Тогда  $X = A\varepsilon_1$ ,  $Y = B\varepsilon_2$  и

$$U = \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} A & O \\ O & B \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = C\varepsilon$$

5. Если  $U = [X', Y']'$  – нормальный вектор и векторы  $X, Y$  некоррелированные (т.е.

$\Sigma_{XY} = \text{Cov}(X, Y) = 0$ ), то векторы  $X, Y$  независимы.

6. Пусть  $U = (X', Y')' \sim N\left(\begin{pmatrix} m_X \\ m_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right)$ . Тогда условная плотность  $f_{Y|X}(y|x)$  является

плотностью нормального распределения с параметрами

$$m_{Y|X}(x) = E(Y | X = x) = \Sigma_{YX} \Sigma_{XX}^{-1} (x - m_X) + m_Y,$$

$$\Sigma_{Y|X}(x) = E((Y - E(Y | X))(Y - E(Y | X))' | X = x) = \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \quad (1)$$

(не зависит от  $x$ ). Иными словами, условное распределение  $Y | X = x$  является нормальным с параметрами (1).

В частности, при  $m = n = 1$  имеем

$$E(Y | X = x) = m_Y + \frac{\text{Cov}(X, Y)}{V(X)}(x - m_X), \quad V(Y | X = x) = V(Y) - \frac{\text{Cov}^2(X, Y)}{V(X)}$$

Равенство  $E(Y | X = x) = \Sigma_{YX} \Sigma_{XX}^{-1} (x - m_X) + m_Y$  можно доказать так. Считая для простоты, что

$m_X = 0, m_Y = 0$ , имеем  $\text{Cov}(Y - \Sigma_{YX} \Sigma_{XX}^{-1} X, X) = \Sigma_{YX} - \Sigma_{YX} = 0$ , т.е.  $(Y - \Sigma_{YX} \Sigma_{XX}^{-1} X)$  и  $X$

некоррелированы и по свойству 3 независимы. Значит,

$$0 = E(Y - \Sigma_{YX} \Sigma_{XX}^{-1} X | X) = E(Y | X) - \Sigma_{YX} \Sigma_{XX}^{-1} E(X | X) = E(Y | X) - \Sigma_{YX} \Sigma_{XX}^{-1} X \Rightarrow$$

$$E(Y | X) = \Sigma_{YX} \Sigma_{XX}^{-1} X$$

Обобщение на ненулевые средние тривиально.

Вывод: для нормальных векторов функция регрессии линейная (аффинная).

### Распределения, связанные с нормальным распределением

Пусть  $Z_1, \dots, Z_n \sim N(0, 1)$  и независимы  $\Leftrightarrow Z = [Z_1, \dots, Z_n]' \sim N(0, I_n)$ .

**Определение.** Распределение случайной величины  $\chi^2(n) = \sum_{i=1}^n Z_i^2 = Z'Z$  называется хи-

квадрат распределением с  $n$  степенями свободы (degrees of freedom).

Нетрудно проверить, что  $E(\chi^2(n)) = n$ ,  $V(\chi^2(n)) = 2n$ .

Пусть  $Z_0 \sim N(0, 1)$ , и  $Z_0$  не зависит от  $Z$ .

**Определение.** Распределение случайной величины  $t(n) = \frac{Z_0}{\sqrt{1/n \sum_{i=1}^n Z_i^2}} = \frac{N(0,1)}{\sqrt{1/n \chi^2(n)}}$

называется  $t$ -распределением (распределением Стьюдента) с  $n$  степенями свободы.

Пусть  $U_1, \dots, U_m \sim N(0,1)$ , независимы (т.е.  $U = [U_1, \dots, U_m]' \sim N(0, I_m)$ ) и не зависят от  $Z$ .