

**RACHUNEK
PRAWDOPODOBIENSTWA
WYKŁAD 7.**

**ZMIENNA LOSOWA
DWUWYMIAROWA
(WIELOWYMIAROWA)**

Jeśli X_i $i = 1, 2, \dots, n$ są zmiennymi losowymi w ustalonej przestrzeni probabilistycznej $(\Omega, \mathcal{S}, P)$ to ciąg $X = (X_1, X_2, \dots, X_n)$ nazywamy **zmienną losową n -wymiarową (wektorem losowym)**.

Zauważmy, że w tym przypadku każdemu zdarzeniu elementarnemu przyporządkujemy ciąg n liczb rzeczywistych.

$$X : \Omega \rightarrow R^n \quad (\text{funkcja borelowska})$$

W szczególności gdy $n = 2$ mamy **dwuwymiarową zmienną losową** (X, Y) .

Zmienne losowe wielowymiarowe służą do modelowania takich doświadczeń losowych, których wyniki opisuje się układem wielu liczb rzeczywistych np. losowo wybranego człowieka możemy m.in. scharakteryzować trzema liczbami: wzrostem, wagą i wiekiem.

Zmienna losowa 2-wymiarowa.

$$P_{(X,Y)} : \mathcal{B}(R^2) \rightarrow [0, 1]$$

- rozkład prawdopodobieństwa zmiennej losowej (X,Y) .

$$P_{(X,Y)}(A) = P\left((X,Y)^{-1}(A)\right), \quad \forall A \in \mathcal{B}(R^2)$$

Rozkłady prawdopodobieństwa zmiennych losowych X , Y nazywamy **rozkładami brzegowymi**.

Rozkład prawdopodobieństwa zmiennej losowej (X, Y) nazywamy **rozkładem łącznym**.

Dystrybuanta

$$F(x, y) = P(X < x, Y < y)$$

Własności dystrybuanty zmiennej losowej (X, Y) .

a) F jest niemalejąca względem każdego argumentu,

$$\text{b) } \forall x \left(\lim_{y \rightarrow -\infty} F(x, y) = 0 \right); \quad \forall y \left(\lim_{x \rightarrow -\infty} F(x, y) = 0 \right);$$
$$\left(\lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} F(x, y) = 1 \right),$$

c) F jest lewostronnie ciągła względem każdego argumentu,

$$\forall x_1 \leq x_2; \forall y_1 \leq y_2;$$
$$\text{d) } F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) =$$
$$= P(x_1 \leq X < x_2; y_1 \leq Y < y_2) \geq 0$$

Jeśli $F(x, y)$ jest dystrybuantą zmiennej losowej (X, Y) to funkcje

$$F_X(x) = \lim_{y \rightarrow \infty} F(x, y) = F(x, \infty);$$

$$F_Y(y) = \lim_{x \rightarrow \infty} F(x, y) = F(\infty, y)$$

są dystrybuantami odpowiednich rozkładów brzegowych.

Zmienne losowe X, Y są **niezależne** gdy dla dowolnych zbiorów borelowskich A, B na prostej mamy

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Zmienne losowe X, Y są **niezależne** wtedy i tylko wtedy, gdy dla dowolnych x, y rzeczywistych

$$F(x, y) = F_X(x)F_Y(y)$$

Dwuwymiarowa zmienna losowa (X, Y) ma **rozkład skokowy** jeśli zmienne losowe X i Y mają skończony lub przeliczalny zbiór wartości.

Rozkład zmiennej losowej (X, Y) (łączny rozkład zmiennych X i Y) określa się za pomocą funkcji prawdopodobieństwa lub dystrybuanty.

Funkcją prawdopodobieństwa skokowej zmiennej losowej (X, Y) przyjmującej wartości (x_i, y_j) jest

$$p_{ij} = P(X = x_i, Y = y_j) \quad i, j = 1, 2, \dots$$

przy czym $p_{ij} \geq 0$ oraz $\sum_i \sum_j p_{ij} = 1$

Dystrybuantą $F(x, y)$ skokowej zmiennej losowej (X, Y) jest funkcja rzeczywista

$$F(x, y) = \sum_{x_i < x} \sum_{y_j < y} p_{ij}$$

Funkcję prawdopodobieństwa skokowej zmiennej losowej (X, Y) przyjmującej wartości (x_i, y_j) można zapisać w postaci tablicy:

$X \backslash Y$	y_1	y_2		y_l	$p_{i\cdot}$
x_1	p_{11}	p_{12}		p_{1l}	$p_{1\cdot}$
x_2	p_{21}	p_{22}		p_{2l}	$p_{2\cdot}$
....					
x_k	p_{k1}	p_{k2}		p_{kl}	$p_{k\cdot}$
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$		$p_{\cdot l}$	1

gdzie

$x_1, x_2, \dots, x_k$ – wartości zmiennej losowej X ,

$y_1, y_2, \dots, y_l$ – wartości zmiennej losowej Y ,

$p_{\cdot j}$ – sumy prawdopodobieństw w kolumnach,

$$p_{\cdot j} = \sum_i p_{ij}$$

$p_{i\cdot}$ – sumy prawdopodobieństw w wierszach,

$$p_{i\cdot} = \sum_j p_{ij} .$$

Uwaga. $\sum_{i,j} p_{ij} = 1.$

Rozkładem brzegowym zmiennej losowej X nazywamy rozkład określony funkcją prawdopodobieństwa:

x_i	x_1	x_2	$\dots$	x_k
$p_{i\cdot}$	$p_{1\cdot}$	$p_{2\cdot}$	$\dots$	$p_{k\cdot}$

Rozkładem brzegowym zmiennej losowej Y nazywamy rozkład określony funkcją prawdopodobieństwa:

y_i	y_1	y_2		y_l
$p_{\cdot i}$	$p_{\cdot 1}$	$p_{\cdot 2}$		$p_{\cdot l}$

Jeśli zmienna losowa (X, Y) jest skokowa to zmienne losowe X i Y są **niezależne** gdy **dla każdej pary** (x_i, y_j) ($i, j = 1, 2, \dots$) spełniony jest warunek:

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

Warunek ten można również zapisać w postaci

$$p_{ij} = p_{i.} \cdot p_{.j}$$

Przykład.

Rzucamy dwa razy kostką.

X - liczba parzystych oczek w pierwszym rzucie, tzn. $X = 0$ lub $X = 1$.

Y - liczba jedynek w obu rzutach, tzn. $Y = 0$ lub $Y = 1$, lub $Y = 2$.

Funkcja rozkładu prawdopodobieństwa tej zmiennej losowej dana jest tabelką:

$X \backslash Y$	0	1	2	$p_{i\cdot}$
0	10/36	7/36	1/36	18/36
1	15/36	3/36	0	18/36
$p_{\cdot j}$	25/36	10/36	1/36	1

Rozkłady brzegowe wyznaczone są przez brzegowe wartości tej tabeli.

Rozkład brzegowy zmiennej losowej X :

x_i	0	1
$p_{i\cdot}$	$18/36$	$18/36$

Rozkład brzegowy zmiennej losowej Y :

y_j	0	1	2
$p_{\cdot j}$	$25/36$	$10/36$	$1/36$

Przykład.

Funkcje rozkładu prawdopodobieństwa dane tabelkami:

$X \backslash Y$	0	1	2	$p_{i\cdot}$
0	$1/8$	0	$3/8$	$4/8$
1	$1/4$	$1/4$	0	$4/8$
$p_{\cdot j}$	$3/8$	$1/4$	$3/8$	1

$X \backslash Y$	0	1	2	$p_{i\cdot}$
0	$3/8$	0	$1/8$	$4/8$
1	0	$1/4$	$1/4$	$4/8$
$p_{\cdot j}$	$3/8$	$1/4$	$3/8$	1

mają identyczne rozkłady brzegowe.

Wniosek.

Na ogół rozkłady brzegowe nie wyznaczają rozkładu łącznego jednoznacznie.

W przypadku zmiennych losowych niezależnych rozkłady brzegowe wyznaczają rozkład łączny jednoznacznie.

(X, Y) nazywamy **zmienną losową ciągłą** jeśli jej dystrybuanta da się przedstawić w postaci

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt$$

dla pewnej nieujemnej funkcji f zwanej **gęstością**.

Uwaga.

1. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1, \quad f(x, y) \geq 0$

2. W punktach ciągłości funkcji f zachodzi:

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$$

3. Dla $A \in \mathcal{B}(R^2)$ mamy

$$P_{(X, Y)}(A) = \iint_A f(x, y) dx dy.$$

Mając gęstość rozkładu łącznego gęstości rozkładów brzegowych wyznaczamy następująco

Jeśli $f(x, y)$ jest gęstością zmiennej losowej (X, Y) to funkcje

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

są gęstościami odpowiednich rozkładów brzegowych.

Jeśli łączny rozkład (X, Y) jest ciągły, to zmienne losowe X, Y są **niezależne** wtedy i tylko wtedy, gdy dla dowolnych x, y rzeczywistych

$$f(x, y) = f_X(x)f_Y(y)$$

Przykład.

Funkcja $f(x, y)$ jest gęstością zmiennej losowej (X, Y) .

$$f(x, y) = \begin{cases} c & \text{dla } 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & \text{dla innych } x, y \end{cases}$$

Przez całkowanie lub z interpretacji geometrycznej wynika, że $c = 0,25$ (bo pole rozpatrywanego kwadratu wynosi 4).

Przez całkowanie lub z interpretacji geometrycznej wynika, że dystrybuanta tego rozkładu ma postać

$$F(x, y) = \begin{cases} 0 & x \leq 0 \vee y \leq 0 \\ 0,25xy & 0 < x \leq 2, 0 < y \leq 2 \\ 0,5x & 0 < x \leq 2, y > 2 \\ 0,5y & 0 < y \leq 2, x > 2 \\ 1 & x > 2, y > 2 \end{cases}$$

Rozkłady brzegowe to rozkłady jednostajne na przedziale $[0, 2]$.

Zauważmy, że zmienne losowe X, Y są niezależne.

Informacja o zmiennych losowych n - wymiarowych.

$(\Omega, \mathcal{S}, P)$ - ustalona przestrzeń probabilistyczna.

$X = (X_1, X_2, \dots, X_n)$ - zmienna losowa n - wymiarowa (wektor losowy).

Dystrybuanta

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n)$$

X nazywamy **zmienną losową skokową**, jeśli jej zbiór wartości jest skończony lub przeliczalny.

X nazywamy **zmienną losową ciągłą** jeśli jej dystrybuanta da się przedstawić w postaci

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_n} f(u_1, \dots, u_n) du_1 \dots du_n$$

dla pewnej nieujemnej funkcji f zwanej gęstością.

Uwaga.

$$1. \int \int \dots \int_{R^n} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$$

2. W punktach ciągłości funkcji f zachodzi:

$$\frac{\partial^{(n)} F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} = f(x_1, \dots, x_n)$$

3. Dla $A \in \mathcal{B}(R^n)$ mamy

$$P_X(A) = \int \int \dots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Funkcja charakterystyczna zmiennej losowej n - wymiarowej.

$$\varphi(t) = \varphi(t_1, \dots, t_n) = E(e^{itX}) = E(\exp(i(t_1 X_1 + \dots + t_n X_n))).$$

Funkcja charakterystyczna zmiennej losowej 2 - wymiarowej.

$$\varphi(t) = \varphi(t_1, t_2) = E(e^{itX}) = E(\exp(i(t_1 X_1 + t_2 X_2))).$$

Rozkłady warunkowe. (numery zmiennych umowne).

Jeśli $P_{1,...,k}(X_1 = x_{1j}, ..., X_k = x_{kj}) > 0$ to rozkład zmiennej losowej skokowej (n - k)

wymiarowej określonej wzorem:

$$P(X_{k+1} = x_{k+1,j}, ..., X_n = x_{nj} | X_1 = x_{1j}, ..., X_k = x_{kj}) = \frac{P(X_1 = x_{1j}, ..., X_n = x_{nj})}{P_{1,...,k}(X_1 = x_{1j}, ..., X_k = x_{kj})}$$

nazywamy rozkładem warunkowym zmiennej losowej $(X_{k+1}, ..., X_n)$ pod warunkiem, że $(X_1 = x_{1j}, ..., X_k = x_{kj})$.

Rozkłady warunkowe dla $n = 2$.

Zmienna losowa skokowa.

Rozkład warunkowy Y pod warunkiem $X = x_i$

$$P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i\bullet}}$$

$j = 1, 2, \dots, l$ dla $p_{i\bullet} > 0$

Rozkład warunkowy X pod warunkiem $Y = y_j$

$$P(X = x_i | Y = y_j) = \frac{p_{ij}}{p_{\bullet j}}$$

$$i = 1, 2, \dots, k \quad \text{dla } p_{\bullet j} > 0$$

Przykład

Funkcja rozkładu prawdopodobieństwa zmiennej losowej dwuwymiarowej (X, Y) dana jest tabelką:

$X \backslash Y$	- 1	1	$p_{i\cdot}$
- 1	1/6	1/6	1/3
0	1/3	0	1/3
1	1/6	1/6	1/3
$p_{\cdot j}$	2/3	1/3	1

rozkład warunkowy zmiennej losowej X pod warunkiem, że $Y = 1$ jest określony przez funkcję prawdopodobieństwa

$$P(X = -1|Y = 1) = (1/6)/(1/3) = 1/2,$$

$$P(X = 0|Y = 1) = 0/(1/3) = 0,$$

$$P(X = 1|Y = 1) = (1/6)/(1/3) = 1/2,$$

Jeśli gęstość $f_{1,...,k} > 0$ to rozkład zmiennej losowej ciągłej $(n - k)$ wymiarowej określonej wzorem:

$$f(x_{k+1}, \dots, x_n \mid x_1, \dots, x_k) = \frac{f(x_1, \dots, x_n)}{f(x_1, \dots, x_k)}$$

nazywamy rozkładem warunkowym zmiennej losowej $(X_{k+1}, \dots, X_n)$ pod warunkiem, że $(X_1 = x_1, \dots, X_k = x_k)$.

Zmienna losowa ciągła.

Gęstość rozkładu warunkowego Y pod warunkiem $X = x_0$

$$f(y | X = x_0) = \frac{f(x_0, y)}{f_X(x_0)}$$

dla $f_X(x_0) > 0$

Gęstość rozkładu warunkowego X pod warunkiem $Y = y_0$

$$f(x|Y = y_0) = \frac{f(x, y_0)}{f_Y(y_0)}$$

dla $f_Y(y_0) > 0$

Przykład.

Funkcja $f(x, y)$ jest gęstością zmiennej losowej (X, Y) .

$$f(x, y) = \begin{cases} 0,25 & \text{dla } 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & \text{dla innych } x, y \end{cases}$$

gęstość rozkładu warunkowego $X|Y = 1$ ma dla $0 < x < 2$ postać $0,25/0,5 = 0,5$; zatem

$$f(x|Y=1) = \begin{cases} 0 & x \notin (0,2) \\ 0,5 & x \in (0,2) \end{cases}$$

Niezależność zmiennych losowych n - wymiarowych.

Zmienne losowe $X_1, X_2, \dots, X_n$ są niezależne jeśli

$$F(x_1, \dots, x_n) = F_1(x_1) \cdot F_2(x_2) \dots \cdot F_n(x_n)$$

dla dowolnych $x_1, x_2, \dots, x_n \in \mathbb{R}^n$.

gdzie F_i - dystrybuanty rozkładów brzegowych jednowymiarowych.

Dla zmiennych losowych skokowych
odpowiedni warunek ma postać:

$$P(X_1 = x_{1j}, \dots, X_n = x_{nj}) = P_1(X_1 = x_{1j}) \cdot \dots \cdot P_n(X_n = x_{nj})$$

dla dowolnych $x_{1j}, \dots, x_{nj} \in R^n$

Dla zmiennych losowych ciągłych odpowiedni warunek ma postać:

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \dots \cdot f_n(x_n)$$

dla dowolnych $x_1, x_2, \dots, x_n \in \mathbb{R}^n$.

Jeśli zmienne losowe $X_1, X_2, \dots, X_n$ są niezależne to funkcje od nich też są niezależne.

WYBRANE PARAMETRY ZMIENNEJ LOSOWEJ DWUWYMIAROWEJ

Kowariancją zmiennych losowych (X, Y) nazywamy wielkość

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - EX)(Y - EY)] = \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

Dla zmiennej losowej skokowej (X, Y) mamy:

$$\text{Cov}(X, Y) = \sum_{i=1}^k \sum_{j=1}^l (x_i - EX)(y_j - EY) p_{ij} \quad \text{- definicja}$$

$$E(XY) = \sum_{i=1}^k \sum_{j=1}^l x_i y_j p_{ij}$$

$$\text{Cov}(X, Y) = \sum_{i=1}^k \sum_{j=1}^l x_i y_j p_{ij} - EX \cdot EY \quad \text{- własność}$$

Dla zmiennej losowej ciągłej (X, Y) mamy:

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY) f(x, y) dx dy$$

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy$$

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - EX \cdot EY$$

Uwaga

- a) Dla zmiennych losowych niezależnych $\text{Cov}(X, Y) = 0$, zatem **zmienne losowe niezależne są nieskorelowane** (odwrotna własność nie zachodzi – patrz przykład),
- b) $\text{Cov}(X, X) = D^2X$,
- c) $D^2(X + Y) = D^2X + D^2Y + 2\text{Cov}(X, Y)$,
 X, Y – **dowolne** zmienne losowe

Unormowaną kowariancję nazywamy **współczynnikiem korelacji** między zmiennymi X i Y :

$$\rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{(DX) \cdot (DY)}$$

Współczynnik korelacji mierzy „siłę” zależności liniowej między zmiennymi X i Y .

Własności współczynnika korelacji:

- a) $-1 \leq \rho(X, Y) \leq 1$
- b) dla niezależnych zmiennych losowych współczynnik korelacji jest równy zero,
- c) jeżeli współczynnik korelacji jest dodatni, to między zmiennymi X i Y istnieje zależność liniowa dodatnia, co oznacza, że ze wzrostem wartości jednej zmiennej rosną średnie wartości drugiej zmiennej,
- d) jeżeli współczynnik korelacji jest ujemny, to między zmiennymi X i Y istnieje zależność liniowa ujemna, co oznacza, że ze wzrostem wartości jednej zmiennej maleją średnie wartości drugiej zmiennej,
- e) jeżeli współczynnik korelacji jest równy 1 lub -1 , to między zmiennymi X i Y istnieje **funkcyjna** zależność liniowa,

Jeżeli współczynnik korelacji jest równy 0 to mówimy, że zmienne losowe X i Y są **nieskorelowane**.

Macierz

$$K = \begin{bmatrix} D^2 X & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & D^2 Y \end{bmatrix}$$

nazywamy **macierzą kowariancji**

Macierz $R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$
nazywamy **macierzą korelacji**

Przykład

Funkcja rozkładu prawdopodobieństwa
zmiennej losowej dwuwymiarowej
(X, Y) dana jest tabelką:

$X \backslash Y$	- 1	1	$p_{i\cdot}$
- 1	1/6	1/6	1/3
0	1/3	0	1/3
1	1/6	1/6	1/3
$p_{\cdot j}$	2/3	1/3	1

Obliczymy współczynnik korelacji między
tymi zmiennymi.

Rozkład brzegowy zmiennej losowej X :

x_i	-1	1
$p_{i.}$	$1/3$	$1/3$

$$EX = 0$$

Rozkład brzegowy zmiennej losowej Y :

y_j	-1	1
$p_{.j}$	$2/3$	$1/3$

$$EY = -1/3$$

Ponieważ

$$E(XY) = (-1) \cdot (-1) \cdot 1/6 + (-1) \cdot 1 \cdot 1/6 + 1 \cdot (-1) \cdot 1/6 + 1 \cdot 1 \cdot 1/6 = 0,$$

$$EX \cdot EY = 0; \quad \text{Cov}(X, Y) = 0 \quad \text{to} \quad \rho = 0.$$

Zatem zmienne X, Y są nieskorelowane.

Uwaga. Zauważmy, że powyższe zmienne losowe chociaż są zależne to są nieskorelowane.

Zakładamy, że macierz kowariancji K istnieje.

Regresja I rodzaju Y względem X = zbiór punktów $(x, E(Y|x))$.

Regresja I rodzaju X względem Y = zbiór punktów $(E(X|y), y)$.

Gdzie $E(Y|x)$, $E(X|y)$ to warunkowe wartości oczekiwane.

Linie regresji I rodzaju tylko w szczególnych przypadkach są liniami prostymi.

Twierdzenie.

$E((Y - \varphi(X))^2)$ osiąga wartość najmniejszą gdy $\varphi(x) = E(Y | x)$ z prawdopodobieństwem 1.

Jeśli poszukujemy funkcji liniowej minimalizującej wyrażenie

$$E((Y - \varphi(X))^2)$$

to otrzymamy prostą regresji zwaną prostą regresji II rodzaju.

Regresja II rodzaju Y względem X = prosta

$$y = \rho \frac{\sigma_Y}{\sigma_X} x + m_Y - \rho \frac{\sigma_Y}{\sigma_X} m_X .$$

tzn.

$$y = \beta_0 + \beta_1 x$$

gdzie

$$\beta_1 = \rho \frac{\sigma_Y}{\sigma_X}$$

$$\beta_0 = m_Y - \beta_1 m_X$$

Regresja II rodzaju X względem Y = prosta

$$x = \rho \frac{\sigma_X}{\sigma_Y} y + m_X - \rho \frac{\sigma_X}{\sigma_Y} m_Y .$$

Powyższe pojęcia regresji można uogólnić na przypadek n - wymiarowych zmiennych losowych.

W szczególności hiperpłaszczyzna regresji II rodzaju Zmiennej X_1 względem zmiennych $X_2, X_3, \dots, X_n$ ma równanie

$$x_1 - EX_1 = a_{12}(x_2 - EX_2) + \dots + a_{1n}(x_n - EX_n)$$

$$a_{1i} = -\frac{K_{1i}}{K_{11}}$$

gdzie K_{1i} są dopełnieniami algebraicznymi elementów k_{1i} macierzy kowariancji K .

Parametry zmiennej losowej
n – wymiarowej (mogą nie istnieć).

Wartość oczekiwana $E(X) = [EX_1, EX_2, \dots, EX_n]$.

Wariancja

$$D^2(X) = [D^2 X_1, D^2 X_2, \dots, D^2 X_n]$$

Moment (zwyuczajny) rzędu $l_1 + l_2 + \dots + l_n$

$$m_{l_1 l_2 \dots l_n} = E(X_1^{l_1} X_2^{l_2} \dots X_n^{l_n}),$$

Np. dla (X, Y)

$$m_{11} = E(X \cdot Y),$$

$$m_{10} = E(X), \quad m_{20} = E(X^2)$$

$$m_{01} = E(Y), \quad m_{02} = E(Y^2)$$

$$D^2 X = m_{20} - m_{10}^2, \quad D^2 Y = m_{02} - m_{01}^2$$

$$\rho = \frac{m_{11} - m_{10}m_{01}}{\sqrt{m_{20} - m_{10}^2} \sqrt{m_{02} - m_{01}^2}}$$

Moment centralny rzędu $l_1 + l_2 + \dots + l_n$

$$\mu_{l_1 l_2 \dots l_n} = E \left((X_1 - EX_1)^{l_1} \dots (X_n - EX_n)^{l_n} \right),$$

Np.

$$\mu_{11} = Cov(X, Y)$$

Macierz kowariancji $K = [k_{ij}]$, gdzie

$$k_{ij} = \text{cov}(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)] = \\ = E(X_i X_j) - E(X_i)E(X_j)$$

Uwaga $k_{ii} = D^2 X_i$, jest wariancją i - tej składowej.

Macierz K jest kwadratowa, symetryczna i słabo dodatnio określona (w szczególności ma wyznacznik nieujemny).

Macierz korelacji

$$R = [\rho_{ij}], \quad \text{gdzie } \rho_{ij} = \frac{\text{cov}(X_i, X_j)}{DX_i \cdot DX_j} =$$

Uwaga $\rho_{ii} = 1$.

FUNKCJE ZMIENNYCH LOSOWYCH 2 WYMIAROWYCH.

(X_1, X_2) - dana zmienna losowa ciągła o gęstości f .

$Y = g(X_1, X_2)$ g - borelowska,
Dystrybuanta tej zmiennej losowej ma postać

$$G(y) = \iint_{(g(x_1, x_2) < y)} f(x_1, x_2) dx_1 dx_2$$

gęstość $g(y)$ wyznaczamy przez różniczkowanie.

Przykład.

$$Y = X_1 \cdot X_2 ,$$

$$G(y) = \iint_{x_1 \cdot x_2 < y} f(x_1, x_2) dx_1 dx_2 =$$

$$= \int_{-\infty}^0 \left(\int_{y/x_1}^{\infty} f(x_1, x_2) dx_2 \right) dx_1 + \int_0^{\infty} \left(\int_{-\infty}^{y/x_1} f(x_1, x_2) dx_2 \right) dx_1$$

wtedy
$$g(y) = \int_{-\infty}^0 \frac{1}{x} f\left(x, \frac{y}{x}\right) dx + \int_0^{\infty} \frac{1}{x} f\left(x, \frac{y}{x}\right) dx$$

Przykład.

(X_1, X_2) - zmienna losowa o rozkładzie jednostajnym w kwadracie $(0, 1) \times (0, 1)$.

Wyznaczyć rozkład pola prostokąta o bokach x_1, x_2 tzn. zmiennej losowej $Y = X_1 \cdot X_2$.

$$G(y) = \iint_{x_1 \cdot x_2 < y} dx_1 dx_2 = 1 - \iint_{x_1 \cdot x_2 \geq y} dx_1 dx_2 =$$

$$= 1 - \int_y^1 \left(\int_{y/x_1}^1 dx_2 \right) dx_1 = y(1 - \ln y)$$

dla $0 < y \leq 1$

$$\text{stąd} \quad G(y) = \begin{cases} 0 & \text{dla } y \leq 0 \\ y(1 - \ln y) & \text{dla } 0 < y \leq 1 \\ 1 & \text{dla } 1 < y \end{cases}$$

zatem $g(y) = -\ln y$ dla $0 < y \leq 1$

Przykład.

$$Y = X_2/X_1 ,$$

$$G(y) = \int_{-\infty}^0 \left(\int_{y \cdot x_1}^{\infty} f(x_1, x_2) dx_2 \right) dx_1 + \int_0^{\infty} \left(\int_{-\infty}^{y \cdot x_1} f(x_1, x_2) dx_2 \right) dx_1$$

wtedy

$$g(y) = - \int_{-\infty}^0 x f(x, yx) dx + \int_0^{\infty} x f(x, yx) dx$$

Przykład.

X_1, X_2 - niezależne zmienne losowe.

$X_1 - N(0, \sigma_1), \quad X_2 - N(0, \sigma_2).$

Niech $\tilde{X}_1 = \frac{X_1}{\sigma_1}, \quad \tilde{X}_2 = \frac{X_2}{\sigma_2}$ (mają rozkład $N(0, 1)$).

$$\tilde{g}(\tilde{y}) = - \int_{-\infty}^0 x \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} \cdot \frac{e^{-\tilde{y}x^2/2}}{\sqrt{2\pi}} dx + \int_0^{\infty} x \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} \cdot \frac{e^{-\tilde{y}x^2/2}}{\sqrt{2\pi}} dx = \frac{1}{\pi(1+\tilde{y}^2)}$$

(rozkład Cauchy'ego)

i korzystając funkcji liniowej od zmiennej

losowej $Y = \tilde{Y} \frac{\sigma_1}{\sigma_2}$ mamy

$$g(y) = \frac{1}{\pi \frac{\sigma_2}{\sigma_1} \left(1 + y^2 \left(\frac{\sigma_1}{\sigma_2} \right)^2 \right)}$$

Przykład.

$$Y = X_1 + X_2 ,$$

$$G(y) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{y-x_1} f(x_1, x_2) dx_2 \right) dx_1$$

wtedy

$$g(y) = \int_{-\infty}^{\infty} f(x, y-x) dx = \int_{-\infty}^{\infty} f(y-x, x) dx$$

Uwaga.

Jeśli X_1, X_2 - niezależne zmienne losowe to gęstość sumy wyraża się splotem gęstości brzegowych (p. dalej).

Przykład.

$$Y = X_1 - X_2 ,$$

wtedy

$$g(y) = \int_{-\infty}^{\infty} f(x, x-y) dx = \int_{-\infty}^{\infty} f(x-y, x) dx$$

Suma niezależnych zmiennych losowych.

Własności:

- 1) X, Y niezależne skokowe zmienne losowe o funkcjach prawdopodobieństwa $P(X = x_i)$, $P(Y = y_j)$; wtedy funkcja prawdopodobieństwa zmiennej losowej $X + Y$ wyraża się wzorem:

$$P(Z = z_k) = \sum P(X = x_i)P(Y = z_k - x_i);$$

$(z_k = x_i + y_j)$

2) X, Y niezależne ciągłe zmienne losowe o gęstościach f_1 i f_2 ; wtedy gęstość zmiennej losowej $X + Y$ wyraża się wzorem:

$$f(x) = \int_{-\infty}^{\infty} f_1(t) f_2(x-t) dt = \int_{-\infty}^{\infty} f_1(x-t) f_2(t) dt$$

(splot gęstości składników).

FUNKCJE ZMIENNYCH LOSOWYCH n - WYMIAROWYCH.

Przykład.

System składa się n układów z których każdy ma czas bezawaryjnej pracy określony rozkładem wykładniczym X_i o parametrze a_i niezależnym od pozostałych układów. Wyznaczyć rozkład bezawaryjnego czasu pracy całego systemu (system działa jeśli pracuje chociaż jeden układ).

$$Y = X_1 + X_2 + X_3 + \dots + X_n,$$

Przez indukcję pokazuje się, że Y ma rozkład o gęstości:

$$g(y) = (-1)^{n-1} \prod_{i=1}^n a_i \sum_{j=1}^n \frac{e^{-a_j y}}{\prod_{\substack{k=1 \\ k \neq j}}^n (a_j - a_k)} \quad \text{dla } y > 0$$

wtedy

$$G(y) = (-1)^{n-1} \prod_{i=1}^n a_i \sum_{j=1}^n \frac{1 - e^{-a_j y}}{\prod_{\substack{k=1 \\ k \neq j}}^n (a_j - a_k)}$$

Otrzymany rozkład nazywamy uogólnionym rozkładem Erlanga n - tego rzędu T_n .

$$E(T_n) = E\left(\sum_{k=1}^n T_i\right) = \sum_{k=1}^n \frac{1}{a_k}$$

$$D^2(T_n) = D^2\left(\sum_{k=1}^n T_i\right) = \sum_{k=1}^n \frac{1}{a_k^2}$$

gdy $a_1 = a_2 = \dots = a_n = \lambda$ to

$$g(y) = \frac{\lambda(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \lambda P(n-1, \lambda t) \quad t > 0$$

dla $y > 0$

gdzie $P(n-1, \lambda t)$ jest rozkładem Poissona.

Przykład.

$$Y = \min(X_1, X_2)$$

$$G(y) = F_1(y) + F_2(y) - F(y, y)$$

wtedy

$$g(y) = f_1(y) + f_2(y) - \int_{-\infty}^y f(y, x) dx - \int_{-\infty}^y f(x, y) dx$$

Uwaga.

Jeśli X_1, X_2 - niezależne zmienne losowe to:

$$G(y) = F_1(y) + F_2(y) - F_1(y) \cdot F_2(y)$$

$$g(y) = f_1(y)(1 - F_2(y)) + f_2(y)(1 - F_1(y))$$

Jeśli X_1, X_2 - niezależne zmienne losowe o takim samym rozkładzie to:

$$G(y) = F(y)(2 - F(y))$$

wtedy $g(y) = 2f(y)(1 - F(y))$

Przykład.

$$Y = \max(X_1, X_2)$$

$$G(y) = F(y, y)$$

wtedy

$$g(y) = \int_{-\infty}^y f(y, x)dx + \int_{-\infty}^y f(x, y)dx$$

Uwaga.

Jeśli X_1, X_2 - niezależne zmienne losowe o takim samym rozkładzie to:

$$G(y) = F^2(y)$$

wtedy $g(y) = 2f(y)F(y)$

ROZKŁAD NORMALNY 2-WYMIAROWY

Zmienna losowa (X, Y) o rozkładzie normalnym 2-wymiarowym zależy od pięciu parametrów: $m_1, m_2, \sigma_1, \sigma_2, \rho$.

$m_1 = EX; m_2 = EY; \sigma_1 = DX; \sigma_2 = DY;$

$\rho =$ współczynnik korelacji.

Współczynnik korelacji musi spełniać warunek: $\rho^2 \neq 1$.

Macierz kowariancji K ma wtedy postać

$$K = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Gęstość rozkładu normalnego 2-wymiarowego $N(m_1, m_2, \sigma_1, \sigma_2, \rho)$ można zapisać następująco:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}$$

Twierdzenie

Dowolny rozkład brzegowy normalnego rozkładu 2-wymiarowego jest rozkładem normalnym.

Twierdzenie

Jeśli składowe normalnego rozkładu 2-wymiarowego są nieskorelowane to są niezależne.

Rozkład normalny n - wymiarowy.

K - macierz kowariancyjna, niech $\det K \neq 0$.
Zmienna losowa n - wymiarowa ma rozkład normalny n - wymiarowy gdy gęstość tej zmiennej losowej wyraża się wzorem:

$$\begin{aligned} f(x) = f(x_1, x_2, \dots, x_n) &= \frac{\sqrt{|L|}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{j,k=1}^n l_{jk} (x_j - m_j)(x_k - m_k)\right) = \\ &= \frac{\sqrt{|L|}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} (x - m)^T L (x - m)\right) \end{aligned}$$

gdzie

$m_i = E(X_i)$ dla $i = 1, 2, \dots, n$

$L = [l_{jk}]$ $j, k = 1, 2, \dots, n$ jest macierzą odwrotną do K .

Dla $n = 2$ warunek $|K| \neq 0$ jest równoważny warunkowi $\rho^2 \neq 1$.

Ponieważ macierz K ma wtedy postać

$$\begin{aligned} K &= \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \text{ to} \\ L &= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \end{aligned}$$

Zatem gęstość rozkładu normalnego 2-wymiarowego $N(m_1, m_2, \sigma_1, \sigma_2, \rho)$ można zapisać następująco:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}$$

Powyższa funkcja gęstości ma stałą wartość $f(x, y) = h$ na elipsie:

$$\frac{(x - m_1)^2}{\sigma_1^2} - 2\rho \frac{(x - m_1)(y - m_2)}{\sigma_1 \sigma_2} + \frac{(y - m_2)^2}{\sigma_2^2} = \text{const} = \lambda^2$$

o środku w punkcie (m_1, m_2) , gdzie

$$\lambda^2 = -2(1 - \rho^2) \ln(h 2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}).$$

Dla $\rho \neq 0$ osie główne mają równania:

$$y - m_2 = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2 \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2\sigma_1^2\sigma_2^2}}(x - m_1)$$

Dla $\rho = 0$ osie rozpatrywanej elipsy są równoległe do osi układu współrzędnych.

Zauważmy, że gdy $\rho^2 \rightarrow 1$ to jedna oś się wydłuża, a druga skraca, zależność między zmiennymi staje się ściśle liniowa.

Osie powyższej elipsy tworzą z osią OX kąty α i $\alpha + \pi/2$ gdzie

$$\operatorname{tg} 2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}$$

Funkcja charakterystyczna:

$$\varphi(t) = \exp\left(im^T t - \frac{1}{2}t^T Kt\right)$$

gdy $n = 2$ to

$$\varphi(t_1, t_2) = \exp\left(i(t_1 m_1 + t_2 m_2) - \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)\right)$$

Twierdzenie.

Dowolny rozkład brzegowy normalnego rozkładu n -wymiarowego jest rozkładem normalnym.

Twierdzenie.

Jeśli składowe normalnego rozkładu n -wymiarowego są parami nieskorelowane to są niezależne.

Twierdzenie.

Dowolny rozkład warunkowy normalnego rozkładu n -wymiarowego jest rozkładem normalnym. Warunkowa wartość oczekiwana i warunkowa wariancja są równe:

$$E(X_n | X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) = \frac{1}{K_{nn}} \sum_{i=1}^{n-1} x_i K_{in}$$

gdzie K_{ij} - dopełnienie algebraiczne elementu k_{ij} macierzy K .

$$D^2(X_n | X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) = \frac{|K|}{K_{nn}}$$

Uwaga.

Dla $n = 2$ gęstość rozkładu warunkowego $Y|x$ jest równa:

$$\begin{aligned} f_y(y|x) &= \frac{f(x,y)}{f_x(x)} = \\ &= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)\sigma_2^2}\left[y-m_2-\rho\frac{\sigma_2}{\sigma_1}(x-m_1)\right]^2\right\} \end{aligned}$$

oraz

$$E(Y | X = x) = m_2 - \rho \frac{\sigma_2}{\sigma_1} (x - m_1),$$

$$D^2(Y | X = x) = \sigma_2^2 (1 - \rho^2)$$

Przykład.

X, Y - niezależne zmienne losowe o rozkładzie normalnym, $EX = 1$, $EY = 2$, $D^2X = 4$, $D^2Y = 9$.

Wyznaczyć gęstość rozkładu zmiennej losowej (X, Y) , obliczyć

a) $P(1 < X < 2; 1 < Y < 4)$,

b) $P(X > 3)$.

$$f(x, y) = \frac{1}{12\pi} \exp \left\{ -\frac{1}{2} \left[\frac{(x-1)^2}{4} + \frac{(y-2)^2}{9} \right] \right\}$$

a)

$$P(1 < X < 2; 1 < Y < 4) =$$

$$= \frac{1}{12\pi} \int_1^2 \int_1^4 \exp \left\{ -\frac{1}{2} \left[\frac{(x-1)^2}{4} + \frac{(y-2)^2}{9} \right] \right\} dx dy =$$

$$= \frac{1}{12\pi} \int_1^2 \exp \left[-\frac{1}{2} \left(\frac{x-1}{2} \right)^2 \right] dx \int_1^4 \exp \left[-\frac{1}{2} \left(\frac{y-2}{3} \right)^2 \right] dy =$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{1/2} e^{-\frac{1}{2}t^2} dt \frac{1}{\sqrt{2\pi}} \int_{-1/3}^{2/3} e^{-\frac{1}{2}u^2} du =$$

$$= [\Phi(1/2) - \Phi(0)] [\Phi(2/3) - \Phi(-1/3)] =$$

$$= (0,6915 - 0,5)(0,7486 - (1 - 0,6293)) =$$

$$= 0,0718$$

b)

$$\begin{aligned}P(X > 3) &= \frac{1}{12\pi} \int_3^{\infty} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-1)^2}{4} + \frac{(y-2)^2}{9}\right]\right\} dx dy = \\&= \frac{1}{2\sqrt{2\pi}} \int_3^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-1}{2}\right)^2\right] dx \frac{1}{3\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{y-2}{3}\right)^2\right] dy = \\&= \frac{1}{\sqrt{2\pi}} \int_1^{\infty} e^{-\frac{1}{2}t^2} dt = \Phi(\infty) - \Phi(1) = 1 - 0,8413 = \\&= 0,1587.\end{aligned}$$

Przykład.

Wyznaczyć gęstość rozkładu normalnego (X, Y, Z) jeśli rozkład ten ma zerowy wektor wartości oczekiwanych i macierz kowariancji:

$$K = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 3 & 1 \\ 2 & 1 & 2 \end{bmatrix}$$

Rozwiązanie.

$$\det K = 2. \quad K^{-1} = \frac{1}{2} \begin{bmatrix} 5 & -4 & -3 \\ -4 & 4 & 2 \\ -3 & 2 & 3 \end{bmatrix}$$

zatem

$$f(x, y, z) = \frac{1}{\sqrt{2}(2\pi)^{3/2}} \exp \left\{ -\frac{1}{2} \left(\frac{5}{2} x^2 - 4xy - 3xz + 2y^2 + 2yz + \frac{3}{2} z^2 \right) \right\}$$

.

Uwaga.

Dla rozkładu normalnego 2 wymiarowego (X, Y) takiego, że $EX = EY = 0$, $DX = \sigma_X$, $DY = \sigma_Y$, którego składowe są nieskorelowane (każdy rozkład normalny może mieć taką postać po obrocie układu współrzędnych o kąt α) prawdopodobieństwo, że wartości zmiennej losowej (X, Y) należą do elipsy $\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} = k^2$ jest równe $1 - e^{-\frac{k^2}{2}}$

Rozkład funkcji od rozkładu normalnego.

$(X_1, X_2, X_3, \dots, X_n)$ - rozkład normalny.

$Y = g(X_1, X_2, X_3, \dots, X_n)$, należy wyznaczyć rozkład Y .

Przykład.

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n + b$$

$n = 2$

$$g(y) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{\frac{-(y-m_y)^2}{2\sigma_y^2}}$$

gdzie: $m_y = a_1m_1 + a_2m_2 + b$,

$$\sigma_y^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2a_1 a_2 r \sigma_1 \sigma_2 ,$$

Przez indukcję można pokazać, że dla dowolnego n zmienna losowa Y ma rozkład normalny o parametrach:

$$m_y = a_1m_1 + a_2m_2 + \dots + a_nm_n + b,$$

$$\sigma_y^2 = \sum a_i^2 \sigma_i^2 + 2 \sum_{i < j} a_i a_j r_{ij} \sigma_i \sigma_j ,$$

Przykład.

$X_1, \dots, X_n$ - niezależne, o rozkładzie $N(0, 1)$.

$$Y_n = X_1^2 + \dots + X_n^2$$

ma **rozkład chi kwadrat**

$$n \in N \quad f(y) = \begin{cases} \frac{y^{\frac{n}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$EX = n; \quad D^2X = 2n$$

Przykład.

$X - N(m, \sigma), \quad Y = e^X$

Ma rozkład **logarytmiczno-normalny**.

Nazwa pochodzi stąd, że $X = \ln Y$ ma rozkład normalny.

L.Kowalski 30.04.2011