ZMIENNE LOSOWE WIELOWYMIAROWE

(Ω, S, P) - ustalona przestrzeń probabilistyczna.
X = (X_1, X_2, ..., X_n) - zmienna losowa n - wymiarowa (wektor losowy, ciąg losowy).

$X : \Omega \to R^n$ (funkcja borelowska)

$P_X : \mathrm{B}(R^n) \to [0,\ 1]$ - rozkład zmiennej losowej X.

Dystrybuanta

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

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

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

dla pewnej nieujemnej funkcji f zwanej gęstością.

Uwaga.

1. W punktach ciągłości funkcji f zachodzi:

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

2. Dla $A \in \mathrm{B}(R^n)$ mamy $P_X(A) = \int\int_A \ldots \int f(x_1,\ \ldots, x_n) dx_1 \ldots dx_n$.

Funkcja charakterystyczna zmiennej losowej n - wymiarowej.
$\varphi(t) = \varphi(t_1,\ \ldots, t_n) = E\left(e^{itX}\right) = E\left(\exp(i(t_1 X_1 + \ldots + t_n X_n))\right)$.

Rozkłady warunkowe.
Jeśli $P_{1,\ldots,k}(X_1 = x_{1j},\ \ldots, X_k = x_{kj}) > 0$ to rozkład zmiennej losowej skokowej (n - k) wymiarowej określonej wzorem:

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

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

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

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

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

Niezależność zmiennych losowych.
Zmienne losowe $X_1, X_2, \ldots, X_n$ są niezależne jeśli
$F(x_1,\ \ldots, x_n) = F_1(x_1)\cdot F_2(x_2)\ldots\cdot F_n(x_n)$ dla dowolnych $x_1, x_2, \ldots, x_n \in R^n$.
gdzie F_i - dystrybuanty rozkładów brzegowych jednowymiarowych.
Dla zmiennych losowych skokowych odpowiedni warunek ma postać:
$P(X_1 = x_{1j},\ \ldots, X_n = x_{nj}) = P_1(X_1 = x_{1j})\cdot\ldots\cdot P_n(X_n = x_{nj})$
dla dowolnych $x_{1j},\ \ldots, x_{nj} \in R^n$
Dla zmiennych losowych ciągłych odpowiedni warunek ma postać:
$f(x_1,\ \ldots, x_n) = f_1(x_1)\cdot f_2(x_2)\ldots\cdot f_n(x_n)$
dla dowolnych $x_1, x_2, \ldots, x_n \in R^n$.

Przypadek n = 2.

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

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

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 \ 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
$x_1, x_2, \ldots, x_k$ – wartości zmiennej losowej X,
$y_1, y_2, \ldots, y_l$ – 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	...	x_k
$p_{i\cdot}$	$p_{1\cdot}$	$p_{2\cdot}$	...	$p_{k\cdot}$

Rozkładem brzegowym zmiennej losowej Y nazywamy rozkład określony funkcją prawdopodobieństwa:

y_j	y_1	y_2	....	y_l
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	......	$p_{\cdot l}$

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, ...) 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

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

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 \ 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 \ 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 \ 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

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

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

Uwaga.

3. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, \quad y)dxdy = 1$

4. W punktach ciągłości funkcji f zachodzi:
$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$$

5. Dla $A \in \mathrm{B}(R^2)$ mamy $P_{(X,Y)}(A) = \iint_A f(x,y)dxdy$.

Mając gęstość rozkładu łącznego gęstości rozkładów brzegowych wyznaczamy następująco.
Jeśli f(x, y) jest gęstością zmiennej losowej (X,Y) to funkcje

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

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).

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

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

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

Rozkłady brzegowe to rozkłady jednostajne na przedziale [0, 2].
Zauważmy, że zmienne losowe *X*, *Y* są niezależne.

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

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

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

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

Parametry (mogą nie istnieć)
Wartość oczekiwana $E(X)=[EX_1,\ EX_2,\ ...,EX_n]$.
Wariancja $\mathrm{D}^2(X)=[\mathrm{D}^2X_1,\ \mathrm{D}^2X_2,\ ...,\mathrm{D}^2X_n]$.

Moment (zwyczajny) rzędu $l_1 + l_2 + ...+ l_n$

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

Moment centralny rzędu $l_1 + l_2 + ...+ l_n$

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

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

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

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

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

Macierz korelacji R = [ρ_{ij}], gdzie $\rho_{ij} = \dfrac{\operatorname{cov}(X_i, X_j)}{DX_i \cdot DX_j} =$

Uwaga $\rho_{ii} = 1$.

Przypadek zmiennej losowej dwuwymiarowej

Kowariancją zmiennych losowych (*X*, *Y*) nazywamy wielkość

$$\operatorname{Cov}(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - E(X)E(Y)$$

Dla zmiennej losowej skokowej (*X*, *Y*) mamy:

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

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

Dla zmiennej losowej ciągłej (*X*, *Y*) mamy:

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, \ y)dxdy$$

$$\operatorname{Cov}(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, \ y)dxdy - EX \cdot EY$$

Uwaga

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

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

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

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

Własności współczynnika korelacji:

a) $-1 \le \rho(X,Y) \le 1$
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 $$K = \begin{bmatrix} D^2X & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & D^2Y \end{bmatrix}$$

nazywamy **macierzą kowariancji**

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

X \ 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	0	1
$p_{i.}$	1/3	1/30	1/3

$EX = 0$

Rozkład brzegowy zmiennej losowej Y:

y_j	– 1	1
$p_{.j}$	2/3	1/3

$EY = – 1/3$

Ponieważ
$E(XY) = (– 1)\cdot(– 1)\cdot 1/6 + – 1)\cdot 1\cdot 1/6 + 1\cdot (– 1)\cdot 1/6 + 1\cdot 1\cdot 1/6 = 0$,
$EX\cdot EY = 0$; Cov(X, Y) = 0 to $\rho = 0$.
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 $\varphi(x) = E(Y \mid x)$ 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 to prosta $y = \rho \frac{\sigma_Y}{\sigma_X} x + m_Y - \rho \frac{\sigma_Y}{\sigma_X} m_X$.

Regresja II rodzaju X względem Y to 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, ...,X_n ma równanie

$$x_1 - EX_1 = a_{12}(x_2 - EX_2) + ...+ a_{1n}(x_n - EX_n) \qquad a_{1i} = -\frac{K_{1i}}{K_{11}}$$

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

Wielowymiarowy rozkład Bernoulliego.

Dla danych k ∈ N, p = $[p_1, p_2, ...,p_n]^T$ takiego, że $0 \le \sum_{i=1}^{n} p_i < 1$ oraz

i = $[i_1, i_2, ...,i_n]^T$ gdzie $i_j \in \{0, 1, ..., n\}$ $\sum_{j=1}^{n} i_j \le k$ określamy

$$P(X = i) = \frac{k!}{i_0!i_1!i_2!... i_n!} p_0^{i_0} \cdot p_1^{i_1} \cdot ... \cdot p_n^{i_n}$$

gdzie $p_0 = 1 - \sum_{i=1}^{n} p_i$; $i_0 = k - \sum_{j=1}^{n} i_j$.

Przykład.

Badanie systemu telekomunikacyjnego polega na wielokrotnych próbach uzyskania połączenia. Rozpatrujemy trzy możliwe wyniki każdego połączenia:

- A_0 - połączenie bez zakłóceń,
- A_1 - połączenie z zakłóceniami,
- A_2 - brak połączenia.

Wiadomo, że $P(A_0) = 0{,}7$; $P(A_1) = 0{,}2$; $P(A_2) = 0{,}1$.

Wykonano 50 prób łączności, obliczyć prawdopodobieństwo tego, że w tych próbach co najwyżej raz nie uzyskamy połączenia i co najwyżej raz uzyskamy połączenie z zakłóceniami.

X - liczba prób z brakiem łączności,

Y - liczba prób z połączeniami z zakłóceniami.

$P(X = i, Y = j) = \frac{50!}{i!\, j!(50 - i - j)!} 0{,}7^{50-i-j} 0{,}2^i 0{,}1^j$ gdzie i, j = 0, 1, ..., 50; i + j ≤ 50.

Zatem

$$P(X \le 1, Y \le 1) = P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 1, Y = 0) + P(X = 1, Y = 1) =$$
$$0{,}7^{50} + 50 \cdot 0{,}1 \cdot 0{,}7^{49} + 50 \cdot 0{,}2 \cdot 0{,}7^{49} + 50 \cdot 49 \cdot 0{,}1 \cdot 0{,}2 \cdot 0{,}7^{48} \approx 0{,}0000022.$$

Wielowymiarowy rozkład wielomianowy.

Jeśli w definicji rozkładu Bernoulliego mamy $p_o = 0$, $\sum_{j=1}^{n} i_j = k$ to otrzymany rozkład nazywamy rozkładem wielomianowym.

Wielowymiarowy rozkład Poissona.

Dla danego $\lambda = [\lambda_1, \lambda_2, ..., \lambda_n]^T$ oraz

$i = [i_1, i_2, ..., i_n]^T$ gdzie $i_j \in \{0, 1, ..., n\}$ określamy

$$P(X = i) = \frac{\lambda_1^{i_1}}{i_1!} \cdot ... \cdot \frac{\lambda_n^{i_n}}{i_n!} e^{-\lambda_0} \qquad \text{gdzie } \lambda_0 = \sum_{i=1}^{n} \lambda_i \quad .$$

Rozkład normalny n - wymiarowy.

K - macierz kowariancyjna, niech $\det K \ne 0$.

Zmienna losowa n - wymiarowa ma rozkład normalny n - wymiarowy gdy gęstość tej zmiennej losowej wyraża się wzorem:

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

gdzie

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

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

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

Ponieważ macierz K ma wtedy postać $K = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ to

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

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:

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

o środku w punkcie (m_1, m_2), gdzie $\lambda^2 = -2\left(1-\rho^2\right)\ln\left(h2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)$.
Dla $\rho \neq 0$ osie główne mają równania:

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

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

Zauważmy, że gdy $\rho^2 \to 1$ 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

$$\varphi(t_1,t_2) = \exp\left(i\left(t_1m_1 + t_2m_2\right) - \frac{1}{2}\left(\sigma_1^2t_1^2 + 2\rho\sigma_1\sigma_2t_1t_2 + \sigma_2^2t_2^2\right)\right)$$

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:

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

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

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

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

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

oraz

$$E(Y \mid X = x) = m_2 - \rho\frac{\sigma_2}{\sigma_1}(x - m_1),\quad D^2(Y \mid X = x) = \sigma_2^2\left(1-\rho^2\right)$$

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

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

$= \frac{1}{\sqrt{2\pi}}\int_0^{1/2} e^{-\frac{1}{2}t^2}dt\,\frac{1}{\sqrt{2\pi}}\int_{-1/3}^{2/3} e^{-\frac{1}{2}u^2}du =$ [Φ(1/2) - Φ(0)] [Φ(2/3) - Φ(-1/3)] =

= (0,6915 - 0,5)(0,7486 - (1 - 0,6293)) = 0,0718

b)

P(X > 3) = $\frac{1}{12\pi}\int_3^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-1)^2}{4}+\frac{(y-2)^2}{9}\right]\right\}dxdy =$

$= \frac{1}{2\sqrt{2\pi}}\int_3^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-1}{2}\right)^2\right]dx\,\frac{1}{3\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{y-2}{3}\right)^2\right]dy =$

$\frac{1}{\sqrt{2\pi}}\int_1^{\infty} e^{-\frac{1}{2}t^2}dt =$ Φ(∞) - Φ(1) = 1 - 0,8413 = 0,1587.

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.

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

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 = σ_X, DY = σ_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 $1-e^{-\frac{k^2}{2}}$

ZADANIA

Zadanie 1.1

Zmienna losowa (*X*, *Y*) ma rozkład określony tabelą:

X \ *Y*	– 1	0
1	0,4	0,3
0	0,1	0,2

Wyznaczyć macierz korelacji. Obliczyć współczynnik korelacji między tymi zmiennymi. Czy *X, Y* są skorelowane? Czy *X, Y* są niezależne?

Zadanie 1.2

Zmienna losowa (*X*, *Y*) ma rozkład określony tabelą:

X \ *Y*	0	1	2
5	0,0	0,0	0,1
6	0,1	0,2	0,1
7	0,3	0,1	0,1

Wyznacz rozkład zmiennej losowej *X*. Wyznacz rozkład zmiennej losowej *Y*.

Obliczyć współczynnik korelacji między tymi zmiennymi.

Czy *X*, *Y* są skorelowane? Czy *X*, *Y* są niezależne?

Zadanie 1.3

Dla zmiennej losowej z poprzedniego zadania wyznacz i narysuj

a) linie regresji I rodzaju,

b) proste regresji II rodzaju.

Zadanie 1.4

Zmienna losowa (X, Y) ma rozkład określony tabelą:

Y / X	0	1	2
5	0	0	0,1
6	0,1	0,2	0,1
7	0,3	0,1	0,1

a) wyznaczyć F(1; 1), F(6; 2), F(7; 1),

b) obliczyć $P(|X| \geq 6; |Y| \leq 1)$,

c) wyznacz rozkłady warunkowe $X \mid Y = 1$; $Y \mid X = 5$,

d) obliczyć wartości oczekiwane zmiennych z punktu c).

Zadanie 1.5

(X, Y) jest zmienną losową o gęstości

$$f(x, y) = \begin{cases} c & \text{dla } (x, y) \in D \\ 0 & \text{dla } (x, y) \notin D \end{cases}$$

gdzie D jest trójkątem o wierzchołkach (0; 0); (1; 0); (1; 1).

a) wyznaczyć c,

b) wyznaczyć F(1; 0,5),

c) wyznaczyć gęstości rozkładów brzegowych,

d) wyznaczyć gęstość rozkładu $X \mid Y = 0{,}5$,

e) obliczyć EX, EY,

f) obliczyć cov(X, Y),

g) obliczyć współczynnik korelacji,

h) Czy X, Y są nieskorelowane? Czy są niezależne?

i) wyznacz prostą regresji Y względem X,

Zadanie 1.6

Zmienna losowa (X, Y) ma macierz kowariancji: $K = \begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix}$.

Ile wynosi współczynnik korelacji między X i Y?

Zadanie 1.7

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

$$f(x,y,z)=\begin{cases} c & \text{dla } 0\le x\le 1,\ 0\le y\le 2,\ 0\le z\le 3 \\ 0 & \text{dla innych x, y, z}\end{cases}$$

a) wyznaczyć c,

b) wyznaczyć gęstości brzegowe jedno i dwuwymiarowe,

c) wyznaczyć gęstość rozkładu warunkowego X| Y = 1, Z = 1,

d) wyznaczyć gęstość rozkładu warunkowego (X, Y)| Z = 1,

e) czy X, Y, Z są niezależne?

f) wyznacz wektor wartości oczekiwanych tej zmiennej losowej.

Zadanie 1.8

Wyznaczyć wartość parametru *c* aby funkcja $f(x,y)=c\cdot\exp\left\{-\frac{1}{2}\left(x^2+2xy+5y^2\right)\right\}$

była gęstością 2 wymiarowego rozkładu normalnego.

Wyznaczyć parametry m_1, m_2, σ_1, σ_2, ρ.

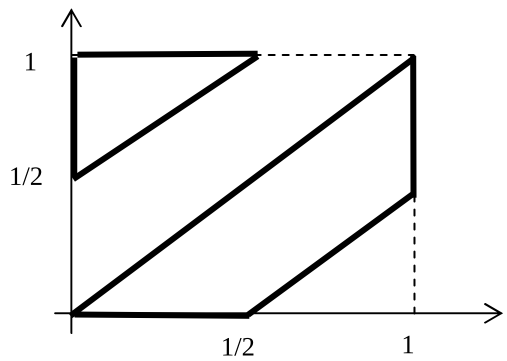
Zadanie 1.9

(X, Y) ma rozkład o gęstości $f(x,y)=\frac{1}{300\pi}\exp\left\{-\frac{1}{2}\left[\frac{(x-30)^2}{100}+\frac{(y-40)^2}{225}\right]\right\}$.

Czy X, Y są skorelowane? Czy X, Y są niezależne?

Zadanie 1.10

Zmienna losowa (X,Y) ma stałą gęstość na zaznaczonym zbiorze



Sprawdź, że rozkłady brzegowe mają rozkład jednostajny na przedziale (0, 1).

Sprawdź, że X, Y są zależne.

Zadanie 1.11

Zmienne losowe X, Y są niezależne i mają rozkłady jednostajne odpowiednio w przedziałach [0, 3] i [-2, 2].

Wyznacz gęstość rozkładu łącznego (X, Y).

Zadanie 1.12

(*X*, *Y*) ma rozkład o dystrybuancie

$$F(x,y)=\begin{cases}1-e^{-3x}-e^{-4y}+e^{-3x-4y} & \text{dla } \ x>0,\ y>0\\ 0 & \text{dla innych x, y}\end{cases}$$

Wyznacz gęstość zmiennej losowej (X, Y).

Zadanie 1.13

Gęstość 2 wymiarowego rozkładu normalnego wyraża się funkcją

$$f(x,y)=\frac{1}{\pi}\cdot\exp\left\{-\frac{1}{2}\left(x^2+2xy+5y^2\right)\right\}$$

Zapisać gęstość rozkładu brzegowego $f_1(x)$ i określić jego parametry.

Zapisać gęstość rozkładu warunkowego $f_2(y|x)$ i określić jego parametry.

Zadanie 1.14

Wyznaczyć wartość parametru c aby funkcja

$$f(x,y)=c\cdot\exp\left\{-\frac{1}{2}\left(x^2-xy+y^2\right)\right\}$$

była gęstością 2 wymiarowego rozkładu normalnego.

Wyznaczyć macierz kowariancji tej zmiennej losowej.

Zadanie 1.15

Wyznaczyć gęstość rozkładu normalnego (X, Y, Z) jeśli rozkład ten ma wektor wartości oczekiwanych E(X, Y, Z) = [1, -1, 0]^T i macierz kowariancji:

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

Zadanie 1.16

Funkcja

$$f(x, y, z) = \frac{1}{2\pi\sqrt{230\pi}} \exp\left\{-\frac{1}{230}\left(39x^2 + 36y^2 + 26z^2 - 44xy + 36xz - 38yz\right)\right\}$$

jest gęstością 3 wymiarowego rozkładu normalnego.

Wyznaczyć wektor wartości oczekiwanych i macierz kowariancji tej zmiennej losowej.

Zadanie 1.17

Rzucamy 4 razy monetą.

X - liczba orłów uzyskanych w tych rzutach,

Y - liczba serii orłów.

a) Wypisać wszystkie zdarzenia elementarne w tym doświadczeniu losowym.
b) Wyznaczyć rozkład zmiennej losowej (X, Y),
c) Wyznaczyć rozkłady brzegowe i i ich wartości oczekiwane,
d) Wyznaczyć i narysować linie regresji I rodzaju,
e) Wyznaczyć i narysować proste regresji II rodzaju.
f) Czy X i Y są niezależne ? czy są skorelowane?

4. 03.2010