

DWUWYMIAROWA ZMIENNA LOSOWA

Niech X i Y będą zmiennymi losowymi określonymi na zbiorze zdarzeń elementarnych Ω .

Zmienna losowa dwuwymiarowa (lub dwuwymiarowy wektor losowy) to para (X, Y) zmiennych losowych jednowymiarowych.

Zmienna losowa dwuwymiarowa przyporządkowuje zdarzeniom elementarnym pary liczb rzeczywistych.

Dystrybuantą dwuwymiarowej zmiennej losowej (X, Y) nazywamy funkcję F , dwóch zmiennych rzeczywistych x, y określoną następująco:

$$F(x, y) = P(X < x, Y < y) \quad \text{dla } (x, y) \in \mathbb{R}^2.$$

Własności dystrybuanty zmiennej losowej (X, Y)

$$1. \quad \forall_{x \in \mathbb{R}} \lim_{y \rightarrow -\infty} F(x, y) = 0, \quad \forall_{y \in \mathbb{R}} \lim_{x \rightarrow -\infty} F(x, y) = 0$$

$$2. \quad \lim_{y \rightarrow \infty, x \rightarrow \infty} F(x, y) = 1$$

$$3. \quad \text{Dla dowolnych punktów: } (x_1, x_2), (y_1, y_2)$$

takich, że $x_1 \leq x_2, y_1 \leq y_2$, zachodzi

$$F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) = P(x_1 \leq X < x_2, y_1 \leq Y < y_2) \geq 0.$$

4. Dystrybuanta jest funkcją niemalejącą i co najmniej lewostronnie ciągłą względem każdego z argumentów x lub y .

Uwaga.

Każda funkcja dwóch zmiennych spełniająca warunki 1, 2, 3, 4 może być traktowana jak o dystrybuantę pewnej dwuwymiarowej zmiennej losowej (X, Y)

ZMIENNA LOSOWA DWUWYMIAROWA TYPU SKOKOWEGO

Dwuwymiarowa zmienna losowa przyjmuje skończony lub przeliczalny zbiór wartości.

Funkcja prawdopodobieństwa zmiennej losowej dwuwymiarowej skokowej (X, Y) jest określona wzorem

$$p_{ij} = P(X=x_i, Y=y_j), i, j=1, 2, 3, \dots$$

przy czym

$$p_{ij} > 0 \text{ oraz } \sum_i \sum_j p_{ij} = 1.$$

lub tabelą

$x_i \backslash y_j$	x_1	x_2	x_3	
y_1	p_{11}	p_{21}	p_{31}		...
y_2	p_{12}	p_{22}	p_{32}		...
y_3	p_{13}	p_{23}	p_{33}		
....

Dystrybuanta zmiennej losowej dwuwymiarowej typu skokowego (X, Y) jest to funkcja F dwóch zmiennych rzeczywistych x i y określona wzorem

$$F(x, y) = P(X < x, Y < y) = \sum_{x_i < x, y_j < y} p_{ij} \quad \text{dla } x, y \in \mathbb{R}.$$

Przykład

Losujemy liczbę ze zbioru $\{1,2,3,4,5,6\}$. Zmienna losowa X przyjmuje 1, gdy wylosowano liczbę parzystą lub wartość 0, gdy wylosowano liczbę nieparzystą. Zmienna losowa Y przyjmuje wartość 1, gdy wylosowano liczbę podzielną przez 3 lub wartość 0, gdy wylosowano liczbę niepodzielną przez 3.

(a) Wyznaczyć rozkład zmiennej losowej (X, Y) .

X:	0	1	0	1	0	1
$\Omega=\{$	1,	2,	3,	4,	5,	6}
Y:	0	0	1	0	0	1

Zbiór wartości zmiennej losowej dwuwymiarowej (X, Y) :

$$\{(0,0),(0,1),(1,0),(1,1)\}$$

Rozkład j zmiennej (X, Y)

$x_i y_j$	0	1
0	$p_{11}=1/3$	$p_{12}=1/6$
1	$p_{21}=1/3$	$p_{22}=1/6$

gdzie $p_{11}=P(X=0, Y=0)=P(X=1, Y=5)=\frac{2}{6}=\frac{1}{3}$,

$$p_{21}=P(X=0, Y=1)=P(\{3\})=\frac{1}{6}$$

$$p_{12}=P(X=1, Y=0)=P(\{2,4\})=\frac{2}{6}=\frac{1}{3}, \quad p_{22}=P(X=1, Y=1)=P(\{6\})=\frac{1}{6}$$

(b) Wyznacz dystrybuantę zmiennej losowej (X, Y)

xy	$(-\infty; 0]$	$(0; 1]$	$(1, \infty)$
$(-\infty; 0]$	0	0	0
$(0; 1]$	0	1/3	1/2
$(1, \infty)$	0	2/3	1

Rozpatrzmy przypadki

1. $x \leq 0$ lub $y \leq 0$ $F(x, y) = 0$

2. $0 < x \leq 1$ i $0 < y \leq 1$ $F(x, y) = P(X=0, Y=0) = 1/3$

3. $0 < x \leq 1$ i $1 < y < \infty$ $F(x, y) = P(X=0, Y=0) + P(X=0, Y=1) = 1/3 + 1/6 = 1/2$

4. $0 < x < \infty$ i $0 < y \leq 1$ $F(x, y) = P(X=0, Y=0) + P(X=1, Y=0) = 1/3 + 1/3 = 2/3$

5. $0 < x < \infty$ i $1 < y < \infty$

$$F(x, y) = P(X=0, Y=0) + P(X=1, Y=0) + P(X=0, Y=1) + P(X=1, Y=1) = \\ 1/3 + 1/3 + 1/6 + 1/6 = 1$$

c) Oblicz $P(X=-1, Y=-1), P(X=0, Y=-1), P(X=1, Y=1)=1/6$,

$$F(-1, -1), F(0, -1), F(1, 1)$$

$$P(X=-1, Y=-1)=0, \quad P(X=0, Y=-1)=0, \quad P(X=1, Y=1)=1/6$$

$$F(-1, -1)=0, \quad F(0, -1)=0, \quad F(1, 1)=1/3.$$

ROZKŁADY BRZEGOWE ZMIENNYCH LOSOWYCH TYPU SKOKOWEGO

Rozkład prawdopodobieństwa zmiennej losowej dwuwymiarowej skokowej (X, Y) wyznacza rozkładskokowy zmiennej losowej X i rozkład skokowy zmiennej losowej Y .

Rozkłady brzegowe zmiennych losowych skokowych X i Y są to rozkłady prawdopodobieństwa tych zmiennych wyznaczone na podstawie dwuwymiarowej zmiennej losowej.

Jeśli (X, Y) jest dwuwymiarową zmienną losową skokową o funkcji prawdopodobieństwa

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

1. X jest zmienną losową skokową o funkcji prawdopodobieństwa (nazywaną **funkcją prawdopodobieństwabrzegowegozmiennejlosowej** (X, Y)) postaci:

$$P(X = x_i) = p_{i\cdot} \quad i = 1, 2, \dots, \text{ gdzie}$$

$$p_{i\cdot} = \sum_j p_{ij} = P(X = x_i, Y = y_1) + P(X = x_i, Y = y_2) + P(X = x_i, Y = y_3) + \dots \text{ lub tabelą}$$

x_i	x_1	x_2	x_3
$p_{i\cdot}$	$p_{1\cdot}$	$p_{2\cdot}$	$p_{3\cdot}$

I dystrybuancie rozkładu brzegowego

$$F_X(x) = \sum_{x_i < x} p_{i\cdot} \text{ dla } x \in \mathbb{R}.$$

2. Y jest zmienną losową skokową o funkcji prawdopodobieństwa (nazywaną **funkcją prawdopodobieństwabrzegowegozmiennejlosowej** (X, Y)) postaci

$$P(Y = y_j) = p_{\cdot j} \quad j = 1, 2, \dots, \text{ gdzie}$$

$$p_{\cdot j} = \sum_i p_{ij} = P(X = x_1, Y = y_j) + P(X = x_2, Y = y_j) + P(X = x_3, Y = y_j) + \dots$$

lub tabelą

y_j	y_1	y_2	y_3
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$

I dystrybuancie rozkładu brzegowego $F_Y(x) = \sum_{y_k < y} p_{\cdot k} \text{ dla } x \in \mathbb{R}.$

ROZKŁADY WARUNKOWE ZMIENNYCH LOSOWYCH TYPU SKOKOWEGO

Niech (X, Y) będzie dwuwymiarową zmienną losową skokową o funkcji prawdopodobieństwa

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

Funkcja prawdopodobieństwa warunkowego $X|Y=y_j$ wyraża się wzorem

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

w którym zdarzenie $Y=y_j$ jest ustalone, a x_i przebiega wszystkie wartości zmiennej losowej X .

Funkcja prawdopodobieństwa warunkowego $Y|X=x_i$ wyraża się wzorem

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

w którym zdarzenie $X=x_i$ jest ustalone, a y_j przebiega wszystkie wartości zmiennej losowej Y .

Przykład

Zmienna losowa X oznacza cenę komputera (w tys.zł), a Y liczba awarii tego komputera w czasie T . Zmienna losowa dwuwymiarowa (X, Y) ma funkcję prawdopodobieństwa przedstawioną w tabeli

$y_j \backslash x_i$	0	1	2	3	4	5
2	0	0,01	0,02	0,02	0,06	0,06
3	0,01	0,02	0,03	0,02	0,05	0,04
4	0,02	0,03	0,04	0,04	0,04	0
5	0,03	0,05	0,05	0,01	0,03	0
6	0,04	0,07	0,04	0,01	0	0
7	0,05	0,08	0,03	0	0	0

(a) Wyznaczyć rozkłady brzegowe tych zmiennych.

(b) Wyznaczyć rozkład awarii komputerów kosztujących 7 tys. zł oraz rozkład ceny komputerów mających 4 awarie w ciągu czasu T .

(a) Rozkład brzegowy z.l. X (struktura komputerów wg ceny)

x_i	2	3	4	5	6	7
$p_{i\cdot}$	$p_{1\cdot}$	$p_{2\cdot}$	$p_{3\cdot}$	$p_{4\cdot}$	$p_{5\cdot}$	$p_{6\cdot}$

gdzie $p_{1\cdot} = \sum_{j=1}^6 p_{1j} = P(X = x_1, Y = y_1) + P(X = x_1, Y = y_2) + P(X = x_1, Y = y_3) + P(X = x_1, Y = y_4) +$

$P(X = x_1, Y = y_5) + P(X = x_1, Y = y_6) = 0 + 0,01 + 0,02 + 0,02 + 0,06 + 0,06 = 0,17$

$p_{2\cdot} = \sum_{j=1}^6 p_{2j} = 0,17$ $p_{3\cdot} = \sum_{j=1}^6 p_{3j} = 0,17$ $p_{4\cdot} = \sum_{j=1}^6 p_{4j} = 0,17$ $p_{5\cdot} = \sum_{j=1}^6 p_{5j} = 0,16$ $p_{6\cdot} = \sum_{j=1}^6 p_{6j} = 0,16$

Rozkład brzegowy z.l. Y (struktura komputerów wg liczby awarii)

y_j	0	1	2	3	4	5
$p_{\cdot j}$	0,15	0,26	0,21	0,1	0,18	0,1

	0	1	2	3	4	5	X
2	0	0,01	0,02	0,02	0,06	0,06	0,17
3	0,01	0,02	0,03	0,02	0,05	0,04	0,17
4	0,02	0,03	0,04	0,04	0,04	0	0,17
5	0,03	0,05	0,05	0,01	0,03	0	0,17
6	0,04	0,07	0,04	0,01	0	0	0,16
7	0,05	0,08	0,03	0	0	0	0,16
Y	0,15	0,26	0,21	0,1	0,18	0,1	

Interpretacja $P(X = 2) = 0,17$ $P(X = 6) = 0,16$, $P(Y = 3) = 0,1$ $P(Y = 4) = 0,18$

(b) Aby znaleźć rozkład awarii komputerów kosztujących 7 tys. zł tak naprawdę musimy wyznaczyć rozkład warunkowy zmiennej $Y/X=7$

$P(Y=0/X=7)=0,05/0,16= 5/16$

$P(Y=1/X=7)=0,08/0,16= 8/16$

$P(Y=2/X=7)=0,03/0,16=3/16$

Rozkład awarii komputerów kosztujących 7 tys. zł (funkcja prawdopodobieństwa warunkowa $Y|X=7$) ($p_{\cdot 7} = 0,05+0,08+0,03=0,16$. $0,05/0,16=5/16, \dots$

y_j	0	1	2	3	4	5
$P(Y=y_j X=7)$	5/16	8/16	3/16	0	0	0

Aby znaleźć rozkład ceny komputerów mających 4 awarie w ciągu czasu T musimy wyznaczyć rozkład warunkowy zmiennej $X|Y=4$.

x_i	2	3	4	5	6	7
$P(X=x_i Y=4)$	$1/3=0.06/0.18$	$5/18$	$2/9$	$1/6$	0	0

DWUWYMIAROWA CIĄGŁA ZMIENNA LOSOWA

Zmienna losowa dwuwymiarowa ciągła jest zmienną, której dystrybuantę F można przedstawić w postaci

$$F(x, y) = \int_{-\infty}^x \left[\int_{-\infty}^y f(t, u) du \right] dt \text{ dla } x, y \in \mathbb{R}$$

gdzie f jest funkcją nieujemną dwóch zmiennych rzeczywistych zwaną *gęstością prawdopodobieństwa* zmiennej losowej (X, Y) .

Własności gęstości dwuwymiarowej zmiennej losowej (X, Y) :

1. f jest funkcją nieujemną: $f(x, y) \geq 0$ dla $x, y \in \mathbb{R}$,
2. $\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx = 1$,
3. $P(a < X < b, c < Y < d) = \int_a^b \left[\int_c^d f(x, y) dy \right] dx$
4. Jeśli F jest dystrybuantą zmiennej losowej dwuwymiarowej ciągłej to funkcja

$$f(x, y) = \begin{cases} \frac{\partial^2}{\partial x \partial y} F(x, y) & \text{gdy pochodna ta istnieje} \\ 0 & \text{w p.p.} \end{cases}$$

jest gęstością tej zmiennej.

Uwaga

Każda funkcja spełniająca warunki 1 i 2 jest gęstością pewnej zmiennej losowej ciągłej.

Uwaga

Znamy rozkład prawdopodobieństwa dwuwymiarowej zmiennej losowej ciągłej, jeśli znamy dystrybuantę tej zmiennej lub gęstość.

Przykład

Sprawdzić czy funkcja

$$f(x, y) = \begin{cases} x + y & \text{dla } 0 < x < 1, 0 < y < 1, \\ 0 & \text{dla pozostałych.} \end{cases}$$

jest gęstością zmiennej losowej dwuwymiarowej ciągłej (X, Y) .

Funkcja jest gęstością jeżeli spełnia dwa warunki:

1) f jest funkcją nieujemną?

Oczywiście, ponieważ $f(x, y) = 0$ lub $f(x, y) > 0$ gdy $0 < x < 1$ i $0 < y < 1$

2) Spełniony warunek $\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx = 1$

$$\begin{aligned} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx &= \int_0^1 \left[\int_0^1 (x + y) dy \right] dx \\ &= \int_0^1 \left[xy + \frac{y^2}{2} \Big|_0^1 \right] dx = \\ &= \int_0^1 \left[x(1 - 0) + \frac{1^2 - 0^2}{2} \right] dx = \\ &= \int_0^1 \left[x + \frac{1}{2} \right] dx = \frac{x^2}{2} + \frac{1}{2}x \Big|_0^1 = \frac{1^2 - 0^2}{2} + \frac{1}{2}(1 - 0) = \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

Warunek 2 jest spełniony, więc (X, Y) jest gęstością.

ROZKŁADY BRZEGOWE ZMIENNYCH LOSOWYCH TYPU CIĄGŁEGO

Uwaga

Rozkład prawdopodobieństwa zmiennej losowej dwuwymiarowej ciągłej (X, Y) wyznacza rozkład ciągły zmiennej losowej X i rozkład ciągły zmiennej losowej Y .

Rozkłady brzegowe zmiennych losowych ciągłych X i Y są to rozkłady prawdopodobieństwa tych zmiennych wyznaczone za pomocą dwuwymiarowej zmiennej losowej ciągłej (X, Y) .

Jeśli (X, Y) jest zmienną losową dwuwymiarową ciągłą o gęstości $f(x, y)$ to

1. **Gęstość brzegowa** z. l. X jest postaci

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \text{ dla } x \in \mathbb{R},$$

a **dystrybuanta brzegowa** zmiennej losowej X

$$F_X(x) = \int_{-\infty}^x f_X(u) du \text{ dla } x \in \mathbb{R}.$$

2. **Gęstość brzegowa** z. l. Y jest postaci

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \text{ dla } y \in \mathbb{R},$$

A **dystrybuanta brzegowa** zmiennej losowej Y

$$F_Y(y) = \int_{-\infty}^y f_Y(u) du \text{ dla } y \in \mathbb{R}.$$

ROZKŁADY WARUNKOWE ZMIENNYCH LOSOWYCH TYPU CIĄGŁEGO

Niech (X,Y) będzie dwuwymiarową z. l. ciągłą o gęstości $f(x,y)$.

Gęstość warunkowa zmiennej losowej $X/Y=y$ wyraża się wzorem

$$f_X(x|Y = y) = \frac{f(x,y)}{f_Y(y)} \text{ dla } x \in \mathbb{R} \text{ oraz } f_Y(y) > 0.$$

Gęstość warunkowa zmiennej losowej $Y/X=x$ wyraża się wzorem

$$f_Y(y|X = x) = \frac{f(x,y)}{f_X(x)} \text{ dla } y \in \mathbb{R} \text{ oraz } f_X(x) > 0.$$

Przykład

Zmienna losowa ciągła (X,Y) ma rozkład o gęstości

$$f(x,y) = \begin{cases} \frac{3}{4}xy^2 & \text{dla } 0 < x < 1, 0 < y < 2 \\ 0 & \text{dla pozostałych} \end{cases}$$

Wyznaczyć gęstości brzegowe. Obliczyć $F(0.5,0.75)$, $P(0.25 < X < 0.5, 0 < Y < 0.5)$.

Gęstość brzegowa zmiennej losowej X jest postaci:

$$1. \quad x \in (0,1) \quad f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy = \int_0^2 \frac{3}{4}xy^2 dy = \frac{3}{4}x \frac{y^3}{3} \Big|_0^2 = \frac{x}{4}(2^3 - 0^3) = 2x,$$

$$2. \quad x \notin (0,1) \quad f_X(x) = 0.$$

Stąd
$$f_X(x) = \begin{cases} 2x & \text{dla } 0 < x < 1, \\ 0 & \text{dla pozostałych.} \end{cases}$$

Gęstość brzegowa zmiennej losowej Y wyraża się wzorem:

$$1. \quad y \in (0,2) \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx = \int_0^1 \frac{3}{4}xy^2 dx = \frac{3}{4}y^2 \frac{x^2}{2} \Big|_0^1 = \frac{3y^2}{8}(1^2 - 0^2) = \frac{3}{8}y^2,$$

$$2. \quad y \notin (0,2) \quad f_Y(y) = 0.$$

Stąd
$$f_Y(y) = \begin{cases} \frac{3}{8}y^2 & \text{dla } 0 < y < 2, \\ 0 & \text{dla pozostałych.} \end{cases}$$

$$F(0.5,0.75) = P(X < 0.5, Y < 0.75) =$$

$$\begin{aligned} \int_{-\infty}^{0.5} \int_{-\infty}^{0.75} f(x,y) dy dx &= \int_0^{0.5} \int_0^{0.75} \frac{3}{4}xy^2 dy dx = \int_0^{0.5} \frac{3}{4}x \frac{y^3}{3} \Big|_0^{0.75} dx \\ &= \frac{27}{256} \int_0^{0.5} x dx = \frac{27}{256} \frac{x^2}{2} \Big|_0^{0.5} = \frac{27}{2048} \end{aligned}$$

$$P(0.25 < X < 0.5, 0 < Y < 0.5) =$$

$$\int_{0.25}^{0.5} \int_0^{0.5} \frac{3}{4}xy^2 dy dx = \int_{0.25}^{0.5} \frac{3}{4}x \frac{y^3}{3} \Big|_0^{0.5} dx = \frac{1}{32} \int_{0.25}^{0.5} x dx = \frac{1}{32} \frac{x^2}{2} \Big|_{0.25}^{0.5} = \frac{3}{1024}.$$

WSPÓŁCZYNNIK KORELACJI

Kowariancja zmiennych losowych X i Y nazywamy wielkość

$$\begin{aligned} cov(X, Y) &= E[(X - EX)(Y - EY)] = EXY - EXEY = \\ &= \begin{cases} \sum_i \sum_j x_i x_j p_{ij} - \sum_i x_i p_{i\cdot} \cdot \sum_j y_j p_{\cdot j} & \text{dla zm. los. skończonych} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dx dy - \int_{-\infty}^{\infty} xf_X(x) dx \cdot \int_{-\infty}^{\infty} yf_Y(y) dy & \text{dla zm. los. ciągłych} \end{cases} \end{aligned}$$

Współczynnik korelacji zmiennych losowych X i Y

$$\rho = \frac{cov(X, Y)}{DXY}.$$

Własności współczynnika korelacji

1. Jeśli zmienne losowe X i Y są niezależne, to współczynnik korelacji jest równy zeru.
2. Współczynnik korelacji jest liczbą między -1 a 1 ($-1 \leq \rho \leq 1$).
3. Wartość bezwzględna współczynnika korelacji jest równa 1 wtedy i tylko wtedy, gdy zmienne losowe są liniowo zależne.
4. Jeżeli wektor losowy ma rozkład dwuwymiarowy normalny to zmienne losowe są niezależne wtedy i tylko wtedy gdy korelacja jest zerowa.

Przykład

Zmienna losowa dwuwymiarowa skokowa ma funkcje prawdopodobieństwa określoną następująco:

y_j, x_i	-1	0	1
1	$\frac{1}{11}$	$\frac{3}{11}$	$\frac{2}{11}$
3	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{2}{11}$

y_j, x_i	-1	0	1	
1	$\frac{1}{11}$	$\frac{3}{11}$	$\frac{2}{11}$	$\frac{6}{11}$
3	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{5}{11}$
	$\frac{3}{11}$	$\frac{4}{11}$	$\frac{4}{11}$	

Obliczyć współczynnik korelacji zmiennych losowych X i Y .

Współczynnik korelacji z. l. X i Y wyraża się wzorem

$$\rho = \frac{cov(X, Y)}{DXDY}$$

gdzie

$$cov(X, Y) = EXY - EXEY = \sum_i \sum_j x_i x_j p_{ij} - \sum_i x_i p_{i\cdot} \cdot \sum_j y_j p_{\cdot j}$$

$$EXY = \sum_i \sum_j x_i x_j p_{ij} = -1 \cdot 1 \cdot \frac{1}{11} - 1 \cdot 3 \cdot \frac{2}{11} + 0 \cdot 1 \cdot \frac{3}{11} + 0 \cdot 3 \cdot \frac{2}{11} + 1 \cdot 1 \cdot \frac{2}{11} + 3 \cdot 1 \cdot \frac{2}{11} = \frac{1}{11}$$

$$EX = \sum_{i=1}^3 x_i p_{i\cdot} = -1 \cdot \frac{3}{11} + 0 \cdot \frac{4}{11} + 1 \cdot \frac{4}{11} = \frac{1}{11}$$

$$EY = \sum_{j=1}^2 y_j p_{\cdot j} = 1 \cdot \frac{6}{11} + 3 \cdot \frac{5}{11} = \frac{21}{11}$$

$$\text{cov}(X, Y) = EXY - EXEY = \frac{1}{11} - \frac{1}{11} \cdot \frac{21}{11} = -\frac{10}{121}$$

$$DX = \sqrt{D^2X} = \sqrt{EX^2 - (EX)^2} = \sqrt{\frac{7}{11} - \frac{1}{121}} = \sqrt{\frac{76}{121}} = \frac{\sqrt{76}}{11}$$

$$EX^2 = (-1)^2 \cdot \frac{3}{11} + 0^2 \cdot \frac{4}{11} + 1^2 \cdot \frac{4}{11} = \frac{7}{11}$$

$$DY = \sqrt{D^2Y} = \sqrt{EY^2 - (EY)^2} = \sqrt{\frac{51}{11} - \frac{441}{121}} = \sqrt{\frac{120}{121}} = \frac{\sqrt{120}}{11}$$

$$EY^2 = 1^2 \cdot \frac{6}{11} + 3^2 \cdot \frac{5}{11} = \frac{21}{11}$$

$$\rho = \frac{\text{cov}(X, Y)}{DXDY} = \frac{-\frac{10}{121}}{\frac{\sqrt{76} \cdot \sqrt{120}}{11 \cdot 11}} \approx -0.1.$$

Wobec tego, zmienne losowe X i Y są zależne.

NIEZALEŻNOŚĆ ZMIENNYCH LOSOWYCH

Zmienne losowe X i Y nazywamy niezależnymi, jeżeli

$$F(x,y)=F_X(x)\cdot F_Y(y) \quad \text{dla } x,y\in\mathbb{R},$$

gdzie $F(x,y)$ jest dystrybuanta zmiennej losowej dwuwymiarowej (X,Y)

$F_X(x)$ -dystrybuanta zmiennej losowej X ,

$F_Y(y)$ -dystrybuanta zmiennej losowej Y .

Zmienne losowe skokowe X i Y są niezależnymi wtedy i tylko wtedy gdy

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

dla każdego (x_i, y_j) będącego wartością zmiennej losowej dwuwymiarowej (X,Y) .

Zmienne losowe ciągłe X i Y są niezależnymi wtedy i tylko wtedy gdy

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

gdzie $f(x,y)$ jest gęstością zmiennej losowej dwuwymiarowej (X,Y)

$f_X(x)$ -gęstość brzegowa zmiennej losowej X ,

$f_Y(y)$ -gęstość brzegowa zmiennej losowej Y .

Zmienne losowe X i Y nie będące zmiennymi losowymi niezależnymi nazywamy zmiennymi losowymi zależnymi.

Przykład

Zmienna losowa ciągła (X,Y) ma rozkład o gęstości $f(x,y) =$

$$\begin{cases} \frac{3}{4}xy^2 & \text{dla } 0 < x < 1, 0 < y < 2. \\ 0 & \text{dla pozostałych.} \end{cases}$$

Zbadać niezależność zmiennych losowych X i Y .

Gęstość brzegowa zmiennej losowej X jest postaci

$$f_X(x) = \begin{cases} 2x & \text{dla } 0 < x < 1 \\ 0 & \text{dla pozostałych} \end{cases}$$

Gęstość brzegowa zmiennej losowej Y jest postaci

$$f_Y(y) = \begin{cases} \frac{3}{8}y^2 & \text{dla } 0 < y < 2 \\ 0 & \text{dla pozostałych} \end{cases}$$

Zmienne losowe X i Y są niezależne jeśli $f(x,y)=f_X(x) \cdot f_Y(y)$

Niech $x \notin (0,1)$ i $y \notin (0,2)$ wówczas $0=f(x,y)=f_X(x) \cdot f_Y(y)=0 \cdot 0$

Niech $x \in (0,1)$ i $y \in (0,2)$ wówczas $f(x,y) = \frac{3}{4}xy^2 = 2x \cdot \frac{3}{8}y^2 = f_X(x) \cdot f_Y(y)$

Zmienne X i Y są więc niezależne.

Centralne twierdzenia graniczne

Jedną z najważniejszych własności rozkładu normalnego jest fakt, że (przy pewnych założeniach) rozkład sumy dużej liczby zmiennych losowych jest w przybliżeniu normalny. Jest to tak zwane centralne twierdzenie graniczne skrót CTG.

W praktyce twierdzenie to ma zastosowanie, jeśli chcemy rozkładem normalnym przybliżyć inny rozkład.

Założenie.

(Ω, F, P) - przestrzeń probabilistyczna,

X_1, X_2, \dots, X_n – ciąg niezależnych zmiennych losowych. Zmienne X_i mają taki sam rozkład, a ich wspólna wartość oczekiwana m oraz wariancja σ^2 istnieją i są skończone, przy czym $\sigma > 0$ (ten ostatni warunek oznacza, że zmienne losowe nie są stałymi). Niech

$$S_n = X_1 + X_2 + \dots + X_n$$

Twierdzenie (Centralne tw. graniczne dla sum)

Rozkład zmiennej losowej S_n jest asymptotycznie równy rozkładowi normalnemu $N(nm, \sigma\sqrt{n})$.

Przykład.

Wielokrotnie rzucamy kostką. Suma uzyskanych oczek S jest zmienną losową mającą, zgodnie z cytowanym twierdzeniem, w przybliżeniu rozkład $N(nm, \sigma\sqrt{n})$, gdzie m oraz σ są odpowiednio wartością oczekiwaną oraz odchyleniem standardowym zmiennej losowej X , reprezentującej wynik pojedynczego rzutu, a n jest liczbą wykonanych rzutów. Ponieważ X ma rozkład dyskretny i przyjmuje wartości 1,2,3,4,5,6 z jednakowym prawdopodobieństwem $\frac{1}{6}$, więc

$$m = 3.5 \text{ oraz } \sigma = \frac{\sqrt{105}}{6} \approx 1.7078251.$$

$$3.5 = \frac{1 + 2 + 3 + 4 + 5 + 6}{6}$$

$$\frac{105}{36}$$

$$= \frac{(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2}{6}$$

Przykład.

Rzucono 1000 razy kostką. Obliczyć prawdopodobieństwo tego, że "szóstka" wypadła więcej niż 150 razy.

$$X_i = \begin{cases} 1 & \text{gdy wypadnie z } i \text{ tym razem } 6 \text{ z pr. } 1/6 \\ 0 & \text{gdy nie wypadnie z } i \text{ tym razem } 6 \text{ z pr. } 5/6 \end{cases}$$

Mamy 1000 niezależnych prób. Prawdopodobieństwo sukcesu (szóstka) wynosi 1/6. Rozkład dwumianowy (Bernoulliego).

Z centralnego twierdzenia granicznego suma szóstek S_{1000} ma w przybliżeniu rozkład $N(np, \sqrt{npq})$. Wobec tego

$$\begin{aligned} P(S_{1000} > 150) &= 1 - P(S_{1000} \leq 150) \approx 1 - \Phi_{1000 \cdot \frac{1}{6}, \sqrt{1000 \cdot \frac{1}{6} \cdot \frac{5}{6}}}(150) \\ &= 1 - \Phi\left(\frac{150 - \frac{1000}{6}}{\sqrt{\frac{5000}{36}}}\right) \approx 1 - \Phi(-1.41) = \Phi(1.41) \approx 0.9207, \end{aligned}$$

Przykład.

Jakie jest prawdopodobieństwo, że przy 1000 rzutach monetą, różnica między liczbą reszek i orłów będzie wynosić co najmniej 100?

Rozkład Bernoulliego z prawdopodobieństwem sukcesu $1/2$.

Należy obliczyć:

$$P(|S_{1000} - (1000 - S_{1000})| \geq 100) = P(|S_{1000} - 500| \geq 50).$$

Łatwo obliczyć, że prawdopodobieństwo zdarzenia wynosi

$$\begin{aligned} F_{S_{1000}}(550) - F_{S_{1000}}(450) &\approx \Phi_{500, 5\sqrt{10}}(550) - \Phi_{500, 5\sqrt{10}}(450) \\ &= \Phi(\sqrt{10}) - \Phi(-\sqrt{10}) = 2\Phi(\sqrt{10}) - 1 \approx 2\Phi(3.16227766) - 1 \approx 0.9984346. \end{aligned}$$

Wobec tego szukane prawdopodobieństwo jest w przybliżeniu wynosi 0.0016 .

$$\sqrt{10} = (550 - 500)/(5\sqrt{10})$$

TWIERDZENIE (CENTRALNE TW. GRANICZNE)

Rozkład zmiennej losowej $\frac{S_n}{\sqrt{n}}$ jest asymptotycznie równy rozkładowi normalnemu $N(m, \sigma)$.