5 DWU- I WIELOWYMIAROWE ZMIENNE LOSOWE

5.1. DWUWYMIAROWA ZMIENNA LOSOWA I JEJ ROZKŁAD PRAWDOPODOBIEŃSTWA

5.1.1. Dystrybuanta dwuwymiarowej zmiennej losowej. Przypuśćmy, że badamy pewną zbiorowość ze względu na dwie cechy np. śrubki ze względu na średnicę przekroju i długość, ludzi ze względu na ciężar i wzrost, włókna bawełny – na długość i wytrzymałość, wiek małżonków itp. Zdarzeniu elementarnemu (losowo wybrana śrubka, człowiek, włókno bawełny, małżeństwo) zostały przyporządkowane pary uporządkowanych liczb rzeczywistych. Podobnie będzie przy rzucie dwiema kostkami sześciennymi, gdzie każdemu rzutowi jest przyporządkowana dokładnie jedna para liczb (i, k), gdzie $i, k=1, \ldots, 6$. Te przykładowo podane pary zmiennych będziemy nazywali *dwuwymiarowymi zmiennymi losowymi.*

Niech X oraz Y będą zmiennymi losowymi określonymi niekoniecznie na tej samej przestrzeni probabilistycznej. Parę (X, Y) zmiennych losowych X, Y nazywamy *dwuwymiarową zmienną losową* lub *dwuwymiarowym wektorem losowym,* a X oraz Y jej *współrzędnymi.*

Dystrybuantą dwuwymiarowej zmiennej losowej (X, Y) nazywamy funkcję F zmiennych x, y, która dla każdej pary liczb rzeczywistych $(x, y) \in \boldsymbol{R}^2$ przyjmuje wartości równe prawdopodobieństwu zdarzenia polegającego na tym, że zmienna losowa X przyjmie wartość mniejszą od x i zmienna losowa Y przyjmie wartość mniejszą od y:

$$F(x, y)=P(X<x,\ Y<y) \quad \text{dla} \quad (x, y) \in \boldsymbol{R}^2. \tag{5.1.1}$$

F nazywamy także *dystrybuantą łącznej zmiennej losowej* (X, Y).

Własności dystrybuanty.

a) $\bigwedge\limits_{x \in \boldsymbol{R}} \lim\limits_{y \to -\infty} F(x, y)=0, \quad \bigwedge\limits_{y \in \boldsymbol{R}} \lim\limits_{x \to -\infty} F(x, y)=0,$

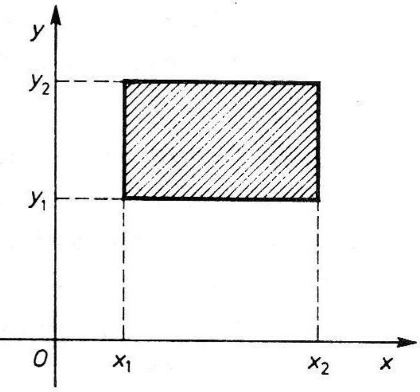
b) $\lim\limits_{\substack{x \to \infty \\ y \to \infty}} F(x, y)=1,$

(5.1.2)

c) Dla dowolnych punktów: (x_1, y_1), (x_2, y_2) takich, że $x_1 \leqslant x_2$ i $y_1 \leqslant y_2$, zachodzi nierówność:

$$F(x_2, y_2)-F(x_2, y_1)-F(x_1, y_2)+F(x_1, y_1) \geqslant 0.$$

Warunek ten okaże się oczywisty, jeśli zauważymy, że lewa strona ostatniej nierówności jest równa $P(x_1 \leqslant X<x_2,\ y_1 \leqslant Y<y_2)$ (rys. 5.1).



Rys. 5.1. $P(x_1 \leqslant X < x_2, y_1 \leqslant Y < y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geqslant 0$.

d) Dystrybuanta jest funkcją niemalejącą i co najmniej lewostronnie ciągłą względem każdego z argumentów x bądź y.

Każda funkcja dwóch zmiennych spełniająca warunki a) - d) może być traktowana jako dystrybuanta dwuwymiarowej zmiennej losowej (X, Y). W zastosowaniach najczęściej spotykamy dwuwymiarowe zmienne losowe typu skokowego bądź typu ciągłego.

5.1.2. Dwuwymiarowe zmienne losowe typu skokowego. Dwuwymiarową zmienną losową (X, Y), która przyjmuje skończoną, bądź przeliczalną liczbę wartości (x_i, y_k), każdą odpowiednio z prawdopodobieństwem:

$$P(X=x_i,\ Y=y_k)=p_{ik} \quad \text{dla} \quad i, k \in N, \tag{5.1.3}$$

przy czym $\sum_i \sum_k p_{ik}=1$, nazywamy *dwuwymiarową zmienną losową skokową* (*dyskretną*) (zad. 5.2). Funkcję (5.1.3), która wartościom (x_i, y_k) przyporządkowuje odpowiednie prawdopodobieństwa p_{ik}, nazywamy *funkcją prawdopodobieństwa dwuwymiarowej zmiennej losowej* (X, Y).

Znając funkcję (5.1.3) można wyznaczyć dystrybuantę dwuwymiarowej zmiennej losowej (X, Y) i odwrotnie. Jeśli dwuwymiarowa zmienna losowa (X, Y) przyjmuje skończoną liczbę wartości, to wygodnie jest umieścić wartości funkcji prawdopodobieństwa (5.1.3) w tabelce dwudzielczej:

	x_i				
y_k	x_1	x_2	...	x_m	$p_{.k}$
y_1	p_{11}	p_{21}	...	p_{m1}	$p_{.1}$
y_2	p_{12}	p_{22}	...	p_{m2}	$p_{.2}$
...	...	...	...	...	...
y_s	p_{1s}	p_{2s}	...	p_{ms}	$p_{.s}$
$p_{i.}$	$p_{1.}$	$p_{2.}$	...	$p_{m.}$	1

Powiemy, że został wyznaczony rozkład prawdopodobieństwa dwuwymiarowej zmiennej losowej skokowej, gdy znana jest jej dystrybuanta albo funkcja prawdopodobieństwa (5.1.3).

ZADANIE 5.1. Z talii 52 kart wylosowano 1 kartę. Niech zmienna losowa X przyjmuje wartości równe liczbie wylosowanych asów, zaś Y – liczbie wylosowanych pików. Wyznaczyć rozkład prawdopodobieństwa dwuwymiarowej zmiennej losowej (X, Y).

Rozwiązanie. Ponieważ losujemy tylko 1 kartę, każda ze zmiennych X, Y może przyjmować z dodatnim prawdopodobieństwem tylko dwie wartości 0 albo 1.

Oznaczmy przez:

A zdarzenie polegające na wylosowaniu asa pik,

B zdarzenie polegające na wylosowaniu karty pikowej, która nie jest asem,

C zdarzenie polegające na wylosowaniu asa, który nie jest pikiem,

D zdarzenie polegające na wylosowaniu karty, która nie jest asem i nie jest pikiem,

$$P(X=0,\ Y=0)=P(D)=\tfrac{36}{52}, \quad P(X=1,\ Y=0)=P(C)=\tfrac{3}{52},$$
$$P(X=0,\ Y=1)=P(B)=\tfrac{12}{52}, \quad P(X=1,\ Y=1)=P(A)=\tfrac{1}{52}.$$

	x_i	
y_k	0	1
0	$\frac{36}{52}$	$\frac{3}{52}$
1	$\frac{12}{52}$	$\frac{1}{52}$

5.1.3. Rozkłady brzegowe zmiennych losowych typu skokowego. Oznaczmy:

$$p_{i.}=\sum_k p_{ik} \quad \text{dla} \quad i\in N, \tag{5.1.4}$$

$$p_{.k}=\sum_i p_{ik} \quad \text{dla} \quad k\in N. \tag{5.1.5}$$

Zauważmy, że $p_{i.}=P(X=x_i,\ Y=y_1)+P(X=x_i,\ Y=y_2)+\ldots$ jest prawdopodobieństwem tego, że zmienna losowa X przyjmie wartość równą x_i, bez względu na to, którą z wartości: $y_1, y_2, \ldots$ przyjmuje zmienna losowa Y, oraz, że $\sum_i p_{i.}=1$, a więc funkcja:

$$P(X=x_i)=p_{i.}, \quad i\in N \tag{5.1.6}$$

wyznacza rozkład prawdopodobieństwa zmiennej losowej X, nazywany rozkładem brzegowym (marginalnym) zmiennej losowej X w rozkładzie dwuwymiarowej zmiennej losowej (X, Y) (jak widać z tabelki – $p_{i.}$ zapisuje się na brzegu tabelki). Podobnie określamy rozkład brzegowy zmiennej losowej Y jako rozkład prawdopodobieństwa określony wzorem:

$$P(Y=y_k)=p_{.k}, \quad k\in N. \tag{5.1.7}$$

Oznaczmy przez F_1 i F_2 dystrybuanty rozkładów brzegowych zmiennych losowych X i Y odpowiednio. Jeśli (X, Y) jest dwuwymiarową zmienną losową skokową, to

$$F_1(x) = \sum_{x_i < x} p_{i.} \quad \text{dla} \quad x \in \boldsymbol{R}, \tag{5.1.8}$$

$$F_2(y) = \sum_{y_k < y} p_{.k} \quad \text{dla} \quad y \in \boldsymbol{R}. \tag{5.1.9}$$

ZADANIE 5.2. Pewien mechanizm składa się z dwóch kół zębatych: dużego i małego. Warunki techniczne przy montażu urządzenia zostają naruszone, jeśli w obu kołach występują dodatnie odchylenia grubości zębów od nominalnego wymiaru. Robotnik dysponuje 2 kołami zębatymi dużymi: „plusowym" i „minusowym" i dwoma małymi „plusowym" i „minusowym". Rozważmy zero-jedynkowe zmienne losowe X i Y: zmienna losowa X przyjmuje wartość 1, jeśli robotnik wylosuje duże koło „plusowe" i 0 jeśli duże koło „minusowe". Analogicznie określona jest zmienna losowa Y w przypadku koła małego. a) Wyznaczyć i naszkicować dystrybuantę F dwuwymiarowej zmiennej losowej (X, Y). b) Obliczyć prawdopodobieństwo naruszenia warunków technicznych przy montażu mechanizmu.

Rozwiązanie. a) Dwuwymiarowa zmienna losowa (X, Y) przyjmuje wartości $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$ z prawdopodobieństwami:

$$P(X=0,\, Y=0) = P(A \cap B) = P(A)P(B) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4},$$

$$P(X=1,\, Y=0) = P(A' \cap B) = P(A')P(B) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4},$$

$$P(X=0,\, Y=1) = P(A \cap B') = P(A)P(B') = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4},$$

$$P(X=1,\, Y=1) = P(A' \cap B') = P(A')P(B') = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4},$$

gdzie A jest zdarzeniem polegającym na wylosowaniu dużego koła zębatego „minusowego", B — małego koła „minusowego", zaś A', B' są odpowiednio zdarzeniami przeciwnymi do A i B. Zestawmy otrzymane wyniki w tabelce:

	x_i	
y_k	0	1
0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$

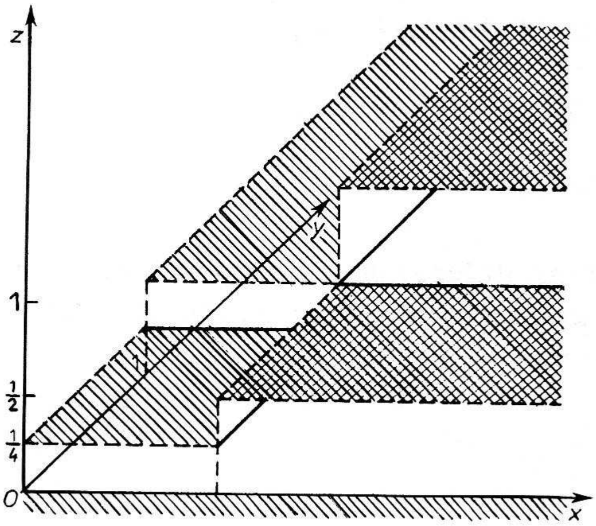
Stosując wzór (5.1.1) wyznaczymy dystrybuantę F dwuwymiarowej zmiennej losowej (X, Y): dla wygody umieśćmy jej wartości w tabelce:

	x		
y	$(-\infty, 0\rangle$	$(0, 1\rangle$	$(1, \infty)$
$(-\infty, 0\rangle$	0	0	0
$(0, 1\rangle$	0	$\frac{1}{4}$	$\frac{1}{2}$
$(1, \infty)$	0	$\frac{1}{2}$	1

Szukana dystrybuanta (rys. 5.2) jest więc określona wzorami:

$$F(x, y)=\begin{cases} 0 & \text{dla} \quad x\leqslant 0 \vee y\leqslant 0, \\ \frac{1}{4} & \text{dla} \quad 0<x\leqslant 1 \wedge 0<y\leqslant 1, \\ \frac{1}{2} & \text{dla} \quad x>1 \wedge 0<y\leqslant 1, \\ \frac{1}{2} & \text{dla} \quad 0<x\leqslant 1 \wedge y>1, \\ 1 & \text{dla} \quad x>1 \wedge y>1. \end{cases}$$

b) Naruszenie warunków technicznych montażu nastąpi wtedy, gdy robotnik wybierze losowo duże koło zębate „plusowe" i małe „plusowe"; jest to koniunkcja zdarzeń niezależnych A', B', a więc $P(A' \cap B')=P(A')P(B')=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$.



Rys. 5.2. Dystrybuanta dwuwymiarowej zmiennej losowej typu skokowego (zad. 5.2)

ZADANIE 5.3. Dwuwymiarowa zmienna losowa (X, Y) ma rozkład określony w tabelce:

	x_i		
y_k	1	2	3
2	0,1	0,2	0,3
4	0,1	0,1	0,2

Wyznaczyć dystrybuantę rozkładu brzegowego zmiennej losowej Y.

Rozwiązanie. Aby wyznaczyć rozkład brzegowy (5.1.7) zmiennej Y sumujemy prawdopodobieństwa w tabelce dwudzielczej w wierszach i tak:

$$P(Y=2)=0{,}1+0{,}2+0{,}3=0{,}6,$$

$$P(Y=4)=0{,}1+0{,}1+0{,}2=0{,}4.$$

Rozkład prawdopodobieństwa zmiennej Y jest więc rozkładem dwupunktowym:

y_k	2	4
$p_{\cdot k}$	0,6	0,4

Dystrybuanta (5.1.9) tego rozkładu jest określona wzorami:

$$F_2(y)=\begin{cases}0 & \text{dla} \quad y\leqslant 2,\\ 0{,}6 & \text{dla} \quad 2<y\leqslant 4,\\ 1 & \text{dla} \quad y>4.\end{cases}$$

5.1.4. Rozkłady warunkowe zmiennych losowych typu skokowego. Załóżmy, że wszystkie prawdopodobieństwa $p_{.k}$ rozkładu brzegowego zmiennej losowej Y są dodatnie i rozważmy

$$P(X=x_i|Y=y_k)=\frac{p_{ik}}{p_{\cdot k}}. \tag{5.1.10}$$

Zauważmy, że:

1) $\bigwedge\limits_{i,k\in N} 0\leqslant\frac{p_{ik}}{p_{.k}}\leqslant 1$, ponieważ licznik jest jednym ze składników mianownika (5.1.5),

2) $\bigwedge\limits_{k}\sum\limits_{i} P(X=x_i|Y=y_k)=1$, ponieważ

$$\sum_i P(X=x_i|Y=y_k)=\sum_i\frac{p_{ik}}{p_{.k}}=\frac{1}{p_{.k}}\sum_i p_{ik}=\frac{p_{.k}}{p_{.k}}=1.$$

Wzór (5.1.10) określa zatem rozkład prawdopodobieństwa zmiennej losowej X przy ustalonej wartości $Y=y_k$; nazywamy go *rozkładem warunkowym* zmiennej losowej X, pod warunkiem, że Y przyjmuje ustaloną wartość równą y_k. Podobnie można określić rozkład prawdopodobieństwa zmiennej losowej Y przy warunku, że X przyjmuje wartość x_i:

$$P(Y=y_k|X=x_i)=\frac{p_{ik}}{p_{i.}}, \qquad k\in N, \tag{5.1.11}$$

jeśli $\bigwedge\limits_{i} p_{i.}>0$.

Dystrybuanty rozkładów warunkowych oznaczmy odpowiednio przez: $F(x|y_k)$, $F(y|x_i)$. Podobnie jak dla innych rozkładów skokowych można je wyznaczyć według wzorów:

$$F(x|y_k)=P(X<x|Y=y_k)=\sum_{x_i<x}\frac{p_{ik}}{p_{.k}}, \tag{5.1.12}$$

$$F(y|x_i)=P(Y<y|X=x_i)=\sum_{y_k<y}\frac{p_{ik}}{p_{i.}}. \tag{5.1.13}$$

ZADANIE 5.4. Niech dwuwymiarowa zmienna losowa (X, Y) ma rozkład podany w tabelce:

	x_i				
y_k	2	3	3,5	4	5
2	$\frac{3}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	0	0
3	$\frac{1}{35}$	$\frac{10}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	0
3,5	0	$\frac{1}{35}$	$\frac{5}{35}$	$\frac{1}{35}$	$\frac{1}{35}$
4	0	0	0	$\frac{3}{35}$	$\frac{2}{35}$
5	0	0	0	$\frac{1}{35}$	$\frac{1}{35}$

gdzie X jest oceną klasówki z matematyki losowo wybranego ucznia pewnej klasy, Y zaś oceną klasówki z fizyki. Wyznaczyć rozkład warunkowy zmiennej losowej Y pod warunkiem, że $X=4$.

Rozwiązanie. Korzystamy ze wzoru (5.1.4). Sumując w tabelce dwudzielczej prawdopodobieństwa w kolumnie czwartej odpowiadającej $X=4$, otrzymamy: $p_{4.}=P(X=4)=\frac{6}{35}$. Rozkład warunkowy zmiennej losowej Y przy warunku $X=4$ otrzymamy ze wzoru (5.1.11)

$$P(Y=y_k|X=4)=\frac{P(X=4,\ Y=y_k)}{P(X=4)}=\frac{P(X=4,\ Y=y_k)}{\frac{6}{35}},$$

$$P(Y=2|X=4)=0,\quad P(Y=3|X=4)=\tfrac{1}{6},\quad P(Y=3{,}5|X=4)=\tfrac{1}{6},\quad P(Y=4|X=4)=\tfrac{3}{6},$$
$$P(Y=5|X=4)=\tfrac{1}{6}.$$

* * *

Innym typem dwuwymiarowych zmiennych losowych są tzw. *zmienne losowe typu ciągłego*, np. wytrzymałość i długość odcinka drutu wybranego losowo z pewnej partii, albo włókna bawełny wylosowanego z określonej beli, wzrost i masa człowieka z pewnej grupy ludzi.

5.1.5. Dwuwymiarowa zmienna losowa typu ciągłego. *Dwuwymiarową zmienną losową* (X, Y) nazywamy *typu ciągłego*, jeśli istnieje nieujemna funkcja f taka, że dystrybuanta tej zmiennej losowej da się przedstawić jako całka

$$F(x,y)=\int_{-\infty}^{x}\Big[\int_{-\infty}^{y}f(u,v)\,dv\Big]du \quad \text{dla} \quad (x,y)\in\boldsymbol{R}^2; \tag{5.1.14}$$

funkcję f nazywamy *gęstością rozkładu prawdopodobieństwa.* Z (5.1.14) i interpretacji całki wynika, że dystrybuantę F w punkcie (x_0, y_0) można traktować jako objętość bryły ograniczonej powierzchnią S o równaniu $z=f(x,y)$, płaszczyznami $x=x_0$, $y=y_0$ i tą częścią płaszczyzny Oxy, dla której $x<x_0$ i $y<y_0$ (rys. 5.3).