Rozkład wielowymiarowy

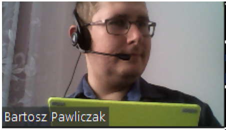
Korepetycje.intro23wertyk@gmail.com Bartosz Pawliczak

2021-05-19

Materiały ogólnodostępne przygotowane przeze mnie:
https://github.com/bpawliczak/intro23wertyk_public

Pełne zestawienie materiałów (tylko dla kursantów):
https://github.com/bpawliczak/intro23wertyk

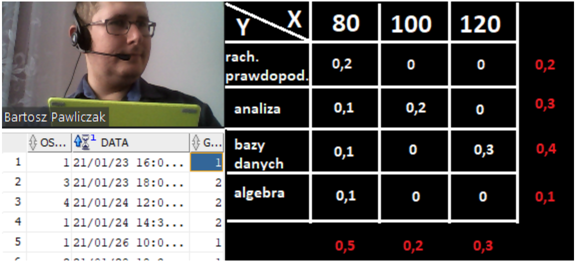
Więcej informacji i inspiracji:
https://www.facebook.com/intro23wertyk



	OS...	DATA	G...
1	1	21/01/23 16:0...	1
2	3	21/01/23 18:0...	2
3	4	21/01/24 12:0...	2
4	1	21/01/24 14:3...	2
5	1	21/01/26 10:0...	1

Y \ X	80	100	120	
rach. prawdopod.	0,2	0	0	0,2
analiza	0,1	0,2	0	0,3
bazy danych	0,1	0	0,3	0,4
algebra	0,1	0	0	0,1
	0,5	0,2	0,3	

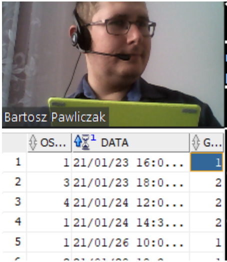
Rysunek: Korepetycje i stawka godzinowa (równoważnie: czy były realizowane w jedną, dwie, czy większą liczbę osób). Jaka stawka jest najpopularniejsza?



Y \ X	80	100	120	
rach. prawdopod.	0,2	0	0	0,2
analiza	0,1	0,2	0	0,3
bazy danych	0,1	0	0,3	0,4
algebra	0,1	0	0	0,1
	0,5	0,2	0,3	

Rysunek: Korepetycje i stawka godzinowa (równoważnie: czy były realizowane w jedną, dwie, czy większą liczbę osób). Jaka stawka jest najpopularniejsza?

Rozkład dyskretny możemy zapisać za pomocą tabelki, z której odczytujemy prawdopodobieństw konkretnego zdarzenia (np. jakie jest prawdopodobieństwo, że udzielanymi korepetycjami były bazy danych po stawce 80 zł/h), czyli **rozkład łączny**.



Y \ X	80	100	120	
rach. prawdopod.	0,2	0	0	0,2
analiza	0,1	0,2	0	0,3
bazy danych	0,1	0	0,3	0,4
algebra	0,1	0	0	0,1
	0,5	0,2	0,3	

Rysunek: Korepetycje i stawka godzinowa (równoważnie: czy były realizowane w jedną, dwie, czy większą liczbę osób). Jaka stawka jest najpopularniejsza?

Rozkład dyskretny możemy zapisać za pomocą tabelki, z której odczytujemy prawdopodobieństw konkretnego zdarzenia (np. jakie jest prawdopodobieństwo, że udzielanymi korepetycjami były bazy danych po stawce 80 zł/h), czyli **rozkład łączny**.
Możemy też sprawdzać rozkład względem konkretnej zmiennej, np. prawdopodobieństwo, że wybierając dowolną godzinę koreptycji natrafimy akurat na analizę. Jest to **rozkład brzegowy**.

Zmienne niezależne

Dwie zmienne X i Y są niezależne, jeżeli dla dowolnych zbiorów liczb rzeczywistych A i B zachodzi:

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

Zmienne niezależne

Dwie zmienne X i Y są niezależne, jeżeli dla dowolnych zbiorów liczb rzeczywistych A i B zachodzi:

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

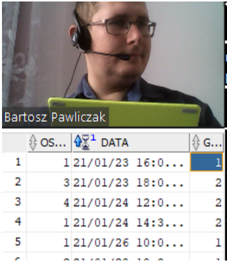
Jeżeli choć w jednym wypadku nie zachodzi równość, mówimy, że zmienne są zależne.

Zmienne niezależne

Dwie zmienne X i Y są niezależne, jeżeli dla dowolnych zbiorów liczb rzeczywistych A i B zachodzi:

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

Jeżeli choć w jednym wypadku nie zachodzi równość, mówimy, że zmienne są zależne.



Y \ X	80	100	120	
rach. prawdopod.	0,2	0	0	0,2
analiza	0,1	0,2	0	0,3
bazy danych	0,1	0	0,3	0,4
algebra	0,1	0	0	0,1
	0,5	0,2	0,3	

Rysunek: Czy te zmienne są niezależne?

Warunkowy rozkład

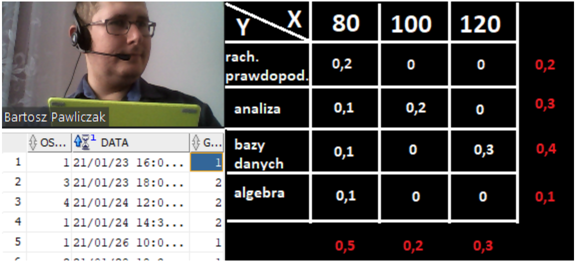
Dla dyskretnych zmiennych rozkład warunkowy ma postać:

$$\mathbb{P}(Y = y_i | X = x_j) = \frac{\mathbb{P}(X = x_j, Y = y_i)}{\mathbb{P}(X = x_j)}$$

Warunkowy rozkład

Dla dyskretnych zmiennych rozkład warunkowy ma postać:

$$\mathbb{P}(Y = y_i | X = x_j) = \frac{\mathbb{P}(X = x_j, Y = y_i)}{\mathbb{P}(X = x_j)}$$

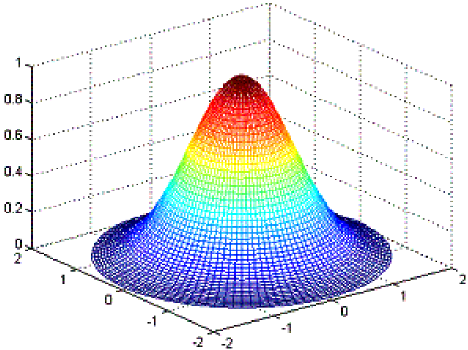


Y \ X	80	100	120	
rach. prawdopod.	0,2	0	0	0,2
analiza	0,1	0,2	0	0,3
bazy danych	0,1	0	0,3	0,4
algebra	0,1	0	0	0,1
	0,5	0,2	0,3	

Rysunek: Jaką interpretację ma rozkład warunkowy na przedstawionym przypadku?

W jaki sposób intrepretować rozkład ciągły?

W jaki sposób intrepretować rozkład ciągły? Można rozpatrywać punkty w układzie współrzędnych OXY, a wartość prawdopodobieństwa traktować jako wysokość w danym punkcie.



Rysunek: Jeżeli przedstawiony wykres wyznaczałby ciągły rozkład dwuwymiarowy, to ile wynosiłaby wartość $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

Rozkład brzegowy w rozkładzie ciągłym

Niech łączny rozkład zmiennych losowoych X i Y będzie zdefiniowany przez funkcję gęstości $g(x, y)$. Rozkłady brzegowe to odpowiednio:

Rozkład brzegowy w rozkładzie ciągłym

Niech łączny rozkład zmiennych losowoych X i Y będzie zdefiniowany przez funkcję gęstości $g(x, y)$. Rozkłady brzegowe to odpowiednio:

$$g_X(x) = \int_{-\infty}^{\infty} g(x, y) dy,$$

Rozkład brzegowy w rozkładzie ciągłym

Niech łączny rozkład zmiennych losowych X i Y będzie zdefiniowany przez funkcję gęstości $g(x, y)$. Rozkłady brzegowe to odpowiednio:

$$g_X(x) = \int_{-\infty}^{\infty} g(x, y)dy,$$

$$g_Y(y) = \int_{\infty}^{\infty} g(x, y)dx.$$

Rozkład brzegowy w rozkładzie ciągłym

Niech łączny rozkład zmiennych losowoych X i Y będzie zdefiniowany przez funkcję gęstości $g(x, y)$. Rozkłady brzegowe to odpowiednio:

$$g_X(x) = \int_{-\infty}^{\infty} g(x, y)dy,$$

$$g_Y(y) = \int_{\infty}^{\infty} g(x, y)dx.$$

Analogicznie jak w przypadku zmiennych dyskretnych, rozkład ciągłych $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$. (definicję tę wykorzystuje się przy liczeniu prawdopodobieństwa, ale też warunkowej watości oczekiwanej)

Rozkład brzegowy w rozkładzie ciągłym

Niech łączny rozkład zmiennych losowoych X i Y będzie zdefiniowany przez funkcję gęstości $g(x, y)$. Rozkłady brzegowe to odpowiednio:

$$g_X(x) = \int_{-\infty}^{\infty} g(x, y)dy,$$

$$g_Y(y) = \int_{\infty}^{\infty} g(x, y)dx.$$

Analogicznie jak w przypadku zmiennych dyskretnych, rozkład ciągłych $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$. (definicję tę wykorzystuje się przy liczeniu prawdopodobieństwa, ale też warunkowej watości oczekiwanej)
A niezależność: $g(x, y) = g_X(x)g_Y(y)$

Uzasadnij, że $g(x, y) = \frac{3}{2}xy,\ ,\ x + y \leq 2,\ x > 0,\ y > 0$ jest funkcją gęstości. Następnie wyznacz:

1. $g(x|y)$,
2. $\mathbb{P}(X > 1|Y = \frac{1}{2})$,
3. $\mathbb{P}(X \in (\frac{1}{2}, 1)|Y = \frac{1}{2})$,
4. $\mathbb{E}(\frac{X}{Y})$.

Uzasadnij, że $g(x, y) = \frac{3}{2}xy$, , $x + y \leq 2$, $x > 0$, $y > 0$ jest funkcją gęstości. Następnie wyznacz:

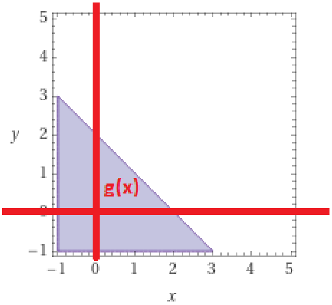
1. $g(x|y)$,
2. $\mathbb{P}(X > 1|Y = \frac{1}{2})$,
3. $\mathbb{P}(X \in (\frac{1}{2}, 1)|Y = \frac{1}{2})$,
4. $\mathbb{E}(\frac{X}{Y})$.

Aby sprawdzić, że $g(x, y)$ jest funkcją gęstości, stwierdzamy, że $x, y > 0$, a ponadto $\int_0^2 \int_0^{-x+2} g(x, y)\, dy\, dx$. Skąd takie granice całkowania?

Uzasadnij, że $g(x,y) = \frac{3}{2}xy,\ ,\ x+y \leq 2,\ x>0,\ y>0$ jest funkcją gęstości. Następnie wyznacz:

1. $g(x|y)$,
2. $\mathbb{P}(X > 1 | Y = \frac{1}{2})$,
3. $\mathbb{P}(X \in (\frac{1}{2}, 1) | Y = \frac{1}{2})$,
4. $\mathbb{E}(\frac{X}{Y})$.

Aby sprawdzić, że $g(x,y)$ jest funkcją gęstości, stwierdzamy, że $x, y > 0$, a ponadto $\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx$. Skąd takie granice całkowania?



Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx = \ldots$$

Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx = \dots$$

I kolejno:

Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx = \dots$$

I kolejno:

1. $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$, gdzie $g_Y(y) = \int_{-\infty}^{\infty} g(x,y)dx$,

Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx = \dots$$

I kolejno:

1. $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$, gdzie $g_Y(y) = \int_{-\infty}^{\infty} g(x,y)dx$, mamy $g_Y(y) = \int_0^{-y+2} \frac{3}{2}xydx$ i ostatecznie

Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx = \dots$$

I kolejno:

1. $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$, gdzie $g_Y(y) = \int_{-\infty}^{\infty} g(x,y)dx$, mamy $g_Y(y) = \int_0^{-y+2} \frac{3}{2}xydx$ i ostatecznie $g_{X|Y}(xy) = \frac{2x}{(2-y)^2}$,

Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx = \dots$$

I kolejno:

1. $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$, gdzie $g_Y(y) = \int_{-\infty}^{\infty} g(x,y)dx$, mamy $g_Y(y) = \int_0^{-y+2} \frac{3}{2}xydx$ i ostatecznie $g_{X|Y}(xy) = \frac{2x}{(2-y)^2}$,
2. $\mathbb{P}(X > 1|Y = \frac{1}{2}) = \int_1^{\frac{3}{2}} g_{x|\frac{1}{2}}(x)\,dx$,

Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\,dy\,dx = \ldots$$

I kolejno:

1. $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$, gdzie $g_Y(y) = \int_{-\infty}^{\infty} g(x,y)dx$, mamy $g_Y(y) = \int_0^{-y+2} \frac{3}{2}xydx$ i ostatecznie $g_{X|Y}(xy) = \frac{2x}{(2-y)^2}$,
2. $\mathbb{P}(X > 1|Y = \frac{1}{2}) = \int_1^{\frac{3}{2}} g_{x|\frac{1}{2}}(x)\,dx$,
3. $\mathbb{P}(X \in (\frac{1}{2},1)|Y = \frac{1}{2}) = \int_{\frac{1}{2}}^{1} g_{x|\frac{1}{2}}(x)\,dx$,

Zatem przystępujemy do wyliczeń:

$$\int_0^2 \int_0^{-x+2} g(x,y)\, dy\, dx = \dots$$

I kolejno:

1. $g_{X|Y}(x|y) = \frac{g(x,y)}{g_Y(y)}$, gdzie $g_Y(y) = \int_{-\infty}^{\infty} g(x,y)dx$, mamy $g_Y(y) = \int_0^{-y+2} \frac{3}{2}xydx$ i ostatecznie $g_{X|Y}(xy) = \frac{2x}{(2-y)^2}$,
2. $\mathbb{P}(X > 1 | Y = \frac{1}{2}) = \int_1^{\frac{3}{2}} g_{x|\frac{1}{2}}(x)\, dx$,
3. $\mathbb{P}(X \in (\frac{1}{2}, 1) | Y = \frac{1}{2}) = \int_{\frac{1}{2}}^{1} g_{x|\frac{1}{2}}(x)\, dx$,
4. $\mathbb{E}(\frac{X}{Y}) = \int_0^2 \int_0^{-x+2} \frac{x}{y} g(x,y)\, dy\, dx$.