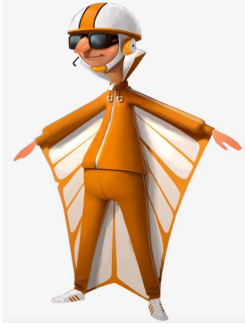


Wektor



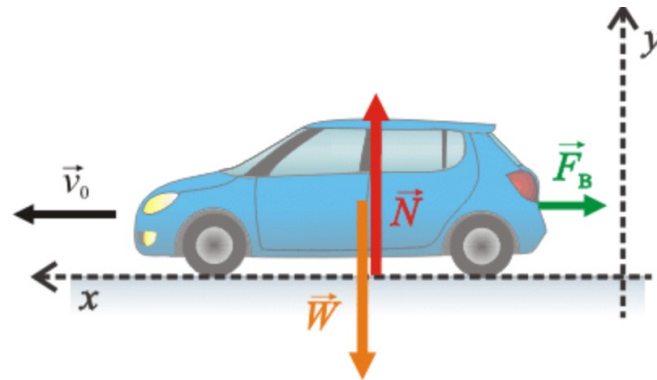
Programista

Tablica wypełniona liczbami

```
array([1, 2, 3])
```

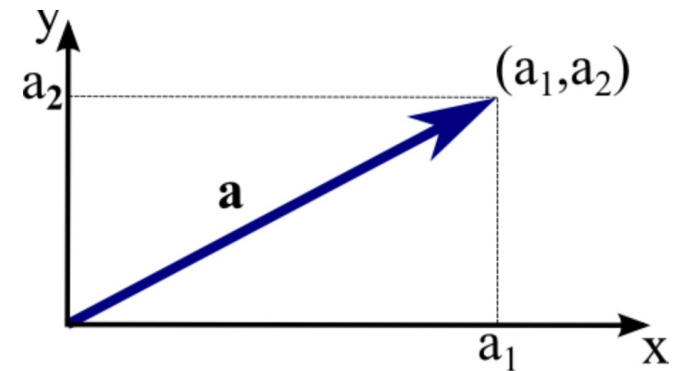
Fizyk

Obiekt opisywany za pomocą długości, kierunku i zwrotu.



Matematyk

Współrzędne punktu w przestrzeni.



Macierze

TROCHĘ TEORII

Macierz to, podobnie jak wektor, tablica liczb, tyle że dwuwymiarowa. Identyfikacja konkretnej liczby jest możliwa za pomocą dwóch indeksów określających położenie.

Macierze są zapisywane w postaci prostokątnej tablicy i są oznaczane zazwyczaj dużą literą alfabetu, co pokazano na przykładzie poniższej macierzy

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ a_{31} & a_{32} & \dots & a_{3m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}.$$

Elementy aji nazywamy elementami macierzy. Przykładowo element a_{12} znajduje się w pierwszym wierszu i drugiej kolumnie macierzy, bowiem wiersze numerujemy "od góry", a kolumny – "od lewej strony". Prezentowana macierz ma wymiary $m \times n$.

Macierz 2x3 (2 wiersze 3 kolumny)

```
B = np.array([[3,7,4], [9,3,5]])  
B  
  
array([[3, 7, 4],  
       [9, 3, 5]])
```

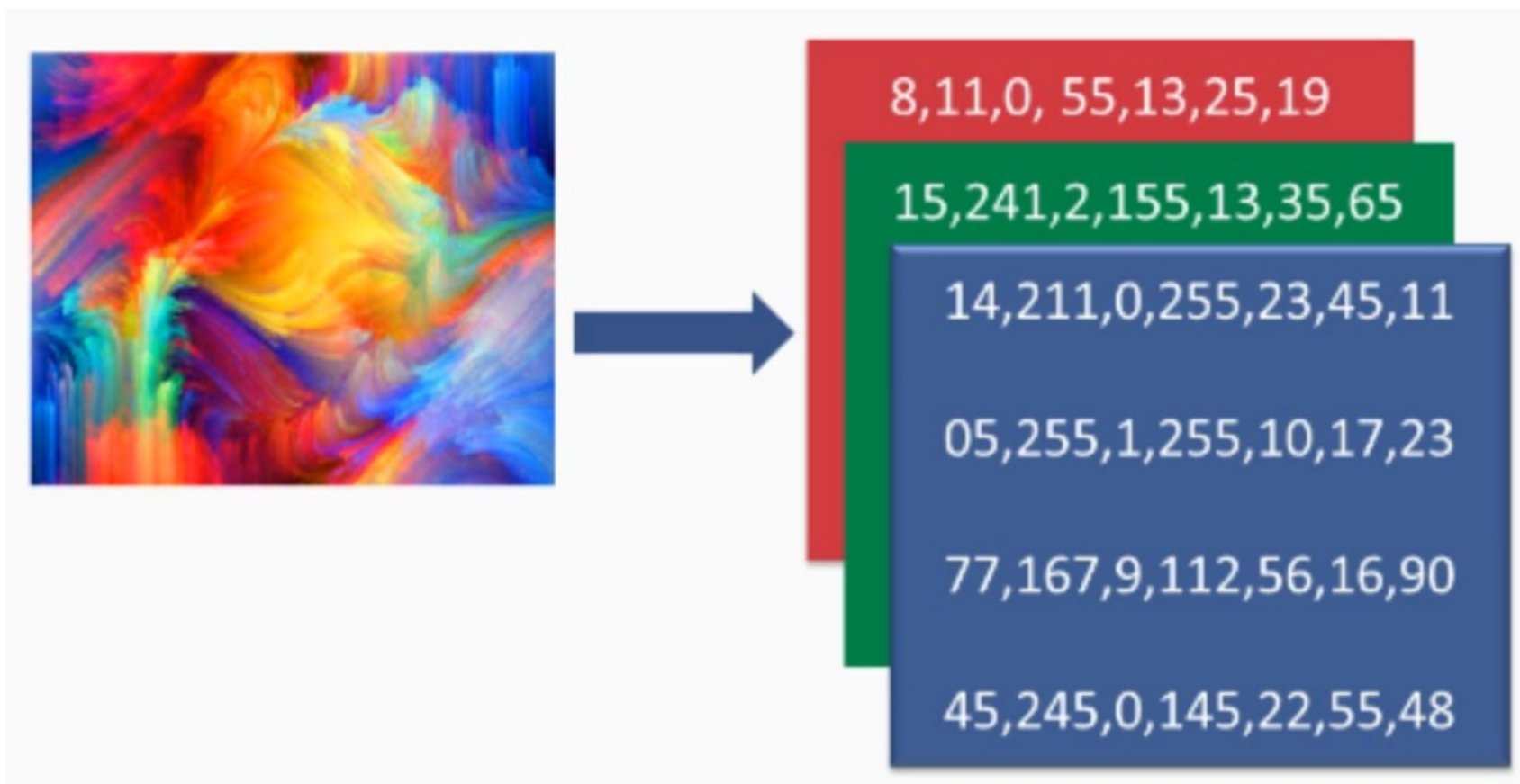
Tworzenie macierzy za pomocą np.arange()

```
C = np.arange(2,8)  
C  
  
array([2, 3, 4, 5, 6, 7])
```





Przy użyciu np.linspace(start, end, count) – (0,1,5) – 5 liczb od 0 do 1 w równych odległościach

```
D = np.linspace(0,1,5)  
D  
  
array([ 0. ,  0.25,  0.5 ,  0.75,  1.  ])
```

Tensor - przykład



Tensor - reprezentacja

Scalar	Vector	Matrix	Tensor
			
9	<code>array([9, -8, 5])</code>	<code>array([[4, 7, -4], [-8, 4, 9], [7, 6, 6]])</code>	<code>array([[[7, 0, -10], [7, -2, 7], [-3, -1, -8]], [[-2, 8, -9], [6, 6, -8], [-9, -6, -6]], [[-6, -7, -10], [-2, -6, 6], [0, -4, 4]])]</code>

Mnożenie macierzy

$$\mathbf{X} = \begin{bmatrix} [5 & 6 & 1 & 2] \\ [8 & 7 & 6 & 3] \\ [5 & 0 & 6 & 4] \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} [3 & 0 & 4 & 9] \\ [4 & 6 & 5 & 8] \\ [7 & 0 & 1 & 5] \end{bmatrix}$$

$$\mathbf{Z} = \mathbf{X} * \mathbf{Y} = \begin{bmatrix} [15 & 0 & 4 & 18] \\ [32 & 42 & 30 & 24] \\ [35 & 0 & 6 & 20] \end{bmatrix}$$

Każdy element macierzy X jest pomnożony przez odpowiadający mu element macierzy Y

$$x_{11} = 5 \quad y_{11} = 3 \quad z_{11} = x_{11} \times y_{11} = 15$$

$$x_{12} = 6 \quad y_{12} = 0 \quad z_{12} = x_{12} \times y_{12} = 0$$

$$X_{3 \times 4}$$

$$Y_{3 \times 4}$$

↔ Wymiar obu macierzy musi być taki sam

Mnożenie macierzy (algebraiczne)

$$\mathbf{X} = \begin{bmatrix} [9 & 2 & 2] \\ [4 & 0 & 0] \\ [9 & 3 & 9] \end{bmatrix}$$

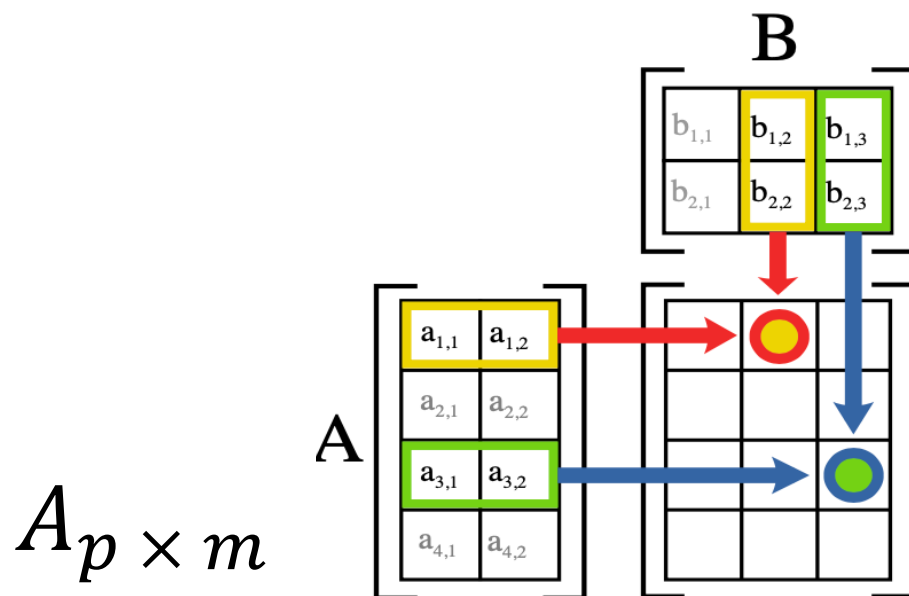
$$\mathbf{Y} = \begin{bmatrix} [8 & 1 & 1] \\ [9 & 6 & 8] \\ [7 & 4 & 8] \end{bmatrix}$$

$$\mathbf{Z} = \mathbf{X} \cdot \mathbf{Y} = \begin{bmatrix} [104 & 29 & 41] \\ [32 & 4 & 4] \\ [162 & 63 & 105] \end{bmatrix}$$

$$z_{11} = x_{11} \times y_{11} + x_{12} \times y_{21} + x_{13} \times y_{31}$$

$$z_{12} = x_{11} \times y_{12} + x_{12} \times y_{22} + x_{13} \times y_{32}$$

$$z_{21} = x_{21} \times y_{11} + x_{22} \times y_{21} + x_{23} \times y_{31}$$



$$\mathbf{B}_{m \times q}$$

← pierwszy wymiar macierzy \mathbf{A} musi być taki sam jak drugi wymiar macierzy \mathbf{B}



The Matrix is everywhere. It is all around us. Even now, in this very room.

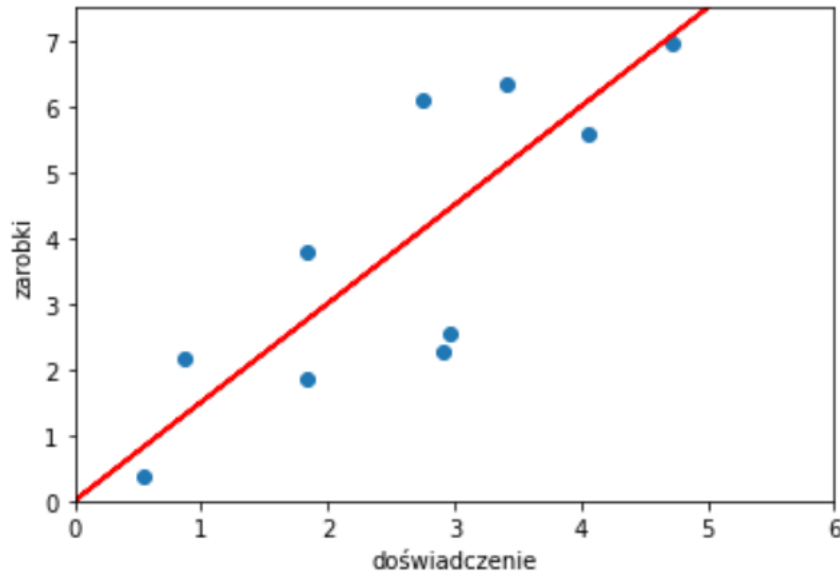
-Morpheus

Mnożenie macierzy (*geometria* regresji)

$$y = AX + b + e$$

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

$$\text{zarobki} = a \times \text{doświadczenie} + b$$



b	doswaidczenie	zarobki
1	4.057534	5.586908
1	2.750049	6.105262
1	2.964347	2.533063
1	3.410334	6.332447
1	1.828527	1.840721
1	0.548475	0.355877
1	0.876145	2.176086
1	4.729183	6.966537
1	1.834105	3.784180
1	2.904804	2.281207

$$\text{zarobki}_1 = a \times \text{doświadczenie}_1 + b$$

$$\text{zarobki}_2 = a \times \text{doświadczenie}_2 + b$$

$$zarobki_3 = a \times do\acute{s}wiadczenie_3 + b$$

$$(\dots)$$

b doświadczenie						zarobki
1	4.057534	* współczynniki b a	=	a_1 * 4.06 + b	=	5.586908
1	2.750049			a_2 * 2.75 + b		6.105262
1	2.964347			a_3 * 2.96 + b		2.533063
1	3.410334			a_4 * 3.41 + b		6.332447
1	1.828527			a_5 * 1.83 + b		1.840721
1	0.548475			a_6 * 0.55 + b		0.355877
1	0.876145			a_7 * 0.88 + b		2.176086
1	4.729183			a_8 * 4.73 + b		6.966537
1	1.834105			a_9 * 1.83 + b		3.784180
1	2.904804			a_10 * 2.9 + b		2.281207

Mnożenie macierzy (reprezentacja regresji)

$y = AX + b, \quad b = 0 \qquad y = AX$

$X =$

długość	szerokość	wysokość
79.564798	139.113256	165.247195
67.174099	47.367705	286.349772
118.450682	10.869266	83.118642
14.156648	25.846663	189.236549
200.551913	95.144694	37.003291
63.339889	99.574222	137.854002
156.675029	204.190017	16.619945

$y =$

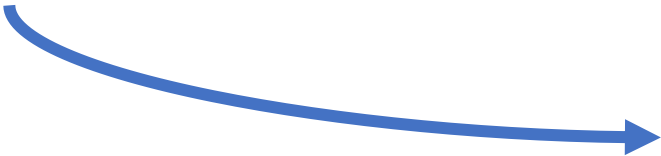
opóźnienie
0.601009
3.031186
4.160437
1.326778
3.453975
1.106210
0.091691

$A =$

współczynniki
a_1
a_2
a_3

współczynniki
a_1
a_2
a_3

$op\acute{o}znienie = a_1 \times dl\acute{u}gość + a_2 \times szerokość + a_3 \times wysokość$



$y = AX =$

długość	szerokość	wysokość	opóźnienie
79.564798	139.113256	165.247195	0.601009
67.174099	47.367705	286.349772	3.031186
118.450682	10.869266	83.118642	4.160437
14.156648	25.846663	189.236549	1.326778
200.551913	95.144694	37.003291	3.453975
63.339889	99.574222	137.854002	1.106210
156.675029	204.190017	16.619945	0.091691

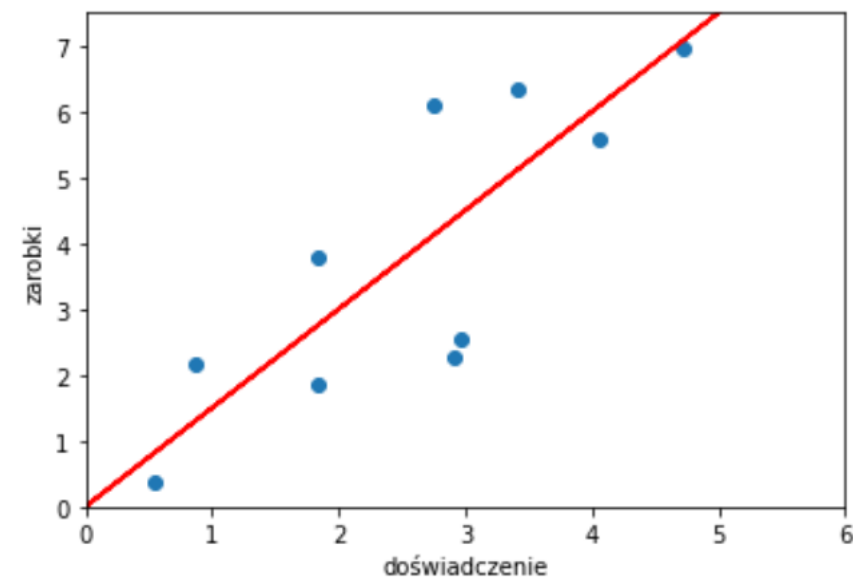
Metoda najmniejszych kwadratów

$$Y = \beta X + e$$

$$\hat{Y} = \beta X$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

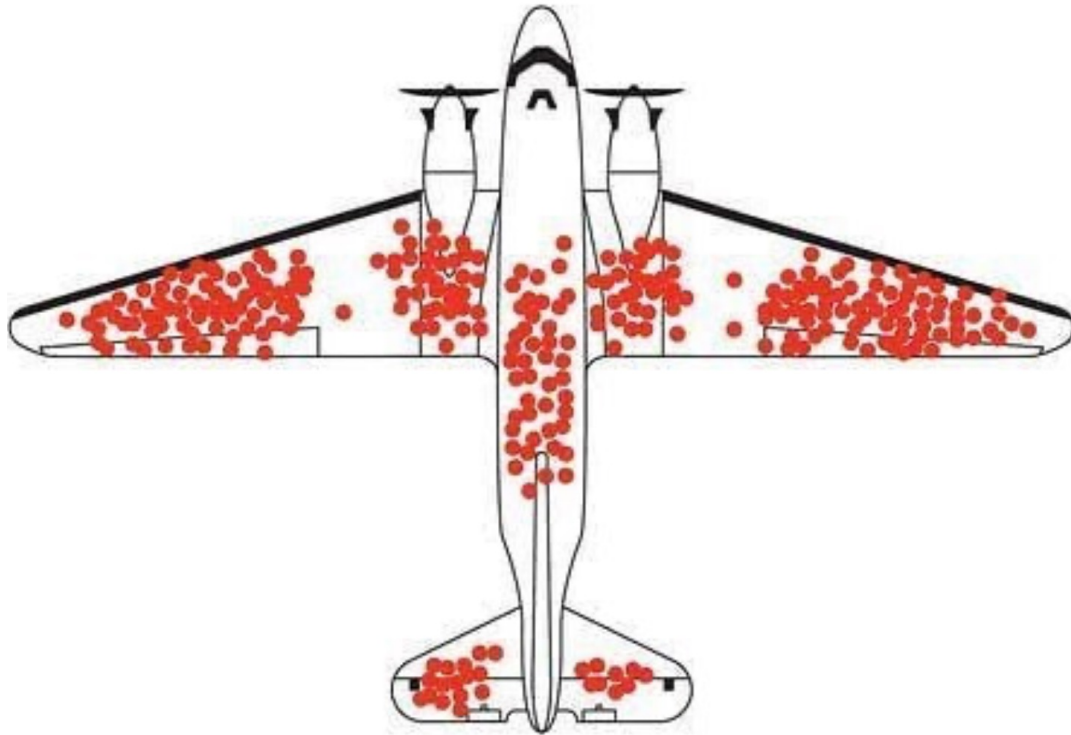
$$\mathbf{X} = \begin{pmatrix} x_{1,0} & x_{1,1} & \dots & x_{1,p-1} \\ x_{2,0} & x_{2,1} & \dots & x_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,0} & x_{n,1} & \dots & x_{n,p-1} \end{pmatrix}.$$



$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Case study: *Dopancerznie* samolotów

https://en.wikipedia.org/wiki/Survivorship_bias#In_the_military



The red dots indicate areas of combat damage received by surviving WWII bombers. Where would you add armor to increase survivability? The statistician Abraham Wald recommended reinforcing the areas *without* damage. Since these data came from surviving aircraft only, bombers hit in undotted areas were the ones that did not make it back.

Case study: Produkcja czołgów

An empirical Approach to Economic Intelligence in World War II,
Journal of the American Statistical Association, Vol. 42, No. 237
(Mar., 1947), pp. 72–91.

https://en.wikipedia.org/wiki/German_tank_problem



Produkcja czołgów (dane historyczne):

Miesiąc	est. statystyków	est. agentów	dane prod.
Czerwiec 1940	169	1000	122
Czerwiec 1941	244	1550	271
Sierpień 1942	327	1550	342

Paradoks Simpsona

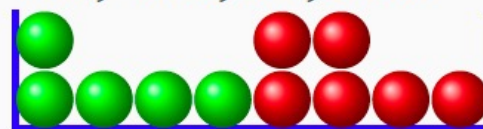
Przyjęci na studia (Univ. of California, Berkeley)

	Zgłoszeń	Przyjętych
Mężczyźni	2691	(1198) 45%
Kobiety	1835	(614) 33%

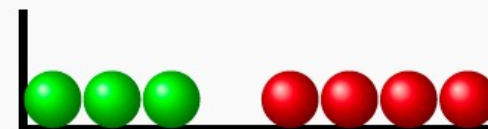
Przyjęci z podziałem na kierunki

K.	Mężczyźni		Kobiety	
	Zgłoszeń	Przyjętych	Zgłoszeń	Przyjętych
A	825	(512) 62%	108	(89) 82%
B	560	(353) 63%	25	(17) 68%
C	325	(120) 37%	593	(219) 37%
D	417	(138) 33%	375	(131) 35%
E	191	(53) 28%	393	(134) 34%
F	373	(22) 6%	341	(24) 7%

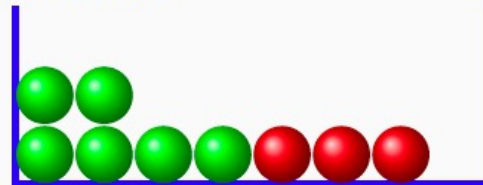
Wybieramy koszyk, z koszyka losujemy kulę, wygrywa zielona.
Który koszyk wybrać?



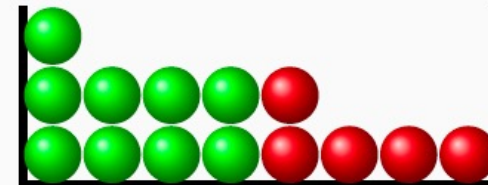
$$\frac{5}{11} > \frac{3}{7}$$



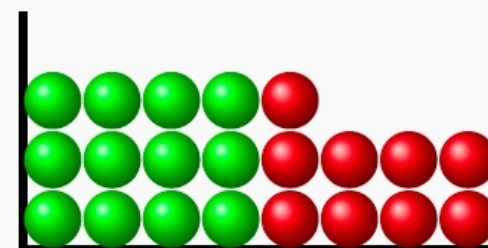
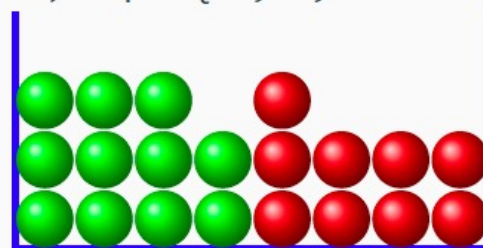
A teraz?



$$\frac{6}{9} > \frac{9}{14}$$



A jak połączymy zawartość koszyków?

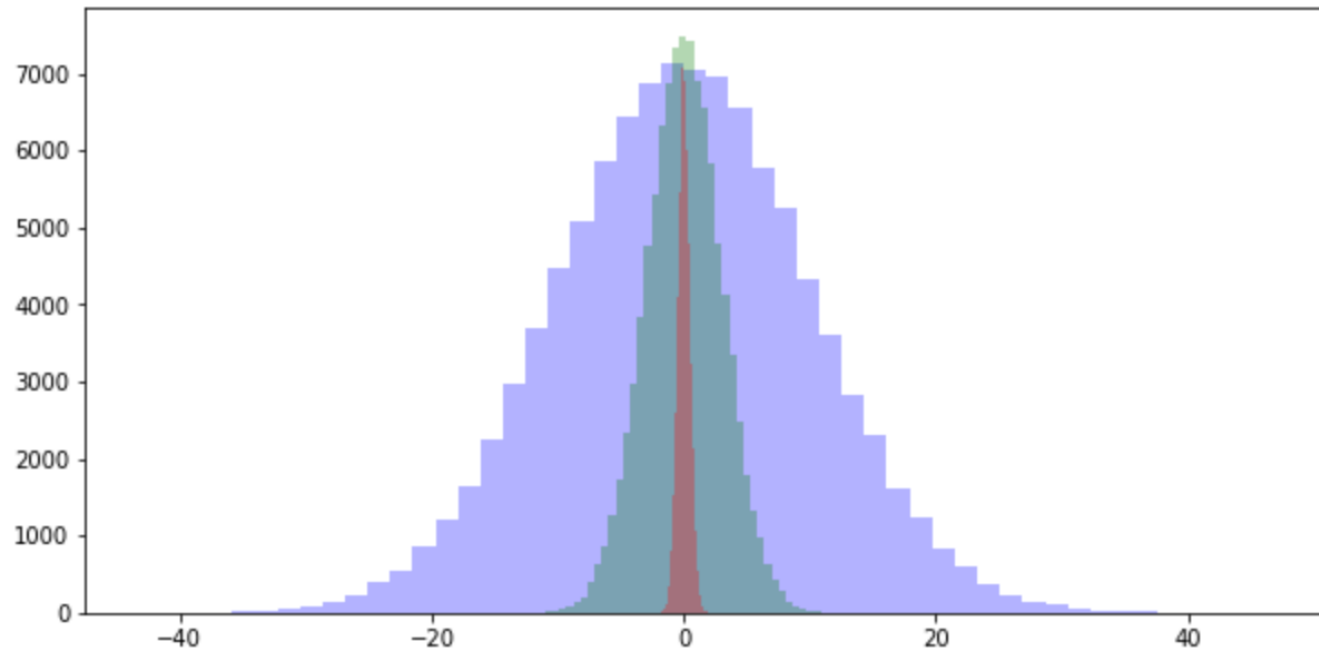


Wariancja

✓
0s



```
N1 = np.random.normal(loc=0.0, scale= 0.5, size=100_000)
N2 = np.random.normal(loc=0.0, scale= 3.0, size=100_000)
N3 = np.random.normal(loc=0.0, scale=10.0, size=100_000)
plt.figure(figsize=(10, 5))
plt.hist(N3, bins=50, alpha=0.3, color='blue')
plt.hist(N2, bins=50, alpha=0.3, color='green')
plt.hist(N1, bins=50, alpha=0.3, color='red')
plt.show()
```



Podstawowe statystyki – wzory!

wartość
oczekiwana:

$$\mathbb{E}X = \sum_{i=1}^n x_i p_i$$

wariancja:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

odchylenie
standardowe:

$$\begin{aligned}\sigma(X) &= \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]} \\ &= \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}\end{aligned}$$

kowariancja: $\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)]$.

Współczynnik
korelacji Pearsona :

$$r_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

statystyka:

Wariancja populacji:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Wariancja próbki:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Wariancja-znana średnia:

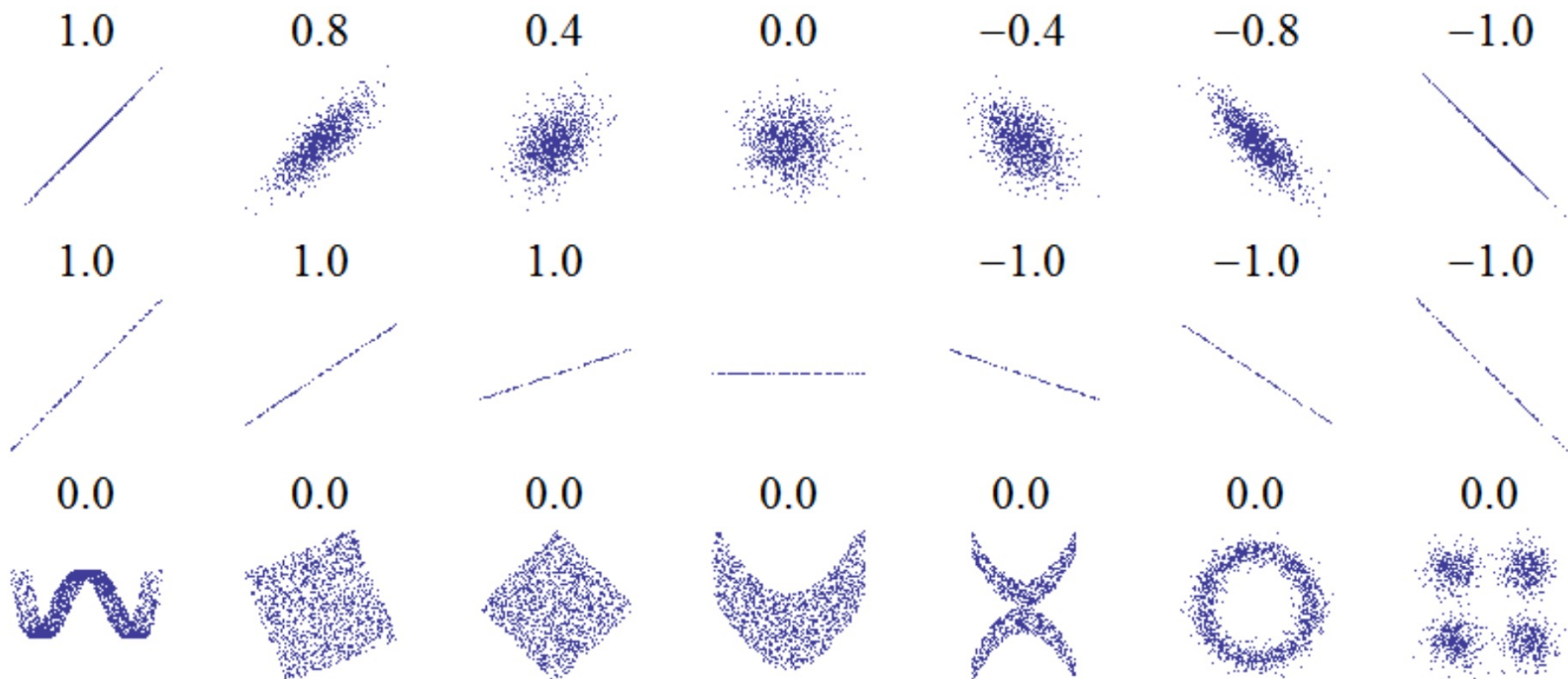
$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

\bar{x} – średnia

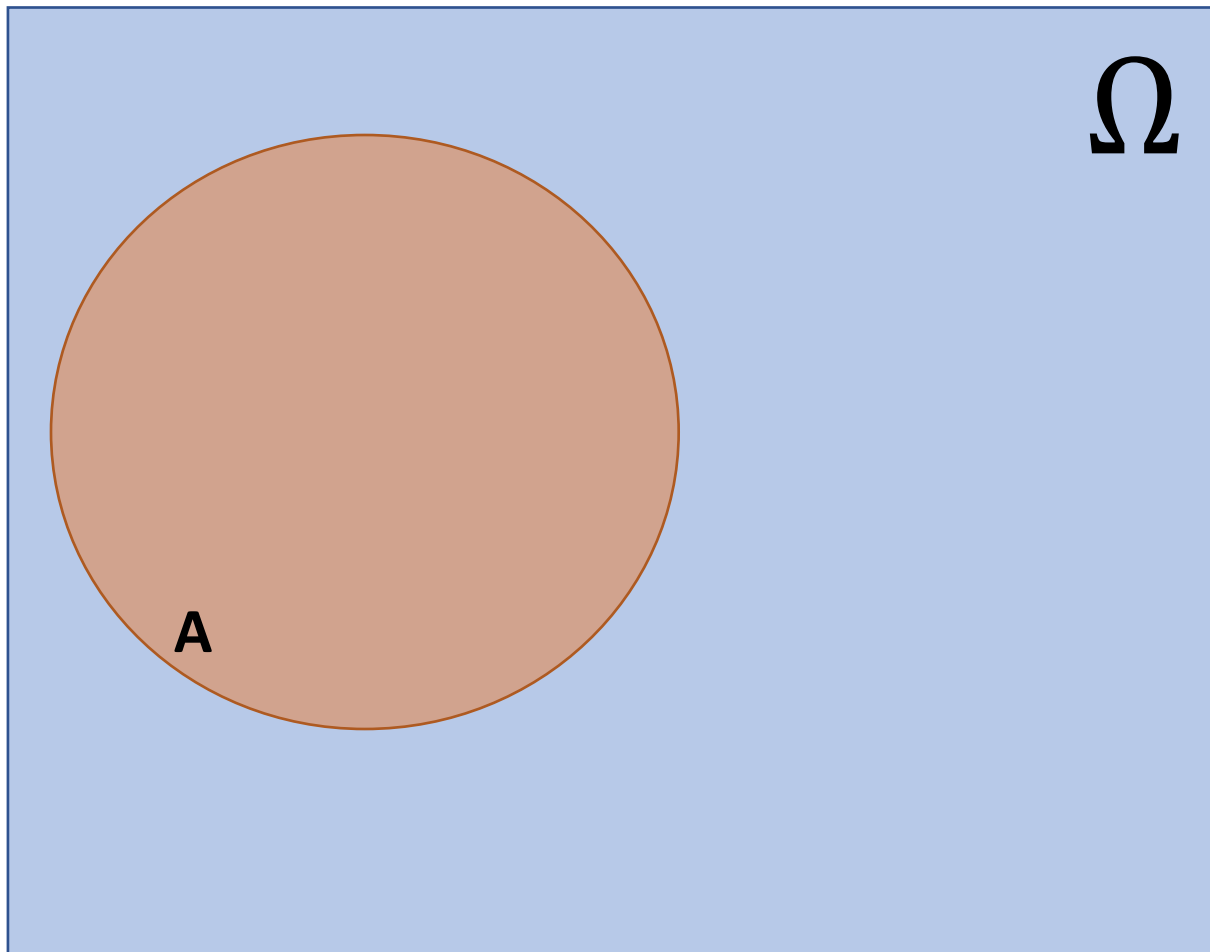
μ – wartość oczekiwana

Współczynnik korelacji Pearsona

współczynnik określający poziom zależności liniowej między zmiennymi losowymi

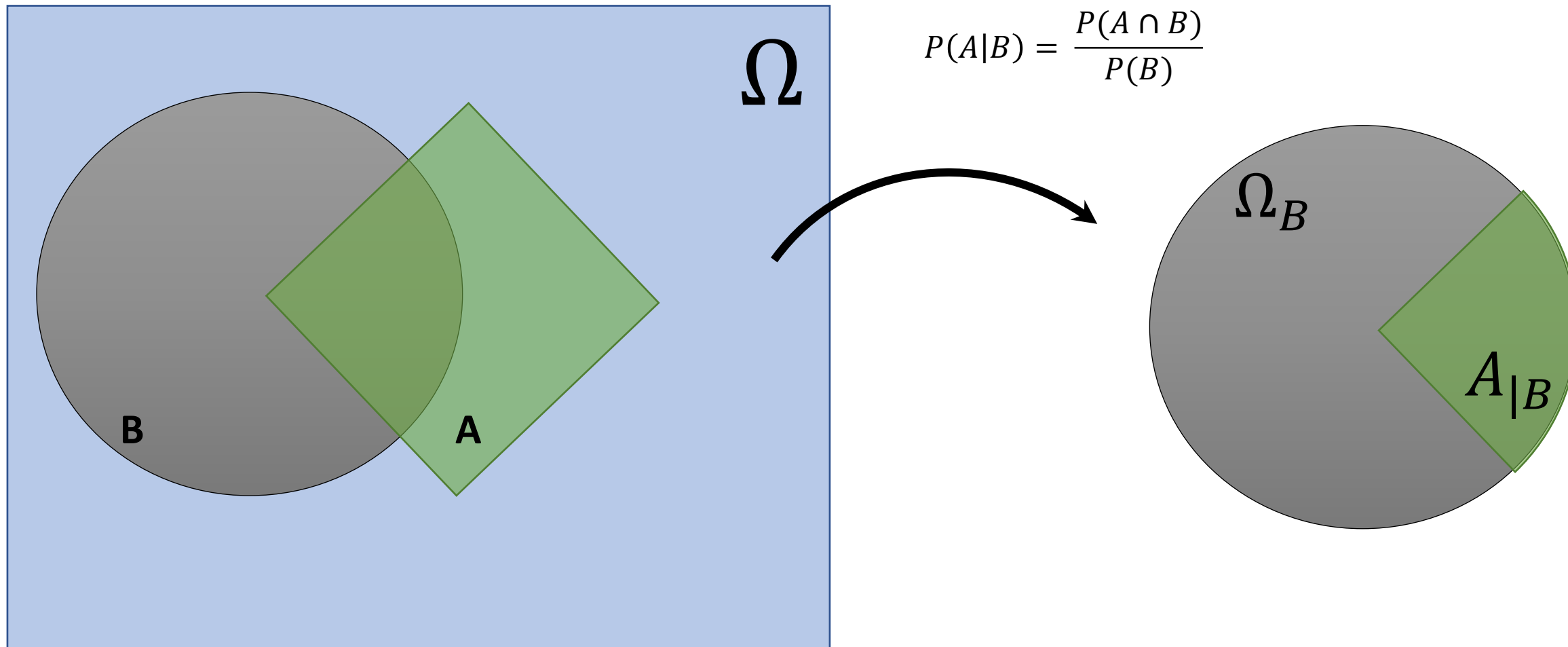


Prawdopodobieństwo

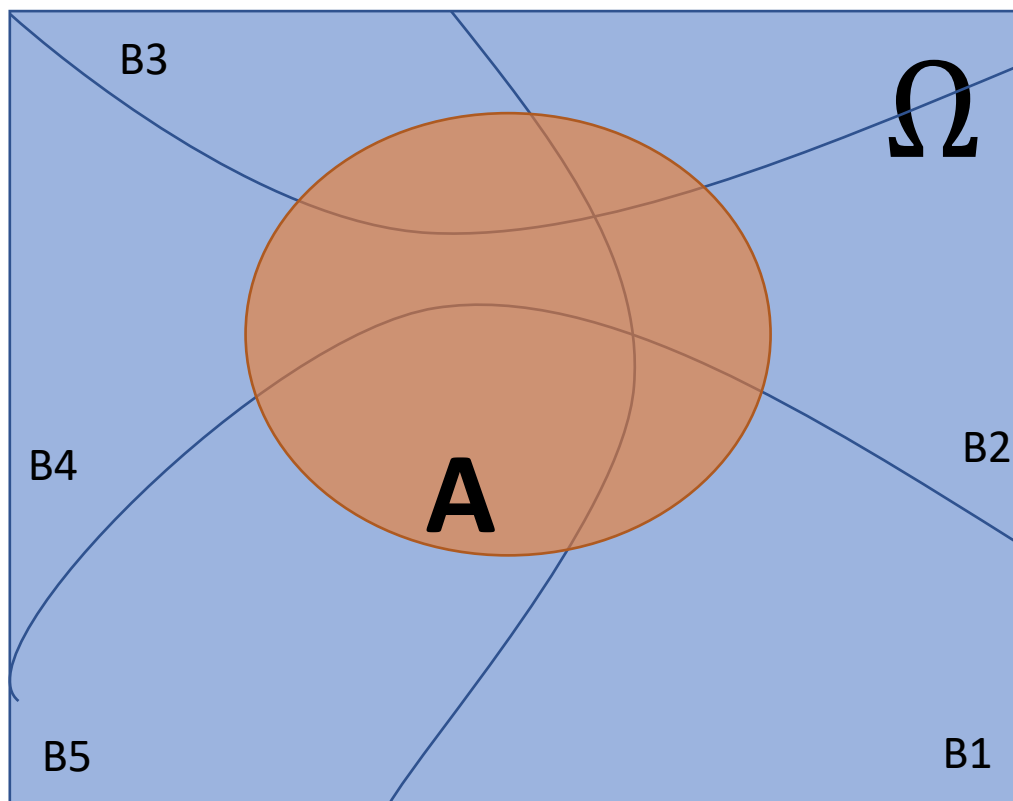


$$P(A) = \frac{|A|}{|\Omega|}$$

Prawdopodobieństwo warunkowe



Prawdopodobieństwo całkowite



$$\begin{aligned} P(A) = & P(A|B1)P(B1) \\ & + P(A|B2)P(B2) \\ & + P(A|B3)P(B3) \\ & + P(A|B4)P(B4) \\ & + P(A|B5)P(B5) \end{aligned}$$

$$P(\Omega) = P(B1) + P(B2) + P(B3) + P(B4) + P(B5)$$

Twierdzenie Bayesa

przykład praktyczny (prawdopodobieństwo choroby)

Dany jest test o wrażliwości 80%, co oznacza, że jeśli choroba występuje to test będzie pozytywny z prawdopodobieństwem 0.8.

Prawdopodobieństwo wystąpienia choroby to 0.004.

Test zwraca 10% fałszywie pozytywnych wyników.

$$P(test_{pozytywny} | pacjent_{chory}) = 0.8$$

$$P(pacjent_{chory} | test_{pozytywny}) = ?$$

$$P(pacjent_{chory}) = 0.004$$

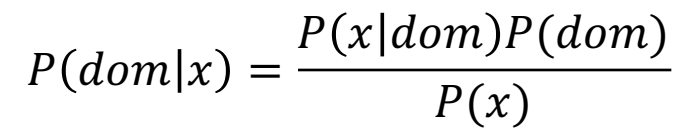
$$P(test_{pozytywny} | pacjent_{zdrowy}) = 0.1$$

$$P(pacjent_{chory} | test_{pozytywny}) = \frac{P(test_{pozytywny} | pacjent_{chory})P(pacjent_{chory})}{P(test_{pozytywny})}$$

$$= \frac{P(test_{pozytywny} | pacjent_{chory})P(pacjent_{chory})}{P(test_{pozytywny} | pacjent_{chory})P(pacjent_{chory}) + P(test_{pozytywny} | pacjent_{zdrowy})P(pacjent_{zdrowy})}$$

$$= \frac{0.8 \times 0.004}{0.8 \times 0.004 + 0.1 \times 0.996} = 0.031$$

przykład praktyczny (Naïve Bayes)



$$P(dom) = \frac{|dom|}{|obserwacje|} = \frac{20}{30}$$

$$P(x|dom) = \frac{|dom \text{ w otoczeniu}|}{|dom|} = \frac{1}{20}$$

$$\begin{aligned} P(x) &= P(x|dom)P(dom) \\ &\quad + P(x|mieszkanie)P(mieszkanie) \\ &= \frac{1}{20} \frac{20}{30} + \frac{4}{10} \frac{10}{30} = \frac{4}{30} \end{aligned}$$

$$P(dom|x) = 0.25$$

Zmienna Losowa

Zmienna losowa to funkcja:

$$X: \Omega \longrightarrow E \subset \mathbb{R}$$

$$\Omega = \left\{ \begin{array}{c} \text{2 złote} \\ \text{Rzeczpospolita Polska} \\ \text{2018} \end{array} ; \begin{array}{c} \text{Rzeczpospolita Polska} \\ \text{2018} \end{array} \right\}$$

