infoShare ACADEMY

# Statystyka

infoShareAcademy.com

# Statystyka

## Agenda

1. **Eksperyment**
2. Operacje na zbiorach
3. Zdarzenie
4. Prawdopodobieństwo
5. Twierdzenie Bayesa
6. Zmienna losowa
7. Statystyczna analiza danych
8. Rozkłady

# Statystyka

## Agenda

1. Eksperyment
2. **Operacje na zbiorach**
3. Zdarzenie
4. Prawdopodobieństwo
5. Twierdzenie Bayesa
6. Zmienna losowa
7. Statystyczna analiza danych
8. Rozkłady

# Statystyka

## Agenda

1. Eksperyment
2. Operacje na zbiorach
3. **Zdarzenie**
4. Prawdopodobieństwo
5. Twierdzenie Bayesa
6. Zmienna losowa
7. Statystyczna analiza danych
8. Rozkłady



infoShare ACADEMY

# Statystyka

## Agenda

1. Eksperyment
2. Operacje na zbiorach
3. Zdarzenie
4. **Prawdopodobieństwo**
5. **Twierdzenie Bayesa**
6. Zmienna losowa
7. Statystyczna analiza danych
8. Rozkłady



infoShare ACADEMY

# Statystyka

## Agenda

1. Eksperyment
2. Operacje na zbiorach
3. Zdarzenie
4. Prawdopodobieństwo
5. Twierdzenie Bayesa
6. **Zmienna losowa**
7. **Statystyczna analiza danych**
8. **Rozkłady**

# Statystyka

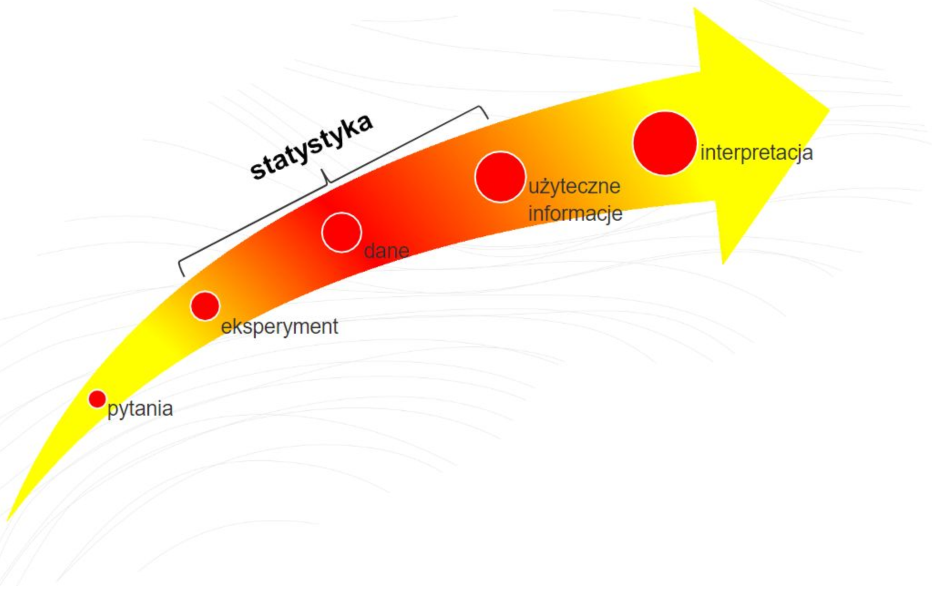
## Czym jest?

**Statystyka** – nauka zajmująca się analizowaniem i liczbowym opisem zjawisk masowych.

**Statystyka matematyczna** – dział matematyki zajmujący się metodami wnioskowania o populacji na podstawie losowych prób z tej populacji.

Słownik języka polskiego PWN.

# Statystyka

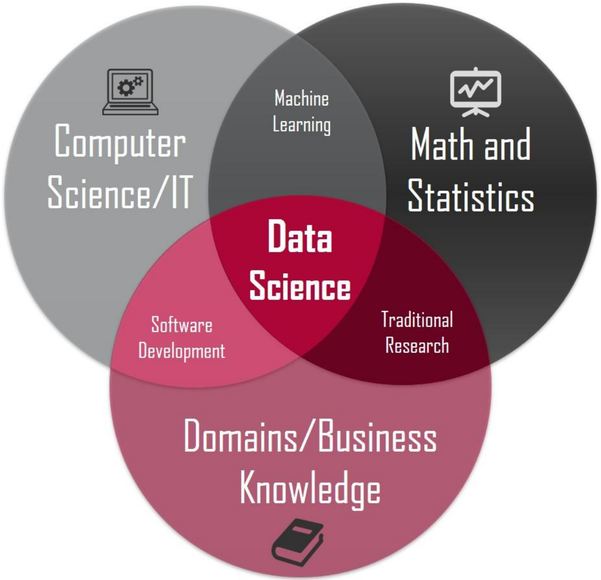





info Share ACADEMY

# Statystyka

## Statystyka i Data Science




Źródło grafiki: towardsdatascience.com

# Statystyka

## Podstawowa klasyfikacja

# Statystyka

## Eksperyment

Uzyskanie wyniku lub wyników doświadczenia losowego. Np.:

- wynik rzutu kostką do gry,
- odpowiedź respondenta na pytanie,
- wynik badania medycznego (np. pomiar stężenia glukozy we krwi),
- wyniki sprzedaży produktu,
- ...

# Statystyka

## Zbiór

Skończona lub nieskończona kolekcja obiektów, w której kolejność nie ma znaczenia, oraz ignorowany jest fakt, że element występuje więcej niż raz w zbiorze.

Fakt przynależności danego elementu do zbioru zapisuje się jako $a \in A$.

infoShare ACADEMY

# Statystyka

## Zbiór (przykład)

Te zbiory są identyczne:

$$A=\{a,1,2,3\}$$

$$B=\{2,a,1,3\}$$

$$C=\{3,2,1,a\}$$

# Statystyka

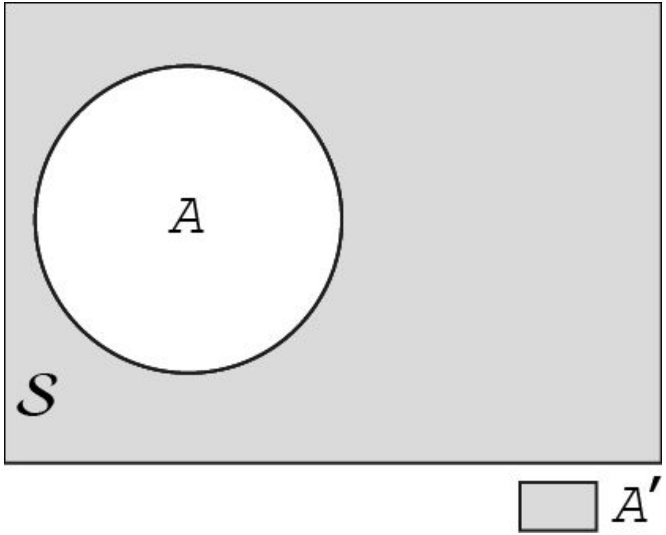
## Zbiór (przykład)

Przykłady zbiorów:

- A = {⚀, ⚁, ⚂, ⚃, ⚄, ⚅},
- pracownicy przedsiębiorstwa,
- stany kontrolki na stronie internetowej,
- klienci firmy,
- produkty w sklepie internetowym,
- stan produktu w sklepie internetowym.

# Statystyka

## Zbiory: podstawowe oznaczenia



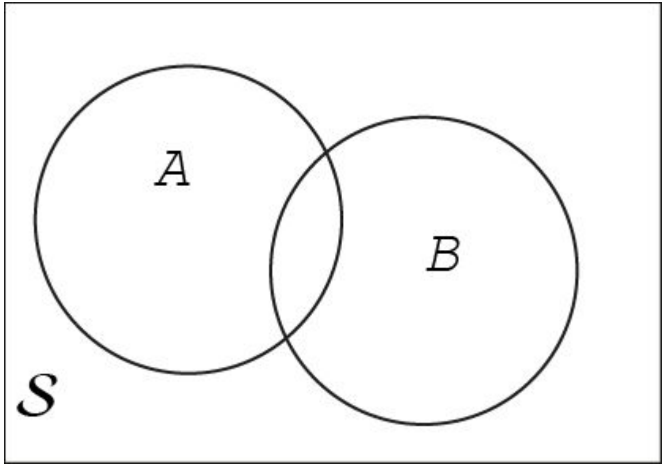

∅ – zbiór pusty

S – zbiór uniwersum

A′ – dopełnienie zbioru A (wszystko poza zbiorem A)

# Statystyka

## Zbiory: podstawowe operacje



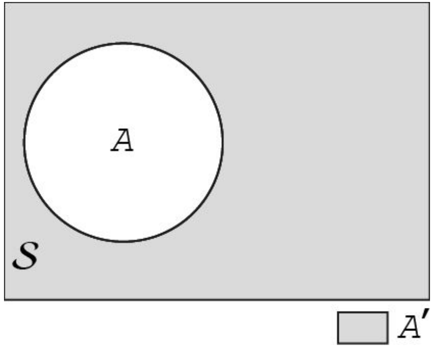

U – suma zbiorów

∩ – iloczyn zbiorów

- – różnica zbiorów

# Statystyka

## Dopełnienie zbioru




A′ jest zbiorem zawierającym wszystkie elementy S, poza tymi, które znajdują się w zbiorze A.

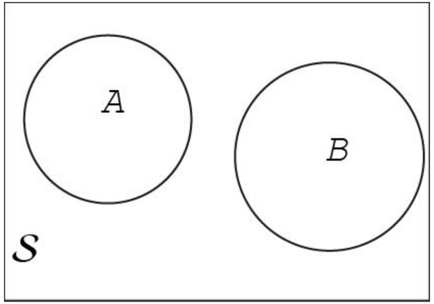
Zbiory A i A′ są parami rozłączne: A∩A′=∅e

Sumą zbiorów A oraz A′ jest S: A∪A′=Se

Inne oznaczenia spotykane w literaturze: AC and Aҕ.

# Statystyka

## Zbiory parami rozłączne

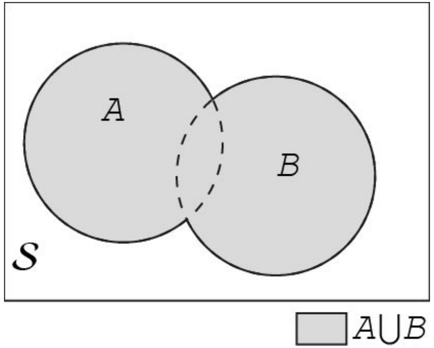



Zbiory A i B są parami rozłączne jeśli ich iloczynem jest zbiór pusty.

$$A \cap B = \varnothing$$

# Statystyka

## Suma (alternatywa) zbiorów

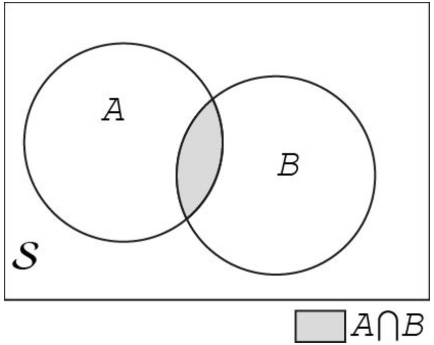



Alternatywą zbiorów A i B jest zbiór zawierający elementy ze zbioru A lub B.

$$A \cup B$$

# Statystyka

## Iloczyn (przecięcie) zbiorów

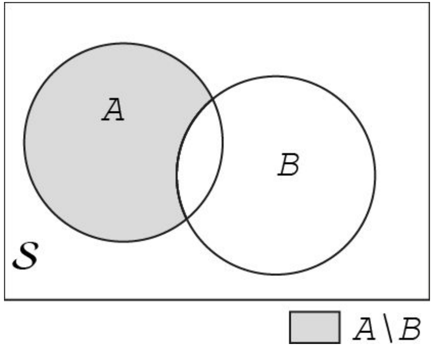



Iloczynem zbiorów A i B jest zbiór zawierający elementy ze zbioru A i B.

$$A \cap B$$

# Statystyka

## Różnica zbiorów

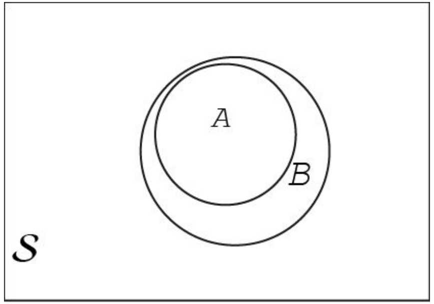



Różnicą zbiorów A i B jest zbiór zawierający elementy ze zbioru A, które nie znajdują się w zbiorze B.

$$A - B$$

# Statystyka

## Zawieranie zbiorów

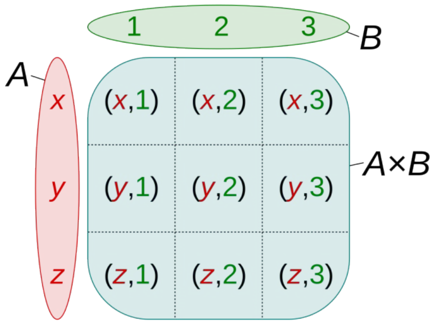



$A \subseteq B$

Zbiorów A jest podzbiorem B jeśli wszystkie jego elementy znajdują się w zbiorze B.

# Statystyka

## Iloczyn kartezjański zbiorów




Iloczynem kartezjańskim zbiorów nazywamy zbiór uporządkowany punktów (krotek) (a,b), gdzie a∈A i b∈B.

$$A \times B$$

# Statystyka

## Iloczyn kartezjański zbiorów (przykład)

$A=\{1,2,3\}$; $B=\{a,b\}$

$$A \times B = \{ (1,a),(1,b),\\ (2,a),(2,b),\\ (3,a),(3,b) \}$$

infoShare ACADEMY

# Statystyka

## Zdarzenie

Zdarzenia są to zbiory i elementy zbiorów w rachunku prawdopodobieństwa.

# Statystyka

## Zdarzenie elementarne

Pojęcie pierwotne rachunku prawdopodobieństwa (tzn. nie posiada formalnej definicji).

Potocznie: najmniejszy niepodzielny wynik doświadczenia losowego.

# Statystyka

## Przestrzeń zdarzeń elementarnych Ω

Zbiór wszystkich możliwych wyników doświadczenia losowego.

Jest to odpowiednik zbioru uniwersum.

# Statystyka

## Przestrzeń zdarzeń elementarnych Ω

infoShare ACADEMY

Przestrzeń zdarzeń elementarnych pojedynczego rzutu kostką (czyli wszystkie możliwe wyniki takiego eksperymentu).

Ω={ ⚀ ⚁ ⚂ ⚃ ⚄ ⚅ }

# Statystyka

## Przestrzeń zdarzeń elementarnych Ω (przykład)

Przestrzeń zdarzeń elementarnych dwukrotnego rzutu monetą (czyli wszystkie możliwe wyniki takiego eksperymentu).

# Statystyka

## Przestrzeń zdarzeń elementarnych Ω (przykład)

Przestrzeń zdarzeń elementarnych eksperymentu polegającego na wypełnieniu ankiety: układ wszystkich możliwych odpowiedzi na każde pytanie.

$$
\begin{aligned}
\Omega = \{&(a,a,a,\ldots,a),\\
&(b,a,a,\ldots,a),\\
&(c,a,a,\ldots,a),\\
&(d,a,a,\ldots,a),\\
&(a,b,a,\ldots,a),\\
&\ldots,\\
&(d,d,d,\ldots,c),\\
&(d,d,d,\ldots,d)\}
\end{aligned}
$$



infoShare ACADEMY

# Statystyka

## Zdarzenie losowe

infoShare ACADEMY

Eksperyment nazywa się losowym jeśli przy powtarzaniu go w tych samych warunkach jego wynik nie może być przewidziany, a wszystkie możliwe wyniki mogą zostać określone przed jego przeprowadzeniem.

# Statystyka

## Zdarzenie losowe (przykład)

infoShare ACADEMY

- Rzut monetą.
- Liczba wypadków samochodowych w danym dniu.
- Partia, na którą zagłosuje pytana osoba.
- Wystąpienie katastrof naturalnych.
- Cena akcji na rynku.
- Liczba klientów, która zakupi produkt.
- …

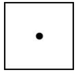
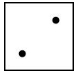
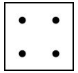
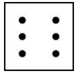
# Statystyka

## Zdarzenie losowe: typy

Wyróżnia się dwa typy zdarzeń losowych:

- proste,
- złożone.

Np. w eksperymencie polegającym na rzucie kostką:

- zdarzenie proste: otrzymano 
- zdarzenie złożone: otrzymano parzystą liczbę oczek, czyli: ,  lub 

# Statystyka

## Zdarzenie pewne

Zdarzenie, w którego skład wchodzą wszystkie elementy przestrzeni zdarzeń elementarnych (czyli zdarzy się na pewno).

Zdarzenie pewne oznacza się symbolem Ω.

infoShare ACADEMY

# Statystyka

## Zdarzenie niemożliwe

Zdarzenie, które nie zawiera żadnych elementów.

Zdarzenie niemożliwe oznacza się symbolem zbioru pustego, czyli $\emptyset$.

# Zadanie 11.1

## Praca ze zbiorami (instrukcja)

infoShare ACADEMY

Pewna firma rozważa dołączanie do zakupu pakietu próbek kosmetyków. Do wyboru jest 10 różnych produktów, przy czym pakiet testowy ma się składać z próbek 3 różnych produktów. Ile jest możliwych pakietów próbek? Wypisz wszystkie rodzaje pakietów gratisowych.

# Statystyka

## Prawdopodobieństwo

infoShareAcademy.com

infoShare ACADEMY

# Statystyka

## Aksjomatyczna definicja prawdopodobieństwa

Aksjomat I. Dla każdego $A \in \Omega$

$$P(A) \geq 0$$

Aksjomat II. Prawdopodobieństwo zdarzenia pewnego

$$P(\Omega) = 1$$

Aksjomat III. Dla każdej pary $A_1$, $A_2$... ze zdarzeń rozłącznych parami zachodzi

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

info Share ACADEMY

# Statystyka

**Aksjomatyczna definicja prawdopodobieństwa**

- Prawdopodobieństwo zdarzenia niemożliwego wynosi zero $P$ $(\emptyset) = 0$
- Prawdopodobieństwo zdarzenia przeciwnego równe jest $P$ $(A')=1-P(A)$
- Dla dowolnych dwóch zdarzeń zachodzi $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Jeśli A C B, to $P(A) \leq P(B)$
- Dla każdego zdarzenia A $P(A) \leq 1$

# Statystyka

**Klasyczna interpretacja prawdopodobieństwa (przypadek szczególny)**

infoShare ACADEMY

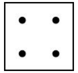
Niech wyniki doświadczenia losowego będą jednakowo prawdopodobne i niech możliwych wyników tego doświadczenia będzie *N*.

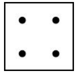
Jeżeli zdarzenie *A* składa się z *k* elementów (czyli *k* zdarzeń elementarnych, tzw. zdarzeń sprzyjających), to:

$$P(\mathrm{A}) = \frac{\mathrm{k}}{\mathrm{N}}$$

# Statystyka

## Klasyczna interpretacja prawdopodobieństwa

Prawdopodobieństwo otrzymania  wynosi:

$P($  $) = \frac{1}{6}$

# Statystyka

## Klasyczna interpretacja prawdopodobieństwa

Prawdopodobieństwo otrzymania liczby oczek mniejszej niż 3 wynosi:

$$P(\boxed{\text{⚀}} \cup \boxed{\text{⚁}}) = P(\boxed{\text{⚀}}) + (\boxed{\text{⚁}}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$


info Share ACADEMY

# Statystyka

## Klasyczna interpretacja prawdopodobieństwa

Niech przestrzeń zdarzeń elementarnych składa się z $N$ elementów

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$$

i niech prawdopodobieństwo zajścia zdarzenia elementarnego $\omega_i$ wynosi $P(\omega_i)$. Wówczas, dla każdego zdarzenia $A \subset \Omega$

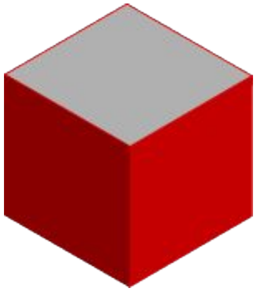
$$P(A) = \Sigma_{\omega i \in A} P(\omega_i)$$

infoShare ACADEMY

# Statystyka

## Klasyczna interpretacja prawdopodobieństwa (przykład)

Dana jest kostka sześcienna jak na rysunku (dwie ściany szare, cztery czerwone).
Określ prawdopodobieństwa zdarzeń polegających na tym, że w wyniku pojedynczego rzutu otrzyma się ścianę:



- czerwoną,
- szarą

$$\Omega = \{\bullet, \bullet\}$$

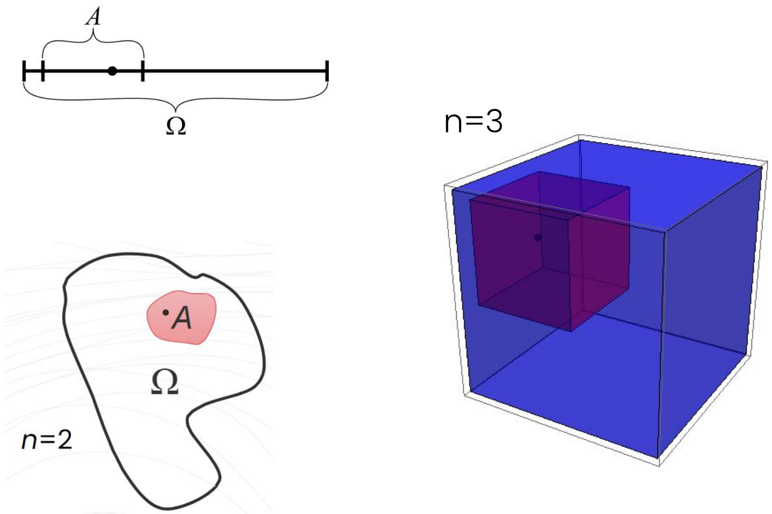
$$P(\bullet)=4/6$$

$$P(\bullet)=2/6$$

infoShare ACADEMY

# Statystyka

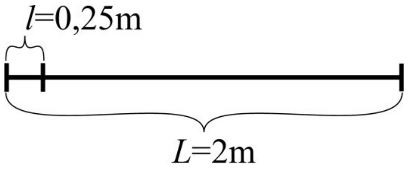
## Geometryczna interpretacja prawdopodobieństwa (przykład)




infoShare ACADEMY

# Statystyka

## Geometryczna interpretacja prawdopodobieństwa (przykład)



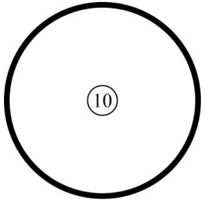

Dany jest odcinek o długości $L$=2m.

Jakie jest prawdopodobieństwo zdarzenia $A$, że losowo wybrany punkt znajdzie się w zbiorze $l=\{x: x>0 \text{ i } x<0{,}25 \text{ m}\}$?

$$P(A)=\frac{l}{L}=\frac{0{,}25}{2}=\frac{1}{8}$$

# Statystyka

## Geometryczna interpretacja prawdopodobieństwa (przykład)




Jakie jest prawdopodobieństwo, że łucznik trafi w „10” jeśli średnica tarczy wynosi D a średnica obszaru „10” d. Założyć, że łucznik zawsze trafia w tarczę, a szansa trafienia w którykolwiek jej punkt jest taka sama.

$$P(A)=\frac{A_{10}}{L_T}=\frac{\frac{\pi d^2}{4}}{\frac{\pi D^2}{4}}=\frac{d^2}{D^2}$$



infoShare ACADEMY

# Statystyka

## Statystyczna interpretacja prawdopodobieństwa

Prawdopodobieństwo zdarzenia $A$ równe jest

$$P(\mathrm{A}) = \lim_{N \to \infty} \frac{k}{N}$$

$k$ – liczba eksperymentów w których otrzymano zdarzenie sprzyjające

$N$ – liczba wszystkich eksperymentów

# Statystyka

## Prawdopodobieństwo warunkowe

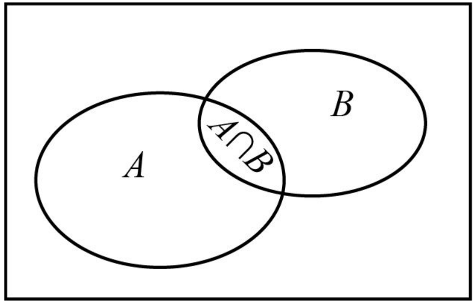
info Share ACADEMY

Niech A i B będą dowolnymi zdarzeniami, $A \subset \Omega$ i $B \subset \Omega$, przy czym prawdopodobieństwo zdarzenia $A$ jest dodatnie $P(A) > 0$. Prawdopodobieństwo warunkowe zdarzenia $B$ pod warunkiem zajścia zdarzenia $A$ jest dane wzorem:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

# Statystyka

## Prawdopodobieństwo warunkowe



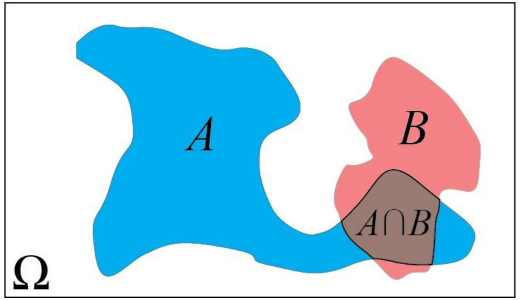

Prawdopodobieństwo wystąpienia zdarzenia $B$, pod warunkiem, że wystąpiło zdarzenie $A$ jest równe prawdopodobieństwu wystąpienia zdarzenia $A \cap B$ pod warunkiem wystąpienia zdarzenia $A$.

$$P\ B\ \ A = P(A \cap B|\mathrm{A})$$

infoShareAcademy.com

# Statystyka

## Prawdopodobieństwo warunkowe (objaśnienia)




Jeśli wystąpiło jakieś zdarzenie $A$, a rozważane jest inne zdarzenie $B$, oznacza to, że przestrzeń zdarzeń elementarnych została zredukowana z Ω do $A$.

# Statystyka

## Prawdopodobieństwo warunkowe (objaśnienia)

Prawdopodobieństwo (nie warunkowe) zdarzenia i prawdopodobieństwo warunkowe są jakościowo innymi pojęciami:

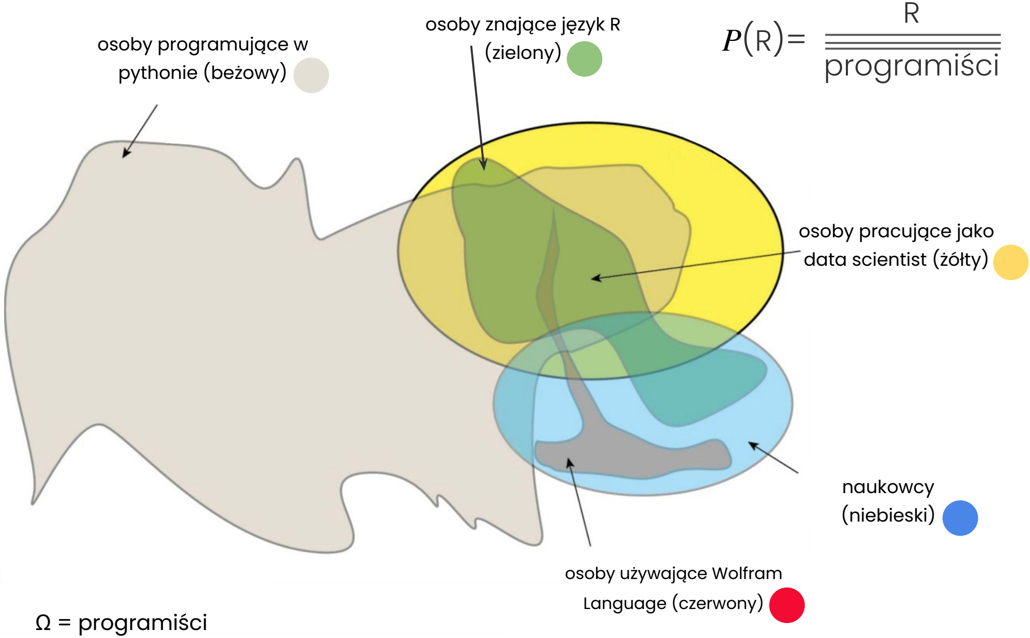
- Prawdopodobieństwo znalezienia programisty, który zna język R.
- Prawdopodobieństwo znalezienia programisty, który zna język R jeśli pracuje jako Data Scientist.

infoShare ACADEMY

# Statystyka

## Prawdopodobieństwo warunkowe (objaśnienia)

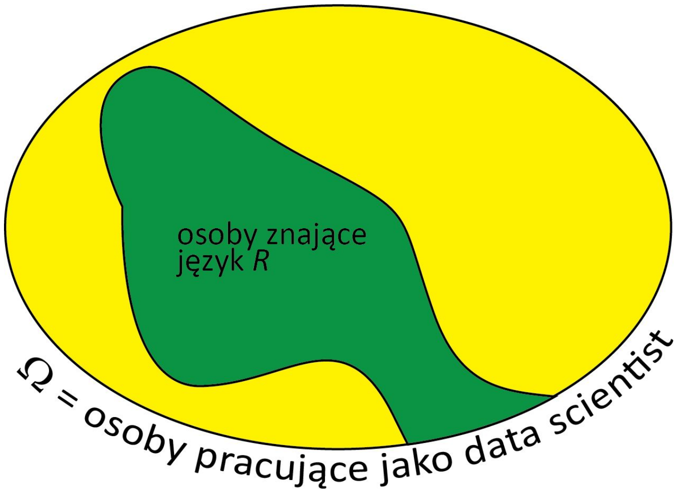
$$P(R) = \frac{R}{\text{programiści}}$$




Ω = programiści

# Statystyka

**Prawdopodobieństwo warunkowe (objaśnienia)**




$$P(\mathrm{R}|\mathrm{DS}) = \frac{P(\mathrm{DS} \cap \mathrm{R})}{P(\mathrm{DS})}$$

Inna przestrzeń zdarzeń elementarnych!

# Zadanie 11.2 (instrukcja)

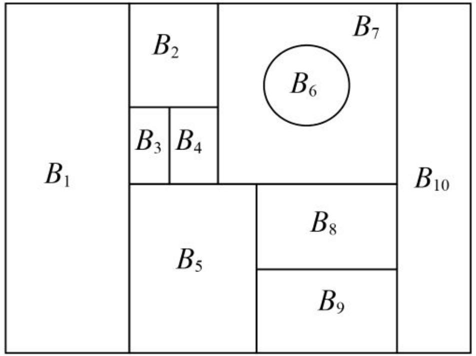
info Share ACADEMY

Na podstawie danych ze strony kaggle (https://www.kaggle.com/datasets/jaimevalero/developers-and-programming-languages) dotyczących użycia języków programowania oraz środowisk programistycznych obliczyć prawdopodobieństwo, że respondent:

- używa języka Python,
- używa środowiska Jupyter,
- używa środowiska Jupyter oraz Pythona,
- używa środowiska Jupyter jeśli wiadomo, że używa Pythona,
- używa Pythona jeśli wiadomo, że używa środowiska Jupyter.

# Statystyka

## Prawdopodobieństwo warunkowe (objaśnienia)




Zdarzenia $B_1$, $B_2$, ..., $B_k$ tworzą podział przestrzeni zdarzeń elementarnych Ω (lub układ zupełny zdarzeń),

jeżeli

$$B_i \cap B_j = \varnothing, \text{ gdy } i \neq j$$

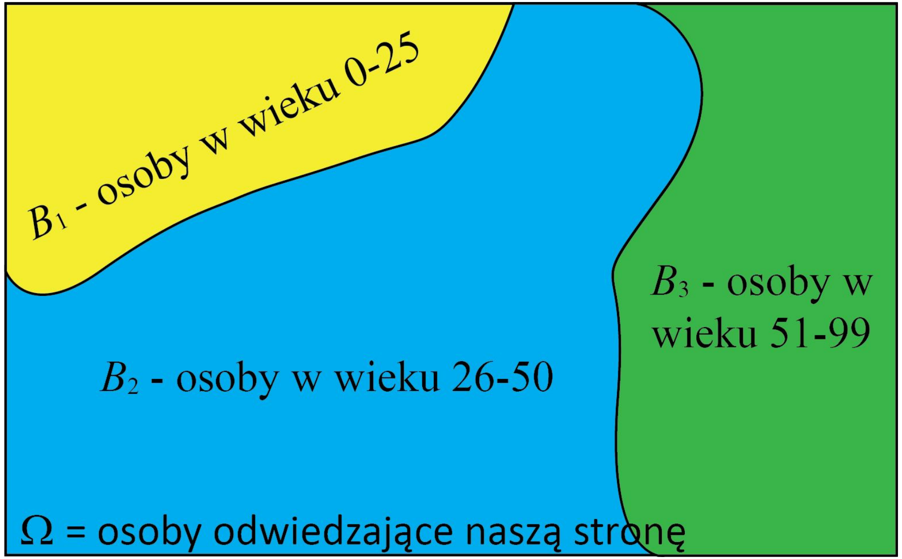
oraz

$$B_1 \cup B_2 \cup \ldots \cup B_k = \Omega$$

infoShare ACADEMY

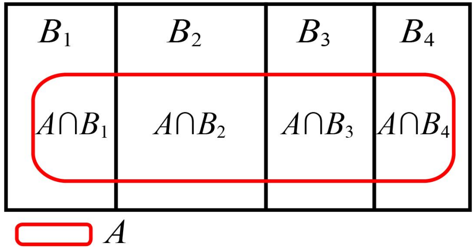
# Statystyka

## Zupełny podział przestrzeni zdarzeń elementarnych



# Statystyka

## Prawdopodobieństwo zupełne



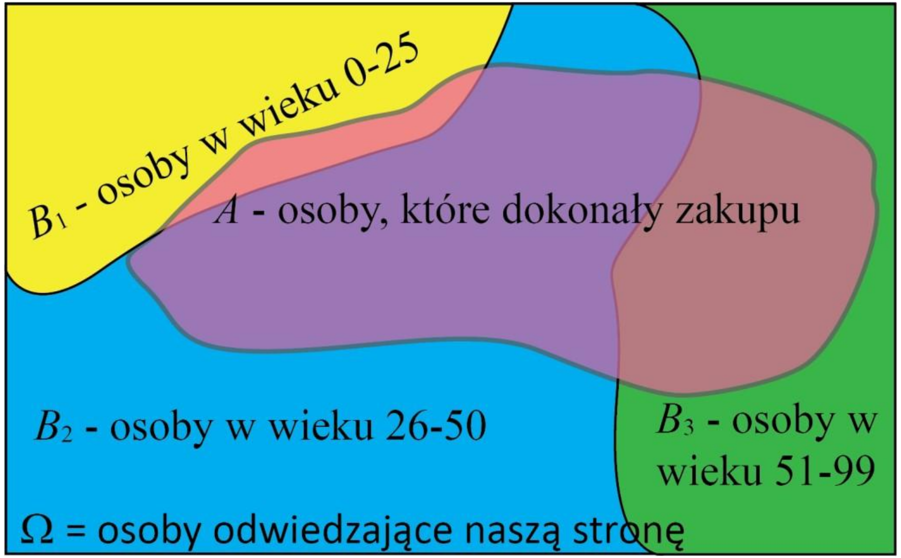

Jeżeli zdarzenia $B_1, B_2, \dots, Bk$ tworzą podział przestrzeni zdarzeń elementarnych W, prawdopodobieństwa zajścia każdego z nich są większe od zera $P(B_i) \neq 0$, dla $i = 1,2, \dots, k$, to dla każdego zdarzenia

$A \subset \Omega$.

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)$$

# Statystyka

## Prawdopodobieństwo (przykład)



# Statystyka

## Prawdopodobieństwo zupełne (przykład)

Klienci pewnego sklepu są klasyfikowani do trzech grup wiekowych: $B_1$ – młodzież, $B_2$ – dorośli, $B_3$ – starsi. Prawdopodobieństwo dokonania zakupu w każdej z tych grup wynosi odpowiednio 0,25, 0,7 oraz 0,5. Udział tych grup wiekowych wśród odwiedzających wynosi odpowiednio 0,2, 0,5 i 0,3 Oblicz prawdopodobieństwo, że osoba odwiedzająca stronę dokona zakupu $(A)$.

$$P(A) = \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i) = 0{,}25 \cdot 0{,}2 + 0{,}7 \cdot 0{,}5 + 0{,}5 \cdot 0{,}3 = 0,55$$

# Statystyka

## Twierdzenie Bayesa

infoShare ACADEMY

Jeżeli zdarzenia $B_1, B_2, \ldots, B_k$ tworzą podział przestrzeni zdarzeń elementarnych $\Omega$, prawdopodobieństwa zajścia każdego z nich są większe od zera $P(Bi) \neq 0$, dla $i = 1,2, \ldots, k$, to dla każdego zdarzenia $A \subset \Omega$, że $P(A) \neq 0$

$$P(B_m|A) = \frac{P(A|B_m) \cdot P(B_m)}{P(A)} = \frac{P(A|B_m) \cdot P(B_m)}{\sum_{j=1}^{k} P(A|B_j) \cdot P(B_j)}$$

# Statystyka

## Twierdzenie Bayesa (objaśnienie)

infoShare ACADEMY

Twierdzenie Bayesa pozwala na aktualizację informacji (prawdopodobieństw różnych zdarzeń) w oparciu o nowe dowody np.:

- prawdopodobieństwo, że osoba jest stałym klientem $P(B_i)$,
- prawdopodobieństwo, że osoba jest stałym klientem jeśli zalogowała się na konto sklepu (aktualizacja informacji) $P(B_i|A)$.

# Statystyka

## Twierdzenie Bayesa (objaśnienie)

infoShare ACADEMY

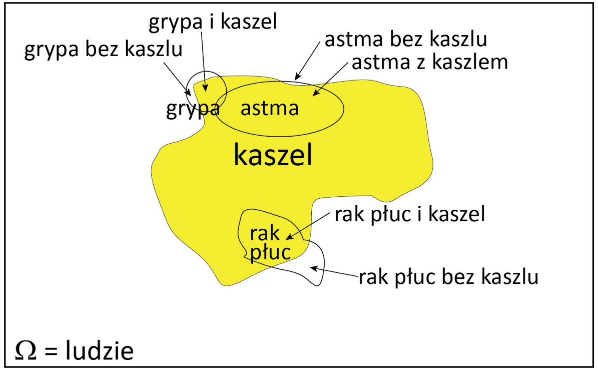
Twierdzenie Bayesa jest używane do określania możliwych przyczyn wystąpienia zjawisk: pewien stan wystąpił, jaka jest szansa, że jego przyczyną jest pewne wyszczególnione zdarzenie.

$$P(przyczyna_i|stan) = \frac{P(stan|przyczyna_i) \cdot P(przyczyna_i)}{P(stan)}$$

Uwaga: określenie co jest stanem, a co przyczyną jest umowne, choć zwykle wynika z logicznego opisu problemu.

# Statystyka

## Twierdzenie Bayesa (objaśnienie)




Prawdopodobieństwo, że pacjent ma grypę:

P(grypa)

Prawdopodobieństwo, że pacjent ma grypę, jeśli ma kaszel:

P(grypa|kaszel) (aktualizacja wiedzy – kaszel)

# Statystyka

## Twierdzenie Bayesa (objaśnienie 2)

infoShare ACADEMY

Twierdzenie Bayesa jest używane do określania efektywności testów medycznych, algorytmów klasyfikujących itp. ponieważ możliwe jest wnioskowanie o zachowaniu testów w rzeczywistych przypadkach tj. uwzględnieniu proporcji np. osób chorych itp.

# Statystyka

## Twierdzenie Bayesa (przykład)

1 na 1000 klientów pewnego banku okazuje się być osobą nieuczciwą.

Dział fraud detection tego banku opracował zestaw algorytmów poprawnie wykrywający oszusta z prawdopodobieństwem 0,99 (tzn. 99 na 100 oszustów udaje się wykryć). Niestety zdarza się, że system klasyfikuje osoby zupełnie uczciwe jako oszustów z prawdopodobieństwem 0,02 (2 na 100 uczciwych klientów jest niesłusznie oskarżonych). Oblicz prawdopodobieństwo, że osoba, którą system zaklasyfikował jako oszusta rzeczywiście jest nieuczciwa.

# Statystyka

## Twierdzenie Bayesa (przykład)

infoShare ACADEMY

Oznaczmy:

$A$ – zaklasyfikowanie klienta jako oszusta,

$B_1$ – klient jest nieuczciwy, $B_2$ – klient jest uczciwy,

$A|B_1$ – klient został zaklasyfikowany jako oszust, jeśli jest nieuczciwy,

$A|B_2$ – klient został zaklasyfikowany jako oszust, jeśli jest uczciwy,

$B_1|A$ – klient jest nieuczciwy i został zaklasyfikowany jako oszust,

$B_2|A$ – klient jest uczciwy, a został zaklasyfikowany jako oszust.

$P(B_1) = 0{,}001$ $\quad$ $P(B_2)=0{,}999$

$P(A|B_1)=0{,}99$ $\quad$ $P(A|B_2)=0{,}02$

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = 0{,}047$$

# Statystyka

## Twierdzenie Bayesa (przykład)

infoShare ACADEMY

$B_1|A$ – klient jest rzeczywiście nieuczciwy, jeśli został zaklasyfikowany jako oszust, $B_2|A$ – klient jest uczciwy, a został zaklasyfikowany jako oszust.

$P(B_1) = 0{,}001$ $\quad$ $P(B_2)=0{,}999$

$P(A|B_1)=0{,}99$ $\quad$ $P(A|B_2)=0{,}02$

$$P(B_1|A) = 0{,}047$$

$$P(B_2|A) = 1 - P(B_1|A) = 0{,}953$$

**Co to dokładnie oznacza ?!**

# Statystyka

## Twierdzenie Bayesa (przykład, interpretacja)

Jeżeli model skutecznie wykrywa 99% oszustów, a niesłusznie oskarża 2% uczciwych klientów, to przy średniej ilości nieuczciwych klientów wynoszącej 0,1%, aż 95,3% osób jest uczciwych mimo, że zostały zaklasyfikowane jako oszuści, a zaledwie 4,7% rzeczywiście okaże się oszustami jeśli system ich tak zaklasyfikował.

Czyli:
Jeśli bank zaklasyfikował klienta jako oszusta, istnieje 95,3% szans, że nie jest oszustem.

**Jak temu zaradzić?**

| | | obserwacja | |
|---|---|---|---|
| Predykcja | | uczciwy | oszust |
| | uczciwy | 97902 | 1 |
| | oszust | 1998 | 99 |
| 100000 - liczba analizowanych | | | |
| 100 - liczba oszustów | | | |
| 99900 - liczba uczciwych | | | |

infoShare ACADEMY

# Statystyka

## Nikt nigdy nie zapyta Was o:

- prawdopodobieństwo warunkowe, że programista używa Pythona jeśli wiadomo, że używa Jupytera,
- prawdopodobieństwo zupełne wylosowania kuli czarnej z jednej z urn, w których proporcje kul białych i czarnych wynoszą odpowiednio…

# Statystyka

## Ale mogą zapytać:

- ile osób (statystycznie rzecz biorąc) można natychmiast wdrożyć do projektów w Pythonie, skoro w naszych oddziałach raporty przygotowuje się używając głównie Jupytera,

- na ile (statystycznie rzecz biorąc) klient jest skłonny kupić coś w naszym sklepie internetowym, skoro badania pokazały, że tyle i tyle osób dokonuje zakupów przy przekierowaniu z takiej a takiej strony, a tyle z takiej,

- czy opracowany właśnie test do wykrywania raka nadaje się do badań przesiewowych.

infoShare ACADEMY

# Statystyka

## Zmienna losowa

info Share ACADEMY

Dowolną funkcję o wartościach rzeczywistych, określoną na zbiorze zdarzeń elementarnych Ω, nazywamy zmienną losową.

$$RV(\Omega) \to X$$

# Statystyka

## Zmienna losowa (objaśnienie)

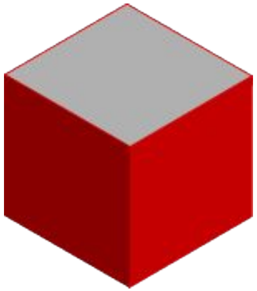
infoShare ACADEMY

Dana jest kostka sześcienna jak na rysunku
(dwie ściany szare, cztery czerwone).
Określ zmienną losową dla tego eksperymentu.

$\Omega = \{$●, ●$\}$ (zdarzenia)

$\downarrow$ (zmienna losowa)

$X = \{1, 2\}$ (liczby rzeczywiste)



# Statystyka

## Zmienna losowa (objaśnienie)

infoShare ACADEMY

1. Zmienna losowa przyporządkowuje liczby rzeczywiste zdarzeniom.
2. Dowolnemu zdarzeniu z przestrzeni zdarzeń elementarnych Ω może zostać przyporządkowana dowolna liczba.
3. Zmienne losowe oznacza się zwykle wielkimi literami. Wartości osiągane przez zmienne losowe oznaczane są małymi literami, np.:

$$P(X < x)$$

czyta się jako: „prawdopodobieństwo, że zmienna losowa $X$ przyjmie wartość mniejszą niż $x$”.

# Statystyka

## Rodzaje zmiennych losowych

Ze względu na typy wartości:

- dyskretne,
- ciągłe,
- mieszane.

Ze względu na ilość wymiarów:

- jednowymiarowe,
- wielowymiarowe.



infoShare ACADEMY

# Statystyka

## Prawdopodobieństwo przyjęcia wartości przez zmienną losową

Następujące zbiory zdarzeń elementarnych są zdarzeniami losowymi:

$$\{\omega\colon X(\omega)\epsilon A\},$$

gdzie $A$ jest zbiorem borelowskim na prostej, np.:

$$A=\{x_0\},\ A=[a,b],$$
$$A=[x_0,+\infty),\ \text{itp.}$$

Prawdopodobieństwo, że zmienna losowa przyjmie wartość ze zbioru $A\in\mathbb{R}$ wynosi:

$$P\ X\epsilon A\ = P[\{\omega\colon X(\omega)\epsilon A\}]$$

infoShare ACADEMY

# Statystyka

## Dystrybuanta

Niech $X$ będzie dowolną zmienną losową. Dystrybuantą zmiennej losowej $X$ nazywa się funkcję $F$ określoną dla dowolnego x$\in \mathbb{R}$, takiego, że

$$F\ x = P(X \leq x)$$

Własności dystrybuanty:

1. jest funkcją niemalejącą,
2. $\lim F(x) = 0$; $\lim F(x) = 1$,
3. jest przynajmniej lewostronnie ciągła.

# Statystyka

## Zmienna losowa dyskretna

Zmienna losowa $X$ jest dyskretna jeśli istnieje skończony lub przeliczalny zbiór $\Omega = \{x1, \ldots, xn, \ldots\}$ jej wartości $x1, \ldots, xn, \ldots$ takich, że:

$$P(X = x_i) = P_i > 0, \qquad i \in \mathbb{N}$$

$$\sum_{i=1}^{\infty} p_i = 1$$

# Statystyka

## Rozkład prawdopodobieństwa

Funkcję $p$ określoną na zbiorze $\Omega = \{x1, \dots, xn, \dots\}$ następująco

$$p(x_i) = P(X = x_i) = p_i, \qquad x_i \in \Omega$$

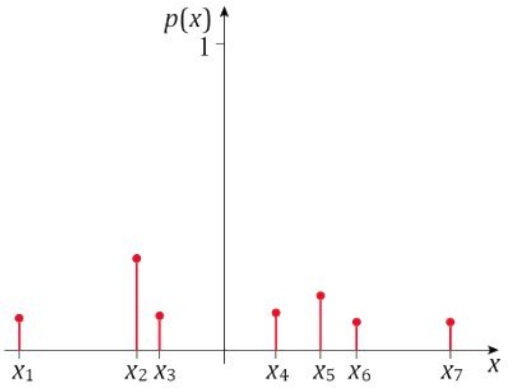
$$\sum_{i=1}^{\infty} p_i = 1$$

nazywa się funkcją rozkładu prawdopodobieństwa (masową funkcją prawdopodobieństwa).

# Statystyka

## Rozkład prawdopodobieństwa

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | ... | $x_n$ | ... |
| $p_i$ | $p_1$ | $p_2$ | ... | $p_n$ | ... |



# Statystyka

## Dystrybuanta dyskretnej zmiennej losowej

Dla dyskretnej zmiennej losowej, dystrybuanta $F\ x = P(X \leq x)$ przyjmuje postać

$$F(x) = P(X \leq x) = \sum_{x_i : x_i \leq x} p_i$$

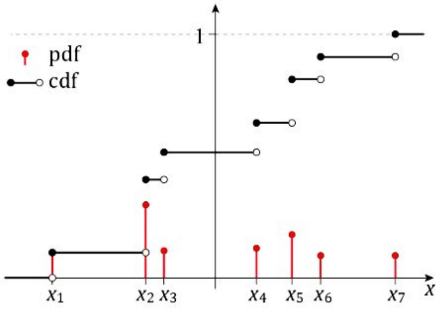


info Share ACADEMY

# Statystyka

## Dystrybuanta dyskretnej zmiennej losowej

| $x_i$ | $(-\infty, x_1)$ | $[x_1, x_2)$ | $[x_2, x_3)$ | ... | $[x_{n-1}, x_n)$ | ... |
|---|---|---|---|---|---|---|
| $F(x)$ | 0 | $p_1$ | $p_1 + p_2$ | ... | $p_1 + p_2 + \cdots + p_n$ | ... |




infoShareAcademy.com

info Share ACADEMY

# Statystyka

## Wartość oczekiwana zmiennej losowej dyskretnej

Wartość oczekiwana zmiennej losowej dyskretnej $X$, której rozkład prawdopodobieństwa to $p(x)$ jest następująco zdefiniowaną liczbą

$$E(X) = \mu_X = \sum_{i=1}^{\infty} x_i \cdot p(x_i)$$

Gdy Ω jest zbiorem skończonym to

$$E(X) = \mu_X = \sum_{i=1}^{k} x_i \cdot p(x_i)$$

# Statystyka

## Wartość wariancji zmiennej losowej dyskretnej

Wariancją zmiennej losowej dyskretnej $X$, której rozkład prawdopodobieństwa to $p(x)$ jest następująco zdefiniowaną liczbą

$$V(X) = \sigma_X^2 = \sum_{i=1}^{\infty} (x_i - E(X))^2 \cdot p(x_i)$$

Gdy Ω jest zbiorem skończonym to

$$V(X) = \sigma_X^2 = \sum_{i=1}^{k} (x_i - E(X))^2 \cdot p(x_i)$$

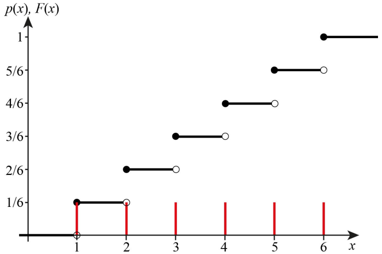


info Share ACADEMY

# Statystyka

## Zmienna losowa dyskretna (przykład)

| Ω | ⚀ | ⚁ | ⚂ | ⚃ | ⚄ | ⚅ |
|---|---|---|---|---|---|---|
| $X(\omega_i)$ | 1 | 2 | 3 | 4 | 5 | 6 |
| $p(x_i)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |
| $F(x_i)$ | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 1 |





info Share
ACADEMY

# Statystyka

## Wartość oczekiwana (przykład)

| $X(\omega_i)$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $p(x_i)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

$$E(X) = \mu_X = \sum_{i=1}^{6} x_i \cdot p(x_i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3\tfrac{1}{2}$$

# Statystyka

## Wariancja (przykład)

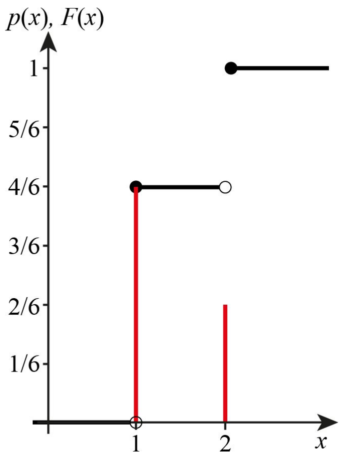
$$E(X) = 3{,}5$$

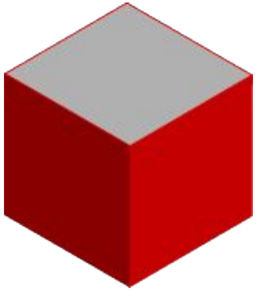
$$V(X) = \sum_{i=1}^{6} (x_i - E(X))^2 \cdot p(x_i) =$$

$$(1-3{,}5)^2 \cdot \frac{1}{6} + (2-3{,}5)^2 \cdot \frac{1}{6} + \cdots + (6-3{,}5)^2 \cdot \frac{1}{6} = 2{,}917$$

# Statystyka

## Zmienna losowa dyskretna (przykład)



| Ω | ● | ● |
|---|---|---|
| $X(\omega_i)$ | 1 | 2 |
| $p(x_i)$ | 4/6 | 2/6 |
| $F(x_i)$ | 4/6 | 1 |

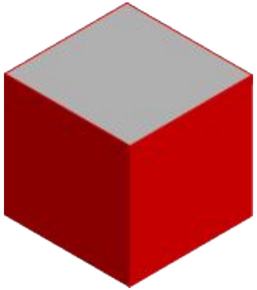
infoShare ACADEMY

# Statystyka

## Wartość oczekiwana (przykład)

| | | |
|---|---|---|
| $X(\omega_i)$ | 1 | 2 |
| $p(x_i)$ | 4/6 | 2/6 |

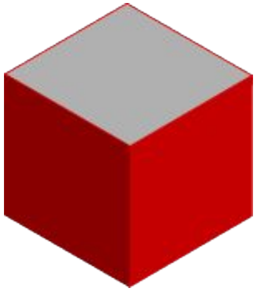
$$E(X) = \mu_X = \sum_{i=1}^{2} x_i \cdot p(x_i) = 1 \cdot \frac{4}{6} + 2 \cdot \frac{2}{6} = \frac{4}{3}$$



infoShare ACADEMY

# Statystyka

## Wariancja (przykład)



| $X(\omega_i)$ | 1 | 2 |
|---|---|---|
| $p(x_i)$ | 4/6 | 2/6 |

$$E(X) = \frac{4}{3}$$

$$V(X) = \sum_{i=1}^{2}(x_i - E(X))^2 \cdot p(x_i) = \left(1 - \frac{4}{3}\right)^2 \cdot \frac{4}{6} + \left(2 - \frac{4}{3}\right)^2 \cdot \frac{2}{6} = \frac{2}{9}$$

# Zadanie 11.3 (instrukcja)

1. Znaleźć rozkład prawdopodobieństwa, dystrybuantę, wartość oczekiwaną i wariancję zmiennej losowej X – sumy oczek otrzymanej w wyniku rzutu dwiema kostkami do gry.

# Statystyka

## Zmienna losowa ciągła

Zmienną losową $X$ nazywa się ciągłą, jeśli dla dowolnej nieujemnej funkcji $f$ i dowolnych liczb rzeczywistych $a$ i $b$, takich że $-\infty \leq a < b \leq \infty$, zachodzi

$$P(a \leq X \leq b) = \int_a^b f(t)dt$$

oraz

$$\int_{-\infty}^{\infty} f(t)dt = 1$$

infoShare ACADEMY

# Statystyka

## Gęstość rozkładu prawdopodobieństwa

infoShare ACADEMY

Dowolna nieujemna funkcja $f(x)$, gdzie $x \in \mathbb{R}$, dla której zachodzi

$$\int_{-\infty}^{\infty} f(t)dt = 1$$

jest gęstością rozkładu prawdopodobieństwa.

# Statystyka

## Dystrybuanta zmiennej losowej ciągłej

Dla ciągłej zmiennej losowej, dystrybuanta $F(x) = P(X \leq x)$ przyjmuje postać

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

zatem zachodzi również

$$F' x = f(x)$$

infoShare ACADEMY

# Statystyka

## Wartość oczekiwana zmiennej losowej ciągłej

Wartość oczekiwana zmiennej losowej ciągłej $X$, której gęstość rozkładu prawdopodobieństwa to $f(x)$ jest następująco zdefiniowaną liczbą

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

# Statystyka

## Wariancja zmiennej losowej ciągłej

Wariancją zmiennej losowej ciągłej $X$, której rozkład prawdopodobieństwa to $f(x)$ jest następująco zdefiniowaną liczbą

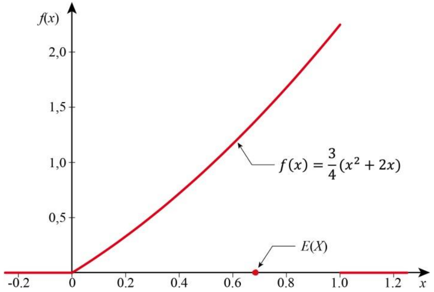
$$V(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)dx$$

# Statystyka

## Wartość oczekiwana zmiennej losowej ciągłej (przykład)

Zmienna losowa $X$ ma w przedziale [0,1] ma gęstość rozkładu prawdopodobieństwa $f(x) = \frac{3}{4}(x^2 + 2x)$ i zero poza nim.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{-\infty}^{0} x \cdot 0\,dx + \int_{0}^{1} x \cdot \frac{3}{4}(x^2 + 2x)\,dx + \int_{1}^{\infty} x \cdot 0\,dx$$




$$0 + \frac{3}{4}\int_0^1 (x^3 + 2x^2)\,dx + 0 = \frac{3}{4}\left(\int_0^1 x^3 dx + 2\int_0^1 x^2 dx\right) =$$

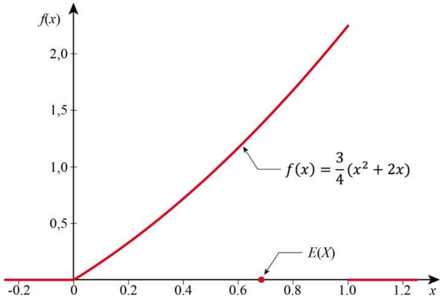
$$= \frac{3}{4}\left(\left[\frac{1}{4}x^4\right]_0^1 + 2\left[\frac{1}{3}x^3\right]_0^1\right) = \frac{11}{16}$$

# Statystyka

## Wariancja zmiennej losowej ciągłej (przykład)

Zmienna losowa $X$ ma w przedziale [0,1] ma gęstość rozkładu prawdopodobieństwa $f(x) = (x + 2x)$ i zero poza nim.

$$E(X) = \frac{11}{16}$$




$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)\,dx$$

$$= \int_0^1 \left(x - \frac{11}{16}\right)^2 \cdot \frac{3}{4}\,(x^2 + 2x)\,dx = \frac{67}{1280}$$

# Zadanie 11.4 (instrukcja)

infoShare ACADEMY

Zmienna losowa $X$ ma rozkład normalny $N(125,15)$. W języku Python:

1. Utwórz instancję tego rozkładu,
2. Narysuj gęstość rozkładu prawdopodobieństwa i dystrybuantę w zakresie $\mu$ - $4\sigma$; $\mu + 4\sigma$ ,
3. Oblicz prawdopodobieństwo, że zmienna losowa osiągnie wartość z zakresu 130 do 145,
4. Oblicz wartość oczekiwaną,
5. Oblicz odchylenie standardowe tej zmiennej,
6. Oblicz kwantyl 0,25 tej zmiennej losowej.

# Statystyka

## Podstawowe charakterystyki zmiennej losowej

infoShareAcademy.com

info Share ACADEMY

# Statystyka

## Charakterystyki zmiennej losowej

Opisują położenie, rozproszenie, kształt i asymetrię zmiennej losowej.

Większość parametrów opisujących zmienną losową jest momentem lub kombinacją momentów zmiennej losowej.

# Statystyka

## Przykłady rozkładów teoretycznych opisujących rzeczywiste procesy

infoShare ACADEMY

Momentem zwykłym rzędu $k$ zmiennej losowej $X$ nazywamy liczbę

$$m_k = \mathrm{E}\left(X^k\right)$$

Wartość oczekiwana jest momentem zwykłym rzędu pierwszego:

$$E(X) = m_1.$$

# Statystyka

## Moment zwykły zmiennej losowej

Momentem centralnym rzędu $k$ zmiennej losowej $X$ nazywamy liczbę

$$\mu_k = \mathrm{E}\left((X - \mathrm{E}(X))^k\right) = \mathrm{E}\left((X - m_1)^k\right)$$

Moment centralny rzędu drugiego zawsze jest równy zeru.

Wariancja jest momentem centralnym rzędu drugiego: $V(X) = \mu_2$.



infoShare ACADEMY

# Statystyka

## Moment dyskretny zmiennej losowej

Moment zwykły:

$$m_k = \mathrm{E}(X^k) = \sum_{x \in \mathbb{R}} p(x_i) \cdot x_i^k$$

Moment centralny:

$$\mu_k = \mathrm{E}\left((X - m_1)^k\right) = \sum_{x \in \mathbb{R}} p(x_i) \cdot (x_i - m_1)^k$$

# Statystyka

## Moment dyskretnej zmiennej losowej (przykład)

| $x$ | $0$ | $3$ | $4$ |
|---|---|---|---|
| $p(x)$ | $0,7$ | $0,2$ | $0,1$ |

Dana jest zmienna losowa dyskretna $X$ o rozkładzie prawdopodobieństwa przedstawionym w tabeli.

Znajdź jej momenty zwykłe i centralne pierwszego i drugiego rzędu.

$$m_1 = \mathrm{E}(X) = \sum_{x \in \mathbb{R}} p(x_i) \cdot x_i \;= 0{,}7 \cdot 0 + 0{,}2 \cdot 3 + 0{,}1 \cdot 4 = 1$$

$$m_2 = \mathrm{E}(X^2) = \sum_{x \in \mathbb{R}} p(x_i) \cdot x_i^2 = 0{,}7 \cdot 0^2 + 0{,}2 \cdot 3^2 + 0{,}1 \cdot 4^2 = 3{,}4$$

# Statystyka

## Moment dyskretnej zmiennej losowej (przykład)

| $x$ | 0 | 3 | 4 |
|---|---|---|---|
| $p(x)$ | 0,7 | 0,2 | 0,1 |

Dana jest zmienna losowa dyskretna $X$ o rozkładzie prawdopodobieństwa przedstawionym w tabeli.

Znajdź jej momenty zwykłe i centralne pierwszego i drugiego rzędu.

$$\mu_1 = \mathrm{E}(X - m_1) = \sum_{x \in \mathbb{R}} p(x_i) \cdot (x_i - m_1) = 0{,}7 \cdot (0-1) + 0{,}2 \cdot (3-1) + 0{,}1 \cdot (4-1) = 0$$

$$\mu_2 = \mathrm{E}(X - m_1)^2 = \sum_{x \in \mathbb{R}} p(x_i) \cdot (x_i - m_1)^2$$

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

# Statystyka

## Moment ciągłej zmiennej losowej

Moment zwykły:

$$m_k = \mathrm{E}(X^k) = \int_{-\infty}^{\infty} x^k \cdot f(x)dx$$

Moment centralny:

$$\mu_k = \mathrm{E}\left((X - m_1)^k\right) = \int_{-\infty}^{\infty} (x - m_1)^k \cdot f(x)dx$$

infoShare ACADEMY

# Statystyka

## Parametry położenia

Charakterystyczne wielkości związane z położeniem rozkładu, np.:

- wartość oczekiwana,
- kwantyle (w tym kwartyle, percentyle, itd.),
- moda,
- ...

infoShare ACADEMY

# Statystyka

## Kwantyl

Kwantylem rzędu $p$ ($0<p<1$) zmiennej losowej $X$, której dystrybuantą jest funkcja $F$ jest taka liczba $x_p$, dla której zachodzi

$$F([x_p]^-) \leq p \leq F([x_p]^+)$$

Kwantyl rzędu $p$ dzieli rozkład prawdopodobieństwa w stosunku $p$: 1 - $p$

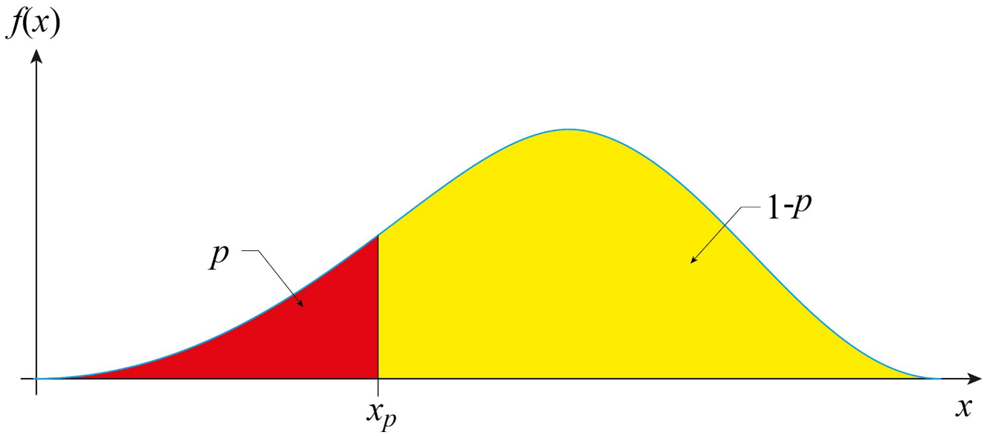
Mediana jest kwantylem rzędu 0,5.

Kwartyle są kwantylami rzędu 0,25, 0,5 i 0,75.

info Share ACADEMY

# Statystyka

## Kwantyl






infoShare ACADEMY

# Statystyka

## Moda

Wartość zmiennej losowej, dla której funkcja rozkładu lub funkcja gęstości rozkładu prawdopodobieństwa jest największa.

Moda jest wartością o największym prawdopodobieństwie wystąpienia. Rozkład może nie mieć mody lub mieć ich wiele.

# Statystyka

## Parametry rozproszenia

infoShare ACADEMY

Charakterystyczne wielkości związane rozrzutem wartości przyjmowanych przez zmienną losową:

- wariancja / odchylenie standardowe,
- odchylenie przeciętne,
- rozstęp,
- rozstęp międzykwartylowy,
- ...

# Statystyka

## Odchylenie przeciętne

infoShare ACADEMY

Odchyleniem przeciętnym zmiennej losowej $X$ nazywamy wartość

$$d_1 = E \quad X - E \; X$$

# Statystyka

## Odchylenie przeciętne zmiennej losowej dyskretnej

Odchyleniem przeciętnym dyskretnej zmiennej losowej $X$ nazywamy wartość

$$d_1 = \sum_{i=1}^{\infty} |(x_i - E(X)| \cdot p(x_i)$$

Jeśli Ω jest zbiorem skończonym zawierającym $k$ elementów

$$d_1 = \sum_{i=1}^{k} |(x_i - E(X)| \cdot p(x_i)$$

infoShare ACADEMY

# Statystyka

## Odchylenie przeciętne zmiennej losowej ciągłej

Odchyleniem przeciętnym ciągłej zmiennej losowej $X$ nazywamy wartość

$$d_1 = \int_{-\infty}^{\infty} |x - E(X)| \cdot f(x)dx$$



infoShare ACADEMY

# Statystyka

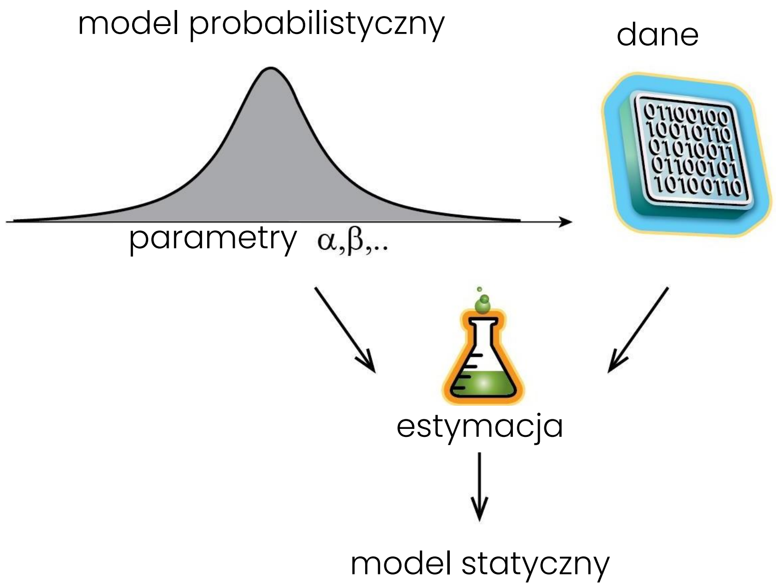
## Parametry kształtu

Charakterystyczne wielkości związane z położeniem rozkładu, np.:

- kurtoza,
- współczynnik asymetrii,
- …

# Statystyka

## Model statystyczny



# Statystyka

## Statystyczna analiza danych

Zebrane dane charakteryzuje się parametrami identycznymi jak te używane do charakteryzowania rozkładów, np.:

- wartość oczekiwana (średnia),
- odchylenie standardowe,
- kwantyle (mediana, IQR),
- rozstęp.



infoShare ACADEMY

# Statystyka

## Skąd się biorą teoretyczne rozkłady prawdopodobieństwa?

infoShareAcademy.com

infoShare ACADEMY

# Statystyka

**Czy Data Scientist potrzebuje teoretycznych rozkładów prawdopodobieństwa opisujących rzeczywiste zjawiska?**

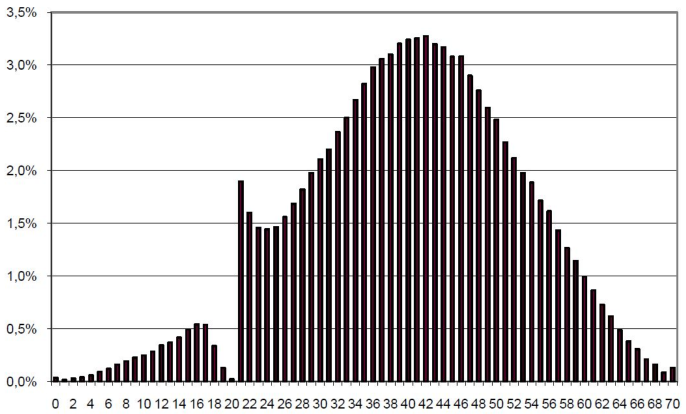
infoShareAcademy.com

infoShare ACADEMY

# Statystyka

## Rozkład ilości punktów na egzaminie maturalnym w 2010 r.

Zaliczenie uzyskuje się od 21 punktów.

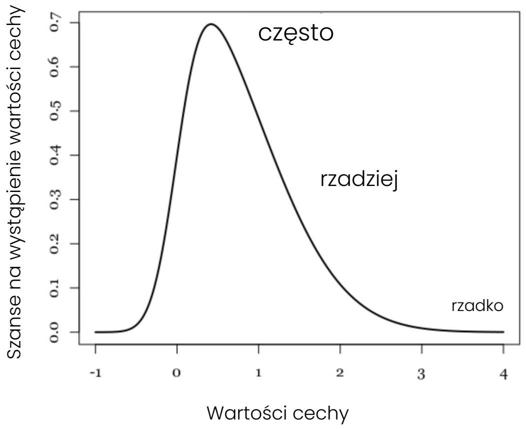



Wykres 1. Rozkład wyników na poziomie podstawowym

Źródło:www.cke.edu.pl

# Statystyka

## Przykłady rozkładów teoretycznych opisujących rzeczywiste procesy




infoShare ACADEMY

# Statystyka

## Rozkład dwupunktowy

Niech przestrzenią zdarzeń elementarnych będzie

$$\Omega = \{\omega_1, \omega_2\}$$

dziedziną zmiennej losowej

$$D = \{x, x_2\}$$

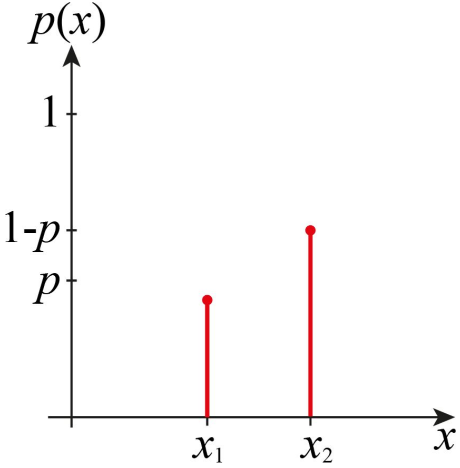
oraz

$$P(\omega \colon X(\omega) = x_1 = p$$

$$P(\omega \colon X(\omega) = x_2 = 1-p$$

gdzie

$$0 < p < 1$$




infoShareAcademy.com

# Statystyka

## Zdarzenie losowe: typy

Niech przestrzenią zdarzeń elementarnych będzie

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$

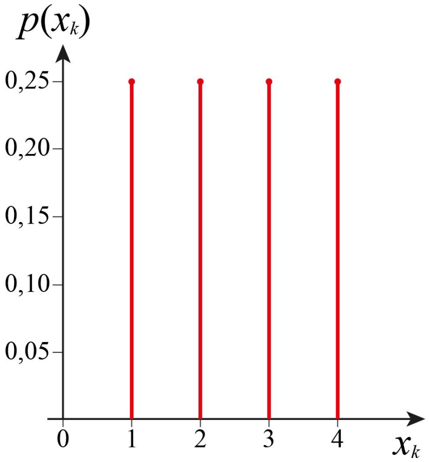
dziedziną zmiennej losowej

$D = \{x_1, x_2, \ldots, x_k\}$

oraz

$P(\{\omega : X(\omega) = x_k, k = 1,2, \ldots, N\}) =$

$$p = \frac{1}{N}$$

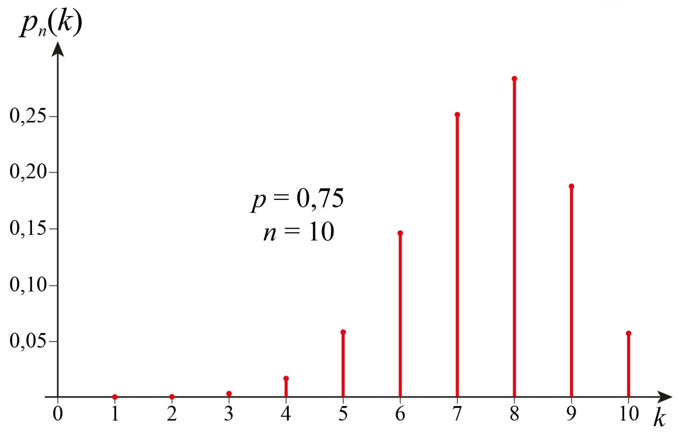


# Statystyka

## Rozkład dwumianowy

Zakładając schemat losowania zgodny ze schematem Bernoullego, prawdopodobieństwo zajścia $k$ razy zdarzenia $A$ w $n$ próbach równe jest:

$$p_n(k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$




infoShareAcademy.com

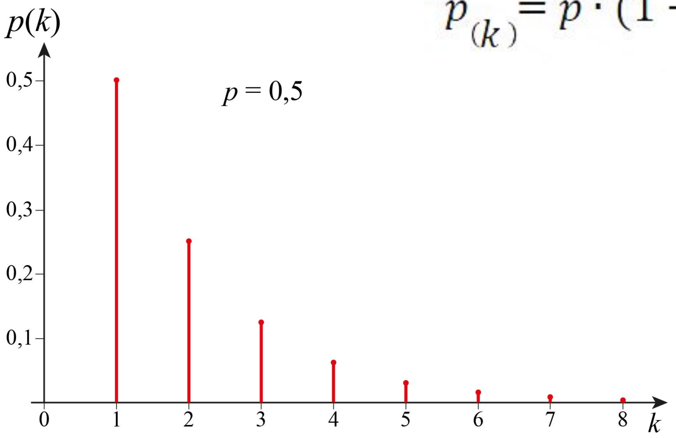
infoShare ACADEMY

# Statystyka

## Rozkład geometryczny

Zakładając schemat losowania zgodny ze schematem Bernoullego, prawdopodobieństwo zajścia zdarzenia $A$ pierwszy raz w próbie $k$ -tej równe jest:

$$p_{(k)} = p \cdot (1-p)^{k-1}$$






infoShare ACADEMY

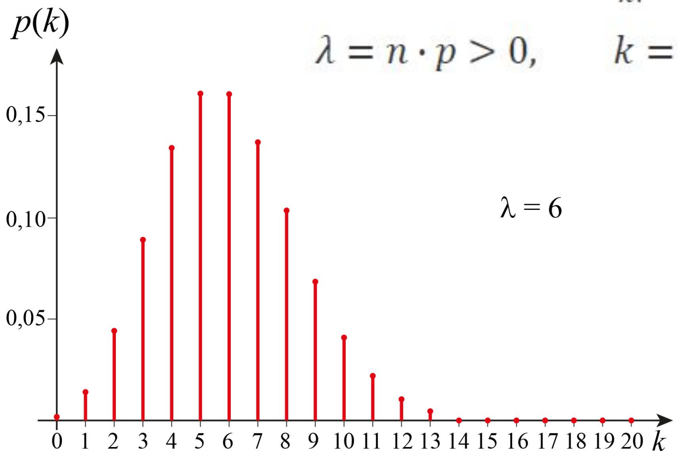
# Statystyka

## Rozkład Poissona

Zakładając schemat losowania zgodny ze schematem Bernoullego, mówimy, że zmienna losowa ma rozkład Poissona, jeśli $D = \{0, 1, 2, \dots\}$

$$p(k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

$$\lambda = n \cdot p > 0, \qquad k = 0, 1, 2, \dots$$

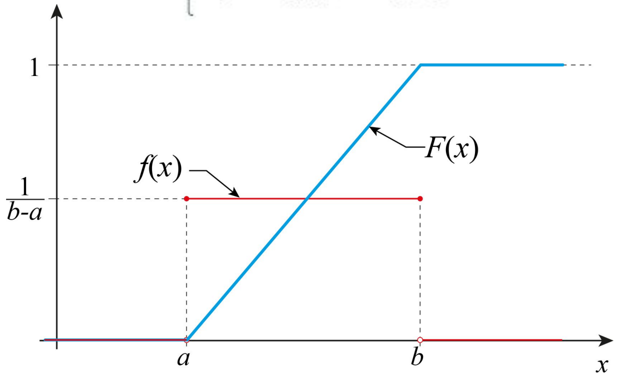


# Statystyka

## Ciągły rozkład jednostajny

Zmienna losowa X, ma rozkład jednostajny (prostokątny, równomierny) w przedziale <a,b>, jeśli jego gęstość opisana jest wzorem

$$f(x)=\begin{cases} 0 & \text{dla} \quad x<a \\ \frac{1}{b-a} & \text{dla} \quad a \le x \le b \\ 0 & \text{dla} \quad x>b \end{cases}$$



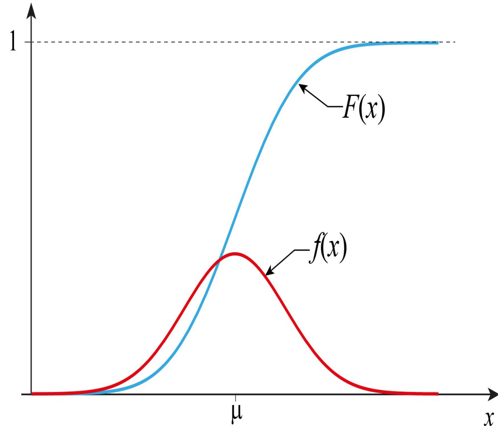

infoShareAcademy.com

# Statystyka

## Rozkład normalny

Zmienna losowa $X$, ma rozkład normalny o parametrach μ i $\sigma > 0$ jeśli jej gęstość prawdopodobieństwa opisana jest wzorem

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$




gdzie:

μ – parametr położenia,

σ – parametr rozproszenia.

infoShareAcademy.com

info Share ACADEMY

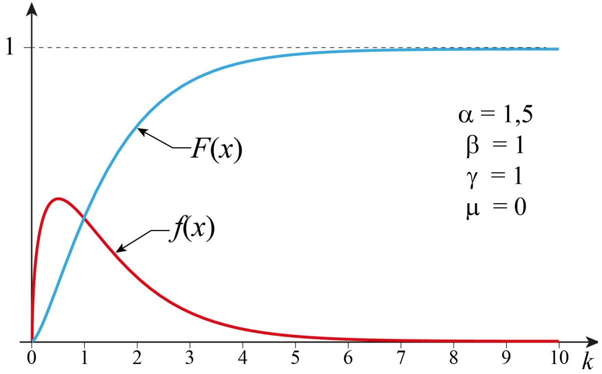
# Statystyka

## Rozkład gamma

Zmienna losowa $X$, ma uogólniony rozkład gamma o parametrach $\alpha > 0$, $\beta > 0$, $\gamma > 0$ i $\mu \in \mathbb{R}$ jeśli jej gęstość prawdopodobieństwa opisana jest wzorem:

$$f(x) = \begin{cases} 0 & dla \quad x \le \mu \\ \dfrac{\gamma \cdot \left(\frac{x-\mu}{\beta}\right)^{\alpha \cdot \gamma - 1}}{\beta \cdot \Gamma(\alpha)} \cdot e^{-\left(\frac{x-\mu}{\beta}\right)^{\gamma}} & dla \quad x > \mu \end{cases} \qquad \text{gdzie} \qquad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \cdot e^{-x} dx$$

$\alpha, \gamma$ – parametry kształtu, $\beta$ – parametr skalujący, $\mu$– parametr położenia.




infoShareAcademy.com

info Share ACADEMY

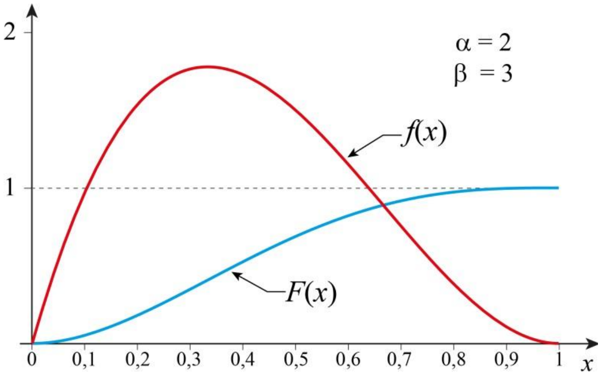
# Statystyka

## Rozkład beta

Zmienna losowa $X$, ma rozkład beta o parametrach $\alpha > 0$ , $\beta > 0$ jeśli jej gęstość prawdopodobieństwa opisana jest wzorem:

$$f(x) = \begin{cases} 0 & dla \quad x \notin (0,1) \\ \frac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} & dla \quad x \in (0,1) \end{cases}$$

gdzie $B(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$




infoShareAcademy.com

infoShare ACADEMY

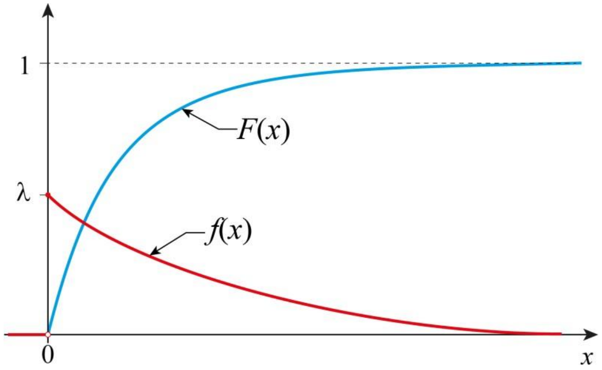
# Statystyka

## Rozkład wykładniczy

Zmienna losowa $X$, ma rozkład wykładniczy o parametrze λ

jeśli jej gęstość prawdopodobieństwa opisana jest wzorem:

$$f(x) = \begin{cases} 0 & dla \quad x < 0 \\ \lambda \cdot e^{-\lambda \cdot x} & dla \quad x > 0 \end{cases}$$




infoShareAcademy.com

info Share ACADEMY

# Statystyka

## Podsumowanie

infoShareAcademy.com

infoShare ACADEMY