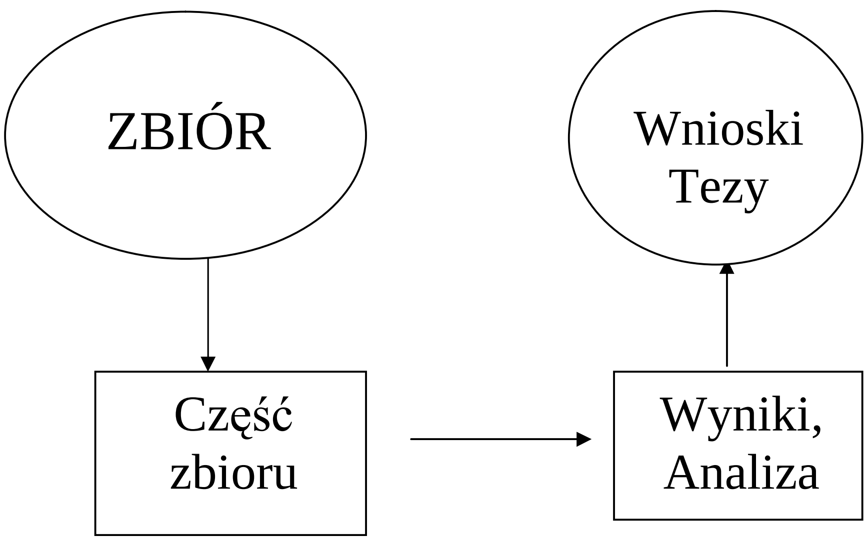
STATYSTYKA MATEMATYCZNA - dział matematyki stosowanej oparty na rachunku prawdopodobieństwa; zajmuje się badaniem zbiorów na podstawie analizy ich części. Nauka, której przedmiotem zainteresowania są metody pozyskiwania i prezentacji, a przede wszystkim analizy danych opisujących zjawisk

(*C. Gauss, P. Laplace, Bernoulli,* R. Fisher, J. Spława-Neyman)

Wikipedia 2008



Zbiorem (populacją) jest interesujący nas typ obiektu i jego właściwości, które chcemy poznać. Częścią zbioru jest próba, którą obserwujemy podczas badania.

Statystyka jest stosowana w wielu dziedzinach wiedzy, w niektórych doczekała się własnej terminologii i wyspecjalizowanych metod. Wytworzyły się dziedziny z pogranicza statystyki i innych nauk.

Należą do nich:

- Biometria
- Demografia
- Ekonometria
- Fizyka statystyczna
- Psychologia statystyczna
- Socjologia statystyczna
- Statystyka gospodarcza

Obecnie standardem w naukach eksperymentalnych jest potwierdzanie istnienia obserwowanych zależności przy pomocy metod statystyki matematycznej. Pozwala to odróżnić rzeczywiste wyniki od przypadkowej zbieżności – odróżnienie efektów działających praw od efektu przypadku (prawidłowości w zakresie zjawisk masowych)

Od starożytności do około połowy XIX wieku termin *statystyka* oznaczał *podany w tabelarycznej formie zbiór danych na temat stanu państwa*. W pewnym momencie posiadanie podstawowych danych stało się niewystarczające, szczególnie przy coraz szybciej rozwijającej się gospodarce światowej.

Konieczne stało się nie tylko ulepszanie metod pozyskiwania danych, ale również ich opisu i analizy. Zbiegło się to w czasie z szybkim rozwojem metod matematycznych, szczególnie teorii prawdopodobieństwa.

Obecnie (od połowy XX wieku) rozwój statystyki jest związany z możliwością wykorzystania komputerów do przeprowadzana skomplikowanych (lub żmudnych) obliczeń.

ELEMENTY RACHUNKU PRAWDOPODOBIEŃSTWA

Rachunek prawdopodobieństwa – analiza praw rządzących zdarzeniami losowymi.

Pojęcia pierwotne – zdarzenie elementarne ω oraz zbiór zdarzeń elementarnych Ω.

Zdarzenie losowe A – podzbiór zbioru zdarzeń elementarnych Ω.

Prawdopodobieństwo (klasyczna definicja Laplace'a): Jeżeli Ω składa się z n jednakowo prawdopodobnych zdarzeń elementarnych, to prawdopodobieństwo zdarzenia A składającego się z k zdarzeń elementarnych wyraża się wzorem:

$$P(A) = \frac{k}{n}$$

Prawdopodobieństwo (aksjomatyczna definicja Kołmogorowa):

1. $P(A) \in \langle 0, 1 \rangle$
2. $P(\Omega) = 1$
3. $P(A \cup B) = P(A) + P(B)$ dla A i B rozłącznych

Punkt 3 jest prawdziwy również dla nieskończonej liczby zbiorów rozłącznych

Prawdopodobieństwo zbioru pustego

$$P(\varnothing)=0$$

Przykład: rzucamy kostką idealną – zawsze coś wypadnie i wszystkie możliwości są równie prawdopodobne. Prawdopodobieństwo, że nic nie wypadnie jest równe zeru

Prawdopodobieństwo zdarzenia A' (przeciwnego do zdarzenia *A*):

$$P(A') = 1 - P(A)$$

Przykład: rzucamy kostką. A-wypadnie „6”.
Wtedy A’-nie wypadło „6”.
P(A)=1/6 więc P(A’)=1–1/6=5/6

Zdarzenia rozłączne. Zdarzenia A i B są rozłączne, jeżeli

$$(A \cap B) = \emptyset$$

Wtedy

$$P(A \cap B) = 0$$

A i A' są rozłączne. Wypadnie „1" i wypadnie „6" na kostce też są rozłączne.
C- wypadnie parzysta liczba i D- wypadnie mniej niż 4 nie są rozłączne, bo wynik 2 należy zarówno do zdarzenia C jak i D.

Prawdopodobieństwo **sumy** i **iloczynu** zbiorów:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

przykład: $C \cup D=\{1,2,3,4,6\}$ $C \cap D=\{2\}$
5/6=3/6+3/6–1/6

Prawdopodobieństwo warunkowe zajścia zdarzenia *B* pod warunkiem realizacji zdarzenia *A*:

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

przykład: A-wypadnie parzysta liczba oczek, B-wypadnie conajwyżej 3. P(A)=3/6, P(B)=3/6) ale P(B/A)=1/3 – wiemy że wypadło jedno z A={2,4,6} więc P({2})=1/3. Korzystając ze wzoru: $A \cap B$={2}, więc P($A \cap B$)=1/6, P(A)=3/6 => P(B/A)=$^{1/6}/_{3/6}$

Zdarzenia niezależne: Zdarzenia *A* i *B* są niezależne, jeżeli

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

Równoważnie

$$\mathrm{P}(B/A) = \mathrm{P}(B)$$

$$\mathrm{P}(A/B) = \mathrm{P}(A)$$

przykład: wypadnięcie orła w rzucie monetą i wypadnięcie parzystej liczby oczek w rzucie kostką są niezależne. A- parzysta liczba, B- conajwyżej 4 też są niezależne.

Schemat Bernoulliego

- ciąg *n* niezależnych zdarzeń,
- w pojedynczym zdarzeniu realizuje się sukces lub porażka,
- prawdopodobieństwo sukcesu jest stałe, równe *p*

$$P(k\ sukcesów) = \binom{n}{k} p^k (1-p)^{n-k}$$

q = 1-p

k = 0, 1, 2, ,n

Przykład: Prawdopodobieństwo spełnienia normy dla klasy I przez pewien produkt wynosi 70%. Jakie jest prawdopodobieństwo, że przy kontroli jakości na 5 losowo wybranych produktów normę spełni przynajmniej 4 z nich? (53%)

$$P_{n=5}(k=4) = \frac{5!}{4!*1!} 0{,}7^4 0{,}3^1 \approx 36\%$$

$$P_{n=5}(k=5) = \frac{5!}{5!*0!} 0{,}7^5 0{,}3^0 \approx 17\%$$

ZMIENNA LOSOWA (CECHA)

Zmienna losowa: funkcja o wartościach rzeczywistych, określona na zbiorze zdarzeń elementarnych Ω. Innymi słowy zmienna losowa X jest liczbową prezentacją wyniku doświadczenia losowego.

$$\Omega \longrightarrow R$$

Skala nominalna ang: *nominal scale* (cecha jakościowa): wartości liczbowe stanowią tylko etykiety. Nie wynika z nich żadna zależność (porządek, różnice, ilorazy nie mają merytorycznego znaczenia) np. rodzaj przetworu mlecznego, ID osoby. Skala nominalna dychotomiczna – możliwe są tylko dwie wartości.

Skala porządkowa ang: *Ordinal scale*, (przypisanie rangi każdej jednostce populacji): zmienne przyjmują wartości, dla których dane jest uporządkowanie (kolejność), ale ani różnica ani iloraz między dwiema wartościami nie mają merytorycznego znaczenia (np. klasa jakościowa, wykształcenie).

Skala interwałowa ang: *Interval scale* różnice między dwiema jej wartościami mają interpretację, jednak dzielenie dwóch wartości (iloraz) nie ma merytorycznego sensu (określona jest jednostka, jednak punkt zero jest wybrany umownie np. miary temperatury).

Skala absolutna Wymaga ustalenia jednostki miary i wyznaczenia naturalnego punktu zerowego np. liczba produktów.

Skala ilorazowa ang: *Ratio measurement*: merytorycznie uzasadnione jest położenie zera, zaś jednostka jest umowna (nie wynika z natury zjawiska). Nie wprowadza to ograniczeń w stosowaniu operacji arytmetycznych do wyników pomiaru (np. średnia geometryczna i harmoniczna). przykładem jest temperatura w stopniach Kelwina.

Zmienna jakościowa (skala nominalna): przyjmuje wartości nie będące liczbami (np. kolor, płeć, smakowitość).

Zmienna ilościowa skokowa (interwałowa, ilorazowa lub absolutna): przyjmuje pewne wartości liczbowe i nie przyjmuje wartości pośrednich (np. liczba bakterii, liczba kolonii drożdży, liczba pracowników). Nazywana jest również zmienną **dyskretną**.

Zmienna ilościowa ciągła (interwałowa, ilorazowa lub absolutna): przyjmuje wartości z pewnego przedziału liczbowego (np. wzrost, plon, kaloryczność).

zmienna losowa

zm. jakościowa **zm. ilościowa**

zm. skokowa **zm. ciągła**

Rozkład zmiennej losowej skokowej *X*: zbiór wartości zmiennej losowej x_i oraz prawdopodobieństwa p_i z jakimi te wartości są przyjmowane (dla każdej możliwej x_i z X, odpowiednio w_i z W, ma pewne p_i realizacji).

$$p_i = \mathrm{P}(X = x_i), \qquad p_i \in <0,1>, \quad \sum_i p_i = 1$$

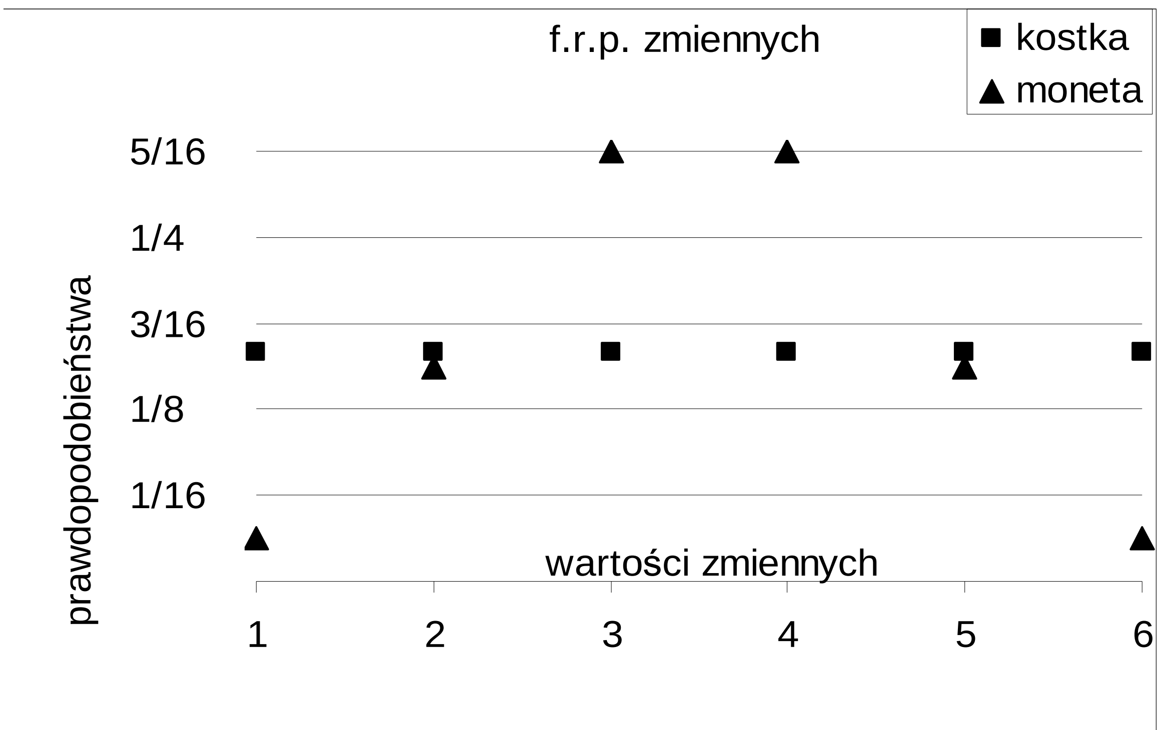
Przykład. Jednokrotny rzut kostką. Zmienna losowa X to liczba wyrzuconych oczek.

Rozkład: liczba oczek w rzucie kostką

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Rozkład: jeden plus liczba orłów w 5 rzutach

x_i	1	2	3	4	5	6
p_i	1/32	5/32	10/32	10/32	5/32	1/32



Rozkład zmiennej losowej ciągłej *X* określony jest funkcją gęstości prawdopodobieństwa f(x). Funkcja ta posiada własności:

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- $P(a < x < b) = \int_{a}^{b} f(x)dx$

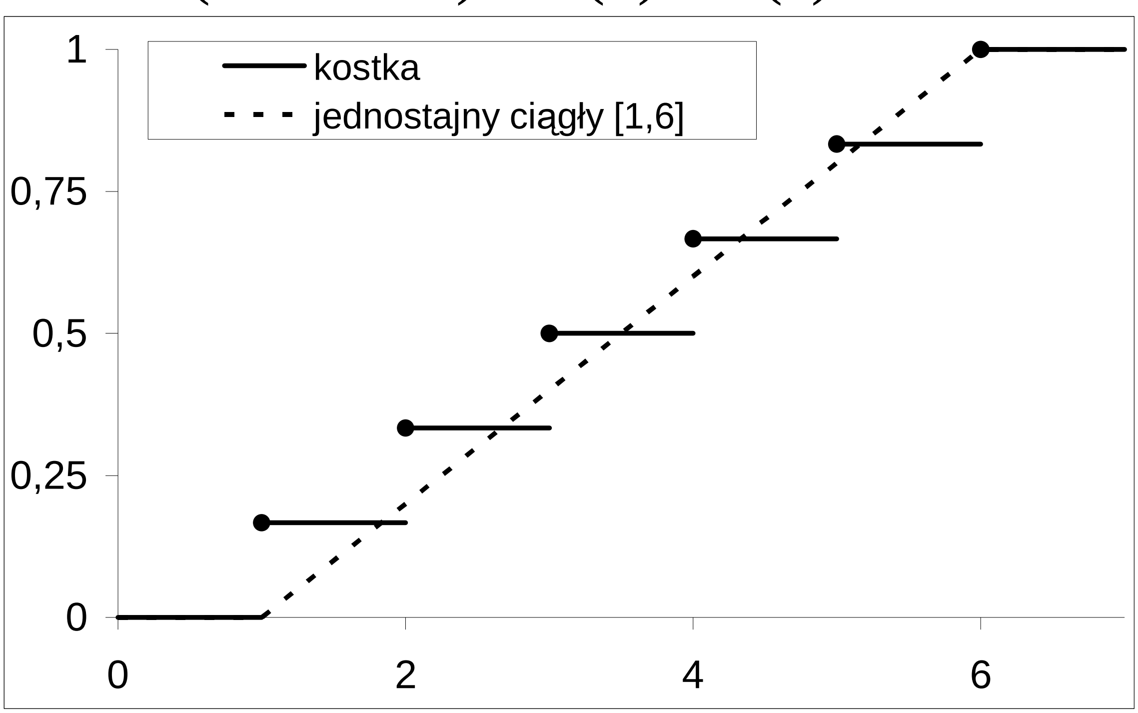
Zdefiniowane prawdopodobieństwo realizacji jakiejś wartości z dowolnie wybranego przedziału, np. dla rozkładu jednostajnego ciągłego (równomiernego ciągłego) na [0,1] prawdopodobieństwo realizacji wartości z przedziału (będącego podzbiorem [0,1]) równe jest długości przedziału, więc P([0; 0,2])=0,2.

Rozkład zmiennej losowej, obydwu typów, można przedstawić przy pomocy funkcji dystrybuanty F(x). Dystrybuanta F jest funkcją określoną na zbiorze liczb rzeczywistych R wzorem:

$$F(x_0) = P(X \le x_0) = \begin{cases} \sum\limits_{x_i}^{x_i \le x_0} P(x = x_i) & \textit{dla zmiennej skokowej} \\ \int\limits_{-\infty}^{x_0} f(x)dx & \textit{dla zmiennej ciaglej} \end{cases}$$

Funkcja dystrybuanty posiada własności:

- $0 \le F(x) \le 1$
- $F(-\infty) = 0$, $F(\infty) = 1$
- Jest funkcją niemalejącą
- $P(a < X \le b) = F(b) - F(a)$



Parametry zmiennej losowej

Wartość oczekiwana (średnia) EX zmiennej losowej X jest liczbą charakteryzującą położenie zbioru jej wartości:

$$EX = \begin{cases} \sum x_i p_i & \textit{dla zmiennej skokowej} \\ \int x\ f(x)\ dx & \textit{dla zmiennej ciaglej} \end{cases}$$

Wariancja D^2X i odchylenie standardowe DX zmiennej losowej X są liczbami charakteryzującymi rozrzut zbioru jej wartości wokół wartości średniej EX:

$$D^2X = \begin{cases} \sum (x_i - EX)^2\ p_i & \textit{dla zmiennej skokowej} \\ \int (x - EX)^2\ f(x)\ dx & \textit{dla zmiennej ciaglej} \end{cases}$$

$$DX = \sqrt{D^2X}$$

przykład: dla rzutu kostką
EX$=1 * {}^1/_6 + 2 * {}^1/_6 + 3 * {}^1/_6 + 4 * {}^1/_6 + 5 * {}^1/_6 + 6 * {}^1/_6 = 3{,}5$; **$D^2$X** $= (1–3{,}5)^2 * {}^1/_6 + (2–3{,}5)^2 * {}^1/_6 + (3–3{,}5)^2 * {}^1/_6 + (4–3{,}5)^2 * {}^1/_6 +$
$+ (5–3{,}5)^2 * {}^1/_6 + (6–3{,}5)^2 * {}^1/_6 = 2 + {}^{11}/_{12}$