

**Poziom istotności** - jest to prawdopodobieństwo popełnienia błędu I rodzaju (zazwyczaj oznaczane symbolem  $\alpha$ ). Określa również maksymalne ryzyko błędu, jakie badacz jest skłonny zaakceptować. Wybór wartości  $\alpha$  zależy od badacza, natury problemu i od tego jak dokładnie chce on weryfikować swoje hipotezy, najczęściej przyjmuje się  $\alpha = 0,05, 0,03$  lub  $0,01$ .

**Poziom ufności** – w metrologii i statystyce: prawdopodobieństwo  $1-\alpha$  związane z przedziałem ufności. Poziom ufności bywa często wyrażany w procentach.

Wartość poziomu ufności jest jednym ze składowych wyniku pomiaru umieszczanym na świadectwie wzorcowania. Wynik pomiaru podaje się zwykle dla poziomu ufności  $p=95\%$ . Oznacza to 95-procentowe prawdopodobieństwo, że wynik pomiaru zawiera się w przedziale domkniętym ograniczonym niepewnością rozszerzoną pomiaru.

**Błąd pierwszego rodzaju** ('false positive') - w statystyce pojęcie z zakresu weryfikacji hipotez statystycznych - błąd polegający na odrzuceniu hipotezy zerowej, która w rzeczywistości jest prawdziwa. Błąd pierwszego rodzaju znany też jest jako: *błąd pierwszego typu*, *błąd przyjęcia* lub *alfa-błąd*. Oszacowanie prawdopodobieństwa popełnienia błędu pierwszego rodzaju oznaczamy symbolem  $\alpha$  (mała grecka litera alfa) i nazywamy *poziomem istotności testu*.

**Błąd drugiego rodzaju** (*błąd drugiego typu*, *błąd przyjęcia*, *beta-błąd*) - w statystyce pojęcie z zakresu weryfikacji hipotez statystycznych - polegające na **nieodrzućeniu** hipotezy zerowej, która jest w rzeczywistości **falszywa**. Oszacowanie prawdopodobieństwo popełnienia błędu drugiego rodzaju oznaczamy symbolem  $\beta$  (mała grecka litera beta), a jego dopełnienie do jedności nazywane jest mocą testu.

**Standaryzacja** jest w statystyce rodzajem normalizacji zmiennej losowej, w wyniku której zmienna uzyskuje średnią wartość oczekiwaną zero i wariancję jeden. Na ogół odejmuje się od zmiennej średnią z jej próby i dzieli wynik przez odchylenie standardowe z próby. Bardziej złożone metody standaryzacji zmieniają dodatkowo rozkład zmiennej na normalny.

**Wartość oczekiwana (przeciętna, średnia), nadzieja matematyczna** – w rachunku prawdopodobieństwa wartość opisująca spodziewany (średnio) wynik doświadczenia losowego. Wartość oczekiwana to inaczej pierwszy moment zwykły. Estymatorem wartości oczekiwanej rozkładu cechy w populacji jest średnia arytmetyczna.

Niech  $X$  będzie zmienną losową typu dyskretnego. **Wartością oczekiwaną** nazywa się sumę iloczynów wartości tej zmiennej losowej oraz prawdopodobieństw, z jakimi są one przyjmowane.

Formalnie, jeżeli dyskretna zmienna losowa  $X$  przyjmuje wartości  $x_1, x_2, \dots, x_n$  z prawdopodobieństwami wynoszącymi odpowiednio  $p_1, p_2, \dots, p_n$ , to wartość oczekiwana  $\mathbb{E}X$  zmiennej losowej  $X$  wyraża się wzorem

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

Jeżeli zmienna  $X$  przyjmuje przeliczalnie wiele wartości, to wzór na jej wartość oczekiwaną ma  $\infty$  w miejsce  $n$  (istnieje ona tylko wtedy, gdy szereg jest zbieżny).

Jeżeli  $X$  jest zmienną losową typu ciągłego zdefiniowaną na przestrzeni probabilistycznej  $(\Omega, \mathcal{F}, \mathbb{P})$ , to **wartość oczekiwaną** zmiennej losowej  $X$  definiuje się jako całkę

$$\mathbb{E}X = \int_{\Omega} X d\mathbb{P}$$

o ile powyższa całka istnieje, czyli jest skończona:

$$\mathbb{E}|X| = \int_{\Omega} |X| d\mathbb{P} < +\infty$$

Dowodzi się, że jeśli  $X$  jest zmienną losową o funkcji gęstości prawdopodobieństwa  $f(x)$ , to jej wartość oczekiwana wynosi

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

Jeżeli  $Y = \varphi(X)$  jest funkcją mierzalną, to

$$\mathbb{E}Y = \mathbb{E}(\varphi(X)) = \int_{\mathbb{R}} \varphi(x) f(x) dx$$

Jeśli istnieją  $\mathbb{E}X$  oraz  $\mathbb{E}Y$ , to:

- $\mathbb{E}c = c$ , gdzie  $c$  jest funkcją stałą (wynika z jednorodności sumy/szeregu/całki),
- $\forall_{a,b} \mathbb{E}(aX + b) = a\mathbb{E}X + b$  (wynika z liniowości sumy/szeregu/całki),
- jeżeli  $X, Y$  są niezależne, to  $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$ ,
- jeżeli  $X \geq 0$  prawie wszędzie, to  $\mathbb{E}X \geq 0$ ,

$$\bullet \quad \mathbb{E}|X| \geq |\mathbb{E}X|$$

**Wariancja** to w statystyce klasyczna miara zmienności. Intuicyjnie utożsamiana ze zróżnicowaniem zbiorowości jest średnią arytmetyczną kwadratów odchyłeń (różnic) poszczególnych wartości cechy od wartości oczekiwanej.

Wariancja zmiennej losowej  $X$  zdefiniowana jest wzorem:

$$\text{Var}[X] = \mathbb{E}[(X - \mu)^2]$$

gdzie  $\mathbb{E}[\cdot]$  jest wartością oczekiwaną zmiennej losowej. Innym, często prostszym sposobem wyznaczania wariancji jest wzór:  $D^2(X) = E(X^2) - [E(X)]^2$ .

Wariancja jest momentem centralnym drugiego rzędu zmiennej losowej.

Wariancję dla szeregu szczegółowego wyznacza się ze wzoru:

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

a dla szeregu rozdzielczego:

$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i \cdot (x_i - m)^2$$

Wariancja próby losowej o wartościach  $x_i$ , gdzie  $i = 1, 2, 3, \dots$ , jest następująca:

$$\sigma^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Wariancję dla populacji można estymować za pomocą n-elementowej próby losowej. Estymator największej wiarygodności:

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

jest estymatorem obciążonym wariancji. Innymi słowy, gdybyśmy z populacji losowali próbkę wielokrotnie i obliczali jego wyniki, to ich średnia nie byłaby równa wariancji w całej populacji. Dlatego też częściej używa się nieobciążonego estymatora:

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

**Gęstość prawdopodobieństwa** - w statystyce nazywa się w ten sposób funkcję rozkładu prawdopodobieństwa ciągłej zmiennej losowej. Całka oznaczona z takiej funkcji równa jest prawdopodobieństwu wystąpienia zdarzenia zawartego w granicach całkowania.

$$P(x \in [a, b]) = \int_a^b \rho(x) dx$$

Estymator jest **nieobciążony**, jeśli wartość oczekiwana rozkładu estymatora jest równa wartości szacowanego parametru:

$$E(T_n) = \theta$$

Jeśli różnica pomiędzy wartością oczekiwaną rozkładu estymatora a wartością szacowanego parametru jest zależna funkcyjnie od estymatora:

$$E(T_n) - \theta = b(T_n)$$

to estymator nazywamy **obciążonym**, zaś samą różnicę nazywamy **obciążeniem estymatora**.

Estymator nazywamy **zgodnym**, jeśli jest stochastycznie zbieżny do szacowanego parametru:

$$\lim_{n \rightarrow \infty} P\{|T_n - \theta| < \epsilon\} = 1$$

Oznacza to, że jeśli rośnie liczebność próby, rośnie też prawdopodobieństwo, że oszacowanie przy pomocy estymatora będzie przyjmować wartości coraz bliższe wartości szacowanego parametru. Inaczej: zwiększając liczebność próby, zmniejszamy ryzyko popełnienia błędu.

**Funkcja rozkładu** – funkcja podająca prawdopodobieństwo lub gęstość prawdopodobieństwa wystąpienia danej wartości  $X$ .

- ciągła funkcja rozkładu  $f(x)$  – gęstość prawdopodobieństwa:

$$f(X) = \frac{dF(X)}{dX} \quad \int_{X_{min}}^{X_{max}} f(X) dX = 1$$

gdzie:

$$0 \leq F(x) \leq 1 \text{ – ciągła dystrybucja prawdopodobieństwa}$$

- dyskretna funkcja rozkładu  $P(X_i)$  – prawdopodobieństwo dla  $n$  możliwych wartości  $X_i$ :

$$\sum_{i=1}^n P_i = 1$$

$$P_i = P(X_i) \text{ gdzie } i=1$$

**Dystrybuanta** – w rachunku prawdopodobieństwa, statystyce i dziedzinach pokrewnych - funkcja rzeczywista, która jednoznacznie wyznacza rozkład prawdopodobieństwa (tj. miarę probabilistyczną określoną na  $\sigma$ -ciele borelowskich podzbiorów prostej<sup>[1]</sup>), a więc zawiera o nim wszystkie informacje. Dystrybuanty są efektywnym narzędziem badania prawdopodobieństwa ponieważ, z matematycznego punktu widzenia, są obiektem *prostszy* niż rozkłady prawdopodobieństwa.

Niech  $P$  będzie rozkładem prawdopodobieństwa na prostej. Funkcję  $F: \mathbb{R} \rightarrow \mathbb{R}$  daną wzorem

$$F(x) = P((-\infty, x])$$

nazywamy **dystrybuantą** (rozkładu  $P$ ).

Funkcja  $F: \mathbb{R} \rightarrow \mathbb{R}$  jest dystrybuantą wtedy i tylko wtedy, gdy jest ona

1. prawostronnie ciągła,
2. niemalejąca,
3.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$ .

**Uwaga 1:** Ponieważ powyższe twierdzenie podaje warunek konieczny i wystarczający na to by funkcja była dystrybuantą, więc czasami to właśnie je przyjmuje się jako definicję. Podejście takie może być korzystniejsze, gdyż nie trzeba odwoływać się do pojęcia rozkładu, pochodzącego z teorii miary. Ponieważ przyjęliśmy definicję używaną częściej w praktyce, więc zdanie "funkcja  $F$  jest dystrybuantą wtedy i tylko wtedy" zawiera ciche założenie, że istnieje rozkład, którego ta funkcja jest dystrybuantą.

**Uwaga 2:** W starszej rosyjskiej literaturze dystrybuantę definiuje się jako funkcję **lewostronnie ciągłą** spełniającą warunki 2. i 3. Czytelnik powinien zawsze upewnić się jaką definicję przyjmuje autor książki. Różnica ta jest istotna przy rozważaniu rozkładów dyskretnych, ponieważ w ich przypadku zbiory jednoelementowe **nie** muszą być miary zero.

**Uwaga 3:** Dystrybuanta  $F$  wyznacza pewien rozkład  $P$  jednoznacznie i na odwrót, więc gdy zachodzi potrzeba całkowania pewnej funkcji borelowskiej  $g$  względem rozkładu  $P$ , to możemy mówić, że całkujemy ją względem dystrybuanty  $F$ , co zapisujemy:

$$\int g dP = \int g dF$$

**Dyskretny rozkład prawdopodobieństwa** to w probabilistyce rozkład prawdopodobieństwa zmiennej losowej dający się opisać przez podanie wszystkich przyjmowanych przez nią wartości, wraz z prawdopodobieństwem przyjęcia każdej z nich. Funkcja przypisująca prawdopodobieństwo do konkretnej wartości zmiennej losowej jest nazywana **funkcją rozkładu prawdopodobieństwa** (*probability mass function*, pmf). Zachodzi:

$$\sum_u \Pr(X = u) = 1$$

gdzie  $u$  przebiega zbiór możliwych wartości zmiennej  $X$

Jeśli zmienna losowa jest dyskretna, wówczas zbiór wszystkich wartości, które przyjmuje z niezerowym prawdopodobieństwem jest skończony lub przeliczalny, gdyż suma nieprzeliczalnie wielu dodatnich liczb rzeczywistych jest zawsze rozbieżna do nieskończoności.

Zwykle ten zbiór przyjmowanych wartości jest topologicznie zbiorem izolowanych punktów. Istnieją jednak zmienne dyskretne, dla których zbiór przyjmowanych wartości jest zbiór gęsty.

Równoważnie dyskretną zmienną losową można zdefiniować jako zmienną losową, której dystrybuanta jest funkcją schodkową:

$$\sum_{x_k \in S} \left[ F(x_k) - \lim_{x \rightarrow x_k^-} F(x) \right] = 1$$

Rozkład Poissona, Bernoullego, rozkład dwupunktowy, rozkład geometryczny są najbardziej znanymi rozkładami dyskretnymi.

**Rozkład Poissona** (czyt. *plasona*) – rozkład dyskretny przedstawiający liczbę wystąpień zjawiska w czasie  $t$ , w określonej liczbie prób, jeśli wystąpienia te są niezależne od siebie. Rozkład ma zastosowanie do obliczenia przybliżonej wartości prawdopodobieństwa w rozkładzie dwumianowym przy dużej liczbie prób i niskim prawdopodobieństwie sukcesu.

Rozkład Poissona jest określany przez jeden parametr  $\lambda$ , który ma interpretację wartości oczekiwanej. Parametr ten jest równy prawdopodobieństwu uzyskania sukcesu w pojedynczej próbie pomnożony przez liczbę prób.

Własności rozkładu Poissona:

$$\frac{e^{-\lambda} \lambda^k}{k!},$$

- prawdopodobieństwo:  $\frac{e^{-\lambda} \lambda^k}{k!}$ ,
- wartość oczekiwana:  $\lambda$ ,
- wariancja:  $\lambda$ ,
- współczynnik skośności:  $\lambda^{-1/2}$ ,
- kurtოza:  $\lambda^{-1}$ ,
- funkcja charakterystyczna:  $\phi_X(t) = e^{\lambda(e^{it} - 1)}$

Zmienna losowa  $X$  ma rozkład Poissona, jeśli:

$$\mathbb{P}(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots; \quad \lambda > 0$$

**Próba Bernoullego** - eksperyment losowy z dwoma możliwymi wynikami, określanymi zazwyczaj jako *sukces* oraz *porażka*. Za przykłady prób Bernoullego matematycy uważają:

- pojedynczy rzut monetą (rezultat: orzeł lub reszka)
- narodziny dziecka (rezultat: chłopiec lub dziewczynka).

Matematycznie próbę taką modeluje się przy pomocy zmiennej losowej o **rozkładzie Bernoulliego**, czasem nazywanym **rozkładem dwupunktowym**. Zmienna taka przyjmuje tylko dwie wartości:

- 1, interpretowane jako *sukces*, z prawdopodobieństwem  $p$
- 0, interpretowane jako *porażka*, z prawdopodobieństwem  $1 - p$ .

Niech  $X$  będzie zmienną losową o rozkładzie Bernoulliego:

- rozkład prawdopodobieństwa:

$$\mathbb{P}(X = 1) = p,$$

$$\mathbb{P}(X = 0) = 1 - p$$

- wartość oczekiwana:  $\mathbb{E}X = p$
- wariancja:  $\text{Var } X = p(1 - p)$ .

Ciąg niezależnych prób Bernoulliego nazywa się procesem Bernoulliego.

**Rozkład dwupunktowy** - rozkład dyskretny prawdopodobieństwa w którym zmienna losowa przyjmuje tylko dwie wartości. Jest on na przykład rezultatem doświadczenia (zwanego próbą Bernoulliego), w wyniku którego określone zdarzenie  $A$  wystąpi lub nie wystąpi. Na ogół zdarzeniom elementarnym wchodzącym w skład zdarzenia  $A$  przyporządkowana jest wartość 1 zmiennej losowej, a innym zdarzeniom elementarnym liczba 0.

Wówczas jeżeli

$$P(A) = p$$

to

$$P(\bar{A}) = 1 - p = q$$

gdzie  $\bar{A}$  oznacza zdarzenie przeciwne, oraz

$$P(X = 1) = p$$

$$P(X = 0) = q$$

**Rozkład geometryczny** jest dyskretnym rozkładem prawdopodobieństwa opisującym prawdopodobieństwo zdarzenia, że proces Bernoulliego odniesie pierwszy sukces dokładnie w  $n$ -tej próbie.

W szczególności, prawdopodobieństwo, że zmienna losowa  $X$  o rozkładzie geometrycznym przyjmie wartość  $n$  dane jest wzorem:

$$P(X = n) = p(1 - p)^{n-1} \quad n = 1, 2, \dots$$

$$\mathbb{E}X = \frac{1}{p}$$

Wartość oczekiwana:

$$D^2 X = \frac{1 - p}{p^2}$$

Wariancja:

$$\phi_X(t) = \frac{pe^{it}}{1 - (1 - p)e^{it}}$$

Funkcja charakterystyczna:

**Rozkład geometryczny** to szczególny przypadek ujemnego rozkładu dwumianowego dla  $r = 1$ .

Ciągłym odpowiednikiem rozkładu geometrycznego jest rozkład wykładniczy. Ma własność braku pamięci.

**Rozkład wykładniczy** to rozkład zmiennej losowej opisujący sytuację, w której obiekt może przyjmować stany  $X$  i  $Y$ , przy czym obiekt w stanie  $X$  może ze stałym prawdopodobieństwem przejść w stan  $Y$  w jednostce czasu. Prawdopodobieństwo wyznaczone przez ten rozkład to prawdopodobieństwo przejścia ze stanu  $X$  w stan  $Y$  w czasie  $\delta t$ .

Dystrybucja tego rozkładu to prawdopodobieństwo, że obiekt jest w stanie  $Y$ .

Innymi słowy, jeżeli w jednostce czasu ma zajść  $1/\lambda$  niezależnych zdarzeń, to *rozkład wykładniczy* opisuje odstępy czasu pomiędzy kolejnymi zdarzeniami.

Jest on określony jednym parametrem  $\lambda$ .

Własności rozkładu wykładniczego:

- gęstość prawdopodobieństwa:

$$p(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0, \\ 0 & , x < 0. \end{cases}$$

- dystrybucja:

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & , x \geq 0, \\ 0 & , x < 0. \end{cases}$$

- mediana:  $\frac{\ln(2)}{\lambda}$
- wartość oczekiwana:  $\frac{1}{\lambda}$
- wariancja:  $\frac{1}{\lambda^2}$

**Ciągły rozkład prawdopodobieństwa** - rozkład prawdopodobieństwa dla którego dystrybuenta jest funkcją ciągłą. Równoważnie można powiedzieć, że zmienna losowa  $X$  posiadająca taki rozkład ma  $Pr[X = a] = 0$  dla wszystkich  $a \in \mathbb{R}$ . O ile w przypadku rozkładu dyskretnego zdarzenie o zerowym prawdopodobieństwie jest niemożliwe, o tyle dla rozkładów ciągłych nie jest to prawda, gdyż wówczas żadna wartość nie byłaby możliwa. Ten paradoks można zrozumieć, uświadomiwszy sobie przez analogię, że choć pojedynczy punkt ma zerowe rozmiary, jednak odcinek złożony z nieskończonej liczby takich punktów ma już niezerową długość.

**Centralne twierdzenie graniczne** to twierdzenie matematyczne mówiące, że jeśli  $X_i$  są niezależnymi zmiennymi losowymi o jednakowym rozkładzie, takiej samej wartości oczekiwanej  $\mu$  i skończonej wariancji  $\sigma^2$ , to zmienna losowa o postaci

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$$

zbiega według rozkładu do standardowego rozkładu normalnego gdy  $n$  rośnie do nieskończoności.

**Centralne twierdzenie graniczne** znane też pod nazwą **Twierdzenie Lindeberga-Lévy'ego** orzeka:

$$\sum_{k=1}^n D^2 X_{n,k} = 1$$

Niech  $(X_{n,k})$  będzie schematem serii, w którym  $EX_{n,k} = 0$  dla  $k \leq n$  dla każdego  $n$  mamy  $\sum_{k=1}^n D^2 X_{n,k} = 1$ . Jeśli spełniony

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n EX_{n,k}^2 \mathbf{1}_{\{|X_{n,k}| > \epsilon\}} = 0$$

jest warunek Lindeberga, tj. dla każdego  $\epsilon > 0$  zachodzi

, wtedy

$$\sum_{k=1}^n X_{n,k} \xrightarrow{D} N(0, 1)$$

**Mediana** (zwana też **wartością środkową** lub **drugim kwartylem**) to w statystyce wartość cechy w szeregu uporządkowanym, powyżej i poniżej której znajduje się jednakowa liczba obserwacji. Mediana jest kwantylem rzędu 1/2.

Aby obliczyć medianę ze zbioru  $n$  obserwacji, sortujemy je w kolejności od najmniejszej do największej i numerujemy od 1 do  $n$ . Następnie, jeśli  $n$  jest nieparzyste, medianą jest wartość obserwacji w środku (czyli obserwacji numer  $(n+1)/2$ ). Jeśli natomiast  $n$  jest parzyste, wynikiem jest średnia arytmetyczna między dwiema środkowymi obserwacjami, czyli obserwacją numer  $n/2$  i obserwacją numer  $(n/2)+1$ .

Niekiedy używane są też inne wersje mediany:

- Wersja w której dla parzystego  $n$  zamiast średniej arytmetycznej losuje się jedną z dwóch obserwacji: numer  $n/2$  lub  $n/2+1$ . Taka mediana nie wyprowadza wyniku poza zbiór dotychczasowych wartości. Znajduje zastosowanie szczególnie przy obróbce dwubarwnych map bitowych. Klasyczna mediana wymagałaby wówczas wprowadzenia obok istniejących kolorów białego i czarnego także koloru szarego.
- **Mediana ważona** w której każda obserwacja  $a_i$  ma przypisaną wagę  $w_i$ . Jeśli  $w_i$  są liczbami naturalnymi, jej obliczenie sprowadza się do obliczenia klasycznej mediany, w której obserwacja  $a_i$  jest wzięta pod uwagę  $w_i$  razy.

Mediana znalazła szerokie zastosowanie w statystyce jako średnia znacznie bardziej odporna na elementy odstające niż średnia arytmetyczna. Używana jest także w grafice komputerowej i przetwarzaniu dźwięku w celu odszumiania - na obrazie zachowuje ona ostre krawędzie przy jednoczesnym usunięciu szumów.

Odporność na elementy odstające jest na ogół zaletą, jednak czasem może być uważane za wadę — nawet olbrzymie zmiany skrajnych obserwacji nie wpływają na jej wartość. Stąd pojawiły się propozycje pośrednie pomiędzy nimi, takie jak średnia ucinana, stosowana na przykład w konkursach tańca na lodzie.

**Prawdopodobieństwo** to funkcja  $P(X)$ , która przyporządkowuje każdemu elementowi zbioru zdarzeń losowych pewną nieujemną wartość rzeczywistą i ma następujące własności:

- $P(\Omega) = 1$ , gdzie  $\Omega$  jest przestrzenią zdarzeń elementarnych
- prawdopodobieństwo sumy przeliczalnego zbioru zdarzeń parami rozłącznych jest równe sumie prawdopodobieństw tych zdarzeń:

$$P(A_1 \cup \dots \cup A_n \cup \dots) = P(A_1) + \dots + P(A_n) + \dots$$

Wartość  $P(X)$  nazywa się **prawdopodobieństwem zdarzenia  $X$** .

Ważniejsze własności prawdopodobieństwa:

- $P(A) \geq 0$
- $P(\emptyset) = 0$  (**UWAGA:** odwrotna implikacja nie jest prawdziwa -  $P(A)=0$  nie implikuje  $A=\emptyset$ )
- $A \subset B \Rightarrow P(A) \leq P(B)$

- $P(A) \leq 1$
- $A \subset B \Rightarrow P(B|A) = 1$
- $P(A) + P(A') = 1$ , gdzie  $A'$  oznacza zdarzenie losowe przeciwne do  $A$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

Sposób liczenia prawdopodobieństwa z poprzedniego przykładu podał po raz pierwszy Pierre Simon de Laplace w roku 1812.

Definicję tę nazywamy *klasyczną*:

**Prawdopodobieństwem** zajścia zdarzenia  $A$  nazywamy iloraz liczby zdarzeń sprzyjających zdarzeniu  $A$  do liczby wszystkich możliwych przypadków, zakładając, że wszystkie przypadki wzajemnie się wykluczają i są jednakowo możliwe.

Definicję tę można zapisać również w bardziej formalny sposób:

Oznaczmy zbiór wszystkich możliwych przypadków przez  $\Omega$ . Elementami zbioru  $\Omega$  są zdarzenia elementarne  $\omega$ , zaś zbiór  $\Omega$  to zbiór zdarzeń elementarnych. Zbiór zdarzeń sprzyjających  $A$  będzie w takim wypadku podzbiorem zbioru  $\Omega$ :  $A \subset \Omega$ .

**Prawdopodobieństwo** zajścia zdarzenia  $A$  możemy zapisać w postaci:

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

gdzie  $|A|$  oznacza liczbę elementów (moc) zbioru  $A$ , zaś  $|\Omega|$  liczbę elementów (moc) zbioru  $\Omega$ .

*Przykład:* Rzucamy sześcienną kostką. Jakie jest prawdopodobieństwo, że liczba oczek będzie większa od 5?

*Odpowiedź:* Zbiór zdarzeń elementarnych  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , zatem liczba możliwych zdarzeń  $|\Omega| = 6$ . Zbiór zdarzeń sprzyjających  $A = \{6\}$ , liczba zdarzeń sprzyjających  $|A| = 1$ . Prawdopodobieństwo zajścia zdarzenia wynosi:

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

**Prawdopodobieństwo warunkowe** to podstawowe pojęcie teorii prawdopodobieństwa. W zasadzie każde zadanie z rachunku prawdopodobieństwa da się zapisać przy użyciu prawdopodobieństwa warunkowego.

Wyobraźmy sobie, że mamy dwie urny. W pierwszej są same białe kule. W drugiej same czarne. Najpierw wybieramy losowo urnę, a później losujemy kolejno dwie kule. Niech  $A$  oznacza zdarzenie, że pierwsza kula jest biała.  $B$  oznacza zdarzenie, że druga kula jest biała.

$P(B) = \frac{1}{2}$ , bo wybór urny determinuje wybór koloru kuli. Jeśli wiemy, że zaszło zdarzenie  $A$ , to druga wylosowana kula będzie biała, więc prawdopodobieństwo zajścia zdarzenia  $B$ , gdy wiemy, że zaszło zdarzenie  $A$  (oznaczane przez  $P(B|A)$ ) jest równe 1.

**Prawdopodobieństwem warunkowym** zajścia zdarzenia  $A$  pod warunkiem zajścia zdarzenia  $B$ , gdzie  $P(B) > 0$  nazywamy liczbę

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

jest to iloraz prawdopodobieństwa części wspólnej zdarzeń  $A, B$  i prawdopodobieństwa zdarzenia  $B$ .

Rzucamy trzema kostkami. Jakie jest prawdopodobieństwo, że na żadnej kostce nie wypadła szóstka, jeśli na każdej kostce wypadła inna liczba oczek?

Niech  $A$  oznacza zdarzenie, że nie wypadła szóstka.  $B$  zdarzenie, że na każdej kostce wypadła inna liczba oczek.

$$P(A \cap B) = \frac{5 \cdot 4 \cdot 3}{\bar{\bar{\Omega}}} \quad P(B) = \frac{6 \cdot 5 \cdot 4}{\bar{\bar{\Omega}}} \quad P(A|B) = \frac{1}{2}.$$

Chcemy użyć wzoru z definicji.

Jeżeli zdarzenia  $A$  i  $B$  są niezależne (tj.  $P(A \cap B) = P(A)P(B)$ ), to  $P(A|B) = P(A)$ .

Zdarzenia  $A_1$  i  $A_2$  są **niezależne**, jeśli

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2).$$

Def. ogólna: Zdarzenia  $A_1, A_2, \dots, A_n$ , gdzie  $n \geq 2$ , są niezależne, jeśli

$$\bigwedge_{2 \leq k \leq n} \bigwedge_{i_1, \dots, i_k \in \{1, \dots, n\}} P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$$

**Twierdzenie Bayesa** to twierdzenie teorii prawdopodobieństwa. Ma ono bardzo prostą postać, jednak staje się bardzo istotne przy pewnej jego interpretacji.

Niech :

$$X \subset \bigcup_{j=1}^n T_j \wedge T_i T_j = \emptyset, i \neq j$$

Wtedy:

$$P(T_i|X) = \frac{P(T_i)P(X|T_i)}{P(X)}.$$

$$P(T)P(X|T) = P(X)P(T|X) \iff P(T|X) = \frac{P(T)P(X|T)}{P(X)}.$$