

ZMIENNA LOSOWA

Definicja 1

Niech Ω będzie przestrzenią zdarzeń elementarnych.

Zmienną losową nazywamy dowolną funkcję liczbową określoną na przestrzeni zdarzeń elementarnych:

$$X: \Omega \ni \omega \mapsto X(\omega) \in \mathbb{R}.$$

Uwaga

Zmienne losowe oznaczamy dużymi literami np. X, Y, Z , natomiast ich wartości oznaczamy literami małymi, np. x, y, z .

Przykłady zmiennych losowych:

X = liczba wyrzuconych oczek na kostce sześciennej.

Y = „liczba osób, które wsiądą na najbliższym przystanku do autobusu, którym właśnie jedziemy”

Z = „liczba mieszkańców Polski za 10 lat”

ale też:

X = „położenie elektronu w studni potencjału”

Y = „czas jaki pozostał do najbliższej awarii mojego komputera”

Z = „czas między przyjściem turysty pod Wawel a zionięciem ogniem przez Smoka

W = „czas między zapaleniem papierosa a nadjechaniem busa” itp.

Definicja 2

Zmienna losowa jest typu skokowego (zmienna dyskretna), gdy jej zbiór wartości jest skończony lub przeliczalny.

Rozkład prawdopodobieństwa zmiennej losowej dyskretnej

Założmy, że zbiór wartości zmiennej losowej X jest skończony i liczy n elementów: $x_1, x_2, \dots, x_n$.

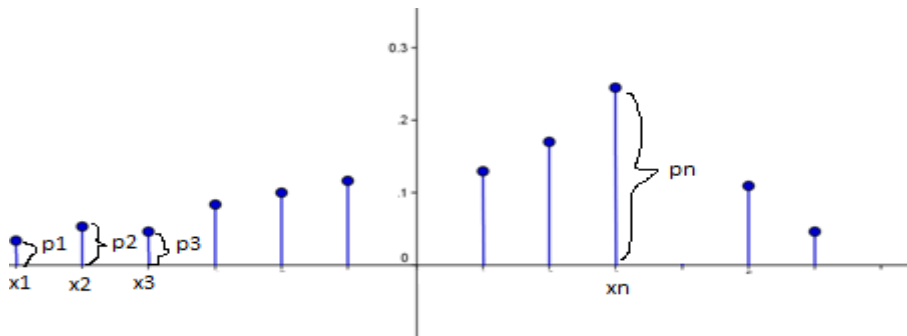
Jeżeli przez p_i oznaczmy prawdopodobieństwo, że zmienna X przyjmie wartość x_i , tzn.

$$p_i = \text{prob}(X = x_i), \text{ dla } i = 1, \dots, n$$

to możemy zdefiniować funkcję rozkładu prawdopodobieństwa:

$$P: x_i \mapsto p_i,$$

która każdej wartości zmiennej losowej X przyporządkowuje prawdopodobieństwo jej przyjęcia.



Z definicji prawdopodobieństwa jest oczywiste, że:

$$\sum_{i=1}^n p_i = 1.$$

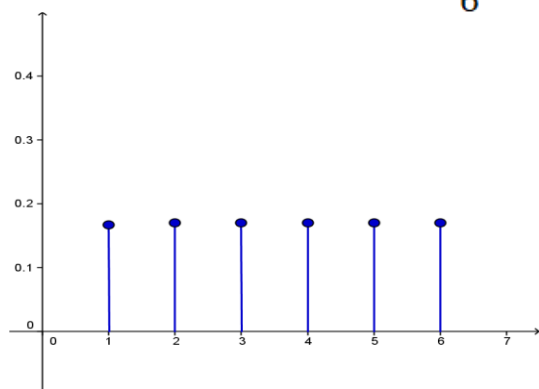
Inny przykład rozkładu prawdopodobieństwa zmiennej losowej dyskretnej:

X = liczba wyrzuconych oczek na kostce sześciennej.

Na podstawie znajomości prawdopodobieństw wystąpienia poszczególnych zdarzeń elementarnych możemy określić prawdopodobieństwa przyjęcia przez zmienną losową X poszczególnych wartości ze zbioru $\{1, \dots, 6\}$.

W tym przypadku prawdopodobieństwa są takie same i wynoszą $\frac{1}{6}$ tzn.

$$\forall x \in \{1, 2, \dots, 6\} P(X = x) = \frac{1}{6}$$



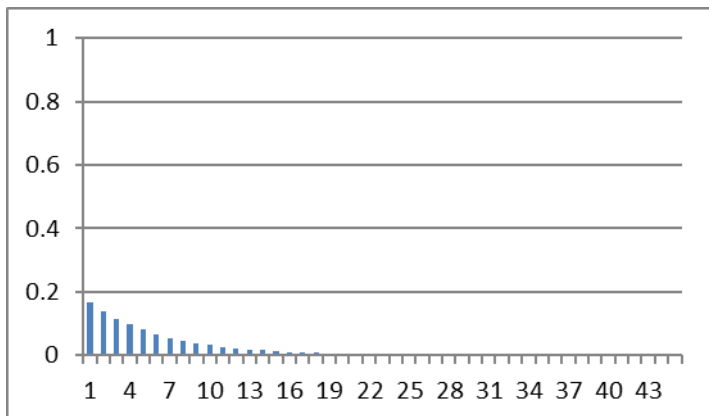
Skokowa zmienna losowa może przyjmować nieskończoną, ale przeliczalną ilość wartości.

Przykładowo, jakie jest prawdopodobieństwo, że „szóstka” pokaże się po raz pierwszy po k rzutach kostką.

Czyli zdarzenia typu: „ pierwsza szóstka wypadnie w pierwszym rzucie”, „ pierwsza szóstka wypadnie w drugim rzucie”, ..., „pierwsza szóstka wypadnie w n -tym rzucie”, ... itd. Gdzie teoretycznie n może być nieskończenie wielkie. Prawdopodobieństwo, że pierwsza szóstka wypadnie w k -tym rzucie jest równe, zgodnie ze schematem Bernoulliego:

$$P(\text{"pierwsza "6" wypadnie w } k - \text{tym rzucie"}) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6},$$

przy czym $\left(\frac{5}{6}\right)^{k-1}$ oznacza, że w pierwszych $k-1$ rzutach mogą się pojawić dowolne wyniki różne od „6” natomiast w k -tym rzucie musi pojawić się „szóstka”, której prawdopodobieństwo wyrzucenia jest równe $\frac{1}{6}$.



Dyskretny rozkład Poissona

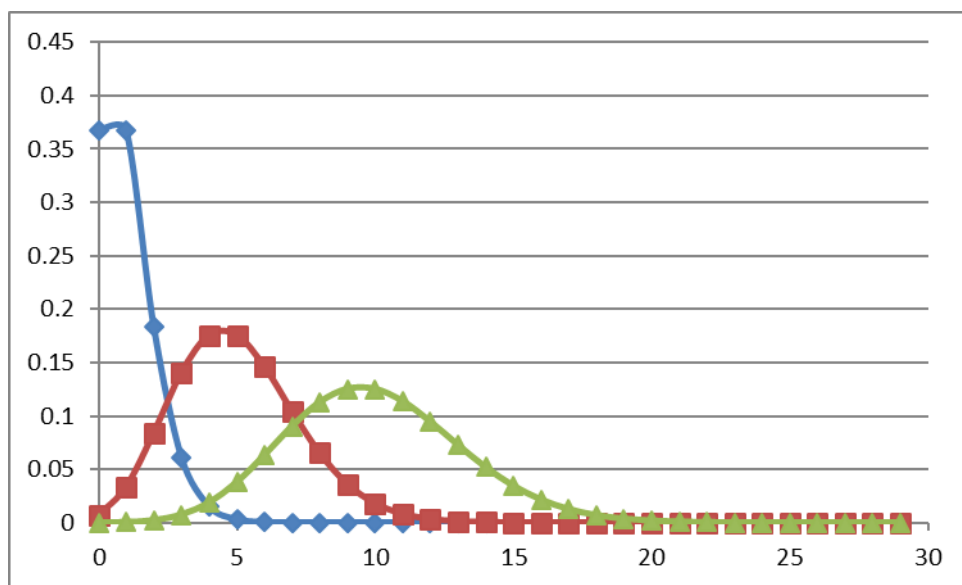
Definicja

Zmienna losowa X ma rozkład Poissona z parametrem $\lambda > 0$ jeżeli

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ dla } k = 0, 1, \dots$$

Jest to prawdopodobieństwo szeregu k wydarzeń mających miejsce w określonym czasie, gdy te wydarzenia występują ze znaną średnią częstotliwością λ i w sposób niezależny od czasu, jaki upłynął od ostatniego zajścia takiego zdarzenia.

Przykłady rozkładu Poissona dla $\lambda=1, 5, 10$:



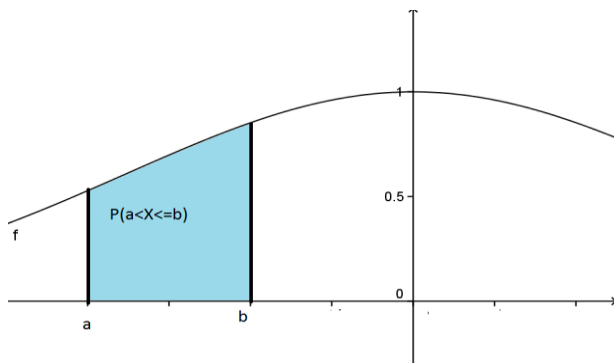
Definicja 3

Zmienna losowa jest typu ciągłego, gdy jej zbiór wartości jest przedziałem lub sumą przedziałów.

Gęstość prawdopodobieństwa zmiennej losowej ciągłej:

X jest zmienną losową ciągłą, jeżeli istnieje funkcja $f(x)$ taka, że

$$\forall a, b \in \mathbb{R} P(X \in (a, b]) = \int_a^b f(x) dx.$$



Z definicji prawdopodobieństwa wynika, że:

- f jest nieujemna
- $\int_{-\infty}^{+\infty} f(x) dx = 1$

Przykład gęstości prawdopodobieństwa ciągłej zmiennej losowej:

Smok wawelski, a w zasadzie jego pomnik stojący nad Wisłą, obecnie ziele ogień co 10 minut. Możemy się więc zastanowić nad problemem turysty przybywającego pod Wawel o dowolnej godzinie i czekającego aż pojawi się ogień. Jak długo będzie musiał czekać? Turysta nie wie, o której godzinie smok ziele ogień. Wie tylko, że odbywa się to co 10 minut. Zatem czas oczekiwania na zionięcie zawiera się w przedziale od 0 do 10 minut:

$$\Omega = \{\omega = \text{"smok będzie zionął po czasie } x" : x \in [0,10]\}.$$

Zmienna losowa zdefiniowana jest wtedy następująco:

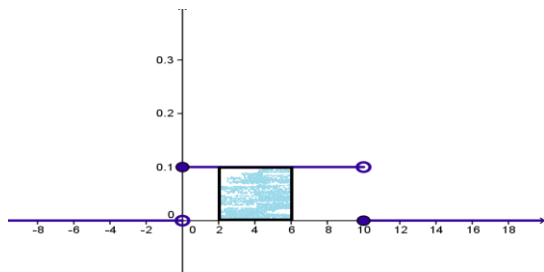
$$X: \Omega \ni \omega \mapsto X(\omega) = \text{"czas oczekiwania na zionięcie smoka"} \in \mathbb{R}.$$

Gęstość prawdopodobieństwa musi być równa zero dla $x < 0$ (zanim turysta poszedł) i dla $x > 10$ (jak zobaczy smoka zielejącego ogniem, to pójdzie). Ponieważ turysta nie wie, kiedy smok ziele, każdy czas oczekiwania jest „równie prawdopodobny”. Oznacza to, że dla każdego x z przedziału $[0,10]$ funkcja gęstości przyjmuje takie same wartości:

$$\forall x \in [0,10] \quad f(x) = c.$$

Wiemy jednak, że $\int_{-\infty}^{+\infty} f(x) dx = 1$, a ponieważ funkcja f zeruje się poza przedziałem $[0,10]$ to:

$$1 = \int_0^{10} f(x) dx = \int_0^{10} c dx = 10c. \text{ Oznacza to, że } c = \frac{1}{10}$$



Mając gęstość prawdopodobieństwa można wyliczyć np. prawdopodobieństwo, że czas oczekiwania na pojawienie się ognia będzie w przedziale od 2 do 5 minut:

$$P(X \in [2,5]) = \int_2^5 f(x) dx = \int_2^5 \frac{1}{10} dx = \frac{1}{10} x \Big|_2^5 = \frac{3}{10}.$$

Rozkład normalny

Rozkład normalny jest najważniejszym rozkładem występującym w statystyce.

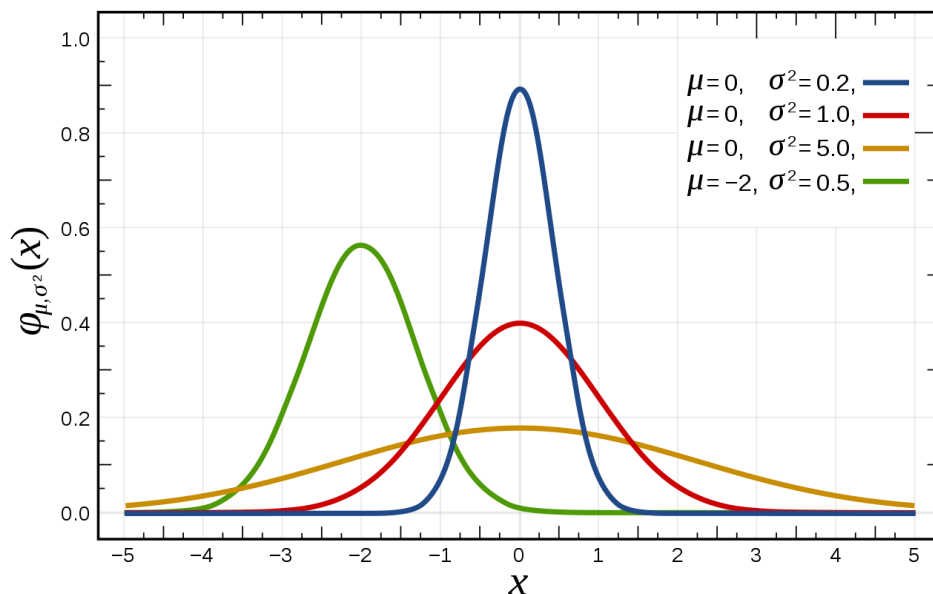
Definicja:

Zmienna losowa X ma rozkład normalny o parametrach μ i σ , gdy jej gęstość prawdopodobieństwa jest dana wzorem

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

Przykłady rozkładu normalnego o różnych parametrach wartości oczekiwanej i szerokości.

Rozkład w kolorze czerwonym to tzw. standardowy rozkład normalny.



Dystrybuanta

Rozkład zmiennej losowej można scharakteryzować również za pomocą dystrybuanty.

Definicja:

Dystrybuantą zmiennej losowej X nazywamy funkcję $F: \mathbb{R} \rightarrow [0,1]$, taką, że

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

Uwaga

Dystrybuantą zmiennej losowej jest funkcja, która każdej liczbie rzeczywistej przyporządkowuje prawdopodobieństwo, że zmienna losowa X przyjmie wartości nie większe niż x .

W przypadku, gdy X jest zmienną losową skokową o wartościach $x_1, x_2, \dots$ i prawdopodobieństwach $p_1, p_2, \dots$ oznacza to, że wartość dystrybuanty w punkcie x jest równa sumie prawdopodobieństw p_i przyjęcia przez zmienną X wartości x_i nie większych niż x

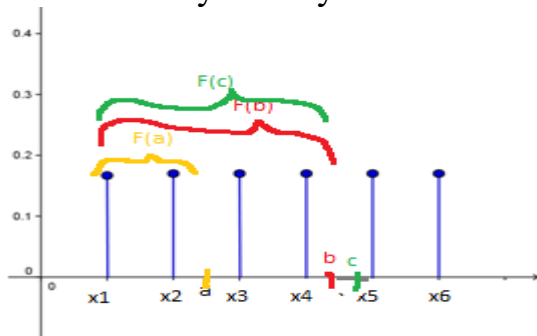
$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i,$$

Jeżeli natomiast X jest zmienną losową ciągłą o gęstości f to prawdopodobieństwo jest definiowane jako odpowiednia całka, zatem też i wartość dystrybuanty będzie tak zdefiniowana:

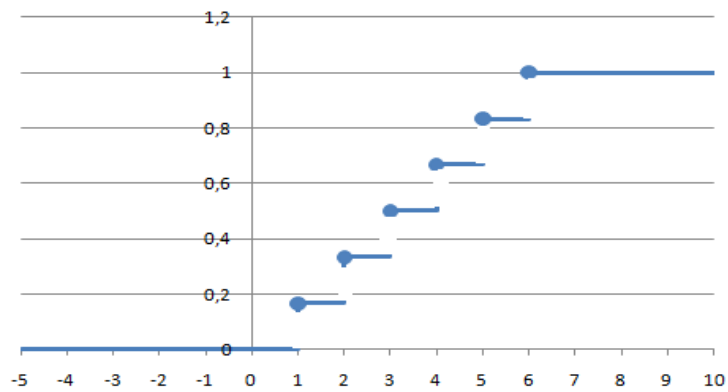
$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Dystrybuanta dyskretnej zmiennej losowej

X = liczba wyrzuconych oczek na kostce sześcienniej.



Dystrybuanta dyskretnej zmiennej losowej jest funkcją lewostronnie ciągłą:

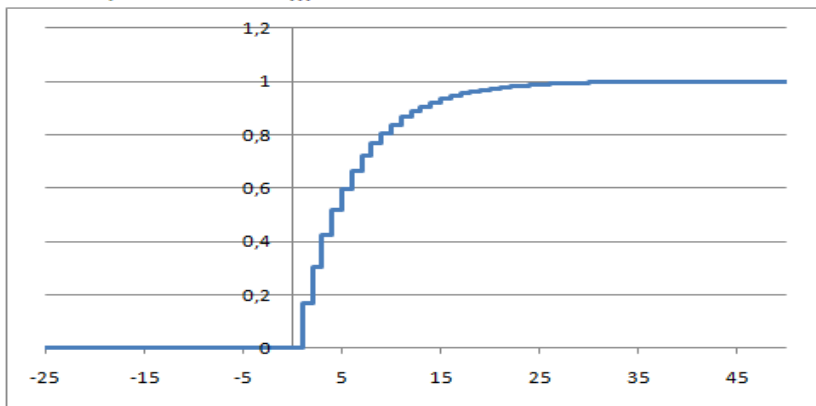


X = po którym razie wypadnie pierwsza szóstka

$$P(X = k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}, \text{ dla } k = 1, 2, \dots,$$

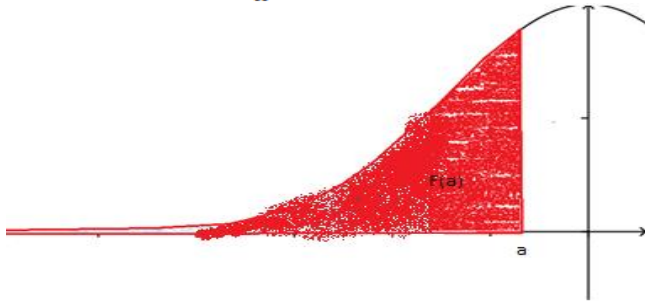
Wtedy dystrybuanta

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{6}, & x \in [1, 2) \\ \frac{1}{6} + \frac{1}{6} \cdot \frac{5}{6}, & x \in [2, 3) \\ \frac{1}{6} \cdot \left(1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2\right), & x \in [3, 4) \\ \vdots \\ 1 - \left(\frac{5}{6}\right)^n, & x \in [n, n+1) \end{cases}$$



Dystrybuanta ciągłej zmiennej losowej:

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



Natomiast pochodna dystrybuanty $F(x)$ daje z powrotem rozkład prawdopodobieństwa $f(x)$.

Czekanie na Smoka – ciągła zmienna losowa:

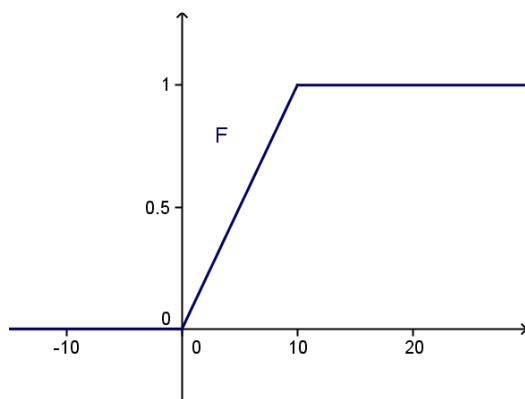
$$f(x) = \begin{cases} 0 & \text{dla } x \in (-\infty, 0) \\ \frac{1}{10} & \text{dla } x \in [0, 10] \\ 0 & \text{dla } x \in (10, +\infty). \end{cases}$$

Dystrybuanta dla przedziału $[0, 10]$:

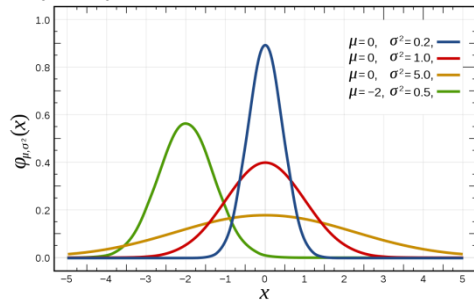
$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt = \int_0^x \frac{1}{10} dt = \frac{1}{10} t \Big|_0^x = \frac{1}{10} x.$$

Pełna dystrybuanta:

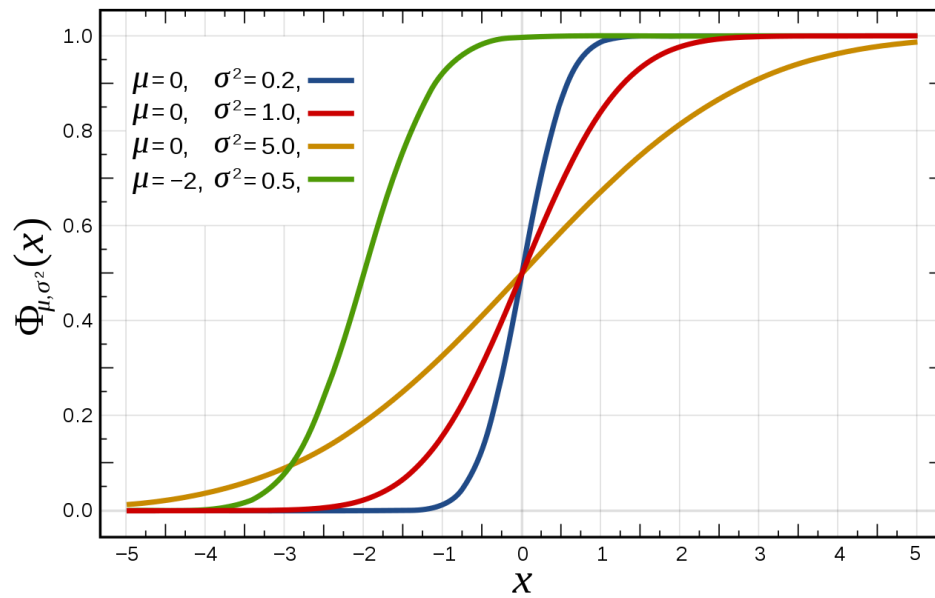
$$F(x) = \begin{cases} 0 & \text{dla } x \in (-\infty, 0) \\ \frac{1}{10} x & \text{dla } x \in [0, 10] \\ 1 & \text{dla } x \in (10, +\infty). \end{cases}$$



Dystrybuanta rozkładu normalnego:



$$F(x) = \int_{-\infty}^x e^{-\frac{(t-m)^2}{2\sigma^2}} dt = (1 + \operatorname{erf}(x/\sqrt{2}))/2$$



CHARAKTERYSTYKI ZMIENNYCH LOSOWYCH

1. Wartość oczekiwana

Wartość oczekiwana opisuje średnią wartość przyjmowaną przez zmienną losową.

Definicja

Niech X będzie dowolną zmienną losową o skończonym zbiorze wartości, np. $x_1, x_2, \dots, x_n$ i prawdopodobieństwami $p_i = P(X = x_i)$ dla $i=1, \dots, n$.

Wartość oczekiwana skokowej zmiennej losowej X jest równa:

$$EX = \sum_{i=1}^n x_i \cdot p_i$$

Definicja

Niech X będzie dowolną zmienną losową ciągłą o gęstości f .

Wartość oczekiwana zmiennej losowej X jest równa:

$$EX = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

Uwaga

Istnieją zmienne losowe, które nie posiadają wartości oczekiwanej, gdy suma szeregu (całka) w definicji wartości oczekiwanej zmiennej losowej skokowej (ciągłej) nie jest bezwzględnie zbieżna.

Przykład-promocja chipsów

Producent ogłasza, że w co dziesiątej paczce chipsów znajduje się banknot o nominale 100 złotych. Niech wartościami rozważanej przez nas zmiennej losowej Z będzie liczba paczek, które trzeba kolejno kupić i zjeść, aby w końcu znaleźć jeden (pierwszy) banknot.

Jej rozkład jest następujący:

$$P(Z = k) = \frac{1}{10} \cdot \left(\frac{9}{10}\right)^{k-1},$$

gdzie k oznacza numer pierwszej kupionej paczki, w której pojawia się wygrana.

Wartość oczekiwana zmiennej losowej Z jest równa:

$$EZ = \sum_{i=1}^{+\infty} z_i \cdot p_i = \sum_{k=1}^{+\infty} k \cdot \frac{1}{10} \cdot \left(\frac{9}{10}\right)^{k-1} = \frac{1}{10} \cdot \sum_{k=1}^{+\infty} k \cdot \left(\frac{9}{10}\right)^{k-1}.$$

Aby obliczyć sumę $\sum_{k=1}^{+\infty} k \cdot \left(\frac{9}{10}\right)^{k-1}$ rozważmy szereg potęgowy $g(x) = \sum_{k=1}^{+\infty} k \cdot x^{k-1}$, którego promień zbieżności R jest równy 1. Oznacza to, że dla $x \in (-1, 1)$ szereg ten jest bezwzględnie zbieżny. Możemy więc dla x z tego przedziału stosować twierdzenie o całkowaniu szeregu potęgowego.

$$\int g(x) dx = \int \left(\sum_{k=1}^{+\infty} k \cdot x^{k-1} \right) dx = \sum_{k=1}^{+\infty} \left(\int k \cdot x^{k-1} dx \right) = \sum_{k=1}^{+\infty} \left(k \cdot \frac{1}{k} \cdot x^k \right) = \sum_{k=1}^{+\infty} x^k = \frac{x}{1-x}.$$

Przy czym ostatnie równość wynika z zastosowania wzoru na sumę szeregu geometrycznego.

Z tego wynika, że

$$g(x) = \left(\frac{x}{1-x} \right)' = \frac{1-x+x}{(1-x)^2} = \frac{1}{(1-x)^2}.$$

W szczególności, podstawiając $x = \frac{9}{10}$, otrzymujemy:

$$EZ = \frac{1}{10} \cdot \sum_{k=1}^{+\infty} k \cdot \left(\frac{9}{10}\right)^{k-1} = \frac{1}{10} \cdot \frac{1}{\left(1 - \frac{9}{10}\right)^2} = 10.$$

Wynik ten oznacza, że aby zdobyć jeden promocyjny banknot przeciętny uczestnik promocji musi kupić 10 paczek chipsów.

Własności wartości oczekiwanej:

- $E(c)=c$
- X – zmienna losowa o wartości oczekiwanej EX , $a \in \mathbb{R}$, to
 $E(a \cdot X) = a \cdot EX$
- X – zmienna losowa o wartości oczekiwanej EX , $b \in \mathbb{R}$, to
 $E(X + b) = b + EX$
- X, Y – zmienne losowe, to
 $E(X + Y) = EX + EY$.

Uwaga

Zaprezentowane powyżej własności wartości oczekiwanej oznaczają, że jest ona liniowa i można je zapisać następująco. Jeżeli $a, b, c \in \mathbb{R}$ oraz X i Y są zmiennymi losowymi, to wówczas

$$E(aX + bY + c) = aEX + bEY + c.$$

W szczególności oznacza to, że jeżeli X i Y są dowolnymi zmiennymi losowymi, to:

$$E(X - Y) = EX - EY,$$

czyli wartość oczekiwana różnicy zmiennych losowych jest równa różnicy ich wartości oczekiwanych.

Twierdzenie

Jeżeli X i Y są niezależnymi zmiennymi losowymi, to

$$E(X \cdot Y) = EX \cdot EY.$$

Oznacza to, że wartość oczekiwana iloczynu niezależnych zmiennych losowych jest równa iloczynowi wartości oczekiwanych tych zmiennych.

Definicja

Momentem rzędu k ($k=1,2,\dots$) zmiennej losowej X nazywamy wyrażenie

$$m_k = E(X^k).$$

Uwaga

Korzystając z definicji wartości oczekiwanej moment rzędu k zmiennej losowej X możemy obliczyć ze wzoru

$$m_k = E(X^k) = \begin{cases} \sum_i^n x_i^k p_i & \text{gdy } X \text{ jest zmienną losową skokową} \\ \int_{-\infty}^{+\infty} x^k f(x) dx & \text{gdy } X \text{ jest ciągłą zmienną losową} \end{cases}$$

2. Wariancja

Wariancja zmiennej losowej (drugi moment centralny) jest równa wartości oczekiwanej kwadratu odchylenia zmiennej losowej od jej wartości średniej, tzn. jeżeli X jest zmienną losową o skończonej wartości oczekiwanej E , to jej wariancja jest równa

$$D^2 X = E(X - EX)^2$$

Oznacza to, że wariancja mierzy średni kwadrat odchylenia wartości zmiennej losowej od jej wartości oczekiwanej. Dzięki temu uzyskujemy informację o rozrzucie wartości zmiennej losowej wokół jej wartości oczekiwanej. W praktyce używa się pierwiastka z wariancji

Definicja

Niech X będzie dowolną zmienną losową skokową o wartościach x_i i prawdopodobieństwach p_i . Wariancja zmiennej losowej X jest równa:

$$D^2 X = \sum_{i=1}^{+\infty} (x_i - EX)^2 \cdot p_i$$

Definicja

Niech X będzie dowolną zmienną losową ciągłą o gęstości $f(x)$.

Wariancja oczekiwana zmiennej losowej X jest równa:

$$D^2 X = \int_{-\infty}^{+\infty} (x - EX)^2 \cdot f(x) dx$$

3. Skośność

Trzeci moment centralny jest równy wartości oczekiwanej sześcianu odchylenia zmiennej losowej od jej wartości średniej

$$D^3 X = E(X - EX)^3$$

Skośność czyli trzeci moment centralny podzielony przez odchylenie standardowe do sześcianu jest miarą niesymetryczności rozkładu, np. ujemna wartość oznacza rozkład „o grubym ogonie” z lewej strony.

4. Kurtoza

Kurtoza to czwarty moment centralny znormalizowany przez kwadrat wariancji,

Czwarty moment centralny:

$$D^4 X = E(X - EX)^4$$

Kurtoza jest miarą skupienia rozkładu (szybkiego zanikania „ogonów”).

Kurtoza rozkładu normalnego = 3, mniejsza kurtoza oznacza „grubsze ogony”.