Wykład 3.

Zmienne losowe i ich rozkłady

dr hab.inż. Katarzyna Zakrzewska, prof.AGH, Katedra Elektroniki, WIET AGH

Plan:

- Pojęcie zmiennej losowej
- Ilościowy opis zmiennych losowych
- Przykładowe rozkłady zmiennych losowych

Pojęcie zmiennej losowej

Zmienna losowa jest to funkcja X, która przypisuje liczbę rzeczywistą x danemu wynikowi eksperymentu losowego.

$$\Omega = \{e_1, e_2, \ldots\}$$

$$X : \Omega \rightarrow R$$

$$X(e_i) = x_i \in R$$

Przykłady:

1) Rzut monetą: zdarzeniu ‘orzeł’ przypisujemy 1; zdarzeniu reszka przypisujemy 0.
2) Analog. losowanie wyrobów: zdarzeniu ‘brak’ (wadliwy) - 0, dobry – 1
3) Rzut kostką wyrzucenie ‘1’ – 1, ‘2’ – 2 itd…
4) Odcinek [a, b] na osi liczbowej – wybór punktu o współrzędnej ‘x’ przypisujemy np. wartość x ; wartość $\sin^2(3x+17)$ itp.…

Zmienna losowa

dyskretna

Gdy wartości zmiennej losowej X są izolowanymi punktami na osi liczbowej (obejmują skończony przedział wartości)

- *Rzut monetą*
- *Błędy przy transmisji*
- *Wadliwe układy z linii produkcyjnej.*
- *Ilość połączeń przychodzących w ciągu 5 minut*

ciągła

Gdy wartości zmiennej losowej stanowią wszystkie punkty odcinka (obejmują przedział liczb rzeczywistych)

- *Natężenie prądu w przewodniku*
- *Temperatura*
- *Ciśnienie*

Ilościowy opis zmiennych losowych

- Rozkład zmiennej losowej lub rozkład prawdopodobieństwa (tylko dla zmiennych dyskretnych)
- Funkcja gęstości prawdopodobieństwa (tylko dla zmiennych ciągłych)
- Dystrybuanta (funkcja rozkładu dla zmiennych dyskretnych i ciągłych)
- Wielkości charakteryzujące (wartość oczekiwana, wariancja, kwantyle, itp.)

Rozkład zmiennej losowej

Rozkładem zmiennej losowej (rozkładem prawdopodobieństwa dla zmiennych dyskretnych) nazywamy zbiór par (x_i, p_i) gdzie x_i jest wartością zmiennej losowej X a p_i jest prawdopodobieństwem, że zmienna losowa X przyjmuje wartość x_i

Przykład 3.1

Rozkład prawdopodobieństwa dla jednokrotnego rzutu monetą. Zdarzeniu polegającemu na wyrzuceniu orła przypisujemy $x_1=1$; zdarzeniu polegającemu na wyrzuceniu reszki $x_2=0$. Zatem:

$$x_1 = 1 \quad p(X = 1) = p(x_1) = \frac{1}{2}$$

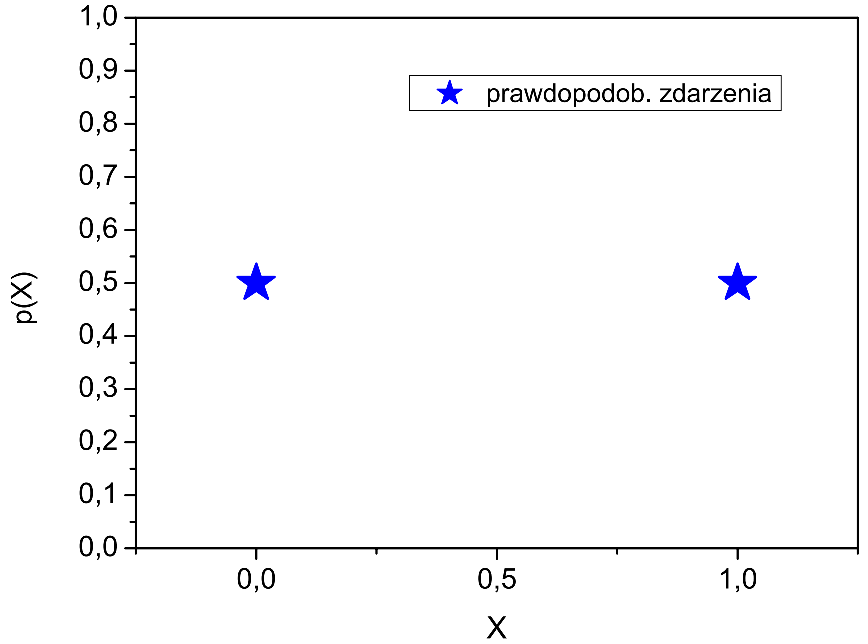
$$x_2 = 0 \quad p(X = 0) = p(x_2) = \frac{1}{2}$$

Rozkład zmiennej losowej

Przykład 3.1 cd

Rozkład prawdopodobieństwa dla jednokrotnego rzutu monetą jest następującym zbiorem par:

$$\{(1, \frac{1}{2}), (0, \frac{1}{2})\}$$



Zmienna losowa jest w tym przypadku skokowa (dyskretna) a jej rozkład jest też skokowy (dyskretny).

Funkcja gęstości prawdopodobieństwa

Funkcję gęstości prawdopodobieństwa wprowadza się dla zmiennych ciągłych; ma ona związek z prawdopodobieństwem:

$$f(x)dx \equiv P(x \le X < x + dx)$$

Własności funkcji gęstości prawdopodobieństwa:

$$1.\ \ f(x) \ge 0$$

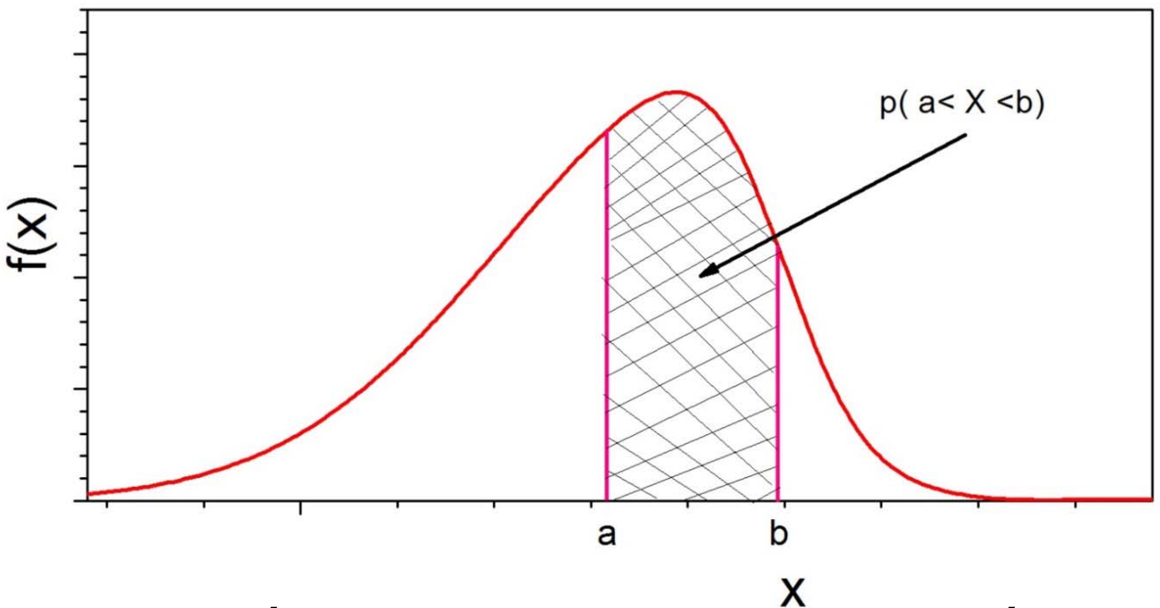
$$2.\ \ f(x)\ jest\ unormowana\ \int_{-\infty}^{+\infty} f(x)dx = 1$$

3. f(x) ma wymiar 1/x

Funkcja gęstości prawdopodobieństwa

Z definicji funkcji gęstości prawdopodobieństwa f(x) wynika praktyczny sposób obliczania prawdopodobieństwa, że wartość zmiennej losowej znajduje się w przedziale [a,b]:

$$P(a < X < b) = \int_a^b f(x)dx$$

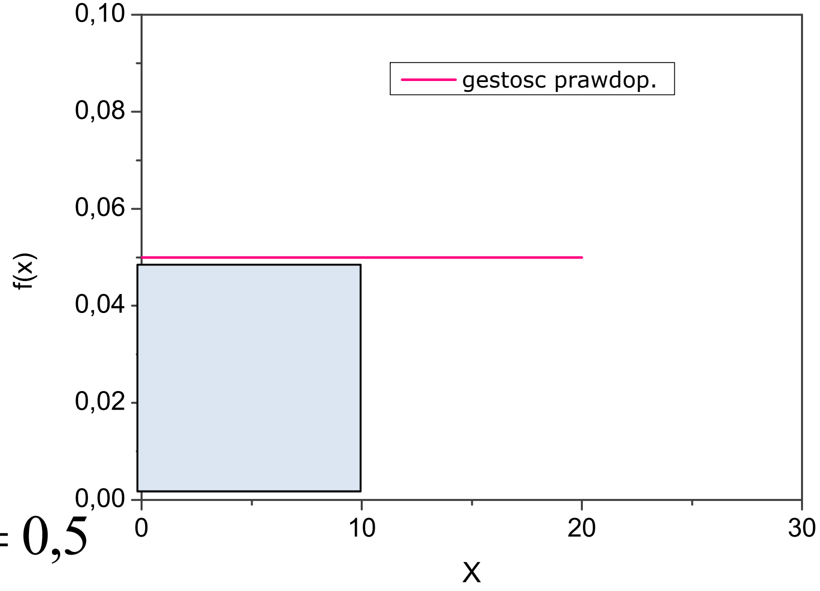


Nie ma sensu pytać, jakie jest prawdopodobieństwo, że x=a

Funkcja gęstości prawdopodobieństwa

Przykład 3.2

Oznaczmy przez X zmienną losową ciągłą, która opisuje natężenie prądu w cienkim przewodzie miedzianym (w jednostkach mA). Załóżmy, że zakres X wynosi [0, 20 mA] i funkcja gęstości prawdopodobieństwa jest dana jest jako f(x)=0,05 w tym przedziale. Oblicz prawdopodobieństwo, że zmierzone natężenie prądu jest mniejsze niż 10 mA.



$$P(0 \leq X < 10) = \int_0^{10} f(x)dx = \int_0^{10} 0{,}05\, dx = 0{,}5$$

Ilościowy opis zmiennych losowych

Dystrybuantą (funkcja rozkładu, ang. cumulative distribution function – CDF) F(x) nazywamy prawdopodobieństwo tego, że zmienna losowa X przyjmie wartość mniejszą od x (co najwyżej daną wartość)

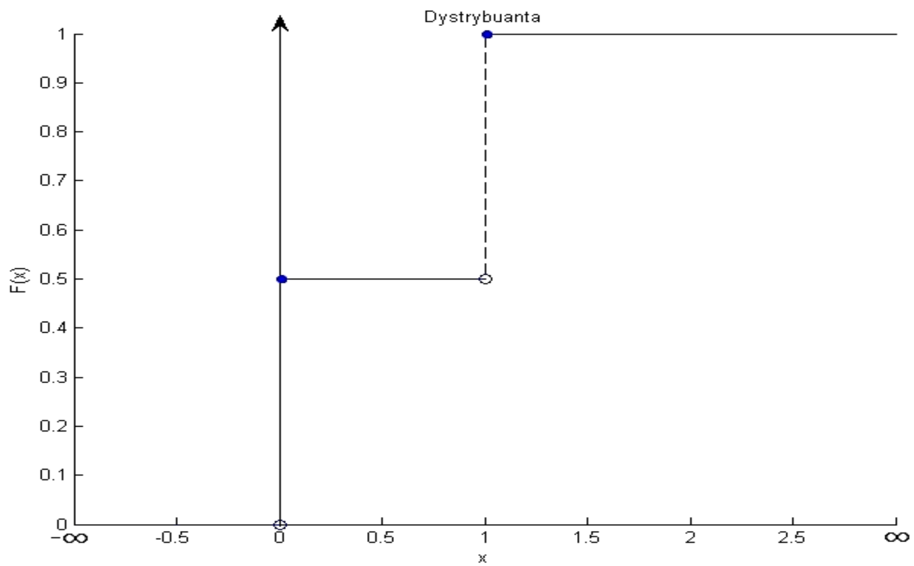
$$F(x) = P(X \le x)$$

Przykład 3.1 cd

Dystrybuanta dla rzutu monetą:

$$F(x=0) = P(X \le 0) = \frac{1}{2}$$

$$F(x=1) = P(X \le 1) = 1$$



Własności dystrybuanty

1. $0 \le F(x) \le 1$

2. $F(-\infty) = 0$

3. $F(+\infty) = 1$

4. $x \le y \;\Rightarrow\; F(x) \le F(y)$

Jest funkcją niemalejącą

5. F(x) nie posiada wymiaru

6. $f(x) = \dfrac{dF(x)}{dx}$ Związek gęstości prawdopodobieństwa z dystrybuantą (dla zmiennej ciągłej)

Dystrybuanta dla zmiennej dyskretnej

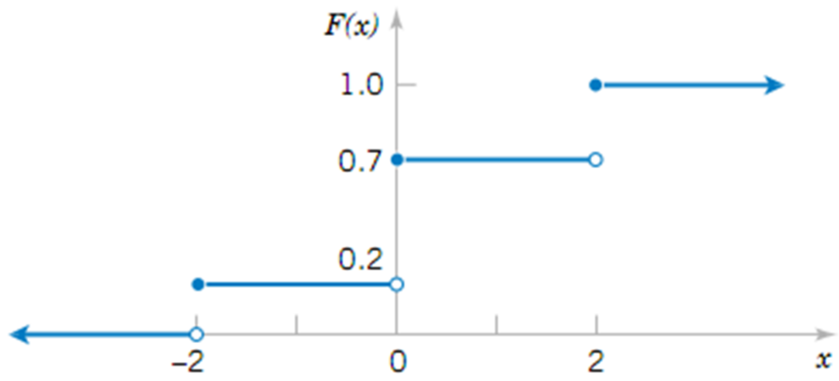
$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

***f* (x_i) – rozkład prawdopodobieństwa**

Przykład 3.3

Na podstawie następujących wartości dystrybuanty F(x) znajdź funkcję rozkładu prawdopodobieństwa f(x)

$$F(x) = 0 \; dla \;\; x < -2$$
$$0{,}2 \;\; dla \;\; -2 \le x < 0$$
$$0{,}7 \;\; dla \;\; 0 \le x < 2$$
$$1 \;\; dla \;\; 2 \le x$$



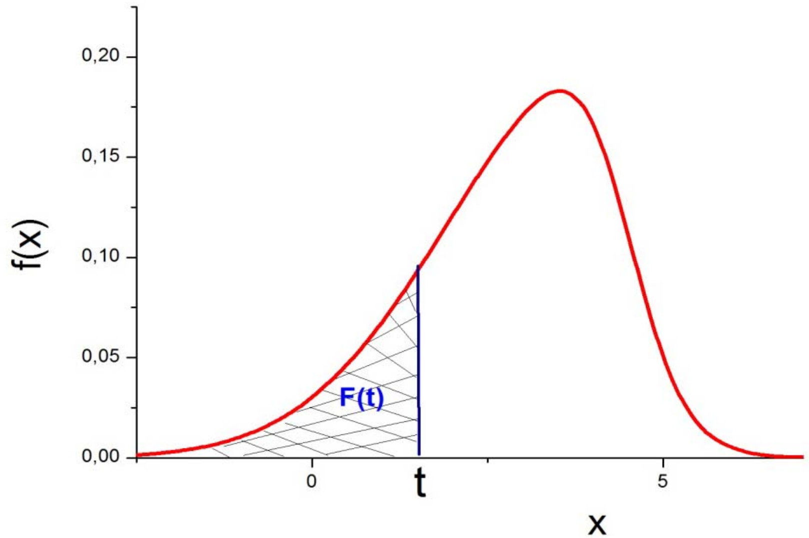
Na podstawie rysunku, jedynymi punktami dla których f(x)≠0 są -2, 0, 2.

$$f(-2) = 0{,}2 - 0 = 0{,}2 \quad f(0) = 0{,}7 - 0{,}2 = 0{,}5 \quad f(2) = 1{,}0 - 0{,}7 = 0{,}3$$

Dystrybuanta dla zmiennej ciągłej

$$F(t) = P(X \le t) = \int_{-\infty}^{t} f(u)\,du$$

Dystrybuanta zmiennej ciągłej jest niemalejącą funkcją ciągłą a oblicza się ją jako pole pod wykresem funkcji gęstości prawdopodobieństwa.



Numeryczne miary opisowe

MIARY (parametry) OPISOWE

Położenia

- Kwantyle (np. mediana)
- Moda
- Wartość oczekiwana (średnia, nadzieja matematyczna)

Rozproszenia

- Wariancja (Odchylenie standardowe)
- Rozstęp

Charakterystyki rozkładu prawdopodobieństwa

Fraktyl (kwantyl) x_q jest to wartość zmiennej losowej, dla której dystrybuanta przyjmuje wartość q.

$$F(x_q) = P(X \leq x_q) = \int_{-\infty}^{x_q} f(u)\,du = q$$

Najczęściej stosowanym kwantylem jest mediana czyli $x_{0.5}$.

W przykładzie 3.2 natężenie prądu 10 mA jest medianą rozkładu.

Przykład 3.4

Dla dyskretnego rozkładu eksperymentalnego o wynikach: 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 mediana wynosi 22 bo jest wartość środkowa uporządkowanego zbioru wartości (albo średnia arytmetyczna dwóch środkowych wielkości).

Charakterystyki rozkładu prawdopodobieństwa

Moda (wartość modalna) jest to taka wartość zmiennej losowej, dla której rozkład prawdopodobieństwa (lub funkcja gęstości prawdopodobieństwa) osiąga maksimum.

Rozkłady jednomodalne mają jedną modę (wielomodalne – więcej niż jedną)

W przykładzie 3.4 dla dyskretnego rozkładu eksperymentalnego o wynikach: 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 moda wynosi 21 bo jest wartość, która pojawia się najczęściej w zbiorze wyników.

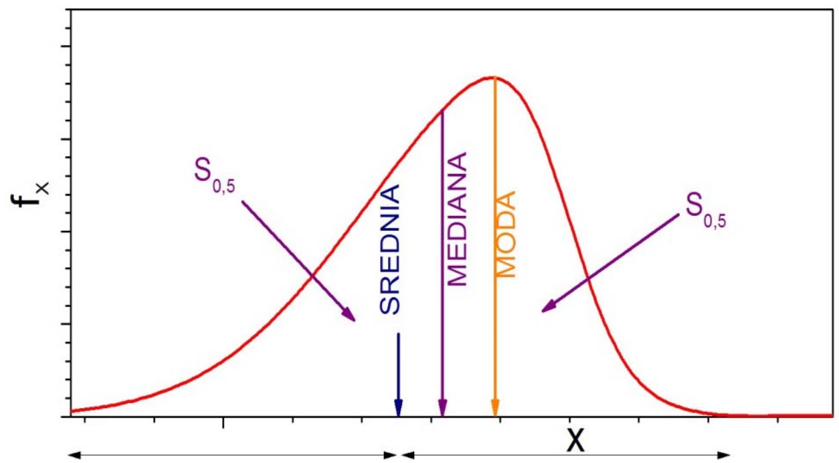
Wartość średnia

Średnia arytmetyczna:

x_i - elementy zbioru n – elementowego (niekoniecznie różne):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

W przykładzie 3.4 dla dyskretnego rozkładu eksperymentalnego o wynikach: 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 wartość średnia wynosi 22,7.



Średnia arytmetyczna

Jeżeli wiele elementów ma w zbiorze tę samą wartość, to dzielimy zbiór na klasy zawierające identyczne elementy o liczebnościach n_k:

Przykład 3.5

x_k	n_k	f_k
10,2	1	0,0357
12,3	4	0,1429
12,4	2	0,0714
13,4	8	0,2857
16,4	4	0,1429
17,5	3	0,1071
19,3	1	0,0357
21,4	2	0,0714
22,4	2	0,0714
25,2	1	0,0357
Razem	28	

$$\overline{x} = \frac{\sum_{k=1}^{p} n_k x_k}{n} = \sum_{k=1}^{p} f_k x_k$$

gdzie: $f_k = \frac{n_k}{n}, p - liczba\ \ klas\ (p \leq n)$

Warunek normalizacji $\sum_k f_k = 1$

$$\overline{x} = x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n =$$

$$= 10{,}2 \cdot 0{,}04 + 12{,}3 \cdot 0{,}14 + \ldots + 25{,}2 \cdot 0{,}04$$

$$\overline{x} = 15{,}77$$

Charakterystyki rozkładu prawdopodobieństwa

Moment rozkładu rzędu k względem punktu x_0

$$m_k(x_0) \equiv \sum_i (x_i - x_0)^k \, p(x_i)$$ dla zmiennych dyskretnych

$$m_k(x_0) \equiv \int (x - x_0)^k f(x)\, dx$$ dla zmiennych ciągłych

Najważniejszymi momentami są te, które są liczone względem początku układu współrzędnych czyli względem $x_0=0$ (m_k) oraz momenty liczone względem $X_0=m_1$ tj. względem pierwszego momentu względem początku układu współrzędnych (m_1 nazywamy wartością oczekiwaną, wartością średnią lub nadzieją matematyczną) – to są momenty centralne μ_k.

Charakterystyki rozkładu prawdopodobieństwa

Wartość oczekiwana oznaczana jako: m_1, E(X), μ, $\overline{x}$, $\hat{x}$

$$E(X) = \sum_i x_i p_i$$ dla zmiennych dyskretnych

$$E(X) \equiv \int x\, f(x)\, dx$$ dla zmiennych ciągłych

E(X) jest współrzędną punktu, który byłby środkiem masy rozkładu prawdopodobieństwa (lub pola pod funkcją gęstości prawdopodobieństwa f(x)) gdyby p_i traktować jak masy (lub odpowiednio f(x) jak fizyczną gęstość).

Własności E(X)

E(X) jest operatorem liniowym co oznacza, że:

1. $$E(\sum_i C_i X_i) = \sum_i C_i E(X_i)$$

co prowadzi w konsekwencji do:

$$E(C) = C$$

$$E(CX) = CE(X)$$

$$E(X_1+X_2) = E(X_1)+E(X_2)$$

2. Dla niezależnych zmiennych $X_1, X_2, \ldots X_n$

$$E(\prod_i X_i) = \prod_i E(X_i)$$

Warunkiem koniecznym i wystarczającym by zmienne były niezależne jest

$$f(X_1, X_2, \ldots, X_n) = f_1(X_1) f_2(X_2) \cdot \ldots \cdot f_n(X_n)$$

Własności E(X)

3. Dla funkcji zmiennej X; Y= Y(X) wartość oczekiwana E(Y) może być znaleziona przy pomocy rozkładu zmiennej X bez konieczności szukania rozkładu f(y)

$$E(Y) = \sum_i \; y(x_i)\, p_i$$ dla zmiennych dyskretnych

$$E(Y) \equiv \int y(x)\; f(x)\, dx$$ dla zmiennych ciągłych

Można zauważyć, że dowolny moment $m_k(x_0)$ może być potraktowany jako wartość oczekiwana funkcji $Y(X)=(X-x_0)^k$

$$m_k(x_0) \equiv \int (x - x_0)^k\, f(x)\, dx = E((x - x_0)^k)$$

Charakterystyki rozkładu prawdopodobieństwa

Wariancja (dyspersja) oznaczana jako: $\sigma^2(X)$, var(X), V(X), D(X).
Pierwiastek z wariancji nazywamy *odchyleniem standardowym* $\sigma(x)$

$$\sigma^2(X) \equiv \sum_i p_i (x_i - E(X))^2$$ dla zmiennych dyskretnych

$$\sigma^2(X) \equiv \int f(x)(x - E(X)^2\, dx$$ dla zmiennych ciągłych

Wariancja (lub odchylenie standardowe) jest miarą rozrzutu zmiennej losowej wokół wartości oczekiwanej.

W analizie danych doświadczalnych utożsamiamy wartość oczekiwaną pomiarów wykonanych w obecności błędów przypadkowych z wartością rzeczywistą mierzonej wielkości. Miarą błędu przypadkowego jest odchylenie standardowe bo ono określa rozrzut wyników wokół wartości rzeczywistej.

Własności $\sigma^2(X)$

Wariancję można obliczyć stosując wartości oczekiwane:

1. $$\sigma^2(X) = E(X^2) - E^2(X)$$

co prowadzi w konsekwencji do:

$$\sigma^2(C) = 0$$

$$\sigma^2(CX) = C^2\,\sigma^2(X)$$

$$\sigma^2(C_1X + C_2) = C_1^2\,\sigma^2(X)$$

2. Dla niezależnych zmiennych $X_1, X_2, \dots X_n$

$$\sigma^2\left(\sum_i C_i X_i\right) = \sum_i C_i^2 \sigma^2(X)$$

Nierówność Czebyszewa

Interpretacja wariancji wynika z nierówności Czebyszewa:

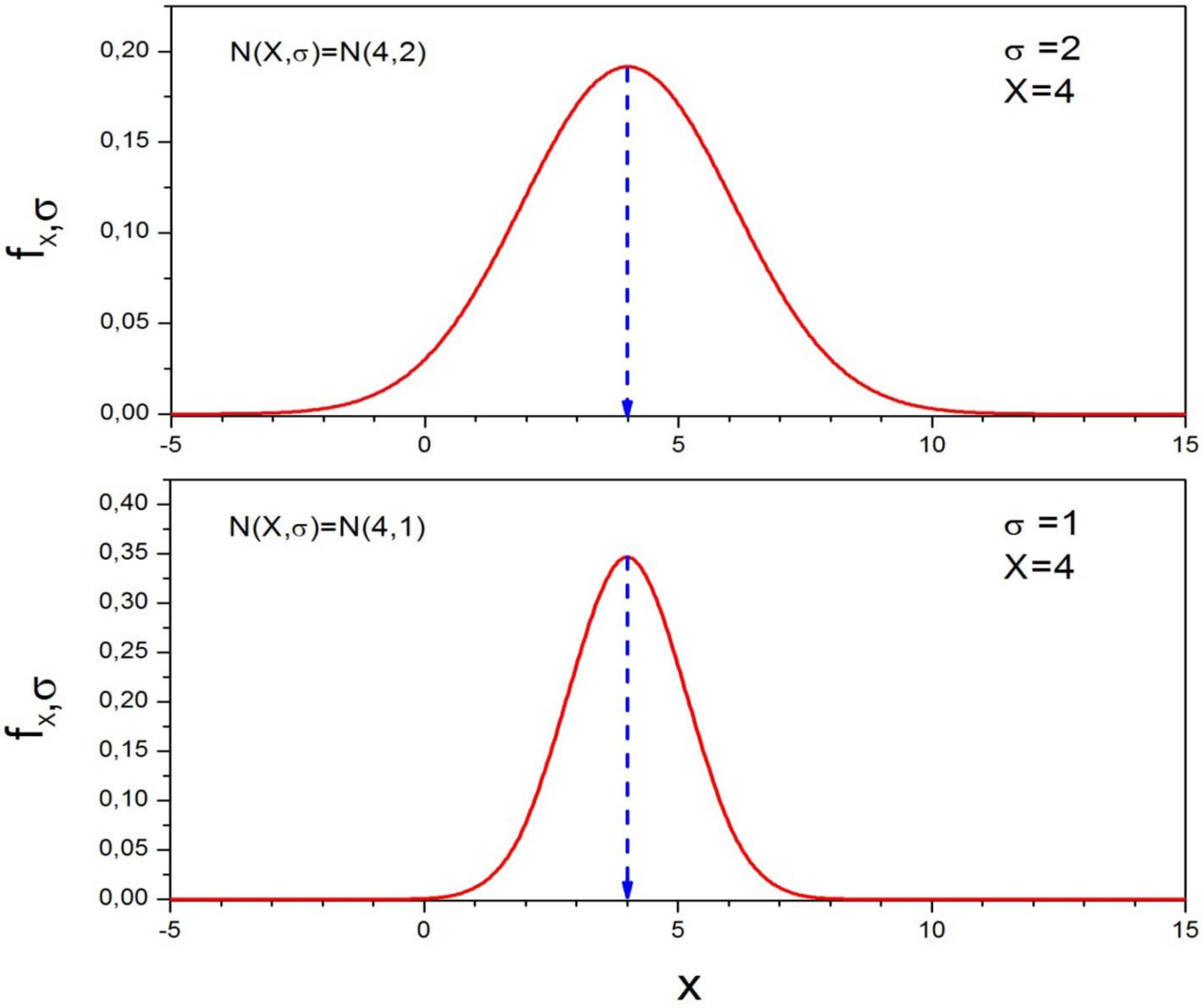
$$P\left(\left|X - E(X)\right| \geq a.\sigma(X)\right) \leq \frac{1}{a^2}$$

Twierdzenie:

Prawdopodobieństwo odchylenia wartości zmiennej losowej od oczekiwanej E(X) o a-krotną wartość odchylenia standardowego jest mniejsze bądź równe $1/a^2$

Twierdzenie to jest słuszne dla wszystkich rozkładów, które mają wariancję a zatem i wartość oczekiwaną. Liczba a jest dowolną, dodatnią liczbą rzeczywistą.

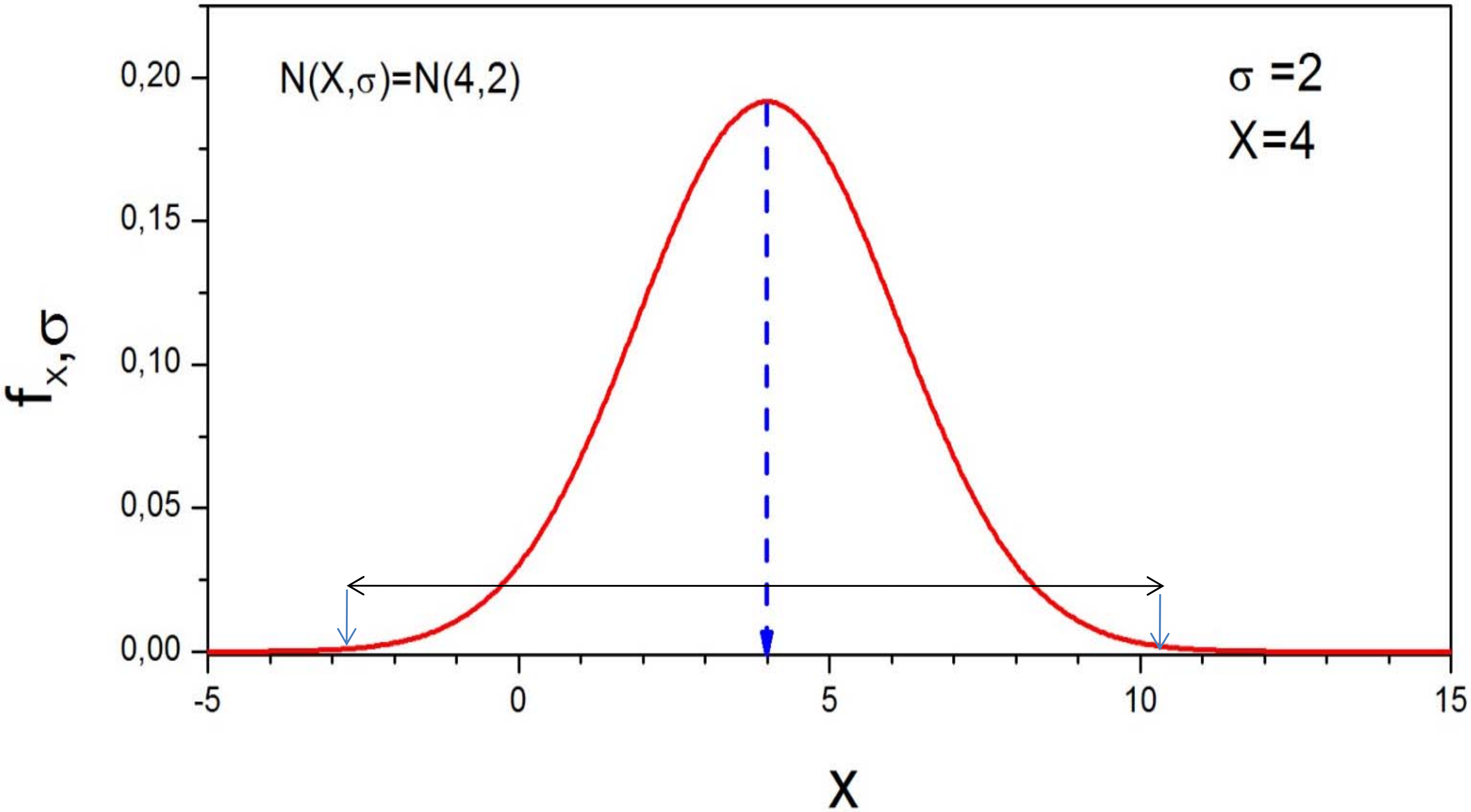
Wariancja jako miara rozproszenia



DUŻE ROZPROSZENIE

MNIEJSZE ROZPROSZENIE

Rozstęp jako miara rozproszenia



ROZSTĘP = x_{max} - x_{min}

Praktyczne sposoby obliczania wariancji

Wariancja z próby (n-elementowej):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right]$$

$$\bar{x} - średnia$$

Wariancja z populacji (N-elementowej):

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

$$\mu - średnia\ \ z\ populacji\ (oczekiwana\ )$$

Praktyczne sposoby obliczania odchylenia standardowego

Odchylenie standardowe próby (lub: błąd standardowy):

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

Odchylenie standardowe (populacji):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2}$$

Przykładowe rozkłady dla dyskretnej zmiennej losowej

Rozkład dwupunktowy (zero-jedynkowy), np. rzut monetą wylosowanie reszki (braku orła, porażki) x=0, wylosowanie orła (dobrego wyrobu, sukcesu) x=1, p - prawdopodobieństwo sukcesu, jego rozkład:

x_i	0	1
p_i	1-p	p

Dwumianowy (ang.binomial, Bernoulliego)

$$p_k = \binom{n}{k} \cdot p^k (1-p)^{n-k} \quad , k = 0,1,\ldots,n$$

gdzie 0<p<1; X={0, 1, 2, … k} k - liczba sukcesów w losowaniu n-krotnym ze zwracaniem

dla k=1 jest to rozkład dwupunktowy

Rozkład dwumianowy (Bernoulliego) - założenia

Eksperyment losowy składa się z *n* prób Bernoulliego, takich że:

1. **Każda próba jest niezależna od innych.**
2. **Każda próba może mieć tylko dwa wyniki: „sukces” i „porażkę” (binarne!).**
3. **Prawdopodobieństwo „sukcesu” wynosi *p* i jest wartością stałą.**

Pytamy o prawdopodobieństwo p_k zdarzenia, że zmienna losowa X będzie równa ilości otrzymanych k-sukcesów przy n próbach.

$$p_k = \binom{n}{k} \cdot p^k (1-p)^{n-k} \quad , \; k = 0, 1, \ldots, n$$

Trójkąt Pascala

W rozkładzie występuje symbol $\binom{n}{k} = \frac{n!}{(n-k)!k!}$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

dwumian Newtona

$n=0$	$\binom{0}{0}=1$						
$n=1$	$\binom{1}{0}=1$	$\binom{1}{1}=1$					
$n=2$	$\binom{2}{0}=1$	$\binom{2}{1}=2$	$\binom{1}{1}=1$				
$n=3$	$\binom{3}{0}=1$	$\binom{3}{1}=3$	$\binom{3}{2}=3$	$\binom{3}{3}=1$			
$n=4$	$\binom{4}{0}=1$	$\binom{4}{1}=4$	$\binom{4}{2}=6$	$\binom{4}{3}=4$	$\binom{4}{4}=1$		
$n=5$	$\binom{5}{0}=1$	$\binom{5}{1}=5$	$\binom{5}{2}=10$	$\binom{5}{3}=10$	$\binom{5}{4}=5$	$\binom{5}{5}=1$	
$n=6$	$\binom{6}{0}=1$	$\binom{6}{1}=6$	$\binom{6}{2}=15$	$\binom{6}{3}=20$	$\binom{6}{4}=15$	$\binom{6}{5}=6$	$\binom{6}{6}=1$

Trójkąt Pascala

						1							**n = 0**
					1	+	1						**n = 1**
				1		2		1					**n = 2**
			1		3		3		1				**n = 3**
		1		4		6		4		1			**n = 4**
	1		5		10		10		5		1		**n = 5**
1		6		15		20		15		6		1	**n = 6**

Rozkład Bernoulliego

Przykład 3.6

Prawdopodobieństwo, że w danym zakładzie produkcyjnym dzienne zużycie wody nie przekroczy pewnego ustalonego poziomu wynosi p=3/4. Monitorujemy zużycie wody w zakładzie przez 6 dni.

Wyznaczyć prawdopodobieństwo, że w ciągu 6 dni obserwacji, zużycie nie przekroczy ustalonego poziomu odpowiednio w 0, 1, 2, …, 6 dniach.

Tutaj sukcesem jest odpowiednie zużycie wody w jednym dniu.

Dane:

$$p = \frac{3}{4} \qquad q = \frac{1}{4} \qquad N = 6 \qquad k = 0, 1, \ldots, 6$$

Rozkład Bernoulliego

Do rozwiązania zadania wykorzystujemy właściwości dwumianu Newtona i trójkąt Pascala.

$$k = 0 \qquad P(k=0) = \binom{6}{0} \cdot \left(\frac{3}{4}\right)^0 \cdot \left(\frac{1}{4}\right)^6$$

$$k = 1 \qquad P(k=1) = \binom{6}{1} \cdot \left(\frac{3}{4}\right)^1 \cdot \left(\frac{1}{4}\right)^5$$

$$k = 2 \qquad P(k=2) = \binom{6}{2} \cdot \left(\frac{3}{4}\right)^2 \cdot \left(\frac{1}{4}\right)^4$$

$$k = 3 \qquad P(k=3) = \binom{6}{3} \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^3$$

$$k = 4 \qquad P(k=4) = \binom{6}{4} \cdot \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right)^2$$

$$k = 5 \qquad P(k=5) = \binom{6}{5} \cdot \left(\frac{3}{4}\right)^5 \cdot \left(\frac{1}{4}\right)^1$$

$$k = 6 \qquad P(k=6) = \binom{6}{6} \cdot \left(\frac{3}{4}\right)^6 \cdot \left(\frac{1}{4}\right)^0$$

Rozkład Bernoulliego

$$k = 0 \qquad P(0) = 1 \cdot 1 \cdot \frac{1}{4^6} \cong 0.00024$$

$$k = 1 \qquad P(1) = 6 \cdot \frac{3}{4} \cdot \frac{1}{4^5} = \frac{6 \cdot 3}{4^6} = 18 \cdot P(0) \cong 0.004$$

$$k = 2 \qquad P(2) = 15 \cdot \left(\frac{3}{4}\right)^2 \cdot \frac{1}{4^4} = \frac{15 \cdot 9}{4^6} = 135 \cdot P(0) \cong 0.033$$

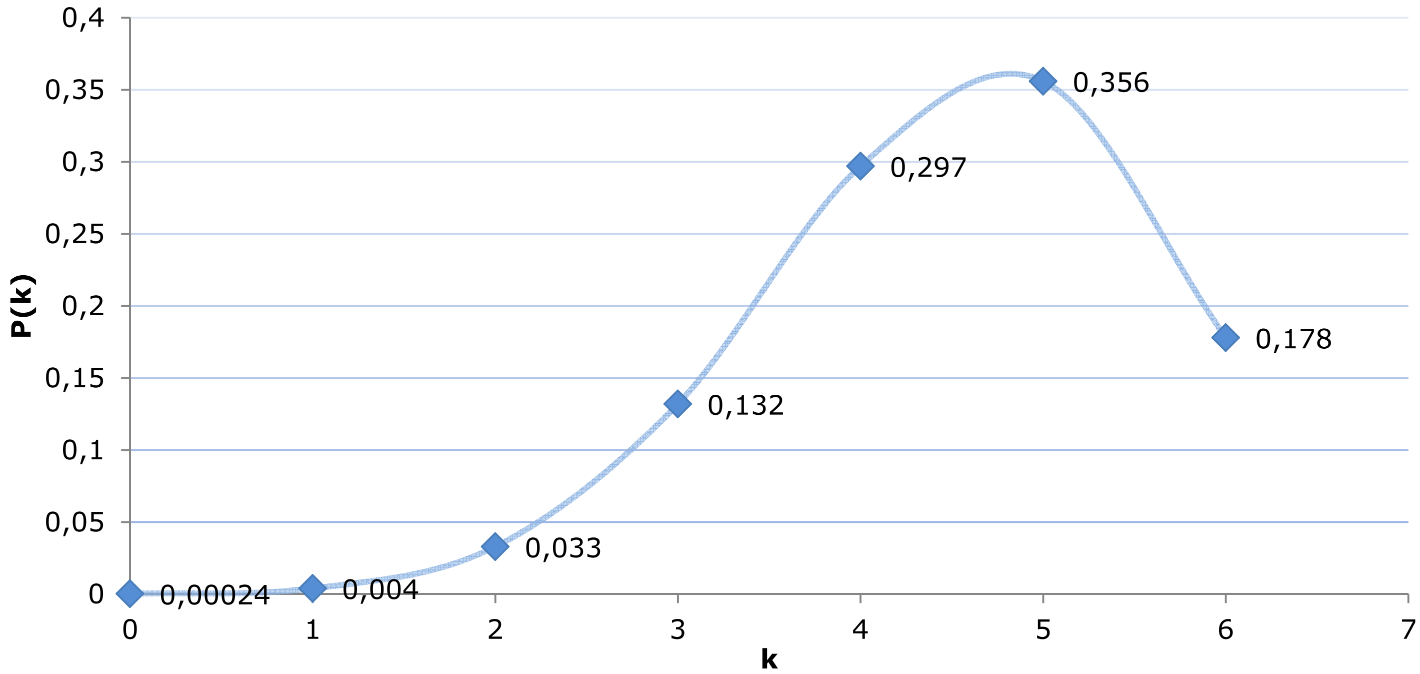
$$k = 3 \qquad P(3) = 20 \cdot \left(\frac{3}{4}\right)^3 \cdot \frac{1}{4^3} = \frac{20 \cdot 9 \cdot 3}{4^6} = 540 \cdot P(0) \cong 0.132$$

$$k = 4 \qquad P(4) = 15 \cdot \left(\frac{3}{4}\right)^4 \cdot \frac{1}{4^2} = \frac{15 \cdot 9 \cdot 9}{4^6} = 1215 \cdot P(0) \cong 0.297$$

$$k = 5 \qquad P(5) = 6 \cdot \left(\frac{3}{4}\right)^5 \cdot \frac{1}{4^1} = \frac{6 \cdot 9 \cdot 9 \cdot 3}{4^6} = 1458 \cdot P(0) \cong 0.356$$

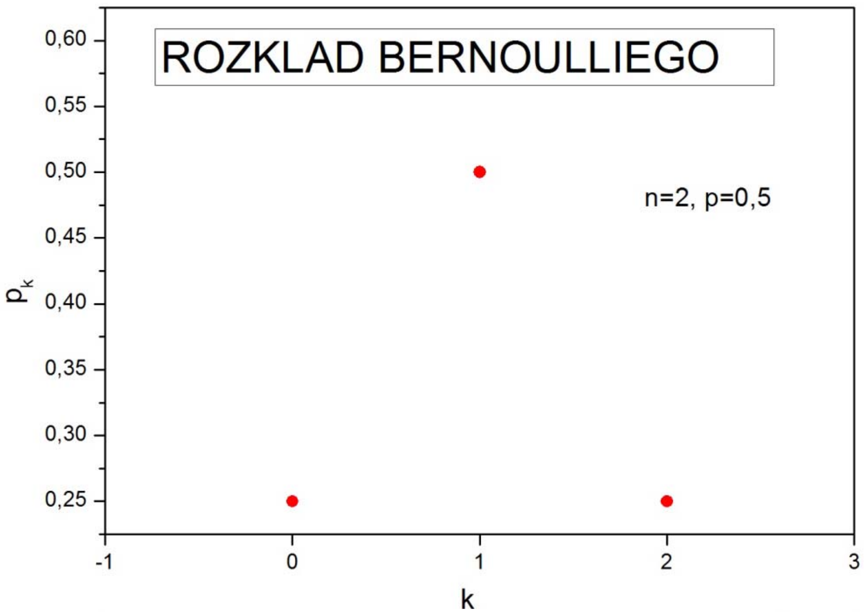
$$k = 6 \qquad P(6) = 1 \cdot \left(\frac{3}{4}\right)^6 \cdot \frac{1}{4^0} = \frac{9 \cdot 9 \cdot 9}{4^6} = 729 \cdot P(0) \cong 0.178$$

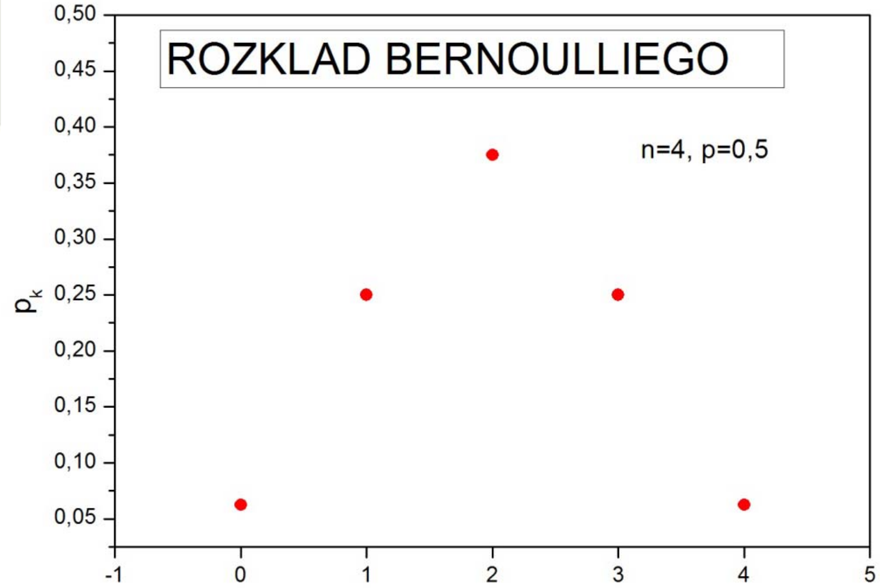
Rozkład Bernoulliego

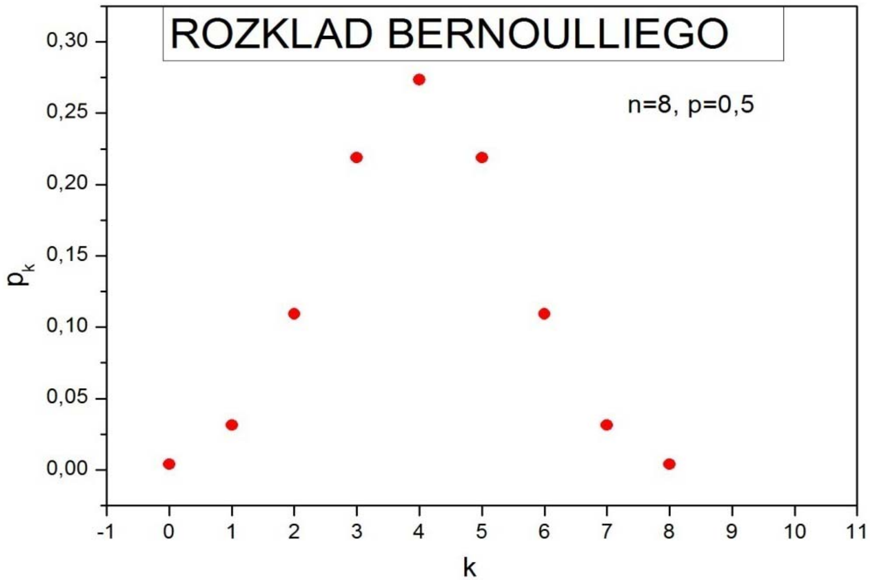


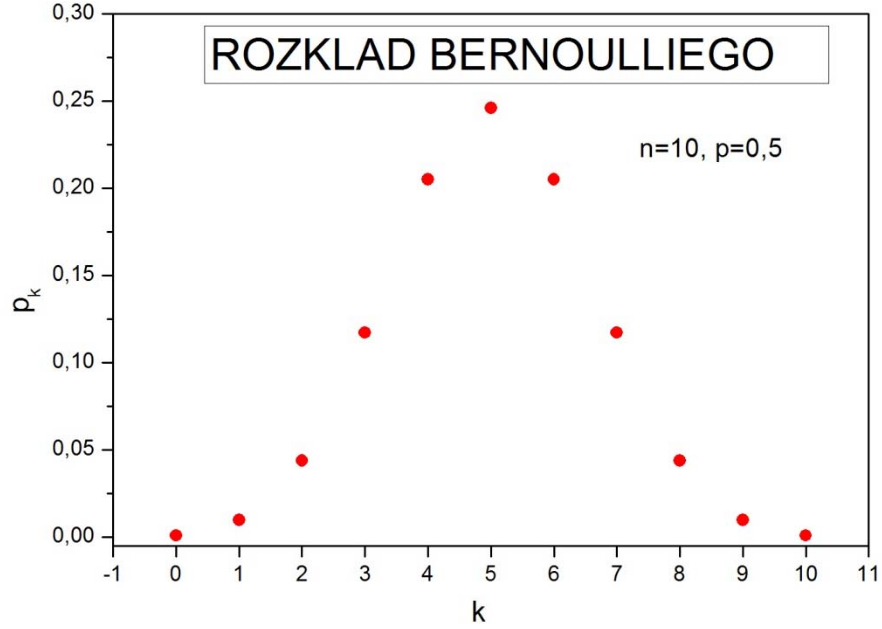
Największe prawdopodobieństwo uzyskujemy dla k=5 co oznacza, że prawdopodobieństwo zdarzenia, że poziom wody w zakładzie w ciągu 5 dni nie przekroczy ustalonego poziomu dziennego jest największe.

Rozkład Bernoulliego

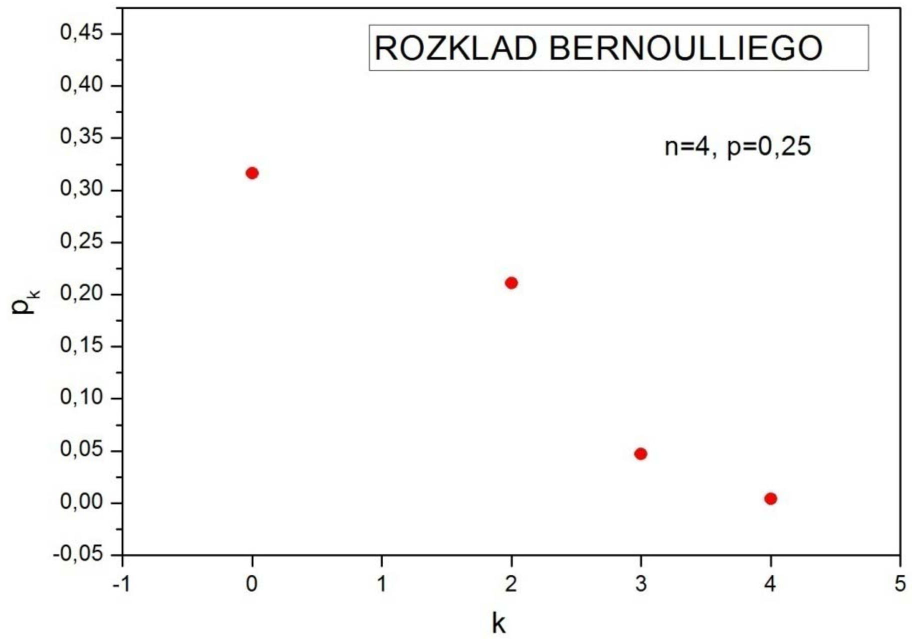


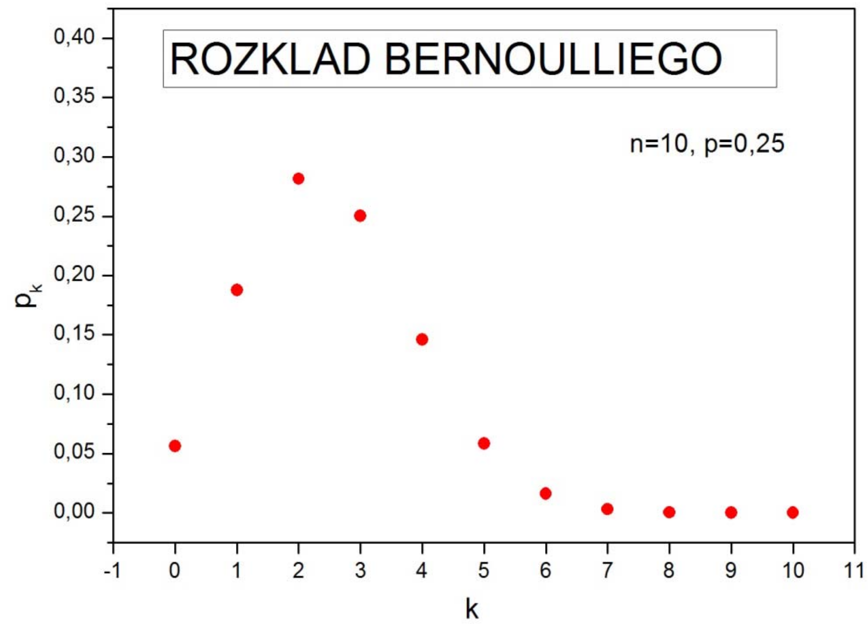






Rozkład Bernoulliego





Wartość oczekiwana w rozkładzie Bernoulliego

$$E(X) = \mu = np$$

Wariancja w rozkładzie Bernoulliego

$$V(X) = \sigma^2 = np(1-p)$$

Błędy w transmisji bitów

Przykład 3.7

Przy przesyłaniu informacji przez kanał cyfrowej transmisji zdarzają się błędy pojedynczych bitów. Załóżmy, że prawdopodobieństwo, że pojedynczy bit dotrze do konsumenta z błędem wynosi p=0,1 (i chociaż obiektywnie nie jest to sukces, to tutaj p nazwiemy prawdopodobieństwem sukcesu)

Załóżmy, że kolejne akty transmisji są niezależne. Niech X oznacza zmienną losową, której wartości są równe ilości bitów przesłanych z błędem, w sekwencji kolejnych 4 bitów.

Oznaczmy E błąd bitu, O brak błędu. Wynik transmisji OEOE oznacza, że drugi i czwarty bit są błędne, X=2; kolejność nie jest istotna czyli EEOO też odpowiada X=2

Błędy w transmisji bitów

Przykład 3.7 cd

Zdarzenie opisane zmienną losową X=2 to zbiór następujących wyników:

{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE}

Jakie jest prawdopodobieństwo P(X=2) zdarzenia, że dwa bity w sekwencji czterech zostaną przesłane z błędem?

Zdarzenia są niezależne więc

$P(EEOO)=P(E)P(E)P(O)P(O)=(0,1)^2\ (0,9)^2\ = 0,0081$

Zdarzenia są wzajemnie wykluczające i mają to samo prawdopodobieństwo wystąpienia więc

$P(X=2)=6\ P(EEOO)= 6\ (0,1)^2\ (0,9)^2\ = 6\ (0,0081)=0.0486$

Błędy w transmisji bitów

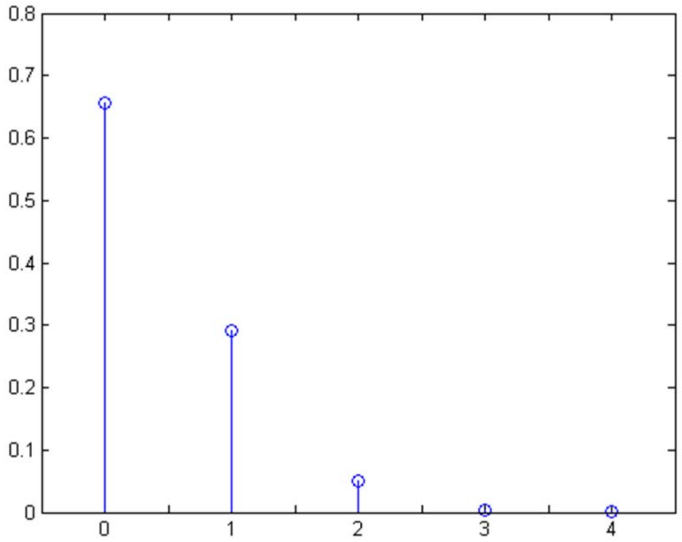
Przykład 3.7 cd

$$\binom{4}{2} = \frac{4\,!}{(2)!2!} = 6$$

A zatem P(X=2)=6 $(0,1)^2$ $(0,9)^2$ dane jest rozkładem Bernoulliego

$$P(X = x) = \binom{4}{x} \cdot p^x (1-p)^{4-x} \quad , \; x = 0,1,2,3,4, \; p = 0,1$$

P(X = 0) = 0,6561
P(X = 1) = 0,2916
P(X = 2) = 0,0486
P(X = 3) = 0,0036
P(X = 4) = 0,0001



Rozkład Poissona

Posłużmy się przykładem 3.7 transmisji n bitów przez kanał cyfrowy. Niech zmienna losowa X będzie przyjmowała wartości równe ilości bitów przesłanych z błędem.

Jeżeli prawdopodobieństwo p zdarzenia przesłania błędnego bitu jest stałe i kolejne akty transmisji są niezależne, to X ma rozkład dwumianowy (Bernoulliego).

Wprowadźmy parametr $\lambda=pn$ (E(X) dla tego rozkładu równa się λ)

$$P(X=x)=\binom{n}{x}\cdot p^{x}(1-p)^{n-x}=\binom{n}{x}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

Rozkład Poissona

Załóżmy, że n wzrasta a p maleje tak, że λ=pn pozostaje stałe. Rozkład przechodzi w rozkład Poissona.

$$\lim_{n\to\infty} P(X=x) = \lim_{n\to\infty} \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}$$

Ze względu na to, że liczba przesyłanych bitów zmierza do nieskończoności, liczba błędów może być równa dowolnej nieujemnej liczbie całkowitej. Zakres możliwych wartości X sięga od 0 do ∞

Rozkład Poissona stosujemy pod pewnymi warunkami dla zmiennej losowej X, która jest równa liczbie zdarzeń (zliczeń) w danym przedziale (przy podziale na podprzedziały) w eksperymencie losowym zwanym procesem Poissona.

Proces Poissona

Załóżmy, że dany przedział liczb rzeczywistych może być <u>podzielony na podprzedziały o małej długości</u> takiej że:

1. Prawdopodobieństwo więcej niż jednego zliczenia w tym podprzedziale jest równe zero.
2. Prawdopodobieństwo jednego zliczenia w podprzedziale jest **takie samo** dla wszystkich podprzedziałów i **proporcjonalne do jego długości**
3. Zliczanie w każdym podprzedziale jest **niezależne** od innych podprzedziałów

   Eksperyment losowy które spełnia te warunki nazywamy procesem Poissona. Zmienną losową X która jest równa liczbie zliczeń w przedziale nazywamy zmienną losową Poissona. Funkcja gęstości prawdopodobieństwa f(x) jest zależna od parametru λ

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}\ \text{, gdzie } x = 0,1,2,\ldots$$

Rozkład Poissona

Przykład 3.8

Podczas inspekcji cienkiego miedzianego przewodnika stwierdzono występowanie uszkodzeń. Oznaczmy przez X zmienną losową równą liczbie uszkodzeń (zliczeń) na długości L przewodnika i załóżmy, że średnia liczba uszkodzeń na całej długości wynosi λ. Należy znaleźć funkcję gęstości prawdopodobieństwa zmiennej X.

Dzielimy długość L (kilka milimetrów) na n podprzedziałów o bardzo małej długości np. 1 mikrometr.

- prawdopodobieństwo, że na tym podprzedziale wystąpi więcej niż jedno uszkodzenie, jest zaniedbywalnie małe
- Założenie, że uszkodzenia są losowe pozwala przyjąć, że na każdym podprzedziale prawdopodobieństwo uszkodzenia jest takie samo i wynosi p
- Zakładamy, niezależność zdarzeń na podprzedziałach

Rozkład Poissona

Przykład 3.8

Można w tym przykładzie zatem modelować rozkład zmiennej losowej X rozkładem dwumianowym:

$$E(X) = \lambda = np$$

czyli

$$p = \frac{\lambda}{n}$$

Prawdopodobieństwo, że podprzedział zawiera wadę wynosi λ/n i gdy n jest bardzo duże, p jest bardzo małe. Rozkład uszkodzeń to rozkład Poissona.

Rozkład Poissona

Rozkład Poissona to jeden z nielicznych rozkładów, w którym wartość oczekiwana jest równa wariancji:

$$E(X) = np = \lambda$$

Z wariancji w rozkładzie Bernoulliego $V(X) = \sigma^2 = np(1-p)$

przy dużym n i małym p, otrzymujemy

$$V(X) = \sigma^2 = \lim_{n\to\infty, p\to 0}(np - np^2) = np = \lambda$$

czyli wariancję w rozkładzie Poissona.

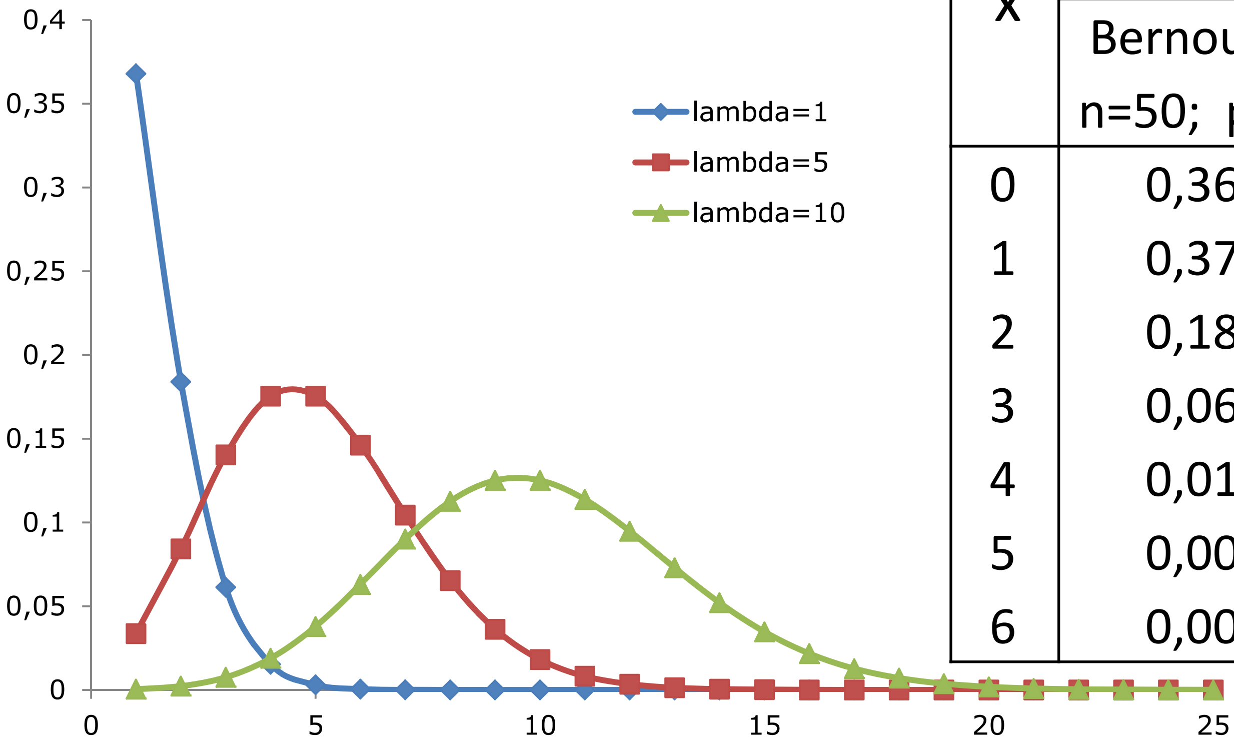
Rozkład Poissona

Rozkład Poissona ma wiele zastosowań zwłaszcza w eksperymentach fizyki jądrowej i atomowej, np. rozpadach jąder atomowych, aktach emisji cząstek, itp. Przedział, o którym mówiliśmy może być przedziałem czasu (często), wycinkiem powierzchni, elementem objętości. Rozkład może być stosowany do systemów z dużą liczbą możliwych zdarzeń, z których każde jest bardzo rzadkie (**prawo rzadkich zdarzeń**).

Przykłady zdarzeń, które mogą być modelowane rozkładem Poissona:

- Historyczne – liczba zabitych przez kopnięcie konia każdego roku w korpusie kawalerii w Prusach (W.Bortkiewicz 1868-1931)
- Liczba połączeń telefonicznych przychodzących do centrali na minutę
- Liczna mutacji w danym odcinku DNA po ekspozycji na pewną dawkę promieniowania
- Odsetek komórek, które zostaną zakażone dla danej liczebności zakażeń
- W elektronice szum Poissona (śrutowy); ziarnistość przy powiększaniu fotografii, zastosowania molekularne

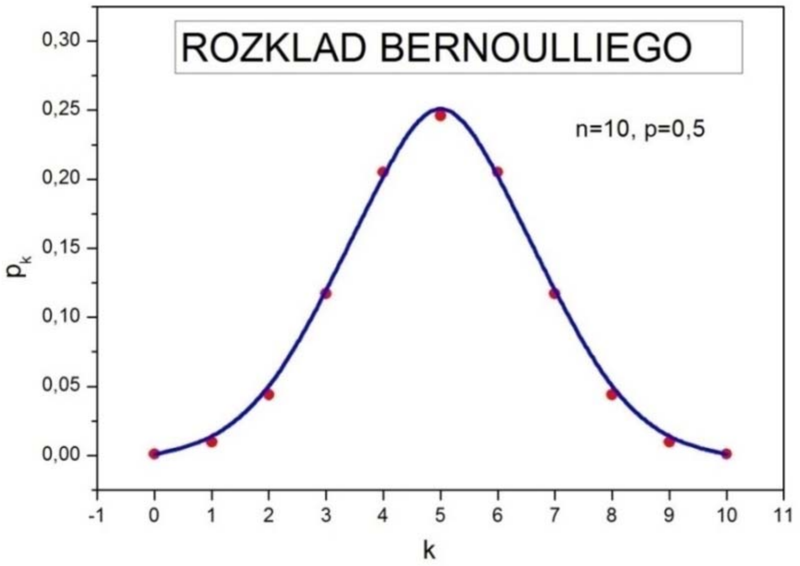
Rozkład Poissona



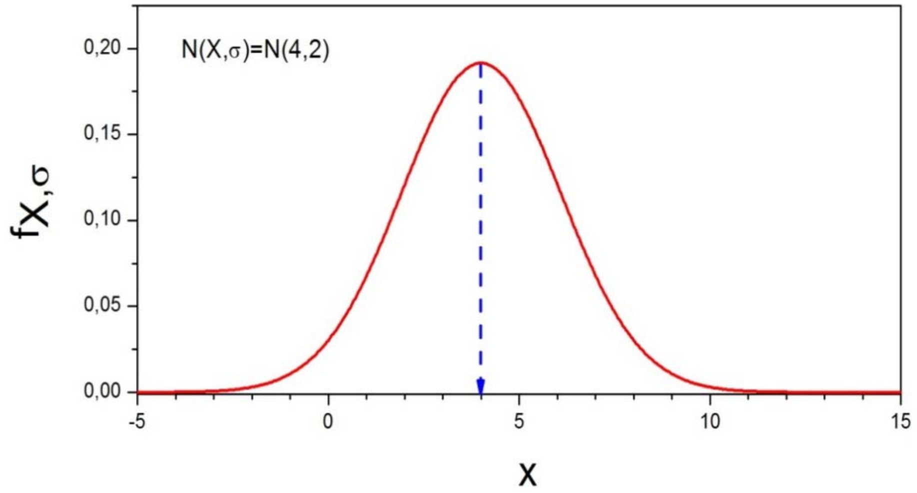
x	Funkcja rozkładu	
	Bernoulliego n=50; p=0,02	Poissona: λ=1
0	0,364	0,368
1	0,372	0,368
2	0,186	0,184
3	0,061	0,061
4	0,014	0,015
5	0,003	0,003
6	0,000	0,001

Rozkład dyskretny o nieskończonej liczbie wartości (x- dowolna liczba całkowita $x \geq 0$. Dla dużych n rozkład Bernoulliego `upodabnia się` do rozkładu Poissona

Rozkład normalny (Gaussa)



ROZKŁAD GRANICZNY
(rozkład normalny)



Najczęściej spotykanym rozkładem zmiennej losowej jest rozkład normalny (zwany rozkładem Gaussa).

Centralne twierdzenie graniczne sformułowane po raz pierwszy w 1733 r. przez de Moivre'a.

Jeżeli powtarzamy wielokrotnie eksperyment losowy, rozkład zmiennej losowej, będącej średnią (lub sumą) wszystkich wyników zmierza do rozkładu normalnego przy bardzo dużej liczbie powtórzeń eksperymentu.

Rozkład normalny (Gaussa)

Zmienna losowa X charakteryzująca się funkcją gęstości prawdopodobieństwa f(x) daną wzorem:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu^{2}}{2\sigma^{2}}\right] \text{, gdzie} -\infty < x < +\infty$$

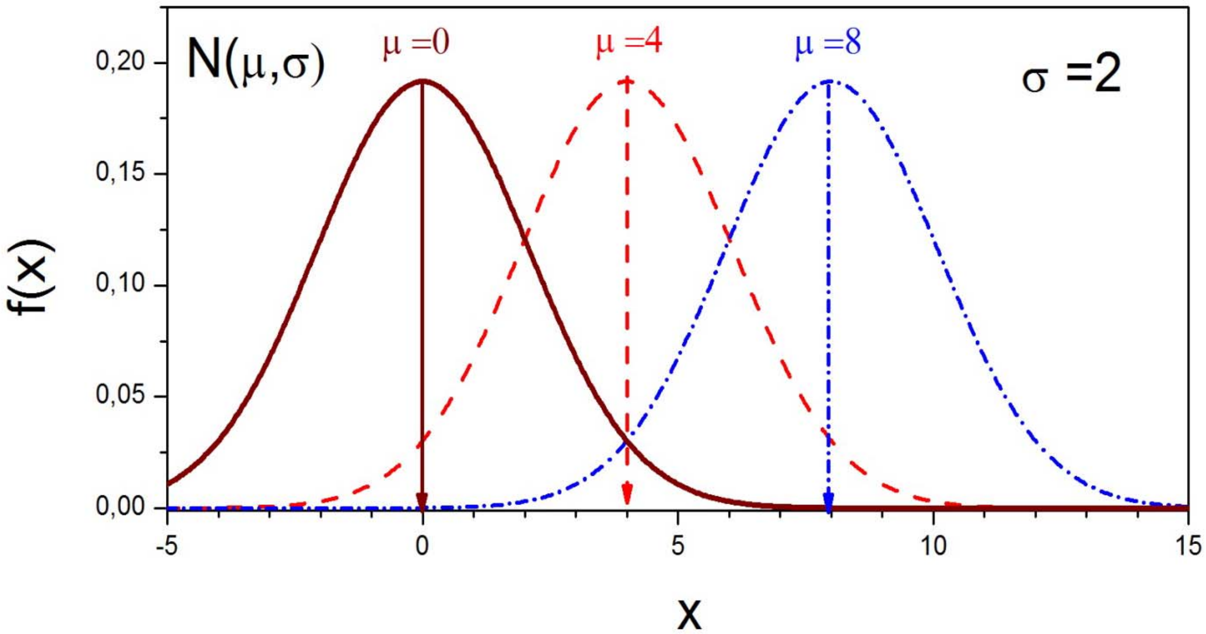
nazywana jest zmienną o rozkładzie normalnym i tylko dwóch parametrach

$$-\infty < \mu < +\infty, \quad \sigma > 1$$

Można pokazać, że wartość oczekiwana E(X)=μ a wariancja V(X)=σ^2

Używa się zapisu N(μ,σ)

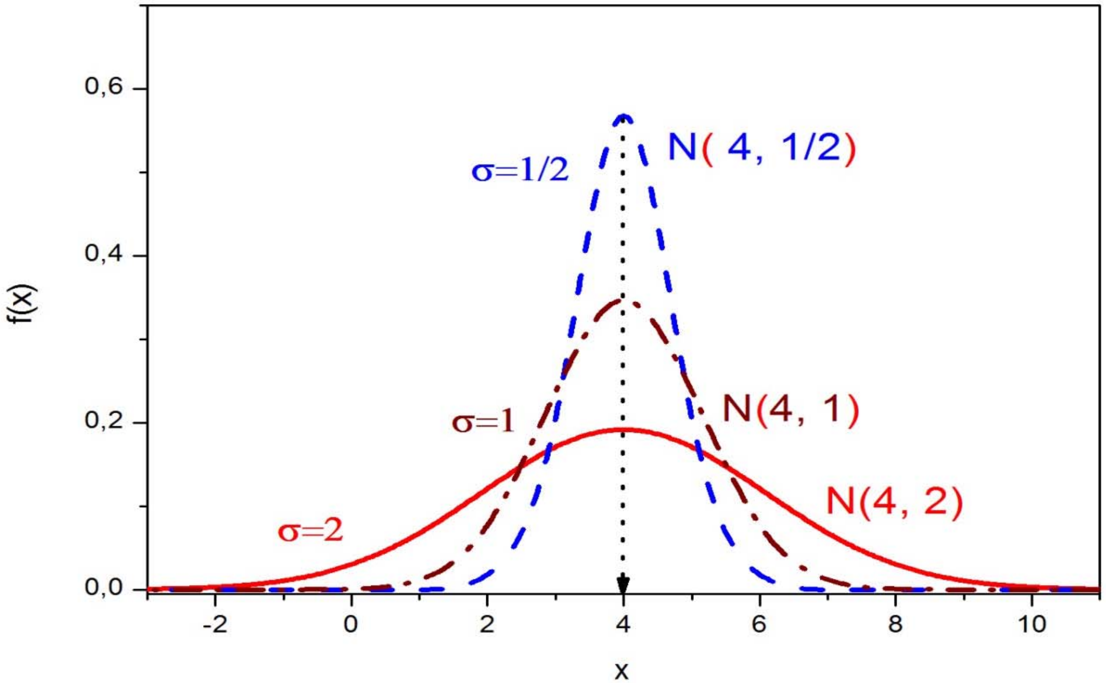
Rozkład normalny (Gaussa)



Wartość oczekiwana, położenie maksimum gęstości prawdopodobieństwa (moda) i mediana pokrywają się ($x=\mu$). Rozkład jest symetryczny (krzywa Gaussa = krzywa dzwonowa).

Wariancja jest miarą szerokości rozkładu. Punkty o współrzędnych $x=+\sigma$ i $x=-\sigma$ są punktami przegięcia.

Rozkład normalny (Gaussa)



Rozkład normalny jest rozkładem błędów przypadkowych i wyników wielu eksperymentów fizycznych. Miarą błędu pomiaru jest odchylenie standardowe σ. Pomiar o większym σ charakteryzuje się większym rozrzutem wyników wokół wartości średniej a zatem mniejszą precyzją.

Standardowy rozkład normalny

Zmienna losowa Z charakteryzująca się funkcją gęstości prawdopodobieństwa N(z) daną wzorem:

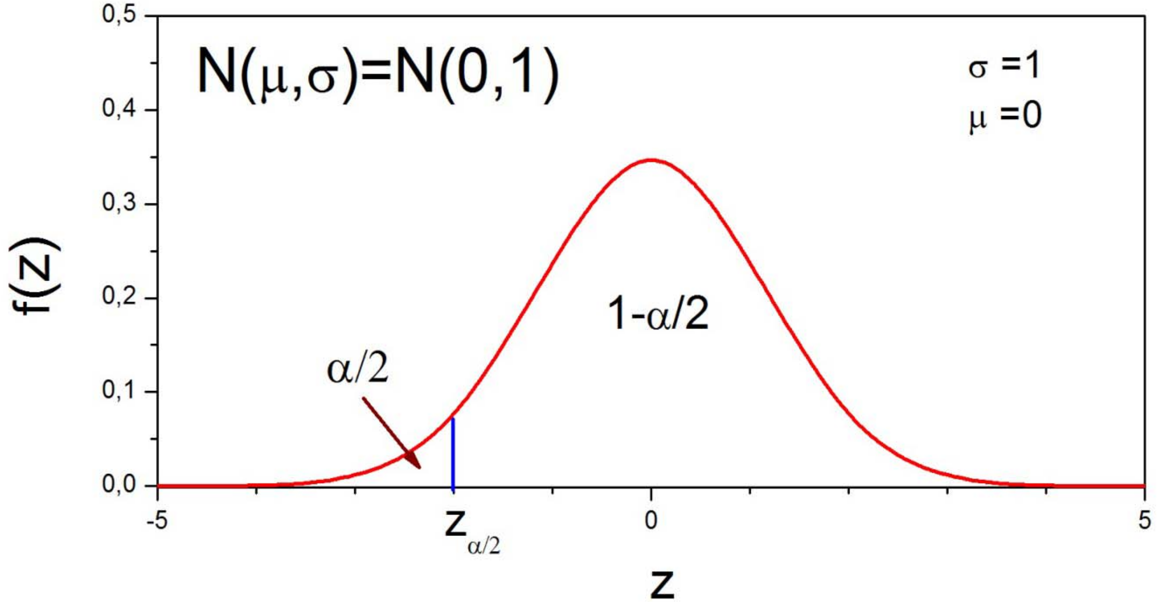
$$N(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right], \text{ gdzie } -\infty < z < +\infty$$

nazywana jest zmienną standaryzowaną tj. o standardowym rozkładzie normalnym N(0,1)

$$E(Z) = 0, \quad V(Z) = 1$$

Definicja zmiennej standardowej $Z = \dfrac{X - \mu}{\sigma}$

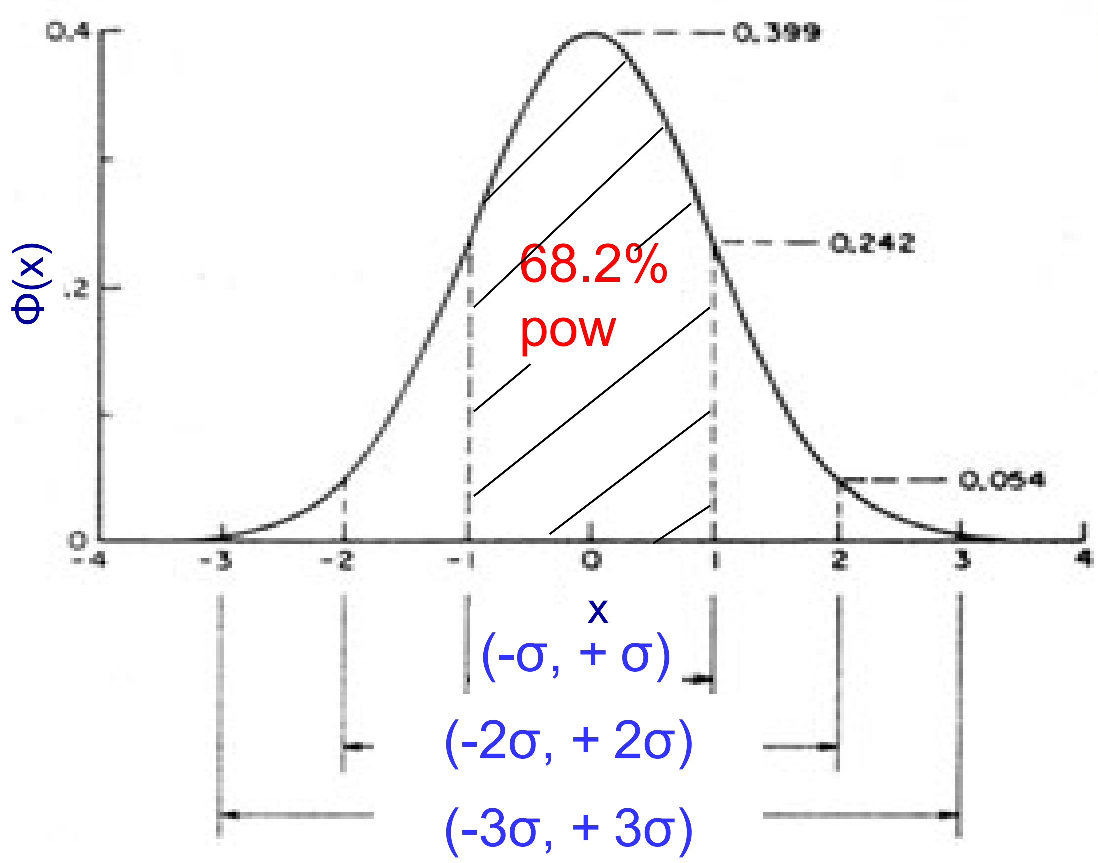
Standardowy rozkład normalny



KORZYŚCI STANDARYZACJI:

- Stwarza możliwość tablicowania funkcji gęstości prawdopodobieństwa i dystrybuanty dla N(0,1). Można stworzyć zmienną o rozkładzie N(μ,σ) przez prostą transformację $X = \sigma * Z + \mu$
- Przez standaryzację sprowadzamy wszystkie wartości oryginalnej zmiennej losowej do obszaru w pobliżu zera a jednostką jest odchylenie standardowe. Dzięki temu można porównywać rozkłady wielkości różniące się znacznie położeniem centrum i skalą wartości

Obliczanie prawdopodobieństwa w rozkładzie Gaussa



$P(\mu-\sigma < X < \mu+\sigma) = 0{,}6827$ (około 2/3 wyników),

$P(\mu-2\sigma < X < \mu+2\sigma) = 0{,}9545$

$P(\mu-2\sigma < X < \mu+2\sigma) = 0{,}9973$ (prawie wszystkie)

Przykład 3.9

Seria wyników (próba) $x_1, x_2, \ldots. x_n$ obarczonych niepewnością przypadkową jest duża gdy $30 < n \leq 100$. W próbie takiej wyniki się powtarzają: n_k jest liczbą pomiarów, w których wystąpił wynik x_k,

n_k/n jest częstością występowania wyniku

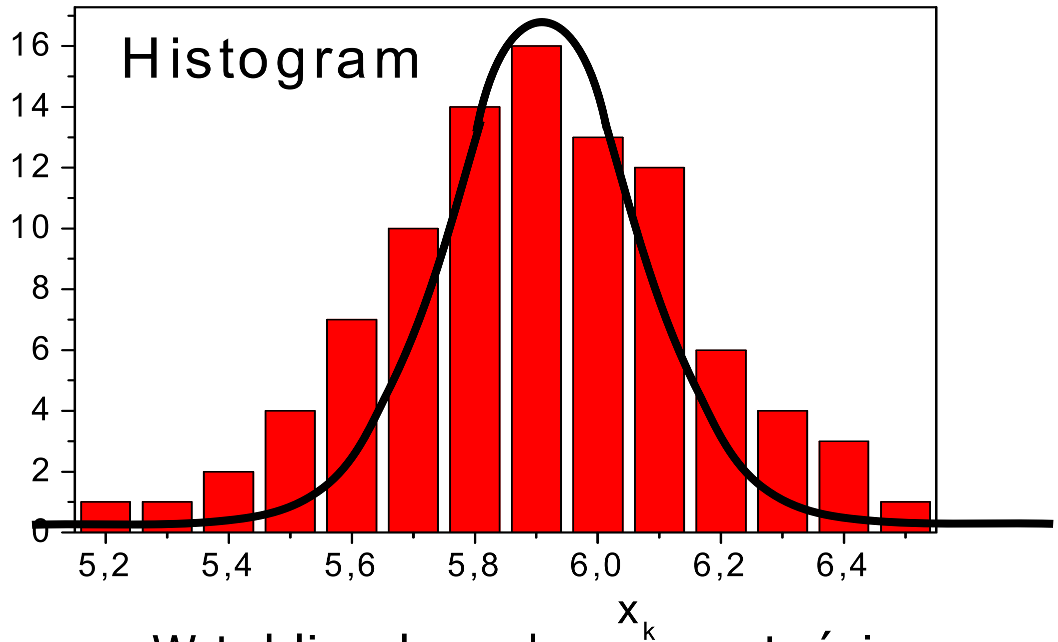
x_k	n_k	n_k/n
5,2	1	0,011
5,3	1	0,011
5,4	2	0,021
5,5	4	0,043
5,6	7	0,075
5,7	10	0,106
5,8	14	0,149
5,9	16	0,170
6,0	13	0,138
6,1	12	0,128
6,2	6	0,064
6,3	4	0,043
6,4	3	0,032
6,5	1	0,011
Suma	94	

Opracowanie serii pomiarów bezpośrednich dużej próby

Średnia arytmetyczna – estymator wartości oczekiwanej

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = 5{,}9$$



Estymator odchylenia standardowego

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

σ=0,2

W tablicach szukamy wartości N(0,1) dla zmiennej Z; porównujemy z histogramem

$$u(\bar{x}) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$$

Tworzymy zmienną standardową Z o wartościach z_i

$$z_i = \frac{x_i - 5{,}9}{0{,}2}$$

Centralne twierdzenie graniczne – intuicyjne sformułowanie dla wielu zmiennych losowych

Zmienna Z będąca standaryzowaną sumą niezależnych zmiennych losowych będzie miała standardowy rozkład normalny gdy liczba składników w sumie dąży do nieskończoności oraz w sumie nie występują zmienne o wariancjach dominujących w stosunku do reszty składników.

To twierdzenie powoduje, że rozkład normalny jest wyróżnionym rozkładem