Wybrane rozkłady zmiennych losowych

Statystyka

Rozkład dwupunktowy

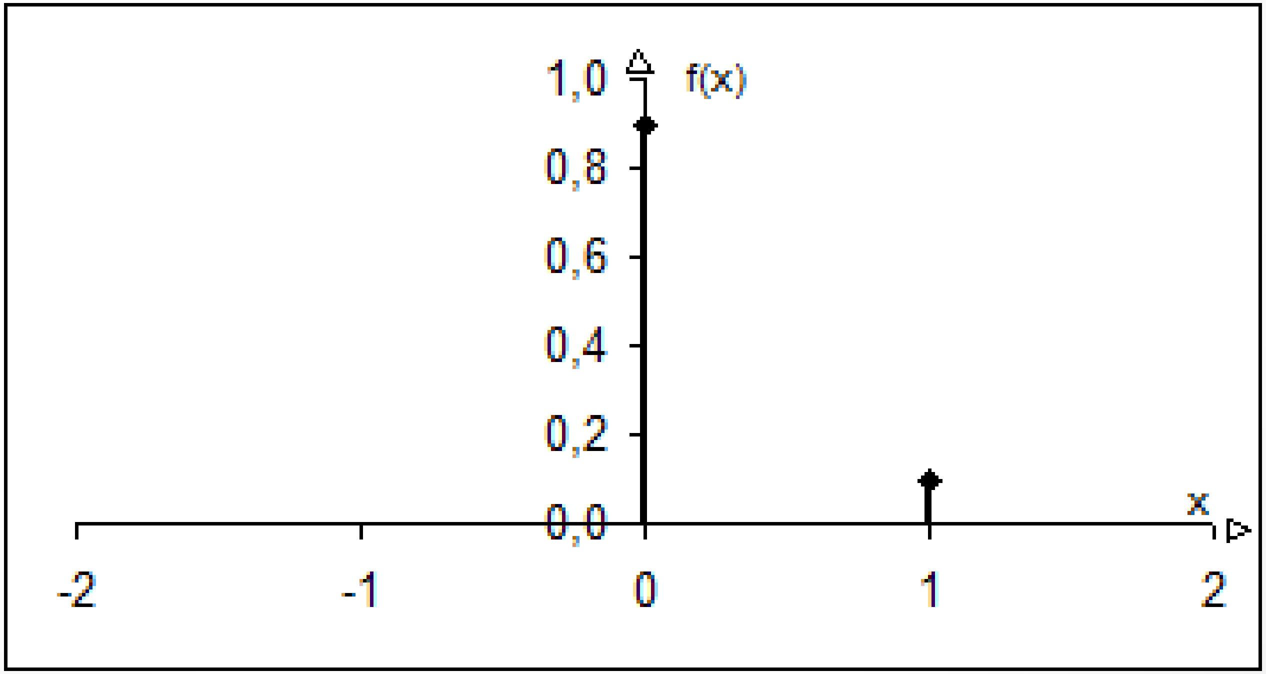
Zmienna losowa przyjmuje tylko dwie wartości:
wartość 1 z prawdopodobieństwem p
i wartość 0 z prawdopodobieństwem 1- p

x_i	p_i
0	1-p
1	p
suma	1

Rozkład dwupunktowy

Funkcja rozkładu prawdopodobieństwa

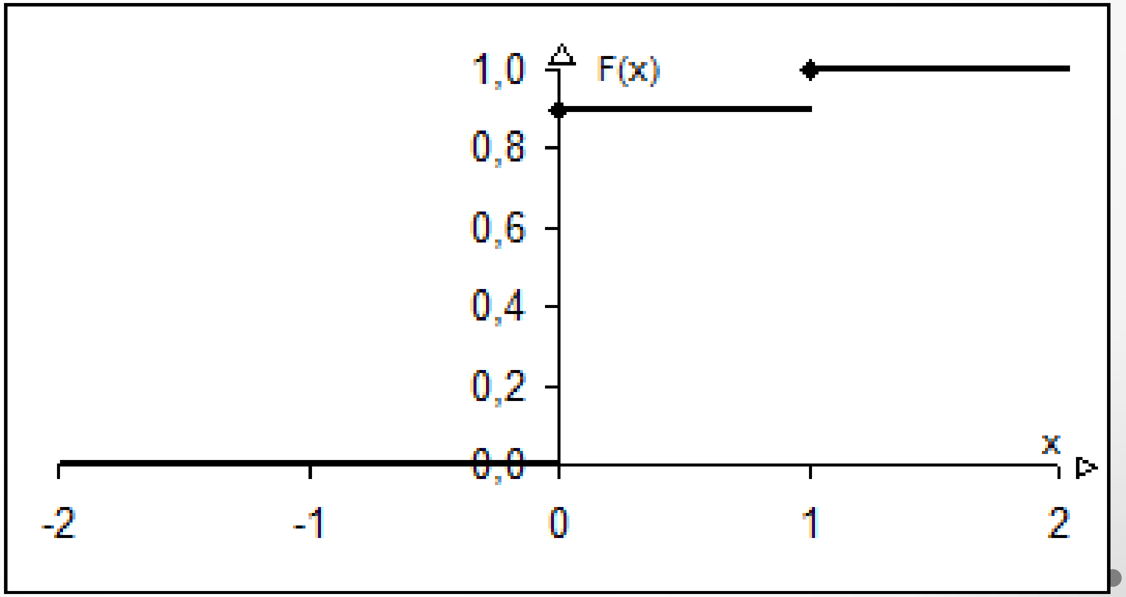
$$f(x) = \begin{cases} 1-p & dla \quad x = 0 \\ p & dla \quad x = 1 \end{cases} \quad \text{oraz} \quad 0 < p < 1$$



Rozkład dwupunktowy

Dystrybuanta

$$F(x)=\begin{cases} 0 & dla \quad \mathrm{x}<0 \\ 1-p & dla \quad 0\le \mathrm{x}<1 \\ 1 & dla \quad \mathrm{x}\ge 1 \end{cases}$$

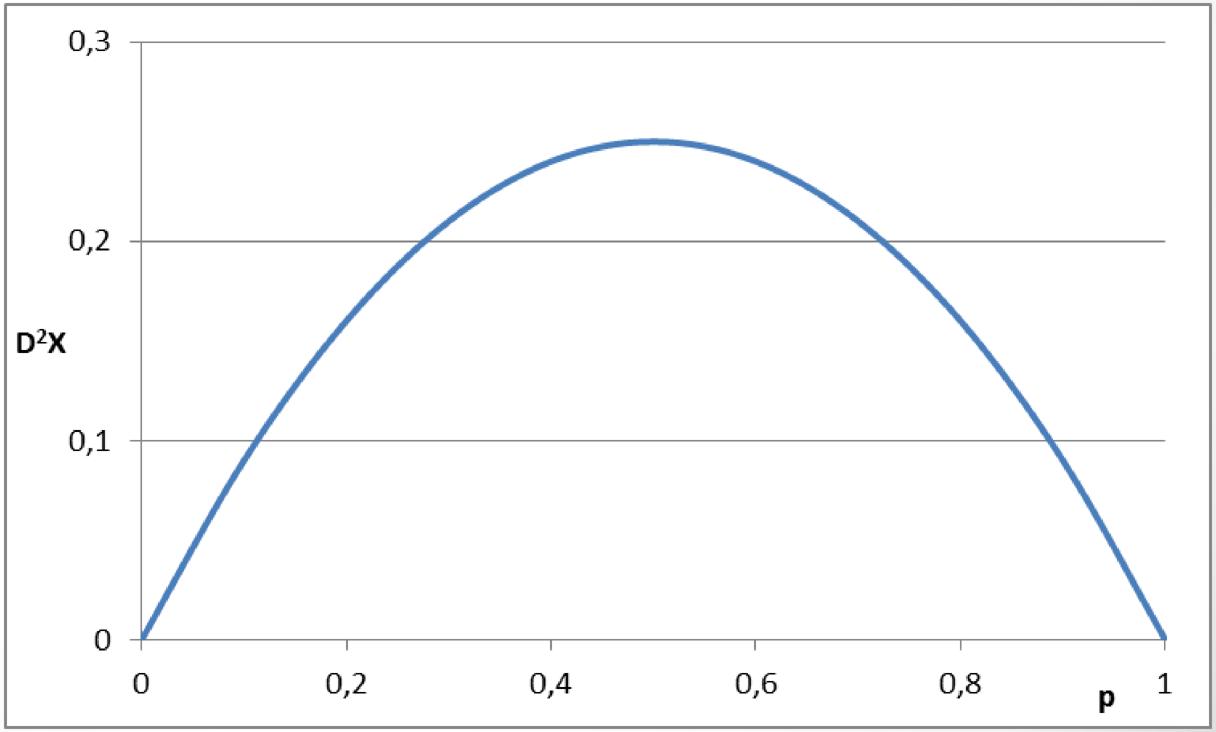


Rozkład dwupunktowy

Parametry

$$EX = p$$

$$D^2X = p(1-p)$$



Rozkład dwupunktowy

Przykładem cechy opisywanej przy pomocy rozkładu dwupunktowego może być:
stan zdrowia (zdrowy i chory),
płeć (samiec czy samica),
przeżywalność (żyje lub nie żyje),
rogatość (rogaty czy bezrożny)
i wiele innych.

Przykładem może być również para alleli w *locus*, przy założeniu istnienia tylko dwóch alleli

Rozkład dwumianowy (Bernoulli'ego)

Wykonujemy n niezależnych doświadczeń w nie zmienionym schemacie. W każdym z doświadczeń może pojawić się sukces („1”) z prawdopodobieństwem p albo porażka („0”) z prawdopodobieństwem 1-p. Wartościami zmiennej losowej jest liczba sukcesów k = 0,1,...,n uzyskanych w takiej serii doświadczeń. Rozkład dwumianowy jest zatem sumą n zmiennych zero-jedynkowych.

Rozkład dwumianowy (Bernoulli'ego)

Prawdopodobieństwo każdej wartości zmiennej losowej obliczamy posługując się wzorem:

$$P_{n,p}(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \qquad \text{gdy } 0<p<1 \text{ oraz } k=0,1\ldots n$$

Trójkąt Pascala

$\binom{n}{k}$

										1										
									1		1									
								1		2		1								
							1		3		3		1							
						1		4		6		4		1						
					1		5		10		10		5		1					
				1		6		15		20		15		6		1				
			1		7		21		35		35		21		7		1			
		1		8		28		56		70		56		28		8		1		
	1		9		36		84		112		112		84		36		9		1	
1		10		45		120		196		224		196		120		45		10		1

Przykład Rozkład dwumianowy (Bernoulli'ego)

Wiadomo, że wśród cieląt 70% jest odpornych na dany rodzaj infekcji. Niech zmienną losową będzie liczba odpornych spośród czterech cieląt.

$$P(X=0)=\binom{4}{0}\cdot(0{,}7)^0\cdot(0{,}3)^4=1\cdot1\cdot0{,}0081=0{,}0081$$

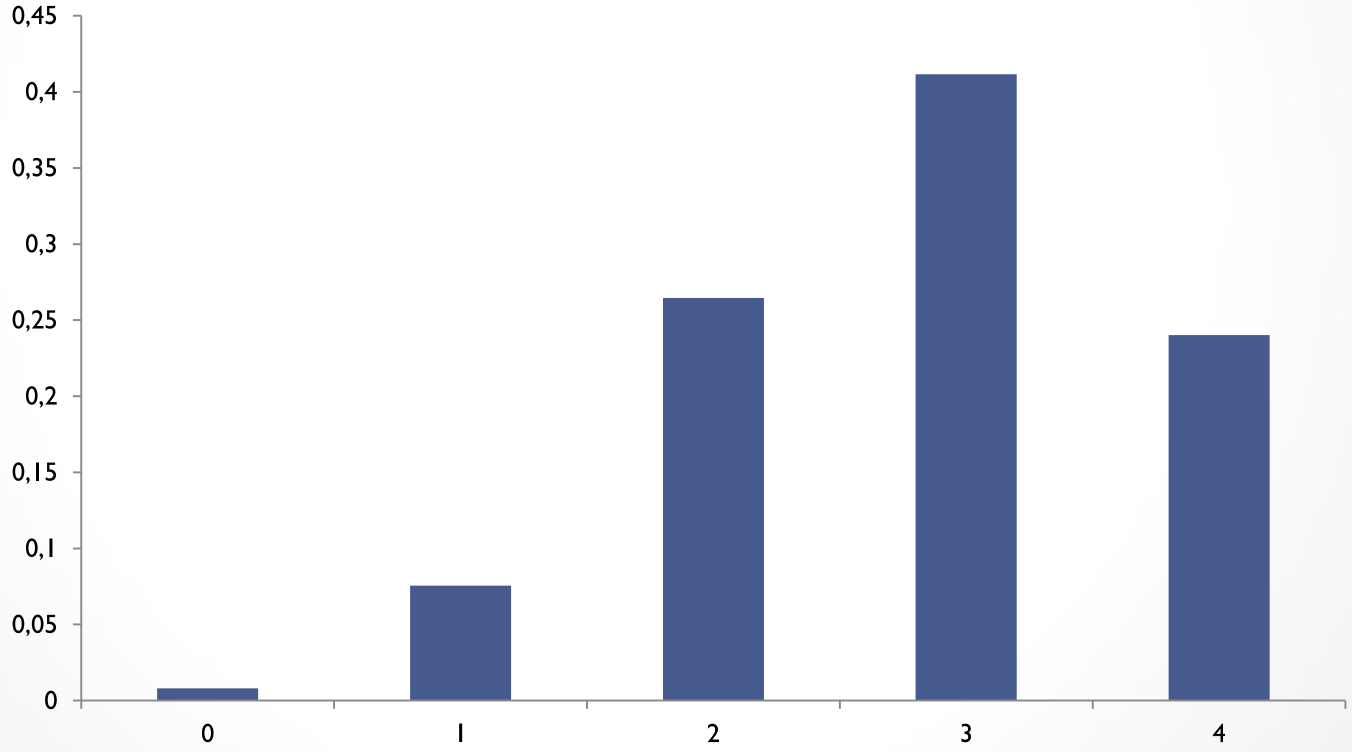
$$P(X=1)=\binom{4}{1}\cdot(0{,}7)^1\cdot(0{,}3)^3=4\cdot0{,}7\cdot0{,}027=0{,}0756$$

$$P(X=2)=\binom{4}{2}\cdot(0{,}7)^2\cdot(0{,}3)^2=6\cdot0{,}49\cdot0{,}09=0{,}2646$$

$$P(X=3)=\binom{4}{3}\cdot(0{,}7)^3\cdot(0{,}3)^1=4\cdot0{,}343\cdot0{,}3=0{,}4116$$

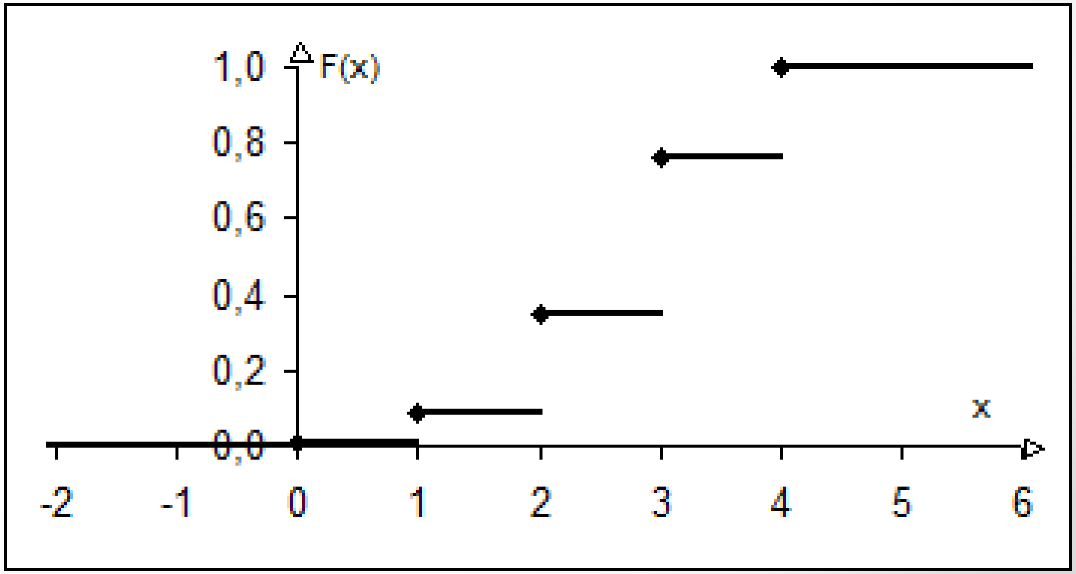
$$P(X=4)=\binom{4}{4}\cdot(0{,}7)^4\cdot(0{,}3)^0=1\cdot0{,}2401\cdot1=0{,}2401$$

Rozkład dwumianowy (Bernoulli'ego)



Rozkład dwumianowy (Bernoulli'ego)

$$F(x) = \begin{cases} 0 & dla \quad x < 0 \\ 0{,}0081 & dla \quad 0 \le x < 1 \\ 0{,}0837 & dla \quad 1 \le x < 2 \\ 0{,}3483 & dla \quad 2 \le x < 3 \\ 0{,}7599 & dla \quad 3 \le x < 4 \\ 1 & dla \quad x \ge 4 \end{cases}$$



Rozkład dwumianowy (Bernoulli'ego)

Parametry zmiennej o tym rozkładzie

$$EX = n \cdot p \qquad D^2X = n \cdot p \cdot (1 - p)$$

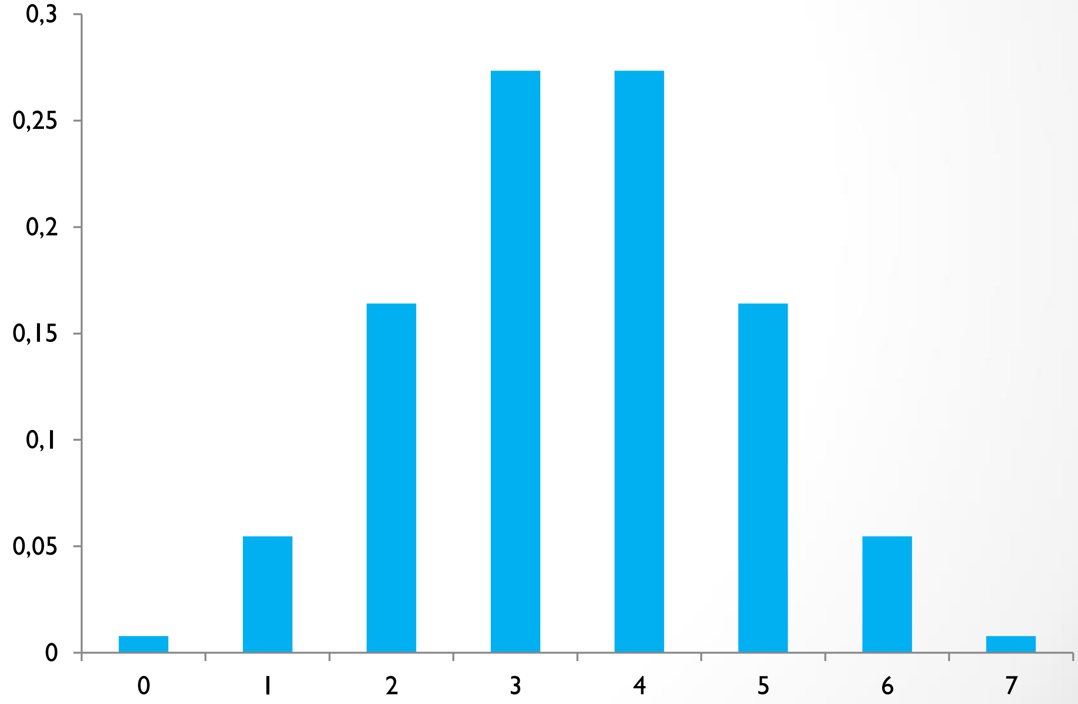
$$\boxed{EX = 4 \cdot 0{,}7 = 2{,}8}$$

$$\boxed{D^2X = 4 \cdot 0{,}7 \cdot (1 - 0{,}7) = 0{,}84}$$

Rozkład dwumianowy (Bernoulli'ego)

p=0,5
n=7

x_i	p_i
0	0,0078125
1	0,0546875
2	0,1640625
3	0,2734375
4	0,2734375
5	0,1640625
6	0,0546875
7	0,0078125



Rozkład jest symetryczny

Rozkład geometryczny

Rozkład geometryczny – wynik ciągu niezależnych doświadczeń, które powtarzane są tak długo, aż pojawi się sukces z prawdopodobieństwem p. Seria składa się zatem z (k + 1) doświadczeń, w tym k - porażek i jednego sukcesu. Wartością zmiennej losowej jest liczba doświadczeń poprzedzających sukces, (tzn. czas oczekiwania na sukces). Zmienna losowa ma nieskończenie wiele wartości.

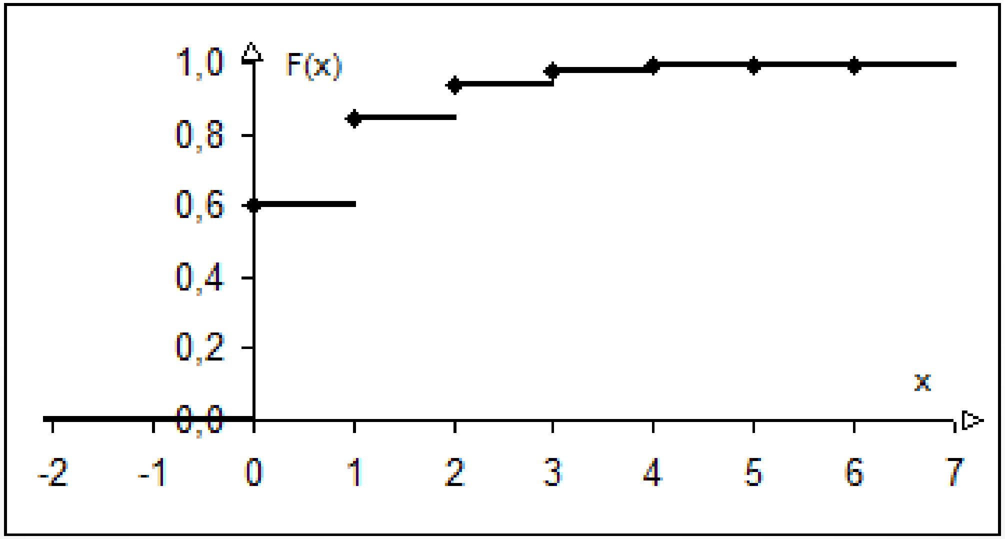
$$P_p(k) = (1-p)^k p \qquad \text{oraz} \;\; 0 < p < 1 \quad k = 0, 1, 2...$$

Przykład Rozkład geometryczny

Załóżmy, że każdorazowe szczepienie powoduje odporność u 60% zwierząt, a efekty kolejnych szczepień są niezależne. Ile razy powinno się szczepić cielęta, aby uzyskać co najmniej 95% odpornych zwierząt?

x_i	$P(X=x_i)$
0	0,6
1	0,24
2	0,096
3	0,0384
4	0,01536
5	0,00614
...	...
...	...
...	...

x	F(x)
$(-\infty ; 0)$	0
$\langle 0 ; 1)$	0,6
$\langle 1 ; 2)$	0,84
$\langle 2 ; 3)$	0,936
$\langle 3 ; 4)$	0,9744
$\langle 4 ; 5)$	0,98976
$\langle 5; 6)$	0,9959
...	...
...	...



Rozkład Poisson'a

Rozkład Poisson'a jest rozkładem granicznym dla ciągu zmiennych losowych mających rozkład dwumianowy. Wraz ze wzrostem długości serii (n) maleje prawdopodobieństwo sukcesu (p), tak, że np = const.

$$np \to \lambda \quad \text{tzn.} \quad \lim_{n \to \infty} P_{n,p}(k) = P_\lambda(k)$$

Dla rozkładu granicznego seria niezależnych doświadczeń musi być długa (minimum 100), a prawdopodobieństwo sukcesu niewielkie. Parametrem rozkładu Poisson'a jest $\lambda = np$, iloczyn długości serii (n) i prawdopodobieństwa sukcesu (p). Wartościami zmiennej, tak jak w rozkładzie Bernoulli'ego, jest liczba sukcesów k = 0,1,2,.... Zbiór wartości jest nieskończony i przeliczalny. Prawdopodobieństwo oblicza się ze wzoru:

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

Rozkład Poisson'a

Przykładem zjawisk, które można opisywać rozkładem Poisson'a jest liczba wypadków w jednostce czasu, liczba bakterii w danej objętości, liczba zachorowań na rzadkie choroby, czy liczba awarii jakiegoś urządzenia w danym przedziale czasu.

Jedynym parametrem rozkładu Poisson'a jest λ. Jest ona zarówno wartością oczekiwaną jak i wariancją zmiennej losowej:

$$EX = \lambda \qquad D^2 X = \lambda$$

PRZYKŁAD Rozkład Poisson'a

Wiadomo, że prawdopodobieństwo pojawienia się genetycznej wady włosa w okrywie lisów jest równe 0,004. Właściciel fermy złożonej z 800 lisów chce uzyskać informację jakie jest prawdopodobieństwo, że na jego fermie znajdzie co najmniej trzy lisy z wadą.

długość serii n=800,
prawdopodobieństwa pojawiania się wady p = 0,004
Zatem $\lambda = 800 \cdot 0{,}004 = 3{,}2$

$$P(X \geq 3) = 1 - P(X < 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2)) =$$
$$= 1 - (0{,}040762 + 0{,}130439 + 0{,}208702) = 1 - 0{,}379903 = 0{,}620097$$

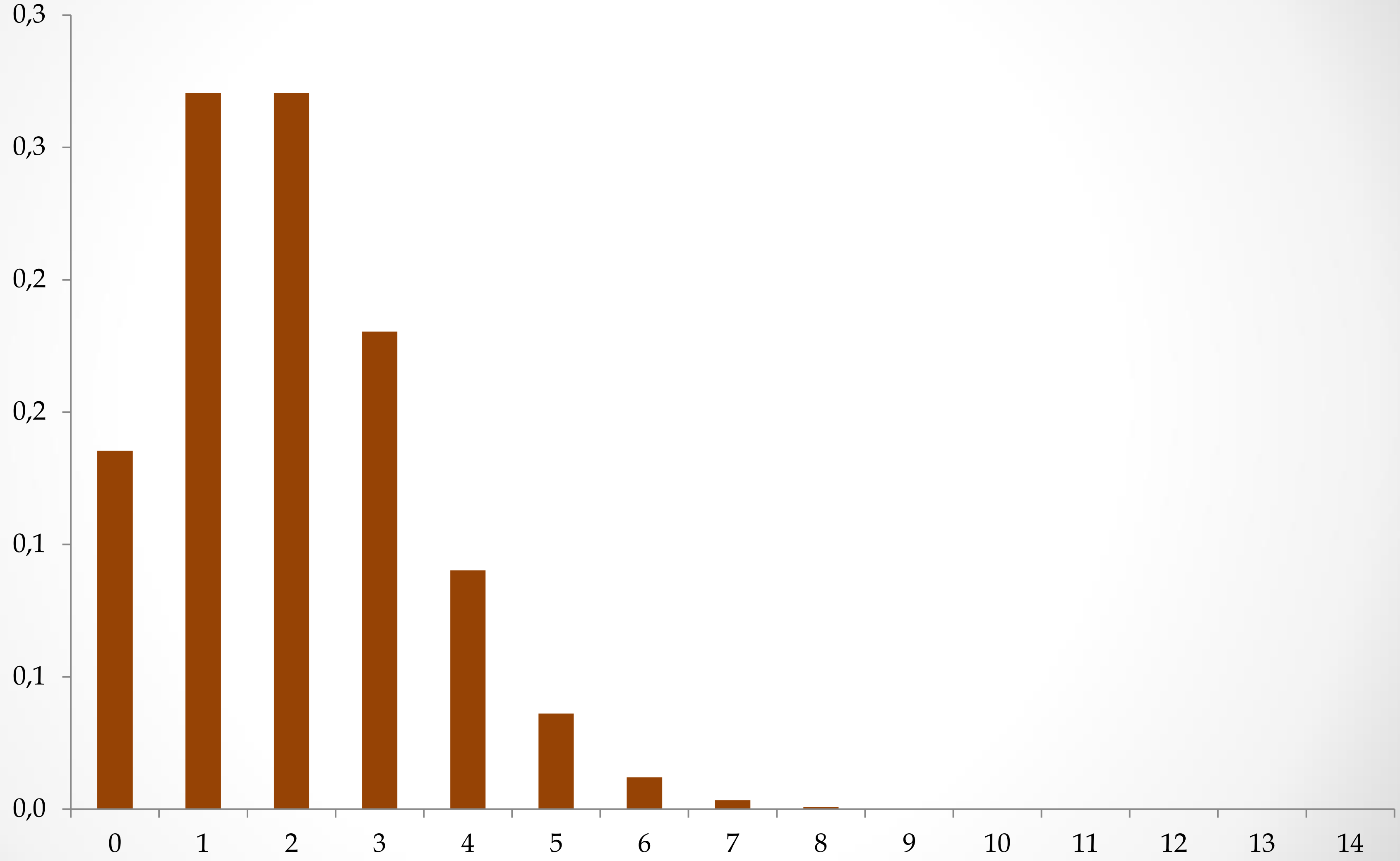
$$P(X = 0) = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{3{,}2^0}{0!} e^{-3{,}2} = 0{,}040762$$

Rozkład Poisson'a

$\lambda = 2$

xi	pi
0	0,13533528
1	0,27067057
2	0,27067057
3	0,18044704
4	0,09022352
5	0,03608941
6	0,01202980
7	0,00343709
8	0,00085927
9	0,00019095
10	0,00003819
11	0,00000694
12	0,00000116
13	0,00000018
14	0,00000003

Rozkład Poisson'a



Rozkład Poisson'a

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
0	,9048	,8187	,7408	,6703	,6065	,5488	,4966	,4493	,4066	,3679	,3329	,3012
1	,0905	,1637	,2222	,2681	,3033	,3293	,3476	,3595	,3659	,3679	,3662	,3614
2	,0045	,0164	,0333	,0536	,0758	,0988	,1217	,1438	,1647	,1839	,2014	,2169
3	,0002	,0011	,0033	,0072	,0126	,0198	,0284	,0383	,0494	,0613	,0738	,0867
4		,0001	,0003	,0007	,0016	,0030	,0050	,0077	,0111	,0153	,0203	,0260
5				,0001	,0002	,0004	,0007	,0012	,0020	,0031	,0045	,0062
6							,0001	,0002	,0003	,0005	,0008	,0012
7										,0001	,0001	,0002

k \ λ	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	2,1	2,2	2,3	2,4
0	,2725	,2466	,2231	,2019	,1827	,1653	,1496	,1353	,1225	,1108	,1003	,0907
1	,3543	,3452	,3347	,3230	,3106	,2975	,2842	,2707	,2572	,2438	,2306	,2177
2	,2303	,2417	,2510	,2584	,2640	,2678	,2700	,2707	,2700	,2681	,2652	,2613
3	,0998	,1128	,1255	,1378	,1496	,1607	,1710	,1804	,1890	,1966	,2033	,2090
4	,0324	,0395	,0471	,0551	,0636	,0723	,0812	,0902	,0992	,1082	,1169	,1254
5	,0084	,0111	,0141	,0176	,0216	,0260	,0309	,0361	,0417	,0476	,0538	,0602
6	,0018	,0026	,0035	,0047	,0061	,0078	,0098	,0120	,0146	,0174	,0206	,0241
7	,0003	,0005	,0008	,0011	,0015	,0020	,0027	,0034	,0044	,0055	,0068	,0083
8	,0001	,0001	,0001	,0002	,0003	,0005	,0006	,0009	,0011	,0015	,0019	,0025
9					,0001	,0001	,0001	,0002	,0003	,0004	,0005	,0007
10									,0001	,0001	,0001	,0002

Przykład

Średnia liczba bakterii w kropli substancji jest równa 0,5. Z ilu kropli należy utworzyć próbkę substancji, aby z prawdopodobieństwem 0,95 znalazła się w niej co najmniej jedna bakteria?

Rozkład liczby bakterii w kropli wody jest rozkładem Poisson'a o λ = 0,5, w dwóch kroplach wody liczba bakterii będzie miała rozkład o λ = 1,0 itd.

Trzeba obliczyć liczbę kropli, aby P(X>0) było równe 95%.

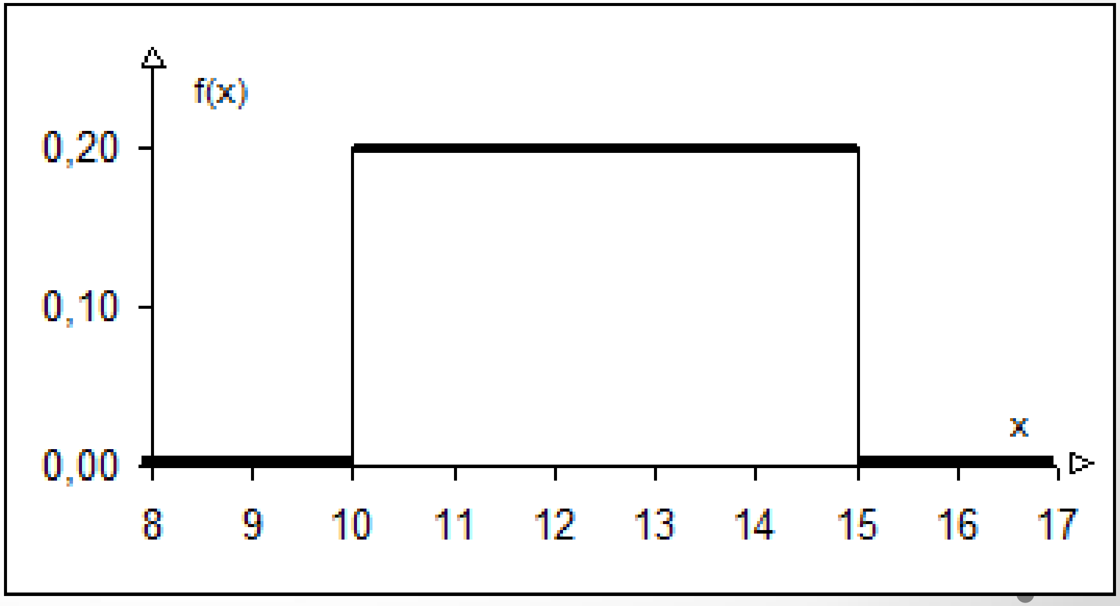
P(X>0) = 1- P(X=0)
Skoro P(X>0) = 0,95 to (PX=0) = 1 – P(X>0) = 1 – 0,95 = 0,05

L. kropli	1	2	3	4	5	6	7	8	9	10
λ	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
P(X=0)	0,6065	0,3679	0,2231	0,1353	0,0821	0,0498	0,0302	0,0183	0,0111	0,0067

Rozkład jednostajny

Zmienna losowa, która przyjmuje dowolną wartość z przedziału (a ; b) z jednakowym prawdopodobieństwem

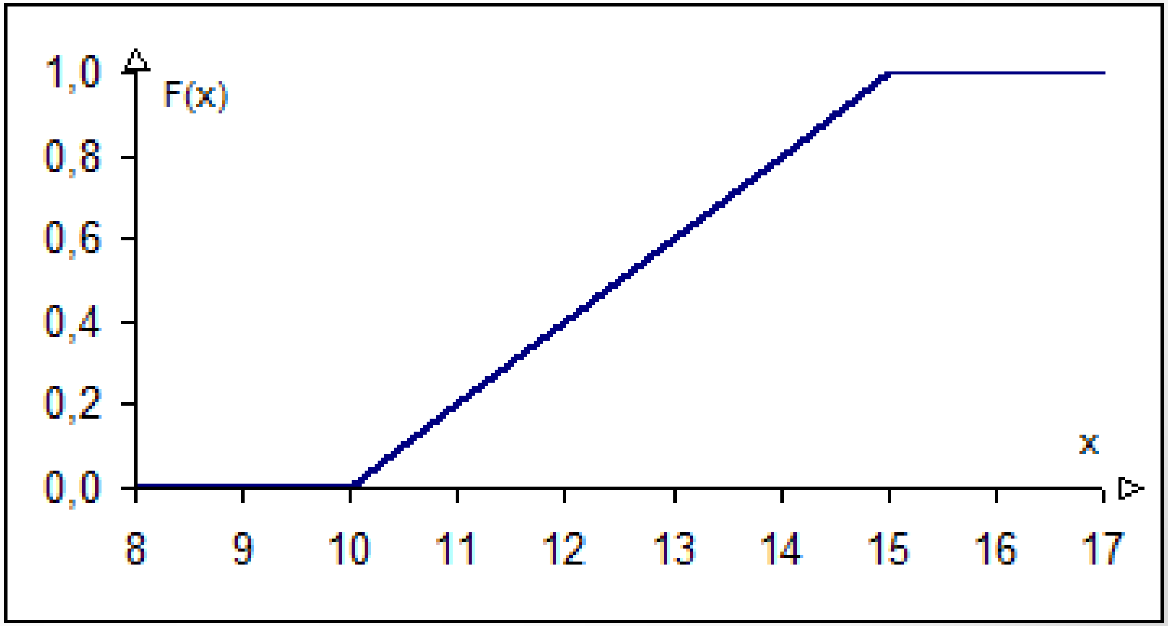
$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{dla } \ x \in (a;b) \\ 0 & \text{dla } \ x \notin (a;b) \end{cases}$$



Rozkład jednostajny

Dystrybuanta

$$F(x)=\begin{cases}0 & \text{dla } x\le a\\ \dfrac{x-a}{b-a} & \text{dla } x\in(a;b)\\ 1 & \text{dla } x\ge b)\end{cases}$$



Rozkład jednostajny

Parametry

$$EX = \frac{1}{2}(a+b)$$

$$D^2X = \frac{1}{12}(b-a)^2$$

Rozkład normalny

Rozkład normalny z wielu względów jest najważniejszym rozkładem zmiennej losowej ciągłej.

Cechy bezpośrednio mierzone, takie jak masy, wysokości, długości, powierzchnie, objętości, wydajności mają rozkład normalny.

Błędy losowe mają rozkład normalny.

Ponadto wiele rozkładów dąży do rozkładu normalnego w określonych warunkach, czyli rozkład ten jest graniczny dla wielu innych rozkładów.

Rozkład normalny

Funkcja gęstości rozkładu normalnego określona dla wszystkich liczb rzeczywistych

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

Dwa parametry rozkładu normalnego:

$$EX = m \quad D^2X = \sigma^2 \quad DX = \sigma$$

Rozkład normalny

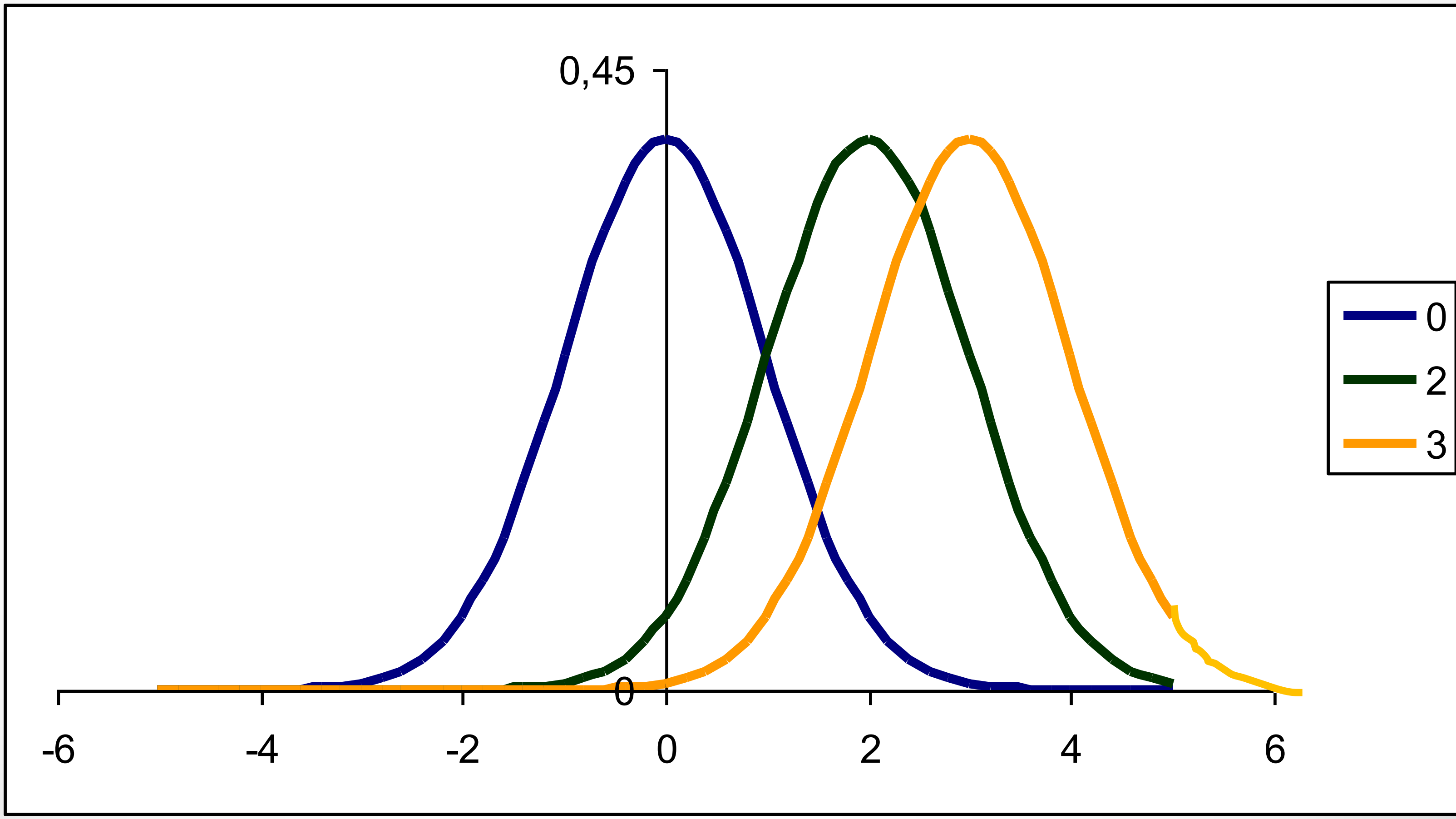
Funkcja gęstości rozkładu normalnego określona dla wszystkich liczb rzeczywistych

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

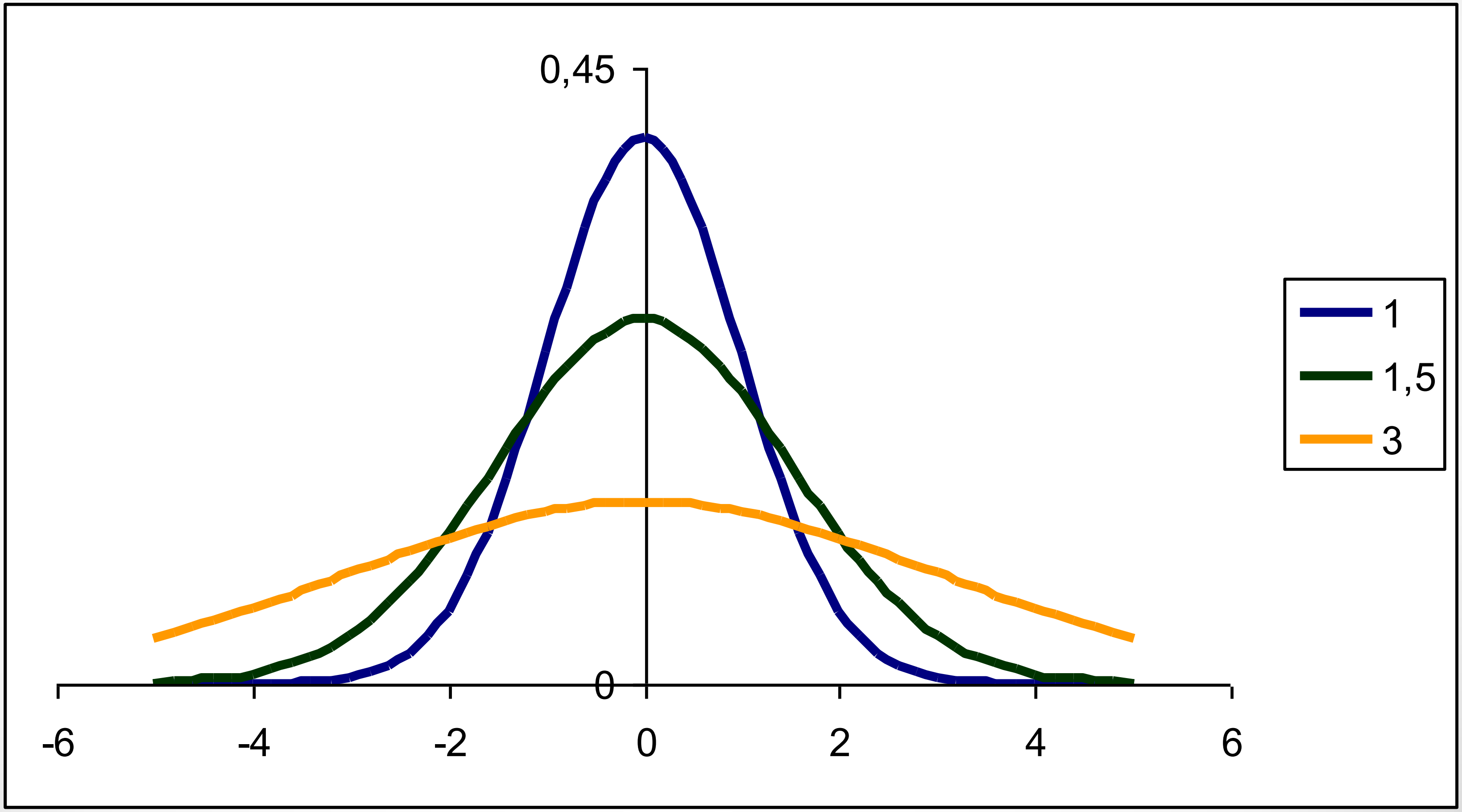
Dwa parametry rozkładu normalnego:

$$EX = m \quad D^2X = \sigma^2 \quad DX = \sigma$$

Rozkład normalny – różne *m*



Różne wartości odchylenia standardowego



Rozkład normalny

- każdy rozkład jest jednoznacznie określony przez swoje dwa parametry: wartość oczekiwaną m i odchylenie standardowe σ, co zapisujemy: X~N(m ; σ),
- jest symetryczny względem prostej x=m, funkcja gęstości osiąga maksimum dla EX, stąd wartość oczekiwana, modalna i mediana są sobie równe
- jest określony dla liczb rzeczywistych
- prawdopodobieństwo występowania wartości zmiennej losowej w przedziałach liczbowych o końcach wyznaczonych przez parametry rozkładu (m i σ) jest jednakowe dla każdej zmiennej o rozkładzie normalnym. Zasada ta nazywana jest regułą „trzech sigm” tzn:

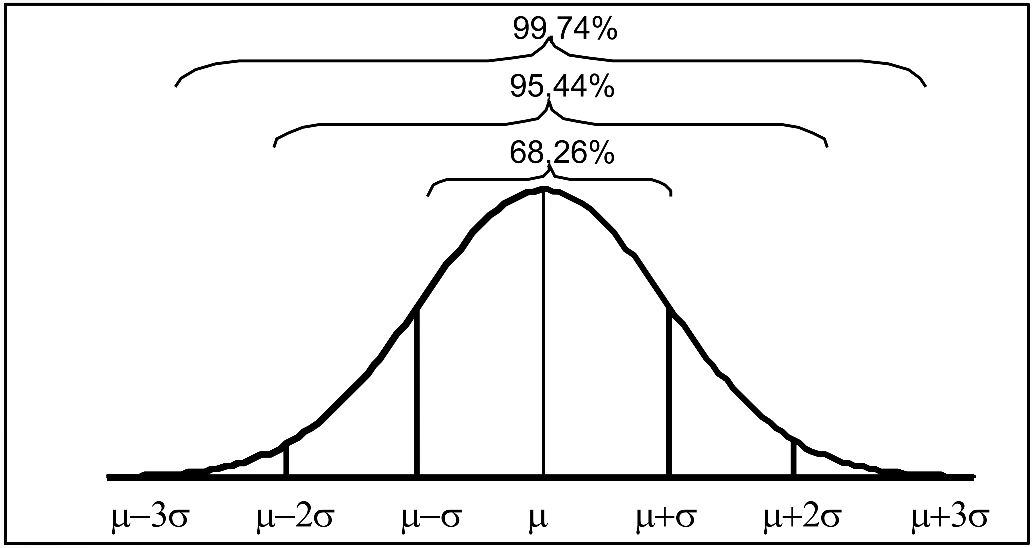
$$P(m-\sigma < X < m+\sigma) = 0{,}6826$$

$$P(m-2\cdot\sigma < X < m+2\cdot\sigma) = 0{,}9544$$

$$P(m-3\cdot\sigma < X < m+3\cdot\sigma) = 0{,}9974$$

Rozkład normalny

reguła „trzech sigm"



$$P(m-\sigma < X < m+\sigma) = 0{,}6826$$
$$P(m-2\cdot\sigma < X < m+2\cdot\sigma) = 0{,}9544$$
$$P(m-3\cdot\sigma < X < m+3\cdot\sigma) = 0{,}9974$$

Podobnie wartość dystrybuanty dla wartości zmiennej losowej wyznaczonej przez parametry jej rozkładu jest jednakowa dla każdej zmiennej o rozkładzie normalnym, tzn:

Jeżeli $X_1 \sim N(m_1\,;\,\sigma_1)$ *oraz* $X_2 \sim N(m_2\,;\,\sigma_2)$

to $F(X_1 = m_1 + k\cdot\sigma_1) = F(X_2 = m_2 + k\cdot\sigma_2)$.

Standaryzacja

- Przekształcenie zwane standaryzacją, jest liniowym przekształceniem, w wyniku którego uzyskujemy zmienną losową o wartości oczekiwanej równej zero i odchyleniu standardowym równym jeden.
- Jeśli zmienna $X \sim N(m;\sigma)$ to zmienna U zdefiniowana jako ma rozkład normalny o parametrach 0 i 1 , czyli zmienna $U \sim N(0\,;1)$.

$$U = \frac{X - m}{\sigma}$$

Standaryzacja

x_i	p_i					
1	0,25		0,25	0,25		
2	0,2		0,4	0,8	D^2X	9,848
5	0,25		1,25	6,25		
7	0,15		1,05	7,35	DX	3,138
10	0,15		1,5	15		
	1	EX	4,45	29,65		

u_i	p_i					
-1,099	0,25		-0,2749	0,30217		
-0,781	0,2		-0,1561	0,12191	D^2X	1
0,175	0,25		0,04382	0,00768		
0,813	0,15		0,12189	0,09905	DX	1
1,769	0,15		0,26529	0,46919		
	1	EX	0	1		

Przykład

Obliczyć P(7<X<16) wiedząc, że zmienna X~N(10 ; 4).

dla zm. los. ciągłej P(7<X<16) = F(X=16) - F(X=7)

$$F(X=16)=F(U=\frac{16-10}{4})=F(U=1{,}5)=0{,}9332$$

$$F(X=7)=F(U=\frac{7-10}{4})=F(U=-0{,}75)=1-F(U=0{,}75)=1-0{,}7734=0{,}2266$$

$$P(7<X<16)=F(X=16)-F(X=7)=0{,}9332-0{,}2266=0{,}7066$$

Przykład

Wzrost mężczyzn podlega rozkładowi normalnemu o średniej 180 cm, przy czym 2,5% mężczyzn jest niższych niż 170,2 cm. Jaki procent mężczyzn jest wyższy niż 185 cm?

$$P(X < 170{,}2) = F(X = 170{,}2) = 0{,}025$$

$$F(X = 170{,}2) = F(U = \frac{170{,}2 - 180}{\sigma}) = 0{,}025$$

$$\text{z tablic} \quad U = \frac{-9{,}8}{\sigma} = -1{,}96 \Rightarrow \sigma = 5$$

$$P(X > 185) = 1 - F(X = 185) = 1 - F\left(U = \frac{185 - 180}{5}\right) = 1 - F(U = 1) = 1 - 0{,}8413 = 0{,}1587$$

Dystrybuanta rozkładu normalnego

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	,50000	,50399	,50798	,51197	,51595	,51994	,52392	,52790	,53188	,53586
0,1	,53983	,54380	,54776	,55172	,55567	,55962	,56356	,56749	,57142	,57535
0,2	,57926	,58317	,58706	,59095	,59483	,59871	,60257	,60642	,61026	,61409
0,3	,61791	,62172	,62552	,62930	,63307	,63683	,64058	,64431	,64803	,65173
0,4	,65542	,65910	,66276	,66640	,67003	,67364	,67724	,68082	,68439	,68793
0,5	,69146	,69497	,69847	,70194	,70540	,70884	,71226	,71566	,71904	,72240
0,6	,72575	,72907	,73237	,73565	,73891	,74215	,74537	,74857	,75175	,75490
0,7	,75804	,76115	,76424	,76730	,77035	,77337	,77637	,77935	,78230	,78524
0,8	,78814	,79103	,79389	,79673	,79955	,80234	,80511	,80785	,81057	,81327
0,9	,81594	,81859	,82121	,82381	,82639	,82894	,83147	,83398	,83646	,83891
1,0	,84134	,84375	,84614	,84849	,85083	,85314	,85543	,85769	,85993	,86214
1,1	,86433	,86650	,86864	,87076	,87286	,87493	,87698	,87900	,88100	,88298
1,2	,88493	,88686	,88877	,89065	,89251	,89435	,89617	,89796	,89973	,90147
1,3	,90320	,90490	,90658	,90824	,90988	,91149	,91308	,91466	,91621	,91774
1,4	,91924	,92073	,92220	,92364	,92507	,92647	,92785	,92922	,93056	,93189
1,5	,93319	,93448	,93574	,93699	,93822	,93943	,94062	,94179	,94295	,94408
1,6	,94520	,94630	,94738	,94845	,94950	,95053	,95154	,95254	,95352	,95449
1,7	,95543	,95637	,95728	,95818	,95907	,95994	,96080	,96164	,96246	,96327
1,8	,96407	,96485	,96562	,96638	,96712	,96784	,96856	,96926	,96995	,97062
1,9	,97128	,97193	,97257	,97320	,97381	,97441	,97500	,97558	,97615	,97670


Gyps fulvus