Rozkłady prawdopodobieństwa.

Rozkłady dyskretne i ciągłe.

W przypadku rozkładu dyskretnego określamy wartości prawdopodobieństwa dla przeliczalnej (skończonej lub nieskończonej) liczby wartości zmiennej losowej.
Np.
$x_1, x_2, ...x_n$ – wszystkie możliwe wartości zmiennej losowej
$P(x_1), P(x_2), ...P(x_n)$ – prawdopodobieństwa, z którymi zmienna losowa przyjmuje kolejne wartości.
Z definicji prawdopodobieństwa wynika, że

$$\sum_{i=1}^{n} P(x_i) = 1$$

Wartość średnia (oczekiwana) zmiennej losowej

$$\mu = \sum_{i=1}^{n} P(x_i) \cdot x_i$$

Mediana zmiennej losowej

$$P(x < \mu_{1/2}) = P(x \geq \mu_{1/2}) = \tfrac{1}{2}$$

Moda (wartość modalna) zmiennej losowej

$$P(\mu_{\max}) \geq P(x \neq \mu_{\max})$$

Wariancja zmiennej losowej

$$\sigma^2 = \sum P(x_i) \cdot (x_i - \mu)^2 = \sum P(x_i) \cdot x_i^2 - \mu^2$$

Dyspersja zmiennej losowej

$$\sigma = \sqrt{\sum P(x_i) \cdot (x_i - \mu)^2}$$

W przypadku rozkładu ciągłego zbiór wartości zmiennej losowej pokrywa się ze zbiorem liczb rzeczywistych.

Zamiast rozkładu prawdopodobieństwa wprowadza się rozkład gęstości prawdopodobieństwa zmiennej losowej ciągłej.

$$p(x) \text{ - } x \in \{R\}$$

$$\int_{x=-\infty}^{x=\infty} p(x) \cdot dx = 1$$

$$P(a < x < b) = \int_{x=a}^{x=b} p(x) \cdot dx$$

$$P(x_0 < x < x_0 + \delta x) \cong p(x) \cdot \delta x$$

$$\mu = \int p(x) \cdot x \cdot dx$$

$$\sigma^2 = \int p(x) \cdot (x - \mu)^2 \cdot dx = \int p(x) \cdot x^2 \cdot dx - \mu^2$$

Rozkład równomierny dyskretny

$$x \in \{1,2,3,4,5,6\}, \;\; P(x_i) = p = \tfrac{1}{6}$$

$$\mu = \sum_{i=1}^{6} P(x_i) \cdot x_i \; = p(1+2+3+4+5+6) = \frac{21}{6} = 3{,}5$$

$$\mu_{1/2} = 4$$

$$\sigma^2 = \sum_{i=1}^{6} p \cdot (x_i - \mu)^2 = \sum_{i=1}^{6} \frac{1}{6} \cdot (x_i - 3{,}5)^2 = \frac{17{,}5}{6} = 2{,}916(6)$$

$$\sigma = 1{,}7078...$$

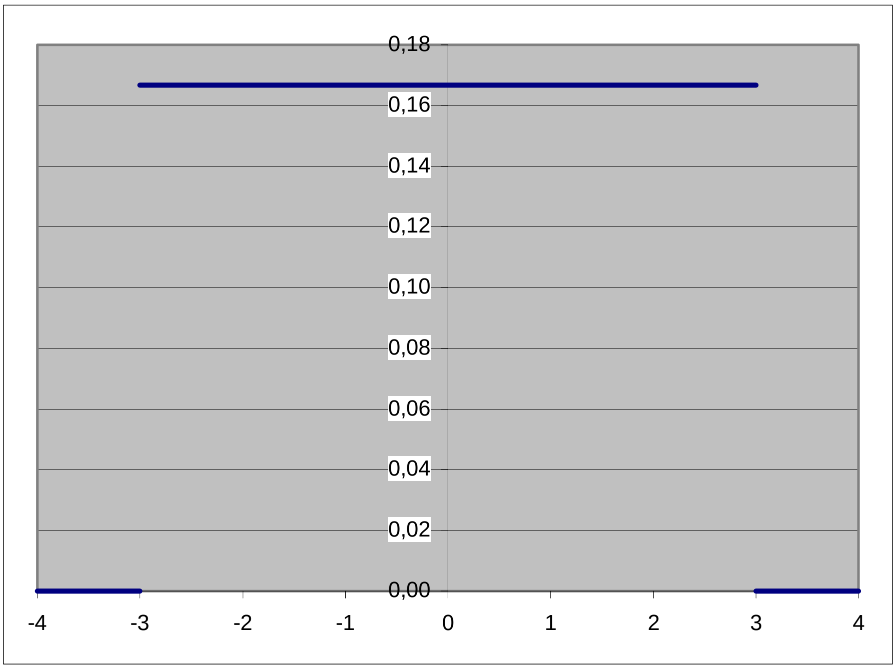
$$\mu = 3{,}5$$

$$\sigma^2 = 2{,}916(6)$$

$$\sigma = 1{,}7078...$$

Rozkład równomierny ciągły

$$p(x)=\begin{cases}0, & x\notin\langle a,b\rangle\\ \dfrac{1}{b-a}, & x\in\langle a,b\rangle\end{cases}$$



$$\mu=\int_a^b\frac{x}{b-a}\cdot dx=\left.\frac{x^2}{2(b-a)}\right|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2}$$

$$\sigma^2=\frac{1}{b-a}\int_a^b x^2\cdot dx-\mu^2=\frac{1}{b-a}\cdot\left.\frac{x^3}{3}\right|_a^b-\mu^2=$$

$$\frac{1}{b-a}\cdot\frac{b^3-a^3}{3}-\frac{(a+b)^2}{4}=\frac{1}{b-a}\cdot\frac{(b-a)(b^2+ab+a^2)}{3}-\frac{(a+b)^2}{4}=$$

$$\frac{4(b^2+ab+a^2)-3(a+b)^2}{12}=\frac{b^2-2ab+a^2}{12}=\frac{(b-a)^2}{12}$$

$$a=0$$

$$b=6$$

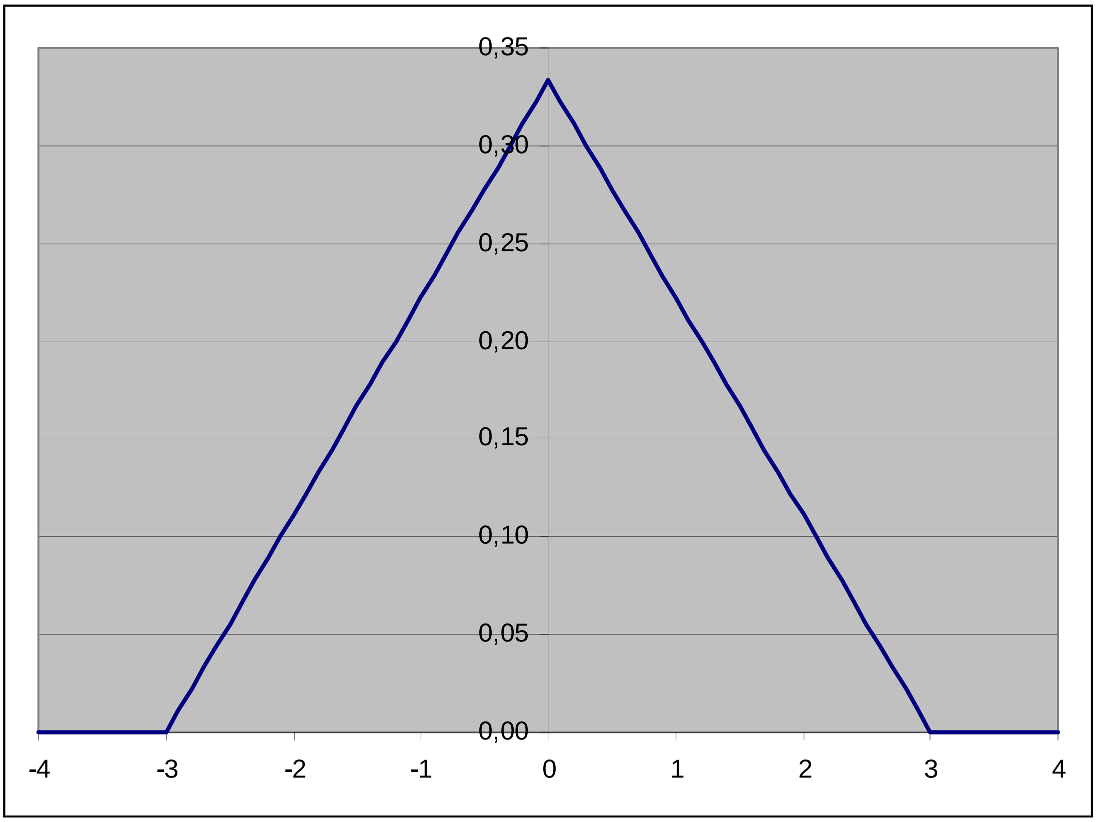
$$\mu=\frac{a+b}{2}=3$$

$$\sigma^2=\frac{(b-a)^2}{12}=3$$

$$\sigma=\frac{b-a}{2\sqrt{3}}=\sqrt{3}=1{,}7320...$$

Rozkład trójkątny

$$p(x) = \begin{cases} 0, \ \ x \notin \left\langle -\frac{a}{2}, \frac{a}{2} \right\rangle \\ \frac{2}{a}\left(1 + \frac{2}{a}x\right), \ \ x \in \left\langle -\frac{a}{2}, 0 \right\rangle \\ \frac{2}{a}\left(1 - \frac{2}{a}x\right), \ \ x \in \left(0, \frac{a}{2} \right\rangle \end{cases}$$



Ze względu na symetrię $\mu = 0$

$$\sigma^2 = \int_{-a/2}^{0} p_1(x)x^2 dx + \int_{0}^{a/2} p_2(x)x^2 dx = 2\int_{0}^{a/2} p_2(x)x^2 dx$$

$$\sigma^2 = \frac{a^2}{24}$$

$$\sigma = \frac{a}{2\sqrt{6}}$$

dla $a = 6$

$$\sigma^2 = 1{,}5$$

$$\sigma = 1{,}2247\ldots$$

Rozkład dwumianowy

Prawdopodobieństwo sukcesu w jednej próbie wynosi *p*,
a porażki $q = 1 - p$.

Liczba sukcesów *x* w serii *n* prób jest liczbą losową o rozkładzie dwumianowym.

$$P(x; n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Na podstawie tożsamości

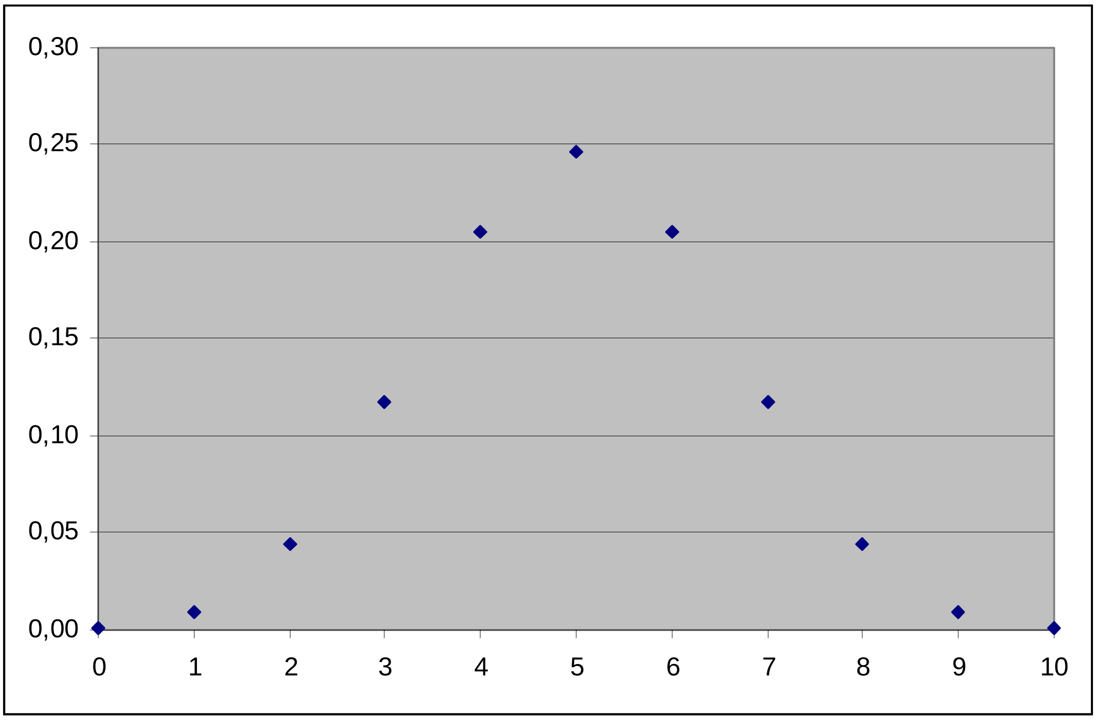
$$(p+q)^n = \sum_{x=0}^{n} \left[\binom{n}{x} p^x q^{n-x} \right]$$

można natychmiast wykazać, że wartości prawdopodobieństwa są unormowane, to znaczy

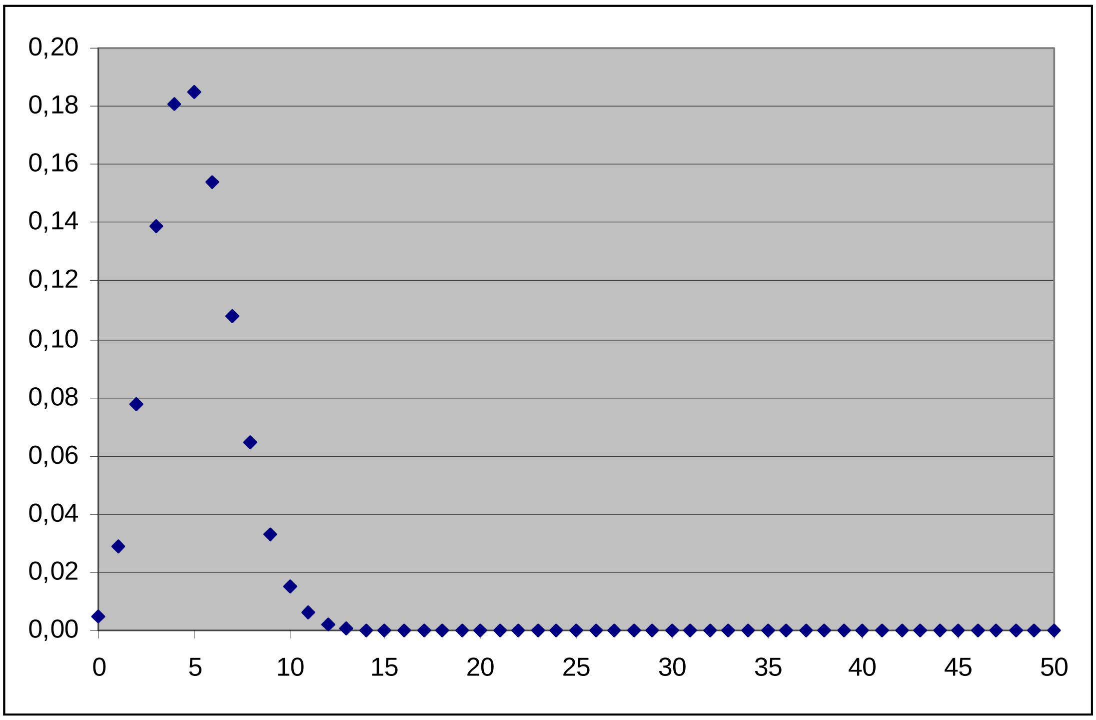
$$\sum_{x=0}^{n} P(x; n, p) = 1$$

$$\mu = \sum_{x=0}^{n} \left[x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \right] = np$$

$$\sigma^2 = \sum_{x=0}^{n} \left[(x-\mu)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \right] = np(1-p)$$



$p = 0{,}5;\ n = 10$



$p = 0{,}1;\ n = 50$

Rozkład Poissona

Rozkład Poissona jest granicznym przypadkiem rozkładu dwumianowego przy spełnieniu warunków

$$n \to \infty, \; p \to 0, \; \mu = np = \text{const}$$

$$\lim_{\substack{p\to 0\\ n\to\infty}} P_D(x;n,p) = P_P(x;\mu) \equiv \frac{\mu^x}{x!}e^{-\mu}$$

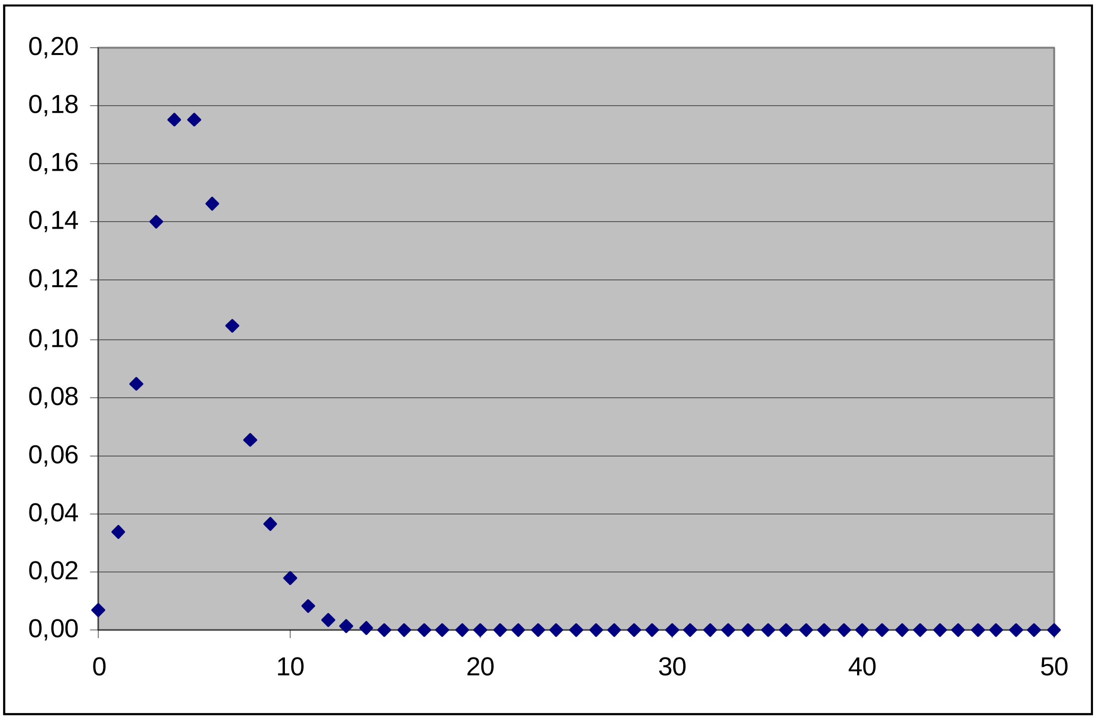
Wartość średnia wynosi oczywiście

$$\sum_{x=0}^{\infty} x\frac{\mu^x}{x!}e^{-\mu} = \mu$$

Wariancja rozkładu Poissona jest równa średniej

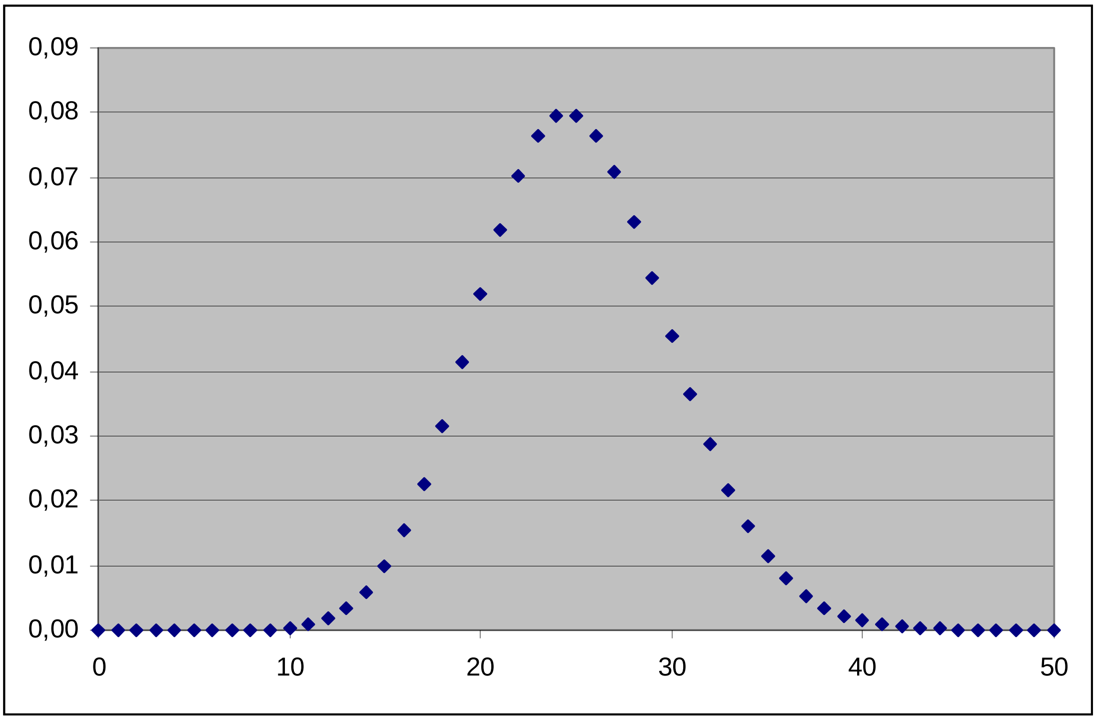
$$\sigma^2 = \sum_{x=0}^{\infty}(x-\mu)^2\frac{\mu^x}{x!}e^{-\mu} = \mu$$

Rozkład Poissona



$\mu = 5$

Rozkład Poissona



$\mu = 25$

Rozkład Poisson'a otrzymujemy bezpośrednio w przypadku tzw. ciągłej rejestracji zdarzeń, jeżeli spełniony jest postulat Poisson'a. Postulat Poisson'a stwierdza, że prawdopodobieństwo pojawienia się kolejnego zdarzenia w przedziale czasu $(t, t+dt)$ nie zależy od liczby zdarzeń zarejestrowanych w odstępie $(0,t)$ i wynosi $\mu \cdot dt$, gdzie μ jest średnią liczbą zdarzeń na jednostkę czasu.

Niech $P_r(t)$ jest prawdopodobieństwem zarejestrowania r zdarzeń w czasie t, a μ jest średnią liczbą zdarzeń jak wyżej. Wtedy

$$P_r(t+dt) = P_{r-1}(t)\cdot P_1(dt) + P_r(t)\cdot P_0(dt) = \\ P_{r-1}(t)\cdot \mu \cdot dt + P_r(t)\cdot(1-\mu\cdot dt)$$

$$P_r(t+dt) - P_r(t) = \left(P_{r-1}(t)\cdot\mu - P_r(t)\cdot\mu\right)\cdot dt$$

stąd

$$\frac{P_r(t+dt)-P_r(t)}{dt} = \frac{dP_r(t)}{dt} = -\mu\cdot P_r(t) + \mu\cdot P_{r-1}(t)$$

gdzie z definicji

$$P_{r-1} = 0 \text{ dla } r = 0.$$

Otrzymujemy zatem

$$\frac{dP_0(t)}{dt} = -\mu\cdot P_0(t)$$

co po rozwiązaniu daje

$$P_0(t) = e^{-\mu\cdot t}$$

Następnie

$$\frac{dP_1(t)}{dt} = -\mu\cdot P_1(t) + \mu\cdot e^{-\mu\cdot t}$$

a po scałkowaniu

$$P_1(t) = \mu\cdot t\cdot e^{-\mu\cdot t}.$$

Metodą indukcji można udowodnić, że

$$P_n(t) = \frac{(\mu\cdot t)^n}{n!}\cdot e^{-\mu\cdot t}$$

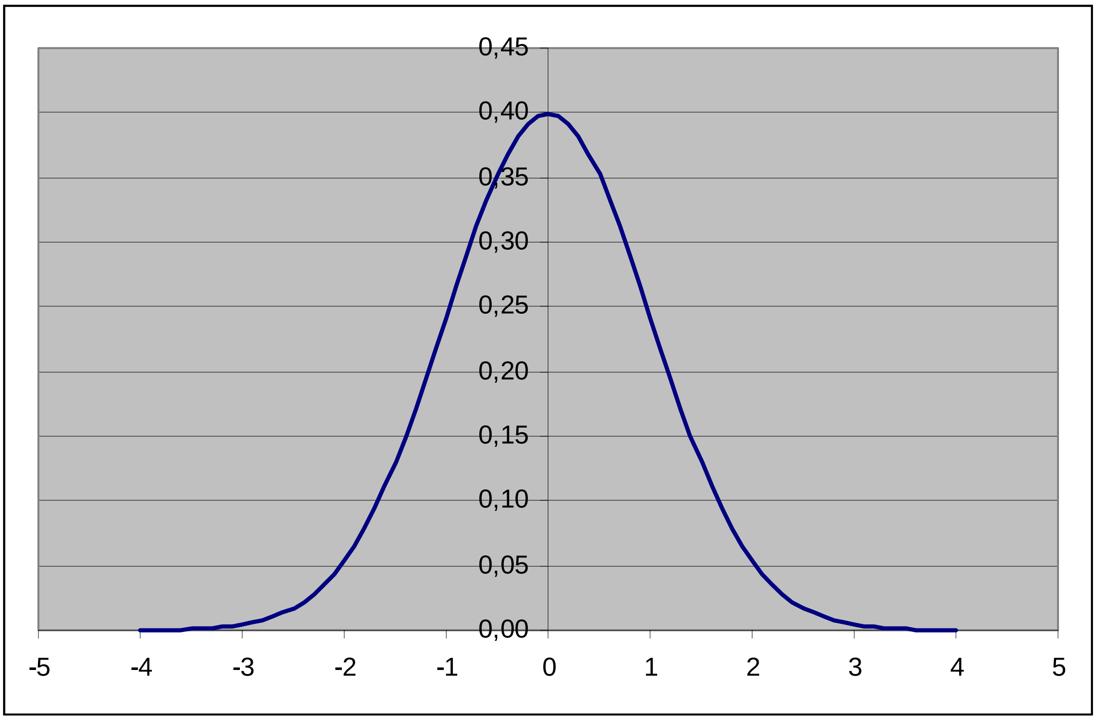
Czyli rozkład identyczny z rozkładem Poisson'a dla średniej liczby zdarzeń (sukcesów) $\mu\cdot t$.

Rozkład normalny (Gaussa)

$$p_N(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Standardowy rozkład Gaussa

$$p_N(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$



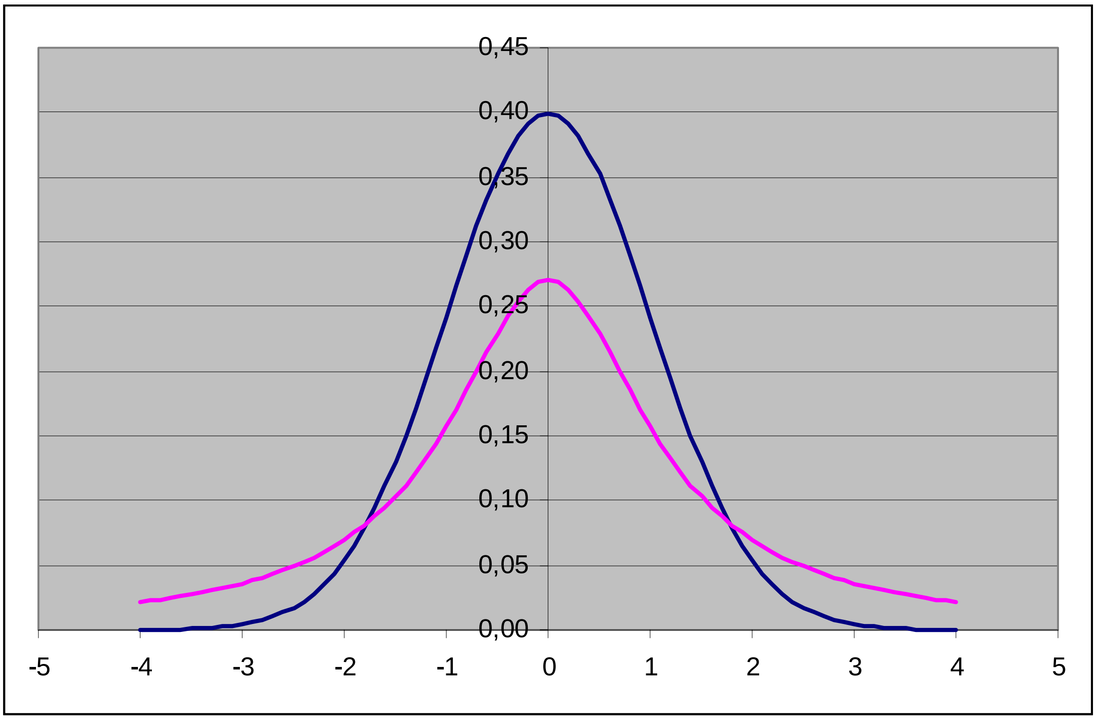
$$P_N(|x-\mu| < \Delta x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-\Delta x}^{\mu+\Delta x} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

Δz	$P_N(\lvert z\rvert < \Delta z)$	$1 - P_N(\lvert z\rvert < \Delta z)$
0,5	0,382925	0,617075
1,0	0,682689	0,317311
1,5	0,866386	0,133614
2,0	0,954500	0,045500
2,5	0,987581	0,012419
3,0	0,997300	0,002700
3,5	0,999535	0,000465
4,0	0,999937	0,000063

Rozkład Lorentza (Cauchy'ego)

$$p_L(x;\mu,\Gamma) = \frac{1}{\pi}\frac{\Gamma/2}{(x-\mu)^2+(\Gamma/2)^2}$$

Wariancja rozkładu Lorentza nie istnieje.



$\mu = 0,\ \Gamma = 2{,}354$

Rozkład wykładniczy

Funkcja rozkładu gęstości prawdopodobieństwa jest określona dla nieujemnego rzeczywistego argumentu:

$$p_e(x) = \lambda e^{-\lambda x} \text{ dla } x \geq 0,\ \lambda > 0$$

gdzie λ jest parametrem rozkładu.

Wartość średnia tego rozkładu wynosi

$$\mu = \lambda \int_0^\infty x e^{-\lambda x} dx = \lambda \left(-\frac{1}{\lambda}\left(x + \frac{1}{\lambda}\right) e^{-\lambda x} \right)\Bigg|_0^\infty = \left(x + \frac{1}{\lambda}\right) e^{-\lambda x}\Bigg|_\infty^0 = \frac{1}{\lambda}$$

a wariancja

$$\sigma^2 = \lambda \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

Przykład dla zmiennej dyskretnej (rozkład geometryczny)

n – liczba rzutów kostką do wyrzucenia sześciu oczek

szukamy $P_g(n)$ prawdopodobieństwa wyrzucenia sześciu oczek w *n*-tym rzucie $n = 1,2,3,\ldots$

$$P_g(1) = \frac{1}{6},\ P_g(2) = \frac{5}{6}\cdot\frac{1}{6},\ P_g(3) = \left(\frac{5}{6}\right)^2\cdot\frac{1}{6},\ \ldots\ P_g(n) = \left(\frac{5}{6}\right)^{n-1}\cdot\frac{1}{6}$$

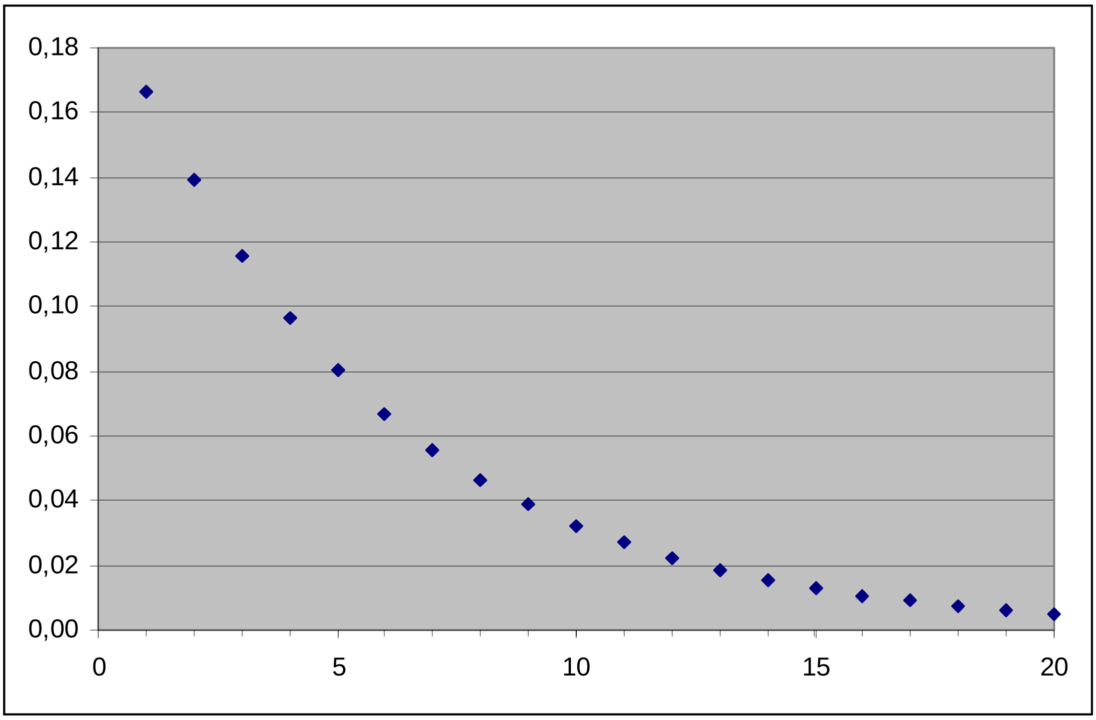
albo

$$P_g(n) = \frac{1}{6}e^{-(n-1)\ln\frac{6}{5}}$$

przy czym $\ln\left(\frac{6}{5}\right)$ nie różni się wiele od $\frac{1}{6}$ (różnica jest mniejsza od 10%).

Wyrzucenie sześciu (lub jakiejkolwiek innej ustalonej liczby) oczek jest najbardziej prawdopodobne w pierwszym rzucie.

Rozkład geometryczny dla rzutów kostką



Przykład dla zmiennej ciągłej

W przypadku zdarzeń podlegających rozkładowi Poisson'a prawdopodobieństwo nie zarejestrowania żadnego zdarzenia w czasie t wynosi

$$P_0(t) = e^{-\mu \cdot t}$$

natomiast prawdopodobieństwo zarejestrowania jednego zdarzenia w czasie dt bezpośrednio potem wynosi

$$\mu \cdot dt$$

Zatem prawdopodobieństwo tego, że zdarzenie zostanie zarejestrowane po czasie t (od momentu rozpoczęcia obserwacji lub od poprzedniego zdarzenia, lub od dowolnie ustalonej chwili) wynosi

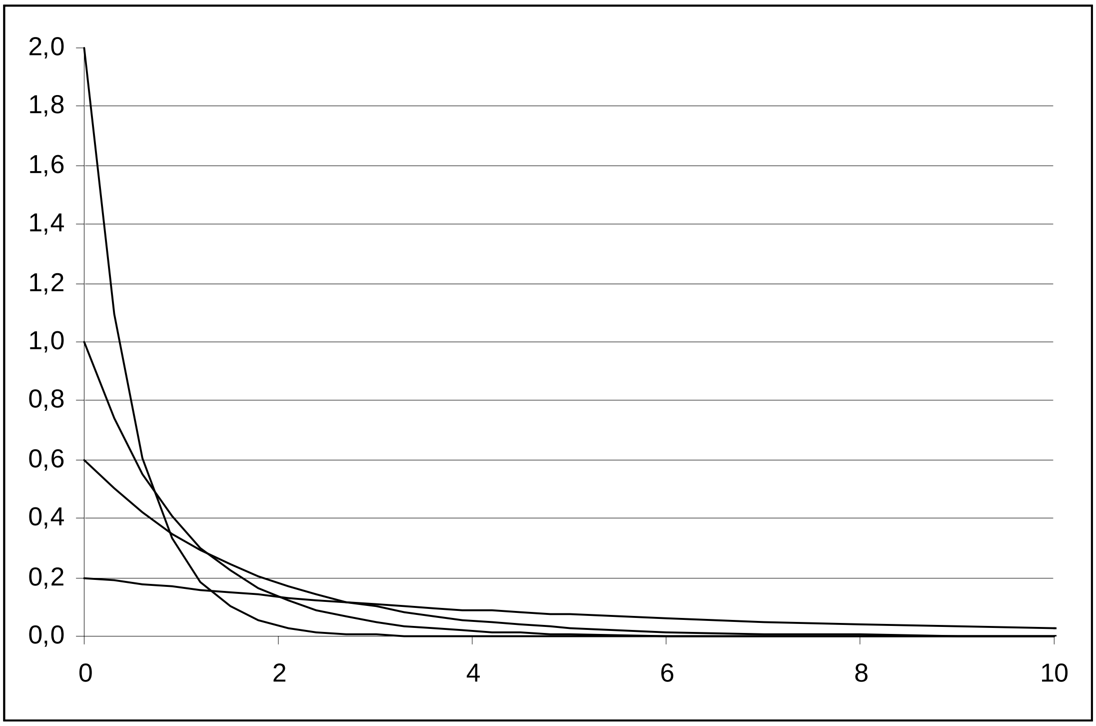
$$p(t) \cdot dt = \mu \cdot e^{-\mu \cdot t} dt$$

Rozkład gęstości prawdopodobieństwa odstępów między kolejnymi zdarzeniami podlegającymi rozkładowi Poisson'a jest wykładniczy

$$p(t) = \mu \cdot e^{-\mu \cdot t},$$

dla którego charakterystyczne jest to, że najbardziej prawdopodobne jest pojawienie się kolejnego zdarzenia bezpośrednio po poprzednim.

Rozkład wykładniczy $p(t) = \mu e^{-\mu \cdot t}$



dla μ = 0,2; 0,6; 1,0; 2,0