Rachunek prawdopodobieństwa i statystyka

Wykład 7

Rozkład Erlanga

Znajdziemy rozkład czasów oczekiwania na *n*-te zdarzenie. Łączny czas oczekiwania na *n* zdarzeń dany jest przez: $\mathbf{t} = \sum_{i=1}^{n} \mathbf{t}_i$ gdzie $\mathcal{E}(t_i;\lambda) = \lambda \exp(-\lambda t_i)$

Dla $n = 2$ mamy $t = t_1 + t_2$ oraz (wybieramy $v = t_2$, a więc $t_1 = u - v$ i $t_2 = v$ oraz $|J| = 1$):

$$\mathcal{E}_2(t;\lambda) = \int_{-\infty}^{\infty} \mathcal{E}(t-v;\lambda)\mathcal{E}(v;\lambda)\,dv = \lambda^2 \int_0^t \exp(-\lambda(t-v))\exp(-\lambda v)\,dv =$$

$$= \lambda^2 \exp(-\lambda t)\int_0^t dv = \lambda(\lambda t)\exp(-\lambda t)$$

Podobnie znajdujemy: $\mathcal{E}_3(t;\lambda) = \int_0^t \mathcal{E}(t-v;\lambda)\mathcal{E}_2(v;\lambda)\,dv =$

$$= \lambda^2 \int_0^t (\lambda v)\exp(-\lambda(t-v))\exp(-\lambda v)\,dv = \lambda^3 \exp(-\lambda t)\int_0^t v\,dv = \lambda \frac{(\lambda t)^2}{2}\exp(-\lambda t)$$

Rozkład czasu oczekiwania na *n* zdarzeń ma postać (rozkład Erlanga):

$$\mathcal{E}_n(t;\lambda) = \lambda \frac{(\lambda t)^{n-1}}{\Gamma(n)}\exp(-\lambda t)$$

Zadanie: Udowodnij powyższy wzór metodą indukcji matematycznej.

Rozkład Erlanga

Przykład: Przychodzimy do urzędu gdzie są dwa okienka. Przed jednym czeka jedna osoba, a przed drugim dwie. Jakie jest p-two, że ustawiając się w dłuższej kolejce dotrzemy do okienka szybciej?

$$\mathcal{E}(t_1;\lambda)=\lambda\exp(-\lambda t_1) \qquad \mathcal{E}_2(t_2;\lambda)=\lambda^2 t_2\exp(-\lambda t_2)$$

Łączny rozkład czasów oczekiwania dany jest przez:

$$f(t_1,t_2;\lambda)=\mathcal{E}(t_1;\lambda)\mathcal{E}_2(t_2;\lambda)=\lambda^3 t_2\exp(-\lambda(t_1+t_2))$$

Szukane p-two wynosi:

$$\mathbf{P}(t_2<t_1)=\iint\limits_{t_2<t_1} f(t_1,t_2;\lambda)\,dt_1dt_2=\lambda^3\iint\limits_{t_2<t_1} t_2\exp(-\lambda(t_1+t_2))\,dt_1dt_2=$$

$$=\lambda^3\int_0^\infty t_2\exp(-\lambda t_2)\left(\int_{t_2}^\infty \exp(-\lambda t_1)dt_1\right)dt_2=\lambda^2\int_0^\infty t_2\exp(-2\lambda t_2)\,dt_2=\frac{1}{4}$$

Zadanie: Pokaż, że w ogólnym przypadku dwóch kolejek ($m>n$) p-two to dane jest przez:

$$\mathbf{P}(t_m<t_n)=\frac{1}{2^m}\sum_{k=0}^{n-1}\binom{m-1+k}{m-1}\frac{1}{2^k}$$

Rozkład Poissona

Rozważmy proces podlegający rozkładowi wykładniczemu.

Ze względu na „brak pamięci" po każdym sygnale (np. przejazd samochodu) historia powtarza się od nowa, tak jakbyśmy oczekiwali na pierwsze zdarzenie.

Jaki jest rozkład p-twa wystąpienia *k* sygnałów w ustalonym czasie t obserwacji?

Niech zmienna losowa $T_k = t_1 + t_2 + \ldots + t_k$ oznacza moment przybycia k-tego sygnału.

Szukamy p-twa $P((T_k \le t) \cap (T_{k+1} > t))$, gdzie zmienna losowa $T_{k+1} = T_k + t_{k+1}$.

Łączną funkcję gęstości p-twa niezależnych zmiennych losowych T_k i T_{k+1} znajdujemy pamiętając, że zmienna T_k podlega rozkładowi Erlanga:

$$\lambda \frac{(\lambda T_k)^{k-1}}{(k-1)!} \exp(-\lambda T_k) \cdot \lambda \exp(-\lambda t_{k+1}) = \lambda^2 \frac{(\lambda T_k)^{k-1}}{(k-1)!} \exp(-\lambda (T_k + t_{k+1})) = \lambda^2 \frac{(\lambda T_k)^{k-1}}{(k-1)!} \exp(-\lambda T_{k+1})$$

Ze względu na jednostkowy jakobian przejścia od zmiennych T_k, t_{k+1} do zmiennych T_k, T_{k+1} łączna funkcja gęstości p-twa ma postać:

$$f(T_k, T_{k+1}; \lambda) = \lambda^2 \frac{(\lambda T_k)^{k-1}}{(k-1)!} \exp(-\lambda T_{k+1}) \qquad 0 \le T_k < T_{k+1} < \infty$$

Szukane p-two jest więc równe: $\mathrm{P}((\mathrm{T}_k \le t) \cap (\mathrm{T}_{k+1} > t)) =$

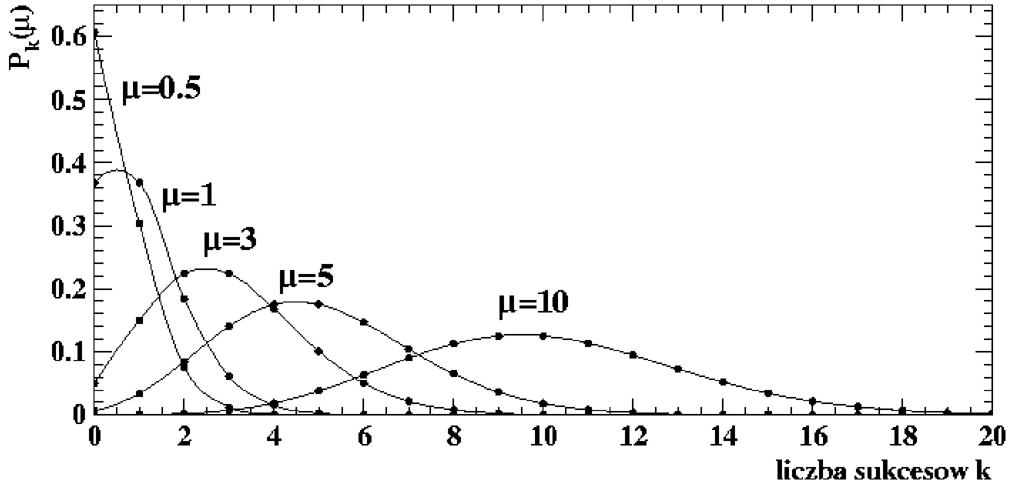
$$= \int_0^t \int_t^\infty f(T_k, T_{k+1}; \lambda)\, dT_{k+1} dT_k = \lambda^2 \int_0^t \frac{(\lambda T_k)^{k-1}}{(k-1)!} dT_k \int_t^\infty \exp(-\lambda T_{k+1})\, dT_{k+1} = \frac{(\lambda t)^k}{k!} \exp(-\lambda t)$$

Rozkład Poissona

Oznaczając przez $\mu = \lambda t$ otrzymujemy rozkład Poissona liczby sygnałów k=0,1,... w określonym przedziale czasu:

Math Player

$$\mathcal{P}_k(\mu) = \frac{\mu^k}{k!}\exp(-\mu)$$

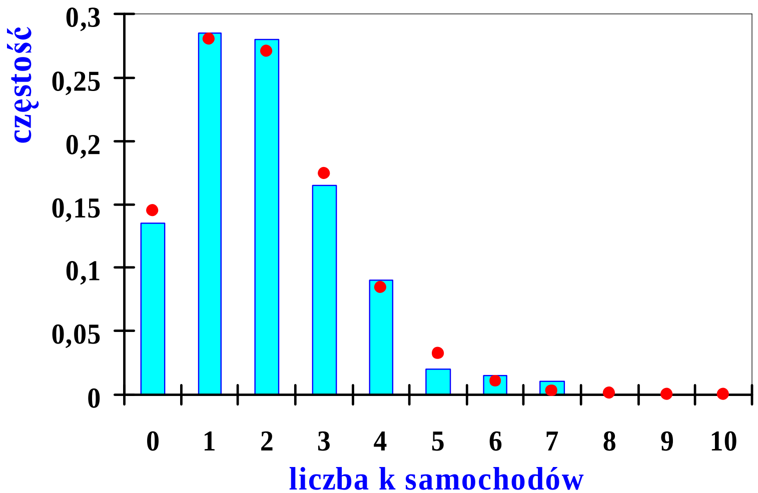


Wartość oczekiwana i wariancja:

$$\mathcal{E}[\mathrm{k}] = \sum_{k=0}^{\infty} k\mathcal{P}_{\mathrm{k}}(\mu) = \mu$$

$$\mathcal{V}[\mathrm{k}] = \sum_{k=0}^{\infty} k^2\mathcal{P}_{\mathrm{k}}(\mu) - \mu^2 = \mu$$

Błąd liczby zliczeń jest równy pierwiastkowi z tej liczby: $\mathcal{D}[\mathrm{k}] = \sqrt{\mu}$ $\qquad \frac{\mathcal{D}[\mathrm{k}]}{\mathcal{E}[\mathrm{k}]} = \frac{1}{\sqrt{\mu}}$



Przykład: Wyniki dotyczące obserwacji liczby k samochodów w 10 sekundowych przedziałach czasu. Parametr rozkładu Poissona:

$$\mu = \lambda t = 0.193\ [1/\mathrm{s}] \times 10\ [\mathrm{s}] = 1.93$$

Rozkład Poissona

Związek rozkładu dwumianowego z rozkładem Poissona.

Jakie jest p-two wystąpienia dokładnie k zdarzeń w przedziale czasu $[0, t]$?

$$\mathcal{B}_k(n,p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \frac{1}{k!} n(n-1)(n-2)\ldots(n-k+1)\, p^k (1-p)^{n-k}$$

Jeśli p-two sukcesu w pojedynczej próbie dane jest przez $p = \lambda\, t/n$ to dostajemy:

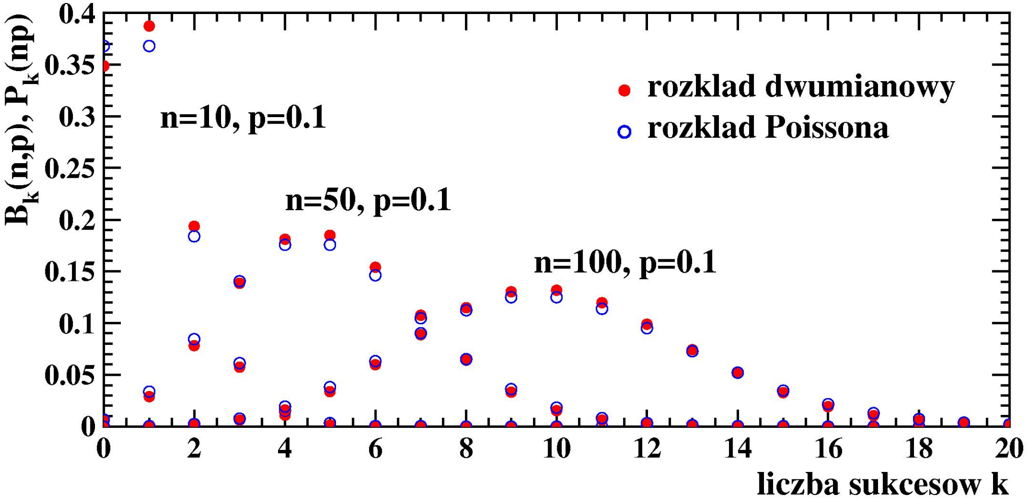
$$\mathcal{B}_k(n,p) = \frac{1}{k!} n(n-1)(n-2)\ldots(n-k+1)\left(\frac{\lambda t}{n}\right)^k \left(1-\frac{\lambda t}{n}\right)^{n-k} =$$

$$= \frac{1}{k!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)(\lambda t)^k \left(1-\frac{\lambda t}{n}\right)^{n}\left(1-\frac{\lambda t}{n}\right)^{k}$$

Dokonując przejścia granicznego dostajemy:

$$\mathcal{B}_k(n,p) \xrightarrow[n\to\infty]{} \mathcal{P}_k(\lambda t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Zastosowanie do rzadkich zdarzeń (np. liczba katastrof, błedów drukarskich, rzadkich rozpadów jądrowych, itp.)



Rozkład Poissona

Eksperyment Poissona polega na zliczaniu zdarzeń (sukcesów) w określonej jednostce czasu lub przestrzeni generowanych przez proces Poissona o następujących cechach:

- istnieje eksperymentalnie określona stała λ, będąca średnią częstością pojawiania się sukcesów w ustalonej jednostce czasu lub przestrzeni,
- liczby sukcesów występujące w dowolnych ustalonych i nie przekrywających się podjednostkach są wzajemnie niezależne,
- jeśli ustaloną jednostkę podzielimy na bardzo małe podjednostki h, to p-two dokładnie jednego sukcesu w każdej z podjednostek jest małe i takie samo, i w granicy dąży do λh,
- p-two więcej niż jednego sukcesu w bardzo małej podjednostce dąży do zera.

Przykład: Producent kabli chce używać rozkładu Poissona do oceny liczby defektów. Badając wiele 4-metrowych odcinków stwierdza, że średnio zawierają one 4 defekty. Czy spełnione są warunki procesu Poissona?

Przykład: Pod koniec każdego tygodnia zlicza się w klinice w dużym mieście liczby nowych przypadków zakaźnej choroby. W ciągu ostatnich trzech tygodni zanotowano odpowiednio 2, 10 i 30 przypadków. Czy spełnione są warunki procesu Poissona?

Rozkład normalny (Gaussa)

Wyprowadzenie rozkładu Gaussa w modelu Laplace'a błędów pomiarowych.

Rozważmy pomiar wielkości μ, który jest zaburzany przez n losowych efektów o wielkości ε każdy, zarówno zaniżających jak i zawyżających pomiar:

$$\mathrm{P}(\varepsilon) = \mathrm{P}(-\varepsilon) = \frac{1}{2}$$

W wyniku pomiaru otrzymujemy jedną z wielkości:

$$x_k = \mu + k\varepsilon + (n-k)(-\varepsilon) = \mu + (-n+2k)\varepsilon \qquad k = 0,1,2,\dots,n$$

których rozkład p-twa dany jest przez: $\mathcal{B}_k(n, p=0.5) = \binom{n}{k}\left(\frac{1}{2}\right)^n$

Wartość oczekiwana i wariancja zmiennej x_k:

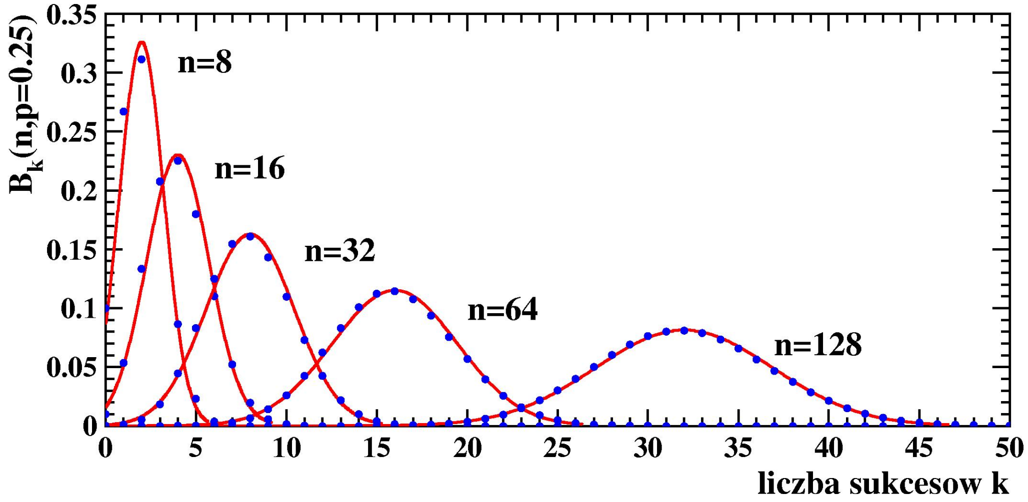
$$\mathcal{E}[x_k] = \langle \mu + (-n+2k)\varepsilon \rangle = \mu + (-n + 2\langle k \rangle)\varepsilon = \mu + \left(-n + 2\frac{1}{2}n\right)\varepsilon = \mu$$

$$\mathcal{V}[x_k] = \varepsilon^2 \mathcal{V}[-n+2k] = 4\varepsilon^2 \mathcal{V}[k] = 4\varepsilon^2 n \frac{1}{2}\left(1 - \frac{1}{2}\right) = \varepsilon^2 n$$

Rozkład normalny - wyprowadzenie

Zachowanie graniczne rozkładu dwumianowego dla dużych n:

$$\mathcal{B}_k(n,p) \cong \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right)$$



co w naszym przypadku prowadzi do:

$$\mathcal{B}_k(n,p) \cong \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right) =$$

$$= \frac{2}{\sqrt{2\pi n}} \exp\left(-\frac{\left(k-\frac{1}{2}n\right)^2}{2\frac{1}{4}n}\right) = \left\| k = \frac{x_k-\mu}{2\varepsilon} + \frac{n}{2} \right\| = \frac{2\varepsilon}{\sqrt{2\pi}\sqrt{n\varepsilon}} \exp\left(-\frac{(x_k-\mu)^2}{2n\varepsilon^2}\right)$$

Przechodząc z ε do zera, natomiast z n i k do nieskończoności, ale tak aby wariancja dążyła do stałej $n\varepsilon^2 \to \sigma^2$ dostajemy gęstość p-twa zmiennej x:

$$\frac{\mathcal{B}_k(n,p=0.5)}{2\varepsilon} \xrightarrow[n\to\infty,\ \varepsilon\to 0]{} \mathcal{N}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Math Player

Własności rozkładu normalnego

Wartość oczekiwana i wariancja:

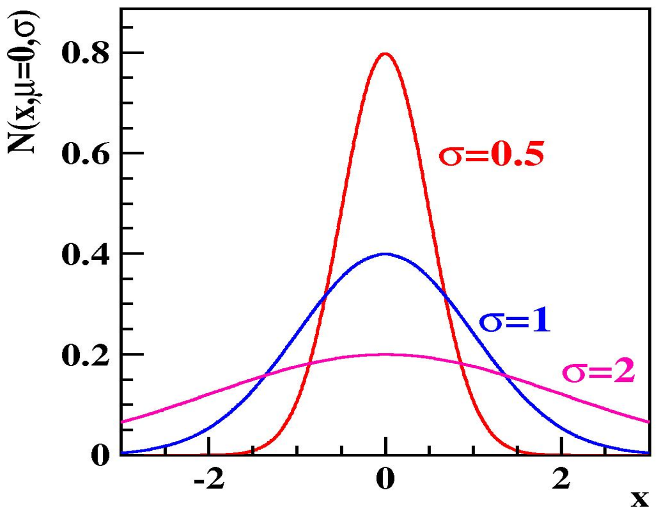
$$\mathcal{E}[x] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = \mu \qquad \mathcal{V}[x] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = \sigma^2$$

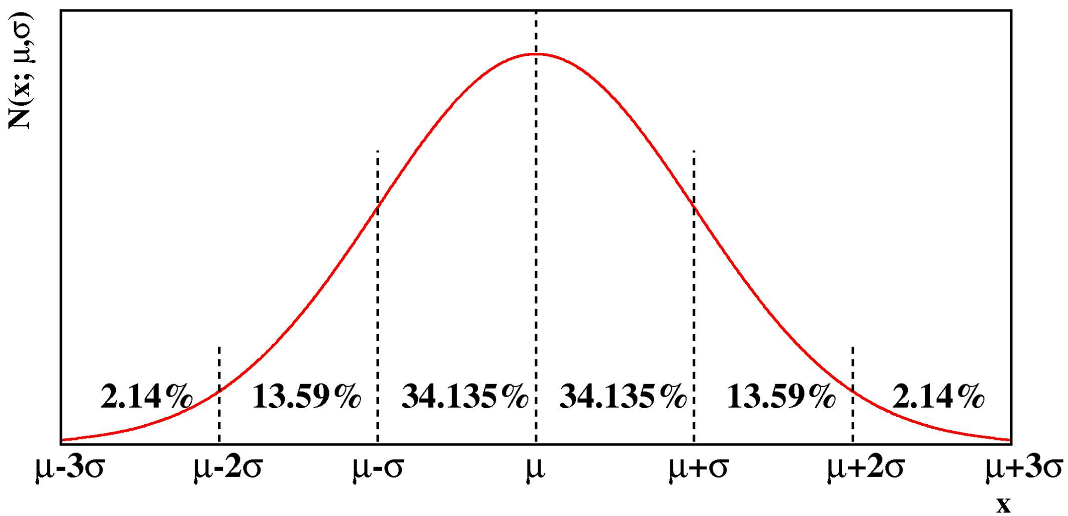
Wszystkie nieparzyste momenty centralne znikają ze względu na symetrię, natomiast parzyste dane są przez:

$$\left\langle (x-\mu)^{2k} \right\rangle = \frac{(2k)!}{2^k k!}\sigma^{2k}$$

Dla k = 2 mamy $\langle(x-\mu)^4\rangle = 3\sigma^4$, co oznacza, że współczynniki asymetrii $\gamma_1 = \dfrac{\mu_3}{\mathcal{D}^3[x]}$

i spłaszczenia $\gamma_2 = \dfrac{\mu_4}{\mathcal{V}^2[x]} - 3$ przyjmują wartości zerowe.

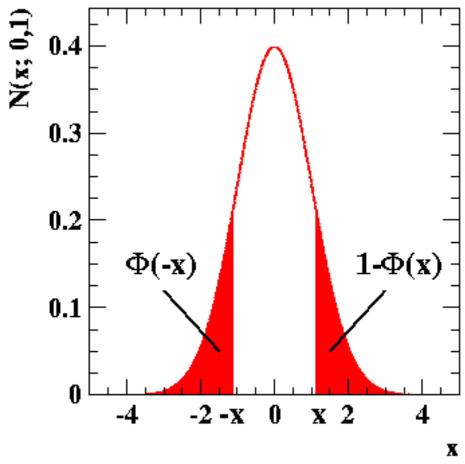


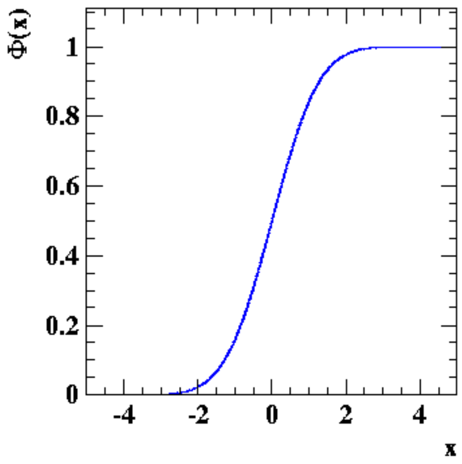


Dystrybuanta rozkładu normalnego

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) du \cong$$

$$\cong \begin{cases} \frac{1}{2} + \frac{1}{2}\sqrt{1 - \exp\left(-\frac{2x^2}{\pi}\right)} & \text{dla} \quad x > 0 \\ \frac{1}{2} - \frac{1}{2}\sqrt{1 - \exp\left(-\frac{2x^2}{\pi}\right)} & \text{dla} \quad x < 0 \end{cases}$$





X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0,5040	0.5080	0.5120	0.5160	0,5199	0,5239	0,5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5713
0.2	0.5793	0.5832	0.5861	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.2911	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0,8461	0,8485	0,8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1										

Math Player

Math Player

Rozkład normalny - przykłady

<u>Przykład:</u> Biolog chce ocenić wpływ snu zimowego na masę ciała wiewiórek. W tym celu waży 1000 dorosłych osobników płci męskiej w pod koniec lata i wczesną wiosną. Okazuje się, że pomiary wykonane w lecie mają rozkład normalny o średniej μ=400 g i odchyleniu standardowym σ=100 g. Jakie jest p-two, że losowo wybrana wiewiórka waży w lecie pomiędzy 350 g i 450 g?

$$P(350\,g < x < 450\,g) = P\left(\frac{350\,g - 400\,g}{100\,g} < z \equiv \frac{x-\mu}{\sigma} < \frac{450\,g - 400\,g}{100\,g}\right) =$$

$$= P(-0.5 < z < 0.5) = \Phi(0.5) - \Phi(-0.5) = 2\Phi(0.5) - 1 = 2 \times 0.6915 - 1 = 0.3830$$

<u>Przykład:</u> Wyniki testu IQ przeprowadzonego w pewnej populacji mają rozkład normalny o średniej μ=100 i odchyleniu standardowym σ=16. Ile wynosi wynik testu poniżej którego wypada 85% populacji?

$$P(x < x_{85}) = P\left(z \equiv \frac{x-\mu}{\sigma} < \frac{x_{85} - 100}{16}\right) = P(z < z_{85}) = \Phi(z_{85}) = 0.85$$

Z tablic odczytujemy: $z_{85} = 1.04 \quad \Rightarrow \quad \frac{x_{85} - 100}{16} = 1.04 \quad \Rightarrow \quad x_{85} = 116.64 \cong 117$

<u>Przykład:</u> Suma niezależnych zmiennych z rozkładu z rozkładu Gaussa o parametrach μ i σ:

$$\frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty} \exp\left(-\frac{(z-t-\mu)^2}{2\sigma^2}\right)\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)dt = \frac{1}{\sqrt{2\pi}\left(\sqrt{2}\sigma\right)}\exp\left(-\frac{(z-2\mu)^2}{2\left(\sqrt{2}\sigma\right)^2}\right) = \mathcal{N}\left(z; 2\mu, \sqrt{2}\sigma\right)$$

Dwuwymiarowy rozkład normalny

Gęstość p-twa dwuwymiarowego rozkładu normalnego:

Math Player Math Player

$$\mathcal{N}(x,y;\mu_x,\mu_y,\sigma_x,\sigma_y)=$$

$$=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\}$$

Elipsy kowariancji:

$$\frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]=C^2$$

wartość wykładnika	wielokrotność dyspersji	udział p-twa
0.5 (C=1)	1	39.3%
2.0 (C=2)	2	86.5%
4.5 (C=3)	3	98.9%

Kąt nachylenia dłuższej osi elipsy:

$$\tan 2\alpha=\frac{2\rho\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

Proste regresji II-go rodzaju:

$$y=\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

$$y=\mu_y+\frac{1}{\rho}\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$