Rachunek prawdopodobieństwa i statystyka

Wykład 9

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:

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

W wyniku pomiaru otrzymujemy jedną z wielkości:

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

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:

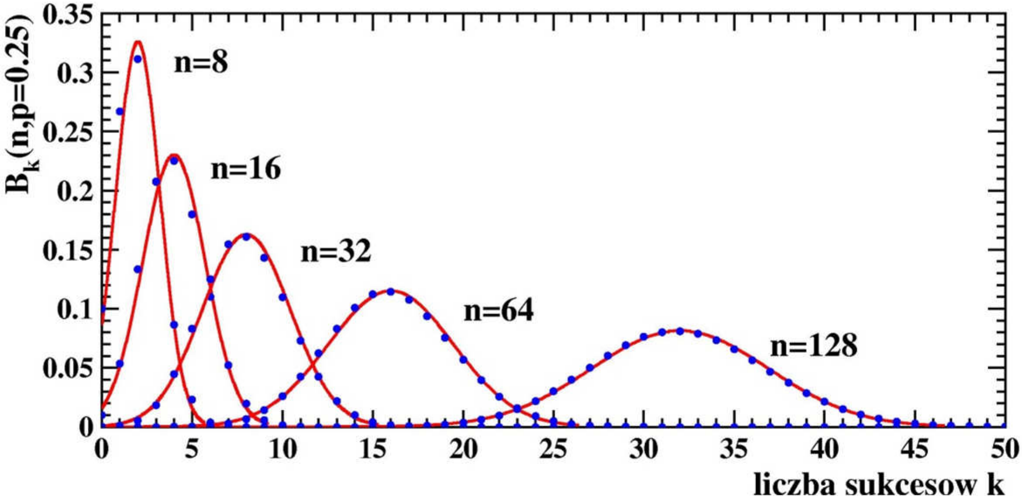
$$\mathcal{E}[x_k] = \langle \mu + (-n+2k)\varepsilon \rangle = \mu + (-n+2\langle k \rangle)\varepsilon = \mu + \left(-n + 2\cdot\frac{1}{2}n\right)\varepsilon = \mu$$

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

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:

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

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

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:

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

Własności rozkładu normalnego

Wartość oczekiwana i wariancja:

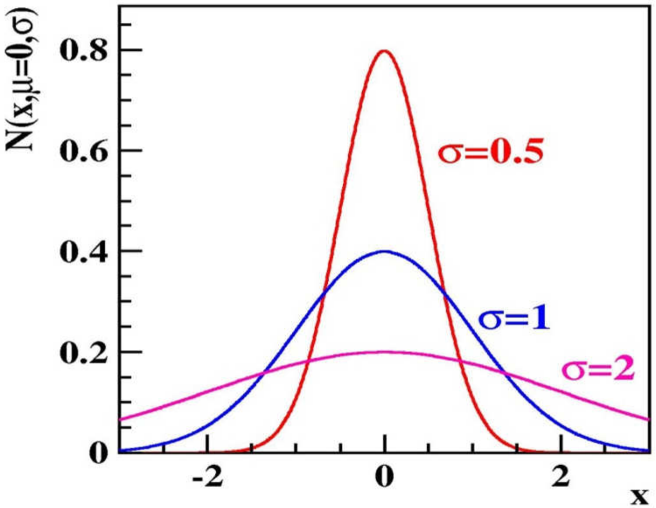
$$\mathcal{E}[x] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = \mu \qquad \mathcal{V}[x] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = \sigma^2$$

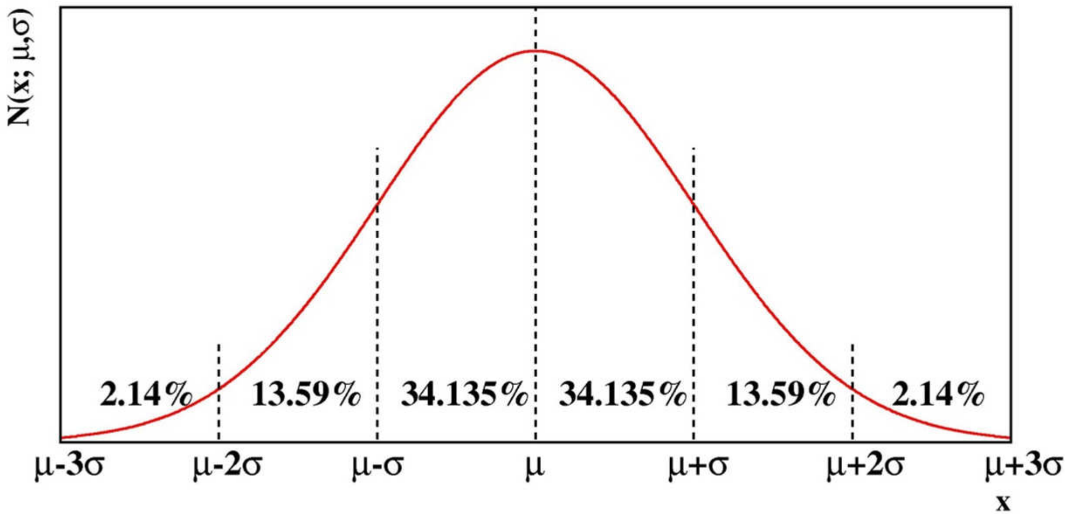
Wszystkie nieparzyste momenty centralne znikają ze względu na symetrię, natomiast parzyste dane są przez:

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

Math Player

Dla k = 2 mamy $\langle(x-\mu)^4\rangle=3\sigma^4$, co oznacza, że współczynniki asymetrii $\gamma_1 = \frac{\mu_3}{\mathcal{D}^3[x]}$ i spłaszczenia $\gamma_2 = \frac{\mu_4}{\mathcal{V}^2[x]} - 3$ przyjmują wartości zerowe.

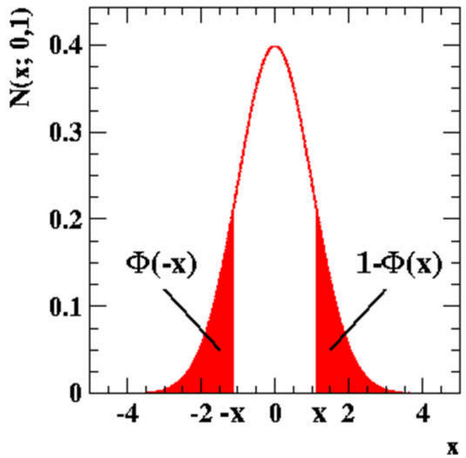


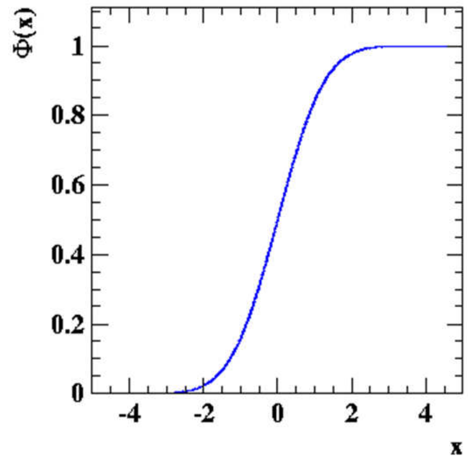


Dystrybuanta rozkładu normalnego

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

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





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

Przykład: 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?

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

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

Przykład: 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: $z_{85} = 1.04 \quad \Rightarrow \quad \dfrac{x_{85} - 100}{16} = 1.04 \quad \Rightarrow \quad x_{85} = 116.64 \cong 117$

Przykład: 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

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

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

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%

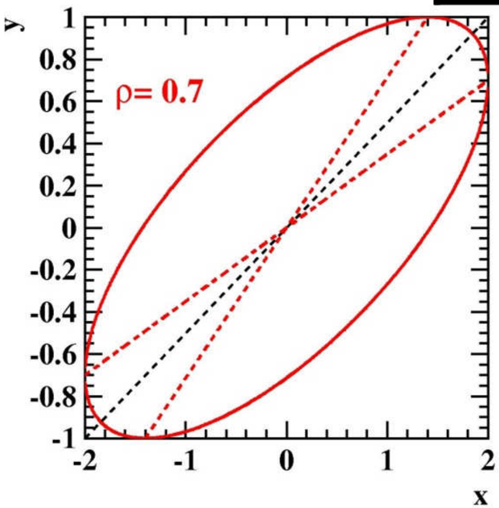
Kąt nachylenia dłuższej osi elipsy:

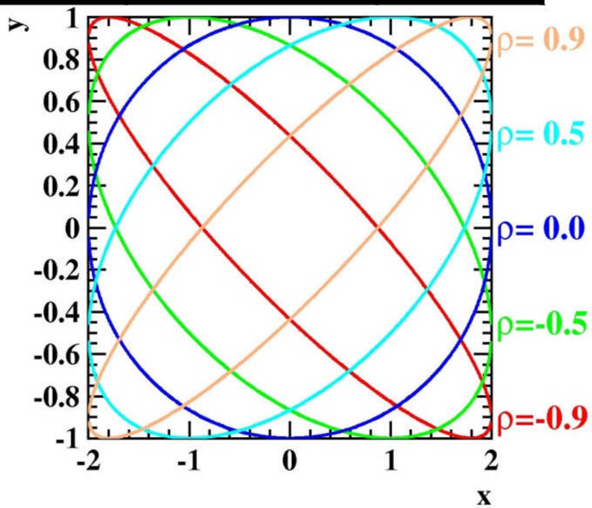
$$\tan 2\alpha=\frac{2\rho\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

Proste regresji II-go rodzaju:

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

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





Nierówność Chebyshev'a

Niech $x_1, x_2, \ldots, x_n$ będą niezależnymi i pochodzącymi z tego samego rozkładu (o wartości oczekiwanej μ i dyspersji σ) zmiennymi losowymi.

Wartość średnia: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \Rightarrow \quad \mathcal{E}[\bar{x}] = \mathcal{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}[x_i] = \frac{1}{n} n\mathcal{E}[x] = \mathcal{E}[x] \equiv \mu$

$$\mathcal{V}[\bar{x}] = \mathcal{V}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathcal{V}[x_i] = \frac{1}{n^2} n\mathcal{V}[x] \equiv \frac{\sigma^2}{n}$$

Twierdzenie: (Nierówność Chebyshev'a). Dla dowolnej zmiennej losowej X i dowolnej liczby $\varepsilon > 0$ zachodzi:

$$P\left(\left|x - \mathcal{E}[x]\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathcal{V}[x]$$

Dowód: $\mathcal{V}[x] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \geq$

$$\geq \int_{|x-\mu|\geq\varepsilon} (x-\mu)^2 f(x)\,dx \geq \int_{|x-\mu|\geq\varepsilon} \varepsilon^2 f(x)\,dx = \varepsilon^2 P\left(|x-\mu| \geq \varepsilon\right)$$

Uwaga: „Większość p-twa dla dowolnej zmiennej losowej skoncentrowana jest wokół wartości oczekiwanej w zakresie kilku dyspersji":

$$P\left(|x-\mu| < k\sigma\right) = 1 - P\left(|x-\mu| \geq k\sigma\right) \geq 1 - \frac{\mathcal{V}[x]}{k^2\sigma^2} = 1 - \frac{1}{k^2}$$

Prawo wielkich liczb

Przykład: Zastosowanie nierówności Chebyshev'a do rozkładu wykładniczego z parametrem $\lambda = 1$.

$$\mathbf{P}\left(|x-\mu|<k\sigma\right)=\mathbf{P}\left(|x-1|<k\right)=\mathbf{P}\left(1-k<\mathbf{x}<1+k\right)=\mathbf{P}\left(\mathbf{x}<1+k\right)=1-\mathbf{e}^{-k-1}$$

k	1	2	3	4
Chebyshev	0	0.750	0.889	0.938
P(|x-m|<k)	0.865	0.950	0.982	0.993

Zastosujmy nierówność Chebyshev'a do wartości średniej:

$$\mathbf{P}\left(|\overline{\mathbf{x}}-\mu|\geq\varepsilon\right)\leq\frac{1}{\varepsilon^2}\mathcal{V}[\overline{\mathbf{x}}]=\frac{\sigma^2}{n\varepsilon^2}$$

Math Player

Twierdzenie: (Prawo wielkich liczb). Dla dwu dowolnych liczb δ i ε istnieje taka liczba naturalna N, że dla wszystkich liczb naturalnych n > N zachodzi:

$$\mathbf{P}\left(|\overline{\mathbf{x}}_n-\mu|>\varepsilon\right)<\delta \quad\Leftrightarrow\quad \lim_{n\to\infty}\mathbf{P}\left(|\overline{\mathbf{x}}_n-\mu|>\varepsilon\right)=0$$

Math Player

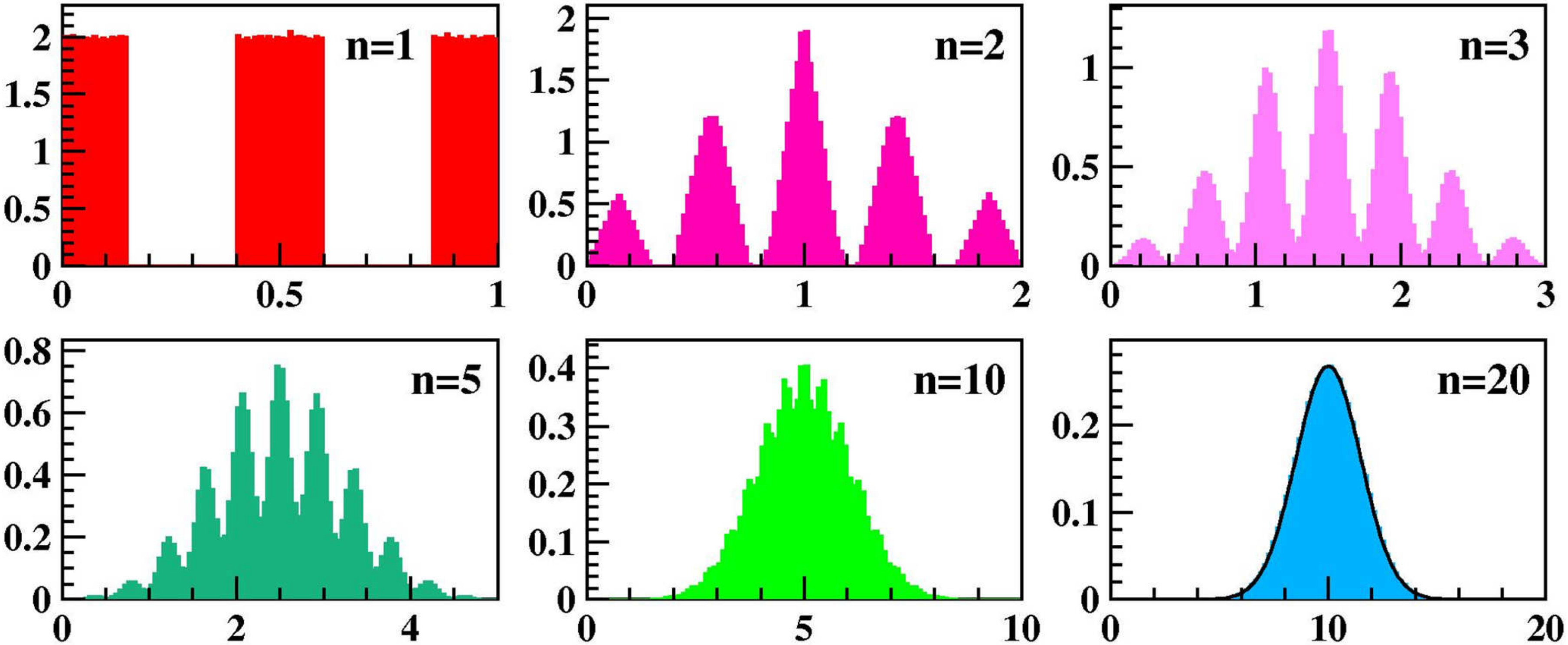
Centralne twierdzenie graniczne

Jeśli dany jest ciąg niezależnych zmiennych losowych $x_1, x_2, \ldots, x_n$ pochodzących z dowolnego rozkładu, o skończonych wartości oczekiwanej μ i dyspersji σ,

to rozkład gęstości zmiennej losowej $z_n = \dfrac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}$ gdzie $\bar{x}_n = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

Math Player

dąży do standaryzowanego rozkładu Gaussa.



Suma zmiennych poissonowskich

Przykład: Rozważmy sumę dwóch niezależnych zmiennych losowych $\mathbf{k}$ oraz $\mathbf{j}$ z rozkładu Poissona o tym samym parametrze μ. Łączny rozkład p-twa zmiennych $\mathbf{k}$ oraz $\mathbf{j}$ ma postać:

$$\mathrm{P}_{k,j}(\mu) = \mathcal{P}_k(\mu)\,\mathcal{P}_j(\mu) = \frac{\mu^k}{k!}e^{-\mu}\frac{\mu^j}{j!}e^{-\mu} = \frac{\mu^k\mu^j}{k!\,j!}e^{-2\mu}$$

Rozkład zmiennej losowej $\mathbf{m} = \mathbf{k}+\mathbf{j}$ otrzymujemy z powyższego łącznego rozkładu sumując po wszystkich parach $(\mathbf{k}, \mathbf{j})$ takich których suma jest stała i równa $\mathbf{m}$:

$$\mathrm{P}_m = \sum_{k+j=m} \mathrm{P}_{k,j}(\mu) = e^{-2\mu}\sum_{k=0}^{m}\frac{\mu^k\mu^{m-k}}{k!(m-k)!} = \frac{1}{m!}e^{-2\mu}\sum_{k=0}^{m}\binom{m}{k}\mu^k\mu^{m-k} = \frac{(2\mu)^m}{m!}e^{-2\mu}$$

Zmienna losowa $\mathbf{m}$ będąca sumą n niezależnych zmiennych losowych z rozkładu Poissona, będzie więc miała rozkład:

$$\mathcal{P}_m(n\mu) = \frac{(n\mu)^m}{m!}e^{-n\mu}$$

Ten sam rezultat otrzymamy korzystając z FGP:

$$g_m(t) = e^{\mu(t-1)} \quad\Rightarrow\quad g_{m_1+\ldots+m_n}(t) = \prod_{k=1}^{n} g_{m_k}(t) = e^{n\mu(t-1)}$$

Zachowanie graniczne r. Poissona

$$\mathcal{P}_k(\mu) = \frac{\mu^k}{k!}e^{-\mu} \cong \frac{\mu^k}{\sqrt{2\pi k}\,k^{k+\frac{1}{2}}e^{-k}}e^{-\mu} = \frac{\mu^{k+\frac{1}{2}}e^{k-\mu}}{\sqrt{2\pi\mu}\;k^{k+\frac{1}{2}}} = \frac{e^{k-\mu}}{\sqrt{2\pi\mu}}\left(\frac{k}{\mu}\right)^{-k-\frac{1}{2}} = \qquad \boxed{k \equiv \mu + x}$$

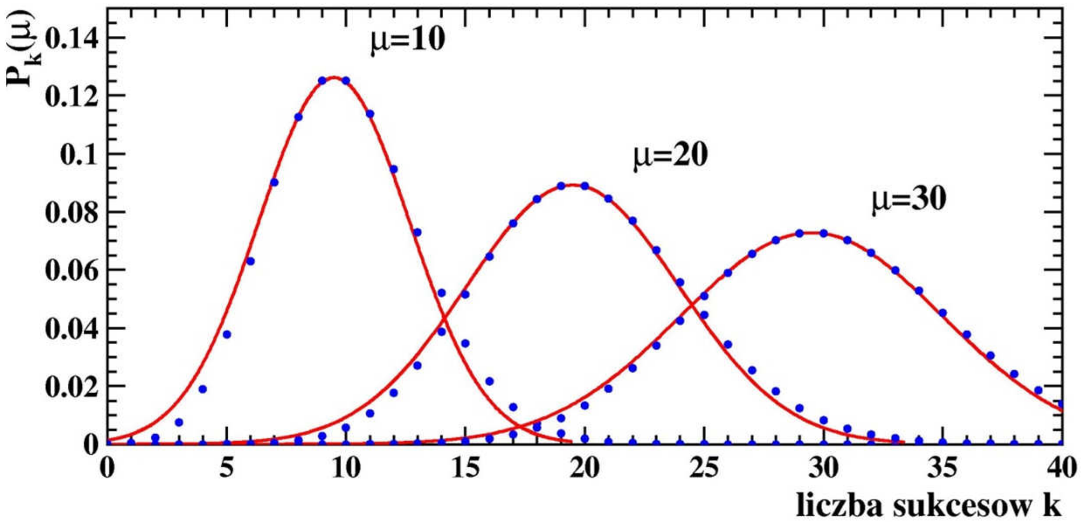
$$= \frac{e^x}{\sqrt{2\pi\mu}}\left(\frac{\mu+x}{\mu}\right)^{-\mu-x-\frac{1}{2}} = \frac{1}{\sqrt{2\pi\mu}}\exp\left(x - \left(\mu + x + \frac{1}{2}\right)\ln\left(1+\frac{x}{\mu}\right)\right) \cong \qquad \boxed{\ln(1+z) \cong z - \frac{1}{2}z^2}$$

$$\cong \frac{1}{\sqrt{2\pi\mu}}\exp\left(x - \left(\mu + x + \frac{1}{2}\right)\left(\frac{x}{\mu} - \frac{1}{2}\left(\frac{x}{\mu}\right)^2\right)\right) \cong \qquad \boxed{\mu(\mu-1) \cong \left(\mu - \frac{1}{2}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\mu}}\exp\left(-\frac{x^2+x}{2\mu}\right) =$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(k-(\mu-0.5)\right)^2}{2\sigma^2}\right)$$

$$\boxed{\sigma = \sqrt{\mu}}$$



Zachowanie graniczne r. dwumianowego

<u>Przykład:</u> P-two, że w czasie T przestanie świecić jedna żarówka jest równe $p = 0.1$. Jakie jest p-two, że w czasie T spośród $n = 100$ żarówek przestanie świecić od 7 do 19 przy założeniu, że żarówki przepalają się niezależnie?

Obliczenia bezpośrednio z rozkładu dwumianowego są czasochłonne:

$$\mathrm{P}(7 \le \mathrm{k} \le 19) = \sum_{k=7}^{19} \mathcal{B}_k(n = 100, p = 0.1) = \sum_{k=7}^{19} \binom{100}{k} (0.1)^k (1 - 0.1)^{100-k} = 0.8809$$

Korzystając z tw. de Moivre'a – Laplace'a mamy:

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

$$\mathrm{P}(7 \le \mathrm{k} \le 19) = \mathrm{P}\left(\frac{7 - 0.5 - np}{\sqrt{npq}} < \frac{\mathrm{k} - np}{\sqrt{npq}} < \frac{19 + 0.5 - np}{\sqrt{npq}}\right) =$$

$$= \mathrm{P}\left(\frac{7 - 0.5 - 10}{\sqrt{9}} < \mathrm{z} < \frac{19 + 0.5 - 10}{\sqrt{9}}\right) = \mathrm{P}(-1.17 < \mathrm{z} < 3.17) =$$

$$= \Phi(3.17) - \Phi(-1.17) = \Phi(3.17) + \Phi(1.17) - 1 = 0.9992 + 0.8790 - 1 = 0.8782$$

Bez „0.5” otrzymalibyśmy: $\mathrm{P}(7 \le \mathrm{k} \le 19) = \mathrm{P}(-1 \le \mathrm{z} \le 3) = \Phi(3) - \Phi(-1) \cong 0.8400$