Komputerowa Analiza Danych Doświadczalnych

Prowadząca:
dr inż. Hanna Zbroszczyk

e-mail: *gos@if.pw.edu.pl*
Tel: +48 22 234 58 51

Konsultacje:
Poniedziałek: 10.15 – 11.00
Piątek: 11.15-12.00

www: http://www.if.pw.edu.pl/~gos/students/kadd

Politechnika Warszawska
Wydział Fizyki
Pok. 117b (wejście przez 115)

Zmienne losowe

Jeśli w tych samych warunkach eksperymentalnych dokonujemy n pomiarów danej wielkości i dostajmy różne wyniki, to znaczy, że są one obarczone niepewnościami przypadkowymi, traktujemy je jako zdarzenia losowe.

Zdarzeniu losowemu przypisywana jest określona wartość **zmiennej losowej**.

Typy zmiennych losowych:

- jednowymiarowe,
- dwuwymiarowe,
- ...
- n-wymiarowe

Rodzaje zmiennych losowych:

- skokowe (dyskretne),
- ciągłe.

Zmienne losowe będziemy oznaczali znakami: *x*,*y*,..

Dystrybuanta zmiennej losowej

Dana jest zmienna losowa *x* oraz liczba x, która może przybierać dowolną wartość.

Prawdopodobieństwo zajścia zdarzenia *x*<x jest funkcją x i nosi nazwę

dystrybuanty zmiennej losowej *x*.

$$F(x)=P(\mathit{x}<x)$$

Jeśli *x* może przybierać tylko skończoną liczbę wartości, to dystrybuanta jest funkcją skokową.

Dystrubuanta dla zmiennej losowej typu skokowego:

$$F(x)=P(\mathit{x}<x)=\sum_{i=1}^{k} P_i$$

Dystrybuanta dla zmiennej losowej typu ciągłego, gdzie f(x) jest funkcją gęstości:

$$F(x)=P(\mathit{x}<x)=\int_{-\infty}^{x} f(x')dx'$$

Pojęcie prawdopodobieństwa

Własności dystrybuanty:

1) funkcja niemalejąca

2) $\lim_{x\to-\infty} F(x)=0$ oraz $\lim_{x\to\infty} F(x)=1$

3) w przypadku, kiedy dystrubuanta F(x) jest ciągła oraz ma 1-szą pochodną:

$$F'(x)=\frac{dF(x)}{dx}=f(x)$$

f(x) jest funkcją gęstości prawdopodobieństwa zmiennej losowej *x*

f(x) jest miarą prawdopodobieństwa zajścia zdarzenia $x \leqslant \mathit{x} \leqslant x+dx$

Prawdopodobieństwo:

$$P(\mathit{x}<a)=F(a)=\int_{-\infty}^{a} f(x)\,dx$$

$$P(a\leqslant \mathit{x}\leqslant b)=\int_{a}^{b} f(x)\,dx=F(b)-F(a)$$

$$\int_{-\infty}^{\infty} f(x)\,dx=1$$

Położenia kątowe wskazówek zegara – zmienna losowa ciągła

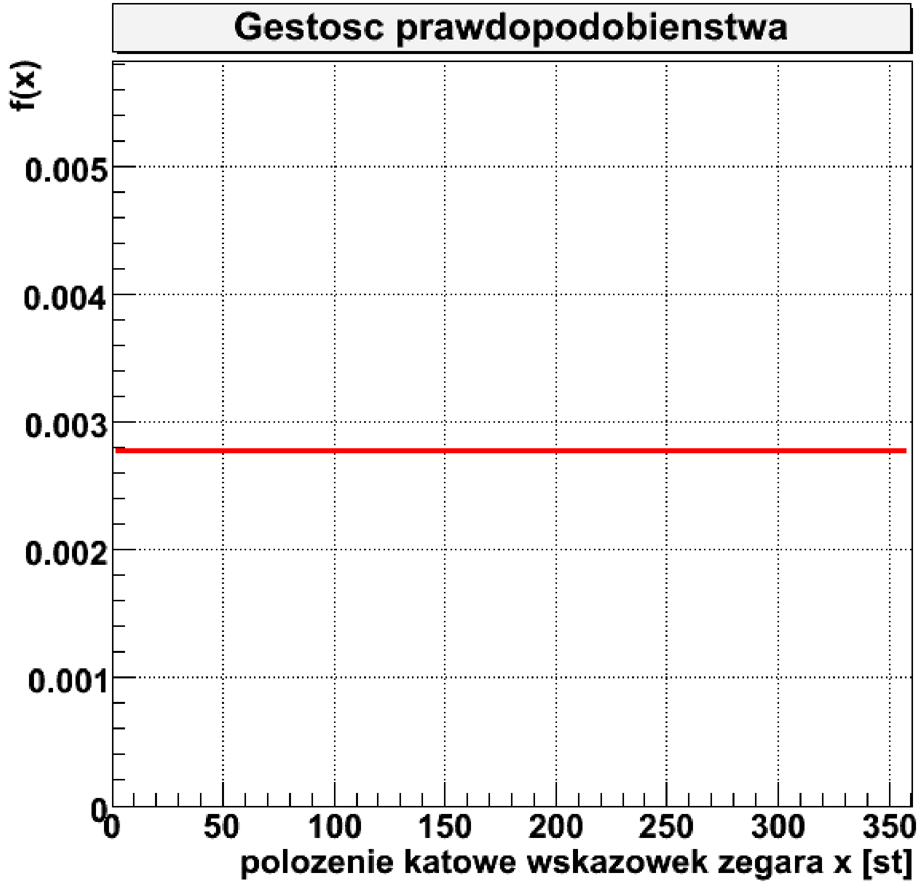
$$f(x)=\frac{1}{360};\, x\in[0,360]$$

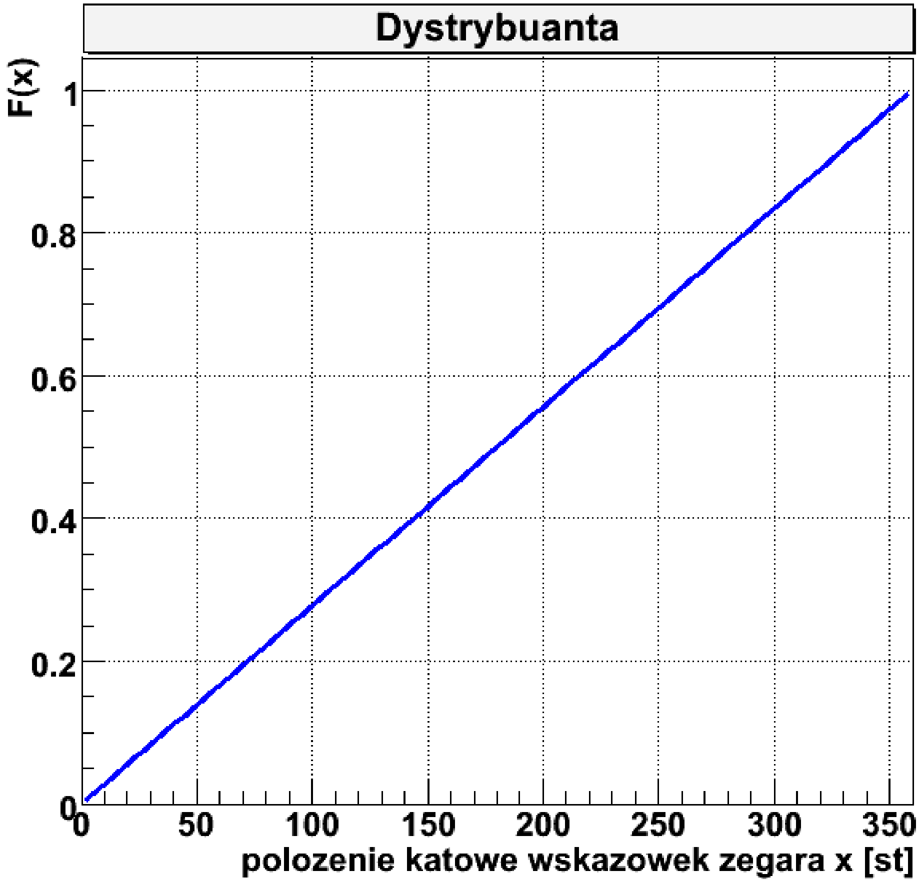
$$f(x)=0;\, x\notin[0,360]$$

$$F(x)=0;\, x<0$$

$$F(x)=\int_0^x f(x')dx'=\frac{1}{360}\int_0^x dx'=\frac{1}{360}x$$

$$F(x)=1;\, x>360$$





Rozkład normalny – zmienna losowa ciągła

$$f(x)=\frac{1}{\sqrt{2\pi}b}\exp\left(\frac{(x-a)^2}{2b^2}\right);\ x\in R$$

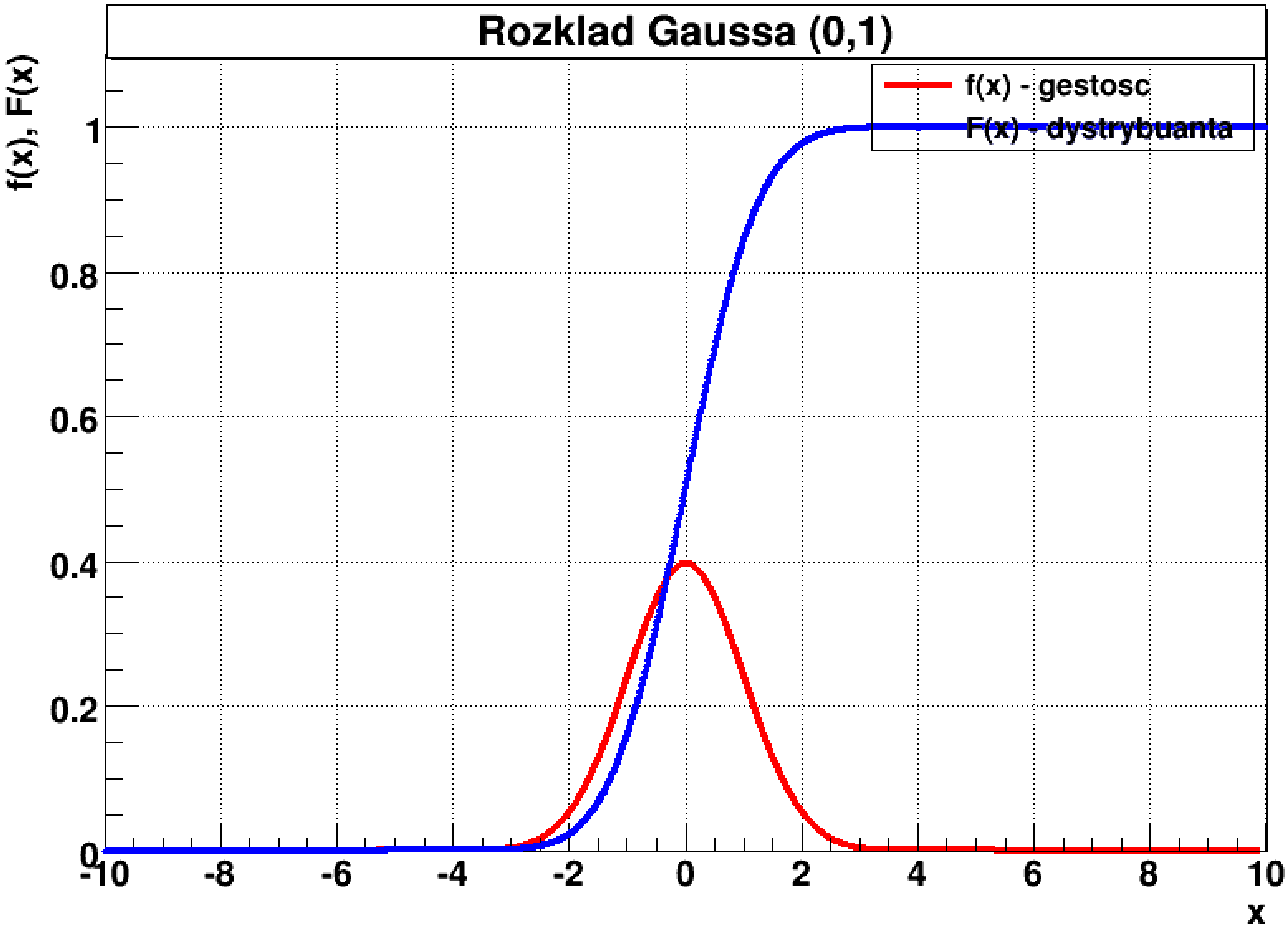
$$F(x)=\int_{-\infty}^{x} f(x')\,dx'$$

$$\lim_{x\to-\infty} F(x)=0$$

$$\hat{x}=a$$

$$\lim_{x\to\infty} F(x)=1$$

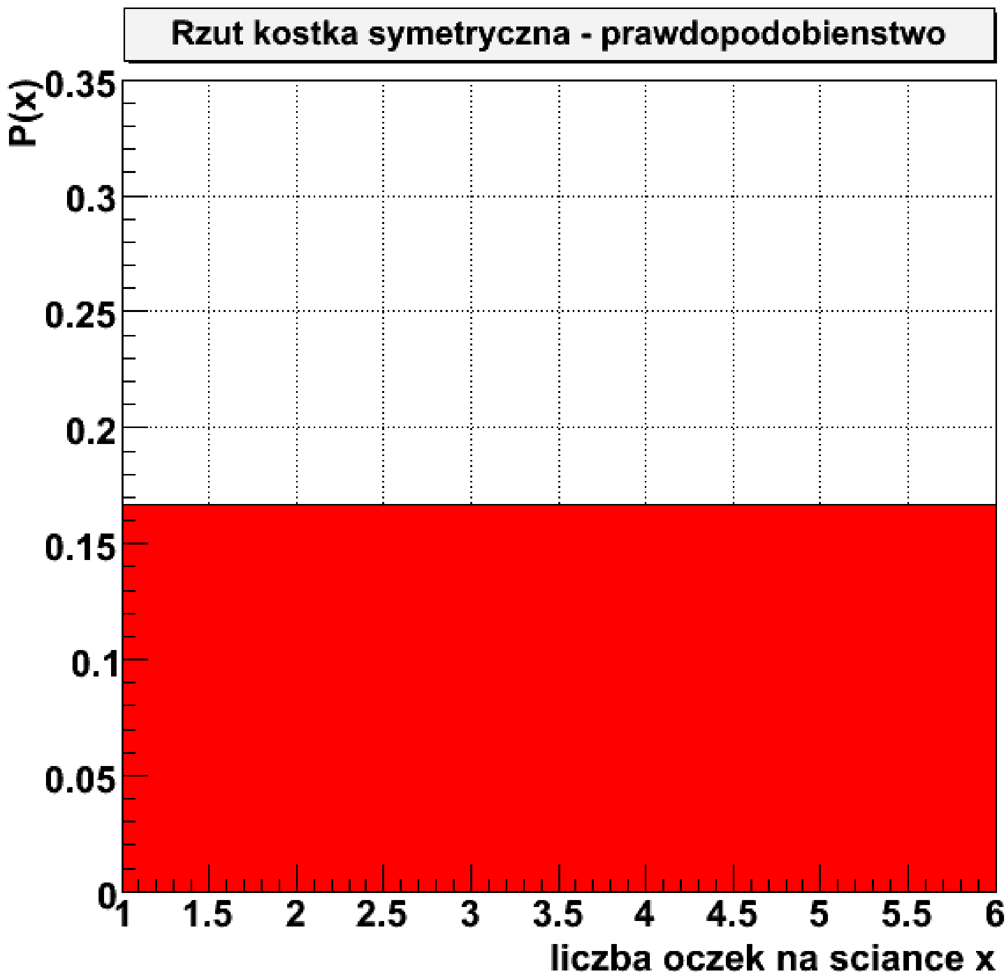
$$\sigma=b$$

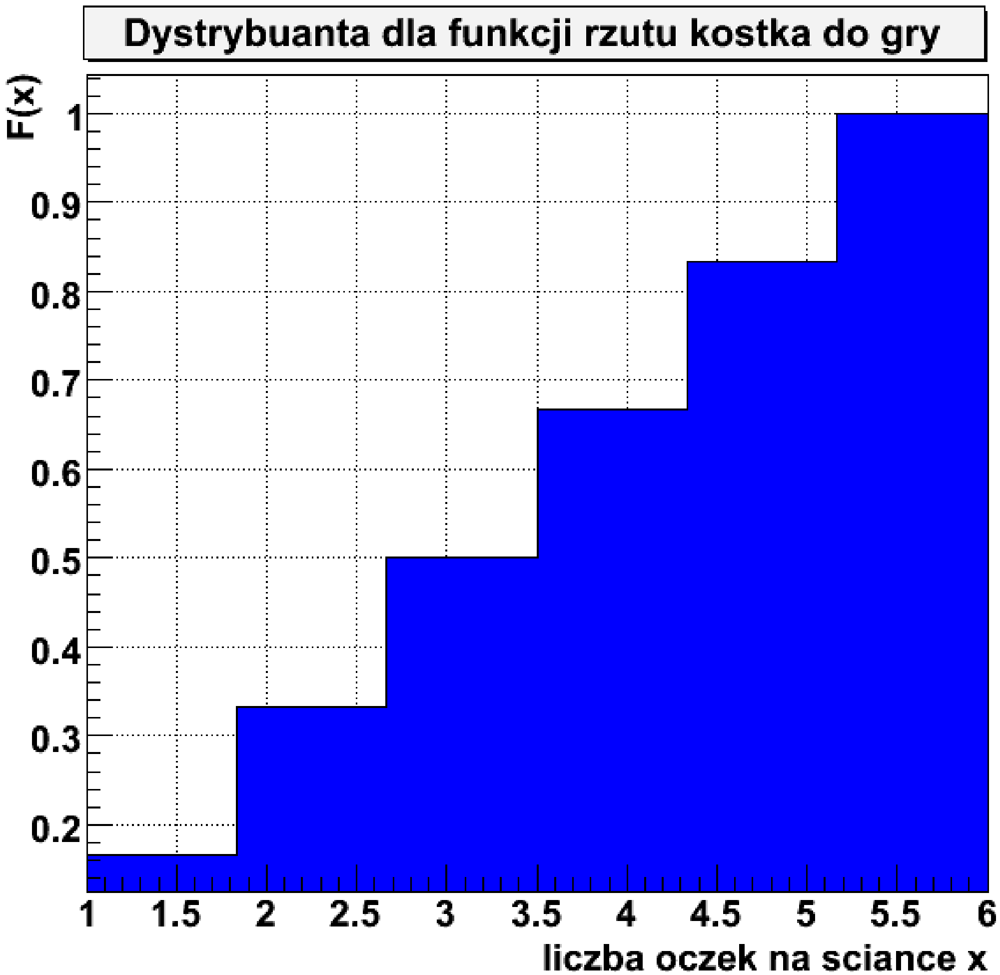


Rzut kostką – zmienna losowa dyskretna

$P(x_i)=\frac{1}{6};i=\{1,2,3,4,5,6\}$

$F(x_i)=\frac{1}{6}i;i=\{1,2,3,4,5,6\}$





Wartość średnia

Funkcje zmiennej losowej *x* są zmienną losową. $y = \mathrm{H}(x)$

Zmienna losowa y ma swoją dystrybuantę oraz gęstość prawdopodobieństwa.

Wartość średnia (wartośc oczekiwana, wartość przeciętna) dla zmiennej loswej typu skokowego:

$$\mathrm{E}(x) = \hat{\mathrm{x}} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}\, \mathrm{P}(x = \mathrm{x_i})$$

Wartość średnia dla funkcji *y*= H(*x*)

$$\mathrm{E}\{\mathrm{H}(x)\} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{H}(\mathrm{x_i})\, \mathrm{P}(x = \mathrm{x_i})$$

Wartość średnia (oczekiwana, przeciętna) dla zmiennej loswej typu ciągłego:

$$\mathrm{E}(x) = \hat{\mathrm{x}} = \int_{-\infty}^{\infty} \mathrm{x}\, \mathrm{f}(\mathrm{x})\, \mathrm{dx}$$

Wartość średnia dla funkcji *y*=H(*x*)

$$\mathrm{E}\{\mathrm{H}(x)\} = \int_{-\infty}^{\infty} \mathrm{H}(\mathrm{x})\, \mathrm{f}(\mathrm{x})\, \mathrm{dx}$$

Momenty

W przypadku, kiedy H(x) = (x-c)^l

Momenty rzędu l względem c: wartości oczekiwane

$$\mu_l = E\{(x-c)^l\}$$

Ważne momenty:

Moment centralny: moment wartości średniej $\mu_l = E\{(x-\hat{x})^l\}$

Najniższe momenty: $\mu_0 = 1$ $\mu_1 = 0$

Wariancja zmiennej losowej x

$$\mu_2 = \sigma^2(x) = \mathrm{var}(x) = E\{(x-\hat{x})^2\}$$

Najniższy moment, który zawiera informację

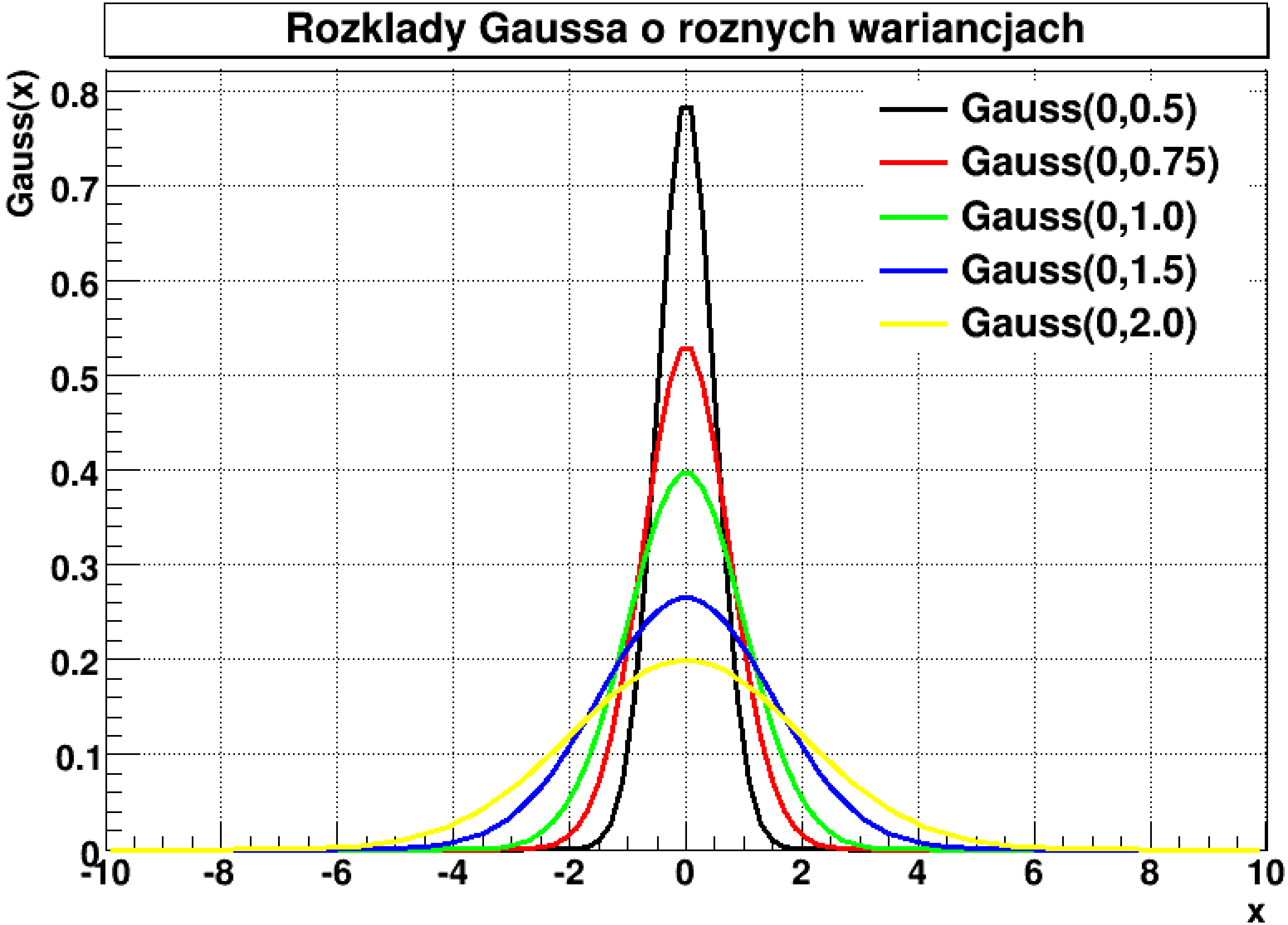
o średnim odchyleniu zmiennej losowej x od swojej wartości średniej.

$$\sigma^2(x) = \mathrm{var}(x) = \int_{-\infty}^{\infty} (x-\hat{x})^2 f(x)\, dx$$

$$\sigma^2(x) = \mathrm{var}(x) = \sum_{i=1}^{n} (x_i-\hat{x})^2 P(x = x_i)$$

Rozkład normalny o różnych wariancjach (dyspersjach)

$$f(x)=\frac{1}{\sqrt{2\pi}b}\exp\left(\frac{(x-a)^2}{2b^2}\right);\ x\in R$$



Wariancja, dyspersja, skośność

Wariancja jest miarę szerokości rozkładu gęstości prawdopodobieństwa w pobliżu wartości średniej. Jeżeli wariancja jest mała, to wyniki poszczególnych pomiarów leżą w pobliżu wartości średniej, jeśli natomiast jest ona duża, to wyniki są bardziej rozproszone wokół średniej.

Odchylenie standardowe (dyspersja) $\sigma(x)=\sqrt{\sigma^2(x)}$

jest miarą odchylenia wyników od wartości oczekiwanej.

Odchylenie standardowe ma ten sam wymiar co zmienna losowa x, jest utoższamiane z błędem pomiaru.

$$\sigma(x)=\Delta x$$

Skośność – trzeci moment względem wartości średniej μ_3 (współczynnik skośności, skośność)

Współczynnik asymetrii rozkładu zmiennej losowej *x* $\gamma=\frac{\mu_3}{\sigma^3}$ Zawiera on informacje o możliwych różnicach pomiędzy dodatnimi i ujemnymi odchyleniami od wartości średniej. Dla rozkładów symetrycznych skośność jest równa zeru.

Własności wartości średniej oraz wariancji

Własności: $\mathrm{H}(x)=\mathrm{c}\,x \qquad \mathrm{c}=\mathrm{const}$

Wartość oczekiwana: $\mathrm{E}(\mathrm{c}\,x)=\mathrm{c}\,\mathrm{E}(x)$

$$\sigma^2(\mathrm{c}\,x)=\mathrm{c}^2\,\sigma^2(x)$$

$$\sigma^2(x)=\mathrm{E}\{(x-\hat{\mathrm{x}})^2\}=\mathrm{E}\{x^2-2\,x\,\hat{\mathrm{x}}+\hat{\mathrm{x}}^2\}=\mathrm{E}(x^2)-\hat{\mathrm{x}}^2=\mathrm{E}(x^2)-(\mathrm{E}(x))^2$$

Rozpatrzmy funkcję: $u=\frac{x-\hat{\mathrm{x}}}{\sigma(x)}$

wariancja: $\sigma^2(u)=\frac{1}{\sigma^2(x)}\mathrm{E}\{(x-\hat{\mathrm{x}})^2\}=\frac{\sigma^2(x)}{\sigma^2(x)}=1$

Zmienna standaryzowana – zmienna losowa o wartości oczekiwanej 0 oraz odchyleniu standardowym 1.

Wartość modalna, mediana

Wartość modalną (wartość najbardziej prawdopodobna, moda) x_m, to wartość zmiennej losowej odpowiadającą maksimum prawdopodobieństwa $P(x = x_m) = \max$

Rozkład jednomodalny – gęstość prawdopodobieństwa ma jedno maksimum

Rozkład wielomodalny – gęstość prawdopodobieństwa ma więcej niż jedno maksimum

Dla rozkładów gęstości prawdopodobieństwa mających pierwszą oraz drugą pochodną wartość modalna odpowiada maksimum rozkładu, jest określona przez warunki:

$$\frac{df(x)}{dx} = 0 \qquad \frac{d^2 f(x)}{dx^2} < 0$$

Mediana $x_{0.5}$ rozkładu, to taka wartość zmiennej losowej, dla której dystrybuanta przyjmuje wartość 0.5:

Dla ciągłej gęstości prawdopodobieństwa: $F(x_{0.5}) = P(x < x_{0.5}) = 0.5$

mediana dzieli cały zakres zmiennej loswej na dwa obszary o równym prawdopodbieństwie.

$$\int_{-\infty}^{x_{0.5}} f(x)\,dx = 0.5$$

Kwantyle

Kwantyl x_q: $F(x_q)=\int_{-\infty}^{x_q} f(x)dx=q \qquad q\in[0,1]$

Kwantyl górny $x_{0.25}$: $F(x_{0.25})=\int_{-\infty}^{x_{0.25}} f(x)dx=0.25$

Kwantyl dolny $x_{0.75}$: $F(x_{0.75})=\int_{-\infty}^{x_{0.75}} f(x)dx=0.75$

Mediana $x_{0.5}$: $F(x_{0.5})=\int_{-\infty}^{x_{0.5}} f(x)dx=0.5$

Decyle: $x_{0.1}$, $x_{0.2}$, $x_{0.3}$,...

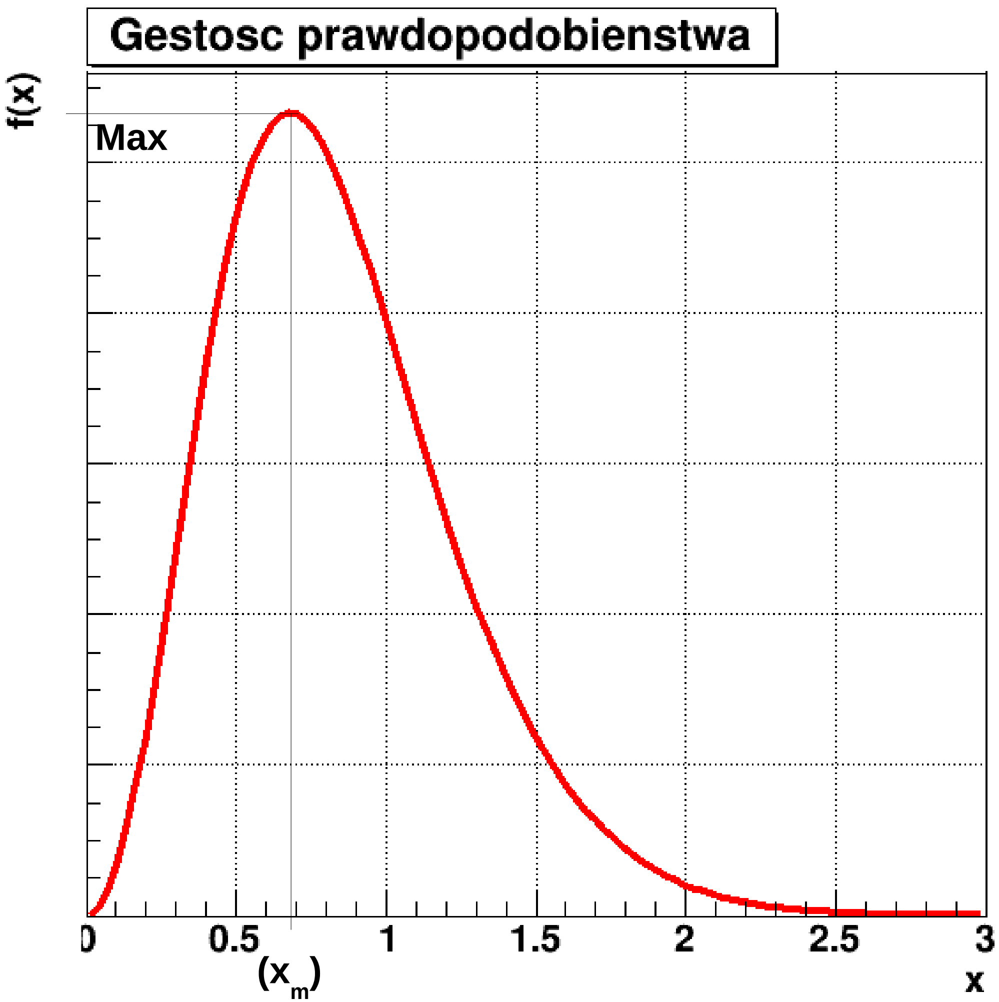
Dla rozkładów jednomodalnych, o ciągłej funkcji gęstości prawdopodobieństwa, symetrycznych Wokół wartości średniej – wartość średnia, wartość modalna oraz mediana są identyczne.

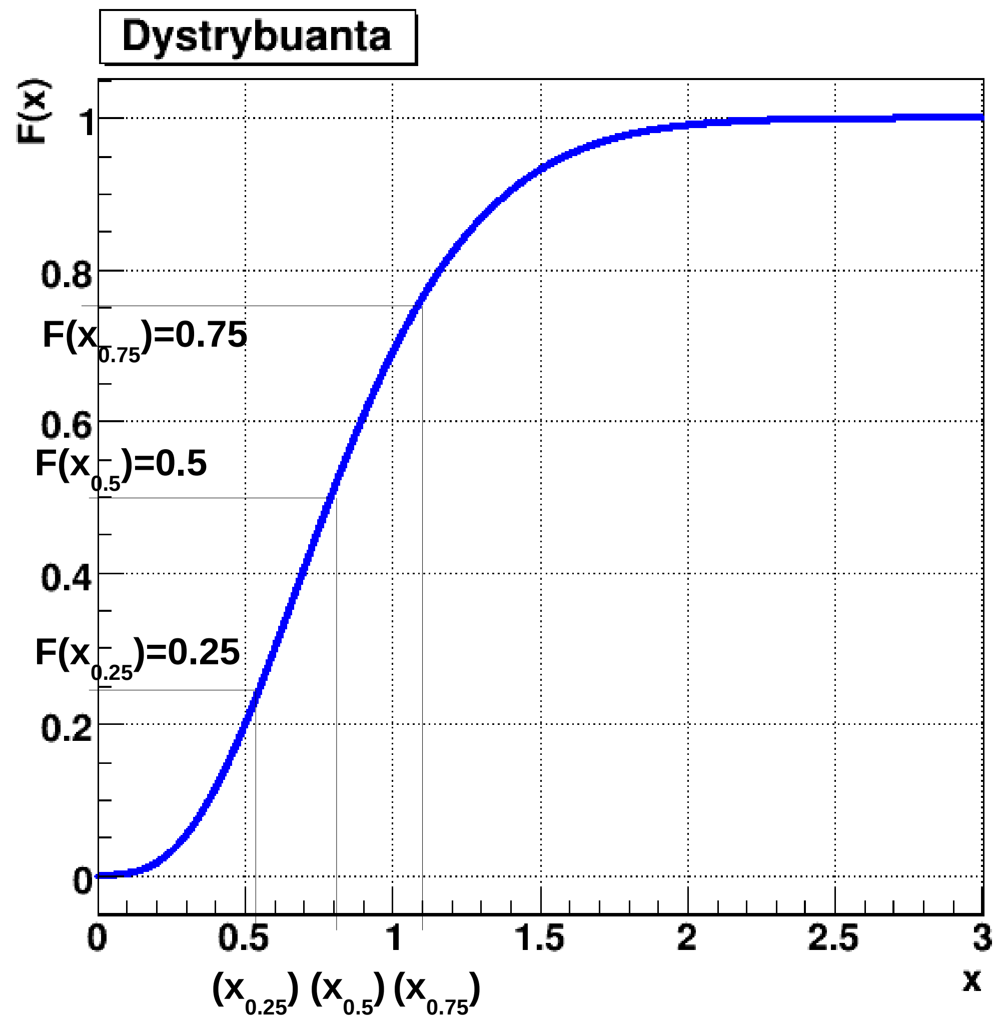
Kwantyle - ilustracja

$$F(x_{0.5}) = \int_{-\infty}^{x_{0.5}} f(x)\,dx = 0.5$$ Mediana

$$F(x_{0.25}) = \int_{-\infty}^{x_{0.25}} f(x)\,dx = 0.25$$ Kwantyl dolny

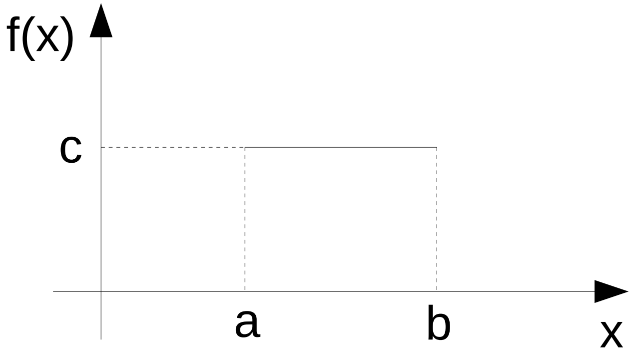
$$F(x_{0.75}) = \int_{-\infty}^{x_{0.75}} f(x)\,dx = 0.75$$ Kwantyl górny





Rozkład jednostajny

Gęstość prawdopodobieństwa:



$$f(x)=c\,;\,x\in(a,b)$$

$$f(x)=0\,;\,x\notin(a,b)$$

Rozkład jednostajny:

$$\int_{-\infty}^{\infty} f(x)\,dx=c\int_a^b dx=c(b-a)=1$$

$$f(x)=\frac{1}{b-a}\,;\,x\in(a,b)$$

$$f(x)=0\,;\,x\notin(a,b)$$

Dystrybuanta:

$$F(x)=0\,;\,x\le a$$

$$F(x)=\frac{1}{b-a}\int_a^x dx=\frac{x-a}{b-a}\,;\,x\in(a,b)$$

$$F(x)=1\,;\,x\ge b$$

Wartość oczekiwana:

$$E(x)=\hat{x}=\frac{1}{b-a}\int_a^b x\,dx=\frac{1}{2}\frac{1}{b-a}(b^2-a^2)=\frac{b+a}{2}$$

Wariancja:

$$\sigma^2(x)=\int_a^b (x-\hat{x})^2 f(x)\,dx=\int_a^b \left(x-\frac{b+a}{2}\right)^2\frac{1}{b-a}dx=\frac{1}{b-a}\int_a^b\left(x-\frac{b+a}{2}\right)^2 dx$$

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

Rozkład dwumianowy

Prawdopodobieństwo: $p_n(k)=\frac{n!}{k!(n-k)!}p^k q^{n-k}$ $p\in[0,1]$ $q=1-p$

k sukcesów w n niezależnych próbach przeprowadzonych w identycznych warunkach

p – prawdopodobieństwo sukcesu w pojedynczej próbie

q= 1-p – prawdopodobieństwo porażki w pojedynczej próbie

Wartość oczekiwana w pojedynczej próbie:

$$E(x_i)=1\,p+0\,q=p$$

Wariancja w pojedynczej próbie:

$$\mathrm{var}(x_i)=E\{(x_i-p)^2\}=(1-p)^2 p+(0-p)^2 q=pq$$

Wartość oczekiwana:

$$E(x)=np$$

Wariancja:

$$\mathrm{var}(x)=npq$$

Odchylenie standardowe:

$$\sigma(x)=\sqrt{(\mathrm{var}(x))}=\sqrt{(npq)}$$

Rozkład dwumianowy

Prawdopodobieństwo: $p_n(k)=\frac{n!}{k!(n-k)!}p^k q^{n-k}$ $\quad p\in[0,1]$ $\quad q=1-p$

Wartość oczekiwana:

$$E(x)=\sum_{k=0}^{n} k\, p_n(k)=\sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$E(x)=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}=np(p+q)^{n-1}=np$$

Wariancja:

$$V(x)=E(x^2)-(E(x))^2$$

$$E(x^2)=\sum_{k=0}^{n} k^2\frac{n!}{k!(n-k)!}p^k q^{n-k}=\sum_{k=1}^{n}((k-1)+1)k\frac{n!}{k!(n-k)!}p^q q^{n-k}$$

$$E(x^2)=\sum_{k=2}^{n}\frac{n!}{(k-2)!(n-k)!}p^k q^{n-k}+\sum_{k=1}^{n}\frac{n!}{(k-1)!(n-k)!}p^k q^{n-k}$$

$$E(x^2)=n(n-1)p^2\sum_{k=2}^{n}\frac{(n-2)!}{(k-2)!(n-k)!}p^k q^{n-k}+np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}$$

$$E(x^2)=n(n-1)p^2(p+q)^{n-2}+np(p+q)^{n-1}=n(n-1)p^2+np$$

$$\mathrm{var}(x)=n(n-1)p^2+np-(np)^2=npq$$

Odchylenie standardowe:

$$\sqrt{\mathrm{var}(x)}=\sqrt{(npq)}$$