Metody Statystyczne

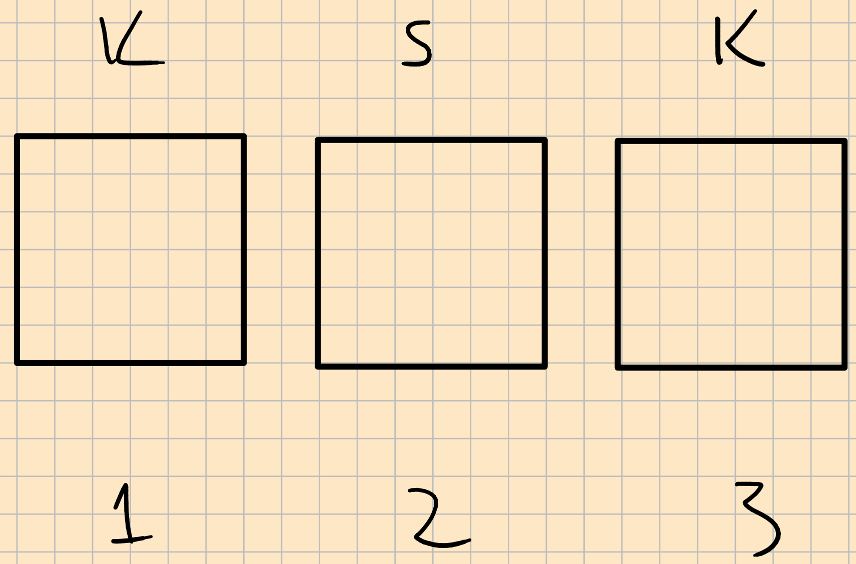
wykład 1

17 X 2020

kacper.topolnicki@uj.edu.pl

- problem „Monty Hall"
- aksjomaty prawdopodobieństwa (Kołmogorow)
- własności prawdopodobieństwa
  - prawo Bayes'a
- zmienne losowe
- ciągłe zmienne losowe, rozkłady prawdop.
  - funkcje gęstości prawdopodobieństwa
- własnościach rozkładów prawdopodobieństwa

PARADOKS MONTIEGO HALLA



K – koza
S – samochód

przebieg teleturnieju:
- u. wybiera zasłonę
  $1 \vee 2 \vee 3$
- M.H. odsłania kozę
- u. może zmienić
  wybór

Czy warto zmienić wybór?

AKSJOMATY PRAWDOPODOBIEŃSTWA

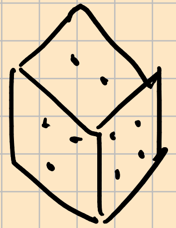
(Kołmogorow 1933)

– (Ω, F, P)

zbiór wszystkich „zdarzeń elementarnych"

przestrzeń wszystkich „zdarzeń losowych"

funkcja przypisująca „zdarzeniu losowemu" „prawdopodobieństwo" $[0, 1]$



(Ω, F, P)

Ω = { ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ }

F - podzbiory Ω

F = { ∅, ..., ..., }

zbiór pusty

P({⚀}) = 1/6, P({⚁}) = 1/6, P({⚂}) = 1/6

P({⚃}) = 1/6, P({⚄}) = 1/6, P({⚅}) = 1/6

P({⚁, ⚃, ⚅}) = ? (1/2)

(Ω, F, P)

- $\forall E \in F : P(E) \in \mathbb{R} \wedge P(E) \geq 0$

- $\underbrace{P(\Omega)} = 1$

prawdopodobieństwo zajścia przynajmniej jednego zd. el.

- $E_i \in F \qquad \forall_{i,j}\ E_i \cap E_j = \emptyset$:

$$P(E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_N) = \sum_{i=1}^{N} P(E_i)$$

WŁAŚCIWOŚCI PRAWDOPODOBIEŃSTWA

- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(\emptyset) = 0$
- $P(A^*) = P(\Omega \setminus A) = 1 - P(A)$

  $$A^* = \Omega \setminus A$$
- $\forall E \in F: \quad 0 \leq P(E) \leq 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

ZADANIE:

P({⚀}) = $\frac{1}{6}$, P({⚁}) = $\frac{1}{6}$, P({⚂}) = $\frac{1}{6}$

P({⚃}) = $\frac{1}{6}$, P({⚄}) = $\frac{1}{6}$, P({⚅}) = $\frac{1}{6}$

$+$

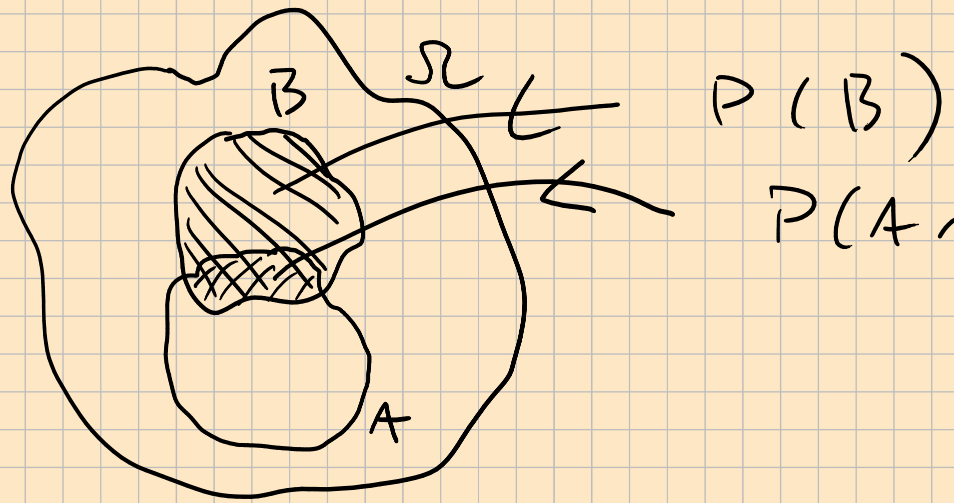
aksjomaty, właściwości

$\downarrow$

P({⚁, ⚃, ⚅}) = ?

DODATKOWE WŁASNOŚCI:

$$P(A|B) \equiv \frac{P(A\cap B)}{P(B)}$$



$$-P(A\cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$$

$P(A|B) = P(A\cap B)/P(B)$ $\quad$ $P(B|A) = P(B\cap A)/P(A)$

$$A\cap B = B\cap A$$

$$P(A\cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$$

$$- \underbrace{P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)\, P(A)}{P(B)}}_{\text{twierdzenie Bayesa}}$$

MH - BEZ ZMIANY WYBORU

A - samochód za 1

B - koza za 2

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{3}{6}} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$

$$P(B|A) = \frac{1}{2}$$

$$P(A) = \frac{1}{3}$$

$$P(B) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 = \frac{3}{6}$$

MH - ZE ZMIANĄ WYBORU

C_i - samochód za zasłoną i

X_i - pierwszy wybór gracza $= i$

H_i - odsłonięcie zasłony i

tw. Bayesa

$$P\Big(C_2 \mid \underbrace{(H_3 \cap X_1)}_{\substack{\text{gracz } 1 \\ \text{odsłonięta } 3}}\Big) = \frac{P(C_2 \cap H_3 \cap X_1)}{P(H_3 \cap X_1)} = \frac{P(H_3 \mid (C_2 \cap X_1))\, P(C_2 \cap X_1)}{P(H_3 \cap X_1)} = \ldots$$

samochód za 2

gdy: gracz 1, odsłonięta 3

||

prawdopodobieństwo zwycięstwa

$$\ldots = \frac{P(H_3 \mid (C_2 \cap X_1)) \; P(C_2 \cap X_1)}{P(H_3 \mid X_1) \; P(X_1)} =$$

$$P(H_3 \mid (C_1 \cap X_1)) = \frac{1}{2} \qquad P(H_3 \mid X_1) = \frac{1}{2}$$

$$P(H_3 \mid (C_2 \cap X_1)) = 1 \qquad P(C_i \cap X_i) = \underbrace{P(C_i) \cdot P(X_i)}_{\text{niezależne}}$$

$$P(H_3 \mid (C_3 \cap X_1)) = 0 \qquad P(X_i) = \frac{1}{3}$$

$$= \frac{1 \cdot \frac{1}{3} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3} \approx 0.6666\ldots$$

INTERPRETACJE PRAWDOPODOBIEŃSTWA

Kołmogorow

- prawdopodobieństwo obiektywne (frequentist probability)
- prawdopodobieństwo subiektywne (Bayesian probability)
  - „likelyhood"

ZMIENNE LOSOWE

- zmienna, której wartość zależy od wyniku procesu losowego
- formalnie:

  zmienna losowa jest funkcja:

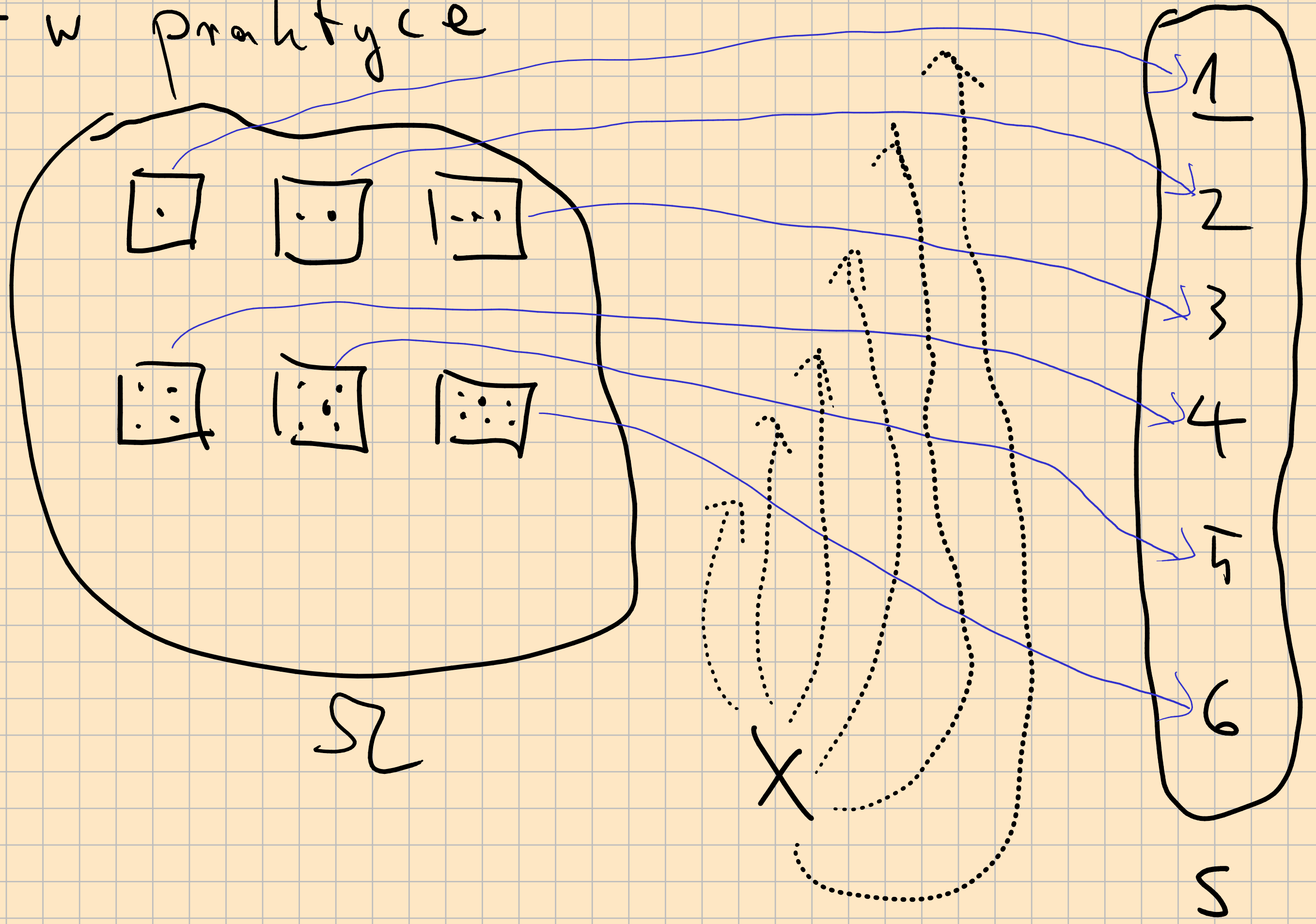
  $$X : \Omega \longrightarrow E$$

  „przestrzeń mierzalna"

  np. $E = \mathbb{R}$

$$\underbrace{P(\overbrace{X \in S}^{S \subseteq E})}_{\text{prawdopodobieństwo zmiennej losowej}} = P\left(\{\omega \in \Omega \mid X(\omega) \in S\}\right)$$

- w praktyce



X(⚀) = 1

X(⚁) = 2

X(⚂) = 3

X(⚃) = 4

X(⚄) = 5

X(⚅) = 6

$$P(X \text{ jest parzyste}) =$$

$$= P(\{⚁, ⚃, ⚅\}) = \frac{1}{2}$$

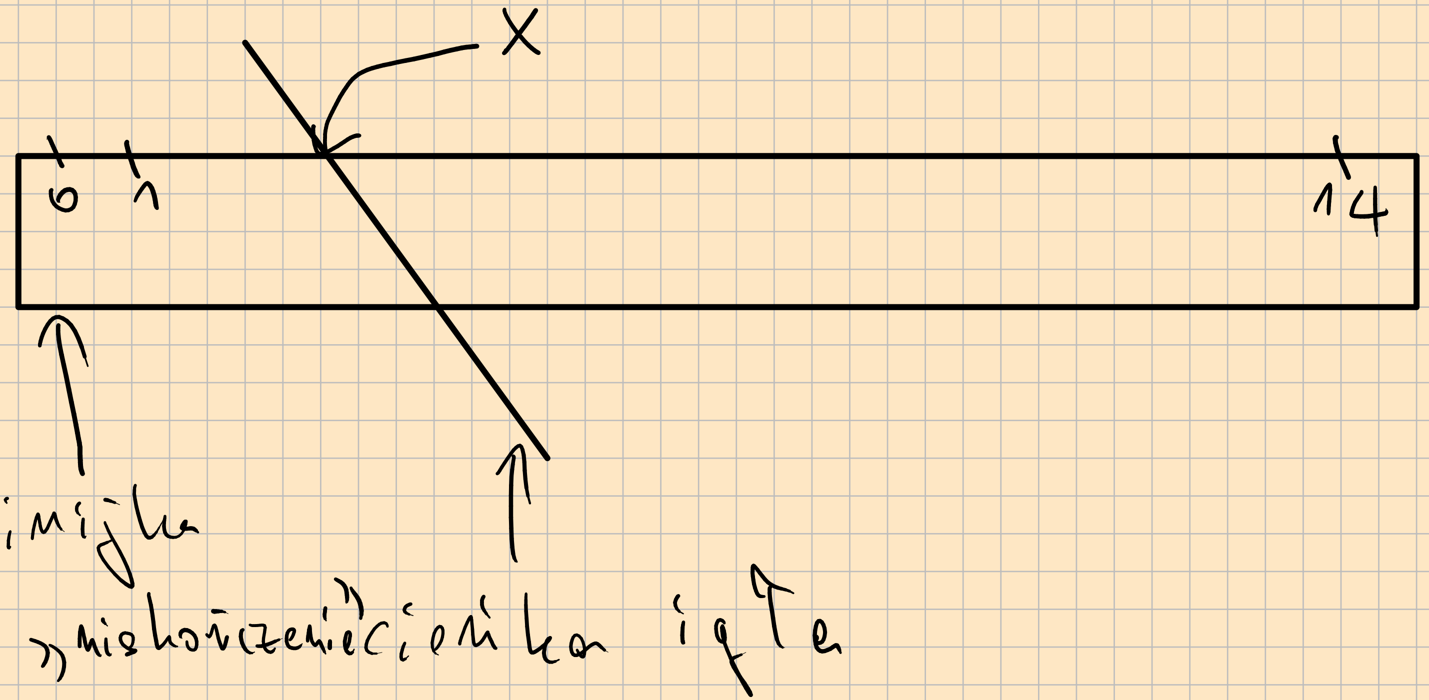
$$X(\ldots) = 2 \quad 4 \quad 6$$

$$P(X=2) = P(\{⚁\}) = \frac{1}{6}$$

CIĄGŁE ZMIENNE LOSOWE

- do tej pory skończone zbiory (rzut kostką)

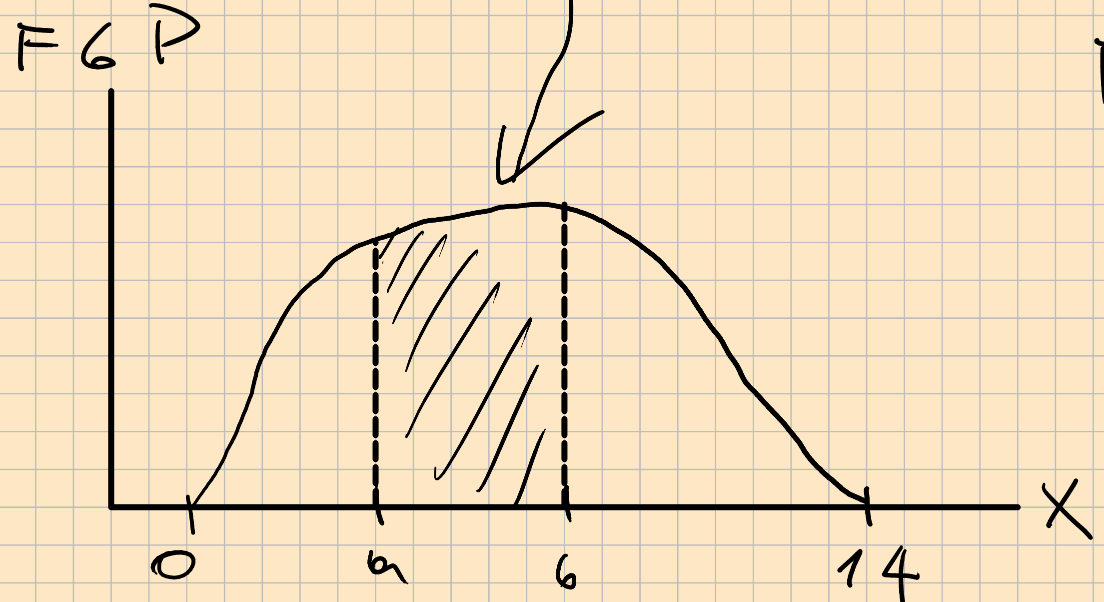
-



$$P(x=3)=?=0$$

— funkcja gęstości prawdopodobieństwa

$$P(a \le X \le b) = \int_a^b \underbrace{f_X(u)}_{\text{FGP}}\, du$$



- normalizacja:

$$\int_0^{14} f_X(u)\,du = 1$$

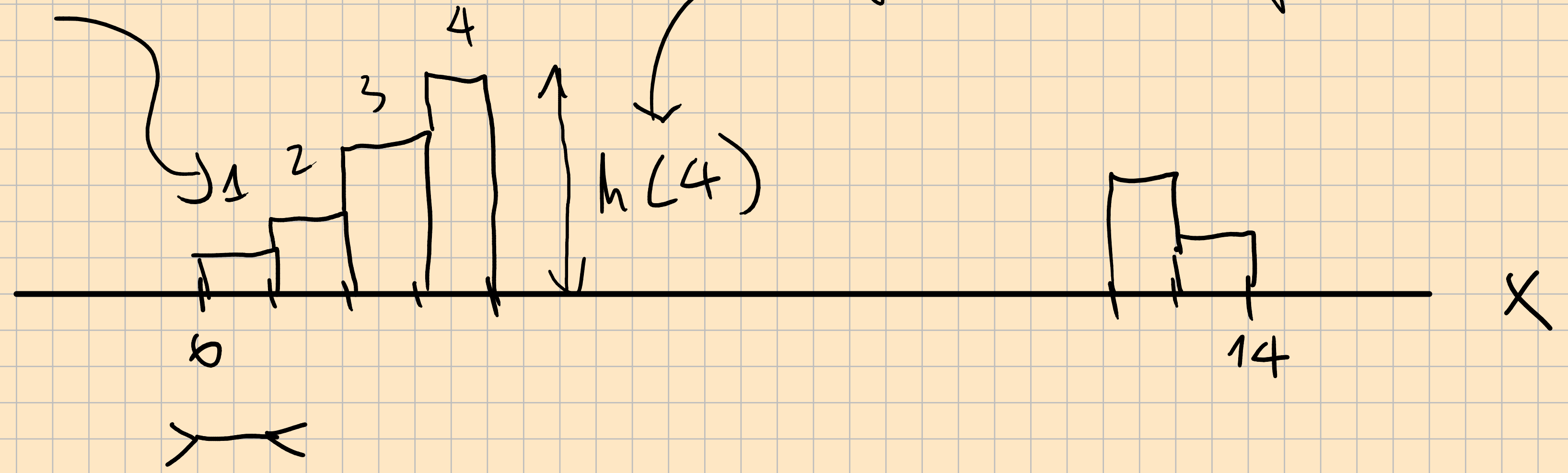
$$P(y < X \leq y + dy) = f_X(y)\,dy$$

$$\forall_y \; P(y \leq X \leq y + dy) > 0 \Rightarrow f_X(y) > 0$$

HISTOGRAMY

numer „binu"

wysokość słupka nr 4



w - szerokość „binu"

A) Jakie jest prawdopodobieństwo, że X „wpadnie" w „bin" o numerze i?

$$P\left((i-1)w \le x \le i \cdot w\right) = \int_{(i-1)w}^{i \cdot w} f_X(u)\, du$$

B) Ile będzie „zliczeń" w „binie" i?

$$N \cdot P\left((i-1) \cdot w \le x \le i \cdot w\right)$$

całkowita liczba przypadków

c) Jak ma się FGP do l. "zliczeń" w binie numer i?

$$P\big((i-1)w \le X \le i\cdot w\big) = \int_{(i-1)w}^{i\cdot w} f_X(u)\,du \approx$$

$$\approx f_X\Big(\underbrace{\tfrac{1}{2}\big((i-1)\cdot w + i\cdot w\big)}_{\text{wartość FGP w środku "binu" } i}\Big)\cdot w$$

$$f_X(\ldots) \approx \frac{P(\ldots)}{w}$$

Jak wybrać wysokość słupka:

$\underbrace{(A),\ (B),\ (C)}$

większość bibliotek rysujących histogramy pozwala na wybór sposobu liczenia wysokości słupków

A) $h(6) = P(\ldots \leq X < \ldots)$

B) $h(6) \approx N \cdot P(\ldots < X \leq \ldots)$

C) $h(6) = \underbrace{\frac{P(\ldots < X \leq \ldots)}{w}}_{FGP}$

OPIS FGP

- wartości oczekiwane

$$E(X) = \int_{-\infty}^{+\infty} x \, f_X(u) \, du$$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) \, f_X(u) \, du$$

stała

$$g(x) = C$$

$$E(C) = \int_{-\infty}^{+\infty} C \, f_X(u) \, du = C \cdot \underbrace{\int_{-\infty}^{+\infty} f_X(u) \, du}_{1} = C$$

$g(x) = a \cdot x + b$, a oraz b - stałe

$$E(ax+b) = \int_{-\infty}^{+\infty} (a \cdot x + b)\, f_x(u)\, du =$$

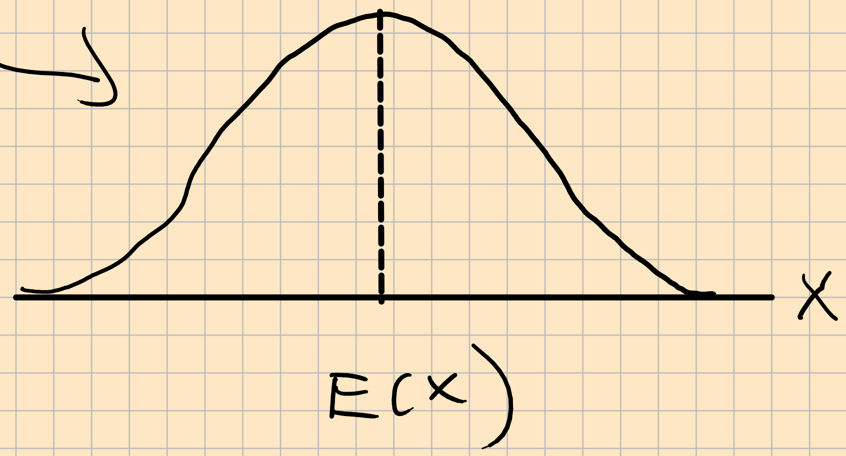
$$= \int_{-\infty}^{+\infty} a \cdot \underbrace{x \; f_x(u)}\, du + \underbrace{\int_{-\infty}^{+\infty} b \; f_x(u)\, du}_{b} =$$

$$= a \int_{-\infty}^{+\infty} x\, f_x(u)\, du + b =$$

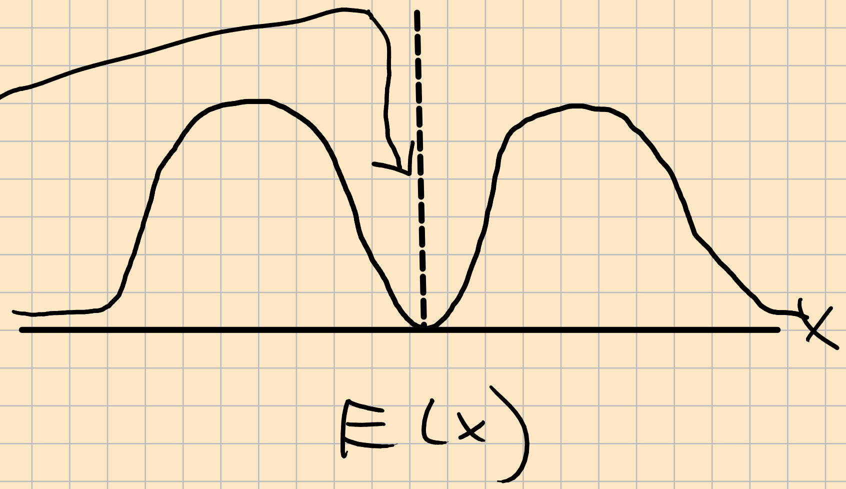
$$= a\,E(x) + b$$

UWAGA

FGP



FGP



nigdy nie znajdziemy
x w tej okolicy

- dystrybuanta

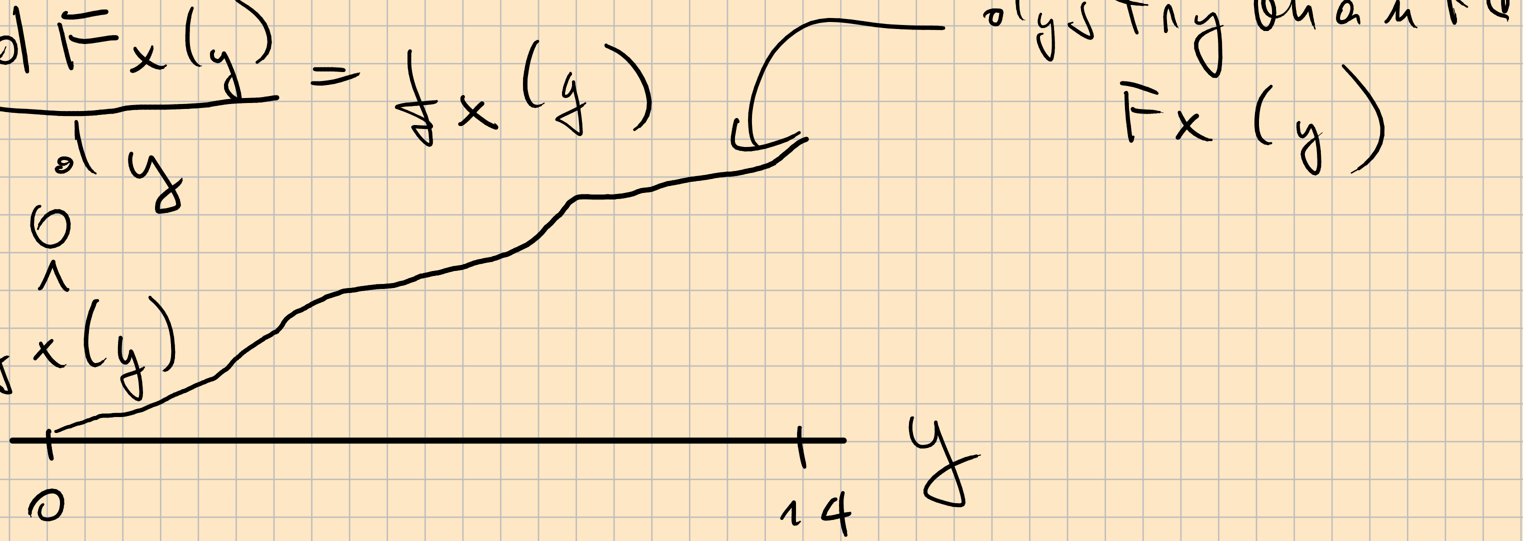
$$F_X(y) = \int_{-\infty}^{y} du \; f_X(u) =$$

$$= P(X \leq y)$$

- zawsze rośnie

$$\frac{dF_X(y)}{dy} = f_X(y)$$

$$dF_X(y) = \underset{\substack{\vee\\ 0}}{dy}\; \overset{\substack{0\\ \wedge}}{f_X(y)}$$



- wariancja

$$Var(X) = \sigma^2(X) = E\left((X - E(X))^2\right)$$

$$Var(g(X)) = \sigma^2(g(X)) = E\left(\left(g(X) - E(g(X))\right)^2\right)$$

$$g(X) = \overbrace{c}^{\text{stała}}$$

$$var(c) = \int_{-\infty}^{+\infty} (c - \underbrace{E(c)}_{c})^2 f_X(u)\, du =$$

$$= \int_{-\infty}^{+\infty} (c - c)\, f_X(u)\, du = 0$$

$$g(x) = a \cdot x + b$$, a oraz b - stałe

$$\mathrm{var}(ax+b) = \int_{-\infty}^{+\infty} du \left((a \cdot u + b) - \underbrace{E(a \cdot u + b)}_{a \cdot E(x) + b} \right)^2 f_x(u) =$$

$$= \int_{-\infty}^{+\infty} du \left(a \cdot u + \cancel{b} - a \cdot E(x) - \cancel{b} \right)^2 f_x(u)$$

$$= \int_{-\infty}^{+\infty} du \; a^2 \cdot (u - E(x))^2 f_x(u) = a^2 \int_{-\infty}^{+\infty} (u - E(x))^2 f_x(u) du =$$

$$= a^2 \cdot \mathrm{var}(x)$$

$$\mathrm{var}(x) = \int_{-\infty}^{+\infty} du\, (u - E(x))^2 f_x(u) =$$

$$= \int_{-\infty}^{+\infty} du\, \left(u^2 - 2 \cdot u \cdot E(x) + E^2(x)\right) f_x(u) =$$

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

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

- odchylenie standardowe

$$\sigma(x) = \sqrt{Var(x)}$$

$$\sigma(g(x)) = \sqrt{Var(g(x))}$$

- momenty rozkładu

$$\overline{\mu}_k = E(x^k) \qquad (\text{wzgl } 0)$$

$$\mu_k = E((x - \underset{\substack{\parallel \\ E(x)}}{\mu_1})^k)$$

- skośność $\beta_1 = \dfrac{\mu_3}{\sigma(x)^3}$

- kurtoza $\beta_2 = \dfrac{\mu_4}{\sigma(x)^2} - 3$

- kwantyl x_p

x_p jest kwantylem rzędu p

gdy $F_X(x_p) = p$

- moda