

metody statystyczne 1

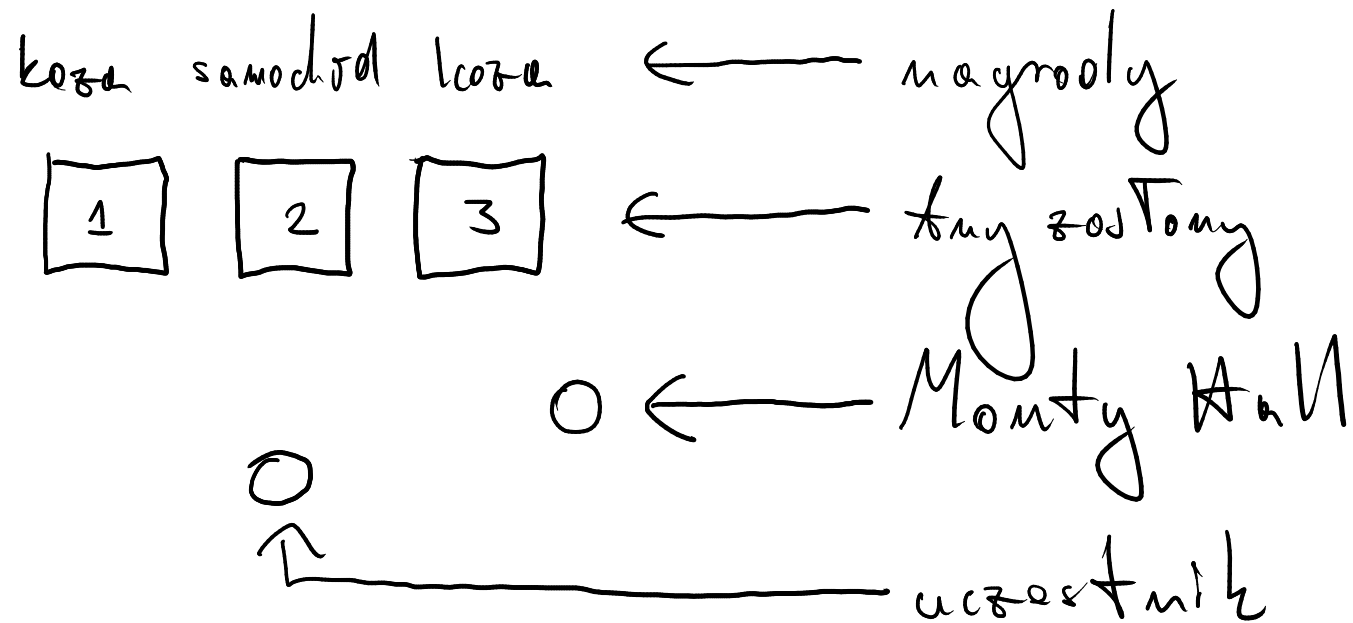
kacper.topolnicki@uj.edu.pl

Plan

- problem Montiego Halla
- aksjomaty prawdopodobieństwa
- właściwości prawdopodobieństwa
- twierdzenie Bayesa
- interpretacje prawdopodobieństwa
- dyskretne zmienne losowe
- ciągłe zmienne losowe
- funkcja gęstości prawdopodobieństwa
- właściwości rozkładów prawdopodobieństwa

Paradoks Montiego Halla

- schematycznie:



- przebieg:

- 1) uczestnik wybiera zastanę
 - 2) Monty Hall odsłania kozę
 - 3) uczestnik może zmienić zastanę
 - 4) wszystkie zastany idą wygrać
- czy uczestnik powinien zmienić wybór

monty-hall.nb

link dostepny na stronie

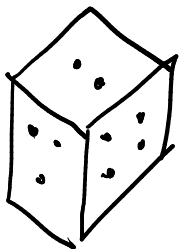
aksjomaty prawdziwości

- Колхозников 1933

$$(\Omega, F, \overset{\vee}{P})$$

\uparrow \uparrow
 pnesłmen' wszystkich "zdanien' losowych"
 zbior wszystkich "zdanien' elementarnych"

Funkcje opisujące "złotemu losowaniu" "prawdopodobieństwo" $\in [0, 1]$



$$\Omega = \{\square, \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}\}$$

zbiór partycji \rightarrow $\underbrace{\text{elementy elementarne}}_{\text{niepusty zbiór orzek}} \leftarrow \text{podzbiory } \Omega$

$$\mathcal{F} = \{\emptyset, \{\square, \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot \\ \hline \end{array}\}, \dots\}$$

zbiór losów

$$\left. \begin{aligned} P(\{\square\}) &= \frac{1}{6} \\ P(\{\begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \end{array}\}) &= \frac{1}{6} \\ P(\{\begin{array}{|c|c|} \hline \cdot & \cdot & \cdot \\ \hline \end{array}\}) &= \frac{1}{6} \\ P(\{\begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}\}) &= \frac{1}{6} \\ P(\{\begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}\}) &= \frac{1}{6} \\ P(\{\begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}\}) &= \frac{1}{6} \end{aligned} \right\}$$

aksjomaty (Ω, \mathcal{F}, P)

dwie orzeczki lub trzy orzeczki lub cztery orzeczki

$$\begin{aligned} P(\{\begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}\}) \\ P(\{\begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot \\ \hline \end{array}, \begin{array}{|c|c|} \hline \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}\}) \\ \dots \end{aligned}$$

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

np. $\{\square, \dots, \boxtimes\}$ np. $\{\emptyset, \{\square, \boxtimes, \boxtimes\}, \dots\}$ l. mierzwiłte

$$\bullet P(\Omega) = 1$$

[prawdopodobieństwo zajścia przynajmniej jednego zdarzenia elementarnego]

$$\bullet E_i \in \mathcal{F} \quad \forall_{i,j} E_i \cap E_j = \emptyset : P(E_1 \cap E_2 \cap \dots \cap E_N) = \sum_{i=1}^N P(E_i)$$

zbiory rozłączne

może to wystarczy aby policzyć:

$P(\{\square, \boxtimes, \boxtimes\})$

$P(\{\square, \boxtimes, \boxtimes\}) \quad ?$

$\uparrow = \sum_{i=1}^N P(E_i)$

liczenie prawdopodobieństwa zdarzeń rozłącznych

uŹródło: *Wojciechowski* *Przewodopodobieństwo*

- $P(\emptyset) = 0$
 \uparrow zbitr pusty

- $A \subseteq B \Rightarrow P(A) \leq P(B)$

- $P(\Omega \setminus A) = 1 - P(A)$
 \uparrow A^*

- $\forall E \in \mathcal{F}: 0 \leq P(E) \leq 1$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$\uparrow$$

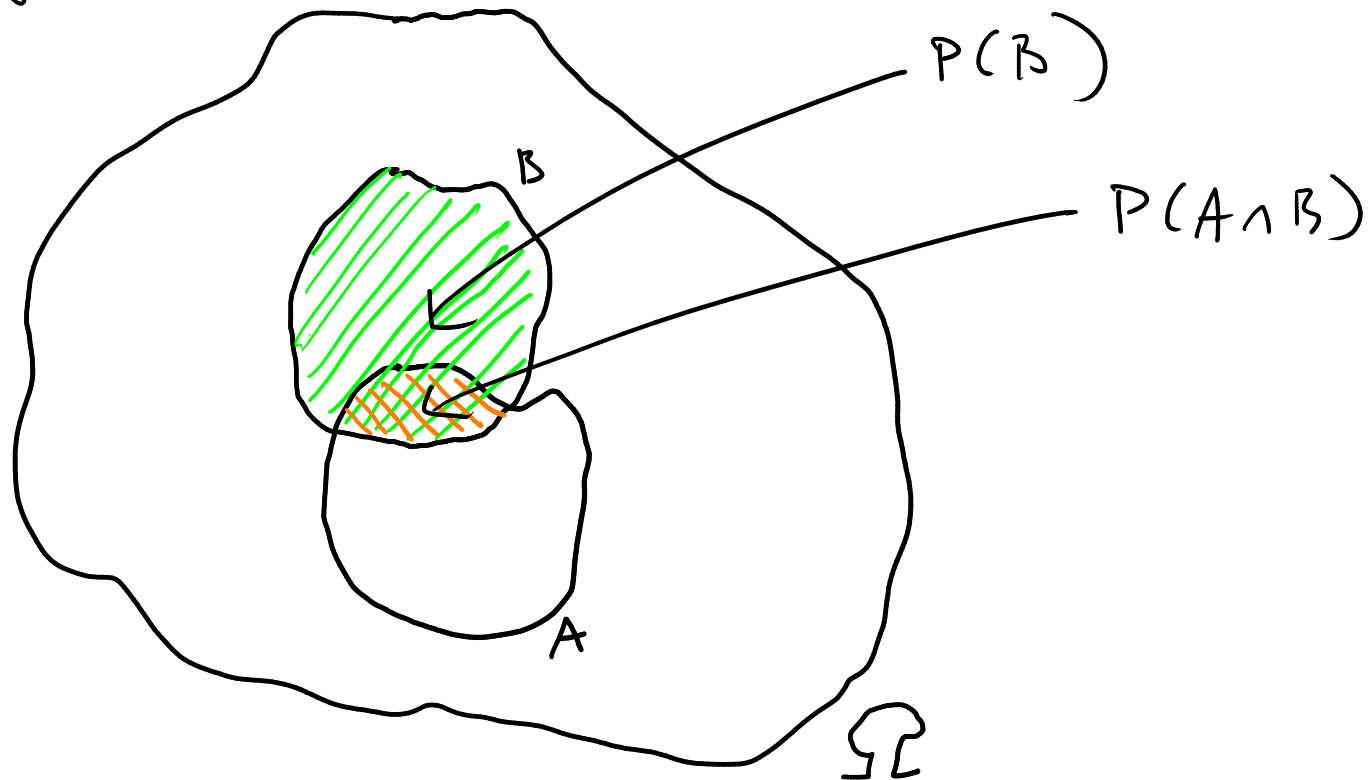
gdy rozłączne

$$P(A \cap B) = P(\emptyset) = 0$$

Award zenit Bayese

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

prawdopodobieństwo warunkowe



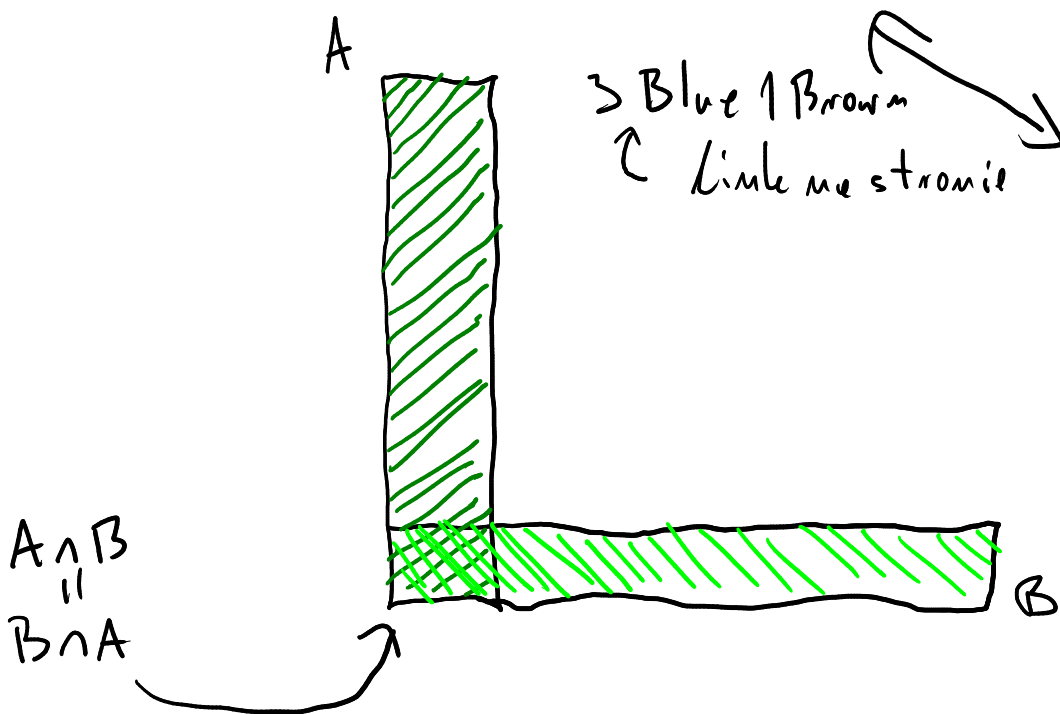
$$\left. \begin{array}{l} P(A \cap B) = P(A|B) P(B) \\ P(B \cap A) = P(B|A) P(A) \end{array} \right\} \Rightarrow P(A|B) P(B) = P(B|A) P(A)$$

\parallel
 $A \cap B$

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

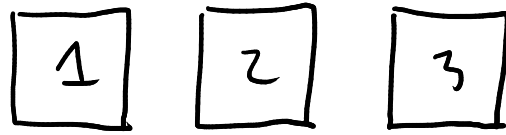
← *Lehrsatz von Bayes*

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



- Monty Hall bez zmiany:

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



odstaniecie koza

wybór

A - samochód ze 1

$\frac{1}{3}$ B - koza odstanie ze 2

tu Bayes

$\frac{1}{2}$

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

↑ samochód ze 1

pool warunkiem odstanie kozy ze 2, prawdziwe prawdopodobieństwo wygranej

samochód ze 1

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

↑ samochód ze 2 odstanie ze 2 lub 3

$+ \frac{1}{3} \cdot 0 +$

uwaga! nie możemy odstanieć 2!

$+ \frac{1}{3} \cdot 1 = \frac{3}{6} = \frac{1}{2}$

↑ samochód ze 3

nie możemy odstanieć 1, 3

- Monty Hall ze zmianą

C_i - samochód ze zastaną i

X_i - pierwszy wybór gracza to i

H_i - odstępniacze zastany i

$$P(C_2 | (H_3 \cap X_1)) = \frac{P(C_2 \cap H_3 \cap X_1)}{P(H_3 \cap X_1)} = \frac{P(H_3 | (C_2 \cap X_1)) P(C_2 \cap X_1)}{P(H_3 \cap X_1)} =$$

↑
samochód ze zastaną 2
gracz wybiera 1
odstępniacze zastany 3
gracz zmienia wybór zastany na 2

$$= \frac{P(H_3 | (C_2 \cap X_1)) P(C_2 \cap X_1)}{P(H_3 | X_1) P(X_1)} = \dots$$

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

$$P(H_3 | X_1) = \frac{1}{2}$$

$$P(H_3 | (C_2 \cap X_1)) = \frac{1}{2}$$

$$P(H_3 | (C_3 \cap X_1)) = 0$$

$$P(C_i \cap X_j) = P(C_i) P(X_j) \quad \text{niezależnie}$$

$$= \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

przełożenie gier
z wyliczeniami

1, 2, 3 → 2, 1, 3

1, 2, 3 → 1, 3, 2

$$\dots = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3}} = \frac{6}{9} = \frac{2}{3}$$

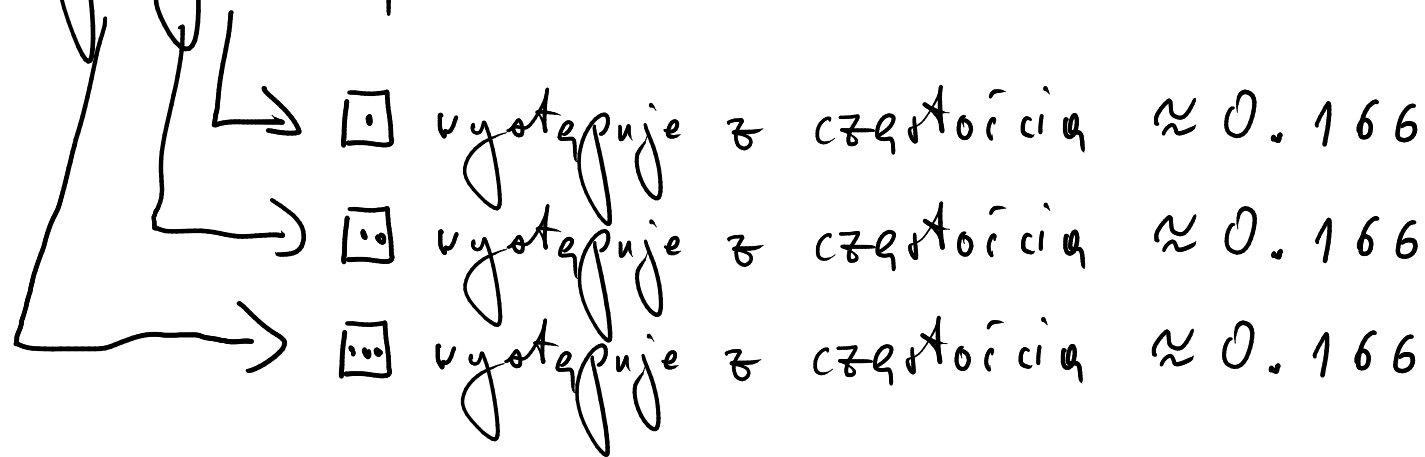
$$P(X_i) = \frac{1}{3}$$

interpretacje prawdopodobieństwa

- prawdopodobieństwo obiektywne (frequentist probability)



powtarzający się proces



...

częstość kojarzymy z prawdopodobieństwem

- prawdopodobieństwo subiektywne (Bayesian probability)

$$p(h|e) = \frac{p(e|h) p(h)}{p(e)}$$

$p(h|e)$
 \uparrow dowód
 \uparrow hipoteza

posterior probability

prior probability

prawdopodobieństwo odzwierciedla posiadaną wiedzę

na początku:
 $p(h)$

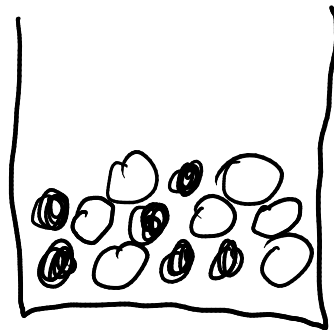
prawo Bayesa

na końcu:
 $p(h|e)$

pytanie:

$$P(h = (m, N-m) \mid e = m_i) =$$

\swarrow \searrow
 $h. \text{ czarnych}$ $h. \text{ białych}$



N -kul

m -czarnych

$N-m$ białych

możemy wyliczyć

$$= \frac{P(e = m_i \mid h = (m, N-m)) \cdot P(h = (m, N-m))}{P(e = m_i)}$$

możemy wyliczyć

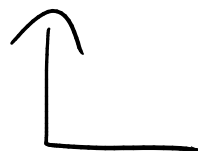
= warunek normalizacji

odzwierciedla wiedzę przed wylosowaniem

$$\sum_h P(h|e) = 1$$

- losujemy kulę $m_1 = c, b$
- na podstawie wylosowanych (m_1, m_2, \dots) stawiamy hipotezy dotyczące m
- znamy N

pages.m6



link dostępny na stronie

- o lotzego me koniec nie mamy ostrego piku
dla wtorekij liczby?

- informacja o bieżącej liczbie kul w pudle

można spróbować dodać

Emienne losowe

- zmienna, której wartości zależą od
wyniku procesu losowego

- formalnie:

funkcja $X: \Omega \rightarrow E$

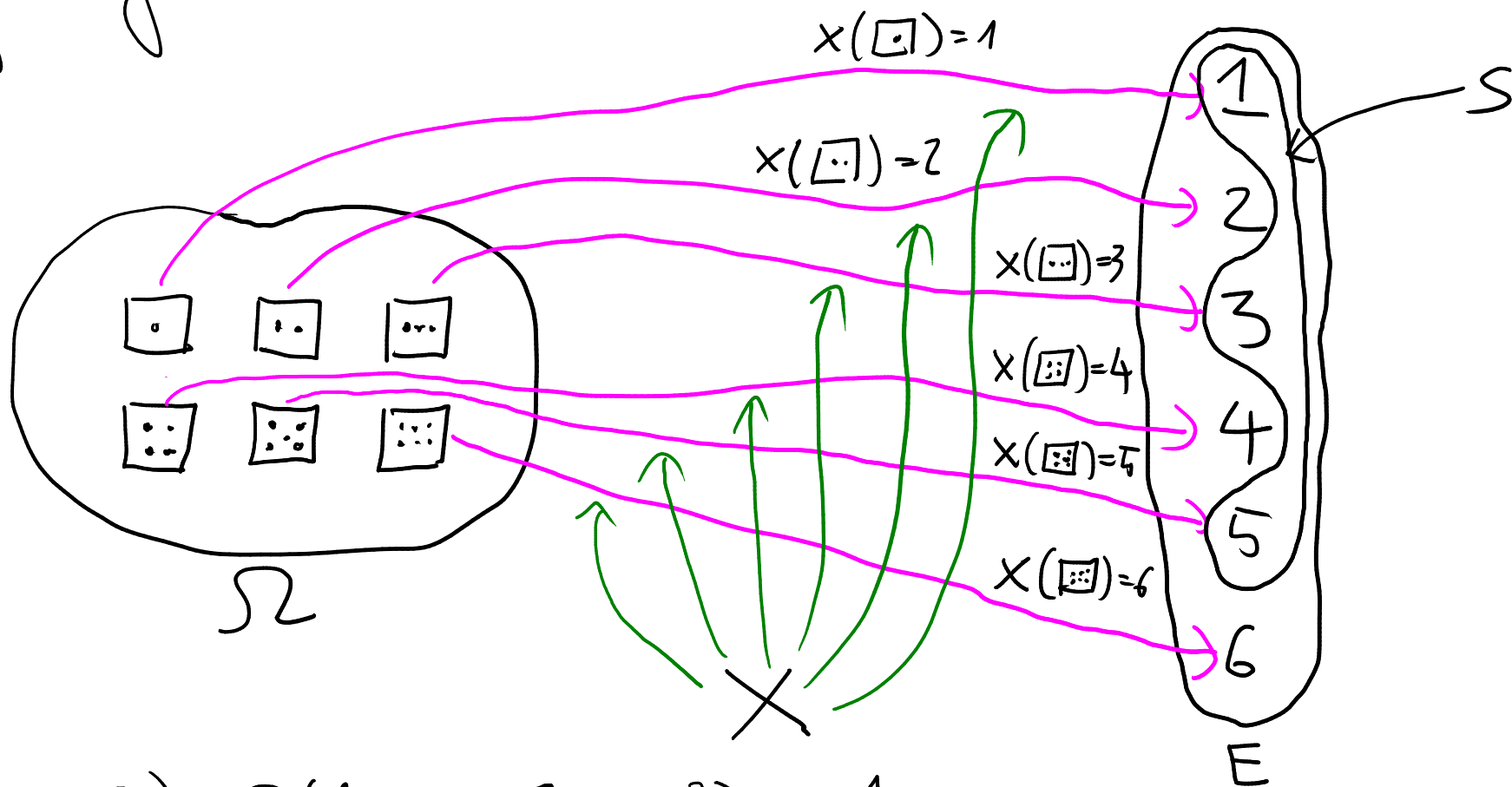
\uparrow przestrzeń miernotw. $E = \mathbb{R}$

- prawdopodobieństwo zmiennych losowych:

$$P(y \in S) = P(\{\omega \in \Omega \mid X(\omega) \in S\})$$

\uparrow
 $S \subset E$

w praktyce:

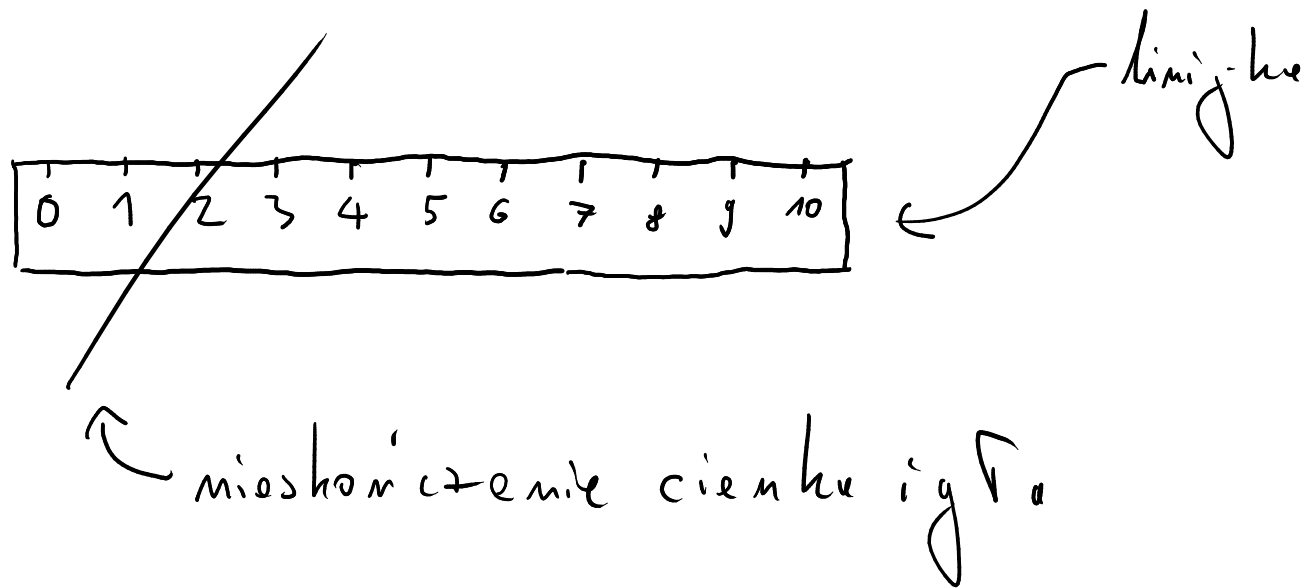


$$P(y \in \{1, 2, 3\}) = P(\{ \text{1 dot}, \text{2 dots}, \text{3 dots} \}) = \frac{1}{2}$$

\uparrow
 S

$X(\text{1 dot}) = 1 \in S$ $X(\text{2 dots}) = 2 \in S$ $X(\text{3 dots}) = 3 \in S$

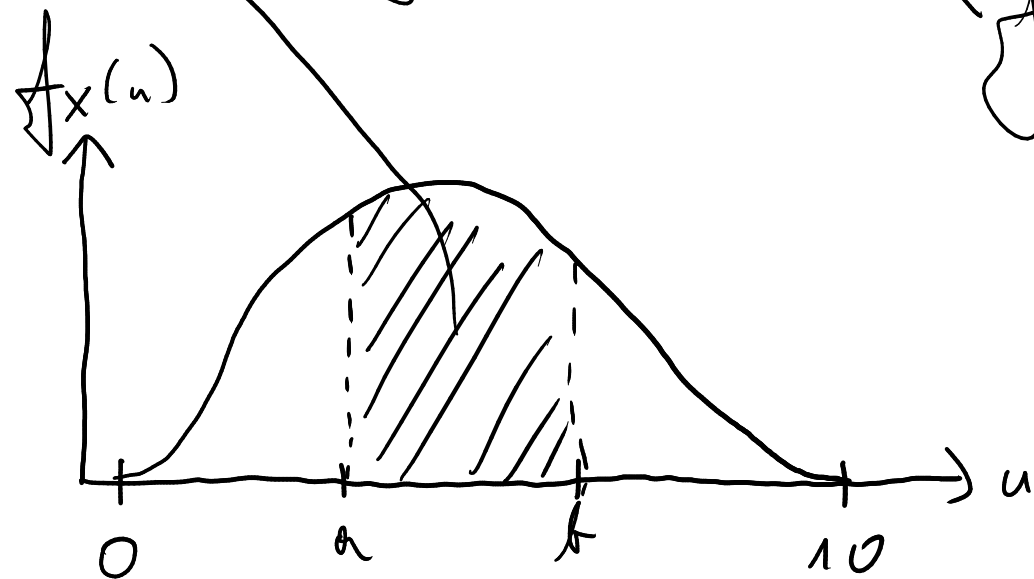
ciężkie ziemne łozowe



$$P(y \in \{3\}) = ? = 0$$

$$P(a \leq y \leq b) = \int_a^b f_X(u) du$$

$y \in S = [a, b]$



funkcja gęstości prawdopodobieństwa
dla zmiennej losowej X

$$P(0 \leq y \leq 10) = 1$$

$$\underbrace{\int_0^{10} f_X(u) du}_\text{normalizacja}$$

właściwości funkcji gęstości: probabilistyczne:

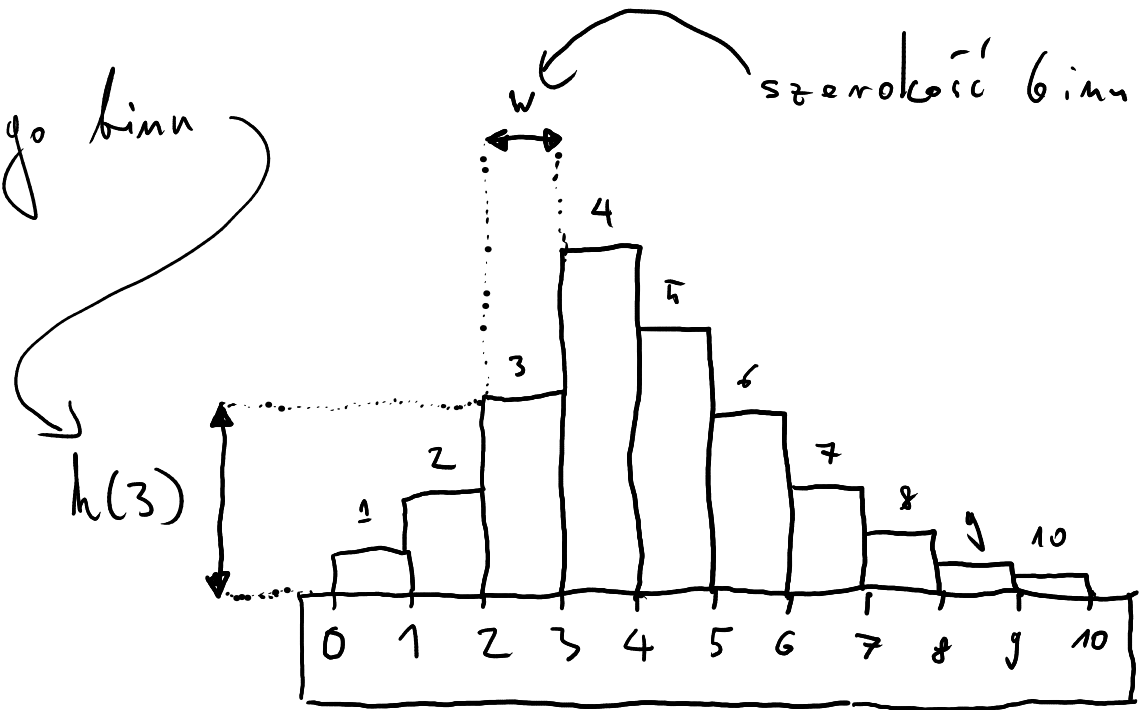
$$\bullet \int_{u_{\min}}^{u_{\max}} f(u) du = 1$$

$$\bullet P(z \leq y \leq z + dz) = f(z) dz$$

$$\bullet \forall_z P(z \leq y \leq z + dz) > 0 \Rightarrow \forall_z f(z) > 0$$

histogramy

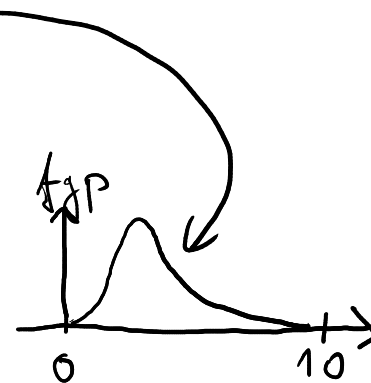
wysokość danego binu



jak dobrać wysokość binów?

- prawdopodobieństwo, że y wpadnie w bin o numerze i :

$$P((i-1)w \leq y \leq i \cdot w) = \int_{(i-1)w}^{i \cdot w} f(u) du$$



- ile będzie zliczeń w binie i :

$$N \cdot P((i-1)w \leq y \leq i \cdot w)$$

cofkałowa liczba próbek

- przybliżenie funkcji gęstości prawdopodobieństwa

$$P((i-1)w \leq y \leq i \cdot w) \approx w \cdot f\left(\frac{1}{2}((i-1)w + i \cdot w)\right) \quad f(u) \approx \frac{P(u)}{w}$$

wysokość
stopy

często można
wybrać
w białkach
do tworzenia
wykresu

opis funkcji gęstości prawdopodobieństwa

wartości oczekiwane

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

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

zmienna losowa $\mathbb{R} \rightarrow \mathbb{R}$

$$g(x) = c \quad \text{state}$$

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

$$g(x) = ax + 6$$

$$E(ax + 6) = \int_{-\infty}^{+\infty} (au + 6) f_X(u) du =$$

$$= \underbrace{\int_{-\infty}^{+\infty} au f_X(u) du}_{=aE(X)} + \underbrace{\int_{-\infty}^{+\infty} 6 f_X(u) du}_6 =$$

$$= a \int_{-\infty}^{+\infty} u f_X(u) du + 6 =$$

$$= aE(X) + 6$$

Agp. mb

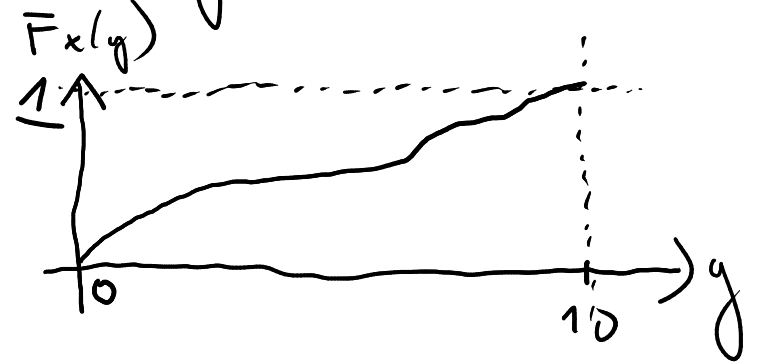
link me stronie

dystribucja

$$F_x(y) = \int_{-\infty}^y du f_x(u) = P(X \leq y)$$

zawsze rośnie

$$\frac{dF_x(y)}{dy} = f_x(y)$$



varianze

$$\begin{aligned}\text{var}(X) &= \int_{-\infty}^{+\infty} du (u - \mathbb{E}(X))^2 \downarrow x(u) - \\ &= \int_{-\infty}^{+\infty} du (u^2 - 2u\mathbb{E}(X) + \mathbb{E}^2(X)) \downarrow x(u) = \\ &= \mathbb{E}(X^2) - 2\mathbb{E}^2(X) + \mathbb{E}^2(X) = \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\end{aligned}$$

odchylenie standardowe

$$\sigma(X) = \sqrt{\text{var}(X)}$$

$$\sigma(g(X)) = \sqrt{\text{var}(g(X))}$$

momenty rozkładu

$$\mu_k = E(X^k) \leftarrow \text{względem } 0$$

$$\mu_k = E\left((X - \underbrace{\mu_1}_{E(X)})^k\right)$$

skośności

$$\beta_1 = \frac{\mu_3}{\sigma(X)^3}$$

kurtosis

$$\beta_2 = \frac{\mu_4}{\sigma(X)^4} - 3$$

kwantyl

mod

