



SPOTKANIA Z MATEMATYKĄ STOSOWANĄ

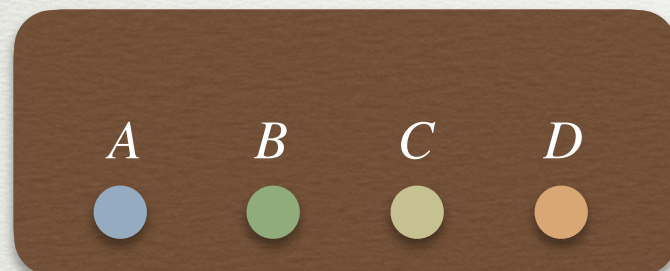
Instytut Matematyki, Politechnika Łódzka

WARTOŚĆ OCZEKIWANA W PODEJMOWANIU DECYZJI

Grzegorz Skalski

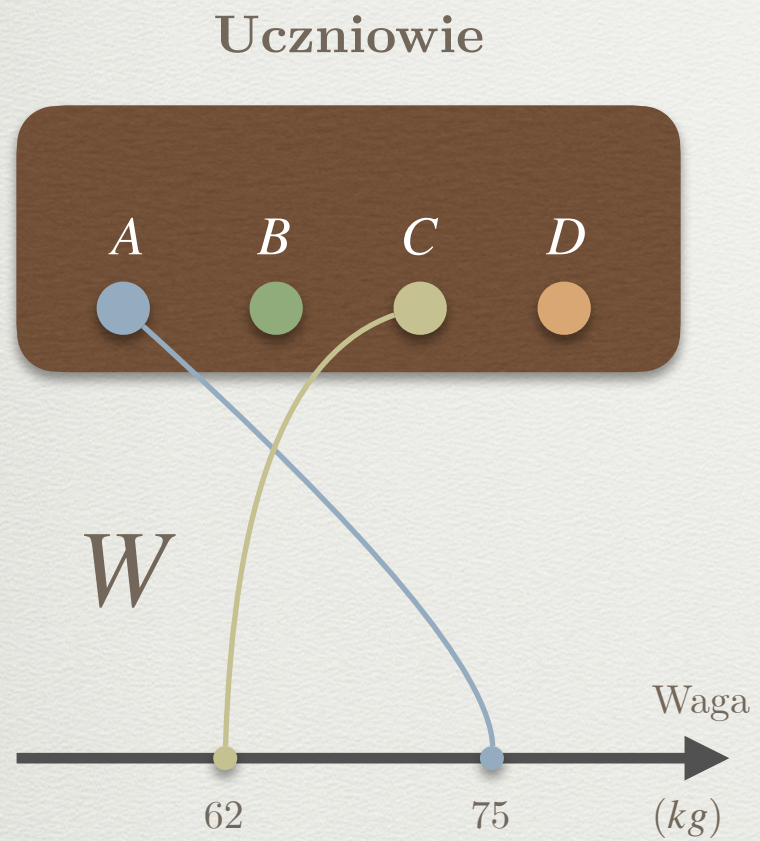
Zmienna losowa — idea

Uczniowie

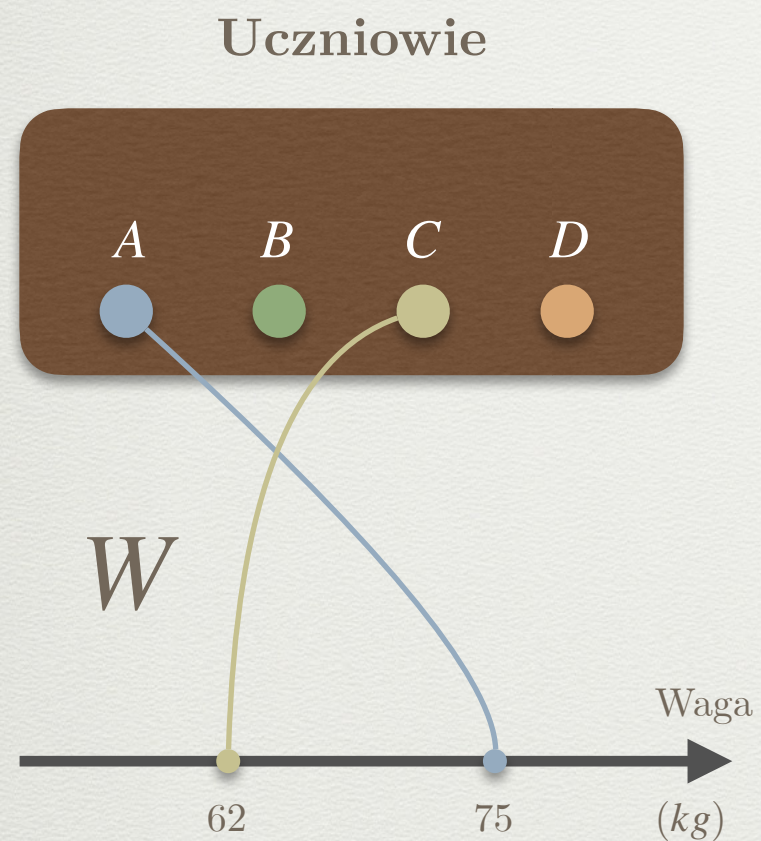


Eksperyment polega na losowym wyborze jednego ucznia i zarejestrowaniu jego różnych cech: wagi, wzrostu, liczby spóźnień w ciągu zeszłego tygodnia, itp. Zbiór zdarzeń elementarnych, czyli możliwych wyników losowania oznaczamy jako Ω . W naszym przypadku utożsamiamy go ze zbiorem wszystkich uczniów.

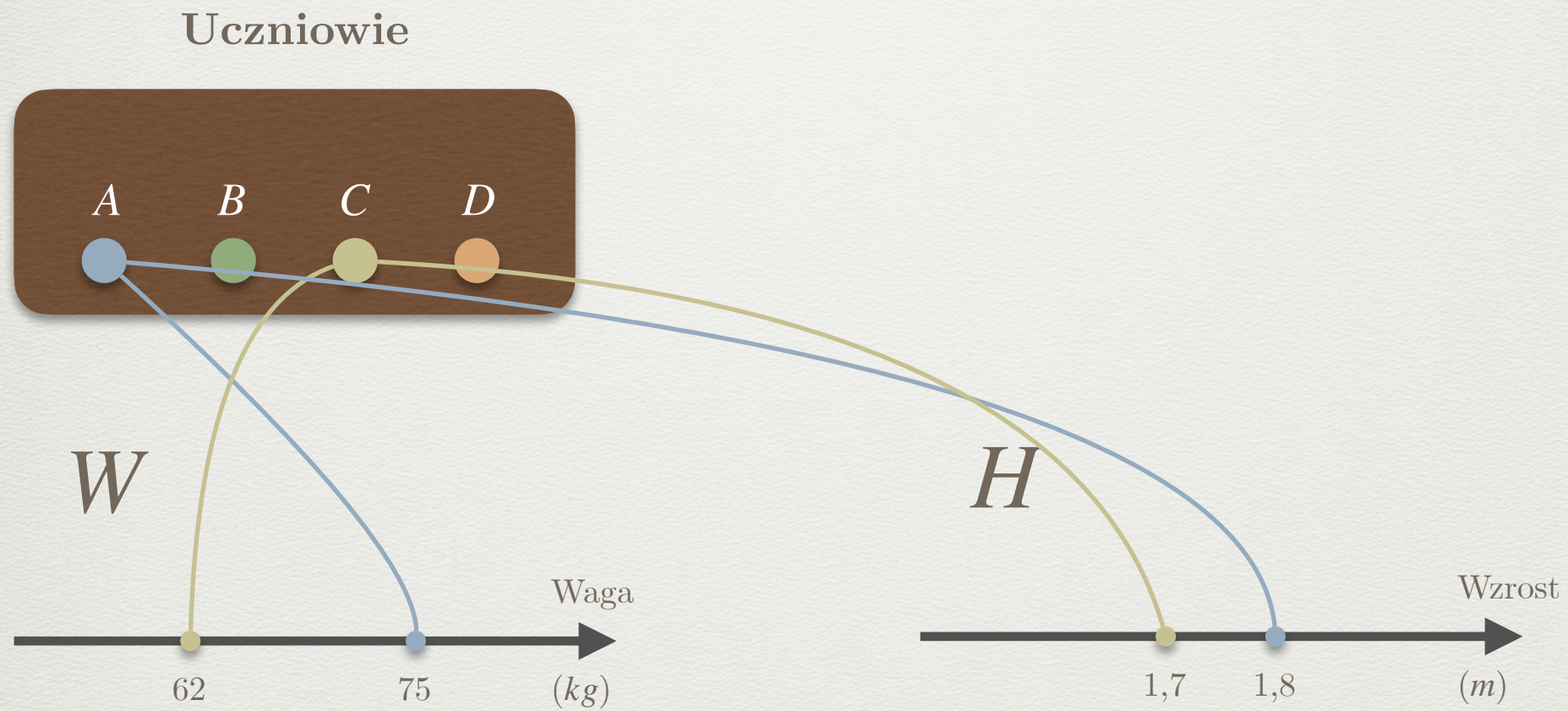
Zmienna losowa — idea



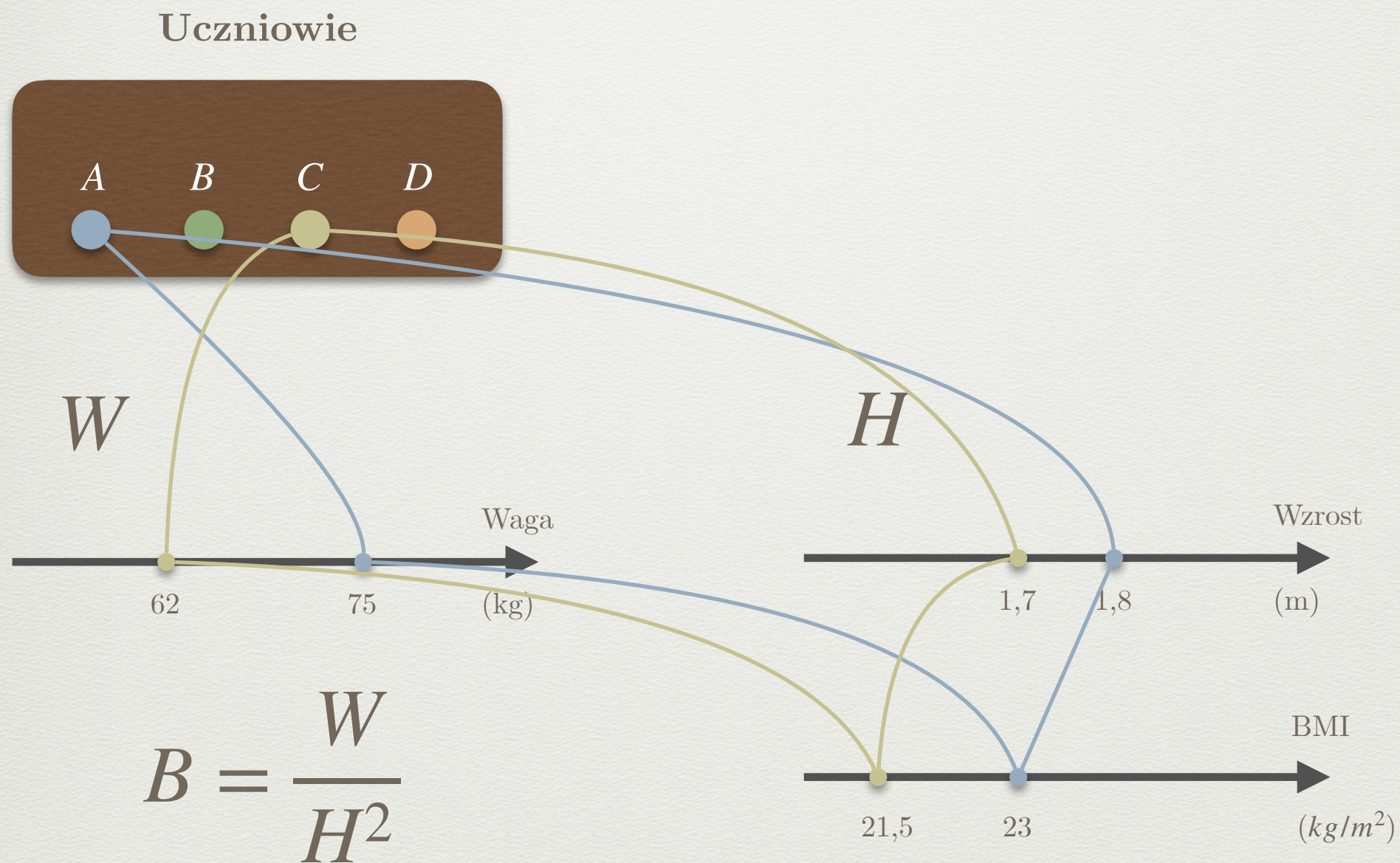
Zmienna losowa — idea



Zmienna losowa — idea



Zmienna losowa — idea



Zmienna losowa — formalizm

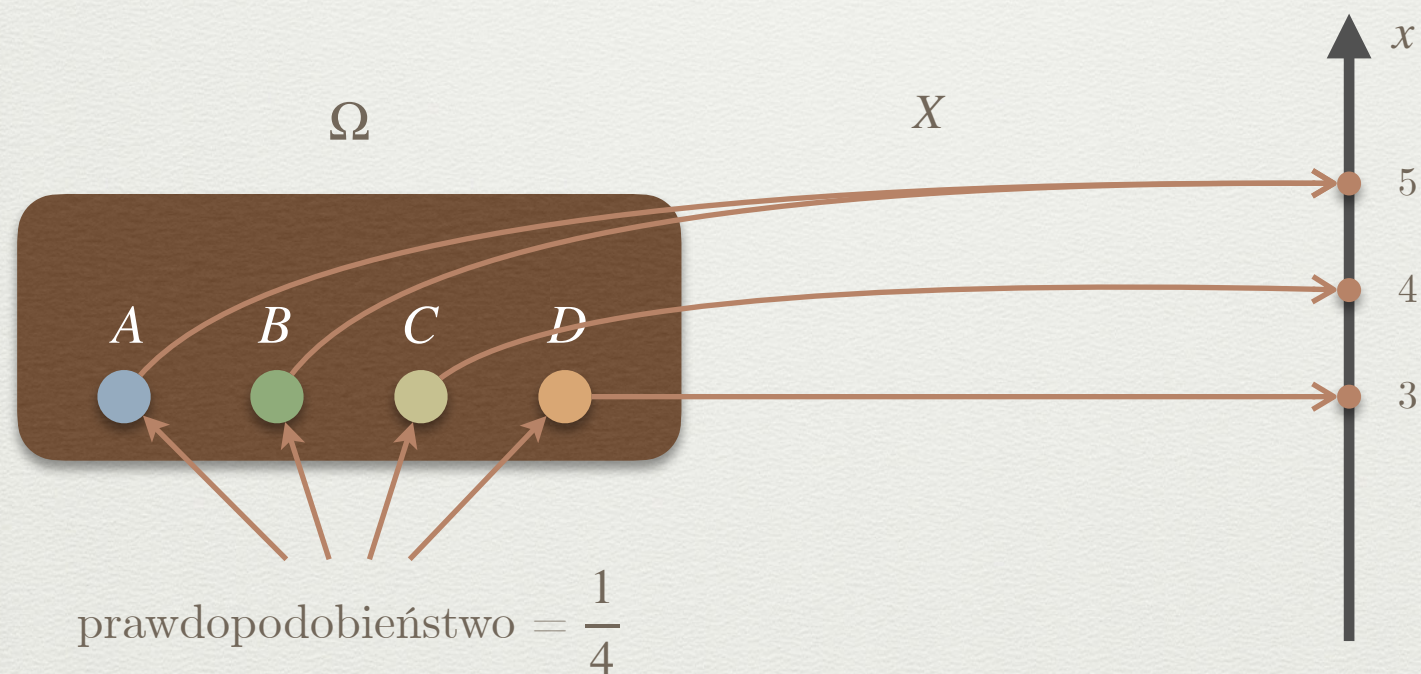
- Zmienna losowa przypisuje wartość (liczbę) każdemu możliwemu wynikowi eksperymentu.
- Matematycznie, zmienna losowa to funkcja działająca z przestrzeni zdarzeń elementarnych Ω w zbiór liczb rzeczywistych $\mathbb{R}$.
- Może ona przyjmować tylko wartości dyskretne (np. liczby całkowite 0, 1, 2, 3) wtedy mówimy i zmiennej *skokowej* albo *dyskretnej* lub wszystkie wartości z pewnego zakresu (np. przedziału liczbowego (0,2]) i wówczas mówimy o zmiennej *ciągłej*.
- Zmienną losową oznaczamy zwykle dużą literą X , a jej konkretną wartość małą literą x . Zatem piszemy $X: \Omega \rightarrow \mathbb{R}$, a zapis $X = x$ odczytujemy: Zaobserwowano wartość x eksperymentu losowego związanego z badaniem cechy X .
- Na tej samej przestrzeni zdarzeń elementarnych Ω możemy określić wiele różnych zmiennych losowych.
- Funkcja wiążąca ze sobą takie zmienne losowe jest również zmienną losową, np. $X + Y$, $X \cdot Y$, X^2 .

Rozkład prawdopodobieństwa

- Przyjęcie konkretnej wartości przez zmienną losową związane jest z wynikami eksperymentu losowego,.
- Zatem prawdopodobieństwo przyjęcia tej wartości również zależy od wyników tego eksperymentu, np. od tego, który uczeń zostanie wylosowany.
- Można powiedzieć, że na zbiorze wartości danej zmiennej losowej rozłożone zostało prawdopodobieństwo.

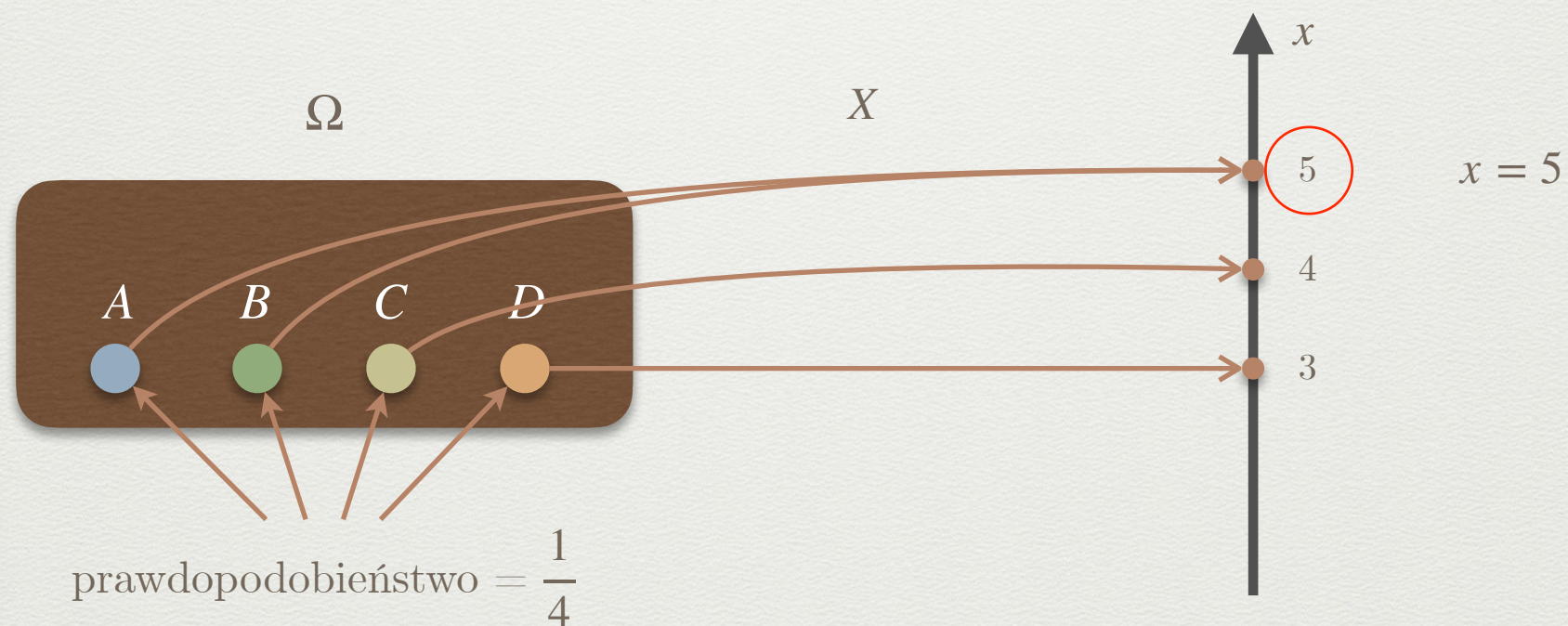
Funkcja rozkładu prawdopodobieństwa

- *Funkcja rozkładu prawdopodobieństwa* zmiennej losowej $X: \Omega \rightarrow \mathbb{R}$ opisuje prawdopodobieństwa z jakimi zmienna losowa przyjmuje swoje konkretne wartości.



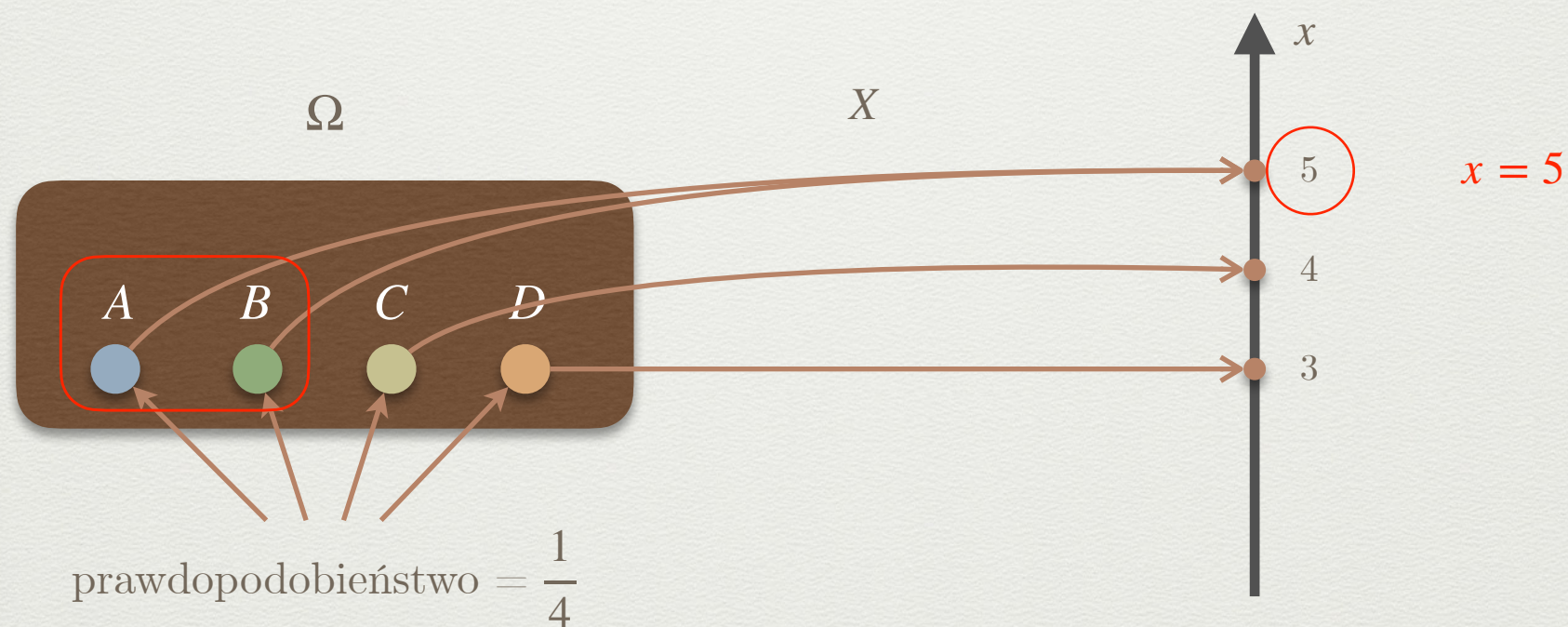
Funkcja rozkładu prawdopodobieństwa

- *Funkcja rozkładu prawdopodobieństwa* zmiennej losowej $X: \Omega \rightarrow \mathbb{R}$ opisuje prawdopodobieństwa z jakimi zmienna losowa przyjmuje swoje konkretne wartości.



Funkcja rozkładu prawdopodobieństwa

- *Funkcja rozkładu prawdopodobieństwa* zmiennej losowej $X: \Omega \rightarrow \mathbb{R}$ opisuje prawdopodobieństwa z jakimi zmienna losowa przyjmuje swoje konkretne wartości.



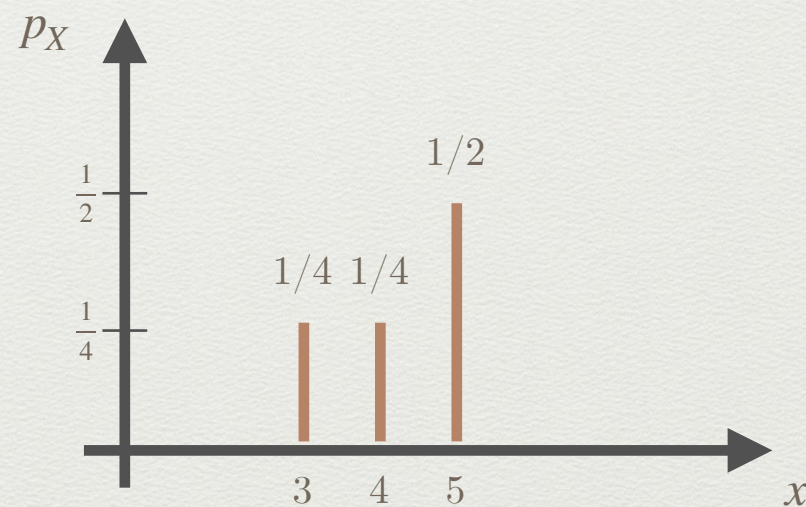
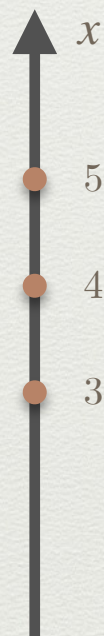
$$\{\omega \in \Omega: X(\omega) = 5\} = \{A, B\}$$

$$p_X(5) = \frac{2}{4} = \frac{1}{2}$$

Funkcja rozkładu prawdopodobieństwa

- *Funkcja rozkładu prawdopodobieństwa* zmiennej losowej $X: \Omega \rightarrow \mathbb{R}$ opisuje prawdopodobieństwa z jakimi zmienna losowa przyjmuje swoje konkretne wartości.

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega: X(\omega) = x\})$$



Funkcja rozkładu prawdopodobieństwa

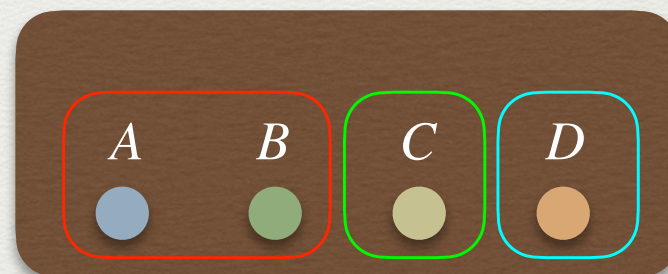
- *Funkcja rozkładu prawdopodobieństwa* zmiennej losowej $X: \Omega \rightarrow \mathbb{R}$ opisuje prawdopodobieństwa z jakimi zmienna losowa przyjmuje swoje konkretne wartości.

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega: X(\omega) = x\})$$

Własności

- $p_X(x) \geq 0$,
- $\sum_x p_X(x) = 1$.

Ω



Przykłady

Dla ustalonej liczby $c \in \mathbb{R}$ możemy mówić o zmiennej losowej stałej $X(\omega) = c$ dla $\omega \in \Omega$. Taka zmienna losowa nosi często nazwę *zdegenerowanej* lub *pewnej*, gdyż przyjmie ona swoją jedyną wartość c z prawdopodobieństwem 1. Oznaczamy ją symbolem c . Mówimy, że taka zmienna losowa ma *jednopunktowy rozkład prawdopodobieństwa*

x_i	c
$p_X(x_i)$	1

X = Liczba pasażerów samochodów, którzy byli na Marsie

Przykłady

Mówimy, że zmienna losowa X ma rozkład *zero-jedynkowy*, gdy przyjmuje tylko dwie wartości 0 i 1, przy czym wartość 1 przyjmuje z prawdopodobieństwem p , a wartość 0 z prawdopodobieństwem $q = 1 - p$.

x_i	0	1
$p_X(x_i)$	$1 - p$	p

$X =$ Liczba pomyłek sapera

Przykłady

Mówimy, że zmienna losowa X ma rozkład *Bernoulliego* (*dwumianowy*) z parametrami n i $p \in [0,1]$, gdy przyjmuje tylko $n + 1$ wartości: $0, 1, 2, \dots, n$, przy czym funkcja rozkładu prawdopodobieństwa wyraża się wzorem:

$$p_X(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x \in \{0, 1, \dots, n\}.$$

x_i	0	1	...	n
$p_X(x_i)$	$(1 - p)^n$	$np(1 - p)^{n-1}$	...	p^n

X = Liczba nasion, które wykiełkowały z n -elementowego worka

Przykłady

Mówimy, że zmienna losowa X ma rozkład *Poissona* z parametrem $\lambda > 0$, gdy przyjmuje nieskończenie wiele wartości: $0, 1, 2, \dots$, przy czym funkcja rozkładu prawdopodobieństwa wyraża się wzorem:

$$p_X(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \in \{0, 1, \dots\}.$$

x_i	0	1	2	...
$p_X(x_i)$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	...

X = Liczba samochodów, które przejechały przez skrzyżowanie w niedzielę

Przykłady

Mówimy, że zmienna losowa X ma rozkład *geometryczny* z parametrem $p \in (0,1)$, gdy przyjmuje nieskończenie wiele wartości: $1,2,\dots$, przy czym funkcja rozkładu prawdopodobieństwa wyraża się wzorem:

$$p_X(x) = P(X = x) = (1 - p)^{x-1}p, \quad x \in \{1,2,\dots\}.$$

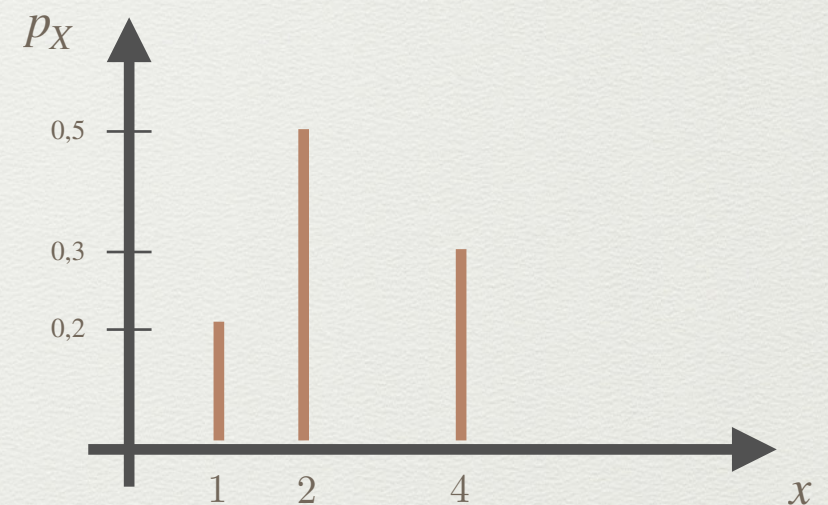
x_i	1	2	3	...
$p_X(x_i)$	p	$(1 - p)p$	$(1 - p)^2p$	...

X = Numer rzutu monetą, w którym pierwszy raz wypadła reszka

Przykłady

Przypuśćmy, że gramy w pewną grę, której wartości $x_i : 1,2,4$ opisuje zmienna losowa X . Jej rozkład prawdopodobieństwa możemy na przykład opisać w poniższej tabeli

x_i	1	2	4
$p_X(x_i)$	0,2	0,5	0,3



Średnia

W każdym z powyższych przykładów możemy zadać podobne pytanie:

- Średnio, ilu pasażerów było na Marsie?
- Średnio, ilu saperów się myli?
- Średnio, ile wzejdzie nasion z jednego worka?
- Średnio za którym razem wypadnie reszka?
- Średnio, ile wygramy w opisaną grę?

Wartość oczekiwana

- Załóżmy, że gramy w opisaną grę losową, powiedzmy, 1000 razy.

x_i	1	2	4
$p_X(x_i)$	0,2	0,5	0,3



- Interpretując prawdopodobieństwa jako częstości, 200 razy powinniśmy otrzymać nagrodę 1, 500 razy — nagrodę 2 i 300 razy — nagrodę 4. Zatem średni zysk z gry powinien wynieść

$$\frac{1 \cdot 200 + 2 \cdot 500 + 4 \cdot 300}{1000} = 1 \cdot 0,2 + 2 \cdot 0,5 + 4 \cdot 0,3 = 2,4.$$

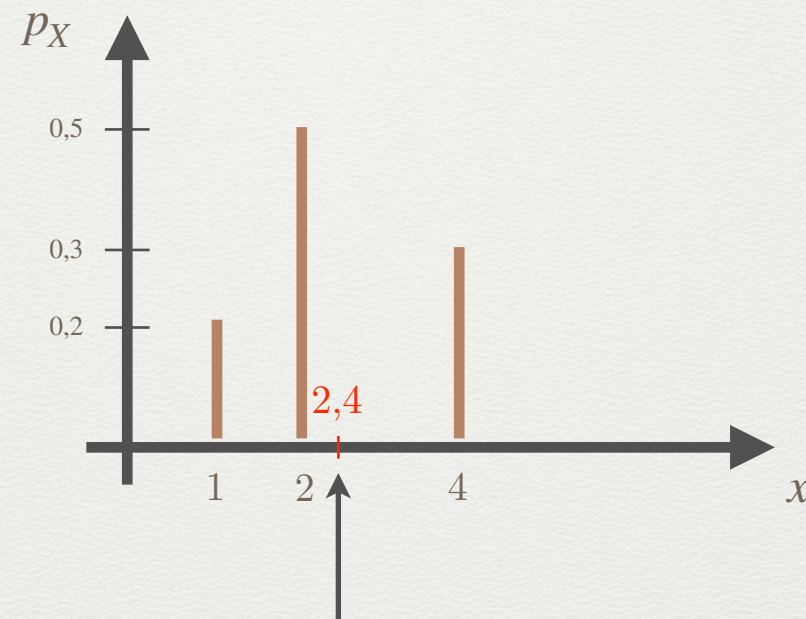
Wartość oczekiwana

- *Wartością oczekiwaną* zmiennej losowej X o rozkładzie prawdopodobieństwa p_X nazywamy liczbę

$$E[X] = \sum_x x p_X(x) = x_1 p_X(x_1) + x_2 p_X(x_2) + \dots$$

- Interpretacja: średnia wartość liczbowa, której oczekujemy w dużej liczbie niezależnych powtórzeń eksperymentu.
- **Uwaga:** Jeśli w powyższej definicji występuje suma nieskończona, zakładamy dodatkowo, że $\sum_x |x| p_X(x) < +\infty$.

Wartość oczekiwana



Wygodnie jest traktować wartość oczekiwaną $E[X]$ jako środek ciężkości rozkładu prawdopodobieństwa.

Podjmowanie decyzji za pomocą wartości oczekiwanej

Problem z Quizem

Przykład ten, po odpowiednim uogólnieniu, stanowi pierwowzór problemu optymalnej kolejności rozwiązywania ciągu zadań o niepewnych nagrodach.

Rozważmy grę (quiz), w której pewna osoba odpowiada na dwa pytania:

Pytanie A

Poprawna odpowiedź z
prawdopodobieństwem 0,8
i nagrodą 100 zł.

Pytanie B

Poprawna odpowiedź z
prawdopodobieństwem 0,5
i nagrodą 200 zł.

Jeśli na pierwsze pytanie odpowie poprawnie, otrzyma odpowiednią nagrodę i może odpowiedzieć na drugie. Jeśli odpowie błędnie, nie otrzymuje nagrody i nie może już odpowiedzieć na drugie pytanie. Poprawna odpowiedź na drugie pytanie również skutkuje nagrodą, błędna — jej brakiem. Na które pytanie osoba powinna odpowiedzieć najpierw?

Problem z Quizem

Okazuje się, że rozsądnej decyzji dostarcza wartość oczekiwana. Pytanie brzmi: na które pytanie (A czy B) odpowiedzieć wpierw, by zmaksymalizować wartość oczekiwaną całkowitej nagrody otrzymanej w grze?

Odpowiedź nie jest oczywista, gdyż musimy utrzymać balans: przystępując do odpowiedzi na trudniejsze, ale też wartościowsze Pytanie B ponosimy ryzyko nieprzystąpienia do odpowiedzi na prostsze Pytanie A.

Pytanie A

Poprawna odpowiedź z
prawdopodobieństwem 0,8
i nagrodą 100 zł.

Pytanie B

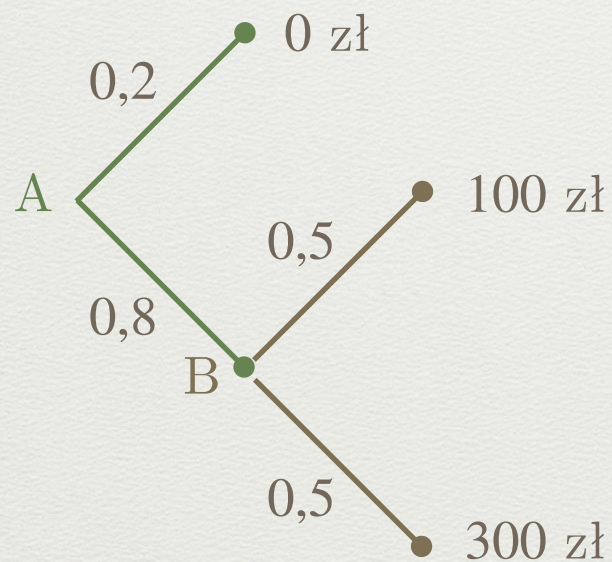
Poprawna odpowiedź z
prawdopodobieństwem 0,5
i nagrodą 200 zł.

Problem z Quizem

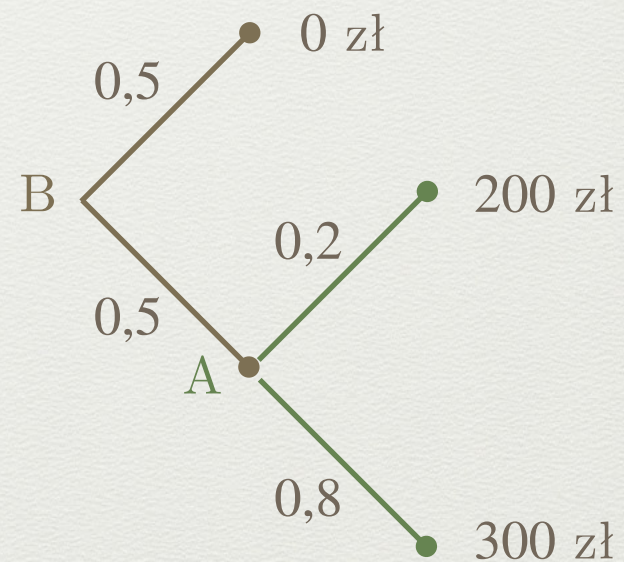
Rozważmy zmienną losową

$X =$ Całkowita nagroda w grze

i wyznaczmy jej wartość oczekiwaną w dwóch możliwych scenariuszach:



Wybrano pierwsze pytanie A



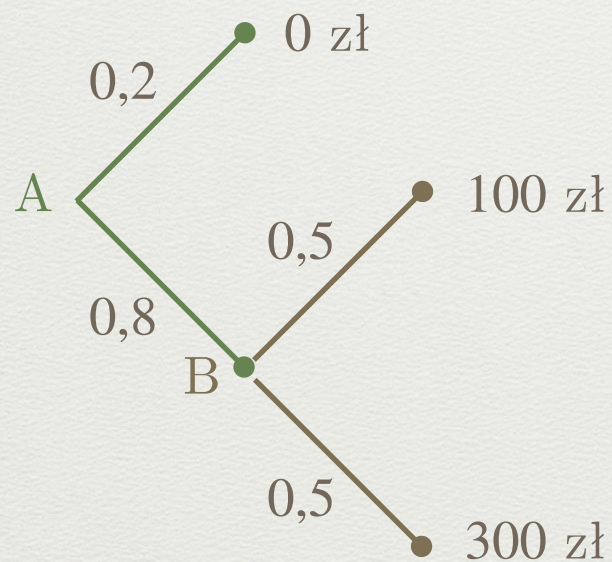
Wybrano pierwsze pytanie B

Problem z Quizem

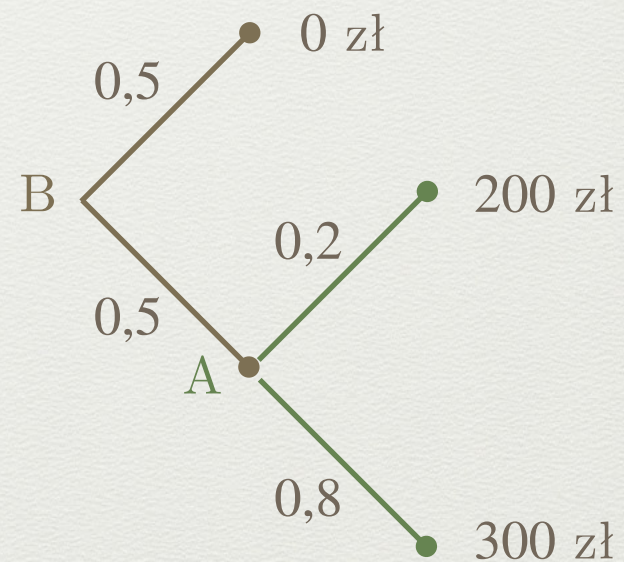
Rozważmy zmienną losową

$X =$ Całkowita nagroda w grze

i wyznaczmy jej wartość oczekiwaną w dwóch możliwych scenariuszach:



Wybrano pierwsze pytanie A



Wybrano pierwsze pytanie B

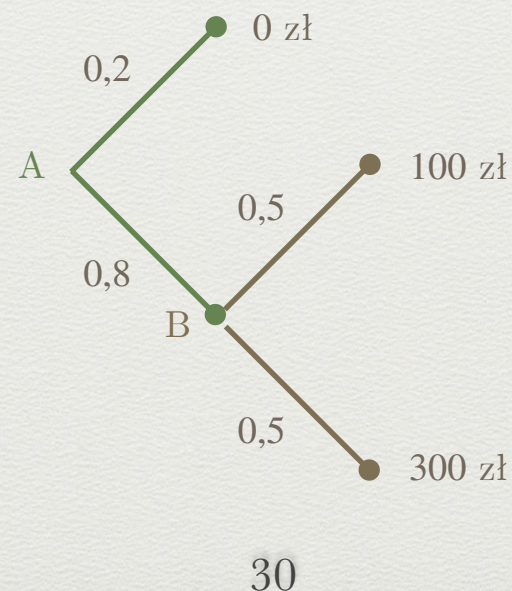
Scenariusz 1

Jeśli jako pierwsze wybierzemy Pytanie A, to funkcja rozkładu prawdopodobieństwa zmiennej losowej X będzie miała postać

$$p_X(0) = 0,2, \quad p_X(100) = 0,8 \cdot 0,5, \quad p_X(300) = 0,8 \cdot 0,5.$$

A zatem wartość oczekiwana zmiennej losowej X będzie równa

$$E[X] = 0,8 \cdot 0,5 \cdot 100 + 0,8 \cdot 0,5 \cdot 300 = 160 \text{ zł}.$$



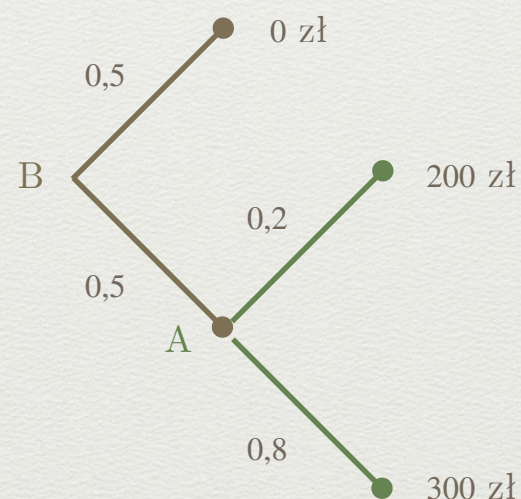
Scenariusz 2

Jeśli jako pierwsze wybierzemy Pytanie B, to funkcja rozkładu prawdopodobieństwa zmiennej losowej X będzie miała postać

$$p_X(0) = 0,5, \quad p_X(200) = 0,5 \cdot 0,2, \quad p_X(300) = 0,5 \cdot 0,8.$$

A zatem wartość oczekiwana zmiennej losowej X będzie równa

$$E[X] = 0,5 \cdot 0,2 \cdot 200 + 0,5 \cdot 0,8 \cdot 300 = 140 \text{ zł}.$$



Decyzja

Ponieważ w pierwszym scenariuszu oczekiwana wartość całkowitej nagrody jest wyższa (160 zł), więc ten scenariusz jest preferowany. Dana osoba powinna wybrać ten właśnie scenariusz.

Scenariusz 1

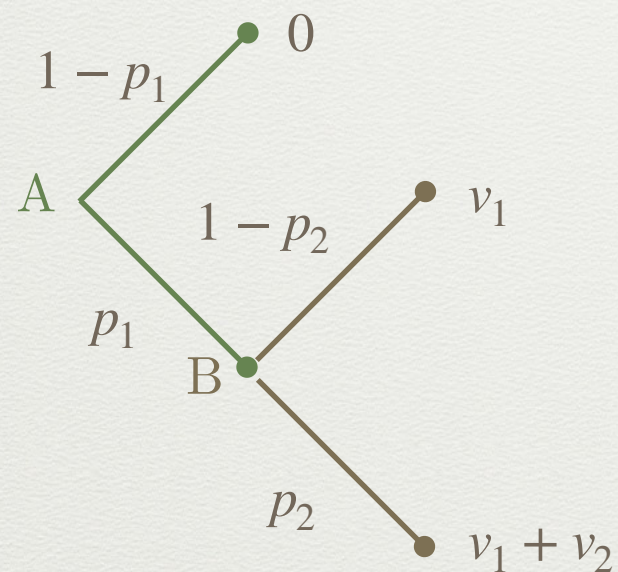
$$E[X] = 160$$

Scenariusz 2

$$E[X] = 140$$

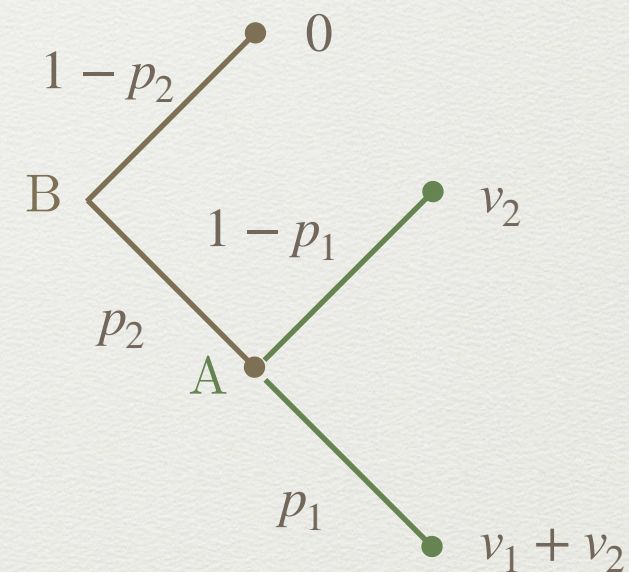
Problem z Quizem

Uogólnijmy naszą analizę. Oznaczmy przez p_1 i p_2 prawdopodobieństwa poprawnej odpowiedzi na pytania A i B, odpowiednio, a przez v_1 i v_2 — odpowiadające im nagrody.



Wybrano pierwsze pytanie A

$$\begin{aligned} E[X] &= p_1(1 - p_2)v_1 + p_1p_2(v_1 + v_2) \\ &= p_1v_1 + p_1p_2v_2 \end{aligned}$$



Wybrano pierwsze pytanie B

$$\begin{aligned} E[X] &= p_2(1 - p_1)v_2 + p_2p_1(v_1 + v_2) \\ &= p_2v_2 + p_1p_2v_1 \end{aligned}$$

Problem z Quizem

Zatem odpowiedź na pytanie A będzie optymalna (z maksymalną nagrodą) wtedy i tylko wtedy, gdy

$$p_1v_1 + p_1p_2v_2 \geq p_2v_2 + p_1p_2v_1$$

lub równoważnie, gdy

$$\frac{p_1v_1}{1-p_1} \geq \frac{p_2v_2}{1-p_2}.$$

Zatem, najlepiej jest tak uporządkować pytania, by wartości wyrażenia

$$\frac{pv}{1-p}, \quad p - \text{prawdopodobieństwo prawidłowej odpowiedzi, } v - \text{wartość nagrody}$$

maleły.

Uogólnienie

Rozważmy quiz, w którym uczestnik otrzymuje listę n pytań $A_1, \dots, A_n$ i może na nie odpowiedzieć w dowolnej kolejności. Prawdopodobieństwo poprawnej odpowiedzi na i -te pytanie wynosi p_i z nagrodą za poprawną odpowiedź v_i . Przy pierwszej błędnej odpowiedzi, quiz się kończy, a uczestnik zachowuje swoje dotychczasowe nagrody. Problem polega na wyborze takiej strategii odpowiedzi na pytania, by zmaksymalizować wartość oczekiwaną całkowitej nagrody. Pokaż, że najlepiej odpowiadać kolejno na pytania $A_{k_1}, \dots, A_{k_n}$ tak, by ciąg

$$a_i = \frac{p_{k_i} v_{k_i}}{1 - p_{k_i}}$$

był nierosnący, tzn. $a_{i+1} \leq a_i$ dla $i = 1, \dots, n$.