

# Rachunek prawdopodobieństwa 2R

Kycia

## Spis różnorodności treściowalnych

<b>06.10.23 : Warunkowa wartość oczekiwana</b>	<b>3</b>
1.1. Prawdopodobieństwo warunkowe . . . . .	3
1.2. Konstrukcja warunkowej wartości oczekiwanej . . . . .	3
1.3. Prawdopodobieństwo warunkowe . . . . .	6
<b>09.10.23 : Własności WWO</b>	<b>8</b>
2.1. Istnienie i jedyność . . . . .	8
2.2. Własności wwo . . . . .	10
<b>23.10.23 : Interpretacje geometryczne WWO</b>	<b>15</b>
3.1. Regularne rozkłady warunkowe . . . . .	17
<b>30.10.23 : Martynały</b>	<b>20</b>
4.1. Transformata martynałowa . . . . .	23

## Wykład 06.10.23 : Warunkowa wartość oczekiwana

### 1.1 Prawdopodobieństwo warunkowe

Tak jak zwykle do tej pory, pracować będziemy na przestrzeni probabilistycznej  $(\Omega, \mathcal{F}, \mathbb{P})$ .

Przypomnijmy definicję **prawdopodobieństwa warunkowego** z Rachunku Prawdopodobieństwa 1 (i z liceum). Dla zdarzenia  $A \in \mathcal{F}$  takiego, że  $\mathbb{P}[A] \in (0, 1)$  definiujemy prawdopodobieństwo warunkowe jako

$$\mathbb{P}[B | A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}.$$

Wartość ta informuje nas o zajściu B wtedy, gdy jesteśmy pewni, że A zaszło. Ale co, jeśli nasza wiedza dotycząca A jest mniej pewna? To znaczy, *co jeśli*  $\mathbb{P}[A] = 0$ ? Dość naturalne wydaje się rozważenie zdarzenia przeciwnego i zsumowania obu prawdopodobieństw:

$$\mathbb{1}_A \mathbb{P}[B | A] + \mathbb{1}_{A^c} \mathbb{P}[B | A^c].$$

Zauważmy od razu, że wyrażenie  $\mathbb{1}_A \mathbb{P}[B | A]$  jest zmienną losową.

W przypadku, gdy mamy dwa zbiory,  $A_1, A_2 \in \mathcal{F}$ , i chcemy zbadać  $\mathbb{P}[B | A_1 \cap A_2]$  możemy powyższe rozumowanie rozszerzyć na wszystkie możliwe kombinacje  $A_1, A_2$  i ich dopełnień:

$$\mathbb{1}_{A_1 \cap A_2} \mathbb{P}[B | A_1 \cap A_2] + \mathbb{1}_{A_1 \cap A_2^c} \mathbb{P}[B | A_1 \cap A_2^c] + \mathbb{1}_{A_1^c \cap A_2} \mathbb{P}[B | A_1^c \cap A_2] + \mathbb{1}_{A_1^c \cap A_2^c} \mathbb{P}[B | A_1^c \cap A_2^c].$$

Działanie jak wyżej daje pełną informację o każdym zdarzeniu z ciała generowanego przez zdarzenia  $A_1$  i  $A_2$ . Nazywamy je **rozbiciem** względem  $\sigma$ -ciała generowanego przez  $A_1$  i  $A_2$ .

Analogicznie możemy zdefiniować  $\mathbb{E}[X | A]$  dla całkowalnej zmiennej losowej X (tzn.  $\mathbb{E}[|X|] < \infty$ ):

$$\mathbb{E}[X | A] = \int_{\Omega} X(\omega) \mathbb{P}[d\omega | A] = \frac{1}{\mathbb{P}[A]} \mathbb{E}[X \mathbb{1}_A],$$

gdzie całka wyżej tłumaczy się na całkę po X względem miary  $\mathbb{P}[B | A]$ .

Uzasadnimy, dlaczego wzór wyżej jest zasadną definicją prawdopodobieństwa warunkowego przy ograniczonej wiedzy o zdarzeniu A.

### 1.2 Konstrukcja warunkowej wartości oczekiwanej

Zanim zdefiniujemy **warunkową wartość oczekiwaną [wwo]** zmiennej losowej X, zaczniemy od przyjrzenia się bliżej motywacji i konstrukcji stojącej za tym pojęciem.

Niech Z będzie całkowalną zmienną losową przyjmującą przeliczalnie wiele wartości. Zdefiniujmy funkcję

$$h(z) = \begin{cases} \mathbb{E}[X | Z = z] & \mathbb{P}[Z = z] > 0 \\ 0 & \text{wpp} \end{cases}$$

oraz zmienną losową  $Y = h(Z)$ . Weźmy dowolny  $C \in \text{Bor}(\mathbb{R})$  i zbadajmy  $\mathbb{E}[Y \mathbb{1}_{\{Z \in C\}}]$ . Zaczniemy od skorzystania z faktu, że Z przyjmuje przeliczalnie wiele wartości, więc możemy zapisać sumę po



nich wszystkich

$$\begin{aligned}\mathbb{E}[Y\mathbb{1}_{\{Z \in C\}}] &= \sum_{z \in C} h(z) \mathbb{P}[Z = z] = \\ &\stackrel{*}{=} \sum_{z \in C} \mathbb{E}[X\mathbb{1}_{\{Z=z\}}] \frac{1}{\mathbb{P}[Z = z]} \mathbb{P}[Z = z] = \\ &= \sum_{z \in C} \mathbb{E}[X\mathbb{1}_{\{Z=z\}}] = \\ &= \mathbb{E}\left[\sum_{z \in C} X\mathbb{1}_{\{Z=z\}}\right] = \\ &= \mathbb{E}[X\mathbb{1}_{\{Z \in C\}}]\end{aligned}$$

Równość  $\star$  wynika ze sposobu w jaki zdefiniowaliśmy  $\mathbb{E}[X | A]$  w poprzednim podrozdziale.

Zauważmy, że dowolne zdarzenie  $F \in \sigma(Z)$  jest postaci  $F = \{z \in C\}$  dla pewnego  $C \in \text{Bor}(\mathbb{R})$ . Wyprowadziliśmy więc równość:

$$\mathbb{E}[Y\mathbb{1}_F] = \mathbb{E}[X\mathbb{1}_F] \quad F \in \sigma(Z).$$

Pozostaje zapytać, co z tej zależności wynika?

Dla  $F = \Omega$  dostajemy

$$\mathbb{E}[h(Z)] = \mathbb{E}[Y] = \mathbb{E}[X].$$

### Dygresja.

W tym miejscu kuszące byłoby rozpisanie  $Y = h(Z)$  wprost z definicji, tzn.  $h(Z) = \mathbb{E}[X | Z = Z]$ , ale jest to całkowitą brednią. W definicji funkcji  $h$  podanej na samym początku przykładu  $z$  jest teoretycznym punkcikiem, natomiast przy definiowaniu  $Y = h(Z)$  ów  $Z$  jest już obserwowaną przez nas, konkretną zmienną losową. W takim razie, bardziej poprawny byłby zapis

$$h(Z(\omega)) = \mathbb{E}[X | \{\omega' : Z(\omega') = Z(\omega)\}].$$

## Przykład(y) 1.1

1. Ze zbioru  $\{1, 2, \dots, 10\}$  losujemy w sposób jednostajny liczbę i oznaczamy ją jako  $N$ . W drugim losowaniu, również w sposób jednostajny, wybieramy liczbę ze zbioru  $\{1, \dots, N\}$  i nazywamy ją  $M$ . Chcemy znaleźć średnią wartość liczby  $M$ . Oczywiście, nie jest trudno zrobić to metodami poznanymi na poprzednich przygodach probabilistycznych, jednak w tym przypadku użyjemy konstrukcji wyżej.

Funkcja  $h$  będzie wyglądać następująco:

$$h(n) = \mathbb{E}[M | N = n] = \sum_{1 \leq i \leq n} \frac{i}{n} = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

czyli  $h(N) = \frac{N+1}{2}$ .

Stosując notację jak wyżej, mamy

$$\begin{aligned}Z &= N \\ X &= M\end{aligned}$$

czyli podstawiając do wzoru:

$$\begin{aligned}\mathbb{E}[M] &= \mathbb{E}[h(N)] = \mathbb{E}\left[\frac{N+1}{2}\right] = \\ &= \frac{1}{2}(\mathbb{E}[N] + 1) = \frac{1}{2}\left(\sum_{1 \leq i \leq 10} \frac{i}{10} + 1\right) = \\ &= \frac{1}{2}\left(\frac{11}{2} + 1\right) = \frac{13}{4}\end{aligned}$$

Rozbicie jak wyżej można w elegancki sposób zamienić w bardziej abstrakcyjną definicję warunkowej wartości oczekiwanej.

### Definicja 1.1.

Niech  $\mathcal{G} \subseteq \mathcal{F}$  będzie  $\sigma$ -ciałem, a  $X$  całkowalną zmienną losową.

Zmienną losową  $Y$  nazywamy **warunkową wartością oczekiwaną** [wwo]  $X$  pod warunkiem  $\mathcal{G}$ , jeśli następujące warunki są spełnione:

(W1)  $Y$  jest  $\mathcal{G}$ -mierzalne

(W2)  $(\forall G \in \mathcal{G}) \mathbb{E}[X1_G] = \mathbb{E}[Y1_G]$

Nasuwać się teraz pytania o poprawność  $Y$  zdefiniowanego jak wyżej. Czy zawsze istnieje i czy jest on jedyny?

### Przykład(y) 1.2

1. Niech  $\mathcal{G} = \sigma(Z)$ , gdzie  $Z$  jest zmienną losową przyjmującą przeliczalnie wiele wartości. Wówczas  $Y = h(Z)$  dla  $h(z) = \mathbb{E}[X | Z = z]$  jest wwo  $X$  względem  $\mathcal{G}$ .

### Twierdzenie 1.1 : poprawność wwo.

Dla  $\sigma$ -ciała  $\mathcal{G} \subseteq \mathcal{F}$  i całkowalnej zmiennej losowej  $X$  **istnieje jedyna zmienna losowa**  $Y$  będąca wwo  $X$  względem  $\mathcal{G}$ . Będziemy ją oznaczać

$$\mathbb{E}[X | \mathcal{G}] = Y.$$

Jeśli  $Y, Y'$  są wwo  $X$  względem  $\mathcal{G}$ , to  $Y = Y'$  prawie wszędzie.

### Dowód

Dowód na następnym wykładzie.



### Uwaga 1.2.

O wwo  $X$  pod warunkiem  $\mathcal{G}$  należy myśleć jako o przybliżeniu  $X$  na podstawie informacji zawartych w  $\mathcal{G}$  (więcej na wykładzie 3).

### Przykład(y) 1.3

1. Jeśli  $X$  i  $\mathcal{G}$  są niezależne, to znaczy dla każdego  $B \in \text{Bor}(\mathbb{R})$  i dla każdego  $G \in \mathcal{G}$  zachodzi

$$\mathbb{P}[X \in B, G] = \mathbb{P}[X \in B] \mathbb{P}[G],$$

to wtedy  $\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X] = Y$ .

Warunek (W1) jest oczywiście spełniony, bo  $Y$  jest funkcją stałą, więc jego przeciwobraz to całość lub  $\emptyset$  (czyli jest  $\mathcal{G}$ -mierzalny). Warunek (W2) sprawdzamy dla dowolnego  $G \in \mathcal{G}$ :

$$\mathbb{E}[X 1_G] = \mathbb{E}[X] \mathbb{E}[1_G] = \mathbb{E}[\mathbb{E}[X] 1_G] = \mathbb{E}[Y 1_G].$$

2. Rozważmy pokrycie  $\Omega$  rozłącznymi zbiorami  $\{A_n\}_{n \in \mathbb{N}}$ , gdzie  $A_i \in \mathcal{F}$  dla każdego  $i$ . Niech  $\mathcal{G} = \sigma(A_i : i \in \mathbb{N})$  będzie  $\sigma$ -ciałem rozpinanym przez to pokrycie. Wówczas

$$\mathbb{E}[X | \mathcal{G}] = \sum_{i \in \mathbb{N}} 1_{A_i} \mathbb{E}[X | A_i]$$

Spełnianie pierwszego warunku jest oczywiste, bo mamy doczynienia z funkcją prostą. Warunek (W2) wystarczy sprawdzić dla atomów, czyli  $G = A_j$ , bo wszystkie zmienne losowe  $\mathcal{G}$ -mieralne są stałe na  $A_j$ .

$$\begin{aligned} \mathbb{E}\left[\left[\sum 1_{A_i} \mathbb{E}[X | A_i]\right] 1_{A_j}\right] &= \mathbb{E}\left[\mathbb{E}[X | A_j] 1_{A_j}\right] = \\ &= \mathbb{E}\left[1_{A_j} \frac{\mathbb{E}[X 1_{A_j}]}{\mathbb{P}[A_j]}\right] = \\ &= \mathbb{E}\left[1_{A_j}\right] \frac{\mathbb{E}[X 1_{A_j}]}{\mathbb{P}[A_j]} = \mathbb{E}[X 1_{A_j}], \end{aligned}$$

gdyż  $\mathbb{E}[1_{A_j}] = \mathbb{P}[A_j]$ .

3. Jeśli w przykładzie wyżej weźmiemy  $A_1 = A$ ,  $A_2 = A^c$  i  $A_3 = A_4 = \dots = \emptyset$  oraz  $\mathcal{G} = \sigma(A)$ , to dostajemy to samo co na samym początku tego wykładu:

$$\mathbb{E}[X | \mathcal{G}] = 1_A \mathbb{E}[X | A] + 1_{A^c} \mathbb{E}[X | A^c].$$

### 1.3 Prawdopodobieństwo warunkowe

#### Definicja 1.2 : prawdopodobieństwo warunkowe.

Dla  $\sigma$ -ciała  $\mathcal{G} \subseteq \mathcal{F}$  definiujemy **prawdopodobieństwo warunkowe** pod warunkiem  $\mathcal{G}$  jako

$$\mathbb{P}[A | \mathcal{G}] = \mathbb{E}[1_A | \mathcal{G}]$$

Prawdopodobieństwo  $\mathbb{P}[A | \mathcal{G}]$  jest zmienną losową taką, że:

→  $\mathbb{P}[A | \mathcal{G}]$  jest  $\mathcal{G}$ -mierzalna (ze względu na wwo w definicji)



$$\mathbb{E} [\mathbb{P} [A | \mathcal{G}] \mathbb{1}_G] = \mathbb{E} [\mathbb{1}_A \mathbb{1}_G] = \mathbb{P} [A \cap G]$$

### Przykład(y) 1.4

1. Niech  $E_1, E_2$  będą niezależnymi zmiennymi losowymi z rozkładem  $\exp(1)$ . Chcemy się zastanowić jak wygląda prawdopodobieństwo

$$\mathbb{P} [E_1 + E_2 > t | \sigma(E_1)]$$

dla  $t > 0$ . Ponieważ liczymy to prawdopodobieństwo względem  $\sigma(E_1)$ , to tak naprawdę wszystkie informacje o  $E_1$  mamy w ręku, gdyż tę zmienną obserwujemy. Czyli  $E_1$  możemy w takim przypadku potraktować jako zwykłą stałą i zgadnąć, że

$$\mathbb{P} [E_1 + E_2 > t | \sigma(E_1)] = e^{-(t-E_1)}.$$

Dla pewności, przeliczymy cały ten przykład wprost z definicji, żeby przekonać się że strzał był poprawny.

Niech  $G \in \sigma(E_1)$ , wtedy zgodnie z wcześniejszą obserwacją istnieje pewne  $C \in \text{Bor}(\mathbb{R})$  takie, że  $G$  jest postaci  $G = \{E_1 \in C\}$ . Policzmy  $\mathbb{E} [\mathbb{P} [\{E_1 + E_2 > t\} | \sigma(E_1)]]$  gdyż jak wyżej zauważyliśmy,  $\mathbb{P} [A | \mathcal{G}]$  jest zmienną losową. Mamy więc

$$\begin{aligned} \mathbb{E} [\mathbb{P} [E_1 + E_2 > t | \sigma(E_1)] \mathbb{1}_G] &\stackrel{*}{=} \mathbb{P} [\{E_1 + E_2 > t\} \cap G] = \\ &= \mathbb{P} [\{E_1 + E_2 > t\} \cap \{E_1 \in C\}] = \\ &= \iint_{\substack{C \times \mathbb{R}_+ \\ x+y>t}} e^{-x} e^{-y} dx dy = \\ &= \int_C e^{-x} \underbrace{\left[ \int_{x+y>t} e^{-y} dy \right]}_{**} dx = \\ &= \int_C e^{-x} e^{-(t-x)+} dx = \mathbb{E} [e^{-(t-E_1)+} \mathbb{1}_{\{E_1 \in C\}}] = \mathbb{E} [e^{-(t-E_1)+} \mathbb{1}_G] \end{aligned}$$

Równość  $*$  wynika z uwagi pod definicją prawdopodobieństwa warunkowego. Całka  $**$  jest równa 1 gdy  $x > t$  (gdyż wtedy dla każdego  $y$  mamy  $x + y > t$ ), natomiast dla  $x \leq t$  wynosi ona  $e^{-(t-x)}$ .

## Wykład 09.10.23 : Własności WWO

Na tym wykładzie zajmiemy się dowodzeniem własności wwo, w tym pokażemy jej istnienie i jedyność.

### 2.1 Istnienie i jedyność

#### Lemat 2.1 : WWO jest całkowalna.

To znaczy, że mając całkowalną zmienną losową  $X$  oraz  $\sigma$ -ciąto  $\mathcal{G} \subseteq \mathcal{F}$ , to zachodzi  $\mathbb{E} [|\mathbb{E} [X | \mathcal{G}]|] < \infty$ .

#### Dowód

Rozważmy zbiór

$$A = \{\omega : \mathbb{E} [X | \mathcal{G}] (\omega) > 0\} = \{\omega : \mathbb{E} [X | \mathcal{G}] \in (0, \infty)\} = [\mathbb{E} [X | \mathcal{G}]]^{-1} ((0, \infty))$$

jako przeciwobraz zbioru  $(0, \infty) \in \text{Bor}(\mathbb{R})$  przez funkcję  $\mathcal{G}$ -mierzalną  $\mathbb{E} [X | \mathcal{G}]$  wiemy, że  $A \in \mathcal{G}$ . Ponieważ  $\mathbb{E} [X | \mathcal{G}]$  jest wwo  $X$  pod warunkiem  $\mathcal{G}$ , to musi warunek (W2):

$$0 \leq \mathbb{E} [\mathbb{E} [X | \mathcal{G}] 1_A] = \mathbb{E} [X 1_A] \leq \mathbb{E} [|X| 1_A] < \infty$$

bo  $X$  jest całkowalna.

Analogicznie postępujemy dla zbioru  $A^c$ :

$$0 \leq \mathbb{E} [-\mathbb{E} [X | \mathcal{G}] 1_{A^c}] = \mathbb{E} [-X 1_{A^c}] \leq \mathbb{E} [|X| 1_{A^c}] < \infty.$$

Zauważmy, że

$$|\mathbb{E} [X | \mathcal{G}]| = \mathbb{E} [X | \mathcal{G}] 1_A - \mathbb{E} [X | \mathcal{G}] 1_{A^c}$$

Dodając obie te nierówności (i korzystając z liniowości wartości oczekiwanej) uzyskujemy

$$0 \leq \mathbb{E} [\mathbb{E} [X | \mathcal{G}] 1_A] - \mathbb{E} [\mathbb{E} [X | \mathcal{G}] 1_{A^c}] = \mathbb{E} [\mathbb{E} [X | \mathcal{G}] 1_A - \mathbb{E} [X | \mathcal{G}] 1_{A^c}] = \mathbb{E} [|\mathbb{E} [X | \mathcal{G}]|] < \infty$$



#### Lemat 2.2 : jedyność p.w..

Niech  $\mathcal{G} \subseteq \mathcal{F}$  będzie  $\sigma$ -ciątem. Jeśli  $Y$  i  $Y'$  są obie wersjami  $\mathbb{E} [X | \mathcal{G}]$ , to  $Y = Y'$  p.w..

#### Dowód

Ustalmy  $\varepsilon > 0$  i rozważmy zdarzenie

$$A_\varepsilon = \{Y - Y' > \varepsilon\} \in \mathcal{G}$$

które jest  $\mathcal{G}$ -mierzalne, bo  $Y$  i  $Y'$  takie są.



$$\begin{aligned}
\varepsilon \mathbb{P}[A_\varepsilon] + \mathbb{E}[Y' \mathbb{1}_{A_\varepsilon}] &= \mathbb{E}[\varepsilon \mathbb{1}_{A_\varepsilon}] + \mathbb{E}[Y' \mathbb{1}_{A_\varepsilon}] = \\
&= \mathbb{E}[(\varepsilon + Y') \mathbb{1}_{A_\varepsilon}] \leq \\
&\stackrel{*}{\leq} \mathbb{E}[Y \mathbb{1}_{A_\varepsilon}] \stackrel{(W2)}{=} \mathbb{E}[X \mathbb{1}_{A_\varepsilon}] = \\
&= \mathbb{E}[Y' \mathbb{1}_{A_\varepsilon}]
\end{aligned}$$

gdzie  $*$  wynika z tego, że na zbiorze  $A_\varepsilon$   $Y > Y' + \varepsilon$ .

Dostajemy więc, że

$$\varepsilon \mathbb{P}[A_\varepsilon] + \mathbb{E}[Y' \mathbb{1}_{A_\varepsilon}] \leq \mathbb{E}[Y' \mathbb{1}_{A_\varepsilon}]$$

co po przeniesieniu  $\mathbb{E}$  na jedną stronę daje

$$\varepsilon \mathbb{P}[A_\varepsilon] \leq 0$$

a ponieważ  $\varepsilon > 0$ , to musi być  $\mathbb{P}[A_\varepsilon] = 0$ .

Wówczas

$$\mathbb{P}[Y > Y'] = \underbrace{\mathbb{P}\left[\left(\exists n\right) Y \geq Y' + \frac{1}{n}\right]}_{\mathbb{P}\left[A_{\frac{1}{n}}\right]} = \mathbb{P}\left[\bigcup A_{\frac{1}{n}}\right] = \lim \mathbb{P}\left[A_{\frac{1}{n}}\right] = 0$$

ponieważ  $A_{\frac{1}{n}} \subseteq A_{\frac{1}{n+1}}$ .

Zamieniając miejscami  $Y$  i  $Y'$  w dowodzie dostaniemy  $\mathbb{P}[Y' > Y] = 0$ , czyli obie możliwości są miary zero.



### Twierdzenie 2.3 : o istnieniu WWO.

Niech  $\mathcal{G} \subseteq \mathcal{F}$  będzie  $\sigma$ -ciałem, a  $X$  jest całkowalną zmienną losową. Istnieje zmienna losowa  $Y$  spełniająca oba postulaty wwo  $X$  pod warunkiem  $\mathcal{G}$ .

Jest to Twierdzenie 1.1 z poprzedniego wykładu.

Zanim jednak przejdziemy do dowodu 2.3, przypomnijmy *twierdzenie Radona-Nikodyma* z teorii miary:

#### Dygresja : twierdzenie Radona-Nikodyma.

Niech  $\mu$  i  $\nu$  będą  $\sigma$ -miarami na przestrzeni  $(\Omega, \mathcal{G})$  takimi, że  $\nu$  jest *absolutnie ciągła* względem  $\mu$  [ $\nu \ll \mu$ ], tzn  $\mu(A) = 0 \Rightarrow \nu(A) = 0$ . Wówczas istnieje  $\mathcal{G}$ -mierzalna funkcja  $f : \Omega \rightarrow \mathbb{R}$  taka, że

$$\nu(A) = \int_A f(x) \mu(dx)$$

Funkcję  $f$  jak wyżej często oznaczamy  $f = \frac{d\nu}{d\mu}$  i nazywamy *pochodną Radona-Nikodyma*.

### Dowód

Wracając do dowodu twierdzenia 2.3. Najpierw pokażemy prostszy przykład, gdy  $X \geq 0$ , a potem uogólnimy go do dowolnego  $X$ .

Założmy, że  $X \geq 0$  p.w. Wtedy możemy rozważyć miary  $\mu = \mathbb{P} \upharpoonright \mathcal{G}$  oraz  $\nu(A) = \mathbb{E}[X1_A]$ . Od razu widać, że w takim ułożeniu  $\nu \ll \mu$ , więc na mocy twierdzenia Radona-Nikodyma istnieje  $f$   $\mathcal{G}$ -mierzalna taka, że

$$\mathbb{E}[f1_A] = \int_A f(\omega) \mu(d\omega) = \nu(A) = \mathbb{E}[X1_A].$$

Funkcja  $f$  spełnia (W1) z definicji wwo, bo jest  $\mathcal{G}$ -mierzalna, a (W2) jest potwierdzone przez rachunek wyżej. Czyli  $f$  jest wwo  $X$  pod warunkiem  $\mathcal{G}$ .

Niech teraz  $X$  będzie dowolną zmienną losową. Możemy ją rozbić jako

$$X = X^+ - X^-,$$

gdzie  $X^+ = \max(0, X) \geq 0$  oraz  $X^- = -\min(0, X) \geq 0$ . Do obu tych zmiennych możemy zastosować pierwszą część dowodu, by dostać zmienne  $\mathbb{E}[X^+ | \mathcal{G}]$  oraz  $\mathbb{E}[X^- | \mathcal{G}]$ . Wystarczy zauważyć, że dzięki liniowości  $\mathbb{E}$  możemy w prosty sposób pokazać

$$\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X^+ | \mathcal{G}] - \mathbb{E}[X^- | \mathcal{G}]$$



## 2.2 Własności wwo

### Twierdzenie 2.4 : o arytmetyce wwo.

Niech  $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$  będą  $\sigma$ -ciałami, a  $X, X_1, X_2$  całkowalnymi zmiennymi losowymi

1.  $\mathbb{E}[\mathbb{E}[X | \mathcal{G}]] = \mathbb{E}[X]$
2. Jeśli  $X \geq 0$ , to również  $\mathbb{E}[X | \mathcal{G}] \geq 0$
3.  $\mathbb{E}[aX_1 + bX_2 | \mathcal{G}] = a\mathbb{E}[X_1 | \mathcal{G}] + b\mathbb{E}[X_2 | \mathcal{G}]$
4.  $|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}[|X| | \mathcal{G}]$
5. Jeśli  $\mathcal{G}_1 \subseteq \mathcal{G}_2$ , to wówczas

$$\mathbb{E}[\mathbb{E}[X | \mathcal{G}_1] | \mathcal{G}_2] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}_2] | \mathcal{G}_1] = \mathbb{E}[X | \mathcal{G}_1]$$

To znaczy, że mając informacje o  $X$  w dwóch zawartych w sobie ciałach, to mniejsze zawsze wygrywa.

6. Jeśli  $Y$  jest  $\mathcal{G}$ -mierzalna i  $XY$  jest całkowalna, to  $\mathbb{E}[XY | \mathcal{G}] = Y\mathbb{E}[X | \mathcal{G}]$ , czyli  $Y$  możemy traktować jako stałą.

### Dowód



1. Wystarczy wstawić  $G = \Omega$  w warunek (W2):

$$\mathbb{E} [\mathbb{E} [X | \mathcal{G}]] = \mathbb{E} [\mathbb{E} [X | \mathcal{G}] 1_{\Omega}] \stackrel{(W2)}{=} \mathbb{E} [X 1_{\Omega}] = \mathbb{E} [X]$$

2. Wynika z dowodu twierdzenia o istnieniu, bo  $\frac{d\nu}{d\mu} = \mathbb{E} [X | \mathcal{G}]$ . Gdyby  $A = \{\omega : \mathbb{E} [X | \mathcal{G}] < 0\}$ , to wówczas

$$\mathbb{E} [X 1_A] = \nu(A) = \int_A \mathbb{E} [X | \mathcal{G}] (\omega) \mu(d\omega) = \mathbb{E} [\mathbb{E} [X | \mathcal{G}] 1_A] < 0$$

ale przecież  $X \geq 0 \Rightarrow \mathbb{E} [X 1_A] \geq 0$ , więc  $A = \emptyset$ .

3. Można to zrobić na dwa sposoby: licząc wszystko pokolei, albo można sprawdzić, czy  $Y = a\mathbb{E} [X_1 | \mathcal{G}] + b\mathbb{E} [X_2 | \mathcal{G}]$  spełnia warunki wwo tej samej zmiennej co  $\mathbb{E} [aX_1 + bX_2 | \mathcal{G}]$ . Wówczas obie te zmienne są równe prawie wszędzie.

Warunek  $\mathcal{G}$ -mierzalności dla  $Y$  jest spełniony, bo  $Y$  jest kombinacją liniową dwóch funkcji  $\mathcal{G}$ -mierzalnych. Wystarczy więc sprawdzić warunek (W2). W tym celu ustalmy  $A \in \mathcal{G}$ .

$$\begin{aligned} \mathbb{E} [Y 1_A] &\stackrel{*}{=} a\mathbb{E} [\mathbb{E} [X_1 | \mathcal{G}] 1_A] + b\mathbb{E} [\mathbb{E} [X_2 | \mathcal{G}] 1_A] = \\ &= a\mathbb{E} [X_1 1_A] + b\mathbb{E} [X_2 1_A] = \\ &= \mathbb{E} [(aX_1 + bX_2) 1_A] = \mathbb{E} [\mathbb{E} [aX_1 + bX_2 | \mathcal{G}] 1_A] \end{aligned}$$

4. Wiemy, że  $-|X| \leq X \leq |X|$ . Korzystając z punktu 2 dostajemy

$$0 \leq X + |X| \Rightarrow 0 \leq \mathbb{E} [|X| | \mathcal{G}] + \mathbb{E} [X | \mathcal{G}] \Rightarrow -\mathbb{E} [|X| | \mathcal{G}] \leq \mathbb{E} [X | \mathcal{G}]$$

$$0 \leq |X| - X \Rightarrow 0 \leq \mathbb{E} [|X| | \mathcal{G}] - \mathbb{E} [X | \mathcal{G}] \Rightarrow \mathbb{E} [X | \mathcal{G}] \leq \mathbb{E} [|X| | \mathcal{G}]$$

Po złożeniu tych dwóch nierówności:

$$-\mathbb{E} [|X| | \mathcal{G}] \leq \mathbb{E} [X | \mathcal{G}] \leq \mathbb{E} [|X| | \mathcal{G}]$$

wiemy, że  $-\mathbb{E} [X | \mathcal{G}] \leq \mathbb{E} [X | \mathcal{G}] \leq \mathbb{E} [X | \mathcal{G}]$ , więc musi być

$$|\mathbb{E} [X | \mathcal{G}]| \leq \mathbb{E} [|X| | \mathcal{G}].$$

5. Zaczniemy od sprawdzenia, że  $\mathbb{E} [\mathbb{E} [X | \mathcal{G}_2] | \mathcal{G}_1] = \mathbb{E} [X | \mathcal{G}_1]$ . Wybierzmy  $A \in \mathcal{G}_1 \subseteq \mathcal{G}_2$ :

$$\mathbb{E} [\mathbb{E} [X | \mathcal{G}_1] 1_A] = \mathbb{E} [X 1_A] = \mathbb{E} [\mathbb{E} [X | \mathcal{G}_2] 1_A]$$

co potwierdza warunek (W2).  $\mathcal{G}_1$ -mierzalność  $\mathbb{E} [X | \mathcal{G}_2]$  jest oczywista, gdyż  $\mathcal{G}_1 \subseteq \mathcal{G}_2$ ,  $\mathbb{E} [X | \mathcal{G}_2]$  jest  $\mathcal{G}_2$ -mierzalne, a po obcięciu do  $\mathcal{G}_1$  dostajemy funkcję  $\mathcal{G}_1$ -mierzalną.

Pozostaje nam sprawdzić czym jest  $\mathbb{E} [\mathbb{E} [X | \mathcal{G}_1] | \mathcal{G}_2]$ . Roboczo nazwiemy  $Y = \mathbb{E} [X | \mathcal{G}_1]$ . Jest to funkcja  $\mathcal{G}_1$ -mierzalna, ale dzięki  $\mathcal{G}_1 \subseteq \mathcal{G}_2$  mamy też  $\mathcal{G}_2$ -mierzalność. W takim razie (tak jak w jednym z przykładów z pierwszego wykładu)  $\mathbb{E} [Y | \mathcal{G}_2] = Y$ . Piszac bez używania litery  $Y$  dostajemy

$$\mathbb{E} [\mathbb{E} [X | \mathcal{G}_1] | \mathcal{G}_2] = \mathbb{E} [X | \mathcal{G}_1]$$

6. Ćwiczenie, a poniżej moja próba.

Jeśli  $Y$  jest  $\mathcal{G}$ -mierzalne, to  $Y\mathbb{E} [X | \mathcal{G}]$  też takie jest jako iloczyn dwóch funkcji  $\mathcal{G}$ -mierzalnych. Pozostaje sprawdzić warunek (W2).



Zaczniemy od  $Y = \sum a_i 1_{A_i}$  dla  $A_i \in \mathcal{G}$ , czyli od funkcji prostej. Wybierając  $A \in \mathcal{G}$  możemy ograniczyć się do zbiorów  $A_i$ , gdyż są one rozłączne i na dowolnym innym zbiorze  $Y = 0$ . Mamy więc

$$\begin{aligned} \mathbb{E} [\mathbb{E} [XY | \mathcal{G}] 1_{A_i}] &\stackrel{(W2)}{=} \mathbb{E} [XY 1_{A_i}] = \mathbb{E} [a_i X 1_{A_i}] = a_i \mathbb{E} [X 1_{A_i}] = \\ &= a_i \mathbb{E} [\mathbb{E} [X | \mathcal{G}] 1_{A_i}] = \mathbb{E} [(a_i \mathbb{E} [X | \mathcal{G}]) 1_{A_i}] = \\ &= \mathbb{E} [(Y \mathbb{E} [X | \mathcal{G}]) 1_{A_i}] \end{aligned}$$

Czyli  $\mathbb{E} [XY | \mathcal{G}] = Y \mathbb{E} [X | \mathcal{G}]$  dla przypadku gdy  $Y$  jest funkcją prostą.

Jeśli teraz  $Y$  jest dowolną nieujemną funkcją mierzalną, to istnieje ciąg funkcji prostych

$$s_1 \leq s_2 \leq \dots \leq s_n \leq \dots \quad \lim s_i = Y$$

Wówczas dla dowolnego  $A \in \mathcal{G}$

$$\begin{aligned} \mathbb{E} [\mathbb{E} [XY] 1_A] &= \mathbb{E} [XY 1_A] = \mathbb{E} [X \lim s_i 1_A] \stackrel{*}{=} \lim \mathbb{E} [X s_i 1_A] = \\ &\stackrel{**}{=} \lim \mathbb{E} [s_i \mathbb{E} [X | \mathcal{G}] 1_A] = \mathbb{E} [\lim s_i \mathbb{E} [X | \mathcal{G}] 1_A] = \\ &= \mathbb{E} [Y \mathbb{E} [X | \mathcal{G}] 1_A] \end{aligned}$$

\* można zrobić na mocy twierdzenia o monotoniczności ciągu  $s_i$  dla zwykłej  $\mathbb{E}$ , natomiast \*\* stosuje poprzedni przypadek  $Y$ .

Pozostaje przypadek, gdy  $Y$  jest dowolną  $\mathcal{G}$ -mierzalną zmienną losową. Wówczas możemy rozbić  $Y = Y^+ - Y^-$  i skorzystać z liniowości wwo:

$$\mathbb{E} [XY | \mathcal{G}] = \mathbb{E} [XY^+ | \mathcal{G}] - \mathbb{E} [XY^- | \mathcal{G}] = Y^+ \mathbb{E} [X | \mathcal{G}] - Y^- \mathbb{E} [X | \mathcal{G}] = Y \mathbb{E} [X | \mathcal{G}]$$



### Twierdzenie 2.5 : o zbieżności i ciągłości.

Niech  $\mathcal{G} \subseteq \mathcal{F}$  będzie  $\sigma$ -ciałem, a  $X, X_1, X_2, \dots$  będzie ciągiem całkowalnych zmiennych losowych. Wówczas

1. Jeśli  $0 \leq X_1 \leq X_2 \leq \dots$  oraz  $X_n \nearrow X$ , to  $\mathbb{E} [X_n | \mathcal{G}] \nearrow \mathbb{E} [X | \mathcal{G}]$  p.w. (twierdzenie o zbieżności monotonicznej)
2. Jeśli  $X \geq 0$ , to  $\mathbb{E} [\liminf_n X_n | \mathcal{G}] \leq \liminf_n \mathbb{E} [X_n | \mathcal{G}]$  (lemat Fatou).
3. Jeśli  $|X_n| \leq Y$  oraz  $Y$  jest całkowalny i  $X_n \rightarrow X$  p.w., to  $\mathbb{E} [X_n | \mathcal{G}] \rightarrow \mathbb{E} [X | \mathcal{G}]$  (twierdzenie o zbieżności ograniczonej)

#### Dowód

1. Zauważamy, że ciąg  $\mathbb{E} [X_n | \mathcal{G}]$  jest niemalejący i ograniczony przez  $\mathbb{E} [X | \mathcal{G}]$  (na mocy punktu 2 z poprzedniego twierdzenia).

Niech  $Y = \lim \mathbb{E} [X_n | \mathcal{G}]$ . Wystarczy, że pokażemy  $Y = \mathbb{E} [X | \mathcal{G}]$  p.w., czyli sprawdzimy warunki (W1) i (W2). Oczywiście, warunek (W1) wynika z faktu, że granica ciągu funkcji  $\mathcal{G}$ -mierzalnych jest nadal  $\mathcal{G}$ -mierzalna. Dla sprawdzenia warunku (W2) wybierzmy  $A \in \mathcal{G}$

$$\begin{aligned} \mathbb{E} [Y 1_A] &= \mathbb{E} [\lim \mathbb{E} [X_n | \mathcal{G}] 1_A] \stackrel{*}{=} \lim \mathbb{E} [\mathbb{E} [X_n | \mathcal{G}] 1_A] = \\ &= \lim \mathbb{E} [X_n 1_A] = \mathbb{E} [\lim X_n 1_A] = \mathbb{E} [X 1_A] \end{aligned}$$

czyli  $Y = \mathbb{E} [X | \mathcal{G}]$  p.w.

2. Zaczniemy od dwóch obserwacji:

➡ Dla ciągu  $\{a_n\}$   $\liminf a_n$  to najmniejszy z jego punktów skupienia, równoważnie:

$$\liminf_n a_n = \lim_{k \rightarrow \infty} \inf_{n \geq k} a_n$$

✿ Dla dowolnej przeliczalnej rodziny zmiennych losowych  $\{Z_n\}_{n \in T}$  i dla dowolnego  $t \in T$  mamy

$$\begin{aligned} \inf_{s \in T} Z_s &\leq Z_t \\ \mathbb{E} \left[ \inf_{s \in T} Z_s \right] &\leq \mathbb{E} [Z_t] \\ \mathbb{E} \left[ \inf_{s \in T} Z_s \right] &\leq \inf_{t \in T} \mathbb{E} [Z_t] \end{aligned}$$

(co jest tak naprawdę wersją lematu Fatou dla  $\mathbb{E}$  z RP1R).

Stosując obserwację ➡ w przejściach  $*$ , obserwację ✿ w przejściu  $***$  oraz pkt 1. ( $\inf_{n>k} X_n \leq \inf_{n>k+1} X_n$ ) w  $***$ , dostajemy

$$\begin{aligned} \mathbb{E} [\liminf_n X_n | \mathcal{G}] &\stackrel{*}{=} \mathbb{E} \left[ \lim_{k \rightarrow \infty} \inf_{n>k} X_n | \mathcal{G} \right] \stackrel{***}{=} \lim_k \mathbb{E} \left[ \inf_{n>k} X_n | \mathcal{G} \right] \leq \\ &\stackrel{**}{\leq} \lim_{k \rightarrow \infty} \inf_{n>k} \mathbb{E} [X_n | \mathcal{G}] \stackrel{*}{=} \liminf_n \mathbb{E} [X_n | \mathcal{G}] \end{aligned}$$

3. Rozważmy zmienne  $X'_n = Y + X_n$ . Ponieważ  $|X_n| \leq Y$ , to  $Y + X_n \geq 0$ .

$$\begin{aligned} \mathbb{E} [Y + \liminf X_n | \mathcal{G}] &= \mathbb{E} [\liminf (X_n + Y) | \mathcal{G}] \stackrel{3.}{\leq} \liminf \mathbb{E} [Y + X_n | \mathcal{G}] = \\ &= \mathbb{E} [Y] + \liminf \mathbb{E} [X_n | \mathcal{G}] \end{aligned}$$

To daje nam, że  $\mathbb{E} [\liminf X_n | \mathcal{G}] \leq \liminf \mathbb{E} [X_n | \mathcal{G}]$ .

Postępując analogicznie dla  $X''_n = Y - X_n$  (które dalej jest  $\geq 0$ ) dostaniemy  $\mathbb{E} [\limsup X_n | \mathcal{G}] \geq \limsup \mathbb{E} [X_n | \mathcal{G}]$ :

$$\begin{aligned} \mathbb{E} [Y - \limsup X_n | \mathcal{G}] &= \mathbb{E} [\limsup (Y - X_n) | \mathcal{G}] = \\ &= -\mathbb{E} [\liminf (X_n - Y) | \mathcal{G}] \stackrel{3.}{\geq} -\liminf \mathbb{E} [X_n - Y | \mathcal{G}] = \\ &= \limsup \mathbb{E} [Y - X_n | \mathcal{G}] \end{aligned}$$

Ale wiemy, że  $\liminf X_n = X$  oraz  $\limsup X_n = X$ , czyli

$$\liminf \mathbb{E} [X_n | \mathcal{G}] \geq \mathbb{E} [\liminf X_n | \mathcal{G}] = \mathbb{E} [X | \mathcal{G}] = \mathbb{E} [\limsup X_n | \mathcal{G}] \geq \limsup \mathbb{E} [X_n | \mathcal{G}]$$

ale przecież  $\liminf \mathbb{E} [X_n | \mathcal{G}] \leq \limsup \mathbb{E} [X_n | \mathcal{G}]$ , czyli musi być

$$\liminf \mathbb{E} [X_n | \mathcal{G}] = \mathbb{E} [X | \mathcal{G}] = \limsup \mathbb{E} [X_n | \mathcal{G}]$$

i ponieważ  $\liminf = \limsup = \lim$  to mamy

$$\lim \mathbb{E} [X_n | \mathcal{G}] = \mathbb{E} [X | \mathcal{G}].$$



### Twierdzenie 2.6.

Niech  $\mathcal{G} \subseteq \mathcal{F}$  będzie  $\sigma$ -ciałem. Załóżmy, że

🐟 X jest  $\mathcal{G}$ -mierzalna

🐟 Y jest niezależna od  $\mathcal{G}$

🐟 funkcja  $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$  jest mierzalna taka, że

$$\mathbb{E} [\psi(X, Y)] < \infty.$$

Wówczas

$$\mathbb{E} [\psi(X, Y) | \mathcal{G}] = \Psi(X),$$

gdzie funkcja  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  jest zdefiniowana jako  $\Psi(X) = \mathbb{E} [\psi(X, Y)]$ .

### Dowód

Jeśli  $\psi(x, y) = xy$ , to korzystając z punktu 3. twierdzenia 2.4 dostajemy

$$\psi(x) = \mathbb{E} [xY] = x\mathbb{E} [Y]$$





## Wykład 23.10.23 : Interpretacje geometryczne WWO

Rozważmy zmienną losową  $X$  taką, że  $\mathbb{E}[X^2] < \infty$ . Interesuje nas zagadnienie regresji, mianowicie obserwując inną zmienną losową  $Z$  chcemy móc  $X$  aproksymować. Przez przybliżanie  $X$  rozumiemy przybliżanie średniokwadratowe.

Szukamy więc funkcji mierzalnej  $h_0 : \mathbb{R} \rightarrow \mathbb{R}$  takiej, żeby

$$\mathbb{E}[(X - h_0(Z))^2] = \inf_{h: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E}[(X - h(Z))^2]$$

### Fakt 3.1.

Dla każdej zmiennej losowej  $Y$  mierzalnej względem  $\sigma(Z)$  (co w skrócie będziemy notować  $Y \in \sigma(Z)$ ) istnieje  $h$  takie, że  $Y = h(Z)$ .

### Dowód

Zadanie.



$$\mathbb{E}[(X - h(Z))^2] = \inf_{Y \in \sigma(Z)} \mathbb{E}[(X - Y)^2]$$

mając pewną wiedzę o przestrzeniach Hilberta jest do tego dość naturalne podejście: rzut ortogonalny.

Dla  $\sigma$ -ciała  $\mathcal{G} \subseteq \mathcal{F}$  będziemy rozważać

$$L^2(\mathcal{G}) = \{Y \in \mathcal{G} : \mathbb{E}[Y^2] < \infty\}$$

wówczas  $L^2(\mathcal{G})$  jest przestrzenią Euklidesową z iloczynem skalarnym  $\langle X, Y \rangle = \mathbb{E}[XY]$ , który z kolei zadaje normę

$$\|X\| = \sqrt{\mathbb{E}[X^2]}.$$

Używając tego języka będziemy wiedzieli jak szukać  $Y$  z początku wykładu, ale najpierw fakt pomocniczy do twierdzenia które nadejdzie lada moment.

### Fakt 3.2 : warunkowa wersja nierówności Cauchy'ego-Schwartz.

Dla zmiennych  $X, Y$  takich, że  $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$  i  $\sigma$ -ciała  $\mathcal{G} \subseteq \mathcal{F}$  zachodzi

$$\mathbb{E}[|XY| \mid \mathcal{G}] \leq \mathbb{E}[X^2 \mid \mathcal{G}]^{\frac{1}{2}} \mathbb{E}[Y^2 \mid \mathcal{G}]^{\frac{1}{2}}$$

### Dowód

Zadanie



Z tego nierówności w fakcie 3.2 wynika, że dla  $Y = 1$  mamy

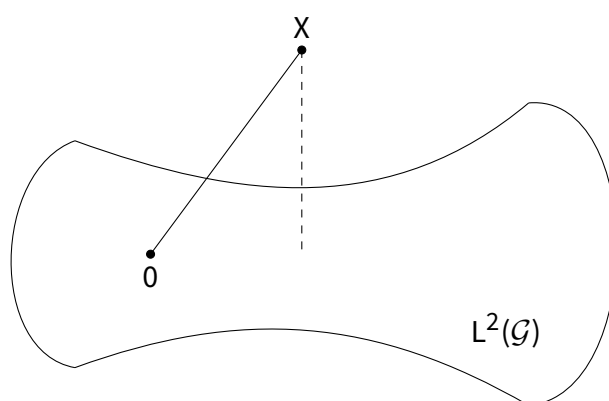
$$\mathbb{E} [|X| | \mathcal{G}]^2 \leq \mathbb{E} [X^2 | \mathcal{G}] \Rightarrow \mathbb{E} [\mathbb{E} [X | \mathcal{G}]^2] < \infty$$

### Twierdzenie 3.3 : wwo jest rzutem ortogonalnym na $L^2(\mathcal{G})$ .

Niech  $X$  będzie zmienną losową taką, że  $\mathbb{E} [X^2] < \infty$ , a  $\mathcal{G} \subseteq \mathcal{F}$  jest  $\sigma$ -ciałem. Wówczas

$$\mathbb{E} [X | \mathcal{G}] \in L^2(\mathcal{G})$$

jest *rzutem ortogonalnym*  $X$  na  $L^2(\mathcal{G})$ .



Dokładniej,  $\mathbb{E} [X | \mathcal{G}]$  daje minimum zbioru  $\{\mathbb{E} [(X - Y)^2] : Y \in L^2(\mathcal{G})\}$ . Z faktu 3.1 dla  $\mathcal{G} = \sigma(Z)$ ,  $\mathbb{E} [X | \mathcal{G}] = h_0(Z)$  dla pewnego  $h_0$ .

#### Dowód

Dla  $Y \in L^2(\mathcal{G})$  mamy

$$\begin{aligned} \mathbb{E} [(X - Y)^2] &= \mathbb{E} \left[ ((X - \mathbb{E} [X | \mathcal{G}]) - \overbrace{(Y - \mathbb{E} [X | \mathcal{G}])}^{=Y'})^2 \right] = \\ &= \mathbb{E} [(X - \mathbb{E} [X | \mathcal{G}])^2] - 2\mathbb{E} [(X - \mathbb{E} [X | \mathcal{G}])Y'] + \mathbb{E} [(Y')^2] = \\ &\stackrel{*}{=} \mathbb{E} [(X - \mathbb{E} [X | \mathcal{G}])^2] + \mathbb{E} [(Y')^2] \end{aligned}$$

Zauważmy, że

$$\begin{aligned} \mathbb{E} [Y'X | \mathcal{G}] &= Y' \mathbb{E} [X | \mathcal{G}] \\ \mathbb{E} [\mathbb{E} [Y'X | \mathcal{G}]] &= \mathbb{E} [Y'X] = \mathbb{E} [Y' \mathbb{E} [X | \mathcal{G}]] \end{aligned}$$



### Przykład(y) 3.1

1. Niewiele mający z tym co przed chwilą było. Niech  $\Omega = [0, 1]$ ,  $\mathcal{F} = \text{Bor}([0, 1])$  i  $\mathbb{P} = \lambda$ . Chcemy rozważyć  $t \in (0, 1)$  oraz  $\mathcal{G} = \sigma(\text{Bor}([0, t]))$ .

Dla całkowalnej zmiennej losowej  $X$  szukamy  $\mathbb{E}[X | \mathcal{G}]$ .

Dowolny  $G \in \mathcal{G}$  ma postać  $G = A \cup B$ , gdzie  $A \in \text{Bor}([0, t])$  i  $B \in \{(t, 1], \emptyset\}$ . W takim razie, jeśli  $Y \in \mathcal{G}$ , to  $Y$  jest stała na  $(t, 1]$ . Czyli żeby  $Y = \mathbb{E}[X | \mathcal{G}]$  to zapewne będzie postaci:

$$Y(\omega) = X(\omega)\mathbb{1}_{[0, t]}(\omega) + c\mathbb{1}_{(t, 1]}(\omega)$$

gdzie  $c$  jest pewną stałą.

Musimy sprawdzić, czy (i kiedy)  $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[Y\mathbb{1}_G]$ . Widać od razu, że dla  $G = A \cup B$  jak wyżej, mamy

$$X\mathbb{1}_A = Y\mathbb{1}_A,$$

czyli  $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$ . Zostaje nam uzgodnić część  $B$  kiedy jest on niepusty:

$$\mathbb{E}[X\mathbb{1}_B] = \int_t^1 X(s)ds$$

$$\mathbb{E}[Y\mathbb{1}_B] = c(1 - t),$$

czyli  $c$  musi być średnią  $X$  na przedziale  $(t, 1]$ :

$$c = \frac{1}{1 - t} \int_t^1 X(s)ds.$$

### 3.1 Regularne rozkłady warunkowe

Dla zmiennej losowej  $Y$  i całkowalnej zmiennej losowej  $X$  napis

$$\mathbb{E}[X | Y] := \mathbb{E}[X | \sigma(Y)]$$

będzie wwo  $X$  względem  $\sigma$ -ciała generowanego przez  $Y$ .

#### Zadanie dla dociekliwych:

Jeśli  $\mathbb{P}[Y = y] > 0$  dla  $y \in \mathbb{R}$ , to biorąc  $\omega \in \{Y = y\}$  dostajemy:

$$\mathbb{E}[X | Y](\omega) = \mathbb{E}[X | Y = y] = \frac{1}{\mathbb{P}[Y = y]} \mathbb{E}[X\mathbb{1}_{\{Y=y\}}]$$

#### Definicja 3.1 : wwo $X$ pod warunkiem $\{Y = y\}$ .

Po pierwsze zauważamy, że istnieje funkcja  $h : \mathbb{R} \rightarrow \mathbb{R}$  taka, że  $\mathbb{E}[X | Y] = h(Y)$ . (\*)

Niech  $X$  i  $Y$  będą dowolnymi zmiennymi losowymi, przy czym od  $X$  wymagamy całkowalności. Dla  $y \in \mathbb{R}$  warunkową wartość oczekiwaną  $X$  pod warunkiem  $\{Y = y\}$  definiujemy przez

$$\mathbb{E}[X | Y = y] = h(y)$$

gdzie  $h$  spełnia warunek (\*).

Podobnie definiujemy prawdopodobieństwo zbioru  $A$  pod warunkiem  $\{Y = y\}$ :

$$\mathbb{P}[A | Y = y] = \mathbb{E}[\mathbb{1}_A | Y = y]$$



## Przykład(y) 3.2

1. Jeżeli  $X$  i  $Z$  są niezależne, to chcemy zapytać o

$$\mathbb{P}[X + Z \in B \mid X = x] \stackrel{?}{=} \mathbb{P}[Z + x \in B] = \mu_Z(B - x)$$

Wystawiając tę wartość w terminach całego  $X$ :

$$\mathbb{P}[X + Z \in B \mid X] = \mathbb{E}[\mathbb{1}_{X+Z \in B} \mid X] \stackrel{*}{=} \mathbb{E}[\phi(X, Z) \mid X] = \Phi(X)$$

w  $*$  definiujemy:  $\phi(x, z) = \mathbb{1}_{x+z \in B}$ . Dla ustalonego  $x$  mamy więc:

$$h(x) = \mathbb{E}[\phi(x, Z)] = \mathbb{E}[\mathbb{1}_{x+Z \in B}] = \mathbb{P}[Z + x \in B]$$

2. Niech wektor losowy  $(X, Y)$  ma gęstość łączną  $f(x, y)$ . Wówczas

$$\mathbb{P}[X \in B \mid Y = y] = \int_B f_{X|Y}(x, y) dx,$$

gdzie

$$f_{X|Y}(x, y) = \frac{f(x, y)}{\int f(t, y) dt}.$$

3. Rozważmy  $\mathbb{P}[E_1 \in \bullet \mid E_1 + E_2 = y]$  dla  $E_1, E_2$  niezależnych o rozkładzie  $\text{Exp}(1)$ .

Przyłożymy do tego przypadku wzór z przykładu 2. Wektor losowy  $(E_1, E_1 + E_2)$  ma rozkład losowy o łącznej gęstości  $f(x, y) = e^{-x}e^{-(y-x)}\mathbb{1}_{y \geq x \geq 0}$ .

$$\int f(s, y) ds = \int_0^y e^{-y} ds = y,$$

czyli

$$f_{E_1|E_1+E_2}(x, y) = \frac{1}{y}\mathbb{1}_{y \geq x \geq 0}$$

co daje rozkład jednostajny:

$$\mathbb{P}[E_1 \in \bullet \mid E_1 + E_2 = y] = U[0, y](\bullet)$$

Można zadać sobie pytanie, czy

$$\mathbb{P}[A \mid Y = y] = \mathbb{E}[\mathbb{1}_A \mid Y = y]$$

zawsze zadaje miarę probabilistyczną? Okazuje się, że tak faktycznie jest.

### Definicja 3.2 : regularny rozkład warunkowy.

Niech  $X$  będzie zmienną losową, a  $\mathcal{G} \subseteq \mathcal{F}$  będzie  $\sigma$ -ciałem. Funkcja

$$\kappa_{X, \mathcal{G}} : \Omega \times \text{Bor}(\mathbb{R}) \rightarrow [0, 1]$$

nazywa się **regularnym rozkładem warunkowym** [rrw]  $X$  pod warunkiem  $\mathcal{G}$ , jeżeli

(R1) Dla każdego  $B \in \text{Bor}(\mathbb{R})$  zmienna losowa  $\kappa_{X, \mathcal{G}}(\bullet, B)$  jest  $\mathcal{G}$ -mierzalna

(R2) Dla każdej  $\omega \in \Omega$   $\kappa_{X, \mathcal{G}}(\omega, \bullet)$  jest rozkładem prawdopodobieństwa na prostej.

(R3) Dla każdego  $B \in \text{Bor}(\mathbb{R})$  i  $\mathbb{P}$ -p.w.  $\omega \in \Omega$

$$\mathbb{P}[X \in B \mid \mathcal{G}](\omega) = \kappa_{X,\mathcal{G}}(\omega, B)$$

### Twierdzenie 3.4 : rrw istnieje.

Dla każdego  $X$  i dla każdego  $\mathcal{G}$  istnieje rrw  $X$  pod warunkiem  $\mathcal{G}$

#### Dowód

W notatkach



### Fakt 3.5.

Dla funkcji  $f : \mathbb{R} \rightarrow \mathbb{R}$  i zmiennej losowej  $X$  takiej, że  $\mathbb{E}[|f(X)|] < \infty$  mamy

$$\mathbb{E}[f(X) \mid \mathcal{G}] = \int (x) \kappa_{X,\mathcal{G}}(\omega, dx)$$

Pisząc to mówię "weź  $f(x)$  i weź miarę  $\kappa$  w punkcie  $\omega$  i scałkuj  $\kappa$ ".

#### Dowód

Ćwiczenia.



Jeżeli  $G = \sigma(Z)$ , to pojęcie rrw troszkę się upraszcza (z naciskiem na trochę):

### Definicja 3.3.

Dla zmiennych losowych  $X, Y$  funkcję  $\kappa_{X,Y} : \mathbb{R} \times \text{Bor}(\mathbb{R}) \rightarrow [0, 1]$  nazywamy rrw  $X$  pod warunkiem  $Y$ , jeżeli:

(P1) Dla każdego  $B \in \text{Bor}(\mathbb{R})$  funkcja  $\kappa_{X,Y}(\bullet, B)$  jest borelowska

(P2) Dla każdego  $y \in \mathbb{R}$ ,  $\kappa_{X,Y}(y, \bullet)$  jest rozkładem prawdopodobieństwa na prostej.

(P3)  $\mathbb{P}[X \in B \mid Y = y] = \kappa_{X,Y}(y, B)$

## Wykład 30.10.23 : Martyngały

Mają coś wspólnego z jazdą konną (podobno).

### Przykład(y) 4.1

1. Rozważmy dowolną grę i dla uproszczenia niech polega ona na rzucaniu monetą, na której orzeł wypada z  $\mathbb{P} = p \in (0, 1)$ . Obstawiamy według zasady double or nothing:

🐟 jeśli wypada orzeł, to podwajamy nasz kapitał

🐟 jeżeli wypada reszka, to tracimy wszystko

Czy taka gra jest sprawiedliwa?

Rozważmy ciąg niezależnych zmiennych losowych  $\{\xi_k\}_{k \in \mathbb{N}}$  o tym samym rozkładzie

$$\mathbb{P}[\xi_k = 2] = 1 - \mathbb{P}[\xi_k = 0] = p.$$

Wówczas ciąg

$$X_n = \xi_n \cdot \xi_{n-1} \cdot \dots \cdot \xi_1 \cdot X_0$$

reprezentuje stan konta po  $n$ -tym rzucie, gdzie  $X_0$  to jakaś stała.

Rozważmy  $\sigma$ -ciało  $\mathcal{F}_n = \sigma(\xi_n, \dots, \xi_1, X_0)$  generowane przez pierwszych  $n$  rzutów i stan początkowy. Zadajmy sobie teraz pytanie, ile wynosi

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]?$$

Mamy zależność rekurencyjną  $X_{n+1} = \xi_{n+1}X_n$ , stąd możemy powiedzieć, że

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\xi_{n+1}X_n \mid \mathcal{F}_n]$$

samo  $X_n$  jest w  $\mathcal{F}_n$ , więc możemy je wyciągnąć przed  $\mathbb{E}$ . Dodatkowo,  $\xi_{n+1}$  jest niezależne od  $\mathcal{F}_n$ , więc

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n \mathbb{E}[\xi_{n+1} \mid \mathcal{F}_n] = X_n \mathbb{E}[\xi_{n+1}] = (2p) \cdot X_n$$

🐟 Jeżeli  $p > \frac{1}{2}$ , to wówczas

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$$

i wtedy taka gra jest korzystna, bo z coraz to kolejnym rzutem oczekiwania rosną.

🐟 Jeżeli  $p < \frac{1}{2}$ , to wówczas

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n$$

i gra jest korzystna, ale dla kasyna a nie gracza.

🐟 Jeżeli  $p = \frac{1}{2}$ , to wówczas

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$$

i w takim przypadku powiemy, że gra jest sprawiedliwa.

Ten ostatni, uczciwy przypadek to jest jeden ze sposobów, na które możemy myśleć o martyngałach.

### Definicja 4.1 : o martyngałach słów kilka.

🐟 Wstępującą rodzinę  $\sigma$ -ciat

$$\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}},$$



$\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ , nazywamy **filtracją**

🐟 Powiemy, że ciąg zmiennych losowych  $\{X_n\}_{n \in \mathbb{N}}$  jest  **$\mathbb{F}$ -adaptowalny**, jeżeli  $X_n \in \mathcal{F}_n$ .

🐟 Adaptowalny i całkowalny ciąg  $\{X_n\}$  nazywamy **nadmartyngatem**, jeśli

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n$$

🐟 Adaptowalny i całkowalny ciąg  $\{X_n\}$  nazywamy **podmartyngatem**, jeśli

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$$

🐟 Z kolei ciąg  $\{X_n\}$  jest **martyngatem**, jeśli jest jednocześnie nadmartyngatem i podmartyngatem, czyli zachodzi równość

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$$

## Przykład(y) 4.2

1. Niech  $\{\eta_k\}_{k \in \mathbb{N}}$  będą niezależne takie, że  $\mathbb{E}[\eta_k] = 0$  dla każdego  $k \in \mathbb{N}$ . Wówczas jako filtrację możemy rozważyć

$$\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$$

a jako nowy ciąg zmiennych losowych zdefiniujemy jako  $M_0 = 0$  i

$$M_n = \sum_{k=1}^n \eta_k.$$

Tak zdefiniowany ciąg  $\{M_n\}$  jest  $\mathbb{F} = \{\mathcal{F}_n\}$ -martyngatem:

$$\begin{aligned} \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[\eta_{n+1} + M_n \mid \mathcal{F}_n] = \\ &= \mathbb{E}[\eta_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[M_n \mid \mathcal{F}_n] = \\ &= \mathbb{E}[\eta_{n+1}] + M_n = 0 + M_n = M_n \end{aligned}$$

2. Dla dowolnej filtracji  $\mathbb{F} = \{\mathcal{F}_n\}$  i całkowalnej zmiennej losowej  $X$  rozważmy

$$M_n = \mathbb{E}[X \mid \mathcal{F}_n].$$

Wówczas

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}[X \mid \mathcal{F}_n] = M_n$$

### Uwaga 4.1.

Jeżeli  $\{X_n\}$  jest martyngatem, to

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$$

czyli mamy dwie zmienne losowe które są sobie równe, czyli

$$\mathbb{E}[X_{n+1}] = \mathbb{E}[\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]] = \mathbb{E}[X_n]$$

W szczególności, jeśli zastosujemy indukcję, to dostaniemy, że dla dowolnego  $n \in \mathbb{N}$

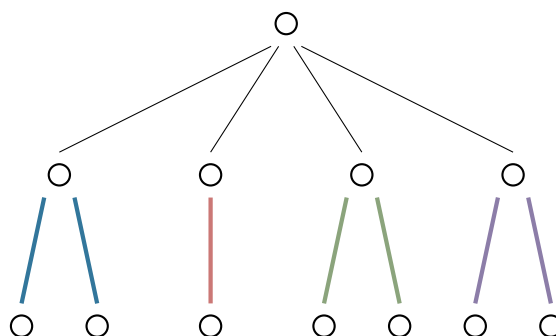
$$\mathbb{E}[X_n] = \mathbb{E}[X_0]$$

### Przykład(y) 4.3

#### 1. Proces Gaultona-Watsona

Rozważmy populację, w której osobniki rozmnażają się bezpłciowo, niezależnie od siebie. Można myśleć o tym jako o obserwacji populacji pantofelków z pomiarami w jakiś określonych odstępach czasu.

Myślimy o tym jako o drzewie, w którym liczba krawędzi z danego wierzchołka oznacza liczbę potomstwa, a ilość wierzchołków na danej głębokości oznacza ilość pantofelków po  $n$ -tym pokoleniu.



Niech  $\mu$  będzie dowolnym rozkładem prawdopodobieństwa na  $\mathbb{N} = \{0, 1, \dots\}$ . Rozważmy zmienne losowe losowo indeksowane parami liczb naturalnych  $\{Y_{n,k}\}_{n,k \in \mathbb{N}}$ . Kładziemy

$$Z_1 = 1$$

$$Z_{n+1} = \sum_{k=1}^{Z_n} Y_{n+1,k}$$

$Z_1$  to liczba pantofelków na samym początku,  $Z_2 = Y_{2,1}$  to liczba dzieci pierwszego pantofelka, z kolei

$$Z_3 = Y_{3,1} + Y_{3,2} + Y_{3,3} + Y_{3,4},$$

co odpowiada kolorom na rysunku. To znaczy, że  $Y_{n+1,k}$  to liczba potomstwa w generacji  $n+1$  zrodzona z  $k$ -tego pantofelka w generacji  $n$ .

Filtracją będzie dla nas ciąg o elementach  $\mathcal{F}_n = \sigma(Y_{j,k} : k \in \mathbb{N}, j \leq n)$ . Chcemy zapytać się o wartość oczekiwaną  $Z_{n+1}$

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = \mathbb{E}\left[\sum_{k=1}^{Z_n} Y_{n+1,k} | \mathcal{F}_n\right] = h(Z_n)$$

gdzie

$$h(z) = \mathbb{E}\left[\sum_{k=1}^z Y_{n+1,k}\right] = z \cdot \underbrace{\mathbb{E}[Y_{n,k}]}_m = m \cdot z,$$

bo wszystkie  $Y_{n,k}$  mają taką samą średnią. Oznacza to, że

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = m \cdot Z_n.$$

Jeżeli  $m < 1$ , to dostajemy w ten sposób nadmartyngał, jeśli  $m > 1$  to mamy podmartyngał, a w krytycznym przypadku  $m = 1$ , to  $\{Z_n\}$  jest martyngałem.

Jeśli pomnożymy

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = mZ_n$$

oboma stronami przez  $m^{-n-1}$ , to dostajemy

$$\mathbb{E}[m^{-n-1}Z_{n+1} | \mathcal{F}_n] = m^{-n}Z_n$$

i wtedy  $W_n = m^{-n}Z_n$  jest zawsze martyngałem, bo

$$\mathbb{E}[W_{n+1} | \mathcal{F}_n] = W_n$$

## 4.1 Transformata martyngałowa

Stan konta gracza wynosi  $X_n$  po  $n$ -tej sprawiedliwej grze. Przychodzi drugi gracz i obstawia on wyniki w grze tego pierwszego. Wypłata drugiego gracza wynosi  $B_n \cdot (X_n - X_{n-1})$ , tzn. za każdy przychód pierwszego gracza dostaje jakąś część tej wygranej.

Dla ciągu funkcji  $B_n \in \mathcal{F}_{n+1} = \sigma(X_0, \dots, X_{n-1})$ . Żeby było łatwiej, niech drugi gracz zaczyna z tym samym kapitałem co pierwszy. Stan konta drugiego gracza po  $n$ -tej grze wynosi

$$W_n = \sum_{k=1}^n B_k \cdot (X_k - X_{k-1}) + X_0.$$

Tak zdefiniowany ciąg  $\{Q_n\}$  jest również martyngałem, bo

$$\begin{aligned} \mathbb{E}[W_{n+1} | \mathcal{F}_n] &= \mathbb{E}[B_{n+1} \cdot (X_{n+1} - X_n) | \mathcal{F}_n] + \mathbb{E}[W_n | \mathcal{F}_n] = \\ &= B_{n+1} \mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] + W_n = \\ &= B_{n+1}(\mathbb{E}[X_{n+1} | \mathcal{F}_n] - X_n) + W_n = W_n \end{aligned}$$

bo  $X_n$  sam w sobie był martyngałem, więc  $X_n = \mathbb{E}[X_{n+1} | \mathcal{F}_n]$ .

### Definicja 4.2.

Niech  $\mathbb{F}$  będzie filtracją. Zmienną losową  $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$  nazywamy  **$\mathbb{F}$ -czasem zatrzymania**, jeżeli zdarzenie  $\{T = n\}$  jest mierzalne względem  $\mathcal{F}_n$  dla każdego  $n \in \mathbb{N}$ .

## Przykład(y) 4.4

1. Rzucamy 10-krotnie monetą. Zdefiniujmy zmienną losową

$$X_n = \begin{cases} 1 & \text{orzeł w } n\text{-tym} \\ 0 & \text{wpp} \end{cases}$$

Filtracją niech będzie  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ . Jeśli  $T$  będzie momentem wypadnięcia pierwszego orła, a  $S$  - wypadnięcia ostatniego orła, to  $T$  jest czasem zatrzymania, bo

$$\{T = n\} = \{X_1 = 0, X_2 = 0, \dots, X_n = 1\} \in \mathcal{F}_n$$

a  $S$  nim nie jest, bo wymaga informacji wybiegającej w przyszłość:

$$\{S = n\} = \{X_n = 1, X_{n+1} = 0, \dots\} \notin \mathcal{F}_n$$



2. Rozważmy  $\mathbb{F}$ -adaptowalny ciąg zmiennych losowych  $\{X_n\}$ . Dla  $B \in \text{Bor}(\mathbb{R})$  kładziemy

$$T(B) = \inf\{n : X_n \in B\}.$$

Tak zdefiniowane  $T$  jest czasem zatrzymania:

$$\{T = n\} = \{X_1 \notin B, \dots, X_{n-1} \notin B, X_n \in B\} \in \mathcal{F}_n$$

3. Jeżeli  $T = n_0$  dla pewnego  $n_0 \in \mathbb{N}$ , to taka stała funkcja nadal jest czasem zatrzymania, bo

$$\{T = n\} = \begin{cases} \emptyset & n \neq n_0 \\ \Omega & n = n_0 \end{cases}$$