Rachunek prawdopodobieństwa 2R

Kycia

Spis rozmaitości treściowalnych

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

$$\mathbb{P}\left[B \mid A\right] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}.$$

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:

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

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

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

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

Działanie jak wyżej daje pełną informacje o każdym zdarzeniu z ciała generowanego przez zdarzenia A_1 i A_2. Nazywamy je **rozbiciem** względem σ-ciała generowanego przez A_1 i A_2.

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

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

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

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

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

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

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

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

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

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

Pozostaje zapytać, co z tej zależności wynika?

Dla $F = \Omega$ dostajemy

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

Dygresja.
W tym miejscu kuszące byłoby rozpisanie $Y = h(Z)$ wprost z definicji, tzn. $h(Z) = \mathbb{E}\left[X \mid Z = Z\right]$, 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

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

Przykład(y) 1.1

1. Ze zbioru $\{1, 2, ..., 10\}$ losujemy w sposób jednostajny liczbę i oznaczamy ją jako N. W drugim losowaniu, również w sposób jednostajny, wybieramy liczbę ze zbioru $\{1, ..., N\}$ 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:

   $$h(n) = \mathbb{E}\left[M \mid N = n\right] = \sum_{1\le i\le n}\frac{i}{n} = \frac{1}{n}\cdot\frac{n(n+1)}{2} = \frac{n+1}{2}$$

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

   Stosując notację jak wyżej, mamy

   $$\begin{gathered} Z = N \\ X = M \end{gathered}$$

czyli podstawiając do wzoru:

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

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 σ-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) $(\forall\ G \in \mathcal{G})\ \mathbb{E}\left[X\mathbb{1}_G\right] = \mathbb{E}\left[Y\mathbb{1}_G\right]$

Nasuwają 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 $h(z) = \mathbb{E}\left[X \mid Z = z\right]$ jest wwo X względem $\mathcal{G}$.

Twierdzenie 1.1 : poprawność wwo.
Dla σ-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

$$\mathbb{E}\left[X \mid \mathcal{G}\right] = Y.$$

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 $\mathbb{E}\left[X \mid \mathcal{G}\right] = \mathbb{E}[X] = Y$.

   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 $G \in \mathcal{G}$:
$$\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[X]\,\mathbb{E}[\mathbb{1}_G] = \mathbb{E}[\mathbb{E}[X]\,\mathbb{1}_G] = \mathbb{E}[Y\mathbb{1}_G].$$

2. Rozważmy pokrycie Ω 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 σ-ciałem rozpinanym przez to pokrycie. Wówczas
$$\mathbb{E}\left[X \mid \mathcal{G}\right] = \sum_{i\in\mathbb{N}} \mathbb{1}_{A_i}\mathbb{E}\left[X \mid A_i\right]$$
Spełnianie pierwszego warunku jest oczywiste, bo mamy doczynienia z funkcją prostą. Warunek (W2) wystarczy sprawdzić dla atomów, czyli $G = A_i$, bo wszystkie zmienne losowe $\mathcal{G}$-mierzalne są stałe na A_i.
$$\begin{aligned}\mathbb{E}\left[\left[\sum \mathbb{1}_{A_i}\mathbb{E}\left[X \mid A_i\right]\right]\mathbb{1}_{A_j}\right] &= \mathbb{E}\left[\mathbb{E}\left[X \mid A_j\right]\mathbb{1}_{A_j}\right] = \\ &= \mathbb{E}\left[\mathbb{1}_{A_j}\frac{\mathbb{E}\left[X\mathbb{1}_{A_j}\right]}{\mathbb{P}\left[A_j\right]}\right] = \\ &= \mathbb{E}\left[\mathbb{1}_{A_j}\right]\frac{\mathbb{E}\left[X\mathbb{1}_{A_j}\right]}{\mathbb{P}\left[A_j\right]} = \mathbb{E}\left[X\mathbb{1}_{A_j}\right],\end{aligned}$$
gdyż $\mathbb{E}\left[\mathbb{1}_{A_j}\right] = \mathbb{P}\left[A_j\right]$.

3. Jeśli w przykładzie wyżej weźmiemy $A_1 = A$, $A_2 = A^c$ i $A_3 = A_4 = \ldots = \emptyset$ oraz $\mathcal{G} = \sigma(A)$, to dostajemy to samo co na samym początku tego wykładu:
$$\mathbb{E}\left[X \mid \mathcal{G}\right] - \mathbb{1}_A\mathbb{E}\left[X \mid A\right] + \mathbb{1}_{A^c}\mathbb{E}\left[X \mid A^c\right].$$

1.3 Prawdopodobieństwo warunkowe

Definicja 1.2 : prawdopodobieństwo warunkowe.
Dla σ-ciała $\mathcal{G} \subseteq \mathcal{F}$ definiujemy **prawdopodobieństwo warunkowe** pod warunkiem $\mathcal{G}$ jako
$$\mathbb{P}\left[A \mid \mathcal{G}\right] = \mathbb{E}\left[\mathbb{1}_A \mid \mathcal{G}\right]$$

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

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

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

Przykład(y) 1.4

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

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

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

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

Dla pewności, przerachujemy 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 $C \in \mathrm{Bor}(\mathbb{R})$ takie, że G jest postaci $G = \{E_1 \in C\}$. Policzymy $\mathbb{E}\left[\mathbb{P}\left[\{E_1 + E_2 > t\} \mid \sigma(E_1)\right]\right]$ gdyż jak wyżej zauważyliśmy, $\mathbb{P}\left[A \mid \mathcal{G}\right]$ jest zmienną losową. Mamy więc

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

Równość $\star$ wynika z uwagi pod definicją prawdopodobieństwa warunkowego. Całka $\star\star$ 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 σ-ciało $\mathcal{G} \subseteq \mathcal{F}$, to zachodzi $\mathbb{E}\left[\left|\mathbb{E}\left[X \mid \mathcal{G}\right]\right|\right] < \infty$.

Dowód
Rozważmy zbiór

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

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

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

bo X jest całkowalna.

Analogicznie postępujemy dla zbioru A^c:

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

Zauważmy, że

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

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

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

Lemat 2.2 : jedyność p.w..
Niech $\mathcal{G} \subseteq F$ będzie σ-ciałem. Jeśli Y i Y′ są obie wersjami $\mathbb{E}\left[X \mid \mathcal{G}\right]$, 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.

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

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

Dostajemy więc, że

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

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

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

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

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

Twierdzenie 2.3 : o istnieniu WWO.
Niech $\mathcal{G} \subseteq \mathcal{F}$ będzie σ-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 μ i ν będą σ-miarami na przestrzeni $(\Omega, \mathcal{G})$ takimi, że ν jest *absolutnie ciągła* względem μ $[\nu \ll \mu]$, tzn $\mu(A) = 0 \Rightarrow \nu(A) = 0$. Wówczas istnieje $\mathcal{G}$-mierzalna funkcja $f : \Omega \to \mathbb{R}$ 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.

Załóżmy, że $X \geq 0$ p.w. Wtedy możemy rozważyć miary $\mu = \mathbb{P} \restriction \mathcal{G}$ oraz $\nu(A) = \mathbb{E}[X\mathbb{1}_A]$. Od razu widać, że w takim ułożeniu $\nu \ll \mu$, więc na mocy twierdzenia Radona-Nikodyma istnieje f $\mathcal{G}$-mierzalna taka, że

$$\mathbb{E}\left[f\mathbb{1}_A\right] = \int_A f(\omega)\mu(d\omega) = \nu(A) - \mathbb{E}\left[X\mathbb{1}_A\right].$$

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 $\mathbb{E}\left[X^+ \mid \mathcal{G}\right]$ oraz $\mathbb{E}\left[X^- \mid \mathcal{G}\right]$. Wystarczy zauważyć, że dzięki liniowości $\mathbb{E}$ możemy w prosty sposób pokazać

$$\mathbb{E}\left[X \mid \mathcal{G}\right] = \mathbb{E}\left[X^+ \mid \mathcal{G}\right] - \mathbb{E}\left[X^- \mid \mathcal{G}\right]$$

2.2 Własności wwo

Twierdzenie 2.4 : o arytmetyce wwo.
Niech $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$ będą σ-ciałami, a X, X_1, X_2 całkowalnymi zmiennymi losowymi

1. $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right]\right] = \mathbb{E}[X]$
2. Jeśli $X \geq 0$, to również $\mathbb{E}\left[X \mid \mathcal{G}\right] \geq 0$
3. $\mathbb{E}\left[aX_1 + bX_2 \mid \mathcal{G}\right] = a\mathbb{E}\left[X_1 \mid \mathcal{G}\right] + b\mathbb{E}\left[X_2 \mid \mathcal{G}\right]$
4. $\left|\mathbb{E}\left[X \mid \mathcal{G}\right]\right| \leq \mathbb{E}\left[|X| \mid \mathcal{G}\right]$
5. Jeśli $\mathcal{G}_1 \subseteq \mathcal{G}_2$, to wówczas
$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}_1\right] \mid \mathcal{G}_2\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}_2\right] \mid \mathcal{G}_1\right] = \mathbb{E}\left[X \mid \mathcal{G}_1\right]$$
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 $\mathbb{E}\left[XY \mid \mathcal{G}\right] = Y\mathbb{E}\left[X \mid \mathcal{G}\right]$, czyli Y możemy traktować jako stałą.

Dowód

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

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] \mathbb{1}_\Omega\right] \overset{(W2)}{=} \mathbb{E}\left[X\mathbb{1}_\Omega\right] = \mathbb{E}\left[X\right]$$

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

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

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

3. Można to zrobić na dwa sposoby: licząc wszystko pokolei, albo można sprawdzić, czy $Y = a\mathbb{E}\left[X_1 \mid \mathcal{G}\right] + b\mathbb{E}\left[X_2 \mid \mathcal{G}\right]$ spełnia warunki wwo tej samej zmiennej co $\mathbb{E}\left[aX_1 + bX_2 \mid \mathcal{G}\right]$. 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 $A \in \mathcal{G}$.

$$\begin{aligned}\mathbb{E}\left[Y\mathbb{1}_A\right] &\overset{\star}{=} a\mathbb{E}\left[\mathbb{E}\left[X_1 \mid \mathcal{G}\right] \mathbb{1}_A\right] + b\mathbb{E}\left[\mathbb{E}\left[X_2 \mid \mathcal{G}\right] \mathbb{1}_A\right] = \\ &= a\mathbb{E}\left[X_1\mathbb{1}_A\right] + b\mathbb{E}\left[X_2\mathbb{1}_A\right] = \\ &= \mathbb{E}\left[(aX_1 + bX_2)\mathbb{1}_A\right] = \mathbb{E}\left[\mathbb{E}\left[aX_1 + bX_2 \mid \mathcal{G}\right] \mathbb{1}_A\right]\end{aligned}$$

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

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

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

   Po złożeniu tych dwóch nierówności:

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

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

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

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

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

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

   Pozostaje nam sprawdzić czym jest $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}_1\right] \mid \mathcal{G}_2\right]$. Roboczo nazwiemy $Y = \mathbb{E}\left[X \mid \mathcal{G}_1\right]$. 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) $\mathbb{E}\left[Y \mid \mathcal{G}_2\right] = Y$. Pisząc bez używania litery Y dostajemy

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}_1\right] \mid \mathcal{G}_2\right] = \mathbb{E}\left[X \mid \mathcal{G}_1\right]$$

6. Ćwiczenie, a poniżej moja próba.

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

Zacznijmy od $Y = \sum a_i \mathbb{1}_{A_i}$ 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

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

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

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

$$s_1 \leq s_2 \leq \ldots \leq s_n \leq \ldots \qquad \lim s_i = Y$$

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

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[XY\right] \mathbb{1}_A\right] = \mathbb{E}\left[XY\mathbb{1}_A\right] = \mathbb{E}\left[X \lim s_i \mathbb{1}_A\right] &\overset{\star}{=} \lim \mathbb{E}\left[X s_i \mathbb{1}_A\right] = \\ &\overset{\star\star}{=} \lim \mathbb{E}\left[s_i \mathbb{E}\left[X \mid \mathcal{G}\right] \mathbb{1}_A\right] = \mathbb{E}\left[\lim s_i \mathbb{E}\left[X \mid \mathcal{G}\right] \mathbb{1}_A\right] = \\ &= \mathbb{E}\left[Y\mathbb{E}\left[X \mid \mathcal{G}\right] \mathbb{1}_A\right]\end{aligned}$$

$\star$ można zrobić na mocy twierdzenia o monotoniczności ciągu s_i dla zwykłej $\mathbb{E}$, natomiast $\star\star$ 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:

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

Twierdzenie 2.5 : o zbieżności i ciągłości.
Niech $\mathcal{G} \subseteq \mathcal{F}$ będzie σ-ciałem, a $X, X_1, X_2, \ldots$ będzie ciągiem całkowalnych zmiennych losowych. Wówczas

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

Dowód

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

Niech $Y = \lim \mathbb{E}\left[X_n \mid \mathcal{G}\right]$. Wystarczy, że pokażemy $Y = \mathbb{E}\left[X \mid \mathcal{G}\right]$ 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 $A \in \mathcal{G}$

$$\mathbb{E}\left[Y\mathbb{1}_A\right] = \mathbb{E}\left[\lim \mathbb{E}\left[X_n \mid \mathcal{G}\right] \mathbb{1}_A\right] \stackrel{\star}{=} \lim \mathbb{E}\left[\mathbb{E}\left[X_n \mid \mathcal{G}\right] \mathbb{1}_A\right] =$$
$$= \lim \mathbb{E}\left[X_n\mathbb{1}_A\right] = \mathbb{E}\left[\lim X_n\mathbb{1}_A\right] = \mathbb{E}\left[X\mathbb{1}_A\right]$$

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

2. Zaczniemy od dwóch obserwacji:

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

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

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

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

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

Stosując obserwację w przejściach $\star$, obserwację w przejsciu $\star\star$ oraz ptk 1. ($\inf_{n>k} X_n \leq \inf_{n>k+1} X_n$) w $\star\star\star$, dostajemy

$$\mathbb{E}\left[\liminf_n X_n \mid \mathcal{G}\right] \stackrel{\star}{=} \mathbb{E}\left[\lim_k \inf_{n>k} X_n \mid \mathcal{G}\right] \stackrel{\star\star\star}{=} \lim_k \mathbb{E}\left[\inf_{n>k} X_n \mid \mathcal{G}\right] \leq$$
$$\stackrel{\star\star}{\leq} \lim_k \inf_{n>k} \mathbb{E}\left[X_n \mid \mathcal{G}\right] \stackrel{\star}{=} \liminf_n \mathbb{E}\left[X_n \mid \mathcal{G}\right]$$

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

$$\mathbb{E}\left[Y + \liminf X_n \mid \mathcal{G}\right] = \mathbb{E}\left[\liminf(X_n + Y) \mid \mathcal{G}\right] \stackrel{3.}{\leq} \liminf \mathbb{E}\left[Y + X_n \mid \mathcal{G}\right] =$$
$$= \mathbb{E}\left[Y\right] + \liminf \mathbb{E}\left[X_n \mid \mathcal{G}\right]$$

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

Postępując analogicznie dla $X_n" = Y - X_n$ (które dalej jest ≥ 0) dostaniemy $\mathbb{E}\left[\limsup X_n \mid \mathcal{G}\right] \geq \limsup \mathbb{E}\left[X_n \mid \mathcal{G}\right]$:

$$\mathbb{E}\left[Y - \limsup X_n \mid \mathcal{G}\right] = \mathbb{E}\left[\limsup(Y - X_n) \mid \mathcal{G}\right] =$$
$$= -\mathbb{E}\left[\liminf(X_n - Y) \mid \mathcal{G}\right] \stackrel{3.}{\geq} -\liminf \mathbb{E}\left[X_n - Y \mid \mathcal{G}\right] =$$
$$= \limsup \mathbb{E}\left[Y - X_n \mid \mathcal{G}\right]$$

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

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

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

$$\liminf \mathbb{E}\left[X_n \mid \mathcal{G}\right] = \mathbb{E}\left[X \mid \mathcal{G}\right] = \limsup \mathbb{E}\left[X_n \mid \mathcal{G}\right]$$

i ponieważ lim inf = lim sup = lim to mamy

$$\lim \mathbb{E}\left[X_n \mid \mathcal{G}\right] = \mathbb{E}\left[X \mid \mathcal{G}\right].$$

Twierdzenie 2.6.
Niech $\mathcal{G} \subseteq \mathcal{F}$ będzie σ-ciałem. Załóżmy, że

- X jest $\mathcal{G}$-mierzalna
- Y jest niezależna od $\mathcal{G}$
- funkcja $\psi : \mathbb{R}^2 \to \mathbb{R}$ jest mierzalna taka, że

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

Wówczas

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

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

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

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

Wykład 23.10.23 : Interpretacje geometryczne WWO

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

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

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

> **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.

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

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

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

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

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

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

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-Schwartza.**
> Dla zmiennych X, Y takich, że $\mathbb{E}\left[X^2\right], \mathbb{E}\left[Y^2\right] < \infty$ i σ-ciała $\mathcal{G} \subseteq \mathcal{F}$ zachodzi
>
> $$\mathbb{E}\left[|XY| \mid \mathcal{G}\right] \le \mathbb{E}\left[X^2 \mid \mathcal{G}\right]^{\frac{1}{2}} \mathbb{E}\left[Y^2 \mid \mathcal{G}\right]^{\frac{1}{2}}$$

Dowód
Zadanie

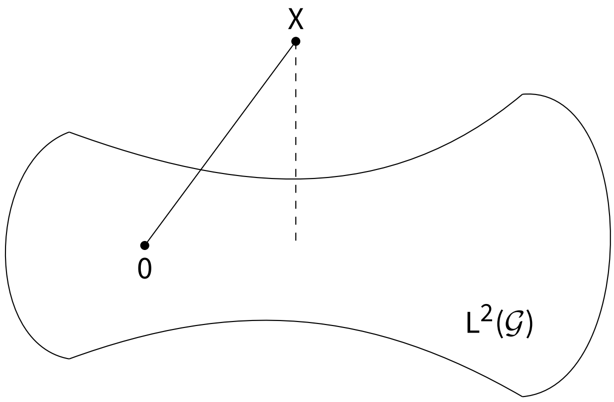
Z tego nierówności w fakcie 3.2 wynika, że dla Y = 1 mamy

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

Twierdzenie 3.3 : wwo jest rzutem ortogonalnym na $L^2(\mathcal{G})$.
Niech X będzie zmienną losową taką, że $\mathbb{E}\left[X^2\right] < \infty$, a $\mathcal{G} \subseteq \mathcal{F}$ jest σ-ciałem. Wówczas

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

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



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

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

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

Zauważmy, że

$$\mathbb{E}\left[Y'X \mid \mathcal{G}\right] = Y'\mathbb{E}\left[X \mid \mathcal{G}\right]$$
$$\mathbb{E}\left[\mathbb{E}\left[Y'X \mid \mathcal{G}\right]\right] = \mathbb{E}\left[Y'X\right] = \mathbb{E}\left[Y'\mathbb{E}\left[X \mid \mathcal{G}\right]\right]$$

Przykład(y) 3.1

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

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

   Dowolny $G \in \mathcal{G}$ ma postać $G = A \cup B$, gdzie $A \in \mathrm{Bor}([0, t])$ i $B \in \{(t, 1], \emptyset\}$. W takim razie, jeśli $Y \in \mathcal{G}$, to Y jest stała na $(t, 1]$. Czyli żeby $Y = \mathbb{E}\left[X \mid \mathcal{G}\right]$ 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 losowe X napis

$$\mathbb{E}\left[X \mid Y\right] := \mathbb{E}\left[X \mid \sigma(Y)\right]$$

będzie wwo X względem σ-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:

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

Definicja 3.1 : wwo X pod warunkiem $\{Y = y\}$.
Po pierwsze zauważamy, że istnieje funkcja $h : \mathbb{R} \to \mathbb{R}$ taka, że $\mathbb{E}\left[X \mid Y\right] = h(Y)$. ($\star$)

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

$$\mathbb{E}\left[X \mid Y = y\right] = h(y)$$

gdzie h spełnia warunek ($\star$).

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

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

Przykład(y) 3.2

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

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

   Wysławiając tę wartość w terminach całego X:

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

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

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

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

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

   gdzie

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

3. Rozważmy $\mathbb{P}\left[E_1\in\bullet \mid E_1+E_2=y\right]$ dla E_1, E_2 niezależnych o rozkładzie 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:

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

Można zadać sobie pytanie, czy

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

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 σ-ciałem. Funkcja

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

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

(R1) Dla każdego $B\in\mathrm{Bor}(\mathbb{R})$ 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 $B \in Bor(\mathbb{R})$ i $\mathbb{P}$-p.w. $\omega \in \Omega$

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

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

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

Pisząc to mówię "weź f(x) i weź miarę κ w punkcie ω i scałkuj κ".

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 $\kappa_{X,Y} : \mathbb{R} \times Bor(\mathbb{R}) \to [0, 1]$ nazywamy rrw X pod warunkiem Y, jeżeli:

(P1) Dla każdego $B \in Bor(\mathbb{R})$ 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) $\mathbb{P}\left[X \in B \mid Y = y\right] = \kappa_{X,Y}(y, B)$