

Rachunek prawdopodobieństwa 2R

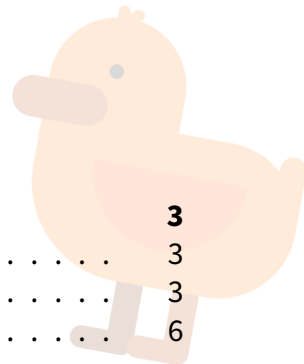
Kycia



Spis różnorodności treściwalnych

Warunkowa wartość oczekiwana

1.1	Prawdopodobieństwo warunkowe	3
1.2	Konstrukcja warunkowej wartości oczekiwanej	3
1.3	Prawdopodobieństwo warunkowe	6



1 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 zjściu 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 **rozbićciem** względem σ -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. Zdefiniujemy 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 $*$ 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:

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

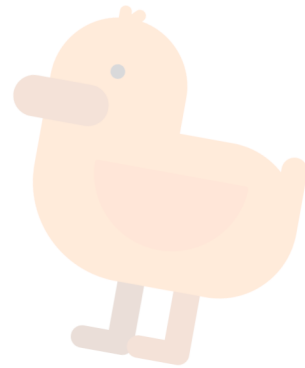
czyli $h(N) = \frac{N+1}{2}$.

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

Nasuują 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 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

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

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 Ω rozłącznymi zbiorami $\{A_i\}_{i \in \mathbb{N}}$, 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

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

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

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

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 σ -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)

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

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

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

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

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

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 $e^{-(t-x)}$.