RACHUNEK PRAWDOPODOBIEŃSTWA

Piotr Dyszewski

2020

Przejdźmy od słów do czynów.
Chciałem powiedzieć kilka słów.

Rejs

Spis treści

1

Prawdopodobieństwo dyskretne

W latach siedemdziesiątych ubiegłego wieku na łamach *American Statistician* [Ame75] ukazała się zagadka oparta na amerykańskim teleturnieju *Let's Make a Deal*. W założeniach problemu prowadzący Monty Hall stawia przed graczem trzy bramki. Za jedną z bramek znajduje się sportowy samochód a za pozostałymi dwoma koza. Gracz wybiera jedną z trzech bramek po czym prowadzący odsłania jedną z pozostałych bramek z kozą. Następnie Monty Hall składa graczowi propozycję zmiany wybranej przez siebie bramki. Pytanie brzmi: *Czy gracz powinien przystać na tę propozycję?* Autorka artykułu Marilyn vos Savant opisała w swoje rubryce rozwiązanie, które sugerowało, że gracz zawsze powinien przystać na propozycję prowadzącego i zmienić bramkę. Vos Savant otrzymała w odpowiedzi tysiące listów, niektóre wręcz oprysкliwe, w których ogromna większość optowała za tym, że gracz powinien pozostać przy swoim pierwszym wyborze. Większość kłopotów ze zadania brała się z barku określenia przestrzeni probabilistycznej dla naszego problemu.

Ten i wiele innych przykładów pokazują, jak ważne jest prawidłowe zdefiniowanie naszego problemu. W tym przypadku wymaga to określenia przestrzeni probabilistycznej. Na pierwszy rzut oka taka procedura może wydawać się zbędna. Przykładowo, czytelnik z dobrze rozwiniętą intuicją nie potrzebuje określania abstrakcyjnych bytów aby wyznaczyć prawdopodobieństwa wypadnięcia karety w grze w pokera. Jednak gdy problemy stają się coraz bardziej złożone, poprawnie zdefiniowana przestrzeń probabilistyczna staje się coraz bardziej przydatna. Koniec końców nie jest możliwe uprawianie teorii prawdopodobieństwa na uniwersyteckim poziomie bez wprowadzenia odpowiedniego formalizmu. Pierwszy dział poświęcimy właśnie na wprowadzenie odpowiedniego języka, który następnie zastosujemy do rozwiązania problemy Montego Halla.

1.1 Opis doświadczenia losowego

Za każdym razem kiedy rzucamy kością czy losujemy kartę z talii wykonujemy pewnego rodzaju eksperyment losowy. W pierwszej części wykładu zajmiemy się ich badaniem. Zaczniemy od sposobu matematycznej reprezentacji wyniku naszego eksperymentu. Każdy możliwy wynik eksperymentu nazywać będziemy zdarzeniem elementarnym. Przykładowo dla rzutu kością sześcienną zdarzeniem elementarnym będzie liczba wyrzuconych oczek, czyli dowolna z liczb 1,2, . . . 6. Zbiór wszystkich zdarzeń elementarnych oznaczać będziemy przez Ω i będziemy nazywać *przestrzenią zdarzeń elementarnych*. **Zakładać będziemy, że Ω jest zbiorem co najwyżej przeliczalnym**, czyli skończonym bądź nieskończonym o przeliczalnej liczbie elementów. Wówczas elementy Ω można wpisać w następujący sposób

$$\Omega = \{\omega_k : k \in \mathbb{N}\}, \qquad \mathbb{N} = \{1, 2, 3, \ldots\}$$

jeżeli Ω jest nieskończona oraz

$$\Omega = \{\omega_k : k \in [N]\}, \qquad [N] = \{1, 2, \ldots, N\}, \quad N \in \mathbb{N}$$

jeśli Ω jest skończona. Wybór przestrzeni Ω zależy zawsze od rodzaju eksperymentu losowego, który chcemy opisać. Na tym etapie przestrzeń zdarzeń elementarnych jest zbiorem wszystkich możliwych wyników naszego eksperymentu i nie mówi jeszcze nic o prawdopodobieństwach tych wyników.

Przykład 1.1
Powiedzmy, że nasz eksperyment losowy polega na rzucie parą kości sześciennych. Wynikiem naszego eksperymentu jest para elementów zbioru $D = \{$⚀,⚁$, \ldots,$⚅$\}$. Istotne jest aby rozróżnić wyniki otrzymane na obu kościach, wszak kości nie muszą być symetryczne. Zbiorem zdarzeń elementarnych jest

$$\Omega = D^2 = \{\text{⚀⚀}, \text{⚀⚁}, \text{⚀⚂}, \ldots, \text{⚅⚅}\}.$$

W tym przypadku przestrzeń zdarzeń elementarnych jest skończona i składa się z $|\Omega| = 36$ elementów.

Przykład 1.2
Tym razem rzucamy monetą do momentu otrzymania orła. Przestrzeń zdarzeń elementarnych Ω składa się ze skończonych ciągów $\{O, R\}$ kończących się znakiem O i jednego nieskończonego ciągu R. Dokładniej

$$\Omega = \{R\}^{\mathbb{N}} \cup \bigcup_{k=0}^{\infty} \{R\}^k \times \{O\} = \{RRRR\ldots\} \cup \{O, RO, RRO, RRRO, \ldots\}.$$

Tutaj niekończony ciąg $RRR\ldots$ opisuje scenariusz, w którym nigdy nie wyrzucimy orła.

W większości eksperymentów losowych nie interesują nas poszczególne jego wyniki, lecz pewien ich wspólny aspekt. Dla przykładu w rzucie kością sześcienną istotna dla nas może być tylko parzystość uzyskanego wyniku. Innymi słowy bardziej od poszczególnych zdarzeń elementarnych interesować nas będę ich zbiory.

Definicja 1.3
Niech Ω będzie co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. *Zdarzeniem* nazywać będziemy dowolny podzbiór $A \subseteq \Omega$.

Duch wprowadzanego formalizmu wymaga od nas ścisłego definiowania zdarzeń. Należy pamiętać jednak, że formalizm nie jest celem samym w sobie. Często będziemy pozwalać sobie na definiowanie zdarzeń w sposób opisowy o ile nie będzie prowadziło to do nieścisłości.

Przykład 1.4
Wróćmy do rzutu parą kości, gdzie przestrzenią zdarzeń elementarnych jest $\Omega = D^2$. Wówczas zdarzeniem jest

$$\begin{aligned} A =& \{\text{suma oczek na obu kościach jest równa 7}\} \\ =& \{\text{⚀⚅}, \text{⚁⚄}, \ldots, \text{⚅⚀}\} \subseteq \Omega = D^2. \end{aligned}$$

Natomiast zbiór

$$\{\text{⚀⚀}, \text{⚁⚁}, \ldots, \text{⚅⚅}\}$$

jest zdarzeniem $\{$wypadł dublet$\}$.

Przykład 1.5
Alex układa swoją dwudziestosześciotomową encyklopedię na półce. Eksperyment polega na umieszczeniu poszczególnych tomów encyklopedii w losowej kolejności. Wtedy zbiór zdarzeń elementarnych możemy utożsamić ze zbiorem permutacji. Dla $n \in \mathbb{N}$ definiujemy

$$S_n = \{\pi\colon [n] \to [n] \;:\; \pi \text{ jest } 1-1 \text{ i ''na''}\}.$$

Każdy element $\omega \in S_n$ możemy utożsamić z macierzą

$$\omega = \begin{pmatrix} \omega(1) & \omega(2) & \omega(3) & \ldots & \omega(n) \\ 1 & 2 & 3 & \ldots & n \end{pmatrix},$$

gdzie $\omega(1), \omega(2), \ldots, \omega(n)$ są parami różnymi liczbami ze zbioru $[n]$. Wówczas w naszym przykładzie $\Omega = S_{26}$. Przykładowo ułożenie tomów w kolejności alfabetycznej odpowiada permutacji identycznościowej $\mathrm{id} \in \Omega$ czyli takiej, że $\mathrm{id}(i) = i$ dla $i \in [26]$. Innymi słowy

$$\mathrm{id} = \begin{pmatrix} 1\,2\,3 \ldots n \\ 1\,2\,3 \ldots n \end{pmatrix}.$$

Przykład 1.6
Andrzej układa sagę o Wiedźminie na półce. Wówczas tomy możemy ponumerować liczbami $1, 2, 3, 4, 5$ według daty wydania. Zbiorem zdarzeń elementarnych jest zatem $\Omega = S_5$. Rozważmy zdarzenie A polegające na tym, że pierwsze dwa tomy ułożone są chronologicznie. Wówczas A składa się z permutacji o reprezentacji

$$\begin{pmatrix} 1\,2 * * * \\ 1\,2\,3\,4\,5 \end{pmatrix}.$$

Dokładniej

$$A = \left\{ \begin{pmatrix} 1\,2\,3\,4\,5 \\ 1\,2\,3\,4\,5 \end{pmatrix}, \begin{pmatrix} 1\,2\,3\,5\,4 \\ 1\,2\,3\,4\,5 \end{pmatrix}, \begin{pmatrix} 1\,2\,3\,4\,5 \\ 1\,2\,4\,3\,5 \end{pmatrix}, \begin{pmatrix} 1\,2\,4\,5\,3 \\ 1\,2\,3\,4\,5 \end{pmatrix}, \begin{pmatrix} 1\,2\,5\,3\,4 \\ 1\,2\,3\,4\,5 \end{pmatrix}, \begin{pmatrix} 1\,2\,5\,4\,3 \\ 1\,2\,3\,4\,5 \end{pmatrix} \right\}.$$

Wprowadzone przez nas **Definicja 1.3** zdarzenia jest niezwykle pojemna. W skutek czego jeżeli $A_1, A_2 \dots$ jest ciągiem zdarzeń, czyli dla każdego $k \in \mathbb{N}$, A_k jest zdarzeniem, to zdarzeniem jest również $\bigcup_{k=1}^{\infty} A_k$ oraz $\bigcap_{k=1}^{\infty} A_k$. Dodatkowo, jeżeli A jest zdarzeniem, to jest nie również *zdarzenie przeciwne do zdarzenia A*, zadane wzorem

$$A' = \Omega \setminus A.$$

Zdarzenie A' polega na tym, że nie zachodzi zdarzenie A. Jest jeszcze jedna relacja teoriomnogościowa, z której będziemy korzystać. Powiemy, że *zdarzenie B pociąga zdarzenie A*, jeżeli $B \subseteq A$.

Przykład 1.7
Powiedzmy, że rzucamy tylko jedną kością. Wówczas $\Omega = D = \{\text{⚀}, \text{⚁}, \dots, \text{⚅}\}$. Niech

$$A = \{\text{wypadła liczba parzysta}\} = \{\text{⚁}, \text{⚃}, \text{⚅}\}.$$

Zdarzenie przeciwne do zdarzenia A to

$$A' = \{\text{wypadła liczba nieparzysta}\} = \{\text{⚀}, \text{⚂}, \text{⚄}\}.$$

Z kolei zdarzenie

$$B = \{\text{wypadła szóstka}\} = \{\text{⚅}\}$$

pociąga zdarzenie A, czyli $B \subseteq A$.

1.2 Miara probabilistyczna

Jesteśmy przyzwyczajeni do intuicyjnego mylenia o prawdopodobieństwie. Dla przykładu w eksperymencie polegającym na rzucie dwoma monetami od razu napiszemy, że

$$\mathbb{P}\left[\text{pierwszej kości wypadnie 3 a na drugiej 5}\right] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

W pewnym sensie o liczbie $\mathbb{P}[A]$ myślimy jak o prawdopodobieństwie zdarzenia A opierając się na słownym opisie darzenia A. Nie ma w tym nic złego, lecz w wielu przypadkach wygodniej jest myśleć o zdarzeniu A jak o podzbiorze Ω a o liczbie $\mathbb{P}[A]$ jak o masie zbioru A.

Teoria rachunku prawdopodobieństwa rozpoczyna się od pierwotnego pojęcia przestrzeni losowej, która jest zbiorem Ω składającym się z wszystkich rzeczy, które mogą się zdarzyć w danym problemie, wraz z regułą przypisującą prawdopodobieństwo $\mathbb{P}[\omega]$ każdemu zdarzeniu elementarnemu $\omega \in \Omega$.

Definicja 1.8
Niech Ω będzie co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. *Funkcją prawdopodobieństwa* nazywamy odwzorowanie $p\colon \Omega \to [0,1]$ takie, że

$$\sum_{\omega\in\Omega} p(\omega) = 1.$$

Przykład 1.9
Dla $\Omega = \{R\}^{\mathbb{N}} \cup \bigcup_{k=0}^{\infty}\{R\}^k \times \{O\}$, cztyli jak w **Przykładzie 1.2**, funkcją prawdopodobieństwa jest $p(O) = \frac{1}{2}$, $p(OR) = \frac{1}{4}$ oraz ogólnie $p(\omega_k) = 2^{-k}$, gdzie ω_k składa się z $k-1$ reszek i orła. oraz $p(\omega_\infty) = 0$, gdzie ω_∞ składa się z samych reszek. Rzeczywiście, wystarczy sprawdzić

$$\sum_{\omega\in\Omega} p(\omega) = p(\omega_\infty) + \sum_{k=1}^{\infty} p(\omega_k) = 0 + 1.$$

To, że możemy określić miarę probabilistyczną jak funkcję zdarzeń elementarnych jest szczególną konsekwencją tego, że chwilowo ograniczamy się do co najwyżej przeliczalnych przestrzenie probabilistycznych Ω. W ogólnym kontekście miara probabilistyczna jest funkcją zbioru, czyli odwzorowaniem są podzbiory Ω, które umówiliśmy się nazywać zdarzeniami. Zaczniemy od konstruktywnej definicji dyskretnej miary probabilistycznej.

Definicja 1.10 (Konstruktywna)
Niech Ω będzie co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. *Miarą probabilistyczną* nazywamy odwzorowanie $\mathbb{P}\colon 2^{\Omega} \to [0,1]$ zadane wzorem

$$\mathbb{P}[A] = \sum_{\omega \in A} p(\omega),$$

dla zdarzenia A oraz funkcji prawdopodobieństwa $p(\cdot)$.

Przykład 1.11
Jeśli rzucamy kością sześcienną, to zbiorem zdarzeń elementarnych jest $D = \Omega = \{\epsdice{1}, \epsdice{2}, \ldots, \epsdice{6}\}$. Jeżeli kość jest dobrze wyważona, to

$$\mathbb{P}_0[\{\epsdice{1}\}] = \mathbb{P}_0[\{\epsdice{2}\}] = \ldots = \mathbb{P}_0[\{\epsdice{6}\}] = \frac{1}{6}.$$

Wówczas miara $\mathbb{P}_0$ opisuje rzut jedną wyważoną kością sześcienną. Możemy też zakładać, że jedna kość jest niewyważona, czyli

$$\begin{aligned}
&\mathbb{P}_1[\{\epsdice{2}\}] = \mathbb{P}_1[\{\epsdice{6}\}] = \frac{1}{4} \\
&\mathbb{P}_1[\{\epsdice{1}\}] = \mathbb{P}_1[\{\epsdice{3}\}] = \mathbb{P}_1[\{\epsdice{4}\}] = \mathbb{P}_1[\{\epsdice{5}\}] = \frac{1}{8}.
\end{aligned}$$

Wtedy $\mathbb{P}_1$ opisuje rzut niewyważoną kością. Miary $\mathbb{P}_0$ i $\mathbb{P}_1$ są różne i te same zdarzenia będą miały względem nich różne prawdopodobieństwa. Dla przykładu dla zdarzenia

$$A = \{\text{wypadła liczba parzysta}\} = \{\epsdice{2}, \epsdice{4}, \epsdice{6}\}$$

otrzymujemy $\mathbb{P}_0[A] = \frac{1}{3}$ oraz $\mathbb{P}_1[A] = \frac{5}{8}$.

Jak się okazuje miary probabilistyczne są addytywnymi funkcjami zbioru. Dodatkowo każda taka funkcja zbioru jest zadana przez funkcję prawdopodobieństwa.

Fakt 1.12
Niech Ω będzie co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. Załóżmy, że $\mathbb{P}$ jest miarą probabilistyczną na Ω zadaną przez funkcję prawdopodobieństwa p. Wówczas

(A1) $\mathbb{P}[\Omega] = 1$
(A2) $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$ dla rozłącznych zdarzeń A i B.
(A3) Jeżeli $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$ są zdarzeniami takimi, że $\bigcap_{j=1}^{\infty} B_j = \emptyset$, to

$$\lim_{n\to\infty} \mathbb{P}[B_n] = 0.$$

Dodatkowo jeżeli $\mathbb{P}\colon 2^\Omega \to [0,1]$ spełnia warunki (i), (ii) i (iii), to $\mathbb{P}$ jest miarą probabilistyczną zadaną przez funkcję prawdopodobieństwa $p(\omega) = \mathbb{P}[\{\omega\}]$.

Dowód. Niech p będzie funkcją prawdopodobieństwa zadającą $\mathbb{P}$. Wówczas

$$\mathbb{P}[\Omega] = \sum_{\omega\in\Omega} p(\omega) = 1$$

co dowodzi (A1). Aby uzasadnić (A2) napiszmy dla rozłącznych zbiorów A i B,

$$\mathbb{P}[A\cup B] = \sum_{\omega A\cup B} p(\omega) = \sum_{\omega\in A} p(\omega) + \sum_{\omega B} p(\omega) = \mathbb{P}[A] + \mathbb{P}[B].$$

Z rozłączności zbiorów A i B skorzystaliśmy w drugiej równości. Ten argument uzasadnia (A2). Jeżeli Ω jest zbiorem skończonym własność $(A3)$ jest oczywista, ponieważ wówczas istnieje j_0 takie, że $B_{j_0} = \emptyset$. Załóżmy, że Ω jest zbiorem nieskończonym przeliczalnym. Wówczas zdarzenia elementarne możemy ponumerować

$$\Omega = \{\omega_k \,:\, k\in\mathbb{N}\}.$$

Aby pokazać, że $\lim_{n\to\infty} \mathbb{P}[B_n] = 0$ ustalmy dowolny $\varepsilon > 0$. Skoro szereg $\sum_{k=1}^\infty p(\omega_k)$ jest zbieżny istnieje k_0 takie, że

$$\sum_{k=k_0}^{\infty} p(\omega_k) < \varepsilon.$$

Warunek $\bigcap_{j=1}^\infty B_j = \emptyset$ pociąga istnienie j_0 takiego, że $\{\omega_1, \omega_2, \ldots, \omega_{k_0-1}\} \cap B_{j_0} = \emptyset$, czyli $B_{j_0} \subseteq \{\omega_{k_0}, \omega_{k_0+1}, \ldots\}$. Stąd dla $j \ge j_0$,

$$\mathbb{P}[B_j] = \sum_{\omega\in B_j} p(\omega) \le \sum_{\omega\in B_{j_0}} p(\omega) \le \sum_{k=k_0}^{\infty} p(\omega_k) < \varepsilon.$$

To dowodzi (A3).

Załóżmy teraz, że $\mathbb{P}\colon 2^\Omega \to [0,1]$ spełnia własności (A1), (A2) i (A3). Określmy funkcję $p\colon \Omega \to [0,1]$ wzorem $p(\omega) = \mathbb{P}[\{\omega\}]$. Ustalmy dowolne zdarzenie skończone lub nieskończone $A = \{\omega_k \,:\, k \le N\}$, $N \in \mathbb{N}\cup\{\infty\}$. Korzystając n razy z własności (A2) otrzymujemy

$$
\begin{aligned}
\mathbb{P}[A] &= \mathbb{P}[\{\omega_1\}] + \mathbb{P}[\{\omega_2, \omega_3, \dots, \}] \\
&= \mathbb{P}[\{\omega_1\}] + \mathbb{P}[\{\omega_2\}] + \mathbb{P}[\{\omega_3, \omega_3, \dots, \}] \\
&= \sum_{k=1}^{n} \mathbb{P}[\{\omega_k\}] + \mathbb{P}[\{\omega_j \,:\, n+1 \leq j \leq N\}].
\end{aligned}
$$

Jeżeli $N < \infty$ to dla $n = N$ otrzymujemy

$$
\mathbb{P}[A] = \sum_{k=1}^{n} \mathbb{P}[\{\omega_k\}] = \sum_{\omega \in A} \mathbb{P}[\{\omega\}] = \sum_{\omega \in A} p(\omega).
$$

Natomiast w przypadku nieskończonego A, biorąc granicę $n \to \infty$

$$
\begin{aligned}
\mathbb{P}[A] = \lim_{n\to\infty} \mathbb{P}[A] &= \lim_{n\to\infty} \left(\sum_{k=1}^{n} \mathbb{P}[\{\omega_k\}] + \mathbb{P}[\{\omega_j \,:\, n+1 \leq j\}] \right) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}[\{\omega_k\}] + \lim_{n\to\infty} \mathbb{P}[\{\omega_j \,:\, n+1 \leq j\}] \\
&= \sum_{k=1}^{\infty} \mathbb{P}[\{\omega_k\}] + 0 = \sum_{k=1}^{\infty} p(\omega_k) = \sum_{\omega \in A} p(\omega),
\end{aligned}
$$

gdzie w ostatniej linijce skorzystaliśmy z własności (A3). Własność (A1) mówi, że p jest funkcją prawdopodobieństwa. To dowodzi, że $\mathbb{P}$ jest miarą probabilistyczną zadaną przez p. □

Powyższy fakt pozwala na drugą, bardziej abstrakcyjną definicję miary probabilistycznej.

Definicja 1.13 (Aksjomatyczna)
Niech Ω będzie co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. *Miarą probabilistyczną* na Ω nazywamy funkcją $\mathbb{P} \colon 2^{\Omega} \to [0,1]$ taką, że

(A1) $\mathbb{P}[\Omega] = 1$
(A2) $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$ dla rozłącznych zdarzeń A i B.
(A3) Jeżeli $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$ są zdarzeniami takimi, że $\bigcap_{j=1}^{\infty} B_j = \emptyset$, to

$$
\lim_{n\to\infty} \mathbb{P}[B_n] = 0.
$$

Fakt 1.12 mówi, że **Definicje 1.10** i **1.13** równoważne. W zależności od kontekstu będziemy odwoływali się do jednej z nich. W przypadku twierdzeń o ogólnej naturze wystarczy nam **Definicja 1.13**. W przypadku, gdy dowód ogólnego twierdzenia nie będzie możliwy, ponieważ przykładowo wymaga znajomości teorii miary, ograniczymy się do dyskretnych przestrzeni zdarzeń elementarnych i odwołamy się do **Definicji 1.10**.

Przykład 1.14
Rzucamy symetryczną kością dwudziestościenną. Wówczas przestrzenią zdarzeń elementarnych jest

$$\Omega = \{1, 2, \ldots, 20\}.$$

Jakie jest prawdopodobieństwo, że wylosowana liczba jest nieparzysta lub dzieli się przez 4? Dla liczby p oznaczmy zdarzenie

$$D_p = \{n \in [20] \mid n \text{ dzieli się przez } p\} \subseteq \Omega.$$

Szukamy $\mathbb{P}[N \cup D_4]$, gdzie $N = \{1, 3, \ldots, 19\}$ Zauważmy, że zdarzenia N i D_4 są rozłączne. Mamy $\mathbb{P}[N] = \frac{1}{2}$ i $\mathbb{P}[D_4] = \frac{1}{5}$. Stąd

$$\mathbb{P}[N \cup D_4] = \mathbb{P}[N] + \mathbb{P}[D_4] = \frac{7}{10}.$$

Definicja 1.15
Dyskretną przestrzenią probabilistyczną nazywamy parę $(\Omega, \mathbb{P})$, gdzie Ω jest zbiorem co najwyżej przeliczalnym a $\mathbb{P}$ jest miarą probabilistyczną na Ω.

Przykład 1.16 (Problem Monty'ego Halla)
Bramkę z samochodem oznaczymy etykietą 1. Pozostałe zaś 2 i 3. Nasz eksperyment opiszemy za pomocą pary liczb (x, y), gdzie x jest etykietą bramki, którą początkowo wybraliśmy a y jest etykietą bramki, którą odsłania prowadzący. Skoro prowadzący odsłania tylko bramki 2 lub 3, to przestrzenią zdarzeń elementarnych jest

$$\Omega = \{(1, 2), (1, 3), (2, 3), (3, 2)\}.$$

Mając określoną przestrzeń zdarzeń przejdźmy do wybrania odpowiedniej miary probabilistycznej. Zauważmy najpierw, że przy pierwszym wyborze, każda z bramek ma jednakowe prawdopodobieństwo. Dokładniej każde ze zdarzeń

$$A_i = \{\text{gracz początkowo wybrał bramkę z etykietą } i\}$$

mamy $\mathbb{P}[A_i] = \frac{1}{3}$. Oznacza to, że

$$\mathbb{P}[(2, 3)] = \mathbb{P}[(3, 2)] = \frac{1}{3} \quad \text{oraz} \quad \mathbb{P}[(1, 2)] + \mathbb{P}[(1, 3)] = \frac{1}{3}.$$

W przypadku, gdy wybieramy bramkę z etykietą 1 prowadzący ma do wyboru bramki 2 i 3. Zakładamy, że prowadzący wybiera każdą z nich z jednakowym prawdopodobieństwem. Oznacza to, że

$$\mathbb{P}[(1,2)] = \mathbb{P}[(1,3)] = \frac{1}{6}.$$

Przejdźmy teraz do opisy szukanych zdarzeń. Niech

$$A = \{\text{wygrywamy przy zmianie bramki}\} = \{(2,3), (3,2)\}$$

oraz

$$B = \{\text{wygrywamy bez zmiany bramki}\} = \{(1,2), (1,3)\}.$$

Wówczas $\mathbb{P}[A] = \frac{2}{3}$ oraz $\mathbb{P}[B] = \frac{1}{3}$.

Przejdziemy teraz do opisu podstawowych reguł i zależności liczenia prawdopodobieństw.

Fakt 1.17
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Dla zdarzeń A, B zachodzi

(i) $\mathbb{P}[A \setminus B] = \mathbb{P}[A] - \mathbb{P}[A \cap B]$
(ii) $\mathbb{P}[B'] = 1 - \mathbb{P}[B]$
(iii) $\mathbb{P}[\varnothing] = 0$
(iv) $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$

Dowód. Zauważmy, że zbiory $A \setminus B$ i $A \cap B$ są rozłączne, więc

$$\mathbb{P}[A] = \mathbb{P}[(A \setminus B) \cup (A \cap B)] = \mathbb{P}[A \setminus B] + \mathbb{P}[A \cap B],$$

co dowodzi (i). Podstawiając $A = \Omega$ do wzoru (i) otrzymujemy

$$\mathbb{P}[B'] = \mathbb{P}[\Omega \setminus B] = \mathbb{P}[\Omega] - \mathbb{P}[\Omega \cap B] = 1 - \mathbb{P}[B],$$

czyli wzór (ii). Dla $B = \Omega$ powyższa równość daje

$$\mathbb{P}[\varnothing] = \mathbb{P}[\Omega'] = 1 - \mathbb{P}[\Omega] = 1 - 1 = 0.$$

Aby wreszcie uzasadnić (iv) napiszmy

$$\begin{aligned} \mathbb{P}[A \cup B] = \mathbb{P}[B \cup (A \setminus B)] &\overset{(A2)}{=} \mathbb{P}[B] + \mathbb{P}[A \setminus B] \\ &\overset{(i)}{=} \mathbb{P}[B] + \mathbb{P}[A] - \mathbb{P}[A \cap B]. \end{aligned}$$

□

Przykład 1.18
Losujemy jedną z liczb $1, 2, \ldots 1000$ w taki sposób, że wylosowanie każdej z nich jest jednakowo prawdopodobne. Wówczas przestrzenią zdarzeń elementarnych jest

$$\Omega = \{1, 2, \ldots, 1000\}.$$

Jakie jest prawdopodobieństwo, że wylosowana liczba dzieli się przez 3 lub 5? Dla liczby p oznaczmy zdarzenie

$$D_p = \{n \in [1000] \mid n \text{ dzieli się przez } p\} \subseteq \Omega.$$

Szukamy $\mathbb{P}[D_3 \cup D_5]$. Zauważmy, że

$$D_3 = \{3m \mid 1 \leq m \leq 333\},$$

więc $\mathbb{P}[D_3] = \frac{333}{1000}$. Podobnie $\mathbb{P}[D_5] = \frac{200}{1000} = \frac{1}{5}$. Skoro $D_3 \cap D_5 = D_{15}$, to $\mathbb{P}[D_3 \cap D_5] = \frac{66}{1000} = \frac{33}{500}$. Ostatecznie na mocy **Faktu 1.17** (ii) mamy

$$\mathbb{P}[D_3 \cup D_5] = \mathbb{P}[D_3] + \mathbb{P}[D_5] - \mathbb{P}[D_3 \cap D_5] = \frac{467}{1000}.$$

Fakt 1.19 (Wzór włączeń i wyłączeń)
Niech $n \in \mathbb{N}$ i niech $A_1, A_2, \ldots, A_n$ będą zdarzeniami. Wówczas

$$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] = \sum_{1 \leq j_1 \leq n} \mathbb{P}[A_{j_1}] - \sum_{1 \leq j_1 < j_2 \leq n} \mathbb{P}[A_{j_1} \cap A_{j_2}]$$
$$+ \sum_{1 \leq j_1 < j_2 < j_3 \leq n} \mathbb{P}[A_{j_1} \cap A_{j_2} \cap A_{j_3}] - \ldots (-1)^{n-1} \mathbb{P}\left[\bigcap_{j=1}^{n} A_j\right].$$

W dowodzie wykorzystamy indykator zdarzenia A dany wzorem

$$\mathbb{1}_A(\omega) = \begin{cases} 0, \omega \notin A \\ 1, \omega \in A \end{cases}.$$

Dowód. Przypomnijmy wzór, z którego korzystaliśmy już w dowodzie poprzedniego faktu. Dla zdarzenia B,

$$\sum_{\omega \in \Omega} \mathbb{1}_B(\omega) \mathbb{P}[\{\omega\}] = \mathbb{P}[B]. \tag{1.1}$$

Niech $A = \bigcup_{j=1}^{n} A_j$. Naszym drugim krokiem będzie uzasadnienie wzoru

$$\mathbb{1}_A(\omega) = 1 - \prod_{j=1}^{n}(1 - \mathbb{1}_{A_j}(\omega)). \tag{1.2}$$

Ustalmy $\omega \in \Omega$.

Jeżeli $\omega \in A$, to lewa strona (1.2) jest równa 1. Skoro $\omega \in A$, to $\omega \in A_j$ dla pewnego j a co za tym idzie $1 - \mathbb{1}_{A_j}(\omega) = 0$, więc i prawa strona (1.2) jest równa 1.

Jeżeli $\omega \notin A$, to oczywiście lewa strona (1.2) jest równa 0. Skoro $\omega \notin A$, to $\omega \notin A_j$ dla każdego j. W takim razie $\mathbb{1}_{A_j}(\omega) = 0$ dla każdego j a co za tym idzie $\prod_{j=1}^{n}(1 - \mathbb{1}_{A_j}(\omega)) = 1$, więc prawa strona (1.2) jest równa 0. To dowodzi (1.2).

Innym sposobem dowodu (1.2) jest wykorzystanie praw De Morgana i **Faktu 1.17**(*ii*) aby napisać

$$\mathbb{1}_A = 1 - \mathbb{1}_{A'} = 1 - \mathbb{1}_{\bigcap_{j=1}^n A_j'} = 1 - \prod_{j=1}^{n} \mathbb{1}_{A_j'} = 1 - \prod_{j=1}^{n}(1 - \mathbb{1}_{A_j}).$$

Rozpisując prawą stronę (1.2) otrzymujemy

$$\begin{aligned}\mathbb{1}_A(\omega) =& 1 - \prod_{j=1}^{n}(1 - \mathbb{1}_{A_j}(\omega)) = \sum_{1 \le j_1 \le n} \mathbb{1}_{A_{j_1}}(\omega) - \sum_{1 \le j_1 < j_2 \le n} \mathbb{1}_{A_{j_1} \cap A_{j_2}}(\omega) \\ &+ \sum_{1 \le j_1 < j_2 < j_3 \le n} \mathbb{1}_{A_{j_1} \cap A_{j_2} \cap A_{j_3}}(\omega) - \ldots + (-1)^{n-1} \mathbb{1}_{\bigcap_{j=1}^n A_j}(\omega).\end{aligned}$$

Korzystając z tego wzoru możemy napisać

$$\begin{aligned}\mathbb{P}[A] = \sum_{\omega \in \Omega} \mathbb{1}_A(\omega)\mathbb{P}[\{\omega\}] &= \sum_{1 \le j_1 \le n} \sum_{\omega \in \Omega} \mathbb{1}_{A_{j_1}}(\omega)\mathbb{P}[\{\omega\}] \\ &- \sum_{1 \le j_1 < j_2 \le n} \sum_{\omega \in \Omega} \mathbb{1}_{A_{j_1} \cap A_{j_2}}(\omega)\mathbb{P}[\{\omega\}] \\ &+ \sum_{1 \le j_1 < j_2 < j_3 \le n} \sum_{\omega \in \Omega} \mathbb{1}_{A_{j_1} \cap A_{j_2} \cap A_{j_3}}(\omega)\mathbb{P}[\{\omega\}] - \ldots \\ &\ldots + (-1)^{n-1} \sum_{\omega \in \Omega} \mathbb{1}_{\bigcap_{j=1}^n A_j}(\omega)\mathbb{P}[\{\omega\}].\end{aligned}$$

Odwołując się teraz do wzoru (1.1) możemy uprościć ostatni wzór do postaci

$$\begin{aligned}\mathbb{P}[A] =& \sum_{1 \le j_1 \le n} \mathbb{P}[A_{j_1}] - \sum_{1 \le j_1 < j_2 \le n} \mathbb{P}[A_{j_1} \cap A_{j_2}] \\ &+ \sum_{1 \le j_1 < j_2 < j_3 \le n} \mathbb{P}[A_{j_1} \cap A_{j_2} \cap A_{j_3}] - \ldots (-1)^{n-1} \mathbb{P}\left[\bigcap_{j=1}^{n} A_j\right],\end{aligned}$$

która jest naszą tezą. □

Przykład 1.20
Tym razem losujemy jedną z liczb $1, 2, \ldots 2000$. Wówczas $\Omega = [2000]$. Jakie jest prawdopodobieństwo, że wylosowana liczba jest podzielna przez 9, 11, 13 lub 15? Przypomnijmy, że dla liczby naturalnej p,

$$D_p = \{k \in [2000] \mid k \text{ dzieli się przez } p\}.$$

Szukamy $\mathbb{P}[D_9 \cup D_{11} \cup D_{13} \cup D_{15}]$. Mamy

$$\mathbb{P}[D_9] = \frac{222}{2000}, \quad \mathbb{P}[D_{11}] = \frac{181}{2000}, \quad \mathbb{P}[D_{13}] = \frac{153}{2000}, \quad \mathbb{P}[D_{15}] = \frac{133}{2000}.$$

Aby wyliczyć prawdopodobieństwa poszczególnych przekrojów, możemy skorzystać z tożsamości $D_p \cap D_q = D_{pq}$ dla względnie pierwszych p i q aby otrzymać

$$\begin{aligned}
\mathbb{P}[D_9 \cap D_{11}] &= \mathbb{P}[D_{99}] = \frac{20}{2000}, \quad \mathbb{P}[D_9 \cap D_{13}] = \mathbb{P}[D_{117}] = \frac{17}{2000}, \\
\mathbb{P}[D_9 \cap D_{15}] &= \mathbb{P}[D_{45}] = \frac{44}{2000}, \quad \mathbb{P}[D_{11} \cap D_{13}] = \mathbb{P}[D_{143}] = \frac{13}{2000}, \\
\mathbb{P}[D_{11} \cap D_{15}] &= \mathbb{P}[D_{162}] = \frac{12}{2000}, \quad \mathbb{P}[D_{13} \cap D_{15}] = \mathbb{P}[D_{195}] = \frac{10}{2000}.
\end{aligned}$$

Podobnie dla względnie pierwszych p, q, r, $D_p \cap D_q \cap D_r = D_{pqr}$ oraz

$$\begin{aligned}
\mathbb{P}[D_9 \cap D_{11} \cap D_{13}] &= \mathbb{P}[D_{1287}] = \frac{1}{2000}, \quad \mathbb{P}[D_9 \cap D_{11} \cap D_{15}] = \frac{4}{2000}, \\
\mathbb{P}[D_9 \cap D_{13} \cap D_{15}] &= \mathbb{P}[D_{585}] = \frac{3}{2000}, \quad \mathbb{P}[D_{11} \cap D_{13} \cap D_{15}] = \mathbb{P}[D_{2145}] = 0.
\end{aligned}$$

Wreszcie $D_9 \cap D_{11} \cap D_{13} \cap D_{15} = \emptyset$. Wzór włączeń i wyłączeń zapisuje się jako

$$\begin{aligned}
\mathbb{P}[D_9 \cup D_{11} \cup D_{13} \cup D_{15}] =& \mathbb{P}[D_9] + \mathbb{P}[D_{11}] + \mathbb{P}[D_{13}] + \mathbb{P}[D_{15}] \\
& - \mathbb{P}[D_9 \cap D_{11}] - \mathbb{P}[D_9 \cap D_{13}] - \mathbb{P}[D_9 \cap D_{15}] \\
& - \mathbb{P}[D_{11} \cap D_{13}] - \mathbb{P}[D_{11} \cap D_{15}] - \mathbb{P}[D_{13} \cap D_{15}] \\
& + \mathbb{P}[D_9 \cap D_{11} \cap D_{13}] + \mathbb{P}[D_9 \cap D_{11} \cap D_{15}] \\
& + \mathbb{P}[D_9 \cap D_{13} \cap D_{15}] + \mathbb{P}[D_{11} \cap D_{13} \cap D_{15}] \\
& - \mathbb{P}[D_9 \cap D_{11} \cap D_{13} \cap D_{15}].
\end{aligned}$$

Daje to

$$\begin{aligned}
\mathbb{P}[D_9 \cup D_{11} \cup D_{13} \cup D_{15}] =& \frac{222 + 181 + 153 + 133}{2000} \\
& - \frac{20 + 17 + 44 + 13 + 12 + 10}{2000} + \frac{1 + 4 + 3 + 0}{2000} - 0 \\
=& \frac{581}{2000}.
\end{aligned}$$

Na koniec udowodnimy techniczny fakt, który będzie przydatny w dalszej części wyk ładu.

Fakt 1.21
Niech Ω będzie co najwyżej przeliczalną przestrzenią probabilistyczną. Wówczas dla ciągu zdarzeń $A_1, A_2, \ldots$ zachodzi

(i) jeżeli $A_1, A_2, \ldots$ są parami rozłączne, to

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \mathbb{P}[A_n]$$

(ii)Jeżeli $A_1 \subseteq A_2 \subseteq \ldots$, to

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] = \lim_{n\to\infty} \mathbb{P}[A_n]$$

Dowód. Zaczynając od (i) zauważmy, że dla każdego n,

$$\mathbb{P}[A_n] = \sum_{\omega\in A_n} \mathbb{P}[\{\omega\}] = \sum_{\omega\in\Omega} \mathbb{1}_{A_n}(\omega)\mathbb{P}[\{\omega\}].$$

Korzystając z tego, że $\mathbb{P}[\{\omega\}] \geq 0$ możemy zamienić kolejność sumowania i otrzymać

$$\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \sum_{n=1}^{\infty} \sum_{\omega\in\Omega} \mathbb{1}_{A_n}(\omega)\mathbb{P}[\{\omega\}] = \sum_{\omega\in\Omega} \mathbb{P}[\{\omega\}] \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega).$$

Korzystając z rozłączności zbiorów A_n,

$$\sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = \mathbb{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega). \tag{1.3}$$

Szczegółowy dowód (1.3) pozostawiamy jako ćwiczenie. Aby uzasadnić (ii) definiujemy zbiory $\tilde{A}_1 = A_1$, $\tilde{A}_2 = \tilde{A}_2 \setminus A_1$, $\ldots \tilde{A}_k = A_k \setminus A_{k-1}$. Wówczas zbiory $\tilde{A}_k$ są rozłączne oraz

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \tilde{A}_n.$$

Korzystając z punktu (i) mamy

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] = \mathbb{P}\left[\bigcup_{n=1}^{\infty} \tilde{A}_n\right] = \sum_{n=1}^{\infty} \mathbb{P}[\tilde{A}_n].$$

Odwołując się do definicji sumy szeregu

$$\sum_{n=1}^{\infty} \mathbb{P}[\tilde{A}_n] = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{P}[\tilde{A}_n] = \lim_{N\to\infty} \mathbb{P}[A_1] + \sum_{n=2}^{N} \mathbb{P}[A_n \setminus A_{n-1}]$$
$$= \lim_{N\to\infty} \mathbb{P}[A_1] + \sum_{n=2}^{N} \mathbb{P}[A_n] - \mathbb{P}[A_{n-1}] = \lim_{N\to\infty} \mathbb{P}[A_N].$$

□

1.3 Prawdopodobieństwo klasyczne

Wprowadzony przez nas formalizm pozwoli nam w przyszłości skutecznie analizować modele stochastyczne, które będziemy rozważać w przyszłości. Ważne jest jednak aby wprowadzony język współgrał z naszą intuicją, czyli klasyczną definicją prawdopodobieństwa.

Fakt 1.22
Niech Ω będzie skończoną przestrzenią zdarzeń elementarnych. Załóżmy, że każde zdarzenie elementarne jest jednakowo prawdopodobne, czyli że istnieje $p \in [0, 1]$ takie, że $\mathbb{P}[\{\omega\}] = p$ dla każdej $\omega \in \Omega$. Wówczas $p = \frac{1}{|\Omega|}$ oraz miara probabilistyczna $\mathbb{P}$ zadana jest wzorem

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|}$$

dla każdego zdarzenia A.

Dowód. Aby wyznaczyć wartość p wystarczy zauważyć, że

$$1 = \mathbb{P}[\Omega] = \sum_{\omega\in\Omega} \mathbb{P}[\{\omega\}] = \sum_{\omega\in\Omega} p = |\Omega| p.$$

Stąd $p = \frac{1}{|\Omega|}$. Ustalmy teraz dowolne zdarzenie $A \subseteq \Omega$. Mamy

$$\mathbb{P}[A] = \sum_{\omega\in A} \mathbb{P}[\{\omega\}] = \sum_{\omega\in A} p = |A| p = \frac{|A|}{|\Omega|}.$$

□

Przykład 1.23 (Paradoks urodzin)
Odpowiemy na pytanie jakie jest prawdopodobieństwo, że w grupie $r \in \mathbb{N}$

osób znajdą się dwie które obchodzą urodziny tego samego dnia. Każdemu dniu w roku możemy przypisać liczbę ze zbioru [356]. Jeżeli grupa liczy $r \in \mathbb{N}$ znajomych, to przestrzenią zdarzeń elementarnych jest zbiór ciągów liczb naturalnych ze zbioru $[356] = \{1, 2, \ldots 356\}$, czyli

$$\Omega = [356]^r = \left\{ (x_1, x_2, \ldots, x_r) \,:\, x_j \in [356],\, i \le r \right\}.$$

Zakładać będziemy, że każda konfiguracja jest jednakowo prawdopodobna. Chcąc zbadać prawdopodobieństwo zdarzenia, że istnieją dwie osoby, które mają urodziny tego samego dnia wygodniej będzie zbadać zdarzenie przeciwne. Niech A będzie zdarzeniem, że każda z osób ma urodziny innego dnia. Możemy formalnie zapisać

$$A = \left\{ (x_1, x_2, \ldots, x_r) \in \Omega \,:\, x_i \neq x_j,\, i \neq j \right\}.$$

Wówczas

$$|A| = 356 \cdot (356 - 1) \cdot (356 - 2) \cdot \ldots \cdot (356 - r + 1)$$

a co za tym idzie

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|} = \left(1 - \frac{1}{356}\right) \cdot \left(1 - \frac{2}{356}\right) \cdot \ldots \cdot \left(1 - \frac{r-1}{356}\right).$$

Jak duże powinno być r, aby $\mathbb{P}[A] < \frac{1}{2}$? Korzystając z oszacowania $1 - x \le e^{-x}$ dla $x \ge 0$ mamy

$$\mathbb{P}[A] \le \exp\left\{ -\frac{1}{356} - \frac{2}{356} - \ldots - \frac{r-1}{356} \right\}.$$

Chcąc uzyskać, że prawa strona jest mniejsza nić $\frac{1}{2}$ wystarcza aby

$$\frac{r(r-1)}{2 \cdot 356} \ge \frac{(r-1)^2}{2 \cdot 356} \ge \log(2).$$

Łatwo sprawdzić, że ostatnia nierówność spełniona jest już dla $r = 24$.

Przykład 1.24
Rzucamy n razy monetą. Wówczas $\Omega = \{O, R\}$. Niech

$$A_k = \{\text{wypadło dokładnie } k \text{ orłów}\}.$$

$$|A_k| = \binom{n}{k}.$$

oraz $\mathbb{P}[A_k] = \binom{n}{k} 2^{-n}$.

Przykład 1.25 (Problem podrzucanych kapeluszy)
Grupa n kibiców wygrywającej drużyny piłkarskiej podrzuciła wysoko w powietrze swoje kapelusze. Kapelusze spadły przypadkowo, z tym że każdy z n kibiców złapał jeden kapelusz. Zastanówmy się, jakie jest prawdopodobieństwo zdarzenia A polegającego na tym, że żaden z kibiców nie złapał swojego kapelusza. Niech

$$A_j = \{\, j\text{-ty kibic złapał swój kapelusz}\}.$$

Wówczas $A' = \bigcup_{j=1}^{n} A_j$. Korzystając ze wzoru włączeń i wyłączeń mamy

$$\begin{aligned}\mathbb{P}\left[A'\right] = &\sum_{1\le j_1\le n} \mathbb{P}[A_{j_1}] - \sum_{1\le j_1<j_2\le n} \mathbb{P}[A_{j_1}\cap A_{j_2}] \\ &+ \sum_{1\le j_1<j_2<j_3\le n} \mathbb{P}[A_{j_1}\cap A_{j_2}\cap A_{j_3}] - \ldots + (-1)^{n-1}\mathbb{P}\left[\bigcap_{j=1}^{n} A_j\right].\end{aligned}$$

Zauważmy, że

$$\begin{aligned}\mathbb{P}[A_{j_1}] &= \frac{(n-1)!}{n!} \\ \mathbb{P}[A_{j_1}\cap A_{j_2}] &= \frac{(n-2)!}{n!} \\ \mathbb{P}[A_{j_1}\cap A_{j_2}\cap A_{j_3}] &= \frac{(n-3)!}{n!} \\ &\cdots \\ \mathbb{P}\left[\bigcap_{j=1}^{n} A_j\right] &= \frac{1}{n!}.\end{aligned}$$

Stąd

$$\mathbb{P}\left[A'\right] = 1 - \frac{1}{n-1} + \frac{1}{(n-1)(n-2)} - \ldots + (-1)^{n-1}\frac{1}{(n-1)!}.$$

Korzystając ze wzoru na prawdopodobieństwo zdarzenia przeciwnego

$$\begin{aligned}\mathbb{P}\left[A\right] =& \binom{n}{2}\frac{(n-2)!}{n!} - \binom{n}{3}\frac{(n-3)!}{n!} + \ldots + (-1)^{n}\frac{1}{(n-1)!} \\ =& \sum_{k=2}^{n}\frac{(-1)^k}{k!} = \sum_{k=0}^{n}\frac{(-1)^k}{k!}.\end{aligned}$$

W szczególności liczba możliwości, na jakie kibice mogą złapać kapelusze wynosi

$$|A| = |\Omega|\mathbb{P}[A] = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

W przyszłości będziemy stosowali oznaczenie

$$!n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}. \tag{1.4}$$

1.4 Przykłady

Przykład 1.26
rzucamy koscia az wypadnie 6.

Przykład 1.27
Przykład przestrzeni probabilistycznej i addytywnej funkcji zbioru, która nie jest ciągła z góry na zbiorze pustym. Niech Ω

Przykład* 1.28 (Metoda probabilistyczna)
Jak pokazać, że istnieje graf o zadanych własnościach? Najprościej jest podać konkretny przykład, jednak często jest to bardzo trudne. Jednym ze sposobów ominięcia tego problemu jest tak zwana metoda probabilistyczna. Polega ona na wylosowaniu obiektu, na przykład grafu, i pokazanie, że wylosowany graf ma zadaną własność z dodatnim prawdopodobieństwem.

Liczbą Ramseya $R(k)$ dla $k \in \mathbb{N}$ nazywamy najmniejszą liczbę n, taką że dla dowolnego pokolorowania krawędzi na czerwono i niebiesko n-wierzchołkowego grafu na pełnego K_n istnieje k-wierzchołkowa klika w kolorze czerwonym lub k-wierzchołkowa kilka w kolorze niebieskim. Pokażmy, że

$$R(k) > \lfloor 2^{k/2} \rfloor, \quad k \geq 3.$$

Rozważmy losowe 2-pokolorowanie krawędziowe K_n w którym w którym każde kolorowanie występuje z jednakowym prawdopodobieństwem. Każde kolorowanie to funkcja $E \to \{C, N\}$, gdzie E jest zbiorem krawędzi grafu K_n. Wówczas $\Omega = \{C, N\}^E$, gdzie $|E| = \binom{n}{2}$, z miarą probabilistyczną

$$\mathbb{P}[\{\omega\}] = \frac{1}{|\Omega|} = 2^{-\binom{n}{2}}.$$

Ustalmy $k \leq n = \lfloor 2^{k/2} \rfloor$. Niech R będzie k-elementowym podzbiorem wierzchołków. Rozważmy

$$A_R \text{ - graf rozpięty na wierzchołkach } R \text{ jest monochromatyczny.}$$

Wówczas

$$\mathbb{P}[A_R] = 2 \cdot 2^{-\binom{k}{2}}.$$

Mamy

$$\mathbb{P}[\text{istnieje } k\text{-elementowy podzbiór monochromatyczny}] \leq \binom{n}{k} 2^{1-\binom{k}{2}}$$
$$< \frac{2^{1+k/2}}{k!} \cdot \frac{n^k}{2^{k^2/2}} < 1$$

a co za tym idzie

$$\mathbb{P}[\text{nie istnieje } k\text{-elementowy podzbiór monochromatyczny}] > 0.$$

Istnieje więc kolorowanie, dla którego nie istnieje k-elementowy podzbiór monochromatyczny, czyli $R(k,k) > n$.

1.5 Zadania

Zadanie 1.1
Rzucamy jedną kością, a następnie rzucamy dwukrotnie monetą. Opisz przestrzeń zdarzeń elementarnych Ω. Załóżmy, że wszystkie wyniki mają to samo prawdopodobieństwo. Znajdź prawdopodobieństwo następujących zdarzeń: wypadła 6 i co najmniej jedna reszka; wypadła liczba parzysta i orzeł przy drugim rzucie monetą; wypadła przynajmniej jedna reszka i wypadła liczba mniejsza niż 5.

Zadanie 1.2
Pokaż, że $\mathbb{P}[A \setminus B] = \mathbb{P}[A] - \mathbb{P}[A \cap B]$

Zadanie 1.3
Załóżmy, że w problemie Monte'go Hala uczestnik wybiera jedną z czterech bramek.

Zadanie 1.4
Pokaż, że

$$\sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = \mathbb{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega). \tag{1.5}$$

2

Zmienne losowe

Często skonstruowana przez nas przestrzeń probabilistyczna staje się zbyt duża do efektywnej analizy. Dla przykładu w eksperymencie polegającym na rzucie dużą liczbą kości, powiedzmy siedemnastu, interesować nas może tylko suma wyrzuconych oczek. W praktyce zamiast analizować poszczególne zdarzenia elementarne wygodniej jest analizować funkcje tych zdarzeń.

Definicja 2.1
Niech Ω będzie co najwyżej przeliczaną przestrzenią zdarzeń elementarnych. *Zmienna losową* nazywamy dowolną funkcję $X\colon \Omega \to \mathbb{R} \cup \{\infty\}$ zdefiniowaną na przestrzeni zdarzeń elementarnych.

Zmienne losowe same w sobie nie upraszczają problemów probabilistycznych. Stanowią one jedynie język, którym będziemy się w przyszłości posługiwać.

Przykład 2.2
Powiedzmy, że rzucamy parą kości. Niech X równe sumie wyrzuconych oczek jest zmienną losową. Aby ją formalnie zdefiniować, niech $\Omega = [6] \times [6]$ będzie przestrzenią opisującą wyniki na obu kościach. Wówczas $X\colon \to \mathbb{Z}$ zadane jest wzorem $X(a, b) = a + b$.

Wartości, które może przyjąć zmienna losowa $X\colon \Omega \to \mathbb{R}$ będziemy oznaczać przez x_k. Dokładniej $X[\Omega] = \{x_k : k \leq T\}$ dla $T \in \mathbb{N}$ jeżeli Ω jest skończona lub $T = \infty$ jeżeli Ω jest przeliczalna. Dla każdego ustalonego x_k definiujemy zdarzenie, że X jest równe x_k przez

$$\{X = x_k\} = X^{-1}[\{x_k\}] = \{\omega \in \Omega \mid X(\omega) = x_k\}.$$

W przyszłości zobaczymy, że liczby $\mathbb{P}[X = x_k]$ pomogą znaleźć uproszczony język do opisu zmiennej losowej X.

Przykład 2.3
W grze w kości, dla X równej sumie wyrzuconych oczek mamy $\Omega = [6] \times [6]$ oraz $X(a,b) = a + b$. Prawdopodobieństwo $\mathbb{P}_{00}$ przypisujące jednakowe prawdopodobieństwo każdemu zdarzeniu elementarnemu odpowiada naszej sytuacji. Wówczas

$$\mathbb{P}_{00}[X = 5] = \frac{5}{36}.$$

Przykład 2.4
Rzucamy kością aż do momentu otrzymania orła. Wówczas $\Omega = \{R\}^{\mathbb{N}} \cup \bigcup_{k=0}^{\infty} \{R\}^k \times \{O\}$ jak w **Przykładzie 1.2**. Niech X oznacza liczbę rzutów w naszym eksperymencie. Innymi słowy X jest liczbą rzutów potrzebną do otrzymania orła. Wówczas $X(O) = 1$, $X(RO) = 2$ oraz ogólnie $X(\omega_k) = k$ jeżeli ω_k składa się z $k-1$ reszek i orła. W szczególności $X(\omega_\infty) = \infty$.

2.1 Wartość oczekiwana

Najprostszym opisem zmiennej losowej X byłaby jedna liczba. Aby umotywować nasz przyszły wybór rozważmy następujący przykład. Powiedzmy, że mamy do dyspozycji niesymetryczną kostkę sześcienną na której 1 wypada z prawdopodobieństwem p_1, 2 wypada z prawdopodobieństwem p_2 itd. gdzie $p_1 + p_2 + \ldots + p_6 = 1$. Powiedzmy, że rzucamy tą kością $N \in \mathbb{N}$ razy i zastanówmy się, jaka będzie średnia z uzyskanych wyników. Niech X_k dla $k \leq N$ opisuje wynik k-tego rzutu. Załóżmy, że w naszym eksperymencie n_1 razy wpadła 1, n_2 razy wypadła 2 itd. Szukaną średnią możemy zapisać jako

$$\frac{X_1 + X_2 + \ldots + X_N}{N} = 1 \cdot \frac{n_1}{N} + 2 \cdot \frac{n_2}{N} + 3 \cdot \frac{n_3}{N} + 4 \cdot \frac{n_4}{N} + 5 \cdot \frac{n_5}{N} + 6 \cdot \frac{n_6}{N},$$

gdzie $n_k = |\{j \leq N \,:\, X_j = k\}|$ dla $k \in [6]$. Na poziomie intuicyjnym jasne jest, że dla dużego N, $\frac{n_k}{N} \approx p_k$. W przyszłości ściśle uzasadnimy ten fakt, ale na chwilę obecną pozostańmy na poziomie intuicji. Mamy

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 + 5 \cdot p_5 + 6 \cdot p_6.$$

Liczba po prawej stronie, co średnia ważona dla rzutu jedną kością. Istotnie, mamy $\mathbb{P}[X_1 = 1] = p_1$, $\mathbb{P}[X_1 = 2] = p_2$, $\ldots$, $\mathbb{P}[X_1 = 6] = p_6$. Możemy wreszcie napisać

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \sum_{k=1}^{6} k \cdot \mathbb{P}[X_1 = k]. \tag{2.1}$$

Otrzymujemy średnią wartości X_1 z wagą, która jest prawdopodobieństwem poszczególnych wyników. Taką średnią nazywamy wartością oczekiwaną X_1.

Definicja 2.5
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Powiemy, że zmienna losowa X jest *całkowalna*, jeżeli

$$\sum_{\omega \in \Omega} |X(\omega)| \mathbb{P}[\{\omega\}] < \infty.$$

Dla całkowalnej zmiennej losowej $X \colon \Omega \to \mathbb{R}$ jej *wartością oczekiwaną* nazywamy

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}[\{\omega\}].$$

W powyższej definicji, jeżeli $X(\omega_0) = \infty$ oraz $\mathbb{P}[\{\omega_0\}] = 0$ to interpretujemy $X(\omega_0)\mathbb{P}[\{\omega_0\}] = 0$. Niebawem przekonamy się, że powyższa definicja wartości oczekiwanej zadaje dokładnie prawą stronę równania (2.1).

Przykład 2.6
Wróćmy do eksperymentu czekania na pierwszego orła. Tutaj $\Omega = \{R\}^{\mathbb{N}} \cup \bigcup_{k=0}^{\infty} \{R\}^k \times \{O\}$. Niech X będzie zmienną losową opisującą liczbę rzutów potrzebnych do otrzymania orła ($X(\omega_k) = k$). Wówczas X jest całkowalna, bo

$$\sum_{\omega \in \Omega} |X(\omega)| \mathbb{P}[\{\omega\}] = \sum_{k=1}^{\infty} \frac{k}{2^k} < \infty$$

oraz $\mathbb{E}X = 2$. Z kolei zmienna zadana przez $Y = 3^X$, innymi słowy $X(\omega_k) = 3^k$ nie jest całkowalna. Istotnie

$$\sum_{\omega \in \Omega} |Y(\omega)| \mathbb{P}[\{\omega\}] = \sum_{k=1}^{\infty} \left(\frac{3}{2}\right)^k = \infty.$$

Przykład 2.7
Niech X oznacza liczbę otrzymanych oczek przy rzucie jedną wyważoną kością. Wówczas

$$\mathbb{E}X = X(\text{⚀})\frac{1}{6} + X(\text{⚁})\frac{1}{6} + X(\text{⚂})\frac{1}{6} + X(\text{⚃})\frac{1}{6} + X(\text{⚄})\frac{1}{6} + X(\text{⚅})\frac{1}{6} = \frac{7}{2}.$$

Ten przykład pokazuje, że nazwa *wartość oczekiwana* jest co najmniej niefortunna. Przy naszym eksperymencie nie możemy oczekiwać jego wartości oczekiwanej.

Fakt 2.8
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Niech $\{x_1, x_2, \dots\}$ będzie zbiorem wartości całkowalnej zmiennej losowej X, czyli $X[\Omega] = \{x_k : k \leq T\}$ dla $T \in \mathbb{N} \cup \{\infty\}$. Wówczas

$$\mathbb{E}X = \sum_{k=1}^{T} x_k \mathbb{P}[X = x_k].$$

Dowód. Niech $E_k = \{X = k\}$. Wówczas E_k są parami rozłączne oraz $\bigcup_{k=1}^{T} E_k = \Omega$ a co za tym idzie

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}[\{\omega\}] = \sum_{k=1}^{T} \sum_{\omega \in E_k} X(\omega)\mathbb{P}[\{\omega\}].$$

Skorzystaliśmy tutaj z całkowalności zmiennej losowej X: dla szeregu bezwzględnie zbieżnego zmiana kolejności sumowania nie wpływa na jego wartość. Dla $\omega \in E_k$ mamy $X(\omega) = x_k$, więc

$$\mathbb{E}X = \sum_{k=1}^{T} \sum_{\omega \in E_k} x_k \mathbb{P}[\{\omega\}] = \sum_{k=1}^{T} x_k \sum_{\omega \in E_k} \mathbb{P}[\{\omega\}].$$

Na koniec skorzystamy z $\mathbb{P}[E_k] = \sum_{\omega \in E_k} \mathbb{P}[\{\omega\}]$ aby otrzymać

$$\mathbb{E}X = \sum_{k=1}^{T} x_k \sum_{\omega \in E_k} \mathbb{P}[\{\omega\}] = \sum_{k=1}^{T} x_k \mathbb{P}[X = x_k].$$

□

Przykład 2.9
Rzucamy n razy monetą. Jako przestrzeń zdarzeń elementarnych wybieramy

$$\Omega = \{O, R\}^n = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) : \omega_k \in \{O, R\}\}.$$

Wówczas ω_k reprezentuje wynik k-tego rzutu. Zbadajmy zmienną losową X która jest liczbą otrzymanych orłów. Wówczas przykładowo $X(RR\dots R) = 0$, $X(RORR\dots R) = 1$ oraz ogólnie

$$X((\omega_1, \omega_2, \dots, \omega_n)) = |\{j \in [n] : \omega_j = O\}|.$$

Zakładając, że każdy wynik jest jednakowo prawdopodobny otrzymujemy, dla $k \leq n$,

$$\mathbb{P}[X = k] = \frac{|\{X = k\}|}{|\Omega|} \frac{\binom{n}{k}}{|\Omega|} = \binom{n}{k} 2^{-n}.$$

Korzystając z **Faktu 2.8** otrzymujemy

$$\begin{aligned}\mathbb{E}X &= \sum_{k=0}^{n} k\mathbb{P}[X=k] = \sum_{k=0}^{n} k\binom{n}{k}2^{-n} = \sum_{k=1}^{n} k\binom{n}{k}2^{-n} \\ &\overset{(*)}{=} \sum_{k=1}^{n} n\binom{n-1}{k-1}2^{-n} = \frac{n}{2}\sum_{j=0}^{n-1}\binom{n-1}{j}2^{-(n-1)} \overset{(**)}{=} \frac{n}{2}.\end{aligned}$$

Skorzystaliśmy tutaj ze wzorów

$$\binom{n}{k} \overset{(*)}{=} \frac{n}{k}\binom{n-1}{k-1}, \qquad \sum_{j=0}^{m}\binom{m}{j} \overset{(**)}{=} 2^m$$

dla $k, m \geq 1$.

Fakt 2.10
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną i niech X, Y będą całkowalnymi zmiennymi losowymi. Wówczas dla dowolnych stałych $\alpha, \beta \in \mathbb{R}$, zmienna losowa $\alpha X + \beta Y$ jest całkowalna oraz zachodzi wzór

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}X + \beta\mathbb{E}Y.$$

Zanim zobaczymy dowód powyższego faktu zaznajomimy się z jednym przykładem.

Przykład 2.11
Rozważmy grę, w której rzucamy dwiema dobrze wyważonymi kośćmi. Pierwsza kość jest sześcienna a druga dwudziestościanna. Nasz wynik obliczany jest według następującego wzoru

$$6 \cdot (\text{wynik na pierwszej kości})^2 - 2 \cdot (\text{wynik na drugiej kości}).$$

Jaka jest wartość oczekiwana wyniku? Naszą przestrzenią jest $\Omega = [6] \times [20]$ z jednakowym prawdopodobieństwem dla każdego zdarzenie elementarnego. Rozważmy zmienne losowe

$$X(a,b) = a^2 \qquad Y(a,b) = b.$$

Wtedy wynik zapisuje się jako $Z = 3X - 2Y$. Dla pierwszej zmiennej mamy $\{X = k\} = \{(a,b) \,:\, k = a^2\}$ a co za tym idzie

$$\mathbb{P}[X = k^2] = \frac{1}{6}, \qquad k \in [6]$$

oraz $\mathbb{E}[X] = \frac{91}{6}$. Podobnie

$$\mathbb{P}[Y = j] = \frac{1}{20}, \qquad j \in [20]$$

oraz $\mathbb{E}[Y] = \frac{21}{2}$. **Fakt 2.10** daje

$$\mathbb{E}[Z] = 3\mathbb{E}[X] - 2\mathbb{E}[Y] = 70.$$

*Dowód **Faktu 2.10**.* Oznaczmy $Z = \alpha X + \beta Y$. Mamy

$$\sum_{\omega\in\Omega} |Z(\omega)|\mathbb{P}[\{\omega\}] \leq |\alpha| \sum_{\omega\in\Omega} |X(\omega)|\mathbb{P}[\{\omega\}] + |\beta| \sum_{\omega\in\Omega} |Y(\omega)|\mathbb{P}[\{\omega\}] < \infty$$

co pokazuje, że Z jest całkowalna. Podobnie

$$\begin{aligned}\mathbb{E}Z &= \sum_{\omega\in\Omega} Z(\omega)\mathbb{P}[\{\omega\}] \\ &= \alpha \sum_{\omega\in\Omega} X(\omega)\mathbb{P}[\{\omega\}] + \beta \sum_{\omega\in\Omega} Y(\omega)\mathbb{P}[\{\omega\}] \\ &= \alpha\mathbb{E}X + \beta\mathbb{E}Y.\end{aligned}$$

□

Zauważmy, że jeżeli X jest zmienną losową a $\varphi\colon \mathbb{R}\cup\{\infty\} \to \mathbb{R}\cup\{\infty\}$ to zmienną losową staje się również $Z = \varphi \circ X = \varphi(X)$. Bardzo często przyda nam się sposób na efektywne obliczanie wartości oczekiwanej zmiennej losowej Z.

Fakt 2.12
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną a X zmienną losową. Niech dodatkowo $\varphi\colon \mathbb{R}\cup\{\infty\} \to \mathbb{R}\cup\infty$ będzie funkcją taką, że zmienna losowa $Z(\omega) = \varphi(X(\omega))$ jest całkowalna. Wówczas

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[Z] = \sum_{k=1}^{T} \varphi(x_k)\mathbb{P}[X = x_k],$$

gdzie $\{x_j : j \leq T\}$, dla $T \in \mathbb{N}\cup\{\infty\}$, jest zbiorem wartości zmiennej losowej X.

Dowód. Z **Definicji 2.5**

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[Z] = \sum_{\omega\in\Omega} \varphi(X(\omega))\mathbb{P}[\{\omega\}].$$

Podobnie jak w dowodzie **Faktu 2.8**, dla zbiorów $E_k = \{X = x_k\}$ mamy $\Omega = \bigcup_{k=1}^T E_k$ oraz

$$\mathbb{E}[\varphi(X)] = \sum_{\omega\in\Omega} \varphi(X(\omega))\mathbb{P}[\{\omega\}] = \sum_{k=1}^{T} \sum_{\omega\in E_k} \varphi(X(\omega))\mathbb{P}[\{\omega\}].$$

Tutaj skorzystaliśmy z całkowalności zmiennej Z. Zauważmy, że jeżeli $\omega \in E_k$, to $X(\omega) = x_k$. Korzystając z tej obserwacji możemy napisać

$$\mathbb{E}[\varphi(X)] = \sum_{k=1}^{T} \sum_{\omega\in E_k} \varphi(x_k)\mathbb{P}[\{\omega\}] = \sum_{k=1}^{T} \varphi(x_k) \sum_{\omega\in E_k} \mathbb{P}[\{\omega\}].$$

Na koniec wystarczy skorzystać z tego, że $\sum_{\omega\in E_k} \mathbb{P}[\{\omega\}] = \mathbb{P}[E_k] = \mathbb{P}[X = x_k]$ żeby otrzymać

$$\mathbb{E}[\varphi(X)] = \sum_{k=1}^{T} \varphi(x_k) \sum_{\omega\in E_k} \mathbb{P}[\{\omega\}] = \sum_{k=1}^{T} \varphi(x_k)\mathbb{P}[X = x_k].$$

$\square$

Przykład 2.13
Rzucamy n razy symetryczną monetą. Oznaczmy przez Z liczbę uzyskanych orłów (**Przykład 2.9**). Niech $s \in (0,1]$ znajdziemy zwartą postać dla $\mathbb{E}\left[s^X\right]$. Stosując **Fakt 2.12** dla funkcji $\varphi(x) = s^x$ otrzymujemy

$$\mathbb{E}\left[s^X\right] = \sum_{k=0}^{n} s^k \mathbb{P}[X = k].$$

Korzystając z otrzymanego wcześniej wzoru $\mathbb{P}[X = k] = \binom{n}{k}2^{-n}$ otrzymujemy

$$\mathbb{E}\left[s^X\right] = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{s}{2}\right)^k 2^{-n+k} = \left(\frac{s}{2} + \frac{1}{2}\right)^n.$$

Powyższy wzór bywa bardzo pomocny. Różniczkując stronami po zmiennej s otrzymujemy

$$\mathbb{E}\left[Xs^{X-1}\right] = \sum_{k=0}^{n} \binom{n}{k} ks^{k-1}2^{-n} = \frac{n}{2}\left(\frac{s}{2} + \frac{1}{2}\right)^{n-1}. \tag{2.2}$$

Podstawiając $s = 1$ otrzymujemy znany już nam wzór

$$\mathbb{E}[X] = \frac{n}{2}.$$

Jeżeli zróżniczkujemy (2.2) po zmiennej s otrzymujemy

$$\mathbb{E}\left[X(X-1)s^{X-2}\right] = \sum_{k=0}^{n} \binom{n}{k} k(k-1)s^{k-2}2^{-n} = \frac{n(n-1)}{4}\left(\frac{s}{2}+\frac{1}{2}\right)^{n-2}.$$

Dla $s = 1$ daje to

$$\mathbb{E}[X(X-1)] = \frac{n(n-1)}{4}.$$

Skoro $\mathbb{E}[X(X-1)] = \mathbb{E}[X^2 - X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]$ otrzymujemy

$$\mathbb{E}\left[X^2\right] = \frac{n(n+1)}{4}.$$

Fakt 2.14

Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Wówczas

(i) Jeżeli X jest zmienną losową taką, że dla pewnej stałej $a \in \mathbb{R}$ zachodzi $X(\omega) = a$ dla wszystkich $\omega \in \Omega$, to

$$\mathbb{E}[X] = a.$$

(ii) Jeżeli X i Y są zmiennymi losowymi takimi, że $X(\omega) \leq Y(\omega)$ dla wszystkich $\omega \in \Omega$, to

$$\mathbb{E}[X] \leq \mathbb{E}[Y].$$

Dowód. Jeżeli $X(\omega) = a$ dla wszystkich $\omega \in \Omega$, to

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}[\{\omega\}] = \sum_{\omega\in\Omega} a\mathbb{P}[\{\omega\}] = a \cdot \sum_{\omega\in\Omega} \mathbb{P}[\{\omega\}] = a \cdot 1 = a.$$

Jeśli natomiast $X(\omega) \leq Y(\omega)$ dla wszystkich $\omega \in \Omega$, to

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}[\{\omega\}] \leq \sum_{\omega\in\Omega} Y(\omega)\mathbb{P}[\{\omega\}] = \mathbb{E}[Y].$$

□

Za każdym razem kiedy dokonujemy takiego przybliżenia warto jest jest znać jego błąd. Wartość oczekiwana nie jest tutaj wyjątkiem. W przyszłości bardzo istotna będzie dla nas informacja jak bardzo zmienna losowa X różni się od swojej wartości oczekiwanej $\mathbb{E}X$.

Definicja 2.15

Wariancję zmiennej losowej X definiujemy przez

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}X)^2].$$

Powyższa definicja jest w zupełności ogólna i będzie nam służyć w przyszłości dla przestrzeni probabilistycznych, które nie są dyskretne. Pomijając chwilowo przyszłe zastosowania **Definicji 2.15**, na dyskretnej przestrzeni probabilistycznej $(\Omega, \mathbb{P})$ zachodzi wzór

$$\mathbb{V}[X] = \sum_{\omega \in \Omega} (X(\omega) - \mathbb{E}X)^2 \mathbb{P}[\{\omega\}]. \tag{2.3}$$

Przykład 2.16
Przypuśćmy, że właśnie otrzymaliśmy propozycję nie do odrzucenia; ktoś podarował nam dwa losy na pewną loterię. Organizatorzy loterii sprzedają 100 losów na cotygodniowe losowanie. Każdy z losów jest wybierany w jednorodnym procesie losowym, to znaczy, że każdy los może być wybrany z takim samym prawdopodobieństwem—i szczęśliwy właściciel wybranego losu wygrywa sto milionów dolarów. Pozostałe 99 losów nic nie wygrywa.

Możemy teraz wykorzystać nasz prezent na dwa sposoby: albo kupujemy dwa losy na to samo losowanie, albo kupimy po jednym losie na dwa różne losowania. Która strategia jest lepsza? Jeżeli przez X oznaczymy wygraną dla pierwszej strategii, to

$$\mathbb{E}X = 0 \cdot \frac{98}{100} + 100 \cdot \frac{2}{100} = 2.$$

Dla drugiej strategii Y otrzymujemy

$$\mathbb{E}Y = 0 \cdot \frac{99}{100} + 100 \cdot \frac{1}{100} + 0 \cdot \frac{99}{100} + 100 \cdot \frac{1}{100} = 2.$$

Wartości oczekiwane są takie same. Mimo to obydwie strategie wyglądają różnie. Nie patrzmy jednak na wartości oczekiwane i przeanalizujmy dokładnie rozkłady zmiennych losowych X oraz Y

	0	100	200
X	$0,98$	$0,02$	0
Y	$0,9801$	$0,0198$	$0,0001$

Wariancja ma służyć właśnie do analizy pojęcia rozproszenia zmiennej losowej. W przypadku X wariancja wynosi

$$\mathbb{V}[X] = 0,98 \cdot (0-2)^2 + 0,02 \cdot (100-2)^2 = 196$$

a w przypadku Y

$$\mathbb{V}[Y] = 0,9801 \cdot (0-2)^2 + 0,0198 \cdot (100-2)^2 + 0,0001 \cdot (200-2)^2 = 198$$

Tak jak oczekiwaliśmy, druga wariancja jest odrobinę większa, ponieważ rozkład losowy w przypadku Y jest odrobinę bardziej rozproszony.

Wzór (2.3) jest w praktyce nieporęczny. Okazuje się. $\mathbb{V}[X]$ można wyrazić w zgrabny sposób, który jest dodatkowo wygodny w rachunkach.

Fakt 2.17
Niech X będzie całkowalną zmienną losową. Wówczas

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2. \tag{2.4}$$

Dowód. Oznaczmy $m = \mathbb{E}[X] \in \mathbb{R}$. Wówczas

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}\left[X^2 - 2Xm + m^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[2mX] + \mathbb{E}\left[m^2\right] \\ &= \mathbb{E}\left[X^2\right] - 2m\mathbb{E}[X] + m^2 = \mathbb{E}\left[X^2\right] - m^2.\end{aligned}$$

□

Przykład 2.18
Niech X oznacza liczbę uzyskanych orłów w serii n rzutów symetryczną monetą. Wiemy, że $\mathbb{E}[X] = \frac{n}{2}$ oraz $\mathbb{E}\left[X^2\right] = \frac{n(n+1)}{4}$ a co za tym idzie

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 = \frac{n}{4}.$$

Wniosek 2.19
Niech X będzie całkowalną zmienną losową. Wówczas zachodzi nierówność

$$\mathbb{E}\left[X^2\right] \geq (\mathbb{E}X)^2.$$

Dowód. Jako, że X przyjmuje wartości rzeczywiste, $(X(\omega) - \mathbb{E}X)^2 \geq 0$ a co za tym idzie

$$\mathbb{V}[X] = \sum_{\omega\in\Omega} (X(\omega) - \mathbb{E}X)^2 \mathbb{P}[\{\omega\}] \geq 0.$$

Z drugiej strony $0 \leq \mathbb{V}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$, co pociąga naszą tezę. □

Twierdzenie 2.20 (Nierówność Czebyszewa)
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Dla całkowalnej zmiennej losowej X i dowolnej stałej $c > 0$ zachodzi

$$\mathbb{P}[|X - \mathbb{E}X| \geq c] \leq \frac{\mathbb{V}X}{c^2}.$$

Dowód.

$$\begin{aligned}\mathbb{V}X &= \sum_{\omega\in\Omega}(X(\omega)-\mathbb{E}X)^2\mathbb{P}[\omega] \geq \sum_{|X(\omega)-\mathbb{E}X|\geq c}(X(\omega)-\mathbb{E}X)^2\mathbb{P}[\omega] \\ &\geq c^2 \sum_{|X(\omega)-\mathbb{E}X|\geq c}\mathbb{P}[\omega] \geq c^2\mathbb{P}[|X-\mathbb{E}X|\geq c].\end{aligned}$$

□

Przykład 2.21 (problem podrzucanych kapeluszy raz jeszcze)
Grupa n kibiców wygrywającej drużyny piłkarskiej podrzuciła wysoko w powietrze swoje kapelusze. Kapelusze spadły przypadkowo, z tym że każdy z n kibiców złapał jeden kapelusz. Wówczas prawdopodobieństwo, że dokładnie k kibiców złapie swój kapelusz wynosi

$$P(n,k) = \frac{1}{n!}\binom{n}{k}!(n-k) = \frac{!(n-k)}{k!(n-k)!},$$

gdzie $!(n-k)$ jest liczbą nieporządków na $n-k$ elementach daną przez (1.4). Wyrażając te wyniki w formalizmie, który właśnie poznaliśmy, rozważmy przestrzeń $\Omega = S_n$ wszystkich permutacji n liczb $\{1,2,...,n\}$, gdzie wylosowanie każdej permutacji jest jednakowo prawdopodobne. Wówczas

$$\mathbb{P}[\{\omega\}] = \frac{1}{n!}, \quad \omega \in S_n.$$

Zmienna losowa

$$F_n(\omega) = \text{ liczba punktów stałych } \omega, \qquad \omega \in S_n$$

wyznacza liczbę poprawnie spadających kapeluszy. Wiemy, że

$$\mathbb{P}[F_n = k] = P(n,k) = \frac{!(n-k)}{k!(n-k)!}.$$

Ile wynosi wartość średnią F_n i jej odchylenie standardowe? Niech

$$F_{n,k}(\pi) = [\text{pozycja } k \text{ jest punktem stałym w } \pi] = \begin{cases} 1 & \pi(k) = k \\ 0 & \text{poza tym} \end{cases}.$$

Wówczas

$$F_n(\pi) = F_{n,1}(\pi) + F_{n,2}(\pi) + \ldots + F_{n,n}(\pi)$$

a co za tym idzie

$$\mathbb{E}F_n = \mathbb{E}F_{n,1} + \mathbb{E}F_{n,2} + \ldots + \mathbb{E}F_{n,n}.$$

Zauważmy, że

$$\mathbb{P}[F_{n,k} = 1] = \frac{1}{n}, \qquad \mathbb{P}[F_{n,k} = 0] = 1 - \mathbb{P}[F_{n,k} = 1] = \frac{n-1}{n}.$$

Stąd $\mathbb{E}F_{n,k} = \frac{1}{n}$ a co za tym idzie $\mathbb{E}F_n = 1$. Tak więc średnio jeden kapelusz będzie na swojej pozycji początkowej. Aby wyznaczyć wariancję obliczmy najpierw

$$\begin{aligned}\mathbb{E}F_n^2 &= \sum_{k=1}^{n} \mathbb{E}F_{n,k}^2 + \sum_{j \neq k} \mathbb{E}[F_{n,k}F_{n,j}] = \sum_{k=1}^{n} \frac{1}{n} + \sum_{j \neq k} \frac{(n-2)!}{n!} \\ &= 1 + \binom{n}{2} \frac{2}{n(n-1)} = 2.\end{aligned}$$

Wariancja wynosi zatem $\mathbb{V}[F_n] = 1$. Korzystając z Nierówność Czebyszewa,

$$\mathbb{P}[|F_n - 1| \geq 2] \leq \frac{\mathbb{V}[F_n]}{4} = \frac{1}{4}.$$

Losowa permutacja $n \geq 2$ elementów ma 1 ± 1 stałych punktów

2.2 Funkcje tworzące

Powiemy teraz nieco więcej o podejściu zaprezentowanym w **Przykładzie 2.13**. W tej części będziemy rozważać zmienną losową o wartościach naturalnych, czyli funkcję $X \colon \Omega \to \mathbb{N}$. Dla zmiennej losowej X definiujemy funkcję tworzącą $G_X \colon (0,1) \to \mathbb{R}$ wzorem

$$G_X(s) = \mathbb{E}\left[s^X\right].$$

Przykład 2.22
Niech X będzie liczbą otrzymanych orłów w n rzutach monetą. Jak wyliczyliśmy w **Przykładzie 2.13** funkcja tworząca X jest dana wzorem

$$G_X(s) = \left(\frac{s}{2} + \frac{1}{2}\right)^n.$$

Fakt 2.23
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Dla całkowalnej zmiennej losowej X o wartościach naturalnych i stowarzyszonej z nią funkcji tworzącej G_X zachodzą wzory

$$\mathbb{E}[X] = G_X'(1-) \quad \text{oraz} \quad \mathbb{E}\left[X^2\right] = G_X''(1-) + G_X'(1-).$$

Dowód. Zauważmy najpierw, że szereg

$$G_X(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}[X = n]$$

jest zbieżny jednostajnie dla $s \in (0,1)$ a zatem można wejść z pochodną pod sumę i otrzymać

$$G'_X(s) = \sum_{n=1}^{\infty} n s^{n-1} \mathbb{P}[X = n].$$

Gdy $s \to 1$ z twierdzenia Abela otrzymujemy

$$G'_X(1-) = \lim_{s\to 1} G'_X(s) = \sum_{n=1}^{\infty} n\mathbb{P}[X = n] = \mathbb{E}X.$$

Podobnie pokazujemy, że

$$G''_X(1-) = \lim_{s\to 1} G'_X(s) = \sum_{n=2}^{\infty} n(n-1)\mathbb{P}[X = n] = \mathbb{E}X^2 - \mathbb{E}[X].$$

□

Następny przykład wyjaśnia konieczność

Przykład 2.24
Niech z jednakowym prawdopodobieństwem X będzie jedną z liczb $\{1, 2, \ldots, 10\}$. Wówczas licząc wprost z definicji $\mathbb{E}[X] = \frac{11}{2}$. Sprawdźmy funkcję tworzącą X.

$$G_X(s) = \mathbb{E}\left[s^X\right] = \frac{1}{10}\left(s + s^2 + s^3 + \ldots + s^{10}\right) = \frac{s - s^{11}}{6(1-s)}.$$

Stąd

$$G'_X(s) = \frac{10s^{11} - 11s^{10} + 1}{6(1-s)^2}.$$

Przechodząc z s do 1,

$$G'_X(1-) = \lim_{s\to 1} \frac{10s^{11} - 11s^{10} + 1}{6(1-s)^2} = \lim_{s\to 1} \frac{110s^{10} - 110s^9}{-12(1-s)}$$

2.3 Wektory losowe

Gdy rozważamy więcej niż jedną zmienną losową często wygodnie jest je grupować w wektor. Przykładowo o dwóch zmiennych losowych możemy myśleć jak o odwzorowaniu $\Omega \to \mathbb{R}^2$.

Definicja 2.25
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. *Wektorem losowym* nazywamy dowolną funkcję $\mathbb{X} \colon \Omega \to \mathbb{R}^n$.

Jeżeli $\mathbb{X}$ jest wektorem losowym dwuwymiarowym, to jest on postaci

$$\mathbb{X}(\omega) = (X(\omega), Y(\omega)) \qquad \omega \in \Omega,$$

gdzie X, Y są zmiennymi losowymi. Zanim zobaczymy zastosowanie wektorów losowych udowodnimy jeden pomocniczy fakt. Rozważmy dowolną funkcję $F \colon \mathbb{R}^2 \to \mathbb{R}$. Wówczas $F \circ \mathbb{X} \colon \Omega \to \mathbb{R}$ jest zmienną losową.

Fakt 2.26
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Załóżmy, że $\mathbb{X} = (X, Y) \colon \Omega \to \mathbb{R}^2$ jest dwuwymiarowym wektorem losowym takim. Niech $x_1, x_2, \ldots$ będzie zbiorem wartości zmiennej X, czyli $Y[\Omega] = \{x_j \;:\; j \le T_X\}$. Podobnie a $Y[\Omega] = \{y_k \;:\; k \le T_Z\}$. Niech $F \colon \mathbb{R}^2 \to \mathbb{R}$ będzie funkcją taką, że

$$\mathbb{E}[|F(\mathbb{X})|] = \sum_{\omega \in \Omega} |F(\mathbb{X}(\omega))| \mathbb{P}[\omega] < \infty.$$

Wówczas zachodzi wzór na wartość oczekiwaną zmiennej losowej $F(\mathbb{X})$,

$$\mathbb{E}[F(\mathbb{X})] = \sum_{j=1}^{T_X} \sum_{k=1}^{T_Z} F(x_j, y_k) \mathbb{P}[X = x_j,\, Y = y_k].$$

Dowód. Niech $E_{j,k} = \{X = x_j,\, Y = y_k\}$. Wówczas $\Omega = \bigcup_{j=1}^{T_X} \bigcup_{k=1}^{T_Y} E_{j,k}$. Mamy

$$
\begin{aligned}
\mathbb{E}[|F(\mathbb{X})|] &= \sum_{\omega\in\Omega} F(\mathbb{X}(\omega))\mathbb{P}[\{\omega\}] = \sum_{j=1}^{T_X}\sum_{k=1}^{T_Z}\sum_{\omega\in E_{j,k}} F(\mathbb{X}(\omega))\mathbb{P}[\{\omega\}] \\
&= \sum_{j=1}^{T_X}\sum_{k=1}^{T_Z}\sum_{\omega\in E_{j,k}} F(x_j,y_k)\mathbb{P}[\{\omega\}] = \sum_{j=1}^{T_X}\sum_{k=1}^{T_Z} F(x_j,y_k)\sum_{\omega\in E_{j,k}}\mathbb{P}[\{\omega\}] \\
&= \sum_{j=1}^{T_X}\sum_{k=1}^{T_Z} F(x_j,y_k)\mathbb{P}[X=x_j,\ Y=y_k].
\end{aligned}
$$

□

Przykład 2.27
?

Definicja 2.28
Załóżmy, że $(\Omega,\mathbb{P})$ jest dyskretną przestrzenią probabilistyczną. Niech $\mathbb{X}\colon\Omega\to\mathbb{R}$ będzie wektorem losowym takim, że

$$\mathbb{E}[|X|],\ \mathbb{E}[|Y|],\ \mathbb{E}[|XY|]\ <\infty.$$

Kowariancją X i Y nazywamy

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right].$$

Korelacją X i Y nazywamy

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}.$$

Przykład 2.29
W celu ilustracji korelacji między zmiennymi losowymi X i Y rozważmy następujące zagadnienie. Załóżmy, że poza naszymi oczami ma miejsce eksperyment losowy, którego wynik (czyli ω) nie jest nam znany. To co obserwujemy, to wartość zmiennej $X(\omega)$. Przy pomocy znanej wartości $X(\omega)$ chcemy przybliżyć nieznaną nam wartość $Y(\omega)$. Naszym przybliżeniem będzie

$$\widehat{Y}(\omega) = aX(\omega)+b$$

dla pewnych stałych $a,b\in\mathbb{R}$. Naszym kryterium doboru stałych a i b będzie minimalizacja błędu średniokwadratowego, czyli

$$k(a,b) = \mathbb{E}\left[\left(Y-\widehat{Y}\right)^2\right].$$

Mamy

$$\begin{aligned}
k(a,b) &= \mathbb{E}\left[(Y-aX-b)^2\right] \\
&= \mathbb{E}\left[((Y-\mathbb{E}Y)-a(X-\mathbb{E}X)+(\mathbb{E}Y-a\mathbb{E}X-b))^2\right] \\
&= \mathbb{V}[Y]+a^2\mathbb{V}[X]+(\mathbb{E}Y-a\mathbb{E}X-b)^2-2a\mathrm{Cov}(X,Y) \\
&= \left(a\mathbb{V}(X)-\frac{\mathrm{Cov}(X,Y)}{\mathbb{V}(X)}\right)^2+\mathbb{V}[Y](1-\rho(X,Y)^2)+(\mathbb{E}Y-a\mathbb{E}X-b)^2
\end{aligned}$$

Wyrażenie to osiąga minimum dla

$$a_0 = \frac{\mathrm{Cov}(X,Y)}{\mathbb{V}[X]} = \sqrt{\frac{\mathbb{V}[Y]}{\mathbb{V}[X]}}\rho(X,Y),$$

$$b_0 = \mathbb{E}[Y]-a\mathbb{E}[X] = \mathbb{E}[Y]-\sqrt{\frac{\mathbb{V}[Y]}{\mathbb{V}[X]}}\rho(X,Y)\mathbb{E}[X].$$

Dla tych parametrów błąd średniokwadratowy naszego przybliżenia wynosi

$$k(a_0,b_0) = \mathbb{V}[Y](1-\rho(X,Y)^2).$$

Twierdzenie 2.30
Załóżmy, że $(\Omega,\mathbb{P})$ jest dyskretną przestrzenią probabilistyczną. Niech $\mathbb{X}\colon \Omega \to \mathbb{R}$ będzie wektorem losowym takim, że

$$\mathbb{E}\left[X^2\right], \mathbb{E}\left[Y^2\right] < \infty.$$

wówczas

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

Dowód. Dla $t \in \mathbb{R}$ rozważmy funkcję

$$f(t) = \mathbb{E}\left[(X+tY)^2\right]$$

Zauważmy, że $f(t) \geq 0$ oraz

$$f(t) = t^2\mathbb{E}[Y^2]+2t\mathbb{E}[XY]+\mathbb{E}[X^2]$$

f jako nieujemny trójmian kwadratowy musi mieć niedodatni wyróżnik. Stąd

$$\Delta_f = 4\mathbb{E}[XY]^2-4\mathbb{E}[Y^2]\mathbb{E}[X^2].$$

Warunek $\Delta_f \leq 0$ jest równoważny naszej tezie. □

Przykład 2.31
?

2.4 Przykłady

Przykład* 2.32
Przejdźmy do następnego przykładu: Ile razy trzeba rzucić monetą, żeby otrzymać dwa orły z rzędu? Okazuje się, że skrupulatne określenie przestrzeni probabilistycznej będzie w tym przykładzie bardzo pomocne. Zdarzenia elementarne składają się ze skończonych ciągów symboli O i R które nie zawierają wzorca OO z wyjątkiem ostatnich dwóch pozycji. Innymi słowy

$$\Omega = \{OO, ROO, OROO, RROO, \ldots\}.$$

Interesuje nas zmienna losowa

$$X(\omega) = k, \qquad \omega \text{ jest ciągiem długości } k.$$

Przykładowo $X(ROO) = 3$, $X(OROO) = 4$ oraz $X(ORROROO) = 7$. Naszym celem będzie wyznaczenie funkcji tworzącej X, $G_X(s) = \mathbb{E}\left[s^X\right]$. Przypomnijmy, że

$$\mathbb{P}[\{\omega\}] = p(\omega) = 2^{-k}, \qquad \omega \text{ jest ciągiem długości } k.$$

Powinniśmy sprawdzić, że p jest funkcją prawdopodobieństwa. Zauważmy, że Ω ma następującą strukturę. Każdy ciąg z $\Omega \setminus \{OO\}$ zaczyna się wzorcem R bądź OR po którym następuje iny ciąg, w który nie zawiera wzorca OO poza ostatnią pozycją (czyli inny element Ω). Możemy napisać

$$\Omega = \{OO\} \cup \{R\} \times \Omega \cup \{OR\} \times \Omega.$$

Mamy

$$\begin{aligned}
G_X(s) &= \sum_{\omega \in \Omega} s^{X(\omega)} \mathbb{P}[\{\omega\}] \\
&= \frac{s^2}{2^2} + \sum_{\omega \in \{R\} \times \Omega} s^{X(\omega)} \mathbb{P}[\{\omega\}] + \sum_{\omega \in \{OR\} \times \Omega} s^{X(\omega)} \mathbb{P}[\{\omega\}] \\
&= \frac{s^2}{2^2} + \sum_{\omega \in \Omega} s^{X(R,\omega)} \mathbb{P}[\{R, \omega\}] + \sum_{\omega \in \Omega} s^{X(OR,\omega)} \mathbb{P}[\{OR, \omega\}] \\
&= \frac{s^2}{2^2} + \frac{s}{2} \sum_{\omega \in \Omega} s^{X(\omega)} \mathbb{P}[\{\omega\}] + \frac{s^2}{2^2} \sum_{\omega \in \Omega} s^{X(\omega)} \mathbb{P}[\{\omega\}].
\end{aligned}$$

To prowadzi do wzoru

$$G_X(s) = \frac{\frac{s^2}{4}}{1 - \frac{s}{2} - \frac{s^2}{4}} = \frac{s^2}{4 - 2s - s^2}.$$

Podstawiając $s = 1$ przekonujemy się, że p jest funkcją prawdopodobieństwa

$$1 = G_X(1) = \sum_{\omega \in \Omega} \mathbb{P}[\{\omega\}].$$

Mamy też

$$G'_X(s) = \frac{2(4-s)s}{(-s^2 - 2s + 4)^2}$$

Stąd $\mathbb{E}[X] = 6$.

2.5 Zadania

Zadanie 2.1
Znajdź zbiór $\{X = k\}$ dla ...

Zadanie 2.2
Pokaż, że

$$\sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = \mathbb{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega). \tag{2.5}$$

3

Niezalezność i prawdopodobieństwo warunkowe

Jednym z problemów jest zrozumienie szeroko pojętej struktury eksperymentów losowych. Jednym ze sposobów nakreślenia tej struktury jest rozumienia jak zajście jednego zdarzenia wpływa na prawdopodobieństwo zajścia innych zdarzeń.

3.1 Prawdopodobieństwo warunkowe

Rozważmy rzut dwiema kośćmi sześciennymi. Wiemy już, że odpowiednia przestrzeń zdarzeń elementarnych to

$$\Omega = D^2 = \{\epsdice{1}\epsdice{1}, \epsdice{1}\epsdice{2}, \epsdice{1}\epsdice{3}, \dots, \epsdice{6}\epsdice{6}\}.$$

Jeżeli rozważymy zdarzenie $A =$ suma oczek na obu kościach wynosi 6, to A jest zbiorem danym przez

$$A = \{\epsdice{1}\epsdice{5}, \epsdice{2}\epsdice{4}, \epsdice{3}\epsdice{3}, \epsdice{2}\epsdice{4}, \epsdice{5}\epsdice{1}\}.$$

Zakładać będziemy, że kości są dobrze wyważone. Do opisu tego eksperymentu posłużymy się prawdopodobieństwem $\mathbb{P}$, które każdemu zdarzeniu elementarnemu przypisuje takie samo prawdopodobieństwo. Pracujemy więc na przestrzeni probabilistycznej $(\Omega, \mathbb{P})$. Wówczas $\mathbb{P}[A] = \frac{5}{36}$. Złóżmy teraz, że posiadamy dodatkową informację, że na pierwszej kości wypadło jedno oczko. Wówczas powinniśmy zmienić rozważaną przestrzeń zdarzeń elementarnych, a mianowicie

$$B = \{\epsdice{1}\epsdice{1}, \epsdice{1}\epsdice{2}, \epsdice{1}\epsdice{3}, \epsdice{1}\epsdice{4}, \epsdice{1}\epsdice{5}, \epsdice{1}\epsdice{6}\}.$$

Musimy zmienić również sposób przypisywania prawdopodobieństwa. Naturalnie każdemu zdarzeniu z przestrzeni B prawdopodobieństwo $\mathbb{P}_1$ przypisuje każdemu zdarzeniu elementarnemu jednakowe prawdopodobieństwo. Wówczas zdarzenie $A_1 =$ suma oczek jest równa 6 to

$$A_1 = A \cap B = \{\epsdice{1}\epsdice{5}\}$$

co daje $\mathbb{P}_1[A_1] = \frac{1}{6}$. Zauważmy, że

$$\mathbb{P}_1[A_1] = \mathbb{P}_1[A \cap B] = \frac{|A \cap B|}{|B|} = \frac{|A \cap B| \cdot |\Omega|^{-1}}{|B| \cdot |\Omega|^{-1}} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Okazuje się, że nie prawdopodobieństwo $\mathbb{P}_1$ w nowej przestrzeni probabilistycznej można wyrazić w terminach pierwotnie rozważanego prawdopodobieństwa $\mathbb{P}$. Ta konstrukcja jest spotykana na tak często, że wyrażenie występujące po prawej stronie ostatniego wzoru ma swoją specjalną nazwę.

Definicja 3.1
Niech B będzie zdarzeniem takim, że $\mathbb{P}[B] > 0$. Prawdopodobieństwem warunkowym pod warunkiem B nazywamy

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \tag{3.1}$$

dla zdarzenia $A \subseteq \Omega$.

Należy się przekonać, że $\mathbb{P}[\cdot \mid B]$ jest miarą probabilistyczną na Ω. Sprawdzimy, że $\mathbb{P}[\cdot \mid B]$ spełnia konstruktywną definicję miary probabilistycznej **Definicję 1.10**. Jako ćwiczenia warto sprawdzić, że $\mathbb{P}[\cdot \mid B]$ spełnia aksjomatyczną definicję miary probabilistycznej **Definicję 1.13**.

Fakt 3.2
Załóżmy, że $(\Omega, \mathbb{P})$ jest dyskretną przestrzenią probabilistyczną. Niech B będzie zdarzeniem o dodatnim prawdopodobieństwie. Wówczas prawdopodobieństwo warunkowe pod warunkiem B zadane wzorem (3.1) jest miarą probabilistyczną na Ω.

Dowód. Musimy sprawdzić, że

$$\forall \omega \in \Omega \quad \mathbb{P}[\omega \mid B] \in [0,1], \qquad \sum_{\omega \in \Omega} \mathbb{P}[\omega \mid B] = 1.$$

Pierwszy warunek wynika z tego, że

$$0 \leq \mathbb{P}[\{\omega\} \cap B] \leq \mathbb{P}[B].$$

Drugi warunek wynika z następującego rachunku

$$
\begin{aligned}
\sum_{\omega\in\Omega} \mathbb{P}[\omega \mid B] &= \frac{1}{\mathbb{P}[B]} \sum_{\omega\in\Omega} \mathbb{P}[\{\omega\} \cap B] \\
&= \frac{1}{\mathbb{P}[B]} \sum_{\omega\in B} \mathbb{P}[\{\omega\} \cap B] + \frac{1}{\mathbb{P}[B]} \sum_{\omega\in B'} \mathbb{P}[\{\omega\} \cap B] \\
&= \frac{1}{\mathbb{P}[B]} \sum_{\omega\in B} \mathbb{P}[\{\omega\}] + \frac{1}{\mathbb{P}[B]} \sum_{\omega\in B'} \mathbb{P}[\emptyset] = 1 + 0 = 1.
\end{aligned}
$$

□

Przykład 3.3
Przykład z rozkładem niejednostajnym.

Fakt 3.4 (Reguła łańcucha)
Niech Ω będzie co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. Niech $A_1, A_2, \ldots A_n$ będą zdarzeniami takimi, że

$$\mathbb{P}\left[\bigcap_{j=1}^{n} A_j\right] > 0.$$

Wówczas dla $B_1 = A_1$, $B_k = \bigcap_{j=1}^{k} A_j$ dla $k \geq 2$ zachodzi wzór

$$\mathbb{P}\left[\bigcap_{j=1}^{n} A_j\right] = \mathbb{P}[B_1] \prod_{k=2}^{n} \mathbb{P}[A_k \mid B_{k-1}]. \tag{3.2}$$

Załóżmy, ze $n = 3$ wówczas reguła łańcucha zapisuje się jako

$$\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_1]\mathbb{P}[A_2|A_1]\mathbb{P}[A_3 \mid A_1 \cap A_2].$$

Zanim udowodnimy wzór (3.2) zobaczymy jego zastosowanie.

Przykład 3.5

Fakt 3.6
Załóżmy, że Ω jest co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. Niech $A_1, A_2, \ldots A_n$ będą wzajemnie wykluczającymi się zdarzeniami takimi, że

$$\bigcup_{j=1}^{n} A_j = \Omega.$$

Wówczas dla dowolnego zdarzenia B,

$$\mathbb{P}[B] = \sum_{j=1}^{n} \mathbb{P}[A_j]\mathbb{P}[B \mid A_j].$$

Dowód. Zauważmy, że zdarzenia $\{B \cap A_j\}_{j=1}^{n}$ wykluczają się wzajemnie. Stąd

$$\mathbb{P}[B] = \mathbb{P}\left[\bigcup_{j=1}^{n} B \cap A_j\right] = \sum_{j=1}^{n} \mathbb{P}[B \cap A_j].$$

Z kolei z reguły łańcucha $\mathbb{P}[B \cap A_j] = \mathbb{P}[A_j]\mathbb{P}[B \mid A_j]$, co po podstawieniu do powyższego wzoru pociąga tezę. $\square$

Przykład 3.7

Fakt 3.8
Załóżmy, że Ω jest co najwyżej przeliczalną przestrzenią zdarzeń elementarnych. Niech $A_1, A_2, \ldots A_n$ będą wzajemnie wykluczającymi się zdarzeniami takimi, że

$$\bigcup_{j=1}^{n} A_j = \Omega.$$

Wówczas dla dowolnych zdarzeń B i H,

$$\mathbb{P}[B \mid H] = \frac{\mathbb{P}[B]\mathbb{P}[H \mid B]}{\sum_{j=1}^{n} \mathbb{P}[A_j]\mathbb{P}[H \mid A_j]}.$$

Dowód. Ze wzoru na prawdopodobieństwo całkowite

$$\mathbb{P}[H] = \sum_{j=1}^{n} \mathbb{P}[A_j]\mathbb{P}[H \mid A_j].$$

Dodatkowo $\mathbb{P}[B \cap H] = \mathbb{P}[B]\mathbb{P}[H \mid B]$ co razem daje

$$\mathbb{P}[B \mid H] = \frac{\mathbb{P}[B \cap H]}{\mathbb{P}[H]} = \frac{\mathbb{P}[B]\mathbb{P}[H \mid B]}{\sum_{j=1}^{n} \mathbb{P}[A_j]\mathbb{P}[H \mid A_j]}.$$

$\square$

Przykład 3.9

3.2 Stochastyczna niezależność

Wróćmy do naszego ulubionego przykładu polegającym na rzucie dwoma kośćmi. Wówczas przestrzenią probabilistyczną jest $\Omega = D^2$ z miarą jednostajną probabilistyczną $\mathbb{P}$ na Ω. Rozważmy dwa zdarzenia

$$A = \{\text{na pierwszej kości wypadła liczba podzielna przez } 3\}$$

oraz

$$B = \{\text{na drugiej kości wypadła liczba mniejsza niż } 6\}.$$

Wtedy $\mathbb{P}[A] = \frac{1}{2}$ oraz $\mathbb{P}[B] = \frac{5}{6}$. Z kolei prawdopodobieństwo A pod warunkiem B wynosi

$$\mathbb{P}[A \mid B] = \frac{1}{2} = \mathbb{P}[A].$$

Innymi słowy zajście zdarzenia B nie wpływa na prawdopodobieństwo zdarzenia A. Powyższa równość może być zapisana jako

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

Definicja 3.10
Powiemy, że zdarzenie A i B są *niezależne*, jeżeli

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

Definicja 3.11
Powiemy, że zdarzenie A, B i C są *niezależne*, jeżeli

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B], \quad \mathbb{P}[B \cap C] = \mathbb{P}[B]\mathbb{P}[C], \quad \mathbb{P}[C \cap B] = \mathbb{P}[C]\mathbb{P}[B]$$

oraz

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

3.3 Niezależne zmienne losowe

Definicja 3.12
Zmienne losowe X, Y nazywamy *niezależnymi* jeżeli dla każdych $x, y \in \mathbb{R}$,

$$\mathbb{P}[X = x,\ Y = y] = \mathbb{P}[X = x]\mathbb{P}[Y = y].$$

Fakt 3.13
Niech $(\Omega, \mathbb{P})$ będzie dyskretną przestrzenią probabilistyczną. Załóżmy, że X, Y są niezależnymi zmiennymi losowymi. Wówczas

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

Dowód. □

4

Dyskretne rozkłady prawdopodobienstwa

Definicja 4.1
Dyskretnym rozkładem prawdopodobieńtwa na $\mathbb{R}$ nazywamy funkcję $p\colon \mathbb{R} \to [0,1]$ taką dla której istnije ciąg skończony lub nieskończony ciąg $x_1, x_2, \ldots$ taki, że

$$\sum_{k=1}^{\infty} p(x_k) = 1.$$

4.1 Rozkład geometryczny

4.2 Rozkład dwumianowy

4.3 Rozkład Poissona

5

Aproksymacja i rozklady ciagle

Niekiedy dokładne obliczenie wartości oczekiwanej może być problematyczne. W przypadku, gdy przestrzeń próbek jest bardzo duża, suma reprezentująca wartość oczekiwaną może być bardzo ciężka do dokładnego obliczenia. W takich przypadkach możemy posłużyć się jej przybliżoną wartością. Aby zilustrować takie przybliżenie rozważmy X_n - zmienną losową o rozkładzie danym wzorem

$$\mathbb{P}[X_n = k] = \frac{2k}{n(n-1)}, \quad k \in [n].$$

Rozważmy zagadnienie przybliżonego obliczenia wartości oczekiwanej

$$\mathbb{E}\left[f\left(\frac{X_n}{n}\right)\right]$$

dla ciągłej funkcji $f\colon [0,1] \to \mathbb{R}$. Rozpisując

$$\mathbb{E}\left[f\left(\frac{X_n}{n}\right)\right] = \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\frac{2k}{n(n-1)} \to \int_0^1 f(x) 2x dx.$$

Podobnie chcąc obliczyć przybliżoną wartość prawdopodobieństwa

$$\mathbb{P}\left[\alpha < \frac{X_n}{n} \leq \beta\right]$$

dla $0 < \alpha < \beta < 1$ możemy posłużyć się przybliżeniem całkowym

$$\mathbb{P}\left[\alpha < \frac{X_n}{n} \leq \beta\right] = \sum_{k=[\alpha n]+1}^{[\beta n]} \frac{2k}{n(n-1)} \to \int_\alpha^\beta 2x dx.$$

Funkcja graniczna $2x$ jest dobrana dość sztucznie i jej główna zaleta polega na prostocie powyższych argumentów. Istnieje jednak bardzo ważny rozkład, który występuje jako granica dla bardzo szerokiej klasy ciągów zmiennych losowych.

Definicja 5.1
Gęstością rozkładu normalnego nazywamy funkcję $\varphi\colon \mathbb{R} \to \mathbb{R}$ daną wzorem

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{5.1}$$

Słowo *gęstość* odnosi się do szczególnej klasy rozkładów ciągłych, które omówimy szerzej w niedalekiej przyszłości. Skupimy się na opisie podstawowych własności funkcji φ.

Fakt 5.2
Niech φ będzie gęstością rozkładu normalnego zadaną (5.1). Wówczas

$$\int_{-\infty}^{+\infty} \varphi(x) dx = 1.$$

Dowód. Zauważmy, że

$$I = \int_{-\infty}^{+\infty} e^{-x^2/2} dx < \infty.$$

Korzystając ze współrzędnych biegunowych

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} e^{-x^2/2} dx \cdot \int_{-\infty}^{+\infty} e^{-y^2/2} dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2/2-y^2/2} dxy \\ &= \int_0^{2\pi} \int_0^{+\infty} re^{-r^2/2} dr d\theta = 2\pi. \end{aligned}$$

Stąd $I = \sqrt{2\pi}$. $\square$

W przyszłości będziemy korzystali z przybliżenia dla $\alpha < \beta$ oraz pewnych zmiennych losowych S_n^*,

$$\mathbb{P}\left[\alpha < S_n^* \leq \beta\right] \sim \int_\alpha^\beta \varphi(x) dx.$$

Całkę po prawej stronie wygodnie będzie nam zapisywać w postaci różnicy

$$\int_\alpha^\beta \varphi(x) dx = \int_{-\infty}^\beta \varphi(x) dx - \int_{-\infty}^\alpha \varphi(x) dx.$$

Definicja 5.3
Dystrybuantą rozkładu normalnego nazywamy funkcję $\Phi\colon \mathbb{R} \to [0,1]$ daną wzorem

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt.$$

5.1 Twierdzenie de Moivre'a-Laplace'a

Lemat 5.4 (Wzór Stirlinga)

$$n! \sim \sqrt{2\pi n}\left(\frac{e}{n}\right)^n \tag{5.2}$$

Dowód. s □

Niech S_n będzie liczbą sukcesów w schemacie n prób Bernoulliego z prawdopodobieństwem sukcesu $p \in (0,1)$.

$$\mathbb{P}[S_{2n} = 2k] = \binom{2n}{n+k} p^{n+k}(1-p)^{n-k}$$

$$\binom{2n}{n+k} \sim 2^{2n}\left(1+\frac{k}{n}\right)^{-n-k} \cdot \left(1-\frac{k}{n}\right)^{-n+k} \cdot (\pi n)^{-1/2}\left(1+\frac{k}{n}\right)^{-1/2}\left(1-\frac{k}{n}\right)^{-1/2}$$

take $2k = (2p-1)2n + \sqrt{2n}x$.

$$2^{n+k}p^{n+k}\left(1+\frac{k}{n}\right)^{-n-k} = \left(1+\sqrt{\frac{1-p}{2p}}\frac{x}{2\sqrt{n}}\right)^{-n-k}$$

$$2^{n-k}(1-p)^{n-k}\left(1-\frac{k}{n}\right)^{-n+k} = \left(1-\sqrt{\frac{p}{2(1-p)}}\frac{x}{2\sqrt{n}}\right)^{-n+k}$$

Twierdzenie 5.5 (de Moivre'a - Laplace'a)
Niech S_n będzie zmienną losową o rozkładzie $b(n,p)$. Wówczas

$$\mathbb{P}\left[a < \frac{S_n - pn}{\sqrt{np(1-p)}} \leq b\right] \to \Phi(b) - \Phi(a).$$

5.2 Rozkłady ciągłe

Jak przekonaliśmy się w poprzedniej części tego rozdziału niekiedy warto jest zastosować przybliżenie całkowe

$$\mathbb{P}[a < X \leq b] \approx \int_a^b f(x)dx.$$

Definicja 5.6
Gęstością nazywamy dowolną funkcję $f\colon \mathbb{R} \to [0, +\infty)$ taką, że

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Definicja 5.7
Ciągłym rozkładem prawdopodobieństwa nazywamy funkcję μ określoną na przedziałach $I \subseteq \mathbb{R}$ taką, że

$$\mu(I) = \int_I f(x)dx$$

dla pewnej gęstości f, gdzie I jest przedziałem postaci $I = (a, b)$.

6

Rozkłady prawdopodobieństwa i funkcje charaktertstyczne

Definicja 6.1
Rozkładem prawdopodobieństwa nazywamy funkcję μ określoną na przedziałach $I \subseteq \mathbb{R}$, taką, że

$$\mu(I) = \mu_c(I) + \mu_d(I)$$

gdzie μ_c jest ciągłym rozkładem prawdopodobieństwa a μ_d jest dyskretnym rozkładem prawdopodobieństwa.

6.1 Funkcje charakterystyczne

Przypomnijmy, że dla $x \in \mathbb{R}$,

$$e^{it} = \cos(x) + i\sin(x).$$

Definicja 6.2
Niech μ będzie rozkładem prawdopodobieństwa. Funkcją charakterystyczną rozkładu μ nazywamy $\varphi_\mu \colon \mathbb{R} \to \mathbb{R}$ daną wzorem

$$\varphi_\mu(t) = \int_{-\infty}^{+\infty} e^{itx}\mu(dx).$$

7

Spacery Losowe

Literatura

Ame75. Letters to the editor. *The American Statistician*, 29(1):67–71, 1975.