

Rachunek Prawdopodobieństwa 1

– notatki z wykładu –

(wersja robocza)

dr Marcin Pitera
Uniwersytet Jagielloński

23 lutego 2022

Spis treści

Wstęp	3
1 Przestrzeń probabilistyczna	3
1.1 Definicja przestrzeni probabilistycznej	4
1.2 Podstawowe własności przestrzeni probabilistycznych	5
1.3 Przykłady dyskretnych przestrzeni probabilistycznych	6
1.4 Przestrzeń probabilistyczna na podzbiórze \mathbb{R}^n i prawdopodobieństwo geometryczne	11
1.5 Iloczyn kartezjański przestrzeni probabilistycznych	13
2 Prawdopodobieństwo warunkowe i niezależność	15
2.1 Prawdopodobieństwo warunkowe	15
2.2 Niezależność zdarzeń	16
2.3 Lemat Borela-Cantelliego	18
3 Zmienne losowe i ich rozkłady	20
3.1 Definicja zmiennej losowej i wektora losowego	20
3.2 Rozkład prawdopodobieństwa i σ -algebra generowana przez zmienną losową	22
3.3 Rozkłady ciągłe i gęstość prawdopodobieństwa	23
3.4 Rozkłady dyskretne i masa prawdopodobieństwa	25
3.5 Dystrybuanta zmiennej losowej	26
3.6 Związek między dystrybuantą, a gęstością	28
4 Wektory losowe i zależność między zmiennymi losowymi	30
4.1 Dystrybuanta wielowymiarowa	30
4.2 Dystrybuanty brzegowe i twierdzenie Sklara	32
4.3 Niezależność zmiennych losowych	33
5 Wartość oczekiwana i inne charakterystyki zmiennych losowych	36
5.1 Wartość oczekiwana i jej podstawowe własności	36
5.2 Wariancja i odchylenie standardowe	40
5.3 Kowariancja i korelacja	41
5.4 Momenty i inne charakterystyki	43
5.5 Nierówności probabilistyczne wykorzystujące momenty	45

6	Przegląd i własności wybranych rozkładów	48
6.1	Rozkłady dyskretne	48
6.2	Rozkłady ciągłe	51
6.3	Podstawowe własności rozkładu normalnego	55
6.4	Wybrane własności i zależności między rozkładami	57
7	Prawa wielkich liczb i centralne twierdzenie graniczne	61
7.1	Prawo 0-1 Kołmogorowa i zbieżność szeregów zmiennych losowych	61
7.2	Prawa wielkich liczb	63
7.3	Centralne Twierdzenie Graniczne (CTG)	67
8	Zbieżność zmiennych losowych i ich rozkładów	72
8.1	Rodzaje zbieżności zmiennych losowych	72
8.2	Zależność między różnymi typami zbieżności	73
8.3	Przykłady	77
8.4	Własności i różne charakteryzacje zbieżności według rozkładu	78
A	Wybrane fakty z matematyki dyskretnej i kombinatoryki	84
B	Wybrane fakty z analizy oraz teorii miary i całki	84
C	Notacja	85

Wstęp

Wykład ten wprowadza podstawowe narzędzia i metody rachunku prawdopodobieństwa. Materiał wyłożony na wykładzie stanowi fundament wielu dziedzin matematyki stosowanej i finansowej. W szczególności, znajomość materiału prezentowanego na tym kursie jest niezbędna do studiowania wielu przedmiotów na studiach matematycznych. Dotyczy to m.in. statystyki, procesów stochastycznych, czy ekonometrii.

Kurs ten zakłada znajomość podstaw teorii mnogości, analizy matematycznej funkcji jednej zmiennej, algebry liniowej z geometrią i topologią. Zakładamy, że w trakcie trwania wykładu studenci uczestniczą równolegle w zajęciach związanych z analizą matematyczną funkcji wielu zmiennych, czy teorią miary i całki. Dzięki temu w miarę upływu czasu będziemy w naszym kursie korzystać z powyższego faktu używając poznanych na tych kursach pojęć i metod. W związku z tym zdecydowaliśmy się pominąć niektóre dowody twierdzeń stanowiących podstawę rachunku prawdopodobieństwa, gdyż są one omawiane (i dowodzone) w trakcie wymienionych kursów.

Kurs ten można traktować jako kontynuację prowadzonego przez wiele lat na Uniwersytecie Jagiellońskim autorskiego wykładu prof. dra hab. Jerzego Ombacha. Z kursu tego został zaczerpnięty m.in. układ treści, redakcja wybranych dowodów, przykłady, czy wypowiedzi twierdzeń. Wykład ten został (w nieznacznie zmienionej formie) opublikowany w 2018 r. w [Omb19]. Innym dobrym (alternatywnym) źródłem wiedzy, z którego również częściowo korzystaliśmy w trakcie przygotowywania tego kursu, są dwa podręczniki wprowadzające do teorii rachunku prawdopodobieństwa napisane przez prof. dra hab. Jacka Jakubowskiego oraz dra Rafała Sztencła, tj. [JS04] oraz [JS17]. Bardziej dociekliwym studentom polecamy również książki [Bil12] oraz [ADD00].

Uwaga: Głównym celem tego wykładu jest wyłożenie teoretycznej części podstaw teorii prawdopodobieństwa. Mając jednak na uwadze, że jest to wykład wstępny, w kursie tym stosunkowo dużo miejsca poświęciliśmy na prezentację przykładów, mających na celu wyrobienie intuicji probabilistycznych. Przykłady te powinny być uzupełnione o zadania przerobione na ćwiczeniach. Zachęcamy również do spojrzenia na wybrane książki (zbiory zadań) w których można znaleźć wiele ciekawych zadań. Oprócz wyżej wymienionych pozycji, zadania z rachunku prawdopodobieństwa można znaleźć np. w [GS01a] lub [CZ13].

1 Przestrzeń probabilistyczna

Pojęciem stojącym u podstaw teorii prawdopodobieństwa jest tzw. *przestrzeń probabilistyczna* oznaczana często przez $(\Omega, \Sigma, \mathbb{P})$. Trójka obiektów przestrzeni probabilistycznej to: Ω – **zbiór zdarzeń elementarnych**, czyli pewien z góry ustalony zbiór (przestrzeń), Σ – **zbiór zdarzeń (mierzalnych)**, czyli zbiór podzbiorów Ω , które możemy zmierzyć oraz \mathbb{P} – **prawdopodobieństwo**, czyli funkcja przypisująca zdarzeniom mierzalnym z Σ jakąś wartość liczbową. Warto tutaj podkreślić, że *zdarzeniem* nazywamy wyłącznie podzbiory Ω które możemy zmierzyć, tzn. te które należą do Σ . W szczególności, zdarzenia elementarne (jednoelementowe zbiory $\omega \in \Omega$) nie muszą zbyć zdarzeniami. Zanim formalnie wprowadzimy definicję przestrzeni mierzalnej oraz przestrzeni probabilistycznej, przedstawmy prosty przykład, zob. Przykład 1.1.

Przykład 1.1 (Przykład prostej przestrzeni probabilistycznej). Rozważmy rzut standardową monetą. Zbiorem zdarzeń elementarnych może być tutaj zbiór $\Omega = \{O, R\}$. Przestrzeń zdarzeń mierzalnych to $\Sigma = \mathcal{P}(\Omega) = \{\emptyset, \{O, R\}, \{O\}, \{R\}\}$. Zdrowy rozsądek podpowiada nam, że miara (prawdopodobieństwo) zdarzeń, które możemy zmierzyć powinno wynosić $\mathbb{P}(\{O\}) = \frac{1}{2}$, $\mathbb{P}(\{R\}) = \frac{1}{2}$, $\mathbb{P}(\emptyset) = 0$ oraz $\mathbb{P}(\{O, R\}) = 1$. Możemy to zapisać zbiorczo wzorem $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$.

1.1 Definicja przestrzeni probabilistycznej

Oczywiście, aby ogólna definicja przestrzeni probabilistycznej miała sens należy najpierw nałożyć pewne ograniczenia na rodzinę zbiorów mierzalnych, a następnie na miarę. Zaczniemy od zdefiniowania pojęcia σ -algebry, która definiuje nam rodziny podzbiorów zadanego zbioru Ω , które mają dobrą strukturę.¹

Definicja 1.2 (σ -algebra i przestrzeń mierzalna). Niech Ω będzie przestrzenią (zbiorem). Rodzinę Σ podzbiorów Ω nazywamy **σ -algebrą** (lub σ -ciałem), jeżeli

- 1) (Zbiór Ω należy do Σ) $\Omega \in \Sigma$;
- 2) (Dopełnienie zbioru z Σ jest w Σ) $A \in \Sigma \implies \Omega \setminus A \in \Sigma$;
- 3) (Suma przeliczalna zbiorów z Σ jest w Σ) $A_1, A_2, \dots \in \Sigma \implies \bigcup_{i=1}^{\infty} A_i \in \Sigma$.

Parę (Ω, Σ) złożoną ze zbioru Ω i określonej na niej σ -algebry Σ nazywa się **przestrzenią mierzalną**.

Znając zbiór zdarzeń elementarnych Ω oraz rodzinę zdarzeń Σ które możemy zmierzyć, chcielibyśmy przypisać każdemu zdarzeniu pewną (nieujemną) wartość liczbową odpowiadającą prawdopodobieństwu zdarzenia.

Definicja 1.3 (Miara probabilistyczna). Niech dana będzie przestrzeń mierzalna (Ω, Σ) oraz funkcja $\mathbb{P}: \Sigma \rightarrow \mathbb{R}$. Funkcję \mathbb{P} nazywamy **miarą probabilistyczną** (prawdopodobieństwem), gdy spełnia ona następujące własności:

- 1) (Nieujemność) $\mathbb{P}(A) \geq 0$, dla dowolnego $A \in \Sigma$;
- 2) (Unormowanie) $\mathbb{P}(\Omega) = 1$;
- 3) (Przeliczalna addytywność) Jeżeli $A_1, A_2, \dots \in \Sigma$ są parami rozłączne, to

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Łącząc Definicję 1.2 z Definicją 1.3 dostajemy definicję przestrzeni probabilistycznej.

Definicja 1.4 (Przestrzeń probabilistyczna). Trójkę $(\Omega, \Sigma, \mathbb{P})$, w której (Ω, Σ) jest przestrzenią mierzalną, a \mathbb{P} określoną na niej miarą probabilistyczną (prawdopodobieństwem) nazywamy **przestrzenią probabilistyczną**.

Na koniec warto jeszcze wspomnieć, że zazwyczaj σ -algebrę (i określoną na niej miarę probabilistyczną) definiuje się wychodząc od pewnej ustalonej rodziny podzbiorów Ω .

Definicja 1.5 (Najmniejsza σ -algebra). Niech Ω będzie przestrzenią, a $\mathcal{F} \subset \mathcal{P}(\Omega)$ ustaloną rodziną jej podzbiorów. Wtedy **najmniejszą σ -algebrę zawierającą w sobie rodzinę \mathcal{F}** oznaczamy przez $\sigma(\mathcal{F})$.

Własności najmniejszych σ -algebr będą omówione na teorii miary i całki. Zauważmy tylko, że $\sigma(\mathcal{F})$ jest iloczynem wszystkich σ -algebr zawierających \mathcal{F} .

¹Więcej informacji na temat σ -algebr będzie podanych na wykładzie z teorii miary i całki.

1.2 Podstawowe własności przestrzeni probabilistycznych

Mając zadaną przestrzeń probabilistyczną, możemy się zastanowić, jakie własności musi ona spełniać. Zbiór (wybranych) podstawowych własności przedstawiony jest w Propozycji 1.6 oraz Propozycji 1.7.

Propozycja 1.6 (Podstawowe własności prawdopodobieństwa – ogólne). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną. Wtedy zachodzą następujące własności:

- 1) $\emptyset \in \Sigma$ oraz $\mathbb{P}(\emptyset) = 0$;
- 2) $A_1, A_2, \dots \in \Sigma \implies \bigcap_{i=1}^{\infty} A_i \in \Sigma$
- 3) $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ dla parami rozłącznych zdarzeń A_1, \dots, A_n , $n \in \mathbb{N}$.
- 4) Jeżeli $A \subset B$, to $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.
- 5) $\mathbb{P}(A) = 1 - \mathbb{P}(\Omega \setminus A)$.
- 6) Jeżeli $A \subset B$, to $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- 7) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Dowód.

- 1) Pierwsza część wynika bezpośrednio z 1) i 2) w Definicji 1.2. Z przeliczalnej addytywności miary probabilistycznej dostajemy $\mathbb{P}[\emptyset] = \mathbb{P}(\bigcup_{i=1}^{\infty} \emptyset) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$, co implikuje $\mathbb{P}[\emptyset] = 0$.
- 2) Wynika to wprost z praw de Morgana (ćw. do domu).
- 3) Skończona addytywność wynika z przeliczalnej addytywności – wystarczy w definicji przyjąć $A_k = \emptyset$ dla $k > n$.
- 4) Wystarczy zauważyć, że $B = A \cup (B \setminus A)$, $A \cap (B \setminus A) = \emptyset$, oraz skorzystać z przeliczalnej addytywności.
- 5) Wynika wprost z 4) (ćw. do domu).
- 6) Wynika wprost z 4) (ćw. do domu).
- 7) Wynika wprost z 3); wystarczy $A \cup B$ przedstawić jako sumę rozłącznych zbiorów $A \setminus B$, $B \setminus A$ oraz $A \cap B$.

□

Propozycja 1.7 (Podstawowe własności – ciągi zdarzeń). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną oraz niech $A_1, A_2, \dots \in \Sigma$ będzie dowolnym ciągiem zdarzeń. Wtedy zachodzą następujące własności:

- 1) $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.
- 2) Jeżeli $\mathbb{P}(A_i) = 0$, dla $i \in \mathbb{N}$, to $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = 0$.
- 3) Jeżeli $\mathbb{P}(A_i) = 1$, dla $i \in \mathbb{N}$, to $\mathbb{P}(\bigcap_{i=1}^{\infty} A_i) = 1$.
- 4) Jeżeli $A_1 \subset A_2 \subset A_3 \subset \dots$, to $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_{i=1}^{\infty} A_i)$.
- 5) Jeżeli $A_1 \supset A_2 \supset A_3 \supset \dots$, to $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_{i=1}^{\infty} A_i)$.

Dowód.

1) Definiujemy rodzinę $(B_i)_{i \in \mathbb{N}}$ daną przez

$$B_i := \begin{cases} A_1 & i = 1, \\ A_i \setminus (A_1 \cup A_2 \cup \dots \cup A_{i-1}) & i \neq 1. \end{cases}$$

Łatwo zauważyć, że $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ oraz zdarzenia z rodziny $(B_i)_{i \in \mathbb{N}}$ są parami rozłączne. Łącząc to z faktem, że dla $i \in \mathbb{N}$ zachodzi $B_i \subset A_i$, dostajemy

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

2) Wynika bezpośrednio z 1) (ćw. do domu).

3) Wynika bezpośrednio z 1) oraz praw de Morgana (ćw. do domu).

4) Niech $A_0 = \emptyset$. Zakładając $A_1 \subset A_2 \subset A_3 \subset \dots$ i rozumując podobnie jak w 1) dostajemy

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \setminus A_{i-1})\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \setminus A_{i-1}) = \lim_{n \rightarrow \infty} \left[\sum_{i=1}^n \mathbb{P}(A_i \setminus A_{i-1}) \right] = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

5) Wynika z 4) i praw de Morgana (ćw. do domu). □

1.3 Przykłady dyskretnej przestrzeni probabilistycznych

W tym podrozdziale przedstawimy przykłady dyskretnej przestrzeni probabilistycznych.

1.3.1 Schemat klasyczny

Najprostszym przykładem przestrzeni probabilistycznej jest tzw. *schemat klasyczny* w którym rozważamy skończoną liczbę zdarzeń elementarnych wraz z ich zbiorem potęgowym (zbiorem wszystkich podzbiorów), w którym zakładamy, że każde zdarzenie jest równoprawdopodobne.

Definicja 1.8 (Schemat klasyczny). Niech $n \in \mathbb{N}$ oraz niech $\Omega = \{\omega_1, \dots, \omega_n\}$. Wtedy **schematem klasycznym** nazywamy przestrzeń probabilistyczną $(\Omega, \Sigma, \mathbb{P})$ dla której $\Sigma = \mathcal{P}(\Omega)$ oraz miara \mathbb{P} jest zadana przez

$$\mathbb{P}(A) = \frac{\#A}{n}, \quad A \in \mathcal{P}(\Omega). \quad (1.1)$$

Łatwo sprawdzić, że trójka $(\Omega, \Sigma, \mathbb{P})$ zadana w Definicji 1.8 w istocie tworzy przestrzeń probabilistyczną (ćw. do domu). Miara zadana w (1.4) zakłada, że każdy element ze zbioru Ω występuje z równym prawdopodobieństwem, tzn.

$$\mathbb{P}(\{\omega_1\}) = \dots = \mathbb{P}(\{\omega_n\}) = \frac{1}{n},$$

a $\#A$ odpowiada ilości tzw. *sprzyjających zdarzeń elementarnych*, tzn. zdarzeń elementarnych $\omega \in \Omega$, które należą do A . Na dobrą sprawę zamiast warunku (1.4) wystarczyłoby zadać miarę na zdarzeniach $A_i = \{\omega_i\}$ – każdy inny zbiór z $\mathcal{P}(\Omega)$ można przedstawić jako sumę zbiorów $(A_i)_{i=1}^n$, a następnie skorzystać ze skończonej addytywności miary. Warto również zauważyć, że mamy tutaj tzw. *pełną informację* o eksperymencie, tzn. jesteśmy w stanie zmierzyć prawdopodobieństwo wystąpienia

każdego ze zdarzeń elementarnych; w schemacie klasycznym Σ jest najmniejszą σ -algebrą, która zawiera w sobie wszystkie zdarzenia elementarne. Często zamiast używać zapisu $\Sigma = \mathcal{P}(\Omega)$ będziemy pisać $\Sigma = \sigma(\{\omega_1\}, \dots, \{\omega_n\})$. Zbiór podzbiorów $\{\{\omega_1\}, \dots, \{\omega_n\}\}$ w pewnym sensie wyznacza nam tutaj tzw. *atomy* naszej σ -algebry, czyli rozłączne podzbiory Ω z których możemy odtworzyć całą Σ . Liczba elementów Σ w schemacie klasycznym równa jest 2^n (liczba n -wyrazowych wariacji zbioru $\{0, 1\}$ w którym każda wariacja identyfikuje, które atomy należą do powiązanego zbioru).

Przykład 1.9 (Dwukrotny rzut kostką symetryczną z pełną informacją). Załóżmy, że chcemy stworzyć model dwukrotnego rzutu symetryczną kostką w schemacie klasycznym. Przykładową przestrzenią probabilistyczną może być tutaj trójka $(\Omega, \Sigma, \mathbb{P})$, gdzie

$$\begin{aligned}\Omega &:= \{(i, j), \text{ dla } i, j = 1, 2, 3, 4, 5, 6\}, \\ \Sigma &:= \mathcal{P}(\Omega), \\ \mathbb{P}(A) &:= \frac{\#A}{36}, \quad A \in \mathcal{P}(\Omega).\end{aligned}$$

Aby obliczyć prawdopodobieństwo, że w obu rzutach wypadła parzysta liczba oczek, wystarczy obliczyć tzw. ilość elementarnych zdarzeń sprzyjających. Dla $B := \{(i, j) \in \Omega : i, j \in \{2, 4, 6\}\}$ dostajemy $\mathbb{P}(B) = \frac{\#B}{36} = \frac{3 \cdot 3}{36} = \frac{1}{4}$.

Przykład 1.10 (Losowanie karty). Załóżmy, że chcemy stworzyć model wylosowania jednej karty z 24 elementowej talii kart. Przykładowymi dwoma przestrzeniami probabilistyczną mogą być tutaj $(\Omega, \Sigma, \mathbb{P})$ oraz $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}})$, gdzie

$$\begin{aligned}\Omega &:= \{1, \dots, 24\}, & \tilde{\Omega} &:= \{(i, j) : i \in \{\clubsuit, \diamond, \heartsuit, \spadesuit\}, j \in \{9, 10, J, Q, K, A\}\} \\ \Sigma &:= \mathcal{P}(\Omega), & \tilde{\Sigma} &:= \mathcal{P}(\tilde{\Omega}), \\ \mathbb{P}(A) &:= \frac{\#A}{24}, \quad A \in \mathcal{P}(\Omega) & \tilde{\mathbb{P}}(\tilde{A}) &:= \frac{\#\tilde{A}}{24}, \quad \tilde{A} \in \mathcal{P}(\tilde{\Omega}).\end{aligned}$$

W pierwszym przypadku mamy przestrzeń 24 elementów w którym każdy odpowiada jakiejś karcie. W drugim przypadku wprowadzamy strukturę w elementach, aby łatwiej liczyć prawdopodobieństwo interesujących nas zdarzeń. Warto tutaj zauważyć dwie rzeczy: (1) istnieje wiele przestrzeni probabilistycznych, które mogą opisywać ten sam eksperyment; (2) dobranie właściwej przestrzeni probabilistycznej może nam ułatwić liczenie prawdopodobieństwa interesujących nas zdarzeń. Przykładowo, zdarzenie wylosowanie trefla łatwiej opisać w drugiej przestrzeni definiując zbiór $\tilde{B} := \{(i, j) \in \tilde{\Omega} : i = \clubsuit\}$; w pierwszej przestrzeni musielibyśmy zidentyfikować elementy odpowiadające kartom trefl i je ręcznie wypisać. Przykładowo, jeżeli karty są ponumerowane według kolorów zaczynając od kart treflowych, to dostalibyśmy $B := \{1, 2, 3, 4, 5, 6\}$.

1.3.2 Schemat klasyczny z niepełną informacją

W poprzednim rozdziale zakładaliśmy, że σ -algebra jest zbiorem potęgowym Ω i jesteśmy w stanie policzyć prawdopodobieństwo każdego podzbioru Ω . Oczywiście tak być nie musi, i σ -algebra może być istotnie mniejsza od $\mathcal{P}(\Omega)$.

Definicja 1.11 (Schemat klasyczny z niepełną informacją). Niech $n \in \mathbb{N}$ oraz niech $\Omega = \{\omega_1, \dots, \omega_n\}$. Wtedy **schematem klasycznym z niepełną informacją** nazywamy przestrzeń probabilistyczną $(\Omega, \Sigma, \mathbb{P})$ dla której $\Sigma \subset \mathcal{P}(\Omega)$ oraz miara \mathbb{P} jest zadana przez

$$\mathbb{P}(A) = \frac{\#A}{n}, \quad A \in \Sigma. \quad (1.2)$$

Gdy mamy do czynienia z niepełną informacją, to σ -algebrę najłatwiej jest opisać używając *atomów* ją budujących, tzn. rozłącznych podzbiorów Ω z których możemy odtworzyć całą Σ . Załóżmy, że A_1, \dots, A_k , gdzie $k < n$, to podzbiory Ω spełniające warunki $\bigcup_{i=1}^k A_i = \Omega$ oraz $A_i \cap A_j = \emptyset$, gdy $i \neq j$. Wtedy możemy zdefiniować Σ jako

$$\Sigma := \{A : A = \bigcup_{i \in I} A_i, \text{ gdzie } I = \mathcal{P}(\{1, \dots, k\})\}.$$

Przykład 1.12 (Rzut monetą bez informacji o wyniku). Załóżmy, że chcemy stworzyć model rzutu monetą bez żadnych informacji o wyniku. Przykładową przestrzenią probabilistyczną może być tutaj $(\Omega, \Sigma, \mathbb{P})$, gdzie $\Omega := \{0, 1\}$, $\Sigma := \{\Omega, \emptyset\}$, $\mathbb{P}(\Omega) = 1 - \mathbb{P}(\emptyset) = 1$. Nie mając informacji o rzucie (czyli np. o tym czy moneta nie jest sfałszowana) przestrzeń jest trywialna, Ω jest jedynym atomem, a liczba elementów Σ to 2^1 - możemy policzyć tylko prawdopodobieństwo zdarzenie pewnego i pustego.

Przykład 1.13 (Trzy rzuty monetą z informacją, czy wypadł orzeł). Załóżmy, że chcemy stworzyć model trzykrotnego rzutu monetą mając daną tylko informację, czy w którymś rzucie wypadł orzeł. Możemy zdefiniować przykładową przestrzenią probabilistyczną $(\Omega, \Sigma, \mathbb{P})$, gdzie

$$\begin{aligned} \Omega &:= \{(i, j, k), \text{ dla } i, j, k \in \{O, R\}\}, \\ \Sigma &:= \sigma(A_1, \Omega \setminus A_1), \text{ gdzie } A_1 = \{(i, j, k) \in \Omega : i = O \vee j = O \vee k = O\} \\ \mathbb{P}(A) &:= \frac{\#A}{8}, \quad A \in \Sigma. \end{aligned}$$

Mamy tutaj dwa atomy, a liczba elementów Σ to 2^2 , które odpowiadają zdarzeniom wypadnięcia orła, niewypadnięcia orła, zdarzeniu pewnemu i pustemu. Łatwo policzyć $\mathbb{P}(A_1) = 1 - \mathbb{P}(\Omega \setminus A_1) = 1 - \frac{1}{8} = \frac{7}{8}$. W tej przestrzeni nie jesteśmy w stanie policzyć np. prawdopodobieństwa, czy w pierwszym rzucie wypadł orzeł, gdyż powiązany zbiór zdarzeń elementarnych nie jest mierzalny.

Przykład 1.14 (Dwukrotny rzut kostką z informacją o sumie oczek). Załóżmy, że chcemy stworzyć model dwukrotnego rzutu symetryczną kostką z informacją o sumie oczek. Przypominając, że $\sigma(A_1, \dots, A_k)$ oznacza najmniejszą σ -algebrę zawierającą w sobie zbiory A_1, \dots, A_k , możemy zdefiniować przykładową przestrzenią probabilistyczną $(\Omega, \Sigma, \mathbb{P})$, gdzie

$$\begin{aligned} \Omega &:= \{(i, j), \text{ dla } i, j = 1, 2, 3, 4, 5, 6\}, \\ \Sigma &:= \sigma(A_k, k = 2, \dots, 12), \text{ gdzie } A_k := \{(i, j) \in \Omega : i + j = k\} \\ \mathbb{P}(A) &:= \frac{\#A}{36}, \quad A \in \Sigma. \end{aligned}$$

Mamy tutaj 11 atomów, liczba elementów Σ to 2^{11} , co jest istotnie mniejsze od liczby elementów zbioru potęgowego, tzn. 2^{36} . W tak zadanym eksperymencie, jesteśmy w stanie policzyć np. prawdopodobieństwo zdarzenia, że suma liczby oczek jest mniejsza od 4. Dla $B = A_2 \cup A_3 \cup A_4$ dostajemy

$$\mathbb{P}(B) = \mathbb{P}(\{(1, 1)\} \cup \{(1, 2), (2, 1)\} \cup \{(1, 3), (3, 1), (2, 2)\}) = \frac{6}{36} = \frac{1}{6}.$$

Warto zauważyć, że przy tak zdefiniowanej σ -algebrze nie jesteśmy w stanie policzyć np. prawdopodobieństwa, że w pierwszym rzucie wypadła jedynka, gdyż zdarzenie $B = \{(1, 1), (1, 2), \dots, (1, 6)\}$ nie należy do Σ .

Uwaga 1.15 (Niepełna informacja, a prawdopodobieństwo zdarzeń). Niepełna informacja odnosi się do tego, jakie informacje o modelu otrzymujemy. W szczególności w schemacie z pełną informacją jesteśmy (również) w stanie policzyć prawdopodobieństwa zbiorów opisanych w poprzednich przykładach. Co więcej, z praktycznego punktu widzenia do obliczania prawdopodobieństw lepiej używać schematu klasycznego (o ile to tylko możliwe) zamiast schematu z niepełną informacją.

1.3.3 Schemat dyskretny skończony

W schemacie klasycznym miara probabilistyczna zliczała nam ilość zdarzeń elementarnych w danym zbiorze. Oczywiście zdarzenia elementarne nie muszą mieć równego prawdopodobieństwa, a miarę można zadać w sposób bardziej ogólny.

Definicja 1.16 (Schemat dyskretny). Niech $n \in \mathbb{N}$ oraz niech $\Omega = \{\omega_1, \dots, \omega_n\}$. Wtedy **schematem dyskretnym skończonym** nazywamy przestrzeń probabilistyczną $(\Omega, \Sigma, \mathbb{P})$ dla której $\Sigma = \mathcal{P}(\Omega)$ oraz miara \mathbb{P} jest zadana przez $\mathbb{P}(\{\omega_i\}) = p_i$, tj.

$$\mathbb{P}(A) = \sum_{i: \omega_i \in A} p_i, \quad A \in \Sigma. \quad (1.3)$$

gdzie $(p_i)_{i=1}^n$ to ciąg nieujemnych liczb rzeczywistych spełniających równość $\sum_{i=1}^n p_i = 1$,

W Definicji 1.16 wystarczy zadać miarę tylko na zdarzeniach elementarnych (które są tutaj atomami) – ze skończonej addytywności jesteśmy w stanie podać prawdopodobieństwo zajścia dowolnego zbioru z Σ , tj. dla $A \in \Sigma$ dostajemy

$$\mathbb{P}(A) = \mathbb{P}(\{\omega_i \in \Omega : \omega_i \in A\}) = \sum_{i: \omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{i: \omega_i \in A} p_i.$$

Oczywiście, podobnie jak w przypadku klasycznym, można zdefiniować schemat dyskretny z niepełną informacją, ale nie będziemy tutaj tego robić. Warto też zauważyć, że schemat dyskretny uogólnia nam schemat klasyczny: dla $n \in \mathbb{N}$ i $p_i = \frac{1}{n}$ schemat dyskretny jest schematem klasycznym.

Przykład 1.17 (Rzut fałszywą monetą). Załóżmy, że chcemy stworzyć model rzutu sfalszowaną monetą, w której prawdopodobieństwo wypadnięcia orła wynosi $\frac{3}{4}$. Eksperyment ten opisuje przestrzeń $(\Omega, \Sigma, \mathbb{P})$, gdzie $\Omega := \{O, R\}$, $\Sigma := \mathcal{P}(\Omega)$, a miara zadana jest przez $\mathbb{P}(\{O\}) = \frac{3}{4}$ i $\mathbb{P}(\{R\}) = \frac{1}{4}$.

Przykład 1.18 (Trzy rzuty monetą z informacją, czy wypadł orzeł). Załóżmy, że chcemy stworzyć model trzykrotnego rzutu monetą z informacją, czy w którymś rzucie wypadł orzeł. Zamiast definiować przestrzeń z niepełną informacją, jak w Przykładzie 1.13, możemy zdefiniować uproszczoną przestrzeń probabilistyczną $(\Omega, \Sigma, \mathbb{P})$ zadaną przez

$$\begin{aligned} \Omega &:= \{0, 1\}, \\ \Sigma &:= \mathcal{P}(\Omega) \\ \mathbb{P}(\{0\}) &:= \frac{7}{8} = 1 - \mathbb{P}(\{1\}). \end{aligned}$$

Prawdopodobieństwo zdarzeń elementarnych jest tutaj różne, tzn. $\mathbb{P}(\{0\}) \neq \mathbb{P}(\{1\})$, więc nie jest to schemat klasyczny.

Przykład 1.19 (Dwukrotny rzut kostką z informacją o sumie oczek). Podobnie jak w poprzednim przykładzie, można zmodyfikować przestrzeń probabilistyczną z Przykładu 1.14, tak aby było w niej 11 nierównoprawdopodobnych zdarzeń elementarnych (ćw. do domu).

1.3.4 Schemat dyskretny (nieskończony)

W poprzednich przypadkach zakładaliśmy, że ilość elementów w zbiorze Ω musi być skończona. Tak oczywiście być nie musi. Najprostszy przykład w którym zbiór Ω jest nieskończony stanowi dyskretny schemat nieskończony.

Definicja 1.20 (Schemat dyskretny nieskończony). Niech $\Omega = \{\omega_1, \omega_2, \dots\}$. Wtedy **schematem dyskretnym (nieskończonym)** nazywamy przestrzeń probabilistyczną $(\Omega, \Sigma, \mathbb{P})$ dla której $\Sigma = \mathcal{P}(\Omega)$ oraz miara \mathbb{P} jest zadana przez $\mathbb{P}(\{\omega_i\}) = p_i$, tj.

$$\mathbb{P}(A) = \sum_{i: \omega_i \in A} p_i, \quad A \in \Sigma. \quad (1.4)$$

gdzie $(p_i)_{i=1}^{\infty}$ to ciąg nieujemnych liczb rzeczywistych spełniających równość $\sum_{i=1}^{\infty} p_i = 1$.

Oczywiście, schemat nieskończony jest w pewnym sensie uogólnieniem schematu skończonego - dla $i > n$ możemy zadać $p_i = 0$, co odtworzy nam schemat dyskretny; elementy o prawdopodobieństwie zero w pewnym sensie nie są wtedy istotne.

Przykład 1.21 (Seria rzutów kostką). Załóżmy, że chcemy stworzyć model w którym rzucamy kostką aż do uzyskania liczby oczek równej sześć. Teoretycznie, liczba tych rzutów może być dowolnie duża, co sugeruje użycie schematu dyskretnego nieskończonego. Odpowiada temu przestrzeń $(\Omega, \Sigma, \mathbb{P})$ zadana przez $\Omega = \mathbb{N}$, $\Sigma = \mathcal{P}(\Omega)$ oraz $\mathbb{P}(\{i\}) = p_i = \frac{1}{6} \left(\frac{5}{6}\right)^{i-1}$. Oczywiście $\sum_{i=1}^{\infty} p_i = \frac{1}{6} \cdot \frac{1}{1-\frac{5}{6}} = 1$.

Warto zauważyć, że wartość $\mathbb{P}(\{i\}) = p_i$ można uzyskać stosując schemat klasyczny, tzn. dla przestrzeni określającej i -rzutów kostką szansa, że 6 wypadnie po raz pierwszy w i -tym rzucie wynosi $\frac{1}{6} \left(\frac{5}{6}\right)^{i-1}$.

1.3.5 Przestrzenie probabilistyczne związane z problemami kombinatorycznymi

Wiele problemów probabilistycznych sprowadza się do losowania (jednego lub wielu) elementów z zadanych zbiorów. Przydają się tutaj podstawowe wzory kombinatoryczne, które ułatwią nam określenie prawdopodobieństwa na danej przestrzeni, zob. Dodatek A. Przykładowo, jeżeli interesuje nas **losowanie n elementów z N elementowego zbioru X** , to przestrzeń probabilistyczną można określić jako

$$\Omega := X^n, \quad \Sigma := \mathcal{P}(\Omega), \quad \mathbb{P}(A) := \frac{\#A}{\#\Omega} = \frac{\#A}{N^n}.$$

W przypadku **losowania n elementów bez zwracania z zbioru N elementowego X** możemy określić

$$\Omega := \{A \subset X: \#A = n\}, \quad \Sigma := \mathcal{P}(\Omega), \quad \mathbb{P}(A) := \frac{\#A}{\#\Omega} = \frac{\#A}{\binom{N}{n}}.$$

Przykład 1.22 (Losowanie kul ze zwracaniem). Załóżmy, że mamy urnę z kulą białą i czarną. Z urny losujemy dwa razy kulę ze zwracaniem. Chcemy odpowiedzieć na pytanie: *Jakie jest prawdopodobieństwo, że wylosujemy co najmniej raz kulę białą?* Definiujemy $\Omega = \{C, B\}^2$ i używając schematu klasycznego obliczamy prawdopodobieństwo zdarzenia sprzyjającego dostając

$$\mathbb{P}(\{(x, y) \in \Omega: x = B \vee y = B\}) = \mathbb{P}(\{(B, B)\} \cup \{(B, C)\} \cup \{(C, B)\}) = \frac{3}{4}.$$

Gdybyśmy zdecydowali się nie rozróżniać kolejności losowania, to moglibyśmy zdefiniować $\tilde{\Omega} = \{(B, C), (B, B), (C, C)\}$, gdzie każde zdarzenie odpowiada wynikowi losowania (bez uwzględnienia kolejności). W takim wypadku użycie schematu klasycznego nie byłoby poprawne, gdyż prawdopodobieństwo wylosowanie kul obu kolorów jest (dwukrotnie) większe niż prawdopodobieństwo wylosowania kul o jednym (zadany kolorze), tzn. $\tilde{\mathbb{P}}(\{(B, C)\}) = 2\tilde{\mathbb{P}}(\{(B, B)\}) = 2\tilde{\mathbb{P}}(\{(C, C)\})$.

Więcej przykładów związanych z losowaniem zostanie przedstawionych na ćwiczeniach.

1.4 Przestrzeń probabilistyczna na podzbiorze \mathbb{R}^n i prawdopodobieństwo geometryczne

Oprócz przestrzeni dyskretnej, możemy również badać bardziej skomplikowane struktury opisane na przykład na podzbiorze \mathbb{R}^n . O ile zdefiniowanie zbioru zdarzeń elementarnych $\Omega \subset \mathbb{R}^n$ nie jest trudne, o tyle właściwe zadanie σ -algebry oraz miary wymaga znajomości podstaw teorii miary i całki. Opiszemy teraz pokrótce jak to zrobić w klasycznym przypadku; więcej informacji na ten temat będzie podanych na wykładzie z teorii miary i całki.

Jak już wspomnieliśmy, zazwyczaj σ -algebrę zadaje się wychodząc od tzw. zbiorów generujących na \mathbb{R}^n . W szczególności, jeżeli \mathcal{F} jest rodziną wszystkich zbiorów otwartych na \mathbb{R}^n , to $\sigma(\mathcal{F})$ nazywamy **σ -algebrą zbiorów borelowskich** i oznaczamy przez $\mathcal{B}(\mathbb{R}^n)$. σ -algebrę zbiorów borelowskich można zdefiniować na wiele równoważnych sposobów. Dla \mathbb{R} mamy $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F})$, gdzie \mathcal{F} jest: (1) rodziną zbiorów domkniętych; (2) rodziną przedziałów otwartych; (3) rodziną przedziałów jednostronnie domkniętych; (4) rodziną przedziałów $\{(\infty, b)\}$ dla $b \in \mathbb{R}$. Jest to o tyle ważne, że zazwyczaj miarę probabilistyczną również wystarczy zadać na zbiorach generujących, co upraszcza znacznie definicje.

Uwaga 1.23 (Zbiory nieborelowskie). O ile z praktycznego punktu widzenia struktura $\mathcal{B}(\mathbb{R}^n)$ jest wystarczająco bogata do określenia większości modeli probabilistycznych, o tyle należy pamiętać, że istnieją podzbiory \mathbb{R}^n niebędące zbiorami borelowskimi, tzn. $\mathcal{P}(\mathbb{R}^n) \setminus \mathcal{B}(\mathbb{R}^n) \neq \emptyset$. Więcej informacji na ten temat zostanie podanych na wykładzie z teorii miary i całki.

Mając daną przestrzeń mierzalną $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ możemy na niej określić miarę. Najczęstszym wyborem jest tzw. miara Lebesgue'a, oznaczana przez \mathcal{L}_n , która w naturalny sposób uogólnia pojęcia długości w \mathbb{R} , pola w \mathbb{R}^2 , objętości w \mathbb{R}^3 , itd. Dla \mathbb{R} i zbiorów generujących postaci $A = (a, b)$ miarę zadajemy poprzez $\mathcal{L}_1(A) = |b - a|$. Oczywiście tak zdefiniowana miara nie jest unormowana (tzn. $\mathcal{L}_1(\mathbb{R}) = \infty$), więc nie jest miarą probabilistyczną. W związku z tym większość modeli probabilistycznych określa się na zwartych podzbiorach \mathbb{R} o skończonej mierze, np. definiując $\Omega = [0, 1]$ i rozważając zbiory borelowskie na Ω . W podobny sposób dla \mathbb{R}^n można zdefiniować przestrzeń $\Omega = [0, 1]^n$.

Przykład 1.24 (Rzut kostką symetryczną). Załóżmy, że chcemy stworzyć model rzutu symetryczną kostką na $\Omega = [0, 1]$. Przykładową przestrzenią probabilistyczną może być tutaj trójka $(\Omega, \Sigma, \mathbb{P})$, gdzie $\Omega = [0, 1]$, $\Sigma = \sigma([0, \frac{1}{6}), [\frac{1}{6}, \frac{2}{6}), \dots, [\frac{4}{6}, \frac{5}{6}), [\frac{5}{6}, 1])$ oraz $\mathbb{P}(A) = \mathcal{L}_1(A)$ dla $A \in \Sigma$. Zauważmy, że liczba elementów Σ to 2^6 oraz np. $\mathbb{P}([\frac{1}{6}, \frac{2}{6})) = \frac{2}{6} - \frac{1}{6} = \frac{1}{6}$, co może odpowiadać prawdopodobieństwu wypadnięcia dwójki.

Uwaga 1.25 (Standardowa przestrzeń probabilistyczna). Wszystkie rozważane przez nas modele (przykłady) można ustandaryzować i przedstawić (równoważnie) używając $\Omega = [0, 1]$ wraz z σ -algebrą będącą podzbiorem $\mathcal{B}([0, 1])$ i określoną na niej miarą Lebesgue'a. Przestrzenie tego typu nazywa się *przestrzeniami Lebesgue'a-Rokhlina* lub *standardowymi przestrzeniami probabilistycznymi*. O ile nie wszystkie przestrzenie można w ten sposób zapisać,² o tyle w praktyce są one wystarczające do modelowania większości zjawisk.

Popularnymi modelami probabilistycznymi są modele związane z tzw. **prawdopodobieństwem geometrycznym**. Niech $\Omega \in \mathcal{B}(\mathbb{R}^n)$ będzie zbiorem o skończonej mierze, tzn. $\mathcal{L}_n(\Omega) < \infty$, oraz

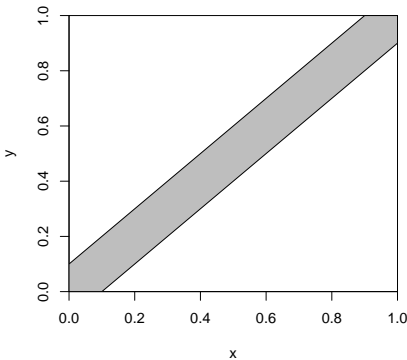
²Dokładne wytłumaczenie co oznacza tutaj standaryzacja wykracza poza materiał tego wykładu. Mówiąc skrótowo, nie wszystkie przestrzenie probabilistyczne są izomorficzne (modulo 0) ze standardową przestrzenią probabilistyczną.

niech $\Sigma = \mathcal{B}(\Omega) = \{A \subset \Omega : A \in \mathcal{B}(\mathbb{R}^n)\}$. Wtedy przez prawdopodobieństwo geometryczne rozumiemy miarę probabilistyczną zadaną przez

$$\mathbb{P}(A) := \frac{\mathcal{L}_n(A)}{\mathcal{L}_n(\Omega)}, \quad A \in \Sigma.$$

Łatwo sprawdzić, że tak zadana trójka $(\Omega, \Sigma, \mathbb{P})$ tworzy przestrzeń probabilistyczną (ćw. do domu).

Przykład 1.26 (Losowanie dwóch liczb z przedziału $(0,1)$). Załóżmy, że chcemy stworzyć model w którym losujemy niezależnie od siebie i w sposób jednostajny dwie liczby z przedziału $[0,1]$ a następnie policzyć prawdopodobieństwo, że wartość bezwzględna z różnicy tych liczb jest mniejsza niż 0.1. Definiujemy $\Omega = [0, 1]^2$, $\Sigma = \mathcal{B}(\Omega)$ oraz zadajemy prawdopodobieństwo poprzez $\mathbb{P}(A) = \mathcal{L}_2(A)$ dla $A \in \Sigma$. Zdarzeniem nas interesującym jest $B := \{(x, y) \in \Omega : |x - y| \leq 0.1\}$. Elementy B łatwo jest zobrazować na rysunku poglądowym, co ułatwia obliczenia prawdopodobieństwa (miary zbioru) B . Dostajemy



$$\mathbb{P}(B) = 1 - \mathbb{P}(\Omega \setminus B) = 1 - 2 \cdot (0.5 \cdot 0.9 \cdot 0.9) = 0.19.$$

Warto też zwrócić uwagę, że przy prawdopodobieństwie geometrycznym należy uważać, co rozumiemy przez *wybór losowy* i jaki jest jego związek np. z miarą Lebesgue'a. Ilustruje to tzw. paradoks Bertranda.

Przykład 1.27 (Paradoks Bertranda). Rozważmy następujący problem: *Na ustalonym okręgu jednostkowym skonstruowano losowo cięciwę. Jaka jest szansa, że cięciwa będzie dłuższa niż bok trójkąta równobocznego wpisanego w ten okrąg?* Problem ten moglibyśmy teoretycznie rozwiązać na trzy sposoby:

- *Sposób 1:* Długość cięciwy jest wyznaczona jednoznacznie przez kąt $\angle AOB$, gdzie A i B to wierzchołki cięciwy, a O to środek okręgu. Długość boku trójkąta równobocznego odpowiada wyborowi kątów $120^\circ = \frac{2}{3}\pi$ oraz $240^\circ = \frac{4}{3}\pi$. Definiujemy $\Omega = [0, 2\pi]$, $\Sigma = \mathcal{B}(\Omega)$ i $\mathbb{P}(A) = \mathcal{L}_1(A)/2\pi$. Zdarzeniem sprzyjającym jest tutaj zbiór $A = (\frac{2}{3}\pi, \frac{4}{3}\pi)$, którego prawdopodobieństwo wynosi $\frac{1}{3}$.
- *Sposób 2:* Narysujmy dowolny promień i rozważmy zbiór cięciw prostopadłych do tego promienia. Długość cięciwy jest wyznaczona jednoznacznie przez odległość $|OC|$, gdzie C jest punktem przecięcia cięciwy z promieniem. Definiujemy $\Omega = [0, 1]$, $\Sigma = \mathcal{B}(\Omega)$ i $\mathbb{P}(A) = \mathcal{L}_1(A)$. Długość boku trójkąta równobocznego odpowiada punktowi C dokładnie na środku promienia. Zdarzeniem sprzyjającym jest tutaj zbiór $A = [\frac{1}{2}, 1]$, którego prawdopodobieństwo wynosi $\frac{1}{2}$.
- *Sposób 3:* Wybieramy dowolny punkt w środku okręgu, który będzie odpowiadać środkowi cięciwy; każdy wybrany punkt odpowiada dokładnie jednej cięciwie. Definiujemy przestrzeń probabilistyczną jako okrąg jednostkowy z σ -algebrą zbiorów borelowskich i unormowaną miarą \mathcal{L}_2 na tym okręgu. Cięciwa jest dłuższa od boku trójkąta gdy leży wewnątrz okręgu o promieniu $\frac{1}{2}$ (okrąg wpisany w trójkąt równoboczny). Zdarzeniem sprzyjającym jest tutaj dopełnienie okręgu o promieniu $\frac{1}{2}$, którego prawdopodobieństwo wynosi $1 - (\frac{1}{2})^2 = \frac{3}{4}$.

Jak widać każdy sposób daje nam inny wynik. Istotą problemu jest tutaj nieprecyzyjne sformułowanie w pytaniu, gdyż nie wiemy co oznacza *losowa cięciwa*. Każdy z przedstawionych metod zadaje nam inny rozkład wylosowanych cięciw co prowadzi do innego wyniku.

Uwaga 1.28 (Wybór losowy). Paradoks Bertranda (Przykład 1.27) pokazuje nam jak ważne jest właściwe postawienie problemu i dobranie do niego odpowiedniej przestrzeni probabilistycznej. W szczególności, gdy mówimy o *losowaniu*, czy wybraniu czegoś *w sposób losowy* zawsze należy upewnić się co przez to rozumiemy.

W2
+
W3

1.5 Iloczyn kartezjański przestrzeni probabilistycznych

Zdefiniujemy teraz iloczyn kartezjański przestrzeni probabilistycznych. Najważniejszą częścią konstrukcji jest zbudowanie (najmniejszej) σ -algebry zawierającej w sobie zadaną rodzinę zbiorów, tzn. iloczyny zbiorów mierzalnych; jak zwykle, dla rodziny zbiorów \mathcal{F} najmniejszą σ -algebrę zawierającą rodzinę \mathcal{F} oznaczamy przez $\sigma(\mathcal{F})$.

Definicja 1.29 (Iloczyn kartezjański przestrzeni probabilistycznych). Niech $(\Omega_1, \Sigma_1, \mathbb{P}_1)$ oraz $(\Omega_2, \Sigma_2, \mathbb{P}_2)$ będą przestrzeniami probabilistycznymi. Wtedy przestrzeń probabilistyczną $(\Omega, \Sigma, \mathbb{P})$ zadaną przez

$$\begin{aligned}\Omega &= \Omega_1 \times \Omega_2 \\ \Sigma &= \sigma(\{A_1 \times A_2 : A_1 \in \Sigma_1, A_2 \in \Sigma_2\}); \\ \mathbb{P}(A_1 \times A_2) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2).\end{aligned}$$

nazywamy **iloczynem kartezjańskim** przestrzeni $(\Omega_1, \Sigma_1, \mathbb{P}_1)$ oraz $(\Omega_2, \Sigma_2, \mathbb{P}_2)$. Dla uproszczenia często stosuje się oznaczenie $\Sigma = \Sigma_1 \otimes \Sigma_2$ oraz $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$.

W Definicji 1.29 zdefiniowaliśmy miarę \mathbb{P} na zbiorach generujących $A_1 \times A_2$, gdzie $A_1 \in \Sigma_1$ oraz $A_2 \in \Sigma_2$. Można pokazać, że miara ta rozszerza się w sposób jednoznaczny na Σ , Dowód wykracza jednak poza zakres tego kursu; zob. Twierdzenie 18.2 w [Bil12]. Iloczyn kartezjański można też łatwo zdefiniować dla n -różnych przestrzeni (wprost albo w sposób rekurencyjny, co daje ten sam wynik). Zamiast wprowadzać formalną definicję posłużymy się przykładem.

Przykład 1.30 (n rzutów kostką). W Przykładzie 1.9 zdefiniowaliśmy przestrzeń odpowiadającą podwójnemu rzutowi kostką. Model ten można rozszerzyć na n rzutów kostką. Dla $i = 1, 2, \dots, n$ definiujemy $(\Omega_i, \Sigma_i, \mathbb{P}_i)$ jako $\Omega_i = \{1, 2, 3, 4, 5, 6\}$, $\Sigma_i = \mathcal{P}(\Omega_i)$ oraz $\mathbb{P}_i(A_i) = \frac{\#A_i}{6}$, $A_i \in \Sigma_i$. Wtedy przestrzeń określającą n -rzutów kostką można zapisać jako $(\Omega, \Sigma, \mathbb{P})$, gdzie $\Omega = \Omega_1 \times \dots \times \Omega_n$, σ -algebra zadana jest przez $\Sigma = \mathcal{P}(\Omega_1 \times \dots \times \Omega_n)$, a miara to $\mathbb{P}(A) = \frac{\#A}{6^n}$, dla $A \in \Sigma$.

W przypadku w którym chcemy stworzyć model iloczynu przestrzeni probabilistycznych o tej samej strukturze (np. powtórzenie tego samego eksperymentu n -razy) będziemy pisać często w uproszczeniu $(\Omega^n, \Sigma^n, \mathbb{P}^n)$.

Przykład 1.31 (Przestrzeń prób Bernoulliego). Standardowym przykładem wykorzystującym iloczyn kartezjański jest tzw. *przestrzeń prób Bernoulliego*. Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią taką, że $\Omega = \{0, 1\}$, $\Sigma = \mathcal{P}(\Omega)$ oraz $\mathbb{P}(\{1\}) = p$ i $\mathbb{P}(\{0\}) = 1 - p$, dla pewnego ustalonego $p \in (0, 1)$. Wtedy iloczyn $(\Omega^n, \Sigma^n, \mathbb{P}^n)$ nazywa się przestrzenią Bernoulliego (ang. *sample space*) dla n -prób. W szczególności zdarzenie odpowiadające k sukcesom w n próbach możemy zapisać jako

$$B_k := \{(\omega_1, \dots, \omega_n) \in \Omega^n : \sum_{i=1}^n \omega_i = k\}.$$

Zauważając, że dla każdego $\omega = (\omega_1, \dots, \omega_n) \in B_k$ zachodzi

$$\mathbb{P}^n(\{\omega\}) = \mathbb{P}(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}(\{\omega_n\}) = p^k \cdot (1 - p)^{n-k},$$

możemy łatwo policzyć prawdopodobieństwo B_k zliczając liczbę jego elementów, która wynosi $\binom{n}{k}$. Dla każdego $k = 1, 2, \dots, n$ dostajemy

$$\mathbb{P}^n(B_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

W Przykładzie 1.31 skonstruowaliśmy przestrzeń odpowiadającą skończonej ilości prób. Rozważając tzw. zbiory cylindryczne oraz nieskończone produkty miar możemy to rozszerzyć do nieskończonej ilości rzutów. Dokładny procedura konstrukcji wykracza jednak poza materiał tego wykładu. Warto jednak zauważyć, że eksperyment taki można w stosunkowo łatwy sposób zbudować na przestrzeni $\Omega = [0, 1]$ odpowiednio (w sposób indykcyjny) zadając na nim atomy; zob. Rozdział 4.2 (Przykład 2) w [JS04].

2 Prawdopodobieństwo warunkowe i niezależność

W rozdziale tym zajmiemy się tzw. prawdopodobieństwem warunkowym oraz niezależnością zdarzeń. W dużym skrócie chcemy zbadać, jak informacja o zajściu jednego zdarzenia wpływa na prawdopodobieństwo zajścia innego zdarzenia.

2.1 Prawdopodobieństwo warunkowe

Wiedząc, że zaszło jakieś zdarzenie uzyskujemy dodatkową informację, która może wpłynąć na prawdopodobieństwo zdarzenia, które chcemy obliczyć.

Definicja 2.1 (Prawdopodobieństwo warunkowe). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną oraz niech $B \in \Sigma$ spełnia $\mathbb{P}[B] > 0$. Wtedy dla dowolnego $A \in \Sigma$ definiujemy **prawdopodobieństwo warunkowe** względem zdarzenia B jako miarę probabilistyczną $\mathbb{P}(\cdot|B) : \Sigma \rightarrow [0, 1]$ zadaną przez

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Intuicyjnie rzecz ujmując, $\mathbb{P}(A|B)$ podaje nam prawdopodobieństwo zajścia zdarzenia A , gdy wiemy, że zaszło zdarzenie B . Wtedy, zamiast rozważać całą przestrzeń probabilistyczną, wystarczy zawęzić się do zbioru zdarzeń elementarnych w B i na nich liczyć (warunkowe) prawdopodobieństwo. Podajmy teraz prosty przykład, jak informacja może wpłynąć na prawdopodobieństwo.

Przykład 2.2 (Informacja o parzystości wyrzuconych oczek w rzucie kostką). Niech $\Omega = \{1, 2, \dots, 6\}$ oraz $(\Omega, \Sigma, \mathbb{P})$ odpowiada schematowi klasycznemu. Bezwarunkowe prawdopodobieństwo wypadnięcia szóstki wynosi $\frac{1}{6}$, tzn $\mathbb{P}(A) = \frac{1}{6}$, gdzie $A = \{6\}$. Niech $B := \{2, 4, 6\}$ będzie zdarzeniem odpowiadającym wyrzuceniu parzystej liczby oczek. Wiedząc, że wyrzucona liczba oczek jest parzysta, warunkowe prawdopodobieństwo wyrzucenia 6 wzrasta. Istotnie, dostajemy

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/6}{3/6} = \frac{1}{3}.$$

Z prawdopodobieństwem warunkowym związany jest wzór na tzw. prawdopodobieństwo całkowite.

Twierdzenie 2.3 (Prawdopodobieństwo całkowite). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną oraz niech $B_1, \dots, B_n \in \Sigma$, $n \in \mathbb{N}$, spełniają warunki

- 1) $\mathbb{P}(B_i) > 0$, $i = 1, \dots, n$;
- 2) $B_i \cap B_j = \emptyset$, dla $i \neq j$;
- 3) $B_1 \cup \dots \cup B_n = \Omega$.

Wtedy dla każdego zdarzenia $A \in \Sigma$ zachodzi wzór

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Dowód. Ponieważ $A = A \cap \Omega = A \cap (\bigcup_{i=1}^n B_i) = \bigcup_{i=1}^n (A \cap B_i)$ oraz zdarzenia $A \cap B_i$, $i = 1, \dots, n$, są rozłączne, dostajemy

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap B_i) = \sum_{i=1}^n \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \mathbb{P}(B_i) = \sum_{i=1}^n \mathbb{P}(A|B_i) \mathbb{P}(B_i).$$

□

Przykład 2.4 (Informacja o parzystości wyrzuconych oczek w rzucie kostką – c.d.). Używając notacji z Przykładu 2.2 wiemy, że B i B' spełniają założenia Propozycji 2.3. Dostajemy

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B')\mathbb{P}(B') = \frac{1}{3} \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{6}.$$

Kolejnym ważnym wzorem związanym z prawdopodobieństwem warunkowym jest tzw. wzór Bayesa.

Twierdzenie 2.5 (Wzór Bayesa). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną oraz niech $B_1, \dots, B_n \in \Sigma$, $n \in \mathbb{N}$, będą zdarzeniami spełniającymi warunki Propozycji 2.3. Wtedy dla każdego $j = 1, \dots, n$ dostajemy

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Dowód tego faktu pozostawiamy jako ćwiczenie do domu. Wzór Bayesa pozwala na odwrócenie warunkowania i pokazanie jak informacja zawarta w A wpływa na prawdopodobieństwo poszczególne zdarzeń B_1, \dots, B_n .

Uwaga 2.6 (Prawdopodobieństwo a priori i a posteriori). W analizie Bayesowskiej prawdopodobieństwo $\mathbb{P}(B_j)$ często nazywa się prawdopodobieństwem *a priori*, a prawdopodobieństwo warunkowe $\mathbb{P}(B_j|A)$ prawdopodobieństwem *a posteriori*. Rozróżnienie to jest wprowadzone, aby pokazać, jak informacja o zdarzeniu A wpływa na (warunkowe) prawdopodobieństwo zajścia zdarzenia B_j .

2.2 Niezależność zdarzeń

W poprzednim podrozdziale zastanawialiśmy się, jak zadane zdarzenie może wpływać na prawdopodobieństwo innego zdarzenia. Szczególnym przypadkiem jest tutaj sytuacja, gdy zadane zdarzenie nie ma wpływu na prawdopodobieństwo innego zdarzenia.

Definicja 2.7 (Zdarzenia niezależne). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną. Zdarzenia $A \in \Sigma$ oraz $B \in \Sigma$ nazywamy **zdarzeniami niezależnymi**, gdy

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Zdarzenia A i B nazywamy *zależnymi*, gdy nie są niezależne.

Z definicji niezależności wynika niezależność zdarzenia pewnego (oraz jego dopełnienia) z każdym zdarzeniem, oraz zależność zdarzeń rozłącznych o dodaniu prawdopodobieństwie, co jest omówione w dwóch kolejnych uwagach.

Uwaga 2.8 (Niezależność zdarzenia pewnego). Łatwo zauważyć, że zdarzenie $A \in \Sigma$ takie, że $\mathbb{P}(A) = 1$ jest niezależne z dowolnym innym zdarzeniem, gdyż dla dowolnego $B \in \Sigma$ zachodzi

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) = \mathbb{P}(B) = \mathbb{P}(B)\mathbb{P}(A).$$

W podobny sposób można pokazać, że zdarzenie o zerowym prawdopodobieństwie jest niezależne z dowolnym zdarzeniem.

Uwaga 2.9 (Zależność zdarzeń rozłącznych). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną. Załóżmy, że zdarzenia $A \in \Sigma$ oraz $B \in \Sigma$ są rozłącznymi zdarzeniami o dodatnim prawdopodobieństwie. Wtedy zdarzenia A i B są zależne, gdyż $\mathbb{P}(A \cap B) = 0$ podczas, gdy $\mathbb{P}(A) \cdot \mathbb{P}(B) > 0$.

Jak można się domyślić, niezależność zdarzeń jest ściśle związana z prawdopodobieństwem warunkowym.

Uwaga 2.10 (Niezależność a prawdopodobieństwo warunkowe). W szczególności, gdy $\mathbb{P}(B) > 0$, to zdarzenia się niezależne wtedy o tylko wtedy, gdy zachodzi $\mathbb{P}(A|B) = \mathbb{P}(A)$, tj. jest informacja zawarta w zdarzeniu B nie wpływa na prawdopodobieństwo zdarzenia A . Jeżeli zachodzi natomiast $\mathbb{P}(A|B) > \mathbb{P}(A)$ (odp. $<$) to często mówi się, że zdarzenie B *sprzyja* (odp. *nie sprzyja*) zdarzeniu A .

Związek ten można zilustrować na przykładzie.

Przykład 2.11 (Przykład zdarzeń zależnych i niezależnych). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną odpowiadającą dwoma rzutom kostką, zob. Przykład 1.9. Wtedy łatwo sprawdzić, stosując nieformalny zapis, że zdarzenie $A = \{\text{w pierwszym rzucie wypadła szóstka}\}$, oraz zdarzenie $B = \{\text{w drugim rzucie wypadła jedynka}\}$, są niezależne. Istotnie

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

W szczególności dostajemy $\mathbb{P}(A|B) = \frac{1/36}{1/6} = \frac{1}{6} = \mathbb{P}(A)$. Przykładem zdarzeń zależnych mogą być z kolei zdarzenia $C := \{\text{suma oczek jest większa niż 10}\}$ oraz $D := \{\text{w drugim rzucie wypadła szóstka}\}$. Dostajemy

$$\mathbb{P}(C \cap D) = \frac{2}{36} < \frac{3}{36} \cdot \frac{1}{6} = \mathbb{P}(C) \cdot \mathbb{P}(D).$$

W szczególności $\mathbb{P}(C|D) = \frac{2}{6} > \frac{3}{36} = \mathbb{P}(C)$, czyli D sprzyja C .

Uwaga 2.12 (Zdarzenia sprzyjające). Można w łatwy sposób wykazać (ćw. do domu), że relacja *sprzyjania* jest symetryczna, tzn. jeżeli zdarzenia A sprzyja B , to również B sprzyja A .

Przykład 2.13 (Iloczyn kartezjański przestrzeni probabilistycznych). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie iloczynem kartezjańskim przestrzeni $(\Omega_1, \Sigma_1, \mathbb{P}_1)$ i $(\Omega_2, \Sigma_2, \mathbb{P}_2)$. Niech $A_1 \in \Sigma_1$ oraz $A_2 \in \Sigma_2$. Zdefiniujmy zdarzenia brzegowe zadane przez $Z_1 := A_1 \times \Omega_2$ oraz $Z_2 := \Omega_1 \times A_2$. Łatwo pokazać, że zdarzenia Z_1 oraz Z_2 są niezależne. Istotnie, dostajemy

$$\mathbb{P}(Z_1 \cap Z_2) = \mathbb{P}((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) = \mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1) \cdot \mathbb{P}_2(A_2). \quad (2.1)$$

Zauważmy następnie, że $\mathbb{P}_1(A_1) = \mathbb{P}_1(A_1) \cdot \mathbb{P}_2(\Omega_2) = \mathbb{P}(A_1 \times \Omega_2) = \mathbb{P}(Z_1)$ oraz w podobny sposób $\mathbb{P}_2(A_2) = \mathbb{P}(Z_2)$. Łącząc to z (2.1) dostajemy niezależność, tzn. równość $\mathbb{P}(Z_1 \cap Z_2) = \mathbb{P}(Z_1) \cdot \mathbb{P}(Z_2)$. Intuicyjnie rzecz ujmując traktując przestrzenie brzegowe jako niezależne eksperymenty (modele) dostajemy, że nawet pełna wiedza o pierwszym eksperymencie, nie daje nam żadnej informacji o drugim eksperymencie (i na odwrót). Zakładając, że $\mathbb{P}_1(A_1) > 0$ mamy $\mathbb{P}(Z_2|Z_1) = \mathbb{P}(Z_2) = \mathbb{P}_2(A_2)$.

Definicję niezależności można uogólnić na większą ilość zdarzeń.

Definicja 2.14 (Łączna niezależność zdarzeń). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną. Dla ustalonego $n \in \mathbb{N}$ zdarzenia $A_1, \dots, A_n \in \Sigma$ nazywamy **zdarzeniami łącznie niezależnymi**, gdy dla każdego podzbioru tych zdarzeń, ozn. A_{i_1}, \dots, A_{i_k} , zachodzi

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_k}).$$

Przeliczalną liczbę zdarzeń A_1, A_2, \dots nazywamy zdarzeniami łącznie niezależnymi, gdy dla każdego $n \in \mathbb{N}$ zdarzenia A_1, \dots, A_n są łącznie niezależne.

Warto zaznaczyć, że o ile łączna niezależność zdarzeń w sposób oczywisty implikuje niezależność (par) tych zdarzeń, o tyle implikacja w drugą stronę nie jest zawsze prawdziwa. Ilustruje to następujący przykład.

Przykład 2.15 (Niezależność par zdarzeń, a łączna niezależność). Rozważmy model rzutu kostką czworościenną i dla $i = 1, 2, 3$ definiujemy $A_i := \{\text{wylosowano liczbę oczek równą } 1 \text{ lub } i + 1\}$. Łatwo pokazać, że $\mathbb{P}(A_i) = \frac{1}{2}$ oraz $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = \frac{1}{4}$, a więc zdarzenia te są parami niezależne. Z drugiej strony dostajemy $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} > \frac{1}{8} = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$, a więc zdarzenia te są łącznie zależne. Rozważając prawdopodobieństwo warunkowe dostajemy $\mathbb{P}(A_1 | A_2 \cap A_3) = 1 > \frac{1}{2} = \mathbb{P}(A_1)$, co pokazuje, że zdarzenie A_1 nie jest niezależne od zdarzenia $A_1 \cap A_2$.

Więcej przykładów zdarzeń zależnych i niezależnych zostanie omówionych na ćwiczeniach.

2.3 Lemat Borela-Cantelliego

Lemat Borela-Cantelliego jest związany z analiza sytuacji, w której chcemy zbadać, czy z danej przeliczalnej rodziny zdarzeń zaszło nieskończenie wiele zdarzeń. Formalnie jest to związane z tzw. granicą górną i dolną ciągu zdarzeń.

Definicja 2.16 (Granica górna i dolna ciągu zdarzeń). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną oraz niech $A_1, A_2, \dots \in \Sigma$ będzie ciągiem określonych na niej zdarzeń. Wtedy **granicą górną** (łac. *limes superior*) oraz **granicą dolną** (łac. *limes inferior*) nazywamy zdarzenia

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n, \quad \text{oraz} \quad \liminf_{n \rightarrow \infty} A_n := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n$$

Intuicyjnie rzecz ujmując $\limsup_{n \rightarrow \infty} A_n$ identyfikuje nam zdarzenia elementarne, które należą do nieskończenie wielu zbiorów $(A_n)_{n=1}^{\infty}$ podczas, gdy $\liminf_{n \rightarrow \infty} A_n$ identyfikuje zdarzenia, które należą do prawie wszystkich $(A_n)_{n=1}^{\infty}$.

Propozycja 2.17 (Inna reprezentacja granicy górnej i dolnej). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną oraz niech $A_1, A_2, \dots \in \Sigma$ będzie ciągiem określonych na niej zdarzeń. Wtedy

- 1) $\limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = \infty\}$;
- 2) $\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \sum_{n=1}^{\infty} \mathbb{1}_{A_n'}(\omega) < \infty\}$;
- 3) $(\limsup_{n \rightarrow \infty} A_n)' = \liminf_{n \rightarrow \infty} A_n'$.

Dowód Propozycji 2.17 pozostawiamy jako ćwiczenie do domu; zob. Rozdział 4.3 (Uwaga 2) w [JS04].

Twierdzenie 2.18 (Lemat Borela–Cantelliego). *Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną oraz niech $A_1, A_2, \dots \in \Sigma$ będzie ciągiem określonych na niej zdarzeń. Wtedy*

- 1) *Jeżeli $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, to $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$;*
- 2) *Jeżeli $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ oraz $(A_n)_{n=1}^{\infty}$ są łącznie niezależne, to $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$.*

Dowód.

1) Zauważmy, że rodzina zdarzeń (B_m) , gdzie $B_m := \bigcup_{n=m}^{\infty} A_n$, jest rodziną zstępującą. Korzystając z Propozycji 1.7, punkt 5), dostajemy

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \mathbb{P}\left(\bigcap_{m=1}^{\infty} B_m\right) = \lim_{m \rightarrow \infty} \mathbb{P}(B_m) \leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0.$$

2) Korzystając z Propozycji 2.17, punkt 3), wystarczy pokazać, że $\mathbb{P}(\liminf_{n \rightarrow \infty} A'_n) = 0$. Ponieważ dla $C_m := \bigcap_{n=m}^{\infty} A'_n$ zachodzi

$$\mathbb{P}(\liminf_{n \rightarrow \infty} A'_n) = \mathbb{P}\left(\bigcup_{m=1}^{\infty} C_m\right),$$

wystarczy pokazać, że $\sum_{m=1}^{\infty} \mathbb{P}(C_m) = 0$, czyli, że każdy składnik sumy jest równy zero. Ustalmy więc $m \in \mathbb{N}$. Korzystając z Propozycji 1.7, punkt 4), niezależności zdarzeń, oraz nierówności $1 - x \leq e^{-x}$, dostajemy

$$\mathbb{P}(C_m) = \lim_{N \rightarrow \infty} \prod_{n=m}^N \mathbb{P}(A'_n) = \lim_{N \rightarrow \infty} \prod_{n=m}^N (1 - \mathbb{P}(A_n)) \leq \lim_{N \rightarrow \infty} \prod_{n=m}^N e^{-\mathbb{P}(A_n)} = \lim_{N \rightarrow \infty} e^{-\sum_{n=m}^N \mathbb{P}(A_n)} = 0,$$

co kończy dowód. □

Lemat Borela–Cantelliego daje nam receptę na wyznaczania zdarzeń, które zachodzą z prawdopodobieństwem 1 w powtarzalnych eksperymentach.

Przykład 2.19 (Nieskończona ilość rzutów monetą). Niech przestrzeń probabilistyczna $(\Omega, \Sigma, \mathbb{P})$ odpowiada modelowi nieskończonej ilości rzutów monetą. Wtedy, prawdopodobieństwo, że orzeł wypadnie 100 razy pod rząd nieskończenie wiele razy wynosi 1. Aby skorzystać z lematu Borela–Cantelliego, zdefiniujmy ciąg zdarzeń $(A_n)_{n=1}^{\infty}$, gdzie A_n oznacza zdarzenie wypadnięcia 100 orłów w rzutach o numerach od $100n + 1$ do $100(n + 1)$. Zdarzenie mówiące, że orzeł wypadnie 100 razy nieskończenie wiele razy – oznaczmy je przez B – zawiera granicę górną zdarzeń (A_n) , tzn. zachodzi

$$B \supset \limsup_{n \rightarrow \infty} A_n.$$

Ponadto, zdarzenia (A_n) są niezależne oraz $\mathbb{P}(A_n) = \frac{1}{2^{100}}$, co implikuje $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. Korzystając z lematu Borela–Cantelliego wiemy więc, że $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$, co implikuje $\mathbb{P}(B) = 1$.

3 Zmienne losowe i ich rozkłady

W rozdziale tym przedstawimy definicję zmiennej losowej, jej rozkładu oraz innych podstawowych własności. W poprzednim rozdziale pokazaliśmy dużo przykładów przestrzeni probabilistycznych, które miały skomplikowaną strukturę. Często z wynikiem doświadczenia losowego wiąże się jakaś liczba albo ciąg liczb, a sama struktura przestrzeni nie jest dla nas aż tak ważna. W takim wypadku często łatwiej jest dokonać analizy eksperymentu poprzez wprowadzenie tzw. *zmiennej losowej* i operowanie na podzbiorach wartości w \mathbb{R} (lub \mathbb{R}^n w przypadku wielowymiarowym).

3.1 Definicja zmiennej losowej i wektora losowego

Na początku wprowadźmy formalną definicję zmiennej losowej, czyli funkcji (mierzalnej), która każdemu zdarzeniu elementarnemu przypisuje jakąś wartość liczbową.

Definicja 3.1 (Zmienna losowa). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną. Odwzorowanie $X: \Omega \rightarrow \mathbb{R}$ nazywamy zmienną losową, jeżeli dla każdego $A \in \mathcal{B}(\mathbb{R})$ dostajemy $X^{-1}(A) = \{\omega \in \Omega: X(\omega) \in A\} \in \Sigma$, tj. jeżeli przeciwobraz każdego zbioru mierzalnego w $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ jest mierzalny w (Ω, Σ) .

Będziemy stosować zapis $\{X \in A\} := X^{-1}(A)$, czy $\{X \leq k\} := \{\omega \in \Omega: X(\omega) \in (-\infty, k]\}$ na oznaczenie przeciwobrazów zadanych zbiorów.

Przykład 3.2 (Suma reszek w dwóch rzutach symetryczną monetą). Niech $(\Omega^2, \Sigma^2, \mathbb{P}^2)$ będzie przestrzenią probabilistyczną odpowiadającą iloczynowi kartezjańskiemu $\Omega = \{0, R\}$ ze schematem klasycznym, co odpowiada 2-krotnemu rzutowi monetą. Wtedy zmienną losową opisującą sumę reszek, które wypadły w 2-rzutach, można zadać poprzez $X(\omega_1, \omega_2) := \mathbb{1}_{\{R\}}(\omega_1) + \mathbb{1}_{\{R\}}(\omega_2)$ dla $(\omega_1, \omega_2) \in \Omega^2$. Możemy również zdefiniować zmienne $X_1(\omega_1, \omega_2) = \mathbb{1}_{\{R\}}(\omega_1)$ oraz $X_2(\omega_1, \omega_2) = \mathbb{1}_{\{R\}}(\omega_2)$ oraz zapisać $X = X_1 + X_2$. Prawdopodobieństwo tego, że wypadły 2 reszki w 2 rzutach możemy wtedy zapisać jako $\mathbb{P}(X = 2)$.

Przykład 3.3 (Losowanie dwóch liczb i ich bezwzględna różnica). W Przykładzie 1.26 losowaliśmy dwie liczby z przedziału $[0, 1]$ i liczyliśmy prawdopodobieństwo, że wartość bezwzględna z różnicy tych liczb jest mniejsza niż 0.1. Na przestrzeni $(\Omega, \Sigma, \mathbb{P})$, gdzie $\Omega = [0, 1]^2$, $\Sigma = \mathcal{B}(\Omega)$ oraz prawdopodobieństwo zadane jest poprzez $\mathbb{P}(A) = \mathcal{L}_2(A)$, dla $A \in \Sigma$, można definiować zmienną, która bezpośrednio określa wartość bezwzględną z różnicy losowań. Zmienna ta zadana jest przez $X(x, y) = |x - y|$. Wystarczy wtedy policzyć prawdopodobieństwo $\mathbb{P}(X < 0.1)$.

Definicję zmiennej losowej można uogólnić na więcej wymiarów, tzn. rozważać wektory losowe postaci $\mathbf{X} = (X_1, \dots, X_n)$ dla każdego $n \in \mathbb{N}$.³

Definicja 3.4 (Wektor losowy). Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną. Odwzorowanie $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ nazywamy wektorem losowym, jeżeli dla każdego $A \in \mathcal{B}(\mathbb{R}^n)$ dostajemy $\mathbf{X}^{-1}(A) \in \Sigma$, tj. jeżeli przeciwobraz każdego zbioru mierzalnego w $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ jest mierzalny w (Ω, Σ) .

³Z teoretycznego punktu widzenia wyróżnianie *zmiennych losowych* i *wektorów losowych* nie ma (formalnego) znaczenia; wektor losowy jest zmienną losową dla $n = 1$. Robi się to jednak ze względów historycznych oraz aby wyróżnić przypadek wielowymiarowy.

Łatwo zauważyć, że definicja *zmiennej losowej* (oraz *wektora losowego*) w teorii nie zależy od wyboru miary \mathbb{P} . W praktyce jednak będzie nas interesował *rozkład prawdopodobieństwa* tych funkcji, który będzie już zależał od wyjściowej miary. Czasami w literaturze wprowadza się dodatkowe pojęcie *funkcji losowej*, w przypadku w którym definiujemy funkcję X na przestrzeni mierzalnej (Ω, Σ) bez miary probabilistycznej.

Uwaga 3.5 (Równość prawie na pewno). Przez większą część wykładu będziemy ze sobą utożsamiać zmienne losowe, które są sobie równe z prawdopodobieństwem jeden; równość $X = Y$ będziemy rozumieć jako $\mathbb{P}(X = Y) = \mathbb{P}(\{\omega \in \Omega: X(\omega) = Y(\omega)\}) = 1$. Czasami do znaku równości będziemy dodawać oznaczenia *p.n.* (lub *a.s.*), aby podkreślić przyjętą konwencję. Jeżeli ważna będzie równość dla wszystkich zdarzeń elementarnych (a nie tylko prawie na pewno), to będziemy się starali to również wyróżniać. Widać, że dla równości prawie na pewno istotna jest miara probabilistyczna względem której zdefiniowane są zmienne. Czasami zamiast rozważać zmienną losową (wektor losowy) X rozważa się *klasy abstrakcji*, gdzie każdy element wiąże ze sobą wszystkie zmienne, które są sobie równe prawie na pewno.⁴

Uwaga 3.6 (Sprawdzanie mierzalności). W praktyce, aby sprawdzić czy coś jest zmienną losową wystarczy pokazać mierzalność funkcji $X: \Omega \rightarrow \mathbb{R}$ na zbiorach generujących σ -algebrę zbiorów borelowskich $\mathcal{B}(\mathbb{R})$. Przykładowo, aby sprawdzić, czy zadane odwzorowanie $X: \Omega \rightarrow \mathbb{R}$ jest zmienną losową wystarczy pokazać, że $X^{-1}(A) \in \Sigma$ dla każdego A postaci $(-\infty, a]$, gdzie $a \in \mathbb{R}$. Podobnie, w przypadku wielowymiarowym, wystarczy rozważać zbiory mierzalne postaci $A = (-\infty, a_1] \times \dots \times (-\infty, a_n]$.

Łatwo zauważyć, że obłożenie wektora losowego funkcją borelowską⁵ $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ daje nam wektor losowy.

Propozycja 3.7 (Obłożenie wektora losowego funkcją borelowską). Niech $X: \Omega \rightarrow \mathbb{R}^n$ będzie n -wymiarowym wektorem losowym określonym na $(\Omega, \Sigma, \mathbb{P})$. Wtedy dla każdego $m \in \mathbb{N}$ i funkcji borelowskiej $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$, funkcja $\phi(\mathbf{X}): \Omega \rightarrow \mathbb{R}^m$ jest wektorem losowym.^a

^aFormalnie, definiujemy $\phi(\mathbf{X})(\omega) := (\phi \circ \mathbf{X})(\omega)$.

Dowód. Niech $B \in \mathcal{B}(\mathbb{R}^m)$. Ponieważ ϕ jest funkcją borelowską, to $\phi^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$. Korzystając z tego \mathbf{X} jest wektorem losowym dostajemy więc $(\phi(\mathbf{X}))^{-1}(B) = \mathbf{X}^{-1}(\phi^{-1}(B)) \in \Sigma$. \square

Niech X będzie zmienną losową. Z Propozycji 3.7 wynika w szczególności, że funkcje $Y = X^2$, czy $Z = e^{|X|}$ również są zmiennymi losowymi. Dodawanie, odejmowanie, czy mnożenie zmiennych losowych również daje nam zmienne losowe. Ogólniej rzecz biorąc, każda funkcja $Y: \Omega \rightarrow \mathbb{R}^n$, którą można zdefiniować wychodząc od (przeliczalnego) wektora losowego X_1, X_2, \dots i korzystając z funkcji elementarnych oraz działań przeliczalnych (\lim , $\lim \sup$, $\lim \inf$, itd.) będzie zmienną losową.⁶

W rozdziale tym część teorii zaprezentujemy tylko dla przypadku jednowymiarowego, tj. *zmiennych losowych*. Należy mieć na uwadze, że duża część wyników w sposób oczywisty (lub prawie oczywisty) przynosi się na większą liczbę wymiarów.

⁴Rozważając klasy abstrakcji można wprowadzić lepszą strukturę np. w przestrzeni (klas abstrakcji) zmiennych losowych całkowalnych (z normą L^1).

⁵Funkcją dla której przeciwobrazy zbiorów borelowskich są borelowskie.

⁶Trzeba tutaj jednak uważać: np. $\sup_{a \in \mathbb{R}} X_a$, nie musi być zmienną losową, nawet gdy X_a jest zmienną losową dla każdego $a \in \mathbb{R}$. Warto zwrócić uwagę, że mamy tu supremum na zbiorze nieprzeliczalnym.

3.2 Rozkład prawdopodobieństwa i σ -algebra generowana przez zmienną losową

Ze zmienną losową w naturalny sposób wiążą się dwa obiekty: σ -algebra generowana przez zmienną losową oraz rozkład prawdopodobieństwa (na \mathbb{R}), który generuje zmienną losową. Zaczniemy od definicji σ -algebry generowanej przez zmienną losową.

Definicja 3.8 (σ -algebra generowana przez zmienną losową). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy **σ -algebrę generowaną przez zmienną losową X** nazywamy najmniejszą σ -algebrą podzbiorów Ω , względem której X jest mierzalna; oznaczamy ją przez $\sigma(X)$. Łatwo pokazać, że

$$\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}.$$

Z postaci σ -algebry z Definicji 3.8 dostajemy $\sigma(X) \subseteq \Sigma$.

Przykład 3.9 (Suma reszek w dwóch rzutach symetryczną monetą c.d.). Dla zmiennej X określonej w Przykładzie 3.2 dostajemy $\sigma(X) = \sigma(\{X = 0\}, \{X = 1\}, \{X = 2\})$, co daje nam 3-atomową przestrzeń probabilistyczną. Warto zwrócić uwagę, że $\sigma(X) \subset \mathcal{P}(\Omega)$, a w szczególności $\#\sigma(X) = 2^3 < 2^4 = \#\mathcal{P}(\Omega)$.

Przykład 3.10 (Losowanie dwóch liczb i ich bezwzględna różnica c.d.). Dla zmiennej X określonej w Przykładzie 3.3 $\sigma(X)$ jest σ -algebrą generowaną przez zbiory postaci $A_t := \{(x, y) \in \Omega : |x - y| < t\}$, $t \in \mathbb{R}$. Taka σ -algebra jest istotnie mniejsza od Σ , gdyż na nie należy do niej na przykład rodzina zbiorów $B_t := \{(x, y) \in \Omega : x < t\}$, $t \in \mathbb{R}$, która jest borelowska.⁷

Zdefiniujmy teraz *rozkład prawdopodobieństwa zmiennej losowej*, który w naturalny sposób umożliwia "transport" prawdopodobieństwa z abstrakcyjnej przestrzeni (Ω, Σ) na przestrzeń $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Zaczniemy od wprowadzenia ogólnej definicji *rozkładu prawdopodobieństwa*.

Definicja 3.11 (Rozkład prawdopodobieństwa). Przez **rozkład prawdopodobieństwa** na \mathbb{R} będziemy rozumieć każdą miarę probabilistyczną $\mu : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}$ określoną na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Rozkładów prawdopodobieństwa na \mathbb{R} jest oczywiście bardzo dużo. Do każdej zmiennej losowej można przyporządkować jeden rozkład.

Definicja 3.12 (Rozkład prawdopodobieństwa zmiennej losowej). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy **rozkładem prawdopodobieństwa zmiennej losowej X** nazywamy rozkład prawdopodobieństwa, oznaczany przez μ_X , dla którego zachodzi

$$\mu_X(B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).^a$$

^aW literaturze można również spotkać notację \mathbb{P}_X zamiast μ_X .

Z Definicji 3.12 widzimy, że $\mu_X(B) = \mathbb{P}(\{X \in B\})$ oznacza prawdopodobieństwo, że wartości zmiennej X będą w zbiorze B . Rozkład prawdopodobieństwa zmiennej losowej X dostarcza nam zazwyczaj wszystkich potrzebnych informacji o zmiennej X , choć jest on w istocie miarą probabilistyczną na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ i nie musi być związany z konkretną zmienną losową, czy powiązaną przestrzenią probabilistyczną $(\Omega, \Sigma, \mathbb{P})$. Dla uproszczenia, mając zadany rozkład μ , będziemy pisać $X \sim \mu$, jeżeli będziemy wymagać, aby $\mu_X = \mu$.

⁷Intuicyjnie rzecz ujmując, mamy tylko informację o tym jaka jest odległość bezwzględna między x i y . Z tej informacji nie da się wywnioskować jaka jest wartość x .

Przykład 3.13 (Suma reszek w dwóch rzutach symetryczną monetą c.d.). Dla zmiennej X określonej w Przykładzie 3.2 jej rozkład skupiony jest na zbiorze dyskretnym $\{0, 1, 2\}$ poprzez $\mu_X(\{0\}) = \mu_X(\{2\}) = \frac{1}{4}$ oraz $\mu_X(\{1\}) = \frac{1}{2}$.

Przykład 3.14 (Losowanie dwóch liczb i ich bezwzględna różnica c.d.). Dla zmiennej X określonej w Przykładzie 3.3 jej rozkład skupiony jest na zbiorze $[0, 1]$. Przykładowo, dla zbiorów mierzalnych $A_t = [0, t]$, $t \in [0, 1]$, dostajemy $\mu_X(A_t) = \mathbb{P}(X \leq t) = 1 - 2 \cdot (0.5 \cdot (1 - t) \cdot (1 - t)) = 1 - (1 - t)^2$.

W dalszej części tego rozdziału zastanowimy się, jak dobrze charakteryzować (pewne) rozkłady prawdopodobieństwa oraz związane z nimi obiekty. Można to zrobić na przykład poprzez funkcję gęstości prawdopodobieństwa, czy dystrybuantę zmiennej losowej.

3.3 Rozkłady ciągłe i gęstość prawdopodobieństwa

Zajmijmy się teraz pewnymi szczególnymi typami rozkładów prawdopodobieństwa dla których istnieje tzw. *funkcja gęstości prawdopodobieństwa*.

Definicja 3.15 (Rozkład ciągły i gęstość rozkładu). Niech μ będzie rozkładem prawdopodobieństwa na \mathbb{R} . Jeżeli dla pewnej funkcji $f: \mathbb{R} \rightarrow \mathbb{R}$, całkownej w sensie Lebesgue'a, zachodzi

$$\mu(A) = \int_A f(x)dx, \quad A \in \mathcal{B}(\mathbb{R}),$$

to rozkład μ nazywamy **rozkładem ciągłym**, a funkcję f nazywamy **funkcją gęstości prawdopodobieństwa** rozkładu μ .

Podobnie jak wcześniej, do (pewnych) zmiennych losowych, możemy przyporządkować gęstość. Warto tutaj podkreślić, że istnieją zmienne losowe, dla których rozkład (gęstość) nie istnieje.

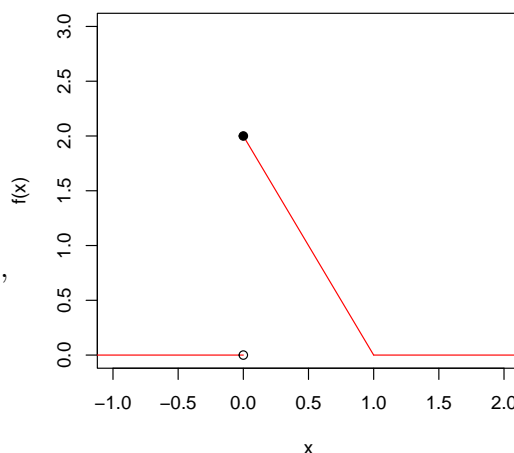
Definicja 3.16 (Gęstość zmiennej losowej i absolutna ciągłość). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Jeżeli rozkład zmiennej X , tj. μ_X , jest ciągły z funkcją gęstości prawdopodobieństwa f_X to mówimy, że zmienna losowa X jest **absolutnie ciągła** oraz ma **gęstość** f_X .

Przez f_X będziemy zazwyczaj oznaczać gęstość zmiennej X .

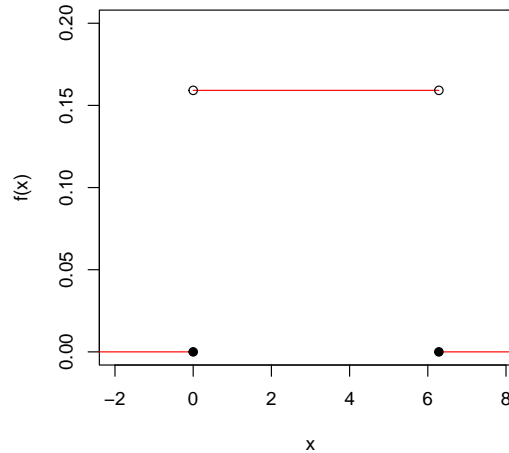
Przykład 3.17 (Losowanie dwóch liczb i ich bezwzględna różnica c.d.). Zmienna losowa X określona w Przykładzie 3.3 ma rozkład ciągły, który zadany jest przez gęstość $f_X(x) = \mathbb{1}_{[0,1]}(x)2(1-x)$. Istotnie, dla zbioru $A_t = [0, t]$, $t \in [0, 1]$, dostajemy

$$\mu_X(A_t) = \int_0^t 2(1-x)dx = 2t - t^2 = 1 - (1-t)^2,$$

co pokrywa się z rozkładem prawdopodobieństwa podanym w Przykładzie 3.14.



Przykład 3.18 (Rozkład jednostajny na odcinku $(0, 2\pi)$). Rozważając paradoks Bertranda losowaliśmy kąt z przedziału $(0, 2\pi)$. Odpowiada to tzw. rozkładowi jednostajnemu na odcinku. Na przestrzeni probabilistycznej $(\Omega, \Sigma, \mathbb{P})$, gdzie $\Omega = (0, 2\pi)$, $\Sigma = \mathcal{B}((0, 2\pi))$ oraz $\mathbb{P}(\cdot) = \mathcal{L}_1(\cdot)/2\pi$ określimy zmienną losową przez $X(\omega) = \omega$. Zmienna taka ma rozkład $\mu_X(\cdot) = \mathbb{P}(\cdot)$ dla którego istnieje funkcją gęstości zadana przez $f_X(x) = \frac{1}{2\pi} \cdot \mathbb{1}_{(0, 2\pi)}(x)$, $x \in \mathbb{R}$.



Naturalnym pytaniem wydaje się scharakteryzowanie wszystkich funkcji $f: \mathbb{R} \rightarrow \mathbb{R}$ będących gęstością dla pewnego rozkładu prawdopodobieństwa.

Twierdzenie 3.19 (Własności funkcji gęstości prawdopodobieństwa). Niech f będzie gęstością dla pewnego rozkładu prawdopodobieństwa μ na \mathbb{R} . Wtedy

- 1) $\int_{\mathbb{R}} f(x) dx = 1$;
- 2) $f \geq 0$ p.n. (tzn. $f(x) \geq 0$ dla $x \in \mathbb{R} \setminus A$, gdzie A jest zbiorem Lebesgue'a miary zero);
- 3) f jest wyznaczona jednoznacznie z dokładnością do zbiorów Lebesgue'a miary zero.

Ponadto, każda funkcja $f: \mathbb{R} \rightarrow \mathbb{R}$ spełniająca warunki 1) i 2) jest gęstością prawdopodobieństwa dla pewnego rozkładu μ na \mathbb{R} .

Dowód.

1) Ponieważ μ jest rozkładem prawdopodobieństwa na \mathbb{R} , dostajemy $\int_{\mathbb{R}} f(x) dx = \mu(\mathbb{R}) = 1$.

2) Niech $A := \{x \in \mathbb{R}: f(x) < 0\}$. Dla $k \in \mathbb{N}$ definiujemy $A_k := \{x \in \mathbb{R}: f(x) < -\frac{1}{k}\}$. Dla każdego $k \in \mathbb{N}$ mamy $0 \leq \mu(A_k) = \int_{A_k} f(x) dx \leq -\frac{1}{k} \mathcal{L}_1(A_k)$, co daje nam $\mathcal{L}_1(A_k) = 0$. Stąd dostajemy $\mathcal{L}_1(A) = \mathcal{L}_1(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} \mathcal{L}_1(A_k) = 0$, co kończy dowód.

3) Niech f_1 i f_2 będą funkcjami gęstości rozkładu μ . Podobnie jak wcześniej, definiujemy zbiory $B := \{x \in \mathbb{R}: f_1(x) - f_2(x) < 0\}$ oraz $B_k := \{x \in \mathbb{R}: f_1(x) - f_2(x) < -\frac{1}{k}\}$. Dla każdego $k \in \mathbb{N}$ mamy $0 = \mu(B_k) - \mu(B_k) = \int_{B_k} (f_1(x) - f_2(x)) dx \leq -\frac{1}{k} \mathcal{L}_1(B_k)$, co daje nam $\mathcal{L}_1(B_k) = 0$ i w konsekwencji $\mathcal{L}_1(B) = 0$. W podobny sposób, zamieniając f_1 z f_2 , dostajemy $\mathcal{L}_1(\tilde{B}) = 0$ dla $\tilde{B} := \{x \in \mathbb{R}: f_1(x) - f_2(x) > 0\}$.

Pokazane, że funkcja spełniająca 1) i 2) jest gęstością jest rutynowym zadaniem z teorii miary i pominiemy tego dowód (ćw. do domu). \square

Uwaga 3.20 (Charakteryzacja gęstości). Z Twierdzenia 3.19 wynika, że funkcja f jest funkcją gęstości prawdopodobieństwa wtedy i tylko wtedy, gdy f jest nieujemna (p.n.) i całkuje się do jedności.

3.4 Rozkłady dyskretne i masa prawdopodobieństwa

Kolejnym szczególnym typem rozkładów prawdopodobieństwa są tzw. *rozkłady dyskretne*.

Definicja 3.21 (Rozkład dyskretny i funkcja masy prawdopodobieństwa). Niech μ będzie rozkładem prawdopodobieństwa na \mathbb{R} . Jeżeli istnieje zbiór przeliczalny $S \subset \mathbb{R}$ taki, że $\mu(S) = 1$, to rozkład μ nazywamy **rozkładem dyskretnym**. Mówimy, że rozkład μ ma **funkcję masy prawdopodobieństwa** $p : \mathbb{R} \rightarrow \mathbb{R}$, zadaną przez $p(x) = \mu(\{x\})$.

Podobnie jak wcześniej, pewne zmiennych losowe są związane z rozkładami dyskretnymi i możemy do nich przyporządkować funkcję masy prawdopodobieństwa.

Definicja 3.22 (Dyskretna zmienna losowa i jej masa prawdopodobieństwa). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Jeżeli rozkład zmiennej X , tj. μ_X , jest dyskretny z funkcją masy prawdopodobieństwa p_X to mówimy, że zmienna losowa X jest **dyskretna** oraz ma **masę prawdopodobieństwa** p_X .

Masę prawdopodobieństwa dyskretnej zmiennej losowej X oznaczamy zazwyczaj przez p_X . Oczywiście dla rozkładu dyskretnego funkcją masy prawdopodobieństwa jednoznacznie identyfikuje nam rozkład prawdopodobieństwa. Funkcję masy prawdopodobieństwa często reprezentuje się w postaci tablicowej jako ciąg par $((x_i, p_i))_{i \in I}$, gdzie $I = \mathbb{N}$ lub $I = \{1, 2, \dots, n\}$ dla $n \in \mathbb{N}$.

Przykład 3.23 (Suma reszek w dwóch rzutach symetryczną monetą c.d.). Zmienna X wyznaczająca sumę reszek w dwóch rzutach monetą jest zmienną dyskretną; zob. Przykład 3.2. Funkcja masy prawdopodobieństwa zadana jest przez $p(x) = \frac{1}{4} \mathbb{1}_{\{0\}}(x) + \frac{1}{2} \mathbb{1}_{\{1\}}(x) + \frac{1}{4} \mathbb{1}_{\{2\}}(x)$ lub w postaci tablicowej przez

$$\begin{array}{c|ccc} x_i & 0 & 1 & 2 \\ \hline p_i & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}.$$

Uwaga 3.24 (Obłożenie zmiennej dyskretnej funkcją). Znając rozkład (masę prawdopodobieństwa) dyskretnej zmiennej losowej możemy stosunkowo łatwo policzyć rozkład (masę prawdopodobieństwa) zmiennej $Y = \phi(X)$, gdzie $\phi : \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją borelowską. Przykładowo, dla ściśle rosnącej funkcji ϕ wiemy, że masę X można wyrazić w postaci tablicowej jako $((\tilde{x}_i, p_i))_{i \in I}$, gdzie $\tilde{x}_i = \phi(x_i)$.

Przykład 3.25 (Suma reszek w dwóch rzutach monetą symetryczną c.d.). Rozważmy zmienną X z Przykładu 3.23 oraz zadajmy $Y := |X - 1|$. Zauważając, że $\mathbb{P}(\{Y = 1\}) = \mathbb{P}(\{X = 0\} \cup \{X = 2\})$ dostajemy funkcję masy prawdopodobieństwa zmiennej Y zadaną przez $p(x) = \frac{1}{2} \mathbb{1}_{\{0\}}(x) + \frac{1}{2} \mathbb{1}_{\{1\}}(x)$.

Istnieją oczywiście rozkłady, które nie są ani ciągłe ani dyskretne. Na teorii miary i całki dowiemy się, jak dokonać dekompozycji rozkładu prawdopodobieństwa na część *absolutnie ciągłą* i część *singularną*, którą dla tzw. rozkładów regularnych można zdekomponować na część *dyskretną* oraz część *singularną ciągłą*. Podsumowuje to tzw. *Twierdzenie Lebesgue'a o dekompozycji* (ang. Lebesgue's decomposition theorem). Jego sformułowanie oraz dowód wykracza poza materiał tego wykładu, zob. Twierdzenie 19.61 w [HS65].

Przykład 3.26 (Przykład rozkładu, który nie jest ani dyskretny, ani ciągły). Niech X ma rozkład jednostajny na odcinku $[0, 10]$ oraz niech $Y = \min(X, 5)$. Łatwo pokazać, że zmienna losowa Y nie ma rozkładu dyskretnego, gdyż np. $\mathbb{P}(Y = 5) = \frac{1}{2} > 0$ oraz $\mathbb{P}(Y < t) = \mathbb{P}(X < t)$ dla $t \leq 5$ (a X ma rozkład ciągły).

3.5 Dystrybuanta zmiennej losowej

Oprócz funkcji gęstości, czy funkcji masy prawdopodobieństwa, istnieją inne metody identyfikowania rozkładu prawdopodobieństwa. Najważniejsza z nich jest powiązana z tzw. *dystrybuantą*, czyli funkcją która jednoznacznie wyznacza rozkład i zależy od jednego parametru.

Definicja 3.27 (Dystrybuanta rozkładu). Niech μ będzie rozkładem prawdopodobieństwa na \mathbb{R} . **Dystrybuantą rozkładu** μ nazywamy funkcję $F_\mu: \mathbb{R} \rightarrow \mathbb{R}$, która zadana jest przez

$$F_\mu(t) := \mu((-\infty, t]), \quad t \in \mathbb{R}.$$

Jak widać, dystrybuanta podaje nam wartości miary μ na zbiorach generujących postaci $(-\infty, t]$. Pomimo tego, że rodzina ta jest indeksowana tylko jednym parametrem $t \in \mathbb{R}$, generuje nam on całą σ -algebrę zbiorów borelowskich i wystarcza do wyznaczenia miary μ . Tak, jak poprzednio, zwińmy teraz dystrybuanty rozkładu ze zmiennymi losowymi.

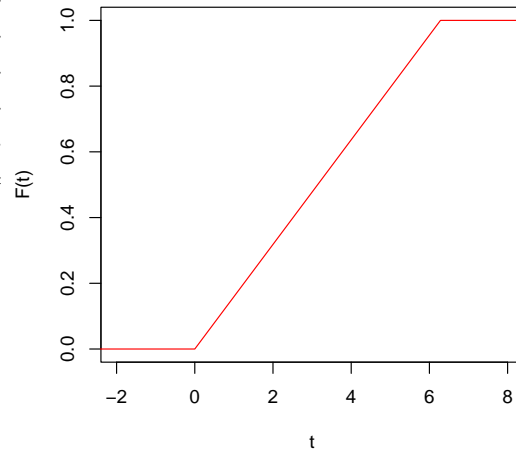
Definicja 3.28 (Dystrybuanta zmiennej losowej). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. **Dystrybuantą zmiennej losowej** X nazywamy funkcję określoną przez

$$F_X(t) := \mathbb{P}(X \leq t), \quad t \in \mathbb{R}.$$

Oczywiście z Definicji 3.28 dostajemy od razu, że dystrybuanta zmiennej losowej jest w istocie dystrybuantą rozkładu prawdopodobieństwa związanego z tą zmienną, tzn. zachodzi $F_X(t) = F_{\mu_X}(t)$ dla $t \in \mathbb{R}$.

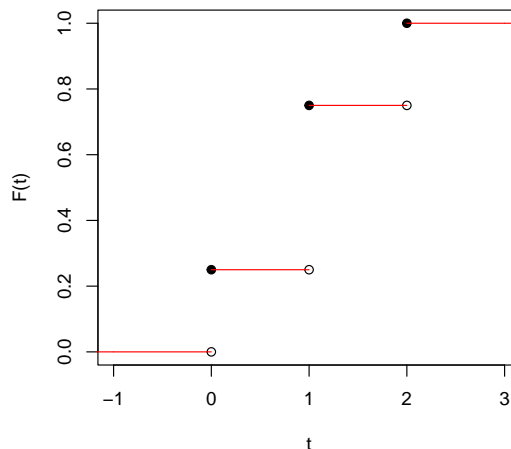
Przykład 3.29 (Dystrybuanta rozkładu jednostajnego na odcinku $(0, 2\pi)$). W Przykładzie 3.18 rozważaliśmy zmienną losową X o rozkładzie z funkcją gęstości $f_X(x) = \frac{1}{2\pi} \cdot \mathbb{1}_{(0, 2\pi)}(x)$, $x \in \mathbb{R}$. Dystrybuantę takiego rozkładu łatwo wyliczyć. Istotnie, zauważając, że dla $t \in [0, 2\pi]$ mamy $F_X(t) = \mu_X((-\infty, t]) = \int_{-\infty}^t f_X(t) dt = \frac{1}{2\pi} \int_0^t 1 dt = \frac{t}{2\pi}$, dostajemy

$$F_X(t) = \begin{cases} 0 & t < 0, \\ \frac{t}{2\pi} & t \in [0, 2\pi], \\ 1 & t > 2\pi. \end{cases}$$



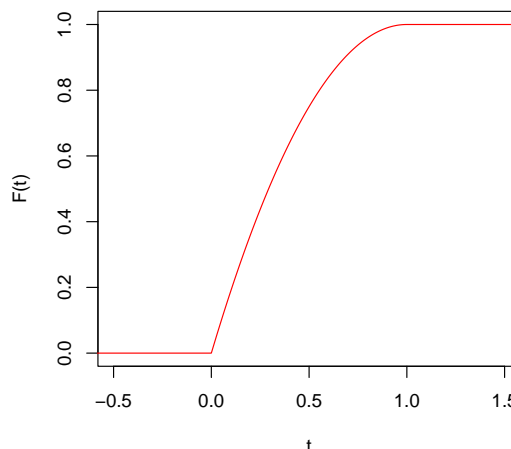
Przykład 3.30 (Suma reszek w dwóch rzutach symetryczną monetą c.d.). W Przykładzie 3.2 rozważaliśmy zmienną losową opisującą sumę reszek w dwóch rzutach monetą. Jej funkcja masy prawdopodobieństwa wynosiła $p(x) = \frac{1}{4}\mathbb{1}_{\{0\}}(x) + \frac{1}{2}\mathbb{1}_{\{1\}}(x) + \frac{1}{4}\mathbb{1}_{\{2\}}(x)$, skąd dostajemy wzór na dystrybuantę

$$F_X(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{4} & t \in [0, 1) \\ \frac{3}{4} & t \in [1, 2) \\ 1 & t \geq 2. \end{cases}$$



Przykład 3.31 (Losowanie dwóch liczb i ich bezwzględna różnica c.d.). W Przykładzie 3.14 policzyliśmy wartość $\mathbb{P}(X \leq t)$ dla $t \in [0, 1]$. Dostajemy stąd wprost wzór na dystrybuantę zmiennej X . Dystrybuanta ta wynosi

$$F_X(t) = \begin{cases} 0 & t < 0 \\ 1 - (1 - t)^2 & t \in [0, 1] \\ 1 & t > 1. \end{cases}$$



Podobnie jak w przypadku rozkładu chcielibyśmy określić jakie własności powinna spełniać funkcja $F: \mathbb{R} \rightarrow \mathbb{R}$ aby być dystrybuantą.

Twierdzenie 3.32 (Własności dystrybuanty). Niech F_μ będzie dystrybuantą dla pewnego rozkładu prawdopodobieństwa μ na \mathbb{R} . Wtedy

- 1) F_μ jest funkcją niemalejącą;
- 2) F_μ jest funkcją prawostronnie ciągłą;
- 3) $\lim_{t \rightarrow \infty} F_\mu(t) = 1$ oraz $\lim_{t \rightarrow -\infty} F_\mu(t) = 0$.

Ponadto, każda funkcja $F: \mathbb{R} \rightarrow \mathbb{R}$ spełniająca warunki 1), 2) i 3) jest dystrybuantą dla pewnego rozkładu μ na \mathbb{R} .

Dowód.

- 1) Wynika to wprost z własności miary, gdyż dla $t \leq s$ mamy $(-\infty, t] \subseteq (-\infty, s]$.

2) Niech $t \in \mathbb{R}$ oraz niech $(t_n)_{n \in \mathbb{N}}$ będzie nierosnącym ciągiem liczb spełniającym $t_n \searrow t$, $n \rightarrow \infty$. Zauważając, że rodzina zbiorów $(T_n)_{n \in \mathbb{N}}$, gdzie $T_n := (-\infty, t_n]$, jest zstępująca oraz korzystając z Propozycji 1.7 (dla miary μ) dostajemy

$$F_\mu(t) = \mu((-\infty, t]) = \mu\left(\bigcap_{t=1}^{\infty} T_n\right) = \lim_{n \rightarrow \infty} \mu((-\infty, t_n]) = \lim_{n \rightarrow \infty} F_\mu(t_n).$$

3) Postępujemy podobnie jak w punkcie 2), biorąc rodzinę $T_z := (-\infty, z]$, $z \in \mathbb{Z}$, oraz zauważając, że $\bigcap_{t=1}^{\infty} T_z = \emptyset$ i $\bigcup_{t=1}^{\infty} T_z = \mathbb{R}$; gdy z rośnie, to rodzina $(T_z)_{z \in \mathbb{Z}}$ jest wstępująca, a gdy z maleje, to rodzina $(T_z)_{z \in \mathbb{Z}}$ jest zstępująca.

Pokazanie, że funkcja spełniająca 1), 2) i 3) jest dystrybuantą pewnego rozkładu jest rutynowym zadaniem z teorii miary i pominiemy tego dowód (ćw. do domu). \square

Warto nadmienić, że istnieje wzajemna jednoznaczność pomiędzy rozkładami prawdopodobieństwa, a dystrybuantami. Innymi słowy, jeżeli μ i ν są rozkładami o równych dystrybuantach $F_\mu = F_\nu$, to rozkłady μ i ν są sobie równe (tzn. $\mu(A) = \nu(A)$ dla dowolnego $A \in \mathcal{B}(\mathbb{R})$). O ile funkcja gęstości służyła do charakteryzacji tylko pewnej podklasy rozkładów (absolutnie ciągłych), o tyle dystrybuanta charakteryzuje nam już jednoznacznie każdy możliwy rozkład (z dokładnością do zbiorów miary zero).

Uwaga 3.33 (Obłożenie zmiennej funkcją rosnącą, a dystrybuantą). Niech X będzie zmienną określoną na $(\Omega, \Sigma, \mathbb{P})$, oraz niech $Y = \varphi(X)$, gdzie $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ jest borelowską funkcją ściśle rosnącą. Łatwo zauważyć, że dystrybuantę zmiennej Y można przedstawić przy pomocy dystrybuanty zmiennej X , gdyż dla dowolnego $t \in \mathbb{R}$ zachodzi $F_Y(t) = \mathbb{P}(\varphi(X) \leq t) = \mathbb{P}(X \leq \varphi^{-1}(t)) = F_X(\varphi^{-1}(t))$. W szczególności, dla $Y = aX + b$, gdzie $a > 0$ i $b \in \mathbb{R}$ dostajemy

$$F_Y(t) = \mathbb{P}(aX + b \leq t) = \mathbb{P}(X \leq \frac{t-b}{a}) = F_X(\frac{t-b}{a}), \quad t \in \mathbb{R}.$$

3.6 Związek między dystrybuantą, a gęstością

Jeżeli mamy do czynienia ze zmienną losową X o rozkładzie absolutnie ciągłym z gęstością f_X , to oczywiście dystrybuantę tej zmiennej możemy wyliczyć ze wzoru

$$F_X(t) = \int_{-\infty}^t f_X(x) dx.$$

Powyższy wzór można też odwrócić, tzn. dla absolutnie ciągłych zmiennych losowych X pochodna dystrybuanty daje nam gęstość, czyli zachodzi $F'_X(t) = f_X(t)$ (p.w.).

Oczywiście nie zawsze rozważając pochodną F' , a następnie ją całkując dostaniemy dystrybuantę F – będzie tak tylko dla funkcji F , które są absolutnie ciągłe (dlatego też tak nazywa się powiązane zmienne losowe); zob. strona 438 w [Bil12], gdzie w szczególności podana jest definicja absolutnej ciągłości. Warto tutaj zaznaczyć, że aby dostać $F'_X(t) = f_X(t)$ nie wystarczy założyć ciągłości (w zwykłym sensie) funkcji F_X . Przedstawmy na koniec twierdzenie, które pokazuje nam, jak szybko sprawdzić, kiedy F' jest gęstością.

Twierdzenie 3.34. Niech F będzie dystrybuantą dla której F' istnieje prawie wszędzie oraz $\int_{-\infty}^{\infty} F'(x) dx = 1$. Wtedy F' jest gęstością prawdopodobieństwa rozkładu o dystrybuancie F .

Dowód Twierdzenia 3.34 opiera się na standardowych technikach teorii miary i całki. Jego dowód można znaleźć np. w Rozdziale 5.4 w [JS04].

Uwaga 3.35 (Gładka transformacja absolutnie ciągłej zmiennej losowej). Zakładając, że X jest absolutnie ciągła, a funkcja $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ jest odpowiedni gładka, zmienna losowa $Y := \varphi(X)$ będzie również absolutnie ciągła i w stosunkowo łatwy sposób możemy policzyć jej gęstość. Przykładowo, zakładając, że funkcja φ jest ściśle rosnącą funkcją klasy C^1 , całkując przez podstawienie, dostajemy

$$F_Y(t) = P(X \leq \varphi^{-1}(t)) = \int_{\varphi^{-1}((-\infty, t])} f_X(x) dx = \int_{-\infty}^t f(\varphi^{-1}(u)) \cdot |\varphi^{-1}(u)| du,$$

co daje nam wzór na gęstość $f_Y(t) = f(\varphi^{-1}(u)) \cdot |\varphi^{-1}(u)|$. Więcej szczegółów na ten temat można znaleźć w Rozdziale 5.5 w [JS04].

4 Wektory losowe i zależność między zmiennymi losowymi

W poprzednim rozdziale zdefiniowaliśmy wielowymiarowy odpowiednik zmiennej losowej, tj. wektor losowy. Większość definicji z Rozdziału 3 przenosi się w naturalny sposób na przypadek wielowymiarowe. W rozdziale tym przedstawimy niektóre własności dystrybuant wielowymiarowych oraz wprowadzimy pojęcie *niezależności* zmiennych losowych. Podobnie jak wcześniej z każdym wektorem losowym można związać wielowymiarowy rozkład prawdopodobieństwa, czyli miarę probabilistyczną $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}^n$ określoną na $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

4.1 Dystrybuanta wielowymiarowa

Podobnie jak w przypadku zmiennych losowych, rozkłady wektorów losowych można charakteryzować na różne sposoby dostając w szczególności klasy absolutnie ciągłych wektorów losowych, czy dyskretnych wektorów losowych. Z każdym rozkładem można też (w sposób jednoznaczny) związać wielowymiarową dystrybuantę.

Definicja 4.1 (Dystrybuanta wektora losowego). Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie wektorem losowym określonym na $(\Omega, \Sigma, \mathbb{P})$. **Dystrybuantą wektora losowego \mathbf{X}** nazywamy funkcję $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$ określoną przez

$$F_{\mathbf{X}}(t_1, \dots, t_n) := \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n), \quad (t_1, \dots, t_n) \in \mathbb{R}^n.$$

^aZapis $\mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n)$ to uproszczone oznaczenie $\mathbb{P}(\{X_1 \leq t_1\} \cap \dots \cap \{X_n \leq t_n\})$.

Oczywiście, podobnie jak wcześniej, dostajemy $F_{\mathbf{X}} = F_{\mu_{\mathbf{X}}}$, gdzie jest rozkładem prawdopodobieństwa na \mathbb{R}^n związanym z \mathbf{X} .

Przykład 4.2 (Dwa rzuty monetą symetryczną). Niech $(\Omega^2, \Sigma^2, \mathbb{P}^2)$ będzie przestrzenią probabilistyczną odpowiadającą iloczynowi kartezjańskiemu $\Omega = \{0, R\}$ ze schematem klasycznym, co odpowiada 2-krotnemu rzutowi monetą. Wtedy wektorem losowym opisującym wynik obu rzutów będzie wektor $\mathbf{X} = (X_1, X_2)$, gdzie $X_1(\omega_1, \omega_2) = \mathbb{1}_{\{R\}}(\omega_1)$ oraz $X_2(\omega_1, \omega_2) = \mathbb{1}_{\{R\}}(\omega_2)$. Dystrybuanta takiego wektora wyraża się wzorem $F_{\mathbf{X}}(t_1, t_2) = F_{X_1}(t_1) \cdot F_{X_2}(t_2)$ co można wprost zapisać jako

$$F_{\mathbf{X}}(t_1, t_2) = \begin{cases} 0 & t_1 < 0 \vee t_2 < 0 \\ \frac{1}{4} & t_1 \in [0, 1) \wedge t_2 \in [0, 1) \\ \frac{1}{2} & (t_1 \in [0, 1) \wedge t_2 \geq 1) \vee (t_1 \geq 1 \wedge t_2 \in [0, 1)) \\ 1 & t_1 \geq 1 \wedge t_2 \geq 1. \end{cases}$$

Podobnie jak w przypadku jednowymiarowym, można również scharakteryzować funkcje będące dystrybuantami.

Twierdzenie 4.3 (Własności dystrybuanty). Niech F_μ będzie dystrybuantą dla pewnego rozkładu prawdopodobieństwa μ na \mathbb{R}^n . Wtedy

- 1) F_μ jest funkcją niemalejącą względem każdego argumentu;
- 2) F_μ jest tzw. funkcją *n-rosnącą*, tzn. dla $t, s \in \mathbb{R}^n$ takich, że $t \leq s$ (po wszystkich współrzędnych), zachodzi

$$\sum_{\epsilon \in U} (-1)^{\sum_{i=1}^n \epsilon_i} F_\mu(\epsilon \cdot t + (1 - \epsilon) \cdot s) \geq 0$$

gdzie mnożenie (\cdot) jest po współrzędnych^a oraz $U = \{\epsilon \in \mathbb{R}^n : \epsilon_k \in \{0, 1\} \text{ dla } k = 1, 2, \dots, n\}$;

- 3) F_μ jest funkcją prawostronnie ciągłą;
- 4) $F_\mu(t_1, \dots, t_n) \rightarrow 1$ jeżeli $\inf_{i=1, \dots, n} t_i \rightarrow \infty$ oraz $F_\mu(t) = 0$ jeżeli $\inf_{i=1, \dots, n} t_i \rightarrow -\infty$.

Ponadto, każda funkcja $F: \mathbb{R}^n \rightarrow \mathbb{R}$ spełniająca warunki 1), 2), 3) i 4) jest dystrybuantą dla pewnego rozkładu μ na \mathbb{R}^n .

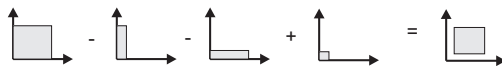
^atzn. $\epsilon \cdot t + (1 - \epsilon) \cdot s = (\epsilon_1 t_1 + (1 - \epsilon_1) s_1, \dots, \epsilon_n t_n + (1 - \epsilon_n) s_n)$ dla $t = (t_1, \dots, t_n)$ i $s = (s_1, \dots, s_n)$

Dowód Twierdzenia 4.3 wykracza poza materiał tego wykładu, zob. Rozdział 5.3 w [JS04]. Podobnie jak wcześniej istnieje wzajemna jednoznaczność pomiędzy rozkładami prawdopodobieństwa, a dystrybuantami. Innymi słowy, jeżeli μ i ν są rozkładami o równych dystrybuantach $F_\mu = F_\nu$, to rozkłady μ i ν są sobie równe (tzn. $\mu(A) = \nu(A)$ dla dowolnego $A \in \mathcal{B}(\mathbb{R}^n)$).

Uwaga 4.4 (Funkcje *n-rosnące*). Warunek drugi w Twierdzeniu 4.3 jest związany z tzw. objętością funkcji (ang. *function volume*). Dla $n = 2$ i $t \leq s$ (po współrzędnych) warunek ten sprowadza się do nierówności

$$F(s_1, s_2) - F(t_1, s_2) - F(s_1, t_2) + F(t_1, t_2) \geq 0,$$

co można traktować jako "objętość" funkcji na kwadracie $[t_1, s_1] \times [t_2, s_2]$. Każdy element powyżej można traktować jako "objętość" funkcji na pewnym prostokącie, co graficznie można zilustrować jako



Więcej szczegółów na ten temat można znaleźć w książce [Nel07]

Dla dyskretnych dwuwymiarowych wektorów losowych, częstą praktyką na scharakteryzowanie rozkładu jest podanie tablicy zawierającej prawdopodobieństwa przyjęcia przez wektor poszczególnych wartości. Podobnie jak to miało miejsce w Rozdziale 3.4, mając dany wektor losowy $\mathbf{X} = (X, Y)$, rozkład dyskretny można scharakteryzować zadając tablicę $(p_{ij})_{i \in I, j \in J}$, gdzie I odpowiada ilości wartości, które może przyjąć zmienna X , J odpowiada ilości wartości, które może przyjąć zmienna Y , a p_{ij} to prawdopodobieństwo, że X przyjmie i -tą wartość oraz jednocześnie Y przyjmie swoją j -tą wartość.

Przykład 4.5 (Dwa rzuty monetą symetryczną c.d.). Rozkład dyskretnego wektora losowego z Przykładu 4.2 można scharakteryzować zadając tablicę

$X \backslash Y$	0	1
0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$

Z tablicy tej łatwo obliczyć prawdopodobieństwo różnych zdarzeń, np. $\mathbb{P}(X = 0, Y = 1) = \frac{1}{4}$ lub $\mathbb{P}(X = 0) = \mathbb{P}(X = 0, Y \in \{0, 1\}) = \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 0, Y = 1) = \frac{1}{2}$.

4.2 Dystrybuanty brzegowe i twierdzenie Skłara

Mając dany wektor losowy $\mathbf{X} = (X_1, \dots, X_n)$ naturalne wydaje się rozważanie rozkładów jego współrzędnych i badanie ich związku z dystrybuantą.

Definicja 4.6 (Dystrybuanta brzegowa wektora losowego). Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie wektorem losowym określonym na $(\Omega, \Sigma, \mathbb{P})$. Wtedy dla każdego $i = 1, 2, \dots, n$, przez i -tą **dystrybuantę brzegową** wektora \mathbf{X} rozumiemy funkcję $F_i : \mathbb{R} \rightarrow \mathbb{R}$ zadaną przez

$$F_i(t) = \mathbb{P}(X_i \leq t), \quad t \in \mathbb{R}.$$

Warto zauważyć, że możemy nieformalnie zapisać $F_i(t) = \mathbb{F}_{\mathbf{X}}(T_i)$, gdzie $T_i \in \bar{\mathbb{R}}^n$ to wektor, który na i -tej pozycji ma t_i a na pozostałych pozycjach ∞ .^a

^aAby ująć wszystko formalnie, musielibyśmy rozszerzyć funkcję $\mathbb{F}_{\mathbf{X}}$ z \mathbb{R}^n do $\bar{\mathbb{R}}^n$, zob. Definicja 4.1.

Widać więc, że z każdą dystrybuantą wielowymiarową możemy związać n jednowymiarowych dystrybuant brzegowych odpowiadających *brzegowym zmiennym losowym* X_1, \dots, X_n . Oczywiście znając same dystrybuanty jednowymiarowe, nie jesteśmy w stanie odtworzyć dystrybuanty wielowymiarowej, co będą ilustrować dwa kolejne przykłady.

Często, w przypadku dwuwymiarowych rozkładów dyskretnych, dodaje się informację o rozkładach brzegowych do tablicy (masy prawdopodobieństwa). Ilustruje to Przykład 4.2.

Przykład 4.7 (Dwa rzuty monetą symetryczną c.d.). Tablicę opisującą rozkład dyskretnego wektora losowego $\mathbf{X} = (X, Y)$ z Przykładu 4.5 można uzupełnić o informację o rozkładach brzegowych dostając

$X \backslash Y$	0	1	X
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
Y	$\frac{1}{2}$	$\frac{1}{2}$	

Warto zauważyć, że wartość w ostatniej kolumnie (odp. ostatnim rzędzie) to nic innego, jak suma wszystkich wartości w danym rzędzie (odp. kolumnie), co istotnie odpowiada rozkładowi brzegowemu. Przykładowo, $\mathbb{P}(Y = 0) = \mathbb{P}(X \in \{0, 1\}, Y = 0) = \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 1, Y = 0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Przykład 4.8 (Rozkłady brzegowe nie wyznaczają jednoznacznie rozkładu). Rozważmy wektor losowy $\mathbf{X} = (X, Y)$ z Przykładu 4.5 oraz wektor $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{Y})$, gdzie $\tilde{X} = \tilde{Y} = X$. Rozkład wektora $\tilde{\mathbf{X}}$ można przedstawić w postaci tablicy

$\tilde{X} \backslash \tilde{Y}$	0	1	\tilde{X}
0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	0	$\frac{1}{2}$	$\frac{1}{2}$
\tilde{Y}	$\frac{1}{2}$	$\frac{1}{2}$	

Widać, że rozkłady wektorów losowych \mathbf{X} oraz $\tilde{\mathbf{X}}$ są różne, choć rozkłady brzegowe się pokrywają. Rozkłady brzegowe nie pozwalają nam więc na pełne określenie rozkładu brzegowego.

Analizując przykład oraz pojawia się naturalne pytanie: Co jesteśmy w stanie powiedzieć o dystrybucji wektora losowego, znając jego dystrybuanty brzegowe oraz, czy można formalnie związać dystrybuantę łączną z brzegowymi. Okazuje się, że można dokonać dekompozycji dystrybuanty na rozkłady brzegowe oraz tzw. *funkcje copula* (ang. *copula*) opisującą zależność między zmiennymi brzegowymi.⁸

Twierdzenie 4.9 (Twierdzenie Sklara). Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie wektorem losowym określonym na $(\Omega, \Sigma, \mathbb{P})$. Niech F oznacza dystrybuantę wektora \mathbf{X} oraz niech F_i oznacza dystrybuantę brzegową zmiennej losowej X_i , $i = 1, \dots, n$. Wtedy istnieje funkcja $C : [0, 1] \rightarrow [0, 1]$ taka, że

$$F(t_1, \dots, t_n) = C(F_1(t_1), \dots, F_n(t_n)).$$

Funkcja C jest określona jednoznacznie, jeżeli dystrybuanty brzegowe są ciągłe.⁹ Warto wspomnieć, że funkcja C w istocie opisuje zależność między zmiennymi brzegowymi i jest ona dystrybuantą dla której brzegi mają rozkład jednostajny na odcinku $[0, 1]$; zależność wyrażona jest poprzez unormowanie brzegów. Prawdziwe jest też twierdzenie odwrotne, które mówi, że znając funkcje brzegowe oraz funkcję copula rozkładu jesteśmy w stanie odtworzyć jego dystrybuantę. Dowód Twierdzenia 4.9 oraz więcej własności funkcji C (wraz z ich dowodami) można znaleźć w [Nel07].

4.3 Niezależność zmiennych losowych

W Rozdziale 2 podaliśmy definicję niezależności zdarzeń $A, B \in \Sigma$ dla ustalonej przestrzeni $(\Omega, \Sigma, \mathbb{P})$. Definicja ta w naturalny sposób przenosi się na zmienne losowe, co pozwoli nam mówić formalnie np. o niezależnych rzutach monetą.

Definicja 4.10 (Niezależność zmiennych losowych). Niech X_1, \dots, X_n będą zmiennymi losowymi określonymi na $(\Omega, \Sigma, \mathbb{P})$. Zmienne losowe X_1, \dots, X_n nazywamy **zmiennymi niezależnymi**, gdy dla każdego ciągu zbiorów borelowskich $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ zachodzi

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdot \dots \cdot \mathbb{P}(X_n \in B_n).$$

Mówimy, że zmienne losowe są **zależne**, gdy nie są niezależne.

Mówiąc inaczej, zmienne są niezależne, gdy dowolne zdarzenia generowane przez te zmienne są niezależne; podobnie jak wcześniej, definicję niezależności można rozszerzyć na przeliczalne rodziny zmiennych losowych.

Uwaga 4.11 (Niezależność zmiennych losowych, a niezależność σ -algebr generowanych przez zmienne). Definicji 4.10 można wyrazić również w języku niezależności σ -algebr generowanych przez zmienne losowe; dla uproszczenia ograniczmy się do przypadku $n = 2$. Niech $(\Omega, \Sigma, \mathbb{P})$ będzie przestrzenią probabilistyczną, a Σ_1 i Σ_2 określonymi na niej σ -algebrami (takimi, że $\Sigma_1, \Sigma_2 \subset \Sigma$). Mówimy, że σ -algebry Σ_1 i Σ_2 są niezależne, gdy dla dowolnego $A_1 \in \Sigma_1$ oraz $A_2 \in \Sigma_2$ zachodzi $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$. Korzystając z tej definicji wiemy, że zmienne losowe X_1 oraz X_2 są niezależne wtedy i tylko wtedy, gdy σ -algebry generowane przez te zmienne, tj. $\sigma(X_1)$ oraz $\sigma(X_2)$ są niezależne.

⁸W polskiej literaturze występuje wiele nazw funkcji copula, np. kopuła, funkcja łącząca, funkcja łącznikowa.

⁹W przeciwnym przypadku jest ona określona jednoznacznie na zbiorze wartości funkcji

Pokażmy teraz dwa przykłady.

Przykład 4.12 (Przykład niezależnych zmiennych losowych). Rozważmy wektor losowy $\mathbf{X} = (X, Y)$ = określone w Przykładzie 4.2. Zmienne losowe X oraz Y są niezależne, tzn. dla każdego $B, C \in \mathcal{B}(b\mathbb{R})$ zachodzi $\mathbb{P}(X \in B, Y \in C) = \mathbb{P}(X \in B) \cdot \mathbb{P}(Y \in C)$. Jeżeli $B \cap \{0, 1\} = \emptyset$ lub $B \cap \{0, 1\} = B$, to dostajemy $\mathbb{P}(X \in B) = 0$ lub $\mathbb{P}(X \in B) = 1$, czyli równość $\mathbb{P}(X \in B, Y \in C) = \mathbb{P}(X \in B) \cdot \mathbb{P}(Y \in C)$ jest spełniona dla każdego $C \in \mathcal{B}(\mathbb{R})$. Podobnie, jeżeli $C \cap \{0, 1\} = \emptyset$ lub $C \cap \{0, 1\} = C$, to równość jest spełniona dla każdego $B \in \mathcal{B}(\mathbb{R})$. W pozostałych sytuacjach dostajemy $\mathbb{P}(X \in B, Y \in C) = \frac{1}{4}$ oraz $\mathbb{P}(X \in B) = \mathbb{P}(Y \in C) = \frac{1}{2}$, co kończy dowód.

Przykład 4.13 (Przykład zależnych zmiennych losowych). Niech X_1 oraz X_2 oznacza liczbę oczek w dwóch (niezależnych rzutach kostką). Definiujemy $Y_1 = \min(X_1, X_2)$ oraz $Y_2 = \max(X_1, X_2)$. Wtedy zmienne losowe Y_1 oraz Y_2 są od siebie zależne, gdyż np. $\mathbb{P}(Y_1 = 6, Y_2 = 1) = 0$ podczas, gdy $\mathbb{P}(Y_1 = 6) = \mathbb{P}(Y_2 = 6) = \frac{1}{36} \neq 0$.

Niezależność zmiennych losowych można opisać na wiele sposobów, przy wykorzystaniu ich rozkładu łącznego, czy dystrybuanty.

Propozycja 4.14 (Warunki równoważne na niezależność zmiennych losowych). Niech X_1, \dots, X_n będą zmiennymi losowymi określonymi na $(\Omega, \Sigma, \mathbb{P})$. Wtedy następujące warunki są równoważne

- 1) zmienne losowe X_1, \dots, X_n są niezależne;
- 2) $F_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_n}(t_n)$ dla $(t_1, \dots, t_n) \in \mathbb{R}^n$.
- 3) $\mu_{(X_1, \dots, X_n)} = \mu_{X_1} \otimes \dots \otimes \mu_{X_n}$;

Dowód.

[1] \Rightarrow 2)] Wystarczy rozważyć zbiory postaci wziąć $B_i = (-\infty, t]$ dla $i = 1, \dots, n$ oraz $t \in \mathbb{R}$.

[2] \Rightarrow 3)] Niech \tilde{F} będzie dystrybuanta rozkładu iloczynu kartezjańskiego $\mu_{X_1} \otimes \dots \otimes \mu_{X_n}$. Wtedy

$$\tilde{F}(t_1, \dots, t_n) = (\mu_{X_1} \otimes \dots \otimes \mu_{X_n})((-\infty, t_1] \times \dots \times (-\infty, t_n]) = \mu_{X_1}((-\infty, t_1]) \cdot \dots \cdot \mu_{X_n}((-\infty, t_n]),$$

co daje nam $\tilde{F}(t_1, \dots, t_n) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_n}(t_n) = F_{(X_1, \dots, X_n)}(t_1, \dots, t_n)$. Równość dystrybuant \tilde{F} oraz $F_{(X_1, \dots, X_n)}$ implikuje równość powiązanych rozkładów, co kończy dowód.

[3] \Rightarrow 1)] Dla dowolnych $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ zachodzi

$$\begin{aligned} \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) &= \mu_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) \\ &= \mu_{(X_1)}(B_1) \cdot \dots \cdot \mu_{(X_n)}(B_n) \\ &= \mathbb{P}(X_1 \in B_1) \cdot \dots \cdot \mathbb{P}(X_n \in B_n), \end{aligned}$$

co kończy dowód. □

Z Propozycji 4.14 można również wywnioskować warunki równoważne na niezależność dla rozkładów ciągłych i dyskretnych, które są oparte o masę prawdopodobieństwa, czy gęstość. Istotnie, jeżeli rozkład (X_1, \dots, X_n) jest dyskretny, to warunkiem równoważnym jest równość

$$\mathbb{P}(X_1 = a_1, \dots, X_n = a_n) = \mathbb{P}(X_1 = a_1) \cdot \dots \cdot \mathbb{P}(X_n = a_n),$$

gdzie $(a_i) \in \mathbb{R}$ to zbiór wartości wektora X_i . Jeżeli rozkład wektora (X_1, \dots, X_n) jest z kolei ciągły o gęstości f , to warunkiem równoważnym na niezależność jest równość (p.n.) funkcji

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n),$$

dla $(x_1, \dots, x_n) \in \mathbb{R}^n$, gdzie f_i jest gęstością zmiennej X_i , $i = 1, \dots, n$.¹⁰

Na koniec przedstawmy propozycję, która mówi, że obłożenie niezależnych zmiennych losowych funkcjami borelowskimi nie wpływa na ich niezależność;

Propozycja 4.15. Niech X_1, \dots, X_n będą niezależnymi zmiennymi losowymi określonymi na $(\Omega, \Sigma, \mathbb{P})$ oraz niech $\varphi_i : \mathbb{R} \rightarrow \mathbb{R}$ będą funkcjami borelowskimi, $i = 1, \dots, n$. Wtedy, zmienne losowe $\varphi_1(X_1), \dots, \varphi_n(X_n)$ są niezależne.

Propozycję 4.15 można udowodnić np. analizując jak obłożenie zmiennej losowej funkcją borelowską wpływa na generowaną σ -algebrę. Po pełny dowód tej propozycji odsyłamy do Rozdziału 5.8 w [JS04]. Warto zaznaczyć, że podobny wynik jest prawdziwy, gdy obkładamy funkcją borelowską odpowiednie wektory losowe (np. powstałe z niezależnych ciągów zmiennych losowych).

W6
-
W7

¹⁰Jeżeli wektor losowy (X_1, \dots, X_n) jest absolutnie ciągły, to jest brzegi też są absolutnie ciągłe, co wynika np. z twierdzenia Fubinięgo; więcej informacji na ten temat będzie podanych na wykładzie z teorii miary i całki.

5 Wartość oczekiwana i inne charakterystyki zmiennych losowych

Czasami zamiast rozważać cały rozkład zmiennej losowej (czy wektora losowego) wystarczy podać jej pewne charakterystyki, czy parametry, które identyfikują jednoznacznie pewien rozkład. Takich charakterystyk opisujących jest oczywiście bardzo dużo – najczęściej używanymi są *wartość oczekiwana* opisująca nam średnią wartość, którą przyjmuje zmienna losowa oraz *wariancja* opisująca rozrzut średnio-kwadratowy wokół średniej. W przypadku wielowymiarowym chcielibyśmy z kolei opisać zależność między zmiennymi przy pomocy charakterystyk liczbowych, czemu może służyć tzw. *kowariancja* lub jej unormowany odpowiednik nazywany *korelacją*.

5.1 Wartość oczekiwana i jej podstawowe własności

Zanim przejdziemy do ogólnej definicji wartości oczekiwanej, aby lepiej zrozumieć intuicję, podajmy jej szczególną postać dla rozkładów dyskretnych skończonych i ciągłych.

Definicja 5.1 (Wartość oczekiwana dla rozkładu dyskretnego). Niech X będzie dyskretną zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$ przyjmującą wartości ze zbioru przeliczalnego $S = \{x_1, x_2, \dots\}$. Wtedy, **wartość oczekiwana dyskretnej zmiennej losowej X** – o ile istnieje – oznaczamy przez $\mathbb{E}(X)$ i definiujemy jako

$$\mathbb{E}(X) := \sum_{i=1}^{\infty} x_i \cdot \mathbb{P}(X = x_i). \quad (5.1)$$

We wzorze (5.1) mnożymy każdą wartość zmiennej losowej przez prawdopodobieństwo jej wystąpienia, co daje nam średnią wartość jaką przyjmuje nasza zmienna. Dla rozkładów dyskretnych skończonych rozważamy oczywiście skończoną sumę w (5.1). Najlepiej ilustruje to prosty przykład.

Przykład 5.2 (Średnia liczba oczek w rzucie kostką). Niech zmienna losowa X ma rozkład odpowiadający liczbie oczek w standardowym rzucie kostką, tzn. wartości zmiennej są w zbiorze $S := \{1, \dots, 6\}$ oraz $\mathbb{P}(X = x) = \frac{1}{6}$ dla każdego $x \in S$. Wtedy

$$\mathbb{E}(X) = \sum_{i=1}^6 i \cdot \mathbb{P}(X = i) = \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{21}{6} = 3.5.$$

Innymi słowy, wykonując wiele rzutów i zapisując ich wynik na kartce, a następnie biorąc średnią, otrzymalibyśmy liczbę bliską 3.5 (a w granicy równą dokładnie 3.5); oczywiście wartość ta nie musi być (bezpośrednio) osiągnięta w rzutach.

Definicja 5.3 (Wartość oczekiwana dla rozkładu ciągłego). Niech X będzie ciągłą zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$ z funkcją gęstości f . Wtedy, **wartość oczekiwana ciągłej zmiennej losowej X** – o ile istnieje – oznaczamy przez $\mathbb{E}(X)$ i definiujemy jako

$$\mathbb{E}(X) := \int_{-\infty}^{\infty} x \cdot f(x) dx. \quad (5.2)$$

Wzór (5.2) można traktować jako odpowiednik (5.1) dla rozkładów ciągłych, podobnie jak to ma miejsce w standardowej definicji całki.

Przykład 5.4 (Losowanie liczby z odcinka $(0, 1)$). Niech zmienna losowa X ma rozkład jednostajny na odcinku $(0, 1)$ wyrażony funkcją gęstości $f(x) = \mathbb{1}_{(0,1)}(x)$. Wartość oczekiwana tej zmiennej wynosi

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \mathbb{1}_{(0,1)}(x) dx = \int_0^1 x dx = 0.5.$$

W ogólnym przypadku wartość oczekiwaną definiujemy korzystając z definicji całki względem miary. Definicja ta powinna być podana i dokładnie omówiona na wykładzie z teorii miary i całki, zob. Dodatek B.

Definicja 5.5 (Wartość oczekiwana). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, **wartość oczekiwaną zmiennej losowej** X oznaczamy przez $\mathbb{E}(X)$ – o ile istnieje – i definiujemy jako

$$\mathbb{E}(X) := \int_{\Omega} X d\mathbb{P}. \quad (5.3)$$

Łatwo sprawdzić, że wartości oczekiwane dla rozkładów ciągłych i dyskretnych podane w Definicji 5.1 oraz Definicji 5.3 pokrywają się z ogólnym wzorem podanym w Definicji 5.5. Z teorii miary i całki wiemy, że wartość oczekiwana to nic innego jak całka względem miary probabilistycznej. Oczywiście nie musi ona istnieć lub może być nieskończona. Warunkiem koniecznym i wystarczającym na istnienie i skończoność wartości oczekiwanej jest warunek $\int_{\Omega} |X| d\mathbb{P} < \infty$.¹¹ Zmienne losowe o skończonej wartości oczekiwanej będziemy nazywać czasami *całkowalnymi*, a zbiór całkowalnych zmiennych losowych na przestrzeni $(\Omega, \Sigma, \mathbb{P})$ oznaczać będziemy przez $L^1(\Omega, \Sigma, \mathbb{P})$.¹²

Uwaga 5.6 (Wartość oczekiwana a rozkład prawdopodobieństwa zmiennej losowej). Wartość oczekiwana zmiennej losowej jest jednoznacznie wyznaczona przez rozkład tej zmiennej. Istotnie dla całkowalnej zmiennej losowej X zachodzi równość $\int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} \text{Id} d\mu_X$, gdzie Id to funkcja tożsamościowa, co pokazuje nam jak związać wartość oczekiwaną z rozkładem.

Uwaga 5.7 (Różne oznaczenia całki względem miary). W literaturze można spotkać wiele konwencji zapisu całki względem miary probabilistycznej. Również w tym wykładzie, mając daną zmienną X , zamiast używać zapisu jak w Definicji 5.5 będziemy pisać $\int_{\Omega} X(\omega)P(d\omega)$. Podobnie, zamiast $\int_{\mathbb{R}} \text{Id} d\mu_X$ możemy używać zapisu $\int_{\mathbb{R}} x\mu_X(dx)$ lub $\int_{\mathbb{R}} x d\mu_X(x)$.

Wartość oczekiwana spełnia szereg własności, które powinniśmy już znać z analizy matematycznej oraz teorii miary i całki. Podsumujmy teraz podstawowe z nich bez podania dowodu.

¹¹Warto tutaj przypomnieć, że całka z funkcji nieujemnej jest zawsze dobrze określona, choć może być nieskończona.

¹²Formalnie rzecz biorąc elementami przestrzeni $L^1(\Omega, \Sigma, \mathbb{P})$ są klasy abstrakcji zmiennych losowych, które są sobie równe prawie wszędzie, por. Uwaga 3.5.

Propozycja 5.8 (Podstawowe własności wartości oczekiwanej). Niech $X, Y \in L^1(\Omega, \Sigma, \mathbb{P})$. Wtedy

- 1) (monotoniczność) Jeżeli $X \geq Y$, to $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- 2) (nierówność modułowa) $\mathbb{E}|X| \geq |\mathbb{E}(X)|$;
- 3) (addytywność) $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ dla dowolnych $a, b \in \mathbb{R}$;^a
- 4) (lemat Fatou) Jeżeli $X_n \geq 0$, to $\mathbb{E}(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n)$;
- 5) (twierdzenie Lebesgue'a o zbieżności monotonicznej) Jeżeli (X_n) jest niemalejącym ciągiem nieujemnych zmiennych losowych, to $\mathbb{E}(\lim_{n \rightarrow \infty} X_n) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n)$;
- 6) (twierdzenie Lebesgue'a o zbieżności zmajoryzowanej) Jeżeli (X_n) jest ciągiem zmiennych takim, że $|X_n| \leq Z$ dla pewnej całkowalnej zmiennej losowej Z , to $\mathbb{E}(\lim_{n \rightarrow \infty} X_n) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n)$;

^aW szczególności, dla wektora $aX + bY$ istnieje wartość oczekiwana.

Warto zauważyć, że addytywność operatora wartości oczekiwanej przenosi się na większą liczbę zmiennych losowych i pozwala nam w łatwy sposób obliczyć wartość oczekiwaną w sytuacji, gdy powtarzamy eksperyment, sumując łączny wynik.

Przykład 5.9 (Średnia liczba oczek w 10 rzutach kostką). Załóżmy, że chcemy poznać średnią wartość liczby oczek, które wypadną w 10 rzutach kostką. Oznaczając przez X_i ilość oczek w i -tym rzucie dostajemy

$$\mathbb{E}(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} \mathbb{E}(X_i) = 10 \cdot \mathbb{E}(X_1) = 10 \cdot 3.5 = 35.$$

Często będziemy chcieli obliczyć średnią wartość zmiennej losowej, po obłożeniu jej przez funkcję borelowską. Podajmy teraz ogólne twierdzenie, które pozwala to wyliczyć; dowód tego twierdzenia opiera się na standardowej metodzie stopniowego komplikowania funkcji i powinien być podany na teorii miary i całki w ogólnym przypadku.

Twierdzenie 5.10 (Wartość oczekiwana zmiennej losowej obłożonej funkcją borelowską). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$ oraz niech $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją borelowską. Wtedy, wartość oczekiwaną zmiennej losowej $\varphi(X)$ – o ile istnieje – możemy policzyć ze wzoru

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}} \varphi(x) d\mu_X(x).$$

W przypadku zmiennych o rozkładzie dyskretnym lub ciągłym można zmodyfikować wzór z Twierdzenia 5.10. Jeżeli zmienna losowa X ma rozkład dyskretny skupiony na zbiorze $S := \{x_1, x_2, \dots\}$, to – o ile powiązana wartość oczekiwana $\varphi(X)$ istnieje – zachodzi równość

$$\mathbb{E}(\varphi(X)) = \sum_{i=1}^{\infty} \varphi(x_i) \cdot \mathbb{P}(X = x_i). \quad (5.4)$$

Jeżeli natomiast zmienna losowa ma rozkład ciągły o gęstości f , to – o ile powiązana wartość oczekiwana istnieje – zachodzi

$$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx. \quad (5.5)$$

Przykład 5.11 (Wartość oczekiwana pola prostokąta). Załóżmy, że chcemy obliczyć pole prostokąta o bokach X oraz $(1 - X)$, gdzie X ma rozkład jednostajny na zbiorze $(0, 1)$. Zadanie to sprowadza się do obłożenia zmiennej X funkcją $\varphi(x) = x(1 - x)$, $x \in \mathbb{R}$, i policzeniu jej wartości oczekiwanej. Korzystając ze wzoru (5.5) dostajemy

$$\mathbb{E}(X(1 - X)) = \int_{-\infty}^{\infty} x(1 - x)1_{(0,1)}(x) dx = \int_0^1 x dx - \int_0^1 x^2 dx = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

W przypadku nieujemnych zmiennych losowych, można szybko policzyć wartość oczekiwaną znając jej dystrybuantę.

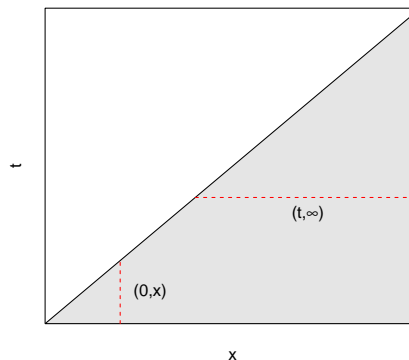
Propozycja 5.12 (Wartość oczekiwana nieujemnej zmiennej losowej). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$ taką, że $X \geq 0$. Wtedy wartość oczekiwaną zmiennej losowej X można policzyć ze wzoru

$$\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(t)) dt = \int_0^{\infty} \mathbb{P}(X > t) dt.$$

Dowód. Niech $X \geq 0$. Wtedy, korzystając z twierdzenia Fubiniego, dostajemy

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} x d\mu_X(x) = \int_0^{\infty} \left(\int_0^x 1 dt \right) d\mu_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[0,x]}(t) 1_{[0,\infty)}(x) dt d\mu_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[0,\infty)}(t) 1_{[t,\infty)}(x) d\mu_X(x) dt \\ &= \int_0^{\infty} \left(\int_t^{\infty} d\mu_X(x) \right) dt = \int_0^{\infty} \mathbb{P}(X > t) dt \\ &= \int_0^{\infty} (1 - F_X(t)) dt. \end{aligned}$$

Warto zauważyć, że funkcja $t \rightarrow \mathbb{P}(X > t)$ ma co najwyżej przeliczalnie wiele punktów nieciągłości, więc użycie ostrej nierówności nie wpływa tutaj na wartość całki.



□

Na koniec przedstawimy użyteczną często własność, które pozwala nam szybko policzyć wartość oczekiwaną iloczynu zmiennych losowych, gdy są one niezależne.

Propozycja 5.13 (Wartość oczekiwana iloczynu niezależnych zmiennych losowych). Niech $X, Y \in L^1(\Omega, \Sigma, \mathbb{P})$ będą niezależnymi zmiennymi losowymi. Wtedy

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Dowód. Pokażmy najpierw, że wartość oczekiwana iloczynu istnieje i jest skończona, tzn. $\int_{\Omega} |XY| dP < \infty$. Korzystając z twierdzenia Fubiniego, dwuwymiarowego odpowiednika Twierdzenia 5.10 (dla $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$ danej przez $\varphi(x, y) = |xy|$) oraz warunku 3) w Propozycji 4.14 dostajemy

$$\begin{aligned} \mathbb{E}(|XY|) &= \int_{\mathbb{R}^2} |xy| \mu_{(X,Y)}(dx dy) = \int_{\mathbb{R}} \int_{\mathbb{R}} |x| |y| \mu_X(dx) \mu_Y(dy) \\ &= \int_{\mathbb{R}} |x| \mu_X(dx) \int_{\mathbb{R}} |y| \mu_Y(dy) = \mathbb{E}(|X|)\mathbb{E}(|Y|) < \infty. \end{aligned}$$

Wartość oczekiwana iloczynu więc istnieje, i jest dana wzorem

$$\mathbb{E}(XY) = \int_{\mathbb{R}^2} xy\mu_{(X,Y)}(dx dy) = \dots = \int_{\mathbb{R}} x\mu_X(dx) \int_{\mathbb{R}} y\mu_Y(dy) = \mathbb{E}(X)\mathbb{E}(Y).$$

□

Na koniec dodajmy jeszcze, że w przypadku wielowymiarowym, przez wartość oczekiwaną z wektora losowego będziemy rozumieć wektor wartości średnich jego brzegów.

Definicja 5.14 (Wartość oczekiwana wektora losowego). Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie wektorem losowym określonym na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, **wartością oczekiwaną wektora \mathbf{X}** oznaczaną przez $\mathbb{E}(\mathbf{X})$ – o ile istnieje – nazywamy wektor z \mathbb{R}^n zadany przez

$$\mathbb{E}(\mathbf{X}) := (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n)). \quad (5.6)$$

5.2 Wariancja i odchylenie standardowe

W poprzednim rozdziale zdefiniowaliśmy wartość oczekiwaną zmiennej losowej, którą można traktować jako pewną charakterystykę liczbową rozkładu. Oczywiście podanie samej wartości oczekiwanej (jednej liczby) nie charakteryzuje nam rozkładu. Aby móc w lepszym stopniu go opisać przy pomocy charakterystyk, często określa się w jakim stopniu rozkład jest rozrzuty (rozproszony) wokół swojej średniej.

Definicja 5.15 (Wariancja i odchylenie standardowe). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, **wariancję zmiennej losowej X** – o ile istnieje – oznaczamy przez $D^2(X)$ i definiujemy jako

$$D^2(X) := \mathbb{E}((X - \mathbb{E}(X))^2). \quad (5.7)$$

Przez *odchylenie standardowe* zmiennej losowej X rozumiemy pierwiastek z jej wariancji i oznaczamy je jako $D(X) := \sqrt{D^2(X)}$.

Jak widać z Definicji 5.15, wariancji to nic innego jak średni kwadrat odchylenia od wartości średniej. Warunkiem koniecznym i wystarczającym na istnienie wariancji jest oczywiście istnienie całki z kwadratem zmiennej, tzn. warunek $\int_{\Omega} X^2 dX < \infty$.¹³ Przestrzeń całkowalnych z kwadratem zmiennych losowych zadanych na $(\Omega, \Sigma, \mathbb{P})$ będziemy oznaczać przez $L^2(\Omega, \Sigma, \mathbb{P})$. Pokażemy teraz podstawowe własności wariancji.

Propozycja 5.16 (Podstawowe własności wariancji). Niech $X \in L^2(\Omega, \Sigma, \mathbb{P})$. Wtedy

- 1) (alternatywny wzór) $D^2(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
- 2) (nieujemność) $D^2(X) \geq 0$, przy czym $D^2(X) = 0$ wtedy i tylko wtedy, gdy $X = x$ (p.n.) dla pewnej stałej $x \in \mathbb{R}$;
- 3) (homogeniczność) $D^2(cX) = c^2 D^2(X)$ dla $c \in \mathbb{R}$;
- 4) (niezmienniczość względem przesunięć) $D^2(X + a) = D^2(X)$ dla $a \in \mathbb{R}$;

¹³Warunek ten oczywiście implikuje istnienie wartości średniej, tzn. warunek $\int_{\Omega} X^2 dX < \infty$; aby to sprawdzić, wystarczy skorzystać z nierówności $|x| \leq x^2 + 1$, $x \in \mathbb{R}$.

Dowód.

1) Korzystając wprost z definicji i własności wartości średniej dostajemy

$$D^2(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

2) Nieujemność wynika wprost z definicji, gdyż całka z funkcji nieujemnej jest nieujemna. Załóżmy teraz, że $D^2(X) = 0$ i oznaczmy $\mathbb{E}(X) = x$. Wtedy wartość oczekiwana z nieujemnej funkcji $(X - x)^2$ wynosi 0, co implikuje już $(X - x)^2 = 0$ (p.n.) i w konsekwencji $X = x$ (p.n.); dowód tego pozostawiamy jako ćwiczenie do domu.

3) Korzystając z 1) i własności wartości oczekiwanej dostajemy

$$D^2(cX) = \mathbb{E}((cX)^2) - (\mathbb{E}(cX))^2 = c^2(\mathbb{E}(X^2) - (\mathbb{E}(X))^2)$$

4) Wprost z definicji dostajemy

$$D^2(X + a) = \mathbb{E}((X + a - \mathbb{E}(X + a))^2) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

□

Wariancję można często łatwo wyliczyć dla rozkładów ciągłych i dyskretnych, co ilustrują dwa kolejne przykłady.

Przykład 5.17 (Suma orłów w dwóch rzutach monetą). Rozważmy model w którym rzucamy dwiema monetami. Niech zmienna losowa X opisuje sumę reszek tj. $\mathbb{P}[X = 0] = \frac{1}{2}\mathbb{P}[X = 1] = \mathbb{P}[X = 2] = \frac{1}{4}$. Wariancja zmiennej losowej X wynosi wtedy

$$D^2(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = (0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4}) - 1^2 = \frac{1}{2}.$$

Przykład 5.18 (Losowanie liczby z rozkładu jednostajnego na $(0,1)$). Niech X ma rozkład jednostajny na odcinku $(0,1)$. Korzystając z faktu, że $\mathbb{E}(X) = \frac{1}{2}$, dostajemy

$$D^2(X) = \mathbb{E}((X - \frac{1}{2})^2) = \int_0^1 (x - \frac{1}{2})^2 dx = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}$$

Czasami, podobnie jak w przypadku wartości oczekiwanej, przez wariancję wektora losowego \mathbf{X} rozumiemy wektor złożony z wariancji jego brzegów. Nie będziemy tego jednak definiować, gdyż częściej podaje się tzw. macierz wariancji-kowariancji, której przekątna odpowiada wariancjom brzegowym – macierz ta zostanie zdefiniowana w następnym podrozdziale.

5.3 Kowariancja i korelacja

Do tej pory zajmowaliśmy się współczynnikami liczbowymi opisującymi rozkład danej zmiennej losowej. Przedstawimy teraz współczynnik, którym można opisać zależność między dwoma zmiennymi losowymi.

Definicja 5.19 (Kowariancja). Niech X i Y będą zmiennymi losowymi określonymi na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, **kowariancję zmiennych losowych X i Y** – o ile istnieje – oznaczamy przez $\text{Cov}(X, Y)$ i definiujemy jako

$$\text{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))). \quad (5.8)$$

Kowariancję można traktować, jako swego rodzaju uogólnienie wariancji, które bierze pod uwagę zależność między zmiennymi losowymi. Własności kowariancji są przedstawione poniżej.

Propozycja 5.20 (Podstawowe własności kowariancji). Niech $X, Y, Z \in L^2(\Omega, \Sigma, \mathbb{P})$. Wtedy

- 1) (alternatywny wzór) $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;
- 2) (związek z wariancją) $\text{Cov}(X, X) = \text{Var}(X)$;
- 3) (zerowa kowariancja dla niezależnych zmiennych losowych) $\text{Cov}(X, Y) = 0$, gdy X oraz Y są niezależne;
- 4) (symetria) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- 5) (homogeniczność i niezależność od przesunięć) $\text{Cov}(aX + b, cY + d) = ab \text{Cov}(X, Y)$ dla $a, b, c, d \in \mathbb{R}$;
- 6) (dwuliniowość) $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.

Dowód Propozycji 5.20 opiera się na wykorzystaniu podstawowych własności operatora wartości oczekiwanej i w dużej mierze jest bardzo podobny do dowodu Propozycji 5.16. Zauważmy także, że punkt 3) wynika niemal wprost z Propozycji 5.13. Pełny dowód Propozycji 5.20 pozostawiamy jako ćwiczenie do domu.

Pokażmy teraz, jak przy pomocy kowariancji można policzyć wariancję sumy zmiennych losowych.

w7
ws

Propozycja 5.21 (Wariancja sumy dwóch zmiennych losowych). Niech $X, Y \in L^2(\Omega, \Sigma, \mathbb{P})$. Wtedy istnieje wariancja sumy tych zmiennych i jest ona określona wzorem

$$D^2(X + Y) = D^2(X) + D^2(Y) + 2 \text{Cov}(X, Y).$$

Dowód. Korzystając z równości $D^2(X + Y) = \text{Cov}(X + Y, X + Y)$, dwuliniowości operatora kowariancji oraz symetrii dostajemy

$$\text{Cov}(X + Y, X + Y) = \text{Cov}(X, X) + \text{Cov}(Y, Y) + 2 \text{Cov}(X, Y) = D^2(X) + D^2(Y) + 2 \text{Cov}(X, Y),$$

co kończy dowód. □

Uwaga 5.22 (Wariancja sumy większej ilości zmiennych losowych). Wzór z Propozycji 5.21 można łatwo uogólnić na n zmiennych losowych. Jeżeli X_1, \dots, X_n są zmiennymi o skończonej wariancji, to wariancja ich sumy ma postać

$$D^2(X_1 + \dots + X_n) = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n D^2(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

W szczególności warto zauważyć, że dla (parami) niezależnych zmiennych losowych dostajemy

$$D^2(X_1 + \dots + X_n) = D^2(X_1) + \dots + D^2(X_n). \quad (5.9)$$

Warto tutaj wyraźnie podkreślić, że z równości (5.9) nie wynika niezależność (nawet parami) zmiennych losowych X_1, \dots, X_n .

Kowariancja oddaje nam w pewien sposób zależność między zmiennymi losowymi, ale oczywiście jej wartość zależy od wariancji zmiennych losowych. W szczególności, przekształcenia liniowe (skalowanie) zmiennych losowych zmieniają ich kowariancję. Aby tego uniknąć i w pewnym sensie unormować ten współczynnik, wprowadza się tzw. korelację.

Definicja 5.23 (Kowariancja). Niech $X, Y \in L^2(\Omega, \Sigma, \mathbb{P})$. Wtedy, **korelację zmiennych losowych X i Y** oznaczamy przez $\text{Cor}(X, Y)$ lub $(\rho_{X,Y})$ i definiujemy jako

$$\text{Cor}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{D^2(X)D^2(Y)}}. \quad (5.10)$$

Łatwo pokazać, że dla $a, d > 0$ oraz $b, c \in \mathbb{R}$ mamy $\text{Cor}(aX + b, cY + d) = \text{Cor}(X, Y)$ więc operator korelacji jest niezmienniczy względem (rosnących) przekształceń liniowych. Łatwo również pokazać (korzystając np. z nierówności Schwarz'a), że dla dowolnych zmiennych losowych X, Y (całkowalnych z kwadratem) dostajemy $\text{Cor}(X, Y) \in [-1, 1]$, przy czym wartości graniczne -1 i 1 są osiągnięte tylko w przypadku, gdy jedna zmienna jest liniową transformacją drugiej zmiennej. W związku z tym współczynnik korelacji często nazywa się *współczynnikiem korelacji liniowej*. Gdy $\text{Cor}(X, Y) = 0$, to mówimy, że zmienne X oraz Y są *liniowo niezależne*; w szczególności, jeżeli zmienne X i Y są niezależne, to są liniowo niezależne (ale nie na odwrót!).

Na koniec zdefiniujemy jeszcze wielowymiarowe odpowiedniki zarówno kowariancji, jak i korelacji.

Definicja 5.24 (Macierz wariancji-kowariancji wektora losowego). Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie wektorem losowym określonym na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, **macierz wariancji-kowariancji wektora \mathbf{X}** – o ile istnieje – oznaczamy przez $\Sigma_{\mathbf{X}}$ i definiujemy jako

$$\Sigma_{\mathbf{X}} := \begin{bmatrix} \text{Cov}(X_1, X_1) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \dots & \text{Cov}(X_n, X_n) \end{bmatrix}.$$

Przez **macierz korelacji** wektora \mathbf{X} będziemy natomiast rozumieć macierz $R_{\mathbf{X}}$ zadaną przez

$$R_{\mathbf{X}} := \begin{bmatrix} \text{Cor}(X_1, X_1) & \dots & \text{Cor}(X_1, X_n) \\ \text{Cor}(X_2, X_1) & \dots & \text{Cor}(X_2, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cor}(X_n, X_1) & \dots & \text{Cor}(X_n, X_n) \end{bmatrix}.$$

Warto zauważyć, że $\text{diag}(\Sigma_{\mathbf{X}}) = (D^2(X_1), \dots, D^2(X_n))$, gdzie $\text{diag}(\Sigma_{\mathbf{X}})$ to wektor złożony z wyrazów na przekątnej macierzy. Dostajemy również $\text{diag}(R_{\mathbf{X}}) = (1, \dots, 1)$, gdyż dla każdego $i = 1, \dots, n$ zachodzi $\text{Cor}(X_i, X_i) = 1$. Łatwo też pokazać, że obie macierze są symetryczne, nieujemnie określone, oraz jeżeli rząd tych macierzy nie jest pełny (macierz jest zdegenerowana), to niektóre ze zmiennych brzegowych można przedstawić jako kombinacje liniowe innych zmiennych brzegowych.

5.4 Momenty i inne charakterystyki

Omówione przez nas charakterystyki – średnia, wariancja, odchylenie standardowe, kowariancja i korelacja – nie są oczywiście jedynymi charakterystykami, które można związać z rozkładem. Często charakterystyki są związane z momentami wyższych rzędów.

Definicja 5.25 (Momenty zmiennej losowej). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, dla $p > 0$, wprowadzamy następująca charakterystyki – o ile istnieją – zmiennej X :

- **Moment (zwykły) rzędu p** zdefiniowany jako $\mathbb{E}(X^p)$;
- **Moment absolutny rzędu p** zdefiniowany jako $\mathbb{E}(|X|^p)$;
- **Moment centralny rzędu p** zdefiniowany jako $\mathbb{E}((X - \mathbb{E}[X])^p)$.

Podobnie jak w przypadku średniej i wariancji, przestrzeń zmiennych losowych na $(\Omega, \Sigma, \mathbb{P})$ dla których istnieje moment rzędu p , tzn. zachodzi $\int_{\Omega} |X|^p d\mathbb{P}$, oznaczamy przez $L^p(\Omega, \Sigma, \mathbb{P})$. Warto zauważyć, że dla dowolnego $p < q$ dostajemy $L^p(\Omega, \Sigma, \mathbb{P}) \supset L^q(\Omega, \Sigma, \mathbb{P})$.¹⁴ Często też wprowadza się przypadki graniczne, w których $L^0(\Omega, \Sigma, \mathbb{P})$ oznacza przestrzeń wszystkich zmiennych losowych, natomiast $L^\infty(\Omega, \Sigma, \mathbb{P})$ oznacza przestrzeń ograniczonych zmiennych losowych. Wyższe momenty często służą do zdefiniowania innych charakterystyk związanych z kształtem rozkładu zmiennej losowej. Wprowadźmy teraz dwa z nich, które mierzą tzw. skośność i spłaszczenie rozkładu

Definicja 5.26 (Współczynnik skośności i kurtozy). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy

- **Współczynnikiem skośności (asymetrii) zmiennej losowej X** – o ile istnieje – nazywamy charakterystykę $\alpha_3(X)$ zadaną przez

$$\alpha_3(X) := \frac{\mathbb{E}((X - \mathbb{E}(X))^3)}{(D^2(X))^{3/2}}.$$

- **Współczynnikiem kurtozy (spłaszczenia) zmiennej losowej X** – o ile istnieje – nazywamy natomiast charakterystykę $\alpha_4(X)$ zadaną przez

$$\alpha_4(X) := \frac{\mathbb{E}((X - \mathbb{E}(X))^4)}{(D^2(X))^2}.$$

Oprócz powyższych charakterystyk z rozkładem można związać wiele innych liczb. Przykładem może być tutaj chociażby pomiar tzw. entropii rozkładu, czy wyznaczenie ryzyka rozkładu (mierzonego np. poprzez wyliczenia punktu, dla której dystrybuanta przyjmuje określoną wartość). Omówienie tych charakterystyk i ich własności wykracza jednak poza zakres tego wykładu. Na koniec wprowadźmy jeszcze jedną popularną charakterystykę rozkładu, zwaną *medianą*. Jest to liczba, która dzieli rozkład na pół w ten sposób, aby wpadnięcie do każdej z tych części było możliwie jak najbardziej równoprawdopodobne.

¹⁴Wprowadzamy tutaj pewne uproszczenie, gdyż często definiując przestrzeń $L^p(\Omega, \Sigma, \mathbb{P})$ definiuje się na niej od razu normę poprzez zadania p -tego absolutnego momentu zmiennej. Traktujmy jednak te przestrzenie na razie jako przestrzenie bez ustalonej normy

Definicja 5.27 (Mediana). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. **Medianą zmiennej losowej X** nazywamy każdą liczbę $m \in \mathbb{R}^a$ dla której spełnione są nierówności

$$\mathbb{P}(X \leq m) \geq \frac{1}{2} \quad \text{oraz} \quad \mathbb{P}(X \geq m) \leq \frac{1}{2}.$$

^aDla każdej zmiennej losowej istnieje (co najmniej jedna) mediana, ale w zdegenerowanych przypadkach może być ich więcej

5.5 Nierówności probabilistyczne wykorzystujące momenty

Z analizy znamy wiele nierówności całkowych, które pomagają w dowodach, bądź oszacowaniach. Nierówności te oczywiście można opisać w języku zmiennych losowych i ich momentów. Przypomnijmy teraz niektóre z nich oraz wprowadźmy dwie nierówności charakterystyczne dla rachunku prawdopodobieństwa

Propozycja 5.28 (Nierówność Schwarz). Niech $X, Y \in L^2(\Omega, \Sigma, \mathbb{P})$. Wtedy

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2). \quad (5.11)$$

Dowód. Jeżeli $\mathbb{E}(X^2) = 0$ lub $\mathbb{E}(Y^2) = 0$ to nierówność jest spełniona. Załóżmy, że obie te wartości są silnie większe od 0. Zdefiniujemy zmienne $\tilde{X} := X/\sqrt{\mathbb{E}(X^2)}$ i $\tilde{Y} := Y/\sqrt{\mathbb{E}(Y^2)}$ oraz zauważmy, że nierówność $(|\tilde{X}| - |\tilde{Y}|)^2 \geq 0$ implikuje

$$\tilde{X}^2 + \tilde{Y}^2 \geq 2|\tilde{X}\tilde{Y}|. \quad (5.12)$$

Obkładając (5.12) wartością oczekiwaną, dostajemy $1 \geq \mathbb{E}(|\tilde{X}\tilde{Y}|)$, co jest równoważne nierówności $\sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)} \leq \mathbb{E}(XY)$, która z kolei jest równoważna (5.11). \square

Propozycja 5.29 (Nierówność Jensena). Niech $X \in L^1(\Omega, \Sigma, \mathbb{P})$ oraz niech $g : \mathbb{R} \rightarrow \mathbb{R}$ będzie (mierzalną) funkcją wypukłą dla której $g(X) \in L^1(\Omega, \Sigma, \mathbb{P})$. Wtedy

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X)). \quad (5.13)$$

Dowód. Z analizy matematycznej wiemy, że dla każdego $x_0 \in \mathbb{R}$ istnieje stała $\alpha_{x_0} \in \mathbb{R}$ taka, że dla dowolnego $y \in \mathbb{R}$ dostajemy $g(y) \geq g(x_0) + (y - x_0)\alpha_{x_0}$, co wynika z wypukłości funkcji g . W szczególności podstawiając $x_0 := \mathbb{E}(X)$ oraz dla każdego $\omega \in \Omega$ podstawiając $y := X(\omega)$ dochodzimy do nierówności

$$g(X) \geq g(\mathbb{E}(X)) + (X - \mathbb{E}(X))\alpha_{\mathbb{E}(X)}.$$

Obkładając obie strony wartością oczekiwaną otrzymujemy tezę. \square

Propozycja 5.30 (Nierówność Höldera). Niech $p, q > 1$ będą do siebie sprzężone, tj. $\frac{1}{p} + \frac{1}{q} = 1$. Niech $X \in L^p(\Omega, \Sigma, \mathbb{P})$ oraz $Y \in L^q(\Omega, \Sigma, \mathbb{P})$. Wtedy

$$\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{1/p} \mathbb{E}(|Y|^q)^{1/q}. \quad (5.14)$$

Przykładowy dowód nierówności Höldera opiera się na zastosowaniu nierówności Younga do zmiennych $\tilde{X} := X \mathbb{E}(|X|^p)^{1/p}$ oraz $\tilde{Y} := Y \mathbb{E}(|Y|^q)^{1/q}$ i obłożeniu obu stron wartością oczekiwaną, podobnie jak w nierówności Schwartza. Pozostawiamy go jako ćwiczenie domowe.

Udowodnijmy teraz nierówność Czebyszewa, która będzie przydatna w dowodzeniu wielu twierdzeń rachunku prawdopodobieństwa oraz prostych oszacowaniach.

Propozycja 5.31 (Nierówność Czebyszewa). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, dla każdego $\epsilon > 0$, zachodzi

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{\mathbb{E}|X|}{\epsilon}. \quad (5.15)$$

Dowód. Ustalmy $\epsilon > 0$. Oszacowując z dołu wartość zmiennej losowej $|X|$ dostajemy

$$\epsilon \cdot \mathbb{P}(|X| \geq \epsilon) \leq \int_{\{|X| \geq \epsilon\}} |X| d\mathbb{P} \leq \mathbb{E}|X|.$$

□

Nierówność Czebyszewa łatwo uogólnić na wyższe momenty, co daje nam tzw. nierówność Markowa.

Propozycja 5.32 (Nierówność Markowa). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Wtedy, dla każdego $p > 0$ oraz $\epsilon > 0$, zachodzi

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{\mathbb{E}(|X|^p)}{\epsilon^p}. \quad (5.16)$$

Dowód nierówności Markowa jest prostym wnioskiem z nierówności Czebyszewa i pozostawiamy go jako ćwiczenie do domu. Warto też zauważyć, że dla całkowalnych zmiennych losowych z nierówności Markowa wynika też nierówność

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{D^2(X)}{\epsilon^2}, \quad (5.17)$$

którą czasami nazywa się nierównością *Czebyszewa-Bianaymé*. Ogólnie mówiąc, o ile nierówność Czebyszewa jest przydatna w dowodach, o tyle często daje nam bardzo niedokładne oszacowania – warto zauważyć, że oszacowania w dowodzie Propozycji 5.31 mogą być często nieoptymalne ma oszacowanie wartości $\mathbb{P}(|X| \geq \epsilon)$.

Na koniec tego rozdziału przytoczymy jeszcze jedną ważną nierówność dla sum niezależnych zmiennych losowych, które pozwala nam kontrolować to, jak bardzo możemy odchylić się od (teoretycznej) średniej. Nierówność ta jest często używana (np. wraz z nierównością Czebyszewa) w dowodach twierdzeń na kryteria zbieżności szeregów zmiennych losowych.

Propozycja 5.33 (Nierówność Kołmogorowa). Niech $n \in \mathbb{N}$ oraz niech X_1, \dots, X_n będzie ciągiem niezależnych zmiennych losowych takim, że $\mathbb{E}(X_i) = 0$ oraz $D^2(X_i) < \infty$, $i = 1, \dots, n$. Wtedy dla każdego $\epsilon > 0$ zachodzi

$$\mathbb{P}\left(\max_{i=1, \dots, n} |X_1 + \dots + X_i| \geq \epsilon\right) \leq \frac{D^2(X_1 + \dots + X_n)}{\epsilon^2}.$$

Dowód. Ustalmy $n \in \mathbb{N}$, $\epsilon > 0$ oraz wprowadźmy oznaczenie $S_i := X_1 + \dots + X_i$ dla $i = 1, \dots, n$. Niech $A_i := \{|S_i| \geq \epsilon\}$. Dla $B_1 := A_1$ oraz $B_i := A_i \setminus \{A_1 \cup \dots \cup A_{i-1}\}$, $i = 2, \dots, n$ dostajemy

$$A := \left\{ \max_{i=1, \dots, n} |X_1 + \dots + X_i| \geq \epsilon \right\} = \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i.$$

Ponieważ zdarzenia $(B_i)_{i=1}^n$ są parami rozłączne i sumują się do A oraz $\mathbb{E}(S_n) = 0$, dostajemy

$$D^2(S_n) = \mathbb{E}(S_n^2) \geq \mathbb{E}(1_A \cdot S_n^2) = \sum_{i=1}^n \mathbb{E}(1_{B_i} \cdot S_n^2). \quad (5.18)$$

Dla każdego $i = 1, \dots, n$, zauważając, że zmienne losowe $1_{B_i} \cdot S_i$ oraz $S_n - S_i$ są niezależne oraz zachodzi $\mathbb{E}(S_n - S_i) = 0$, dostajemy

$$\begin{aligned} \mathbb{E}(1_{B_i} \cdot S_n^2) &= \mathbb{E}(1_{B_i} \cdot (S_n - S_i + S_i)^2) \\ &= \mathbb{E}(1_{B_i} \cdot (S_n - S_i)^2) + 2\mathbb{E}(1_{B_i} \cdot S_i)\mathbb{E}(S_n - S_i) + \mathbb{E}(1_{B_i} S_i^2) \\ &\geq \mathbb{E}(1_{B_i} S_i^2) \\ &\geq \mathbb{P}(B_i)\epsilon^2, \end{aligned}$$

co w połączeniu z (5.18) daje nam $D^2(S_n) \geq \sum_{k=1}^n \mathbb{P}(B_k)\epsilon^2 = \mathbb{P}(A)\epsilon^2$ i kończy dowód. \square

6 Przegląd i własności wybranych rozkładów

Poznaliśmy już definicję zmiennej losowej, jej rozkładu, oraz podaliśmy wzoru na podstawowe charakterystyki. Z praktycznego punktu widzenia chcielibyśmy jednak poznać przykłady rodzin rozkładów prawdopodobieństwa, które mogą być użyteczne. W rozdziale tym przedstawimy przegląd podstawowych (jednowymiarowych) rozkładów prawdopodobieństwa wraz z podaniem ich podstawowych charakterystyk; pomijamy charakterystyki takie, jak skośność, kurtoza, czy mediana

6.1 Rozkłady dyskretne

Zacznijmy od przeglądu rozkładów dyskretnych, które często pojawiają się w problemach probabilistycznych. Dla każdego rozkładu podane jest jego oznaczenie oraz tabela z podstawowymi powiązаныmi obiektami i (istniejącymi) charakterystykami.

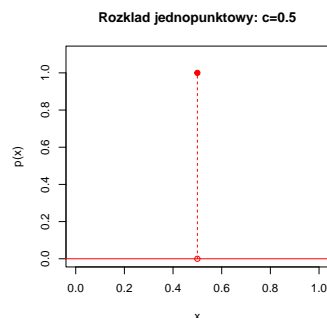
Rozkład jednopunktowy

Na początek przedstawiamy najprostszą możliwą sytuację, w której mamy do czynienia z rozkładem jednopunktowym.

Definicja 6.1 (Rozkład jednopunktowy). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład jednopunktowy** dla $c \in \mathbb{R}$, jeżeli $\mathbb{P}(X = c) = 1$. Rozkład taki oznaczamy przez δ_c i zapisujemy jako $X \sim \delta_c$.^a

^aRozkład ten nazywany jest też często *deltą Diraca* – stąd oznaczenie.

Parametry	$c \in \mathbb{R}$ (punkt)
Masa prawd.	$p(x) = 1_c(x), x \in \mathbb{R}$
Średnia	c
Wariancja	0

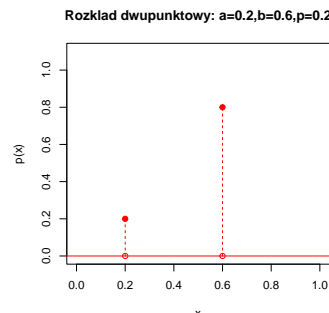


Rozkład dwupunktowy

Rozkład dwupunktowy może służyć do modelowania prostych eksperymentów w których interesuje nas tylko sukces i porażka. Za pomocą odpowiednich sum rozkładów dwupunktowych można również odtworzyć wiele znanych rozkładów.

Definicja 6.2 (Rozkład dwupunktowy). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład dwupunktowy** dla (różnych) $a, b \in \mathbb{R}$ oraz $p \in (0, 1)$, gdy $\mathbb{P}(X = a) = p$ i $\mathbb{P}(X = b) = q$, gdzie $q = 1 - p$.

Parametry	$a, b \in \mathbb{R}$ (punkty) $p \in (0, 1)$ (prawdopodobieństwo dla a) $q = 1 - p$ (prawdopodobieństwo dla b)
Masa prawd.	$p(x) = p \cdot 1_a(x) + q \cdot 1_b(x), x \in \mathbb{R}$
Średnia	$p \cdot a + q \cdot b$
Wariancja	$pq(a - b)^2$



Rozkład dwumianowy

Rozkład dwumianowy informuje nas o liczbie sukcesów, gdy powtarzamy eksperyment zadaną ilość razy (n). Rozkład ten można w prosty sposób wyprowadzić wychodząc z rozkładu dwupunktowego na $\{0, 1\}$ i rozważając sumy niezależnych zdarzeń.

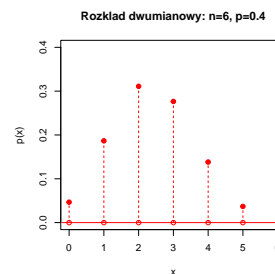
Definicja 6.3 (Rozkład dwumianowy). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład dwumianowy** dla $n \in \mathbb{N}$ oraz $p \in (0, 1)$, gdy

$$\mathbb{P}(X = k) = B(k, n, p) := \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Rozkład taki oznaczamy przez $B(n, p)$ i zapisujemy jako $X \sim B(n, p)$.^a

^aW polskiej literaturze można też spotkać alternatywną nazwę *rozkład Bernoulliego*, choć w angielskiej literaturze nazwa ta jest zarezerwowana dla rozkładu dwupunktowego na zbiorze $\{0, 1\}$.

Parametry	$n \in \mathbb{N}$ (ilość prób) $p \in (0, 1)$ (prawdopodobieństwo sukcesu)
Masa prawd.	$p(x) = \sum_{k=0}^n 1_k(x) B(k, n, p), x \in \mathbb{R}$
Średnia	np
Wariancja	npq



Rozkład Poissona

Rozkład Poissona można traktować jako przypadek graniczny rozkładu dwumianowego, gdy zwiększamy liczbę powtórzeń do nieskończoności, jednocześnie zmniejszając prawdopodobieństwo wystąpienia zdarzeń; badamy tutaj prawdopodobieństwo wystąpienia określonej ilości zdarzeń w zadanym z góry przedziale czasowym. Jest to przypadek graniczny rozkładu $B(n, p_n)$, dla którego zachodzi zależność $\lim_{n \rightarrow \infty} n \cdot p_n = \alpha$, gdzie parametr $\alpha > 0$ odpowiada za tzw. intensywność (średnią ilość zdarzeń zachodzącą w unormowanym przedziale).¹⁵ Rozkład Poissona może służyć np. do modelowania ilości pacjentów przybywających na SOR w określonym dniu, czy ilości meteorytów uderzających w księżyc w danym roku.

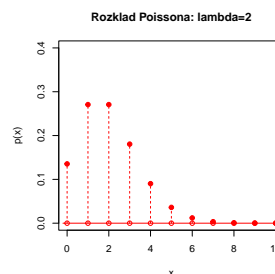
¹⁵Fakt ten zostanie udowodniony w następnym podrozdziale.

Definicja 6.4 (Rozkład Poissona). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład Poissona** dla $\lambda \in \mathbb{R}_+$, gdy

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}.$$

Rozkład taki oznaczamy przez $\text{Pois}(\lambda)$ i zapisujemy jako $X \sim \text{Pois}(\lambda)$.

Parametry	$\lambda \in \mathbb{R}_+$ (intensywność)
Masa prawd.	$p(x) = \sum_{k=0}^{\infty} 1_k(x) \frac{\lambda^k}{k!} e^{-\lambda}, x \in \mathbb{R}$
Średnia	λ
Wariancja	λ



Rozkład geometryczny

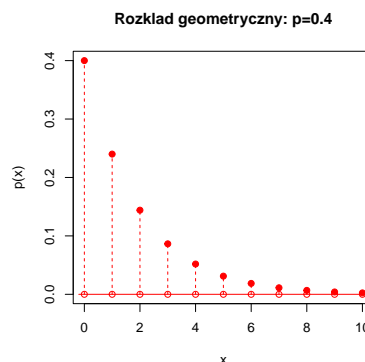
Rozkład ten odpowiada czasowi oczekiwania na pierwszy sukces w eksperymencie w którym mamy do czynienia z nieskończoną ilością prób.

Definicja 6.5 (Rozkład Geometryczny). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład geometryczny** dla $p \in (0, 1)$, gdy

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p, \quad k \in \mathbb{N}.$$

Rozkład taki oznaczamy przez $\text{Geom}(p)$ i zapisujemy jako $X \sim \text{Geom}(p)$.

Parametry	$p \in (0, 1)$ (prawd. sukcesu)
Masa prawd.	$p(x) = \sum_{k=0}^{\infty} 1_k(x) (1 - p)^{k-1} p, x \in \mathbb{R}$
Dystrybuanta	$F(t) = \sum_{k=0}^{\infty} 1_{[k, k+1)}(x) [1 - (1 - p)^k], t \in \mathbb{R}$
Średnia	$1/p$
Wariancja	$(1 - p)/p^2$



Inne rozkłady dyskretne

Oprócz wymienionych tutaj rozkładów dyskretnych, istnieje oczywiście więcej typów rozkładów, które mogą być przydatne w modelowaniu różnych zjawisk. Przykładowo, odnosi się to do *rozkładów*

wielomianowych, rozkładów ujemnych dwumianowych, czy rozkładów hipergeometrycznych. Więcej definicji oraz dokładniejszy opis wszystkich omówionych tutaj rodzin rozkładów można znaleźć w Rozdziale 5.10 w książce [JS04] lub w Rozdziale 3.2 w [CB02].

6.2 Rozkłady ciągłe

Przejdźmy teraz do omówienia wybranych rozkładów ciągłych, które często pojawiają się w problemach probabilistycznych. Dla każdego rozkładu podane jest jego oznaczenie oraz tabela z podstawowymi powiązаныmi obiektami i (istniejącymi) charakterystykami.

Rozkład jednostajny

Rozkład ten najczęściej odpowiada problemom w których mamy *wylosować* jakąś liczbę z danego odcinka. Często jest to blok budulcowy dla wszystkich innych rozkładów – poprzez obłożenie rozkładu jednostajnego odpowiednią funkcją jesteśmy w stanie dostać dowolny interesujący nas rozkład.

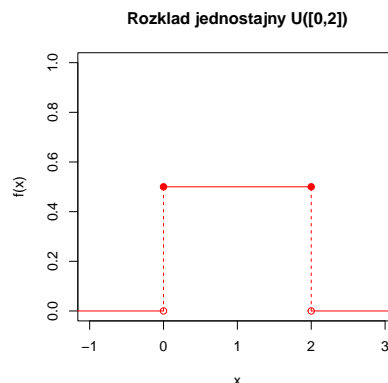
Definicja 6.6 (Rozkład jednostajny). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład jednostajny** na odcinku (a, b) , gdzie $a, b \in \mathbb{R}$ oraz $a < b$, gdy dla zmiennej X istnieje gęstość i jest ona wyrażona wzorem

$$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x), \quad x \in \mathbb{R}.$$

Rozkład taki oznaczamy przez $U([a, b])$ i zapisujemy jako $X \sim U([a, b])$.

Rozkład jednostajny można w intuicyjny sposób uogólnić na dowolny zbiór mierzalny A (o dodatniej i skończonej mierze Lebesguea) poprzez rozważanie gęstości $f(x) = (\mathcal{L}_1(A))^{-1} 1_A(x)$. Będziemy używać zapisu $U(\{x_1, \dots, x_n\})$ na oznaczenie rozkładu *jednostajnego* na zbiorze dyskretnym $\{x_1, \dots, x_n\}$ w którym przyjęcie każdej wartości jest równoprawdopodobne i wynosi $\frac{1}{n}$.

Parametry	$a, b \in \mathbb{R}, a < b$ (krańce przedziału)
Funkcja gęstości	$f(x) = \frac{1}{b-a} 1_{[a,b]}(x), x \in \mathbb{R}$
Dystrybuanta	$F(t) = \begin{cases} 0 & x < a \\ \frac{t-a}{b-a} & t \in [a, b] \\ 1 & x > b \end{cases}$
Średnia	$(a + b)/2$
Wariancja	$(b - a)^2/12$



Rozkład wykładniczy

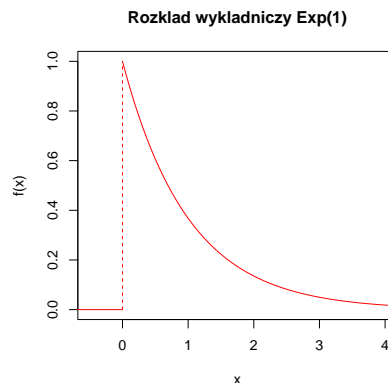
Rozkład ten jest ciągłym odpowiednikiem rozkładu geometrycznego. Podobnie jak w przypadku rozkładu Poissona, przy zastosowaniu przejścia granicznego $n \cdot p_n = \lambda$, możemy pytać o to kiedy nastąpi pierwszy sukces na unormowanym odcinku.

Definicja 6.7 (Rozkład wykładniczy). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład wykładniczy** z parametrem $\lambda > 0$, gdy dla zmiennej X istnieje gęstość i jest ona wyrażona wzorem

$$f_X(x) = \lambda e^{-\lambda x} 1_{(0, \infty)}(x), \quad x \in \mathbb{R}.$$

Rozkład taki oznaczamy przez $\text{Exp}(\lambda)$ i zapisujemy jako $X \sim \text{Exp}(\lambda)$.

Parametry	$\lambda > 0$ (intensywność)
Funkcja gęstości	$f(x) = \lambda e^{-\lambda x} 1_{(0, \infty)}(x), x \in \mathbb{R}$
Dystrybuanta	$F(t) = (1 - e^{-\lambda t}) 1_{(0, \infty)}(x), t \in \mathbb{R}$
Średnia	$1/\lambda$
Wariancja	$1/\lambda^2$



Rozkład Gamma

Rozkład Gamma z parametrami $\alpha \in \mathbb{N}$ i $\beta > 0$ służy do opisu sumy a niezależnych zmiennych losowych o rozkładzie wykładniczym z parametrem b . Rozkład ten uogólnia się na dowolne $a \in \mathbb{R}$.

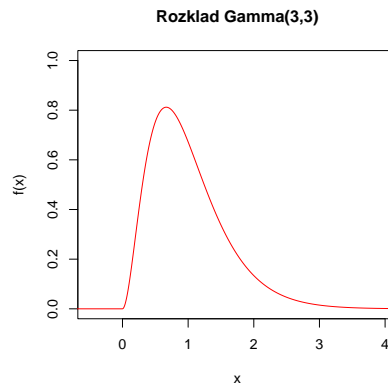
Definicja 6.8 (Rozkład Gamma). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład gamma** z parametrami $\alpha, \beta > 0$, gdy dla zmiennej X istnieje gęstość i jest ona wyrażona wzorem

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{(0, \infty)}(x), \quad x \in \mathbb{R}.^a$$

Rozkład taki oznaczamy przez $\Gamma(\alpha, \beta)$ i zapisujemy jako $X \sim \Gamma(\alpha, \beta)$.

^a Γ to funkcja Gamma. W szczególności dla $\alpha \in \mathbb{N}$ mamy $\Gamma(\alpha) = (\alpha - 1)!$.

Parametry	$\alpha > 0$ (parametr kształtu) $\beta > 0$ (odwrotny parametr skali)
Funkcja gęstości	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{(0,\infty)}(x), x \in \mathbb{R}$
Średnia	α/β
Wariancja	α/β^2



Rozkład normalny

Rozkład normalny (zwany także *rozkładem Gaussa* albo *krzywą Gaussa*) jest jednym z najważniejszych rozkładów w rachunku prawdopodobieństwa i służy do opisu wielu zjawisk statystycznych. Jego częste występowanie w przyrodzie umotywowane jest faktem, że jest on rozkładem granicznym, gdy rozpatrujemy średnią wartość sumy niezależnych zmiennych losowych z $L^2(\Omega, \Sigma, \mathbb{P})$ o tym samym rozkładzie, co np. może służyć do przybliżonego opisu rozkładu cech w populacji takich jak wzrost czy waga.¹⁶

Definicja 6.9 (Rozkład normalny). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład normalny** z parametrami $\mu \in \mathbb{R}$ i $\sigma > 0$, gdy dla zmiennej X istnieje gęstość i jest ona wyrażona wzorem

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Rozkład taki oznaczamy przez $\mathcal{N}(\mu, \sigma)$ i zapisujemy jako $X \sim \mathcal{N}(\mu, \sigma)$.^a

^aParametr μ odpowiada średniej, a parametr σ odchyleniu standardowemu. W literaturze często stosowany jest alternatywny zapis $\mathcal{N}(\mu, \sigma^2)$, gdzie do parametryzacji użyta jest wariancja, zamiast odchylenia standardowego.

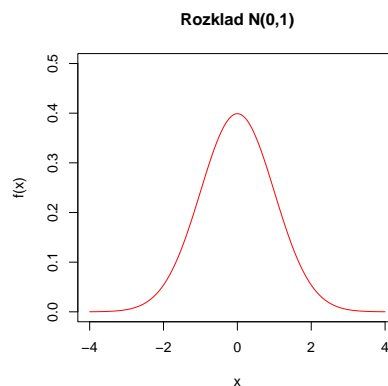
Łatwo zauważyć, że rozkład normalny uzyskany jest poprzez przekształcenie tzw. *standardowego rozkładu normalnego*, tzn. rozkładu $\mathcal{N}(0, 1)$, funkcją liniową. Istotnie, jeżeli zachodzi $X \sim \mathcal{N}(0, 1)$, to dostajemy $Y \sim \mathcal{N}(\mu, \sigma)$ dla $Y = \sigma X + \mu$. Dystrybuantę standardowego rozkładu normalnego będziemy oznaczać przez Φ , a gęstość standardowego rozkładu normalnego przez φ , tzn.

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \Phi(t) := \int_{-\infty}^t \varphi(x) dx.$$

Niestety nie istnieje dobry analityczny wzór na dystrybuantę rozkładu normalnego – wartości te odczytuje się zazwyczaj z tablic rozkładu normalnego.

¹⁶Więcej szczegółów na ten temat poznamy później, w rozdziale poświęconym centralnemu twierdzeniu granicznemu.

Parametry	$\mu \in \mathbb{R}$ (parametr średniej) $\sigma > 0$ (parametr odchylenia)
Funkcja gęstości	$f(x) = 1/\sigma \cdot \varphi((x - \mu)/\sigma), x \in \mathbb{R}$
Dystrybuanta	$F(t) = \Phi((t - \mu)/\sigma), t \in \mathbb{R}$
Średnia	μ
Wariancja	σ^2



Rozkład χ^2

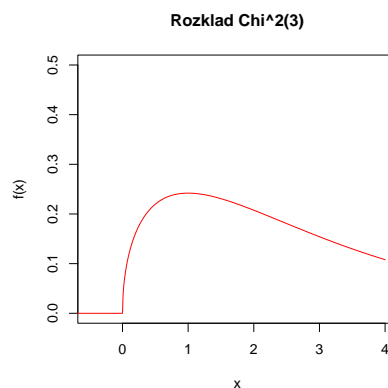
Rozkład χ^2 (wym. *hi kwadrat*) o $n \in \mathbb{N}$ stopniach swobody jest rozkładem sumy niezależnych zmiennych losowych o standardowym rozkładzie normalnym. Rozważanie takiej sumy jest pomocne w statystyce np. gdy konstruujemy tzw. *przedziały ufności*, czy *testujemy* hipotezy statystyczne.

Definicja 6.10 (Rozkład χ^2). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład** χ^2 z liczbą stopni swobody $n \in \mathbb{N}$, gdy dla zmiennej X istnieje gęstość i jest ona wyrażona wzorem

$$f_X(x) = \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} 1_{[0,\infty)}(x).$$

Rozkład taki oznaczamy przez $\chi^2(n)$ i zapisujemy jako $X \sim \chi^2(n)$.

Parametry	$n \in \mathbb{N}$ (liczba stopni swobody)
Funkcja gęstości	$\frac{1}{2^{n/2} \cdot \Gamma(n/2)} x^{n/2-1} e^{-x/2} 1_{[0,\infty)}(x), x \in \mathbb{R}$
Średnia	n
Wariancja	$2n$



Rozkład t-Studenta

Rozkład t-Studenta jest kolejnym ważnym rozkładem, który służy np. do opisu błędów estymatorów. Rozkład t-Studenta można też traktować jako uogólnienie rozkładu normalnego, w którym dopuszczamy możliwość tzw. *grubych ogonów*. Można go też wyrazić za pomocą rozkładu normalnego oraz rozkładu χ^2 .¹⁷

¹⁷Więcej informacji co to dokładnie znaczy zostanie podanych w późniejszej części wykład

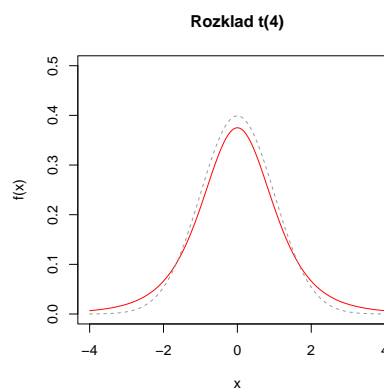
Definicja 6.11 (Rozkład t-Studenta). Niech X będzie zmienną losową określoną na $(\Omega, \Sigma, \mathbb{P})$. Zmienna X ma **rozkład t-studenta** z parametrem $\nu > 0$, gdy dla zmiennej X istnieje gęstość i jest ona wyrażona wzorem

$$f_X(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Rozkład taki oznaczamy przez t_ν i zapisujemy jako $X \sim t_\nu$.

Dla uproszczenia zdefiniowaliśmy rozkład t-Studenta nie wprowadzając dodatkowych parametrów związanych z lokacją i skalą. Można to zrobić, podobnie jak w przypadku rozkładu normalnego, aby umożliwić modelowanie nieunormowanych zjawisk.

Parametry	$\nu > 0$ (liczba stopni swobody)
Funkcja gęstości	$f(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
Średnia	0 (istnieje, gdy $\nu > 1$)
Wariancja	$\frac{\nu}{\nu-2}$ (istnieje, gdy $\nu > 2$)



Inne rozkłady ciągłe

Oprócz wymienionych tutaj rozkładów ciągłych, istnieje oczywiście więcej typów rozkładów, które mogą być przydatne w modelowaniu różnych zjawisk. Przykładowo, odnosi się to do *rozkładów beta*, *rozkładów α -stabilnych*, *rozkładów Cauchy'ego*, czy *rozkładów Laplace'a*. Więcej definicji oraz dokładniejszy opis omówionych tutaj rodzin rozkładów można znaleźć w Rozdziale 5.10 w książce [JS04] lub w Rozdziale 3.3 w [CB02].

6.3 Podstawowe własności rozkładu normalnego

W rozdziale tym omówimy pewne podstawowe własności rozkładu normalnego, które mogą być dla nas przydatne w dalszej części wykładu. Dużo z tych własności jest charakterystycznych dla większej klasy rozkładów (np. symetrycznych), ale postanowiliśmy zebrać własności dla rozkładu normalnego, gdyż są one często wykorzystywane np. w statystyce. Dla przypomnienia, w poprzednim rozdziale wprowadziliśmy oznaczenia

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \Phi(t) = \int_{-\infty}^t \varphi(x) dx,$$

których będziemy używać w tym rozdziale. Zaczniemy od ogólnej obserwacji, która pozwala na szybkie oszacowanie szansy na wpadnięcie zmiennej losowej do zadanych symetrycznych przedziałów.

Uwaga 6.12 (Reguła trzech sigm). Odczytując wartości z tablic, można sprawdzić, że dla zmiennej losowej o rozkładzie normalnym $X \sim N(\mu, \Sigma)$ zachodzi

$$\mathbb{P}(|X - \mu| \leq \sigma) \approx 68.27\%, \quad \mathbb{P}(|X - \mu| \leq 2\sigma) \approx 95.45\%, \quad \mathbb{P}(|X - \mu| \leq 3\sigma) \approx 99.73\%.$$

Jest to empiryczna reguła na podstawie której można wnioskować, że wartości zmiennej losowej X prawie nigdy nie wypadną poza przedział $[\mu - 3\sigma, \mu + 3\sigma]$, raczej nie wypadną poza przedział $[\mu - 2\sigma, \mu + 2\sigma]$ a prawdopodobieństwo, że znajdą się w przedziale $[\mu - \sigma, \mu + \sigma]$, przekracza istotnie 0.5. Reguły tej często używa się w statystyce, aby lepiej zrozumieć rozkład normalnej zmiennej losowej, czy sprawdzać, czy zadana obserwacja istotnie pochodzi z postulowanego rozkładu. Warto tutaj zauważyć, że prawdopodobieństwa przekroczenia trzech odchyłeń dla dowolnej zmiennej losowej X można oszacować z nierówności Markowa i wynosi ono

$$\mathbb{P}(|X - \mu| \geq 3\sigma) \leq \frac{1}{9},$$

co pokazuje, że w ogólnym przypadku szansa na przekroczenie trzech odchyłeń jest wciąż mała, ale może być istotnie większa, niż w przypadku normalnym.

Przedstawmy następnie zbiór podstawowych własności związanych z rozkładem normalnym, które mogą być dla nas użyteczne.

Propozycja 6.13 (Podstawowe własności rozkładu normalnego). Dla rozkładu normalnego zachodzą następujące własności

- 1) (symetria dystrybuanty) Dla $t \in \mathbb{R}_+$ mamy $\Phi(t) = 1 - \Phi(-t)$ oraz $\mathbb{P}(|X| \leq t) = 2\Phi(t) - 1$.
- 2) (przekształcenie afiniczne) Dla $X \sim \mathcal{N}(0, 1)$ i $Y = \sigma X + \mu$ zachodzi $Y \sim \mathcal{N}(\mu, \Sigma)$ i $F_Y(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$.
- 3) (suma kwadratów niezależnych zmiennych) Niech $X_1, \dots, X_n \sim \mathcal{N}(0, 1)$ będą niezależne. Wtedy

$$X_1^2 + \dots + X_n^2 \sim \chi^2(n).$$

- 4) (momenty) Dla $X \sim \mathcal{N}(0, 1)$ i $p \in \mathbb{N}$ dostajemy

$$\mathbb{E}(X^p) = \begin{cases} 0, & \text{gdy } p \text{ jest nieparzyste} \\ (p-1)!!, & \text{gdy } p \text{ jest parzyste.}^a \end{cases}$$

- 5) Dla $X \sim \mathcal{N}(0, 1)$ zachodzi $\mathbb{E}[e^X] = e^{\frac{1}{2}\sigma^2}$ oraz $\mathbb{E}(|X - \mu|) = \sigma\sqrt{2}/\sqrt{\pi}$.

^aTutaj mamy $(p-1)!! := (p-1)(p-3) \cdot \dots \cdot 1$

Propozycje 6.13 pozostawiamy bez dowodu. Dowód własności 1) i 2) jest prostym ćwiczeniem. Własność 3) często traktuje się jako definicję rozkładu χ^2 – dowód ten powinien być przedstawiony na wykładzie ze statystyki. Własność 4) i 5) dowodzi się wykorzystując metody rachunku całkowego (wielokrotne całkowanie przez części, zamiana zmiennych, itd.).

Uwaga 6.14 (Tablice rozkładu normalnego).

Na koniec warto wspomnieć o tablicach standardowego rozkładu normalnego. Tablice te służą do odczytywania wartości dystrybucji zmiennej o dowolnym rozkładzie normalnym – korzystając z własności 1) i 2) w Propozycji 6.13. Przykładowo, korzystając z tabeli po prawej, wartość $\Phi(1.48)$ można odczytać patrząc na wartość w szesnastym wierszu ("1.4") oraz dziewiątej kolumnie ("0.08"), tzn. $\Phi(1.48) \approx 0.9308$.^a

^aOczywiście w obecnych czasach łatwiej skorzystać z komputera. W środowisku **R** wystarczyłoby wpisać `pnorm(1.48)`.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

6.4 Wybrane własności i zależności między rozkładami

Gdy definiowaliśmy rozkłady ciągłe i dyskretne mówiliśmy, że niektóre z nich są związane z innymi. Sformalizujmy teraz kilka z tych obserwacji.

Propozycja 6.15 (Rozkład dwumianowy jako suma prób). Niech $n \in \mathbb{N}$ oraz niech X_1, \dots, X_n będą niezależnymi zmiennymi losowymi o rozkładzie dwupunktowym na $\{0, 1\}$, $\mathbb{P}(X_1 = 1) = p$. Wtedy dla $S_n = X_1 + \dots + X_n$ zachodzi $S_n \sim B(n, p)$.

Dowód. Ustalmy $k \in \{0, 1, \dots, n\}$ oraz zdefiniujmy $X = (X_1, \dots, X_n)$. Dla każdego ustalonego punktu $x = (x_1, \dots, x_n) \in \{0, 1\}^n$ należącego do $D_k := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = k\}$ dostajemy

$$\mathbb{P}(X = x) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) = p^k (1-p)^{n-k}.$$

Łatwo również zauważyć, że $|D_k| = \binom{n}{k}$ oraz dla $x, y \in D_k$ takich, że $x \neq y$ zachodzi $\{X = x\} \cap \{X = y\} = \emptyset$ oraz $\{S_n = k\} = \bigcup_{x \in D_k} \{X = x\}$. Stąd dostajemy już

$$\mathbb{P}(S_n = k) = |D_k| p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k},$$

co kończy dowód. □

Propozycja 6.16 (Rozkład geometryczny jako oczekiwanie na sukces). Niech $n \in \mathbb{N}$ oraz niech X_1, \dots, X_n będą niezależnymi zmiennymi losowymi o rozkładzie dwupunktowym na $\{0, 1\}$, gdzie $\mathbb{P}(X_1 = 1) = p$. Wtedy dla $T = \min_{n \in \mathbb{N}} \{X_n = 1\}$ zachodzi $T \sim \text{Geom}(p)$.

Dowód. Dla każdego $k \in \mathbb{N}$ z niezależności dostajemy

$$\mathbb{P}(T = k) = \mathbb{P}(\{X_1 = 0\} \cap \{X_{k-1} = 0\} \cap \{X_k = 1\}) = (1-p)^{k-1}p,$$

co kończy dowód. □

Propozycja 6.17 (Rozkład Poissona jako graniczna wartość rozkładu dwumianowego). Niech $(p_n)_{n \in \mathbb{N}}$ będzie ciągiem liczb z $(0, 1)$ takim, że $\lim_{n \rightarrow \infty} np_n = \lambda$ dla ustalonego $\lambda \geq 0$. Niech $X_n \sim B(n, p_n)$ oraz niech $Y \sim \text{Pois}(\lambda)$. Wtedy dla każdego $k \in \mathbb{N}$ zachodzi

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

Dowód. Ustalmy $k \in \mathbb{N}$ oraz oznaczmy $\lambda_n := np_n$. Dla odpowiednio dużego n ($n > k$) dostajemy

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\binom{n}{k}}{n^k} \lambda_n^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k}.$$

Zauważając, że $\binom{n}{k}/n^k \rightarrow 1/k!$ oraz $(1 - \frac{\lambda_n}{n})^{-k} \rightarrow 1$, gdy $n \rightarrow \infty$, dostajemy

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \lim_{n \rightarrow \infty} \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n = \frac{\lambda^k}{k!} e^{-\lambda} = \mathbb{P}(X = k),$$

co kończy dowód. □

Propozycja 6.18 (Rozkład wykładniczy jako ciągły odpowiednik rozkładu geometrycznego). Niech $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, oraz dla $n \in \mathbb{N}$ niech $X_n \sim \text{Geom}(p_n)$ dla $p_n = \lambda/n$. Wtedy dla $Y_n := X_n/n$ oraz każdego $t \in \mathbb{R}$ dostajemy

$$\mathbb{P}(Y_n \leq t) \rightarrow \mathbb{P}(X \leq t), \quad n \rightarrow \infty.$$

Dowód. Dla $t > 0$ oraz $n \rightarrow \infty$ dostajemy

$$\mathbb{P}(Y_n \leq t) = 1 - \mathbb{P}(X_n > nt) = 1 - \mathbb{P}(X_n > \lfloor nt \rfloor) = 1 - (1-p_n)^{\lfloor nt \rfloor} = 1 - \left(1 - \frac{\lambda}{n}\right)^{\lfloor nt \rfloor} \rightarrow 1 - e^{-\lambda t},$$

co kończy dowód, gdyż $\mathbb{P}(X \leq t) = 1 - e^{-\lambda t}$. □

Przedstawmy teraz twierdzenie, które daje nam wzór na gęstość dwóch ciągłych i niezależnych zmiennych losowych.

Propozycja 6.19 (Splot niezależnych zmiennych losowych). Niech X oraz Y będą niezależnymi i ciągłymi zmiennymi losowymi o gęstościach f_X i f_Y . Wtedy zmienna losowa $Z = X + Y$ jest ciągłą zmienną losową o gęstości

$$f_Z(x) = \int_{-\infty}^{\infty} f_X(t) f_Y(x-t) dt. \quad (6.1)$$

Dowód Propozycji 6.19 jest prostym zastosowaniem twierdzenia Fubinięgo połączonym z zamianą zmiennych w całce – powinien być przedstawiony na teorii miary i całki; zob Rozdział 5.8 w [JS04]. Oczywiście istnieje analogiczny wzór na dyskretne zmienne losowe. Zakładając, że X oraz Y są niezależnymi dyskretnymi zmiennymi losowymi o wartościach w \mathbb{Z} , łatwo pokazać, że dla $Z = X + Y$ oraz $k \in \mathbb{Z}$ dostajemy

$$\mathbb{P}(Z = k) = \sum_{i=-\infty}^{\infty} \mathbb{P}(X = i)\mathbb{P}(Y = k - i),$$

co można traktować jako odpowiednik wzoru (6.1) dla rozkładów dyskretnych. Korzystając z wzoru na splot niezależnych zmiennych losowych, można znaleźć rozkład sumy wielu popularnych rozkładów prawdopodobieństwa. W następnej propozycji przedstawiamy niektóre z nich.

Propozycja 6.20 (Wybrane rozkłady sum niezależnych zmiennych losowych). Niech X i Y będą niezależnymi zmiennymi losowymi określonymi na $(\Omega, \Sigma, \mathbb{P})$. Wtedy

- 1) Jeżeli $X \sim N(\mu_1, \sigma_1)$ oraz $Y \sim N(\mu_2, \sigma_2)$, to $X + Y \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$;
- 2) Jeżeli $X \sim B(n_1, p)$ oraz $Y \sim B(n_2, p)$, to $X + Y \sim B(n_1 + n_2, p)$;
- 3) Jeżeli $X \sim \text{Pois}(\lambda_1)$ oraz $Y \sim \text{Pois}(\lambda_2)$, to $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$;
- 4) Jeżeli $X \sim \chi^2(n_1)$ oraz $Y \sim \chi^2(n_2)$, to $X + Y \sim \chi^2(n_1 + n_2)$;

Dowód Propozycji 6.20 opiera się o zastosowanie wzoru na splot rozkładów i stosunkowo prostych rachunkach. Pozostawiamy go jako ćwiczenie do domu. O ile dowód własności 1) można przeprowadzić w oparciu o splot funkcji, o tyle prostszą metodą jest obliczenie tzw. funkcji charakterystycznej sumy; wzór na to będzie podany w dalszej części wykładu.

Uwaga 6.21 (Rozkłady sum niezależnych zmiennych). Korzystając z Propozycji 6.19 łatwo obliczyć rozkłady wielu innych sum, choć oczywiście nie zawsze suma niezależnych rozkładów pozostanie w tej samej klasie rozkładów. Przykładowo, suma n niezależnych zmiennych losowych o rozkładzie wykładniczym z tym samym parametrem λ ma tzw. *rozkład Erlanga*, zob. Rozdział 5.8 (Twierdzenie 20) w [JS04].

Na koniec przytoczmy jeszcze jeden rezultat związany z tzw. własnością braku pamięci (zwaną też własnością Markowa). Okazuje się, że w klasie rozkładów ciągłych na $[0, +\infty)$ rozkład wykładniczy jest jedynym rozkładem mającym tę własność. Podobnie, jedynym rozkładem z własnością braku pamięci skupionym na liczbach naturalnych jest rozkład geometryczny. Dowód jedności pozostawiamy jako ćwiczenie do domu, zob. Rozdział 5.1 (Zadanie 16) w [JS04].

Propozycja 6.22. Niech $X \sim \text{Exp}(\lambda)$ (odp. $X \sim \text{Geom}(p)$). Wtedy X spełnia własność braku pamięci, tzn. dla dowolnego $t, s \in \mathbb{R}_+$ (odp. $t, s \in \mathbb{N}$) zachodzi własność

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s).$$

Dowód. Niech $X \sim \text{Exp}(\lambda)$ oraz niech $t, s \in \mathbb{R}_+$. Ponieważ $\mathbb{P}(X > t) > 0$, dostajemy

$$\begin{aligned} \mathbb{P}(X > t + s | X > t) &= \frac{\mathbb{P}(\{X > t + s\} \cap \{X > t\})}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} = \frac{1 - F_X(t + s)}{1 - F_X(t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = 1 - (1 - e^{-\lambda s}) = 1 - F_X(s) = \mathbb{P}(X > s). \end{aligned}$$

Podobnie dla $X \sim \text{Geom}(p)$ oraz $t, s \in \mathbb{N}$, zauważając, że $\mathbb{P}(X > t) > 0$, dostajemy

$$\mathbb{P}(X > t + s | X > t) = \frac{1 - F_X(t + s)}{1 - F_X(t)} = \frac{(1 - p)^{t+s}}{(1 - p)^t} = (1 - p)^s = \mathbb{P}(X > s),$$

co kończy dowód.

□

7 Prawa wielkich liczb i centralne twierdzenie graniczne

W rozdziale tym przedstawimy zbiór ważnych twierdzeń rachunku prawdopodobieństwa dotyczących asymptotycznego zachowania sum zmiennych losowych o tym samym rozkładzie, tj. prawami wielkich liczb i centralnym twierdzeniem granicznym. Będziemy badać asymptotyczne zachowanie wyrażeń postaci

$$\frac{S_n}{n} \quad \text{oraz} \quad \frac{S_n - \mathbb{E}(S_n)}{D(S_n)}, \quad (7.1)$$

gdzie $n \rightarrow \infty$, gdzie $S_n := X_1 + \dots + X_n$ oraz $(X_i)_{i \in \mathbb{N}}$ jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie.¹⁸ W celu uproszczenia notacji, mając zadany ciąg zmiennych losowych $(X_i)_{i \in \mathbb{N}}$, dla każdego $n \in \mathbb{N}$ przez S_n będziemy zaszę oznaczać sumę pierwszy n elementów ciągu. Podamy też kilka pobocznych rezultatów związanych z ogólnym (asymptotycznym) zachowaniem sum niezależnych zmiennych losowych.

Rozważanie wyrażeń postaci (7.1) jest istotne również z praktycznego punktu widzenia, o czy dowiedzą się Państwo więcej na wykładzie ze statystyki. Zakładając, że mamy daną pewną zmienną X o nieznanym rozkładzie ze skończoną średnią oraz jej niezależne replikacje $(X_i)_{i \in \mathbb{N}}$, wyrażenie S_n/n w (7.1) jest naturalnym estymatorem wartości średniej z X . Przykładowo, mając daną monetę (być może sfałszowaną) i oznaczając przez X_i wynik i -tego rzutu, analizując (7.1) powinniśmy być w stanie coraz lepiej (wraz ze zwiększaniem liczby rzutów) szacować prawdopodobieństwo wypadnięcia orła (czyli rozkładu zmiennej X). Aby sformalizować to rozumowanie, potrzebujemy asymptotycznych twierdzeń związanych z (7.1).

Uwaga 7.1 (Asymptotyczna dynamika sum zmiennych losowych). Dla uproszczenia, w wykładzie tym skupimy się na analizie wyrażeń postaci (7.1) dla ciągu niezależnych zmiennych losowych o tym samym rozkładzie. W literaturze znane są jednak niezliczone warianty przedstawionych w tym rozdziale twierdzeń, które mogą zakładać różne rozkłady składników sumy oraz analizować ogólne wyrażenia postaci

$$\frac{X_1 + \dots + X_n - a_n}{b_n},$$

gdzie $(a_n)_{n \in \mathbb{N}}$ i $(b_n)_{n \in \mathbb{N}}$ są pewnymi ciągami liczbowymi. Często rozważa się też tzw. ergodyczne twierdzenia graniczne, w którym dopuszczamy pewną zależność między zmiennymi losowymi. Jeżeli kogoś interesują bardziej ogólne wyniki związane z prawami wielkich liczb i centralnymi twierdzeniami granicznymi, to warto zajrzeć np. do Rozdziału 7 w [JS04].

7.1 Prawo 0-1 Kołmogorowa i zbieżność szeregów zmiennych losowych

Zanim przejdziemy do analizy wyrażeń postaci (7.1) warto zastanowić się co jesteśmy w stanie powiedzieć o asymptotycznym zachowaniu się (nieuśrednionej) sumy $\sum_{i=1}^{\infty} X_i$, gdy ciąg zmiennych $(X_i)_{i \in \mathbb{N}}$ jest niezależny. Aby to zrobić zastanówmy się, jak sformalizować zdarzenia postaci $\{\sum_{i=1}^{\infty} X_i < \infty\}$ i σ -algebrę, do której one należą.

Z Uwagi 4.11 wiemy, że zmienne losowe są niezależne wtedy i tylko wtedy, gdy σ -algebry generowane przez te zmienne są niezależne. Dla każdego $n \in \mathbb{N}$ oznaczmy przez

$$\mathcal{F}_n := \sigma(X_1, \dots, X_n) \quad \text{oraz} \quad \mathcal{F}_{n, \infty} := \sigma(X_n, X_{n+1}, \dots)$$

σ -algebry, które dają nam wiedzę co zdarzyło się do chwili n (tzn. informację którą niosą łącznie zmienne X_1, \dots, X_n) oraz wiedzę o przyszłości od chwili n (tzn. informację którą niosą łącznie

¹⁸Przypominamy, że ciąg zmiennych losowych $(X_i)_{i \in \mathbb{N}}$ nazywamy niezależnym, gdy dla każdego $n \in \mathbb{N}$, zmienne $(X_i)_{i=1}^n$ są niezależne.

zmienne X_n, X_{n+1}). Zdefiniujmy również σ -ciało *ogonowe* dane przez

$$\mathcal{F}_\infty := \bigcap_{n=1}^{\infty} \mathcal{F}_{n,\infty},$$

które niesie ze sobą informację co zdarzy się w "nieskończenie odległej przyszłości"; warto zauważyć, że ciąg $(\mathcal{F}_{n,\infty})_{n \in \mathbb{N}}$ jest zstępujący. Intuicyjnie rzecz biorąc, jeżeli zmienne $(X_i)_{i \in \mathbb{N}}$ są od siebie niezależne, to chcielibyśmy, aby po długim czasie losowość zanikła i przerodziła się (w nieskończoności) w pewność. Intuicję tę formalizuje tzw. Prawo 0-1 Kołmogorowa.

Twierdzenie 7.2 (Prawo 0-1 Kołmogorowa). Niech $(X_i)_{i=1}^{\infty}$ tworzą ciąg niezależnych zmiennych losowych. Wtedy, dla każdego zdarzenia $A \in \mathcal{F}_\infty$ mamy $\mathbb{P}(A) = 0$ lub $\mathbb{P}(A) = 1$.

Dowód. Z niezależności zmiennych losowych wiemy, że dla każdego $n \in \mathbb{N}$ niezależne są σ -algebry \mathcal{F}_n oraz $\mathcal{F}_{n+1,\infty}$. Ponieważ rodzina $(\mathcal{F}_{n,\infty})_{n \in \mathbb{N}}$ jest zstępująca, więc $A \in \mathcal{F}_\infty$ implikuje $A \in \mathcal{F}_{n,\infty}$ dla każdego $n \in \mathbb{N}$. Dostajemy stąd, że dla każdego $n \in \mathbb{N}$ zdarzenie A jest niezależne od \mathcal{F}_n . W szczególności implikuje to, że zdarzenie A jest niezależne od $\sigma(X_1, X_2, \dots)$.¹⁹ Z drugiej strony, wiemy, że $A \in \mathcal{F}_{1,\infty} = \sigma(X_1, X_2, \dots)$, więc zdarzenie A jest w istocie niezależne od zdarzenia A , czyli $\mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ co daje nam $\mathbb{P}(A) = 0$ lub $\mathbb{P}(A) = 1$. \square

Bezpośrednio z twierdzenia 7.2 dostajemy, że nieskończona suma niezależnych zmiennych losowych jest prawie na pewno zbieżna lub rozbieżna.

Propozycja 7.3 (Zbieżność nieskończonej sumy niezależnych zmiennych losowych). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych. Wtedy szereg $\sum_{i=1}^{\infty} X_i$ jest zbieżny z prawdopodobieństwem 0 lub 1.

Dowód. Niech $A := \{\sum_{i=1}^{\infty} X_i < \infty\}$. Ponieważ dla każdego $n \in \mathbb{N}$ zachodzi $A = \{\sum_{i=n}^{\infty} X_i < \infty\}$, więc dla każdego $n \in \mathbb{N}$ dostajemy $A \in \mathcal{F}_{n,\infty}$. Stąd dostajemy $A \in \mathcal{F}_\infty$. Z Twierdzenia 7.2 wynika więc, że $\mathbb{P}(A) = 0$ lub $\mathbb{P}(A) = 1$, co kończy dowód. \square

Uwaga 7.4 (Zdarzenia asymptotyczne). Prawo 0-1 Kołmogorowa można stosować do wielu zdarzeń, w których pytamy o asymptotyczne własności ciągu niezależnych zmiennych losowych. Przykładowo, łatwo pokazać, że $\{\limsup_{n \rightarrow \infty} A_n = \infty\} \in \mathcal{F}_\infty$ lub $\{\lim_{n \rightarrow \infty} S_n < a\} \in \mathcal{F}_\infty$ dla każdego $a \in \mathbb{R}$.

Sformułujmy teraz jedno stosunkowo łatwe do sprawdzenia kryterium zbieżności, które będzie przydatne w dalszej części wykładu; więcej podobnych kryteriów zbieżności można znaleźć np. w Rozdziale 7.3 w [JS04].

Twierdzenie 7.5 (Kryterium Kołmogorowa zbieżności szeregu sumy niezależnych zmiennych losowych oparte o średnią i wariancję). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych. Wtedy, jeżeli szeregi

$$\sum_{i=1}^{\infty} \mathbb{E}(X_i) \quad \text{oraz} \quad \sum_{i=1}^{\infty} D^2(X_i) \quad \text{są zbieżne,}$$

są zbieżne, to szereg $\sum_{i=1}^{\infty} X_i$ jest zbieżny (p.n.).

¹⁹Aby to formalnie pokazać, trzeba skorzystać z twierdzenia o niezależnych π -układach, zob. Rozdział 5.8 (Twierdzenie 12) w [JS04].

Dowód. Dla $i \in \mathbb{N}$ zdefiniujemy $Y_i := X_i - \mathbb{E}[X_i]$; szereg $\sum_{i=1}^{\infty} \mathbb{E}[X_i]$ jest zbieżny, więc aby udowodnić zbieżność szeregu $\sum_{i=1}^{\infty} X_i$ wystarczy pokazać zbieżność szeregu $\sum_{i=1}^{\infty} Y_i$. Niech $S_n := \sum_{i=1}^n Y_i$. Pokażmy teraz, że dla prawie każdego $\omega \in \Omega$ spełniony jest warunek Cauchy'ego, tzn. dla każdego $\epsilon > 0$ istnieje $k \in \mathbb{N}$ takie, że

$$\sup_{m,n \geq k} |S_m(\omega) - S_n(\omega)| < \epsilon. \quad (7.2)$$

Dla każdego ustalonego $k \in \mathbb{N}$ oraz $\epsilon > 0$, korzystając z nierówności Kołmogorowa (Propozycja 5.33), dostajemy

$$\begin{aligned} \mathbb{P} \left(\sup_{m,n \geq k} |S_m - S_n| \geq \epsilon \right) &= \mathbb{P} \left(\sup_{m,n \geq k} |S_m - S_k + S_k - S_n| \geq \epsilon \right) \\ &\leq \mathbb{P} \left(\sup_{n \geq k} |S_n - S_k| \geq \frac{\epsilon}{2} \right) \\ &= \lim_{m \rightarrow \infty} \mathbb{P} \left(\sup_{m \geq n \geq k} |S_n - S_k| \geq \frac{\epsilon}{2} \right) \\ &= \lim_{m \rightarrow \infty} \mathbb{P} \left(\max_{i=1,2,\dots,m-k} |Y_{k+1} + \dots + Y_{k+i}| \geq \frac{\epsilon}{2} \right) \\ &\leq \lim_{m \rightarrow \infty} \frac{D^2(Y_{k+1} + \dots + Y_{k+m})}{(\epsilon/2)^2} \\ &\leq \frac{4}{\epsilon^2} \sum_{i=k+1}^{\infty} D^2(Y_i). \end{aligned} \quad (7.3)$$

Zauważając, że $\sum_{i=1}^{\infty} D^2(Y_i) = \sum_{i=1}^{\infty} D^2(X_i) < \infty$, co implikuje $\lim_{k \rightarrow \infty} \sum_{i=k+1}^{\infty} D^2(Y_i) = 0$, oraz korzystając z (7.3) dostajemy, że dla każdego ustalonego $\epsilon > 0$ zachodzi

$$\mathbb{P} \left(\bigcap_{k=1}^{\infty} \left\{ \sup_{m,n \geq k} |S_m - S_n| \geq \epsilon \right\} \right) = \lim_{k \rightarrow \infty} \mathbb{P} \left(\sup_{m,n \geq k} |S_m - S_n| \geq \epsilon \right) = 0.$$

Z tego wynika, że $\mathbb{P} \left(\bigcup_{k=1}^{\infty} \{ \sup_{m,n \geq k} |S_m - S_n| < \epsilon \} \right) = 1$, co implikuje warunek (7.2) dla prawie wszystkich $\omega \in \Omega$ przy ustalonym $\epsilon > 0$.²⁰ Powtarzając powyższe rozumowanie dla przeliczalnego ciągu $\epsilon_n \searrow 0$, dostajemy warunek (7.2), co kończy dowód. \square

Propozycja 7.5 pozwoli nam w szybki sposób sprawdzać, czy (wybrane) szeregi zmiennych losowych są zbieżne; będziemy go używać np. w dowodzie mocnego prawa wielkich liczb.

7.2 Prawa wielkich liczb

W rozdziale tym zajmiemy się zbieżnością ciągu zmiennych S_n/n , tzn. wyrażeń postaci (7.1), dla $n \rightarrow \infty$, gdy mamy do czynienia z niezależnymi zmiennymi losowymi o tym samym rozkładzie. Intuicyjnie, chcielibyśmy, aby S_n/n zbiegało do średniej z rozkładu. Zaczniemy od sformułowania tzw. słabego prawa wielkich liczb, które jest prostym wnioskiem z nierówności Czebyszewa.

²⁰Oczywiście stała $k \in \mathbb{N}$ w warunku (7.2) może zależeć od wyboru $\epsilon \in \mathbb{N}$.

Twierdzenie 7.6 (Słabe prawo wielkich liczb). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie ze skończoną wariancją i średnią μ . Wtedy, dla każdego $\epsilon > 0$ zachodzi

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n}{n} - \mu \right| \leq \epsilon \right) = 1. \quad (7.4)$$

^aZbieżność tę nazywa się zbieżnością stochastyczną (według prawdopodobieństwa). Więcej na temat różnych typów zbieżności i zależności między nimi dowiemy się w dalszej części wykładu.

Dowód. Dla każdego $n \in \mathbb{N}$, korzystając z niezależności zmiennych losowych X_1, \dots, X_n , dostajemy

$$\mathbb{E} \left(\frac{S_n}{n} \right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \mu \quad \text{oraz} \quad D^2 \left(\frac{S_n}{n} \right) = \frac{1}{n^2} \sum_{i=1}^n D^2(X_i) = \frac{\sigma^2}{n}$$

gdzie $\mu := \mathbb{E}[X_1]$ oraz $\sigma^2 := D^2(X_1)$. Z nierówności Czebyszewa-Bianaymé (5.17) dla każdego $\epsilon > 0$ dostajemy więc

$$\mathbb{P} \left(\left| \frac{S_n}{n} - \mu \right| \leq \epsilon \right) \geq 1 - \mathbb{P} \left(\left| \frac{S_n}{n} - \mu \right| \geq \epsilon \right) \geq 1 - \frac{D^2(S_n/n)}{\epsilon} = 1 - \frac{\sigma^2}{\epsilon \cdot n} \xrightarrow{n \rightarrow \infty} 1,$$

co kończy dowód. □

Uwaga 7.7 (Uogólnienia słabego prawa wielkich liczb). Twierdzenie 7.6 można w łatwy sposób uogólnić na sytuacje, gdy X_1, X_2, \dots są zależne i/lub mają różne rozkłady. Przykładowymi warunkami gwarantującymi zachodzenie (7.4) jest warunek $\lim_{n \rightarrow \infty} \frac{D^2(S_n)}{n^2} = 0$ lub liniowa niezależność (par) zmiennych losowych połączona ze wspólnie ograniczonym drugim momentem. W obu przypadkach dowód jest analogiczny do dowodu Twierdzenia 7.6 (ćw. do domu).

Przejdźmy teraz do mocnego prawa wielkich liczb, które uogólnia nam słabe prawo wielkich liczb. Na początku pokażemy pomocnicze twierdzenie, w którym nie będziemy zakładać identycznego rozkładu zmiennych, a potem przejdziemy do właściwego twierdzenia.

Twierdzenie 7.8 (Twierdzenie Kołmogorowa). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o skończonych wariancjach spełniających warunek $\sum_{i=1}^{\infty} \frac{D^2(X_i)}{i^2} < \infty$. Wtedy zachodzi

$$\lim_{n \rightarrow \infty} \frac{S_n - \mathbb{E}(S_n)}{n} = 0 \quad \text{p.n.} \quad \text{a}$$

^aZbieżność tę nazywa się zbieżnością prawie na pewno. W szczególności implikuje ona zbieżność stochastyczną. Więcej na temat różnych typów zbieżności i zależności między nimi dowiemy się w dalszej części wykładu.

Dowód. Niech $\hat{X}_i := X_i - \mathbb{E}[X_i]$, $i \in \mathbb{N}$. Łatwo zauważyć, że zmienne $(\hat{X}_i/i)_{i \in \mathbb{N}}$ tworzą ciąg niezależnych zmiennych losowych takich, że

$$\sum_{i=1}^{\infty} \mathbb{E} \left(\frac{\hat{X}_i}{i} \right) = 0 < \infty \quad \text{oraz} \quad \sum_{i=1}^{\infty} D^2 \left(\frac{\hat{X}_i}{i} \right) = \sum_{i=1}^{\infty} \frac{D^2(X_i)}{i^2} < \infty.$$

Korzystając z Propozycji 7.5 (kryterium Kołmogorowa zbieżności) dostajemy więc zbieżność szeregu $\sum_{i=1}^{\infty} \hat{X}_i/i$ (p.n.). Z Lematu Kroneckera (zob. Dodatek B) wynika, że dla każdego $\omega \in \Omega$ dla którego

szereg $\sum_{i=1}^{\infty} \hat{X}_i(\omega)/i$ jest zbieżny dostajemy $\lim_{n \rightarrow \infty} \frac{\hat{X}_1(\omega) + \dots + \hat{X}_n(\omega)}{n} = 0$. Stąd dostajemy już

$$\lim_{n \rightarrow \infty} \frac{S_n - \mathbb{E}[S_n]}{n} = \lim_{n \rightarrow \infty} \frac{\hat{X}_1 + \dots + \hat{X}_n}{n} = 0 \quad \text{p.n.},$$

co kończy dowód. \square

Warto zauważyć, że dla niezależnych zmiennych losowych X_1, X_2, \dots o tym samym rozkładzie i skończonej wariancji warunek $\sum_{i=1}^{\infty} \frac{D^2(X_i)}{i^2} < \infty$ w Twierdzeniu (7.8) jest automatycznie spełniony, więc twierdzenie to uogólnia nam słabe prawo wielkich liczb. Okazuje się jednak, że twierdzenie to można wzmocnić, wymagając tylko skończonej średniej, co jest przedstawione w mocnym prawie wielkich liczb.

Twierdzenie 7.9 (Mocne prawo wielkich liczb). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie ze skończoną średnią μ . Wtedy zachodzi

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu \quad \text{p.n.} \quad (7.5)$$

Dowód. Dla $i \in \mathbb{N}$ definiujemy $Y_i := X_i 1_{\{|X_i| \leq i\}}$. Korzystając z Propozycji 5.12 dostajemy $\mathbb{E}(|X_1|) = \int_0^{\infty} \mathbb{P}(|X_1| > t) dt$, co implikuje

$$\sum_{i=1}^{\infty} \mathbb{P}(X_i \neq Y_i) = \sum_{i=1}^{\infty} \mathbb{P}(|X_i| > i) = \sum_{i=1}^{\infty} \mathbb{P}(|X_1| > i) \leq \mathbb{E}(|X_1|) < \infty.$$

Z Lematu Borella-Cantelliego (Twierdzenie 2.18) dostajemy więc $\mathbb{P}(\limsup_{i \rightarrow \infty} \{X_i \neq Y_i\}) = 0$, co oznacza, że dla prawie wszystkich $\omega \in \Omega$ zachodzi $X_i(\omega) = Y_i(\omega)$ dla dostatecznie dużego $i \in \mathbb{N}$.²¹ Stąd wynika, że

$$\lim_{n \rightarrow \infty} \frac{X_1 \dots + X_n}{n} = \mu \quad (\text{p.n.}) \iff \lim_{n \rightarrow \infty} \frac{Y_1 \dots + Y_n}{n} = \mu \quad (\text{p.n.}).$$

Wystarczy więc pokazać zbieżność średnich empirycznych dla ciągu zmiennych losowych $(Y_i)_{i \in \mathbb{N}}$. Zdefiniujmy $\tilde{S}_n := Y_1 + \dots + Y_n$, $n \in \mathbb{N}$. Zauważając, że $\mathbb{E}(Y_n) \rightarrow \mathbb{E}(X_1)$, $n \rightarrow \infty$, oraz korzystając z Lematu Toeplitza (zob. Dodatek B) dostajemy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}(\tilde{S}_n) = \lim_{n \rightarrow \infty} \frac{\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)}{n} = \mathbb{E}(X_1) = \mu.$$

Wystarczy więc pokazać, że zachodzi

$$\lim_{n \rightarrow \infty} \frac{\tilde{S}_n - \mathbb{E}(\tilde{S}_n)}{n} = 0 \quad \text{p.n.} \quad (7.6)$$

W tym celu wystarczy pokazać, że $(Y_i)_{i \in \mathbb{N}}$ spełnia założenia Twierdzenia 7.8. $(Y_i)_{i \in \mathbb{N}}$ jest ciągiem niezależnych zmiennych losowych (gdyż $(X_i)_{i \in \mathbb{N}}$ jest takim ciągiem) o skończonych wariancjach (gdyż dla każdego $i \in \mathbb{N}$ zachodzi $|Y_i| \leq i$). Wystarczy więc pokazać warunek $\sum_{i=1}^{\infty} \frac{D^2(Y_i)}{i^2} < \infty$. Dostajemy

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{D^2(Y_i)}{i^2} &\leq \sum_{i=1}^{\infty} \frac{\mathbb{E}(1_{\{|X_i| \leq i\}} X_i^2)}{i^2} = \sum_{i=1}^{\infty} \frac{\mathbb{E}(1_{\{|X_1| \leq i\}} X_1^2)}{i^2} = \sum_{i=1}^{\infty} \left[\sum_{k=1}^i \frac{\mathbb{E}(1_{\{k-1 < |X_1| \leq k\}} X_1^2)}{i^2} \right] \\ &= \sum_{k=1}^{\infty} \left[\mathbb{E}(1_{\{k-1 < |X_1| \leq k\}} X_1^2) \cdot \sum_{i=k}^{\infty} \frac{1}{i^2} \right]. \end{aligned}$$

²¹Oczywiście i zależy od ω

Zauważając teraz, że dla $k \geq 2$ mamy

$$\sum_{i=k}^{\infty} \frac{1}{i^2} \leq \sum_{i=k}^{\infty} \frac{1}{i(i-1)} = \left(\sum_{i=1}^{\infty} \frac{1}{i(i+1)} - \sum_{i=1}^{k-1} \frac{1}{i(i+1)} \right) = \left(1 - \left(1 - \frac{1}{k} \right) \right) = \frac{1}{k} < \frac{2}{k}$$

oraz dla $k = 1$ mamy $\sum_{i=k}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6} < 2$, dostajemy

$$\sum_{i=1}^{\infty} \frac{D^2(X_i)}{i^2} \leq \sum_{k=1}^{\infty} \left[\mathbb{E}(1_{\{k-1 < |X_1| \leq k\}} X_1^2) \cdot \frac{2}{k} \right] \leq 2 \sum_{k=1}^{\infty} [\mathbb{E}(1_{\{k-1 < |X_1| \leq k\}} |X_1|)] = 2\mathbb{E}(|X_1|) < \infty,$$

co kończy dowód. \square

Oczywiście założenie skończonej średniej w mocnym prawie wielkich liczb jest konieczne, zob. Uwaga 7.10.

Uwaga 7.10 (Mocne prawo wielkich liczb, a nieskończona średnia). Załóżmy, że X_1, X_2, \dots jest ciągiem niezależnych zmiennych losowych o takim samym rozkładzie o nieskończonej średniej, tzn. $\mathbb{E}[|X_1|] = \infty$. Wtedy zachodzi $\limsup_{n \rightarrow \infty} S_n = \infty$ (p.n.), co jest prostym wnioskiem z lematu Borela-Cantelliego (ćw. do domu).

Mocne prawo wielkich liczb pozwala nam na interpretacje częstościową prawdopodobieństwa.

Uwaga 7.11 (Częstościowa interpretacja prawdopodobieństwa). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie i skończonej średniej, które odpowiadają nieskończonemu powtórzeniu pewnego eksperymentu. W chwili $n \in \mathbb{N}$, możemy zdefiniować *empiryczne prawdopodobieństwo* zajścia zdarzenia $A \in \Sigma$, zależne od wyników uzyskanych do chwili n , jako $f_n(A) = \frac{1_{\{X_1 \in A\}} + \dots + 1_{\{X_n \in A\}}}{n}$. Z mocnego prawa wielkich liczb dostajemy od razu $\lim_{n \rightarrow \infty} f_n(A) = \mathbb{E}(1_{\{X_1 \in A\}}) = \mathbb{P}(X_1 \in A)$.

Uwaga 7.12 (Dystrybuanta empiryczna). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie i skończonej średniej, które odpowiadają nieskończonemu powtórzeniu pewnego eksperymentu. W chwili $n \in \mathbb{N}$, korzystając z podobnego rozumowania jak w Uwadze 7.12, możemy zdefiniować *dystrybuantę empiryczną* daną wzorem

$$F_n(t) := \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq t\}}.$$

Zauważmy, że $F_n(t)$ dla każdego $n \in \mathbb{N}$ jest zmienną losową. Mając dany wynik eksperymentu (dla pewnego $\omega \in \Omega$), $F_n(t)(\omega)$ jest empirycznym przybliżeniem prawdziwej dystrybuanty. Z mocnego prawa wielkich liczb wynika, że wraz ze zwiększaniem liczby eksperymentów przybliżamy w dobry sposób dystrybuantę, tzn. dla $t \in \mathbb{R}$ zachodzi $\lim_{n \rightarrow \infty} F_n(t) = F_{X_1}(t)$ (p.n.). Istotnie, dla ustalonego $t \in \mathbb{R}$, rozważając ciąg niezależnych zmiennych losowych o tym samym rozkładzie dany przez $Z_i := 1_{\{X_i \leq t\}}$ z mocnego prawa wielkich liczb dostajemy $\lim_{n \rightarrow \infty} F_n(t) = \mathbb{E}(1_{\{X_1 \leq t\}}) = \mathbb{P}(X_1 \leq t) = F_{X_1}(t)$. Oczywiście pokazaliśmy tutaj zbieżność rozważając osobno $t \in \mathbb{R}$. Można pokazać, że w istocie zbieżność ta jest jednostajna ze względu na t , co jest jednym z podstawowych twierdzeń statystyki matematycznej (tzw. Twierdzenie Glivenko–Cantelliego), zob. Twierdzenie 20.6 w [Bil12]. Więcej informacji na ten temat pojawi się na wykładzie ze statystyki.

7.3 Centralne Twierdzenie Graniczne (CTG)

W poprzednim rozdziale analizowaliśmy do czego mogą dążyć ciągi średnich wartości niezależnych zmiennych losowych, gdy zmienne te mają taki sam rozkład. W rozdziale tym odpowiemy na pytanie, co można (asymptotycznie) powiedzieć o rozkładzie średniej ze zmiennych losowych, tzn. wartości S_n/n , gdy go odpowiednio unormujemy/zestandardyzujemy. Na początek zastanówmy się, w jaki sposób taka standaryzacja mogłaby wyglądać. Mając dany ciąg niezależnych zmiennych o tym samym rozkładzie, aby uzyskać transformację średniej wartości zmiennej o zerowej średniej i jednostkowej wariancji powinniśmy odjąć wartość $\mathbb{E}(S_n) = n\mathbb{E}(X_1)$ oraz podzielić przez $D(S_n) = \sqrt{n}D(X_1)$.

Definicja 7.13 (Standaryzacja średniej wartości zmiennej losowej). Niech X_1, \dots, X_n tworzą ciąg niezależnych zmiennych losowych o tym samym rozkładzie ze skończoną średnią μ i odchyleniem standardowym σ oraz niech $S_n := X_1 + \dots + X_n$. Wtedy **standaryzacją sumy** S_n nazywamy zmienną losową daną przez

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{D(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \quad (7.7)$$

Wprost z definicji widzimy, że dla każdego $n \in \mathbb{N}$ dostajemy $\mathbb{E}(Z_n) = 0$ oraz $D(Z_n) = 1$. Naturalnym pytaniem wydaje się więc to, jaki asymptotyczny rozkład ma ciąg zmiennych losowych $(Z_n)_{n \in \mathbb{N}}$. Na pytanie to odpowiada centralne twierdzenie graniczne.

Twierdzenie 7.14 (Centralne Twierdzenie Graniczne – Twierdzenie Lindeberga-Levy’ego). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie ze skończoną średnią $\mu \in \mathbb{R}$ i odchyleniem $\sigma > 0$. Wtedy dla dowolnego $t \in \mathbb{R}$ zachodzi

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq t) = \Phi(t),^a \quad (7.8)$$

tj. asymptotycznym rozkładem zestandardyzowanego ciągu zmiennych losowych $(Z_n)_{n \in \mathbb{N}}$ zdefiniowanego w (7.7) jest standardowy rozkład normalny.

^aZbieżność tę nazywa się zbieżnością według rozkładów lub słabą zbieżnością. Więcej na temat różnych typów zbieżności i zależności między nimi dowiemy się w dalszej części wykładu.

Ponieważ klasyczny dowód Twierdzenia 7.14 wymaga użycia funkcji charakterystycznych, których znajomość wykracza poza zakres wykładu z Rachunku Prawdopodobieństwa 1, więc nie będziemy go tutaj przedstawiać. Dowód ten można znaleźć np. w Rozdziale 10 w [JS04]. Warto spojrzeć też na Twierdzenie 5.5.14 w [CB02], gdzie przeprowadzony jest dowód wykorzystujący tzw. funkcje tworzące przy mocniejszych założeniach (dowód ten nie korzysta z metod analizy zespolonej). Z Twierdzenia 7.14 wynikają dwa stosunkowo proste wnioski.

Twierdzenie 7.15 (Centralne Twierdzenie Graniczne dla sumy i średniej). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie ze skończoną średnią $\mu \in \mathbb{R}$ i odchyleniem $\sigma > 0$. Wtedy dla dowolnego $t \in \mathbb{R}$ zachodzi

$$\lim_{n \rightarrow \infty} \left[\mathbb{P}(S_n \leq t) - \Phi_{n\mu, \sigma\sqrt{n}}(t) \right] = 0 \quad \text{oraz} \quad \lim_{n \rightarrow \infty} \left[\mathbb{P}\left(\frac{S_n}{n} \leq t\right) - \Phi_{\mu, \frac{\sigma}{\sqrt{n}}}(t) \right] = 0,^a \quad (7.9)$$

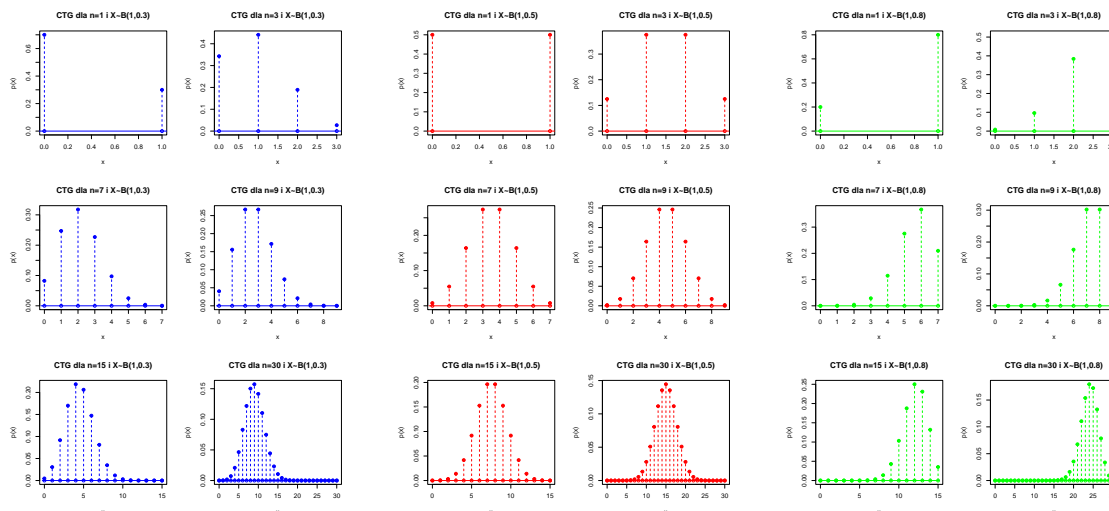
^a $\Phi_{\mu, \sigma}$ oznacza dystrybuantę rozkładu $\mathcal{N}(\mu, \sigma)$.

Wniosek płynący z twierdzeń 7.14–7.15 może być na pierwszy rzut oka zaskakujący: przy stosunkowo prostych założeniach nałożonych na średnią i wariancję otrzymujemy w granicy rozkład normalny unormowanej sumy, co wydaje się tłumaczyć jego uniwersalność tego rozkładu. Sumując więc dużo *małych* błędów (o skończonej wariancji) powinniśmy zawsze dojść do rozkładu normalnego. Doceniając elegancję tego wyniku, należy być jednak świadomym jego ograniczeń - o ile znamy rozkład asymptotyczny, o tyle nie daje nam to dokładnej kontroli nad wielkością błędów i dokładnością przybliżenia. Do tego obecnie, biorąc pod uwagę dużą moc komputerów, często lepszym i bardziej praktycznym wyjściem jest bezpośrednia symulacja rozkładów sum. Pokażemy teraz przykład, które ilustruje Centralne Twierdzenie Graniczne (CTG).

Przykład 7.16 (CTG dla sumy z prób Bernoulliego). Załóżmy, że $(X_i)_{i \in \mathbb{N}}$ jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie $B(1, p)$ dla pewnego $p \in (0, 1)$. Łatwo zauważyć, że dla każdego $n \in \mathbb{N}$ zmienną Z_n można przedstawić jako

$$Z_n := \frac{S_n - np}{\sqrt{p(1-p)n}},$$

gdzie $S_n \sim B(n, p)$. Wykresy dla różnych wartości n oraz $p = 0.3$ (niebieski), $p = 0.5$ (czerwony) o $p = 0.8$ (zielony) przedstawione są poniżej. Jak widać, w każdym przypadku zblizamy się (w różnym tempie) do rozkładu przypominającego rozkład normalny.



W następnej kolejności pokażemy, jak używać CTG do szacowania wartości sum, czy średnich. W większości tych zadań należy skorzystać z tablic rozkładu normalnego.

Przykład 7.17 (Szacowanie wartości na podstawie CTG – rzuty kostką). Centralne Twierdzenie Graniczne może być często przydatne do oszacowań sum. Załóżmy, że rzucono 1000 razy symetryczną kostką do gry i chcemy obliczyć (w przybliżeniu) prawdopodobieństwo tego, że szóstka wypadła więcej niż 150 razy. Dla $n := 1000$ oraz $i = 1, \dots, n$ definiujemy $X_i := 1_{\{Y_i=6\}}$, gdzie $(Y_i)_{i=1}^n$ jest ciągiem niezależnych zmiennych losowych o wspólnym rozkładzie $U(\{1, \dots, 6\})$. Oczywiście ciąg $(X_i)_{i=1}^n$ jest również ciągiem niezależnych zmiennych mających o wspólnym rozkładzie $B(1, 1/6)$.

Dokonując standaryzacji $(X_i)_{i=1}^n$ i korzystając z CTG dostajemy

$$\begin{aligned} \mathbb{P}(S_n > 150) &= \mathbb{P}\left(\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}D(X_1)} > \frac{150 - n\mathbb{E}(X_1)}{\sqrt{n}D(X_1)}\right) = 1 - \mathbb{P}\left(Z_n \leq \frac{150 - n/6}{\sqrt{n}\frac{\sqrt{5}}{6}}\right) \\ &\stackrel{\text{CTG}}{\approx} 1 - \Phi\left(\frac{150 - n/6}{\sqrt{n}\frac{\sqrt{5}}{6}}\right) \approx 1 - \Phi(-1.41) = \Phi(1.41) \approx 0.9207. \end{aligned}$$

Przykład 7.18 (Szacowanie wartości na podstawie CTG – wybory). Rozważmy następujące zadanie: *Aby stwierdzić, jak wielu wyborców popiera obecnie partię ABC, losujemy spośród nich reprezentatywną próbkę i na niej przeprowadzamy badanie. Jak duża powinna być ta próbka, aby uzyskany wynik różnił się od rzeczywistego poparcia dla partii ABC nie więcej niż o 3% z prawdopodobieństwem co najmniej 95%?* Niech $p \in (0, 1)$ oznacza faktyczne (nieznane) poparcie dla ABC. Jeżeli próbka składałaby się z $n \in \mathbb{N}$ osób, gdzie $X_i \sim B(1, p)$, $i = 1, \dots, n$, jest ciągiem niezależnych zmiennych w których sukces utożsamiamy z głosowaniem na ABC, to zmienna $\frac{S_n}{n}$ wyznacza (empiryczne) poparcie na podstawie próbki. Chcemy znaleźć takie $n \in \mathbb{N}$, żeby dla $b := 3\%$ i $\alpha := 5\%$ zachodziło

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq b\right) \geq 1 - \alpha.$$

Korzystając z CTG dostajemy

$$\begin{aligned} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq b\right) &= \mathbb{P}\left(-b \leq \frac{S_n - np}{n} \leq b\right) = \mathbb{P}\left(\frac{-nb}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{nb}{\sqrt{np(1-p)}}\right) \\ &= \mathbb{P}\left(\frac{-b\sqrt{n}}{\sqrt{p(1-p)}} \leq Z_n \leq \frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) \stackrel{\text{CTG}}{\approx} \Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right). \\ &= 2\Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1. \end{aligned}$$

Szukamy więc takiego $n \in \mathbb{N}$ aby zachodziła nierówność

$$2\Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 1 - \alpha.$$

Obłożenie obu stron nierówności funkcją odwrotną do Φ i proste rachunki dają nam równoważną postać

$$n \geq \left(\frac{\Phi^{-1}(1-\frac{\alpha}{2})}{b}\right)^2 (1-p)p.$$

Nie znamy parametru $p \in (0, 1)$ ale widzimy, że maksimum funkcji $p \rightarrow (1-p)p$ na $(0, 1)$ wynosi $\frac{1}{4}$. Podstawiając tę oraz inne wartości do poprzedniej nierówności, oraz odczytując wartość $\Phi^{-1}(0.975)$ z tablic i otrzymujemy więc

$$n \geq \frac{1}{4} \left(\frac{\Phi^{-1}(0.975)}{3\%}\right)^2 \approx \left(\frac{196}{6}\right)^2 \approx 1067.$$

Jeżeli mielibyśmy wstępne informacje o poparciu dla partii ABC – na przykład wiemy, że poparcie to jest mniejsze niż 20% – możemy powyższy wynik znacznie polepszyć. Na przedziale $(0, 0, 2)$, minimum funkcji $(1-p)p$ to 0.16, co oznacza, że $n \geq 683$ jest wystarczającą wielkością próbki.

W Twierdzeniu 7.14 zakładaliśmy, że mamy dany ciąg niezależnych zmiennych losowych o tym samym rozkładzie. W literaturze istnieje wiele wariantów CTG, które pozwalają osłabić ten warunek dopuszczając m.in. różne rozkłady zmiennych losowych bądź zależność między nimi. Przedstawmy teraz dwa warunki, które są wystarczające, aby znormalizowany ciąg serii miał rozkład normalny, gdy nie zakładamy identycznego rozkładu zmiennych losowych. Zaczniemy od wariantu CTG opartego o tzw. *warunek Lapunowa*.

Twierdzenie 7.19 (Centralne Twierdzenie Graniczne – Twierdzenie Lapunowa). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o skończonych średnich $\mu_i \in \mathbb{R}$ i odchyleniach $\sigma_i > 0$, $i \in \mathbb{N}$. Dla $n \in \mathbb{N}$ zdefiniujmy $s_n^2 := \sum_{i=1}^n \sigma_i^2$ i załóżmy, że istnieje $\delta > 0$, takie, że

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}(|X_i - \mu_i|^{2+\delta}) = 0.^a$$

Wtedy dla dowolnego $t \in \mathbb{R}$ zachodzi

$$\lim_{n \rightarrow \infty} \mathbb{P}(\tilde{Z}_n \leq t) = \Phi(t),$$

gdzie $\tilde{Z}_n := \frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i)$, $n \in \mathbb{N}$.

^aMówiąc inaczej, ciąg $(X_i)_{i \in \mathbb{N}}$ spełnia tzw. *warunek Lapunowa* z parametrem $\delta > 0$. W praktyce najczęściej sprawdza się ten warunek dla $\delta = 1$.

Można też zdefiniować wersję CTG ze słabszym warunkiem, nazywanym *warunkiem Lindeberga*.

Twierdzenie 7.20 (Centralne Twierdzenie Graniczne – Twierdzenie Lindeberga). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o skończonych średnich $\mu_i \in \mathbb{R}$ i odchyleniach $\sigma_i > 0$, $i \in \mathbb{N}$. Dla $n \in \mathbb{N}$ zdefiniujmy $s_n^2 := \sum_{i=1}^n \sigma_i^2$ i załóżmy, że dla każdego $\epsilon > 0$ zachodzi

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}((X_i - \mu_i)^2 \cdot 1_{\{|X_i - \mu_i| > \epsilon s_n\}}) = 0^a$$

Wtedy dla dowolnego $t \in \mathbb{R}$ zachodzi

$$\lim_{n \rightarrow \infty} \mathbb{P}(\tilde{Z}_n \leq t) = \Phi(t),$$

gdzie $\tilde{Z}_n := \frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i)$, $n \in \mathbb{N}$.

^aMówiąc inaczej, ciąg $(X_i)_{i \in \mathbb{N}}$ spełnia tzw. *warunek Lindeberga*.

Dowód Twierdzenia 7.19 oraz Twierdzenia 7.20 wykracza poza materiał tego wykładu. Można go znaleźć np. w Rozdziale 27 w [Bil12]. Łatwo zauważyć, że warunek Lapunowa implikuje warunek Lindeberga. Do tego, ciąg zmiennych losowych spełniających założenia klasycznego CTG, tzn. Twierdzenia 7.19, spełnia warunek Lindeberga.

Kolejnym naturalnym pytaniem związanym z CTG jest tempo zbieżności, tzn. jak szybko rozkład unormowanych sum dąży do rozkładu normalnego. Odpowiada na nie częściowo tzw. Twierdzenie Barry’ego-Esseena, które przedstawiamy bez dowodu. Po więcej szczegółów odsyłamy do [PS00].

Twierdzenie 7.21 (Twierdzenie Barry’ego-Esseena). Niech $(X_i)_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o takim samym rozkładzie i skończonym trzecim momencie. Wtedy, dla $n \in \mathbb{N}$ zachodzi

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(Z_n < t) - \Phi(t)| \leq C \frac{\mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)}{\sigma^3 \sqrt{n}}$$

gdzie $\sigma := D(X_1)$ oraz $1/\sqrt{2\pi} \leq C \leq 0.8$ jest pewną stałą.

Na koniec tego rozdziału przedstawmy jeszcze kilka uwag związanych z CTG.

Uwaga 7.22 (CTG dla zależnych zmiennych losowych). Jak już wspomnieliśmy, istnieją również wersje CTG, które dopuszczają zależność między zmiennymi losowymi. W literaturze często nazywa się je *ergodycznymi* wersjami CTG. Po więcej informacji odsyłamy do Rozdziału 27 w [Bil12] oraz Rozdziału w [Hay00].

Uwaga 7.23 (CTG dla zmiennych losowych o nieskończonej wariancji). W Twierdzeniu 7.14 zakładamy, że zmienne losowe mają skończone wariancje. Naturalnym pytaniem wydaje się, co się dzieje z sumami niezależnych zmiennych losowych o tych samym rozkładach, gdy rozkład ten ma nieskończony drugi moment. Na pytanie to odpowiada tak zwane *uogólnione Centralne Twierdzenie Graniczne* które mówi, że klasa tzw. *rozkładów α -stabilnych* odpowiada klasie rozkładów granicznych dla zmiennych losowych o nieskończonej wariancji. Więcej szczegółów na ten temat można znaleźć w Rozdziale 1.8 w [Nol20].

8 Zbieżność zmiennych losowych i ich rozkładów

W poprzednim rozdziale poznaliśmy kilka różnych typów zbieżności zmiennych losowych. W szczególności rozróżniliśmy słabe i mocne prawo wielkich liczb, w którym ciąg zmiennych losowych dążył do swojej granicy na różne sposoby. Dla ciągu $Y_n = S_n/n$ rozważaliśmy zbieżność do granicy $\mu \in \mathbb{R}$ danej przez

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - \mu| \leq \epsilon) = 1 \quad \text{oraz} \quad \lim_{n \rightarrow \infty} Y_n = \mu \quad \text{p.n.}, \quad (8.1)$$

w której pierwsza równość powinna zachodzić dla dowolnego $\epsilon > 0$. Dodatkowo w CTG wymagaliśmy, aby rozkład unormowanej sumy zmiennych losowych $Y_n = Z_n$ dążył do rozkładu normalnego, tj. aby dla każdego $t \in \mathbb{R}$ zachodził warunek

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n \leq t) = \Phi(t). \quad (8.2)$$

Rozważając dowolny ciąg zmiennych losowych $(Y_n)_{n \in \mathbb{N}}$ każda ze zbieżności rozważana w (8.1) oraz (8.2) oznacza co innego. W rozdziale tym spróbujemy usystematyzować te pojęcia, zbadać zależność między nimi, oraz wprowadzić inne typy zbieżności oparte o zbieżności wartości oczekiwanych.

8.1 Rodzaje zbieżności zmiennych losowych

Ciąg zmiennych losowych X_n , jako ciąg funkcji, może dążyć do pewnej zmiennej losowej X . Spróbujmy teraz scharakteryzować różne typy zbieżności i zastanowić się, czy, się one od siebie różnią.

Definicja 8.1 (Różne typy zbieżności zmiennych losowych). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ oraz niech X będzie zmienną losową określoną na tej samej przestrzeni. Mówimy, że ciąg zmiennych (X_n) dąży do X :

1) **prawie na pewno**, ozn. $X_n \xrightarrow{p.n.} X$ lub $X_n \xrightarrow{a.s.} X$, gdy

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

2) **stochastycznie** (według prawdopodobieństwa), ozn. $X_n \xrightarrow{\mathbb{P}} X$ lub $X_n \xrightarrow{s} X$, gdy

$$\forall \epsilon > 0: \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0;$$

3) **według momentu rzędu p** (dla $p > 0$), ozn. $X_n \xrightarrow{L^p} X$, gdy

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = 0;^a$$

4) **według rozkładu** (według dystrybuanty, słaba zbieżność), ozn. $X_n \xrightarrow{d} X$, gdy

$$\forall t \in \mathbb{R} : (F \text{ jest ciągła w } t) \implies \lim_{n \rightarrow \infty} F_n(t) = F(t),$$

gdzie $F(t) := \mathbb{P}(X \leq t)$ oraz $F_n(t) := \mathbb{P}(X_n \leq t)$, $n \in \mathbb{N}$.

^aw przypadku $p = 1$ i $p = 2$ zbieżność tę nazywamy się czasami odpowiednio **zbieżnością według średnich** oraz **zbieżnością średniokwadratową**.

Łatwo zauważyć, że zbieżności 1), 2) oraz 3) są bezpośrednim odpowiednikiem nierówności rozważanych na kursie teorii miary i całki (odp. *zbieżność prawie wszędzie*, *zbieżność według miary*, oraz *zbieżność w normie L^p*). Zbieżność 1) rozważaliśmy przy okazji mocnego prawa wielkich liczb, zbieżność według prawdopodobieństwa przy okazji słabego prawa wielkich liczb, a zbieżność według rozkładu przy okazji CTG. Spróbujmy teraz pokrótce opisać każdy z tych typów zbieżności uwypuklając różnice między nimi, po bardziej szczegółowy opis odsyłamy do Rozdziału 7.1 w [GS01b].

Zbieżność prawie na pewno można traktować jako odpowiednik zbieżności punktowej dostosowanej do rachunku prawdopodobieństwa. Chcielibyśmy, aby poza pewnym zbiorem A o mierze zero (tj. takim, że $\mathbb{P}(A) = 0$), ciąg liczb $X_n(\omega)$ dążył do liczby $X(\omega)$ dla każdego elementu $\omega \in \Omega \setminus A$. Warto tutaj zaznaczyć, że wymaganie zbieżności dla wszystkich elementów zbioru Ω klóciłoby się z Uwagą 3.5 w której stwierdziliśmy, że chcielibyśmy identyfikować ze sobą zmienne losowe, które są sobie równe poza zbiorem miary zero – zbieżność prawie na pewno jest naturalnym następstwem tej konwencji. Warto zauważyć, że w definicji nie odnosimy się bezpośrednio do rozkładów zmiennych losowych, a zbieżność ta jest ściśle związana z przestrzenią probabilistyczną Ω .

Zbieżność stochastyczna odpowiada zbieżności według miary. Mając zmienne X_n i X zbiór punktów odległych od siebie o co najmniej ϵ można zdefiniować jako $A = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}$, co daje nam naturalne oszacowanie z użyciem miary \mathbb{P} , tj. $\mathbb{P}(|X_n - X| > \epsilon) = \int_A 1 d\mathbb{P}$. Chcielibyśmy, aby odległość da dążyła do zera niezależnie od wyboru $\epsilon > 0$, co definiuje już zbieżność stochastyczną.

Zbieżność według momentu rzędu p odpowiada klasycznej zbieżności w normie L^p znanej z analizy matematycznej. Istotnie, norma w przestrzeni L^p zadana przez $\|X\|_p := (\int_{\Omega} |X|^p d\mathbb{P})^{1/p}$ jest związana bezpośrednio ze zbieżnością według momentu rzędu p . Jest to typowa zbieżność oparta o warunek całkowity. Więcej informacji o przestrzeniach L^p można znaleźć w Dodatku C w [JS04].

Zbieżność według rozkładu jest ważnym typem zbieżności w rachunku prawdopodobieństwa i pozwala na formułowanie wniosków podobnych do tych przedstawionych w CTG. Warto zauważyć, że zbieżność ta jest związana tylko z rozkładami zmiennych losowych, a nie ich samymi wartościami – w pewnym sensie zbieżność ta nie zależy od wyboru przestrzeni Ω , itd.. W związku z tym zazwyczaj będzie istniało nieskończenie wiele granic X dla ciągu $(X_n)_{n \in \mathbb{N}}$.²²

8.2 Zależność między różnymi typami zbieżności

W podrozdziale tym zbadamy zależności między różnymi typami zbieżności. Zależności zawsze zachodzące są podsumowane w Twierdzeniu 8.2

Twierdzenie 8.2 (Ogólna zależność między różnymi typami zbieżności zmiennych losowych). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ oraz niech X będzie zmienną losową określoną na tej samej przestrzeni. Wtedy zachodzą następujące implikacje

$$\begin{aligned} (X_n \xrightarrow{p.n.} X) &\implies (X_n \xrightarrow{\mathbb{P}} X) \implies (X_n \xrightarrow{d} X). \\ (X_n \xrightarrow{L^p} X) &\implies (X_n \xrightarrow{\mathbb{P}} X) \implies (X_n \xrightarrow{d} X). \end{aligned}$$

Dodatkowo, dla $p > q \geq 1$ zachodzi $(X_n \xrightarrow{L^p} X) \implies (X_n \xrightarrow{L^q} X)$.

Dowód. Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ oraz niech X będzie zmienną losową określoną na tej samej przestrzeni. Pokażmy po kolei każdą z implikacji.

²²W związku z tym w części literatury zbieżność ta jest przedstawiana osobno i definiowana jako zbieżność rozkładów, a nie zbieżność ciągów zmiennych losowych.

1) $[(X_n \xrightarrow{\mathbb{P}} X) \Rightarrow (X_n \xrightarrow{d} X)]$. Załóżmy, że F jest ciągła w $t \in \mathbb{R}$ oraz ustalmy $\epsilon > 0$. Dostajemy wtedy

$$\begin{aligned} F(t - \epsilon) &= \mathbb{P}(X \leq t - \epsilon) = \mathbb{P}(X \leq t - \epsilon, X_n \leq t) + \mathbb{P}(X \leq t - \epsilon, X_n > t) \\ &\leq \mathbb{P}(X_n \leq t) + \mathbb{P}(|X - X_n| > \epsilon) \\ F(t + \epsilon) &= \mathbb{P}(X \leq t + \epsilon) \geq \mathbb{P}(X_n \leq t, X \leq t + \epsilon) + (\mathbb{P}(X_n \leq t, X > t + \epsilon) - \mathbb{P}(|X_n - X| > \epsilon)) \\ &\geq \mathbb{P}(X_n \leq t) - \mathbb{P}(|X - X_n| > \epsilon), \end{aligned}$$

co daje nam $F(t - \epsilon) - \mathbb{P}(|X - X_n| > \epsilon) \leq F_n(t) \leq F(t + \epsilon) + \mathbb{P}(|X - X_n| > \epsilon)$. Korzystając z założeń, dla $n \rightarrow \infty$, dostajemy więc

$$F(t - \epsilon) \leq \liminf_{n \rightarrow \infty} F_n(t) \leq \limsup_{n \rightarrow \infty} F_n(t) \leq F(t + \epsilon).$$

Następnie, ponieważ funkcja F jest ciągła w t , to dla $\epsilon \searrow 0$ dostajemy równość granicy dolnej i górnej, co implikuje istnienie granicy, tzn. zachodzi

$$\lim_{n \rightarrow \infty} F_n(t) = \lim_{\epsilon \rightarrow 0} F(t \pm \epsilon) = F(t),$$

co kończy dowód.

2) $[(X_n \xrightarrow{L^q} X) \Rightarrow (X_n \xrightarrow{L^p} X)]$. Wiemy, że $q \geq p \geq 1$. Korzystając z nierówności Höldera (Propozycja 5.30) dla $\tilde{X} := |X_n - X|^p$, $\tilde{Y} := 1$ oraz $\tilde{p} := \frac{q}{p} > 1$ (i sprzężonej liczby \tilde{q}) dostajemy

$$\mathbb{E}(|X_n - X|^p) = \mathbb{E}(|\tilde{X}\tilde{Y}|) \leq \mathbb{E}(|\tilde{X}|^{\tilde{p}})^{1/\tilde{p}} \mathbb{E}(|\tilde{Y}|^{\tilde{q}})^{1/\tilde{q}} = \mathbb{E}(|X_n - X|^q)^{p/q} \rightarrow 0, \quad n \rightarrow \infty.$$

co kończy dowód.

3) $[(X_n \xrightarrow{L^p} X) \Rightarrow (X_n \xrightarrow{\mathbb{P}} X)]$. Wobec 2) wystarczy pokazać, że implikacja zachodzi dla $p = 1$. Ustalmy $\epsilon > 0$. Korzystając z nierówności Czebyszewa (Propozycja 5.31) dostajemy

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}(|X_n - X|)}{\epsilon} \rightarrow 0, \quad n \rightarrow \infty.$$

co kończy dowód.

4) $[(X_n \xrightarrow{p.n.} X) \Rightarrow (X_n \xrightarrow{\mathbb{P}} X)]$. Dla każdego $n, m \in \mathbb{N}$ oraz $\epsilon > 0$ definiujemy

$$A_n(\epsilon) := \{|X_n - X| > \epsilon\} \quad \text{oraz} \quad A(\epsilon) := \limsup_{n \rightarrow \infty} A_n(\epsilon).$$

Jeżeli $\omega \in A$ to dla nieskończenie wielu $n \in \mathbb{N}$ zachodzi $|X_n(\omega) - X(\omega)| > \epsilon$, co daje $X_n(\omega) \not\rightarrow X(\omega)$, gdy $n \rightarrow \infty$. Jeżeli więc $X_n \xrightarrow{p.n.} X$, to $\mathbb{P}(A(\epsilon)) = 0$ dla każdego $\epsilon > 0$. Zauważmy teraz, że $B_m(\epsilon) := \bigcup_{n=m}^{\infty} A_n(\epsilon)$ jest zstępującą rodziną zbiorów dającą w granicy $A(\epsilon)$. W związku z tym wiemy, że z $\mathbb{P}(A(\epsilon)) = 0$ wynika $\mathbb{P}(B_m(\epsilon)) \rightarrow 0$, gdy $m \rightarrow \infty$. Zauważając, że $A_n(\epsilon) \subseteq B_m(\epsilon)$, $n \in \mathbb{N}$, dostajemy więc, że dla dowolnego $\epsilon > 0$ zachodzi

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(A_n(\epsilon)) \leq \mathbb{P}(B_m(\epsilon)) \rightarrow 0, \quad n \rightarrow \infty,$$

co kończy dowód. □

W ogólnym przypadku, implikacje w Twierdzeniu 8.2 są jedynymi, które zachodzą, zob. Rozdział 8.3 poświęcony omówieniu wybranych kontrprzykładów. W pewnych specjalnych przypadkach mogą jednak zachodzić inne implikacje, co ilustrują następujące propozycje.

Propozycja 8.3 (Zbieżność według prawdopodobieństwa implikuje zbieżność w L^p dla ograniczonych zmiennych losowych). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ takim, że

$$|X_n| \leq K \quad (p.n.)$$

dla $n \in \mathbb{N}$ i pewnego $K \in \mathbb{R}$. Wtedy, dla X określonej na tej samej przestrzeni, zachodzi implikacja

$$(X_n \xrightarrow{\mathbb{P}} X) \implies (X_n \xrightarrow{L^p} X).$$

Dowód. Jeżeli $|X_n| < K$, $n \in \mathbb{N}$, oraz $X_n \xrightarrow{\mathbb{P}} X$, to dostajemy również $|X| < K$. Dla każdego $\epsilon > 0$ dostajemy więc

$$\begin{aligned} \mathbb{E}(|X_n - X|^p) &= \mathbb{E}(1_{\{|X_n - X| > \epsilon\}} |X_n - X|^p) + \mathbb{E}(1_{\{|X_n - X| \leq \epsilon\}} |X_n - X|^p) \\ &\leq \mathbb{E}(1_{\{|X_n - X| > \epsilon\}} (2K)^p) + \mathbb{E}(1_{\{|X_n - X| \leq \epsilon\}} \epsilon^p) \\ &= (2K)^p \cdot \mathbb{P}(|X_n - X| > \epsilon) + \epsilon^p \cdot \mathbb{P}(|X_n - X| \leq \epsilon) \longrightarrow \epsilon^p, \quad n \rightarrow \infty, \end{aligned}$$

co, wobec dowolności wyboru $\epsilon > 0$, kończy dowód. \square

Propozycja 8.4 (Zbieżność według rozkładu do stałej implikuje zbieżność według prawdopodobieństwa). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ oraz niech $c \in \mathbb{R}$. Wtedy

$$(X_n \xrightarrow{d} c) \implies (X_n \xrightarrow{\mathbb{P}} c).$$

Dowód. Ustalmy $\epsilon > 0$. Zauważając, że $c - \epsilon$ oraz $c + \epsilon$ jest punktem ciągłości dystrybuanty zmiennej losowej $X \equiv c$ dostajemy

$$\begin{aligned} \mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n < c - \epsilon) + \mathbb{P}(X_n > c + \epsilon) \\ &\leq \mathbb{P}(X_n \leq c - \epsilon) + 1 - \mathbb{P}(X_n \leq c + \epsilon) \\ &= F_{X_n}(c - \epsilon) + 1 - F_{X_n}(c + \epsilon) \rightarrow F_X(c - \epsilon) + 1 - F_X(c + \epsilon) = 0, \quad n \rightarrow \infty, \end{aligned}$$

co kończy dowód. \square

Okazuje się również, że z ciągu zbieżnego według prawdopodobieństwa można wybrać ciąg zbieżny prawie na pewno.

Twierdzenie 8.5 (Twierdzenie Riesz). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ oraz niech X będzie zmienną losową określoną na tej samej przestrzeni. Wtedy

$$(X_n \xrightarrow{\mathbb{P}} X) \implies (X_{n_k} \xrightarrow{p.n.} X),$$

dla pewnego podciągu $(n_k)_{k \in \mathbb{N}}$, $n_k \in \mathbb{N}$.

Dowód. Ponieważ $X_n \xrightarrow{\mathbb{P}} X$, $n \rightarrow \infty$, więc dla każdego $k \in \mathbb{N}$ oraz $\epsilon_k := 2^{-k}$ istnieje $n_k \in \mathbb{N}$ takie, że

$$\mathbb{P}(|X_n - X| \geq \epsilon_k) < \epsilon_k \quad \text{dla } n \geq n_k.$$

Założmy, że ciąg $(n_k)_{k \in \mathbb{N}}$ jest rosnący i rozważmy podciąg zmiennych losowych (X_{n_k}) . Zauważając, że

$$\sum_{k=1}^{\infty} \mathbb{P}(|X_{n_k} - X| \geq 2^{-k}) < \infty$$

oraz korzystając z lematu Borela-Cantelliego (Twierdzenie 2.18, 1.) wiemy, że zachodzi równość $\mathbb{P}(\limsup_{k \rightarrow \infty} \{|X_{n_k} - X| \geq 2^{-k}\}) = 0$, tzn. dla prawie każdego $\omega \in \Omega$ skończenie wiele razy (dla skończenie wielu elementów ciągu n_k) zachodzi nierówność $|X_{n_k}(\omega) - X(\omega)| \geq 2^{-k}$. Dla prawie każdego $\omega \in \Omega$ istnieje więc $k(\omega) \in \mathbb{N}$ takie, że dla $k \geq k(\omega)$ zachodzi $|X_{n_k}(\omega) - X(\omega)| < 2^{-k}$, co implikuje $X_{n_k}(\omega) \rightarrow X(\omega)$, $k \rightarrow \infty$. Ponieważ zbieżność ta jest prawdziwa dla prawie wszystkich $\omega \in \Omega$, więc dostajemy $X_{n_k} \xrightarrow{p.n.} X$, $k \rightarrow \infty$, co kończy dowód. \square

Ważnym wnioskiem z powyższego twierdzenia jest charakteryzacja zbieżności według prawdopodobieństwa w oparciu o zbieżność prawie na pewno dla dowolnego podciągu.

Propozycja 8.6 (Zależność między zbieżnością według prawdopodobieństwa a zbieżnością prawie na pewno). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ oraz niech X będzie zmienną losową określoną na tej samej przestrzeni. Wtedy $X_n \xrightarrow{\mathbb{P}} X$, $k \rightarrow \infty$, wtedy i tylko wtedy, gdy z każdego podciągu (X_{n_k}) można wybrać podciąg $(X_{n_{k_m}})$ taki, że $X_{n_{k_m}} \xrightarrow{p.n.} X$, $m \rightarrow \infty$.

Dowód. [\Rightarrow] Skoro $X_n \xrightarrow{\mathbb{P}} X$, $n \rightarrow \infty$, to zachodzi również $X_{n_k} \xrightarrow{\mathbb{P}} X$, $k \rightarrow \infty$ dla każdego podciągu (n_k) . Z Twierdzenia 8.5 można więc wybrać podciąg $(X_{n_{k_m}})$ zbieżny prawie na pewno.

[\Leftarrow] Przeprowadźmy dowód nie wprost. Jeżeli ciąg (X_n) nie jest zbieżny według prawdopodobieństwa, to istnieje $\epsilon > 0$ oraz rosnący podciąg n_k taki, że $\mathbb{P}(|X_{n_k} - X| > \epsilon) > \epsilon$ dla $k \in \mathbb{N}$. Łatwo zauważyć, że z ciągu (X_{n_k}) nie da się wybrać ciągu zbieżnego prawie na pewno podciągu, co prowadzi do sprzeczności. \square

Propozycja 8.7 (Zbieżność według prawdopodobieństwa, a obłożenie ciągu zmiennych losowych funkcją ciągłą). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na przestrzeni $(\Omega, \Sigma, \mathbb{P})$. Niech $f: \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją ciągłą na zbiorze otwartym $A \in \mathcal{B}(\mathbb{R})$ takim, że $\mathbb{P}(X \in A) = 1$, gdzie X jest pewną zmienną losową określoną na tej samej przestrzeni. Wtedy dostajemy

$$(X_n \xrightarrow{\mathbb{P}} X) \implies (f(X_n) \xrightarrow{\mathbb{P}} f(X)).$$

Dowód. Z Propozycji 8.6 wiemy, że wystarczy pokazać, że dla każdego (rosnącego) ciągu liczb naturalnych (n_k) możemy wybrać podciąg (n_{k_m}) dla którego zachodzi $f(X_{n_{k_m}}) \xrightarrow{p.n.} f(X)$, $m \rightarrow \infty$. Ustalmy więc ciąg liczb (n_k) . Z założeń wiemy, że zachodzi również $X_{n_k} \xrightarrow{\mathbb{P}} X$, $k \rightarrow \infty$, więc na mocy Propozycji 8.6 istnieje podciąg (n_{k_m}) taki, że $X_{n_{k_m}} \xrightarrow{p.n.} X$, $m \rightarrow \infty$. Ponieważ f jest funkcją ciągłą na A , więc dla każdego $\omega \in \{X \in A\}$ jeżeli $X_{n_{k_m}}(\omega) \rightarrow X(\omega)$, to $f(X_{n_{k_m}}(\omega)) \rightarrow f(X(\omega))$, co kończy dowód, gdyż $\mathbb{P}(X \in A) = 1$. \square

Na koniec przedstawmy zbiorcze twierdzenie mówiące o przenoszeniu się zbieżności przy sumowaniu ciągów zmiennych losowych.

Propozycja 8.8. Niech $(X_n)_{n \in \mathbb{N}}$ oraz $(Y_n)_{n \in \mathbb{N}}$ będą ciągami zmiennych losowych określonych na przestrzeni $(\Omega, \Sigma, \mathbb{P})$, na której określimy również zmienne X i Y . Wtedy

- 1) jeżeli $X_n \xrightarrow{p.n.} X$ oraz $X_n \xrightarrow{p.n.} X$, to $X_n + Y_n \xrightarrow{p.n.} X + Y$;
- 2) jeżeli $X_n \xrightarrow{L^p} X$ oraz $X_n \xrightarrow{L^p} X$, to $X_n + Y_n \xrightarrow{L^p} X + Y$;
- 3) Jeżeli $X_n \xrightarrow{\mathbb{P}} X$ oraz $X_n \xrightarrow{\mathbb{P}} X$, to $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$;
- 4) z $X_n \xrightarrow{d} X$ oraz $X_n \xrightarrow{d} X$ nie musi wynikać $X_n + Y_n \xrightarrow{d} X + Y$.

Dowód Propozycji 8.8 pozostawiamy jako ćwiczenie do domu. Aby udowodnić własność 2) można skorzystać np. z nierówności Minkowskiego.

8.3 Przykłady

W tym podrozdziale pokażemy szereg przykładów, które pokazują powiązanie między różnymi typami zbieżności. Pokażemy w szczególności, że w ogólnym przypadku żadna implikacja poza tymi pokazanymi w Twierdzeniu 8.2 nie zachodzi.

Przykład 8.9 ($(X_n \xrightarrow{d} X) \not\Rightarrow (X_n \xrightarrow{\mathbb{P}} X)$). Niech $X \sim B(1, \frac{1}{2})$ oraz niech (X_n) będzie dane przez $X_n := 1 - X$. Łatwo zauważyć, że $X_n \sim B(1, \frac{1}{2})$, co implikuje $X_n \xrightarrow{d} X$. Z drugiej strony łatwo zauważyć, że dla $\epsilon = \frac{1}{2}$ zachodzi

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(1 > \epsilon) = 1 \not\rightarrow 0, \quad n \rightarrow \infty.$$

Przykład 8.10 ($(X_n \xrightarrow{L^q} X) \not\Rightarrow (X_n \xrightarrow{L^p} X)$, dla $p > q$). Niech (Y_n) będzie ciągiem zmiennych losowych takim że $Y_n \sim B(1, n^{-1/2 \cdot (p+q)})$ oraz niech $X_n := n \cdot Y_n$. Dla $X \equiv 0$ dostajemy

$$\begin{aligned} \mathbb{E}(|X_n - X|^q) &= \mathbb{P}(X_n = n) \cdot |n - 0|^q + \mathbb{P}(X_n = 0) \cdot |0 - 0|^q = n^{1/2 \cdot (q-p)} \rightarrow 0, \quad n \rightarrow \infty, \\ \mathbb{E}(|X_n - X|^p) &= \mathbb{P}(X_n = n) \cdot |n - 0|^p + \mathbb{P}(X_n = 0) \cdot |0 - 0|^p = n^{1/2 \cdot (p-q)} \rightarrow \infty \neq 0, \quad n \rightarrow \infty. \end{aligned}$$

Przykład 8.11 ($(X_n \xrightarrow{\mathbb{P}} X) \not\Rightarrow (X_n \xrightarrow{L^1} X)$). Niech (Y_n) będzie ciągiem niezależnych zmiennych losowych takim że $Y_n \sim B(1, n^{-2})$ oraz niech $X_n := n^3 Y_n$. Niech $X \equiv 0$. Wtedy dla każdego $\epsilon > 0$ i dostatecznie dużego $n \in \mathbb{N}$ zachodzi $\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(|X_n| > \epsilon) = n^{-2}$, co implikuje $X_n \xrightarrow{\mathbb{P}} X$. Z drugiej strony

$$\mathbb{E}(|X_n - X|) = \mathbb{P}(X_n = n) \cdot |n - 0| + \mathbb{P}(X_n = 0) \cdot |0 - 0| = n^{-2} \cdot n^3 = n \rightarrow \infty \neq 0, \quad n \rightarrow \infty.$$

Przykład 8.12 ($(X_n \xrightarrow{\mathbb{P}} X) \not\Rightarrow (X_n \xrightarrow{p.n.} X)$). Niech (X_n) będzie ciągiem niezależnych zmiennych losowych takim że $X_n \sim B(1, n^{-1})$ oraz niech $X := 0$. Dla każdego $\epsilon > 0$ (i dostatecznie dużego n) dostajemy $\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(X_n > \epsilon) = \frac{1}{n} \rightarrow 0, n \rightarrow \infty$, co daje nam $X_n \xrightarrow{\mathbb{P}} X$. Niech teraz $B_m := \bigcup_{n=m}^{\infty} \{|X_n| > \frac{1}{2}\}$ oraz $B := \limsup_{m \rightarrow \infty} B_m$. Dla każdego $\omega \in B$ dla nieskończenie wielu elementów ciągu (X_n) zachodzi $X_n > \frac{1}{2}$, co implikuje $X_n(\omega) \not\rightarrow X(\omega)$ (p.n). Łatwo zauważyć, że rodzina zbiorów (B_m) jest zstępująca, co daje nam $\mathbb{P}(B) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n)$. Korzystając z

niezależności ciągu (X_n) dostajemy

$$\begin{aligned}
\mathbb{P}(B_m) &= 1 - \mathbb{P}\left(\bigcap_{n=m}^{\infty} \{|X_n| \leq \tfrac{1}{2}\}\right) = 1 - \lim_{k \rightarrow \infty} \mathbb{P}\left(\bigcap_{n=m}^k \{|X_n| \leq \tfrac{1}{2}\}\right) \\
&= 1 - \lim_{k \rightarrow \infty} \prod_{n=m}^k \mathbb{P}(\{|X_n| \leq \tfrac{1}{2}\}) = 1 - \lim_{k \rightarrow \infty} \prod_{n=m}^k \mathbb{P}(\{X_n = 0\}) \\
&= 1 - \lim_{k \rightarrow \infty} \prod_{n=m}^k \left(1 - \frac{1}{n}\right) = 1 - \lim_{k \rightarrow \infty} \prod_{n=m}^k \frac{n-1}{n} \\
&= 1 - \lim_{k \rightarrow \infty} \prod_{n=m}^k \frac{m-1}{k} = 1,
\end{aligned}$$

co daje nam $\mathbb{P}(B) = 1$ i w rezultacie $X_n \not\rightarrow X$ (p.n.), $n \rightarrow \infty$.

Przykład 8.13 ($(X_n \xrightarrow{L^1} X) \not\Rightarrow (X_n \xrightarrow{p.n.} X)$). Dla ciągu zmiennych losowych z Przykładu 8.12 wiemy, że $X_n \not\rightarrow X$ (p.n.), $n \rightarrow \infty$. Z drugiej strony dostajemy

$$\mathbb{E}(|X_n - X|) = \mathbb{E}(X_n) = \mathbb{P}(X_n = 1) = \frac{1}{n} \rightarrow 0.$$

Przykład 8.14 ($(X_n \xrightarrow{p.n.} X) \not\Rightarrow (X_n \xrightarrow{L^1} X)$). Niech $Y \sim U([0, 1])$, $X_n := n \cdot 1_{\{Y \leq 1/n\}}$ oraz $X := 0$. Dla każdego $\omega \in \Omega \setminus \{Y = 0\}$ istnieje $n' \in \mathbb{N}$ takie, że $Y > \frac{1}{n'}$, a więc dla $n \geq n'$ dostajemy $X_n(\omega) = 0$, co daje nam $X_n \rightarrow X$ (p.n.), $n \rightarrow \infty$. Z drugiej strony

$$\mathbb{E}(|X_n - X|) = \mathbb{E}(X_n) = \mathbb{P}(X_n = n) \cdot n = \frac{1}{n} \cdot n = 1 \rightarrow 1 \neq 0, \quad n \rightarrow \infty.$$

8.4 Własności i różne charakteryzacje zbieżności według rozkładu

W rozdziale tym przedstawimy różne charakteryzacje zbieżności według rozkładów i wytłumaczymy, jaki jest ich związek z tzw. *slabą zbieżnością*, którą często definiuje się dla miar probabilistycznych. Pokażemy też Twierdzenie Prochorowa.

Twierdzenie 8.15 (Twierdzenie reprezentacyjne Skorochoda). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na $(\Omega, \Sigma, \mathbb{P})$ oraz niech X będzie zmienną losową określoną na tej samej przestrzeni taką, że $X_n \xrightarrow{d} X$, gdy $n \rightarrow \infty$. Wtedy istnieje (standardowa) przestrzeń probabilistyczne $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}})$ oraz określone na niej zmienne losowe \tilde{X} i (\tilde{X}_n) takie, że:

- 1) $\tilde{X} \sim X$ oraz $\tilde{X}_n \sim X_n$ dla $n \in \mathbb{N}$;^a
- 2) $\tilde{X}_n \xrightarrow{p.n.} \tilde{X}$, gdy $n \rightarrow \infty$.

^aInnymi słowy zmienna \tilde{X} ma taki sam rozkład jak zmienna X , a zmienne \tilde{X}_n mają taki sam rozkład jak zmienne X_n dla $n \in \mathbb{N}$.

Dowód. Niech $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}}) := ((0, 1), \mathcal{B}((0, 1)), \mathcal{L}_1)$ będzie standardową przestrzenią probabilistyczną. Dla $\omega \in \tilde{\Omega}$ oraz $n \in \mathbb{N}$ definiujemy zmienne

$$\tilde{X}(\omega) := \inf\{t \in \mathbb{R} : \omega \leq F(t)\} \quad \text{oraz} \quad \tilde{X}_n(\omega) := \inf\{t \in \mathbb{R} : \omega \leq F_n(t)\}, \quad (8.3)$$

gdzie F jest dystrybuantą zmiennej X , a F_n jest dystrybuantą zmiennej X_n , $n \in \mathbb{N}$. Łatwo zauważyć, że \tilde{X} oraz \tilde{X}_n można traktować jako (uogólnione) odwrotności dystrybuant X oraz X_n , gdyż dla każdego $\omega \in \tilde{\Omega}$ oraz $t \in \mathbb{R}$ zachodzi

$$\omega \leq F(t) \iff \tilde{X}(\omega) \leq t \quad \text{oraz} \quad \omega \leq F_n(t) \iff \tilde{X}_n(\omega) \leq t. \quad (8.4)$$

Stąd dostajemy 1), gdyż $\tilde{\mathbb{P}}(\tilde{X} \leq t) = \mathcal{L}_1((0, F(t)]) = F(t)$ oraz $\tilde{\mathbb{P}}(\tilde{X}_n \leq t) = \mathcal{L}_1((0, F_n(t)]) = F_n(t)$ dla $n \in \mathbb{N}$ i $t \in \mathbb{R}$.

Pokażmy teraz, że \tilde{X} oraz (\tilde{X}_n) spełnia 2). Ustalmy $\omega \in \tilde{\Omega}$. Dla każdego $\epsilon > 0$ istnieje punkt $t \in \mathbb{R}$ taki, że dystrybuanta F jest ciągła w t oraz

$$\tilde{X}(\omega) - \epsilon < t < \tilde{X}(\omega) \quad (8.5)$$

Z (8.4) i (8.5) dostajemy w szczególności $F(t) < \omega$. Ponieważ $\tilde{X}_n \xrightarrow{d} \tilde{X}$, $n \rightarrow \infty$, to dla dostatecznie dużych $n \in \mathbb{N}$ dostajemy również $F_n(t) < \omega$, co daje nam

$$\tilde{X}(\omega) - \epsilon < t < \tilde{X}_n(\omega), \quad (8.6)$$

Stosując graniczne przejście $n \rightarrow \infty$, a następnie $\epsilon \rightarrow 0$, dla (8.6), dostajemy

$$\liminf_{n \rightarrow \infty} \tilde{X}_n(\omega) \geq \tilde{X}(\omega). \quad (8.7)$$

Ustalmy teraz $\omega' \in \Omega$ takie, że $\omega' > \omega$. Dla każdego $\epsilon > 0$ istnieje punkt $t \in \mathbb{R}$ taki, że dystrybuanta F jest ciągła w t oraz

$$\tilde{X}(\omega') < t < \tilde{X}(\omega') + \epsilon. \quad (8.8)$$

Z (8.4) i (8.8), dostajemy w szczególności $\omega < \omega' \leq F(t)$. Ponieważ $\tilde{X}_n \xrightarrow{d} \tilde{X}$, $n \rightarrow \infty$, to dla dostatecznie dużych $n \in \mathbb{N}$ dostajemy również $\omega < F_n(t)$, co daje nam

$$\tilde{X}_n(\omega) \leq t < \tilde{X}(\omega') + \epsilon. \quad (8.9)$$

Stosując graniczne przejście $n \rightarrow \infty$, a następnie $\epsilon \rightarrow 0$, dla (8.9), dostajemy, że dla każdego $\omega' \in \tilde{\Omega}$ takiego, że $\omega' > \omega$, zachodzi

$$\limsup_{n \rightarrow \infty} \tilde{X}_n(\omega) \leq \tilde{X}(\omega'). \quad (8.10)$$

Zauważając, że dla prawie wszystkich $\omega \in \tilde{\Omega}$ zachodzi $\tilde{X}(\omega') \rightarrow \tilde{X}(\omega)$, gdy $\omega' \rightarrow \omega$ (co wynika np. z faktu, że funkcja $\tilde{X} : \tilde{\Omega} \rightarrow \mathbb{R}$ jest niemalejąca więc ma co najwyżej przeliczalnie wiele punktów nieciągłości) oraz łącząc (8.7) z (8.10) dostajemy $\tilde{X}_n(\omega) \rightarrow \tilde{X}(\omega)$, $n \rightarrow \infty$ dla prawie wszystkich $\omega \in \tilde{\Omega}$, co daje nam $\tilde{X}_n \xrightarrow{p.n.} \tilde{X}$, $n \rightarrow \infty$, i kończy dowód. \square

Twierdzenie reprezentacyjne Skorochoda (Twierdzenie 8.15) pozwala nam na sformułowanie wniosku analogicznego do tego z Propozycji 8.7 dla zbieżności według rozkładów.

Propozycja 8.16 (Zbieżność według rozkładów, a obłożenie ciągu zmiennych losowych funkcją ciągłą). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na przestrzeni $(\Omega, \Sigma, \mathbb{P})$ oraz niech $f : \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją ciągłą. Wtedy dostajemy

$$(X_n \xrightarrow{d} X) \implies (f(X_n) \xrightarrow{d} f(X)).$$

Dowód. Niech \tilde{X} oraz $(\tilde{X}_n)_{n \in \mathbb{N}}$ będą zmiennymi losowymi określonymi na standardowej przestrzeni probabilistycznej $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}}) := ((0, 1), \mathcal{B}((0, 1)), \mathcal{L}_1)$ tak, jak w (8.3). Korzystając z ciągłości funkcji $f: \mathbb{R} \rightarrow \mathbb{R}$ dostajemy

$$\{\omega \in \tilde{\Omega}: f(\tilde{X}_n(\omega)) \rightarrow f(\tilde{X}(\omega)), n \rightarrow \infty\} \supseteq \{\omega \in \tilde{\Omega}: \tilde{X}_n(\omega) \rightarrow \tilde{X}(\omega), n \rightarrow \infty\},$$

co daje nam $f(\tilde{X}_n) \xrightarrow{p.n.} f(\tilde{X}), n \rightarrow \infty$. Na mocy Twierdzenia 8.2 dostajemy więc

$$f(\tilde{X}_n) \xrightarrow{d} f(\tilde{X}), \quad n \rightarrow \infty.$$

To już kończy dowód, gdyż $f(\tilde{X}) \sim f(X)$ oraz $f(\tilde{X}_n) \sim f(X_n)$ dla $n \in \mathbb{N}$. \square

Twierdzenie 8.15 pozwala nam też na wprowadzenie innych charakteryzacji zbieżności według rozkładów w oparciu o kryterium całkowe dla tzw. *funkcji testowych*.

Twierdzenie 8.17 (Inna charakteryzacja zbieżności według rozkładów). Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych określonych na przestrzeni $(\Omega, \Sigma, \mathbb{P})$ oraz niech X będzie zmienną na tej przestrzeni. Wtedy następujące warunki są równoważne

- 1) $X_n \xrightarrow{d} X, n \rightarrow \infty$;
- 2) $\mathbb{E}(g(X_n)) \rightarrow \mathbb{E}(g(X)), n \rightarrow \infty$, dla wszystkich ciągłych i ograniczonych funkcji $g: \mathbb{R} \rightarrow \mathbb{R}$;
- 3) $\mathbb{E}(g(X_n)) \rightarrow \mathbb{E}(g(X)), n \rightarrow \infty$, dla wszystkich funkcji $g: \mathbb{R} \rightarrow \mathbb{R}$ postaci $g(x) = f(x)1_{[a,b]}(x)$, gdzie $a, b \in \mathbb{R}$ są punktami ciągłości dystrybuanty F_X oraz $f: \mathbb{R} \rightarrow \mathbb{R}$ jest ciągła na $[a, b]$.

Dowód.

[1] \Rightarrow 2)] Niech $g: \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją ciągłą i ograniczoną. Na mocy Twierdzenia 8.15 istnieją zmienne losowe \tilde{X} oraz (\tilde{X}_n) takie, że $\tilde{X} \sim X, \tilde{X}_n \sim X_n, n \in \mathbb{N}$, oraz $\tilde{X}_n \xrightarrow{p.n.} \tilde{X}, n \rightarrow \infty$. Z Propozycji 8.16 dostajemy $g(\tilde{X}_n) \xrightarrow{p.n.} g(\tilde{X}), n \rightarrow \infty$. Zauważając, że rodzina zmiennych $(g(\tilde{X}_n))_{n \in \mathbb{N}}$ jest jednostajnie ograniczona (gdyż funkcja g jest ograniczona) oraz korzystając z twierdzenia Lebesgue'a o zbieżności zmajoryzowanej (zob. Propozycja 5.8) dostajemy

$$\lim_{n \rightarrow \infty} \mathbb{E}(g(\tilde{X}_n)) = \mathbb{E}\left(\lim_{n \rightarrow \infty} g(\tilde{X}_n)\right) = \mathbb{E}(g(\tilde{X})).$$

Z warunku $\tilde{X} \sim X$ oraz $\tilde{X}_n \sim X_n$ dostajemy $g(\tilde{X}) \sim g(X)$ oraz $g(\tilde{X}_n) \sim g(X_n)$, co daje nam z kolei $\mathbb{E}(g(\tilde{X})) = \mathbb{E}(g(X))$ oraz $\mathbb{E}(g(\tilde{X}_n)) = \mathbb{E}(g(X_n))$ i kończy dowód; wartość oczekiwana zależy tylko od rozkładu zmiennej.

[2] \Rightarrow 3)] Niech C oznacza zbiór punktów ciągłości dystrybuanty F_X . W dowodzie ograniczymy się do funkcji $g: \mathbb{R} \rightarrow \mathbb{R}$ postaci $g(x) = f(x)1_{(-\infty, b]}(x)$, gdzie $b \in C$ oraz $f: \mathbb{R} \rightarrow \mathbb{R}$ jest ciągła na $(-\infty, b]$; dowód w ogólnym przypadku jest analogiczny. Dla ustalonej funkcji g tej postaci oraz $\delta > 0$ wprowadzamy funkcje pomocnicze

$$g_\delta^1(x) := \begin{cases} \left(1 - \frac{|b-x|}{\delta}\right) g(b), & x \in (b, b + \delta) \\ g(x), & x \notin (b, b + \delta). \end{cases} \quad \text{oraz} \quad g_\delta^2(x) := \begin{cases} \left(1 - \frac{|b-x|}{\delta}\right) g(b), & x \in (b - \delta, b + \delta) \\ 0, & x \notin (b - \delta, b + \delta) \end{cases}$$

Łatwo zauważyć, że funkcje g_δ^1 oraz g_δ^2 są ciągłe i ograniczone. Przypominając, postać funkcji $g(x) = f(x)1_{(-\infty, b]}(x), x \in \mathbb{R}$, łatwo zauważyć, że

$$|\mathbb{E}(g(X_n) - g_\delta^1(X_n))| \leq |\mathbb{E}(g_\delta^2(X_n))| \quad \text{oraz} \quad |\mathbb{E}(g(X) - g_\delta^1(X))| \leq |\mathbb{E}(g_\delta^2(X))|,$$

dla $n \in \mathbb{N}$. Korzystając z założenia 2) i stosując przejście graniczne $n \rightarrow \infty$ dostajemy więc

$$|\mathbb{E}(g(X_n) - g(X))| \leq |\mathbb{E}(g_\delta^2(X_n))| + |\mathbb{E}(g_\delta^2(X))| + |\mathbb{E}(g_\delta^1(X_n) - g_\delta^1(X))| \rightarrow 2 |\mathbb{E}(g_\delta^2(X))|.$$

Zauważając, że $\mathbb{P}(X = b) = 0$ (gdyż b jest punktem ciągłości dystrybuanty F_X) oraz stosując przejście graniczne $\delta \rightarrow 0$ dostajemy

$$|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| \leq 2 |\mathbb{E}(g_\delta^2(X))| \leq |g(b)| \cdot \mathbb{P}(b - \delta < X < b + \delta) \rightarrow 0,$$

co kończy dowód tej części twierdzenie.

[3] \Rightarrow 1)] Rozważmy funkcję $f(x) = 1$, $x \in \mathbb{R}$, oraz ustalmy $b \in C$. Dla każdego $a, b', c \in C$, takiego, że $a < b < b' < c$, dostajemy

$$\begin{aligned} \mathbb{P}(X_n \leq b) &\geq \mathbb{P}(a \leq X_n \leq b) = \mathbb{E}(1_{[a,b]}(X_n)f(X_n)) \rightarrow \mathbb{E}(1_{[a,b]}(X)f(X)) = \mathbb{P}(a \leq X \leq b) \\ \mathbb{P}(X_n \geq b') &\geq \mathbb{P}(b' \leq X_n \leq c) = \mathbb{E}(1_{[b',c]}(X_n)f(X_n)) \rightarrow \mathbb{E}(1_{[b',c]}(X)f(X)) = \mathbb{P}(b' \leq X \leq c) \end{aligned}$$

Przechodząc kolejno z a do $-\infty$ oraz z c do ∞ (po wartościach ze zbioru C) dostajemy więc, że istnieje ciąg $\epsilon_n \searrow 0$ taki, że $\mathbb{P}(X_n \leq b) \geq \mathbb{P}(X \leq b) - \epsilon_n$ oraz $\mathbb{P}(X_n > b) \geq \mathbb{P}(X_n \geq b') \geq \mathbb{P}(X \geq b') - \epsilon_n$, co daje nam

$$\mathbb{P}(X \leq b) - \epsilon_n \leq \mathbb{P}(X_n \leq b) \leq \mathbb{P}(X < b') + \epsilon_n,$$

Stosując przejście graniczne $n \rightarrow \infty$ oraz $b' \rightarrow b$ (po wartościach z C) dostajemy więc zbieżność $\mathbb{P}(X_n \leq b) \rightarrow \mathbb{P}(X \leq b)$, dla każdego $b \in C$, co kończy dowód. \square

Uwaga 8.18 (Słaba zbieżność). W wielu podręcznikach zbieżność według rozkładów definiuje się (bezpośrednio dla rozkładów) poprzez warunek 2) z Twierdzenia 8.17, zob. np. Definicja 7 w Rozdziale 8.1 w [JS04]. Zbieżność według rozkładów nazywa się też często *słabą* zbieżnością, co ma swoje korzenie w analizie funkcjonalnej. Ciągłe i ograniczone funkcje testowe f wyznaczają funkcjonały na przestrzeni miar, a od ciągu wartości funkcjonału (dla każdego funkcjonału) wymaga się, aby był zbieżny. Więcej szczegółów na ten temat można znaleźć w [JS04].

Łatwo zauważyć, że gdy $X_n \xrightarrow{d} X$, $n \rightarrow \infty$, oraz X nie jest stałą, to istnieje nieskończenie wiele granic (według rozkładu) dla ciągu $(X_n)_{n \in \mathbb{N}}$. Istotnie, jeżeli zmienne X i \tilde{X} mają taki sam rozkład, to zachodzi $X_n \xrightarrow{d} \tilde{X}$, $n \rightarrow \infty$. Z warunku 2) z Twierdzenia 8.17, wynika, że granica ciągu $(X_n)_{n \in \mathbb{N}}$ jest jedyna w sensie rozkładu prawdopodobieństwa, co pokazuje następująca propozycja.

Propozycja 8.19 (O jednoznaczności granicy według rozkładu). Niech μ i ν będą rozkładami prawdopodobieństwa na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ takimi, że dla każdej ciągłej i ograniczonej funkcji $g: \mathbb{R} \rightarrow \mathbb{R}$ zachodzi

$$\int_{\mathbb{R}} g d\mu = \int_{\mathbb{R}} g d\nu. \quad (8.11)$$

Wtedy $\mu(A) = \nu(A)$ dla $A \in \mathcal{B}(\mathbb{R})$.

Dowód. Bez straty ogólności możemy założyć, że A jest (borelowskim) zbiorem postaci $(-\infty, t]$, $t \in \mathbb{R}$, gdyż zbiory te generują $\mathcal{B}(\mathbb{R})$.²³ Niech X oraz \tilde{X} będą zmiennymi losowymi takimi, że $X \sim \mu$ oraz $\tilde{X} \sim \nu$. Dla $n \in \mathbb{N}$ definiujemy $X_n := X$ oraz $\tilde{X}_n := \tilde{X}$. Łatwo zauważyć, że $X_n \xrightarrow{d} X$

²³Formalnie rzecz biorąc zbiory takie generują tzw. π -układ, więc z Twierdzenie o π i λ układach wystarczy sprawdzić ten warunek, zob. Rozdział 8.2 w [JS04].

oraz $\tilde{X}_n \xrightarrow{d} \tilde{X}$. Z warunku (8.11) wynika, że dla dowolnej funkcji ciągłej i ograniczonej $g: \mathbb{R} \rightarrow \mathbb{R}$ dostajemy

$$\mathbb{E}(g(X_n)) \rightarrow \mathbb{E}(g(X)) = \int_{\mathbb{R}} g d\mu = \int_{\mathbb{R}} g d\nu = \mathbb{E}(g(\tilde{X})), \quad n \in \mathbb{N},$$

co na mocy warunku 2) z Twierdzenia 8.17 daje nam $X_n \xrightarrow{d} \tilde{X}$. Wynika stąd, że dla wszystkich $t \in C$, gdzie C to zbiór punktów ciągłości dystrybuanty \tilde{X} , zachodzi równość

$$F_X(t) = \lim_{n \rightarrow \infty} F_{X_n}(t) = F_{\tilde{X}}(t). \quad (8.12)$$

Założmy teraz, że $t \notin C$. Ponieważ C' jest co najwyżej przeliczalny oraz zarówno $F_{\tilde{X}}$ jak i F_X są prawostronnie ciągłe, więc istnieje taki ciąg $\epsilon_n \searrow 0$, że $t \pm \epsilon_n \in C$, $n \in \mathbb{N}$, oraz

$$F_X(t) = \lim_{n \rightarrow \infty} F_X(t + \epsilon_n) = \lim_{n \rightarrow \infty} F_{\tilde{X}}(t + \epsilon_n) = F_{\tilde{X}}(t). \quad (8.13)$$

Łącząc (8.12) oraz (8.13) dostajemy, że dla każdego $t \in \mathbb{R}$ zachodzi

$$\mu((-\infty, t]) = F_X(t) = F_{\tilde{X}}(t) = \nu((-\infty, t]),$$

co kończy dowód. □

Na koniec tego wykładu przedstawmy Twierdzenie Prochorowa, które mówi nam, kiedy z dowolnego podciągu zmiennych losowych można wybrać podciąg zbieżny według rozkładu do jakiejś zmiennej losowej. Zanim sformułujemy twierdzenie, należy zdefiniować tzw. warunek Prochorowa.

Definicja 8.20 (Warunek Prochorowa). Mówimy, że ciąg zmiennych losowych $(X_n)_{n \in \mathbb{N}}$ spełnia **warunek Prochorowa**, gdy dla każdego $\epsilon > 0$ istnieje taki zbiór zwarty $K_\epsilon \in \mathcal{B}(\mathbb{R})$, że dla $n \in \mathbb{N}$ zachodzi $\mathbb{P}[X_n \in K] > 1 - \epsilon$.^a

^aW literaturze można spotkać też wiele innych nazw tej własności. Alternatywnie, ciąg $(X_n)_{n \in \mathbb{N}}$ – lub ciąg powiązanych miar probabilistycznych – nazywa się *ciasnym* lub *jędrnym* (ang. *tight*).

Okazuje się, że warunek Prochorowa jest warunkiem koniecznym i wystarczającym istnienia granicy (według rozkładu) po podciągu dla dowolnego podciągu zmiennych losowych.

Twierdzenie 8.21 (Twierdzenie Prochorowa). Ciąg zmiennych losowych $(X_n)_{n \in \mathbb{N}}$ spełnia warunek Prochorowa wtedy i tylko wtedy, gdy z każdego jego podciągu można wybrać podciąg zbieżny według rozkładu do jakiejś zmiennej losowej.

Dowód.

(\Rightarrow) Z ciągu $(X_n)_{n \in \mathbb{N}}$ wybieramy podciąg $(X_{n_k})_{k \in \mathbb{N}}$, $k \in \mathbb{N}$. Niech F_k oznacza dystrybuantę zmiennej X_{n_k} oraz niech t_1, t_2, \dots będzie ciągiem zawierającym wszystkie liczby wymierne. Ponieważ ciąg $(F_k(t_1))_{k \in \mathbb{N}}$ jest ciągiem ograniczonym (o wartościach w $[0, 1]$) więc z $(F_k)_{k \in \mathbb{N}}$ możemy wybrać podciąg który jest zbieżny w punkcie t_1 . Z podciągu tego możemy wybrać kolejny podciąg, który jest zbieżny w t_2 . Postępując tak dalej i biorąc m -ty wyraz z uzyskanego w ten sposób m -tego podciągu uzyskujemy podciąg $(F_{a_m})_{m \in \mathbb{N}}$ taki, że dla każdego $t \in \mathbb{Q}$ zachodzi $F_{a_m}(t) \rightarrow F_0(t)$, $m \rightarrow \infty$, gdzie $F_0(t): \mathbb{Q} \rightarrow [0, 1]$ jest pewną funkcją określoną na liczbach wymiernych. Zdefiniujmy teraz funkcję $F: \mathbb{R} \rightarrow [0, 1]$ określoną przez

$$F(t) := \{\inf F_0(w) : w > t\}.$$

Funkcja $F: \mathbb{R} \rightarrow [0, 1]$ spełnia następujące własności: (a) F jest funkcją niemalejącą; (b) F jest funkcją prawostronnie ciągłą. Własność (a) wynika wprost z definicji funkcji F oraz własności podciągu dystrybuant $(F_{a_m})_{m \in \mathbb{N}}$. Ponieważ zbiór liczb wymiernych jest gęsty w \mathbb{R} , to łatwo pokazać, że $F_{a_m}(t) \rightarrow F(t)$ dla wszystkich punktów ciągłości F , co daje nam warunek (b), zob. Lemat 2 w [JS04]. Aby F była dystrybuantą wystarczy więc pokazać, że $\lim_{t \rightarrow -\infty} F(t) = 0$ oraz $\lim_{t \rightarrow \infty} F(t) = 1$.²⁴ W tym celu wykorzystamy warunek Prochorowa. Na mocy warunku Prochorowa, dla każdego $\epsilon > 0$ istnieje takie $M_\epsilon > 0$, że $\pm M_\epsilon$ jest punktem ciągłości funkcji granicznej F oraz

$$F_{a_m}(M_\epsilon) - F_{a_m}(-M_\epsilon) \geq 1 - \epsilon.$$
²⁵

Dla $m \rightarrow \infty$ dostajemy więc $F(M_\epsilon) - F(-M_\epsilon) \geq 1 - \epsilon$. Stosując przejście graniczne $\epsilon \rightarrow 0$ i przypominając, że F jest niemalejąca dostajemy $F(-M_\epsilon) \rightarrow 0$ oraz $F(M_\epsilon) \rightarrow 1$, $\epsilon \rightarrow 0$, co kończy tą część dowodu.

(\Leftarrow) Załóżmy, że ciąg zmiennych losowych $(X_n)_{n \in \mathbb{N}}$ nie spełnia warunku Prochorowa. Wtedy istnieje takie $\epsilon > 0$ oraz podciąg $(n_k)_{k \in \mathbb{N}}$, że dla każdego $k \in \mathbb{N}$ zachodzi

$$\mathbb{P}(X_{n_k} \in [-k, k]) < 1 - \epsilon.$$

Pokażmy teraz, że z ciągu $(X_{n_k})_{k \in \mathbb{N}}$ nie można wybrać podciągu zmiennych losowych zbieżnych według rozkładu. Załóżmy, że istnieje zmienna losowa X taka, że $X_{n_k} \xrightarrow{d} X$, $k \rightarrow \infty$. Ponieważ X jest zmienną losową, więc istnieje takie $k_0 \in \mathbb{R}$, że $\mathbb{P}(X \in [-k_0, k_0]) > 1 - \epsilon$ oraz $\pm k_0$ są punktami ciągłości dystrybuanty F_X . Stąd dostajemy

$$\mathbb{P}(X_{n_k} \in (-k_0, k_0]) = F_{X_{n_k}}(k_0) - F_{X_{n_k}}(-k_0) \rightarrow F_X(k_0) - F_X(-k_0) = \mathbb{P}(X \in (-k_0, k_0]), \quad k \rightarrow \infty,$$

co prowadzi do sprzeczności, gdyż dla dostatecznie dużych $k \in \mathbb{R}$ zachodzi

$$1 - \epsilon > \mathbb{P}(X_{n_k} \in (-k_0, k_0]) \rightarrow \mathbb{P}(X \in (-k_0, k_0]) > 1 - \epsilon.$$

□

Uwaga 8.22 (Twierdzeniu Prochorowa dla rodziny zmiennych losowych oraz miar probabilistycznych). Warunek Prochorowa można również zdefiniować dla rodziny (niekoniecznie przeliczalnej) zmiennych losowych. Jeżeli zastąpimy ciąg zmiennych losowych przez powiązaną rodzinę, teza Twierdzenia 8.21 nie ulegnie zmianie. Warunek Prochorowa jest w istocie warunkiem nałożonym na miary probabilistyczne związane ze zmiennymi losowymi. Można pokazać, że teza Twierdzenia 8.21 zachodzi, gdy rodzinę zmiennych losowych zastąpimy przez rodzinę miar probabilistycznych, a warunek zbieżności według rozkładów przez tzw. słabą zbieżność miar, por. Uwaga 8.18.

Uwaga 8.23 (Zbieżność według rozkładów dla wektorów losowych). Większość rezultatów przedstawionych w tym podrozdziale jest prawdziwa, gdy zamiast zmiennych losowych będziemy rozważać wektory losowe. Dla powiązanych (wielowymiarowych) rozkładów prawdopodobieństwa można w łatwy sposób wprowadzić warunek Prochorowa, czy analizować zbieżność np. w oparciu o warunek 2) z Twierdzenia 8.17. Po więcej szczegółów odsyłamy do Rozdziału 8 w [JS04].

²⁴W ogólnym przypadku tak być nie musi. Biorąc na przykład ciąg $X_n = n$ możemy dostać funkcję $F(t) = 0$ dla $t \in \mathbb{R}$.

²⁵Rozważamy dostatecznie duży zbiór zwarty $K = [-M_\epsilon, M_\epsilon]$.

A Wybrane fakty z matematyki dyskretnej i kombinatoryki

W rozdziale tym przypomnimy kilka wybranych podstawowych faktów związanych z matematyki dyskretnej i kombinatoryki, które zostały użyte podczas dowodów i przykładów (w prawdopodobieństwie klasycznym).

Definicja A.1 (Podstawowe wzory kombinatoryczne). Niech Y będzie zbiorem n -elementowym. Wtedy:

1. Liczba k -wyrazowych wariacji z powtórzeniami zbioru Y to n^k . Odpowiada to liczbie funkcji (ciągów) $f: \{1, \dots, k\} \rightarrow Y$.
2. Liczba k -wyrazowych wariacji bez powtórzeń zbioru Y to $\frac{n!}{(n-k)!}$. Odpowiada to liczbie różnowartościowych funkcji (ciągów) $f: \{1, \dots, k\} \rightarrow Y$.
3. Liczba permutacji zbioru Y to $n!$. Odpowiada to liczbie bijekcji $f: \{1, \dots, n\} \rightarrow Y$.
4. Liczba k -elementowych kombinacji zbioru Y to $\binom{n}{k} := \frac{n!}{k!(n-k)!}$. Odpowiada to liczbie k -elementowych podzbiorów zbioru Y .

B Wybrane fakty z analizy oraz teorii miary i całki

Podstawowe fakty z teorii miary i całki przydatne w tym wykładzie można znaleźć w Dodatku C w [JS04]. Przedstawmy teraz kilka faktów z analizy (w uproszczonej formie), które będą przydatne w dowodach przedstawionych na tym kursie.

Propozycja B.1 (Lemat Toeplitza). Niech $(x_i)_{i \in \mathbb{N}}$ będzie ciągiem liczb rzeczywistych takim, że $\lim_{n \rightarrow \infty} x_n = x$. Wtedy $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_n = x$.

Dowód. Ustalmy $\epsilon > 0$. Wiemy, że istnieje $i_0 \in \mathbb{N}$ takie, że dla każdego $i \geq i_0$ zachodzi $|x_i - x| \leq \epsilon/2$. Ponadto istnieje $i_1 \in \mathbb{N}$, $i_1 > i_0$, takie, że dla każdego $i \geq i_1$ zachodzi $\frac{1}{i_1} \sum_{j=1}^{i_0} |x_j - x| < \epsilon/2$. Dla każdego i większego od i_0 i i_1 dostajemy więc

$$\left| \frac{1}{i} \sum_{j=1}^i x_j - x \right| \leq \frac{1}{i} \sum_{j=1}^i |x_j - x| \leq \frac{1}{i_1} \sum_{j=1}^{i_0} |x_j - x| + \frac{1}{i - i_0} \sum_{j=i_0+1}^i |x_j - x| \leq \epsilon,$$

co kończy dowód. □

Propozycja B.2 (Lemat Kroneckera). Niech $(x_i)_{i \in \mathbb{N}}$ będzie ciągiem liczb rzeczywistych. Jeżeli szeregi $\sum_{i=1}^{\infty} \frac{x_i}{i}$ jest zbieżny, to $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = 0$.

Dowód. Zdefiniujemy $s_0 = 0$ oraz $s_n := \sum_{i=1}^n \frac{x_i}{i}$. Wtedy, dla każdego $n \in \mathbb{N}$, dostajemy

$$\sum_{i=1}^n i \cdot \frac{x_i}{i} = \sum_{i=1}^n i \cdot (s_i - s_{i-1}) = n s_n - \sum_{i=1}^n s_{i-1},$$

Zauważając, że $\lim_{n \rightarrow \infty} s_n = s$, dla pewnego $s \in \mathbb{R}$, oraz $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n s_{i-1} = s$ (co wynika z Lematu Toeplitza), dostajemy

$$\frac{1}{n} \sum_{i=1}^n x_i = s_n - \frac{1}{n} \sum_{i=1}^n s_{i-1} \rightarrow 0, \quad n \rightarrow \infty.$$

□

C Notacja

Ω	Ustalony (z góry) zbiór. Elementy Ω nazywamy <i>zdarzeniami elementarnymi</i> .
Σ	σ -algebra określona na Ω ; zob. Definicja 1.2
\mathbb{P}	Miara probabilistyczna (prawdopodobieństwo) określone na (Ω, Σ)
ω	Zdarzenie elementarne, element zbioru Ω , $\omega \in \Omega$.
A	Zdarzenie (mieralne), $A \in \Sigma$. Używamy też innych dużych liter.
$\#A$	Moc zbioru. Używane zazwyczaj w odniesieniu do zbiorów skończonych.
A', A^c	Zdarzenie <i>przeciwne</i> do A , dla $A \in \Sigma$, tzn. $A' := \Omega \setminus A$.
\emptyset	Zbiór pusty, domyślenie określający zdarzenie puste (niemożliwe), $\emptyset \in \Sigma$.
$\mathcal{P}(\Omega)$	Zbiór potęgowy Ω , tj. zbiór wszystkich możliwych podzbiorów Ω .
$\mathcal{B}(X)$	Rodzina zbiorów Borelowskich na przestrzeni X
\mathcal{L}_n	n -wymiarowa miara Lebesgue'a
$\sigma(\mathcal{F})$	najmniejsza σ -algebra zawierająca w sobie rodzinę zbiorów \mathcal{F} (podzbiorów Ω)
$\sigma(X)$	najmniejsza σ -algebra generowana przez zmienną X
$\mathbb{1}_A$	funkcja charakterystyczna zbioru A
$X^{-1}(A)$	przeciwwobraz funkcji na zbiorze, tzn. $X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}$
μ_X	rozkład prawdopodobieństwa zmiennej losowej X
f_X	funkcja gęstości prawdopodobieństwa zmiennej X
F_X	dystrybuanta zmiennej X
$\mathbb{E}(X)$	wartość oczekiwana zmiennej losowej X
$D^2(X)$	wariancja zmiennej losowej X
$D(X)$	odchylenie standardowe zmiennej losowej X
$\text{Cov}(X, Y)$	kowariancja zmiennych losowych X oraz Y
$\text{Cor}(X, Y)$	korelacja zmiennych losowych X oraz Y ; również $\rho_{X,Y}$
Σ_X	macierz wariancji-kowariancji wektora losowego X
δ_c	rozkład jednopunktowy w punkcie $c \in \mathbb{R}$
$B(n, p)$	rozkład dwumianowy z parametrami $n \in \mathbb{N}$ i $p \in (0, 1)$
$\text{Pois}(\lambda)$	rozkład Poissona z parametrem $\lambda > 0$
$\text{Geom}(p)$	rozkład geometryczny z parametrem $p \in (0, 1)$
$U([a, b])$	rozkład jednostajny ciągły na odcinku $[a, b]$
$U(\{x_1, \dots\})$	rozkład jednostajny dyskretny na zbiorze x_1, \dots
$\text{Exp}(\lambda)$	rozkład wykładniczy z parametrem $\lambda > 0$
$\Gamma(\alpha, \beta)$	rozkład Gamma z parametrami $\alpha, \beta > 0$
$\mathcal{N}(\mu, \sigma)$	rozkład normalny z parametrami $\mu \in \mathbb{R}$ i $\sigma > 0$
$\phi(x)$	gęstość standardowego rozkładu normalnego w punkcji $x \in \mathbb{R}$
$\Phi(x)$	dystrybuanta standardowego rozkładu normalnego w punkcji $x \in \mathbb{R}$
$\chi^2(n)$	rozkład chi-kwadrat o $n \in \mathbb{N}$ stopniach swobody
t_ν	rozkład t -Studenta z parametrem $\nu > 0$
S_n	zazwyczaj suma zadanych zmiennych losowych X_1, X_2, \dots, X_n
$\mathcal{F}_{n,\infty}$	resztkowa σ -algebra $\sigma(X_n, X_{n+1}, \dots)$
\mathcal{F}_∞	ogonowa σ -algebra $\mathcal{F}_\infty := \bigcap_{n=1}^\infty \mathcal{F}_{n,\infty}$
Z_n	zazwyczaj standaryzowacja zmiennej S_n , tzn. $Z_n = \frac{S_n - \mathbb{E}(S_n)}{D(S_n)}$
$X_n \xrightarrow{p,n} X$	zbieżność prawie na pewno
$X_n \xrightarrow{\mathbb{P}} X$	zbieżność stochastyczna
$X_n \xrightarrow{L^p} X$	zbieżność według p -tego momentu
$X_n \xrightarrow{d} X$	zbieżność według rozkładu, słaba zbieżność

Literatura

- [ADD00] R. B. Ash and C. A. Doleans-Dade. *Probability and measure theory*. Academic Press, 2000.
- [Bil12] P. Billingsley. *Probability and Measure: Anniversary Edition*. John Wiley & Sons, 2012.
- [CB02] G. Casella and R. L. Berger. *Statistical inference*. Cengage Learning, 2nd edition, 2002.
- [CZ13] M. Capinski and T. J. Zastawniak. *Probability through problems*. Springer Science & Business Media, 2013.
- [GS01a] G. Grimmett and D. Stirzaker. *One Thousand Exercises in Probability*. Oxford University Press, USA, 2001.
- [GS01b] G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford university press, 3 edition, 2001.
- [Hay00] F. Hayashi. *Econometrics*. Princeton University Press, 2000.
- [HS65] E. Hewitt and K. Stromberg. *Real and abstract analysis: a modern treatment of the theory of functions of a real variable*. Springer-Verlag, 1965.
- [JS04] J. Jakubowski and R. Sztencel. *Wstęp do teorii prawdopodobieństwa*. Script, Warszawa, 2004.
- [JS17] J. Jakubowski and R. Sztencel. *Rachunek prawdopodobieństwa dla (prawie) każdego*. Script, Warszawa, 2017.
- [Nel07] R. B. Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.
- [Nol20] J. P. Nolan. *Univariate stable distributions*. Springer, 2020.
- [Omb19] J. Ombach. *Rachunek prawdopodobieństwa wspomagany komputerowo dla studentów matematyki stosowanej*. Wydawnictwo UJ, Kraków, 2019.
- [PS00] Yu. V. Prokhorov and V. Statulevicius. *Limit theorems of probability theory*, volume 6. Springer Science & Business Media, 2000.