

WYKŁAD Z RACHUNKU PRAWDOPODOBIENSTWA I

ADAM OSEKOWSKI

1. AKSJOMATYKA RACHUNKU PRAWDOPODOBIENSTWA

Przypuśćmy, że wykonujemy pewien eksperyment losowy. Powstaje natychmiast pytanie: w jaki sposób opisać go matematycznie?

Przede wszystkim, na pewno możemy mówić o jego potencjalnych „najdrobniejszych” wynikach, które będziemy nazywać *zdarzeniami elementarnymi*. Zbiór wszystkich zdarzeń elementarnych oznaczamy literą Ω , a do oznaczenia zdarzeń elementarnych będziemy zazwyczaj używać litery ω .

Przykłady:

1. Rzut monetą: możliwe dwa wyniki: $\Omega = \{O, R\}$.
2. Rzut kostką: możliwe sześć wyników: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Często nie interesuje nas konkretny wynik ω , ale to, czy należy on do wcześniej ustalonego podzbioru zbioru Ω . Takie podzbiory nazywamy *zdarzeniami* i oznaczamy literami A, B, C, \dots

Przykłady, c.d.:

3. Rzucamy dwa razy kostką, A - suma oczek wynosi 4. Wówczas

$$\Omega = \{(i, j) : 1 \leq i, j \leq 6\} \quad \text{i} \quad A = \{(1, 3), (2, 2), (3, 1)\}.$$

4. Rzucamy monetą aż do wypadnięcia orła, A - wykonano co najwyżej trzy rzuty. Wówczas

$$\Omega = \{(O), (R, O), (R, R, O), (R, R, R, O), \dots\} \quad \text{i} \quad A = \{(O), (R, O), (R, R, O)\}.$$

5. Obrót tarczy w ruletce, A - strzałka zatrzymuje się w drugiej ćwiartce. Wówczas $\Omega = [0, 2\pi)$ i $A = [\pi/2, \pi]$.

Szczególne zdarzenia, interpretacje działań/relacji na zdarzeniach:

Ω - zdarzenie pewne,

\emptyset - zdarzenie niemożliwe,

$A \cap B$ - zaszły oba zdarzenia A, B ,

$A \cap B = \emptyset$ - zdarzenia się wykluczają (są rozłączne),

$A \cup B$ - zaszło A lub B ,

A' - nie zaszło A (A' nazywamy zdarzeniem przeciwnym do A , bądź dopełnieniem zbioru A),

$A \setminus B = A \cap B'$ - zaszło A i nie zaszło B ,

$A \subseteq B$ - A pociąga za sobą B .

Przypuśćmy, że mamy określony zbiór Ω i chcemy wyróżnić rodzinę \mathcal{F} zdarzeń, które będziemy badać. Pierwszy naturalny pomysł to rozważyć 2^Ω - klasę wszystkich możliwych podzbiorów Ω . Wybór ten dobrze się sprawdza w sytuacji gdy Ω jest zbiorem co najwyżej przeliczalnym. Niestety, dla $|\Omega| > \aleph_0$ klasa 2^Ω jest zbyt duża - pojawiają się kłopoty z określeniem na niej prawdopodobieństwa -

co wymusza wybór pewnej właściwej jej podrodziny. Z drugiej strony, sensowna klasa \mathcal{F} powinna być zamknięta na branie sum, iloczynów i zdarzenia przeciwnego; zakładamy więc, że \mathcal{F} jest pewnym wyróżnionym σ -ciałem podzbiorów Ω . Przypomnijmy odpowiednią definicję.

Definicja 1.1. Rodzinę \mathcal{F} podzbiorów Ω nazywamy σ -ciałem, jeśli

- (i) $\emptyset \in \mathcal{F}$,
- (ii) $A \in \mathcal{F} \Rightarrow A' \in \mathcal{F}$,
- (iii) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Parę (Ω, \mathcal{F}) nazywamy *przestrzenią mierzalną*.

Przechodzimy teraz do określenia prawdopodobieństwa. Aby zyskać nieco intuicji dotyczącej tego obiektu i jego własności, wygodnie najpierw rozważyć tzw. częstość zdarzeń. Załóżmy, iż w pewnym doświadczeniu interesuje nas prawdopodobieństwo zajścia pewnego zdarzenia A . Powtórzmy to doświadczenie n razy i zdefiniujmy

$$\rho_n(A) = \frac{\text{liczba doświadczeń w których zaszło } A}{n}.$$

Jest to częstość względna zajścia zdarzenia A w serii n doświadczeń, i spodziewamy się, iż dla dużych n liczba $\rho_n(A)$ powinna być bliska szansie zajścia zdarzenia A w pojedynczym doświadczeniu. Jak łatwo sprawdzić, ρ_n przyjmuje wartości w przedziale $[0, 1]$ oraz posiada następujące własności:

(i) $\rho_n(\Omega) = 1$,

(ii) jeśli A_1, A_2, \dots są parami rozłączne, to $\rho_n\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \rho_n(A_k)$.

Prowadzi to do następującej definicji.

Definicja 1.2 (Aksjomatyka Kołmogorowa). Niech (Ω, \mathcal{F}) będzie ustaloną przestrzenią mierzalną. Funkcję $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ nazywamy *prawdopodobieństwem*, jeśli

(i) $\mathbb{P}(\Omega) = 1$,

(ii) dla dowolnych parami rozłącznych $A_1, A_2, \dots \in \mathcal{F}$ zachodzi

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

Trójkę $(\Omega, \mathcal{F}, \mathbb{P})$ nazywamy *przestrzenią probabilistyczną*.

Uwagi:

1. Prawdopodobieństwo jest więc miarą unormowaną na (Ω, \mathcal{F}) . Czasami będziemy mówić, że \mathbb{P} jest *miarą probabilistyczną*.

2. Należy pamiętać, iż przy modelowaniu konkretnego doświadczenia losowego wybór przestrzeni probabilistycznej zależy tylko od nas. W wielu sytuacjach z warunków doświadczenia wynikają pewne postulaty, które w mniej czy bardziej jednoznaczny sposób zadają trójkę $(\Omega, \mathcal{F}, \mathbb{P})$; czasami jednak tak nie jest (por. paradoks Bertranda poniżej).

Twierdzenie 1.1 (Podstawowe własności prawdopodobieństwa). *Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz $A, B, A_1, A_2, \dots \in \mathcal{F}$. Wówczas*

$$(i) \mathbb{P}(\emptyset) = 0.$$

$$(ii) \text{ Jeśli } A_1, A_2, \dots, A_n \text{ są parami rozłączne, to } \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i).$$

$$(iii) \mathbb{P}(A') = 1 - \mathbb{P}(A).$$

$$(iv) \text{ Jeśli } A \subseteq B, \text{ to } \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A) \text{ oraz } \mathbb{P}(A) \leq \mathbb{P}(B).$$

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

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

Własność (v) z powyższego twierdzenia można uogólnić na przypadek skończonej liczby zbiorów. Zachodzi następujący fakt.

Twierdzenie 1.2 (Wzór włączeń i wyłączeń). *Jeśli $A_1, A_2, \dots, A_n \in \mathcal{F}$, to*

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots \\ &\quad + (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

Dowody powyższych dwóch twierdzeń są bardzo proste i opierają się na wykorzystaniu aksjomatyki Kołmogorowa. Szczegóły pozostawiamy czytelnikowi.

Twierdzenie 1.3 (Twierdzenie o ciągłości). *Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz $(A_n)_{n=1}^{\infty}$ jest ciągiem zdarzeń.*

(i) *Jeśli ciąg ten jest wstępujący (tzn. $A_1 \subseteq A_2 \subseteq \dots$), to*

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

(ii) *Jeśli ciąg ten jest zstępujący (tzn. $A_1 \supseteq A_2 \supseteq \dots$), to*

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Dowód: (i) Rozważmy ciąg $(B_n)_{n \geq 1}$ zdarzeń, zadany przez

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus A_2, \quad \dots$$

Jak łatwo sprawdzić, zdarzenia B_1, B_2, \dots są parami rozłączne, $\bigcup_{n=1}^k B_n = A_k$ dla dowolnego $k \geq 1$ oraz $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. Zatem

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{k \rightarrow \infty} \sum_{n=1}^k \mathbb{P}(B_n) = \lim_{k \rightarrow \infty} \mathbb{P}\left(\bigcup_{n=1}^k B_n\right) = \lim_{k \rightarrow \infty} \mathbb{P}(A_k), \end{aligned}$$

gdzie w drugim przejściu korzystaliśmy z przeliczalnej addytywności miary \mathbb{P} , a w czwartym skorzystaliśmy z Twierdzenia 1 (ii).

(ii) Ciąg dopełnień $(A'_n)_{n \geq 1}$ jest wstępujący, a zatem, korzystając z (i) oraz z praw de Morgana, mamy

$$\begin{aligned} \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1 - \mathbb{P}\left(\left(\bigcap_{n=1}^{\infty} A_n\right)'\right) \\ &= 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} A'_n\right) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(A'_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n). \quad \square \end{aligned}$$

Przykłady:

1. (Schemat klasyczny, prawdopodobieństwo klasyczne). Załóżmy, że Ω jest zbiorem skończonym, $\mathcal{F} = 2^\Omega$ i wszystkie zdarzenia jednoelementowe są jednakowo prawdopodobne. Wówczas, jak łatwo sprawdzić, dla dowolnego $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

2. Załóżmy, że $\Omega = \{\omega_1, \omega_2, \dots\}$ jest zbiorem co najwyżej przeliczalnym oraz p_1, p_2, \dots - liczby nieujemne o sumie 1. Wówczas wybór $\mathcal{F} = 2^\Omega$ oraz $\mathbb{P}(\{\omega_i\}) = p_i$, $i = 1, 2, \dots$, jednoznacznie zadaje przestrzeń probabilistyczną $(\Omega, \mathcal{F}, \mathbb{P})$: dla każdego $A \in \mathcal{F}$ mamy

$$\mathbb{P}(A) = \sum_i 1_A(\omega_i) p_i,$$

gdzie 1_A to funkcja wskaźnikowa (charakterystyczna) bądź indykator zbioru A :

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

3. (Prawdopodobieństwo geometryczne). Załóżmy, że $\Omega \in \mathcal{B}(\mathbb{R}^d)$, tzn. Ω jest podzbiorem borelowskim \mathbb{R}^d , przy czym $0 < |\Omega| < \infty$ (tu $|\cdot|$ oznacza miarę Lebesgue'a w \mathbb{R}^d). Niech $\mathcal{F} = \mathcal{B}(\Omega)$ będzie σ -ciałem podzbiorów borelowskich Ω , a miara probabilistyczna \mathbb{P} będzie zadana przez

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Wówczas trójka $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną. Przestrzeń tę wykorzystujemy do modelowania doświadczenia polegającego na losowaniu punktu ze zbioru Ω .

4. (Paradoks Bertranda) Z okręgu o promieniu 1 wylosowano cięciwę AB . Jakie jest prawdopodobieństwo tego, że będzie ona dłuższa niż bok trójkąta równobocznego wpisanego w ten okrąg?

Przedstawimy trzy rozwiązania.

I) Ze względu na niezmienniczość okręgu na obroty, wylosowanie cięciwy AB możemy utożsamić z wylosowaniem miary kąta środkowego $\alpha = \angle AOB \in [0, 2\pi)$. Tak więc $\Omega = [0, 2\pi)$, $\mathcal{F} = \mathcal{B}(\Omega)$ oraz \mathbb{P} jest prawdopodobieństwem geometrycznym. Cięciwa spełnia warunki zadania wtedy i tylko wtedy, gdy $\alpha \in (2\pi/3, 4\pi/3)$, a zatem szukane prawdopodobieństwo wynosi

$$\mathbb{P}((2\pi/3, 4\pi/3)) = \frac{|(2\pi/3, 4\pi/3)|}{|[0, 2\pi)|} = \frac{1}{3}.$$

II) Wylosowanie cięciwy można utożsamić z wylosowaniem jej środka. Mamy więc $\Omega = B(0,1)$, $\mathcal{F} = \mathcal{B}(\Omega)$ i \mathbb{P} jest prawdopodobieństwem geometrycznym. Cięciwa będzie spełniała żądane warunki wtedy i tylko wtedy, gdy jej środek będzie leżał wewnątrz koła o promieniu $1/2$ współśrodkowego z danym okręgiem, zatem szukane prawdopodobieństwo wynosi

$$\mathbb{P}([0, 1/2)) = \frac{|B(0, 1/2)|}{|B(0, 1)|} = \frac{1}{4}.$$

III) Tak jak w poprzednim rozwiązaniu, bierzemy pod uwagę położenie środka cięciwy, lecz tym razem patrzemy na jego odległość od środka okręgu. Tak więc $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}(\Omega)$ i \mathbb{P} jest prawdopodobieństwem geometrycznym. Cięciwa będzie spełniała warunki zadania jeśli jej środek będzie odległy od środka okręgu o mniej niż $1/2$. Zatem szukane prawdopodobieństwo wynosi

$$\mathbb{P}([0, 1/2)) = \frac{|[0, 1/2)|}{|[0, 1]|} = \frac{1}{2}.$$

Tak więc widzimy, iż otrzymaliśmy trzy różne wyniki, stąd wyraz „paradoks” powyżej. Sprzeczności jednak tu nie ma - użyliśmy trzech różnych przestrzeni probabilistycznych do opisu tego samego doświadczenia losowego. Ogólnie rzecz ujmując, teoria prawdopodobieństwa nie rozstrzyga, jaki model doświadczenia należy wybrać; pozwala ona obliczać prawdopodobieństwa zdarzeń dopiero w sytuacji, gdy zadano już konkretną przestrzeń probabilistyczną.

2. ZADANIA

1. Na ile sposobów można ustawić w ciąg sześć jedynek, pięć dwójek oraz cztery trójki?
2. Wyznaczyć liczbę rozwiązań równania $x_1 + x_2 + x_3 + x_4 = 50$
 - a) w liczbach całkowitych nieujemnych x_1, x_2, x_3, x_4 ,
 - b) w liczbach całkowitych dodatnich x_1, x_2, x_3, x_4 .
3. Ile jest takich „szóstek” w Totolotku, że żadne dwie z wylosowanych liczb nie są kolejne?
4. Z talii 52 kart wylosowano 13 kart. Jakie jest prawdopodobieństwo tego, że istnieje kolor, w którym a) dokładnie siedem, b) dokładnie sześć kart jest tego samego koloru?
5. Klasa liczy 15 uczniów. Nauczyciel wybiera na każdej lekcji na chybił trafił jednego ucznia do odpowiedzi. Obliczyć prawdopodobieństwo tego, że w ciągu 16 lekcji każdy uczeń będzie przepytany.
6. W szafie jest n par butów. Wyjmujemy na chybił trafił k butów ($k \leq n$). Obliczyć prawdopodobieństwo tego, że
 - a) wśród wyjętych butów jest co najmniej jedna para,
 - b) wśród wyjętych butów jest dokładnie jedna para.
7. (Ω, \mathcal{F}, P) jest przestrzenią probabilistyczną, $A, B, C \in \mathcal{F}$.
 - a) Załóżmy, że $P(A \cup B) = 1/2$, $P(A \cap B) = 1/4$, $P(A \setminus B) = P(B \setminus A)$. Obliczyć $P(A)$ oraz $P(B \setminus A)$.
 - b) Załóżmy, że $A \cup B \cup C = \Omega$, $P(B) = 2P(A)$, $P(C) = 3P(A)$, $P(A \cap B) = P(A \cap C) = P(B \cap C)$. Wykazać, że $1/6 \leq P(A) \leq 1/4$.
 - c) Załóżmy, że $P(A) \geq 2/3$, $P(B) \geq 2/3$, $P(C) \geq 2/3$, $P(A \cap B \cap C) = 0$. Obliczyć $P(A)$.
8. Rozdano 52 karty czterem graczom, po 13 kart każdemu. Jakie jest prawdopodobieństwo, że każdy z graczy ma co najmniej jednego pika?
9. Jest N listów i N zaadresowanych kopert z różnymi adresami. Każdy list odpowiada dokładnie jednemu adresowi i na odwrót. Włożono listy do kopert na chybił trafił, po jednym liście do każdej koperty. Obliczyć prawdopodobieństwo, że żaden list nie trafił do właściwej koperty.
10. Udowodnić, że każde nieskończone σ -ciało jest nieprzeliczalne.
11. Kij o długości 1 złamano losowo w dwóch punktach. Jakie jest prawdopodobieństwo, że z powstałych trzech odcinków można zbudować trójkąt?
12. Na nieskończonej szachownicy o boku 1 rzucono monetę o średnicy $\frac{2}{3}$. Jakie jest prawdopodobieństwo, że a) moneta znajdzie się całkowicie we wnętrzu jednego z pól; b) przetnie się z dwoma bokami szachownicy?
13. Na płaszczyznę podzieloną na nieskończone pasy o szerokości d rzucono losowo igłę o długości ℓ ($\ell < d$). Wyznaczyć prawdopodobieństwo tego, że igła przetnie brzeg któregoś pasa.

3. PRAWDOPODOBIENSTWO WARUNKOWE I NIEZALEŻNOŚĆ ZDARZEŃ

3.1. **Prawdopodobieństwo warunkowe.** Zaczniemy od następującego przykładu.

Przykład: W urnie jest pięć białych kul ponumerowanych liczbami 1, 2, 3, 4, 5 oraz trzy kule czarne ponumerowane liczbami 1, 2, 3. Losujemy kulę i okazuje się, że jest biała. Jakie jest prawdopodobieństwo, że numer na niej jest parzysty?

Oczywista odpowiedź: $2/5$. Z drugiej strony, formalnie, mamy do czynienia ze schematem klasycznym na

$$\Omega = \{(1, b), (2, b), \dots, (5, b), (1, c), (2, c), (3, c)\}.$$

Określmy zdarzenia: A - wylosowano kulę o numerze parzystym, B - wylosowano kulę białą; zatem

$$A = \{(2, b), (4, b), (2, c)\}, \quad B = \{(1, b), (2, b), \dots, (5, b)\}$$

i mamy

$$\frac{2}{5} = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Sugeruje to następującą definicję.

Definicja 3.1. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz A, B są zdarzeniami takimi, że $\mathbb{P}(B) > 0$. Prawdopodobieństwem warunkowym (zajścia) zdarzenia A pod warunkiem (zajścia) zdarzenia B nazywamy liczbę

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

Uwaga: Jak łatwo sprawdzić, przy ustalonym zdarzeniu B takim, że $\mathbb{P}(B) > 0$, prawdopodobieństwo warunkowe $\mathbb{P}(\cdot|B)$ jest nową miarą probabilistyczną na (Ω, \mathcal{F}) .

Twierdzenie 3.1 (Prawdopodobieństwo iloczynu zdarzeń). *Założmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz A_1, A_2, \dots, A_n są zdarzeniami spełniającymi warunek $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) > 0$. Wówczas*

$$\begin{aligned} & \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= \mathbb{P}(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}) \mathbb{P}(A_{n-1} | A_1 \cap A_2 \cap \dots \cap A_{n-2}) \dots \mathbb{P}(A_2 | A_1) \mathbb{P}(A_1). \end{aligned}$$

Dowód. Wystarczy zastosować definicję prawdopodobieństwa warunkowego. \square

Przykład:

W urnie znajduje się $n-1$ białych kul oraz jedna czarna. Losujemy po jednej kuli aż do momentu, gdy wylosujemy czarną kulę. jakie jest prawdopodobieństwo tego, że wykonamy k losowań jeśli a) losujemy bez zwracania b) losujemy ze zwracaniem?

Oznaczmy białe kule przez b_1, b_2, \dots, b_{n-1} , a czarną kulę przez c . Mamy

$$\Omega = \{(c), (b_1, c), (b_2, c), \dots, (b_{n-1}, c), (b_1, b_1, c), \dots\},$$

$\mathcal{F} = 2^\Omega$, a prawdopodobieństwo zadane jest poprzez określenie mas poszczególnych zdarzeń jednoelementowych (por. Przykład 2 z poprzedniego wykładu).

Rozważmy zdarzenie A_i - i -ta kula jest biała, $i = 1, 2, \dots$. Korzystając z powyższego twierdzenia, mamy

$$\begin{aligned} & \mathbb{P}(A'_k \cap A_{k-1} \cap A_{k-2} \cap \dots \cap A_1) \\ &= \mathbb{P}(A'_k | A_{k-1} \cap \dots \cap A_1) \mathbb{P}(A_{k-1} | A_{k-2} \cap \dots \cap A_1) \dots \mathbb{P}(A_2 | A_1) \mathbb{P}(A_1). \end{aligned}$$

a) Z warunków zadania wynika, że

$$\mathbb{P}(A_i | A_{i-1} \cap \dots \cap A_1) = \frac{n-i}{n-i+1}, \quad \mathbb{P}(A'_k | A_{k-1} \cap \dots \cap A_1) = \frac{1}{n-k+1},$$

a zatem szukane prawdopodobieństwo wynosi

$$\frac{1}{n-k+1} \cdot \frac{n-k+1}{n-k+2} \cdot \frac{n-k+2}{n-k+3} \cdot \dots \cdot \frac{n-2}{n-1} \cdot \frac{n-1}{n} = \frac{1}{n}.$$

b) Tym razem mamy

$$\mathbb{P}(A_i | A_{i-1} \cap A_{i-2} \cap \dots \cap A_1) = \frac{n-1}{n},$$

a więc szukane prawdopodobieństwo jest równe

$$\left(1 - \frac{n-1}{n}\right) \cdot \frac{n-1}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{n-1}{n} = \frac{1}{n} \left(\frac{n-1}{n}\right)^{k-1}.$$

Zajmiemy się teraz analizą doświadczeń „wieloetapowych”, w których mamy do czynienia z losowaniem w kilku krokach, a przestrzeń probabilistyczna jest zadana poprzez specyfikację prawdopodobieństw warunkowych związanych z poszczególnymi krokami (por. przykład poniżej). Zaczniemy od definicji.

Definicja 3.2. Mówimy, że rodzina zdarzeń $(B_k)_{k=1}^n$ jest rozbiem (skończonym) zbioru Ω , jeśli $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$ oraz zdarzenia B_1, B_2, \dots, B_n są parami rozłączne. Analogicznie definiujemy rozbiem przeliczalny Ω .

Twierdzenie 3.2 (Wzór na prawdopodobieństwo całkowite). *Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz $A \in \mathcal{F}$ oraz $(B_k)_k$ jest rozbiem Ω (skończonym lub przeliczalnym), takim, że $\mathbb{P}(B_k) > 0$ dla wszystkich k . Wówczas*

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A|B_k)\mathbb{P}(B_k).$$

Dowód. Zdarzenia $A \cap B_1, A \cap B_2, \dots$, są parami rozłączne i dają w sumie A , a zatem

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A \cap B_k) = \sum_k \mathbb{P}(A|B_k)\mathbb{P}(B_k). \quad \square$$

Twierdzenie 3.3 (Wzór Bayesa). *Przy założeniach jak wyżej, jeśli $\mathbb{P}(A) > 0$, to dla każdego k ,*

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n)} \quad \left(= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)} \right).$$

Dowód. Wzór wynika natychmiast z definicji prawdopodobieństwa warunkowego oraz wzoru na prawdopodobieństwo całkowite. \square

Przykład:

Dane są urny I oraz II. W urnie I znajduje się b_1 kul białych oraz c_1 kul czarnych, zaś w urnie II - b_2 kul białych i c_2 kul czarnych. Losujemy urnę, a następnie kulę z tej urny.

a) Jakie jest prawdopodobieństwo, że kula jest biała?

b) Załóżmy, że wyciągnięta kula jest biała. Jakie jest prawdopodobieństwo tego, że losowano z I urny?

Mamy dwa etapy doświadczenia: losowanie urny oraz losowanie kuli z danej urny. Wprowadźmy zdarzenia A - wyciągnięto białą kulę, B_1 - wylosowano urnę I,

B_2 - wylosowano urnę II. Mamy $B_1 \cap B_2 = \emptyset$, $B_1 \cup B_2 = \Omega$, a więc rodzina $(B_i)_{i=1}^2$ jest rozbiem Ω . Z warunków zadania wynika, że

$$\mathbb{P}(B_1) = \mathbb{P}(B_2) = \frac{1}{2} > 0, \quad \mathbb{P}(A|B_1) = \frac{b_1}{b_1 + c_1}, \quad \mathbb{P}(A|B_2) = \frac{b_2}{b_2 + c_2}.$$

a) Korzystając ze wzoru na prawdopodobieństwo całkowite mamy

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) = \frac{1}{2} \left(\frac{b_1}{b_1 + c_1} + \frac{b_2}{b_2 + c_2} \right).$$

b) Na mocy wzoru Bayesa,

$$\mathbb{P}(B_1|A) = \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A)} = \frac{b_1/(b_1 + c_1)}{b_1/(b_1 + c_1) + b_2/(b_2 + c_2)}.$$

3.2. Niezależność zdarzeń. Zaczniemy od intuicji. Załóżmy, że A, B są zdarzeniami, przy czym $\mathbb{P}(B) > 0$. Wówczas zdarzenia A, B powinny być niezależne, jeśli informacja o tym, że zaszło zdarzenie B nie wpływa na prawdopodobieństwo zajścia zdarzenia A ; tzn. niezależność powinna być równoważna równości $\mathbb{P}(A|B) = \mathbb{P}(A)$, czyli $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Przyjmujemy to jako definicję.

Definicja 3.3. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną. Zdarzenia A, B są niezależne, jeśli

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

Uwaga: Jeśli $\mathbb{P}(A) = 0$, to dla dowolnego $B \in \mathcal{F}$ zdarzenia A oraz B są niezależne. Ta sama teza zachodzi gdy $\mathbb{P}(A) = 1$.

Zdefiniujemy teraz niezależność większej liczby zdarzeń. Zaczniemy od przypadku skończonego. Intuicyjnie, zdarzenia A_1, A_2, \dots, A_n są niezależne jeśli każdy podukład tych zdarzeń jest niezależny oraz zdarzenia $A_n, A_1 \cap A_2 \cap \dots \cap A_{n-1}$ są niezależne. Jak łatwo zauważyć, powyższe warunki wymuszają równość

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k})$$

dla dowolnego $k = 2, 3, \dots, n$ i dowolnego rosnącego ciągu $1 \leq i_1 < i_2 < \dots < i_k \leq n$. Przyjmujemy to jako definicję.

Definicja 3.4. Mówimy, że zdarzenia A_1, A_2, \dots, A_n są niezależne, jeśli dla wszystkich $2 \leq k \leq n$ oraz dowolnego ciągu $1 \leq i_1 < i_2 < \dots < i_k \leq n$ zachodzi równość

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}).$$

Definicja 3.5. Mówimy, że zdarzenia A_1, A_2, \dots, A_n są niezależne parami, jeśli dla dowolnych różnych $i, j \in \{1, 2, \dots, n\}$, zdarzenia A_i oraz A_j są niezależne.

Oczywiście niezależność „zespolowa” (czy też „łączna”) zdarzeń A_1, A_2, \dots, A_n pociąga za sobą ich niezależność parami. Implikacja w drugą stronę nie jest prawdziwa, co ilustruje następujący przykład.

Przykład:

Rzucamy dwa razy kostką. Niech A - za pierwszym razem wypadła parzysta liczba oczek, B - za drugim razem wypadła parzysta liczba oczek, C - suma oczek jest parzysta. Bezpośrednio wyliczamy, iż

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}, \quad \mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(C \cap A) = \frac{1}{4},$$

a więc zdarzenia A , B , C są niezależne parami. Nie są jednak niezależne zespolowo: mamy $A \cap B \subset C$, a więc $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = 1/4 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

W przypadku dowolnej (być może nieskończonej) liczby zdarzeń, niezależność definiujemy następująco.

Definicja 3.6. Załóżmy, że $\{A_i\}_{i \in I}$ jest pewną rodziną zdarzeń. Mówimy, iż zdarzenia te są niezależne, jeśli dla każdego n oraz parami różnych $i_1, i_2, \dots, i_n \in I$ zdarzenia $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ są niezależne.

Zdefiniujemy teraz pojęcie niezależności σ -ciał.

Definicja 3.7. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ są σ -ciałami zawartymi w \mathcal{F} . Mówimy, że σ -ciała te są niezależne, jeśli dla dowolnych $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$ zachodzi warunek

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n).$$

Twierdzenie 3.4. Przy założeniach powyższej definicji, σ -ciała $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ są niezależne wtedy i tylko wtedy, gdy dowolne $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$ są niezależne.

Dowód. \Leftarrow oczywiste.

\Rightarrow Mamy dowieść, że dla dowolnego $2 \leq k \leq n$ oraz dowolnego ciągu $1 \leq i_1 < i_2 < \dots < i_k \leq n$ zachodzi równość

$$(*) \quad \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}).$$

Rozważmy zdarzenia B_1, B_2, \dots, B_n dane przez

$$B_i = \begin{cases} A_i & \text{jeśli } i = i_\ell \text{ dla pewnego } \ell, \\ \Omega & \text{w przeciwnym przypadku.} \end{cases}$$

Wówczas $B_1 \in \mathcal{F}_1, B_2 \in \mathcal{F}_2, \dots, B_n \in \mathcal{F}_n$, a zatem

$$\mathbb{P}(B_1 \cap B_2 \cap \dots \cap B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2) \dots \mathbb{P}(B_n),$$

co jest równoważne (*). □

Przykłady:

1. Rzucamy dwa razy kostką. Wprowadźmy standardową przestrzeń probabilistyczną opisującą to doświadczenie (por. poprzedni wykład). Rozważmy σ -ciała

$$\mathcal{F}_1 = \{A \times \{1, 2, 3, 4, 5, 6\} : A \subseteq \{1, 2, 3, 4, 5, 6\}\},$$

$$\mathcal{F}_2 = \{\{1, 2, 3, 4, 5, 6\} \times B : B \subseteq \{1, 2, 3, 4, 5, 6\}\}.$$

Wówczas $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$ i $\mathcal{F}_1, \mathcal{F}_2$ są niezależne: istotnie, dla dowolnych zdarzeń $A \times \{1, 2, 3, 4, 5, 6\} \in \mathcal{F}_1, B \times \{1, 2, 3, 4, 5, 6\} \in \mathcal{F}_2$ mamy

$$\mathbb{P}(A \times \{1, 2, 3, 4, 5, 6\}) = \frac{|A| \cdot 6}{36} = \frac{|A|}{6}, \quad \mathbb{P}(\{1, 2, 3, 4, 5, 6\} \times B) = \frac{6 \cdot |B|}{36} = \frac{|B|}{6}$$

oraz

$$\mathbb{P}((A \times \{1, 2, 3, 4, 5, 6\}) \cap (\{1, 2, 3, 4, 5, 6\} \times B)) = \mathbb{P}(A \times B) = \frac{|A| \cdot |B|}{36}.$$

2. Załóżmy, że $\sigma(A_1), \sigma(A_2), \dots, \sigma(A_n)$ będą σ -ciałami generowanymi przez zdarzenia A_1, A_2, \dots, A_n , odpowiednio (przypomnijmy: $\sigma(A) = \{A, A', \emptyset, \Omega\}$). Wówczas jeśli A_1, A_2, \dots, A_n są niezależne, to $\sigma(A_1), \sigma(A_2), \dots, \sigma(A_n)$ też są

niezależne. Aby to wykazać, musimy sprawdzić, że dla dowolnych $B_i \in \mathcal{F}_i$, $i = 1, 2, \dots, n$, zachodzi

$$\mathbb{P}(B_1 \cap B_2 \cap \dots \cap B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2)\dots\mathbb{P}(B_n).$$

Jeśli co najmniej jedno ze zdarzeń B_i jest zbiorem pustym, to powyższa równość jest spełniona: obie strony są równe 0. Jeśli dla pewnego j mamy $B_j = \Omega$, to możemy to zdarzenie pominąć po obu stronach. Zatem, wystarczy dowiedzieć, że dla dowolnego ciągu $1 \leq i_1 < i_2 < \dots < i_k \leq n$ mamy

$$\mathbb{P}(B_{i_1} \cap B_{i_2} \cap \dots \cap B_{i_k}) = \mathbb{P}(B_{i_1})\mathbb{P}(B_{i_2})\dots\mathbb{P}(B_{i_k}),$$

gdzie dla każdego j , zdarzenie B_{i_j} jest równe A_{i_j} lub A'_{i_j} . Poprzez prostą indukcję, wystarczy wykazać, że

$$\mathbb{P}(A'_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \dots \cap A_{i_k}) = \mathbb{P}(A'_{i_1})\mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\dots\mathbb{P}(A_{i_k}),$$

co natychmiast wynika z niezależności zdarzeń A_1, A_2, \dots, A_n : istotnie,

$$\begin{aligned} & \mathbb{P}(A'_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \dots \cap A_{i_k}) \\ &= \mathbb{P}(A_{i_2} \cap A_{i_3} \cap \dots \cap A_{i_k}) - \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \dots \cap A_{i_k}) \\ &= \mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\dots\mathbb{P}(A_{i_k}) - \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\dots\mathbb{P}(A_{i_k}) \\ &= \mathbb{P}(A'_{i_1})\mathbb{P}(A_{i_2})\mathbb{P}(A_{i_3})\dots\mathbb{P}(A_{i_k}). \end{aligned}$$

Rozważmy teraz następujący problem. Załóżmy, że mamy N doświadczeń, przy czym i -te doświadczenie jest opisywane przez przestrzeń probabilistyczną $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$. W jaki sposób możemy zbudować przestrzeń $(\Omega, \mathcal{F}, \mathbb{P})$ dla doświadczenia polegającego na *niezależnym* przeprowadzeniu tych N doświadczeń?

Oczywiście, jako zbiór Ω bierzemy $\Omega_1 \times \Omega_2 \times \dots \times \Omega_N$. Aby określić \mathcal{F} , zwróćmy uwagę, iż σ -ciało \mathcal{F}_i jest reprezentowane, w kontekście powyższego zbioru Ω , przez klasę

$$\mathcal{F}'_i = \{\Omega_1 \times \Omega_2 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_N : A_i \in \mathcal{F}_i\}.$$

Stąd naturalny pomysł, by wziąć $\mathcal{F} = \sigma(\mathcal{F}'_1, \mathcal{F}'_2, \dots, \mathcal{F}'_N)$, σ -ciało generowane przez $\mathcal{F}'_1, \mathcal{F}'_2, \dots, \mathcal{F}'_N$. Innymi słowy, jako \mathcal{F} bierzemy σ -ciało produktowe $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots \otimes \mathcal{F}_N$. Przejdźmy do określenia miary probabilistycznej \mathbb{P} . Z powyższych postulatów, σ -ciała $\mathcal{F}'_1, \mathcal{F}'_2, \dots, \mathcal{F}'_N$ mają być niezależne, a zatem poszukujemy takiego prawdopodobieństwa \mathbb{P} , że dla dowolnych $A_i \in \mathcal{F}_i$, $i = 1, 2, \dots, N$,

$$\begin{aligned} & \mathbb{P}(A_1 \times A_2 \times \dots \times A_N) = \\ & \mathbb{P}((A_1 \times \Omega_2 \times \dots \times \Omega_N) \cap (\Omega_1 \times A_2 \times \dots \times \Omega_N) \cap \dots \cap (\Omega_1 \times \dots \times \Omega_{N-1} \times A_N)) \\ &= \prod_{i=1}^N \mathbb{P}(\Omega_1 \times \dots \times A_i \times \dots \times \Omega_N). \end{aligned}$$

Ponadto, chcemy by $\mathbb{P}(\Omega_1 \times \dots \times A_i \times \dots \times \Omega_N) = \mathbb{P}_i(A_i)$ dla każdego i . Podsumowując, poszukujemy takiego \mathbb{P} , by dla dowolnych zdarzeń A_1, A_2, \dots, A_N jak wyżej zachodziła równość

$$\mathbb{P}(A_1 \times A_2 \times \dots \times A_N) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)\dots\mathbb{P}_N(A_N).$$

Z teorii miary wiadomo, że istnieje dokładnie jedno takie prawdopodobieństwo \mathbb{P} na $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots \otimes \mathcal{F}_N$, i jest ono równe $\mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \dots \otimes \mathbb{P}_N$ - produktowi miar $\mathbb{P}_1, \mathbb{P}_2, \dots, \mathbb{P}_N$.

Analogiczne rozumowanie można przeprowadzić w przypadku gdy mamy do czynienia z nieskończoną liczbą doświadczeń modelowanych przez przestrzenie probabilistyczne $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$.

Przykład: Schemat Bernoulliego. Załóżmy, iż dla każdego $i = 1, 2, \dots, N$ mamy $\Omega_i = \{0, 1\}$, $\mathcal{F}^i = 2^{\Omega_i}$ oraz $\mathbb{P}_i(\{1\}) = p$, gdzie $p \in [0, 1]$ jest ustalonym parametrem. Widzimy, iż każda pojedyncza przestrzeń $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ modeluje doświadczenie w którym są dwa możliwe wyniki: 0 i 1, interpretowane jako *porażka* i *sukces* (podkreślmy: prawdopodobieństwo sukcesu jest równe p i nie zależy od numeru doświadczenia). Takie pojedyncze doświadczenie nazywamy *próbą Bernoulliego*. Na mocy powyższej konstrukcji, przestrzeń probabilistyczna

$$(\{0, 1\}^N, 2^\Omega, \mathbb{P}) = (\Omega_1 \times \Omega_2 \times \dots \times \Omega_N, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots \otimes \mathcal{F}_N, \mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \dots \otimes \mathbb{P}_N)$$

modeluje ciąg niezależnych N powtórzeń próby Bernoulliego. Ciąg ten nazywamy *schematem Bernoulliego*.

Zauważmy, iż dla dowolnego $\omega = (\omega_1, \omega_2, \dots, \omega_N) \in \Omega$ mamy $\mathbb{P}(\{\omega\}) = p^k(1-p)^{N-k}$, gdzie k jest liczbą jedynek w ciągu ω . Wynika stąd, iż jeśli określimy zdarzenie $A_k = \{\text{liczba sukcesów jest równa } k\}$, to

$$\mathbb{P}(A_k) = \sum_{\omega \in A_k} \mathbb{P}(\{\omega\}) = |A_k| p^k (1-p)^{N-k} = \binom{N}{k} p^k (1-p)^{N-k}.$$

Założmy, że A_1, A_2, \dots są pewnymi zdarzeniami; wówczas $\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ możemy interpretować jako „zaszło nieskończenie wiele spośród zdarzeń A_1, A_2, \dots ”. Okazuje się, że przy pewnych założeniach, zdarzenie to ma prawdopodobieństwo 0 lub 1. Ściślej, zachodzi następujący fakt.

Lemat 3.1 (Borela-Cantelli). *Założmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz $A_1, A_2, \dots \in \mathcal{F}$.*

(i) *Jeśli $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, to*

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0$$

(a *zatem z prawdopodobieństwem 1 zachodzi skończenie wiele spośród A_i*).

(ii) *Jeśli A_1, A_2, \dots są niezależne i $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, to*

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 1.$$

Dowód. (i) Mamy

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) \leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \xrightarrow{n \rightarrow \infty} 0.$$

(ii) Udowodnimy, że zdarzenie przeciwne $\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A'_m$ ma prawdopodobieństwo 0. Wystarczy wykazać, iż $\mathbb{P}(\bigcap_{m=n}^{\infty} A'_m) = 0$ dla wszystkich n (wówczas rozważane zdarzenie przeciwne będzie przeliczalną sumą zbiorów miary 0, a zatem także będzie miało miarę 0). Korzystając z twierdzenia o ciągłości, mamy iż dla dowolnego n ,

$$\mathbb{P}\left(\bigcap_{m=n}^{\infty} A'_m\right) = \mathbb{P}\left(\bigcap_{k=n}^{\infty} \bigcap_{m=n}^k A'_m\right) = \lim_{k \rightarrow \infty} \mathbb{P}\left(\bigcap_{m=n}^k A'_m\right),$$

co na mocy niezależności zdarzeń A_1, A_2, \dots jest równe

$$\lim_{k \rightarrow \infty} \prod_{m=n}^k \mathbb{P}(A'_m) = \lim_{k \rightarrow \infty} \prod_{m=n}^k (1 - \mathbb{P}(A_m)) \leq \limsup_{k \rightarrow \infty} e^{-\sum_{m=n}^k \mathbb{P}(A_m)} = 0. \quad \square$$

Na zakończenie zaprezentujemy pewien przydatny fakt, tzw. *lemat o $\pi - \lambda$ układach*. Aby podać pewną motywację, załóżmy, że μ, ν są miarami probabilistycznymi na pewnej przestrzeni mierzalnej (Ω, \mathcal{F}) . Przypuśćmy, iż chcemy wykazać, iż te miary są sobie równe: mamy więc sprawdzić, czy dla dowolnego $A \in \mathcal{F}$ zachodzi równość

$$\mu(A) = \nu(A).$$

Powstaje bardzo naturalne pytanie: czy wystarczy zweryfikować powyższą tożsamość dla pewnej szczególnej klasy zdarzeń A , np. dla generatorów σ -ciała \mathcal{F} ? Okazuje się że zbiór generatorów nie jest na ogół dobrym wyborem: mianowicie trzeba założyć, że klasa ta jest dodatkowo π -układem.

Definicja 3.8. Załóżmy, że \mathcal{K} jest niepustą klasą podzbiorów Ω . Mówimy, że \mathcal{K} jest π -układem, jeśli klasa ta jest zamknięta ze względu na branie skończonych iloczynów: z tego, że $A, B \in \mathcal{K}$ wynika, że $A \cap B \in \mathcal{K}$.

Definicja 3.9. Załóżmy, że \mathcal{L} jest pewną klasą podzbiorów Ω . Mówimy, że \mathcal{L} jest λ -układem, jeśli są spełnione następujące warunki:

- (i) $\Omega \in \mathcal{L}$,
- (ii) jeśli $A, B \in \mathcal{L}$ i $A \subseteq B$, to $B \setminus A \in \mathcal{L}$,
- (iii) jeśli A_1, A_2, \dots jest wstępującym ciągiem elementów \mathcal{L} , to $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$.

Lemat 3.2 (o $\pi - \lambda$ układach). *Jeśli \mathcal{L} jest λ -układem zawierającym π -układ \mathcal{K} , to \mathcal{L} zawiera także σ -ciało generowane przez \mathcal{K} .*

Dowód. Rozumowanie podzielimy na trzy części.

- 1) Zauważmy, że jeśli \mathcal{L} jest λ -układem oraz $A, B \in \mathcal{L}$ spełniają $A \cap B = \emptyset$, to

$$A \cup B = (A' \setminus B)' \in \mathcal{L},$$

korzystając z (i) i (ii).

- 2) Jeśli \mathcal{L} jest jednocześnie π -układem oraz λ -układem, to jest σ -ciałem. Aby to wykazać, zauważmy, iż jeśli $A, B \in \mathcal{L}$, to

$$A \cup B = A \cup (B \setminus (A \cap B)) \in \mathcal{L},$$

na mocy 1) oraz warunków definiujących π -układ i λ -układ. Wobec tego, przez prostą indukcję, \mathcal{L} będzie zamknięte ze względu na branie skończonych sum, a zatem jeśli A_1, A_2, \dots jest *dowolnym* ciągiem elementów z \mathcal{L} , to

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \left(\bigcup_{k=1}^n A_k \right) \in \mathcal{L}.$$

W ostatnim kroku skorzystaliśmy z tego, że $A_1, A_1 \cup A_2, A_1 \cup A_2 \cup A_3, \dots$ jest wstępującym ciągiem elementów \mathcal{L} .

- 3) Niech Λ będzie klasą wszystkich λ -układów zawierających \mathcal{K} i połóżmy $\mathcal{L}_0 = \bigcap_{\mathbf{L} \in \Lambda} \mathbf{L}$. Wówczas $\mathcal{L}_0 \in \Lambda$ oraz $\mathcal{K} \subseteq \mathcal{L}_0 \subseteq \mathcal{L}$. Wystarczy więc udowodnić, że \mathcal{L}_0 jest σ -ciałem: na mocy 2), wystarczy wykazać, że \mathcal{L}_0 jest π -układem. Weźmy dowolne $A \in \mathcal{K}$ i rozważmy klasę

$$\mathcal{K}_1 = \{B : A \cap B \in \mathcal{L}_0\}.$$

Oczywiście $\mathcal{K}_1 \supseteq \mathcal{K}$, gdyż \mathcal{K} jest π -układem. Ponadto \mathcal{K}_1 jest λ -układem:

- (i) $\Omega \in \mathcal{K}_1$, bo $A \cap \Omega = A \in \mathcal{K} \subseteq \mathcal{L}_0$;
- (ii) jeśli $B_1, B_2 \in \mathcal{K}_1$, $B_1 \subseteq B_2$, to

$$A \cap (B_2 \setminus B_1) = (A \cap B_2) \setminus (A \cap B_1) \in \mathcal{L}_0,$$

a więc $B_2 \setminus B_1 \in \mathcal{K}_1$;

- (iii) $B_1 \subseteq B_2 \subseteq \dots \in \mathcal{K}_1$, to

$$A \cap \left(\bigcup_{n=1}^{\infty} B_n \right) = \bigcup_{n=1}^{\infty} (A \cap B_n) \in \mathcal{L}_0,$$

skąd wynika, iż $\bigcup_{n=1}^{\infty} B_n \in \mathcal{K}_1$.

Zatem \mathcal{K}_1 zawiera \mathcal{L}_0 , gdyż \mathcal{L}_0 jest najmniejszym λ -układem zawierającym \mathcal{K} . Wykazaliśmy więc, że dla dowolnego $A \in \mathcal{K}$ oraz dowolnego $B \in \mathcal{L}_0$, $A \cap B \in \mathcal{L}_0$. Następnie powtarzamy rozumowanie: ustalamy $B \in \mathcal{L}_0$ i definiujemy $\mathcal{K}_2 = \{A : A \cap B \in \mathcal{L}_0\}$. Mamy $\mathcal{K}_2 \supseteq \mathcal{K}$ oraz \mathcal{K}_2 jest λ -układem, skąd wynika, iż $\mathcal{K}_2 \supseteq \mathcal{L}_0$, a więc dla dowolnych $A, B \in \mathcal{L}_0$ zachodzi $A \cap B \in \mathcal{L}_0$. Dowód jest zakończony. \square

Jako zastosowanie, udowodnimy następujący fakt.

Twierdzenie 3.5. *Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną. Przy-
puśćmy, iż rodzinę $\{\mathcal{F}_\gamma\}_{\gamma \in \Gamma}$ niezależnych σ -ciał podzielono na n podrodzin $\{\mathcal{F}_{\gamma^i}\}_{\gamma^i \in \Gamma^i}$,
 $i = 1, 2, \dots, n$. Wówczas σ -ciała*

$$\sigma(\{\mathcal{F}_{\gamma^1}\}_{\gamma^1 \in \Gamma^1}), \sigma(\{\mathcal{F}_{\gamma^2}\}_{\gamma^2 \in \Gamma^2}), \dots, \sigma(\{\mathcal{F}_{\gamma^n}\}_{\gamma^n \in \Gamma^n}),$$

generowane przez poszczególne podrodziny też są niezależne.

Dowód. Przeprowadzimy dowód dla $n = 2$; dla większych n rozumowanie jest analogiczne. Mamy więc dwie rodziny $\{\mathcal{F}_\gamma\}_{\gamma \in \Gamma}$, $\{\mathcal{G}_\delta\}_{\delta \in \Delta}$ niezależnych pod- σ -ciał \mathcal{F} i musimy wykazać, że dla dowolnych $A \in \sigma(\{\mathcal{F}_\gamma\}_{\gamma \in \Gamma})$, $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta \in \Delta})$ zachodzi

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

Na mocy niezależności σ -ciał, wzór (*) zachodzi dla zbiorów postaci

$$A = A_{\gamma^1} \cap A_{\gamma^2} \cap \dots \cap A_{\gamma^k}, \quad A_{\gamma^i} \in \mathcal{F}_{\gamma^i}, \quad i = 1, 2, \dots, k,$$

$$B = B_{\delta^1} \cap B_{\delta^2} \cap \dots \cap B_{\delta^\ell}, \quad B_{\delta^j} \in \mathcal{G}_{\delta^j}, \quad j = 1, 2, \dots, \ell.$$

Ustalmy A jak wyżej i rozważmy klasę $\mathcal{K} = \{B : \text{zachodzi } (*)\}$. Mamy więc $\mathcal{K} \supseteq \{B_{\delta^1} \cap B_{\delta^2} \cap \dots \cap B_{\delta^\ell} : B_{\delta^j} \in \mathcal{G}_{\delta^j}\}$, i ta ostatnia klasa jest, rzecz jasna, π -układem generującym $\sigma(\{\mathcal{G}_\delta\}_{\delta \in \Delta})$. Ponadto \mathcal{K} jest λ -układem:

- (i) $\Omega \in \mathcal{K}$, bo $\mathbb{P}(A \cap \Omega) = \mathbb{P}(A)\mathbb{P}(\Omega)$.
- (ii) jeśli $B_1, B_2 \in \mathcal{K}$ i $B_1 \subseteq B_2$, to $B_2 \setminus B_1 \in \mathcal{K}$: istotnie,

$$\begin{aligned} \mathbb{P}(A \cap (B_2 \setminus B_1)) &= \mathbb{P}((A \cap B_2) \setminus (A \cap B_1)) \\ &= \mathbb{P}(A \cap B_2) - \mathbb{P}(A \cap B_1) \\ &= \mathbb{P}(A)(\mathbb{P}(B_2) - \mathbb{P}(B_1)) = \mathbb{P}(A)\mathbb{P}(B_2 \setminus B_1). \end{aligned}$$

(iii) Jeśli $B_1 \subseteq B_2 \subseteq \dots \in \mathcal{K}$, to $\bigcup_{n=1}^{\infty} B_n \in \mathcal{K}$, gdyż z twierdzenia o ciągłości,

$$\begin{aligned} \mathbb{P} \left(A \cap \left(\bigcup_{n=1}^{\infty} B_n \right) \right) &= \mathbb{P} \left(\bigcup_{n=1}^{\infty} (A \cap B_n) \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(A \cap B_n) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(A) \mathbb{P}(B_n) = \mathbb{P}(A) \mathbb{P} \left(\bigcup_{n=1}^{\infty} B_n \right). \end{aligned}$$

Zatem na mocy lematu o $\pi - \lambda$ układach, \mathcal{K} zawiera $\sigma(\{\mathcal{G}_\delta\}_{\delta \in \Delta})$, czyli (*) zachodzi dla dowolnego zbioru A postaci $A_{\gamma^1} \cap A_{\gamma^2} \cap \dots \cap A_{\gamma^k}$ oraz $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta \in \Delta})$. Następnie, powtarzamy rozumowanie: ustalamy $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta \in \Delta})$ i definiujemy $\mathcal{L} = \{A : \text{zachodzi } (*)\}$. Klasa \mathcal{L} zawiera wszystkie zbiory postaci $A_{\gamma^1} \cap A_{\gamma^2} \cap \dots \cap A_{\gamma^k}$, które tworzą π -układ generujący $\sigma(\{\mathcal{F}_\gamma\}_{\gamma \in \Gamma})$. Tak jak wyżej, sprawdzamy, że \mathcal{L} jest λ układem, a zatem z lematu o $\pi - \lambda$ układach, $\mathcal{L} \supseteq \sigma(\{\mathcal{F}_\gamma\}_{\gamma \in \Gamma})$. Wobec tego (*) zachodzi dla wszystkich $A \in \sigma(\{\mathcal{F}_\gamma\}_{\gamma \in \Gamma})$ oraz $B \in \sigma(\{\mathcal{G}_\delta\}_{\delta \in \Delta})$, skąd dostajemy żadaną niezależność σ -ciał. \square

4. ZADANIA

1. Grupa n osób ($n \geq 3$), wśród których są osoby X , Y i Z , ustawia się losowo w kolejce. Jakie jest prawdopodobieństwo tego, że

- a) X stoi bezpośrednio przed Y , jeśli Y stoi bezpośrednio przed Z ?
- b) X stoi przed Y , jeśli Y stoi przed Z ?

2. Z talii 52 kart losujemy 5 kart bez zwracania. Obliczyć prawdopodobieństwo tego, że mamy dokładnie 3 asy, jeżeli wiadomo, że

- a) mamy co najmniej jednego asa;
- b) mamy asa czarnego koloru;
- c) mamy asa pik;
- d) pierwszą wylosowaną kartą jest as;
- e) pierwszą wylosowaną kartą jest czarny as;
- f) pierwszą wylosowaną kartą jest as pik.

3. W urnie znajdują się trzy białe i cztery czarne kule. Losujemy kulę, wyrzucamy bez oglądania, a następnie losujemy kolejną kulę z urny.

- a) Jakie jest prawdopodobieństwo, że druga kula jest biała?
- b) Załóżmy, że za drugim razem wyciągnięto białą kulę. Jakie jest prawdopodobieństwo, że za pierwszym razem wylosowano czarną kulę?

4. W populacji jest 15% dyslektyków. Jeśli w teście diagnostycznym uczeń popełni 6 lub więcej błędów, to zostaje uznany za dyslektyka. Każdy dyslektyk na pewno popełni co najmniej 6 błędów w takim teście, ale również nie-dyslektyk może popełnić więcej niż 5 błędów – dzieje się tak z prawdopodobieństwem 0,1. Jasio popełnił w teście 6 błędów. Jakie jest prawdopodobieństwo, że jest dyslektykiem? Jakie jest prawdopodobieństwo tego, że w kolejnym teście też popełni co najmniej 6 błędów?

5. W pewnej fabryce telewizorów każdy z aparatów może być wadliwy z prawdopodobieństwem p . W fabryce są trzy stanowiska kontroli i wyprodukowany telewizor trafia na każde ze stanowisk z jednakowym prawdopodobieństwem. i -te stanowisko wykrywa wadliwy telewizor z prawdopodobieństwem p_i ($i = 1, 2, 3$). Telewizory nie odrzucone w fabryce trafiają do hurtowni i tam poddawane są dodatkowej kontroli, która wykrywa wadliwy telewizor z prawdopodobieństwem p_0 .

a) Obliczyć prawdopodobieństwo tego, że dany nowowyprodukowany telewizor znajdzie się w sprzedaży (tzn. przejdzie przez obie kontrole).

b) Przypuśćmy, że telewizor jest już w sklepie. Jakie jest prawdopodobieństwo, że jest on wadliwy?

6. Rzucamy dwa razy kostką. Rozważmy zdarzenia: A – za pierwszym razem wypadła liczba oczek podzielna przez 3; B – suma wyrzuconych oczek jest parzysta; C – za każdym razem uzyskaliśmy tę samą liczbę oczek. Czy zdarzenia A , B są niezależne? Czy A , B , C są niezależne?

7. Na n kartonikach zapisano n różnych liczb rzeczywistych. Kartoniki włożono do pudełka, starannie wymieszano, a następnie losowano kolejno bez zwracania. Niech A_k – k -ta wylosowana liczba jest większa od poprzednich.

- a) Udowodnić, że $\mathbb{P}(A_k) = 1/k$, $k = 1, 2, \dots, n$.
- b) Udowodnić, że zdarzenia A_1, A_2, \dots, A_n są niezależne.

8. Dane są liczby całkowite dodatnie m, n oraz liczby $p, q \in (0, 1)$ spełniające warunek $p + q = 1$. Dowieść, że

$$(1 - p^n)^m + (1 - q^m)^n \geq 1.$$

9. Wyznaczyć najbardziej prawdopodobną liczbę sukcesów w schemacie n prób Bernoulliego z prawdopodobieństwem sukcesu p .

10. Rzucono 10 razy kostką. Jakie jest prawdopodobieństwo tego, że w pierwszym rzucie otrzymano szóstkę, jeśli wiadomo, że

a) otrzymano trzy szóstki?

b) w następnych dziewięciu rzutach otrzymano same szóstki?

11. Rzucamy kostką aż do momentu gdy wyrzucimy piątkę bądź trzy razy szóstkę (łącznie, niekoniecznie pod rząd). Jakie jest prawdopodobieństwo, że rzucimy dokładnie n razy?

12. Prawdopodobieństwo tego, że w urnie znajduje się k kostek, wynosi $\frac{2^k}{k!}e^{-2}$, $k = 0, 1, 2, \dots$. Losujemy kolejno bez zwracania wszystkie kostki z urny i wykonujemy rzuty każdą z nich. Jakie jest prawdopodobieństwo, że uzyskamy l szóstek?

13. Jakie jest prawdopodobieństwo tego, że liczba sukcesów w schemacie Bernoulliego n prób z $p = \frac{1}{2}$ będzie podzielna

a) przez 3?

b) przez 4?

14. Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie przestrzenią probabilistyczną dla schematu n prób Bernoulliego z prawdopodobieństwem sukcesu p . Dla dowolnego $0 \leq k \leq n$, niech A_k oznacza zdarzenie, iż jest dokładnie k sukcesów. Wykazać, że dla dowolnego $B \in \mathcal{F}$ oraz każdego k , prawdopodobieństwo warunkowe $\mathbb{P}(B|A_k)$ nie zależy od p .

15. Rzucamy nieskończenie wiele razy monetą, dla której prawdopodobieństwo wypadnięcia orła wynosi $p \neq 1/2$. Dla $n = 2, 4, \dots$, rozważmy zdarzenie A_n - do rzutu n włącznie wypadło tyle samo orłów co reszek. Udowodnić, że z prawdopodobieństwem 1 zajdzie skończenie wiele spośród zdarzeń A_1, A_2, \dots .

16. Rzucamy nieskończenie wiele razy monetą, dla której prawdopodobieństwo wypadnięcia orła wynosi $p \in (0, 1]$. Udowodnić, że z prawdopodobieństwem 1 wystąpi nieskończenie wiele serii złożonych ze 100 orłów pod rząd.

17. Dane są dwie miary probabilistyczne μ, ν na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

a) Załóżmy, że dla dowolnej liczby $t > 0$ mamy $\nu([-t, t]) = \mu([-t, t])$. Udowodnić, że jeśli $A \in \mathcal{B}(\mathbb{R})$ jest symetryczny względem 0, to $\mu(A) = \nu(A)$.

b) Przypuśćmy, że \mathcal{K} jest pewną klasą generującą $\mathcal{B}(\mathbb{R})$ (tzn. spełniającą $\sigma(\mathcal{K}) = \mathcal{B}(\mathbb{R})$). Czy z tego, że $\mu(A) = \nu(A)$ dla każdego $A \in \mathcal{K}$, wynika, iż $\mu = \nu$?

5. ZMIENNE LOSOWE I ICH ROZKŁADY

Przechodzimy do kluczowego pojęcia rachunku prawdopodobieństwa.

Definicja 5.1. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną. Zmienną losową d -wymiarową (lub zmienną losową o wartościach w \mathbb{R}^d) nazywamy dowolną mierzalną funkcję $X : \Omega \rightarrow \mathbb{R}^d$ (tzn. spełniającą, dla dowolnego $B \in \mathcal{B}(\mathbb{R}^d)$, warunek $\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$). W szczególności, dla $d = 1$ będziemy mówić po prostu „zmienna losowa”.

Twierdzenie 5.1. *Jeśli $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną, X jest d -wymiarową zmienną losową oraz $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ jest funkcją borelowską, to $f(X)$ jest k -wymiarową zmienną losową. W szczególności teza zachodzi więc dla ciągłych funkcji f .*

W szczególności, widzimy iż jeśli $X = (X_1, X_2, \dots, X_d)$, to X_1, X_2, \dots, X_d są jednowymiarowymi zmiennymi losowymi.

Kolejnym ważnym pojęciem jest tzw. rozkład zmiennej losowej. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną, a X jest d -wymiarową zmienną losową. Dla dowolnego $B \in \mathcal{B}(\mathbb{R}^d)$ określamy

$$P_X(B) = \mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

(czyli P_X jest obrazem miary \mathbb{P} przy przekształceniu X). Jak łatwo sprawdzić, $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_X)$ jest nową przestrzenią probabilistyczną.

Definicja 5.2. Miarę P_X nazywamy *rozkładem zmiennej X* .

Definicja 5.3. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną, a X jest d -wymiarową zmienną losową. *Dystrybuantą* tej zmiennej losowej nazywamy funkcję $F_X : \mathbb{R}^d \rightarrow [0, 1]$ daną wzorem

$$\begin{aligned} F_X(x_1, x_2, \dots, x_d) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d) \\ &= P_X((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]). \end{aligned}$$

Dystrybuanta zależy tylko od rozkładu, a więc jest sens mówić o *dystrybuancie rozkładu prawdopodobieństwa*.

Przykłady:

1. Rzucamy raz symetryczną monetą; $\Omega = \{O, R\}$. Niech $X(O) = 1$, $X(R) = -1$. Wówczas

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & \text{jeśli } t < -1, \\ 1/2 & \text{jeśli } -1 \leq t < 1, \\ 1 & \text{jeśli } t \geq 1. \end{cases}$$

2. Wybieramy losowo punkt z koła o promieniu R : mamy zatem $(\Omega, \mathcal{F}, \mathbb{P}) = (K(0, R), \mathcal{B}(K(0, R)), \frac{|\cdot|}{\pi R^2})$. Niech $X(\omega) = \rho(\omega, 0)$ będzie odległością wylosowanego punktu od środka. Wówczas

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & \text{jeśli } t < 0, \\ t^2/R^2 & \text{jeśli } 0 \leq t < R, \\ 1 & \text{jeśli } t \geq R. \end{cases}$$

Twierdzenie 5.2 (Własności dystrybuanty). *Załóżmy, że X jest (jednowymiarową) zmienną losową. Wówczas*

- a) F_X jest niemalejąca.
- b) $\lim_{t \rightarrow \infty} F_X(t) = 1$, $\lim_{t \rightarrow -\infty} F_X(t) = 0$.
- c) F jest prawostronnie ciągła.
- d) dla dowolnego $t \in \mathbb{R}$ istnieje lewostronna granica $F_X(t-) = \lim_{s \uparrow t} F_X(s)$ i

$$F_X(t-) = \mathbb{P}(X < t) = \mathbb{P}(X \in (-\infty, t)).$$

e) F_X jest nieciągła w punkcie t_0 wtedy i tylko wtedy, gdy $\mathbb{P}(X = t_0) > 0$ (taką liczbę t_0 nazywamy wówczas atomem rozkładu). Ścisłej, dla dowolnego $t_0 \in \mathbb{R}$ mamy $\mathbb{P}(X = t_0) = F_X(t_0) - F_X(t_0-)$.

f) dla dowolnych $a < b$ mamy

$$\begin{aligned} \mathbb{P}(a \leq X \leq b) &= \mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a-), \\ \mathbb{P}(a < X \leq b) &= \mathbb{P}(X \in (a, b]) = F_X(b) - F_X(a), \\ \mathbb{P}(a \leq X < b) &= \mathbb{P}(X \in [a, b)) = F_X(b-) - F_X(a-), \\ \mathbb{P}(a < X < b) &= \mathbb{P}(X \in (a, b)) = F_X(b-) - F_X(a). \end{aligned}$$

Dowód. a) Jeśli $t_1 < t_2$, to $(-\infty, t_1] \subset (-\infty, t_2]$, a więc $F_X(t_1) = P_X((-\infty, t_1]) \leq P_X((-\infty, t_2]) = F_X(t_2)$.

b) Dla dowolnego ciągu $(t_n)_{n \geq 1}$ rosnącego do nieskończoności zachodzi równość $\mathbb{R} = \bigcup_n (-\infty, t_n]$ i przedziały pod sumą są wstępujące. Zatem, korzystając z twierdzenia o ciągłości,

$$1 = P_X(\mathbb{R}) = \lim_{n \rightarrow \infty} P_X((-\infty, t_n]) = \lim_{n \rightarrow \infty} F_X(t_n).$$

Analogicznie dowodzimy drugą część.

c) Weźmy $t \in \mathbb{R}$ oraz ciąg $(t_n)_{n \geq 1}$ malejący do t . Ciąg przedziałów $(t, t_n]$ jest zstępujący i $\bigcap_n (t, t_n] = \emptyset$, a więc

$$\begin{aligned} 0 &= P_X \left(\bigcap_n (t, t_n] \right) = \lim_{n \rightarrow \infty} P_X((t, t_n]) = \lim_{n \rightarrow \infty} P_X((-\infty, t_n] \setminus (-\infty, t]) \\ &= \lim_{n \rightarrow \infty} [P_X((-\infty, t_n]) - P_X((-\infty, t])] = \lim_{n \rightarrow \infty} (F_X(t_n) - F_X(t)). \end{aligned}$$

d) Rozumujemy podobnie jak w c).

e) Wynika to wprost z d) oraz równości $\mathbb{P}(X = t_0) = \mathbb{P}(X \leq t_0) - \mathbb{P}(X < t_0)$.

f) Udowodnimy tylko pierwszą równość, pozostałe wykazuje się analogicznie:

$$\mathbb{P}(a \leq X \leq b) = P_X([a, b]) = P_X((-\infty, b]) - P_X((-\infty, a)) = F_X(b) - F_X(a-). \quad \square$$

Uwaga: Jeśli funkcja F spełnia warunki a), b) oraz c) z powyższego twierdzenia, to jest dystrybuantą pewnego rozkładu (dowód pozostawiamy jako ćwiczenie).

Twierdzenie 5.3 (O jednoznaczności). *Dystrybuanta zmiennej losowej d -wymiarowej wyznacza rozkład jednoznacznie.*

Dowód: Załóżmy, że X, Y są zmiennymi losowymi posiadającymi tę samą dystrybuantę. Chcemy wykazać, że $P_X = P_Y$, czyli

$$(*) \quad P_X(B) = P_Y(B)$$

dla dowolnego podzbioru borelowskiego \mathbb{R}^d . Klasa wszystkich zbiorów B spełniających (*) tworzy λ -układ. Z drugiej strony, równość dystrybuant daje, iż

$$\begin{aligned} P_X((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]) &= F_X(x) = F_Y(x) \\ &= P_Y((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]), \end{aligned}$$

a zatem (*) zachodzi dla zbiorów postaci $(-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]$, które tworzą π -układ generujący $\mathcal{B}(\mathbb{R}^d)$. Teza twierdzenia wynika więc natychmiast z lematu o $\pi - \lambda$ układach. \square

Definicja 5.4. Załóżmy, że $X = (X_1, X_2, \dots, X_d)$ jest d -wymiarową zmienną losową oraz $1 \leq i_1 < i_2 < \dots < i_k \leq d$ ($k < d$). Wówczas $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$ też jest zmienną losową i jej rozkład nazywamy k -wymiarowym rozkładem brzegowym rozkładu X .

Uwaga: Rozkłady brzegowe nie wyznaczają na ogół rozkładu łącznego. Rozważmy następujący przykład:

I - rzucamy trzy razy monetą i niech X_1 będzie liczbą reszek przed pojawieniem się pierwszego orła, a X_2 oznacza łączną liczbę orłów.

II - rzucamy dwie serie po trzy razy monetą i niech X'_1 będzie liczbą reszek przed pojawieniem się pierwszego orła w pierwszej serii, a X'_2 oznacza łączną liczbę orłów w drugiej serii.

Jest oczywiste, że zmienne X_1 oraz X'_1 mają ten sam rozkład; tak samo, X_2 oraz X'_2 mają ten sam rozkład. Z drugiej strony, zmienne (X_1, X_2) oraz (X'_1, X'_2) *nie mają* tego samego rozkładu: istotnie,

$$P_{(X_1, X_2)}(\{(3, 3)\}) = 0 \quad \text{oraz} \quad P_{(X'_1, X'_2)}(\{(3, 3)\}) = 2^{-6}.$$

Tak więc rozkłady brzegowe zmiennych (X_1, X_2) oraz (X'_1, X'_2) są identyczne, ale rozkłady łączne są różne.

W dalszej części wykładu będziemy stosować następujące oznaczenie: jeśli X jest d -wymiarową zmienną losową, to

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

jest σ -ciałem zdarzeń generowanym przez X .

Definicja 5.5. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną, a $\{X_i\}_{i \in I}$ jest rodziną zmiennych losowych, przy czym X_i przyjmuje wartości w \mathbb{R}^{d_i} . Mówimy, że zmienne te są niezależne, jeśli σ -ciała generowane przez te zmienne są niezależne. Innymi słowy, zmienne $\{X_i\}_{i \in I}$ są niezależne wtedy i tylko wtedy, gdy dla każdego n , dowolnych parami różnych $i_1, i_2, \dots, i_n \in I$ oraz dowolnych $B_1 \in \mathcal{B}(\mathbb{R}^{d_{i_1}}), \dots, B_n \in \mathcal{B}(\mathbb{R}^{d_{i_n}})$,

$$\mathbb{P}(X_{i_1} \in B_1, X_{i_2} \in B_2, \dots, X_{i_n} \in B_n) = \mathbb{P}(X_{i_1} \in B_1)\mathbb{P}(X_{i_2} \in B_2) \dots \mathbb{P}(X_{i_n} \in B_n).$$

Przykłady:

1. Zdarzenia A_1, A_2, \dots, A_n są niezależne wtedy i tylko wtedy, gdy $1_{A_1}, 1_{A_2}, \dots, 1_{A_n}$ są niezależne. Istotnie, wynika to natychmiast z faktu, iż dla dowolnego

zdarzenia A oraz dowolnego podzbioru borelowskiego prostej mamy

$$\{1_A \in B\} = \begin{cases} A & \text{jeśli } 0 \notin B, 1 \in B, \\ A' & \text{jeśli } 0 \in B, 1 \notin B, \\ \emptyset & \text{jeśli } 0, 1 \notin B, \\ \Omega & \text{jeśli } 0, 1 \in B. \end{cases}$$

2. W schemacie n prób Bernoulliego określmy

$$\begin{aligned} X_i(\omega) &= X_i(\omega_1, \omega_2, \dots, \omega_n) \\ &:= \omega_i = \begin{cases} 1 & \text{jeśli w } i\text{-tej próbie był sukces,} \\ 0 & \text{jeśli w } i\text{-tej próbie była porażka.} \end{cases} \end{aligned}$$

Wówczas X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi. Wyniknie to łatwo z faktów, które przytoczymy w dalszej części wykładu.

Omówimy teraz pewne warunki które są równoważne niezależności zmiennych losowych. Dla prostoty, skupimy się na przypadku jednowymiarowym, ale poniższe twierdzenie pozostaje w mocy także w sytuacji gdy zmienne przyjmują wartości wektorowe.

Twierdzenie 5.4. *Załóżmy, że X_1, X_2, \dots, X_n są zmiennymi losowymi. Następujące warunki są równoważne.*

- 1) X_1, X_2, \dots, X_n są niezależne.
- 2) Dla dowolnych $B_1, B_2, \dots, B_n \in \mathcal{B}(\mathbb{R})$ zdarzenia $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \dots, \{X_n \in B_n\}$ są niezależne.
- 3) $P_{(X_1, X_2, \dots, X_n)} = P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}$,
- 4) Dla dowolnego $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,

$$F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n).$$

Dowód. Wiemy już, iż warunki 1) oraz 2) są równoważne.

2) \Rightarrow 4). Mamy

$$\begin{aligned} F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= \mathbb{P}(X_1 \leq x_1)\mathbb{P}(X_2 \leq x_2) \dots \mathbb{P}(X_n \leq x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n). \end{aligned}$$

4) \Rightarrow 3) Niech X' będzie n -wymiarową zmienną losową o rozkładzie $P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}$. Dla dowolnego $x = (x_1, x_2, \dots, x_n)$ mamy

$$\begin{aligned} F_{X'}(x_1, x_2, \dots, x_n) &= P_{X_1} \otimes \dots \otimes P_{X_n}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]) \\ &= F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n) = F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n). \end{aligned}$$

Na mocy twierdzenia o jednoznaczności, wynika stąd $P_{X'} = P_{(X_1, X_2, \dots, X_n)}$.

3) \Rightarrow 1) Dla dowolnych podzbiorów borelowskich B_1, B_2, \dots, B_n prostej rzeczywistej mamy

$$\begin{aligned} \mathbb{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) &= P_{(X_1, X_2, \dots, X_n)}(B_1 \times B_2 \times \dots \times B_n) \\ &= P_{X_1}(B_1)P_{X_2}(B_2) \dots P_{X_n}(B_n) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2) \dots \mathbb{P}(X_n \in B_n). \end{aligned}$$

Dowód jest zakończony. \square

Przechodzimy teraz do bardziej szczegółowego omówienia typów i przykładów rozkładów prawdopodobieństwa w \mathbb{R}^d .

Definicja 5.6. Mówimy, że d -wymiarowa zmienna losowa X ma dyskretny (skokowy, atomowy) rozkład, jeśli $\mathbb{P}(X \in S_X) = 1$, gdzie

$$S_X = \{x \in \mathbb{R}^d : \mathbb{P}(X = x) > 0\}$$

jest zbiorem *atomów* rozkładu.

Uwagi:

1) Dla dowolnej zmiennej d -wymiarowej X zbiór S_X jest co najwyżej przeliczalny, gdyż

$$S_X = \bigcup_{n \geq 1} \{x \in \mathbb{R}^d : \mathbb{P}(X = x) > 1/n\},$$

i każdy ze zbiorów występujących pod sumą jest skończony.

2) Rozkład dyskretny jest jednoznacznie wyznaczony przez co najwyżej przeliczalny zbiór $S \subset \mathbb{R}^d$ oraz funkcję $p : S \rightarrow [0, 1]$ taką, że $\sum_{x \in S} p(x) = 1$. Istotnie, wówczas

$$P(B) = \sum_{x \in S \cap B} p(x).$$

Odnajmy bardzo prosty fakt.

Twierdzenie 5.5. *Zmienne losowe X_1, X_2, \dots, X_n mają rozkład skokowy wtedy i tylko wtedy, gdy zmienna (X_1, X_2, \dots, X_n) ma rozkład skokowy.*

W przypadku gdy mamy do czynienia ze skończoną rodziną zmiennych dyskretnych, warunek niezależności może być badany za pomocą następującego prostego kryterium.

Twierdzenie 5.6. *Zmienne losowe X_1, X_2, \dots, X_n mające rozkłady dyskretne są niezależne wtedy i tylko wtedy, gdy dla dowolnych $x_1 \in S_{X_1}, x_2 \in S_{X_2}, \dots, x_n \in S_{X_n}$ zachodzi*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \dots \mathbb{P}(X_n = x_n).$$

Dowód. \Rightarrow Oczywiście: postulowana równość jest szczególnym przypadkiem warunku występującego w definicji niezależności.

\Leftarrow Dla ułatwienia zapisu, przeprowadzimy dowód tylko dla $n = 2$; przypadek $n \geq 3$ rozpatruje się analogicznie. Mamy

$$\begin{aligned} \mathbb{P}(X_1 \in B_1, X_2 \in B_2) &= \mathbb{P}(X_1 \in S_{X_1} \cap B_1, X_2 \in S_{X_2} \cap B_2) \\ &= \sum_{x_1 \in S_{X_1} \cap B_1, x_2 \in S_{X_2} \cap B_2} \mathbb{P}(X_1 = x_1, X_2 = x_2) \\ &= \sum_{x_1 \in S_{X_1} \cap B_1} \sum_{x_2 \in S_{X_2} \cap B_2} \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2). \quad \square \end{aligned}$$

Przykłady:

1) *Rozkład skupiony w punkcie* $a \in \mathbb{R}^d$, ozn. δ_a . Zmienna X ma rozkład skupiony w a , jeśli $\mathbb{P}(X = a) = 1$; $S_X = \{a\}$.

2) *Rozkład dwupunktowy skupiony* w $a, b \in \mathbb{R}^d$, $a \neq b$. Zmienna X ma rozkład dwupunktowy skupiony na $\{a, b\}$ jeśli $\mathbb{P}(X = a) = p$ oraz $\mathbb{P}(X = b) = 1 - p$ dla pewnego $p \in (0, 1)$.

3) *Rozkład Bernoulliego (rozkład dwumianowy)* z parametrami n, p ($n = 1, 2, \dots$, $p \in (0, 1)$), ozn. $B(n, p)$. Zmienna X ma rozkład $B(n, p)$, jeśli $\mathbb{P}(X = k) =$

$\binom{n}{k}p^k(1-p)^{n-k}$, $k \in S_X = \{0, 1, 2, \dots, n\}$. Innymi słowy, jeśli mamy dany schemat Bernoulliego składający się z n prób o prawdopodobieństwie sukcesu p , to łączna liczba sukcesów ma rozkład Bernoulliego $B(n, p)$.

Dygresja. Załóżmy, że zmienne X_1, X_2, \dots, X_n zadane są wzorem

$$(*) \quad X_i = \begin{cases} 1 & \text{jeśli w } i\text{-tej próbie był sukces,} \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Wówczas jak łatwo sprawdzić korzystając z Twierdzenia 5.6, zmienne X_1, X_2, \dots, X_n są niezależne. Z drugiej strony, $X_1 + X_2 + \dots + X_n$ jest łączną liczbą sukcesów. Otrzymaliśmy więc następujący fakt:

Założmy, że X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o tym samym rozkładzie (czasami te dwie własności będziemy w skrócie oznaczać przez i.i.d.) zadany przez $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Wówczas zmienna $X_1 + X_2 + \dots + X_n$ ma rozkład $B(n, p)$.

Skorzystaliliśmy tu z następującej prostej obserwacji:

Twierdzenie 5.7. *Jeśli d -wymiarowe zmienne X, Y mają ten sam rozkład oraz $f: \mathbb{R}^d \rightarrow \mathbb{R}^k$ jest funkcją borelowską, to $f(X)$ oraz $f(Y)$ mają ten sam rozkład.*

4) *Rozkład geometryczny z parametrem p ($0 < p < 1$), ozn. $\text{Geom}(p)$.* Zmienna X ma taki rozkład, jeśli

$$\mathbb{P}(X = k) = (1 - p)^k p, \quad k \in S_X = \{0, 1, 2, \dots\}.$$

Interpretacja: załóżmy, iż dany jest schemat Bernoulliego o nieskończonej liczbie prób i prawdopodobieństwie sukcesu p . Niech X oznacza liczbę porażek poprzedzających pojawienie się pierwszego sukcesu. Wówczas X ma rozkład geometryczny z parametrem p . Istotnie, wprowadzając zmienne X_1, X_2, \dots jak w poprzednim przykładzie, możemy zapisać

$$\begin{aligned} \mathbb{P}(X = k) &= \mathbb{P}(X_1 = 0, X_2 = 0, \dots, X_k = 0, X_{k+1} = 1) \\ &= \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 0)\dots\mathbb{P}(X_k = 0)\mathbb{P}(X_{k+1} = 1) = (1 - p)^k p. \end{aligned}$$

Czasami w literaturze rozkład geometryczny jest definiowany nieco inaczej: mianowicie, $Y \sim \text{Geom}(p)$ jeśli dla dowolnej $k \in S_Y = \{1, 2, \dots\}$, mamy $\mathbb{P}(Y = k) = (1 - p)^{k-1} p$. Wówczas zmienna ta ma interpretację jako *czas oczekiwania na pierwszy sukces w nieskończonym schemacie Bernoulliego o prawdopodobieństwie sukcesu p* . Jest ona związana z poprzednią zmienną zależnością $Y = X + 1$.

5) *Rozkład Poissona z parametrem λ ($\lambda > 0$), ozn. $\text{Pois}(\lambda)$.* Zmienna X ma taki rozkład, jeśli

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in S_X = \{0, 1, 2, \dots\}.$$

Rozkład Poissona powstaje przez odpowiednie przejście graniczne dla rozkładów Bernoulliego. Ścisłej, zachodzi następujący fakt (który pozostawiamy bez dowodu).

Twierdzenie 5.8 (Poissona). *Załóżmy, że $(p_n)_{n \geq 1}$ jest ciągiem liczb z przedziału $(0, 1)$ spełniającym warunek $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. Wówczas dla dowolnej liczby $k \in \{0, 1, 2, \dots\}$,*

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Innymi słowy, jeśli X_n ma rozkład $B(n, p)$, X ma rozkład $\text{Pois}(\lambda)$ i $np \approx \lambda$, to rozkłady X_n oraz X są bliskie.

Przechodzimy do kolejnej ważnej rodziny rozkładów.

Definicja 5.7. Mówimy, że zmienna losowa d -wymiarowa X ma rozkład ciągły, jeśli istnieje funkcja borelowska $g : \mathbb{R}^d \rightarrow [0, \infty)$ taka, że dla dowolnego podzbioru borelowskiego $B \subseteq \mathbb{R}^d$,

$$\mathbb{P}(X \in B) = P_X(B) = \int_B g(x) dx.$$

Funkcję g nazywamy wówczas *gęstością rozkładu*.

Uwagi:

1. Jeśli P_X jest ciągły, to $S_X = \emptyset$, ale nie na odwrót (warunkiem koniecznym i dostatecznym jest równość $\mathbb{P}(X \in B) = 0$ dla dowolnego $B \subset \mathbb{R}^d$ o zerowej mierze Lebesgue'a).

2. Jeśli $g : \mathbb{R}^d \rightarrow [0, \infty)$ jest gęstością pewnego rozkładu μ oraz $\tilde{g} : \mathbb{R}^d \rightarrow [0, \infty)$ jest funkcją borelowską, to \tilde{g} jest gęstością μ wtedy i tylko wtedy, gdy $g = \tilde{g}$ p.w.. Istotnie, implikacja \Leftarrow jest oczywista:

$$\mu(A) = \int_A g(x) dx = \int_A \tilde{g}(x) dx.$$

Przechodzimy do implikacji \Rightarrow : mamy $\int_B g = \int_B \tilde{g}$ dla dowolnego zbioru borelowskiego B . Załóżmy, że teza nie zachodzi; wówczas jeden ze zbiorów $\{g > \tilde{g}\}$, $\{g < \tilde{g}\}$ ma dodatnią miarę. Bez straty ogólności załóżmy, że jest to pierwszy zbiór, i oznaczmy go przez B . Jest to zbiór borelowski, i $\int_B g > \int_B \tilde{g}$, sprzeczność.

3. Każda funkcja borelowska $g : \mathbb{R}^d \rightarrow [0, \infty)$ taka, że $\int_{\mathbb{R}^d} g(x) dx = 1$ jest gęstością pewnego rozkładu, zadanego przez $\mu(B) = \int_B g(x) dx$ dla $B \in \mathcal{B}(\mathbb{R}^d)$.

4. Zmienna X ma rozkład ciągły wtedy i tylko wtedy, gdy istnieje funkcja borelowska $g : \mathbb{R}^d \rightarrow [0, \infty)$ taka, że

$$F_X(x_1, x_2, \dots, x_d) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_d} g(y_1, y_2, \dots, y_d) dy_d dy_{d-1} \dots dy_1.$$

Istotnie:

\Rightarrow Mamy $F(x_1, \dots, x_d) = \mathbb{P}((X_1, \dots, X_d) \in (-\infty, x_1] \times \dots \times (-\infty, x_d])$ i wystarczy skorzystać z definicji gęstości.

\Leftarrow Zbiegając z x_1, x_2, \dots, x_d do nieskończoności widzimy, że $\int_{\mathbb{R}^d} g(x) dx = 1$. Niech μ będzie rozkładem prawdopodobieństwa w \mathbb{R}^d , zadanym przez gęstość g (patrz Uwaga 3 powyżej). Wówczas, na mocy poprzedniej implikacji mamy $F_X = F_\mu$, a zatem, z twierdzenia o jednoznaczności, $P_X = \mu$.

5. Jako wniosek z poprzedniej uwagi, otrzymujemy następujący fakt.

Twierdzenie 5.9. Załóżmy, że X jest d -wymiarową zmienną losową o dystrybucie F i niech

$$g(x_1, x_2, \dots, x_d) = \frac{\partial^d F}{\partial x_1 \partial x_2 \dots \partial x_d}(x_1, x_2, \dots, x_d) \quad p.w..$$

Wówczas jeśli $\int_{\mathbb{R}^d} g = 1$, to X ma rozkład ciągły i jego gęstością jest funkcja g .

Przykład:

Losujemy punkt z koła o promieniu R . Niech X oznacza odległość punktu od środka koła. Wówczas, jak już wiemy,

$$F_X(t) = \begin{cases} 0 & \text{dla } t < 0, \\ t^2/R^2 & \text{dla } 0 \leq t < R, \\ 1 & \text{dla } t \geq R. \end{cases}$$

Różniczkując, dostajemy

$$g(t) = \begin{cases} 0 & \text{dla } t < 0 \text{ lub } t \geq R, \\ 2t/R^2 & \text{dla } 0 \leq t < R. \end{cases}$$

Łatwo sprawdzić, że $\int_{\mathbb{R}} g = 1$, a więc g jest gęstością rozkładu zmiennej X .

Twierdzenie 5.10. *Jeśli zmienna $X = (X_1, X_2, \dots, X_d)$ ma rozkład ciągły, to jej rozkłady brzegowe też są ciągłe. Ponadto*

$$g_{X_i}(x) = \int_{\mathbb{R}^{d-1}} g(x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d.$$

Ogólniej, aby otrzymać gęstość wektora $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$, musimy odcałkować gęstość X po wszystkich x_i dla $i \notin \{i_1, i_2, \dots, i_k\}$.

Dowód. Mamy

$$\begin{aligned} \mathbb{P}(X_i \in B) &= \mathbb{P}((X_1, X_2, \dots, X_d) \in \mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i}) = \int_{\mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i}} g \\ &= \int_B \left(\int_{\mathbb{R}^{i-1} \times \mathbb{R}^{d-i}} g(x) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \right) dx_i. \end{aligned}$$

W przypadku gdy rozkład brzegowy jest wielowymiarowy, rozumowanie jest analogiczne. \square

Uwaga: Implikacja w drugą stronę nie zachodzi: wystarczy wziąć dowolną jednowymiarową zmienną losową X o rozkładzie ciągłym i rozważyć wektor (X, X) (który rozkładu ciągłego już nie posiada, gdyż jest skoncentrowany na zbiorze $\{(x, x) : x \in \mathbb{R}\}$, który ma zerową miarę Lebesgue'a).

Ważne przykłady rozkładów ciągłych

1) *Rozkład jednostajny (równomierny) na zbiorze D , ozn. $\mathcal{U}(D)$.* Załóżmy, że $D \in \mathcal{B}(\mathbb{R}^d)$ spełnia warunek $0 < |D| < \infty$. Zmienna losowa X ma rozkład jednostajny na D , jeśli ma gęstość

$$g(x) = \frac{1}{|D|} 1_D(x) = \begin{cases} 1/|D| & \text{dla } x \in D, \\ 0 & \text{dla } x \notin D. \end{cases}$$

Dla dowolnego $B \in \mathcal{B}(\mathbb{R}^d)$ mamy wówczas $\mathbb{P}(X \in B) = \int_B g(x) dx = \frac{|B \cap D|}{|D|}$.

W szczególności, jeśli $d = 1$ oraz $D = [a, b]$, dostajemy rozkład o gęstości $g(x) = \frac{1}{b-a} 1_{[a,b]}(x)$ i dystrybuancie

$$F_X(x) = \int_{-\infty}^x g(s) ds = \begin{cases} 0 & \text{dla } x < a, \\ (x-a)/(b-a) & \text{dla } a \leq x < b, \\ 1 & \text{dla } x \geq b. \end{cases}$$

2) *Rozkład wykładniczy z parametrem λ ($\lambda > 0$), ozn. $\text{Exp}(\lambda)$.* Zmienna losowa X ma taki rozkład, jeśli ma gęstość

$$g(x) = \lambda e^{-\lambda x} 1_{[0, \infty)}(x) = \begin{cases} 0 & \text{dla } x < 0, \\ \lambda e^{-\lambda x} & \text{dla } x \geq 0. \end{cases}$$

Jak łatwo policzyć, wówczas

$$F_X(x) = \begin{cases} 0 & \text{dla } x < 0, \\ 1 - e^{-\lambda x} & \text{dla } x \geq 0. \end{cases}$$

Rozkład wykładniczy służy do modelowania czasu oczekiwania na zjawisko całkowicie losowe. Załóżmy, że nieujemna zmienna losowa X oznacza taki czas oczekiwania, a całkowitą losowość zapisujemy poprzez własność *braku pamięci*:

$$\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > t) \quad \text{dla wszystkich } s, t \geq 0.$$

Oznaczając $f(t) = \mathbb{P}(X > t)$, widzimy, iż powyższe równanie jest równoważne $f(t + s) = f(t)f(s)$. Dodatkowo, f jest nierosnąca, prawostronnie ciągła oraz spełnia $f(0) = 1$, $\lim_{t \rightarrow \infty} f(t) = 0$, skąd już wynika, że $f(t) = e^{-\lambda t}$ dla pewnego $\lambda > 0$, a zatem X ma rozkład wykładniczy.

3) *Rozkład Gaussa (rozkład normalny).* Załóżmy, że m jest ustalonym wektorem w \mathbb{R}^d , a A jest symetryczną i dodatnio określoną macierzą $d \times d$. Zmienna losowa X ma rozkład normalny (z parametrami m i A), jeśli jej gęstość wynosi

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\langle A(x - m), x - m \rangle\right)$$

($\langle \cdot, \cdot \rangle$ oznacza iloczyn skalarny w \mathbb{R}^d). Sprawdźmy, że funkcja g istotnie całkuje się do 1. Z algebry liniowej wiadomo, że istnieje izometria (macierz ortogonalna) $B : \mathbb{R}^d \rightarrow \mathbb{R}^d$ taka, że $B^t A B$ ma postać diagonalną:

$$B^t A B = \begin{bmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & a_d \end{bmatrix}.$$

Podstawiając $x - m = By$, dostajemy

$$\begin{aligned} \int_{\mathbb{R}^d} g(x) dx &= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle ABy, By \rangle\right) |\det B| dy \\ &= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\langle B^{-1} A B y, y \rangle\right) dy \\ &= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} \sum_{k=1}^d a_k y_k^2\right) dy \\ &= \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \prod_{k=1}^d \int_{\mathbb{R}} e^{-a_k y_k^2 / 2} dy_k \\ &= \frac{\sqrt{\det A}}{\sqrt{a_1 a_2 \dots a_d}} \prod_{k=1}^d \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-z_k^2 / 2} dz_k \right) = 1, \end{aligned}$$

gdzie w przedostatnim kroku dokonaliśmy podstawienia $z_k = \sqrt{a_k}y_k$, a w ostatnim skorzystaliśmy z równości

$$\det A = a_1 a_2 \dots a_d \quad \text{oraz} \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2} dx = 1.$$

W szczególnym przypadku $d = 1$, jeśli $m \in \mathbb{R}$ i $\sigma > 0$ (i za macierz A weźmiemy $[\sigma^{-2}]$), otrzymujemy gęstość

$$g(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

Rozkład ten oznaczamy $\mathcal{N}(m, \sigma^2)$. W szczególności, rozkład $\mathcal{N}(0, 1)$ nazywamy standardowym rozkładem normalnym (standardowym rozkładem Gaussa).

W przypadku zmiennych o rozkładzie ciągłym, mamy następujące kryterium niezależności.

Twierdzenie 5.11. *Załóżmy, że g_1, g_2, \dots, g_n są gęstościami. Wówczas zmienne losowe X_1, X_2, \dots, X_n o rozkładach z gęstościami g_1, g_2, \dots, g_n są niezależne wtedy i tylko wtedy, gdy zmienna (X_1, X_2, \dots, X_n) ma gęstość $g(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2) \dots g_n(x_n)$.*

Dowód. \Rightarrow Mamy, na mocy niezależności zmiennych,

$$\begin{aligned} F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) &= F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n) \\ &= \int_{-\infty}^{x_1} g_1(y_1)dy_1 \int_{-\infty}^{x_2} g_2(y_2)dy_2 \dots \int_{-\infty}^{x_n} g_n(y_n)dy_n \\ &= \int_{(-\infty, x_1] \times \dots \times (-\infty, x_n]} g_1(y_1)g_2(y_2) \dots g_n(y_n)dy_1dy_2 \dots dy_n, \end{aligned}$$

skąd natychmiast wynika teza.

\Leftarrow Piszemy ten sam ciąg równości, zaczynając od końca. □

Przykład: Zmienne X_1, X_2, \dots, X_n o rozkładach jednostajnych na D_1, D_2, \dots, D_n są niezależne wtedy i tylko wtedy, gdy (X_1, X_2, \dots, X_n) ma rozkład jednostajny na $D_1 \times D_2 \times \dots \times D_n$. Wynika to natychmiast z tego, że gęstość rozkładu jednostajnego na $D_1 \times D_2 \times \dots \times D_n$ wynosi

$$\frac{1_{D_1 \times D_2 \times \dots \times D_n}(x_1, x_2, \dots, x_n)}{|D_1 \times D_2 \times \dots \times D_n|} = \frac{1_{D_1}(x_1)}{|D_1|} \cdot \frac{1_{D_2}(x_2)}{|D_2|} \cdot \dots \cdot \frac{1_{D_n}(x_n)}{|D_n|},$$

a czynniki po prawej stronie to gęstości rozkładów jednostajnych na D_1, D_2, \dots, D_n , odpowiednio.

W przypadku gdy zmienne losowe są niezależne i mają rozkłady ciągłe, istnieje efektywna metoda liczenia rozkładu ich sumy. Zachodzi następujący fakt.

Twierdzenie 5.12. *Załóżmy, że X_1, X_2 są niezależnymi, jednowymiarowymi zmiennymi losowymi o rozkładach z gęstościami g_1 oraz g_2 , odpowiednio. Wówczas zmienna $X_1 + X_2$ ma rozkład z gęstością*

$$g_1 * g_2(x) = \int_{\mathbb{R}} g_1(x-y)g_2(y)dy.$$

Widzimy więc, że $g_1 * g_2$ to splot gęstości g_1 i g_2 .

Dowód. Dla dowolnego $B \in \mathcal{B}(\mathbb{R})$ mamy

$$\begin{aligned}
\mathbb{P}(X_1 + X_2 \in B) &= P_{(X_1, X_2)}(\{(x, y) : x + y \in B\}) \\
&= \iint_{\{(x, y) : x + y \in B\}} g_{(X_1, X_2)}(x, y) dx dy \\
&= \iint_{\{(x, y) : x + y \in B\}} g_1(x) g_2(y) dx dy \\
&= \iint_{\mathbb{R}^2} 1_B(x + y) g_1(x) g_2(y) dx dy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1_B(x + y) g_1(x) dx \right) g_2(y) dy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1_B(z) g_1(z - y) dz \right) g_2(y) dy \\
&= \int_B \left(\int_{\mathbb{R}} g_1(z - y) g_2(y) dy \right) dz \\
&= \int_B g_1 * g_2(z) dz. \quad \square
\end{aligned}$$

Przykłady:

1) Jeśli X_1, X_2 są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na $[0, 1]$, to $g_1(x) = g_2(x) = 1_{[0,1]}(x)$, a więc $X_1 + X_2$ ma gęstość

$$\begin{aligned}
g_1 * g_2(x) &= \int_{\mathbb{R}} 1_{[0,1]}(x - y) 1_{[0,1]}(y) dy \\
&= \int_{\mathbb{R}} 1_{[x-1, x]}(y) 1_{[0,1]}(y) dy \\
&= \begin{cases} 0 & \text{dla } x < 0, \\ x & \text{dla } 0 \leq x < 1, \\ 2 - x & \text{dla } 1 \leq x < 2, \\ 0 & \text{dla } x \geq 2. \end{cases}
\end{aligned}$$

2) Załóżmy, że X_1, X_2 są niezależnymi zmiennymi losowymi o rozkładach $N(m_1, \sigma_1^2)$, $N(m_2, \sigma_2^2)$: zatem

$$g_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x - m_i)^2}{2\sigma_i^2}\right), \quad i = 1, 2.$$

Wówczas $X_1 + X_2$ ma gęstość

$$\begin{aligned}
g_1 * g_2(x) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{(x - y - m_1)^2}{2\sigma_1^2} - \frac{(y - m_2)^2}{2\sigma_2^2}\right) dy \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{(x - m_1 - m_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right)
\end{aligned}$$

(dowód ostatniej równości pozostawiamy jako ćwiczenie). Zatem $X_1 + X_2 \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. I ogólniej, przez indukcję: jeśli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładach $N(m_1, \sigma_1^2), N(m_2, \sigma_2^2), \dots, N(m_n, \sigma_n^2)$, to $X_1 + X_2 + \dots + X_n$ ma rozkład $N(m_1 + m_2 + \dots + m_n, \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2)$.

6. ZADANIA

1. Rzucamy monetą, dla której prawdopodobieństwo wypadnięcia orła wynosi $p \in (0, 1]$, aż do momentu wyrzucenia k orłów (łącznie, niekoniecznie pod rząd). Niech X oznacza liczbę rzutów. Wyznaczyć rozkład zmiennej X .

2. Rzucamy dwa razy kostką. Niech X, Y oznaczają minimum oraz maksimum z uzyskanych liczb oczek, odpowiednio. Wyznaczyć rozkłady zmiennych X, Y oraz sprawdzić, że zmienne X i $7 - Y$ mają ten sam rozkład.

3. Na skrzyżowaniu ulic na pewnym kierunku światło czerwone świeci się minutę, a światło zielone - pół minuty (zakładamy, że nie ma żółtego światła). Samochód dojeżdża do skrzyżowania (w danym kierunku) w losowym momencie czasowym. Niech X oznacza czas spędzony na skrzyżowaniu; zakładamy, że nie ma korka.

a) Wyznaczyć rozkład zmiennej X .

b) Załóżmy, że po 20 sekundach samochód wciąż nie przejechał skrzyżowania; jakie jest prawdopodobieństwo, że opuści je w przeciągu najbliższych 10 sekund?

4. Dystrybuanta zmiennej losowej X dana jest wzorem

$$F_X(t) = \begin{cases} 0 & \text{dla } t < -1, \\ \frac{1}{2}(t+1) & \text{dla } -1 \leq t < 0, \\ \frac{3}{4} & \text{dla } 0 \leq t < 4, \\ 1 & \text{dla } t \geq 4. \end{cases}$$

Obliczyć $\mathbb{P}(X = -5)$, $\mathbb{P}(2 < X \leq 5)$, $\mathbb{P}(X = 4)$, $\mathbb{P}(-1 < X < 0)$.

5. Zmienna losowa X ma rozkład o dystrybuancie

$$F(t) = \begin{cases} 0 & \text{dla } t < 0, \\ t/2 & \text{dla } 0 \leq t < 2, \\ 1 & \text{dla } t \geq 2. \end{cases}$$

Wyznaczyć dystrybuantę zmiennych $Y = \max(X, 1)$ oraz $Z = \min(X, X^2)$.

6. Niech $F : \mathbb{R} \rightarrow [0, 1]$ będzie funkcją prawostronnie ciągłą, niemalejącą, taką że $\lim_{t \rightarrow \infty} F(t) = 1$ oraz $\lim_{t \rightarrow -\infty} F(t) = 0$. Wykazać, że F jest dystrybuantą pewnej zmiennej losowej w pewnej przestrzeni probabilistycznej.

7. Z talii 52 kart losujemy ze zwracaniem pięć razy po jednej karcie. Niech X oznacza liczbę wyciągniętych pików, Y - liczbę wyciągniętych kierów, a Z - liczbę wyciągniętych waletów. Czy zmienne X, Y są niezależne? Czy zmienne X, Z są niezależne?

8. Zmienne losowe X_1, X_2, \dots, X_n ($n \geq 6$) są niezależne i mają ten sam rozkład, zadany wzorem $\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = 1/2$, $i = 1, 2, \dots, n$.

a) Czy zmienne $X_1 + X_2, X_1 X_2$ są niezależne?

b) Czy zmienne $X_1 + X_2, X_3, X_4 + X_5 X_6$ są niezależne?

c) Czy zmienne $X_1, X_1 X_2, \dots, X_1 X_2 \dots X_n$ są niezależne?

9. Zmienne losowe X, Y są niezależne, przy czym dla $n = 1, 2, \dots$ mamy $\mathbb{P}(X = n) = (1 - p)^{n-1} p$ oraz $\mathbb{P}(Y = n) = (1 - q)^{n-1} q$. Obliczyć $\mathbb{P}(X \leq Y)$.

10. Dla dowolnej liczby $\omega \in [0, 1]$, niech $X_n(\omega)$ oznacza n -tą cyfrę rozwinięcia dwójkowego ω , $n = 1, 2, \dots$ (jeśli ω posiada dwa różne rozwinięcia, to bierzemy to, które zawiera skończoną liczbę jedynek). Wykazać, że X_1, X_2, \dots są niezależnymi zmiennymi losowymi na przestrzeni probabilistycznej $([0, 1], \mathcal{B}([0, 1]), |\cdot|)$.

11. Zmienne X, Y są niezależne, przy czym X nie ma atomów. Udowodnić, że $\mathbb{P}(X = Y) = 0$.

12. Zmienne losowe X_1, X_2, \dots, X_n są niezależne i mają rozkłady Poissona z parametrami $\lambda_1, \lambda_2, \dots, \lambda_n$. Udowodnić, że $X_1 + X_2 + \dots + X_n$ ma rozkład Poissona z parametrem $\lambda_1 + \lambda_2 + \dots + \lambda_n$.

13. Zmienna losowa X jest niezależna od siebie samej. Udowodnić, że istnieje c takie, że $\mathbb{P}(X = c) = 1$.

14. Zmienna losowa X ma rozkład wykładniczy z parametrem 1.

a) Wyznaczyć rozkłady zmiennych $[X]$ oraz $\{X\}$.

b) Czy zmienne te są niezależne?

Uwaga: $[x], \{x\}$ oznaczają część całkowitą i część ułamkową liczby $x \in \mathbb{R}$, odpowiednio.

15. Zmienna losowa X ma rozkład jednostajny na przedziale $[0, 1]$. Wyznaczyć rozkład zmiennej $Y = -\ln X$.

16. Zmienna losowa X ma rozkład normalny $\mathcal{N}(0, 1)$. Wyznaczyć gęstości zmiennych $Y = e^X, Z = X^2$.

17. Tekst broszury zawiera $n = 100000$ znaków. W trakcie pisania (na komputerze) każdy znak może zostać błędnie wprowadzony z prawdopodobieństwem 0,001. Z kolei redaktor znajduje każdy z błędów z prawdopodobieństwem 0,9, po czym tekst wraca do autora, który znajduje każdy z pozostałych błędów z prawdopodobieństwem 0,5. Oszacować prawdopodobieństwo tego, że po obu korektach broszura będzie zawierała nie więcej niż 3 błędy.

18. Zmienne losowe X_1, X_2, \dots, X_n są niezależne i mają rozkłady wykładnicze z parametrami $\lambda_1, \lambda_2, \dots, \lambda_n$, odpowiednio. Wyznaczyć rozkład zmiennej $Y = \max(X_1, X_2, \dots, X_n)$.

19. Zmienna losowa (X, Y) ma rozkład z gęstością

$$g(x, y) = Cxy1_{\{0 \leq x \leq y \leq 1\}}.$$

a) Wyznaczyć C .

b) Obliczyć $\mathbb{P}(X + Y \leq 1)$.

c) Wyznaczyć rozkład zmiennej X/Y .

d) Czy X, Y są niezależne?

e) Czy $X/Y, Y$ są niezależne?

20. Zmienne X, Y są niezależne i mają rozkład jednostajny na przedziale $[-1, 1]$. Obliczyć $\mathbb{P}(X^2 + Y^2 \leq 1)$.

21. Zmienna losowa X ma rozkład Cauchy'ego, tzn. z gęstością

$$g(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

Udowodnić, że zmienne $X, 1/X$ mają ten sam rozkład.

22. Niech $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx, r > 0$. Mówimy, że zmienna X ma rozkład *gamma* z parametrami λ, r (ozn. $\Gamma(\lambda, r)$), jeśli ma gęstość

$$g_{\lambda, r}(x) = \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x} 1_{[0, \infty)}(x).$$

a) Udowodnić, że jeśli X, Y są niezależnymi zmiennymi losowymi, $X \sim \Gamma(\lambda, r), Y \sim \Gamma(\lambda, s)$, to $X + Y \sim \Gamma(\lambda, r + s)$.

b) Udowodnić, że jeśli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładzie $\text{Exp}(\lambda)$, to $X_1 + X_2 + \dots + X_n$ ma rozkład $\Gamma(\lambda, n)$.

c) Udowodnić, że jeśli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładzie $\mathcal{N}(0, 1)$, to $X_1^2 + X_2^2 + \dots + X_n^2$ ma rozkład $\Gamma(1/2, n/2)$ (jest to tzw. rozkład *chi kwadrat o n stopniach swobody*).

23. Zmienne losowe X, Y są niezależne i mają rozkład wykładniczy z parametrem 1. Udowodnić, że zmienne X/Y oraz $X + Y$ są niezależne.

7. PARAMETRY ROZKŁADU ZMIENNEJ LOSOWEJ (WARTOŚĆ OCZEKIWANA I WARIANCJA)

Definicja 7.1. (i) Załóżmy, że X jest jednowymiarową zmienną losową na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$. Mówimy, że X ma wartość oczekiwaną, jeśli istnieje całka $\int_{\Omega} X(\omega) d\mathbb{P}(\omega)$. Całkę tę nazywamy *wartością oczekiwaną (średnią) zmiennej X* i oznaczamy symbolem $\mathbb{E}X$.

(ii) Jeśli $\mathbb{E}|X| < \infty$, to mówimy, że X jest całkowalna i oznaczamy to przez $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

(iii) Analogicznie, załóżmy, że p jest pewną dodatnią liczbą. Jeśli $\mathbb{E}|X|^p < \infty$, to mówimy że X jest całkowalna z p -tą potęgą i oznaczamy to przez $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$.

(iv) Mówimy, że zmienna losowa X jest *ograniczona*, jeśli istnieje liczba u taka, że $\mathbb{P}(|X| \geq u) = 0$. Oznaczenie: $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

Uwaga: Powyższe definicje mają także sens gdy X jest zmienną wielowymiarową: będzie o tym mowa w dalszej części wykładu.

Odnotujmy dwie ważne nierówności:

1) *Nierówność Minkowskiego.* Jeśli X, Y są zmiennymi losowymi oraz $1 \leq p < \infty$, to

$$(\mathbb{E}|X + Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}.$$

Istnieje wersja tej nierówności dla $p = \infty$: mianowicie,

$$\text{ess sup } |X + Y| \leq \text{ess sup } |X| + \text{ess sup } |Y|,$$

gdzie $\text{ess sup } \xi = \inf\{u : \mathbb{P}(\xi \geq u) = 0\}$ to tzw. istotne supremum zmiennej ξ .

2) *Nierówność Höldera.* Załóżmy, że X, Y są zmiennymi losowymi oraz $p, q \in (1, \infty)$ są liczbami harmonicznie sprzężonymi, tzn. spełniającymi równość $\frac{1}{p} + \frac{1}{q} = 1$. Wówczas

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

Uwaga: Bezpośrednio z definicji widzimy, że wartość oczekiwana jest operatorem liniowym: ściślej, jeśli X_1, X_2, \dots, X_n są całkowalnymi zmiennymi losowymi oraz $a_1, a_2, \dots, a_n \in \mathbb{R}$, to zmienna $a_1X_1 + a_2X_2 + \dots + a_nX_n$ jest całkowalna oraz

$$\mathbb{E}(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1\mathbb{E}X_1 + a_2\mathbb{E}X_2 + \dots + a_n\mathbb{E}X_n.$$

Z analizy znamy następujące trzy twierdzenia o przechodzeniu do granicy pod znakiem całki.

Twierdzenie 7.1 (Lebesgue'a o monotonicznym przechodzeniu do granicy). *Założmy, że X_1, X_2, \dots są nieujemnymi i całkowalnymi zmiennymi losowymi, przy czym $X_n(\omega) \leq X_{n+1}(\omega)$, $n = 1, 2, \dots$. Wówczas*

$$\mathbb{E}\left(\lim_{n \rightarrow \infty} X_n\right) = \lim_{n \rightarrow \infty} \mathbb{E}X_n.$$

W szczególności, zmienna $\lim_{n \rightarrow \infty} X_n$ jest całkowalna wtedy i tylko wtedy, gdy $\lim_{n \rightarrow \infty} \mathbb{E}X_n < \infty$.

Twierdzenie 7.2 (Lebesgue'a o zbieżności zmajoryzowanej). *Założmy, że X_1, X_2, \dots są zmiennymi losowymi zmajoryzowanymi przez pewną zmienną całkowalną η : $|X_n| \leq \eta$ dla $n = 1, 2, \dots$. Jeśli istnieje granica $X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)$ dla prawie wszystkich ω (prawie wszystkich w sensie miary \mathbb{P}), to $\lim_{n \rightarrow \infty} \mathbb{E}X_n = \mathbb{E}X$.*

Twierdzenie 7.3 (Lemat Fatou). *Założmy, że X_1, X_2, \dots są nieujemnymi zmiennymi losowymi. Wówczas $\mathbb{E} \liminf_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} \mathbb{E} X_n$.*

Dygresja: Założmy, że X, Y są zmiennymi losowymi. Mówimy, że X i Y są równe prawie na pewno, jeśli zachodzi równość $\mathbb{P}(X \neq Y) = 0$. Założmy, że $p \in [1, \infty]$ i określmy

$$\|X\|_p = \begin{cases} (\mathbb{E}|X|^p)^{1/p} & \text{dla } 1 \leq p < \infty, \\ \text{ess sup } |X| & \text{dla } p = \infty. \end{cases}$$

Jeśli utożsamimy zmienne losowe równe p.n., to $\|\cdot\|_p$ jest normą na $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Co więcej, przestrzeń $L^p(\Omega, \mathcal{F}, \mathbb{P})$ wraz z tą normą jest liniowa i zupełna (czyli jest przestrzenią Banacha).

Uwaga: Na mocy nierówności Höldera, mamy $\|X\|_p \leq \|X\|_{p'}$ jeśli $p < p'$. Dostajemy stąd inkluzję $L^{p'}(\Omega, \mathcal{F}, \mathbb{P}) \subset L^p(\Omega, \mathcal{F}, \mathbb{P})$.

Definicja 7.2. Dla $p \in (0, \infty)$, liczbę $\mathbb{E}|X|^p$ nazywamy p -tym momentem zmiennej X .

Twierdzenie 7.4 (Nierówność Czebyszewa). *Założmy, że X jest zmienną losową oraz $f : [0, \infty) \rightarrow [0, \infty)$ jest funkcją niemalejącą taką, że $f(x) > 0$ dla $x > 0$. Wówczas dla dowolnej liczby $\lambda > 0$,*

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{\mathbb{E}f(|X|)}{f(\lambda)}.$$

W szczególności, biorąc $f(x) = x^p$, $p > 0$, dostajemy

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

Dowód. Mamy

$$\mathbb{E}f(|X|) \geq \mathbb{E}f(|X|)1_{\{|X| \geq \lambda\}} \geq \mathbb{E}f(\lambda)1_{\{|X| \geq \lambda\}} = f(\lambda)\mathbb{P}(|X| \geq \lambda). \quad \square$$

Definicja 7.3. Założmy, że X jest jednowymiarową zmienną losową całkowalną z kwadratem (tzn. $X \in L^2$). Liczbę $\text{Var}X = \mathbb{E}(X - \mathbb{E}X)^2$ nazywamy wariancją zmiennej X .

Jak łatwo sprawdzić, wariancja posiada następujące własności. Przy założeniu, że $X \in L^2$, mamy:

- (a) $\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$,
- (b) $\text{Var}X \geq 0$, przy czym $\text{Var}X = 0$ wtedy i tylko wtedy, gdy X ma rozkład jednopunktowy.
- (c) $\text{Var}(aX + b) = a^2 \text{Var}X$ dla dowolnych liczb rzeczywistych a, b .
- (d) Z nierówności Czebyszewa, dla dowolnej liczby $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq \lambda) \leq \frac{\text{Var}X}{\lambda^2}.$$

Odnotujmy kolejny ogólny fakt.

Twierdzenie 7.5 (O zamianie zmiennych). *Założmy, że X jest d -wymiarową zmienną losową na $(\Omega, \mathcal{F}, \mathbb{P})$, a $f : \mathbb{R}^d \rightarrow \mathbb{R}$ jest funkcją borelowską. Wówczas*

$$\mathbb{E}f(X) = \int_{\mathbb{R}^d} f(x)P_X(dx),$$

o ile jedna z tych całek istnieje.

Dowód. Stosujemy metodę komplikacji funkcji.

(i) Najpierw założmy, że f jest funkcją charakterystyczną pewnego zbioru B : $f = 1_B$. Wówczas dowodzona tożsamość przyjmuje postać $\mathbb{P}(X \in B) = P_X(B)$, która oczywiście jest prawdziwa.

(ii) Jeśli f jest funkcją prostą, tzn. kombinacją liniową funkcji charakterystycznych, to badana równość zachodzi, gdyż jej obie strony zależą od f w sposób liniowy.

(iii) Załóżmy, że $f \geq 0$. Wówczas f jest granicą punktową pewnego niemalejącego ciągu $(f_n)_{n \geq 0}$ nieujemnych funkcji prostych. Na mocy (ii), mamy

$$\mathbb{E}f_n(X) = \int_{\mathbb{R}^d} f_n(x)P_X(dx), \quad n = 1, 2, \dots,$$

a więc wystarczy zbiec z $n \rightarrow \infty$ oraz skorzystać z twierdzenia Lebesgue'a o monotonicznym przejściu do granicy.

(iv) Jeśli f jest dowolna, to rozbijamy ją na różnicę dwóch nieujemnych funkcji borelowskich: $f = f_+ - f_- = f1_{\{f \geq 0\}} + f1_{\{f < 0\}}$, stosujemy (iii) do funkcji f_+ i f_- , a następnie odejmujemy stronami uzyskane dwie tożsamości. Stąd teza. \square

Z powyższego faktu wynikają następujące

Wnioski:

1) Jeśli X jest zmienną losową, to $\mathbb{E}X = \int_{\mathbb{R}} xP_X(dx)$ oraz

$$\text{Var } X = \int_{\mathbb{R}} (x - \mathbb{E}X)^2 P_X(dx) = \int_{\mathbb{R}} x^2 P_X(dx) - (\mathbb{E}X)^2$$

(o ile te wielkości istnieją).

2) Wartość oczekiwana i wariancja zależą tylko od rozkładu.

Jak łatwo widać z powyższego twierdzenia, jeśli X jest d -wymiarową zmienną o rozkładzie dyskretnym, a $f : \mathbb{R}^d \rightarrow \mathbb{R}$ jest funkcją borelowską, to

$$\mathbb{E}f(X) = \int_{\mathbb{R}^d} f(x)P_X(dx) = \sum_{x \in S_X} f(x)P_X(\{x\}) = \sum_{x \in S_X} f(x)\mathbb{P}(X = x),$$

o ile wartość oczekiwana istnieje. Tak więc, w szczególności, dla $d = 1$ mamy

$$\mathbb{E}X = \sum_{x \in S_X} xP_X(\{x\}) = \sum_{x \in S_X} x\mathbb{P}(X = x),$$

$$\text{Var } X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \sum_{x \in S_X} x^2 P_X(\{x\}) - (\mathbb{E}X)^2 = \sum_{x \in S_X} x^2 \mathbb{P}(X = x) - (\mathbb{E}X)^2.$$

W przypadku gdy zmienna ma rozkład ciągły, jej parametry wyznaczamy przy użyciu następującego faktu.

Twierdzenie 7.6. *Założmy, że d -wymiarowa zmienna losowa X ma rozkład z gęstością g . Wówczas dla dowolnej funkcji borelowskiej $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mamy*

$$\mathbb{E}f(X) = \int_{\mathbb{R}^d} f(x)P_X(dx) = \int_{\mathbb{R}^d} f(x)g(x)dx,$$

o ile wartość oczekiwana istnieje.

Dowód. Tak jak wyżej, stosujemy metodę komplikacji funkcji. \square

Wobec tego, jeśli X jest jednowymiarową zmienną losową o rozkładzie z gęstością g , to

$$\mathbb{E}X = \int_{\mathbb{R}} xg(x)dx, \quad \text{Var } X = \int_{\mathbb{R}} x^2g(x)dx - (\mathbb{E}X)^2.$$

Przykłady:

1) Załóżmy, że $P_X = \delta_a$, $a \in \mathbb{R}$. Wówczas $\mathbb{E}X = a \cdot 1 = a$, $\text{Var } X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = a^2 - a^2 = 0$.

2) Przypuśćmy, że P_X to rozkład dwupunktowy, skupiony w $\{a, b\}$, taki że $P_X(\{a\}) = p$, $P_X(\{b\}) = 1 - p$, $0 < p < 1$. Wówczas

$$\mathbb{E}X = a \cdot p + b \cdot (1 - p),$$

$$\text{Var } X = a^2 \cdot p + b^2 \cdot (1 - p) - (ap + b(1 - p))^2 = (a - b)^2 p(1 - p).$$

3) Załóżmy teraz, że $P_X = B(n, p)$: $P_X(\{k\}) = \binom{n}{k} p^k (1 - p)^{n - k}$, $k = 0, 1, \dots, n$. Wówczas, jak za chwilę policzymy,

$$\mathbb{E}X = np \quad \text{oraz} \quad \text{Var } X = np(1 - p).$$

Podójście wprost z definicji jest niewygodne: na przykład, mamy

$$\mathbb{E}X = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n - k}$$

i trzeba „zwinąć” tę sumę. Aby uniknąć rachunków, rozważmy niezależne zmienne losowe X_1, X_2, \dots, X_n o tym samym rozkładzie dwupunktowym zadany przez $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Wówczas, jak już wiemy, $X_1 + X_2 + \dots + X_n$ ma rozkład $B(n, p)$, a zatem, z liniowości wartości oczekiwanej,

$$\mathbb{E}X = \mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}X_1 + \mathbb{E}X_2 + \dots + \mathbb{E}X_n = np.$$

Ponadto,

$$\mathbb{E}X^2 = \mathbb{E}(X_1 + X_2 + \dots + X_n)^2 = \sum_{k=1}^n \mathbb{E}X_k^2 + 2 \sum_{k < \ell} \mathbb{E}X_k X_\ell.$$

Dla dowolnych różnych k, ℓ , zmienna $X_k X_\ell$ ma rozkład dwupunktowy skoncentrowany na $\{0, 1\}$, przy czym

$$\mathbb{P}(X_k X_\ell = 1) = \mathbb{P}(X_k = 1, X_\ell = 1) = \mathbb{P}(X_k = 1) \mathbb{P}(X_\ell = 1) = p^2,$$

na mocy niezależności X_k oraz X_ℓ . Zatem

$$\mathbb{E}X^2 = n \cdot p + 2 \binom{n}{2} p^2 = np + n(n - 1)p^2$$

i w konsekwencji,

$$\text{Var } X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = np - np^2 = np(1 - p).$$

4) Załóżmy następnie, że zmienna X ma rozkład Poissona z parametrem $\lambda > 0$: $P_X(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, 2, \dots$. Wówczas

$$\mathbb{E}X = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

Podobnie obliczamy, iż

$$\text{Var } X = \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda} \frac{\lambda^k}{k!} - (\mathbb{E}X)^2 = \lambda.$$

5) Załóżmy, że $P_X = \mathcal{U}([a, b])$: $g(x) = \frac{1}{b-a} 1_{[a,b]}(x)$. Wtedy

$$\mathbb{E}X = \int_{\mathbb{R}} xg(x)dx = \frac{1}{b-a} \int_a^b xdx = \frac{a+b}{2}$$

oraz

$$\text{Var } X = \frac{1}{b-a} \int_a^b x^2 dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

6) Dalej, przypuśćmy, że $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, tzn. X ma rozkład z gęstością $g(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x)$. Całkując przez części, dostajemy

$$\mathbb{E}X = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = -xe^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$$

oraz, wykonując analogiczne obliczenia, otrzymujemy $\text{Var } X = \frac{1}{\lambda^2}$.

7) Wreszcie, przypuśćmy, że $P_X = N(m, \sigma^2)$, gdzie $m \in \mathbb{R}$ oraz $\sigma > 0$. Wówczas gęstość X dana jest wzorem

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right].$$

Dokonyjąc podstawienia $y = (x-m)/\sigma$, obliczamy, iż

$$\begin{aligned} \mathbb{E}X &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} x \cdot \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\sigma y + m) e^{-y^2/2} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} y e^{-y^2/2} dy + \frac{m}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} dy = m. \end{aligned}$$

Ponadto, ponownie stosując powyższe podstawienie, otrzymujemy

$$\begin{aligned} \text{Var } X &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (x-m)^2 e^{-(x-m)^2/(2\sigma^2)} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} y^2 e^{-y^2/2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} (-ye^{-y^2/2}) \Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} dy = \sigma^2. \end{aligned}$$

Podkreślmy: zatem parametry m i σ^2 pojawiające się w oznaczeniu rozkładu normalnego, to odpowiednio jego średnia i wariancja.

8) Warto tu jeszcze podać jeden przykład. Załóżmy, że X ma rozkład Cauchy'ego, tzn. rozkład z gęstością

$$g(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Wówczas zmienna X nie jest całkowna: mamy

$$\mathbb{E}|X| = \int_{\mathbb{R}} |x| \cdot \frac{1}{\pi} \frac{1}{1+x^2} dx = \infty.$$

Co więcej, wartość oczekiwana X nie istnieje: mamy

$$\mathbb{E}X^+ = \int_{\mathbb{R}} x^+ g(x) dx = \int_0^{\infty} \frac{1}{\pi} \frac{x}{1+x^2} dx = \infty$$

i podobnie $\mathbb{E}X^- = \infty$.

Przechodzimy teraz do związków wartości oczekiwanej i wariancji z niezależnością zmiennych.

Twierdzenie 7.7. *Załóżmy, że X_1, X_2, \dots, X_n to całkowalne i niezależne zmienne losowe. Wówczas zmienna $X_1 X_2 \dots X_n$ jest całkowalna i zachodzi równość*

$$\mathbb{E}X_1 X_2 \dots X_n = \mathbb{E}X_1 \mathbb{E}X_2 \dots \mathbb{E}X_n.$$

Dowód. Wiemy, że $P_{(X_1, X_2, \dots, X_n)} = P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}$. Wobec tego, korzystając z twierdzenia o zamianie zmiennych,

$$\begin{aligned} \mathbb{E}|X_1 X_2 \dots X_n| &= \int_{\mathbb{R}^n} |x_1 x_2 \dots x_n| P_{(X_1, \dots, X_n)}(dx_1 \dots dx_n) \\ &= \prod_{i=1}^n \int_{\mathbb{R}} |x_i| P_{X_i}(dx_i) < \infty, \end{aligned}$$

a więc $X_1 X_2 \dots X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Wystarczy teraz powtórzyć powyższy ciąg równości bez modułów (który ma sens, gdyż, jak właśnie udowodniliśmy, wszystkie wartości oczekiwane istnieją). \square

Uwaga: Twierdzenie odwrotne nie zachodzi. Przykładowo, weźmy zmienne η_1, η_2 o tym samym rozkładzie całkowalnym z kwadratem i połóżmy $X_1 = \eta_1 + \eta_2, X_2 = \eta_1 - \eta_2$. Wówczas $\mathbb{E}X_2 = 0$, a więc $\mathbb{E}X_1 \mathbb{E}X_2 = 0$; ponadto, $\mathbb{E}X_1 X_2 = \mathbb{E}\eta_1^2 - \mathbb{E}\eta_2^2 = 0$, na mocy równości rozkładów. Na ogół zmienne X_1 oraz X_2 nie są jednak niezależne: przykładowo, rozważmy dwukrotny rzut kostką i niech η_i oznacza liczbę oczek w i -tym rzucie, $i = 1, 2$. Wówczas X_1, X_2 są zależne - mają tę samą parzystość.

Przechodzimy do sytuacji wielowymiarowej.

Definicja 7.4. *Załóżmy, że $X = (X_1, X_2, \dots, X_d)$ jest d -wymiarową zmienną losową o całkowalnych współrzędnych (tzn. $\mathbb{E}|X_i| < \infty$ dla $i = 1, 2, \dots, n$). Wartością oczekiwaną X nazywamy wektor $(\mathbb{E}X_1, \mathbb{E}X_2, \dots, \mathbb{E}X_d)$.*

Uwagi:

1) Jeśli X, Y są d -wymiarowymi zmiennymi losowymi mającymi wartość oczekiwaną oraz $\alpha, \beta \in \mathbb{R}$, to $\alpha X + \beta Y$ także posiada wartość oczekiwaną.

2) Zmienna d -wymiarowa X ma wartość oczekiwaną wtedy i tylko wtedy, gdy $\mathbb{E}|X| < \infty$ (gdzie $|\cdot|$ oznacza tu normę euklidesową). Wynika to natychmiast z oszacowania

$$|X_j| \leq |X| \leq \sum_{i=1}^d |X_i|, \quad j = 1, 2, \dots, d.$$

3) Jeśli d -wymiarowa zmienna X ma skończoną wartość oczekiwaną, to $|\mathbb{E}X| \leq \mathbb{E}|X|$. Istotnie, dla dowolnego wektora $a \in \mathbb{R}^d$ o długości 1 mamy

$$\langle \mathbb{E}X, a \rangle = \sum_{j=1}^d \mathbb{E}X_j \cdot a_j = \mathbb{E}\langle X, a \rangle \leq \mathbb{E}|X| |a| = \mathbb{E}|X|$$

i biorąc supremum po a (bądź, alternatywnie, kładąc $a = \mathbb{E}X/|\mathbb{E}X|$), dostajemy żadaną nierówność.

Definicja 7.5. Załóżmy, że X_1, X_2 są zmiennymi losowymi całkowalnymi z kwadratem. Liczbę

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

nazywamy *kowariancją* zmiennych X i Y . W przypadku gdy $\text{Cov}(X, Y) = 0$, mówimy, że zmienne X, Y są *nieskorelowane*.

Jak łatwo sprawdzić, kowariancja posiada następujące własności:

- (a) Przede wszystkim, jest ona dobrze określona, na mocy nierówności Schwarz'a.
- (b) Zachodzi równość $\text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$.
- (c) Dla dowolnej zmiennej $X \in L^2$, $\text{Cov}(X, X) = \text{Var} X$.
- (d) Zachodzi równość $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- (e) Kowariancja jest operatorem dwuliniowym: jeśli $X, Y, Z \in L^2$, to

$$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z).$$

Ponadto, jeśli $X \in L^2$ oraz $a \in \mathbb{R}$, to $\text{Cov}(X, a) = 0$.

Uwaga: Powyższe rozważania pokazują, że jeśli $X, Y \in L^2$ są niezależne, to są nieskorelowane, ale nie na odwrót.

Twierdzenie 7.8. Zmienne X_1, X_2, \dots, X_n są całkowalne z kwadratem. Wówczas

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{k=1}^n \text{Var} X_k + 2 \sum_{k < \ell} \text{Cov}(X_k, X_\ell).$$

W szczególności, jeśli zmienne X_1, X_2, \dots, X_n są nieskorelowane, to

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var} X_1 + \text{Var} X_2 + \dots + \text{Var} X_n.$$

Dowód. Przekształcamy:

$$\begin{aligned} \text{Var}(X_1 + X_2 + \dots + X_n) &= \mathbb{E} \left[\sum_{j=1}^n X_j - \mathbb{E} \left(\sum_{j=1}^n X_j \right) \right]^2 \\ &= \mathbb{E} \left[\sum_{j=1}^n (X_j - \mathbb{E}X_j) \right]^2 \\ &= \sum_{j=1}^n \mathbb{E}(X_j - \mathbb{E}X_j)^2 + 2 \sum_{i < j} \mathbb{E}(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j) \\ &= \sum_{j=1}^n \text{Var} X_j + 2 \sum_{i < j} \text{Cov}(X_i, X_j). \end{aligned}$$

Dowodzi to pierwszej części twierdzenia. Jeśli teraz zmienne są nieskorelowane, to wszystkie kowariancje pojawiające się w drugiej sumie się zerują. Stąd teza. \square

Porównując przypadek jedno- i wielowymiarowy widzimy, iż wartość oczekiwana jednowymiarowej zmiennej losowej jest liczbą, a wartością oczekiwaną wielowymiarowej zmiennej jest wektor. Powstaje naturalne pytanie dotyczące uogólnienia wariancji na przypadek wielowymiarowy. Okazuje się, iż tym uogólnieniem jest tzw. macierz kowariancji.

Definicja 7.6. Załóżmy, że $X = (X_1, X_2, \dots, X_d)$ jest d -wymiarową zmienną losową o współrzędnych całkowalnych z kwadratem. Macierz

$$\Lambda = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_d) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \dots & \text{Cov}(X_2, X_d) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(X_d, X_1) & \text{Cov}(X_d, X_2) & \dots & \text{Cov}(X_d, X_d) \end{bmatrix}$$

nazywamy macierzą kowariancji zmiennej X .

Uwaga: Wartość oczekiwana i macierz kowariancji zmiennej losowej d -wymiarowej zależą tylko od rozkładu.

Twierdzenie 7.9 (Własności macierzy kowariancji). *Macierz kowariancji zmiennej $X = (X_1, X_2, \dots, X_d)$ jest symetryczna i nieujemnie określona.*

Dowód. Symetryczność wynika wprost z własności (d) kowariancji. Aby udowodnić nieujemną określoność, niech $m_i = \mathbb{E}X_i$ oraz weźmy dowolny ciąg liczb rzeczywistych t_1, t_2, \dots, t_d . Rozważmy jednowymiarową zmienną losową $\eta = t_1X_1 + t_2X_2 + \dots + t_dX_d$, która jest całkowalna z kwadratem (gdyż własność tę mają też zmienne X_1, X_2, \dots, X_d). Mamy

$$\begin{aligned} 0 &\leq \text{Var } \eta \\ &= \mathbb{E} \left(\sum_{j=1}^d t_j (X_j - \mathbb{E}X_j) \right)^2 \\ &= \sum_{i,j=1}^d \mathbb{E} [t_i (X_i - \mathbb{E}X_i) \cdot t_j (X_j - \mathbb{E}X_j)] \\ &= \sum_{i,j=1}^d t_i t_j \text{Cov}(X_i, X_j), \end{aligned}$$

co oznacza tezę.

Poczyńmy jeszcze jedną przydatną obserwację. Przypuśćmy, iż macierz kowariancji nie jest *dodatnio* określona, tzn. dla pewnych t_1, t_2, \dots, t_d mamy

$$\sum_{i,j} t_i t_j \text{Cov}(X_i, X_j) = 0.$$

Oznacza to, iż $\eta = t_1X_1 + t_2X_2 + \dots + t_dX_d$ ma rozkład jednopunktowy, tzn. istnieje $c \in \mathbb{R}$ takie, że

$$\mathbb{P}(t_1X_1 + t_2X_2 + \dots + t_dX_d = c) = 1,$$

a zatem z prawdopodobieństwem 1 zmienna X przyjmuje wartości w pewnej $d-1$ -wymiarowej podprzestrzeni afinicznej. \square

Odnotujmy pożyteczny

Wniosek: Zmienna $X = (X_1, X_2, \dots, X_d)$ ma parami nieskorelowane współrzędne wtedy i tylko wtedy, gdy macierz kowariancji jest diagonalna. W szczególności, jeśli współrzędne X_1, X_2, \dots, X_d są niezależne, to X ma diagonalną macierz kowariancji (ale nie na odwrót!).

Przykład: Rozważmy wielowymiarowy rozkład normalny. Niech $m \in \mathbb{R}^d$, niech A będzie symetryczną, dodatnio określoną macierzą $d \times d$ oraz założmy, że $X = (X_1, X_2, \dots, X_d)$ ma rozkład z gęstością

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \exp \left[-\frac{1}{2} \langle A(x - m), (x - m) \rangle \right].$$

Twierdzenie 7.10. *Mamy $\mathbb{E}X = m$, a macierz kowariancji X jest równa A^{-1} .*

Pozostawiamy dowód tego twierdzenia jako ćwiczenie.

Twierdzenie 7.11. *Założmy, że $X = (X_1, X_2, \dots, X_d)$ ma d -wymiarowy rozkład normalny. Wówczas zmienne X_1, X_2, \dots, X_d są niezależne wtedy i tylko wtedy, gdy są nieskorelowane.*

Dowód. \Rightarrow W tę stronę implikacja zachodzi dla dowolnych zmiennych losowych.

\Leftarrow Jeśli współrzędne są nieskorelowane, to, jak wiemy, macierz kowariancji jest przekątniowa:

$$\Lambda = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ & \dots & \dots & \\ 0 & 0 & \dots & \sigma_d^2 \end{bmatrix}.$$

Zatem $A = \Lambda^{-1}$ także jest diagonalna i jej wyrazy na głównej przekątnej to $1/\sigma_1^2, 1/\sigma_2^2, \dots, 1/\sigma_d^2$. Wobec tego

$$\begin{aligned} g(x) &= \frac{\sqrt{\sigma_1^{-2} \sigma_2^{-2} \dots \sigma_d^{-2}}}{(2\pi)^{d/2}} \exp \left[-\frac{1}{2} \sum_{j=1}^d (x_j - m_j) \cdot \sigma_j^{-2} \right] \\ &= \prod_{j=1}^d \left(\frac{1}{\sqrt{2\pi} \sigma_j} e^{-(x_j - m_j)^2 / (2\sigma_j^2)} \right) \\ &= g_1(x_1) g_2(x_2) \dots g_d(x_d), \end{aligned}$$

gdzie g_j to gęstość rozkładu $\mathcal{N}(m_j, \sigma_j^2)$. Stąd niezależność. \square

Rozważmy teraz tzw. zagadnienie regresji liniowej, grające ważną rolę w statystyce. Problem możemy sformułować następująco. Założmy, że mamy zmienne losowe X, Y całkowalne z kwadratem i znamy ich łączny rozkład. Ponadto, przypuścmy, iż obserwujemy wartości zmiennej X , a zmienna Y jest trudniejsza - bądź niemożliwa - do zmierzenia. Powstaje więc interesujące zagadnienie optymalnego przybliżania zmiennej Y za pomocą zmiennej X . Oczywiście, musimy odpowiednio postawić ten problem; będziemy szukać optymalnego przybliżenia *liniowego*, tzn. postaci $aX + b$, $a, b \in \mathbb{R}$, a błąd będziemy mierzyć w sensie średniokwadratowym. Innymi słowy, szukamy stałych $a, b \in \mathbb{R}$, dla których wielkość $f(a, b) = \mathbb{E}(Y - aX - b)^2$ jest najmniejsza.

Aby rozwiązać ten problem, zauważmy, iż przy ustalonym a , funkcja $b \mapsto f(a, b)$ jest trójmianem kwadratowym, który przyjmuje swoją najmniejszą wartość w punkcie $\mathbb{E}(Y - aX)$. Wystarczy więc wyznaczyć najmniejszą wartość funkcji

$$h(a) = f(a, \mathbb{E}(Y - aX)) = \mathbb{E}(Y - \mathbb{E}Y - a(X - \mathbb{E}X))^2 = a^2 \text{Var}X - 2a \text{Cov}(X, Y) + \text{Var}Y.$$

Jeśli zmienna X jest stała p.n. (czyli $\text{Var}X=0$), to wówczas h jest funkcją stałą i widać, że optymalnym liniowym estymatorem zmiennej Y jest jej średnia: $aX + b =$

$aX + (\mathbb{E}Y - a\mathbb{E}X) = \mathbb{E}Y$. Jeśli zaś $\text{Var}X \neq 0$, to h jest trójmianem kwadratowym zmiennej a , przyjmującym swoją najmniejszą wartość w punkcie

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}X}$$

i wówczas

$$b = \mathbb{E}Y - \mathbb{E}X \cdot \frac{\text{Cov}(X, Y)}{\text{Var}X}.$$

Uwagi:

1) Widać, że do powyższych obliczeń nie potrzebowaliśmy całej wiedzy o rozkładzie łącznym zmiennych (X, Y) . Wystarczy nam znajomość średnich i wariancji zmiennych X, Y oraz ich kowariancji.

2) Załóżmy, że wariancje X oraz Y są niezerowe. Dla powyższych (optymalnych) a, b obliczamy, iż

$$f(a, b) = \text{Var} X (1 - \rho^2(X, Y)),$$

gdzie

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var} X \text{Var} Y}}$$

to tzw. *współczynnik korelacji*. Współczynnik ten posiada następujące własności:

- (a) Zachodzi nierówność $-1 \leq \rho(X, Y) \leq 1$,
- (b) Mamy $\rho(X, Y) = \rho(Y, X)$ oraz, dla dowolnych $a, b \in \mathbb{R}$, $\rho(aX + b, Y) = \rho(X, Y)$.
- (c) Jeśli $|\rho(X, Y)| = 1$, to $X = \alpha Y + \beta$ dla pewnych $\alpha, \beta \in \mathbb{R}$; innymi słowy, między X a Y jest zależność liniowa.
- (d) Równość $\rho(X, Y) = 0$ zachodzi wtedy i tylko wtedy, gdy X, Y są nieskorelowane. Wówczas najlepszym przybliżeniem Y jest $\mathbb{E}X$.

8. ZADANIA

1. Dana jest zmienna losowa X taka, że $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{4}$, $\mathbb{P}(X = -3) = \frac{1}{2}$. Obliczyć $\mathbb{E}X$, $\mathbb{E}\frac{1}{X+2}$, $\mathbb{E}\cos(\pi X)$ oraz $\text{Var } X$.

2. Zmienna losowa X ma rozkład Poissona z parametrem 2. Obliczyć $\mathbb{E}6^X$.

3. Zmienna losowa X ma rozkład z gęstością

$$g(x) = \frac{3}{8}x^2 1_{[0,2]}.$$

Obliczyć $\mathbb{E}X$, $\mathbb{E}\frac{1}{1+x^3}$ oraz $\text{Var } X^2$.

4. Zmienna losowa X ma rozkład z dystrybuantą

$$F(t) = \begin{cases} 0 & \text{jeśli } t < 0, \\ t/2 & \text{jeśli } 0 \leq t < 1, \\ 3/4 & \text{jeśli } 1 \leq t < 5, \\ 1 & \text{jeśli } t \geq 5. \end{cases}$$

Wyznaczyć $\mathbb{E}(2X + 1)$.

5. W urnie znajduje się 50 białych kul. Losujemy ze zwracaniem po jednej kuli, przy czym wyciągniętą kulę malujemy na czerwono, jeśli jest biała. Niech X oznacza liczbę czerwonych kul w urnie po 20 losowaniach. Wyznaczyć $\mathbb{E}X$ oraz $\text{Var } X$.

6. Każdy bok i każdą przekątną sześciokąta foremnego malujemy losowo na jeden z trzech kolorów. Wybór każdego koloru jest jednakowo prawdopodobny, kolorowania różnych odcinków są niezależne. Niech X oznacza liczbę jednobarwnych trójkątów o wierzchołkach będących wierzchołkami sześciokąta. Obliczyć $\mathbb{E}X$.

7. Rzucamy kostką aż do momentu, gdy wyrzucimy wszystkie liczby oczek. Obliczyć wartość oczekiwaną i wariancję liczby rzutów.

8. Udowodnić, że dla dowolnej zmiennej losowej nieujemnej X oraz $p > 0$ zachodzi wzór

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

Wynioskować stąd, że jeśli zmienna X ma rozkład dyskretny skoncentrowany na liczbach całkowitych nieujemnych, to

$$\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

9. Liczby $1, 2, \dots, n$ ustawiono losowo w ciąg (a_1, a_2, \dots, a_n) . Niech N oznacza największą taką liczbę, że $a_k > a_{k-1}$ dla $k \leq N$. Obliczyć $\mathbb{E}N$.

10. Dany jest ciąg niezależnych zmiennych losowych X_0, X_1, X_2, \dots o tym samym rozkładzie posiadającym ciągłą dystrybuantę. Niech $\eta = \inf\{n : X_n > X_0\}$. Wyznaczyć rozkład zmiennej η oraz obliczyć $\mathbb{E}\eta$.

11. Kij o długości 1 złamano w punkcie wybranym losowo, z prawdopodobieństwem rozłożonym równomiernie. Obliczyć wartość oczekiwaną stosunku

- długości kawałka lewego do długości kawałka prawego.
- długości kawałka krótszego do długości kawałka dłuższego.

12. Zmienne losowe X, Y spełniają warunki $\text{Var}X = 3, \text{Cov}(X, Y) = -1, \text{Var}Y = 2$. Obliczyć $\text{Var}(4X - 3Y)$ oraz $\text{Cov}(5X - Y, 2X + Y)$.

13. Zmienna losowa X ma wariancję $\sigma^2 < \infty$. Udowodnić, że

$$\mathbb{P}(|X - \mathbb{E}X| > 3\sigma) \leq \frac{1}{9}.$$

14. Zmienne losowe $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ są niezależne i mają ten sam rozkład $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2, k = 1, 2, \dots, n$. Niech a_1, a_2, \dots, a_n będzie ciągiem liczb rzeczywistych i $A = (\sum_{k=1}^n a_k^2)^{1/2}$. Udowodnić, że

$$\mathbb{P}\left(\left|\sum_{k=1}^n a_k \varepsilon_k\right| > t\right) \leq 2 \exp(-t^2/2A^2).$$

15. Zmienne $\varepsilon_1, \varepsilon_2, \dots$ są niezależne i mają ten sam rozkład $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2, k = 1, 2, \dots$. Niech $S_n = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n, n = 1, 2, \dots$. Udowodnić, że

$$\limsup \frac{S_n}{\sqrt{2n \log n}} \leq 1 \quad \text{p.n.}$$

oraz

$$\liminf \frac{S_n}{\sqrt{2n \log n}} \geq -1 \quad \text{p.n..}$$

16. Zmienna losowa X ma następującą własność: dla $n = 1, 2, \dots$ mamy

$$\mathbb{E}|X|^n \leq \binom{2n}{n}.$$

Udowodnić, że $X \in L^\infty$ (tzn. istnieje taka liczba M , że $\mathbb{P}(|X| \leq M) = 1$).

17. Zmienna losowa X ma rozkład normalny w \mathbb{R}^d , o średniej m i macierzy kowariancji Λ . Niech T będzie przekształceniem afinicznym \mathbb{R}^d na $\mathbb{R}^k, k \leq d$. Udowodnić, że TX ma rozkład normalny w \mathbb{R}^k . Wyznaczyć jego średnią oraz macierz kowariancji.

18. Zmienna losowa X ma d -wymiarowy rozkład normalny o gęstości

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}\langle A(x - m), (x - m) \rangle\right].$$

Udowodnić, że $\mathbb{E}X = m$ oraz $\Lambda = A^{-1}$ (Λ oznacza tu macierz kowariancji X).

19. Zmienna losowa (X, Y) ma dwuwymiarowy rozkład normalny o średniej $(0, 0)$ i macierzy kowariancji

$$\Lambda = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

a) Napisać gęstość zmiennej (X, Y) .

b) Wyznaczyć rozkład zmiennej $X + 3Y$.

c) Wyznaczyć taką liczbę $a \in \mathbb{R}$, by zmienne $X + Y, X + aY$ były niezależne.

20. Nadajnik wysyła sygnał ξ , a odbiornik odbiera sygnał $\eta = a\xi + \zeta$, gdzie $a \in \mathbb{R}_+$ jest współczynnikiem wzmocnienia, zaś ζ jest zakłóceniem. Zakładamy, że ξ i ζ są niezależnymi zmiennymi losowymi, przy czym $\mathbb{E}\xi = m, \text{Var}\xi = 1, \mathbb{E}\zeta = 0, \text{Var}\zeta = \sigma^2$. Wyznaczyć współczynnik korelacji ξ i η oraz regresję liniową ξ względem η (tzn. najlepsze liniowe przybliżenie ξ za pomocą η).

9. RÓŻNE RODZAJE ZBIEŻNOŚCI ZMIENNYCH LOSOWYCH

Zajmiemy się teraz zachowaniem granicznym ciągów zmiennych losowych. Zajmiemy jednak od pewnego pożytecznego faktu.

Definicja 9.1. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz $\mathcal{F}_1, \mathcal{F}_2, \dots \subseteq \mathcal{F}$ jest ciągiem σ -ciał. Wówczas σ -ciało

$$\bigcap_{n=1}^{\infty} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \dots)$$

nazywamy σ -ciałem resztkowym.

Przykład: Załóżmy, że X_1, X_2, \dots jest ciągiem zmiennych losowych i niech $\mathcal{F}_n = \sigma(X_n)$ będzie σ -ciałem generowanym przez zmienną X_n , $n = 1, 2, \dots$. Wówczas każde ze zdarzeń

$$\{(X_n)_{n \geq 1} \text{ jest zbieżny}\}, \{\sup_n |X_n| < \infty\}, \left\{ \sum_{n=1}^{\infty} X_n \text{ jest zbieżny} \right\}$$

należy do σ -ciała resztkowego.

Twierdzenie 9.1 (Prawo 0–1 Kolmogorowa). *Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz σ -ciała $\mathcal{F}_1, \mathcal{F}_2, \dots \subseteq \mathcal{F}$ są niezależne. Wówczas dla każdego $A \in \bigcap_{n=1}^{\infty} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots)$ mamy $\mathbb{P}(A) = 0$ lub $\mathbb{P}(A) = 1$.*

Lemat 9.1. *Załóżmy, że $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \dots \subseteq \mathcal{F}$ jest wstępującym ciągiem σ -ciał oraz niech $\mathcal{G} = \sigma(\mathcal{G}_1, \mathcal{G}_2, \dots)$. Wówczas dla dowolnego $A \in \mathcal{G}$ istnieje ciąg $(A_n)_{n \geq 1}$ takich zdarzeń, że $A_n \in \mathcal{G}_n$ dla każdego n oraz $\lim_{n \rightarrow \infty} \mathbb{P}(A \Delta A_n) = 0$ (Δ oznacza tu różnicę symetryczną zbiorów).*

Dowód. Wprowadźmy klasę zbiorów

$$\mathcal{K} = \{A \in \mathcal{G} : \text{istnieje ciąg } (A_n)_{n \geq 1} \text{ jak w sformułowaniu lematu}\}.$$

Oczywiście $\bigcup_{n \geq 1} \mathcal{G}_n \subset \mathcal{K}$, ponadto suma po lewej stronie zawierania jest π -układem. Wystarczy więc wykazać, że \mathcal{K} jest λ -układem. Sprawdzamy:

(i) $\Omega \in \mathcal{K}$ - jest to oczywiste.

(ii) Załóżmy, że $A, B \in \mathcal{K}$ spełniają warunek $A \subseteq B$ i niech $(A_n)_{n \geq 1}, (B_n)_{n \geq 1}$ będą odpowiednimi ciągami przybliżającymi. Wykażemy, zgodnie z intuicją, że ciąg $(B_n \setminus A_n)_{n \geq 1}$ przybliża $B \setminus A$. Oczywiście $B_n \setminus A_n \in \mathcal{G}_n$ dla każdego n . Ponadto, korzystając z tożsamości $\mathbb{P}(A \Delta B) = \mathbb{E}|1_A - 1_B|$, mamy

$$\begin{aligned} \mathbb{P}((B_n \setminus A_n) \Delta (B \setminus A)) &= \mathbb{E}|1_{B_n \setminus A_n} - 1_{B \setminus A}| \\ &= \mathbb{E}|1_{B_n} - 1_{A_n \cap B_n} - 1_B + 1_A| \\ &\leq \mathbb{E}|1_{B_n} - 1_B| + \mathbb{E}|1_A - 1_{A_n} 1_{B_n}|. \end{aligned}$$

Pierwszy składnik jest równy $\mathbb{P}(B_n \Delta B)$, a więc zbiega do 0 gdy $n \rightarrow \infty$. Ponadto,

$$\begin{aligned} \mathbb{E}|1_A - 1_{A_n} 1_{B_n}| &= \mathbb{E}|1_A - (1_{A_n} - 1_A + 1_A)(1_{B_n} - 1_B + 1_B)| \\ &\leq \mathbb{E}|1_A - 1_A 1_B| + \mathbb{E}|1_A| |1_{B_n} - 1_B| \\ &\quad + \mathbb{E}|1_B| |1_{A_n} - 1_A| + \mathbb{E}|1_{A_n} - 1_A| |1_{B_n} - 1_B| \\ &\leq 0 + \mathbb{P}(B \Delta B_n) + \mathbb{P}(A \Delta A_n) + 2\mathbb{P}(A \Delta A_n) \end{aligned}$$

także zbiega do zera gdy n dąży do nieskończoności. Udowodniliśmy zatem, że $B \setminus A \in \mathcal{K}$.

(iii) Jeśli A_1, A_2, \dots jest wstępującym ciągiem elementów z \mathcal{K} , to $\bigcup_{n=1}^{\infty} A_n$ także należy do \mathcal{K} . Istotnie, sumę $\bigcup_{n=1}^{\infty} A_n$ możemy dowolnie dokładnie przybliżyć za pomocą sumy częściowej $\bigcup_{n=1}^N A_n = A_N$, a następnie zbiór A_N przybliżamy odpowiednim ciągiem zdarzeń z σ -ciała \mathcal{G}_n . \square

Dowód twierdzenia Kolmogorowa. Rozważmy wstępujący ciąg σ -ciał dany przez

$$\mathcal{G}_n = \sigma(\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n), \quad n = 1, 2, \dots$$

Weźmy zdarzenie A należące do σ -ciała resztkowego. Oczywiście należy ono także do σ -ciała

$$\sigma(\mathcal{F}_1, \mathcal{F}_2, \dots) = \sigma(\mathcal{G}_1, \mathcal{G}_2, \dots),$$

a zatem, na mocy powyższego lematu, możemy wskazać ciąg $(A_n)_{n \geq 1}$ taki, że $A_n \in \mathcal{G}_n$ oraz $\mathbb{P}(A \Delta A_n) \rightarrow 0$ gdy $n \rightarrow \infty$. Ponieważ $\mathbb{P}(A \Delta A_n) = \mathbb{P}(A \setminus A_n) + \mathbb{P}(A_n \setminus A)$, to każde z tych dwóch prawdopodobieństw także zbiega do 0. Ale

$$\mathbb{P}(A \setminus A_n) = \mathbb{P}(A) - \mathbb{P}(A \cap A_n),$$

zatem $\mathbb{P}(A \cap A_n) \rightarrow \mathbb{P}(A)$. Ponadto,

$$\mathbb{P}(A_n \setminus A) = \mathbb{P}(A_n) - \mathbb{P}(A \cap A_n),$$

a więc, w połączeniu z poprzednią zbieżnością, $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$. Wreszcie, dla dowolnego n mamy

$$A \in \sigma(\mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \dots), \quad A_n \in \mathcal{G}_n,$$

i te σ -ciała są niezależne. Wobec tego A oraz A_n także są niezależne, $\mathbb{P}(A \cap A_n) = \mathbb{P}(A)\mathbb{P}(A_n)$, i lewa strona dąży do $\mathbb{P}(A)$, a prawa do $\mathbb{P}(A)^2$. Stąd $\mathbb{P}(A) \in \{0, 1\}$. \square

Przechodzimy teraz do zbieżności zmiennych losowych.

Definicja 9.2. Załóżmy, że $(X_n)_{n \geq 1}$ jest ciągiem zmiennych losowych o wartościach w \mathbb{R}^d . Mówimy, że

(i) X_n zbiega do X prawie na pewno, jeśli $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$. Oznaczenie: $X_n \rightarrow X$ p.n.

(ii) ($p \geq 1, d = 1$) X_n zbiega do X w L^p jeśli $X_1, X_2, \dots \in L^p$ oraz $\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0$ (przypomnijmy: $\|\xi\|_p = (\mathbb{E}|\xi|^p)^{1/p}$ dla $p < \infty$, $\|\xi\|_{\infty} = \text{esssup}|\xi|$). Oznaczenie: $X_n \rightarrow X$ w L^p .

(iii) X_n zbiega do X według prawdopodobieństwa, jeśli dla każdego $\varepsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. Oznaczenie: $X_n \xrightarrow{\mathbb{P}} X$.

Twierdzenie 9.2. Jeśli $X_n \rightarrow X$ p.n., to $X_n \xrightarrow{\mathbb{P}} X$. Implikacja przeciwna nie zachodzi.

Dowód. Z definicji, ciąg $(X_n)_{n \geq 1}$ zbiega do X prawie na pewno, jeśli

$$\mathbb{P} \left(\bigcap_{\varepsilon > 0} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\} \right) = 1.$$

Jest to równoważne warunkowi, że dla każdego $\varepsilon > 0$,

$$\mathbb{P} \left(\bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\} \right) = 1.$$

Ale ciąg zdarzeń $\left(\bigcap_{n \geq N} \{|X_n - X| < \varepsilon\}\right)_{N \geq 1}$ jest wstępujący; na mocy twierdzenia o ciągłości, powyższa równość oznacza, iż dla dowolnego $\varepsilon > 0$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\bigcap_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\} \right) = 1.$$

Wobec tego tym bardziej

$$\lim_{N \rightarrow \infty} \mathbb{P}(\{\omega : |X_N(\omega) - X(\omega)| < \varepsilon\}) = 1,$$

czyli, po przejściu do zdarzenia przeciwnego, $\mathbb{P}(|X_N - X| \geq \varepsilon) \rightarrow 0$.

Aby udowodnić, że implikacja przeciwna nie zachodzi, rozważmy następujący przykład. Załóżmy, że przestrzeń probabilistyczna to przedział $[0, 1]$ wraz ze swoimi podzbiorami borelowskimi oraz miarą Lebesgue'a. Niech

$$\begin{aligned} X_1 &= 1_{[0,1)}, \\ X_2 &= 1_{[0,1/2)}, \quad X_3 = 1_{[1/2,1)}, \\ X_4 &= 1_{[0,1/4)}, \quad X_5 = 1_{[1/4,1/2)}, \quad X_6 = 1_{[1/2,3/4)}, \quad X_7 = 1_{[3/4,1)}, \\ &\dots \end{aligned}$$

Wówczas ciąg $(X_n)_{n \geq 0}$ zbiega do 0 według prawdopodobieństwa: dla dowolnego ε , $\mathbb{P}(|X_n - 0| > \varepsilon)$ jest potęgą dwójki z coraz mniejszym całkowitym wykładnikiem. Z drugiej strony, dla dowolnego $\omega \in [0, 1)$, liczbowy ciąg $(X_n(\omega))_{n \geq 1}$ nie jest zbieżny; jest to ciąg zawierający nieskończenie wiele zer oraz nieskończenie wiele jedynek. \square

Twierdzenie 9.3. *Jeśli $X_n \rightarrow X$ w L^p , to $X_n \xrightarrow{\mathbb{P}} X$. Implikacja w drugą stronę nie zachodzi.*

Dowód. Na mocy nierówności Czebyszewa, dla dowolnego ε mamy

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p} \xrightarrow{n \rightarrow \infty} 0.$$

Wykażemy, że implikacja w drugą stronę nie zachodzi. Rozpatrzmy tylko $p < \infty$, przypadek $p = \infty$ pozostawiamy czytelnikowi. Rozważmy przestrzeń probabilistyczną $([0, 1], \mathcal{B}(0, 1), |\cdot|)$ oraz ciąg $(X_n)_{n \geq 1}$ zmiennych zadanych wzorem

$$X_n(\omega) = n^{1/p} 1_{[0, 1/n]}(\omega).$$

Wówczas $X_n \xrightarrow{\mathbb{P}} 0$: dla dowolnego ε mamy

$$\mathbb{P}(|X_n - 0| > \varepsilon) \leq 1/n \rightarrow 0.$$

Zatem, gdyby ciąg $(X_n)_{n \geq 1}$ zbiegał w L^p , to do zmiennej skoncentrowanej w zerze (na mocy implikacji którą właśnie udowodniliśmy). Ale

$$\mathbb{E}|X_n - 0|^p = \mathbb{E}|X_n|^p = 1 \not\rightarrow 0. \quad \square$$

Twierdzenie 9.4. *Jeśli $p < p'$ oraz $X_n \rightarrow X$ w $L^{p'}$, to $X_n \rightarrow X$ w L^p .*

Dowód. Wynika to natychmiast z nierówności Höldera: mamy

$$\|X_n - X\|_p \leq \|X_n - X\|_{p'} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Twierdzenie 9.5. a) Ciąg $(X_n)_{n \geq 1}$ jest zbieżny według prawdopodobieństwa wtedy i tylko wtedy, gdy spełnia warunek Cauchy'ego według prawdopodobieństwa:

$$\forall \delta > 0 \forall \varepsilon > 0 \exists N \forall m, n \geq N \mathbb{P}(|X_n - X_m| > \varepsilon) < \delta.$$

b) Jeśli X_n zbiega do X według prawdopodobieństwa, to istnieje podciąg $(n_k)_{k \geq 1}$ taki, że ciąg $(X_{n_k})_{k \geq 1}$ zbiega p.n. do X .

Definicja 9.3. Załóżmy, że $\{X_i\}_{i \in \mathcal{I}}$ jest rodziną całkowalnych zmiennych losowych. Mówimy, że ta rodzina jest jednostajnie (jednakowo) całkowalna, jeśli

$$\sup_{i \in \mathcal{I}} \int_{\{|X_i| \geq r\}} |X_i| d\mathbb{P} \xrightarrow{r \rightarrow \infty} 0.$$

Przykłady:

1) Załóżmy, że istnieje nieujemna całkowalna zmienna η taka, że $|X_i| \leq \eta$ dla wszystkich $i \in \mathcal{I}$. Wówczas $\{X_i\}_{i \in \mathcal{I}}$ jest rodziną jednakowo całkowalną. Istotnie,

$$\sup_{i \in \mathcal{I}} \int_{\{|X_i| \geq r\}} |X_i| d\mathbb{P} \leq \int_{\{\eta \geq r\}} |\eta| d\mathbb{P} \xrightarrow{r \rightarrow \infty} 0,$$

na mocy twierdzenia Lebesgue'a o zmajoryzowanym przejściu do granicy.

2) Każda skończona rodzina zmiennych całkowalnych jest jednakowo całkowalna: wystarczy wykorzystać poprzedni przykład, biorąc $\eta = \sum_{i \in \mathcal{I}} |X_i|$.

3) Dowolna jednostajnie całkowalna rodzina zmiennych losowych, po dodaniu do niej skończonej liczby zmiennych całkowalnych, pozostaje jednostajnie całkowalna.

4) Rozważmy przestrzeń probabilistyczną $([0, 1], \mathcal{B}([0, 1]), |\cdot|)$ oraz ciąg zmiennych X_1, X_2, \dots zadanych przez $X_n(\omega) = n^2 1_{[0, 1/n]}(\omega)$. Rodzina ta nie jest jednostajnie całkowalna: mamy

$$\{|X_n| \geq m^2\} = \{n^2 1_{[0, 1/n]} \geq m^2\} = \begin{cases} \emptyset & \text{dla } n < m, \\ [0, 1/n] & \text{dla } n \geq m, \end{cases}$$

a więc

$$\int_{\{|X_n| \geq m^2\}} |X_n| d\mathbb{P} = \begin{cases} 0 & \text{dla } n < m, \\ n^2 \cdot 1/n & \text{dla } n \geq m, \end{cases}$$

a więc dla każdego r , $\sup_n \int_{\{|X_n| \geq r\}} |X_n| d\mathbb{P} = \infty$.

Z drugiej strony, rodzina zmiennych $\{Y_n\}_{n \geq 1} = \{\sqrt{n} 1_{[0, 1/n]}\}_{n \geq 1}$ jest jednostajnie całkowalna. Powtarzając powyższe rozumowanie widzimy, że

$$\int_{\{|X_n| \geq r\}} |X_n| d\mathbb{P} = \begin{cases} 0 & \text{dla } \sqrt{n} < r, \\ \sqrt{n} \cdot 1/n & \text{dla } \sqrt{n} \geq r. \end{cases}$$

Zatem dla ustalonego r ,

$$\sup_n \int_{\{|X_n| \geq r\}} |X_n| d\mathbb{P} \leq 1/r,$$

co zbiega do 0 gdy $r \rightarrow \infty$.

Udowodnimy teraz pewien równoważny warunek na jednakową całkowalność.

Twierdzenie 9.6. Rodzina $\{X_i\}_{i \in \mathcal{I}}$ jest jednakowo całkowalna wtedy i tylko wtedy, gdy zachodzą następujące dwa warunki:

$$1^\circ \sup_{i \in \mathcal{I}} \mathbb{E}|X_i| < \infty,$$

2° Dla każdego $\varepsilon > 0$ istnieje $\delta > 0$ taka, że jeśli zdarzenie A spełnia $\mathbb{P}(A) < \delta$, to

$$\int_A |X_i| d\mathbb{P} < \varepsilon, \quad i \in \mathcal{I}.$$

Dowód. \Rightarrow Zaczniemy od warunku 2°. Dla każdego $A \in \mathcal{F}$ oraz $i \in \mathcal{I}$ mamy

$$\int_A |X_i| d\mathbb{P} = \int_{A \cap \{|X_i| \geq r\}} |X_i| d\mathbb{P} + \int_{A \cap \{|X_i| < r\}} |X_i| d\mathbb{P} \leq \sup_{i \in \mathcal{I}} \int_{\{|X_i| \geq r\}} |X_i| d\mathbb{P} + r\mathbb{P}(A).$$

Zatem, przy ustalonym $\varepsilon > 0$, bierzemy r takie, by pierwszy składnik był mniejszy niż $\varepsilon/2$ (jest to możliwe na mocy definicji jednakowej całkowalności); następnie, bierzemy $\delta = \varepsilon/(2r)$: wówczas drugi składnik także jest mniejszy niż $\varepsilon/2$. Ponadto, biorąc wyżej $A = \Omega$, dostajemy, iż dla każdego r ,

$$\sup_{i \in \mathcal{I}} \mathbb{E}|X_i| \leq \sup_{i \in \mathcal{I}} \int_{\{|X_i| \geq r\}} |X_i| d\mathbb{P} + r < \infty,$$

co jest żądanym warunkiem 1°.

\Leftarrow Dla dowolnego $i \in \mathcal{I}$ mamy, z nierówności Czebyszewa oraz 1°,

$$\mathbb{P}(|X_i| \geq r) \leq \frac{\mathbb{E}|X_i|}{r} \leq \sup_{i \in \mathcal{I}} \mathbb{E}|X_i|/r < \infty.$$

Następnie, dla dowolnego $\varepsilon > 0$ dobieramy δ z warunku 2°. Powyższy rachunek daje, iż dla dostatecznie dużych r mamy $\sup_{i \in \mathcal{I}} \mathbb{P}(|X_i| \geq r) < \delta$, a zatem z 2°,

$$\sup_{i \in \mathcal{I}} \int_{\{|X_i| \geq r\}} |X_i| d\mathbb{P} < \varepsilon.$$

Oznacza to, iż jest spełniony warunek definiujący jednakową całkowalność. \square

Twierdzenie 9.7. Niech $p \geq 1$ będzie ustaloną liczbą. Ciąg $(X_n)_{n \geq 1}$ jest zbieżny w L^p wtedy i tylko wtedy, gdy jest zbieżny według prawdopodobieństwa oraz rodzina $\{|X_n|^p\}_{n \geq 1}$ jest jednostajnie całkowalna.

Dowód. \Rightarrow Zbieżność według prawdopodobieństwa mamy za darmo; pozostaje wykazać jednostajną całkowalność. Dla dowolnego $A \in \mathcal{F}$,

$$(*) \quad \left(\int_A |X_n|^p d\mathbb{P} \right)^{1/p} = \|X_n 1_A\|_p \leq \|X 1_A\|_p + \|(X - X_n) 1_A\|_p.$$

Dla $A = \Omega$, nierówność (*) daje $\|X_n\|_p \leq \|X\|_p + \|X_n - X\|_p$, a więc $\sup_n \|X_n\|_p < \infty$, co pociąga za sobą warunek 1° z poprzedniego twierdzenia. Aby dowieść 2°, ustalmy $\varepsilon > 0$. Z definicji zbieżności w L^p , istnieje N takie, że $\|X_n - X\|_p < \varepsilon/2$ dla $n \geq N$. Rodzina $\{|X|^p, |X_1 - X|^p, \dots, |X_N - X|^p\}$ jest skończona i zawiera całkowalne zmienne losowe, jest więc jednakowo całkowalna (por. Przykład 2 powyżej) i spełnia warunek 2°: istnieje δ taka, że jeśli $\mathbb{P}(A) < \delta$, to

$$\left(\int_A |X|^p d\mathbb{P} \right)^{1/p} < \varepsilon/2, \quad \left(\int_A |X_i - X|^p d\mathbb{P} \right)^{1/p} < \varepsilon/2, \quad i = 1, 2, \dots, N.$$

Wystarczy teraz połączyć wszystkie powyższe rozważania i (*): jeśli $\mathbb{P}(A) < \delta$, to

$$\sup_n \int_A |X_n|^p d\mathbb{P} \leq \varepsilon^p.$$

\Leftarrow Ponieważ $X_n \rightarrow X$ według prawdopodobieństwa, to możemy wybrać podciąg $(X_{n_k})_{k \geq 1}$ zbieżny do X prawie na pewno. Z lematu Fatou, $X \in L^p$:

$$\mathbb{E}|X|^p = \mathbb{E} \lim_{k \rightarrow \infty} |X_{n_k}|^p \leq \liminf_{k \rightarrow \infty} \mathbb{E}|X_{n_k}|^p \leq \sup_n \mathbb{E}|X_n|^p < \infty.$$

Dalej, mamy

$$\begin{aligned} \|X_n - X\|_p &\leq \|(X_n - X)1_{\{|X_n - X| \geq \alpha\}}\|_p + \|(X_n - X)1_{\{|X_n - X| < \alpha\}}\|_p \\ &\leq \left(\int_{\{|X_n - X| \geq \alpha\}} |X_n|^p d\mathbb{P} \right)^{1/p} + \left(\int_{\{|X_n - X| \geq \alpha\}} |X|^p d\mathbb{P} \right)^{1/p} \\ &\quad + \left(\int_{\{|X_n - X| < \alpha\}} |X_n - X|^p d\mathbb{P} \right)^{1/p}. \end{aligned}$$

Następnie, wybierzmy dowolne $\varepsilon > 0$ i połóżmy $\alpha = \varepsilon/3$. Z warunku 2° dostajemy istnienie takiej δ , że jeśli $\mathbb{P}(A) < \delta$, to

$$\sup_n \int_A |X_n|^p d\mathbb{P} < (\varepsilon/3)^p, \quad \int_A |X|^p < (\varepsilon/3)^p.$$

Ponadto, z definicji zbieżności według prawdopodobieństwa, istnieje N takie, że dla $n \geq N$, $\mathbb{P}(|X_n - X| \geq \alpha) < \delta$. Stąd wynika teza, gdyż dwa pierwsze składniki w powyższym oszacowaniu są mniejsze niż $\varepsilon/3$ oraz

$$\left(\int_{\{|X_n - X| < \alpha\}} |X_n - X|^p d\mathbb{P} \right)^{1/p} \leq \left(\int_{\{|X_n - X| < \alpha\}} \alpha^p d\mathbb{P} \right)^{1/p} \leq \alpha = \varepsilon/3. \quad \square$$

10. ZADANIA

1. Zmienne $(X_n)_{n \geq 1}$ są niezależnymi zmiennymi Rademachera. Udowodnić, że $(X_n)_{n \geq 1}$ nie jest zbieżny p.n.. Czy $(X_n)_{n \geq 1}$ jest zbieżny według prawdopodobieństwa?

2. Dany jest ciąg $(X_n)_{n \geq 1}$ jak poprzednio. Udowodnić, że szereg $\sum_{n=1}^{\infty} 2^{-n} X_n$ jest zbieżny p.n. i wyznaczyć rozkład graniczny.

3. Dane są ciągi $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$ zbieżne według prawdopodobieństwa do X , Y , odpowiednio. Udowodnić, że

- $(X_n + Y_n)_{n \geq 1}$ zbiega do $X + Y$ według prawdopodobieństwa.
- $(X_n Y_n)_{n \geq 1}$ zbiega według prawdopodobieństwa do XY .

4. Dana jest całkowalna zmienna losowa X . Niech dla $n \geq 1$,

$$X_n(\omega) = \begin{cases} -n & \text{jeśli } X(\omega) < -n, \\ X(\omega) & \text{jeśli } |X(\omega)| \leq n, \\ n & \text{jeśli } X(\omega) > n. \end{cases}$$

Czy $(X_n)_{n \geq 1}$ zbiega do X p.n.? Czy zbiega w L^1 ?

5. Dane są ciągi $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$ zbieżne p.n. do zmiennych X , Y . Udowodnić, że jeśli dla każdego n zmienne X_n oraz Y_n mają ten sam rozkład, to X i Y też mają ten sam rozkład.

6. Zmienne X_1, X_2, \dots są niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z parametrem λ .

(a) Udowodnić, że jeśli $\lambda > 1$, to z prawdopodobieństwem 1 mamy $\{X_n < \log n\}$ dla dostatecznie dużych n , natomiast jeśli $\lambda \geq 1$, to z prawdopodobieństwem 1 mamy $X_n \geq \log n$ dla nieskończenie wielu n .

(b) Zbadać zbieżność p.n. ciągu $(X_n / \log n)_{n \geq 2}$.

7. Zmienne X_1, X_2, \dots są niezależne, nieujemne i mają ten sam rozkład, różny od δ_0 . Dowieść, że $\sum_{n=1}^{\infty} X_n = \infty$ z prawdopodobieństwem 1.

8. Zmienne losowe X_1, X_2, \dots są niezależne, mają ten sam rozkład i spełniają warunek $\mathbb{P}(|X_i| < 1) = 1$. Udowodnić, że $\lim_{n \rightarrow \infty} X_1 X_2 \dots X_n = 0$ p.n.

9. Zmienne X_1, X_2, \dots są niezależne i mają ten sam rozkład.

(a) Udowodnić, że ciąg średnich

$$\frac{X_1 + X_2 + \dots + X_n}{n}, \quad n = 1, 2, \dots$$

jest albo zbieżny p.n., albo rozbieżny z prawdopodobieństwem 1.

(b) Udowodnić, że jeśli ten ciąg jest zbieżny p.n., to jego granica ma rozkład jednopunktowy.

10. Dany jest ciąg $(X_n)_{n \geq 1}$ niezależnych zmiennych losowych takich, że dla $n \geq 1$ X_n ma rozkład Poissona z parametrem $1/n$. Czy $(X_n)_{n \geq 1}$ jest zbieżny według prawdopodobieństwa? Czy jest zbieżny p.n.? Czy jest zbieżny w L^2 ? Czy jest zbieżny w $L^{3/2}$?

11. Dany są ciągi zmiennych $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$, przy czym $X_n \rightarrow X$ w L^p oraz $Y_n \rightarrow Y$ w L^q , gdzie $p, q > 1$ spełniają warunek $1/p + 1/q = 1$. Dowieść, że $(X_n Y_n)_{n \geq 1}$ zbiega w L^1 do XY .

12. Jakie warunki musi spełniać niepusty zbiór $\Lambda \subseteq (0, \infty)$, aby rodzina zmiennych losowych $(X_\lambda)_{\lambda \in \Lambda}$, gdzie

(a) $X_\lambda \sim \mathcal{U}([0, \lambda])$,

(b) $X_\lambda \sim \text{Exp}(\lambda)$,

była jednostajnie całkowalna?

13. Dana jest funkcja $G : [0, \infty) \rightarrow [0, \infty)$ taka, że $\lim_{t \rightarrow \infty} \frac{G(t)}{t} = \infty$. Załóżmy, że $(X_i)_{i \in I}$ jest rodziną zmiennych losowych takich, że $\sup_{i \in I} \mathbb{E}G(|X_i|) < \infty$. Udowodnić, że rodzina ta jest jednostajnie całkowalna.

11. PRAWA WIELKICH LICZB

Prawa wielkich liczb mówią o zachowaniu granicznym ciągu średnich arytmetycznych

$$\frac{X_1 + X_2 + \dots + X_n}{n}, \quad n = 1, 2, \dots,$$

przy rozmaitych założeniach dotyczących zmiennych. Zaczniemy od słabego prawa wielkich liczb (SPWL): termin „słabe” bierze się stąd, iż badana jest zbieżność według prawdopodobieństwa.

Twierdzenie 11.1. *Załóżmy, że X_1, X_2, \dots są zmiennymi losowymi całkowalnymi z kwadratem. Jeśli zmienne te są nieskorelowane oraz mają wspólnie ograniczoną wariancję, to*

$$\frac{X_1 + X_2 + \dots + X_n - \mathbb{E}(X_1 + X_2 + \dots + X_n)}{n} \rightarrow 0$$

według prawdopodobieństwa. W szczególności, jeśli zmienne X_i posiadają tę samą wartość oczekiwaną, to

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1.$$

Dowód. Zauważmy, iż

$$\begin{aligned} & \mathbb{E} \left| \frac{X_1 + X_2 + \dots + X_n - \mathbb{E}(X_1 + X_2 + \dots + X_n)}{n} \right|^2 \\ &= \text{Var} \left(\frac{X_1 + X_2 + \dots + X_n - \mathbb{E}(X_1 + X_2 + \dots + X_n)}{n} \right) \\ &= \frac{1}{n^2} \text{Var} \left(\sum_{k=1}^n X_k \right) \\ &= \frac{1}{n^2} \sum_{k=1}^n \text{Var} X_k \leq \frac{\sup_{k \geq 1} \text{Var} X_k}{n}. \end{aligned}$$

Wobec tego dla ustalonego $\varepsilon > 0$ mamy, na mocy nierówności Czebyszewa,

$$\begin{aligned} & \mathbb{P} \left(\left| \frac{X_1 + X_2 + \dots + X_n - \mathbb{E}(X_1 + X_2 + \dots + X_n)}{n} \right| \geq \varepsilon \right) \\ & \leq \frac{\sup_{k \geq 1} \text{Var} X_k}{n\varepsilon^2}. \end{aligned}$$

Wystarczy zauważyć, że ostatnie wyrażenie zbiega do 0 gdy $n \rightarrow \infty$. Wynika stąd żądana zbieżność według prawdopodobieństwa. \square

Jako przypadek szczególny, dostajemy tzw. słabe prawo wielkich liczb Bernoulliego. Mianowicie, rozważmy ciąg $(\xi_n)_{n \geq 1}$ zmiennych losowych (niekoniecznie niezależnych), przy czym dla $n \geq 1$, zmienna ξ_n ma rozkład $B(n, p)$, gdzie $p \in (0, 1)$ jest ustalonym parametrem. Wówczas

$$\frac{\xi_n}{n} \xrightarrow{\mathbb{P}} p.$$

Istotnie, wystarczy wziąć ciąg $(X_n)_{n \geq 1}$ niezależnych (a więc w szczególności nieskorelowanych) zmiennych losowych o tym samym rozkładzie dwupunktowym $\mathbb{P}(X_i =$

1) = $p = 1 - \mathbb{P}(X_i = 0)$. Wówczas $\xi_n \sim X_1 + X_2 + \dots + X_n$, a więc dla $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{\xi_n}{n} - p \right| > \varepsilon \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{X_1 + X_2 + \dots + X_n}{n} - p \right| > \varepsilon \right) = 0.$$

Głównym wynikiem tego rozdziału jest tzw. mocne prawo wielkich liczb (MPWL) (Twierdzenie 11.4 poniżej), które mówi o zbieżności prawie na pewno. Zaczniemy od kilku przygotowawczych faktów.

Twierdzenie 11.2 (Nierówność Kołmogorowa). *Załóżmy, że X_1, X_2, \dots, X_n są niezależnymi i scentrowanymi zmiennymi losowymi całkowalnymi z kwadratem. Wówczas dla dowolnego $\alpha > 0$,*

$$\mathbb{P}(\max_{1 \leq k \leq n} |X_1 + X_2 + \dots + X_k| \geq \alpha) \leq \frac{1}{\alpha^2} \text{Var}(X_1 + X_2 + \dots + X_n).$$

Dowód. Wprowadźmy oznaczenie $S_0 = 0$ oraz $S_k = X_1 + X_2 + \dots + X_k$ dla $k = 1, 2, \dots, n$. Rozważmy zdarzenia

$$A_k = \{|S_j| < \alpha \text{ dla } j < k \text{ oraz } |S_k| \geq \alpha\},$$

$k = 1, 2, \dots, n$. Jak widać, dla dowolnego k mamy $A_k \in \sigma(X_1, X_2, \dots, X_k)$. Ponadto, zdarzenia A_1, A_2, \dots, A_n są parami rozłączne i dają w sumie $B := \{\max_{1 \leq k \leq n} |S_k| \geq \alpha\}$. Dalej,

$$\begin{aligned} \text{Var } S_n &= \mathbb{E} S_n^2 \\ &= \int_B S_n^2 d\mathbb{P} + \int_{B'} S_n^2 d\mathbb{P} \\ &\geq \int_B S_n^2 d\mathbb{P} \\ &= \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P} \\ &= \sum_{k=1}^n \int_{A_k} (S_k + S_n - S_k)^2 d\mathbb{P} \\ &= \sum_{k=1}^n \left[\int_{A_k} S_k^2 d\mathbb{P} + 2 \int_{A_k} (S_n - S_k) S_k d\mathbb{P} + \int_{A_k} (S_n - S_k)^2 d\mathbb{P} \right] \\ &\geq \sum_{k=1}^n \left[\int_{A_k} S_k^2 d\mathbb{P} + 2 \int_{\Omega} (S_n - S_k) S_k 1_{A_k} d\mathbb{P} \right]. \end{aligned}$$

Ale dla dowolnego k zmienne $S_n - S_k$ oraz $S_k 1_{A_k}$ są niezależne (pierwsza z nich to $X_{k+1} + X_{k+2} + \dots + X_n$, a druga z nich zależy wyłącznie od X_1, X_2, \dots, X_k). Stąd

$$\int_{\Omega} (S_n - S_k) S_k 1_{A_k} d\mathbb{P} = \mathbb{E}(S_n - S_k) S_k 1_{A_k} = \mathbb{E}(S_n - S_k) \mathbb{E} S_k 1_{A_k} = 0.$$

Zatem, kontynuując,

$$\text{Var } S_n \geq \sum_{k=1}^n \int_{A_k} S_k^2 d\mathbb{P} \geq \sum_{k=1}^n \int_{A_k} \alpha^2 d\mathbb{P} = \alpha^2 \sum_{k=1}^n \mathbb{P}(A_k) = \alpha^2 \mathbb{P}(B).$$

Dowód jest zakończony. \square

Twierdzenie 11.3. *Załóżmy, że X_1, X_2, \dots jest ciągiem niezależnych, scentrych zmiennych losowych całkowalnych z kwadratem. Jeśli $\sum_{n=1}^{\infty} \text{Var } X_n < \infty$, to $\sum_{n=1}^{\infty} X_n$ jest zbieżny p.n.*

Dowód. Jak łatwo sprawdzić, mamy

$$\begin{aligned} & \mathbb{P} \left(\sum_{n=1}^{\infty} X_n \text{ jest rozbieżny} \right) \\ &= \mathbb{P} \left(\exists_{\gamma \in \mathbb{N}_+} \forall_n \sup_{k \geq 0} |X_n + X_{n+1} + \dots + X_{n+k}| > \frac{1}{\gamma} \right) \\ &= \mathbb{P} \left(\bigcup_{\gamma \in \mathbb{N}_+} \forall_n \sup_{k \geq 0} |X_n + X_{n+1} + \dots + X_{n+k}| > \frac{1}{\gamma} \right). \end{aligned}$$

Wystarczy więc wykazać, że dla każdego $\gamma \in \mathbb{N}_+$,

$$(*) \quad \mathbb{P} \left(\forall_n \sup_{k \geq 0} |X_n + X_{n+1} + \dots + X_{n+k}| > \frac{1}{\gamma} \right) = 0.$$

Ale dla każdego n , powyższe prawdopodobieństwo szacuje się z góry przez

$$\begin{aligned} & \mathbb{P} \left(\sup_{k \geq 0} |X_n + X_{n+1} + \dots + X_{n+k}| > \frac{1}{\gamma} \right) \\ &= \lim_{m \rightarrow \infty} \mathbb{P} \left(\sup_{0 \leq k \leq m} |X_n + X_{n+1} + \dots + X_{n+k}| > \frac{1}{\gamma} \right) \\ &\leq \limsup_{m \rightarrow \infty} \gamma^2 \sum_{k=0}^m \text{Var } X_{n+k} = \gamma^2 \sum_{k=n}^{\infty} \text{Var } X_k. \end{aligned}$$

Jeśli teraz wziąć $n \rightarrow \infty$, to z założenia powyższe wyrażenie zbiega do 0. Tak więc prawdopodobieństwo (*) musi wynosić 0. \square

Lemat 11.1 (Kronecker). *Załóżmy, że $(a_n)_{n \geq 1}$ jest ciągiem liczbowym takim, że szereg $\sum_{n=1}^{\infty} a_n/n$ jest zbieżny. Wówczas $(a_1 + a_2 + \dots + a_n)/n \rightarrow 0$.*

Dowód. Oznaczmy $S_n = \sum_{k=1}^n a_k/k$. Wówczas $a_n = n(S_n - S_{n-1})$ dla wszystkich n oraz

$$\begin{aligned} \frac{a_1 + a_2 + \dots + a_n}{n} &= \frac{S_1 + 2(S_2 - S_1) + \dots + n(S_n - S_{n-1})}{n} \\ &= \frac{nS_n - S_1 - S_2 - \dots - S_{n-1}}{n} \rightarrow 0. \end{aligned} \quad \square$$

Przechodzimy do głównego twierdzenia.

Twierdzenie 11.4 (MPWL Kołmogorowa). *Załóżmy, że X_1, X_2, \dots jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie.*

(a) *Jeśli $X_n \in L^1$ i $m = \mathbb{E}X_1$, to*

$$\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow m \quad \text{p.n.}$$

(b) *Jeśli $X_n \notin L^1$, to*

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} \left| \frac{X_1 + X_2 + \dots + X_n}{n} \right| = \infty \right) = 1.$$

Dowód. (a) Przypuśćmy najpierw, że zmienne X_1, X_2, \dots są całkowalne z kwadratem. Wówczas teza wynika z dwóch powyższych pomocniczych faktów. Istotnie, korzystając z Twierdzenia 11.3, szereg $\sum_{n=1}^{\infty} \frac{X_n - m}{n}$ jest zbieżny p.n., gdyż

$$\text{Var} \frac{X_n - m}{n} = \frac{\text{Var} X_n}{n^2},$$

i wystarczy skorzystać z lematu Kroneckera.

Rozważmy teraz przypadek ogólny. Wprowadźmy nowy ciąg $(X'_n)_{n \geq 1}$ zmiennych losowych, zadanych przez

$$X'_n(\omega) = X_n(\omega) 1_{(-n, n)}(X_n(\omega)) = \begin{cases} X_n(\omega) & \text{jeśli } |X_n(\omega)| < n, \\ 0 & \text{jeśli } |X_n(\omega)| \geq n. \end{cases}$$

Wówczas X'_1, X'_2, \dots są niezależnymi zmiennymi losowymi całkowalnymi z kwadratem. Możemy napisać

$$\frac{X_1 + X_2 + \dots + X_n}{n} - m = I_n + II_n + III_n,$$

gdzie

$$\begin{aligned} I_n &= \frac{X_1 + X_2 + \dots + X_n - (X'_1 + X'_2 + \dots + X'_n)}{n}, \\ II_n &= \frac{X'_1 + X'_2 + \dots + X'_n - (\mathbb{E}X'_1 + \mathbb{E}X'_2 + \dots + \mathbb{E}X'_n)}{n}, \\ III_n &= \frac{\mathbb{E}X'_1 + \mathbb{E}X'_2 + \dots + \mathbb{E}X'_n}{n} - m. \end{aligned}$$

Zbadajmy zachowanie każdego ze składników I_n, II_n, III_n gdy $n \rightarrow \infty$. Na mocy twierdzenia Lebesgue'a o zmajoryzowanym przejściu do granicy,

$$\mathbb{E}X'_n = \mathbb{E}X_n 1_{\{|X_n| < n\}} = \mathbb{E}X_1 1_{\{|X_1| < n\}} \xrightarrow{n \rightarrow \infty} \mathbb{E}X_1 = m,$$

skąd wynika, że $III_n \rightarrow 0$. Następnie, zauważmy, że

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(X_n \neq X'_n) &= \sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq n) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(|X_1| \geq n) \\ &\leq \int_0^{\infty} \mathbb{P}(|X_1| \geq t) dt = \mathbb{E}|X_1| < \infty. \end{aligned}$$

Zatem z lematu Borela-Cantelli, z prawdopodobieństwem 1 zajdzie tylko skończenie wiele spośród zdarzeń $\{X_n \neq X'_n\}$. Innymi słowy, dla prawie wszystkich ω , ciągi $(X_n(\omega))_{n \geq 1}$ oraz $(X'_n(\omega))_{n \geq 1}$ pokrywają się od pewnego miejsca. Stąd $I_n \rightarrow 0$ p.n.

Pozostało już tylko pokazać, że $II_n \rightarrow 0$ p.n. Na mocy lematu Kroneckera, wystarczy udowodnić, że szereg $\sum_{n=1}^{\infty} (X'_n - \mathbb{E}X'_n)/n$ jest zbieżny prawie na pewno.

Skorzystamy z Twierdzenia 11.3. Otóż

$$\begin{aligned}
\text{Var} \left(\frac{X'_n - \mathbb{E}X'_n}{n} \right) &= \frac{1}{n^2} (\mathbb{E}(X'_n)^2 - (\mathbb{E}X'_n)^2) \\
&\leq \frac{1}{n^2} \mathbb{E}(X'_n)^2 \\
&= \frac{1}{n^2} \sum_{k=1}^{\infty} \int_{\{k-1 \leq |X'_n| < k\}} |X'_n|^2 d\mathbb{P} \\
&= \frac{1}{n^2} \sum_{k=1}^n \int_{\{k-1 \leq |X'_n| < k\}} |X'_n|^2 d\mathbb{P} \\
&= \frac{1}{n^2} \sum_{k=1}^n \int_{\{k-1 \leq |X_1| < k\}} |X_1|^2 d\mathbb{P} \\
&\leq \frac{1}{n^2} \sum_{k=1}^n k \int_{\{k-1 \leq |X_1| < k\}} |X_1| d\mathbb{P}.
\end{aligned}$$

Wobec tego,

$$\begin{aligned}
\sum_{n=1}^{\infty} \text{Var} \left(\frac{X'_n - \mathbb{E}X'_n}{n} \right) &\leq \sum_{n=1}^{\infty} \sum_{k=1}^n \frac{k}{n^2} \mathbb{E}|X_1| 1_{\{k-1 \leq |X_1| < k\}} \\
&= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{k}{n^2} \mathbb{E}|X_1| 1_{\{k-1 \leq |X_1| < k\}} \\
&= \sum_{k=1}^{\infty} k \mathbb{E}|X_1| 1_{\{k-1 \leq |X_1| < k\}} \cdot \sum_{n=k}^{\infty} \frac{1}{n^2}.
\end{aligned}$$

Ale

$$\sum_{n=k}^{\infty} \frac{1}{n^2} = \frac{1}{k^2} + \sum_{n=k+1}^{\infty} \frac{1}{n^2} \leq \frac{1}{k} + \int_k^{\infty} \frac{1}{x^2} dx = \frac{2}{k}.$$

Wobec tego, uwzględniając to w powyższych rozważaniach, dostajemy

$$\sum_{n=1}^{\infty} \text{Var} \left(\frac{X'_n - \mathbb{E}X'_n}{n} \right) \leq 2 \sum_{k=1}^{\infty} \mathbb{E}|X_1| 1_{\{k-1 \leq |X_1| < k\}} = 2\mathbb{E}|X_1| < \infty.$$

Stąd teza (a).

(b) Mamy

$$\frac{X_n}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} - \frac{n-1}{n} \frac{X_1 + X_2 + \dots + X_{n-1}}{n-1}.$$

Wynika stąd, że jeśli ciąg $((X_1(\omega) + X_2(\omega) + \dots + X_n(\omega))/n)_{n \geq 1}$ jest ograniczony dla pewnego ω , to ciąg $(X_n(\omega)/n)_{n \geq 1}$ także ma tę własność. Wobec tego, wystarczy wykazać, że

$$\mathbb{P} \left(\text{ciąg} \left(\frac{X_n}{n} \right)_{n \geq 1} \text{ jest nieograniczony} \right) = 1.$$

Mamy

$$\mathbb{P} \left(\left(\frac{X_n}{n} \right) \text{ nieograniczony} \right) = \mathbb{P} \left(\bigcap_{M \in \mathbb{N}} \left\{ \frac{|X_n|}{n} > M \text{ dla nieskończenie wielu } n \right\} \right),$$

a więc teza będzie zachodzić, jeśli udowodnimy, że dla każdego $M \in \mathbb{N}$,

$$\mathbb{P}\left(\frac{|X_n|}{n} > M \text{ dla nieskończenie wielu } n\right) = 1.$$

Zauważmy, że zdarzenia $\{|X_n|/n > M\}$, $n = 1, 2, \dots$, są niezależne; ponadto

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(|X_n|/n > M) &= \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > nM) \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}(kM < |X_1| \leq (k+1)M) \\ &= \sum_{k=1}^{\infty} \sum_{n=1}^k \mathbb{P}(kM < |X_1| \leq (k+1)M) \\ &= \sum_{k=1}^{\infty} k \mathbb{P}(kM < |X_1| \leq (k+1)M) \\ &\geq -1 + \frac{1}{M} \sum_{k=1}^{\infty} (k+1)M \mathbb{P}(kM < |X_1| \leq (k+1)M) \\ &\geq -1 + \frac{1}{M} \sum_{k=1}^{\infty} \int_{\{kM < |X_1| \leq (k+1)M\}} |X_1| d\mathbb{P} \\ &= -1 + \frac{1}{M} \mathbb{E}|X_1| 1_{\{|X_1| > M\}} = \infty. \end{aligned}$$

Zatem, z lematu Borela-Cantelli wynika teza. \square

Omówimy teraz jedno z zastosowań mocnego prawa wielkich liczb, związane z tzw. dystrybuantą empiryczną.

Definicja 11.1. Załóżmy, że X_1, X_2, \dots są niezależnymi zmiennymi o tym samym rozkładzie z dystrybuantą F . Wówczas n -tą dystrybuantą empiryczną nazywamy

$$\begin{aligned} \mathbb{F}_n(t) &= \frac{1_{\{X_1 \leq t\}} + 1_{\{X_2 \leq t\}} + \dots + 1_{\{X_n \leq t\}}}{n} \\ &= \frac{1_{(-\infty, t]}(X_1) + 1_{(-\infty, t]}(X_2) + \dots + 1_{(-\infty, t]}(X_n)}{n}. \end{aligned}$$

Zauważmy, że dla każdego $\omega \in \Omega$, funkcja \mathbb{F}_n jest dystrybuantą (jako funkcja zmiennej t). Poniższe twierdzenie jest jednym z podstawowych wyników statystyki matematycznej.

Twierdzenie 11.5 (Gliwienko-Cantelli). *Jeśli $X_1, X_2, \dots, F, \mathbb{F}_n$ są jak wyżej, to*

$$\sup_{t \in \mathbb{R}} |\mathbb{F}_n(t) - F(t)| \xrightarrow{n \rightarrow \infty} 0$$

prawie na pewno.

W dowodzie wykorzystamy następujący lemat (bez dowodu: pozostawiamy to jako ćwiczenie).

Lemat 11.2. *Załóżmy, że F, F_1, F_2, \dots są dystrybuantami oraz S jest zbiorem punktów nieciągłości funkcji F . Załóżmy, że Q jest gęstym, przeliczalnym podzbiorem \mathbb{R} takim, że $\lim_{n \rightarrow \infty} F_n(t) = F(t)$ dla każdego $t \in Q$. Wówczas jeśli dla każdego $t \in S$ mamy $F_n(t) - F_n(t-) \rightarrow F(t) - F(t-)$, to*

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow{n \rightarrow \infty} 0.$$

Dowód twierdzenia Gliwienki-Cantelli'ego przy założeniu lematu. Ustalmy dowolny gęsty przeliczalny podzbiór $Q \subset \mathbb{R}$ i niech S będzie zbiorem punktów nieciągłości F . Na mocy MPWL, dla każdego $t \in Q$ mamy

$$\mathbb{F}_n(t) \xrightarrow{n \rightarrow \infty} F(t) \quad \text{prawie na pewno,}$$

gdyż $\mathbb{E}1_{(-\infty, t]}(X_1) = \mathbb{P}(X_1 \leq t) = F(t)$. Podobnie, dla dowolnego $t \in S$,

$$\begin{aligned} \mathbb{F}_n(t) - \mathbb{F}_n(t-) &= \frac{1_{\{t\}}(X_1) + 1_{\{t\}}(X_2) + \dots + 1_{\{t\}}(X_n)}{n} \\ &\xrightarrow{n \rightarrow \infty} \mathbb{E}1_{\{t\}}(X_1) = \mathbb{P}(X_1 = t) = F(t) - F(t-). \end{aligned}$$

Zatem zbiór

$$\Omega_0 = \bigcap_{t \in Q} \{\mathbb{F}_n(t) \rightarrow F(t)\} \cap \bigcap_{t \in S} \{\mathbb{F}_n(t) - \mathbb{F}_n(t-) \rightarrow F(t) - F(t-)\}$$

jest pełnej miary, jako przecięcie zbiorów pełnej miary. Zatem, z lematu, dla każdego $\omega \in \Omega_0$ mamy zbieżność jednostajną $\mathbb{F}_n \rightarrow F$. \square

Na zakończenie tego rozdziału, omówimy wstępne wyniki związane ze zbieżnością szeregów niezależnych zmiennych losowych. Zaczniemy od następującego faktu.

Twierdzenie 11.6. *Załóżmy, że X_1, X_2, \dots są niezależnymi, wspólnie ograniczonymi zmiennymi losowymi (tzn. istnieje takie $a > 0$, że $|X_n| \leq a$ z prawdopodobieństwem 1 dla $n = 1, 2, \dots$). Jeśli szereg $\sum_{n=1}^{\infty} X_n$ jest zbieżny prawie na pewno, to jest on także zbieżny w L^p dla dowolnego $p \geq 1$.*

Lemat 11.3 (Nierówność Hoffmana-Joergensena). *Załóżmy, że X_1, X_2, \dots są niezależnymi zmiennymi losowymi (być może o wartościach w \mathbb{R}^d , bądź ogólniej, o wartościach w przestrzeni Banacha) i zdefiniujmy $S_0 = 0$, $S_n = X_1 + X_2 + \dots + X_n$, $n \geq 1$. Wówczas dla dowolnych $s, t, a \geq 0$,*

$$\begin{aligned} &\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > s + t + a\right) \\ &\leq \mathbb{P}\left(\max_{1 \leq k \leq n} |X_k| > a\right) + \mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > s\right) \mathbb{P}\left(\max_{1 \leq k \leq n} |S_n - S_k| > t/2\right). \end{aligned}$$

Dowód. Niech $\tau = \inf\{k \geq 1 : |S_k| > s\}$ (przyjmujemy $\inf \emptyset = \infty$). Zauważmy, że

$$\begin{aligned} &\left\{\max_{1 \leq k \leq n} |S_k| > s + t + a\right\} \\ &\subseteq \left\{\max_{1 \leq k \leq n} |X_k| > a\right\} \cup \bigcup_{j=1}^n \left(\left\{\max_{1 \leq k \leq n} |X_k| \leq a\right\} \cap \{\tau = j\} \cap \left\{\max_{1 \leq k \leq n} |S_k| > s + t + a\right\}\right) \\ &= \left\{\max_{1 \leq k \leq n} |X_k| > a\right\} \cup \bigcup_{j=1}^n A_j. \end{aligned}$$

Ustalmy $\omega \in A_j$. Mamy $S_j(\omega) > s$; ponadto, $|S_{j-1}(\omega)| \leq s$ oraz $|X_j(\omega)| \leq a$, skąd wynika, na mocy nierówności trójkąta, iż $|S_j| \leq s + a$. Następnie, mamy $|S_\ell(\omega)| > s + t + a$ dla pewnego $\ell > j$. Możemy więc napisać

$$\begin{aligned} s + t + a &< |S_\ell| \\ &\leq |S_j| + |S_\ell - S_j| \\ &\leq s + a + |S_n - S_\ell| + |S_n - S_j| \\ &\leq s + a + 2 \max_{j \leq k \leq n} |S_n - S_k|. \end{aligned}$$

Wnioskujemy stąd, że $\max_{j \leq k \leq n} |S_n - S_k| > t/2$. Wobec tego

$$A_j \subset \{\tau = j\} \cap \left\{ \max_{j \leq k \leq n} |S_n - S_k| > t/2 \right\}$$

i przecinane zdarzenia są niezależne: istotnie, pierwsze z nich zależy tylko od zmiennych X_1, X_2, \dots, X_j , podczas gdy drugie zapisuje się w terminach pozostałych zmiennych. Zbierając wszystkie powyższe fakty dostajemy iż

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > s+t+a\right) \leq \mathbb{P}\left(\max_{1 \leq k \leq n} |X_k| > a\right) + \sum_{j=1}^n \mathbb{P}(\tau = j) \mathbb{P}\left(\max_{1 \leq k \leq n} |S_n - S_k| > t/2\right)$$

i wystarczy już tylko zauważyć, że

$$\sum_{j=1}^n \mathbb{P}(\tau = j) = \mathbb{P}(\tau < \infty) = \mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > s\right). \quad \square$$

Dowód Twierdzenia 11.6. Niech, jak wyżej, $S_n = X_1 + X_2 + \dots + X_n$, $n = 1, 2, \dots$ na mocy założeń, $(S_n)_{n \geq 1}$ jest zbieżny p.n., a więc dla każdego $\varepsilon \in (0, 1)$ istnieje m takie, że

$$\mathbb{P}\left(\max_{m \leq k \leq n} |S_n - S_k| > \varepsilon/2\right) < \varepsilon$$

o ile tylko $n > m$. Mamy

$$\begin{aligned} \mathbb{E}|S_n - S_m|^p &= p \int_0^\infty \alpha^{p-1} \mathbb{P}(|S_n - S_m| > \alpha) d\alpha \\ &= \sum_{r=0}^\infty p \int_{(r+1)(\varepsilon+a) > \alpha > r(\varepsilon+a)} \alpha^{p-1} \mathbb{P}(|S_n - S_m| > \alpha) d\alpha \\ &= p \int_0^{\varepsilon+a} \alpha^{p-1} \mathbb{P}(|S_n - S_m| > \alpha) d\alpha \\ &\quad + \sum_{r=1}^\infty p \int_{(r+1)(\varepsilon+a) > \alpha > r(\varepsilon+a)} \alpha^{p-1} \mathbb{P}(|S_n - S_m| > r(\varepsilon+a)) d\alpha. \end{aligned}$$

Ale $S_n - S_m = X_{m+1} + X_{m+2} + \dots + X_n$, więc stosując nierówność Hoffmana-Joergensena do tych zmiennych, z parametrami $s = (r-1)(\varepsilon+a)$, $t = \varepsilon$ oraz $a = a$, dostajemy

$$\begin{aligned} \mathbb{P}\left(\max_{m < k \leq n} |S_k - S_m| > r(\varepsilon+a)\right) &\leq 0 + \mathbb{P}\left(\max_{m < k \leq n} |S_k - S_m| > (r-1)(\varepsilon+a)\right) \times \\ &\quad \times \mathbb{P}\left(\max_{m < k \leq n} |S_n - S_k| > \varepsilon/2\right). \end{aligned}$$

Drugi czynnik szacuje się przez ε , a zatem, przez prostą indukcję, dostajemy, że

$$\mathbb{P}(|S_n - S_m| > r(\varepsilon+a)) \leq \mathbb{P}\left(\max_{m < k \leq n} |S_n - S_m| > r(\varepsilon+a)\right) \leq \varepsilon^r.$$

Wobec tego

$$\begin{aligned} \mathbb{E}|S_n - S_m|^p &\leq p \int_0^\varepsilon \alpha^{p-1} d\alpha + p \int_\varepsilon^{\varepsilon+a} \alpha^{p-1} \mathbb{P}(|S_n - S_m| > \varepsilon) d\alpha \\ &\quad + \sum_{r=1}^{\infty} p \int_{r(\varepsilon+a)}^{(r+1)(\varepsilon+a)} \alpha^{p-1} \varepsilon^r d\alpha \\ &\leq \varepsilon^p + (\varepsilon+a)^p \cdot \varepsilon + \varepsilon \sum_{r=1}^{\infty} \varepsilon^{r-1} ((r+1)^p - r^p) (\varepsilon+a)^p \\ &\leq \varepsilon \cdot C, \end{aligned}$$

gdzie C jest pewną stałą zależącą tylko od p i a . Wobec tego $(S_n)_{m \geq 0}$ spełnia warunek Cauchy'ego w L^p , a więc jest zbieżny w L^p . \square

Wprowadźmy następujące oznaczenie: dla zmiennej losowej X oraz $a > 0$, niech

$$X^a(\omega) = \begin{cases} a & \text{dla } X(\omega) > a, \\ X(\omega) & \text{dla } |X(\omega)| \leq a, \\ -a & \text{dla } X(\omega) < -a. \end{cases}$$

Twierdzenie 11.7 (Kolmogorowa o trzech szeregach). *Załóżmy, że X_1, X_2, \dots są niezależnymi zmiennymi losowymi oraz a jest ustaloną liczbą dodatnią. Wówczas szereg $\sum_{n=0}^{\infty} X_n$ jest zbieżny p.n. wtedy i tylko wtedy, gdy zbieżne są szeregi liczbowe*

$$\sum_{n=1}^{\infty} \mathbb{E}X_n^a, \quad \sum_{n=1}^{\infty} \text{Var} X_n^a, \quad \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > a).$$

W dowodzie wykorzystamy następujący prosty fakt.

Lemat 11.4. *Załóżmy, że X_1, X_2, \dots są zmiennymi losowymi spełniającymi warunek $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > a) < \infty$ dla pewnego $a > 0$. Wówczas szereg $\sum_{n=1}^{\infty} X_n$ jest zbieżny p.n. wtedy i tylko wtedy, gdy szereg $\sum_{n=1}^{\infty} X_n^a$ jest zbieżny p.n.*

Dowód. Na mocy lematu Borela-Cantelli, dla prawie wszystkich ω ciągi $X_n(\omega)$, $X_n^a(\omega)$ pokrywają się od pewnego miejsca. Stąd natychmiast wynika teza. \square

Dowód twierdzenia o trzech szeregach. \Leftarrow Na mocy lematu, wystarczy wykazać, że szereg $\sum_{n=1}^{\infty} X_n^a$ jest zbieżny p.n. Zmienne $X_n^a - \mathbb{E}X_n^a$ są scentrowane, niezależne, ograniczone oraz

$$\sum_{n=1}^{\infty} \text{Var} (X_n^a - \mathbb{E}X_n^a) = \sum_{n=1}^{\infty} \text{Var} X_n^a < \infty,$$

a więc na mocy Twierdzenia 11.3 szereg $\sum_{n=1}^{\infty} (X_n^a - \mathbb{E}X_n^a)$ jest zbieżny p.n. Ponieważ szereg liczbowy $\sum_{n=1}^{\infty} \mathbb{E}X_n^a$ także jest zbieżny, wynika stąd teza.

\Rightarrow Przypuśćmy, wbrew tezie, że $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > a) = \infty$. Wówczas na mocy lematu Borela-Cantelli, z prawdopodobieństwem 1 zachodzi nieskończoność wiele nierówności $|X_n| > a$, co wyklucza zbieżność szeregu $\sum_{n=1}^{\infty} X_n$, sprzeczność. Zatem $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > a) < \infty$, a więc z powyższego lematu szereg $\sum_{n=1}^{\infty} X_n^a$ jest zbieżny p.n. Korzystając z Twierdzenia 11.6, dostajemy, że szereg $\sum_{n=1}^{\infty} X_n^a$ jest zbieżny w L^1 oraz L^2 . Zatem, w szczególności, ciąg liczbowy

$$\left(\mathbb{E} \sum_{n=1}^N X_n^a \right)_{N \geq 1} = \left(\sum_{n=1}^N \mathbb{E}X_n^a \right)_{N \geq 1}$$

jest zbieżny. Ponadto, to samo jest prawdą dla

$$\left(\mathbb{E} \left(\sum_{n=1}^N (X_n^a - \mathbb{E} X_n^a) \right)^2 \right)_{N \geq 1} = \left(\text{Var} \left(\sum_{n=1}^N X_n^a \right) \right)_{N \geq 1} = \left(\sum_{n=1}^N \text{Var} X_n^a \right)_{N \geq 1}.$$

To zaś oznacza tezę. \square

Przykład: Załóżmy, że X_1, X_2, \dots jest ciągiem niezależnych zmiennych losowych takich, że X_n ma rozkład wykładniczy z parametrem λ_n . Wyznamy warunek na ciąg $(\lambda_n)_{n \geq 1}$ który jest równoważny zbieżności prawie na pewno szeregu $\sum_{n=1}^{\infty} X_n$.

Na mocy twierdzenia Kołmogorowa, zbieżność p.n. ma miejsce wtedy i tylko wtedy, gdy

$$(1) \quad \sum_{n=1}^{\infty} \mathbb{P}(|X_n^1| > 1) < \infty, \quad \text{czyli} \quad \sum_{n=1}^{\infty} e^{-\lambda_n} < \infty,$$

$$(2) \quad \sum_{n=1}^{\infty} \mathbb{E} X_n^1 < \infty, \quad \text{czyli} \quad \sum_{n=1}^{\infty} \left(\frac{1 - e^{-\lambda_n}}{\lambda_n} - e^{-\lambda_n} \right) < \infty,$$

$$(3) \quad \sum_{n=1}^{\infty} \text{Var} X_n^1 < \infty, \quad \text{czyli} \quad \sum_{n=1}^{\infty} \left(\frac{1}{\lambda_n^2} - e^{-\lambda_n} - \frac{(1 + \lambda_n)^2}{\lambda_n^2} e^{-2\lambda_n} \right) < \infty.$$

Założmy teraz, że szereg $\sum_{n=1}^{\infty} X_n$ jest zbieżny p.n. Wówczas z (1) mamy $\lambda_n \rightarrow \infty$, co w połączeniu z (2) prowadzi to wniosku, iż $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$.

Wykażemy, że warunek $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$ jest dostateczny. Istotnie, wynika z niego, że $\lambda_n \rightarrow \infty$, a stąd $\sum_{n=1}^{\infty} \frac{1}{\lambda_n^2} < \infty$ (gdyż dla dostatecznie dużych n mamy $\frac{1}{\lambda_n^2} \leq \frac{1}{\lambda_n}$). Dalej, mamy $e^{-\lambda_n} \leq \frac{1}{1 + \lambda_n} \leq \frac{1}{\lambda_n}$ oraz $e^{-2\lambda_n} \leq \frac{1}{\lambda_n^2}$, skąd wynika już zbieżność wszystkich trzech szeregów w (1), (2) oraz (3).

12. ZADANIA

1. Dany jest ciąg $(X_n)_{n \geq 1}$ niezależnych zmiennych losowych o rozkładzie Poissona z parametrem 2. Udowodnić, że ciąg

$$\frac{X_1 X_2 + X_2 X_3 + \dots + X_n X_{n+1}}{n + 2009}, \quad n = 1, 2, \dots,$$

jest zbieżny p.n. i wyznaczyć jego granicę.

2. Dany jest ciąg $(X_n)_{n \geq 1}$ niezależnych zmiennych losowych, przy czym dla $n \geq 1$ zmienna X_n ma rozkład jednostajny na przedziale $(1/n, 1]$. Udowodnić, że ciąg

$$\frac{X_1 + X_2 + \dots + X_n}{n}, \quad n = 1, 2, \dots,$$

jest zbieżny p.n. i wyznaczyć jego granicę.

3. Dany jest ciąg $(X_n)_{n \geq 1}$ niezależnych nieujemnych zmiennych losowych o tym samym rozkładzie. Udowodnić, że jeśli $\mathbb{E}X_1 = \infty$, to

$$\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \infty$$

prawie na pewno.

4. Dany jest ciąg $(A_n)_{n \geq 1}$ niezależnych zdarzeń, $p_n = \mathbb{P}(A_n)$. Udowodnić, że

$$\frac{1_{A_1} + 1_{A_2} + \dots + 1_{A_n}}{n} - \frac{p_1 + p_2 + \dots + p_n}{n} \rightarrow 0$$

według prawdopodobieństwa.

5. Dany jest ciąg $(X_n)_{n \geq 1}$ niezależnych całkowalnych zmiennych losowych o tym samym rozkładzie. Udowodnić, że ciąg

$$\frac{X_1 + X_2 + \dots + X_n}{n}, \quad n \geq 1,$$

zbiega w L^1 do $\mathbb{E}X_1$.

6. Dany jest ciąg $(N_n)_{n \geq 1}$ zmiennych losowych (niekoniecznie niezależnych), przy czym dla $n \geq 1$ zmienna N_n ma rozkład Poissona z parametrem n . Wykazać, że $N_n/n \rightarrow 1$ w L^1 .

7. Zmienne X_1, X_2, \dots są niezależne i mają rozkład jednostajny na $[-1, 1]$. Czy ciąg

$$\frac{X_1 + X_2^2 + \dots + X_n^n}{n}, \quad n = 1, 2, \dots,$$

jest zbieżny p.n.?

8. Obliczyć granice

$$\lim_{n \rightarrow \infty} \frac{1}{2^n} \int_{-1}^1 \int_{-1}^1 \dots \int_{-1}^1 \frac{x_1^2 + x_2^2 + x_3^3 + x_4^4 + \dots + x_n^n}{x_1^2 + x_2^2 + \dots + x_n^2} dx_1 dx_2 \dots dx_n$$

oraz

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 \dots \int_0^1 f(\sqrt[n]{x_1 x_2 \dots x_n}) dx_1 dx_2 \dots dx_n,$$

gdzie $f : [0, 1] \rightarrow \mathbb{R}$ jest ustaloną funkcją ciągłą.

9. Zmienne losowe X_1, X_2, \dots są niezależne, przy czym dla $n \geq 1$ rozkład X_n zadany jest następująco:

$$\mathbb{P}(X_n = 0) = 1/2, \mathbb{P}(X_n = 1) = 1/2 - \frac{1}{4n^2}, \mathbb{P}(X_n = n) = \frac{1}{4n^2}.$$

Udowodnić, że ciąg

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

jest zbieżny prawie na pewno i wyznaczyć jego granicę.

10. Zmienne $\varepsilon_1, \varepsilon_2, \dots$ są niezależne i mają rozkład Rademachera. Dowieść, że dla $\alpha > 1/2$, ciąg

$$\frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n}{n^\alpha}, \quad n = 1, 2, \dots$$

jest zbieżny p.n.

11. Udowodnić następujące twierdzenie o dwóch szeregach: jeśli $(X_n)_{n \geq 1}$ jest ciągiem takich niezależnych zmiennych losowych całkowalnych z kwadratem, że szeregi liczbowe

$$\sum_{n=1}^{\infty} \mathbb{E}X_n, \quad \sum_{n=1}^{\infty} \text{Var} X_n$$

są zbieżne, to szereg $\sum_{n=1}^{\infty} X_n$ jest zbieżny p.n.

12. Dany jest ciąg $(X_n)_{n \geq 1}$ niezależnych zmiennych losowych takich, że

$$\mathbb{P}(X_n = -n) = \mathbb{P}(X_n = n) = \frac{1}{n^3}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{2}{n^3}.$$

Udowodnić, że $\sum_{n=1}^{\infty} X_n$ jest zbieżny p.n.

13. Dany jest ciąg $(\varepsilon_n)_{n \geq 1}$ niezależnych zmiennych Rademachera. Jaki warunek musi spełniać ciąg $(a_n)_{n \geq 1}$, by szereg $\sum_{n=1}^{\infty} a_n \varepsilon_n$ był zbieżny p.n.?

14. Dany jest ciąg (X_n) niezależnych zmiennych losowych takich, że dla $n \geq 1$ zmienna X_n ma rozkład jednostajny na odcinku $[-n, n]$. Dla jakich wartości parametru $p > 0$ szereg

$$\sum_{n=1}^{\infty} \frac{X_n}{n^p}$$

jest zbieżny p.n.?

13. TWIERDZENIE DE MOIVRE'A-LAPLACE'A

Zajmiemy się teraz niezwykle ważnym i użytecznym faktem, pozwalającym przybliżyć rozkład Bernoulliego $B(n, p)$ przez rozkład normalny. Znaczne uogólnienie poniższych wyników będzie podane na wykładzie z Rachunku Prawdopodobieństwa II, przy okazji tzw. Centralnego Twierdzenia Granicznego.

Załóżmy, że $\Phi : \mathbb{R} \rightarrow [0, 1]$ jest dystrybuantą, a g jest gęstością, standardowego rozkładu normalnego:

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

Ponadto, będziemy zakładać, że p jest ustaloną liczbą z przedziału $(0, 1)$, $q = 1 - p$ oraz S_n jest zmienną losową o rozkładzie $B(n, p)$.

Twierdzenie 13.1. *Załóżmy, że k jest liczbą całkowitą taką, że*

$$(*) \quad |k - np| \cdot \frac{\max(p, q)}{npq} \leq 1/2.$$

Wówczas

$$\mathbb{P}(S_n = k) = \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k - np)^2}{2npq} + R(n, k)\right),$$

gdzie

$$R(n, k) \leq \frac{3|k - np|}{4npq} + \frac{|k - np|^3}{3n^2 p^2 q^2} + \frac{1}{3npq}.$$

Dowód. Stosując wzór Stirlinga

$$n! = \sqrt{2\pi n} n^n e^{-n + \theta_n/(12n)}, \quad 0 < \theta_n < 1,$$

dostajemy

$$\begin{aligned} \mathbb{P}(S_n = k) &= \sqrt{\frac{n}{2k\pi(n-k)}} \cdot \binom{np}{k} \left(\frac{nq}{n-k}\right)^{n-k} \cdot \exp\left(\frac{\theta_n}{12n} - \frac{\theta_k}{12k} - \frac{\theta_{n-k}}{12(n-k)}\right) \\ &= I \cdot II \cdot III. \end{aligned}$$

Zbadajmy po kolei czynniki I , II oraz III . Mamy

$$I = \frac{1}{\sqrt{2\pi npq}} \cdot \left(1 + \frac{k - np}{npq} \cdot q\right)^{-1/2} \left(1 - \frac{k - np}{npq} \cdot p\right)^{-1/2} = \frac{1}{\sqrt{2\pi npq}} \cdot e^{R_1(n, k)}.$$

Dla dowolnego $x \geq -1/2$ zachodzi oszacowanie

$$|\log(1 + x) - x| \leq x^2,$$

skąd wynika, że

$$\begin{aligned} R_1(n, k) &= -\frac{1}{2} \left[\log\left(1 + \frac{k - np}{npq} \cdot q\right) + \log\left(1 - \frac{k - np}{npq} \cdot p\right) \right] \\ &= \frac{1}{2} \frac{(p - q)(k - np)}{npq} + R'_1(n, k), \end{aligned}$$

gdzie $|R'_1(n, k)| \leq \frac{1}{2}(p^2 + q^2)(k - np)^2 / (n^2 p^2 q^2) \leq |k - np| / (4npq)$, na mocy założenia (*). W konsekwencji,

$$|R_1(n, k)| \leq \frac{3|k - np|}{4npq}.$$

Następnie, mamy

$$\begin{aligned} \log II &= k \log \left(\frac{np}{k} \right) + (n-k) \log \left(\frac{nq}{n-k} \right) \\ &= -np \cdot \frac{k}{np} \log \left(\frac{k}{np} \right) - nq \cdot \frac{n-k}{nq} \log \left(\frac{n-k}{nq} \right) \\ &= -np \cdot \left(1 + \frac{k-np}{np} \right) \log \left(1 + \frac{k-np}{np} \right) \\ &\quad - nq \cdot \left(1 - \frac{k-np}{nq} \right) \log \left(1 - \frac{k-np}{nq} \right). \end{aligned}$$

Korzystamy teraz z nierówności

$$\left| (1+x) \log(1+x) - x - \frac{x^2}{2} \right| \leq \frac{1}{3} |x|^3,$$

prawdziwej dla $x \geq -1/2$: jeśli (*) jest spełniona, to

$$\begin{aligned} \log II &= -np \left(\frac{k-np}{np} + \frac{1}{2} \left(\frac{k-np}{np} \right)^2 \right) \\ &\quad - nq \left(-\frac{k-np}{nq} + \frac{1}{2} \left(\frac{k-np}{nq} \right)^2 \right) + R'_2(n, k) \\ &= -\frac{1}{2} \frac{(k-np)^2}{npq} + R'_2(n, k), \end{aligned}$$

gdzie

$$|R'_2(n, k)| \leq \frac{1}{3} \left(np \left| \frac{k-np}{np} \right|^3 + nq \left| \frac{k-np}{nq} \right|^3 \right) = \frac{|k-np|^3}{3n^2 p^2 q^2} (p^2 + q^2) \leq \frac{|k-np|^3}{3n^2 p^2 q^2}.$$

Wreszcie, mamy $III = e^{R_3(n, k)}$, gdzie

$$-\left(\frac{1}{12k} + \frac{1}{12(n-k)} \right) < R_3(n, k) < \frac{1}{12n}.$$

Równoważnie,

$$-\frac{1}{12npq} \left(1 + \frac{k-np}{npq} \cdot q \right)^{-1} \left(1 - \frac{k-np}{npq} \cdot p \right)^{-1} < R_3(n, k) < \frac{1}{12n},$$

skąd, na mocy (*), wynika oszacowanie

$$|R_3(n, k)| \leq \frac{1}{3npq}.$$

Łącząc otrzymane wyżej nierówności dla $R_i(n, k)$, dostajemy tezę. \square

Kolejne twierdzenie, tzw. integralne twierdzenie de Moivre'a-Laplace'a, pozwala przybliżać prawdopodobieństwo, że liczba sukcesów należy do ustalonego przedziału.

Twierdzenie 13.2. *Załóżmy, że $a, b \geq 0$ spełniają warunek*

$$(*) \quad |a - np| \cdot \frac{\max(p, q)}{npq} \leq 1/2, \quad |b - np| \cdot \frac{\max(p, q)}{npq} \leq 1/2.$$

Wówczas

$$\mathbb{P}(a \leq S_n \leq b) = \left[\Phi \left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}} \right) - \Phi \left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}} \right) \right] e^{D(n,a,b)},$$

gdzie

$$|D(n, a, b)| \leq \max_{k \in \{a, b\}} \left[\frac{5}{4} \frac{|k - np|}{npq} + \frac{1}{3} \frac{|k - np|^3}{n^2 p^2 q^2} \right] + \frac{1}{3npq} + \frac{1}{8npq}.$$

Dowód. Przypomnijmy, że g oznacza gęstość standardowego rozkładu normalnego. Oznaczmy

$$x_k = \frac{k - np}{\sqrt{npq}}, \quad h = \frac{1}{\sqrt{npq}}.$$

Z twierdzenia o wartości średniej mamy, iż

$$\Phi(x_k + h/2) - \Phi(x_k - h/2) = hg(\xi_k),$$

gdzie $\xi_k \in (x_k - h/2, x_k + h/2)$. Innymi słowy, mamy

$$\begin{aligned} & hg(x_k) \\ &= \exp \left(\frac{1}{2} (\xi_k^2 - x_k^2) \right) \Phi(x_k + h/2) - \Phi(x_k - h/2). \end{aligned}$$

Dalej, $|\xi_k^2 - x_k^2| = |\xi_k + x_k| \cdot |\xi_k - x_k| \leq \frac{1}{2}h(2|x_k| + \frac{1}{2}h) = h|x_k| + \frac{1}{4}h^2$, a zatem

$$hg(x_k) = e^{r_k} [\Phi(x_k + h/2) - \Phi(x_k - h/2)],$$

gdzie $|r_k| \leq \frac{1}{2}h|x_k| + \frac{1}{8}h^2$. W połączeniu z poprzednim twierdzeniem, otrzymujemy zatem

$$\mathbb{P}(S_n = k) = e^{r_k + R(n,k)} [\Phi(x_k + h/2) - \Phi(x_k - h/2)].$$

Niech $d = \max_{k \in \{a, a+1, \dots, b\}} |r_k + R(n, k)|$; dostajemy zatem

$$e^{-d} [\Phi(x_k + h/2) - \Phi(x_k - h/2)] \leq \mathbb{P}(S_n = k) \leq e^d [\Phi(x_k + h/2) - \Phi(x_k - h/2)].$$

Pisząc te nierówności dla $k = a, a + 1, \dots, b$ i sumując, dostajemy tezę. \square

Na zakończenie, sformułujemy (bez dowodu) fakt, który zawiera wygodne oszacowanie na błąd przybliżenia w twierdzeniu de Moivre'a-Laplace'a.

Twierdzenie 13.3. *Przy oznaczeniach jak wyżej, mamy*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(\frac{S_n - np}{\sqrt{npq}} \leq t \right) - \Phi(t) \right| \leq \frac{p^2 + q^2}{\sqrt{npq}}.$$

14. ZADANIA

1. Prawdopodobieństwo urodzenia chłopca wynosi 0,517. Jakie jest prawdopodobieństwo tego, że wśród $n = 10000$ noworodków liczba chłopców nie przewyższy liczby dziewcząt?

2. Rzucamy symetryczną monetą aż do momentu, gdy wyrzucimy 200 orłów (łącznie, niekoniecznie pod rząd). Jakie jest przybliżone prawdopodobieństwo tego, że rzucimy więcej niż 440 razy?

3. Do sklepu meblowego przywieziono 150 biurek I rodzaju oraz 75 biurek II rodzaju. Wiadomo, że biurka I rodzaju cieszą się dwukrotnie większym powodzeniem (tzn. prawdopodobieństwo tego, że klient kupujący biurko zdecyduje się na biurko I rodzaju, wynosi $2/3$). Jakie jest przybliżone prawdopodobieństwo tego, że któryś z pierwszych 200 klientów kupujących biurka nie dostanie takiego modelu, jaki chce?

4. Stwierdzono, iż przeciętnie 30% spośród ogólnej liczby studentów przyjętych na studia kończy je w terminie. Ile osób trzeba przyjąć na pierwszy rok, aby z prawdopodobieństwem co najmniej 0,9 co najmniej 50 osób skończyło studia w terminie?

5. W pewnym doświadczeniu prawdopodobieństwo zajścia zdarzenia A wynosi 0,7. Ile razy trzeba powtórzyć to doświadczenie, żeby z prawdopodobieństwem 0,9 częstość zajścia zdarzenia A nie różniła się od 0,7 o więcej niż 0,1? Czy można coś powiedzieć o potrzebnej liczbie powtórzeń, jeśli nie znamy prawdopodobieństwa zdarzenia A ?

6. a) Rzucamy 4500 razy kostką, dla której prawdopodobieństwo wypadnięcia szóstki wynosi $1/6$. Obliczyć przybliżone prawdopodobieństwo tego, że liczba wyrzuconych szóstek przekroczy 450.

b) Załóżmy, że prawdopodobieństwo wypadnięcia szóstki wynosi $1/1000$. Jakie jest przybliżone prawdopodobieństwo tego, że liczba wyrzuconych szóstek przekroczy 2?

7. Dany jest ciąg $(\varepsilon_n)_{n \geq 1}$ niezależnych zmiennych losowych Rademachera. Dowieść, że ciąg

$$\frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n}{\sqrt{n}}, \quad n = 1, 2, \dots,$$

nie jest zbieżny prawie na pewno.

15. WARUNKOWA WARTOŚĆ OCZEKIWANA

Warunkowa wartość oczekiwana jest jednym z kluczowych pojęć w teorii prawdopodobieństwa. Zaczniemy od sytuacji gdy warunkujemy względem zdarzenia.

Definicja 15.1. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną oraz B jest zdarzeniem o dodatnim prawdopodobieństwie. Niech X będzie całkowalną zmienną losową. Warunkową wartością oczekiwaną X pod warunkiem B nazywamy liczbę

$$\mathbb{E}(X|B) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega|B).$$

Twierdzenie 15.1. *Przy założeniach jak wyżej,*

$$(*) \quad \mathbb{E}(X|B) = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P}.$$

Dowód: Stosujemy standardową metodę komplikacji zmiennej X .

1. Załóżmy najpierw, że $X = 1_A$, gdzie $A \in \mathcal{F}$. Wówczas

$$\mathbb{E}(X|B) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)} \int_B 1_A d\mathbb{P}.$$

2. Z liniowości, dowiedziona równość zachodzi także dla zmiennych prostych (kombinacji liniowych indykatorów zdarzeń).

3. Teraz jeśli X jest nieujemną zmienną losową, to bierzemy niemalejący ciąg $(X_n)_{n \geq 1}$ zmiennych prostych zbieżny prawie na pewno do X . Pisząc $(*)$ dla X_n i zbiegając z $n \rightarrow \infty$ dostajemy $(*)$ dla X , na mocy twierdzenia Lebesgue'a o monotonicznym przejściu do granicy pod znakiem całki.

4. Jeśli X jest dowolną zmienną losową, to rozważamy rozbitcie $X = X_+ - X_-$ i stosujemy $(*)$ dla X_+ oraz X_- ; po odjęciu stronami dostajemy $(*)$ dla X . \square

Rozważmy teraz następujący przykład. Przypuśćmy, że $\{B_i\}_{i=1,2,\dots,n}$ jest rozbitciem Ω na zdarzenia o dodatniej mierze. Niech X będzie całkowalną zmienną losową i zdefiniujmy zmienną η wzorem $\eta(\omega) = \mathbb{E}(X|B_i)$ jeśli $\omega \in B_i$, $i = 1, 2, \dots, n$. Zmienną η interpretujemy jako średnią wartość X jeśli wiemy wszystko o zdarzeniach z σ -ciała generowanego przez rozbitcie $\{B_i\}$. Zmienna η posiada następujące własności:

- 1) η jest mierzalna względem $\sigma(B_1, B_2, \dots, B_n)$ - gdyż jest stała na dowolnym zdarzeniu B_i ,

- 2) Dla każdego $i = 1, 2, \dots, n$ mamy

$$\int_{B_i} \eta d\mathbb{P} = \mathbb{E}(X|B_i) \cdot \mathbb{P}(B_i) = \int_{B_i} X d\mathbb{P},$$

skąd wynika, iż

$$\int_B \eta d\mathbb{P} = \int_B X d\mathbb{P}$$

dla dowolnego $B \in \sigma(B_1, B_2, \dots, B_n)$.

Prowadzi to do definicji warunkowej wartości oczekiwanej względem σ -ciała.

Definicja 15.2. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną, \mathcal{M} jest pod- σ -ciałem \mathcal{F} , a X jest całkowalną zmienną losową. Warunkową wartością oczekiwaną X pod warunkiem \mathcal{M} nazywamy taką zmienną losową η , że są spełnione następujące dwa warunki.

- 1) η jest mierzalna względem \mathcal{M} .

2) Dla każdego $B \in \mathcal{M}$,

$$\int_B \eta d\mathbb{P} = \int_B X d\mathbb{P}.$$

Oznaczenie: $\mathbb{E}(X|\mathcal{M})$.

W szczególności gdy $X = 1_A$, $A \in \mathcal{F}$, to definiujemy prawdopodobieństwo warunkowe zdarzenia A pod warunkiem \mathcal{M} poprzez $\mathbb{P}(A|\mathcal{M}) = \mathbb{E}(1_A|\mathcal{M})$.

Twierdzenie 15.2. *Załóżmy, że X jest całkowalną zmienną losową, a \mathcal{M} jest pod- σ -ciałem \mathcal{F} . Wówczas warunkowa wartość oczekiwana istnieje i jest wyznaczona jednoznacznie z dokładnością do równości p.n.*

Dowód: Dla dowolnego $B \in \mathcal{M}$ definiujemy $\nu(B) = \int_B X d\mathbb{P}$. Funkcja $\nu : \mathcal{M} \rightarrow \mathbb{R}$ jest przeliczalnie addytywną funkcją zbioru. Ponadto jeśli $\mathbb{P}(B) = 0$, to $\nu(B) = 0$ (jest to tzw. absolutna ciągłość ν względem \mathbb{P}). Na mocy twierdzenia Radona-Nikodyma istnieje \mathcal{M} -mierzalna zmienna losowa η będąca gęstością ν względem \mathbb{P} , tzn. taka, że dla wszystkich $B \in \mathcal{M}$,

$$\int_B X d\mathbb{P} = \nu(B) = \int_B \eta d\mathbb{P}.$$

Jednoznaczność jest oczywista: jeśli η_1, η_2 są zmiennymi losowymi spełniającymi 1) oraz 2), to w szczególności, dla każdego $B \in \mathcal{M}$, $\int_B \eta_1 d\mathbb{P} = \int_B \eta_2 d\mathbb{P}$, skąd $\eta_1 = \eta_2$ p.n. \square

Przechodzimy do pojęcia warunkowej wartości oczekiwanej względem zmiennej losowej. Będziemy potrzebować następującego pomocniczego faktu.

Lemat 15.1. *Załóżmy, że Y jest zmienną losową. Wówczas każda zmienna losowa X mierzalna względem $\sigma(Y)$ ma postać $f(Y)$ dla pewnej funkcji borelowskiej f .*

Dowód: Ponownie stosujemy metodę komplikacji zmiennej.

1. Załóżmy, że $X = 1_A$, gdzie $A \in \sigma(Y)$. Wówczas $A = \{Y \in B\}$ dla pewnego B , skąd $X = 1_B(Y)$, czyli jako f możemy wziąć indykator 1_B .

2. Jeśli X jest zmienną prostą, to jako f bierzemy kombinację liniową odpowiednich indykatorów (patrz poprzedni punkt).

3. Załóżmy, że X jest nieujemną zmienną losową. Istnieje niemalejący ciąg (X_n) prostych, $\sigma(Y)$ -mierzalnych zmiennych losowych zbieżny do X . Na mocy 2), mamy $X_n = f_n(Y)$ dla pewnego ciągu funkcyjnego (f_n) . Jak łatwo sprawdzić, wystarczy wziąć

$$f(x) = \begin{cases} \lim_{n \rightarrow \infty} f_n(x) & \text{jeśli granica istnieje,} \\ 0 & \text{jeśli granica nie istnieje.} \end{cases}$$

4. Jeśli teraz X jest dowolną zmienną losową, to mamy $X = X_+ - X_- = f_+(Y) - f_-(Y) = f(Y)$, gdzie f_+, f_- to funkcje borelowskie odpowiadające $\sigma(Y)$ -mierzalnym X_+ oraz X_- . \square

Definicja 15.3. Załóżmy, że X, Y są zmiennymi losowymi, przy czym X jest całkowalna. Definiujemy warunkową wartość oczekiwaną X pod warunkiem Y jako

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

Uwaga: Na mocy lematu mamy $\mathbb{E}(X|Y) = f(Y)$ dla pewnej funkcji borelowskiej f . Liczbę $f(y)$ możemy interpretować jako $\mathbb{E}(X|Y = y)$.

Przykłady:

1. Załóżmy, że X, Y posiadają rozkłady skokowe. Oznaczmy

$$P_Y(y) = \mathbb{P}(Y = y) \quad \text{oraz} \quad P_{(X,Y)}(x, y) = \mathbb{P}(X = x, Y = y).$$

Jeśli h jest dowolną funkcją borelowską taką, że $h(X) \in L^1$, to

$$\mathbb{E}(h(X)|Y) = \sum_{x \in S_X} h(x) \frac{P_{(X,Y)}(x, Y)}{P_Y(Y)}.$$

Aby to wykazać, należy sprawdzić, iż prawa strona (oznaczana dalej przez η) spełnia własności 1) i 2) z definicji $\mathbb{E}(h(X)|\sigma(Y))$. Pierwszy warunek jest jasny - η , jako funkcja Y , jest $\sigma(Y)$ -mierzalna. Zajmijmy się zatem drugim warunkiem. Niech $B \in \sigma(Y)$. Ponieważ Y ma rozkład dyskretny, B jest co najwyżej przeliczalną sumą zdarzeń postaci $\{Y = y\}$ oraz zdarzenia o prawdopodobieństwie 0. Wystarczy więc sprawdzić 2) dla zbiorów B postaci $\{Y = y\}$. Mamy

$$\int_{\{Y=y\}} \eta d\mathbb{P} = \int_{\{Y=y\}} \sum_{x \in S_X} h(x) \frac{P_{X,Y}(x, y)}{P_Y(y)} d\mathbb{P} = \sum_{x \in S_X} h(x) P_{X,Y}(x, y)$$

oraz

$$\int_{\{Y=y\}} h(X) d\mathbb{P} = \sum_{x \in S_X} h(x) \int_{\{Y=y\}} 1_{\{X=x\}} d\mathbb{P} = \sum_{x \in S_X} h(x) P_{X,Y}(x, y).$$

2. Konkretny przykład. Załóżmy, że X, Y są niezależnymi zmiennymi losowymi o rozkładzie Poissona z parametrami λ, μ , odpowiednio. Wyznamy $\mathbb{E}(X|X+Y)$.

Wiadomo, że $X + Y$ ma rozkład Poissona z parametrem $\lambda + \mu$. Stąd

$$P_{X+Y}(k) = \frac{(\lambda + \mu)^k}{k!} e^{-(\lambda + \mu)}, \quad k = 0, 1, 2, \dots$$

Ponadto, jeśli $k \geq \ell \geq 0$, to

$$\begin{aligned} P_{X, X+Y}(\ell, k) &= \mathbb{P}(X = \ell, X + Y = k) = \mathbb{P}(X = \ell) \mathbb{P}(Y = k - \ell) \\ &= \frac{\lambda^\ell}{\ell!} e^{-\lambda} \cdot \frac{\mu^{k-\ell}}{(k-\ell)!} e^{-\mu} \end{aligned}$$

i

$$\frac{P_{X, X+Y}(\ell, k)}{P_{X+Y}(k)} = \frac{k! \lambda^\ell \mu^{k-\ell}}{\ell! (k-\ell)! (\lambda + \mu)^k} = \binom{k}{\ell} \left(\frac{\lambda}{\lambda + \mu} \right)^\ell \left(1 - \frac{\lambda}{\lambda + \mu} \right)^{k-\ell}.$$

Stąd

$$\mathbb{E}(X|X+Y) = \frac{\lambda}{\lambda + \mu} (X + Y).$$

3. Załóżmy, że (X, Y) ma rozkład z gęstością g i niech $g_Y(y) = \int_{\mathbb{R}} g(x, y) dx$ będzie gęstością zmiennej Y . Zdefiniujmy *gęstość warunkową* wzorem

$$g_{X|Y}(x|y) = \begin{cases} \frac{g(x, y)}{g_Y(y)} & \text{jeśli } g_Y(y) \neq 0, \\ 0 & \text{jeśli } g_Y(y) = 0. \end{cases}$$

Wówczas dla dowolnej funkcji borelowskiej $h : \mathbb{R} \rightarrow \mathbb{R}$ takiej, że $h(X) \in L^1$ mamy

$$(*) \quad \mathbb{E}(h(X)|Y) = \int_{\mathbb{R}} h(x) g_{X|Y}(x|Y) dx.$$

Istotnie, sprawdzimy, że prawa strona spełnia warunki 1) i 2) z definicji $\mathbb{E}(h(X)|Y)$. Oczywiście warunek 1) jest spełniony - prawa strona jest funkcją od Y . Przejdźmy do 2). Dla dowolnego $B \in \sigma(Y)$ mamy, iż $B = \{Y \in A\}$ dla pewnego $A \in \mathbb{R}$ oraz

$$\begin{aligned} \int_B h(X) d\mathbb{P} &= \int_{\Omega} 1_{\{Y \in A\}} h(X) d\mathbb{P} = \int_{\mathbb{R}^2} 1_{\{y \in A\}} h(x) g(x, y) dx dy \\ &= \int_{\mathbb{R}} 1_{\{y \in A\}} g_Y(y) \int_{\mathbb{R}} h(x) g_{X|Y}(x|y) dx dy = \int_B \int_{\mathbb{R}} h(x) g_{X|Y}(x|Y) dx d\mathbb{P}. \end{aligned}$$

4. Konkretny przykład. Załóżmy, że (X, Y) ma rozkład jednostajny na trójkącie

$$D = \{(x, y) : 0 \leq x \leq y \leq 1\}.$$

Obliczymy $\mathbb{E}(X|Y)$ oraz $\mathbb{P}(X \leq 1/2|Y)$.

Mamy $g(x, y) = 21_{\{0 \leq x \leq y \leq 1\}}$ oraz

$$g_Y(y) = \int_{\mathbb{R}} g(x, y) dx = 2y1_{[0,1]}(y).$$

Wobec tego, gęstość warunkowa $g_{X|Y}$ zadana jest wzorem

$$g_{X|Y}(x|y) = \begin{cases} \frac{1}{y}1_{[0,y]}(x) & \text{jeśli } y \in (0, 1], \\ 0 & \text{dla pozostałych } y. \end{cases}$$

Stąd

$$\mathbb{E}(X|Y) = \int_{\mathbb{R}} x g_{X|Y}(x|Y) dx = \frac{1}{Y} \int_0^Y x dx = \frac{Y}{2}$$

oraz

$$\begin{aligned} \mathbb{P}(X \leq 1/2|Y) &= \mathbb{E}[1_{(-\infty, 1/2]}(X)|Y] \\ &= \int_{\mathbb{R}} 1_{(-\infty, 1/2]}(x) g_{X|Y}(x|Y) dx \\ &= \frac{1}{Y} \int_0^Y 1_{(-\infty, 1/2]}(x) 1_{[0,Y]}(x) dx \\ &= \begin{cases} 1 & \text{jeśli } Y \leq 1/2, \\ 1/(2Y) & \text{jeśli } Y > 1/2. \end{cases} \end{aligned}$$

Własności warunkowej wartości oczekiwanej

Założmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest ustaloną przestrzenią probabilistyczną i niech \mathcal{M} będzie pewnym pod- σ -ciałem \mathcal{F} . Ponadto, o wszystkich zmiennych losowych warunkowanych zakładamy, że są całkowalne.

0. Mamy $\mathbb{E}(\mathbb{E}(X|\mathcal{M})) = \mathbb{E}X$. Wynika to natychmiast z 2), jeśli weźmiemy $B = \Omega$.

Przykład: Liczba wypadków danego dnia w pewnym mieście ma rozkład Poissona z parametrem 5. Wysokość szkody powstałej w wypadku ma rozkład jednostajny na przedziale $[2, 10]$. Niech X oznacza łączną szkodę danego dnia. Wyznaczyć $\mathbb{E}X$.

Rozwiązanie: Wprowadźmy zmienną losową Y , zadaną jako liczbę wypadków danego dnia. Zmienna Y ma rozkład Poissona z parametrem 5, ponadto, z warunków

zadania, $\mathbb{E}(X|Y) = 6Y$. Istotnie, średnia wysokość szkody powstałej w pojedynczym wypadku wynosi 6, a więc jeśli było Y wypadków, to średnia szkoda wynosi $6Y$. Zatem, korzystając z własności 0.,

$$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}6Y = 30.$$

1. Niech $\alpha, \beta \in \mathbb{R}$. Wówczas

$$\mathbb{E}(\alpha X_1 + \beta X_2 | \mathcal{M}) = \alpha \mathbb{E}(X_1 | \mathcal{M}) + \beta \mathbb{E}(X_2 | \mathcal{M}).$$

Istotnie: sprawdzimy, że prawa strona (oznaczana dalej przez R) spełnia warunki 1) i 2) z definicji $\mathbb{E}(\alpha X_1 + \beta X_2 | \mathcal{M})$. Pierwszy warunek jest oczywisty. Aby sprawdzić drugi zauważmy, że dla dowolnego $B \in \mathcal{M}$,

$$\begin{aligned} \int_B R d\mathbb{P} &= \alpha \int_B \mathbb{E}(X_1 | \mathcal{M}) d\mathbb{P} + \beta \int_B \mathbb{E}(X_2 | \mathcal{M}) d\mathbb{P} = \alpha \int_B X_1 d\mathbb{P} + \beta \int_B X_2 d\mathbb{P} \\ &= \int_B \alpha X_1 + \beta X_2 d\mathbb{P}. \end{aligned}$$

2. Jeśli X jest nieujemną zmienną losową, to $\mathbb{E}(X | \mathcal{M}) \geq 0$ p.n. Istotnie, niech $B = \{\mathbb{E}(X | \mathcal{M}) < 0\}$. Wówczas $B \in \mathcal{M}$ i

$$\int_B \mathbb{E}(X | \mathcal{M}) d\mathbb{P} = \int_B X d\mathbb{P}.$$

Widzimy, że gdyby zdarzenie B miało dodatnie prawdopodobieństwo, to lewa strona byłaby ujemna, a prawa - nieujemna.

3. Mamy

$$(*) \quad |\mathbb{E}(X | \mathcal{M})| \leq \mathbb{E}(|X| | \mathcal{M}) \quad \text{p.n.}$$

Istotnie, na mocy 1. oraz 2. mamy, iż nierówność $X \leq Y$ p.n. pociąga za sobą $\mathbb{E}(X | \mathcal{M}) \leq \mathbb{E}(Y | \mathcal{M})$. Stąd, z prawdopodobieństwem 1,

$$\mathbb{E}(X_1 | \mathcal{M}) \leq \mathbb{E}(|X_1| | \mathcal{M})$$

i

$$-\mathbb{E}(X_1 | \mathcal{M}) \leq \mathbb{E}(|X_1| | \mathcal{M}).$$

Uwaga: Biorąc wartość oczekiwaną obu stron w (*) dostajemy, na mocy 0.,

$$\mathbb{E}(|\mathbb{E}(X | \mathcal{M})|) \leq \mathbb{E}|X|.$$

Innymi słowy, operator liniowy $\mathbb{E}(\cdot | \mathcal{M}) : L^1(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow L^1(\Omega, \mathcal{F}, \mathbb{P})$ jest kontrakcją.

4. Warunkowa wersja twierdzenia Lebesgue'a o monotonicznym przejściu do granicy. Załóżmy, że X_n jest niemalejącym ciągiem nieujemnych zmiennych losowych zbieżnych p.n. do $X \in L^1$. Wówczas $\mathbb{E}(X_n | \mathcal{M}) \uparrow \mathbb{E}(X | \mathcal{M})$ p.n.

Aby to wykazać, zacznijmy od obserwacji iż na mocy 1. i 2., ciąg $(\mathbb{E}(X_n | \mathcal{M}))$ jest z prawdopodobieństwem 1 niemalejący, a więc w szczególności zbieżny. Oznaczmy jego granicę przez η , $\mathbb{E}(X_1 | \mathcal{M}) \leq \eta \leq \infty$. Niech teraz $B \in \mathcal{M}$. Mamy, na mocy 2) oraz bezwarunkowego twierdzenia Lebesgue'a,

$$\int_B X = \lim_{n \rightarrow \infty} \int_B X_n = \lim_{n \rightarrow \infty} \int_B \mathbb{E}(X_n | \mathcal{M}) = \int_B \eta.$$

Ponieważ η jest \mathcal{M} -mierzalna, to z powyższej równości wynika, iż $\eta = \mathbb{E}(X | \mathcal{M})$.

5. Analogicznie dowodzimy warunkowe wersje twierdzenia Lebesgue'a o zmajorzowanym przejściu do granicy pod znakiem całki oraz lematu Fatou.

6. Załóżmy, że X_1 jest mierzalna względem \mathcal{M} . Wówczas

$$(+) \quad \mathbb{E}(X_1 X_2 | \mathcal{M}) = X_1 \mathbb{E}(X_2 | \mathcal{M}) \quad \text{p.n.}$$

W szczególności, biorąc $X_2 \equiv 1$, dostajemy, iż $\mathbb{E}(X_1 | \mathcal{M}) = X_1$.

Sprawdzamy, że prawa strona spełnia warunki 1) oraz 2) z definicji $\mathbb{E}(X_1 X_2 | \mathcal{M})$. Warunek 1) jest oczywisty, pozostaje więc sprawdzić drugi. Zastosujemy metodę komplikacji zmiennej X_1 .

a) Jeśli $X_1 = 1_A$, gdzie $A \in \mathcal{M}$, to dla dowolnego $B \in \mathcal{M}$,

$$\int_B X_1 \mathbb{E}(X_2 | \mathcal{M}) d\mathbb{P} = \int_{A \cap B} \mathbb{E}(X_2 | \mathcal{M}) d\mathbb{P} = \int_{A \cap B} X_2 d\mathbb{P} = \int_B X_1 X_2 d\mathbb{P}.$$

b) Jeśli X_1 jest zmienną prostą, to wzór (+) dostajemy na mocy a) oraz liniowości warunkowych wartości oczekiwanych.

c) Jeśli X_1 jest nieujemną zmienną losową, to istnieje niemalejący ciąg (Y_n) \mathcal{M} -mierzalnych zmiennych prostych, zbieżny p.n. do X_1 . Rozbijmy $X_2 = X_2^+ - X_2^-$ i zastosujmy b) do zmiennych Y_n oraz X_2^+ :

$$\mathbb{E}(Y_n X_2^+ | \mathcal{M}) = Y_n \mathbb{E}(X_2^+ | \mathcal{M}).$$

Zbiegając z $n \rightarrow \infty$ i korzystając z warunkowej wersji twierdzenia Lebesgue'a (własność 4.), dostajemy

$$\mathbb{E}(X_1 X_2^+ | \mathcal{M}) = X_1 \mathbb{E}(X_2^+ | \mathcal{M}).$$

Zastępując X_2^+ przez X_2^- i powtarzając rozumowanie, dostajemy

$$\mathbb{E}(X_1 X_2^- | \mathcal{M}) = X_1 \mathbb{E}(X_2^- | \mathcal{M})$$

i po odjęciu stronami dostajemy (+).

d) Jeśli X_1 jest dowolną zmienną losową, to rozbijamy ją na różnicę $X_1^+ - X_1^-$, stosujemy c) do zmiennych X_1^+ , X_2 , oraz X_1^- , X_2 , i odejmujemy stronami uzyskane równości.

7. Jeśli $\mathcal{M}_1 \subset \mathcal{M}_2$ są pod- σ -ciałami \mathcal{F} , to

$$(=) \quad \mathbb{E}(X | \mathcal{M}_1) = \mathbb{E}(\mathbb{E}(X | \mathcal{M}_2) | \mathcal{M}_1) = \mathbb{E}(\mathbb{E}(X | \mathcal{M}_1) | \mathcal{M}_2).$$

Zacznijmy od obserwacji, iż wyrażenia stojące po skrajnych stronach są równe. Wynika to natychmiast z poprzedniej własności: zmienna losowa $\mathbb{E}(X | \mathcal{M}_1)$ jest mierzalna względem \mathcal{M}_2 . Wystarczy więc udowodnić, że pierwsze dwa wyrazy w (=) są równe. Weźmy $B \in \mathcal{M}_1$. Mamy $B \in \mathcal{M}_2$, a więc

$$\int_B \mathbb{E}(X | \mathcal{M}_1) = \int_B X = \int_B \mathbb{E}(X | \mathcal{M}_2) = \int_B \mathbb{E}(\mathbb{E}(X | \mathcal{M}_2) | \mathcal{M}_1),$$

skąd teza.

8. Załóżmy, że X jest niezależna od \mathcal{M} . Wówczas $\mathbb{E}(X | \mathcal{M}) = \mathbb{E}X$. Istotnie, sprawdzimy, że $\mathbb{E}X$ spełnia warunki 1) i 2) w definicji $\mathbb{E}(X | \mathcal{M})$. Warunek 1) jest oczywisty: $\mathbb{E}X$ jest zmienną losową stałą, a więc mierzalną względem każdego σ -ciała. Niech teraz $B \in \mathcal{M}$. Mamy na mocy niezależności 1_B oraz X ,

$$\int_B \mathbb{E}X d\mathbb{P} = \mathbb{E}1_B \mathbb{E}X = \mathbb{E}(1_B X) = \int_B X d\mathbb{P}.$$

9. Nierówność Jensena. Załóżmy, że $f : \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją wypukłą taką, że $f(X)$ jest zmienną całkowalną. Wówczas

$$\mathbb{E}(f(X) | \mathcal{M}) \geq f(\mathbb{E}(X | \mathcal{M})).$$

Będzie nam potrzebny następujący prosty fakt. Dowód pozostawiamy jako proste ćwiczenie.

Lemat 15.2. *Załóżmy, że $f : \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją wypukłą. Wówczas istnieją ciągi $(a_n), (b_n)$ takie, że dla dowolnego $x \in \mathbb{R}$,*

$$f(x) = \sup_n (a_n x + b_n).$$

Powróćmy do dowodu 9. Dla ciągów $(a_n), (b_n)$, gwarantowanych przez powyższy lemat, mamy $f(X) \geq a_n X + b_n$ dla każdego n . Stąd, na mocy 1. oraz 2., z prawdopodobieństwem 1,

$$\mathbb{E}(f(X)|\mathcal{M}) \geq a_n \mathbb{E}(X|\mathcal{M}) + b_n.$$

Ponieważ ciągi $(a_n), (b_n)$ są przeliczalne, to możemy wziąć supremum po n po prawej stronie i dalej nierówność będzie zachodziła z prawdopodobieństwem 1:

$$\mathbb{E}(f(X)|\mathcal{M}) \geq \sup_n (a_n \mathbb{E}(X|\mathcal{M}) + b_n) = f(\mathbb{E}(X|\mathcal{M})).$$

Uwaga: Jako wniosek, dostajemy, iż dla $p \geq 1$ i $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{E}(|X|^p|\mathcal{M}) \geq [\mathbb{E}(|X||\mathcal{M})]^p.$$

Stąd po wzięciu wartości oczekiwanej obu stron, $\mathbb{E}(|\mathbb{E}(X|\mathcal{M})|^p) \leq \mathbb{E}|X|^p$, czyli

$$\|\mathbb{E}(X|\mathcal{M})\|_p \leq \|X\|_p.$$

Zatem warunkowa wartość oczekiwana $\mathbb{E}(\cdot|\mathcal{M})$ jest kontrakcją w L^p .

Na zakończenie zajmiemy się zagadnieniem regresji nieliniowej. Załóżmy, że X, Y są zmiennymi losowymi całkowalnymi z kwadratem. Obserwujemy zmienną Y i za pomocą tych danych chcemy najlepiej przybliżyć X (w sensie średniokwadratowym) zmienną postaci $h(Y)$. Ścisłej, szukamy funkcji borelowskiej f takiej, że

$$\mathbb{E}(X - f(Y))^2 = \min_h \mathbb{E}(X - h(Y))^2.$$

W przypadku, gdy zawężymy się do klasy funkcji liniowych, prowadzi to do zagadnienia regresji liniowej, rozważanej wcześniej.

Twierdzenie 15.3. *Rozwiązaniem powyższego zagadnienia jest $f(Y) = \mathbb{E}(X|Y)$.*

Dowód. Weźmy dowolną funkcję borelowską h . Mamy

$$\begin{aligned} \mathbb{E}(X - h(Y))^2 &= \mathbb{E}(X - f(Y) + f(Y) - h(Y))^2 \\ &= \mathbb{E}(X - f(Y))^2 + 2\mathbb{E}(X - f(Y))(f(Y) - h(Y)) + \mathbb{E}(f(Y) - h(Y))^2. \end{aligned}$$

Ale zmienna $f(Y) - h(Y)$ jest mierzalna względem $\sigma(Y)$. Zatem korzystając z własności 0., 1. oraz 6.,

$$\begin{aligned} \mathbb{E}(X - f(Y))(f(Y) - h(Y)) &= \mathbb{E}\left\{\mathbb{E}[(X - f(Y))(f(Y) - h(Y))|Y]\right\} \\ &= \mathbb{E}\left\{(f(Y) - h(Y))\mathbb{E}(X - f(Y)|Y)\right\} = 0. \end{aligned}$$

Wobec tego człon środkowy w poprzednim ciągu równości znika i otrzymujemy

$$\mathbb{E}(X - h(Y))^2 \geq \mathbb{E}(X - f(Y))^2.$$

Stąd teza. □

16. ZADANIA

1. Zmienne losowe $\varepsilon_1, \varepsilon_2, \varepsilon_3$ są niezależne i mają ten sam rozkład $\mathbb{P}(\varepsilon_i = -1) = \mathbb{P}(\varepsilon_i = 1) = 1/2, i = 1, 2, 3$. Obliczyć $\mathbb{E}(\varepsilon_1|\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$ oraz $\mathbb{E}(\varepsilon_1\varepsilon_2|e_1 + e_2e_3)$.

2. Zmienne losowe X, Y są niezależne, przy czym X ma rozkład Bernoulliego $B(n, p)$, a Y ma rozkład Bernoulliego $B(m, p)$. Wyznaczyć $\mathbb{E}(X + Y|X)$ oraz $\mathbb{E}(X|X + Y)$.

3. Rzucono kostką, a następnie rzucono nią tyle razy, ile oczek wypadło w pierwszym rzucie. Obliczyć wartość oczekiwaną liczby wyrzuconych trójek.

4. W urnie znajduje się a kul białych, b kul czarnych i c kul czerwonych (a, b, c są dodatnimi liczbami całkowitymi). Losujemy ze zwracaniem po jednej kuli aż do momentu wyciągnięcia kuli czerwonej. Wyznaczyć wartość oczekiwaną liczby losowań w których wyciągnięto białą kulę.

5. Wiadomo, że p procent monet stanowią monety fałszywe, z orłem po obu stronach. Losujemy ze zwracaniem n monet i każdą z nich wykonujemy rzut. Niech F oznacza liczbę losowań, w wyniku których wyciągnięto monetę fałszywą, O - liczba wyrzuconych orłów. Udowodnić, że $\mathbb{E}(F|O) = \frac{2p}{100+p}O$.

6. Zmienna losowa (X, Y) ma gęstość

$$g(x, y) = \frac{x^3}{2}e^{-x(y+1)}1_{\{x>0, y>0\}}.$$

Wyznaczyć $\mathbb{E}(Y|X)$, $\mathbb{E}(Y^2|X^2)$ oraz $\mathbb{P}(Y > 1|X^3 + 1)$.

7. Zmienne losowe X, Y są niezależne i mają rozkład wykładniczy z parametrem 1. Obliczyć $\mathbb{P}(X \in B|X + Y)$ (dla $B \in \mathcal{B}(\mathbb{R})$) oraz $\mathbb{E}(\sin X|X + Y)$.

8. Zmienna losowa X ma rozkład wykładniczy z parametrem 1, zaś Y jest zmienną losową taką, że jeśli $X = x$, to Y ma rozkład wykładniczy z parametrem x .

a) Wyznaczyć rozkład Y .

b) Obliczyć $\mathbb{P}(X > r|Y)$.

9. Zmienna losowa (X, Y) ma rozkład normalny o wartości oczekiwanej 0, $\text{Var}X = \sigma_1^2$, $\text{Var}Y = \sigma_2^2$, $\text{Cov}(X, Y) = c$. Obliczyć $\mathbb{P}(Y \in B|X)$ (dla $B \in \mathcal{B}(\mathbb{R})$) oraz $\mathbb{E}(Y|X)$.

10. Zmienne X_1, X_2, \dots, X_n są niezależne i mają ten sam rozkład o skończonej wartości oczekiwanej. Obliczyć $\mathbb{E}(X_1|X_1 + X_2 + \dots + X_n)$.

11. Załóżmy, że X, Y są zmiennymi losowymi a \mathcal{G} jest σ -ciałem takim, że X jest mierzalne względem \mathcal{G} , a Y jest niezależne od \mathcal{G} . Niech $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ będzie funkcją borelowską taką, że $\phi(X, Y)$ jest całkowalną zmienną losową. Udowodnić, że

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

gdzie $\Phi(x) = \mathbb{E}\phi(x, Y)$.

12. Załóżmy, że X jest całkowalną zmienną losową, a σ -ciało \mathcal{G} jest niezależne od X oraz od σ -ciała \mathcal{M} . Udowodnić, że

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{M})) = \mathbb{E}(X|\mathcal{M}).$$

13. Zmienne X, Y, Z są niezależne, przy czym X ma standardowy rozkład normalny, Y jest nieujemną zmienną ograniczoną, a Z ma rozkład Rademachera. Obliczyć $\mathbb{E}(e^{XY}|Y)$ oraz $\mathbb{E}(e^{XY}|YZ)$.

14. Zmienne N, X_1, X_2, \dots są niezależne, przy czym N ma rozkład Poissona z parametrem 3, a X_n ma rozkład jednostajny na $[0, 1]$, $n = 1, 2, \dots$. Obliczyć $\mathbb{E}(X_1 + X_2 + \dots + X_{N+1})$.