Matematyka stosowana

Wykład z Rachunku Prawdopodobieństwa II

Adam Osękowski
ados@mimuw.edu.pl
http://www.mimuw.edu.pl/~ados

Uniwersytet Warszawski, 2011

Streszczenie. Celem niniejszego skryptu jest wprowadzenie i omówienie podstawowych obiektów i pojęć pojawiających się w drugiej części kursowego wykładu z Rachunku Prawdopodobieństwa.

Wersja internetowa wykładu:
http://mst.mimuw.edu.pl/lecture.php?lecture=rp2

(może zawierać dodatkowe materiały)

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

Projekt współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego.

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Skład w systemie LaTeX, z wykorzystaniem m.in. pakietów beamer oraz listings. Szablony podręcznika i prezentacji: Piotr Krzyżanowski; koncept: Robert Dąbrowski.

Spis treści

1. Zbieżność według rozkładu – zbieżność miar probabilistycznych w przestrzeniach metrycznych

Celem tego rozdziału jest wprowadzenie pewnego nowego typu zbieżności zmiennych losowych, tzw. *zbieżności według rozkładu*. Zacznijmy od pewnych intuicji związanych z tym pojęciem. Jak sama nazwa wskazuje, zbieżność ta odnosi się do rozkładów zmiennych losowych. Zatem, aby ją zdefiniować (na początek, dla rzeczywistych zmiennych losowych), potrzebujemy metody pozwalającej stwierdzić czy dwa rozkłady prawdopodobieństwa na $\mathbb{R}$ są „bliskie". Jeśli tak na to spojrzeć, to automatycznie narzuca się użycie tzw. całkowitej wariacji miary. Ściślej, definiujemy odległość dwóch miar probabilistycznych μ, ν na $\mathbb{R}$ jako całkowitą wariację ich różnicy:

$$||\mu - \nu|| = \sup \sum_{n=1}^{\infty} |\mu(A_n) - \nu(A_n)|,$$

gdzie supremum jest wzięte po wszystkich rozbiciach prostej rzeczywistej na przeliczalną liczbę zbiorów borelowskich $(A_n)_{n=1}^{\infty}$. I teraz mówimy, że X_n zbiega do X jeśli $||P_{X_n} - P_X|| \to 0$ gdy $n \to \infty$.

To podejście jest jednak zbyt restrykcyjne i zbieżność według rozkładu wprowadzimy w inny sposób. W całym niniejszym rozdziale, (E, ρ) jest przestrzenią metryczną, $\mathcal{B}(E)$ oznacza klasę podzbiorów borelowskich E oraz

$$C(E) = \{f : E \to \mathbb{R} \text{ ciągłe i ograniczone}\}.$$

Definicja 1.1. Niech $(P_n)_n$ będzie ciągiem miar probabilistycznych na $\mathcal{B}(E)$ (rozkładów prawdopodobieństwa na E). Mówimy, że ciąg (P_n) jest zbieżny według rozkładu do P (lub słabo zbieżny do P), jeżeli dla każdej funkcji $f \in C(E)$ mamy $\int_E f dP_n \to \int_E f dP$. Oznaczenie: $P_n \Rightarrow P$.

Dowód poprawności definicji: Musimy udowodnić, że jeśli $P_n \Rightarrow P$ oraz $P_n \Rightarrow P'$, to $P = P'$. Innymi słowy, musimy wykazać następujący fakt.

Stwierdzenie 1.1. *Załóżmy, że P, P' są takimi rozkładami w E, że dla każdej funkcji $f \in C(E)$, $\int_E f dP = \int_E f dP'$. Wówczas $P = P'$.*

Przytoczmy pomocniczy fakt z Topologii I.

Lemat 1.1. *Niech F będzie domkniętym podzbiorem E. Wówczas dla każdego $\varepsilon > 0$ istnieje $f \in C(E)$ jednostajnie ciągła spełniająca $0 \leqslant f \leqslant 1$ oraz*

$$f(x) = \begin{cases} 1 & \text{jeśli } x \in F, \\ 0 & \text{jeśli } \rho(x, F) \geqslant \varepsilon. \end{cases}$$

Dowód Stwierdzenia 1.1:. Wystarczy udowodnić, że dla każdego domkniętego $F \subset E$ zachodzi $P(F) = P'(F)$ (teza wynika wówczas prosto z lematu o $\pi - \lambda$ układach). Dla każdego n i $\varepsilon = 1/n$, Lemat 1.1 daje funkcję f_n o odpowiednich własnościach. Widzimy, iż dla każdego $x \in E$, $f_n(x) \to 1_F(x)$, zatem

$$P(F) = \int_E 1_F dP \leftarrow \int_E f_n dP = \int_E f_n dP' \to P'(F). \quad \square$$

Przykłady:

1. Załóżmy, że (a_n) jest ciągiem punktów z $\mathbb{R}^d$ oraz $a \in \mathbb{R}^d$. Wówczas $a_n \to a$ wtedy i tylko wtedy, gdy $\delta_{a_n} \Rightarrow \delta_a$. Istotnie, $a_n \to a$ wtedy i tylko wtedy, gdy dla każdej funkcji $f \in C(E)$ mamy $f(a_n) \to f(a)$, czyli $\int_{\mathbb{R}^d} f d\delta_{a_n} \to \int_{\mathbb{R}^d} f d\delta_a$.
2. Załóżmy, że (P_n) jest ciągiem miar probabilistycznych na $\mathbb{R}$, zadanym przez

$$P_n(\{k/n\}) = 1/n, \qquad k = 1, 2, \ldots, n.$$

Wówczas $P_n \Rightarrow P$, gdzie P jest rozkładem jednostajnym na $[0,1]$. Istotnie, dla dowolnej funkcji $f \in C(\mathbb{R})$,

$$\int_{\mathbb{R}} f dP_n = \sum_{k=1}^{n} f(k/n) \cdot \frac{1}{n} \to \int_0^1 f(x)dx = \int_{\mathbb{R}} f dP.$$

Ważna uwaga: Z tego, że $P_n \Rightarrow P$ **nie wynika**, że dla dowolnego $B \in \mathcal{B}(E)$ mamy $P_n(B) \to P(B)$. Np. weźmy $a \in \mathbb{R}$ oraz ciąg (a_n) liczb rzeczywistych taki, że $a_n > a$ oraz $a_n \to a$. Jak już wiemy, $\delta_{a_n} \to \delta_a$, ale

$$\delta_{a_n}((-\infty, a]) = 0 \not\to 1 = \delta_a((-\infty, a]).$$

Twierdzenie 1.1. *Niech P_n, P $(n = 1, 2, \ldots)$ będą miarami probabilistycznymi na $\mathcal{B}(E)$. Następujące warunki są równoważne.*

a) $P_n \Rightarrow P$.
b) Dla każdej funkcji $f \in C(E)$ jednostajnie ciągłej, $\int_E f dP_n \to \int_E f dP$.
c) Dla każdego domkniętego $F \subset E$, $\limsup_{n\to\infty} P_n(F) \leqslant P(F)$.
d) Dla każdego otwartego $G \subset E$, $\liminf_{n\to\infty} P_n(G) \geqslant P(G)$.
e) Dla każdego $A \in \mathcal{B}(E)$ takiego, że $P(\partial A) = 0$, mamy $\lim_{n\to\infty} P_n(A) = P(A)$.

Dowód:. a) $\Rightarrow$ b) – oczywiste.

b) $\Rightarrow$ c) Ustalmy $\varepsilon > 0$ i niech $F_\varepsilon = \{x \in E : \rho(x, F) \leqslant \varepsilon\}$. Na mocy Lematu 1.1 istnieje $f_\varepsilon \in C(E)$ jednostajnie ciągła, przyjmująca wartości w $[0,1]$, równa 1 na F oraz 0 na F_ε^c. Mamy

$$P_n(F) = \int_F f_\varepsilon dP_n \leqslant \int_E f_\varepsilon dP_n \to \int_E f_\varepsilon dP = \int_{F_\varepsilon} f_\varepsilon dP \leqslant P(F_\varepsilon).$$

Zatem $\limsup_n P_n(F) \leqslant P(F_\varepsilon)$, i z dowolności ε wynika, co trzeba.

c) $\Rightarrow$ a) Wystarczy udowodnić, że dla każdej funkcji $f \in C(E)$,

$$\limsup_n \int_E f dP_n \leqslant \int_E f dP, \tag{1.1}$$

gdyż po zastąpieniu f przez $-f$ dostaniemy $\liminf_n \int_E f dP_n \geqslant \int_E f dP$, a więc w rzeczywistości mamy równość, gdyż $\liminf \leqslant \limsup$.

Zauważmy, że jeśli $f \in C(E)$, to istnieją $a > 0$ oraz $b \in \mathbb{R}$ takie, że $af + b$ przyjmuje wartości w przedziale $(0,1)$. Co więcej, jeśli wykażemy (1.1) dla $af + b$, to nierówność będzie także zachodzić dla f. Innymi słowy, możemy bez straty ogólności założyć, że $0 < f(x) < 1$ dla każdego $x \in E$.

Ustalmy taką funkcję f i weźmy dodatnią liczbę całkowitą k. Rozważmy zbiory

$$A_i = \left\{x \in E : \frac{i-1}{k} \leqslant f(x) < \frac{i}{k}\right\}, \qquad i = 1, 2, \ldots, k.$$

Oczywiście $\bigcup_{i=1}^k A_i = E$ oraz zbiory $A_1, A_2, \ldots, A_k$ są parami rozłączne. Ponadto,

$$L := \sum_{i=1}^k \frac{i-1}{k} P(A_i) \leqslant \int_E f dP = \sum_{i=1}^k \int_{A_i} f dP \leqslant \sum_{i=1}^k \frac{i}{k} P(A_i) =: R.$$

Zauważmy, że

$$A_i = \left\{x : \frac{i-1}{k} \leqslant f(x)\right\} \setminus \left\{x : \frac{i}{k} \leqslant f(x)\right\} =: F_{i-1} \setminus F_i,$$

i $\emptyset = F_k \subset F_{k-1} \subset \ldots F_1 \subset F_0 = E$ jest zstępującym ciągiem zbiorów domkniętych. Zatem $P(A_i) = P(F_{i-1}) - P(F_i)$, $i = 1, 2, \ldots, k$, i podstawiając dostajemy

$$\begin{aligned} L &= \sum_{i=1}^k \frac{i-1}{k}(P(F_{i-1}) - P(F_i)) = \sum_{i=0}^{k-1} \frac{i}{k} P(F_i) - \sum_{i=1}^k \frac{i-1}{k} P(F_i) \\ &= -\frac{k-1}{k} P(F_k) + \frac{1}{k} \sum_{i=1}^{k-1} P(F_i) = \frac{1}{k} \sum_{i=1}^{k-1} P(F_i) \end{aligned}$$

oraz

$$\begin{aligned} R &= \sum_{i=1}^k \frac{i}{k}(P(F_{i-1}) - P(F_i)) = \sum_{i=0}^{k-1} \frac{i+1}{k} P(F_i) - \sum_{i=1}^k \frac{i}{k} P(F_i) \\ &= -P(F_k) + \frac{1}{k} \sum_{i=0}^{k-1} P(F_i) = \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k-1} P(F_i). \end{aligned}$$

Przeprowadzamy analogiczne oszacowania dla $\int_E f dP_n$: w szczególności mamy

$$\int_E f dP_n \leqslant \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k-1} P_n(F_i),$$

skąd wynika, na mocy c),

$$\limsup_n \int_E f dP_n \leqslant \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k-1} \limsup_n P_n(F_i) \leqslant \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k-1} P(F_i) \leqslant \frac{1}{k} + \int_E f dP.$$

Wystarczy tylko zbiec z k do nieskończoności.

c) $\Leftrightarrow$ d): oczywiste po przejściu do dopełnień zbiorów.

c) $\Rightarrow$ e) Załóżmy, że $A \in \mathcal{B}(E)$ spełnia warunek $P(\partial A) = 0$. Ponieważ $\partial A = \overline{A} \setminus \mathrm{int} A$ oraz $\mathrm{int} A \subseteq \overline{A}$, mamy $P(\overline{A}) = P(\mathrm{int} A) = P(A)$. Z drugiej strony, korzystając z c) oraz d), mamy

$$\begin{aligned} P(\overline{A}) &\geqslant \limsup_n P_n(\overline{A}) \geqslant \limsup_n P_n(A) \\ &\geqslant \liminf_n P_n(A) \geqslant \liminf_n P_n(\mathrm{int} A) \geqslant P(\mathrm{int} A), \end{aligned}$$

a zatem wszędzie mamy równości: to oznacza tezę podpunktu e).

e) $\Rightarrow$ c) Weźmy dowolny domknięty zbiór $F \subseteq E$. Dla każdego $\varepsilon > 0$ zbiór $F_\varepsilon = \{x : \rho(x, F) \leqslant \varepsilon\}$ jest domknięty. Ponadto, zbiór $\{\varepsilon > 0 : P(\{x : \rho(x, F) = \varepsilon\}) > 0\}$ jest co najwyżej przeliczalny; zatem istnieje ciąg (ε_n) liczb dodatnich malejący do 0 taki, że $P(\{x : \rho(x, F) = \varepsilon_n\}) = 0$ dla każdego n. Ponieważ $\partial F_\varepsilon \subseteq \{x : \rho(x, F) = \varepsilon\}$, mamy więc $P(\partial F_{\varepsilon_n}) = 0$ dla każdego n, a zatem, korzystając z e), przy ustalonym k,

$$\limsup_n P_n(F) \leqslant \limsup_n P_n(F_{\varepsilon_k}) = P(F_{\varepsilon_k}).$$

Zbiegając z $k \to \infty$, mamy $\varepsilon_k \to 0$ oraz $P(F_{\varepsilon_k}) \to P(F)$, na mocy tego, iż F jest domknięty. $\square$

Stwierdzenie 1.2. *Załóżmy, że P_n, P są rozkładami prawdopodobieństwa w $\mathbb{R}^d$ $(n = 1, 2, \ldots)$, o dystrybuantach F_n, F, odpowiednio. Wówczas $P_n \Rightarrow P$ wtedy i tylko wtedy, gdy $F_n(x) \to F(x)$ dla każdego punktu x, w którym F jest ciągła.*

Dowód:. $\Rightarrow$ Weźmy punkt $x = (x_1, x_2, \ldots, x_d)$ ciągłości dystrybuanty F i niech $A = \{y \in \mathbb{R}^d : y_i \leqslant x_i,\ i = 1, 2, \ldots, d\}$. Zauważmy, iż $P(\partial A) = 0$; w przeciwnym razie F miałaby nieciągłość w punkcie x (istotnie, mielibyśmy

$$\begin{aligned}\lim_{k\to\infty} F(x_1 - \frac{1}{k}, x_2 - \frac{1}{k}, \ldots, x_d - \frac{1}{k}) &= \lim_{k\to\infty} P(\{y \in \mathbb{R}^d : y_i \leqslant x_i - \frac{1}{k}\}) \\ &< P(A) = F(x)\,).\end{aligned}$$

Zatem na mocy podpunktu e) Twierdzenia 1.1, $F_n(x) = P_n(A) \to P(A) = F(x)$.

$\Leftarrow$ Najpierw udowodnimy

Lemat 1.2. *Załóżmy, że E jest przestrzenią metryczną, $\mathcal{K} \subseteq \mathcal{B}(E)$ jest π-układem takim, że każdy zbiór otwarty jest sumą skończoną lub przeliczalną zbiorów z $\mathcal{K}$. Jeśli P_n, P $(n = 1, 2, \ldots)$ są miarami probabilistycznymi na $\mathcal{B}(E)$ takimi, że dla każdego $A \in \mathcal{K}$ mamy $P_n(A) \to P(A)$, to $P_n \Rightarrow P$.*

Dowód. Udowodnimy, że dla każdego zbioru otwartego $G \subseteq E$, $\liminf P_n(G) \geqslant P(G)$. Ustalmy więc zbiór otwarty G oraz $\varepsilon > 0$. Z założeń lematu istnieje skończony ciąg $A_1, A_2, \ldots, A_k$ elementów $\mathcal{K}$ taki, że

$$A_1 \cup A_2 \cup \ldots \cup A_k \subseteq G, \quad \text{oraz} \quad P(G \setminus (A_1 \cup A_2 \cup \ldots \cup A_k)) < \varepsilon.$$

Mamy $P(G\setminus(A_1\cup A_2\cup\ldots\cup A_k)) = P(G) - P(A_1\cup A_2\cup\ldots\cup A_k)$, skąd, na mocy wzoru włączeń i wyłączeń,

$$\begin{aligned}P(G) &< \varepsilon + P(\bigcup_{i=1}^k A_i) = \varepsilon + \sum_{i=1}^k P(A_i) - \sum_{1\leqslant i<j\leqslant k} P(A_i \cap A_j) + \ldots \\ &= \varepsilon + \sum_{i=1}^k \lim_{n\to\infty} P_n(A_i) - \sum_{1\leqslant i<j\leqslant k} \lim_{n\to\infty} P_n(A_i \cap A_j) + \ldots \\ &= \varepsilon + \lim_n P_n(\bigcup_{i=1}^k A_i) \leqslant \varepsilon + \liminf_n P_n(G).\end{aligned}$$

Wystarczy skorzystać z tego, że $\varepsilon > 0$ było dowolne. $\square$

Wracamy do dowodu stwierdzenia. Dla każdego $i = 1, 2, \ldots$ istnieje co najwyżej przeliczalnie wiele hiperpłaszczyzn $H \subset \mathbb{R}^d$ prostopadłych do osi OX_i, o dodatniej mierze P; niech S oznacza dopełnienie sumy wszystkich takich hiperpłaszczyzn (sumujemy także po i). Jak łatwo zauważyć, S jest gęstym podzbiorem $\mathbb{R}^d$ oraz każdy punkt z S jest punktem ciągłości F. Zbiór

$$\mathcal{K} = \{(a, b] = (a_1, b_1] \times (a_2, b_2] \times \ldots (a_d, b_d] : a, b \in S,\ a_i < b_i \text{ dla każdego } i\}$$

jest π-układem i każdy zbiór otwarty jest sumą skończoną lub przeliczalną zbiorów z $\mathcal{K}$. Mamy

$$\begin{aligned}&P_n((a, b]) \\ &= \sum_{\varepsilon_i\in\{0,1\}} (-1)^{d-(\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_d)} F_n(b_1 + \varepsilon_1(b_1 - a_1), \ldots, b_d + \varepsilon_d(b_d - a_d)) \\ &\to \sum_{\varepsilon_i\in\{0,1\}} (-1)^{d-(\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_d)} F(b_1 + \varepsilon_1(b_1 - a_1), \ldots, b_d + \varepsilon_d(b_d - a_d)) \\ &= P((a, b]).\end{aligned}$$

Wystarczy skorzystać z poprzedniego lematu. $\square$

Definicja 1.2. Załóżmy, że X_n, X $(n = 1, 2, \ldots)$ są zmiennymi losowymi o wartościach w E oraz μ jest miarą probabilistyczną na $\mathcal{B}(E)$.

(i) Mówimy, że ciąg (X_n) jest zbieżny według rozkładu do X, jeśli $P_{X_n} \Rightarrow P_X$. Oznaczenie: $X_n \Rightarrow X$ lub $X_n \xrightarrow{\mathcal{D}} X$.

(ii) Mówimy, że ciąg (X_n) jest zbieżny według rozkładu do μ, jeśli $P_{X_n} \Rightarrow \mu$. Oznaczenie $X_n \Rightarrow \mu$ lub $X_n \xrightarrow{\mathcal{D}} \mu$.

Uwagi:

1. W definicji zbieżności według rozkładu, zmienne X_n mogą być określone na różnych przestrzeniach probabilistycznych.
2. Równoważnie, (X_n) zbiega do X według rozkładu wtedy i tylko wtedy, gdy dla każdej funkcji $f \in C(E)$,
$$\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X). \tag{1.2}$$
Ponadto, na mocy podpunktu b) Twierdzenia 1.1, można się ograniczyć w (1.2) do funkcji jednostajnie ciągłych.
3. Słaba zbieżność odnosi się wyłącznie do rozkładów zmiennych losowych. Na przykład, rozważmy ciąg (X_n), zadany na przestrzeni probabilistycznej $([0,1], \mathcal{B}([0,1]), |\cdot|)$ wzorem
$$X_{2n-1} = 1_{[0,1/2]}, \quad X_{2n} = 1_{[1/2,1]}, \qquad n = 1, 2, \ldots.$$
Jak łatwo zauważyć, (X_n) nie jest ani zbieżny prawie na pewno, ani według prawdopodobieństwa. Natomiast z punktu widzenia słabej zbieżności, *jest to ciąg stały*: $P_{X_n} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Ciąg ten zbiega słabo do X_1 oraz do X_2.

Stwierdzenie 1.3. *Załóżmy, że E jest przestrzenią ośrodkową oraz X, X_n, Y_n $(n = 1, 2, \ldots)$ są zmiennymi losowymi o wartościach w E, przy czym dla każdego n, zmienne X_n oraz Y_n są określone na tej samej przestrzeni probabilistycznej. Jeśli $X_n \Rightarrow X$ oraz $\rho(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$, to $Y_n \Rightarrow X$.*

Biorąc $X_n = X$, dostajemy stąd natychmiast następujący fakt.

Wniosek 1.1. *Jeśli (X_n) zbiega do X według prawdopodobieństwa, to zbiega także według rozkładu.*

Dowód Stwierdzenia 1.3. Niech F będzie dowolnym domkniętym podzbiorem przestrzeni E i ustalmy $\varepsilon > 0$. Zbiór $F_\varepsilon = \{x : \rho(x, F) \leqslant \varepsilon\}$ jest domknięty i mamy
$$\begin{aligned} P_{Y_n}(F) &= \mathbb{P}(Y_n \in F,\, \rho(X_n, Y_n) \leqslant \varepsilon) + \mathbb{P}(Y_n \in F,\, \rho(X_n, Y_n) > \varepsilon) \\ &\leqslant \mathbb{P}(X_n \in F_\varepsilon) + \mathbb{P}(\rho(X_n, Y_n) > \varepsilon). \end{aligned}$$
Zatem
$$\limsup_n P_{Y_n}(F) \leqslant \limsup_n P_{X_n}(F_\varepsilon) + 0 \leqslant P_X(F_\varepsilon)$$
i przechodząc z ε do 0 dostajemy $\limsup_n P_{Y_n}(F) \leqslant P_X(F)$. Z dowolności F oraz podpunktu c) Twierdzenia 1.1 wynika teza. □

Definicja 1.3. Niech $\mathcal{P}$ będzie pewnym zbiorem miar probabilistycznych na $\mathcal{B}(E)$. Mówimy, że ten zbiór jest ciasny (jędrny) jeśli dla każdego $\varepsilon > 0$ istnieje zwarty podzbiór K przestrzeni E taki, że $P(K) \geqslant 1 - \varepsilon$ dla każdego $P \in \mathcal{P}$.

Przykład:

Załóżmy, że $(X_i)_{i\in\mathcal{I}}$ jest rodziną zmiennych losowych o wartościach rzeczywistych, takich, że dla pewnego $\alpha > 0$, $a := \sup_{i\in\mathcal{I}} \mathbb{E}|X_i|^\alpha < \infty$. Wówczas rodzina rozkładów $(P_{X_i})_{i\in\mathcal{I}}$ jest ciasna. Istotnie, ustalmy $\varepsilon > 0$ i $L > 0$. Na mocy nierówności Czebyszewa, dla każdego $i \in \mathcal{I}$,

$$P_{X_i}([-L,L]) = \mathbb{P}(|X_i| \leqslant L) = 1 - \mathbb{P}(|X_i| > L) \geqslant 1 - \frac{\mathbb{E}|X_i|^\alpha}{L^\alpha} \geqslant 1 - \frac{a}{L^\alpha} = 1 - \varepsilon,$$

o ile $a/L^\alpha = \varepsilon$; wystarczy więc wziąć $K = [-(a/\varepsilon)^{1/\alpha}, (a/\varepsilon)^{1/\alpha}]$.

Twierdzenie 1.2 (Prochorow)**.**

(i) (Twierdzenie odwrotne) Jeśli $\mathcal{P}$ jest zbiorem ciasnym, to z każdego ciągu elementów $\mathcal{P}$ można wybrać podciąg zbieżny.

(ii) (Twierdzenie proste) Jeśli E jest przestrzenią polską (tzn. ośrodkową i zupełną) i $\mathcal{P}$ ma tę własność, że z każdego ciągu można wybrać podciąg zbieżny, to $\mathcal{P}$ jest zbiorem ciasnym.

Potrzebne nam będą następujące trzy fakty: z Topologii, Analizy Funkcjonalnej oraz Teorii Miary.

Stwierdzenie 1.4. *Załóżmy, że K jest przestrzenią metryczną zwartą. Wówczas $C(K)$ jest ośrodkowa.*

Twierdzenie 1.3 (Riesz)**.** *Załóżmy, że $\varphi : C(K) \to \mathbb{R}$ jest dodatnim funkcjonałem liniowym ciągłym, tzn.*

(i) $\varphi(af + bg) = a\varphi(f) + b\varphi(g)$ dla dowolnych a, $b \in \mathbb{R}$, f, $g \in C(K)$.

(ii) Istnieje stała L taka, że $|\varphi(f)| \leqslant L \sup_{x\in K} |f(x)|$ dla wszystkich $f \in C(K)$.

(iii) Dla dowolnej nieujemnej funkcji $f \in C(K)$ mamy $\varphi(f) \geqslant 0$.

Wówczas istnieje dokładnie jedna miara skończona λ na $\mathcal{B}(K)$ taka, że $\varphi(f) = \int_K f(x)\lambda(dx)$ dla dowolnej funkcji $f \in C(K)$.

Stwierdzenie 1.5 (Regularność)**.** *Załóżmy, że μ jest miarą skończoną na $\mathcal{B}(E)$. Wówczas dla każdego $A \in \mathcal{B}(E)$ istnieje ciąg (F_n) zbiorów domkniętych zawartych w A oraz ciąg (G_n) zbiorów otwartych zawierających A, takie, że $\mu(F_n) \xrightarrow{n\to\infty} \mu(A)$ oraz $\mu(G_n) \xrightarrow{n\to\infty} \mu(A)$.*

Dowód twierdzenia odwrotnego. Załóżmy, że $\mathcal{P}$ jest ciasny. Wobec tego, dla każdego $m = 1, 2, \ldots$ istnieje zwarty podzbiór K_m przestrzeni E taki, że $P(K_m) \geqslant 1 - \frac{1}{m}$ dla wszystkich $P \in \mathcal{P}$. Bez straty ogólności możemy założyć, że ciąg (K_m) jest wstępujący (zastępując ten ciąg, w razie potrzeby, przez ciąg K_1, $K_1 \cup K_2$, $K_1 \cup K_2 \cup K_3$, $\ldots$).

Niech (P_m) będzie ciągiem miar z $\mathcal{P}$. Dla większej przejrzystości dowodu, podzielimy go na kilka części.

1. Na mocy Stwierdzenia 1.4, dla każdego $m = 1, 2, \ldots$, $C(K_m)$ jest przestrzenią ośrodkową. Niech $\{f_{m_r}\}_{r=1,2,\ldots}$ będzie jej przeliczalnym gęstym podzbiorem. Dla każdego m, r, ciąg $(\int_{K_m} f_{m_r} dP_n)_n$ jest ograniczonym ciągiem liczbowym; można z niego wybrać podciąg zbieżny. Stosując metodę przekątniową widzimy, iż istnieje podciąg $(n_1, n_2, \ldots)$ taki, że dla wszystkich m, r, ciąg $(\int_{K_m} f_{m_r} dP_{n_i})_i$ jest zbieżny.

2. Pokażemy, że dla każdego $m = 1, 2, \ldots$ i każdej funkcji $f \in C(K_m)$, ciąg $(\int_{K_m} f dP_{n_i})_i$ jest zbieżny. Ustalmy $\varepsilon > 0$ oraz r takie, że $\sup_{x \in K_m} |f(x) - f_{m_r}(x)| \leqslant \varepsilon/3$. Mamy

$$\begin{aligned}\left|\int_{K_m} f dP_{n_i} - \int_{K_m} f dP_{n_j}\right| \leqslant & \left|\int_{K_m} f dP_{n_i} - \int_{K_m} f_{m_r} dP_{n_i}\right| \\ & + \left|\int_{K_m} f_{m_r} dP_{n_i} - \int_{K_m} f_{m_r} dP_{n_j}\right| \\ & + \left|\int_{K_m} f_{m_r} dP_{n_j} - \int_{K_m} f dP_{n_j}\right|.\end{aligned}$$

Dwa skrajne składniki po prawej stronie szacują się przez $\varepsilon/3$; na przykład, mamy

$$\begin{aligned}\left|\int_{K_m} f dP_{n_i} - \int_{K_m} f_{m_r} dP_{n_i}\right| & \leqslant \int_{K_m} |f - f_{m_r}| dP_{n_i} \\ & \leqslant \sup_K |f - f_{m_r}| P_{n_i}(K_m) \leqslant \varepsilon/3.\end{aligned}$$

środkowy składnik nie przekracza $\varepsilon/3$ o ile i, j są dostatecznie duże; wynika to z definicji podciągu (n_i).

3. Oznaczmy $\varphi_m(f) = \lim_{i\to\infty} \int_{K_m} f dP_{n_i}$, dla $f \in C(K_m)$. Jest oczywiste, że φ spełnia założenia Twierdzenia Riesza. Zatem istnieje miara λ_m na $\mathcal{B}(K_m)$ taka, że $\varphi_m(f) = \int_{K_m} f d\lambda_m$ dla wszystkich $f \in C(K_m)$, $m = 1, 2, \ldots$. Rozszerzmy tę miarę na $\mathcal{B}(E)$, kładąc $\lambda_m(A) = \lambda_m(A \cap K_m)$.

4. Udowodnimy, że dla każdego $A \in \mathcal{B}(E)$ ciąg $(\lambda_m(A))$ spełnia warunek Cauchy'ego. ściślej, wykażemy, że

$$0 \leqslant \lambda_{m_1}(A) - \lambda_{m_2}(A) \leqslant \frac{1}{m_2} \qquad \text{dla } m_1 > m_2 \geqslant 1. \tag{1.3}$$

Najpierw załóżmy, że F jest zbiorem domkniętym i niech $\varepsilon > 0$. Niech f_ε będzie nieujemną funkcją jednostajnie ciągłą pochodzącą z Lematu 1.1. Mamy

$$\begin{aligned}0 & \leqslant \int_{K_{m_1} \setminus K_{m_2}} f_\varepsilon dP_{n_i} = \int_{K_{m_1}} f_\varepsilon dP_{n_i} - \int_{K_{m_2}} f_\varepsilon dP_{n_i} \\ & \leqslant \sup_E |f_\varepsilon| (P_{n_i}(K_{m_1}) - P_{n_i}(K_{m_2})) \leqslant 1 - P_{n_i}(K_{m_2}) \leqslant \frac{1}{m_2}.\end{aligned}$$

Zbiegając teraz z i do nieskończoności dostajemy

$$0 \leqslant \int_{K_{m_1}} f_\varepsilon d\lambda_{m_1} - \int_{K_{m_2}} f_\varepsilon d\lambda_{m_2} = \int_E f_\varepsilon d\lambda_{m_1} - \int_E f_\varepsilon d\lambda_{m_2} \leqslant \frac{1}{m_2}.$$

Weźmy teraz $\varepsilon \to 0$; ponieważ $f_\varepsilon \to 1_A$, otrzymujemy (1.3) dla zbiorów domkniętych, na mocy twierdzenia Lebesgue'a. Aby otrzymać tę nierówność w przypadku ogólnym, posłużymy się regularnością. Dla dowolnego $A \in \mathcal{B}(E)$ istnieją ciągi (F_k') oraz (F_k'') zbiorów domkniętych zawartych w A, takie, że $\lambda_{m_1}(F_k') \to \lambda_{m_1}(A)$ oraz $\lambda_{m_2}(F_k'') \to \lambda_{m_2}(A)$. Korzystając z (1.3) dla zbioru domkniętego $F_k = F_k' \cup F_k''$ i zbiegając z $k \to \infty$ otrzymujemy żądaną nierówność.

5. Wiemy, na mocy poprzedniej części, że ciąg $(\lambda_m(A))_m$ jest zbieżny dla każdego $A \in \mathcal{B}(E)$. Oznaczmy jego granicę przez $\lambda(A)$. Wykażemy, że λ jest miarą probabilistyczną oraz $P_{n_i} \Rightarrow \lambda$. Pierwsza własność wyniknie z następujących trzech faktów.

a) $\lambda(E) = 1$.

b) $\lambda(A_1 \cup A_2) = \lambda(A_1) + \lambda(A_2)$ dla A_1, $A_2 \in \mathcal{B}(E)$ takich, że $A_1 \cap A_2 = \emptyset$.

c) Jeśli $A_1 \supseteq A_2 \supseteq \ldots$ oraz $\bigcap_{k=1}^{\infty} A_k = \emptyset$, to $\lambda(A_k) \to 0$.

Dowód a) Mamy $1 \geqslant P_{n_i}(K_m) = \int_{K_m} 1 dP_{n_i} \geqslant 1 - \frac{1}{m}$. Zbiegając z i do nieskończoności dostajemy $1 \geqslant \lambda_m(E) \geqslant 1 - \frac{1}{m}$, i teraz dążąc z m do nieskończoności otrzymujemy $\lambda(E) = 1$.

Dowód b) Jasne na mocy definicji λ i tego, że λ_m jest miarą dla każdego m.

Dowód c) Na mocy (1.3), mamy $0 \leqslant \lambda(A) - \lambda_m(A) \leqslant \frac{1}{m}$ dla wszystkich $A \in \mathcal{B}(E)$ oraz $m = 1, 2, \ldots$. Zatem, dla dowolnego k,

$$\lambda(A_k) = \lambda(A_k) - \lambda_m(A_k) + \lambda_m(A_k) \leqslant \frac{1}{m} + \lambda_m(A_k).$$

Zbiegając z $k \to \infty$ widzimy, że $\limsup_{k\to\infty} \lambda(A_k) \leqslant 1/m$, co na mocy dowolności m daje $\limsup_k \lambda(A_k) = 0$, czyli $\lim_{k\to\infty} \lambda(A_k) = 0$.

Pozostało już tylko sprawdzić, że $P_{n_i} \Rightarrow \lambda$. Dla usalonej $f \in C(E)$, mamy

$$\begin{aligned}\left|\int_E f dP_{n_i} - \int_E f d\lambda\right| \leqslant \left|\int_{K_m^c} f dP_{n_i}\right| + \left|\int_{K_m} f dP_{n_i} - \int_{K_m} f d\lambda_m\right| \\ + \left|\int_E f d\lambda_m - \int_E f d\lambda\right| = I + II + III.\end{aligned}$$

Na mocy ciasności, $I \leqslant \sup_E |f| \cdot \frac{1}{m}$. Ponadto, z definicji λ_m, $II \to 0$ gdy $m \to \infty$. Wreszcie,

$$III = \left|\int_E f d(\lambda - \lambda_m)\right| \leqslant \sup_E |f|(\lambda(E) - \lambda_m(E)) \leqslant \sup_E |f| \cdot \frac{1}{m}.$$

Zatem $I + II + III \to 0$ gdy $m \to \infty$. Dowód jest zakończony. $\square$

Dowód prostego twierdzenia Prochorowa jest znacznie łatwiejszy i pozostawiamy go jako ćwiczenie (patrz zadanie 13).

Na zakończenie, zaprezentujemy następujące dwa fakty (bez dowodu).

Twierdzenie 1.4 (Skorochod). *Załóżmy, że E jest przestrzenią ośrodkową oraz P_n, P $(n = 1, 2, \ldots)$ są miarami probabilistycznymi na $\mathcal{B}(E)$. Jeśli $P_n \Rightarrow P$, to istnieją zmienne losowe X_n, X $(n = 1, 2, \ldots)$, określone na tej samej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$ takie, że $P_{X_n} = P_n$, $P_X = P$ $(n = 1, 2 \ldots)$ oraz $X_n \to X$ prawie na pewno.*

Twierdzenie 1.5. *Załóżmy, że E jest przestrzenią ośrodkową i niech $\mathcal{M}$ oznacza klasę wszystkich miar probabilistycznych na E. Dla P, $Q \in \mathcal{M}$ definiujemy*

$$\pi(P, Q) = \inf\{\varepsilon > 0 : \forall_{A\in\mathcal{B}(E)}\ Q(A) \leqslant P(A_\varepsilon) + \varepsilon,\ P(A) \leqslant Q(A_\varepsilon) + \varepsilon\}.$$

Wówczas π jest metryką w $\mathcal{M}$ (jest to tzw. metryka Levy-Prochorowa) oraz zbieżność w sensie tej metryki pokrywa się ze zwykłą zbieżnością miar probabilistycznych.

1.1. Zadania

1. Udowodnić, że ciąg $(\mathrm{Exp}(n/(n+1)))$ jest zbieżny według rozkładu do $\mathrm{Exp}(1)$.

2. Dany jest ciąg (X_n) zmiennych losowych zbieżny według rozkładu do zmiennej losowej X. Udowodnić, że ciąg $(\sin X_n)$ jest zbieżny według rozkładu do zmiennej $\sin X$.

3. Czy zmienne losowe posiadające gęstość mogą zbiegać według rozkładu do zmiennej o rozkładzie dyskretnym? Czy zmienne losowe o rozkładach dyskretnych mogą zbiegać do zmiennej o rozkładzie ciągłym?

4. Niech $X_1,\ X_2,\ \ldots$ będą zmiennymi losowymi, przy czym dla $n \geqslant 1$ rozkład zmiennej X_n określony jest następująco:

$$\mathbb{P}\left(X_n = \frac{j}{n}\right) = \frac{2j}{n(n+1)},\quad j = 1,\ 2,\ \ldots,\ n.$$

Udowodnić, że ciąg (X_n) jest zbieżny według rozkładu. Wyznaczyć rozkład graniczny.

5. Niech $B(n,p)$ oznacza rozkład Bernoulliego o n próbach z prawdopodobieństwem sukcesu p, a Pois(λ) - rozkład Poissona z parametrem λ. Wykazać, że jeśli $np_n \to \lambda$, to $B(n,p_n) \Rightarrow Pois(\lambda)$.

6. Zmienne losowe $X_1,\ X_2,\ \ldots$ zbiegają według rozkładu do zmiennej X stałej p.n. Wykazać, że ciąg (X_n) zbiega do X według prawdopodobieństwa.

7. Niech g_n, g oznaczają odpowiednio gęstości rozkładów prawdopodobieństwa μ_n, μ na $\mathbb{R}^N$. Udowodnić, że jeśli $g_n \to g$ p.w., to $\mu_n \Rightarrow \mu$.

8. Niech S będzie przeliczalnym podzbiorem $\mathbb{R}^N$, zaś μ_n, μ - miarami probabilistycznymi skupionymi na S. Wykazać, że jeśli dla każdego $x \in S$ mamy $\mu_n(\{x\}) \to \mu(\{x\})$, to $\mu_n \Rightarrow \mu$.

9. Ciąg dystrybuant (F_n) zbiega punktowo do dystrybuanty ciągłej F. Wykazać, że zbieżność jest jednostajna.

10. Dane są ciągi (X_n), (Y_n) zmiennych losowych, określonych na tej samej przestrzeni probabilistycznej, przy czym (X_n) zbiega według rozkładu do X, a (Y_n) zbiega według rozkładu do zmiennej Y stałej p.n.. Udowodnić, że $(X_n + Y_n)$ zbiega według rozkładu do $X + Y$. Czy teza pozostaje prawdziwa bez założenia o jednopunktowym rozkładzie Y?

11. Dany jest ciąg (X_n) zmiennych losowych przyjmujących wartości w przedziale $[0,1]$. Udowodnić, że jeśli dla każdego $k = 0,\ 1,\ 2,\ \ldots$ mamy $\mathbb{E}X_n^k \xrightarrow{n\to\infty} \frac{1}{k+1}$, to (X_n) jest zbieżny według rozkładu.

12. Załóżmy, że (X_n) jest ciągiem niezależnych zmiennych losowych o rozkładzie Cauchy'ego z parametrem $a > 0$, tzn. z gęstością

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

Udowodnić, że $\frac{1}{n}\max_{k\leqslant n} X_k \Rightarrow \frac{1}{T}$, gdzie T ma rozkład wykładniczy. Wyznaczyć parametr tego rozkładu.

13. Załóżmy, że E jest przestrzenią polską oraz $\mathcal{P}$ jest rodziną miar probabilistycznych na $\mathcal{B}(E)$, taką, że z każdego ciągu jej elementów można wybrać podciąg zbieżny.

(i) Udowodnić, że

$$\forall_{\varepsilon>0}\,\forall_{\delta>0}\,\exists_{x_1,\,x_2,\,\ldots,\,x_n\in E}\,\forall_{P\in\mathcal{P}}\ \ P\Big(\bigcup_{k=1}^{n} B(x_k,\delta)\Big) \geqslant 1-\varepsilon,$$

gdzie $B(x, \delta) = \{y \in E : \rho(x, y) < \delta\}$.

(ii) Wywnioskować z (i) proste twierdzenie Prochorowa (wskazówka: w przestrzeni metrycznej zupełnej zbiór domknięty i całkowicie ograniczony - tzn. dla każdego $\varepsilon > 0$ posiadający skończoną ε-sieć - jest zwarty).

14. Załóżmy, że ciąg (X_n) zbiega według rozkładu do X. Niech $h : \mathbb{R} \to \mathbb{R}$ będzie taką funkcją borelowską, że $\mathbb{P}(X \in \{\text{punkty nieciągłości } h\}) = 0$.

(i) Udowodnić, że $h(X_n) \Rightarrow h(X)$.

(ii) Udowodnić, że jeśli h jest dodatkowo ograniczona, to $\mathbb{E}h(X_n) \xrightarrow{n\to\infty} \mathbb{E}h(X)$.

15. Załóżmy, że ciąg (X_n) zbiega według rozkładu do X. Udowodnić, że

(i) $E|X| \leqslant \liminf_n \mathbb{E}|X_n|$.

(ii) jeśli X_1, X_2, są dodatkowo jednostajnie całkowalne, to $\mathbb{E}X_n \to \mathbb{E}X$.

(iii) jeśli X, X_1, X_2, ... są calkowalne, nieujemne i $\mathbb{E}X_n \xrightarrow{n\to\infty} \mathbb{E}X$, to $X_1, X_2, \ldots$ są jednostajnie całkowalne.

16. Dane są dwa ciągi (X_n) oraz (Y_n) zmiennych losowych, zbieżnych według rozkładu do X oraz Y, odpowiednio.

(i) Czy (X_n, Y_n) zbiega według rozkładu do (X, Y)?

(ii) Jaka jest odpowiedź w (i) jesli dodatkowo przy każdym n zmienne X_n oraz Y_n są niezależne?

2. Funkcje charakterystyczne rozkładów prawdopodobieństwa w $\mathbb{R}^d$

Do tej pory zajmowaliśmy się zmiennymi losowymi o wartościach w $\mathbb{R}^d$ bądź, ogólniej, w przestrzeniach metrycznych (bez żadnej dodatkowej struktury). W tym rozdziale ważną rolę będą pełniły zmienne losowe o wartościach w $\mathbb{C}$.

2.1. Zmienne losowe o wartościach zespolonych.

Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną. Funkcja $X : \Omega \to \mathbb{C}$ jest zmienną losową, jeśli jest zmienną losową przy utożsamieniu $\mathbb{C} = \mathbb{R}^2$ - innymi słowy, jeśli $(X_1, X_2) = (\mathrm{Re}X, \mathrm{Im}X)$ jest zmienną losową w $\mathbb{R}^2$. Jeśli X_1 oraz X_2 są całkowalne (co jest równoważne temu, że $\mathbb{E}|X| = \mathbb{E}\sqrt{X_1^2 + X_2^2} < \infty$), to definiujemy $\mathbb{E}X = \mathbb{E}X_1 + i\mathbb{E}X_2$. Bez trudu dowodzimy, iż mają miejsce następujące fakty.

(i) Mamy $|\mathbb{E}X| \leqslant \mathbb{E}|X|$.

(ii) Zachodzi twierdzenie Lebesgue'a o zmajoryzowanym przejściu do granicy pod znakiem wartości oczekiwanej.

(iii) Dla dowolnych z_1, $z_2 \in \mathbb{C}$ i dowolnych zespolonych zmiennych losowych X, Y takich, że $\mathbb{E}X$, $\mathbb{E}Y$ istnieją, mamy

$$\mathbb{E}(z_1 X + z_2 Y) = z_1 \mathbb{E}X + z_2 \mathbb{E}Y.$$

2.2. Funkcje charakterystyczne

Przechodzimy do definicji głownego pojęcia tego rozdziału.

Definicja 2.1. (i) Załóżmy, że P jest rozkładem prawdopodobieństwa w $\mathbb{R}^d$. Funkcję

$$\varphi_P(t) = \int_{\mathbb{R}^d} e^{i(t,x)} P(dx), \qquad t \in \mathbb{R}^d,$$

nazywamy funkcją charakterystyczną P.

(ii) Załóżmy, że X jest zmienną losową o wartościach w $\mathbb{R}^d$, określoną na $(\Omega, \mathcal{F}, \mathbb{P})$. Wówczas $\varphi_X := \varphi_{P_X}$ nazywamy funkcją charakterystyczną (rozkładu) zmiennej losowej X.

Uwaga: Z twierdzenia o zamianie zmiennych wynika, iż $\varphi_X(t) = \mathbb{E}e^{i(t,X)}$.

Bezpośrednio z definicji widzimy, że funkcja charakterystyczna zmiennej losowej zależy tylko od rozkładu tej zmiennej.

Własności funkcji charakterystycznych.

1) Załóżmy, że X jest d-wymiarową zmienną losową. Wówczas φ_X jest dobrze określona na całym $\mathbb{R}^d$, ponadto $\varphi_X(0) = 1$ oraz

$$|\varphi_X(t)| \leqslant \mathbb{E}|e^{i(t,X)}| = 1 \qquad \text{dla wszystkich } t \in \mathbb{R}^d.$$

2) Załóżmy, że X jest d-wymiarową zmienną losową. Wówczas φ_X jest jednostajnie ciągła na $\mathbb{R}^d$; istotnie, dla $h \in \mathbb{R}^d$,

$$\begin{aligned}\sup_{t\in\mathbb{R}^d} |\varphi_X(t+h) - \varphi_X(t)| &= \sup_{t\in\mathbb{R}^d} |\mathbb{E}e^{i(t+h,X)} - \mathbb{E}e^{i(t,X)}| \\ &\leqslant \sup_{t\in\mathbb{R}^d} \mathbb{E}|e^{i(t+h,X)} - e^{i(t,X)}| \leqslant \mathbb{E}|e^{i(h,X)} - 1| \to 0,\end{aligned}$$

gdy $h \to 0$.

3) Załóżmy, że X jest d-wymiarową zmienną losową. Wówczas φ_X jest dodatnio określona, tzn. dla wszystkich $a_1, a_2, \ldots, a_n \in \mathbb{C}$ oraz $t_1, t_2, \ldots, t_n \in \mathbb{R}^d$,

$$\sum_{j,k} \varphi_X(t_j - t_k) a_j \overline{a_k} \geqslant 0.$$

Istotnie, mamy

$$\begin{aligned}0 &\leqslant \int_{\mathbb{R}^d} \left| \sum_{j=1}^n a_j e^{i(t_j,x)} \right|^2 P_X(dx) = \int_{\mathbb{R}^d} \sum_{j,k} a_j e^{i(t_j,x)} \overline{a_k e^{i(t_k,x)}} P_X(dx) \\ &= \sum_{j,k} a_j \overline{a_k} \int_{\mathbb{R}^d} e^{i(t_j,x)} e^{-i(t_k,x)} P_X(dx) = \sum_{j,k} \varphi_X(t_j - t_k) a_j \overline{a_k}.\end{aligned}$$

Powstaje naturalne pytanie: kiedy funkcja $\varphi : \mathbb{R}^d \to \mathbb{C}$ jest funkcją charakterystyczną pewnego rozkładu? Odpowiedź jest zawarta w następującym twierdzeniu.

Twierdzenie 2.1 (Bochner)**.** *Funkcja $\varphi : \mathbb{R}^d \to \mathbb{C}$ jest funkcją charakterystyczną pewnego rozkładu prawdopodobieństwa w $\mathbb{R}^d$ wtedy i tylko wtedy, gdy jest ciągła, dodatnio określona oraz $\varphi(0) = 1$.*

4) Załóżmy, że X jest d-wymiarową zmienną losową, A jest macierzą $n \times d$ oraz $b \in \mathbb{R}^n$. Wówczas

$$\varphi_{AX+b}(t) = \mathbb{E}e^{i(t,AX+b)} = e^{i(t,b)}\mathbb{E}e^{i(t,AX)} = e^{i(t,b)}\mathbb{E}e^{i(A^Tt,X)} = e^{i(t,b)}\varphi_X(A^Tt).$$

W szczególności, $\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$. Oznacza to, iż jeśli $P_X = P_{-X}$ (rozkład zmiennej jest symetryczny), to φ_X jest rzeczywista.

5) Załóżmy, że X jest rzeczywistą zmienną losową taką, że $\mathbb{E}|X|^k < \infty$ dla pewnej liczby całkowitej dodatniej k. Wówczas φ_X ma k-tą pochodną ciągłą i

$$\varphi_X^{(k)}(t) = i^k \mathbb{E}(e^{itX} X^k).$$

W szczególności, $\varphi_X^{(k)}(0) = i^k \mathbb{E}X^k$.

Weźmy najpierw $k = 1$. Mamy

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathbb{E}\frac{e^{i(t+h)X} - e^{itX}}{h} = \mathbb{E}e^{itX}\left(\frac{e^{ihX} - 1}{h}\right).$$

Zauważmy, że $\lim_{h\to 0} h^{-1}(e^{ihX} - 1) = iX$ oraz

$$\begin{aligned}\left| e^{itX} \frac{e^{ihX} - 1}{h} \right| &\leqslant \frac{|\cos(hX) - 1|}{|h|} + \frac{\sin(hX)}{|h|} \\ &= |X| \left(\left| \sin(hX/2) \frac{\sin(hX/2)}{hX/2} \right| + \frac{|\sin(hX)|}{|hX|} \right) \leqslant 2|X| \in L^1,\end{aligned}$$

zatem z twierdzenia Lebesgue'a wynika teza. Dla $k > 1$ dowód jest analogiczny, opierający się na indukcji.

Zachodzi następujący ogólniejszy fakt: jeśli $X = (X_1, X_2, \ldots, X_d)$ jest d-wymiarową zmienną losową taką, że $\mathbb{E}|X|^k < \infty$, to φ_X ma ciągłe pochodne cząstkowe k-tego rzędu i

$$\frac{\partial^k}{\partial t_1^{j_1} \partial t_2^{j_2} \ldots \partial t_d^{j_d}} \varphi_X(t_1, t_2, \ldots, t_d) = i^k \mathbb{E}(e^{i(t,X)} X_1^{j_1} X_2^{j_2} \ldots X_d^{j_d}).$$

6) Jeśli zmienne $X_1, X_2, \ldots, X_n$ są niezależne, to

$$\varphi_{X_1+X_2+\ldots+X_n}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)\ldots\varphi_{X_n}(t).$$

Istotnie, mamy, iż $e^{i(t,X_1)}, e^{i(t,X_2)}, \ldots, e^{i(t,X_d)}$ są niezależne, skąd

$$\varphi_{X_1+X_2+\ldots+X_n}(t) = \mathbb{E}\prod_{j=1}^{n} e^{i(t,X_j)} = \prod_{j=1}^{n} \mathbb{E}e^{i(t,X_j)} = \prod_{j=1}^{n} \varphi_{X_j}(t).$$

2.3. Przykłady.

(I) Załóżmy najpierw, że X jest d-wymiarową zmienną losową o rozkładzie skokowym i niech S_X oznacza zbiór atomów. Bezpośrednio z definicji mamy, iż

$$\varphi_X(t) = \sum_{x \in S_X} e^{i(t,x)} P_X(\{x\}).$$

W szczególności:
1) Jeśli $P_X = \delta_a$, $a \in \mathbb{R}^d$, to $\varphi_X(t) = e^{i(t,a)}$. Co więcej, jeśli $a = 0$, to $\varphi_X \equiv 1$.
2) Załóżmy, że $P_X =$Pois(λ), $\lambda > 0$. Mamy

$$\varphi_X(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

3) $P_X =$B(n,p). Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Ponieważ $X_1 + X_2 + \ldots + X_n$ ma ten sam rozkład co X, to

$$\begin{aligned}\varphi_X(t) &= \varphi_{X_1+X_2+\ldots+X_n}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)\ldots\varphi_{X_n}(t)\\ &= (\varphi_{X_1}(t))^n = (1 + p(e^{it} - 1))^n.\end{aligned}$$

(II) Załóżmy teraz, że X jest d-wymiarową zmienną losową o rozkładzie ciągłym z gęstością g. Z definicji mamy, iż

$$\varphi_X(t) = \int_{\mathbb{R}^d} e^{i(t,x)} g(x)dx.$$

W szczególności:
4) Jeśli P_X jest rozkładem jednostajnym na przedziale $[a,b]$, to

$$\varphi_X(t) = \frac{1}{b-a}\int_a^b e^{itx}dx = \frac{1}{it(b-a)}(e^{itb} - e^{ita}).$$

Jeśli $b = -a$, to φ_X jest funkcją rzeczywistą i $\varphi_X(t) = \frac{\sin(tb)}{tb}$.

5) Jeśli $P_X = \mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma > 0$, to

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

oraz

$$(*) \qquad \varphi_X(t) = e^{itm} e^{-\sigma^2 t^2/2}$$

(w szczególności, dla standardowego rozkładu normalnego, dostajemy $\varphi(t) = e^{-t^2/2}$).

Istotnie, weźmy X jak wyżej. Zmienna $(X-m)/\sigma$ ma standardowy rozkład normalny i

$$\varphi_X(t) = \varphi_{\sigma \frac{X-m}{\sigma} + m}(t) = \varphi_{(X-m)/\sigma}(\sigma t) e^{itm}.$$

Zatem wystarczy udowodnić wzór (*) dla rozkładu $\mathcal{N}(0,1)$. Załóżmy więc, że X ma ten rozkład i zauważmy najpierw, że φ_X jest funkcją rzeczywistą, gdyż rozkład X jest symetryczny. Zatem

$$\varphi_X(t) = \int_R \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

oraz

$$\begin{aligned} \varphi_X'(t) &= \int_R \sin(tx) \frac{1}{\sqrt{2\pi}} (-x) e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \sin(tx) e^{-x^2/2} \Big|_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_R t \cos(tx) e^{-x^2/2} dx = -t\varphi_X(t). \end{aligned}$$

Dodatkowo, jak wiemy, $\varphi_X(0) = 1$: stąd $\varphi_X(t) = e^{-t^2/2}$.

Ogólniej, jeśli X ma d-wymiarowy rozkład normalny z gęstością

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

(gdzie A to pewna macierz $d \times d$ symetryczna i dodatnio określona, a m jest pewnym wektorem z $\mathbb{R}^d$), to

$$\varphi_X(t) = e^{i(m,t)} e^{(A^{-1}t,t)/2}.$$

Dowód tego faktu przeprowadzimy nieco później.

Przejdziemy teraz do twierdzenia o jednoznaczności: okazuje się, że funkcja charakterystyczna wyznacza rozkład jednoznacznie.

Twierdzenie 2.2 (O jednoznaczności). *Jeśli P, P' są rozkładami prawdopodobieństwa w $\mathbb{R}^d$ takimi, że $\varphi_P(t) = \varphi_{P'}(t)$ dla wszystkich $t \in \mathbb{R}^d$, to $P = P'$.*

Zanim podamy dowód, najpierw sformułujmy wniosek.

Stwierdzenie 2.1. *Zmienne losowe X_1, X_2, ..., X_n są niezależne wtedy i tylko wtedy, gdy*

$$(*) \qquad \varphi_{(X_1, X_2, \ldots, X_n)}(t_1, t_2, \ldots, t_n) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2)\ldots\varphi_{X_n}(t_n)$$

dla wszystkich $(t_1, t_2, \ldots, t_n) \in \mathbb{R}^n$.

Dowód:. $\Rightarrow$ Mamy

$$\begin{aligned}\varphi_{(X_1, X_2, \ldots, X_n)}(t_1, t_2, \ldots, t_n) &= \varphi_{P_{(X_1, X_2, \ldots, X_n)}}(t_1, t_2, \ldots, t_n)\\ &= \varphi_{P_{X_1} \otimes P_{X_2} \otimes \ldots \otimes P_{X_n}}(t_1, t_2, \ldots, t_n)\\ &= \int_{\mathbb{R}^n} \exp\left(i \sum_{j=1}^n t_j x_j\right) P_{X_1}(dx_1) \ldots P_{X_n}(dx_n)\\ &= \prod_{j=1}^n \int_R e^{it_j x_j} P_{X_j}(dx_j) = \prod_{j=1}^n \varphi_{X_j}(t_j),\end{aligned}$$

gdzie w przedostatnim przejściu korzystaliśmy z twierdzenia Fubiniego.

$\Leftarrow$ Korzystając z przed chwilą udowodnionej implikacji, możemy zapisać (*) w postaci

$$\varphi_{P_{(X_1, X_2, \ldots, X_n)}}(t_1, t_2, \ldots, t_n) = \varphi_{P_{X_1} \otimes P_{X_2} \otimes \ldots \otimes P_{X_n}}(t_1, t_2, \ldots, t_n),$$

a więc twierdzenie o jednoznaczności daje

$$P_{(X_1, X_2, \ldots, X_n)} = P_{X_1} \otimes P_{X_2} \otimes \ldots \otimes P_{X_n},$$

czyli niezależność zmiennych $X_1, X_2, \ldots, X_n$. □

W dowodzie twierdzenia o jednoznaczności będziemy potrzebować następującego pomocniczego faktu.

Twierdzenie 2.3 (Weierstrass)**.** *Załóżmy, że $f : \mathbb{R} \to \mathbb{R}$ jest ciągłą funkcją okresową. Wówczas istnieje ciąg (w_n) wielomianów trygonometrycznych o tym samym okresie co f, zbieżny jednostajnie do f. (wielomian trygonometryczny o okresie T to funkcja $w : \mathbb{R} \to \mathbb{R}$ postaci $w(x) = \sum_{k=0}^n [\alpha_k \sin(kx \cdot 2\pi/T) + \beta_k \cos(kx \cdot 2\pi/T)]$.)*

Dowód twierdzenia o jednoznaczności (tylko dla $d = 1$):. Wystarczy udowodnić, że dla dowolnej funkcji $f \in C(\mathbb{R})$ mamy

$$(*) \qquad \int_{\mathbb{R}} f dP = \int_{\mathbb{R}} f dP'.$$

Z założenia, (*) zachodzi dla funkcji $x \mapsto e^{itx}$, $x \in \mathbb{R}$, przy każdym ustalonym $t \in \mathbb{R}$. Zatem, z liniowości, powyższa równość ma miejsce dla dowolnego wielomianu trygonometrycznego; mamy bowiem $\sin(tx) = (e^{itx} - e^{-itx})/(2i)$, $\cos(tx) = (e^{itx} + e^{-itx})/2$. Na mocy twierdzenia Weierstrassa, (*) jest prawdziwa dla dowolnej funkcji ciągłej okresowej. Niech teraz f będzie dowolną funkcją ciągłą i ograniczoną. Istnieje ciąg (f_n) funkcji ciągłych i okresowych o następującej własności:

$$f(x) = f_n(x) \quad \text{dla } x \in [-n, n] \quad \text{oraz} \quad \sup_{x \in \mathbb{R}} |f_n(x)| \leqslant \sup_{x \in \mathbb{R}} |f(x)|.$$

Mamy, na mocy nierówności trójkąta,

$$\begin{aligned}\left|\int_{\mathbb{R}} f dP - \int_{\mathbb{R}} f dP'\right| &\leqslant \int_{\mathbb{R}} |f - f_n| dP + \left|\int_{\mathbb{R}} f_n dP - \int_{\mathbb{R}} f_n dP'\right| + \int_{\mathbb{R}} |f - f_n| dP'\\ &= \int_{[-n,n]^c} |f - f_n| dP + 0 + \int_{[-n,n]^c} |f - f_n| dP'\\ &\leqslant 2 \sup_{x \in \mathbb{R}} |f(x)| \big[P([-n, n]^c) + P'([-n, n]^c)\big] \to 0\end{aligned}$$

gdy $n \to \infty$. Stąd otrzymujemy tezę. W przypadku ogólnym (tzn. dla $d > 1$) dowód jest bardzo podobny; rolę twierdzenia Weierstrassa pełni ogólniejsze twierdzenie Stone'a-Weierstrassa. □

Rozkłady Gaussa (rozkłady normalne) w $\mathbb{R}^d$. Załóżmy, że X ma rozkład normalny w $\mathbb{R}^d$, o wartości ozekiwanej m i macierzy kowariancji Λ. Udowodnimy, że

$$\varphi_X(t) = e^{i(t,m)-(\Lambda t,t)/2}.$$

Istotnie, niech Y_1, Y_2, …, Y_d będą niezależnymi zmiennymi losowymi o standarowym rozkładzie normalnym na $\mathbb{R}$ i niech $Z = BY + m$, gdzie B jest macierzą $d \times d$ i $m \in \mathbb{R}^d$. Mamy

$$\varphi_Y(t) = e^{-|t|^2/2},$$

$$\varphi_Z(t) = e^{i(t,m)}\varphi_Y(B^T t) = e^{i(t,m)-(B^T t, B^T t)/2} = e^{i(t,m)-(BB^T t,t)/2}.$$

Zauważmy, że BB^T jest macierzą symetryczną, nieujemnie określoną. Co więcej, każda macierz symetryczna $d \times d$ nieujemnie określona da się ta zapisać; stąd, dla dowolnej nieujemnie określonej symetrycznej macierzy Λ o wymiarach $d \times d$ i dowolnego wektora $m \in \mathbb{R}^d$, funkcja

$$\varphi(t) = e^{i(t,m)-(\Lambda t,t)/2}$$

jest funkcją charakterystyczną dokładnie jednego rozkładu prawdopodobieństwa w $\mathbb{R}^d$. Rozkłady tej postaci nazywamy rozkładami Gaussa w $\mathbb{R}^d$. Zauważmy, że niektóre rozkłady Gaussa nie mają gęstości.

Bezpośrednio z definicji dostajemy następujące wnioski.

Stwierdzenie 2.2. *Załóżmy, że X ma rozkład Gaussa w $\mathbb{R}^d$, a A jest macierzą $n \times d$ i $m \in \mathbb{R}^n$. Wówczas $AX + m$ ma rozkład Gaussa w $\mathbb{R}^n$.*

Stwierdzenie 2.3. *Jeśli X, Y są niezależnymi zmiennymi losowymi o rozkładach Gaussa w $\mathbb{R}^d$ o wartościach oczekiwanych m_X, m_Y oraz macierzach kowariancji Λ_1, Λ_2, odpowiednio, to $X + Y$ ma rozkład Gaussa w $\mathbb{R}^d$ o wartości średniej $m_X + m_Y$ oraz macierzy $\Lambda_1 + \Lambda_2$.*

Przechodzimy do kolejnego bardzo ważnego faktu, łączącego zbieżność według rozkładu ze zbieżnością funkcji charakterystycznych.

Twierdzenie 2.4 (Lévy - Cramera)**.** *Załóżmy, że P_n ($n = 1, 2, \ldots$) są rozkładami prawdopodobieństwa w $\mathbb{R}^d$.*

(i) Jeśli $P_n \Rightarrow P$, to dla każdego $t \in \mathbb{R}^d$, $\varphi_{P_n}(t) \to \varphi_P(t)$.

(ii) Jeśli dla każdego $t \in \mathbb{R}^d$ mamy $\varphi_{P_n}(t) \to \varphi(t)$, gdzie φ-pewna funkcja ciągła w 0, to $\varphi = \varphi_P$ dla pewnego rozkładu P i $P_n \Rightarrow P$.

Dowód:. (i) Z definicji zbieżności według rozkładu mamy, dla dowolnego $t \in \mathbb{R}^d$,

$$\begin{aligned}\varphi_{P_n}(t) &= \int_{\mathbb{R}^d} \cos(x,t) P_n(dx) + i \int_{\mathbb{R}^d} \sin(t,x) P_n(dx) \\ &\to \int_{\mathbb{R}^d} \cos(x,t) P(dx) + i \int_{\mathbb{R}^d} \sin(t,x) P(dx) = \int_{\mathbb{R}^d} e^{i(t,x)}(dx) = \varphi_P(t).\end{aligned}$$

(ii) Zacznijmy od pomocniczego faktu.

Lemat 2.1. *Jeśli $\varphi_{P_n}(t) \to \varphi(t)$ dla t należących do pewnego otoczenia 0 i φ jest ciągła w 0, to rodzina $\{P_n\}_n$ jest ciasna.*

Dowód lematu:. Wprowadźmy oznaczenie $Q_a = [-a,a]\times[-a,a]\times\ldots[-a,a]\subset\mathbb{R}^d$. Przypuśćmy, wbrew tezie, że rodzina $\{P_n\}$ nie jest ciasna. Wówczas istnieje $\varepsilon>0$ o tej własności, iż przy każdym $k\in\mathbb{N}$ mamy $P_{n_k}(Q_k)<1-\varepsilon$ dla pewnego n_k. Zauważmy, iż $n_k\to\infty$; istotnie, w przeciwnym razie pewna liczba m znalazłaby się w ciągu $(n_k)_k$ nieskończenie wiele razy, co prowadziłoby do nierówności $P_m(\mathbb{R}^d)\leqslant 1-\varepsilon$; sprzeczność.

Ponieważ φ jest ciągła w 0, to $\mathrm{Re}\varphi$ także ma tę własność; ponadto, na mocy zbieżności punktowej, $\mathrm{Re}\varphi(0)=\varphi(0)=1$. Wobec tego istnieje takie $a>0$, że dla każdego $t\in Q_a$ mamy $\varphi_{P_n}(t)\to\varphi(t)$ oraz $\mathrm{Re}\varphi(t)>1-\varepsilon/2$. Dalej,

$$\left|\int_{Q_a}\varphi(t)dt\right|\geqslant\int_{Q_a}\mathrm{Re}\varphi(t)dt\geqslant(1-\varepsilon/2)(2a)^d,$$

$$\begin{aligned}\left|\int_{Q_a}\varphi_{P_{n_k}}(t)dt\right| &= \left|\int_{Q_a}\int_{\mathbb{R}^d}e^{i(t,x)}P_{n_k}(dx)dt\right| = \left|\int_{\mathbb{R}^d}\int_{Q_a}e^{i(t,x)}dtP_{n_k}(dx)\right|\\ &\leqslant \int_{Q_k}\left|\int_{Q_a}e^{i(t,x)}dt\right|P_{n_k}(dx)+\int_{Q_k^c}\left|\int_{Q_a}e^{i(t,x)}dt\right|P_{n_k}(dx)\\ &\leqslant (2a)^dP_{n_k}(Q_k)+T,\end{aligned}$$

gdzie

$$T=\int_{Q_k^c}\left|\int_{Q_a}e^{i(t,x)}dt\right|P_{n_k}(dx)=\int_{Q_k^c}\left|\prod_{j=1}^d\int_{-a}^a e^{it_jx_j}dt_j\right|P_{n_k}(dx).$$

Ustalmy teraz $x\in Q_k^c$. Istnieje współrzędna x_l punktu x większa co do modułu niż k, zatem

$$\prod_{j=1}^d\left|\int_{-a}^a e^{it_jx_j}dt_j\right|\leqslant(2a)^{d-1}\left|\frac{e^{ia_lx_l}-e^{-ia_lx_l}}{ix_l}\right|\leqslant 2(2a)^{a-1}/k.$$

Stąd

$$\begin{aligned}(2a)^dP_{n_k}(Q_k)+T &\leqslant (2a)^dP_{n_k}(Q_k)+2(2a)^{d-1}P_{n_k}(Q_k^c)/k\\ &\leqslant (2a)^d(1-\varepsilon)+2(2a)^{d-1}/k\xrightarrow{k\to\infty}(2a)^d(1-\varepsilon).\end{aligned}$$

Ale na mocy twierdzenia Lebesgue'a, $\int_{Q_a}\varphi_{P_{n_k}}(t)dt\to\int_{Q_a}\varphi(t)dt$; stąd sprzeczność: $(2a)^d(1-\varepsilon/2)<(2a)^a(1-\varepsilon)$. □

Przechodzimy do dowodu części (ii) twierdzenia Levy-Cramera. Powyższy lemat oraz twierdzenie Prochorowa dają istnienie miary probabilistycznej P na $\mathbb{R}^d$ oraz podciągu $(P_{n_k})_k$ zbieżnego słabo do P. Na mocy części (i) twierdzenia Levy-Cramera, mamy $\varphi_{P_{n_k}}(t)\xrightarrow{k\to\infty}\varphi_P(t)$, skąd $\varphi(t)=\varphi_P(t)$. Pozostaje jeszcze tylko udowodnić, że $P_n\Rightarrow P$. Przypuśćmy przeciwnie, iż dla pewnej funkcji $f\in C(R^d)$ mamy $\int_{\mathbb{R}^d}fdP_n\not\to\int_{\mathbb{R}^d}fdP$; stąd, dla pewnego podciągu (m_k),

$$(*)\qquad\int_{\mathbb{R}^d}fdP_{m_k}\to\alpha\neq\int_{\mathbb{R}^d}fdP.$$

Ale na mocy lematu, rodzina (P_{m_k}) także jest ciasna, stąd z twierdzenia Prochorowa istnieje podciąg (m_{k_j}) oraz miara probabilistyczna P' taka, że $P_{m_{k_j}}\Rightarrow P'$. Zatem, korzystając z (i), $\varphi_{P_{m_{k_j}}}(t)\to\varphi_{P'}(t)$, czyli $\varphi_P=\varphi_{P'}$. Sprzeczność z (*) kończy dowód. □

Na zakończenie zaprezentujemy twierdzenie o odwróceniu, pozwalające odczytać gęstość rozkładu za pomocą jego funkcji charakterystycznej. Analogiczny fakt dla zmiennych dyskretnych jest treścią zadania 10 poniżej.

Twierdzenie 2.5. *Załóżmy, że P jest rozkładem prawdopodobieństwa w $\mathbb{R}^d$ o funkcji charakterystycznej φ_P. Wówczas jeśli $\int_{\mathbb{R}^d} |\varphi_P(t)|dt < \infty$, to P ma ciągłą ograniczoną gęstość g daną wzorem*

$$g(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(t,x)} \varphi_P(t)dt.$$

Dowód:. Rozważmy funkcję

$$\begin{aligned} g_\varepsilon(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(t,x)} \varphi_P(t) e^{-|t|^2\varepsilon^2/2} dt \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(t,x)} \int_{\mathbb{R}^d} e^{i(t,y)} P(dy) e^{-|t|^2\varepsilon^2/2} dt \\ &= \frac{1}{(2\pi)^{d/2}\varepsilon^d} \int_{\mathbb{R}^d} \frac{\varepsilon^d}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i(t,y-x)} e^{-|t|^2\varepsilon^2/2} dt P(dy) \\ &= \frac{1}{(2\pi)^{d/2}\varepsilon^d} \int_{\mathbb{R}^d} e^{-|y-x|^2/(2\varepsilon^2)} P(dy), \end{aligned}$$

gdzie w trzecim przejściu skorzystaliśmy z twierdzenia Fubiniego (dzięki czynnikowi $e^{-|t|^2\varepsilon^2/2}$ wyrażenie podcałkowe jest całkowalne), a w czwartym użyliśmy faktu, iż wewnętrzna całka to funkcja charakterystyczna rozkładu $N(0, \varepsilon^{-2})$ w punkcie $y - x$. Jak widać, ostatnie wyrażenie to splot rozkładu P z rozkładem $N(0, \varepsilon^2)$, w punkcie x; innymi słowy, jeśli X, Y są niezależnymi zmiennymi o rozkładach P i $N(0, 1)$, odpowiednio, to $X + \varepsilon Y$ ma rozkład z gęstością g_ε. Na mocy całkowalności funkcji charakterystycznej i twierdzenia Lebesgue'a, mamy $g_\varepsilon(x) \to g(x)$ dla każdego $x \in \mathbb{R}$. Wykażemy teraz, że $\int_{\mathbb{R}^d} g = 1$. Oczywiście, na mocy lematu Fatou, $\int_{\mathbb{R}^d} g \leqslant 1$. By udowodnić nierówność przeciwną, weźmy $\delta > 0$ oraz taką liczbę $M > 0$, by $P((-M, M)) > 1-\delta$. Ponieważ $X + \varepsilon Y \Rightarrow X \sim P$, to

$$1 - \delta \leqslant \liminf_{\varepsilon \to 0+} \mathbb{P}(X + \varepsilon Y \in (-M, M)) = \liminf_{\varepsilon \to 0+} \int_{-M}^{M} g_\varepsilon(x)dx = \int_{-M}^{M} g(x)dx,$$

i z dowolności δ dostajemy, iż g jest gęstością. Wystarczy teraz skorzystać z zadania 7 z pierwszego rozdziału: punktowa zbieżność $g_\varepsilon \to g$ pociąga za sobą, iż $(X + \varepsilon Y)_\varepsilon$ zbiega, przy $\varepsilon \to 0+$, do rozkładu o gęstości g; stąd teza. □

2.4. Zadania

1. Rozstrzygnąć, czy podane niżej funkcje są funkcjami charakterystycznymi i jeśli tak, podać odpowiedni rozkład.

a) $\cos t$, b) $\cos^2 t$, c) $\frac{1}{4}(1 + e^{it})^2$, d) $\frac{1 + \cos t}{2}$, e) $(2 - e^{it})^{-1}$.

2. Niech ϕ_1, ϕ_2, ..., ϕ_n będą funkcjami charakterystycznymi pewnych rozkładów. Udowodnić, iż dowolna kombinacja wypukła tych funkcji jest funkcją charakterystyczną pewnego rozkładu.

3. Dany jest ciąg (X_n) niezależnych zmiennych losowych o tym samym rozkładzie. Zmienna losowa N jest od nich niezależna i ma rozkład Poissona z parametrem λ. Wyznaczyć funkcję charakterystyczną zmiennej $X_1 + X_2 + \ldots + X_N$.

4. Niech ϕ będzie funkcją charakterystyczną pewnego rozkładu. Rostrzygnąć, czy

$$\text{a) } \phi^2, \qquad \text{b) } \mathrm{Re}\phi, \qquad \text{c) } |\phi|^2, \qquad \text{d) } |\phi|$$

są funkcjami charakterystycznymi.

5. Zmienne X, Y są niezależne, przy czym X oraz $X + Y$ mają rozkłady normalne. Udowodnić, że Y ma rozkład normalny lub jest stała p.n..

6. Zmienne losowe X, Y są niezależne, przy czym X ma rozkład jednostajny $U(0, 1)$, a Y ma rozkład zadany przez

$$\mathbb{P}(Y = k) = \frac{1}{n}, \qquad k = 0,\, 1,\, 2,\, \ldots,\, n - 1.$$

Wyznaczyć rozkład zmiennej $X + Y$.

7. Zmienne losowe X_1, X_2, ..., X_n są niezależne i mają ten sam rozkład, przy czym zmienna $X_1 + X_2 + \ldots + X_n$ ma rozkład normalny $\mathcal{N}(0, 1)$. Wyznaczyć rozkład zmiennych X_i.

8. Zmienna losowa X ma rozkład jednostajny $U(-1, 1)$. Czy istnieje niezależna od niej zmienna Y taka, że rozkłady zmiennych $X + Y$ oraz $\frac{1}{2}Y$ są takie same?

9. Funkcja charakterystyczna zmiennej losowej X ma drugą pochodną w zerze. Udowodnić, że $\mathbb{E}X^2 < \infty$.

10. Zmienna losowa X przyjmuje wartości całkowite. Udowodnić, że

$$\mathbb{P}(X = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \phi_X(t) dt, \qquad k \in \mathbb{Z}.$$

11. Udowodnić, że dla $p > 2$ funkcja $\phi(t) = e^{-|t|^p}$ nie jest funkcją charakterystyczną żadnego rozkładu.

12. Udowodnić, że $\phi(t) = e^{-|t|}$ jest funkcją charakterystyczną rozkładu Cauchy'ego w $\mathbb{R}$, tzn. rozkładu o gęstości

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

13. Niech X_1, X_2, ... będą niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na przedziale $[-1, 1]$. Zdefiniujmy

$$Y_n = \frac{\mathrm{sgn}\, X_n}{|X_n|^{1/\alpha}}, \qquad n = 1,\, 2,\, \ldots,$$

gdzie $\alpha \in (0, 2)$ jest ustalone. Udowodnić, że ciąg

$$\frac{Y_1 + Y_2 + \ldots + Y_n}{n^{1/\alpha}}$$

jest zbieżny według rozkładu i wyznaczyć funkcję charakterystyczną rozkładu granicznego.

14. Udowodnić, że jeśli P_n $(n = 1, 2, \ldots)$ są rozkładami Gaussa w $\mathbb{R}^d$ i $P_n \Rightarrow P$, to P jest rozkładem Gaussa.

15. Rzucamy monetą, dla której prawdopodobieństwo wypadnięcia orła wynosi p, aż do momentu, gdy uzyskamy n orłów (łącznie, niekoniecznie pod rząd). Niech X_p oznacza liczbę rzutów. Udowodnić, że $(2pX_p)$ jest zbieżny według rozkładu gdy $p \to 0$.

16. Niech X będzie zmienną losową o funkcji charakterystycznej φ_X. Udowodnić, że następujące warunki są równoważne.
(i) Istnieje $a \neq 0$ takie, że $|\varphi_X(a)| = 1$.
(ii) Istnieją b, $c \in \mathbb{R}$ takie, że zmienna X jest skoncentrowana na zbiorze $\{ck + b : k \in \mathbb{Z}\}$.

17. Dany jest ciąg (X_n) niezależnych zmiennych losowych o tym samym rozkładzie, zadanym przez $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = 1/2$. Wykazać, że szereg $\sum_{n=1}^{\infty} 2^{-n} X_n$ jest zbieżny p.n. i wyznaczyć rozkład sumy tego szeregu.

18. Dla $a \in \mathbb{R}$, niech

$$\varphi_a(t) = \begin{cases} 1 + a|x| & \text{jeśli } |x| \leqslant 1, \\ 1 + a & \text{jeśli } |x| > 1. \end{cases}$$

Dla jakich wartości parametru a funkcja φ_a jest funkcją charakterystyczną rozkładu pewnej zmiennej losowej?

19. Załóżmy, że μ jest rozkładem prawdopodobieństwa o funkcji charakterystycznej φ. Udowodnić, że dla dowolnego $r > 0$ zachodzi nierówność

$$\mu([-r, r]) \geqslant 1 - \frac{r}{2} \int_{-2/r}^{2/r} (1 - \varphi(s))\mathrm{d}s$$

oraz

$$\mu([0, r]) \leqslant r \int_{-\pi/2r}^{\pi/2r} |\varphi(s)|\mathrm{d}s.$$

3. Centralne Twierdzenie Graniczne

Centralne twierdzenie graniczne dotyczy zachowania się rozkładu sum niezależnych zmiennych losowych, przy odpowiedniej normalizacji i pewnych dodatkowych założeniach. Intuicyjnie, suma dużej liczby „małych”, niezależnych zmiennych losowych ma rozkład normalny. Główny wynik niniejszego rozdziału jest następujący.

Twierdzenie 3.1 (Lindeberg)**.** *Załóżmy, że dla każdego n, zmienne X_{1n}, X_{2n}, ..., $X_{r_n n}$ są niezależnymi zmiennymi losowymi o średniej 0, takimi, że*

$$\sum_{k=1}^{r_n} \mathbb{E}X_{kn}^2 \xrightarrow{n\to\infty} 1.$$

Dodatkowo, załóżmy, że jest spełniony warunek Lindeberga

$$(L) \qquad \sum_{k=1}^{r_n} \mathbb{E}X_{kn}^2 1_{\{|X_{kn}|>\varepsilon\}} \xrightarrow{n\to\infty} 0 \quad \text{dla każdego } \varepsilon > 0.$$

Wówczas $X_{1n} + X_{2n} + \ldots + X_{r_n n} \Rightarrow \mathcal{N}(0,1)$.

Powstaje tu naturalne pytanie, co tak naprawdę mówi warunek Lindeberga. Intuicyjnie rzecz biorąc, oznacza on, iż przy n zbiegającym do nieskończoności, zmienne X_{1n}, X_{2n}, ..., $X_{r_n n}$ są „równie małe”. Innymi słowy, w n-tym wierszu nie ma zmiennych losowych, które byłyby dominujące w stosunku do pozostałych. Ściślej, mamy następujące dwie własności.

Wnioski z warunku Lindeberga

1. Mamy $\max_{k\leqslant r_n} |X_{kn}| \xrightarrow{\mathbb{P}} 0$. Istotnie, dla każdego $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}(\max_{k\leqslant r_n} |X_{kn}| > \varepsilon) &= \mathbb{P}\left(\bigcup_{k=1}^{r_n} \{|X_{kn}| > \varepsilon\}\right) \leqslant \sum_{k=1}^{r_n} \mathbb{P}(|X_{kn}| > \varepsilon) \\ &\leqslant \varepsilon^{-2} \sum_{k=1}^{r_n} \mathbb{E}X_{kn}^2 1_{\{|X_{kn}|>\varepsilon\}} \xrightarrow{n\to\infty} 0.\end{aligned}$$

2. Mamy $\max_{k\leqslant r_n} \mathbb{E}X_{kn}^2 \to 0$. Rzeczywiście, dla dowolnego $\varepsilon > 0$,

$$\mathbb{E}X_{kn}^2 = \mathbb{E}X_{kn}^2 1_{\{|X_{kn}|>\varepsilon\}} + \mathbb{E}X_{kn}^2 1_{\{|X_{kn}|\leqslant\varepsilon\}} \leqslant \sum_{l=1}^{r_n} \mathbb{E}X_{ln}^2 1_{\{|X_{ln}|>\varepsilon\}} + \varepsilon^2 \leqslant 2\varepsilon^2,$$

o ile n jest dostatecznie duże.

Sformułujmy teraz nieco inną wersję CTG.

Twierdzenie 3.2. *Załóżmy, że* $X_1, X_2, \ldots,$ *są niezależnymi zmiennymi losowymi całkowalnymi z kwadratem,* $m_n := \mathbb{E}X_n$, $\sigma_n^2 = VarX_n$, $b_n^2 = \sum_{k=1}^n \sigma_n^2$. *Jeśli jest spełniony warunek Lindeberga*

$$(L) \qquad b_n^{-2} \sum_{k=1}^{n} \mathbb{E}|X_k - m_k|^2 1_{\{|X_k - m_k| > \varepsilon b_n\}} \xrightarrow{n\to\infty} 0,$$

to

$$\frac{X_1 + X_2 + \ldots + X_n - m_1 - m_2 - \ldots - m_n}{b_n} \Rightarrow \mathcal{N}(0,1).$$

Dowód. Wynika to bezpośrednio z twierdzenia Lindeberga, przy $r_n = n$, $X_{kn} = (X_k - m_k)/b_n$. □

Powstaje naturalne pytanie: kiedy warunek Lindeberga jest spełniony? Podamy tu kilka własności wymuszających ten warunek.

Stwierdzenie 3.1. *Załóżmy, że* $X_1, X_2, \ldots$ *są niezależne i mają ten sam rozkład o dodatniej wariancji. Oznaczmy* $m = \mathbb{E}X_1$, $\sigma^2 = VarX_1$. *Wówczas warunek Lindeberga jest spełniony i*

$$\frac{X_1 + X_2 + \ldots + X_n - nm}{\sqrt{n}\sigma} \Rightarrow \mathcal{N}(0,1).$$

Dowód:. Wystarczy sprawdzić warunek Lindeberga. Mamy

$$\frac{1}{n\sigma^2} \sum_{k=1}^{n} \mathbb{E}|X_n - m|^2 1_{\{|X_n - m| > \varepsilon\sigma\sqrt{n}\}} = \frac{1}{\sigma^2} \mathbb{E}|X_1 - m|^2 1_{\{|X_1 - m| > \varepsilon\sigma\sqrt{n}\}} \to 0,$$

na mocy twierdzenia Lebesgue'a. □

Sprawdzenie dwóch poniższych warunków pozostawiamy jako ćwiczenie.

Stwierdzenie 3.2. *Załóżmy, że* $X_1, X_2, \ldots$ *są wspólnie ograniczonymi niezależnymi zmiennymi losowymi spełniającymi warunek* $\sum_{k=1}^n VarX_k \to \infty$. *Wówczas spełniony jest warunek Lindeberga.*

Stwierdzenie 3.3 (Lapunow)**.** *Załóżmy, że dla każdego* n, $X_{1n}, X_{2n}, \ldots, X_{r_n n}$ *są niezależnymi, scentrowanymi zmiennymi losowymi spełniającymi warunki*

$$\sum_{k=1}^{r_n} \mathbb{E}X_{kn}^2 \xrightarrow{n\to\infty} 1$$

oraz

$$\sum_{k=1}^{r_n} \mathbb{E}|X_{kn}|^{2+\delta} \xrightarrow{n\to\infty} 0 \qquad \textit{dla pewnego } \delta > 0.$$

Wówczas jest spełniony warunek Lindeberga.

Przechodzimy do dowodu twierdzenia Lindeberga.

Lemat 3.1. *Załóżmy, że* $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ *są liczbami zespolonymi, z których każda ma moduł niewiększy niż* 1. *Wówczas*

$$|a_1 a_2 \ldots a_n - b_1 b_2 \ldots b_n| \leqslant \sum_{k=1}^{n} |a_k - b_k|.$$

Dowód:. Stosujemy indukcję. Dla $n = 1$ nierówność jest oczywista. Dalej, załóżmy, że jest ona prawdziwa dla pewnego n spróbujmy ją udowodnić dla $n+1$. Oznaczając $a = a_1a_2\ldots a_n$, $b = b_1b_2\ldots b_n$, mamy

$$\begin{aligned}|a_1a_2\ldots a_{n+1} - b_1b_2\ldots b_{n+1}| &= |aa_{n+1} - bb_{n+1}|\\ &\leqslant |aa_{n+1} - ab_{n+1}| + |ab_{n+1} - bb_{n+1}|\\ &= |a||a_{n+1} - b_{n+1}| + |b_{n+1}||a-b|\\ &\leqslant \sum_{k=1}^{n+1} |a_k - b_k|,\end{aligned}$$

co kończy dowód. □

Lemat 3.2. *Dla dowolnego $y \in \mathbb{R}$ oraz $k = 0, 1, 2, \ldots$ mamy*

$$\left|e^{iy} - \left(1 + iy + \frac{(iy)^2}{2!} + \ldots + \frac{(iy)^k}{k!}\right)\right| \leqslant \frac{|y|^{k+1}}{(k+1)!}.$$

Dowód:. Stosujemy indukcję. Dla $k = 0$ mamy

$$|e^{iy} - 1| = \left|i\int_0^y e^{ix}dx\right| \leqslant |y|.$$

Dalej, załóżmy, że nierówność zachodzi dla pewnego k. Wówczas

$$\begin{aligned}&\left|e^{iy} - \left(1 + iy + \frac{(iy)^2}{2!} + \ldots + \frac{(iy)^{k+1}}{(k+1)!}\right)\right|\\ &= \left|i\int_0^y e^{ix} - \left(1 + ix + \frac{(ix)^2}{2!} + \ldots + \frac{(ix)^k}{k!}\right)dx\right|\\ &\leqslant \int_0^{|y|}\left|e^{ix} - \left(1 + ix + \frac{(ix)^2}{2!} + \ldots + \frac{(ix)^k}{k!}\right)\right|dx\\ &\leqslant \int_0^{|y|}\frac{x^{k+1}}{(k+1)!}dx = \frac{|y|^{k+2}}{(k+2)!}.\end{aligned}$$

Dowód jest zakończony. □

Dowód twierdzenia Lindeberga:. Oznaczmy $\sigma_{kn} = (\mathbb{E}X_{kn}^2)^{1/2}$, $k = 1, 2, \ldots, r_n$, $n = 1, 2, \ldots$. Na mocy twierdzenia Levy-Cramera wystarczy udowodnić, że dla każdego $t \in \mathbb{R}$, $\varphi_{X_{1n}+X_{2n}+\ldots+X_{r_nn}}(t) \to e^{-t^2/2}$. Ustalmy więc $t \in \mathbb{R}$. Mamy

$$\begin{aligned}A_n :=&|\varphi_{X_{1n}+X_{2n}+\ldots+X_{r_nn}}(t) - e^{-t^2/2}| = \left|\prod_{k=1}^{r_n}\varphi_{X_{kn}}(t) - \prod_{k=1}^{r_n}e^{-\sigma_{kn}^2t^2/2}\right|\\ &+ \underbrace{\left|e^{-t^2\sum_{k=1}^{r_n}\sigma_{kn}^2/2} - e^{-t^2/2}\right|}_{D_n}.\end{aligned}$$

Stosujemy teraz pierwszy z powyższych lematów oraz fakt, iż $e^{-x} = 1 - x + r(x)$, gdzie $r(x)/x \xrightarrow{x\to 0} 0$. W konsekwencji,

$$A_n \leqslant \underbrace{\sum_{k=1}^{r_n}\left|\varphi_{X_{kn}} - 1 + \frac{\sigma_{kn}^2t^2}{2}\right|}_{B_n} + \underbrace{\sum_{k=1}^{r_n}|r(t^2\sigma_{kn}^2/2)|}_{C_n} + D_n.$$

Wystarczy wykazać, że (B_n), (C_n), (D_n) dążą do 0. Zbieżność ciągu (D_n) do 0 jest oczywista na mocy warunku $\sum_{k=1}^{r_n} \sigma_{kn}^2 \to 1$. Zajmijmy się teraz ciągiem (C_n). Ustalmy $\varepsilon > 0$. Istnieje $\delta > 0$ taka,że jeśli $|x| < \delta$, to $|r(x)/x| < \varepsilon$. Jak już wiemy, warunek Lindeberga pociąga za sobą, iż dla dostatecznie dużych n, $\max_{k \leqslant r_n} t^2\sigma_{kn}^2/2 < \delta$, a co za tym idzie,

$$C_n = \sum_{k=1}^{r_n} \left| \frac{r(t^2\sigma_{kn}^2/2)}{t^2\sigma_{kn}^2/2} \right| t^2\sigma_{kn}^2/2 < \frac{t^2\varepsilon}{2} \sum_{k=1}^{r_n} \sigma_{kn}^2 \to \frac{t^2\varepsilon}{2},$$

a więc ciąg (C_n) zbiega do 0. Wreszcie, dla ustalonego $\varepsilon > 0$, korzystając z drugiego z powyższych lematów (z $k = 2$ oraz z $k = 3$),

$$\begin{aligned}
&\left| \varphi_{X_{kn}} - 1 + \frac{\sigma_{kn}^2 t^2}{2} \right| = \left| \mathbb{E}(e^{itX_{kn}} - 1 - itX_{kn} + t^2X_{kn}^2/2) \right| \\
&\leqslant \mathbb{E}\left| e^{itX_{kn}} - 1 - itX_{kn} + t^2X_{kn}^2/2 \right| 1_{\{|X_{kn}| \leqslant \varepsilon\}} \\
&\quad + \mathbb{E}\left| e^{itX_{kn}} - 1 - itX_{kn} + t^2X_{kn}^2/2 \right| 1_{\{|X_{kn}| > \varepsilon\}} \\
&\leqslant \mathbb{E}\frac{|X_{kn}t|^3}{6} 1_{\{|X_{kn}| \leqslant \varepsilon\}} + \mathbb{E}\left| e^{itX_{kn}} - 1 - itX_{kn} \right| 1_{\{|X_{kn}| > \varepsilon\}} + \mathbb{E}\frac{t^2X_{kn}^2}{2} 1_{\{|X_{kn}| > \varepsilon\}} \\
&\leqslant \frac{|t|^3\varepsilon}{6}\sigma_{kn}^2 + 2\mathbb{E}\frac{t^2X_{kn}^2}{2} 1_{\{|X_{kn}| > \varepsilon\}}.
\end{aligned}$$

Zatem

$$B_n \leqslant \frac{|t|^3}{6}\varepsilon \sum_{k=1}^{r_n} \sigma_{kn}^2 + t^2 \sum_{k=1}^{r_n} \mathbb{E}X_{kn}^2 1_{\{|X_{kn}| > \varepsilon\}} \leqslant \frac{2|t|^3}{6}\varepsilon + \varepsilon$$

dla dostatecznie dużych n. Stąd teza. □

Jako wniosek, otrzymujemy

Twierdzenie 3.3 (de Moivre'a-Laplace'a)**.** *Załóżmy, że ξ_n ma rozkład Bernoulliego z parametrami n, p. Wówczas*

$$\frac{\xi_n - np}{\sqrt{np(1-p)}} \Rightarrow \mathcal{N}(0,1).$$

Dowód:. Niech X_1, X_2, ... będą niezależnymi zmiennymi losowymi o tym samym rozkładzie dwupunktowym $\mathbb{P}(X_n = 0) = 1 - p$, $\mathbb{P}(X_n = 1) = p$. Mamy

$$\frac{\xi_n - np}{\sqrt{np(1-p)}} \sim \frac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}}$$

i wystarczy skorzystać z odpowiedniej wersji CTG. □

Sformułujmy teraz uogólnienie twierdzenia Lindeberga. Dowodzi się je w sposób analogiczny.

Twierdzenie 3.4. *Załóżmy, że dla każdego n zmienne X_{1n}, X_{2n}, ..., $X_{r_n n}$ są niezależne i całkowalne z kwadratem. Oznaczmy $m_{kn} := \mathbb{E}X_{kn}$ i przypuśćmy, że*

$$\sum_{k=1}^{r_n} \mathbb{E}X_{kn} \xrightarrow{n\to\infty} m, \qquad \sum_{k=1}^{r_n} VarX_{kn} \xrightarrow{n\to\infty} \sigma^2$$

oraz

$$(L) \qquad \sum_{k=1}^{r_n} \mathbb{E}(X_{kn} - m_{kn})^2 1_{\{|X_{kn}-m_{kn}|>\varepsilon\}} \to 0.$$

Wówczas $X_{1n} + X_{2n} + \ldots + X_{r_n n} \Rightarrow \mathcal{N}(m, \sigma^2)$.

Centralne twierdzenie graniczne pozwala badać zachowanie dystrybuant sum niezależnych zmiennych losowych. Istotnie, zbieżność

$$\frac{X_1 + X_2 + \ldots + X_n - (m_1 + m_2 + \ldots + m_n)}{b_n} \Rightarrow \mathcal{N}(0,1)$$

jest równoważna zbieżności punktowej dystrybuant:

$$\begin{aligned}\mathbb{P}\left(\frac{X_1 + X_2 + \ldots + X_n - (m_1 + m_2 + \ldots + m_n)}{b_n} \leqslant x\right) &\to \Phi(x)\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy.\end{aligned}$$

Co więcej, zbieżność jest jednostajna względem $x \in \mathbb{R}$ (por. zadanie 9 z rozdziału o słabej zbieżności). Zatem dla każdego $\varepsilon > 0$ istnieje numer n_0 taki, że dla $n \geqslant n_0$,

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(X_1 + X_2 + \ldots + X_n \leqslant xb_n + (m_1 + m_2 + \ldots + m_n)) - \Phi(x)| < \varepsilon,$$

czyli

$$\sup_{y\in\mathbb{R}} \left|\mathbb{P}(X_1 + X_2 + \ldots + X_n \leqslant y) - \Phi\left(\frac{y - m_1 - m_2 - \ldots - m_n}{b_n}\right)\right| < \varepsilon.$$

Powstaje naturalne pytanie w jaki sposób wyznaczać n_0 w zależności od ε; innymi słowy, w jaki sposób szacować błąd związany z przybliżeniem dysrtybuanty sumy przez dystrybuantę standardowego rozkładu normalnego.

Twierdzenie 3.5 (Nierówność Berry-Essena). *Załóżmy, że X_1, X_2, ..., są niezależnymi scentrowanymi zmiennymi losowymi o tym samym rozkładzie i niech $\sigma^2 = \mathrm{Var} X_n > 0$, $\rho := \mathbb{E}|X_n|^3 < \infty$. Wówczas*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}\sigma} \leqslant x \right) - \Phi(x) \right| \leqslant \frac{c\rho}{\sigma^3 \sqrt{n}},$$

gdzie jako c można wziąć $0,7655$ (optymalna - czyli najmniejsza możliwa - wartość c nie jest znana).

W szczególności, przy założeniach twierdzenia de Moivre'a-Laplace'a,

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\xi_n - np}{\sqrt{np(1-p)}} \leqslant x \right) - \Phi(x) \right| \leqslant c \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}.$$

Jest to niezwykle użyteczny rezultat z punktu widzenia konkretnych zastosowań: w sposób jawny określa on tempo zbieżności.

3.1. Zadania

1. Sumujemy 10 000 liczb, każdą zaokrągloną z dokładnością do 10^{-m}. Przypuśćmy, że błędy spowodowane przez zaokrąglenia są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na przedziale $(-10^m/2, 10^m/2)$. Znaleźć przedział (możliwie krótki), do którego z prawdopodobieństwem $\geqslant 0,95$ będzie należał błąd całkowity (tzn. po zsumowaniu).

2. Obliczyć

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

3. Udowodnić Stwierdzenie 10 oraz Stwierdzenie 11.

4. Dany jest ciąg (X_n) niezależnych zmiennych losowych, przy czym dla $n \geqslant 1$,

$$\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = \frac{1}{2}\left(1 - \frac{1}{n^2}\right), \ \mathbb{P}(X_n = -n) = \mathbb{P}(X_n = n) = \frac{1}{2n^2}.$$

Udowodnić, że

$$\frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}} \Rightarrow \mathcal{N}(0,1),$$

mimo iż $\lim_{n\to\infty} \mathrm{Var} X_n = 2$.

5. Zmienne losowe $X_1, X_2, \ldots$ są niezależne, przy czym dla $n \geqslant 1$, $\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = 1/2$. Niech $s_n^2 = \sum_{k=1}^n \mathrm{Var} X_k$. Czy ciąg zmiennych losowych

$$\frac{X_1 + X_2 + \ldots + X_n}{s_n}$$

jest zbieżny wedlug rozkładu, a jeśli tak, to do jakiej granicy?

6. Załóżmy, że X jest zmienną losową spełniającą warunki

(i) $\mathbb{E}X^2 < \infty$,
(ii) Jeśli Y, Z są niezależne i mają ten sam rozkład co X, to $X \sim (Y+Z)/\sqrt{2}$.
Wykazać, że X ma rozkład Gaussa o średniej 0.

7. Rzucono 900 razy kostką. Sumujemy oddzielnie parzyste liczby oczek i nieparzyste liczby oczek. Jakie jest przybliżone prawdopodobieństwo tego, że suma parzystych liczb oczek będzie o co najmniej 500 większa od sumy nieparzystych liczb oczek?

8. Liczba studentów przyjętych na pierwszy rok jest zmienną losową o rozkładzie Poissona z parametrem 100. Jeśli ta liczba przekroczy 120, tworzy się 2 grupy wykładowe. Obliczyć przybliżone prawdopodobieństwo (korzystając z CTG), że nie trzeba będzie tworzyć dwóch grup.

9. Dane są dwa ciągi (X_n), (Y_n) niezależnych zmiennych losowych, przy czym zmienne (X_n) mają ten sam rozkład wykładniczy z parametrem 1, a (Y_n) mają rozkład Poissona z parametrem 1. Zbadać zbieżność według rozkładu ciągu

$$\frac{(X_1+X_2+\ldots+X_n)^2-(Y_1+Y_2+\ldots+Y_n)^2}{n\sqrt{n}}.$$

4. Warunkowa wartość oczekiwana

Warunkowa wartość oczekiwana jest jednym z kluczowych pojęć w teorii prawdopodobieństwa. Zacznijmy od sytuacji gdy warunkujemy względem zdarzenia.

Definicja 4.1. 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

$$\mathbb{E}(X|B) = \int_\Omega X(\omega)\mathbb{P}(d\omega|B).$$

Stwierdzenie 4.1. *Przy założeniach jak wyżej,*

$$(*) \qquad \mathbb{E}(X|B) = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P}.$$

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, dowodzona równość zachodzi także dla zmiennych prostych (kombinacji liniowych indykatorów zdarzeń).

3. Teraz jeśli X jest nieujemną zmienną losową, to bierzemy niemalejący ciąg (X_n) zmiennych prostych zbieżny prawie na pewno do X. Pisząc (*) dla X_n i zbiegając z $n \to \infty$ 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 rozbicie $X = X_+ - X_-$ i stosujemy (*) dla X_+ oraz X_-; po odjęciu stronami dostajemy (*) dla X. □

Przechodzimy do definicji warunkowej wartości oczekiwanej względem σ-ciała.

Definicja 4.2. 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 4.1. *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 $\nu(B) = \int_B Xd\mathbb{P}$. Funkcja $\nu : \mathcal{M} \to \mathbb{R}$ 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 $B \in \mathcal{M}$,

$$\int_B X\mathbb{P} = \nu(B) = \int_B \eta d\mathbb{P}.$$

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. □

Uwaga: Warto tu przyjrzeć się warunkowej wartości oczekiwanej zmiennej X względem σ-ciała $\mathcal{M}$ generowanego przez co najwyżej przeliczalne rozbicie (B_n) zbiorów o dodatnim prawdopodobieństwie. Bardzo łatwo wyznaczyć tę zmienną w oparciu o powyższą definicję. Mianowicie, jak widać z warunku 1), $\mathbb{E}(X|\mathcal{M})$ musi być stała na każdym zbiorze B_n, $n = 1, 2, \ldots$; własność 2) natychmiast implikuje, iż $\mathbb{E}(X|\mathcal{M}) = \mathbb{E}(X|B_n)$ na zbiorze B_n. To w jednoznaczny sposób opisuje warunkową wartość oczekiwaną.

Przechodzimy do pojęcia warunkowej wartości oczekiwanej względem zmiennej losowej. Będziemy potrzebować następującego pomocniczego faktu.

Lemat 4.1. *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

$$f(x) = \begin{cases} \lim_{n\to\infty} f_n(x) & \text{jeśli granica istnieje,} \\ 0 & \text{jeśli granica nie istnieje.} \end{cases}$$

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_-. □

Definicja 4.3. 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. Wyznaczymy $\mathbb{E}(X|X+Y)$.

Wiadomo, że $X+Y$ ma rozkład Poissona z parametrem $\lambda+\mu$. Stąd

$$P_{X+Y}(k) = \frac{(\lambda+\mu)^k}{k!} e^{-(\lambda+\mu)}, \qquad k = 0, 1, 2, \ldots.$$

Ponadto, jeśli $k \geqslant \ell \geqslant 0$, 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 $h : \mathbb{R} \to \mathbb{R}$ mamy

$$(*) \qquad \mathbb{E}(h(X)|Y) = \int_{\mathbb{R}} h(x) g_{X|Y}(x|Y) dx.$$

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

$$\begin{aligned} \int_B h(X) d\mathbb{P} &= \int_\Omega 1_{\{Y \in A\}} h(X) d\mathbb{P} = \int_{\mathbb{R}^2} 1_{\{y \in A\}} h(x) g(x,y) dx dy \\ &= \int_{\mathbb{R}} 1_{\{y \in A\}} g_Y(y) \int_{\mathbb{R}} h(x) g_{X|Y}(x|y) dx dy \\ &= \int_B \int_{\mathbb{R}} h(x) g_{X|Y}(x|Y) dx d\mathbb{P}. \end{aligned}$$

Własności warunkowej wartości oczekiwanej

Załóżmy, 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 zakładamy, że są całkowalne.

0. Mamy $\mathbb{E}(\mathbb{E}(X|\mathcal{M})) = \mathbb{E}X$. Wynika to natychmiast z 2), jeśli weźmiemy $B = \Omega$.

1. Niech α, $\beta \in \mathbb{R}$. 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 $\mathbb{E}(X|\mathcal{M}) \geqslant 0$ 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

$$(*) \qquad |\mathbb{E}(X|\mathcal{M})| \leqslant \mathbb{E}(|X||\mathcal{M}) \quad \text{p.n.}$$

Istotnie, na mocy 1. oraz 2. mamy, iż nierówność $X \leqslant Y$ p.n. pociąga za sobą $\mathbb{E}(X|\mathcal{M}) \leqslant \mathbb{E}(Y|\mathcal{M})$. Stąd, z prawdopodobieństwem 1,

$$\mathbb{E}(X_1|\mathcal{M}) \leqslant \mathbb{E}(|X_1||\mathcal{M})$$

i

$$-\mathbb{E}(X_1|\mathcal{M}) \leqslant \mathbb{E}(|X_1||\mathcal{M}).$$

Biorąc wartość oczekiwaną obu stron w (*) dostajemy, na mocy 0.,

$$\mathbb{E}(|\mathbb{E}(X|\mathcal{M})|) \leqslant \mathbb{E}|X|.$$

Innymi słowy, operator liniowy $\mathbb{E}(\cdot|\mathcal{M}) : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{F}, \mathbb{P})$ jest kontrakcją.

4. Warunkowa wersja twierdzenia Lebesgue'a o monotonicznym przejściu do granicy. Załóżmy, że $X_n \uparrow X$. 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, $\mathbb{E}(X_1|\mathcal{M}) \leqslant \eta \leqslant \infty$. Niech teraz $B \in \mathcal{M}$. Mamy, na mocy 2) oraz bezwarunkowego twierdzenia Lebesgue'a,

$$\int_B X = \lim_{n\to\infty} \int_B X_n = \lim_{n\to\infty} \int_B \mathbb{E}(X_n|\mathcal{M}) = \int_B \eta.$$

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 zmajoryzowanym przejściu do granicy pod znakiem całki oraz lematu Fatou.

6. Załóżmy, że X_1 jest mierzalna względem $\mathcal{M}$ oraz X_1X_2 jest całkowalna. Wówczas

$$(+) \qquad \mathbb{E}(X_1X_2|\mathcal{M}) = X_1\mathbb{E}(X_2|\mathcal{M}) \quad \text{p.n.}$$

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 $\mathbb{E}(X_1X_2|\mathcal{M})$. 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 $B \in \mathcal{M}$,

$$\int_B X_1\mathbb{E}(X_2|\mathcal{M})d\mathbb{P} = \int_{A\cap B} \mathbb{E}(X_2|\mathcal{M})d\mathbb{P} = \int_{A\cap B} X_2 d\mathbb{P} = \int_B X_1X_2 d\mathbb{P}.$$

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^+:

$$\mathbb{E}(Y_nX_2^+|\mathcal{M}) = Y_n\mathbb{E}(X_2^+|\mathcal{M}).$$

Zbiegając z $n \to \infty$ i korzystając z warunkowej wersji twierdzenia Lebesgue'a (własność 4.), dostajemy

$$\mathbb{E}(X_1X_2^+|\mathcal{M}) = X_1\mathbb{E}(X_2^+|\mathcal{M}).$$

Zastępując X_2^+ przez X_2^- i powtarzając rozumowanie, dostajemy

$$\mathbb{E}(X_1X_2^-|\mathcal{M}) = X_1\mathbb{E}(X_2^-|\mathcal{M})$$

i po odjęciu stronami dostajemy (+).

d) Jeśli X_1 jest dowolną zmienną losową, to rozbijamy ją na różnicę $X_1^+ - X_1^-$, stoujemy 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

$$\mathbb{E}(X|\mathcal{M}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{M}_2)|\mathcal{M}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{M}_1)|\mathcal{M}_2). \tag{=}$$

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 zmienn:a losową stałą, a więc mierzalną względem każdego σ-ciała. Niech teraz $B \in \mathcal{M}$. Mamy na mocy niezależności 1_B oraz X,

$$\int_B \mathbb{E}X d\mathbb{P} = \mathbb{E}1_B\mathbb{E}X = \mathbb{E}(1_BX) = \int_B X d\mathbb{P}.$$

9. Nierówność Jensena. Załóżmy, że $f : \mathbb{R} \to \mathbb{R}$ jest funkcją wypukłą taką, że $f(X)$ jest zmienną całkowalną. Wówczas

$$\mathbb{E}(f(X)|\mathcal{M}) \geqslant f(\mathbb{E}(X|\mathcal{M})).$$

Będzie nam potrzebny następujący prosty fakt. Dowód pozostawiamy jako proste ćwiczenie.

Lemat 4.2. *Załóżmy, że $f : \mathbb{R} \to \mathbb{R}$ 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 $f(X) \geqslant a_n X + b_n$ dla każdego n. Stąd, na mocy 1. oraz 2., z prawdopodobieństwem 1,

$$\mathbb{E}(f(X)|\mathcal{M}) \geqslant a_n \mathbb{E}(X|\mathcal{M}) + b_n.$$

Poniweaż ciągi (a_n), (b_n) są przeliczalne, to możemy wziąć supremum po n po prawej stronie i dalej nierówno'sć będzie zachodziła z prawdopodobieństwem 1:

$$\mathbb{E}(f(X)|\mathcal{M}) \geqslant \sup_n (a_n \mathbb{E}(X||\mathcal{M}) + b_n) = f(\mathbb{E}(X|\mathcal{M})).$$

Jako wniosek, dostajemy, iż dla $p \geqslant 1$ i $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{E}(|X|^p|\mathcal{M}) \geqslant [\mathbb{E}(|X||\mathcal{M})]^p.$$

Stąd po wzięciu wartości oczekiwanej obu stron, $\mathbb{E}(|\mathbb{E}(X|\mathcal{M})|^p) \leqslant \mathbb{E}|X|^p$, czyli

$$||\mathbb{E}(X|\mathcal{M})||_p \leqslant ||X||_p.$$

Zatem warunkowa wartość oczekiwana $\mathbb{E}(\cdot|\mathcal{M})$ jest kontrakcją w L^p.

4.1. Zadania

1. 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 $\phi : \mathbb{R}^2 \to \mathbb{R}$ będzie funkcją borelowską taką, że $\phi(X, Y)$ jest całkowalną zmienną losową. Udowodnić, że

$$\mathbb{E}[\phi(X, Y)|\mathcal{G}] = \Phi(X),$$

gdzie $\Phi(x) = \mathbb{E}\phi(x, Y)$.

2. Załóżmy, że X jest całkowalną zmienną losową, a σ-ciało $\mathcal{G}$ jest niezależne od X oraz od σ-ciała $\mathcal{M}$. Udowodnić, że

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{M})) = \mathbb{E}(X|\mathcal{M}).$$

3. Zmienna losowa (X, Y) ma gęstość

$$g(x, y) = \frac{x^3}{2} e^{-x(y+1)} 1_{\{x>0,\, y>0\}}.$$

Wyznaczyć $\mathbb{E}(Y|X)$ oraz $\mathbb{E}(Y^2|X)$.

4. Zmienna losowa (X, Y) ma rozkład Gaussa o wartości oczekiwanej 0, $\mathrm{Var}X = \sigma_1^2$, $\mathrm{Var}Y = \sigma_2^2$, $\mathrm{Cov}(X, Y) = c$. Obliczyć $\mathbb{P}(Y \in B|X)$ (dla $B \in \mathcal{B}(\mathbb{R})$) oraz $\mathbb{E}(Y|X)$.

5. 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 $\mathbb{E}(\sin X|X + Y)$.

6. Zmienne losowe ε_1, ε_2, ε_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 $\mathbb{E}(\varepsilon_1\varepsilon_2|e_1 + e_2e_3)$.

7. 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 $\mathbb{E}(F|O) = \frac{2p}{100+p}O$.

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. Losujemy ze zwracaniem po jednej karcie z talii 52 kart tak długo aż wyciągniemy pika. Niech Y oznacza zmienną losową równą liczbie wyciągniętych kart, a X zmienną losową równą liczbie wyciągniętych kierów. Wyznaczyć $\mathbb{E}(Y|X = 4)$ oraz $\mathbb{E}(X|Y = 4)$.

10. Zmienne lsowe X, Y są niezależne i mają rozkład wykładniczy z parametrem 1. Obliczyć $\mathbb{E}(X|X + Y)$ oraz $\mathbb{E}(X|\min(X, Y))$.

5. Martyngały z czasem dyskretnym

Do tej pory, dysponując ciągiem zmiennych losowych, nie wiązaliśmy z ich indeksami żadnej interpretacji. W wielu naturalnych sytuacjach można je interpretować jako współrzędną czasową. W konkretnych przypadkach często X_n opisuje zachowanie układu w chwili n. Tak więc indeks odpowiada za czas.

Załóżmy, że T jest „zbiorem czasów”: to znaczy, jest równy $\{0, 1, 2, \ldots\}$, $\{1, 2, \ldots, \}$, $\{\ldots, -2, -1, 0\}$ lub $\{m, m+1, \ldots, n\}$.

Definicja 5.1. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną, T - jak wyżej. *Filtracją* nazywamy rodzinę $(\mathcal{F}_t)_{t\in T}$, gdzie dla każdego t, $\mathcal{F}_t$ jest σ-ciałem zawartym w $\mathcal{F}$ oraz $\mathcal{F}_t \subseteq \mathcal{F}_s$ jeśli $s \leqslant t$.

Intuicja: σ-ciało $\mathcal{F}_t$ opisuje wszystko co się może zdarzyć do chwili t.

Definicja 5.2. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną wyposażoną w filtrację $(\mathcal{F}_t)_{t\in T}$. Funkcję $\tau : \Omega \to T \cup \{+\infty\}$ nazywamy *momentem zatrzymania*, jeśli dla każdego $t \in T$ mamy $\{\tau = t\} \in \mathcal{F}_n$.

Intuicyjnie, moment zatrzymania jest „sensowną” reguła stopowania: taką, iż decyzję, czy się zatrzymywać, podejmujemy na podstawie zdarzeń z przeszłości i teraźniejszości. Spójrzmy na następujący

Przykład: Rzucamy 10 razy monetą. Niech $X_n = 1$, jeśli w n-tym rzucie wypadł orzeł, i $X_n = 0$ w przeciwnym przypadku. Wprowadźmy σ-ciała $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$, $n = 1, 2, \ldots, 10$ (jest to tzw. *naturalna filtracja względem ciągu* (X_n)) Rozważmy dwie strategie: τ - wycofujemy się, gdy wypadnie orzeł po raz pierwszy, σ - wycofujemy się, gdy orzeł wypada po raz ostatni (jeśli wypadają same reszki, przyjmujemy $\tau = \sigma = 10$). Intuicja podpowiada, iż τ jest sensowną regułą zatrzymania - decyzję o tym, czy się wycofać, czy nie, podejmujemy na podstawie informacji, które dopłynęły do nas do danej chwili. Strategia σ nie jest sensowna: skąd mamy wiedzieć - nie znając przyszłości - czy orzeł, który właśnie wypadł, jest ostatni? Formalny dowód tego, że σ nie jest momentem zatrzymania, pozostawiamy jako ćwiczenie.

Uwaga:

Warunek definiujący moment stopu można zapisać równoważnie w następujący sposób. Funkcja $\tau : \Omega \to T \cup \{+\infty\}$ jest momentem zatrzymania wtedy i tylko wtedy, gdy dla każdego $t \in T$, $\{\tau \leqslant t\} \in \mathcal{F}_t$.

Dowód. $\Rightarrow$ Mamy

$$\{\tau \leqslant t\} = \bigcup_{k=1}^{t} \{\tau = k\} \in \mathcal{F}_t,$$

gdyż dla każdego $k \leqslant t$, $\{\tau = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_t$.

$\Leftarrow$ Mamy $\{\tau = t\} = \{\tau \leqslant t\} \setminus \{\tau \leqslant t-1\}$ i oba zdarzenia należą do $\mathcal{F}_t$. □

Przykłady:

1) $\tau \equiv n$ jest momentem zatrzymania względem każdej filtracji:

$$\{\tau = k\} = \begin{cases} \emptyset & \text{jeśli } n \neq k, \\ \Omega & \text{jeśli } n = k. \end{cases}$$

2) Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną wyposażoną w filtrację $(\mathcal{F}_n)_{n\in T}$. Załóżmy, że $(X_n)_{n\in T}$ jest rodziną zmiennych losowych (*procesem stochastycznym*) o tej własności, że dla każdego n, zmienna X_n jest mierzalna względem $\mathcal{F}_n$ (mówimy, że proces stochastyczny (X_n) jest adaptowany do filtracji (F_n)). Dalej, niech $B \in \mathcal{B}(\mathbb{R})$ oraz

$$\tau_B(\omega) = \inf\{n \in T : X_n(\omega) \in B\},$$

przy czym przyjmijmy konwencję $\inf \emptyset = +\infty$. Funkcja τ_B to *moment pierwszego dojścia procesu* (X_n) *do zbioru* B. Wówczas τ_B jest momentem zatrzymania: dla każdego n,

$$\begin{aligned} \{\tau_B = n\} &= \{X_n \in B \text{ oraz } X_k \notin B \text{ dla } k < n\} \\ &= \{X_n \in B\} \cap \bigcap_{k<n} \{X_k \in B^c\} \in \mathcal{F}_n. \end{aligned}$$

Analogiczny fakt zachodzi, gdy zmienne X_n przyjmują wartości w $\mathbb{R}^d$, albo ogólniej, w przestrzeni metrycznej E.

Definicja 5.3. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną wyposażoną w filtrację $(\mathcal{F}_t)_{t\in T}$ i niech τ będzie momentem zatrzymania. Definiujemy

$$\begin{aligned} \mathcal{F}_\tau &= \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n \text{ dla wszystkich } n\} \\ &= \{A \in \mathcal{F} : A \cap \{\tau \leqslant n\} \in \mathcal{F}_n \text{ dla wszystkich } n\}. \end{aligned}$$

Intuicyjnie, $\mathcal{F}_\tau$ opisuje wszystkie zdarzenia, które mogą zajść do momentu τ.
Uwagi:
1) $\mathcal{F}_\tau$ jest σ-ciałem,
2) jeśli $\tau \equiv n$, to $\mathcal{F}_\tau = \mathcal{F}_n$.

Własności:
1) Jeśli τ_1, τ_2 są momentami zatrzymania, to $\tau_1 \wedge \tau_2 = \min\{\tau_1, \tau_2\}$ oraz $\tau_1 \vee \tau_2 = \max\{\tau_1, \tau_2\}$ też są momentami zatrzymania. Istotnie,

$$\{\tau_1 \wedge \tau_2 \leqslant n\} = \{\tau_1 \leqslant n\} \cup \{\tau_2 \leqslant n\} \in \mathcal{F}_n,$$

$$\{\tau \vee \tau_2 \leqslant n\} = \{\tau_1 \leqslant n\} \cap \{\tau_2 \leqslant n\} \in \mathcal{F}_n.$$

2) Jeśli τ_1, τ_2 są takimi momentami zatrzymania, że $\tau_1 \leqslant \tau_2$, to $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$. Istotnie, jeśli $A \in \mathcal{F}_{\tau_1}$, to dla każdego n,

$$A \cap \{\tau_2 \leqslant n\} = (A \cap \{\tau_1 \leqslant n\}) \cap \{\tau_2 \leqslant n\},$$

i dwa ostatnie przecinane zbiory należą do $\mathcal{F}_n$.

3) Moment zatrzymania τ jest mierzalny względem $\mathcal{F}_\tau$. Istotnie,

$$\{\tau \leqslant a\} \cap \{\tau = n\} = \begin{cases} \emptyset & \text{jeśli } a < n, \\ \{\tau = n\} & \text{jeśli } a \geqslant n \end{cases} \in \mathcal{F}_n.$$

4) Załóżmy, że $(X_t)_{t\in T}$ jest adaptowany do danej filtracji, a τ jest momentem zatrzymania względem tej filtracji spełniającym warunek $\tau < \infty$ (jest to tzw. *skończony moment stopu*. Wówczas zmienna X_τ jest mierzalna względem $\mathcal{F}_\tau$. Istotnie,

$$\{X_\tau \leqslant a\} \cap \{\tau = n\} = \{X_n \leqslant a\} \cap \{\tau = n\} \in \mathcal{F}_n,$$

jako że oba przecinane zdarzenia należą do $\mathcal{F}_n$.

Przechodzimy do definicji głównych pojęć niniejszego rozdziału.

Definicja 5.4. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną wyposażoną w filtrację $(\mathcal{F}_t)_{t\in T}$. Załóżmy, że $(X_t)_{t\in T}$ jest adaptowanym ciągiem całkowalnych zmiennych losowych. Mówimy, że $(X_t, \mathcal{F}_t)_{t\in T}$ jest

a) *martyngałem*, jeśli dla wszystkich $s,\, t \in T$, $s \leqslant t$ zachodzi $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$.

b) *nadmartyngałem*, jeśli dla wszystkich $s,\, t \in T$, $s \leqslant t$ zachodzi $\mathbb{E}(X_t|\mathcal{F}_s) \leqslant X_s$.

c) *podmartyngałem*, jeśli dla wszystkich $s,\, t \in T$, $s \leqslant t$ zachodzi $\mathbb{E}(X_t|\mathcal{F}_s) \geqslant X_s$.

Jeśli filtracja jest ustalona, to mówimy po prostu, że $(X_t)_{t\in T}$ jest martyngałem (nad-, pod-), jeśli zachodzą powyższe warunki.

Uwagi:

a) (X_t) jest martyngałem, wtedy i tylko wtedy, gdy dla dowolnych $s,\, t \in T$, $s < t$, oraz $A \in \mathcal{F}_s$ zachodzi

$$\int_A X_t d\mathbb{P} = \int_A X_s d\mathbb{P}.$$

Analogicznie dla nad- i podmartyngałów.

b) U nas $T = \{0,\, 1,\, 2,\, \ldots\}$, $\{1,\, 2,\, \ldots\}$, $\{m,\, m+1,\, \ldots,\, n\}$, $\{\ldots,\, -2,\, -1,\, 0\}$.

c) (X_t) jest martyngałem wtedy i tylko wtedy, gdy jest nad- i podmartyngałem.

d) (X_t) jest podmartyngałem wtedy i tylko wtedy, gdy $(-X_t)$ jest nadmartyngałem.

e) Jeśli (X_t), (Y_t) są martyngałami względem tej samej filtracji i $a,\, b \in \mathbb{R}$, to $(aX_t + bY_t)$ też jest martyngałem. Analogiczny fakt zachodzi dla nad- i podmartyngałów, o ile $a,\, b > 0$.

f) Jeśli zbiór T jest taki jak w b), to $(X_t)_{t\in T}$ jest martyngałem wtedy i tylko wtedy, gdy dla wszystkich $n \in T$ takich, że $n+1 \in T$, zachodzi $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ (analogiczny fakt zachodzi dla nad- i podmartyngałów).

Dowód:. $\Rightarrow$ oczywiste (szczególny przypadek).

$\Leftarrow$ Załóżmy, że $m,\, n \in T$, $m > n$. Wówczas $\mathcal{F}_n \subseteq \mathcal{F}_{m-1}$, a więc na mocy własności warunkowej wartości oczekiwanej,

$$\mathbb{E}(X_m|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X_m|\mathcal{F}_{m-1})|\mathcal{F}_n) = \mathbb{E}(X_{m-1}|\mathcal{F}_n),$$

i dalej przez indukcję. □

Przykłady:

1) Załóżmy, że $\xi_1,\, \xi_2,\, \ldots$ są niezależnymi, całkowalnymi zmiennymi losowymi o średniej 0. Niech $X_n = \xi_1 + \xi_2 + \ldots + \xi_n$ i $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$, $n = 1,\, 2,\, \ldots$. Wówczas $(X_n, \mathcal{F}_n)_{n=1}^\infty$ jest martyngałem:

$$\begin{aligned}\mathbb{E}(X_{n+1}|\mathcal{F}_n) &= \mathbb{E}(X_n + \xi_{n+1}|\mathcal{F}_n)\\ &= \mathbb{E}(X_n|\mathcal{F}_n) + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = X_n + \mathbb{E}\xi_{n+1} = X_n.\end{aligned}$$

2) Załóżmy, że X jest całkowalną zmienną losową, $(\mathcal{F}_t)_{t\in T}$ jest filtracją i niech $X_t = \mathbb{E}(X|\mathcal{F}_t)$ dla $t \in T$. Wówczas $(X_t, \mathcal{F}_t)_{t\in T}$ jest martyngałem.

Dowód:. Weźmy $s, t \in T$, $s < t$. Mamy, na mocy własności warunkowej wartości oczekiwanej,

$$\mathbb{E}(X_t|\mathcal{F}_s) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_t)|\mathcal{F}_s) = \mathbb{E}(X|\mathcal{F}_s) = X_s. \quad \square$$

Martyngał taki jak w przykładzie 2) nazywamy prawostronnie domkniętym. Czasami nazywa się tak martyngał wraz z domknięciem: $(X_t, \mathcal{F}_t)_{T\cup\{\infty\}}$, gdzie $(X_\infty, \mathcal{F}_\infty) = (X, \mathcal{F})$.

Stwierdzenie 5.1. *Załóżmy, że $(X_t, \mathcal{F}_t)_{t\in T}$ jest martyngałem, a $f : \mathbb{R} \to \mathbb{R}$ jest funkcją wypukłą taką, że $f(X_t)$ jest zmienną całkowalną dla każdego $t \in T$. Wówczas $(f(X_t), \mathcal{F}_t)_{t\in T}$ jest podmartyngałem.*

Dowód:. Załóżmy, że $s, t \in T$, $s < t$. Wówczas, na mocy nierówności Jensena,

$$\mathbb{E}(f(X_t)|\mathcal{F}_s) \geqslant f(\mathbb{E}(X_t|\mathcal{F}_s)) = f(X_s). \quad \square$$

Wniosek 5.1. *Załóżmy, że $(X_t, \mathcal{F}_t)_{t\in T}$ jest martyngałem. Wówczas*

a) Jeśli dla pewnego $p \geqslant 1$ mamy, iż $X_t \in L^p$ dla wszystkich t, to $(|X_t|^p, \mathcal{F}_t)$ jest podmartyngałem.

b) Dla dowolnej liczby rzeczywistej a, proces $(X_t \vee a, \mathcal{F}_t)_{t\in T}$ jest podmartyngałem. W szczególności, (X_t^+), (X_t^-) są podmartyngałami.

Twierdzenie 5.1 (Dooba, „optional sampling”)**.** *Załóżmy, że $(X_n, \mathcal{F}_n)_{n\geqslant 0}$ jest nadmartyngałem (odp., martyngałem). Załóżmy, że τ_1, τ_2 są momentami zatrzymania takimi, że $\tau_1 \leqslant \tau_2$ i τ_2 jest ograniczony. Wówczas mamy $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) \leqslant X_{\tau_1}$ p.n. (odpowiednio, $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) = X_{\tau_1}$ p.n.).*

Dowód:. Załóżmy, że $\tau_2 \leqslant n$. Zauważmy najpierw, iż X_{τ_1}, X_{τ_2} są całkowalne, gdyż $|X_{\tau_i}| \leqslant \max\{|X_1|, |X_2|, \ldots, |X_n|\}$. Zmienna X_{τ_1} jest mierzalna względem $\mathcal{F}_{\tau_1}$, a zatem wystarczy wykazać, że dla każdego $A \in \mathcal{F}_{\tau_1}$,

$$\int_A X_{\tau_2} d\mathbb{P} \leqslant \int_A X_{\tau_1} d\mathbb{P}$$

(odpowiednio, z równością w miejscu nierówności w przypadku martyngałowym).

Załóżmy najpierw, że $\tau_2 - \tau_1 \leqslant 1$. Mamy

$$\begin{aligned}\int_A X_{\tau_1} - X_{\tau_2} d\mathbb{P} &= \int_{A\cap\{\tau_2>\tau_1\}} X_{\tau_1} - X_{\tau_2} d\mathbb{P} \\ &= \sum_{k=0}^{n} \int_{\{\tau_1=k\}\cap A\cap\{\tau_2>k\}} X_k - X_{k+1} \geqslant 0\end{aligned}$$

(odpowiednio, $= 0$). Ostatnia nierówność bierze się stąd, iż $\{\tau_1 = k\} \cap A \cap \{\tau_2 > k\} \in \mathcal{F}_k$.

Weźmy teraz dowolne $\tau_1 \leqslant \tau_2 \leqslant n$. Definiujemy $\tau^{(k)} = \max\{\tau_1, \min\{\tau_2, k\}\}$. Zmienne $\tau^{(k)}$ są momentami zatrzymania, a ponadto

$$\tau_1 = \tau^{(0)} \leqslant \tau^{(1)} \leqslant \ldots \leqslant \tau^{(n)} = \tau_2$$

oraz $\tau^{(k+1)} - \tau^{(k)} \leqslant 1$. Zatem dla każdego $A \in \mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau^{(k)}}$,

$$\int_A X_{\tau_1} = \int_A X_{\tau^{(0)}} \geqslant \int_A X_{\tau^{(1)}} \geqslant \int_A X_{\tau^{(2)}} \geqslant \ldots \geqslant \int_A X_{\tau^{(n)}} = \int_A X_{\tau_2}$$

(z równościami w przypadku martyngałowym). $\square$

Twierdzenie 5.2 (Dooba o zbieżności p.n. nadmartyngałów). *Załóżmy, że proces $(X_n, \mathcal{F}_n)_{n=0,1,2,\ldots}$ jest nadmartyngałem takim, że $\sup_n \mathbb{E}X_n^- < \infty$. Wówczas ciąg (X_n) jest zbieżny p.n. do pewnej zmiennej losowej całkowalnej.*

Wniosek 5.2. *a) Każdy nieujemny nadmartyngał $(X_n, \mathcal{F}_n)$ (tzn. spełniający $X_n \geqslant 0$ p.n. dla wszystkich n) jest zbieżny p.n.*

b) Jeśli $(X_n, \mathcal{F}_n)_{n=0,1,2,\ldots}$ jest podmartyngałem spełniającym $\sup_n \mathbb{E}X_n^+ < \infty$, to (X_n) jest zbieżny p.n.

c) Jeśli $(X_n, \mathcal{F}_n)_{n=0,1,2,\ldots}$ jest nadmartyngałem, to warunek $\sup_n \mathbb{E}X_n^- < \infty$ jest równoważny warunkowi $\sup_n \mathbb{E}|X_n| < \infty$ (tzn. ograniczoności ciągu (X_n) w L^1).

Dowód wniosku:. a) jest oczywiste, b) wynika wprost z twierdzenia Dooba poprzez przejście do procesu $(-X_n, \mathcal{F}_n)$, który jest nadmartyngałem. Zajmijmy się dowodem c). Implikacja $\Leftarrow$ jest oczywista. $\Rightarrow$ Mamy $|X_n| = X_n^+ + X_n^- = X_n + 2X_n^-$, skąd

$$\mathbb{E}|X_n| = \mathbb{E}X_n + 2\mathbb{E}X_n^- \leqslant \mathbb{E}X_0 + 2\sup_n \mathbb{E}X_n^- < \infty. \quad \square$$

W dowodzie twierdzenia o zbieżności będziemy używać następujących obiektów. Załóżmy, że $(x_n)_{n=1,2,\ldots}$ jest ciągiem liczbowym i niech $a < b$ to ustalone liczby rzeczywiste. Określmy

$$\begin{aligned}
\tau_0 &= \inf\{n : x_n < a\},\\
\tau_1 &= \inf\{n > \tau_0 : x_n > b\},\\
&\ldots\\
\tau_{2k} &= \inf\{n > \tau_{2k-1} : x_n < a\},\\
\tau_{2k+1} &= \inf\{n > \tau_{2k} : x_n > b\},\\
&\ldots
\end{aligned}$$

Liczba τ_{2k-1} to *moment k-tego przejścia w górę ciągu (x_n) przez przedział* $[a, b]$. Niech teraz

$$U_a^b = \begin{cases} \sup\{k : \tau_{2k-1} < \infty\} & \text{jeśli } \tau_1 < \infty,\\ 0 & \text{jeśli } \tau_1 = \infty \end{cases}$$

będzie *liczbą przejść w górę ciągu (x_n) przez przedział* $[a, b]$.

Lemat 5.1. *Ciąg liczbowy (x_n) jest zbieżny (być może do $\pm\infty$) wtedy i tylko wtedy, gdy dla wszystkich a, $b \in \mathbb{Q}$, $a < b$, mamy $U_a^b < \infty$.*

Dowód:. $\Rightarrow$ Przypuśćmy wbrew tezie, że (x_n) jest zbieżny oraz że istnieją a, $b \in \mathbb{Q}$ takie, że $a < b$ oraz $U_a^b = \infty$. Wówczas znajdziemy nieskończony podciąg zawierający tylko wyrazy mniejsze od a oraz nieskończony podciąg zawierającego wyrazy tylko większe od b. Sprzeczność.

$\Leftarrow$ Załóżmy, że $\liminf x_n < \limsup x_n$. Wówczas istnieją a, $b \in \mathbb{Q}$ takie, że $\liminf x_n < a < b < \limsup x_n$; mamy wówczas $U_a^b = \infty$. $\square$

Lemat 5.2 (nierówność Dooba dla przejść w górę). *Załóżmy, że $(X_n, \mathcal{F}_n)_{n=0}^m$ jest nadmartyngałem. Wówczas dla dowolnych $a < b$,*

$$\mathbb{E}U_a^b \leqslant \frac{1}{b-a}\mathbb{E}((X_m - a)^-).$$

Dowód:. Załóżmy, że (τ_j) jest ciągiem momentów przejść ciągu (X_n) przez przedział $[a, b]$, i niech U_a^b będzie łączną liczbą przejść. Widzimy, że (τ_j) jest ciągiem momentów zatrzymania (względem filtracji $(\mathcal{F}_n)$) oraz że U_a^b jest zmienną losową. Połóżmy $\tilde{\tau}_j = \tau_j \wedge m$ i wprowadźmy zmienne $Y_k = X_{\tilde{\tau}_{2k+1}} - X_{\tilde{\tau}_{2k}}$, $k = 1, 2, \ldots$. Z definicji widzimy, iż jeśli $0 \leqslant k \leqslant U_a^b(\omega) - 1$, to $Y_k(\omega) > b - a$. Ponadto, jeśli $k = U_a^b(\omega)$, to

$$Y_k(\omega) = X_m - X_{\tilde{\tau}_{2k}} = \begin{cases} 0 & \text{jeśli } \tau_{2k} = \infty, \\ \geqslant X_m - a & \text{jeśli } \tau_{2k} < \infty \end{cases} \geqslant -(X_m - a)^-.$$

Wreszcie, jeśli $k > U_a^b(\omega)$, to $Y_k(\omega) = 0$. Sumując stronami powyższe związki dostajemy

$$\sum_{k=0}^{m} (X_{\tilde{\tau}_{2k+1}} - X_{\tilde{\tau}_{2k}}) \geqslant (b - a)U_a^b - (X_m - a)^-,$$

a zatem, biorąc wartość oczekiwaną,

$$\sum_{k=0}^{m} \mathbb{E}(X_{\tilde{\tau}_{2k+1}} - X_{\tilde{\tau}_{2k}}) \geqslant (b - a)\mathbb{E}U_a^b - \mathbb{E}(X_m - a)^-.$$

Lewa strona jest niedodatnia, na mocy twierdzenia Dooba (optional sampling); dostajemy zatem żądaną nierówność. □

Dowód twierdzenia o zbieżności nadmartyngałów. Ustalmy $a, b \in \mathbb{Q}$, $a < b$. Niech $U_a^b(m)$ będzie łączną liczbą przejść nadmartyngału $(X_n)_{n=1}^m$ w górę przez przedział $[a, b]$. Mamy $U_a^b(m) \uparrow U_a^b$. Na mocy drugiego z powyższych lematów,

$$\begin{aligned} \mathbb{E}U_a^b(m) &\leqslant \frac{1}{b-a}\mathbb{E}((X_m - a)^-) \leqslant \frac{1}{b-a}\mathbb{E}(|X_m| + |a|) \\ &\leqslant \frac{1}{b-a}(\sup \mathbb{E}|X_m| + |a|) < \infty. \end{aligned}$$

Zatem, na mocy twierdzenia Lebesgue'a, $\mathbb{E}U_a^b < \infty$, skąd $U_a^b < \infty$ p.n. Zatem

$$\mathbb{P}(\forall_{a,b \in \mathbb{Q},\, a<b} U_a^b < \infty) = 1$$

i na mocy pierwszego z powyższych lematów, ciąg (X_n) jest zbieżny p.n. Pozostaje tylko wykazać, że granica jest całkowalna; wynika to natychmiast z lematu Fatou:

$$\mathbb{E}|\lim_n X_n| = \mathbb{E}\lim_n |X_n| \leqslant \liminf \mathbb{E}|X_n| \leqslant \sup_n \mathbb{E}|X_n| < \infty. \quad \square$$

Twierdzenie 5.3 (Nierówność maksymalna dla nadmartyngałów)**.** *Załóżmy, że* $(X_n, \mathcal{F}_n)_{n=0,1,2,\ldots}$ *jest nadmartyngałem. Wówczas dla każdego* $\lambda > 0$,

$$\mathbb{P}(\sup_n |X_n| \geqslant \lambda) \leqslant K\frac{\sup_n \mathbb{E}|X_n|}{\lambda},$$

przy czym można wziąć $K = 1$*, jeśli nadmartyngał jest nieujemny (tzn. zmienne losowe* $X_0, X_1, \ldots$ *są nieujemne p.n.), niedodatni, bądź jest martyngałem. W przypadku ogólnym nierówność zachodzi z* $K = 3$.

Dowód:. Zauważmy, iż wystarczy szacować $\mathbb{P}(\sup_n |X_n| > \lambda)$, przez proste przejście graniczne. Mamy

$$\mathbb{P}(\sup_n |X_n| > \lambda) \leqslant \mathbb{P}(\sup_n X_n > \lambda) + \mathbb{P}(\inf_n X_n < -\lambda).$$

Zajmiemy się oddzielnie prawdopodobieństwami występującymi po prawej stronie.

a) Niech $\tau = \inf\{n : X_n > \lambda\}$. Na mocy twierdzenia Dooba (optional sampling),

$$\mathbb{E}X_0 \geqslant \mathbb{E}X_{\tau\wedge n} = \int_{\{\tau\leqslant n\}} X_\tau + \int_{\{\tau>n\}} X_n \geqslant \lambda\mathbb{P}(\max_{k\leqslant n} X_k > \lambda) - \int_{\{\tau>n\}} X_n^-.$$

Stąd

$$\lambda\mathbb{P}(\max_{k\leqslant n} X_k > \lambda) \leqslant \mathbb{E}X_0 + \int_{\{\tau>n\}} X_n^- \leqslant \mathbb{E}X_0 + \sup_n \mathbb{E}|X_n|.$$

Stąd teza (gdy weźmiemy $n \to \infty$) gdy (X_n) jest nieujemny.

b) Rozważmy moment zatrzymania $\tilde{\tau} = \inf\{n : X_n < -\lambda\}$. Z twierdzenia Dooba,

$$\begin{aligned}\mathbb{E}X_n &\leqslant \mathbb{E}X_{\tilde{\tau}\wedge n}\\ &= \int_{\{\tilde{\tau}\leqslant n\}} X_{\tilde{\tau}} + \int_{\{\tilde{\tau}>n\}} X_n \leqslant -\lambda\mathbb{P}(\min_{k\leqslant n} X_k < -\lambda) + \int_{\{\min_{k\leqslant n} X_k \geqslant -\lambda\}} X_n,\end{aligned}$$

skąd

$$(**) \qquad \lambda\mathbb{P}(\min_{k\leqslant n} X_k < -\lambda) \leqslant -\int_{\{\min_{k\leqslant n} X_k < -\lambda\}} X_n \leqslant \sup_n \mathbb{E}X_n^-.$$

Stąd teza, gdy nadmartyngał jest niedodatni. Ponadto, jeśli (X_n) jest martyngałem, to stosujemy powyższą nierówność do niedodatniego nadmartyngału $(-|X_n|, \mathcal{F}_n)$.

W ogólnym przypadku, wystarczy zsumować dwie końcowe nierówności pochodzące z a) i b), dostać nierówność ze stałą 3. □

Jeśli (X_n) jest podmartyngałem, to stosując (**) dla $(-X_n)$ dostajemy

Wniosek 5.3. *Załóżmy, że $(X_n, \mathcal{F}_n)_{n=0}^m$ jest podmartyngałem. Wówczas dla $\lambda > 0$,*

$$\mathbb{P}(\max_{n\leqslant m} X_n > \lambda) \leqslant \frac{1}{\lambda}\int_{\{\max_{n\leqslant m} X_n > \lambda\}} X_n.$$

Twierdzenie 5.4 (Nierówność maksymalna Dooba)**.** *Załóżmy, że $(X_n, \mathcal{F}_n)_{n\geqslant 0}$ jest martyngałem spełniającym warunek $X_n \in L^p$, $n = 0, 1, 2, \ldots$ dla pewnego $p > 1$. Wówczas*

$$||\sup_n |X_n|||_p \leqslant \frac{p}{p-1}\sup_n ||X_n||_p.$$

Dowód:. Niech $Y_n = \max_{k\leqslant n}|X_k|$, $k = 0, 1, 2, \ldots$. Mamy, stosując poprzedni wniosek do podmartyngału $(|X_k|, \mathcal{F}_k)_{k=0,1,2,\ldots,n}$, dostajemy

$$\begin{aligned}
\mathbb{E}Y_n^p &= p\int_0^\infty \lambda^{p-1}\mathbb{P}(Y_n > \lambda)d\lambda \\
&\leqslant p\int_0^\infty \lambda^{p-1}\frac{1}{\lambda}\int_{\{Y_n>\lambda\}}|X_n|d\mathbb{P}d\lambda \\
&= p\int_0^\infty \int_\Omega \lambda^{p-2}1_{\{Y_n>\lambda\}}|X_n|d\mathbb{P}d\lambda \\
&= p\int_\Omega \int_0^{Y_n} \lambda^{p-2}|X_n|d\lambda d\mathbb{P} \\
&= \frac{p}{p-1}\int_\Omega |X_n|Y_n^{p-1}d\mathbb{P} \leqslant \frac{p}{p-1}||X_n||_p||Y_n||_p^{(p-1)/p}.
\end{aligned}$$

Dzieląc obustronnie przez $||Y_n||_p^{(p-1)/p}$ (jeśli ta liczba jest zerem, to otrzymana poniżej nierówność także jest prawdziwa) dostajemy

$$||Y_n||_p \leqslant \frac{p}{p-1}||X_n||_p \leqslant \frac{p}{p-1}\sup_k ||X_k||_p$$

i wystarczy zbiec z $n \to \infty$. □

Twierdzenie 5.5 (Zbieżność martyngałów w L^1). *Załóżmy, że $(X_n, \mathcal{F}_n)_{n\geqslant 0}$ jest martyngałem. następujące warunki są równoważne.*

a) rodzina $\{X_n : n = 0, 1, 2, \ldots\}$ jest jednostajnie całkowalna.

b) (X_n) jest zbieżny w L^1.

c) Istnieje zmienna losowa $X \in L^1$ taka, że $X_n = \mathbb{E}(X|\mathcal{F}_n)$, $n = 0, 1, 2, \ldots$ (czyli martyngał jest prawostronnie domknięty).

Co więcej, jeśli te warunki są spełnione, to (X_n) jest zbieżny p.n. do

$$X_\infty = \mathbb{E}(X|\sigma(\bigcup_n \mathcal{F}_n)) \tag{$*$}$$

i X_∞ jest jedyną zmienną losową mierzalną względem σ-ciała $\sigma(\bigcup_n \mathcal{F}_n)$ taką, że $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$, $n = 0, 1, 2, \ldots$.

Wniosek 5.4 (Twierdzenie Levy'ego). *Jeśli $X \in L^1$ oraz $(\mathcal{F}_n)$ jest filtracją, to*

$$\mathbb{E}(X|\mathcal{F}_n) \xrightarrow{p.n.\ i\ w\ L^1} \mathbb{E}\left[X\middle|\sigma\left(\bigcup_n \mathcal{F}_n\right)\right].$$

Dowód twierdzenia o zbieżności. a)⇒b) Na mocy jednostajnej całkowalności dostajemy, iż $\sup_n \mathbb{E}|X_n| < \infty$. Zatem na mocy twierdzenia Dooba martyngał (X_n) jest zbieżny p.n., a zatem także według prawdopodobieństwa. łącząc to z jednostajną całkowalnością dostajemy zbieżność w L^1.

b)⇒c) Załóżmy, że $X_m \to X_\infty$ w L^1. Dla ustalonego n i $m > n$ mamy $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$. Z drugiej strony, $\mathbb{E}(X_m|\mathcal{F}_n) \to \mathbb{E}(X_\infty|\mathcal{F}_n)$ w L^1, gdyż operator warunkowej wartości oczekiwanej jest kontrakcją w L^1: istotnie,

$$||\mathbb{E}(X_m|\mathcal{F}_n) - \mathbb{E}(X_\infty|\mathcal{F}_n)||_1 \leqslant ||X_m - X_\infty||_1 \to 0.$$

Stąd $\mathbb{E}(X_\infty|\mathcal{F}_n) = X_n$.

c)$\Rightarrow$ a) Pozostawiamy jako ćwiczenie.

Pozostaje wykazać drugą część twierdzenia. Wiemy już, że warunki a), b), c) pociągają za sobą, iż $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$, $n = 0, 1, 2, \ldots$ (gdzie X_∞ jest granicą, w sensie zbieżności w L^1 i p.n., martyngału (X_n)). Oczywiście X_∞ jest mierzalna względem $\sigma(\bigcup_n \mathcal{F}_n)$. Przypuśćmy teraz, że Y jest całkowalną zmienną losową, mierzalną względem tego σ-ciała, dla której $X_n = \mathbb{E}(Y|\mathcal{F}_n)$, $n = 0, 1, 2, \ldots$. Zatem $\mathbb{E}(X_\infty|\mathcal{F}_n) = \mathbb{E}(Y|\mathcal{F}_n)$, skąd dla dowolnego n i dowolnego $A \in \mathcal{F}_n$,

$$\int_A X_\infty d\mathbb{P} = \int_A Y d\mathbb{P}.$$

Klasa $\bigcup_n \mathcal{F}_n$ jest π-układem. Klasa tych zbiorów A, dla których zachodzi powyższa równość, jest λ-układem. Z lematu o π-λ układach mamy, iż powy'rsza równo'sć całek zachodzi dla dowolnego $A \in \sigma(\bigcup_n \mathcal{F}_n)$. Na mocy mierzalności X_∞ oraz Y względem tego σ-ciała, mamy, iż $X_\infty = Y$ p.n.

Wreszcie, pozostaje udowodnić równość (*). Jeśli $X_n = \mathbb{E}(X|\mathcal{F}_n)$, to

$$\begin{aligned} X_n &= \mathbb{E}\left[X_n|\sigma\left(\bigcup_n \mathcal{F}_n\right)\right] = \mathbb{E}\left[\mathbb{E}(X|\mathcal{F}_n)\Big|\sigma\left(\bigcup_n \mathcal{F}_n\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(X\Big|\sigma\left(\bigcup_n \mathcal{F}_n\right)\right)\Big|\mathcal{F}_n\right]. \end{aligned}$$

Na mocy powyższych rozważań o jednoznaczności, dostajemy (*). Dowód jest zakończony. $\square$

Wniosek 5.5 (Prawo $0-1$ Kołmogorowa). *Załóżmy, że $X_1, X_2, \ldots$ są niezależnymi zmiennymi losowymi i $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$ dla $n \geqslant 1$. Wówczas jeśli $A \in \bigcap_{n=0}^\infty \sigma(X_{n+1}, X_{n+2}, \ldots)$, to $\mathbb{P}(A) \in \{0, 1\}$.*

Dowód. Oczywiście 1_A jest mierzalne względem σ-ciała $\sigma\left(\bigcup_{n=1}^\infty \mathcal{F}_n\right)$. Zatem na mocy twierdzenia Levy'ego,

$$\mathbb{E}(1_A|\mathcal{F}_n) \xrightarrow{\text{p.n. i w } L^1} \mathbb{E}\left[1_A\Big|\sigma\left(\bigcup_{n=1}^\infty \mathcal{F}_n\right)\right] = 1_A.$$

Ale z drugiej strony 1_A jest niezależne od $\mathcal{F}_n$, bo $A \in \sigma(X_{n+1}, X_{n+2}, \ldots)$, a to σ-ciało jest niezależne od $\mathcal{F}_n$. Stąd

$$\mathbb{E}(1_A|\mathcal{F}_n) = \mathbb{E}1_A = \mathbb{P}(A) \to 1_A,$$

a zatem $\mathbb{P}(A) = 0$ lub 1. $\square$

Zajmiemy się teraz zbieżnością w L^p dla $p > 1$.

Twierdzenie 5.6. *Załóżmy, że $(X_n, \mathcal{F}_n)_{n=0,1,2,\ldots}$ jest martyngałem i $p > 1$. Następujące warunki są równoważne.*

a) $\sup \mathbb{E}|X_n|^p < \infty$.

b) Rodzina $\{|X_n|^p\}_n$ jest jednostajnie całkowalna.

c) Martyngał (X_n) jest zbieżny w L^p.

d) Istnieje $X \in L^p$ taka, że $X_n = \mathbb{E}(X|\mathcal{F}_n)$.

Jeśli te warunki są spełnione, to (X_n) jest zbieżny p.n. do zmiennej losowej $X_\infty = \mathbb{E}(X|\sigma(\bigcup_n \mathcal{F}_n))$.

Dowód. a)⇒b) Wiemy, że $\mathbb{E}\sup|X_n|^p \leqslant \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}|X_n|^p < \infty$, czyli $\sup|X_n|^p \in L^1$, skąd dostajemy b) (istnienie majoranty całkowalnej).

b)⇒c) Mamy, iż

$$\sup_n \mathbb{E}|X_n| \leqslant \sup_n (\mathbb{E}|X_n|^p)^{1/p} < \infty,$$

a zatem na mocy twierdzenia Dooba o zbieżności nadmartyngałów, (X_n) jest zbieżny p.n.. Dokładając jednostajną całkowalność dostajemy c).

c)⇒d) Mamy $X_n \to X_\infty$ w L^p. Przy ustalonym n oraz $m > n$, $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$. Ponieważ $\mathbb{E}(\cdot|\mathcal{F}_n)$ jest kontrakcją w L^p, więc $\mathbb{E}(X_\infty|\mathcal{F}_n) = X_n$.

d)⇒a) Mamy

$$\mathbb{E}|X_n|^p = \mathbb{E}|\mathbb{E}(X|\mathcal{F}_n)|^p \leqslant \mathbb{E}(\mathbb{E}(|X|^p|\mathcal{F}_n)) = \mathbb{E}|X|^p < \infty. \quad \square$$

5.1. Zadania

1. Załóżmy, że $(\mathcal{F}_n)$ jest filtracją, a (X_n) jest ciągiem zmiennych losowych adaptowanych do tej filtracji. Niech B będzie podzbiorem borelowskim $\mathbb{R}$.

a) Udowodnić, że $\tau_1 = \inf\{n : X_n + n \in B\}$ jest momentem zatrzymania.

b) Udowodnić, że dla dowolnego momentu zatrzymania τ, zmienna $\tau_2 = \inf\{n > \tau : X_n \in B\}$ też jest momentem zatrzymania.

2. Dany jest ciąg $(X_n)_{n=1}^{10}$ niezależnych zmiennych losowych o rozkładzie $\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = 1/2$. Niech

$$\tau = \inf\{n > 1 : X_n > X_{n-1}\}, \qquad \sigma = \sup\{n \geqslant 1 : X_n > X_{n-1}\}$$

(przyjmujemy $\inf\emptyset = \sup\emptyset = \infty$). Czy τ, σ są momentami zatrzymania?

3. Zmienne τ, σ są momentami zatrzymania względem filtracji $(\mathcal{F}_n)_{n=0,1,2,\ldots}$. Czy zmienne τ^2, $\tau+1$, $\tau+\sigma$, $\tau-1$, $\tau \wedge (2\sigma)$ są momentami zatrzymania?

4. Dany jest ciąg (ξ_n) niezależnych zmiennych losowych o tym samym rozkładzie $\mathbb{P}(\xi_n = -1) = \mathbb{P}(\xi_n = 1) = 1/2$. Niech $X_0 = 0$ i $X_n = \xi_1 + \xi_2 + \ldots + \xi_n$ dla $n \geqslant 1$. Niech $(\mathcal{F}_n)$ będzie naturalną filtracją generowaną przez ciąg (X_n).

a) Udowodnić,że (X_n) oraz $(X_n^2 - n)$ są martyngałami.

b) Wyznaczyć taką wartość parametru a, by ciąg $(a^n \cos X_n)$ był martyngałem.

c) Udowodnić, że dla $\lambda > 0$, ciąg $(\exp(\lambda X_n - \lambda^2 n/2))$ jest nadmartyngałem.

5. Załóżmy, że $(X_n)_{n=0}^{\infty}$ jest ciągiem niezależnych zmiennych loswych o tym samym rozkładzie o średniej 0. Niech $Z_0 = 0$, $Z_n = X_0X_1 + X_1X_2 + \ldots + X_{n-1}X_n$ dla $n \geqslant 1$. Udowodnić, że ciąg (Z_n) jest martyngałem.

6. Dany jest ciąg (X_n) adaptowany do filtracji $(\mathcal{F}_n)$. Udowodnić, że ciąg (X_n) jest martyngałem wtedy i tylko wtedy gdy dla każdego ograniczonego momentu zatrzymania τ zachodzi równość $\mathbb{E}X_\tau = \mathbb{E}X_0$.

7. Dany jest martyngał $(X_n, \mathcal{F}_n)_{n=0,1,2,\ldots}$ oraz moment zatrzymania τ. Udowodnić, że $(X_{\tau\wedge n}, \mathcal{F}_n)$ też jest martyngałem.

8. Egzaminator przygotował m zestawów pytań. Studenci kolejno losują kartki z pytaniami, przy czym zestaw raz wyciągnięty nie wraca do ponownego losowania. tudent nauczył się odpowiedzi na k zestawów ($k \leqslant m$). Obserwując przebieg egzaminu chce przystąpić do niego w takim momencie, żeby zmaksymalizować szanse zdania. Czy istnieje strategia optymalna?

9. Gramy w orła i reszkę symetryczną monetą. Przed n-tą grą, opierając się ewentualnie na wynikach poprzednich gier, sami ustalamy stawkę w n-tej grze: wybieramy V_n, $1 \leqslant V_n \leqslant a$, i jeśli wypadnie orzeł dostajemy V_n zł, jeśli reszka - płacimy V_n zł. Niech (S_n) oznacza łączną wygraną po n grach. Udowodnić, że $(S_n)_n$ jest martyngałem (względem naturalnej filtracji).

10. Mamy 10 zł w monetach 1 zł, a potrzebujemy pilnie 20 zł. Jedynym sposobem zdobycia tych pieniędzy jest gra w 3 karty z szulerem (który wygrywa z prawdopodobieństwem 2/3). Szuler gotów jest grać z nami wiele razy o dowolne stawki, jakie jesteśmy w stanie założyć (przyjmijmy dla uproszczenia, że stawka nie przekracza 10 zł). Udowodnić, że niezależnie od wyboru strategii nasze szanse na uzyskanie brakujących 10 zł nie przekraczają $1/3$.

11. (Tożsamość Walda). Dany jest ciąg (X_n) całkowalnych zmiennych losowych o tym samym rozkładzie, adaptowany do filtracji $(\mathcal{F}_n)_{n=1,2,\ldots}$, taki, że zmienna X_{n+1} jest niezależna od $\mathcal{F}_n$. Udowodnić, że dla dowolnego momentu zatrzymania τ takiego, że $\mathbb{E}\tau < \infty$, zachodzi wzór

$$\mathbb{E}(X_1 + X_2 + \ldots + X_\tau) = \mathbb{E}X_1 \cdot \mathbb{E}\tau.$$

12. Załóżmy, że $X_1, X_2, \ldots$ są niezależnymi zmiennymi losowymi o średniej 0, spełniającymi warunek $\sum_{n=1}^\infty \mathrm{Var} X_n < \infty$. Udowodnić, że szereg $\sum_{n=1}^\infty X_n$ jest zbieżny p.n.

W zadaniach 13 - 17 poniżej rozpatrujemy ciąg $X_1, X_2, \ldots$ niezależnych zmiennych losowych o rozkładzie $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = -1)$, i oznaczamy $S_0 = 0$, $S_n = X_1 + X_2 + \ldots + X_n$ dla $n \geqslant 1$. Dla $a, b \in \mathbb{Z}$, $a, b > 0$, niech $\tau_a = \inf\{n : S_n = a\}$ oraz $\tau_{a,b} = \inf\{n : S_n \in \{-a, b\}\}$.

13. Załóżmy, że $p = 1/2$ i niech $\tau = \tau_{a,b}$. Korzystając z teorii martyngałów obliczyć $\mathbb{P}(S_\tau = -a)$, $\mathbb{P}(S_\tau = b)$ oraz $\mathbb{E}\tau$.

14. Rozwiązać zadanie 13 przy założeniu $1/2 < p < 1$.

15. Udowodnić, że $\mathbb{E}\tau_a = \infty$.

16. Załóżmy, że $p = 1/2$ oraz τ jest całkowalnym momentem zatrzymania. Udowodnić, że $\mathbb{E}S_\tau = 0$ oraz $\mathbb{E}S_\tau^2 = \mathbb{E}\tau$.

17. Zbadać zbieżność p.n. oraz w L^p nadmartyngału $(\exp(S_n - n/2))_{n=0}^\infty$ (por. zadanie 4 c)).

18. Zmienne $X_1, X_2, \ldots$, są niezależne i mają ten sam rozkład skoncentrowany na liczbach nieujemnych, różny od $\delta_{\{1\}}$, o średniej 1. Udowodnić, że ciąg $(X_1 X_2 \ldots X_n)$ jest zbieżny p.n., ale nie jest zbieżny w L^1.

19. W pojemniku znajduje się pewna liczba cząstek, z których każda w chwili n z równym prawdopodobieństwem albo dzieli się na dwie, albo ginie. W chwili 0 liczba cząstek wynosi 1. Udowodnić, że z prawdopodobieństwem 1 po pewnym czasie wszystkie cząstki zginą, tzn. w pojemniku nie będzie ani jednej cząstki.

6. Łańcuchy Markowa

6.1. Podstawowe definicje

Zajmiemy się teraz kolejną ważną klasą procesów stochastycznych: łańcuchami Markowa.

Niech E będzie „przestrzenią stanów”: skończonym lub przeliczalnym zbiorem. Jego elementy będziemy oznaczać literami j, k, ... bądź 1, 2, ...

Definicja 6.1. Macierz $P = [p_{ij}]_{(i,j)\in E\times E}$ nazywamy macierzą stochastyczną, jeśli $p_{ij} \in [0,1]$ dla wszystkich i, $j \in E$ oraz $\sum_{j\in E} p_{ij} = 1$ dla każdego $i \in E$.

Definicja 6.2. Załóżmy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest przestrzenią probabilistyczną, E, P są j.w., ustalone. Jednorodnym łańcuchem Markowa o wartościach w E i macierzy przejścia P nazywamy ciąg $(X_n)_{n=0,1,2,\ldots}$ zmiennych losowych takich, że

$$\begin{aligned}\mathbb{P}(X_{n+1} = a_{n+1}|X_n = a_n,\, X_{n-1} = a_{n-1}, \ldots, X_0 = a_0)&\\ = \mathbb{P}(X_{n+1} = a_{n+1}|X_n = a_n) &= p_{a_n a_{n+1}}\end{aligned}$$

dla wszystkich a_0, a_1, ..., a_{n+1} takich, że zdarzenie warunkujące ma dodatnie prawdopodobieństwo. Równoważnie,

$$\mathbb{P}(X_{n+1} = j|X_0, X_1, \ldots, X_n) = \mathbb{P}(X_{n+1} = j|X_n) = p_{X_n j}.$$

Liczba p_{ij} jest prawdopodobieństwem przejścia ze stanu i do stanu j w jednym kroku.

Przykłady:

1) Załóżmy, że $X_0, X_1, X_2, \ldots$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie, przyjmującymi wartości w zbiorze E, przy czym $\mathbb{P}(X_n = j) = p_j$ dla $j \in E$. Wówczas

$$\begin{aligned}\mathbb{P}(X_{n+1} = a_{n+1}|X_n = a_n, \ldots, X_0 = a_0) &= \mathbb{P}(X_{n+1} = a_{n+1}) = p_j\\ &= \mathbb{P}(X_{n+1} = a_{n+1}|X_n = a_n),\end{aligned}$$

a zatem $(X_n)_n$ jest łańcuchem Markowa; mamy

$$P = \begin{bmatrix} p_1 & p_2 & \cdots \\ p_1 & p_2 & \cdots \\ p_1 & p_2 & \cdots \\ \cdots & & \end{bmatrix}.$$

2) (Błądzenie losowe) Załóżmy, że ξ_1, ξ_2, ... są niezależnymi zmiennymi losowymi o rozkładzie $\mathbb{P}(\xi_n = 1) = p$, $\mathbb{P}(\xi_n = -1) = q = 1 - p$. Niech $X_0 = 0$, $X_n = \xi_1 + \xi_2 + \ldots + \xi_n$ dla $n \geqslant 1$. Mamy

$$\begin{aligned}\mathbb{P}(X_{n+1} = a_{n+1}|X_n, X_{n-1}, \ldots, X_0) &= \mathbb{P}(X_n + \xi_{n+1} = a_{n+1}|\xi_0, \xi_1, \ldots, \xi_n)\\ &= \begin{cases} p & \text{jeśli } X_n = a_{n+1} - 1,\\ q & \text{jeśli } X_n = a_{n+1} + 1,\\ 0 & \text{w pozostałych przypadkach}\end{cases}\\ &= \mathbb{P}(X_{n+1} = a_{n+1}|X_n).\end{aligned}$$

3) (Błądzenie z pochłanianiem na brzegu). Załóżmy, że $a, b \in \mathbb{Z}_+$, (X_n) jak poprzednio, przy czym po dojściu do $-a$ bądź b proces zatrzymuje się. Otrzymany ciąg zmiennych losowych także jest łańcuchem Markowa. Mamy $E = \{-a, -a+1, \dots, b-1, b\}$ i

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 \\ 0 & 0 & q & 0 & \dots & 0 & 0 \\ \dots & & & & & & \\ 0 & 0 & 0 & 0 & \dots & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

4) (Błądzenie z odbiciem od brzegu). Niech $a, b \in \mathbb{Z}_+$, (X_n) jak poprzednio, przy czym po dojściu do $-a$ przechodzimy do $-a+1$, po dojściu do b przechodzimy do $b-1$. Wówczas otrzymany proces jest łańcuchem Markowa o macierzy przejścia jak poprzednio, tyle że pierwszy wiersz to $[0, 1, 0, 0, \dots, 0]$, a ostatni - $[0, 0, \dots, 0, 1, 0]$.

5) (Model dyfuzji cząstek). Danych jest n cząstek w dwóch pojemnikach: I i II. Co jednostkę czasu losowo wybrana cząstka przemieszcza się z jednego pojemnika do drugiego.

Niech X_n oznacza liczbę cząstek w pojemniku I w chwili n. Przestrzeń stanów to $E = \{0, 1, 2, \dots, n\}$ oraz

$$p_{i,i+1} = \frac{n-i}{n} \text{ dla } i \leqslant n-1, \;\; p_{i,i-1} = \frac{i}{n} \text{ dla } i \geqslant 1.$$

Pozostałe p_{ij} są równe 0:

$$P = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 1/n & 0 & 1-1/n & \dots & 0 & 0 \\ 0 & 2/n & 0 & \dots & 0 & 0 \\ 0 & 0 & 3/n & \dots & 0 & 0 \\ \dots 0 & 0 & 0 & \dots & 0 & 1/n \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Stwierdzenie 6.1. *Załóżmy, że (X_n) jest łańcuchem Markowa z macierzą przejścia $P = [p_{ij}]$. Wówczas dla każdej ograniczonej funkcji $f : E \to \mathbb{R}$ i dowolnego n,*

$$(*) \qquad \mathbb{E}(f(X_{n+1})|X_0, X_1, \dots, X_n) = \mathbb{E}(f(X_{n+1})|X_n) = \sum_{j \in E} f(j) p_{X_n j}.$$

Dowód:. Mamy

$$\begin{aligned} \mathbb{E}(f(X_{n+1})|X_0, X_1, \dots, X_n) &= \sum_{j \in E} \mathbb{E}(f(X_{n+1}) 1_{\{X_{n+1}=j\}}|X_0, X_1, \dots, X_n \\ &= \sum_{j \in E} f(j) \mathbb{P}(X_{n+1} = j|X_0, X_1, \dots, X_n) \\ &= \sum_{j \in E} f(j) p_{X_n j}. \end{aligned}$$

Stąd równość skrajnych stron w (*). Aby otrzymać prawą równość, wystarczy powtórzyć powyższe rozumowanie z warunkowaniem tylko po zmiennej X_n. $\square$

Twierdzenie 6.1. *Załóżmy, że* (X_n), E, P *- jak wyżej. Wówczas dla wszystkich* $n = 0, 1, 2, \ldots$, $k = 1, 2, \ldots$, $j \in E$,

$$\mathbb{P}(X_{n+k} = j | X_0, X_1, \ldots, X_n) = \mathbb{P}(X_{n+k} = j | X_n) = p^{(k)}_{X_n j},$$

gdzie $[p^{(k)}_{ij}]_{i,j \in E} = P^k$. *Macierz tę możemy interpretować jako* macierz przejścia w k krokach.

Dowód:. Stosujemy indukcję ze względu na k. Dla $k = 1$ dostajemy definicję łańcucha Markowa. Przypuśćmy, że teza zachodzi dla pewnego $k \geqslant 1$. Mamy

$$\begin{aligned}\mathbb{P}(X_{n+k+1} = j | X_0, X_1, \ldots, X_n) &= \mathbb{E}(1_{\{j\}}(X_{n+k+1}) | X_0, X_1, \ldots, X_n) \\ &= \mathbb{E}(\mathbb{E}(1_{\{j\}}(X_{n+k+1}) | X_0, X_1, \ldots, X_{n+1}) | X_0, X_1, \ldots, X_n) \\ &\overset{\text{z.ind.}}{=} \mathbb{E}(p^{(k)}_{X_{n+1} j} | X_0, X_1, \ldots, X_n) \overset{\text{Stw.}}{=} \sum_{i \in E} p^{(k)}_{ij} p_{X_n i} \overset{\text{z.ind.}}{=} p^{(k+1)}_{X_n j}.\end{aligned}$$

□

Wniosek 6.1 (Równanie Chapmana-Kołmogorowa)**.** *Dla wszystkich* k, $n \geqslant 1$ *oraz* $i, j \in E$,

$$p^{k+n}_{ij} = \sum_{i \in E} p^k_{il} p^n_{lj}.$$

Dowód:. Wynika to natychmiast z równości $P^{(k+n)} = P^k P^n$. □

Stwierdzenie 6.2. *Przy założeniach jak wyżej, dla dowolnego* $n \geqslant 0$ *oraz* i_0, i_1, $\ldots$, $i_n \in E$,

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \mathbb{P}(X_0 = i_0) p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.$$

Dowód:. Mamy

$$\begin{aligned}\mathbb{E}(1_{\{i_0\}}(X_0) 1_{\{i_1\}}(X_1) \ldots 1_{\{i_n\}}(X_n)) &= \mathbb{E}[\mathbb{E}(\ldots | X_0, X_1, \ldots, X_{n-1})] \\ &= \mathbb{E}[1_{\{i_0\}}(X_0) 1_{\{i_1\}}(X_1) \ldots 1_{\{i_n\}}(X_{n-1})) \underbrace{\mathbb{P}(X_n = i_n | X_0, X_1, \ldots, X_{n-1})}_{p_{X_{n-1} i_n} = p_{i_{n-1} i_n}}]\end{aligned}$$

itd. □

Definicja 6.3. Rozkład zmiennej X_0 nazywamy rozkładem początkowym. Jest on jednoznacznie wyznaczony przez ciąg $(\pi_i)_{i \in E}$ liczb nieujemnych o sumie 1.

Podane niżej twierdzenie mówi, iż każda macierz stochastyczna i rozkład początkowy prowadzą do pewnego łańcucha Markowa. Twierdzenie to pozostawimy bez dowodu.

Twierdzenie 6.2. *Niech* E *będzie zbiorem co najwyżej przeliczalnym. Wówczas dla każdej macierzy stochatycznej* P *oraz miary probabilistycznej* π *na* E *istnieje przestrzeń probabilistyczna i określony na niej łańcuch Markowa o macierzy przejścia* P *i rozkładzie początkowym* π.

6.2. Klasyfikacja stanów

Mówimy, że stan j jest osiągalny ze stanu i, jeśli $p_{ij}^{(n)} > 0$ dla pewnego $n \geqslant 1$. Mówimy, że stany i oraz j się komunikują, jeśli j jest osiągalny z i oraz i jest osiągalny z j. Stan i jest nieistotny, jeśli istnieje taki stan j, że j jest osiągalny z i oraz i nie jest osiągalny z j. Jak łatwo sprawdzić, korzystając z równania Chapmana-Kołmogorowa, relacja „osiągalności" jest przechodnia: istotnie, jeśli j jest osiągalny z i oraz k jest osiągalny z j, to dla pewnych m, $n \geqslant 1$, $p_{ij}^{(m)} > 0$ i $p_{jk}^{n} > 0$, a zatem

$$p_{ik}^{m+n} = \sum_{l \in E} p_{il}^{m} p_{lk}^{(n)} \geqslant p_{ij}^{(m)} p_{jk}^{(n)} > 0.$$

Aby zilustrować powyższe definicje, odnotujmy, iż dla błądzenia losowego po liczbach całkowitych (przykład 2 powyżej), wszystkie stany wzajemnie się komunikują. Natomiast w przypadku pochłaniania na brzegu (przykład 3), stany $-a+1, -a+2, \ldots, b-1$ są nieistotne.

Zbiór stanów S nazywamy zamkniętym, jeśli dla dowolnego $i \in S$ oraz $j \notin S$, stan j nie jest osiągalny ze stanu i (innymi słowy, startując ze stanu wewnątrz C, z prawdopodobieństwem 1 nie wychodzimy nigdy z tego zbioru). Jeśli stan k ma tę własność, że zbiór $\{k\}$ jest zamknięty (tzn. mamy $p_{kk}^{(n)} = 1$ dla wszystkich n), to stan ten nazywamy pochłaniającym. łańcuch Markowa nazywamy nieprzywiedlnym, jeśli wszystkie stany komunikują się ze sobą. Przykładowo, łańcuch Markowa pojawiający się w modelu dyfuzji cząstek (przykład 5) jest nieprzywiedlny.

Uwaga: Załóżmy, że C jest zamkniętym zbiorem stanów łańcucha Markowa o macierzy przejścia $P = [p_{ij}]_{(i,j)\in E\times E}$. Wówczas możemy rozpatrzyć łańcuch Markowa „zawężony" do C: jako nową przestrzeń stanów bierzemy zbiór C, a macierz przejścia zadana jest przez $[p_{ij}]_{(i,j))]inC\times C}$ (wykreślamy z macierzy P wiersze i kolumny odpowieadające stanom nienależącym do C). Przykładowo, załóżmy, że macierz przejścia łańcucha na $E = \{1, 2, 3, 4\}$ wynosi

$$\begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 1/3 & 1/2 & 1/6 \\ 1/5 & 1/5 & 2/5 & 1/5 \end{bmatrix}.$$

Jak łatwo zauważyć, zbiór $C = \{1, 2\}$ jest zamknięty i indukuje łańcuch Markowa o wartościach w tym zbiorze i macierzy przejścia $\begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$.

Niech

$$F_{kj} = \mathbb{P}(\bigcup_{n=1}^{\infty} \{X_n = j\} | X_0 = k)$$

będzie prawdopodobieństwem tego, że startując z k łańcuch dojdzie kiedyś do stanu j. Mamy

$$F_{kj} = \sum_{n=1}^{\infty} f_{kj}(n),$$

gdzie

$$f_{kj} = \mathbb{P}(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j | X_0 = k)$$

jest prawdopodobieństwem że startując z k łańcuch dochodzi do j po raz pierwszy w chwili n.

Definicja 6.4. Stan j nazywamy powracającym, jeśli $F_{jj} = 1$. Stan j nazywamy chwilowym (tranzytywnym), jeśli $F_{jj} < 1$.

Niech $N_j = \sum_{n=1}^{\infty} 1_{\{X_n=j\}}$ będzie zmienną losową zliczającą ile razy proces (X_n) był w stanie j (nie biorąc pod uwagę zmiennej X_0). Mamy następujący fakt.

Stwierdzenie 6.3. *(i) Stan j jest powracający wtedy i tylko wtedy, gdy $\mathbb{P}(N_j = \infty|X_0 = j) = 1$.*
(ii) Stan j jest chwilowy wtedy i tylko wtedy, gdy $\mathbb{P}(N_j < \infty|X_0 = j) = 1$.

Dowód:. Niech $A_k = \{$proces (X_n) był w stanie j co najmniej k razy$\}$. Oczywiście $A_{k+1} \subseteq A_k$, a zatem z twierdzenia o ciagłości

$$\lim_{k\to\infty} \mathbb{P}(A_k|X_0 = i) = \mathbb{P}(\bigcap_{k=1}^{\infty} A_k|X_0 = i) = \mathbb{P}(N_j = \infty|X_0 = i).$$

Wykażemy, że

$$(*) \qquad \mathbb{P}(A_k|X_0 = i) = F_{ij}F_{jj}^{k-1}.$$

Wówczas dostaniemy (stosując tę równość dla $i = j$), iż $\mathbb{P}(N_j = \infty|X_0 = j) = 1$ wtedy i tylko wtedy, gdy $F_{jj} = 1$ (to teza (i)) oraz $\mathbb{P}(N_j = \infty|X_0 = j) = 0$ wtedy i tylko wtedy, gdy $F_{jj} < 1$ (co jest równoważne tezie (ii)).

Pozostaje więc udowodnić (*). Intuicyjnie ten wynik jest jasny: jeśli mamy k razy odwiedzić stan j (przy założeniu, że startujemy z i), to musimy dojść z i do j, a potem $k-1$ razy powrócić do j po pewnych liczbach kroków. Formalnie, ustalmy $n_1 < n_2 < \ldots < n_k$. Mamy

$$\begin{aligned}
&\mathbb{P}(X_{n_l} = j \text{ dla } l = 1, 2, \ldots, k,\ X_n \neq j \text{ dla } n \neq n_l,\ n \leqslant n_k|X_0 = i) \\
&\quad = \mathbb{P}(X_1 \neq j, X_2 \neq j, \ldots, X_{n_1-1} \neq j, X_{n_1} = j|X_0 = i)\times \\
&\quad\quad \times \prod_{l=2}^{k} [\mathbb{P}(X_{n_{l-1}+1} \neq j, X_{n_{l-1}+2} \neq j, \ldots, X_{n_l-1} \neq j, X_{n_l} = j|X_{n_{l-1}} = j)] \\
&\quad = f_{ij}(n_1) \prod_{l=2}^{n} f_{jj}(n_l - n_{l-1}).
\end{aligned}$$

Podstawmy $m_1 = n_1$, $m_k = n_k - n_{k-1}$. Mamy, dla $i \neq j$,

$$\begin{aligned}
\mathbb{P}(A_k|X_0 = i) &= \sum_{m_1,\ldots,m_k \geqslant 1} f_{ij}(m_1)f_{jj}(m_2)\ldots f_{jj}(m_k) \\
&= \left(\sum_{m_1=1}^{\infty} f_{ij}(m_1)\right) \prod_{l=2}^{k} \left(\sum_{m_l=1}^{\infty} f_{jj}(m_l)\right) = F_{ij}F_{jj}^{k-1}.
\end{aligned}$$

□

Uwaga: W szczególności, $\mathbb{P}(N_j = \infty|X_0 = i) = 0$ jeśli j jest stanem chwilowym. Zatem w przypadku stanu chwilowego, proces odwiedza go skończenie wiele razy niezależnie od punktu startowego.

Podane niżej twierdzenie charakteryzuje stany chwilowe i powracające w terminach macierzy przejścia. Wprowadźmy, dla każdego $j \in E$, liczbę $P_j = \sum_{n=1}^{\infty} p_{jj}^{(n)}$. Na mocy twierdzenia Fubiniego (dla funkcji nieujemnych),

$$P_j = \sum_{n=1}^{\infty} \mathbb{E}(1_{\{j\}}(X_n)|X_0 = j) = \mathbb{E}(N_j|X_0 = j),$$

czyli P_j jest średnim czasem przebywania łańcucha w stanie j (przy założeniu startowania z tego stanu), nie licząc zmiennej X_0.

Twierdzenie 6.3. *(i) Stan j jest chwilowy wtedy i tylko wtedy, gdy $P_j < \infty$.*
(ii) Stan j jest powracający wtedy i tylko wtedy, gdy $P_j = \infty$.

Dowód:. Zauważmy, iż na mocy wzoru na prawdopodobieństwo całkowite,

$$p_{jj}^{(k)} = \sum_{m=0}^{k-1} f_{jj}(k-m) p_{jj}^{(m)},$$

gdzie przyjmujemy $p_{jj}^{(0)} = 1$. Zatem, na mocy twierdzenia Fubiniego (dla funkcji nieujemnych),

$$\begin{aligned}\sum_{k=1}^{n} p_{jj}^{(k)} &= \sum_{k=1}^{n} \sum_{m=0}^{k-1} f_{jj}(k-m) p_{jj}^{(m)} \\ &= \sum_{m=0}^{n-1} p_{jj}^{(m)} \sum_{k=m+1}^{n} f_{jj}(k-m) \leqslant \sum_{m=0}^{n} p_{jj}^{(m)} F_{jj} \\ &= F_{jj} + F_{jj} \sum_{m=1}^{n} p_{jj}^{(m)},\end{aligned}$$

czyli, równoważnie,

$$(1 - F_{jj}) \sum_{m=1}^{n} p_{jj}^{(m)} \leqslant F_{jj}. \tag{Δ}$$

Jeśli stan j jest chwilowy, to (Δ) daje, iż

$$P_j \leqslant \frac{F_{jj}}{1 - F_{jj}} < \infty.$$

I w drugą stronę: jeśli $P_j < \infty$, czyli $\mathbb{E}(N_j | X_0 = j) < \infty$, to $\mathbb{P}(N_j = \infty | X_0 = j) = 0$ i na mocy poprzedniego twierdzenia j jest stanem chwilowym. To dowodzi (i). Część (ii) jest konsekwencją tego, że każdy stan jest albo chwilowy, albo powracający oraz tego, że P_j jest albo skończone, albo nie. □

Przykład: Zbadamy błądzenie losowe po liczbach całkowitych. Mamy

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n \sim \frac{1}{\sqrt{\pi n}} (4pq)^n,$$

na mocy wzoru Stirlinga. Ponadto, $p_{00}^{(2n+1)} = 0$. Stąd

$$\sum_{n=1}^{\infty} p_{00}^{(n)} < \infty$$

wtedy i tylko wtedy,gdy $p \neq 1/2$. Zatem 0 jest stanem powracającym wtedy i tylko wtedy, gdy $p = 1/2$.

W przypadku gdy wszystkie stany komunikują się wzajemnie, stany muszą być tego samego typu.

Twierdzenie 6.4. *Załóżmy, że łańcuch Markowa jest nieprzywiedlny. Wówczas jeśli jeden stan jest chwilowy, to wszystkie są chwilowe; jeśli jeden stan jest powracający, to wszystkie są powracające.*

Możemy więc mówić o łańcuchach określonego typu: chwilowych i powracających.

Dowód. Weźmy dwa stany i, j. Istnieją liczby całowite dodatnie r, s takie, że $\alpha = p_{ij}^{(r)} > 0$, $\beta = p_{ji}^{(s)} > 0$. Dla $n \geqslant 1$ mamy

$$p_{ii}^{(r+s+n)} \geqslant p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} = \alpha\beta p_{jj}(n)$$

i podobnie $p_{jj}(r+s+n) \geqslant \alpha\beta p_{ii}^{(n)}$. Zatem dla $n > r+s$,

$$\frac{1}{\alpha\beta} p_{ii}^{(r+s+n)} \geqslant p_{jj}^{(n)} \geqslant \alpha\beta p_{ii}^{(n-r-s)},$$

czyli asymptotyczne zachowanie ciągów $(p_{ii}^{(n)})_n$ oraz $(p_{jj}^{(n)})_n$ jest takie samo; w szczególności, $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ wtedy i tylko wtedy, gdy $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$. □

Na zakończenie - następujący fakt dotyczący struktury stanów łańcucha Markowa ze względu na stany chwilowe i powracające (bez dowodu).

Stwierdzenie 6.4. *Przestrzeń stanów E łańcucha Markowa możemy jednoznacznie przedstawić w postaci*

$$E = C \cup D_1 \cup D_2 \cup \ldots,$$

gdzie C jest zbiorem stanów chwilowych, a D_i, $i \geqslant 1$ są nieprzywiedlnymi zamkniętymi zbiorami stanów powracających.

Przy danym rozbiciu przestrzeni E jak w powyższym stwierdzeniu, z prawdopodobieństwem 1 łańcuch Markowa zachowuje się następująco. Jeśli startuje on w zbiorze D_i, $i \geqslant 1$, to nigdy go nie opuszcza i odwiedza wszystkie elementy tego zbioru; jeśli startuje on w zbiorze C, to albo pozostaje tam na zawsze (co może mieć miejsce tylko wtedy, gdy C ma nieskończenie wiele elementów), albo po skończonej liczbie kroków trafia do jednego ze zbiorów D_i, i pozostaje tam na zawsze.

6.3. Rozkłady stacjonarne i twierdzenie ergodyczne

Definicja 6.5. Załóżmy, że P jest macierzą stochastyczną. Rozkład π na E nazywamy stacjonarnym (niezmienniczym), jeśli $\pi P = \pi$ (tzn. dla wszystkich $j \in E$, $\sum_{i \in E} \pi_i p_{ij} = \pi_j$.

Rozkład stacjonarny ma następujące własności. Po pierwsze zauważmy, że jeśli π jest rozkładem stacjonarnym, to dla każdego $n \geqslant 1$, $\pi P^n = \pi$ (oczywista indukcja). Innymi słowy, jeśli (X_n) jest łańcuchem Markowa o macierzy przejścia P i rozkładzie początkowym π, to dla $n \geqslant 1$, rozkład X_n jest równy π. Można nawet powiedzieć więcej: dla wszystkich $n \geqslant 1$ oraz dowolnego ciągu $m_1 < m_2 < \ldots < m_k$ ($k \geqslant 1$ również jest dowolne) wektor $(X_{m_1}, X_{m_2}, \ldots, X_{m_k})$ ma ten sam rozkład co $(X_{n+m_1}, X_{n+m_2}, \ldots, X_{n+m_k})$. Istotnie,

$$\mathbb{P}(X_{n+m_1} = j_1, X_{n+m_2} = j_2, \ldots, X_{n+m_k} = j_k)$$

$$= \sum_{i\in E} \pi_i p_{ij_1}^{m_1+n} p_{j_1j_2}^{(m_2-m_1)} \dots p_{j_{k-1}j_k}^{(m_k-m_{k-1})} = \pi_{j_1} p_{j_1j_2}^{(m_2-m_1)} \dots p_{j_{k-1}j_k}^{(m_k-m_{k-1})},$$

co nie zależy od n.

łańcuch o takiej własności nazywamy *stacjonarnym.*

Definicja 6.6. Okresem stanu j nazywamy największą taką liczbę n, że powrót do stanu j jest możliwy tylko po liczbie kroków podzielnej przez n: $o(j)$ =NWD$\{n : p_{jj}^{(n)} > 0\}$.

Stan nazywamy okresowym jeśli $o(j) > 1$ i nieokresowym, jeśli $o(j) = 1$.

Stwierdzenie 6.5. *W nieprzywiedlnym łańcuchu Markowa wszystkie stany mają ten sam okres.*

Wobec tego następująca definicja ma sens.

Definicja 6.7. Nieprzywiedlny łańcuch Markowa (X_n) nazywamy okresowym, jeśli wszystkie jego stany mają okres większy niż 1. W przeciwnym razie łańcuch nazywamy nieokresowym.

Lemat 6.1. *łańcuch jest nieprzywiedlny i nieokresowy wtedy i tylko wtedy, gdy jest spełniony warunek*

$$(O) \qquad \forall_{i,j\in E} \exists_{n_0} \forall_{n\geqslant n_0} p_{ij}^{(n)} > 0.$$

Dowód:. Oczywiście wystarczy tylko udowodnić implikację $\Rightarrow$. Ustalmy i, $j \in E$ oraz liczbę m taką, że $p_{ij}^{(m)} > 0$. Z definicji nieokresowości, istnieją liczby względnie pierwsze $n_1, n_2, \dots, n_k$ takie, że $p_{jj}^{(n_l)} > 0$, $l = 1, 2, \dots, k$. Jeśli n jest dostatecznie duże, to

$$n = a_1n_1 + a_2n_2 + \dots + a_kn_k, \qquad \text{dla pewnych } a_l \in \mathbb{Z}_+,$$

i mamy

$$p_{jj}^{(n)} \geqslant \prod_l p_{jj}^{(a_ln_l)} \geqslant \prod_l (p_{jj}^{(n_l)})^{a_l} > 0.$$

Zatem

$$p_{ij}^{(m+n)} \geqslant p_{ij}^{(m)} p_{jj}^{(n)} > 0$$

o ile $m + n$ jest dostatecznie duże. □

Twierdzenie 6.5. *Załóżmy, że warunek (O) jest spełniony i istnieje rozkład stacjonarny π. Wówczas każdy stan jest powracalny, rozkład stacjonarny jest jednoznaczny oraz dla wszystkich i, $j \in E$,*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j.$$

Uwaga: Jak widać, przy założeniach twierdzenia, $p_{ij}^{(n)}$ „przestaje zależeć od i" o ile n jest duże. Innymi słowy, po dużej liczbie kroków łańcuch „zapomina", z jakiego stanu wystartował.

Dowód:. Dowód przeprowadzimy w pięciu krokach.

1. Wszystkie stany są albo powracalne, albo chwilowe. Załóżmy, że ma miejsce ta druga możliwość. Liczba $\sum_{n=1}^{\infty} p_{ij}^{(n)}$ jest średnim czasem przebywania w stanie j przy założeniu startowania ze stanu i. Na mocy własności Markowa, mamy zatem

$$\sum_{k=1}^{\infty} p_{ij}^{(n)} = F_{ij}P_j < \infty,$$

a zatem $p_{ij}^{(n)} \to 0$ gdy $n \to \infty$. Z drugiej strony, dla każdego $j \in E$,

$$\sum_{i \in E} \pi_i p_{ij}^{(n)} = \pi_j$$

i lewa strona dąży do 0 na mocy tw. Lebesgue'a o zmajoryzowanym przejściu do granicy. Stąd $\pi \equiv 0$ i sprzeczność.

2. Rozważmy nową przestrzeń stanów $E \times E$ oraz macierz przejścia $P^{\otimes 2}$ na tej przestrzeni, o wyrazach $p_{(i,j)(k,l)} = p_{ik}p_{jl}$ (oczywiście jest to macierz stochastyczna). Niech $\pi^{\otimes 2} = (\pi_i \cdot \pi_j)_{(i,j) \in E \times E}$ będzie rozkładem na $E \times E$: jest to rozkład stacjonarny dla $P^{\otimes 2}$. Niech (X_n', X_n'') będzie łańcuchem Markowa z tą macierzą przejścia: (X_n') oraz (X_n'') to dwa niezależne łańcuchy Markowa o macierzach przejścia P, startujące ze stanów i, j, odpowiednio. Ponieważ będziemy zmieniać te punkty startowe, wygodnie nam będzie pracować na miarach probabilistycznych $\mathbb{P}_{ij} = \mathbb{P}(\cdot | X_0' = i, X_0'' = j)$. Jak łatwo sprawdzić, warunek (O) jest spełniony; zatem na mocy kroku 1., z każdego stanu (i,j) można dojść do każdego innego; w szczególności do stanu (k,k). Zatem dla wszystkich i, $j \in E$, $\mathbb{P}_{ij}(X_n' = X_n''$ dla pewnego $n) = 1$.

3. Niech $\tau = \inf\{n : X_n' = X_n''\}$. Definiujemy

$$(Y_n', Y_n'') = \begin{cases} (X_n', X_n'') & \text{dla } n < \tau, \\ (X_n', X_n') & \text{dla } n \geqslant \tau. \end{cases}$$

Z powyższej dyskusji wynika, że dla wszystkich i, $j \in E$,

$$\lim_{n \to \infty} \mathbb{P}_{ij}(Y_n' \neq Y_n'') = 0.$$

Sprawdzimy teraz, że (Y_n'), (Y_n'') są łańcuchami Markowa (względem miary probabilistycznej $\mathbb{P}_{ij}$) z macierzą przejścia P. Ograniczymy się tylko do procesu (Y_n'');w przypadku (Y_n') przekształcenia są analogiczne.

$$\begin{aligned}
\mathbb{P}_{ij}(Y_{n+1}'' = k | X_s', X_s'', s \leqslant n) &= \mathbb{P}_{ij}(Y_{n+1}'' = k, \tau < n | X_s', X_s'', s \leqslant n) \\
&\quad + \mathbb{P}_{ij}(Y_{n+1}'' = k, \tau \geqslant n | X_s', X_s'', s \leqslant n) \\
&= \mathbb{P}_{ij}(X_{n+1}' = k | X_s', X_s'', s \leqslant n) 1_{\{\tau < n\}} \\
&\quad + \mathbb{P}_{ij}(X_{n+1}'' = k | X_s', X_s'', s \leqslant n) 1_{\{\tau \geqslant n\}} \\
&= \mathbb{P}_{ij}(X_{n+1}' = k | X_s', s \leqslant n) 1_{\{\tau < n\}} \\
&\quad + \mathbb{P}_{ij}(X_{n+1}'' = k | X_s'', s \leqslant n) 1_{\{\tau \geqslant n\}} \\
&= p_{X_n' k} 1_{\{\tau < n\}} + p_{X_n'' k} 1_{\{\tau \geqslant n\}} = p_{Y_n'' k}
\end{aligned}$$

i wystarczy obłożyć obie strony warunkową wartością oczekiwaną względem ciągu $Y_0'', Y_1'', \ldots, Y_n''$.

4. Pokażemy, że dla i, j, $k \in E$, $|p_{ik}^{(n)} - p_{jk}^{(n)}| \to 0$. Mamy $|\mathbb{P}(A) - \mathbb{P}(B)| \leqslant \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A)$, więc

$$\begin{aligned}
|p_{ik}^{(n)} - p_{jk}^{(n)}| &= |\mathbb{P}_{ij}(Y_n' = k) - \mathbb{P}_{ij}(Y_n'' = k)| \\
&\leqslant \mathbb{P}_{ij}(Y_n' = k, Y_n'' \neq k) + \mathbb{P}_{ij}(Y_n' \neq k, Y_n'' = k) \\
&\leqslant \mathbb{P}_{ij}(Y_n' \neq Y_n'') \to 0.
\end{aligned}$$

5. Mamy, dla wszystkich $k \in E$, $\sum_{i \in E} \pi_i p_{ik}^{(n)} = \pi_k$, skąd, na mocy poprzedniej części oraz twierdzenia Lebesgue'a,

$$\pi_k - p_{jk}^{(n)} = \sum_{i \in E} \pi_i (p_{ik}^{(n)} - p_{jk}^{(n)}) \to 0.$$

Jednoznaczność rozkładu stacjonarnego jest oczywista: π_k jest wyznaczony jako granice $p_{ik}^{(n)}$. □

Na zakończenie zaprezentujemy następujący fakt. Dowodzi się go używając podobnej argumentacji jak w poprzednim twierdzeniu. Szczegóły pozostawiamy czytelnikowi.

Twierdzenie 6.6. *Jeśli E jest zbiorem skończonym i zachodzi warunek (O), to istnieje rozkład stacjonarny i zachodzi teza poprzedniego twierdzenia.*

6.4. Zadania

1. Niech E będzie pewnym zbiorem przeliczalnym. Dany jest ciąg (X_n) niezależnych zmiennych losowych oraz ciąg funkcyjny (f_n), $f_n : E \times \mathbb{R} \to E$. Definiujemy ciąg (Y_n) wzorem

$$Y_{n+1} = f(Y_n, X_n), \ n = 0, 1, 2, \ldots,$$

gdzie Y_0 jest pewną zmienną losową o wartościach w E. Dowieść, że (Y_n) jest łańcuchem Markowa.

2. Dany jest łańcuch Markowa (X_n) na pewnej przestrzeni E oraz różnowartościowa funkcja $f : E \to E$. Wykazać, że $(f(X_n))$ jest łańcuchem Markowa. Co jeśli f nie jest różnowartościowa?

3. Dany jest ciąg (X_n) niezależnych zmiennych losowych o tym samym rozkładzie $P(X_n = \pm 1) = \frac{1}{2}$. Rozstrzygnąć, które z podanych niżej procesów są łańcuchami Markowa:

$$U_0 = 0, \quad U_n = X_1 + X_2 + \ldots + X_n, n \geqslant 1,$$
$$W_n = X_0 X_1 \cdot X_2 \cdot \ldots X_n, \ n \geqslant 0,$$
$$V_n = (-1)^{U_n}, \ n \geqslant 0,$$
$$Y_n = X_n \cdot X_{n+1}, \ n \geqslant 0,$$
$$Z_n = \frac{X_n + X_{n+1}}{2}, \ n \geqslant 0.$$

4. Dany jest ciąg (X_n) niezależnych zmiennych losowych o tym samym rozkładzie $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = -1)$, $p \in (0, 1)$. Niech $S_n = X_1 + X_2 + \ldots + X_n$, $n = 1, 2, \ldots$. Udowodnić, że ciągi

$$Y_n = |S_n|, \ Z_n = \max_{k \leqslant n} S_k - S_n$$

są łańcuchami Markowa.

5. Rzucamy kostką tak długo, aż pojawi się ciąg 16 lub 66. Jakie jest prawdopodobieństwo, że ciąg 16 pojawi się wcześniej?

6. Rzucamy symetryczną monetą aż do momentu, gdy wyrzucimy serię 4 orłów. Obliczyć wartość oczekiwaną liczby przeprowadzonych rzutów.

7. Macierz przejścia łańcucha Markowa $(X_n)_n$ na przestrzeni $E = \{1, 2, 3, 4\}$ dana jest następująco:

$$\mathbb{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

a) Jakie jest prawdopodobieństwo dojścia w dwóch krokach ze stanu 1 do stanu 2?
b) Zakładając, że $X_0 = 1$ p.n. obliczyć prawdopodobieństwo tego, że X_n będzie w stanie 2 przed stanem 4.
c) Zakładając, że $X_0 = 3$ p.n. obliczyć wartość oczekiwaną czasu dojścia do stanu 2.
d) Wyznaczyć rozkład stacjonarny. Czy łańcuch jest okresowy? Czy jest nieprzywiedlny?

8. Po wierzchołkach pięciokąta ABCDE porusza się pionek. W chwili początkowej znajduje się w punkcie A, a w każdym kolejnym ruchu przesuwa się w sposób niezależny od poprzednich ruchów z prawdopodobieństwem $1/2$ do jednego z sąsiednich wierzchołków. Obliczyć
a) prawdopodobieństwo, że pionek powróci do punktu A przed dotarciem do punktu C,
b) wartość oczekiwaną liczby ruchów, jakie wykona pionek przed powrotem do punktu A.

9. Naukowiec mający r parasoli wędruje między domem a biurem, zabierając ze sobą parasol (jeśli jest on pod ręką) wtedy, gdy pada (prawdopodobieństwo p), lecz nie przy bezdeszczowej pogodzie (prawdopodobieństwo $q = 1 - p$). Niech stanem łańcucha Markowa będzie liczba parasoli znajdujących się pod ręką, bez względu na to, czy naukowiec jest w domu, czy w miejscu pracy. Skonstruować macierz przejścia i znaleźć rozkład stacjonarny. Znaleźć przybliżone prawdopodobieństwo zmoknięcia naukowca w danym (odległym) dniu, a następnie wykazać, że 5 parasoli jest w stanie ochronić go w 95% przed zmoknięciem (dla dowolnego p).

10. Proces (X_n) jest łańcuchem Markowa.
(i) Czy dla dowolnego $n \geqslant 0$, liczb $0 \leqslant i_0 < i_1 < \ldots < i_k = n$ oraz stanów a_0, a_1, ..., a_{k+1} mamy

$$\begin{aligned}\mathbb{P}(X_{n+1} = a_{k+1} | X_{i_k} = a_k, X_{i_{k-1}} = a_{k-1}, \ldots, X_{i_0} = a_0)\\ = \mathbb{P}(X_{n+1} = a_{k+1} | X_{i_k} = a_k)?\end{aligned}$$

(ii) Czy dla dowolnego $n \geqslant 0$, liczb $0 \leqslant i_0 < i_1 < \ldots < i_k = n$ oraz zbiorów A_0, A_1, ..., A_{k+1} mamy

$$\begin{aligned}\mathbb{P}(X_{n+1} \in A_{k+1} | X_{i_k} \in A_k, X_{i_{k-1}} \in A_{k-1}, \ldots, X_{i_0} \in A_0)\\ = \mathbb{P}(X_{n+1} \in A_{k+1} | X_{i_k} \in A_k)?\end{aligned}$$

11. Dany jest łańcuch Markowa (X_n) o macierzy przejścia P, której każdy wiersz jest taki sam. Udowodnić, że zmienne X_0, X_1, ... są niezależne.

12. Dany jest łańcuch Markowa (X_n) startujący ze stanu i. Niech $\tau = \inf\{n \geqslant 1 : X_n \neq i\}$. Udowodnić, że τ ma rozkład geometryczny.

13. Rozważamy błądzenie losowe po $\mathbb{Z}^2$: stan $(i, j) \in \mathbb{Z}^2$ komunikuje się w jednym kroku z każdym ze stanów $(i \pm 1, j)$, $(i, j \pm 1)$ z prawdopodobieństwem $1/4$. Udowodnić, że wszystkie stany są powracalne. Udowodnić, że nie istnieje rozkład stacjonarny.

14. Niech α będzie ustaloną liczbą dodatnią. Dany jest łańcuch Markowa na $E = \{1, 2, \ldots\}$ startujący z 1, o następujących prawdopodobieństwach przejścia: stan $k \in E$ prowadzi w jednym kroku do 1 z prawdopodobieństwem $(k+1)^{-\alpha}$ oraz do $k+1$ z prawdopodobieństwem $1-(k+1)^{-\alpha}$. Czy łańcuch jest okresowy? Czy jest nieprzywiedlny? Dla jakich α łańcuch jest powracalny? Dla jakich α istnieje rozkład stacjonarny?

15. Dany jest spójny graf (W, K) o skończonej liczbie wierzchołków oraz łańcuch Markowa o wartościach w V taki, że z każdego wierzchołka $x \in V$ można w jednym kroku dojść do jednego z wierzchołków sąsiadujących z x. Niech $n(x)$ oznacza liczbę sąsiadów x. Udowodnić, że $\pi_x = n(x)/(2|K|)$ jest rozkładem stacjonarnym.

16. W modelu dyfuzji (przykład 5) powyżej) z $n = 20$, załóżmy, że w chwili 0 nie ma żadnej cząstki w pojemniku I. Wyznaczyć przybliżone prawdopodobieństwo tego, że w chwili 10000 nie będzie żadnej cząstki w I pojemniku.

Literatura