

UNIwersytet Jagiełoński  
Wydział Matematyki i Informatyki

Dawid Tarłowski

# Łańcuchy Markowa i zastosowania

Wykład dla kierunku Matematyka  
Wersja robocza. Proszę nie rozpowszechniać.

KRAKÓW 2020

## 1. ŁAŃCUCHY MARKOWA.

Niech  $(\Omega, \Sigma, \mathbb{P})$  będzie przestrzenią probabilistyczną oraz niech  $(X, d)$  będzie przestrzenią metryczną ośrodkową. Przestrzeń metryczna  $X$  jest naturalnie wyposażona w sigma-algebrę  $\Sigma_X$  zbiorów Borelowskich na  $X$ , tj.  $\Sigma_X = \mathcal{B}(X) = \sigma(\tau_X)$ , gdzie  $\tau_X$  to rodzina otwartych podzbiorów  $X$ . W wielu sytuacjach ciąg zmiennych losowych  $X_n: \Omega \rightarrow X$ ,  $n \in \mathbb{N}$ , będziemy nazywać procesem stochastycznym. Naturalną filtracją procesu  $X_n$  będziemy nazywać filtrację zdefiniowaną  $F_n = \Sigma\{X_0, \dots, X_n\}$ ,  $n \in \mathbb{N}$ .

Wprowadźmy oznaczenie:

$$B(X, \mathbb{R}) = \{f: X \rightarrow \mathbb{R} \mid f \text{ — mierzalna i ograniczona}\}.$$

Niech  $X_n: \Omega \rightarrow X$ ,  $n \in \mathbb{N}$ , będzie ciągiem zmiennych losowych.

**Definicja 1.** Powiemy, że ciąg  $X_n$  jest **łańcuchem Markowa** jeśli dla dowolnej funkcji mierzalnej ograniczonej  $\varphi: X \rightarrow \mathbb{R}$  zachodzi  $\mathbb{P}$ - prawie wszędzie:

$$(1.1) \quad E[\varphi(X_{n+1})|X_0, X_1, \dots, X_n] = E[\varphi(X_{n+1})|X_n] \quad (\text{własność Markowa}).$$

**Fakt 1.** Dla ciągu zm. los.  $X_n$  następujące warunki są równoważne:

- (1)  $E[\varphi(X_{n+1})|X_0, X_1, \dots, X_n] = E[\varphi(X_{n+1})|X_n]$ ,  $\varphi \in B(X, \mathbb{R})$ ,  $n \in \mathbb{N}$ ,
- (2)  $\mathbb{P}[X_{n+1} \in A|X_0, X_1, \dots, X_n] = \mathbb{P}[X_{n+1} \in A|X_n]$ ,  $A \in \mathcal{B}(X)$ ,  $n \in \mathbb{N}$

Zbiór indeksów  $\mathbb{N}$  będziemy interpretować jako czas natomiast przestrzeń  $X$  będziemy często nazywać przestrzenią stanów. Interpretacja własności Markowa jest klarowna: znajomość dotychczasowej historii procesu  $X_0 = x_0, X_1 = x_1, \dots, X_n = x_n$  daje taką samą prognozę przyszłości jak znajomość bieżącego stanu  $X_n = x_n$ . Widać to wyraźnie w poniższych (przykładowych) warunkach równoważnych własności (1.1).

**Fakt 2.** Własność Markowa (1.1) jest równoważna następującym warunkom:

- (1)  $\mathbb{P}[X_{n+k} \in A|X_0, X_1, \dots, X_n] = \mathbb{P}[X_{n+k} \in A|X_n]$ ,  $A \in \mathcal{B}(X)$ ,  $n, k \in \mathbb{N}$ ,
- (2)  $E[\varphi(X_{n+k})|X_0, X_1, \dots, X_n] = E[\varphi(X_{n+k})|X_n]$ ,  $\varphi \in B(X, \mathbb{R})$ ,  $n, k \in \mathbb{N}$ .

**Definicja 2.** Powiemy, że funkcja

$$P: X \times \Sigma_X \ni (x, A) \longrightarrow P(x, A) \in [0, 1]$$

jest **jądrem przejścia**, jeśli

- (1) dla każdego  $x \in X$ ,  $P(x, \cdot)$  jest miarą probabilistyczną na  $\Sigma_X$
- (2) dla każdego  $C \in \Sigma_X$ , funkcja  $x \longrightarrow P(x, C) \in [0, 1]$  jest mierzalna

Niech  $K(X) = \{P: X \times \mathcal{B}(X) \rightarrow [0, 1] \mid P \text{ — jądro przejścia}\}$ .

**Definicja 3.** Powiemy, że ciąg zmiennych losowych  $X_n$  jest **łańcuchem Markowa z jądrem przejścia  $P$**  (z prawdopodobieństwem przejścia  $P$ ), jeśli zachodzi:

$$(1.2) \quad \mathbb{P}[X_{n+1} \in A|X_0, X_1, \dots, X_n] = P(X_n, A), \quad A \in \mathcal{B}(X), n \in \mathbb{N}.$$

Dla łańcucha Markowa z jądrem przejścia  $P$  będziemy zawsze stosować następującą równość:

$$(1.3) \quad P(x, A) = \mathbb{P}[X_{n+1} \in A|X_n = x],$$

dla każdego  $n \in \mathbb{N}$ ,  $x \in X$  oraz wszystkich  $A \in \mathcal{B}(X)$ . By zapewnić pełną precyzję matematyczną oraz jedyność jądra przejścia dla łańcucha Markowa, równość (1.3) wykorzystująca pojęcie rozkładów warunkowych wymagałaby bardziej złożonej definicji łańcucha Markowa niż nasza definicja łańcucha jako ciągu zmiennych losowych. W kolejnym rozdziale zapoznamy się z rekurencyjną postacią łańcucha, która bezpośrednio zapewnia istnienie oraz jednoznaczność jądra przejścia.

Oczywiście warunek (1.2), jako istotnie mocniejszy od bardzo ogólnej własności (1.1), również implikuje że dotychczasowa ewolucja łańcucha w czasie (tj. ewolucja obserwowana do kroku  $n$ ) wpływa na przyszłą ewolucję jedynie poprzez bieżący stan łańcucha  $X_n = x$ . Definicja 3 zakłada jednak istotnie więcej niż jedynie regularną postać dla rozkładów warunkowych występujących w Definicji 1 - zakłada bowiem, że prawdopodobieństwo przejścia  $P$  nie zależy od kroku  $n$  - prawa rządzące ewolucją łańcucha są stałe w czasie. Takie łańcuchy Markowa, których prawdopodobieństwo przejścia zależy tylko od bieżącego stanu łańcucha (nie zależy od kroku  $n \in \mathbb{N}$ ) nazywamy **jednorodnymi łańcuchami Markowa**. Łańcuchy te są naturalnymi modelami dla systemów, których ewolucja przebiega zgodnie z prawami niezmiennymi w czasie oraz zawierającymi losowość. Celem niniejszego wykładu jest przedstawienie ogólnej teorii jednorodnych łańcuchów Markowa z jądrem przejścia (będziemy mówić krótko: łańcuchy Markowa) wraz z uwzględnieniem niektórych zastosowań.

**Uwaga.** W przypadku łańcuchów niejednorodnych rozważa się rodzinę prawdopodobieństw przejścia zależną od czasu, tzn.  $P_n(x, A) = \mathbb{P}(X_{n+1} \in A | X_n = x)$ .

**Przykład 1.** Łańcuchy na dyskretnej skończonej przestrzeni stanów. Niech  $X = \{1, 2, \dots, N\}$  z metryką dyskretną. W takim przypadku jądro przejścia  $P$  łańcucha Markowa  $X_n$  reprezentowane jest w postaci macierzy przejścia  $[P_{i,j}]_{i,j=1,2,\dots,n}$  opisującej prawdopodobieństwa warunkowe

$$P_{i,j} = \mathbb{P}[X_{n+1} = j | X_n = i] = P(i, \{j\}).$$

W przypadku dyskretnym piszemy krótko:  $P(i, j)$  zamiast  $P(i, \{j\})$ .

**Przykład 2.** Poprzedni przykład bezpośrednio uogólnia się na przeliczalną przestrzeń stanów. Rozważmy symetryczne błądzenie losowe po liczbach całkowitych  $X = \mathbb{Z}$  z metryką dyskretną, które definiujemy rekurencyjnie

$$X_{n+1} = X_n + Y_n,$$

gdzie  $X_0, Y_0, Y_1, Y_2, \dots$  to ciąg niezależnych zmiennych losowych oraz zmienne  $Y_n$  mają rozkład dwupunktowy na zbiorze  $\{-1, 1\}$ . Jądro przejścia dane jest formułą

$$P(i, j) = \frac{1}{2} \text{ dla } j \in \{i+1, i-1\} \text{ oraz } P(i, j) = 0 \text{ dla } j \notin \{i+1, i-1\}.$$

**Przykład 3.** Uogólniamy poprzedni przykład do błądzenia losowego na prostej  $X = \mathbb{R}$ :

$$X_{n+1} = X_n + Y_n,$$

gdzie  $Y_0, Y_1, Y_2, \dots$  to ciąg i.i.d, niezależny od punktu startowego  $X_0$ . Nie zakładamy nic o rozkładzie zmiennych  $Y_n: \Omega \rightarrow \mathbb{R}$ , więc błądzenie nie musi być symetryczne. Jądro przejścia można wypisać ogólną formułą:  $P(x, A) = \mathbb{P}(x + Y_1 \in A) = \mathbb{P}_{Y_1}(A - x)$

**Przykład 4.** Układy dynamiczne. Niech  $\varphi: X \rightarrow X$  będzie odwzorowaniem mierzalnym. Teoria dyskretnych układów dynamicznych zajmuje się ewolucją układu  $(X, \varphi)$ , tj. zachowaniem trajektorii  $\varphi^n(x)$ ,  $n \in \mathbb{N}$ ,  $x \in X$ , (zazwyczaj przy założeniu że odwzorowanie  $\varphi$  jest ciągłe). Z naszej perspektywy, taki układ jest szczególnym przypadkiem łańcucha Markowa:  $X_{n+1} = \varphi(X_n)$  - prawa rządzące jego ewolucją są niezmiennie w czasie ale nie zawierają czynnika losowego tak więc prawdopodobieństwa przejścia są zdegenerowane zgodnie ze wzorem:

$$P(x, A) = \mathbb{P}(\varphi(x) \in A) = 1_{\varphi^{-1}(A)}(x).$$

Punkt startowy  $X_0$  może zostać wylosowany z pewnego rozkładu tak więc zmienna  $X_n$  może mieć skomplikowany rozkład prawdopodobieństwa pomimo deterministycznej ewolucji.

**Przykład 5.** Optymalizacja stochastyczna. Niech  $(X, d)$  będzie przestrzenią metryczną. Niech  $Y_n$ ,  $n \in \mathbb{N}$ , będzie ciągiem i.i.d na standardowej prz. probabilistycznej  $([0, 1], \mathcal{B}([0, 1]), \mu)$  oraz

niech  $Q: X \times [0, 1] \rightarrow X$  będzie mierzalne. Niech  $f: X \rightarrow \mathbb{R}$  będzie mierzalną funkcją posiadającą minimum globalne. Definiujemy rekurencyjnie:

$$X_{n+1} = \begin{cases} X_n & \text{if } f(Q(X_n, Y_n)) \geq f(X_n) \\ Q(X_n, Y_n) & \text{if } f(Q(X_n, Y_n)) < f(X_n) \end{cases}$$

**Zadanie 1.** Jakie jądro przejścia posiada algorytm  $X_n$  z powyższego przykładu?

**Obserwacja 1.** Jądro przejścia  $P$  determinuje ewolucję jednowymiarowych rozkładów łańcucha  $\mathbb{X} = \{X_n\}_{n \in \mathbb{N}}$  zgodnie ze wzorem:

$$(1.4) \quad \mathbb{P}_{X_{n+1}}(C) = \int_X P(x, C) \mathbb{P}_{X_n}(dx), \quad C \in \mathcal{B}(X), \quad n \in \mathbb{N}.$$

**Definicja 4.** Jądro przejścia  $P$  determinuje jednoznacznie **operator przejścia (operator Markowa)**  $\hat{P}$  na miarach probabilistycznych

$$\hat{P}: M^1(X) \ni \mu \longrightarrow \hat{P}\mu \in M^1(X)$$

zgodnie ze wzorem

$$(1.5) \quad \hat{P}\mu(C) = \int_X P(x, C) \mu(dx)$$

Zgodnie z Definicją 4, rozkłady jednowymiarowe  $\mathbb{P}_{X_n}$  łańcucha  $\mathbb{X} = \{X_n\}_{n \in \mathbb{N}}$  przyjmują, zgodnie ze wzorem (1.4), następującą postać:

$$\mathbb{P}_{X_{n+1}} = \hat{P}\mathbb{P}_{X_n}.$$

**Obserwacja 2.** Mamy jednoznaczne odwzorowanie:

$$K(X) \ni P \longrightarrow \hat{P} \in \hat{K}(X),$$

gdzie  $\hat{K}(X) = \{\hat{P}: P \in K(X)\}$  to rodzina operatorów przejścia zadanych przez jądro przejścia. Mamy jednoznaczność, ponieważ powyższe odwzorowanie można odwrócić: dla danego  $\hat{P}$  jądro przejścia  $P$  jest wyznaczone jednoznacznie wzorem  $P(x, C) = \hat{P}\delta_x(C)$ , gdzie  $\delta_x$  to miara skupiona w punkcie  $x$ .

Zgodnie z Obserwacją 2 będziemy (zwyczajowo) nadużywać notacji i mówić o operatorze  $P$  zamiast o operatorze  $\hat{P}$  zadanym przez jądro  $P$  - przykładowo, równanie (1.4) zapiszemy jako

$$\mathbb{P}_{X_{n+1}} = P\mathbb{P}_{X_n}.$$

Jest to wygodna notacja, zgodnie z którą jądro przejścia  $P$  będziemy często traktować jako funkcję na miarach probabilistycznych (tzn. jako operator Markowa z Definicji 4). Zauważmy, że

$$\mathbb{P}_{X_n} = P^n \mathbb{P}_{X_0}.$$

gdzie  $P^n = P \circ P \circ \dots \circ P$  to  $n$ -krotne złożenie operatora  $P$  co, jak można pokazać, odpowiada poniższemu (rozwlekłemu) łańcuchowemu zapisowi

$$\begin{aligned} \mathbb{P}_{X_n}(C) &= \int_X P(x, C) \mathbb{P}_{X_{n-1}}(dx) = \int_X \int_X P(x, C) P(y, dx) \mathbb{P}_{X_{n-2}}(dy) = \dots \\ &= \int_X \int_X \dots \int_X P(x_{n-1}, C) P(x_{n-2}, dx_{n-1}) \dots P(x_0, dx_1) \mathbb{P}_{X_0}(dx_0). \end{aligned}$$

**Definicja 5.** Definiujemy indukcyjnie **jądro przejścia w  $n$  krokach**

$$P^0(x, C) = \delta_x(C), \quad P^1(x, C) = P(x, C), \quad P^{n+1}(x, C) = \int_X P(y, C)P^n(x, dy), \quad n \in \mathbb{N}.$$

**Obserwacja 3.** Jeśli operator  $\hat{P}$  jest zadany przez jądro przejścia  $P$ , wtedy  $n$ -krotne złożenie  $\hat{P} \circ \hat{P} \circ \dots \circ \hat{P}$  odpowiada jądro przejścia  $P^n$ .

**Obserwacja 4.** Jądro przejścia w  $n$  krokach odpowiada prawdopodobieństwu warunkowym:

$$P^n(x, C) = \mathbb{P}(X_{n+k} \in C | X_k = x), \quad n, k \in \mathbb{N}.$$

Dla przykładu, zgodnie z powyższymi uwagami, prawdopodobieństwo przejścia w dwóch krokach ze stanu  $x$  do zbioru  $C \in \mathcal{B}(X)$  wynosi

$$\mathbb{P}(X_{n+2} \in C | X_n = x) = P^2(x, C) = \int_X P(y, C)P(x, dy)$$

co możemy również wyrazić w języku operatorów przejścia:

$$\mathbb{P}(X_{n+2} \in C | X_n = x) = P^2 \delta_x(C).$$

W szczególności, jeśli łańcuch startuje z punktu  $x$  (tzn.  $X_0 = x$ ), mamy  $\mathbb{P}_{X_n} = P^n(x, \cdot)$ .

**Przykład 6.** Rozważmy jądro przejścia  $P = [P_{i,j}]_{i,j}$  na przestrzeni dyskretnej  $X = \{1, \dots, n\}$ . Rozkłady prawdopodobieństwa to wektory  $[p_1, \dots, p_n]^T$  a operator przejścia to macierz transponowana  $P^T$  tak więc mamy:

$$\mathbb{P}_{X_{n+1}} = P^T \mathbb{P}_{X_n} = (P^T)^n \mathbb{P}_{X_0} = (P^n)^T \mathbb{P}_{X_0}.$$

Widzimy, że jądro przejścia w  $n$ -krokach to  $P^n$  ( $n$ -ta iteracja macierzy  $P$ ).

W naszym kontekście słynne **równania Chapmana-Kołmogorowa** przybierają postać:

**Twierdzenie 1.** Dla dowolnych  $k, n \in \mathbb{N}$

$$P^{k+n}(x, C) = \int_X P^k(y, C)P^n(x, dy).$$

*Proof.* Na podstawie Obserwacji 3 oraz tożsamości  $P^n(x, C) = P^n \delta_x(C)$ :

$$P^{k+n}(x, C) = P^k P^n \delta_x(C) = \int_X P^k(y, C)P^n(x, dy).$$

□

**Przykład 7.** Jeśli  $X$  jest dyskretna (co najwyżej przeliczalna), równania Chapmana-Kołmogorowa przybierają (oczywistą) postać:

$$P^{k+n}(x, C) = \sum_{y \in X} P^k(y, C)P^n(x, y).$$

**Definicja 6.** Rozkładami skończenie wymiarowymi procesu  $\mathbb{X}$  nazywamy rodzinę rozkładów

$$\{\mathbb{P}_{(X_{t_1}, \dots, X_{t_n})} : n \in \mathbb{N}^+ \text{ oraz } t_1, \dots, t_n \in \mathbb{N}\},$$

gdzie  $\mathbb{P}_{(X_{t_1}, \dots, X_{t_n})}$  to rozkład wektora  $(X_{t_1}, \dots, X_{t_n}) : \Omega \rightarrow X^n$ .

**Propozycja 1.** *Jądro przejścia  $P$  oraz rozkład początkowy  $\mathbb{P}_{X_0}$  jednoznacznie definiują rozkłady skończenie wymiarowe łańcucha  $\mathbb{X}$  zgodnie z regułą ‘łańcuchową’:*

$$\begin{aligned} & \mathbb{P}(X_n \in A_n, X_{n-1} \in A_{n-1}, \dots, X_1 \in A_1, X_0 \in A_0) = \\ & = \int_{A_0} \int_{A_1} \dots \int_{A_{n-1}} P(x_{n-1}, A_n) P(x_{n-2}, dx_{n-1}) \dots P(x_0, dx_1) \mathbb{P}_{X_0}(dx_0). \end{aligned}$$

**Zadanie 2.** Wypisz regułę łańcuchową dla przypadku przestrzeni dyskretnej  $X$  (zamieniamy odpowiednie całki na sumy). Zaczynj od przypadków  $n = 1$  oraz  $n = 2$ .

**Uwaga na koniec rozdziału:** Własności takie jak Propozycja 1 (z której wynika wiele poprzednich stwierdzeń tego rozdziału) dowodzi się indukcyjnie, korzystając z własności tak zwanego regularnego prawdopodobieństwa warunkowego ( jest o tym pojęciu krótki rozdział w podręczniku Wstęp do Rachunku Prawdopodobieństwa ( Jakubowski, Sztencel) na koniec tematu warunkowej wartości oczekiwanej). Po udowodnieniu istnienia rekurencyjnej postaci łańcucha dla dowolnego jądra przejścia  $P$  na przestrzeni polskiej  $X$ , Propozycję 1 będziemy mogli udowodnić korzystając z twierdzenia Fubinięgo.

## 2. REKURENCYJNA POSTAĆ ŁAŃCUCHA.

Podamy teraz ogólną, rekurencyjną postać łańcucha Markowa. Niech  $(X, d)$ ,  $(B, d_B)$  będą przestrzeniami metrycznymi **ośrodkowymi** oraz niech  $Y_n: \Omega \rightarrow B$  będzie ciągiem i.i.d., niezależnym od zmiennej losowej  $X_0: \Omega \rightarrow X$ . Niech  $T: X \times B \rightarrow X$  będzie odwzorowaniem mierzalnym. Definiujemy

$$(2.1) \quad X_{t+1} = T(X_t, Y_t).$$

Ciąg  $\mathbb{X} = \{X_n\}_{n \in \mathbb{N}}$  jest łańcuchem Markowa o jądrze przejścia zdeterminowanym przez odwzorowanie  $T$  oraz rozkład  $\mathbb{P}_{Y_1}$ :

**Propozycja 2.** *Ciąg  $\mathbb{X} = \{X_n\}_{n \in \mathbb{N}}$  jest łańcuchem Markowa o jądrze przejścia*

$$P(x, C) = \mathbb{P}(T(x, Y_1) \in C) = \mathbb{P}_{Y_1}(\{y \in B: T(x, y) \in C\})$$

oraz operatorze przejścia

$$P\mu(C) = (\mu \times \mathbb{P}_{Y_1})(T^{-1}(C)) = \int_X \mathbb{P}(T(x, Y_1) \in C) \mu(dx), \quad \mu \in \mathcal{M}^1(X), C \in \mathcal{B}(X).$$

**Zadanie 3.** Udowodnij powyższą propozycję. Wskazówka: twierdzenie Fubinięgo.

**Przykład 8.** Jeśli  $Y_n$  jest ciągiem i.i.d. oraz  $X_n = Y_0 + \dots + Y_n$ , to mamy rekurencyjną postać:

$$X_{n+1} = X_n + Y_{n+1} =: T(X_n, \hat{Y}_n), \quad \text{gdzie } T(x, y) = x + y \text{ oraz } \hat{Y}_n = Y_{n+1}$$

Gdy mamy postać rekurencyjną łańcucha (2.1), wtedy dla dowolnego  $x \in X$  możemy formalnie zdefiniować łańcuch

$$X_0(x), X_1(x), X_2(x), \dots$$

jako jedyny ciąg spełniający

$$X_0(x) = x \text{ oraz } X_{n+1}(x) = T(X_n(x), Y_n).$$

Przypomnijmy, że  $P_1 \times P_2$  oznacza produkt miar probabilistycznych, który przy założeniu ośrodkowości odpowiednich przestrzeni metrycznych, jest jednoznacznie zdefiniowany warunkiem  $P_1 \times P_2(C \times D) = P_1(C) \cdot P_2(D)$ .

**Definicja 7.** (1) *Przestrzeń metryczną  $(X, d)$  będziemy nazywać **przestrzenią polską** jeśli jest ośrodkowa oraz zupełna.*

- (2) Przestrzeń topologiczną  $(X, \tau_X)$  nazywamy **przestrzenią polską** jeśli istnieje metryka  $d: X \times X \rightarrow \mathbb{R}^+$  generująca topologię  $\tau_X$  taka, że przestrzeń metryczna  $(X, d)$  jest ośrodkowa oraz zupełna.
- (3) Przestrzeń mierzalną  $(X, \mathcal{B}(X))$  zbiorów Borelowskich na przestrzeni polskiej  $(X, \tau_X)$  nazywamy **przestrzenią Borelowską**.
- (4) Odwzorowanie  $h: (X, \Sigma_X) \rightarrow (Y, \Sigma_Y)$  pomiędzy dwiema przestrzeniami Borelowskimi nazwiemy **Borelowskim izomorfizmem** jeśli  $h$  jest mierzalną bijekcją (oraz  $h^{-1}$  jest mierzalne - mierzalność funkcji odwrotnej jest automatycznie spełniona na mocy twierdzenia Souslina)

Poniżej spektakularne twierdzenie Kuratowskiego stwierdzające w szczególności, że dowolne dwie nieprzeliczalne przestrzenie Borelowskie są Borelowsko izomorficzne, czyli nieodróżnialne z punktu widzenia teorii miary.

**Twierdzenie 2** (Kuratowski). *Każda przestrzeń Borelowska  $(X, \Sigma_X)$  jest Borelowsko izomorficzna z jedną z następujących przestrzeni polskich: 1)  $[0, 1]$ , 2)  $\mathbb{N}$ , 3)  $\{a_1, \dots, a_n\}$ .*

Oczywiście, na zbiorze dyskretnym rozważamy zawsze topologię dyskretną (jest to przestrzeń polska, wystarczy rozważyć metrykę dyskretną).

**Dygresja teoretyczna.** Widzimy, że z twierdzenia Kuratowskiego wynika iż w klasie przestrzeni polskich spełniona jest hipoteza continuum - przestrzeń polska ma bijekcję ze zbiorem co najwyżej przeliczalnym lub z odcinkiem.

Na mocy poniższego twierdzenia, dla każdego jądra przejścia na przestrzeni polskiej  $X$  można skonstruować postać rekurencyjną łańcucha Markowa zgodnego z zadaniem jądrem przejścia.

**Twierdzenie 3.** *Niech  $P \in \mathcal{K}(X)$  będzie jądrem przejścia na przestrzeni polskiej  $X$  oraz niech  $\mu$  oznacza miarę Lebesgue'a na  $[0, 1]$  Istnieje odwzorowanie mierzalne  $T: X \times [0, 1] \rightarrow X$  spełniające*

$$P(x, C) = \mu(\{y \in [0, 1]: T(x, y) \in C\}).$$

*Wynika stąd że, jeśli  $Y_n: \Omega \rightarrow [0, 1]$ ,  $n \in \mathbb{N}$ , jest ciągiem i.i.d. o rozkładach  $\mathbb{P}_{Y_1} = \mu$ , niezależnym od  $X_0: \Omega \rightarrow X$ , wtedy proces zadany rekurencyjnie równaniem  $X_{n+1} = T(X_n, Y_n)$ ,  $n \in \mathbb{N}$ , jest łańcuchem Markowa o jądrze przejścia  $P$ .*

Ponieważ jądro przejścia determinuje jednoznacznie całą dynamikę łańcucha, można stwierdzić (mniej precyzyjnie), że powyższe twierdzenie gwarantuje istnienie regularnej rekurencyjnej reprezentacji (2.1) dla dowolnego łańcucha na przestrzeni polskiej  $X$ .

*Proof.* Rozważmy przypadek, gdy przestrzeń polska  $X$  jest nieprzeliczalna. Przypadek dyskretny ma bardzo prostą konstrukcję - bardzo proszę się zastanowić samodzielnie nad dowodem twierdzenia w przypadku dyskretnym, ma to sporą wartość dydaktyczną (w zasadzie przypadek dyskretny wynika również z poniższego Kroku 1, po odpowiednim utożsamieniu przestrzeni  $X$  z liczbami naturalnymi).

Na początek rozważymy sytuację  $(X, \Sigma_X) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  i przeprowadzimy bezpośrednią konstrukcję. W Kroku 2 skorzystamy z twierdzenia Kuratowskiego.

**Krok 1.** Na początek zauważymy, że jeśli miara probabilistyczna  $P \in \mathcal{M}^1(X)$  ma dystrybuantę  $F$  oraz zmienna losowa  $R$  ma rozkład jednostajny na  $(0, 1)$ , wtedy zmienna  $Z = F^{-1}(R)$  również ma dystrybuantę  $F$ , a zatem  $\mathbb{P}_Z = P$  ( $F^{-1}$  jest z pewną dokładnością uogólnieniem pojęcia funkcji odwrotnej do funkcji  $F$ , musimy to zdefiniować). Definiujemy zatem funkcję  $F^{-1}: (0, 1) \rightarrow \mathbb{R}$  wzorem:

$$F^{-1}(r) = \inf\{y \in \mathbb{R}: F(y) \geq r\}.$$

**Zadanie 4.** a) Wykonaj rysunek funkcji  $F^{-1}$  dla przypadku, gdy  $F$  jest dystrybuantą rozkładu dyskretnego. b) Zauważ, że gdy dystrybuanta  $F$  jest ściśle rosnącą i ciągłą, funkcja  $F^{-1}$  jest funkcją odwrotną do  $F: \mathbb{R} \rightarrow (0, 1)$ .

Kontynuujemy (przy założeniu  $\mathbb{P}_R = U(0,1)$  oraz  $Z = F^{-1}(R)$ ). Z definicji  $F^{-1}$  oraz z własności dystrybuanty wynika (ćwiczenie), że dla  $r \in (0,1)$ ,  $t \in \mathbb{R}$ :

$$F^{-1}(r) \leq t \iff r \leq F(t)$$

skąd faktycznie:

$$F_Z(t) = \mathbb{P}(F^{-1}(R) \leq t) = \mathbb{P}(R \leq F(t)) = F(t).$$

Teraz możemy już zdefiniować szukaną funkcję  $T: \mathbb{R} \times (0,1) \rightarrow \mathbb{R}$ . Dla  $x \in \mathbb{R}$  wiemy już, że funkcja

$$T_x(r) := \inf\{y \in \mathbb{R}: P(x, (-\infty, y]) \geq r\}$$

jest wspomnianą "uogólnioną odwrotnością" dystrybuanty rozkładu  $P(x, \cdot)$ , a zatem wystarczy postawić

$$T(x, r) = T_x(r), x \in \mathbb{R}, r \in (0,1).$$

Pozostaje do pokazania mierzalność funkcji  $T$  - zostawiam do rozpisania jako ćwiczenie sprawnościowe (korzystamy z własności jądra przejścia oraz prawostronnej ciągłości dystrybuanty).

**Krok 2.** Teraz niech  $X$  będzie dowolną nieprzeliczalną przestrzenią polską. Z twierdzenia Kuratowskiego wynika, że istnieje Borelowski izomorfizm  $H: X \rightarrow \mathbb{R}$ . Pozwala nam to, dla  $x \in X$ , przenieść miarę  $P(x, \cdot)$  z przestrzeni  $X$  na przestrzeń  $\mathbb{R}$ , definiując miarę probabilistyczną  $\tilde{P}(x, \cdot)$  na  $\mathbb{R}$  wzorem:

$$\tilde{P}(x, A) = P(x, H^{-1}(A)), A \in \mathcal{B}(\mathbb{R}).$$

Teraz analogicznie jak w Kroku 1 definiujemy odwzorowanie  $\tilde{T}_x(r) := \inf\{y \in \mathbb{R}: \tilde{P}(x, (-\infty, y]) \geq r\}$ . Następnie "wracamy" z przestrzeni  $\mathbb{R}$  do przestrzeni  $X$  dzięki odwzorowaniu  $H^{-1}: \mathbb{R} \rightarrow X$ , mianowicie kończymy konstrukcję kładąc:

$$T(x, y) = H^{-1}(\tilde{T}_x(y)), x \in X, y \in (0,1).$$

□

**Uwaga:** Twierdzenie 3 pozostanie prawdziwe, jeśli założenie, że przestrzeń  $X$  jest polska osłabimy do założenia ośrodkowości (startowe założenie naszego wykładu)- dowód tego faktu jest bardzo piękny ale skierowałby nas na inne (głębokie) wody. Nie zmienimy jednak celu naszej podróży, który ukaże się na horyzoncie już w Rozdziale 4.

### 3. MIARY ZNAKOZMIENNE I OPERATOR PRZEJŚCIA.

W pewnych sytuacjach przydatne będzie rozszerzenie operatora przejścia  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  na przestrzeń liniową (nad ciałem  $\mathbb{R}$ )  $\mathcal{M}(X)$  miar skończonych ze znakiem.

**Miarami skończonymi** będziemy nazywać rodzinę miar:

$$\mathcal{M}^+(X) = \{\alpha \cdot \mu: \alpha \geq 0, \mu \in \mathcal{M}^1(X)\},$$

natomiast **miarami skończonymi ze znakiem** będziemy nazywać:

$$\mathcal{M}(X) = \{\mu_1 - \mu_2: \mu_1, \mu_2 \in \mathcal{M}^+(X)\}.$$

Równoważnie, dla funkcji  $\mu: \mathcal{B}(X) \rightarrow \mathbb{R}$ ,

$$(3.1) \quad \mu \in \mathcal{M}(X) \iff \begin{cases} 1^\circ \mu(\emptyset) = 0, \\ 2^\circ \mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n), \text{ jeśli } A_i \cap A_j = \emptyset \text{ dla } i \neq j. \end{cases}$$

Tak więc widzimy, że miary skończone odpowiadają miarom probabilistycznym przemnożonym przez stałą natomiast miary ze znakiem (znakozmienne) to przeliczalnie addytywne funkcje rzeczywiste określone na zbiorach Borelowskich  $\mathcal{B}(X)$ , które zbiorowi pustemu przypisują zero.



Miara  $\mu \in \mathcal{M}(X)$  posiada jednoznaczny rozkład (**rozkład Jordana**) na część dodatnią i ujemną  $\mu^+ \in \mathcal{M}^+(X)$  oraz  $\mu^- \in \mathcal{M}^+(X)$  zgodnie ze wzorem

$$\mu = \mu^+ - \mu^-,$$

gdzie  $\mu^+, \mu^- \in \mathcal{M}^+(X)$  są miarami dodatnimi wzajemnie singularnymi, tzn. takimi, że istnieje rozkład przestrzeni  $X$  na rozłączne mierzalne zbiory  $X^+$  oraz  $X^-$ , takie, że  $\mu^+(X^+) = 1$  oraz  $\mu^-(X^-) = 1$  (zbiory  $X^+, X^-$  tworzą **rozkład Hahna** przestrzeni  $X$ ). Wzajemna singularność miar (równoważnie: prostopadłość miar) oznacza, że są one skupione na osobnych częściach przestrzeni  $X$ .

**Przykład 9.** : Niech  $f: \mathbb{R} \rightarrow \mathbb{R}$  będzie funkcją całkowalną względem miary Lebesgue'a. Możemy zdefiniować miarę znakozmienną  $\mu \in \mathcal{M}(\mathbb{R})$ :

$$\mu(C) = \int_C f(x)(dx), \quad C \in \mathcal{B}(\mathbb{R}).$$

Powyżej  $dx$  oznacza oczywiście całkowanie względem miary Lebesgue'a. Tutaj mamy

$$\mu^+(C) = \int_C f^+(dx) \quad \text{oraz} \quad \mu^-(C) = \int_C f^-(dx),$$

gdzie  $f^+$  oraz  $f^-$  to odpowiednio część dodatnia oraz część ujemna funkcji  $f$ . Rozkład Hahna możemy wziąć  $X^+ = \{x: f(x) \geq 0\}$  oraz  $X^- = \{x: f(x) < 0\}$ .

Przy ustalonym rozkładzie Hahna, dla dowolnego zbioru  $A \in \mathcal{B}(X)$  będziemy pisać:  $A^+ = A \cap X^+$  oraz  $A^- = A \cap X^-$ .

Rozkład Jordana pozwala dla  $\mu \in \mathcal{M}(X)$  szybko zdefiniować całkę z funkcji mierzalnej  $f: X \rightarrow \mathbb{R}$ , oznaczamy oczywiście jako  $\int_X f d\mu$ , równością

$$\int f d\mu = \int f d\mu^+ - \int f d\mu^-,$$

przy założeniu że  $f$  jest całkowalna względem miar  $\mu^+$  oraz  $\mu^-$ .

Rozkład Jordana pozwala również szybko zdefiniować **normę całkowitego wahania miary**:

$$\|\mu\| = \mu^+(X) + \mu^-(X), \quad \mu \in \mathcal{M}(X).$$

Zbieżność miar znakozmiennych w normie całkowitego wahania miary  $\|P_n - P\| \rightarrow 0$  będziemy oznaczać krótko  $P_n \rightarrow P$  i nazywać **zbieżnością silną**. Zauważmy od razu, że miary probabilistyczne to miary dodatnie mające normę równą 1, a zatem są one przecięciem miar dodatnich ze sferą jednostkową w przestrzeni unormowanej  $(\mathcal{M}(X), \|\cdot\|)$ .

Przestrzeń  $(\mathcal{M}(X), \|\cdot\|)$  jest przestrzenią Banacha nad ciałem  $\mathbb{R}$ , co można udowodnić w poniższym zadaniu.

**Zadanie 5.** Udowodnij, że: a)  $(\mathcal{M}(X), \|\cdot\|)$  jest prz. unormowaną, b) jeśli  $P_n$  jest ciągiem Cauchy'ego w  $\mathcal{M}(X)$ , to formuła  $P(A) = \lim_{n \rightarrow \infty} P_n(A)$  definiuje poprawnie miarę  $P$  będącą granicą ciągu  $P_n$ .

**Zadanie 6.** Pokaż, że jeśli miara  $\mu \in \mathcal{M}(X)$  nie jest miarą dodatnią ani miarą ujemną (tzn. miara  $-\mu$  również nie jest dodatnia), to NIE jest prawdą że część dodatnia miary  $\mu$  jest równa funkcji  $A \rightarrow \max\{\mu(A), 0\}$ , analogicznie część ujemna nie odpowiada funkcji  $-\min\{\mu, 0\}$ .

Poniższe twierdzenie mówi, że operator Markowa  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  jednoznacznie rozszerza się do dodatniego  $(P(\mathcal{M}^+(X)) \subset \mathcal{M}^+(X))$  oraz zachowującego normę  $(\|P\mu\| = \|\mu\|, \mu \in \mathcal{M}(X))$  odwzorowania liniowego  $P: \mathcal{M}(X) \rightarrow \mathcal{M}(X)$ . Umożliwia nam to traktować jądro przejścia  $P$  jako operator na miarach ze znakiem (gdy zajdzie taka potrzeba).

**Twierdzenie 4.** Operator  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$ :

(1) *spełnia*

$$P(\alpha\mu_1 + (1 - \alpha)\mu_2) = \alpha P\mu_1 + (1 - \alpha)P\mu_2, \quad \alpha \in [0, 1], \mu_1, \mu_2 \in \mathcal{M}^1(X),$$

(2) *jednoznacznie rozszerza się do dodatnio liniowego operatora  $P: M^+(X) \rightarrow M^+(X)$  na przestrzeni miar skończonych  $\mathcal{M}^+(X)$ :*

$$P(\alpha\mu_1 + \beta\mu_2) = \alpha P\mu_1 + \beta P\mu_2, \quad \mu_1, \mu_2 \in \mathcal{M}^+(X), \alpha, \beta \geq 0$$

(3) *jednoznacznie rozszerza się do liniowego operatora  $P: M(X) \rightarrow M(X)$ , zgodnie ze wzorami:*

$$P\mu = P\mu^+ - P\mu^-, \quad P(\alpha\mu_1 + \beta\mu_2) = \alpha P\mu_1 + \beta P\mu_2, \quad \alpha, \beta \in \mathbb{R}, \mu_1, \mu_2 \in \mathcal{M}(X).$$

*przy czym jest to operator dodatni oraz zachowujący normę:*

$$P(\mathcal{M}^+(X)) \subset \mathcal{M}^+(X) \text{ oraz } \|P\mu\| = \|\mu\|, \quad \mu \in \mathcal{M}(X).$$

Sformułowanie powyższego twierdzenia przybliża jak rozszerza się operator  $P$  na miary znakozmienne. Jest oczywiste, że: 3)  $\Rightarrow$  2)  $\Rightarrow$  1). Twierdzenie to dowodzi się przez dość prosty rachunek.

#### 4. MIARA STACJONARNA, NIEREDUKOWALNOŚĆ, NIEOKRESOWOŚĆ

Zaczynamy omawiać centralną tematykę teorii łańcuchów Markowa - asymptotyczne zachowanie łańcucha. Tematyka ta zazwyczaj sprowadza się do zagadnień związanych ze zbieżnością łańcucha do miary stacjonarnej. W całej sekcji  $\mathbb{X}$  jest łańcuchem Markowa na przestrzeni Borelowskiej  $(X, \mathcal{B}(X))$  o jądrze przejścia  $P \in \mathcal{K}(X)$ .

**Definicja 8.** *Powiemy, że miara  $\pi \in \mathcal{M}^1(X)$  jest miarą stacjonarną (niezmienniczą) łańcucha  $\mathbb{X}$ , jeśli*

$$(4.1) \quad \pi(C) = \int_X P(x, C)\pi(dx), \quad C \in \mathcal{B}(X).$$

Równoważnie,  $\pi \in \mathcal{M}^1(X)$  jest miarą stacjonarną gdy jest punktem stałym dla operatora przejścia  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$ , tzn.

$$P\pi = \pi.$$

Wynika z tego m.in., że łańcuch startujący z rozkładu stacjonarnego  $\pi$  ma wszystkie rozkłady jednowymiarowe równe rozkładowi stacjonarnemu  $\pi$  - łańcuch "pozostaje" w rozkładzie stacjonarnym. Mocniej: wynika z tego stacjonarność łańcucha, zgodnie z Obserwacją 5.

**Definicja 9.** *Powiemy, że proces stochastyczny (ciąg zmiennych losowych)  $Z_n$ ,  $n \in \mathbb{N}$ , jest procesem stacjonarnym, jeśli dla każdego  $k \in \mathbb{N}$  rozkład łączny wektora  $(Z_t, Z_{t+1}, \dots, Z_{t+k})$  nie zależy od  $t$ .*

**Obserwacja 5.** *Łańcuch  $\mathbb{X}$  jest procesem stacjonarnym wtedy i tylko wtedy gdy startuje z rozkładu stacjonarnego, tzn. gdy rozkład początkowy  $\mathbb{P}_{X_0}$  jest rozkładem stacjonarnym łańcucha.*

**Zadanie 7.** Udowodnij powyższą obserwację.

Zauważmy, że jak w przypadku większości pojęć w teorii łańcuchów Markowa, Definicja 8 zależy jedynie od jądra przejścia  $P$ , a nie od samego łańcucha  $\mathbb{X}$ . Poniżej rozszerzamy Definicję 8 na operator  $P$  na miarach skończonych ze znakiem.

**Definicja 10.** *Powiemy, że miara  $\pi \in \mathcal{M}(X)$  jest miarą stacjonarną (niezmienniczą) operatora  $P: \mathcal{M}(X) \rightarrow \mathcal{M}(X)$  jeśli*

$$(4.2) \quad P\pi = \pi \text{ oraz } \pi \neq 0.$$

Podstawowymi zagadnieniami związanymi z dynamiką łańcucha  $\mathbb{X}$  są istnienie i jedyność miary stacjonarnej  $\pi \in \mathcal{M}^1(X)$  oraz zbieżność łańcucha do miary stacjonarnej  $\pi$ . W tym kontekście będziemy rozważać najczęściej **normę całkowitego wahanía miary**  $\|\cdot\|$  w zawężeniu do miar probabilistycznych, która indukuje topologię silniejszą od topologii słabej zbieżności miar probabilistycznych. Przypomnijmy, że zbieżność w tej normie nazywamy silną i oznaczamy krótko  $P_n \rightarrow P$ . Żeby sobie uzmysłwić czym jest zbieżność silna, przypatrzmy się bliżej jej własnościom.

**Propozycja 3.** Dla  $\mu \in \mathcal{M}(X)$  zachodzi

- (1)  $\|\mu\| = \sup\{\sum_{A \in \mathcal{A}} |\mu(A)| : \mathcal{A} \text{ to mierzalne c.n. przeliczalne rozbitcie } X\}$
- (2)  $\|\mu\| = \sup\{|\mu(C)| + |\mu(X \setminus C)| : C \in \mathcal{B}(X)\}$

Z powyższej propozycji z podpunktu (2) można wywnioskować, że:

**Propozycja 4.** Dla dowolnych  $P_1, P_2 \in \mathcal{M}^1(X)$ ,

$$(4.3) \quad \|P_1 - P_2\| = 2 \sup_{C \in \mathcal{B}(X)} |P_1(C) - P_2(C)|$$

Teraz widzimy, że jeśli mamy zbieżność miar probabilistycznych  $P_n \rightarrow P$ , wtedy

$$\sup_{C \in \mathcal{B}(X)} |P_n(C) - P(C)| \rightarrow 0,$$

zatem silna zbieżność rozkładów  $P_n$  do miary  $P$  oznacza zbieżność jednostajną miar  $P_n$  do miary  $P$ . Oczywiście wynika stąd zbieżność punktowa:

$$P_n(A) \rightarrow P(A) \text{ dla każdego } A \in \mathcal{B}(X)$$

oraz zbieżność słaba:

$$P_n \rightarrow P \Rightarrow P_n \xrightarrow{w} P.$$

**Zadanie 8.** a) Korzystając z rozkładu Hahna-Jordana miary znakozmiennej  $\mu$ , udowodnij Propozycję 3. b) Pokaż, jak Propozycja 4 wynika z punktu (2) Propozycji 3.

Przypomnijmy, że dla miar probabilistycznych zbieżność słaba  $P_n \xrightarrow{w} P$  zachodzi, gdy spełniony jest jeden z następujących równoważnych warunków:

- (1)  $\int_X f dP_n \rightarrow \int_X f dP$  dla dowolnej  $f \in BC(X, \mathbb{R})$
- (2)  $P_n(A) \rightarrow P(A)$  dla dowolnego  $A \in \mathcal{B}(X)$  takiego, że  $P(\delta A) = 0$ ,

gdzie  $\delta A$  oznacza brzeg zbioru  $A$ .

**Zadanie 9.** Udowodnij, że gdy  $X$  jest dyskretna (c.n. przeliczalna) oraz  $P, P_1, P_2, \dots \in \mathcal{M}^1(X)$ , wtedy następujące warunki są równoważne:

- (1)  $P_n \rightarrow P$ ,
- (2)  $P_n \xrightarrow{w} P$ ,
- (3)  $P_n(x) \rightarrow P(x)$ ,  $x \in X$

**Powrót:** Wracamy teraz do centralnej tematyki związanej z teorią łańcuchów Markowa: istnienie oraz jedyność miary stacjonarnej, zbieżność do miary stacjonarnej. W przypadku ogólnym łańcuch Markowa może: nie mieć miary stacjonarnej, mieć więcej niż jedną miarę stacjonarną, mieć miarę stacjonarną ale nie wykazywać zbieżności do miary stacjonarnej (lub wykazywać zbieżność tylko z niektórymi punktami startowymi). By zrozumieć takie (skandaliczne?) zachowanie, przeanalizujemy (dokładnie) serię przykładów.

**Przykład 10.** Niech  $X = \{1, 2, 3\}$  oraz niech jądro przejścia  $P$  będzie zadane warunkami:  $P(1, 2) = P(2, 1) = \frac{1}{2}$  oraz  $P(3, 3) = 1$ . Mamy tutaj oczywiście więcej niż jedną miarę stacjonarną:  $\pi_1 = [\frac{1}{2}, \frac{1}{2}, 0]$  oraz  $\pi_2 = [0, 0, 1]$  to miary stacjonarne. Z liniowości operatora przejścia  $P$ , każda kombinacja wypukła tych miar  $\pi = \alpha\pi_1 + (1 - \alpha)\pi_2$ ,  $\alpha \in [0, 1]$ , jest również miarą stacjonarną. Dodatkowo,  $\mathbb{P}_{X_n} \rightarrow \pi_1$  jeśli  $\mathbb{P}_{X_0}(1, 2) = 1$  oraz  $\mathbb{P}_{X_n} = \pi_2$ ,  $n \in \mathbb{N}$ , jeśli  $X_0 = 3$ . W szczególności, jeśli łańcuch startuje z punktu 1 lub 2, zbiega do miary stacjonarnej  $\pi_1$ . Startując z punktu 3, pozostaje w tym punkcie (od razu znajduje się w stanie stacjonarnym  $\pi_2$ ). Jeśli rozkład  $X_0$  jest skupiony na całej przestrzeni  $X$ , wtedy łańcuch zbiega do miary stacjonarnej będącej kombinacją wypukłą (jaką dokładnie?) miar  $\pi_1$  oraz  $\pi_2$ .

Przyczyna istnienia różnych miar stacjonarnych jest następująca: zachowanie łańcucha można rozpatrywać osobno na zbiorach  $\{1, 2\}$  oraz  $\{3\}$ , ponieważ z każdego z tych zbiorów nie ma możliwości przejścia do drugiego zbioru.

Sytuacja z powyższego przykładu jest wyeliminowana dzięki bardzo ważnej koncepcji nieredukowalności łańcucha.

**Definicja 11.** (1) Powiemy, że łańcuch  $\mathbb{X}$  jest **nieredukowalny względem miary**  $\varphi \in \mathcal{M}^1(X)$  jeśli dla każdego  $A \in \mathcal{B}(X)$  spełniającego  $\varphi(A) > 0$  oraz dla każdego  $x \in X$  istnieje  $n > 0$  takie, że  $P^n(x, A) > 0$

(2) Powiemy, że  $\mathbb{X}$  jest **nieredukowalny** jeśli dla pewnej miary  $\varphi \in \mathcal{M}^1(A)$  łańcuch  $\mathbb{X}$  jest nieredukowalny względem  $\varphi$ .

Według powyższej definicji zbiory miary dodatniej  $\varphi$  są osiągalne z dodatnim prawdopodobieństwem z **każdym** stanu przestrzeni  $X$ . W szczególności oznacza to, że nie można rozbić przestrzeni na dwa rozłączne zbiory pochłaniające, zgodnie z późniejszym zadaniem.

**Zadanie 10.** Pokaż, że gdy łańcuch  $\mathbb{X}$  jest nieredukowalny, nie istnieją rozłączne zbiory  $A, B \in \mathcal{M}^1(X)$  pochłaniające, tzn. spełniające  $P(x, A) = 1$  dla  $x \in A$  oraz  $P(x, B) = 1$  dla  $x \in B$ .

**Twierdzenie 5.** Nieredukowalny łańcuch Markowa posiada co najwyżej jedną probabilistyczną miarę stacjonarną.

Zanim przejdziemy do dowodu twierdzenia, udowodnimy następujący lemat.

**Lemat 1.** Dla  $\pi \in \mathcal{M}(X)$ , jeśli  $P\pi = \pi$ , wtedy  $P\pi^+ = \pi^+$  oraz  $P\pi^- = \pi^-$ .

*Proof.* Niech  $\pi = \pi^+ - \pi^-$  będzie miarą stacjonarną oraz niech  $\{X^+, X^-\}$  będzie mierzalnym rozbiem przestrzeni  $X$  takim, że  $\pi^+(X^+) = 1$  oraz  $\pi^-(X^-) = 1$ . Dla ustalenia uwagi, niech  $\pi^+ \neq 0$ . Pokażemy, że  $\pi^+$  jest miarą stacjonarną. Niech  $B \in \mathcal{B}(X)$  oraz niech  $B^+ = B \cap X^+$ . Mamy:

$$\begin{aligned} P\pi^+(B) &\geq P\pi^+(B^+) = \int_X P(x, B^+)\pi^+(dx) = \int_{X^+} P(x, B^+)\pi(dx) \geq \int_X P(x, B^+)\pi(dx) = \\ &= \pi(B^+) = \pi^+(B^+) = \pi^+(B). \end{aligned}$$

Ponieważ  $P\pi^+(X) = \pi^+(X)$ , ostra nierówność  $P\pi^+(B) > \pi^+(B)$  doprowadziłaby do sprzeczności z nierównością  $P\pi^+(X \setminus B) \geq \pi^+(X \setminus B)$ . Zatem  $P\pi^+(B) = \pi^+(B)$ . Wobec dowolności  $B \in \mathcal{B}(X)$ ,  $\pi^+$  jest miarą stacjonarną.  $\square$

**Wniosek 1.** W dowodzie powyższego lematu, biorąc  $B = X^+$ , otrzymujemy

$$\int_{X^+} P(x, X^+)\pi(dx) = \int_X P(x, X^+)\pi(dx),$$

z czego wynika, że

$$P(x, X^+) = 0 \text{ dla } \pi^- - \text{ prawie każdego } x \in X^-.$$

*Proof.* Wykorzystujemy, że  $P\pi = \pi$  oraz  $P\pi^+ = \pi^+$ :

$$\int_{X^+} P(x, X^+) \pi(dx) = \int_X P(x, X^+) \pi^+(dx) = \pi^+(X^+) = \pi(X^+) = \int_X P(x, X^+) \pi(dx).$$

□

**Dowód Twierdzenia 5.** Dla dowodu nie wprost niech  $\pi_1, \pi_2 \in \mathcal{M}^1(X)$  będą dwiema różnymi miarami stacjonarnymi oraz niech  $\pi = \pi_1 - \pi_2$ . Wtedy  $\pi^+, \pi^- \in \mathcal{M}^+(X)$  są dwiema różnymi miarami stacjonarnymi (na mocy Lematu 1) skupionymi odpowiednio na zbiorach rozłącznych  $X^+, X^-$  będących rozkładem Hahna przestrzeni  $X$ . Niech łańcuch  $\mathbb{X}$  będzie  $\varphi$ -nieredukowalny - dla ustalenia uwagi niech  $\varphi(X^+) > 0$ . Z Wniosku 1,  $P(x, X^+) = 0$  dla  $\pi^-$ - prawie wszystkich  $x \in X^-$ . Prostą indukcją dostajemy, że dla dowolnego  $n \in \mathbb{N}$ ,  $P^n(x, X^+) = 0$  dla  $\pi^-$ - prawie wszystkich  $x \in X^-$ , co implikuje:  $\pi^-(\{x \in X^- : \forall n \in \mathbb{N} P^n(x, X^+) = 0\}) = \pi^-(X^-) > 0$ , a to jest sprzeczne z nieredukowalnością łańcucha względem miary  $\varphi$ . □

**Zadanie 11.** Pokaż, że jeśli nieredukowalny łańcuch Markowa posiada miarę stacjonarną  $\pi$ , to jest  $\pi$ -nieredukowalny.

**Uwaga:** Nieredukowalność łańcucha  $\mathbb{X}$  nie jest warunkiem wystarczającym istnienia miary stacjonarnej, co ilustruje poniższy przykład.

**Przykład 11.** Rozważmy błądzenie losowe  $X_n = \sum_{i=1}^n Y_i$ , gdzie ciąg  $Y_n$  jest i.i.d. o rozkładach  $P_{Y_1} = U(\{-1, 1\})$ . Łańcuch  $X_n$  nie ma miary stacjonarnej.

**Zadanie 12.** Pokaż, że powyższe błądzenie losowe jest nieredukowalne. Udowodnij nie-istnienie miary stacjonarnej w powyższym przykładzie.

Tutaj pojęcie charakterystyczne dla przestrzeni dyskretnych: powiemy, że dwa różne stany łańcucha  $x \in X$  oraz  $y \in X$  **komunikują się**, gdy łańcuch potrafi z dodatnim prawdopodobieństwem przejść ze stanu  $x$  do  $y$  oraz ze stanu  $y$  do  $x$

Rozważmy prosty przykład ilustrujący, że nawet w przypadku dyskretnym, z faktu, że łańcuch Markowa jest nieredukowalny, NIE wynika, że dowolne dwa stany się komunikują. Zachodzi oczywiście implikacja odwrotna: jeśli dowolne dwa stany łańcucha się komunikują, wtedy łańcuch jest nieredukowalny.

**Przykład 12.** Niech  $X = \{1, 2, 3\}$ . Załóżmy, że  $P(i, \cdot) = U(\{2, 3\})$  dla  $i = 1, 2, 3$ . Łańcuch jest nieredukowalny, natomiast żaden stan nie komunikuje się ze stanem  $i = 1$ .

Poniżej zadanie, które nie jest na temat, ale warte zrobienia.

**Zadanie 13.** Pokaż, że  $X_0, X_1, X_2, X_3, \dots$  (zmiennie losowe z powyższego przykładu) są niezależne.

Jak zobaczymy poniżej, sama nieredukowalność łańcucha jest niewystarczająca by istnienie miary stacjonarnej implikowało zbieżność łańcucha do miary stacjonarnej.

**Przykład 13.** Niech  $X = \{0, 1\}$  oraz  $P(0, 1) = P(1, 0) = 1$ . Tutaj mamy łańcuch nieredukowalny z jedyną miarą stacjonarną  $\pi = [\frac{1}{2}, \frac{1}{2}]$ . Łańcuch nie jest jednak zbieżny do stanu stacjonarnego jeśli rozkład początkowy nie jest równy mierze stacjonarnej. Dla przykładu, jeśli  $X_0 = 0$ , wtedy  $X_n = n \pmod{2}$ .

Aby uniknąć sytuacji z powyższego przykładu rozważa się bardzo ważne pojęcie okresowości łańcucha.

**Definicja 12.** Powiemy, że łańcuch  $\mathbb{X}$  o rozkładzie stacjonarnym  $\pi$  jest **okresowy** jeśli istnieje naturalne  $d > 1$  oraz mierzalne rozbitcie  $Z, X(1), \dots, X(d)$  przestrzeni  $X$  na podzbiory mierzalne

spełniające  $\pi(Z) = 0$ ,  $\pi(X(1)) > 0$  oraz

$$(4.4) \quad P(x, X(i+1)) = 1, \text{ dla wszystkich } i = 1, \dots, d \text{ oraz } x \in X(i),$$

gdzie  $d+1 := 1$  (dodajemy mod  $d$ ). Powiemy, że łańcuch jest **nieokresowy** jeśli nie jest okresowy.

Jeśli nieredukowalny łańcuch nie ma miary stacjonarnej, to powiemy że jest okresowy, jeśli dla pewnego  $d > 1$  istnieje mierzalne rozbitcie  $Z, X(1), \dots, X(d)$  zgodne z równaniem (4.4) oraz miara  $\varphi \in \mathcal{M}^1(X)$  taka, że łańcuch jest nieredukowalny względem  $\varphi$  oraz zachodzi  $\varphi(X(1)) > 0$ . Z tego łatwo wynika, że  $\varphi(Z) = 0$ .

Zauważmy, że gdy mamy powyższe rozbitcie okresowe oraz ustalimy  $x \in X(i)$ , wtedy ciąg startujący z  $x$  postaci  $X_{n,d}(x)$ ,  $n \in \mathbb{N}$ , jest nieredukowalnym łańcuchem Markowa na przestrzeni  $X(i)$ . Ogólniej, jeśli ustalimy  $x \in X(1)$ , wtedy ciąg  $X_{n,d+l}(x)$  jest nieredukowalnym łańcuchem na przestrzeni  $X(l+1)$ , dla  $l = 0, \dots, d-1$ .

**Okresem** łańcucha okresowego nazywamy największą stałą  $d \in \mathbb{N}$  dla której istnieje rozbitcie  $Z, X(1), \dots, X(d)$  występujące w Definicji 12. Zauważmy, że jest to taka stała, dla której łańcuch  $X_{n,d}(x)$  jest już nieredukowalnym nieokresowym łańcuchem Markowa w zawężeniu do przestrzeni  $X(i)$  dla dowolnych  $i \in \{1, \dots, d\}$  oraz  $x \in X(i)$ .

**Przykład 14.** Niech  $X = \{1, 2, 3\}$  oraz  $P(1, 2) = 1$ ,  $P(2, 3) = 1$ ,  $P(3, 2) = 1$ . Widzimy, że łańcuch jest nieredukowalny z miarą stacjonarną  $\pi = U(\{2, 3\})$ , okresowy o okresie  $d = 2$ , mamy rozbitcie przestrzeni na zbiory  $Z = \{1\}$ ,  $X(1) = \{2\}$ ,  $X(2) = \{3\}$ .

**Zadanie 14.** Pokaż, że błądzenie losowe z Przykładu 11 jest przykładem okresowego łańcucha - jaki ma okres oraz jak rozbija się przestrzeń  $X$ ?

**Zadanie 15.** Pokaż, że powyższa definicja okresu dla łańcucha okresowego jest poprawna (nie da się znaleźć dowolnie dużego  $d$  z definicji okresowości dla ustalonego łańcucha okresowego).

Jak żeglarze, którzy podczas długiej podróży przez nieznanne wody podziwiają majaczący się na horyzoncie nieznaną łódź, tak my teraz jesteśmy gotowi by podziwiać piękno i prostotę następującego twierdzenia o zbieżności łańcuchów Markowa.

**Twierdzenie 6.** Załóżmy, że łańcuch  $\mathbb{X}$  nieredukowalny i nieokresowy. Jeśli istnieje miara stacjonarna  $\pi \in \mathcal{M}^1(X)$ , wtedy  $P^n(x, \cdot) \rightarrow \pi$  w normie całkowitego wahania miary dla  $\pi$  prawie każdego punktu początkowego  $x \in X$ .

Przy bardzo ogólnych założeniach Twierdzenia 6 zbieżność do rozkładu stacjonarnego zachodzi z  $\pi$ -prawie każdego punktu startowego natomiast nie musi zachodzić dla wszystkich punktów startowych  $x \in X$ , co ilustruje poniższy przykład.

**Przykład 15.** Niech  $X = \mathbb{N}$  oraz  $P(1, \cdot) = U(\{1, 2\})$ ,  $P(2, \cdot) = U(\{1, 2\})$ , dla  $n \geq 3$ :  $P(n, 1) = \frac{1}{n^2}$  oraz  $P(n, n+1) = \frac{n^2-1}{n^2}$

**Zadanie 16.** Uzasadnij powyższy przykład, sprawdzając założenia Twierdzenia 6, znajdując miarę stacjonarną, pokazując brak zbieżności z punktów startowych  $3, 4, \dots$

**Przykład 16.** Łańcuch nieredukowalny i nieokresowy może nie mieć miary stacjonarnej. Rozważ błądzenie losowe  $X_n = Y_1 + \dots + Y_n$ , gdzie  $Y_i$  to ciąg i.i.d. o rozkładach  $U([-1, 1])$ .

**Zadanie 17.** Uzasadnij powyższy przykład.

Mocne prawo wielkich liczb sformułujemy dla łańcuchów w postaci rekurencyjnej (2.1), dla których mamy formalnie zdefiniowaną rodzinę łańcuchów  $\{\mathbb{X}(x)\}_{x \in X}$ , gdzie  $\mathbb{X}(x) = X_0(x), X_1(x), \dots$  to łańcuch startujący z  $x$ .

**Mocne prawo wielkich liczb.** Jeśli jądro przejścia łańcucha spełnia założenia Twierdzenia 6, wtedy łańcuch spełnia mocne prawo wielkich liczb, tzn. dla dowolnej funkcji mierzalnej

$f: X \rightarrow \mathbb{R}$  całkowalnej względem  $\pi$  zachodzi dla  $\pi$ - prawie każdego punktu startowego  $x_0 \in X$ :

$$\mathbb{P} \left( \frac{1}{n} \cdot \sum_{i=1}^n f(X_i(x_0)) \xrightarrow{n \rightarrow \infty} \int_X f(x) \pi(dx) \right) = 1.$$

Można zauważyć, że z powyższego wynika mocne prawo wielkich liczb znane nam z kursu rachunku prawdopodobieństwa dla niezależnych zmiennych losowych: jeśli  $Y_n$  jest ciągiem i.i.d. (a zatem łańcuchem Markowa o jądrze  $P(x, \cdot) = \mathbb{P}_{Y_1}$  oraz mierze stacjonarnej  $\pi = \mathbb{P}_{Y_1}$ ) oraz  $E[Y_1] < \infty$ , wtedy  $P(\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{n \rightarrow \infty} E[Y_1]) = 1$ .

**Łańcuchy okresowe.** Wróćmy teraz do sytuacji, gdy łańcuch jest nieredukowalny z miarą stacjonarną  $\pi$ , ale jest okresowy o okresie  $d$ . Analizę zachowania takiego łańcucha można, w dużej mierze, sprowadzić do analizy odpowiedniego łańcucha nieokresowego, korzystając z przedyskutowanego już faktu, że łańcuch o okresie  $d$  można rozbić na  $d$  podciągów będących już łańcuchami nieokresowymi na "swoich" przestrzeniach  $X(i)$ .

1. Przypomnijmy, że jeśli  $\mathbb{X}$  o jądrze  $P$  jest łańcuchem o okresie  $d$  oraz  $\underline{X} = \{X(1), \dots, X(d)\}$  jest rozbiem przestrzeni  $X$  z definicji łańcucha okresowego, wtedy łańcuch  $X_{n,d}$ ,  $n \in \mathbb{N}$ , którego jądro przejścia to oczywiście  $P^d$ , można rozpatrywać osobno na każdej z przestrzeni  $X(i)$ . Faktycznie, weźmy  $x \in X(i)$  - widać, że łańcuch

$$\{X_{nd}(x)\}_{n \in \mathbb{N}} = x, X_d(x), X_{2d}(x), \dots$$

omiya pozostałe zbiory z rozbitcia  $\underline{X}$  oraz, że mamy że  $P^d(z, X(i)) = 1$  dla każdego  $z \in X(i)$ . Oczywiście jest to łańcuch nieredukowalny na  $X(i)$  oraz, z definicji okresu  $d$  łańcucha  $\mathbb{X}$ , jest nieokresowy (przestrzeni  $X(i)$  nie można bardziej rozbić na fragmenty definiujące okresowość). Dodatkowo, ponieważ  $\pi$  jest miarą stacjonarną jądra przejścia  $P$ , jest ona również miarą stacjonarną jądra  $P^d$  (na przestrzeni  $X$ ). Gdy zawężamy się do przestrzeni  $X(i)$  dostajemy bezpośrednio z definicji miary stacjonarnej, że miara zawężona do przestrzeni  $X(i)$ , oznaczamy  $\pi|_{X(i)}$ , jest miarą stacjonarną zawężonego jądra  $P^d|_{X(i)}$ . W szczególności  $\pi^i := \frac{1}{\pi(X(i))} \cdot \pi|_{X(i)}$  jest probabilistyczną miarą stacjonarną łańcucha  $X_{nd}$  zawężonego do przestrzeni  $X(i)$ . Z Twierdzenia 6 dostajemy zatem, że dla dowolnego  $i \in \{1, \dots, d\}$  mamy zbieżność silną

$$X_{nd}(x) \longrightarrow \pi^i \text{ dla } \pi - \text{ prawie każdego } x \in X(i).$$

Zauważmy teraz, że jeśli  $\pi^i \in \mathcal{M}^1(X)$  jest probabilistyczną miarą stacjonarną dla przestrzeni  $X(i)$ , wtedy  $P\pi^i$  jest probabilistyczną miarą stacjonarną na przestrzeni  $X(i+1 \bmod d)$ . Stąd  $\pi^i = P^{i-1}\pi^1$ . Wynika stąd, że  $\frac{1}{d} \left( \sum_{i=0}^{d-1} \pi^i \right)$  jest probabilistyczną miarą stacjonarną jądra  $P$  na przestrzeni  $X$  a zatem, z jednoznaczności miary stacjonarnej łańcucha nieredukowalnego,  $\pi = \frac{1}{d} \left( \sum_{i=0}^{d-1} \pi^i \right)$ . Ta równość implikuje (ponieważ miary  $\pi^i$  są skupione na rozłącznych częściach  $X(i)$  przestrzeni  $X$ ), że,  $\pi(X(i)) = \frac{1}{d}$ . Otrzymujemy zatem  $\pi^i = d \cdot \pi|_{X(i)}$ . W szczególności, na mocy prawa wielkich liczb, dla dowolnego  $i \in 1, \dots, d$  oraz funkcji  $\pi$  całkowalnej  $f: X \rightarrow \mathbb{R}$ ,

$$(4.5) \quad \frac{1}{n} \left( \sum_{j=1}^n f(X_{j-d}(x_0)) \right) \longrightarrow \int_{X(i)} f d\pi^i = d \cdot \int_{X(i)} f d\pi$$

dla  $\pi$ - prawie każdego  $x_0 \in X(i)$ . Mamy również,

$$\sum_{i=1}^{n \cdot d - 1} f(X_k(x_0)) = \sum_{i=1}^{n-1} f(X_{id}(x_0)) + \sum_{i=1}^{n-1} f(X_{id+1}(x_0)) + \dots + \sum_{i=1}^{n-1} f(X_{id+d-1}(x_0)),$$

gdzie powyższe sumy odpowiadają łańuchom  $\{X_{id+l}\}_{i \in \mathbb{N}}$ , które jak już zauważyliśmy spełniają mocne prawo wielkich liczb na odpowiadających sobie przestrzeniach  $X(i)$  (ponieważ są już łańuchami niereukowalnymi i nieokresowymi).

Mamy oczywiście

$$(4.6) \quad \int_X f d\pi = \sum_{l=1}^d \int_{X(l)} f d\pi = \frac{1}{d} \cdot \sum_{l=1}^d \int_{X(l)} f d\pi^l,$$

gdzie każda z całek jest granicą występującą w mocnym prawie wielkich liczb dla odpowiadającego jej łańucha  $X_{id+l}$ . Po sformalizowaniu ostatnich trzech obserwacji, dostaniemy:

**Nieredukowalny łańuch Markowa  $\mathbb{X}$  o okresie  $d$  również spełnia mocne prawo wielkich liczb:**

$$(4.7) \quad \mathbb{P} \left( \frac{1}{n} \cdot \sum_{i=1}^n f(X_i(x_0)) \xrightarrow{n \rightarrow \infty} \int_X f(x) \pi(dx) \right) = 1,$$

gdzie  $x_0$  pochodzi ze zbioru pełnej miary  $\pi$ .

**2.** Można również pokazać, że nieredukowalny okresowy łańuch  $\mathbb{X}$  z miarą stacjonarną  $\pi$  spełnia zbieżność silną do miary stacjonarnej po odpowiednim uśrednieniu, mianowicie zachodzi dla  $\pi$ - prawie każdego  $x$ :

$$\frac{1}{d} \cdot \sum_{i=n+1}^{n+d} P^i(x, \cdot) \longrightarrow \pi.$$

Powyższą równość będzie można pokazać dość łatwo, gdy poznamy trochę więcej własności normy całkowitego wahania miary.

## 5. OPERATORY FELLEROWSKIE, FUNKCJE LAPUNOWA I ISTNIENIE MIARY STACJONARNEJ.

Podamy teraz przykłady sytuacji, w których istnienie miary stacjonarnej dowodzimy metodami topologicznymi - głównym wynikiem tego rozdziału jest Twierdzenie 16. Metody probabilistyczne będziemy stosować w przyszłości. Przy podejściu topologicznym (wykorzystującym takie pojęcia jak ciągłość lub zwartość), zazwyczaj będziemy pracować z topologią słabej zbieżności miar. Zanim rozpoczniemy, przytoczymy podstawowe fakty o słabej zbieżności miar. Przypomnijmy, że  $(X, d)$  jest przestrzenią metryczną polską. Ciąg  $\mu_n \in \mathcal{M}^1(X)$  jest zbieżny słabo do miary  $\mu \in \mathcal{M}^1(X)$ , piszemy  $\mu_n \xrightarrow{w} \mu$ , gdy dla dowolnej funkcji  $f \in BC(X, \mathbb{R})$  zachodzi  $\int_X f(x) \mu_n(dx) \rightarrow \int_X f \mu(dx)$ .

Przypomnijmy

$$BC(X, \mathbb{R}) = \{f: X \rightarrow \mathbb{R} \mid f - \text{ciągła i ograniczona}\}.$$

Zachodzi następujące twierdzenie.

**Twierdzenie 7.** *Ponieważ  $X$  jest przestrzenią polską, przestrzeń  $\mathcal{M}^1(X)$  wraz z topologią słabej zbieżności jest przestrzenią polską. Dodatkowo, jeśli  $X$  jest przestrzenią zwartą, wtedy  $\mathcal{M}^1(X)$  jest przestrzenią zwartą.*

Jak widzimy w Twierdzeniu 7, przestrzeń miar probabilistycznych na  $X$  jest metryzowalna (istnieje metryka zupełna). Zdefiniujmy na początek normę Fortet-Mouriera na  $\mathcal{M}(X)$ :

$$\|\mu_1 - \mu_2\|_F = \sup \left\{ \left| \int_X f d\mu_1 - \int_X f d\mu_2 \right| : |f(x) - f(y)| \leq d(x, y) \text{ oraz } f: X \rightarrow [-1, 1] \right\},$$



gdzie supremum jest brane po wszystkich funkcjach mierzalnych ze stałą Lipschitza 1 o wartościach w  $[-1, 1]$ . Powyższa norma, w zawężeniu do miar probabilistycznych, indukuje wspomnianą w Twierdzeniu 7 metrykę zupełną, o czym mówi kolejne twierdzenie.

W podpunkcie (2) poniższego twierdzenia  $\|\cdot\|_F$  oznacza normę Forter-Mouriera zawężoną do miar skończonych dodatnich natomiast w podpunkcie (3) oznacza metrykę Fortet-Mouriera, tj. normę Forter-Moriera zawężoną do miar probabilistycznych.

**Twierdzenie 8.** *Przypomnijmy, że  $(X, d)$  jest prz. metryczną polską.*

- (1) *Przestrzeń  $(\mathcal{M}(X), \|\cdot\|_F)$  jest przestrzenią unormowaną.*
- (2) *Przestrzeń metryczna  $(\mathcal{M}^+(X), \|\cdot\|_F)$  jest zupełna oraz posiada topologię słabej zbieżności:*

$$\|\mu_n - \mu\|_F \rightarrow 0 \Leftrightarrow \forall f \in BC(X, \mathbb{R}) \quad \int_X f d\mu_n \rightarrow \int_X f d\mu$$

- (3) *Przestrzeń metryczna  $(\mathcal{M}^1(X), \|\cdot\|_F)$  jest zupełna oraz posiada topologię słabej zbieżności*

Poniżej przytaczamy twierdzenie Aleksandrowa dające przykładowe warunki równoważne słabej zbieżności.

**Twierdzenie 9.** *Niech  $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}^1(X)$ , gdzie  $X$  jest przestrzenią polską. Następujące warunki są równoważne:*

- (1)  *$\mu_n \rightarrow \mu$  słabo*
- (2)  *$\limsup_{n \rightarrow \infty} \mu_n(C) \leq \mu(C)$  dla każdego zbioru domkniętego  $C \subset X$*
- (3)  *$\liminf_{n \rightarrow \infty} \mu_n(U) \geq \mu(U)$  dla każdego zbioru otwartego  $U \subset X$*
- (4)  *$\lim_{n \rightarrow \infty} \mu_n(Z) = \mu(Z)$  dla każdego zbioru  $Z \in \mathcal{B}(X)$   $\mu$ -ciągłego (tj.  $\mu(\delta Z) = 0$ )*

Przytoczmy jeszcze jedną metrykę, która metryzuje topologię słabej zbieżności miar probabilistycznych. Definiujemy metrykę Prochorowa, zwaną również metryką Levy'ego-Prochorowa:

$$d_P(\mu_1, \mu_2) = \inf\{\varepsilon > 0: \mu_1(Z) \leq \mu_2(Z(\varepsilon)) + \varepsilon \text{ dla każdego } Z \in \mathcal{B}(X)\},$$

gdzie  $Z(\varepsilon) := \bigcup_{z \in Z} K(z, \varepsilon)$  oznacza  $\varepsilon$ -otoczkę zbioru  $Z$ . Zgodnie z poniższym twierdzeniem, jest to również metryka polska na przestrzeni  $\mathcal{M}^1(X)$ .

**Twierdzenie 10.** *Niech  $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}^1(X)$  oraz niech  $X$  będzie przestrzenią polską. Następujące warunki są równoważne:*

- (1)  *$\mu_n \rightarrow \mu$  słabo*
- (2)  *$\|\mu_n - \mu\|_F \rightarrow 0$*
- (3)  *$d_P(\mu_n, \mu) \rightarrow 0$ .*

*Dodatkowo, przestrzeń  $(\mathcal{M}^1(X), d_P)$  jest przestrzenią metryczną polską.*

Przytoczymy kolejną ważną własność słabej zbieżności miar probabilistycznych - niech  $X_1$  oraz  $X_2$  będą przestrzeniami polskimi oraz niech  $X = X_1 \times X_2$  będzie przestrzenią topologiczną produktową (a zatem również przestrzenią polską! - wystarczy rozważyć na  $X$  metrykę produktową). Produkt miar  $\mu_1 \in \mathcal{M}(X_1)$  i  $\mu_2 \in \mathcal{M}(X_2)$  oznaczamy przez  $\mu_1 \times \mu_2$  - jest to jedyna miara probabilistyczna spełniająca

$$\mu_1 \times \mu_2(C \times D) = \mu_1(C) \times \mu_2(D), \quad C \in \mathcal{B}(X_1), D \in \mathcal{B}(X_2).$$

Na zakończenie przeglądu faktów dotyczących słabej zbieżności, przytoczmy ważne twierdzenie Prochorowa, dające charakteryzację względnej zwartości rodziny miar  $M \subset \mathcal{M}^1(X)$ . Powiemy, że rodzina miar  $M$  jest *ciasna* (można też spotkać: *jedrna*, w j. angielskim: *tight*), jeśli dla każdego  $\varepsilon > 0$  istnieje zbiór zwarty  $K$  dostatecznie duży, by dla każdej miary  $\mu \in M$  zachodziło  $\mu(K) > 1 - \varepsilon$ .

**Twierdzenie 11.** (Prochorow) *Jeśli rodzina miar  $M$  na przestrzeni polskiej  $X$  jest ciasna, wtedy jest relatywnie zwarta, tzn. każdy ciąg miar  $\mu_n \in M$  posiada podciąg zbieżny do pewnej miary  $\mu \in \mathcal{M}^1(X)$ .*

W tym miejscu kończymy przegląd własności słabej topologii miar i przechodzimy do ważnej kwestii istnienia miary stacjonarnej łańcucha Markowa.

**Definicja 13.** *Operator Markowa  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  jest operatorem Fellera, gdy jest on zadany przez jądro przejścia  $P(x, \cdot)$  takie, że:*

$$(5.1) \quad X \ni x \longrightarrow P(x, \cdot) \in \mathcal{M}^1(X) \quad \text{jest funkcją ciągłą,}$$

gdzie  $\mathcal{M}^1(X)$  rozważamy wraz z topologią słabej zbieżności.

W całym niniejszym rozdziale, jeśli nie jest powiedziane inaczej,  $\mathcal{M}^1(X)$  jest domyślnie wyposażona w topologię słabej zbieżności. Poniższa propozycja (dość łatwa do udowodnienia) stwierdza w szczególności, że ciągłość jądra przejścia  $x \rightarrow P(x, \cdot)$  gwarantuje ciągłość operatora Markowa  $\mu \rightarrow P\mu$ .

**Propozycja 5.** *Operator Markowa  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  jest operatorem Fellera, wtu gdy jest on odwzorowaniem ciągłym.*

Korzystając z następującej wersji twierdzenia Schaudera o punkcie stałym:

**Twierdzenie 12** (Schauder). *Jeśli  $\emptyset \neq K \subset Z$  jest zwartym wypukłym podzbiorem przestrzeni unormowanej  $Z$  oraz  $T: K \rightarrow K$  jest odwzorowaniem ciągłym, wtedy  $T$  posiada punkt stały (tj.  $T(x_0) = x_0$  dla pewnego  $x_0 \in K$ ),*

błyskawicznie otrzymamy istnienie miary stacjonarnej  $\pi \in \mathcal{M}^1(X)$  w sytuacji gdy  $P$  jest operatorem Fellerowskim a  $X$  jest prz. metryczną zwartą:

**Twierdzenie 13.** *Jeśli  $(X, d)$  jest przestrzenią metryczną zwartą oraz operator  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  jest Fellerowski, wtedy istnieje  $\pi \in \mathcal{M}^1(X)$ , tze.  $P\pi = \pi$ .*

*Proof.* Przy naszych założeniach, z twierdzeń 7 oraz 8 wynika, że  $\mathcal{M}^1(X)$  jest zwartym podzbiorem przestrzeni unormowanej  $(\mathcal{M}(X), \|\cdot\|_F)$ , natomiast z Propozycji 5 dostajemy, że operator Markowa  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  jest odwzorowaniem ciągłym. Ponieważ zbiór  $\mathcal{M}^1(X)$  jest wypukły (kombinacja wypukła miar probabilistycznych jest miarą probabilistyczną), twierdzenie Schaudera kończy dowód.  $\square$

**Wniosek 2.** *Z Twierdzenia 13 wynika w szczególności, że każdy łańcuch Markowa na skończonej przestrzeni stanów posiada miarę stacjonarną  $\pi \in \mathcal{M}^1(X)$ .*

**Przykład 17.** Niech  $X = [0, 1]$  oraz  $P(x, \cdot) = U([0, x])$ . Istnieje miara stacjonarna (jaka?) Dygresja: Niech  $X = (0, 1)$  oraz niech  $P(x, \cdot) = U((0, x))$ . Nie istnieje miara stacjonarna (dlaczego?)

**Przykład 18.** Niech  $X = [0, 1]$  oraz niech  $S: X \rightarrow X$  będzie ciągłe. Z Twierdzenia 13 dostajemy, że łańcuch  $X_{n+1} = S(X_n)$  ma miarę stacjonarną (ciągłość operatora Markowa zdefiniowanego przez odwzorowanie  $S$  pozostawiamy jako ćwiczenie).

Dla operatora Markowa  $P$  zdefiniujemy teraz następujący operator "średniej"  $A_n: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  (również będący operatorem Markowa):

$$A_n = \frac{1}{n} \cdot \sum_{i=0}^{n-1} P^i, \quad \text{gdzie } P^0 = Id.$$

Niech  $\mathbb{X}$  będzie 1. Markowa o operatorze przejścia  $P$  oraz  $\mu = \mathbb{P}_{X_0}$ . Poniżej bardzo klarowna interpretacja operatora średniej  $A_n$ :

$$(5.2) \quad A_n \mu(C) = E\left[\frac{1}{n} \sum_{i=0}^{n-1} 1_{[X_i \in C]}\right].$$

Widzimy, że jeśli łańcuch aktualnie ma rozkład  $\mu$ , wtedy operator  $A_n$  podaje oczekiwaną liczbę wizyt łańcucha w zbiorze  $C$  w najbliższych  $n$  krokach (uwzględniając krok bieżący). W poniższym lemacie posłużymy się normą  $\|\cdot\|$  całkowitego wahania miary, która zawężona do miar probabilistycznych, przyjmuje zawsze wartość mniejszą bądź równą 2.

**Lemat 2.** *Dla dowolnej  $\mu \in \mathcal{M}^1(X)$ ,  $\|A_n \mu - A_n P \mu\| \rightarrow 0$ .*

*Proof.* Ponieważ  $A_n \mu - A_n P \mu = \frac{1}{n} \cdot (\mu - P^n \mu)$ , mamy

$$\|A_n \mu - A_n P \mu\| = \frac{1}{n} \|\mu - P^n \mu\| \leq \frac{1}{2} (\|\mu\| + \|P \mu\|) \leq \frac{4}{n} \rightarrow 0.$$

□

Z powyższego lematu wynika następujący wniosek.

**Wniosek 3.** *Jeśli podciąg  $A_{n_k} \mu$  jest zbieżny (słabo lub silnie), wtedy podciąg  $A_{n_k} P \mu$  jest zbieżny do tej samej granicy.*

**Twierdzenie 14.** *Jeśli  $P$  jest Fellerowski oraz dla pewnego  $\mu \in \mathcal{M}^1(X)$  ciąg  $A_n \mu$  posiada punkt skupienia  $\pi \in \mathcal{M}^1(X)$  (w słabej topologii), wtedy  $\pi$  jest miarą stacjonarną.*

*Proof.* Jeśli  $A_{n_k} \mu \xrightarrow{w} \pi$ , wtedy z ciągłości  $PA_{n_k} \mu \xrightarrow{w} P\pi$ , ale z liniowości  $P$  oraz Wniosku 3,  $PA_{n_k} \mu = A_{n_k} P \mu \xrightarrow{w} \pi$  skąd, z jednoznaczności granicy (słabej),  $P\pi = \pi$ . □

**Uwaga 1.** *Twierdzenie 14 implikuje Twierdzenie 13.*

**Uwaga 2.** *Ponieważ zawsze zachodzi implikacja  $P_n \mu \rightarrow \pi \Rightarrow A_n \mu \rightarrow \pi$ , widzimy, zgodnie z równaniem (5.2), że łańcuch, który zbiega silnie do miary stacjonarnej odwiedza dowolny zbiór  $C \in \mathcal{B}(X)$  z oczekiwaną częstością odpowiadającą mierze stacjonarnej zbioru  $C$ .*

Oczywiście bezpośrednia weryfikacja istnienia punktów skupienia ciągu  $A_n \mu$  może być trudna, przydatnym narzędziem będą funkcje Lapunowa  $V: X \rightarrow [0, \infty]$ .

Zanim przejdziemy dalej, zauważmy, że dla dowolnej mierzalnej funkcji  $V: X \rightarrow [0, \infty]$ ,

$$\int V(y) P(x, dy) = E_{P(x, \cdot)}[V] = EV[X_1 | X_0 = x],$$

gdzie przez  $E_{P(x, \cdot)}[V]$  oznaczamy wartość oczekiwaną z funkcji  $V$  względem miary  $P(x, \cdot)$ . Do tej pory jądro przejścia  $P$  często traktowaliśmy jako operator na miarach probabilistycznych  $P: \mathcal{M}^1(X) \rightarrow \mathcal{M}^1(X)$  lub ogólniej jako operator na miarach znakozmiennych  $P: \mathcal{M}(X) \rightarrow \mathcal{M}(X)$ . Od teraz, będziemy go również, gdy zajdzie potrzeba, traktować jako operator na zbiorze funkcji mierzalnych  $V: X \rightarrow [0, \infty]$ , które oznaczamy  $M^+(X, [0, \infty])$ .

**Definicja 14.** *Dla dowolnej mierzalnej funkcji  $V: X \rightarrow [0, \infty]$  będziemy pisać*

$$PV(x) := \int V(y) P(x, dy) = \int_X V(y) P_{\delta_x}(dy) = EV[X_1 | X_0 = x],$$

*definiując tym samym operator  $P: M^+(X, [0, \infty]) \ni V \rightarrow PV \in M^+(X, [0, \infty])$ .*

Łatwo pokazać, że odwzorowanie  $P: M^+(X, [0, \infty]) \ni V \rightarrow PV \in M^+(X, [0, \infty])$  jest poprawnie określone i dodatnio liniowe:

$$P(\alpha h_1 + \beta h_2) = \alpha h_1 + \beta h_2, \text{ gdzie } \alpha_1, \alpha_2 \in \mathbb{R}^+, h_1, h_2 \in M^+(X, [0, \infty]).$$

Jest również monotoniczne w następującym sensie:

$$\text{Jeśli } h_1 \leq h_2, \text{ wtedy } Ph_1 \leq Ph_2,$$

gdzie nierówność  $f \leq g$  oznacza  $f(x) \leq g(x)$  dla dowolnego  $x \in X$ .

Możemy teraz udowodnić poniższe twierdzenie, które daje warunki wystarczające istnienia miary stacjonarnej. Gdy korzystamy z Twierdzenia 15, uzasadnienie istnienia miary stacjonarnej dla konkretnego łańcucha sprowadza się do znalezienia funkcji Lapunowa  $V$  spełniającej warunki określone w założeniach Twierdzenia.

**Twierdzenie 15.** *Niech  $P$  będzie fellerowski oraz niech  $V: X \rightarrow [0, \infty]$  będzie mierzalna,  $V(x_0) < \infty$  dla pewnego  $x_0 \in X$ , taka, że*

$$(5.3) \quad PV(x) \leq V(x) - f(x) + b,$$

gdzie  $b \in \mathbb{R}$  jest pewną stałą oraz  $f: X \rightarrow [1, \infty)$  jest pewną funkcją mierzalną taką, że zbiory podpoziomicowe  $\{x \in X: f(x) \leq c\}$  są zwarte. Przy tych założeniach istnieje miara stacjonarna  $\pi \in \mathcal{M}^1(X)$ .

*Proof.* Indukcyjnie pokazujemy, że dla  $n \in \mathbb{N}$ ,

$$(5.4) \quad P^n V(x) + \sum_{k=0}^{n-1} P^k f(x) \leq V(x) + n \cdot b.$$

Dla  $n = 1$  powyższa równość odpowiada równości (5.3) - pamiętajmy, że  $P^0 f = f$  oraz że  $P^{k+1} = P \circ P^k$ . Załóżmy teraz, że (5.4) jest spełniona dla pewnego  $n$  - obkładamy obie strony nierówności operatorem  $P$  i korzystamy z założenia (5.3), otrzymując:

$$P^{n+1}V(x) + \sum_{k=1}^n P^k f(x) \leq PV(x) + nb \leq V(x) - f(x) + (n+1)b,$$

co już dowodzi tezę indukcyjną (pamiętamy, że  $f(x) = P^0 f(x)$ ).

Mając już (5.4), jako, że  $\sum_{k=0}^{n-1} P^k = n \cdot A_n$  oraz  $V \geq 0$ , dostajemy

$$A_n f(x_0) \leq \frac{V(x_0)}{n} + b \leq V(x_0) + b, \quad n \in \mathbb{N}.$$

Stąd, oraz z nierówności Czebyszewa,

$$A_n \delta_{x_0}(\{f \geq c\}) \leq \frac{E_{A_n(x_0, \cdot)}[f]}{c} \leq \frac{V(x_0) + b}{c}.$$

Powyższa nierówność dowodzi ciasności rodziny miar  $\{A_n \delta_{x_0}\}_{n \in \mathbb{N}}$  (przypomnijmy, że  $f$  ma zwarte zbiory podpoziomicowe), zatem na mocy Twierdzenia Prochorowa, ciąg  $A_n \delta_{x_0}$  posiada podciąg słabo zbieżny do pewnej miary  $\pi \in \mathcal{M}^1(X)$ . Twierdzenie 14 kończy dowód.  $\square$

Założenia Twierdzenia 15 nie są zbyt komfortowe do weryfikacji, niemniej jednak jest ono użyteczne, jak dowodzi poniższy przykład.

**Przykład 19.** Rozważmy “błądzenie losowe” po przestrzeni  $X = \mathbb{Z}$  o jądrze przejścia:

$$P(x, x+1) = \frac{1}{3}, P(x, x-1) = \frac{2}{3} \text{ dla } x > 0$$

$$P(x, x-1) = \frac{1}{3}, P(x, x+1) = \frac{2}{3} \text{ dla } x < 0$$

$$P(0, \cdot) = U(\{-1, 1\}).$$

Rozważając:  $V(x) = (\frac{3}{2})^x$ ,  $f(x) = \max\{x, 1\}$  oraz  $b = 100$ , uzyskujemy istnienie miary stacjonarnej na mocy Twierdzenia 15.

**Zadanie 18.** Uzasadnij powyższy przykład (sprawdzając założenia Twierdzenia 15 dla podanych  $V, f, b$  - mam nadzieję że się nie pomyliłem- lub znajdując własne funkcje  $V$  oraz  $f$  oraz stałą  $b$ )

**Zadanie 19.** Rozważmy błądzenie po przestrzeni  $X = \mathbb{N}$  oraz jądrze przejścia  $P(n, n+1) = \frac{1}{n}$  oraz  $P(n, n-1) = \frac{n-1}{n}$  dla  $n \neq 0$  oraz  $P(0, \cdot) = U(\{0, 1\})$ . Pokaż istnienie miary stacjonarnej korzystając z Twierdzenia 15.

**Zadanie 20.** Znajdź miarę stacjonarną dla ostatnich dwóch przykładów. Uwaga - nie sprawdzałem poziomu trudności tego zadania (jeśli okaże się trudniejsze, a jednak komuś się uda zrobić, można dostać więcej niż jeden plus).

Naszym celem będzie teraz udowodnienie twierdzenia ogólniejszego niż Twierdzenie 15:

**Twierdzenie 16.** Niech  $P$  będzie Fellerowskim operatorem przejścia na przestrzeni polskiej lokalnie zwartej  $(X, d)$ . Załóżmy, że istnieje funkcja mierzalna  $V: X \rightarrow [1, \infty]$ ,  $V(x_0) < \infty$  dla pewnego  $x_0 \in X$ , zbiór zwarty  $K \subset X$  oraz stała  $b > 0$  takie, że:

$$(5.5) \quad PV(x) \leq V(x) - 1 + b \cdot 1_K(x)$$

Wtedy  $P$  posiada niezmienniczą miarę  $\pi \in \mathcal{M}^1(X)$ .

Powyższe twierdzenie udowodnimy korzystając z \*-słabej zbieżności miar dodatnich - jest to pojęcie trochę słabsze od słabej zbieżności. Zdefiniujmy teraz funkcje ciągłe znikające w nieskończoności  $C_0(X) \subset CB(X, \mathbb{R})$ :

$$C_0(X) = \{f \in CB(X, \mathbb{R}) : \forall \varepsilon > 0 \exists X \supset K \text{ -zwarty } \forall x \in X \setminus K \ f(x) < \varepsilon\}.$$

W kontekście zbieżności \*-słabej będziemy zawsze zakładać, że przestrzeń polska  $X$  jest lokalnie zwarta. Powiemy, że ciąg  $\mu_n \in \mathcal{M}^+(X)$  jest zbieżny \*-słabo do  $\mu \in \mathcal{M}^+(X)$ , piszemy  $\mu_n \xrightarrow{w*} \mu$ , gdy

$$\int_X f d\mu_n \rightarrow \int_X f d\mu \text{ dla każdej } f \in C_0(X).$$

Widać wyraźnie, że  $\mu_n \xrightarrow{w} \mu \Rightarrow \mu_n \xrightarrow{w*} \mu$ . Należy jednak uważać - zbieżność \*-słaba w ogólnym przypadku nie ma tak dobrych własności jak zbieżność słaba.

**Uwaga 3.** Niech  $X = \mathbb{R}$  oraz  $\mu_n = \delta_n$ . Ciąg nie ma granicy słabej, natomiast ma granicę \*-słabą:  $\mu_n \xrightarrow{w*} 0$ . Zauważmy, że jeśli  $\mu_n \in \mathcal{M}^1(X)$ , wtedy granica słaba ciągu  $\mu_n$  również musi być miarą probabilistyczną. Jak widzimy, własność ta nie musi zachodzić dla zbieżności \*-słabej! (zero oznacza miarę trywialnie równą zero - nie można pomylić z miarą probabilistyczną  $\delta_0$ , która jest skupioną w zerze, ale nie jest zerowa)

**Twierdzenie 17.** Niech  $(X, d)$  będzie lokalnie zwartą prz. polską oraz  $\mu_n \in \mathcal{M}^+(X)$ .

- (1) Jeśli  $\mu_n \xrightarrow{w*} \mu$ , wtedy  $\limsup \mu_n(K) \leq \mu(K)$  dla dowolnego zbioru zwanego  $K$  oraz  $\int_X f d\mu \leq \liminf \int_X f d\mu_n$  dla dowolnej  $f \in CB(X, \mathbb{R})$ . W szczególności,  $\mu(X) \leq \liminf \mu_n(X)$ .
- (2) Jeśli  $\sup \mu_n(X) < \infty$ , wtedy  $\mu_n$  posiada podciąg \*-słabo zbieżny do pewnej  $\mu \in \mathcal{M}^+(X)$ .

Podamy teraz (bez dowodu - w dowodzie musielibyśmy pracować na zbieżności \*-słabej, a to pojęcie jest nam potrzebne jedynie tymczasowo) twierdzenie analogiczne do Twierdzenia 14, w

którym w miejsce zbieżności słabej rozważamy zbieżność \*-słabą. Należy zachować **ostrożność**: punkt skupienia w poniższym twierdzeniu nie musi być miarą probabilistyczną.

**Twierdzenie 18.** *Jeśli  $P$  jest Fellerowski oraz dla pewnego  $\mu \in \mathcal{M}^1(X)$  ciąg  $A_n\mu$  posiada punkt skupienia  $\pi \in \mathcal{M}^+(X)$  w \*-słabej topologii, wtedy  $P\pi = \pi$ .*

Jesteśmy teraz gotowi, by udowodnić główne Twierdzenie tego rozdziału - w dowodzie omi-  
jam rachunki. Osoby atakujące wysoki stopień końcowy zachęcam by uzupełnić dowód - robimy  
to podobnie jak w dowodzie Twierdzenia 15, korzystając z tych samych własności operatora  
 $P: M(X, [0, \infty]) \rightarrow M(X, [0, \infty])$ .

**Dowód Twierdzenia 16** Zdefiniujemy  $\pi_n^x = A_n\delta_x$ . Na mocy Twierdzenia 17 punkt (2), ciąg  
 $\pi_n^{x_0}$  posiada punkt skupienia  $\pi \in \mathcal{M}^+(X)$  w \*-słabej topologii. Jeśli  $\pi \neq 0$ , wtedy  $\frac{\pi}{\pi(X)}$  będzie  
szukaną miarą niezmienniczą. Korzystając z założeń Twierdzenia 16, indukcyjnie pokazujemy,  
że

$$V(x_0) \geq P^n V(x_0) + n - nb \cdot \pi_n^{x_0}(K).$$

Z powyższego dostajemy  $\limsup \pi_n^{x_0}(K) > 0$ , zatem z punktu (1) Twierdzenia 17:

$$\pi(K) \geq \limsup \pi_n^{x_0}(K) > 0,$$

a zatem  $\pi \neq 0$ . Dowód jest zakończony.  $\square$

**Zadanie 21.** Pokaż krótko, jak Twierdzenie 16 implikuje Twierdzenie 13.

Na tym kończymy (tymczasowo) analizę istnienia miary stacjonarnej. W przyszłości pokażemy  
jeszcze warunki wystarczające, przy których problem istnienia miary stacjonarnej sprowadza się  
do twierdzenia Banacha o punkcie stałym.

## 6. MOMENTY STOPU I MOCNA WŁASNOŚĆ MARKOWA

Przypomnijmy, że filtracją nazywamy rosnący ciąg  $\sigma$ -ciał  $F = \{F_n\}_{n \in \mathbb{N}}$  zawartych w  $\sigma$ -ciele  $\Sigma$   
natomiast naturalną filtracją procesu  $X_n$ ,  $n \in \mathbb{N}$ , nazywamy filtrację  $F^X$  zdefiniowaną poprzez  
 $F_n^X = \sigma(X_0, \dots, X_n)$ .

**Definicja 15.** *Powiemy, że zmienna losowa  $\tau: \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  jest momentem stopu względem  
filtracji  $\{F_n\}$  gdy dla każdego  $n \in \mathbb{N}$  zdarzenie  $\{\tau \leq n\}$  należy do sigma-ciała  $F_n$ . Gdy mamy  
ustalony proces  $X_n$ , będziemy mówić krótko, że  $\tau$  jest momentem stopu, gdy jest momentem stopu  
względem naturalnej filtracji tego procesu.*

**Definicja 16.** *Powiemy, że moment stopu  $\tau$  jest skończony, gdy  $\mathbb{P}(\tau < \infty) = 1$ .*

Ponieważ zbiór indeksów  $\mathbb{N}$  interpretujemy jako czas oraz sigma-algebrę  $F_n$  interpretujemy jako  
zbiór możliwych zdarzeń obserwowalnych do momentu  $n$ , zmienna losowa  $\tau: \Omega \rightarrow \mathbb{N} \cup \infty$  jest  
momentem stopu gdy w każdej chwili czasowej  $n$  jesteśmy w stanie określić czy moment  $\tau$  już  
nastąpił, tzn. czy pojawił się do momentu  $n$  (warunek:  $\{\tau \leq n\} \in F_n$ ) - zilustrujmy to prostymi  
przykładami.

**Przykład 20.** Niech  $\mathbb{X}$  oznacza symetryczne błądzenie losowe po  $\mathbb{Z}$ , startujące w zerze. Defini-  
ujemy

$$\tau = \inf\{n \in \mathbb{N}: X_n = 1\}.$$

Tutaj,  $\tau$  jest pierwszym momentem, w którym proces  $\mathbb{X}$  przyjmie wartość 1. Jest to moment  
stopu, ponieważ w każdej chwili  $n$  jesteśmy w stanie określić (przypomnijmy  $F = F^X$ ) czy proces  
 $\mathbb{X}$  odwiedził już "jedynek":

$$\{\tau \leq n\} = \{X_1 = 1 \vee X_2 = 1 \vee \dots \vee X_n = 1\} \in F_n^X.$$

Ponadto z własności symetrycznego błędzenia losowego wynika, że moment stopu  $\tau$  jest skończony.

Rozważmy teraz nawet prostszy przykład.

**Przykład 21.** Niech  $X_n$  oznacza liczbę orłów w pierwszych  $n$  rzutach symetryczną monetą oraz niech

$$\tau = \inf\{n \in \mathbb{N}: X_n = 10\}.$$

Zinterpretujemy rzut monetą jako udział w grze: wypadnięcie orła to zarobienie 1 zł podczas gdy wypadnięcie reszki to zarobek 0 zł. Moment stopu  $\tau$  to pierwsza chwila, w której zyskaliśmy 10 zł biorąc udział w grze- w każdej turze byliśmy w stanie stwierdzić, czy ten warunek już nastąpił. Tutaj mamy

$$\{\tau \leq n\} = \{X_n \geq 10\} \in F_n^X.$$

Bardzo łatwo uzasadnić, że moment stopu  $\tau$  jest skończony - w nieskończonym ciągu rzutów orzeł pojawi się co najmniej 10 razy z prawdopodobieństwem 1.

Momenty stopu odgrywają zasadniczą rolę w matematyce finansowej. Przykładowo, definiują moment realizacji posiadanej opcji lub wyjścia z gry - w każdej chwili jesteśmy w stanie określić, czy zostały spełnione warunki by zakończyć udział w grze. Jeśli proces  $X_n$  określa wypłatę w grze po  $n$  krokach a skończony moment  $\tau$  definiuje moment wyjścia z gry, wtedy nasza końcowa wypłata wynosi  $X_\tau$ , poniżej formalna definicja.

**Definicja 17.** Dla procesu  $X_n$  oraz skończonego momentu stopu  $\tau$  definiujemy zmienną losową  $X_\tau$ :

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega) = \sum_{n \in \mathbb{N}} X_n(\omega) \cdot 1_{\{\tau=n\}}(\omega).$$

W Przykładzie 21 mamy zawsze  $X_\tau = 10$  - w momencie  $\tau$ , który jest pierwszą chwilą gdy nasz zarobek wynosi 10 zł, nasz zarobek wynosi ... 10 zł. A ile wynosi  $X_\tau$  w Przykładzie 20 ?

Uogólniając poprzednie przykłady, definiujemy moment  $\tau_B = \tau_B^1$  pierwszej wizyty procesu w zbiorze  $B \in \mathcal{B}(X)$  oraz, rekurencyjnie, moment  $\tau_B^n$   $n$ - wizyty:

$$\tau_B = \inf\{n \in \mathbb{N}: X_n \in B\} \text{ oraz } \tau_B^{n+1} = \inf\{n \in \mathbb{N}: n > \tau_B^n \wedge X_n \in B\}.$$

Pokazuje się dość łatwo, że  $\tau_B^n$  to momenty stopu (niekoniecznie skończone).

Jak wiemy z poprzedniego rozdziału, łańcuch Markowa jest zdefiniowany poprzez własność Markowa: jeśli w bieżącej chwili  $n$  jesteśmy w stanie  $X_n = x$ , przyszła ewolucja łańcucha "zapomina" dotychczasową historię  $X_0, \dots, X_{n-1}$  i zależy jedynie od bieżącego stanu. Łańcuchy Markowa o czasie dyskretnym, którymi się zajmujemy, spełniają również silniejszą własność, zwaną *mocną własnością Markowa*: gdy  $\tau$  jest skończonym momentem stopu oraz znamy wartość łańcucha w momencie  $\tau$ , przyszła ewolucja łańcucha zależy jedynie od wartości  $X_\tau$ , "zapominając" dotychczasową historię  $X_0, X_1, \dots, X_{\tau-1}$ . By formalnie zdefiniować własność Markowa, musimy najpierw zdefiniować sigma-ciało zdarzeń obserwowalnych do momentu  $\tau$ .

**Definicja 18.** Dla momentu stopu  $\tau$  definiujemy  $\Sigma \supset F_\tau$  -sigma ciało zdarzeń obserwowalnych do momentu  $\tau$ :

$$F_\tau = \{A \in \Sigma: \forall n \in \mathbb{N} A \cap \{\tau \leq n\} \in F_n^X\}.$$

Zauważmy, że zmienna  $\tau = n$ , gdzie  $n \in \mathbb{N}$  jest ustalone, jest momentem stopu, oraz że w tym przypadku  $F_\tau = F_n^X$ , czyli zdarzenia obserwowalne do momentu  $\tau$  to zdarzenia z  $\sigma$ -ciała  $F_n^X = \sigma(X_0, \dots, X_n)$ .

**Twierdzenie 19.** Jeśli łańcuch  $X$  ma jądro przejścia  $P$  oraz  $\tau$  jest skończonym momentem stopu, wtedy:

$$\mathbb{P}(X_{\tau+1} \in A | \Sigma_\tau) = \mathbb{P}(X_{\tau+1} \in A | X_\tau), \quad A \in \mathcal{B}(X), \quad (\text{mocna własność Markowa}).$$

*Dodatkowo,*

$$(6.1) \quad \mathbb{P}(X_{\tau+1} \in A | X_{\tau}) = P(X_{\tau}, A), \quad A \in \mathcal{B}(X).$$

W dowodach często będziemy korzystać z bardzo naturalnej własności (6.1). Wracając do Przykładu 21: skoro w chwili  $\tau$  mamy stan konta  $X_{\tau}$  wynoszący 10 zł, wtedy po kolejnym rzucie z prawdopodobieństwem  $\frac{1}{2}$  będziemy mieć 11 zł a w przeciwnym wypadku pozostaniemy przy 10 zł, niezależnie od dotychczasowego przebiegu gry. Zgodnie z własnością (6.1), zapisujemy to formalnie:

$$\mathbb{P}(X_{\tau+1} \in \cdot | X_{\tau}) = P(X_{\tau}, \cdot) = P(10, \cdot) = U(\{10, 11\}).$$

Oczywiście analogiczną interpretację ma Przykład 20, ale tam z równym prawdopodobieństwem możemy przegrać tyle samo ile możemy wygrać, więc nie będziemy w to grać.

## 7. ŁAŃCUCHY NA DYSKRETNEJ PRZESTRZENI STANÓW.

Zostawiamy chwilowo dotychczasową ogólną sytuację i zakładamy w bieżącym rozdziale, że przestrzeń stanów jest skończona, tj.  $X = \{1, 2, \dots, d\}$ . Ten szczególny przypadek ma dużą wartość dydaktyczną ze względu na zwięzłość i klarowność dowodów. Przypomnijmy, że jądro przejścia jest zadane macierzą  $P(i, j) = p_{i,j}$ ,  $1 \leq i, j \leq n$ . Poniżej podstawowe definicje:

- Definicja 19.**
- (1) *Stan  $y$  jest osiągalny ze stanu  $x$ , jeśli dla pewnego  $n \in \mathbb{N}$  mamy  $P^n(x, y) > 0$ , gdzie  $P^0(x, x) := 1$ . Oznaczamy  $x \rightarrow y$*
  - (2) *Stany  $x, y$  komunikują się, gdy  $x$  jest osiągalny ze stanu  $y$  oraz  $y$  jest osiągalny z  $x$ . Oznaczamy  $x \leftrightarrow y$*
  - (3) *Łańcuch  $X$  jest nieprzywiedlny, gdy dowolne dwa stany się komunikują.*

Jest oczywiste, że łańcuch nieprzywiedlny jest nieredukowalny (aczkolwiek nie ma implikacji w drugą stronę).

**Zadanie 22.** Podaj przykład łańcucha nieredukowalnego o co najwyżej jednej mierze stacjonarnej, który nie jest nieprzywiedlny.

W przypadku skończonym łatwiej scharakteryzować m.in. okresowość łańcucha.

**Definicja 20.** *Okresem stanu  $x \in X$  nazywamy wielkość*

$$o(x) = \text{NWD}\{n \in \mathbb{N} : P^n(x, x) > 0\}.$$

*Stan  $x$  nazwiemy okresowym, gdy  $o(x) > 1$ . Gdy  $o(x) = 1$ , stan  $x$  nazwiemy nieokresowym.*

Pokazuje się, że w łańcuchu nieprzywiedlnym wszystkie stany mają ten sam okres, stąd poprawność poniższej definicji.

**Definicja 21.** *Powiemy, że nieprzywiedlny łańcuch  $\mathbb{X}$  jest okresowy o okresie  $d > 1$ , gdy wszystkie jego stany mają okres  $d$ . W przeciwnym wypadku łańcuch nazwiemy nieokresowym.*

W szczególności, łańcuch nieprzywiedlny będzie nieokresowy gdy dla pewnego  $x$  mamy  $P(x, x) > 0$ . Pokazuje się w przypadku skończonej przestrzeni stanów łańcuch nieprzywiedlny jest nieokresowy, gdy dla dowolnych stanów  $x, y > 0$  istnieje  $n_0 \in \mathbb{N}$  takie, że dla dowolnego  $n \geq n_0$  mamy  $P^n(x, y) > 0$ .

Rozpoczynamy warstwę dowodową dla bieżącego przypadku skończonej przestrzeni stanów - pokażemy postać miary stacjonarnej oraz udowodnimy prawo wielkich liczb. Dowody będą związane z ideą regeneracji łańcucha- wybieramy pewien punkt  $z$  z przestrzeni  $X$ . Będziemy korzystać z faktu, że za każdym razem gdy łańcuch odwiedzi punkt  $z$ , zapomina dotychczasową



historię (z mocnej własności Markowa). Dzięki temu, kolejne wędrówki łańcucha, tj. kawałki trajektorii łańcucha zaczynające się i kończące w punkcie  $z$ , są od siebie niezależne.

Zdefiniujmy moment stopu:

$$T_z = \inf\{n \in \mathbb{N}^+ : X_n = z\}.$$

$T_z$  jest czasem pierwszej wizyty łańcucha w stanie  $z$  (jeśli  $X_0 = z$  wtedy raczej powiemy o czasie pierwszego powrotu do stanu  $z$ ).

**Lemat 3.** *Jeśli łańcuch  $X$  jest nieprzywiedlny, wtedy dla pewnych  $c \in \mathbb{R}$  oraz  $\gamma < 1$*

$$\mathbb{P}[T_z > m] \leq c \cdot \gamma^m,$$

dla dowolnego rozkładu  $X_0$ . W szczególności,  $T_z$  jest skończony oraz spełnia  $E[T_z] < \infty$ .

*Proof. Krok 1.* Załóżmy, że łańcuch jest nieokresowy. Wtedy istnieją  $n \in \mathbb{N}$  oraz  $\delta > 0$  takie, że  $P^n(x, z) > \delta$ ,  $x \in X$ , - wynika to z faktu, że przestrzeń stanów  $X$  jest skończona i da się pokazać, korzystając z nieprzywiedlności i nieokresowości, że dla odpowiednio dużego  $n$  każdy stan będzie osiągalny z dowolnego innego stanu w  $n$  krokach (ćwiczenie na zajęcia). Z równań Chapmana-Kołmogorowa mamy, że  $\mathbb{P}[T_z > kn] < (1 - \delta)^k = \left((1 - \delta)^{\frac{1}{n}}\right)^{kn} =: \gamma^{kn}$ , skąd

$$\mathbb{P}[T_z > m] \leq \mathbb{P}[T_z > \text{floor}\left(\frac{m}{n}\right) \cdot n] \leq \gamma^{\text{floor}\left(\frac{m}{n}\right) \cdot n} \leq \gamma^{\frac{1}{n-1} \cdot m} =: c \cdot \gamma^m.$$

**Krok 2.** Dla przypadku okresowego dowód jest trochę trudniejszy, ale idea jest taka sama.  $\square$

Z powyższego Lematu mamy błyskawiczny wniosek, że dla dowolnego  $z \in X$ ,  $T_z$  jest momentem skończonym, co więcej  $E_x[T_z] < \infty$  dla dowolnego  $x \in X$ , gdzie

$$E_x[T_z] = E[T_z | X_0 = x].$$

Naturalnie, wartość  $E_x[T_z]$  to średni czas oczekiwana (mierzony liczbami kroków) na dotarcie łańcucha do punktu  $z$  przy założeniu, że startował on z punktu  $x$ . W Przykładzie 21 mamy  $E_0[T_1] = 2$  - średni czas oczekiwania na pierwsze zwycięstwo wynosi 2 ponieważ jest to wartość średnia w rozkładzie geometrycznym z parametrem  $p = \frac{1}{2}$ . W Przykładzie 20 można pokazać, że  $E_0[T_1] = \infty$  - chociaż w symetrycznym błędzeniu losowym odwiedzamy każdy punkt z prawdopodobieństwem 1, średni czas oczekiwania na pierwsze odwiedzin "jedyńki" (lub jakiegokolwiek innego punktu) jest nieskończony.

Ogólnie, dla dowolnej zmiennej losowej  $Y$  będziemy pisać

$$E_x[Y] = E[Y | X_0 = x].$$

Twierdzenie Kaca mówi nam, że średnia liczba odwiedzin stanu  $y$  przez nieredukowalny łańcuch pomiędzy dwiema kolejnymi wizytami w punkcie  $z$  jest wprost proporcjonalna do wartości  $\pi(y)$ , gdzie  $\pi$  to miara stacjonarna łańcucha. Wynika stąd równość (7.1) mówiąca, że miara  $\pi(y)$  jest odwrotnie proporcjonalna do średniego czasu oczekiwana na powrót łańcucha do stanu  $y$  po opuszczeniu stany  $y$ . Naturalnie: im większa miara stacjonarna punktu  $y$ , tym częściej ten punkt jest odwiedzany przez łańcuch. Żeby trochę przybliżyć założenia Twierdzenia Kaca, zauważmy, że dla ustalonych  $z$  oraz  $y \neq z$ :

$$\left( \sum_{i=0}^{T_z-1} 1_{\{X_i=y\}} \right) (\omega) = \sum_{i=0}^{T_z(\omega)-1} 1_{\{X_i=y\}}(\omega),$$

a zatem, przy założeniu że łańcuch wystartował z punktu  $z$ , wartość  $\sum_{i=0}^{T_z-1} 1_{\{X_i=y\}}$  podaje ile razy łańcuch odwiedził punkt  $y$  zanim powrócił do punktu  $z$ . Jeśli  $y = z$ , wtedy, przy założeniu że łańcuch startuje z punktu  $z$ , mamy  $\sum_{i=0}^{T_z-1} 1_{\{X_i=y\}} = 1$  bo łańcuch przebywał w punkcie  $z$  przed momentem  $T_z$  tylko w kroku 0.

**Twierdzenie 20.** (Kaca) Niech  $\mathbb{X}$  będzie nieredukowalny oraz niech  $z \in X$ . Miara  $\alpha \in M^+(X)$  zdefiniowana poprzez:

$$\alpha(y) = E_z\left[\sum_{i=0}^{T_z-1} 1_{\{X_i=y\}}\right] = E_z\left[\sum_{i=1}^{T_z} 1_{\{X_i=y\}}\right]$$

jest skończoną miarą stacjonarną łańcucha  $\mathbb{X}$ . W szczególności  $\pi = \frac{1}{\alpha(X)} \cdot \alpha$  jest jedyną probabilistyczną miarą stacjonarną oraz

$$(7.1) \quad \pi(y) = \frac{1}{E_y[T_y]}.$$

*Proof.* Niech  $z \in X$ . Aby uzasadnić skończoność, zauważmy że z addytywności całki  $E(\cdot)$ ,

$$\alpha(X) = \sum_{y \in X} \alpha(y) = E_z\left[\sum_{i=0}^{T_z-1} 1\right] = E_z[T_z] < \infty.$$

Powyżej skorzystaliśmy z Lematu 3. Pozostaje udowodnić stacjonarność. Dla uproszczenia notacji będziemy pisać  $E$  oraz  $T$  zamiast  $E_z$  oraz  $T_z$ . Zauważmy:

$$\alpha(x) = E\left[\sum_{i=0}^{T-1} 1_{\{X_i=x\}}\right] = E\left[\sum_{i=0}^{\infty} 1_{\{X_i=x, T>i\}}\right] = \sum_{i=0}^{\infty} \mathbb{P}(\{X_i = x, T > i\}).$$

Niech  $y \in X$ . Liczymy:

$$\sum_{x \in X} \alpha(x)P(x, y) = \sum_{x \in X} \sum_{i=0}^{\infty} \mathbb{P}(\{X_i = x, T > i\})P(x, y) = \sum_{i=0}^{\infty} \sum_{x \in X} \mathbb{P}(\{X_i = x, T > i\})P(x, y)$$

**Przypadek 1.** Załóżmy  $y \neq z$ . Mamy

$$\sum_{i=0}^{\infty} \sum_{x \in X} \mathbb{P}(\{X_i = x, T > i\})P(x, y) = \sum_{i=0}^{\infty} \mathbb{P}(\{X_{i+1} = y, T > i+1\}) = \alpha(y).$$

**Przypadek 2.** Załóżmy  $y = z$ . Mamy

$$\sum_{i=0}^{\infty} \sum_{x \in X} \mathbb{P}(\{X_i = x, T > i\})P(x, y) = \sum_{i=0}^{\infty} \mathbb{P}(\{X_{i+1} = y, T = i+1\}) = \sum_{i=0}^{\infty} \mathbb{P}(\{T = i+1\}) = 1 = \alpha(y).$$

Dowód stacjonarności jest zakończony. Jedyność miary stacjonarnej wynika z ogólnego Twierdzenia 5.  $\square$

Udowodnimy teraz mocne prawo wielkich liczb. Na początek zauważmy, że dla ustalonego  $z \in X$ , nasz łańcuch odwiedza punkt  $z \in X$  nieskończenie wiele razy. Można to argumentować na różne sposoby. Załóżmy dodatkowo, że łańcuch startuje z punktu  $z$ . Oznaczmy kolejne wizyty w punkcie  $z$

$$T_1 = T_z, \quad T_{n+1} = \inf\{n > T_n : X_n = z.\}$$

Z mocnej własności Markowa wynika, że ewolucja łańcucha po wizycie  $T_n$  nie zależy od wcześniejszej historii, w szczególności dostajemy:

$$E[T_1] = E[T_2 - T_1] = \dots = E[T_{n+1} - T_n],$$

skąd  $E[T_n] = n \cdot E[T_1] < \infty$ , zatem  $\mathbb{P}[T_n < \infty] = 1$ . W efekcie widzimy, że

$$\mathbb{P}[T_n < \infty \text{ dla każdego } n] = \mathbb{P}\left(\bigcap_n \{T_n < \infty\}\right) = 1.$$

Możemy zatem bez zmartwień rozpatrywać niezależne wycieczki łańcucha pomiędzy kolejnymi wizytami w punkcie  $z$  - wykorzystamy je w dowodzie mocnego prawa wielkich liczb. Przypomnijmy z twierdzenia Kaca  $\pi(z) = \frac{1}{E_z[T_z]}$ .

Przy bieżących oznaczeniach oczywiście  $X_{T_k} = z$ ,  $k \in \mathbb{N}$ . Ciąg  $T_i$  dzieli łańcuch  $\mathbb{X}$  (w sposób niedeterministyczny!) na kolejne wycieczki  $B_1, B_2, \dots$  startujące z punktu  $z$ :

$$B_1 = X_0, X_1, \dots, X_{T_1-1}$$

$$B_2 = X_{T_1}, X_{T_1+1}, \dots, X_{T_2-1},$$

które dzięki własności Markowa odbywają się niezależnie od siebie. Umożliwia to metodę dowodową opartą na "regeneracji" łańcucha w momencie trafienia w punkt  $z$ . Korzystając z tej metodologii, udowodnimy teraz poniższe mocne prawo wielkich liczb dla łańcuchów Markowa, zwane też **mocnym twierdzeniem ergodycznym**.

**Twierdzenie 21.** *Jeśli łańcuch  $\mathbb{X}$  jest nieprzywiedlny oraz  $f: X \rightarrow \mathbb{R}$ , wtedy*

$$\mathbb{P} \left[ \frac{\sum_{i=1}^n f(X_i)}{n} \rightarrow E_\pi[f] \text{ przy } n \rightarrow \infty \right] = 1,$$

gdzie  $\pi$  to miara stacjonarna łańcucha  $\mathbb{X}$ .

*Proof.* Dla ustalenia uwagi przyjmijmy, że  $f$  jest dodatnia oraz że łańcuch startuje z naszego ustalonego punktu  $z \in X$ . Na początek postawmy:

$$N(n) = \max\{k \in \mathbb{N} : T_k \leq n\}, n \in \mathbb{N},$$

gdzie  $T_0 := 0$  a zatem  $N(0) = 0$ .  $N(n)$  jest liczbą powrotów do punktu  $z$  do chwili  $n$  (nie liczymy chwili startowej) a  $T_{N(n)}$  jest chwilą ostatniej wizyty w  $z$  do momentu  $n$ . Mamy zatem

$$(7.2) \quad T_{N(n)} \leq n < T_{N(n)+1}.$$

Przypomnijmy, że  $m := E_z[T] < \infty$  oraz że  $T_k$  to suma długości  $k$  niezależnych wycieczek pomiędzy odwiedzinami punktu  $z$ , ( $T_k = T_1 + (T_2 - T_1) + \dots + (T_k - T_{k-1})$ ), stąd ze "zwykłego" mocnego prawa wielkich liczb  $\frac{T_k}{k} \rightarrow m$  oraz, ponieważ  $N(n) \rightarrow \infty$ ,  $\frac{T_{N(n)}}{N(n)} \rightarrow m$ . Z równania (7.2),

$$\frac{T_{N(n)}}{N(n)} \leq \frac{n}{N(n)} < \frac{T_{N(n)+1}}{N(n)},$$

a zatem na mocy twierdzenia o trzech ciągach,  $\frac{n}{N(n)} \rightarrow m$ . Stąd

$$(7.3) \quad \frac{N(n)}{n} \rightarrow \frac{1}{m}.$$

Analogiczne rozumowanie przeprowadzamy dla sum:

$$B_0(f) := \sum_{i=0}^{T_0} f(X_i), \quad B_{k+1}(f) := \sum_{i=T_k+1}^{T_{k+1}} f(X_i), \quad S_n(f) = \sum_{j=0}^n f(X_j).$$

Mamy:

$$\sum_{j=0}^n B_j(f) = \sum_{j=0}^{T_n} f(X_j) = S_{T_n}(f)$$

Dalej, z (7.2), oraz z naszego pomocniczego założenia  $f \geq 0$

$$\sum_{j=0}^{N(n)} B_j(f) = \sum_{j=0}^{T_{N(n)}} f(X_j) \leq S_n(f) \leq \sum_{j=0}^{T_{N(n)+1}} f(X_j) = \sum_{j=0}^{N(n)+1} B_j(f).$$

Z powyższego, analogicznie jak w poprzednim rozumowaniu, z prawa wielkich liczb dla niezależnych zmiennych  $B_j(f)$ , z faktu  $N_n \rightarrow \infty$  oraz z twierdzenia o trzech ciągach dostajemy, że  $\frac{S_n(f)}{N(n)} \rightarrow E[B_1(f)]$ . Stąd oraz z (7.3),

$$\frac{S_n(f)}{n} = \frac{S_n(f)}{N(n)} \cdot \frac{N(n)}{n} \rightarrow \frac{1}{m} \cdot E[B_1(f)].$$

Przypomnijmy, że  $m = E_z[T] = E_z[T_z]$  oraz  $\pi = \frac{\alpha}{m}$ . Twierdzenie Kaca zakończy dowód, ponieważ

$$E[B_1(f)] = E\left[\sum_{x \in X} f(x) \sum_{i=0}^{T_z-1} 1_{\{X_i=x\}}\right] = \sum_{x \in X} \alpha(x) f(x) = \sum_{x \in X} \pi(x) \cdot m f(x) = m \cdot E_\pi[f].$$

Dostajemy zatem, że łańcuch startujący z punktu  $z$  spełnia mocne prawo wielkich liczb. Punkt  $z$  był wybrany dowolnie, wobec tego dowód twierdzenia można uznać za zakończony.  $\square$

Zauważmy, że w dowodzie (ani w założeniach twierdzenia) nie potrzebujemy, by łańcuch był nieokresowy. Przy równie ogólnych założeniach prawdziwe jest **Centralne Twierdzenie Graniczne**, które dla dyskretnych łańcuchów Markowa można prosto sformułować:

**Twierdzenie 22** (Centralne Twierdzenie Graniczne). *Niech  $\mathbb{X}$  będzie łańcuchem nieprzywiedlnym o mierze stacjonarnej  $\pi$ . Zachodzi dla dowolnego rozkładu początkowego i dowolnej funkcji  $f: X \rightarrow \mathbb{R}$  następująca słaba zbieżność do rozkładu normalnego:*

$$\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n f(X_i) - E_\pi[f] \right| \rightarrow N(0, \sigma^2(f)),$$

gdzie wariancja  $\sigma^2(f)$  nie zależy od rozkładu początkowego i można ją zdefiniować jako:

$$\sigma^2(f) = \lim_{n \rightarrow \infty} \frac{1}{n} D^2\left(\frac{1}{n} \sum_{i=1}^n f(X_i)\right).$$

Powyższe twierdzenie również posiada dowody oparte na idei regeneracji ale pomijamy dowód - semestr powoli zbliża się już ku końcowi (a my nie jesteśmy dostatecznie szaleni by przekroczyć dopuszczalne stężenie dowodów w wykładzie). Nie przedstawiliśmy również dowodu Twierdzenia 6 w niniejszym specjalnym przypadku - ten jednak wyniknie błyskawicznie z dalszych, ogólniejszych rozważań.

Na zakończenie przytoczmy twierdzenie dotyczące klasyfikacji stanów dowolnego łańcucha, które elegancko formułuje się w przypadku skończonym (analogicznie jest w przypadku przeliczalnym) i wyjaśnia dlaczego teoria nieprzywiedlnych łańcuchów Markowa jest tak istotna dla przypadku dyskretnego.

**Twierdzenie 23.** *Dla dowolnego łańcucha przestrzeń stanów  $X$  rozbija się na sumę rozłącznych zbiorów  $T, X_1, X_2, \dots$ , gdzie  $T$  jest zbiorem stanów chwilowych, tj. dla dowolnego  $x \in T$  łańcuch startujący z  $x$  powróci do stanu  $x$  co najwyżej skończenie wiele razy z prawdopodobieństwem 1, natomiast każdy ze zbiorów  $X_i$  jest zbiorem absorbującym takim, że łańcuch zawężony do zbioru  $X_i$  jest łańcuchem nieprzewiedlnym.*

Przy oznaczeniach powyższego twierdzenia, łańcuch startujący z punktu  $x \in X_k$  może być analizowany jako łańcuch nieprzywyedlny na zbiorze  $X_k$ . W przypadku skończonym, zbiór  $T$  z prawdopodobieństwem 1 zostajnie opuszczony przez łańcuch, który trafi do któregoś ze zbiorów  $X_k$  i już tam pozostanie. W przypadku przeliczalnym zbiór  $T$  nie ma tak prostej charakteryzacji: łańcuch startujący z  $x \in T$  może trafić do pewnego zbioru  $X_k$  i już tam zostać lub może pozostać na zawsze w zbiorze  $T$ , gdzie każdy z punktów zostanie odwiedzony co najwyżej skończenie wiele razy.

## 8. SPRZĘGANIE ŁAŃCUCHÓW I ZBIĘŻNOŚĆ DO MIARY STACJONARNEJ. ZBIĘŻNOŚĆ GEOMETRYCZNA.

Wracamy do ogólnej teorii, w której zakładamy, że  $X$  jest przestrzenią topologiczną polską. Jest to już ostatni rozdział dotyczący klasycznej teorii łańcuchów Markowa (mam na myśli teorię nieredukowalnych łańcuchów opartą na metodach probabilistycznych). W niniejszym rozdziale zajmijmy się m.in. techniką sprzęgania łańcuchów, która jest w pewnym sensie przeniesiem idei regeneracji łańcucha z przypadku dyskretnego na przypadek ciągły.

Będziemy od tej pory używać następującej "wersji" normy całkowitego wahania miary

$$(8.1) \quad \|P_1 - P_2\| = \sup_{C \in \mathcal{B}(X)} |P_1(C) - P_2(C)|, \quad P_1, P_2 \in \mathcal{M}^1(X).$$

Z dokładnością do przemnożenia przez stałą 2 jest ona równa normie, którą znamy z poprzednich, ogólniejszych rozważań. Nie będziemy już pracować na miarach znakozmiennych dlatego, w zawężeniu do miar probabilistycznych, upraszczamy sobie notację zapominając o stałej 2.

Dla  $A \subset \mathbb{R}$  oznaczamy:

$$M(X, A) = \{f: X \rightarrow A \mid f - \text{mierzalna}\}.$$

Zacniemy od udowodnienia kilku własności:

**Propozycja 6.**  $\|P_1 - P_2\| = \sup_{f \in M(X, [0,1])} \left| \int_X f dP_1 - \int_X f dP_2 \right|.$

*Proof.* Niech  $P_0 = P_1 + P_2$ . Miary  $P_1$  i  $P_2$  są absolutnie ciągłe względem  $P_0$  a zatem mamy z twierdzenia Radona-Nikodyma:  $\int_X f dP_i = \int_X f \frac{dP_i}{dP_0} dP_0$ . Niech  $g_i = \frac{dP_i}{dP_0}$  oznacza pochodną Radona-Nikodyma. Mamy:  $\left| \int_X f dP_1 - \int_X f dP_2 \right| = \left| \int_X f \cdot (g_1 - g_2) dP_0 \right|$ . Ta wielkość przyjmuje wartość maksymalną gdy  $f = 1$  na zbiorze  $A := \{g_1 - g_2 > 0\}$  oraz  $f = 0$  na  $A'$ , a zatem dla  $f = 1_A$ . Mamy

$$\left| \int_X 1_A \cdot (g_1 - g_2) dP_0 \right| = |P_1(A) - P_2(A)|.$$

To kończy dowód, ponieważ dla dowolnego  $B \in \mathcal{B}(X)$ :

$$|P_1(B) - P_2(B)| = \left| \int_X 1_B dP_1 - \int_X 1_B dP_2 \right| \leq \sup_{f \in M(X, [0,1])} \left| \int_X f dP_1 - \int_X f dP_2 \right|$$

□

Z powyższego wyniku następujący wniosek:

**Wniosek 4.**  $\|P_1 - P_2\| = \frac{1}{b-a} \sup_{f \in M(X, [a,b])} \left| \int_X f dP_1 - \int_X f dP_2 \right|$

Pokazuje się, że operator

$$P: (\mathcal{M}(X), \|\cdot\|) \rightarrow (\mathcal{M}(X), \|\cdot\|)$$

jest słabą kontrakcją (tzn. ma stałą Lipschitza równą 1). My pokażemy ten (prosty) fakt w zawężeniu do miar probabilistycznych:

**Propozycja 7.**  $\|PP_1 - PP_2\| \leq \|P_1 - P_2\|$ ,  $P_1, P_2 \in \mathcal{M}^1(X)$ .

*Proof.* Możemy zastosować Propozycję 6:

$$\begin{aligned} \|PP_1 - PP_2\| &= \sup_{f \in \text{Me}(X, [0,1])} \left| \int_X \int_X f(y)P(x, dy)P_1(dx) - \int_X \int_X f(y)P(x, dy)P_2(dx) \right| \leq \\ &\leq \sup_{h \in \text{Me}(X, [0,1])} \left| \int_X h dP_1 - \int_X h dP_2 \right| = \|P_1 - P_2\|, \end{aligned}$$

ponieważ funkcja  $h(x) = \int_X f(y)P(x, dy)$  ma wartości w  $[0, 1]$  gdy  $f$  ma wartości w  $[0, 1]$ .  $\square$

**Wniosek 5.** *Z powyższego wnioskujemy, że jeśli  $\pi \in \mathcal{M}^1(X)$  jest miarą stacjonarną, wtedy*

$$\|P^n(x, \cdot) - \pi\| \leq \|P^{n-1}(x, \cdot) - \pi\|, \quad x \in X, \quad n \in \mathbb{N}^+.$$

Widzimy, że rozkłady łańcucha nie oddalają się od miary stacjonarnej. Powyższy wniosek oczywiście nie daje nam informacji o tempie zbieżności łańcucha do miary stacjonarnej (ani nawet nie implikuje samej zbieżności).

W przypadku skończonym  $X = \{1, \dots, d\}$ , z algebry liniowej (m.in. z teorii Jordana) wynika że rozkłady nieprzywiedlnego nieokresowego łańcucha zbiegają szybko - w tempie geometrycznym - do rozkładu stacjonarnego i tempo tej zbieżności jest ograniczone z góry przez wartość największej wartości własnej nie będącej jedynką. W takim przypadku mamy faktycznie  $\mathbb{P}_{X_n} = (P^T)^n \mathbb{P}_{X_0}$  i możemy skorzystać z postaci Jordana stochastycznej macierzy przejścia  $P$  oraz z odpowiednich twierdzeń algebry liniowej by otrzymać geometryczne tempo zbieżności postaci

$$\|\mathbb{P}_{X_n} - \pi\| \leq r^n \|\mathbb{P}_{X_0} - \pi\|,$$

gdzie stała  $r \in (0, 1)$ .

W przypadku ogólnym ( $X$  - przestrzeń polska) mogą występować różnorodne zachowania, jeśli chodzi o tempo zbieżności łańcucha. Przykładem jakościowo dużego tempa zbieżności jest właśnie wspomniane geometryczne (wykładnicze) tempo zbieżności, występujące w poniższej definicji.

**Definicja 22.** *Łańcuch Markowa z miarą stacjonarną  $\pi \in \mathcal{M}^1(X)$  jest **jednostajnie ergodyczny** (uniformly ergodic) jeśli dla pewnego  $r < 1$  oraz  $M \in \mathbb{R}^+$ :*

$$\|P^n(x, \cdot) - \pi\| \leq M \cdot r^n, \quad \text{dla każdego } x \in X.$$

Z dotychczasowej dyskusji wynika, że nieprzywiedlne nieokresowe łańcuchy Markowa na dyskretnej przestrzeni stanów zawsze są jednostajnie ergodyczne. W modelowaniu matematycznym często jednak rozważa się ciągłą przestrzeń stanów  $X$  i z punktu widzenia zastosowań teorii często ważne jest określenie warunków, przy których łańcuch zbiega szybko do rozkładu stacjonarnego. Jedną z ważnych probabilistycznych technik analizy zbieżności oraz tempa zbieżności nieredukowalnego łańcucha na ciągłej przestrzeni stanów jest technika sprzęgania łańcuchów (coupling, splitting technique). Technika ta bazuje na następującej nierówności: (**coupling inequality**)

$$(8.2) \quad \|\mathbb{P}_Z - \mathbb{P}_Y\| \leq \mathbb{P}(Z \neq Y),$$

gdzie  $Z$  oraz  $Y$  to dowolne zmienne losowe o wartościach w przestrzeni  $X$ . Nierówność (8.2) może być wykorzystana następująco: dla łańcucha  $X_n$  o mierze stacjonarnej  $\pi$  skonstruujemy powiązany z nim łańcuch  $X'_n$  o tym samym jądrze przejścia znajdujący się od początku w stanie stacjonarnym, tj.  $\mathbb{P}_{X'_n} = \pi$ ,  $n \in \mathbb{N}$ . Nierówność (8.2) daje nam wtedy następujące ograniczenie na tempo zbieżności:

$$(8.3) \quad \|P_{X_n} - \pi\| \leq P(X_n \neq X'_n).$$

**Dowód nierówności (8.2).** Niech  $A \in \mathcal{B}(X)$ . Mamy

$$\begin{aligned} |\mathbb{P}(Z \in A) - \mathbb{P}(Y \in A)| &= |\mathbb{P}(Z \in A, Z = Y) + \mathbb{P}(Z \in A, Z \neq Y) - \mathbb{P}(Y \in A, Z = Y) - \mathbb{P}(Y \in A, Z \neq Y)| = \\ &= |\mathbb{P}(Z \in A, Z \neq Y) - \mathbb{P}(Y \in A, Z \neq Y)| \leq P(Z \neq Y). \end{aligned}$$

Wobec powyższego:

$$\|P_Z - P_Y\| = \sup_{A \in \mathcal{B}(X)} |\mathbb{P}(Z \in A) - \mathbb{P}(Y \in A)| \leq P(Z \neq Y). \quad \square.$$

Poniżej przedstawimy jak dla jądra przejścia  $P$  z jedynym rozkładem stacjonarnym  $\pi$  przeprowadzić konstrukcję łańcuchów sprzężonych: łańcucha  $\mathbb{X} = \{X_n\}_{n \in \mathbb{N}}$  oraz łańcucha  $X'_n$ . Konstrukcja nie będzie w pełni formalna - zakładamy bez straty ogólności, że nasza wyjściowa przestrzeń probabilistyczna  $(\Omega, \Sigma, \mathbb{P})$  jest wystarczająco bogata by ją przeprowadzić formalnie. Zanim zaczniemy, potrzebujemy jednak wprowadzić jeszcze pojęcie zbioru małego  $C \in \mathcal{B}(X)$ . Klasa zbiorów małych jest wyznaczona przez własności jądra przejścia  $P$  i nie ma nic wspólnego z metryką na przestrzeni  $X$ . Jest to naturalne m.in. dlatego, że ani prawdopodobieństwo zbioru, ani norma całkowitego wahania miary, nie zależą od naszej metryki  $d$ .

**Definicja 23.** Powiemy, że zbiór  $C \in \mathcal{B}(X)$  jest zbiorem  $(n_0, \varepsilon, \nu)$ -małym, lub inaczej że jest zbiorem  $(n_0, \varepsilon \cdot \nu)$ -małym, jeśli dla stałej  $n_0 \in \mathbb{N}^+$ , stałej  $\varepsilon > 0$  oraz miary probabilistycznej  $\nu$  zachodzi:

$$(8.4) \quad P^{n_0}(x, \cdot) \geq \varepsilon \nu(\cdot), \quad x \in C.$$

Powiemy, że zbiór  $C$  jest zbiorem małym, jeśli dla pewnych stałych  $n_0$  i  $\varepsilon > 0$  oraz pewnej miary  $\nu \in \mathcal{M}^1(X)$  jest on zbiorem  $(n_0, \varepsilon, \nu)$ -małym.

Przyjrzyjmy się warunkowi (8.4), który można zapisać równoważnie jako:  $\inf_{x \in C} P^{n_0}(x, \cdot) \geq \varepsilon \nu(\cdot)$ . Intuicja jest następująca: zbiór  $C$  jest mały, gdy znajdziemy na tyle duże  $n_0 \in \mathbb{N}$ , że wszystkie rozkłady  $P^{n_0}(x, \cdot)$ ,  $x \in C$ , "zazebiają się ze sobą" w tym sensie, że posiadają niezerową część wspólną, tj. są ograniczone z dołu przez pewną niezerową miarę dodatnią  $\varepsilon \cdot \nu$ .

**Przykład 22.** Dla dowolnego jądra przejścia oraz punktu  $x \in X$ , singleton  $\{x\}$  jest zbiorem małym. Warunek (8.4) jest trywialnie spełniony dla stałych  $n_0 = 1$ ,  $\varepsilon = 1$  oraz miary  $\nu = P(x, \cdot)$ .

**Przykład 23.** Dla dowolnego nieprzywiedlnego nieokresowego łańcucha  $\mathbb{X}$  na skończonej przestrzeni  $X = \{1, \dots, d\}$  cała przestrzeń  $X$  jest zbiorem małym. Wynika to z własności, że dla pewnego  $n_0$  dostatecznie dużego wszystkie stany się ze sobą komunikują w  $n_0$  krokach, tzn.  $P^{n_0}(x, y) > 0$  dla dowolnych  $x, y \in X$ . Zatem można łatwo dobrać miarę postaci  $\varepsilon \nu$  definiując dla dowolnego  $z \in X$ ,  $\varepsilon \cdot \nu(z) = \min\{P^{n_0}(x, y) : x, y \in X\}$ .

**Przykład 24.** W błędzeniu losowym po przestrzeni  $X = \mathbb{Z}$  każdy zbiór ograniczony jest zbiorem małym. Każdy zbiór nieograniczony nie jest zbiorem małym.

**Przykład 25.** Jeśli  $X$  jest przestrzenią zwartą oraz dla pewnego  $n \in \mathbb{N}^+$  gęstość  $f$  jądra przejścia  $P^n(x, \cdot)$ , tj. funkcja spełniająca  $P^n(x, A) = \int_A f(x, y) dy$ , jest dodatnia oraz ciągła, wtedy cała przestrzeń  $X$  jest zbiorem małym.

O ile jest oczywiste, że podzbiór zbioru małego jest zbiorem małym, trzeba pamiętać, że w ogólnym przypadku suma zbiorów małych może już nie być zbiorem małym: jeśli łańcuch nie jest nieredukowalny i posiada dwa rozłączne zbiory absorbujące  $A$  oraz  $B$ , wtedy dla wybranych  $a \in A$ ,  $b \in B$ , zbiór  $\{a, b\}$  nie będzie mały - miary  $P^n(a, \cdot)$  oraz  $P^n(b, \cdot)$  mają zerową "część wspólną", ponieważ są skupione na rozłącznych podzbiórach przestrzeni  $X$ .

W poniższej konstrukcji łańcuchów sprzężonych zakładamy, że mamy  $(n_0, \varepsilon, \nu)$ -mały zbiór  $C$ .

### Coupling construction (CC)

Na starcie  $n = 0$  losujemy  $X_0 = x$  z rozkładu początkowego oraz  $X'_0 = x'$  z rozkładu stacjonarnego  $\pi$ .

**Pętla:** Mając dane  $X_n = x$  oraz  $X'_n = x'$ :

- (1) Jeśli  $x = x'$ , losujemy  $X_{n+1} = X'_{n+1}$  z rozkładu  $P(x, \cdot)$ , kładziemy  $n := n + 1$ , wracamy na początek pętli
- (2) Jeśli  $x \in C$  oraz  $x' \in C$ :
  - (a) z prawdopodobieństwem  $\varepsilon$  losujemy  $X_{n+n_0} = X'_{n+n_0}$  z rozkładu  $\nu$ ,
  - (b) z prawdopodobieństwem  $1 - \varepsilon$  losujemy  $X_{n+n_0} = x_{n+n_0}$  oraz  $X'_{n+n_0} = x'_{n+n_0}$  **niezależnie** od siebie,  $X_{n+n_0}$  rozkładu  $\frac{1}{1-\varepsilon}(P^{n_0}(x, \cdot) - \varepsilon \cdot \nu)$ , natomiast  $X'_{n+n_0}$  rozkładu  $\frac{1}{1-\varepsilon}(P^{n_0}(x', \cdot) - \varepsilon \cdot \nu)$ .

Jeśli  $n_0 > 1$ , losujemy wektory  $(X_{n+1}, \dots, X_{n+n_0-1})$  oraz  $(X'_{n+1}, \dots, X'_{n+n_0-1})$  niezależnie, z ich rozkładów warunkowych  $\mathbb{P}(\cdot | X_n = x_n, X_{n+n_0} = x_{n+n_0})$  oraz  $\mathbb{P}(\cdot | X'_n = x'_n, X'_{n+n_0} = x'_{n+n_0})$  by uzupełnić brakujące zmienne w łańcuchu

Po wykonaniu kroku 2) kładziemy  $n := n + n_0$  i wracamy na początek pętli

- (3) Jeśli nie zachodzi (1) ani (2), wtedy losujemy, niezależnie od siebie,  $X_{n+1}$  z rozkładu  $P(x, \cdot)$  oraz  $X'_{n+1}$  z rozkładu  $P(x', \cdot)$ , kładziemy  $n := n + 1$ , wracamy na początek pętli.

Jaki jest **sens powyższej konstrukcji**?

- (1) Po pierwsze, można sprawdzić, że uzyskane ciągi  $X_n$  oraz  $X'_n$  są łańcuchami Markowa o zadanym jądrze przejścia  $P$  (równoważnie: każdy z ciągów  $X_n$  ma rozkłady skończone wymiarowe zadane przez jądro przejścia  $P$ . W Kroku 2(b) dołosowujemy brakujące fragmenty łańcucha z odpowiednich rozkładów warunkowych właśnie po to by zachować prawidłowy rozkład łańcucha). W szczególności,  $P_{X'_n} = \pi$ , ponieważ łańcuch  $X'_n$  startował z rozkładu stacjonarnego. Celem konstrukcji jest uzyskanie łańcuchów mających możliwie duże prawdopodobieństwo zdarzenia  $\{X_n = X'_n\}$  by następnie zastosować nierówność (8.3) do analizy zbieżności łańcucha  $X_n$  do rozkładu stacjonarnego  $\pi$ .
- (2) W Kroku 1 widzimy, że jeśli łańcuchy w którymś kroku przyjmą tę samą wartość, pozostają już sobie równe w dalszych krokach. W chwili startowej  $n = 0$  należy się spodziewać że mają różne wartości (przy naturalnych założeniach - z prawdopodobieństwem jeden startują z różnych punktów). Łańcuchy "działają" niezależnie aż do momentu gdy oba w tym samym kroku trafią w zbiór mały  $C$ .
- (3) W Krokach 2 i 3 widzimy, że jeśli bieżące wartości łańcuchów jeszcze nie są sobie równe, działają one niezależnie aż do momentu, gdy oba równocześnie trafią w zbiór  $C$  - za każdym trafieniem wektora  $(X_n, X'_n)$  w zbiór  $C^2$  z prawdopodobieństwem  $\varepsilon$  zajdzie zdarzenie  $X_{n+1} = X'_{n+1}$ . Stąd, jeśli wiemy, że ciąg  $(X_n, X'_n)$  odwiedzi nieskończenie wiele razy zbiór  $C^2$ , mamy zagwarantowane  $P(X_n = X'_n) \rightarrow 1$  (ponieważ  $\varepsilon > 0$ ). A to na mocy nierówności (8.3) dowodzi zbieżności łańcucha  $X_n$  do miary  $\pi$ .

W przypadku dyskretnym w dowodach stosowaliśmy technikę regeneracji łańcucha po każdym trafieniu w punkt  $z$ . W przypadku ogólnym zazwyczaj nie mamy możliwości trafić w wybrany



singleton  $\{z\}$ , natomiast każdorazowe trafienie pary łańcuchów sprzężonych w zbiór mały  $C$  jest namiastką regeneracji która umożliwi w Kroku 2(a) z dodatnim prawdopodobieństwem zlepienie się łańcuchów.

Przykładowe zastosowanie tej techniki zobaczymy teraz w dowodzie poniższego twierdzenia, dającego prosty warunek dostateczny dla jednostajnej ergodyczności łańcucha.

**Twierdzenie 24.**  $\mathbb{X}$  jest łańcuchem o rozkładzie stacjonarnym  $\pi$ . Rozważmy przypadek specjalny, gdy cała przestrzeń jest zbiorem małym, tzn.  $X$  jest zbiorem  $(n_0, \varepsilon, \nu)$ -małym. Wtedy  $\mathbb{X}$  jest jednostajnie ergodyczny:

$$\|P^n(x, \cdot) - \pi\| \leq (1 - \varepsilon)^{\lfloor \frac{n}{n_0} \rfloor}, \quad x \in X.$$

*Proof.* Stosując konstrukcję sprzęgania łańcuchów (CC) z wykorzystaniem zbioru małego  $C = X$  widzimy, że w  $n = m \cdot n_0$  krokach mieliśmy co najmniej  $m$  szans na zlepienie się łańcuchów - za każdym razem z prawd.  $\varepsilon > 0$ , zatem po  $n = m \cdot n_0$  krokach mamy

$$\mathbb{P}(\{X_n \neq X'_n\}) \leq (1 - \varepsilon)^m.$$

Stąd, korzystając z Wniosku 5, dla dowolnego  $n \in \mathbb{N}$

$$\mathbb{P}(\{X_n \neq X'_n\}) \leq (1 - \varepsilon)^{\lfloor \frac{n}{n_0} \rfloor}.$$

Nierówność 8.2 kończy dowód. □

W szczególności, ponieważ w przypadku dyskretnym  $X = \{1, \dots, d\}$  cała przestrzeń nieredukowalnego nieokresowego łańcucha jest zbiorem małym, właśnie udowodniliśmy twierdzenie o zbieżności łańcucha do miary stacjonarnej z dowolnego punktu początkowego, dając przy okazji informację o tempie zbieżności.

A jak udowodnić Twierdzenie 6 korzystając z techniki łańcuchów sprzężonych? Tutaj trzeba wykorzystać rezultaty wiążące istnienie zbiorów małych z istnieniem miary stacjonarnej - zachodzi następujące twierdzenie.

**Twierdzenie 25.** Dla nieredukowalnego łańcucha Markowa z miarą stacjonarną  $\pi$  istnieje zbiór mały  $C$  taki, że  $\pi(C) > 0$  oraz  $\nu(C) > 0$ , gdzie  $\nu$  to miara występująca w definicji zbioru  $(n, \varepsilon, \nu)$ -małego  $C$ .

By udowodnić Twierdzenie 6 pokazujemy, że w konstrukcji łańcuchów sprzężonych  $(X_n, X'_n)$  zbiór  $C \times C$ , gdzie  $C$  pochodzi z Twierdzenia 25, zostanie odwiedzony przez parę tę łańcuchów nieskończenie wiele razy, co zakończy dowód analogicznie jak w Twierdzeniu 24. Omijamy ten dowód, ponieważ mamy już zielony maj i brak czasu a nie znam zwięzłego dowodu.

Trochę słabszym warunkiem niż jednostajna ergodyczność, określającym jakościowo duże tempo zbieżności, jest geometryczna ergodyczność:

**Definicja 24.** Łańcuch  $\mathbb{X}$  posiadający miarę stacjonarną  $\pi \in \mathcal{M}^1(X)$  jest geometrycznie ergodyczny, jeśli:

$$\|P^n(x, \cdot) - \pi\| \leq M(x) \cdot r^n,$$

gdzie  $r < 1$  oraz  $M(x) < \infty$  dla  $\pi$ -prawie każdego  $x \in X$ .

Podamy teraz warunek wystarczający dla geometrycznej ergodyczności, który wykorzystuje pojęcia zbioru małego oraz funkcję Lapunowa  $V$ , zwany warunkiem (geometrycznego) dryfu.

**Definicja 25.** Powiemy, że łańcuch  $\mathbb{X}$  spełnia **warunek dryfu** w kierunku zbioru  $C$  dla funkcji  $V: X \rightarrow [1, \infty]$  jeśli istnieją stałe  $\lambda \in (0, 1)$  oraz  $b \in \mathbb{R}^+$ , tże:

$$(8.5) \quad PV(x) \leq \lambda V(x) + b \cdot 1_C(x), \quad x \in X.$$

Równoważnie:

$$E[V(X_{n+1}) | X_n = x] \leq \lambda \cdot V(x) + b \cdot 1_C(x).$$

**Zadanie 23.** a) Zauważ, że warunek dryfu (8.5) wymusza, że  $C \neq \emptyset$ . b) Ogólniej, zauważ, że  $C$  jest zbiorem osiągalnym z każdego punktu  $x$  spełniającego  $V(x) < \infty$ .

Możemy teraz sformułować ogólny wynik:

**Twierdzenie 26.** Niech  $\mathbb{X}$  będzie nieredukowalnym nieokresowym łańcuchem z miarą stacjonarną  $\pi \in \mathcal{M}^1(X)$ . Załóżmy, że  $C \in \mathcal{B}(X)$  jest zbiorem  $(n_0, \varepsilon, \nu)$ -małym oraz spełniony jest warunek dryfu (8.5) w kierunku zbioru  $C$  wzdłuż funkcji  $V: X \rightarrow [1, \infty]$  takiej, że  $V(x) < \infty$  dla  $\pi$ -prawie każdego  $x \in X$ . Wtedy  $\mathbb{X}$  jest geometrycznie ergodyczny.

Powyższe twierdzenie również można udowodnić korzystając z konstrukcji łańcuchów sprzężonych - warunek dryfu daje dodatkową kontrolę nad zachowaniem łańcucha skąd, przy użyciu pewnej dość skomplikowanej konstrukcji, możemy oszacować  $P(X_n \neq X'_n)$  uzyskując ostatecznie informację o geometrycznym tempie zbieżności. Nie znam eleganckiego dowodu Twierdzenia 26. Dowód, który znam, prawdopodobnie uczyniłby Państwa mniej szczęśliwymi ludźmi. Aczkolwiek istnieje elegancki i krótki dowód pewnego specjalnego przypadku Twierdzenia 26, autorstwa Martina Hairera (medal Fieldsa 2014), który Państwu podeślę jako materiały uzupełniające do wykładu.

Zakończyliśmy omawiać klasyczną teorię łańcuchów Markowa. Dokonało się.

## 9. ZASTOSOWANIA W STATYSTYCE - METODY MCMC

Jednym z problemów statystyki jest umiejętność wylosowania próbki ze skomplikowanego rozkładu prawdopodobieństwa  $\pi \in \mathcal{M}(\mathbb{R}^n)$  - problem ten może być szczególnie trudny przy dużej liczbie wymiarów  $n$ . Bardzo popularną metodą ominięcia tego problemu stały się metody MCMC (Markov Chain Monte Carlo) polegające na skonstruowaniu odpowiedniego łańcucha Markowa zbieżnego z prawie każdego punktu, najlepiej szybko, do zadanego rozkładu. Dzięki temu wartość łańcucha po  $n$  krokach, dla dostatecznie dużego  $n$ , będzie można traktować z dobrym przybliżeniem jako pochodzącą z rozkładu  $\pi$ . Równie ważnym problemem statystyki jest umiejętność policzenia całki  $\int f d\pi$ , gdzie  $\pi$  jest wspomnianym skomplikowanym rozkładem. Gdybyśmy potrafili uzyskać próbę  $Z_1, Z_2, \dots$  z rozkładu  $\pi$ , wtedy klasyczna metoda Monte Carlo, na podstawie prawa wielkich liczb, daje nam zbieżność z prawdopodobieństwem 1

$$\frac{1}{n} \sum_{i=1}^N f(Z_i) \rightarrow \int f d\pi,$$

zatem, dla odpowiednio dużego  $n$ , skończona suma  $\frac{1}{n} \sum_{i=1}^N f(Z_i)$  będzie dobrym przybliżeniem nieznaney wartości  $\int f d\pi$ . Gdy nie potrafimy uzyskać próby  $Z_1, Z_2, \dots$  ale potrafimy skonstruować nieredukowalny łańcuch Markowa z miarą stacjonarną  $\pi$ , wtedy na mocy prawa wielkich liczb dla łańcuchów Markowa

$$\frac{1}{n} \sum_{i=1}^N f(X_i) \rightarrow \int f d\pi$$

co również pozwala nam aproksymować wartość  $\int f d\pi$ . Wydawać by się mogło że skonstruowanie łańcucha Markowa zbieżnego do skomplikowanego rozkładu jest szczególnie trudnym zadaniem. Jest zupełnie przeciwnie. Podstawowe algorytmy MCMC to algorytm **Metropolisa-Hastinga** oraz algorytm **Gibbsa**. Skupimy się na algorytmie Metropolisa-Hastinga, który jest szerzej stosowany (i bardzo łatwy w implementacji).

Zakładamy, że znamy funkcję nieujemną  $h: \mathbb{R}^n \rightarrow \mathbb{R}$  taką, że po unormowaniu jest ona gęstością rozkładu  $\pi$ , tzn.  $h_\pi = c \cdot h$  jest gęstością rozkładu  $\pi$  dla pewnej stałej  $c \in \mathbb{R}^+$ . Takie sformułowanie problemu (nieznajomość stałej normującej  $c$ ) wynika z zastosowania omawianej

teorii we wnioskowaniu Bayesowskim, gdzie wartość gęstości a posteriori jest w pierwszej kolejności znana jedynie z dokładnością do wspomnianej stałej.

**Algorytm Metropolisa-Hastingsa:** Niech  $Q \in \mathcal{K}(\mathbb{R}^n)$  będzie jądrem przejścia o gęstości  $q$ , tzn. mamy  $Q(x, A) = \int_A q(x, y) dy$ . Krok rekurencyjny wygląda następująco: Jeśli jesteśmy w stanie  $X_n = x$ , wtedy losujemy punkt  $y$  z rozkładu  $Q(x, \cdot)$  a następnie z prawdopodobieństwem  $\alpha(x, y)$  stawiamy  $X_{n+1} = y$  natomiast z prawdopodobieństwem  $1 - \alpha(x, y)$  pozostajemy w punkcie  $x$ , tzn. stawiamy  $X_{n+1} = x$ . Pozostaje jedynie zdefiniować odpowiednio

$$\alpha(x, y) = \min\left\{1, \frac{h(y) \cdot q(y, x)}{h(x) \cdot q(x, y)}\right\},$$

przy czym stawiamy  $\alpha(x, y) := 1$  wszędzie gdzie mianownik w powyższym wyrażeniu się zeruje.

**Uwaga 4.** *Zauważmy, że definicja algorytmu M-H nie wymaga znajomości stałej normalizującej  $c$ . Skonstruowanie algorytmu wymaga jedynie umiejętności liczenia wartości funkcji  $h$  oraz umiejętności losowania z jądra  $Q(x, \cdot)$ .*

Na pierwszy rzut oka wartość prawdopodobieństwa akceptacji nowego punktu jest "zaskakująca", tzn. wzór na funkcję  $\alpha$  wydaje się niejasny. Jego postać wynika z następującego twierdzenia, które dowodzi się bezpośrednim rachunkiem (dość krótkim).

**Twierdzenie 27.** *Rozkład  $\pi \in \mathcal{M}^1(X)$  (zadany gęstością  $h_\pi$ ) jest rozkładem stacjonarnym algorytmu Metropolisa-Hastingsa.*

Powyższe Twierdzenie może być zaskakujące tym bardziej, że nic nie założyliśmy o jądrze przejścia  $Q$  (poza tym że posiada gęstość). Dodatkowo, przy raczej naturalnych założeniach o jądrze przejścia  $Q$ , łatwo pokazuje się, że algorytm M-H jest nieredukowalny oraz nieokresowy, a zatem, na mocy Twierdzenia 6, zbiega on do miary stacjonarnej! Dodatkowo, spełnia on mocne prawo wielkich liczb. Intuicyjnie: jądro przejścia  $Q$  musi umożliwić algorytmowi "swobodną wędrówkę" po przestrzeni  $X$  (zakładamy:  $X \subset \mathbb{R}^n$ ).

Przykładowe założenia (dużo mocniejsze niż konieczne) dają poniżej.

**Twierdzenie 28.** *Jeśli gęstość jądra  $Q$  jest dodatnia na całej przestrzeni, wtedy algorytm M-H jest nieredukowalny (względem miary  $\pi$ ) i nieokresowy. W szczególności, jest zbieżny silnie do miary stacjonarnej oraz spełnia mocne prawo wielkich liczb.*

Zamiast zastanawiać się nad ogólnymi założeniami, jakie trzeba nałożyć na jądro przejścia  $Q$ , rozważmy prosty przypadek, która przyniesie trochę intuicji związanej z algorytmem M-H.

**Przypadek: jądro symetryczne.** Załóżmy że  $X = \mathbb{R}^n$  oraz że jądro przejścia  $Q$  ma symetryczną gęstość, tzn.  $q(x, y) = q(y, x)$  - założenie symetrii jest tutaj naturalne, ponieważ przestrzeń  $\mathbb{R}^n$  jest jednorodna. Przykładowo: jeśli  $Q(x, \cdot)$  to rozkład normalny  $N_n(x, \sigma \cdot I_n)$  lub rozkład jednostajny na kuli  $n$ -wymiarowej  $K(x, r)$ , wtedy jądro jest zadane gęstością symetryczną (łatwo sprawdzić!) W symetrycznym przypadku, po wylosowaniu punktu  $y$  z rozkładu  $Q(x, \cdot)$ , zostaje on zaakceptowany jako wartość algorytmu w kolejnym kroku z prawdopodobieństwem

$$\alpha(x, y) = \min\left\{1, \frac{h(y)}{h(x)}\right\},$$

gdzie  $x$  to stan algorytmu w kroku bieżącym. Tutaj interpretacja jest już prosta: jeśli wartość gęstości  $h(y)$  jest większa niż wartość gęstości w kroku bieżącym  $h(x)$ , wtedy z prawdopodobieństwem 1 przechodzimy do stanu  $y$ . Jeśli wartość  $h(y)$  jest mniejsza niż bieżąca wartość  $h(x)$ , wtedy prawdopodobieństwo jest już mniejsze - jak widzimy, jest tym mniejsze, im mniejsza jest

wartość ułamka  $\frac{h(y)}{h(x)}$ . Krótko mówiąc: miejsca w przestrzeni o większej wartości  $h_\pi$  są chętniej odwiedzane w przestrzeni niż miejsca o mniejszych wartościach  $h_\pi$ . Ponieważ  $h_\pi$  jest gęstością rozkładu stacjonarnego, jest to zgodne z intuicją oraz z teorią: zbiory o większej wartości miary stacjonarnej  $\pi$  są częściej odwiedzane przez algorytm.

A co z tempem zbieżności algorytmu? Jest ono oczywiście zależne zarówno od gęstości docelowej  $h_\pi$ , jak i od jądra przejścia  $Q$ . Tempo zbieżności można badać warunkami dryfu z poprzedniego rozdziału. Jest to jednak problem bardzo trudny (m.in. rachunkowo) - nie ma ogólnego przepisu na funkcję Lapunova  $V$  występującą w warunkach dryfu. Jest to dość charakterystyczna sytuacja dla matematyki stosowanej - mając odpowiednie narzędzia teoretyczne, dość łatwo uzasadniamy poprawność danej metody (w naszym wypadku - zbieżność do miary stacjonarnej). Jednak w kwestii efektywności danej metody, ze względu na złożoność problemu, badania nad tempem zbieżności są prowadzone przede wszystkim numerycznie (co nie powstrzymuje teoretyków przed dalszym rozwojem teorii).

## 10. ŁAŃCUCHY MARKOWA W OPTYMALIZACJI.

Zakładamy, że mamy funkcję mierzalną  $f: X \rightarrow \mathbb{R}$  posiadającą minimum globalne. Bez straty ogólności zakładamy że wartość funkcji w minimum globalnym wynosi zero a zatem definiujemy zbiór minimów globalnych jako

$$A^* = \{x \in A: f(x) = 0\}.$$

Wiele problemów nauki i techniki sprowadza się do znalezienia minimum globalnego funkcji  $f$  (problemy minimalizacji funkcji oraz maksymalizacji funkcji są sobie równoważne z teoretycznego punktu widzenia (maksima funkcji  $f$  to minima funkcji  $-f$ ) i bez straty ogólności zakładamy, że szukamy minimum). Często wystarczy znaleźć dostatecznie dokładne przybliżenie minimum globalnego. W kontekście optymalizacji, funkcję  $f$  będziemy nazywać również funkcją celu (w matematyce finansowej często funkcją kosztu, z angielskiego: 'problem function', 'objective function' lub 'target function'). Gdy problem zadany przez funkcję  $f$  jest odpowiednio skomplikowany (duża liczba wymiarów w przestrzeni  $X$ , duża liczba minimów lokalnych lub brak różniczkowości) zazwyczaj nie potrafimy wyznaczyć analitycznie elementu  $a \in A^*$ . W wielu problemach praktycznych nie znamy nawet wzoru na funkcję  $f$ , natomiast potrafimy policzyć jej wartość dla ustalonego stanu  $x \in X$  (są to tzw. 'black box problem' - przestrzenią stanów jest zazwyczaj wysoko wymiarowy zbiór parametrów konkretnego problemu polegającego na wyborze parametrów optymalnych). Dostępna jest natomiast wielka liczba różnych procedur iteracyjnych, które wykorzystują losowość i są przeznaczone do znajdowania pożądanego aproksymacji minimum globalnego. Przykładowe i powszechne metody to algorytm Simulated Annealing (SA), będący modyfikacją algorytmu Metropolisa-Hastingsa z poprzedniego rozdziału oraz algorytmy inteligencji stadnej inspirowane zachowaniem kolonii mrówek (Ant Colony Algorithm (ACO)), pszczoł (Artificial Bee Colony (ABC)) czy stada ptaków (Particle Swarm Optimization (PSO)). Najbardziej znane są jednak algorytmy ewolucyjne i genetyczne, inspirowane biologicznymi procesami selekcji, mutacji oraz rekombinacji. Istnieje również wiele innych iteracyjnych metod których tutaj nie wymieniam. Wszystkie wspomniane algorytmy możemy modelować ogólnym równaniem:

$$X_{t+1} = T_t(X_t, Y_t), \quad t \in \mathbb{N},$$

gdzie  $X_t$  reprezentuje stan algorytmu w kroku  $t$ ,  $Y_t$  reprezentuje losowość występującą w algorytmie natomiast  $T_t$  to deterministyczna procedura która wykorzystując bieżący stan  $X_t$  oraz losowość  $Y_t$  (oraz powiązane wartości funkcji  $f$ ) przenosi algorytm do kolejnego stanu  $X_{t+1}$ . Powyższe równanie wygląda bardzo znajomo: widzimy, że jeśli mechanizm  $T_t$  nie zmienia się w czasie (czyli  $T_t = T$ ,  $t \in \mathbb{N}$ ) oraz rozkłady  $Y_t$  nie zmieniają się w czasie, równanie modelujące

algorytm definiuje jednorodny łańcuch Markowa. W ogólnym przypadku dostajemy natomiast niejednorodny łańcuch Markowa i tym ogólniejszym przypadkiem nie będziemy się tutaj zajmować (prześle coś dodatkowego w ramach materiałów uzupełniających). By nabrać intuicji, jak wspomniane algorytmy mogą działać poprawnie przy tak ogólnych założeniach dotyczących funkcji  $f$ , rozważmy następujący prosty algorytm.

**Algorytm 1.** Niech  $Q \in \mathcal{K}(X)$  będzie jądrem przejścia oraz niech  $X_0 = x$  będzie punktem startowym algorytmu. Teraz definiujemy rekurencyjnie: znając stan  $X_t = x_t$ , losujemy punkt  $q_t$  z rozkładu  $Q(x_t, \cdot)$  i przechodzimy do kolejnego stanu zgodnie ze wzorem:

$$X_{t+1} = \begin{cases} q_t & \text{jeśli } f(q_t) < f(x_t) \\ x_t & \text{jeśli } f(q_t) \geq f(x_t) \end{cases}.$$

Widzimy, że mechanizm jest prosty: w każdym kroku losujemy kandydata  $q$  na kolejny stan algorytmu z jądra  $Q(x, \cdot)$ , gdzie  $x$  to obecny stan. Następnie, porównujemy wartość funkcji  $f(q)$  z wartością funkcji w bieżącym stanie algorytmu i wybieramy punkt  $q$  na kolejny stan tylko gdy jest on atrakcyjniejszy w sensie funkcji  $f$  niż stan bieżący. Tak więc, wartości funkcji  $f$  ukierunkowują ewolucję ciągu  $x_0, x_1, \dots$  spełniając  $f(x_0) \geq f(x_1) \geq \dots$ . Przy naturalnych założeniach dotyczących funkcji  $f$  zachodzi

$$(10.1) \quad f(x_t) \rightarrow 0 \iff x_t \rightarrow A^*,$$

gdzie zbieżność  $x_t \rightarrow A^*$  oznacza zbieżność  $x_t$  do zbioru  $A^*$ , tzn.

$$d(x_t, A^*) := \inf_{a \in A^*} d(x_t, a) \rightarrow 0 \text{ gdy } t \rightarrow \infty.$$

Warunek (10.1) sugeruje poprawność Algorytmu 1 - funkcja  $f$  ukierunkowuje ewolucję algorytmu  $f$  we właściwy sposób. W poniższej obserwacji zamieszczam warunki wystarczające by spełnić (10.1). Dowodzi się ją korzystając z jednostajnej zbieżności funkcji  $f$  na pewnym zwartym otoczeniu zbioru  $A^*$  oraz z twierdzenia Weierstrassa.

**Obserwacja 6.** *Jeśli dla pewnego  $\delta > 0$  zbiór podpoziomicowy*

$$A_\delta = \{x \in X : f(x) \leq \delta\}$$

*jest zwarty oraz funkcja  $f$  jest ciągła na tym zbiorze, wtedy spełniony jest warunek (10.1). W szczególności, jeśli funkcja  $f$  jest ciągła oraz zbiór minimów  $A^*$  posiada otoczenie zwarte, wtedy spełniony jest warunek (10.1).*

Zdefiniujemy teraz zbieżność globalną algorytmu  $X_t$  - można to zrobić na wiele sposobów, które przy naszych założeniach, okażą się równoważne.

**Definicja 26.** *Powiemy, że algorytm  $X_t$  jest zbieżny globalnie, gdy dla każdego rozkładu początkowego jest zbieżny stochastycznie do zbioru  $A^*$ , co oznaczamy  $X_t \xrightarrow{s} A^*$  i definiujemy:*

$$\forall \varepsilon > 0 \quad \mathbb{P}[d(X_t, A^*) < \varepsilon] \rightarrow 1.$$

W naturalnej dla optymalizacji sytuacji - gdy zbiór minimów globalnych jest singletonem, tj.  $A^* = \{a^*\}$  - zbieżność globalna jest znaną nam dobrze zbieżnością stochastyczną (inaczej: wg prawdopodobieństwa) ciągu  $X_n$  do punktu  $a^*$ . Ponieważ granica w tym przypadku jest stała, zbieżność stochastyczna jest równoważna zbieżności słabej rozkładów  $X_t$  do dystrybuanty  $\delta_{a^*}$ . Poniższa obserwacja uogólnia tę własność stwierdzając, że zbieżność globalna jest równoważna zbieżności rozkładów algorytmu do zbioru  $\mathcal{M}^*$  miar probabilistycznych skupionych na  $A^*$ :

$$\mathcal{M}^* = \{\mu \in \mathcal{M}^1(X) : \mu(A^*) = 1\}.$$

**Obserwacja 7.** *Zachodzi:*

$$X_t \xrightarrow{s} A^* \iff d_P(\mathbb{P}_{X_t}, \mathcal{M}^*) \rightarrow 0,$$

gdzie  $d_P(\mathbb{P}_{X_t}, \mathcal{M}^*)$  to odległość (w metryce Prochorowa  $d_P$ ) rozkładu  $\mathbb{P}_{X_t}$  od zbioru  $\mathcal{M}^*$ .

Od razu zauważmy że w podstawowym przypadku  $A^* = \{a^*\}$  mamy  $\mathcal{M}^* = \{\delta_{a^*}\}$ . Dodatkowo, ponieważ rozkłady algorytmu przy naturalnych założeniach są absolutnie ciągle względem miary Lebesgue'a (w przypadku  $X \subset \mathbb{R}^n$ ), nie zachodzi zbieżność silna do miary granicznej  $\delta_{a^*}$ . Jest to jeden z powodów dla których klasyczna teoria łańcuchów Markowa nie pasuje do niniejszego kontekstu. Nasza metodologia wciąż będzie używać pojęcia funkcji Lapunowa ale wykorzystana technika będzie pochodzić bezpośrednio z teorii układów dynamicznych.

Naturalnie, podstawowe pytania optymalizacji to: 1) Przy jakich warunkach algorytm jest zbieżny globalnie? 2) Jakie jest tempo tej zbieżności? Odpowiemy dość łatwo, na gruncie teoretycznym, na pytanie 1). Jeśli chodzi o pytanie 2), dominują tutaj eksperymenty numeryczne (jest wciąż rozwijana ogromna literatura w tym temacie) natomiast rozważania teoretyczne są bardzo trudne i często prowadzone przy założeniach uproszczonych.

Będziemy pracować przy założeniach trochę ogólniejszych niż występująca w Algorytmie 1 monotoniczność ciągu  $f(X_t)$ ,  $t \in \mathbb{N}$ . Naturalnym uogólnieniem warunku  $f(X_{t+1}) \leq f(X_t)$  jest własność bycia nadmartyngałem:

$$E[f(X_{t+1})|f(X_t), f(X_{t-1}), \dots, f(X_0)] \leq f(X_t) \text{ a. s.}$$

Powyższa nierówność wynika z silniejszej własności

$$E[f(X_{t+1})|X_t, X_{t-1}, \dots, X_0] \leq f(X_t) \text{ a. s.},$$

która z kolei, jeśli  $X_t$  jest łańcuchem Markowa, wynika z następującej własności

$$(10.2) \quad E[f(X_{t+1})|X_t = x] \leq f(x), \quad x \in X.$$

Równanie (10.2) jest oczywiście własnością nałożoną na jądro przejścia naszego łańcucha Markowa, którą możemy równoważnie zapisać

$$Pf(x) \leq f(x), \quad x \in X.$$

Poniżej dowodzimy powyższe uwagi

**Lemat 4.** *Założmy, że  $X_t$  jest łańcuchem Markowa spełniającym  $E(f(X_t)) < +\infty$ ,  $t \in \mathbb{N}$ . Jeśli równanie (10.2) jest spełnione, wtedy ciąg  $f(X_t)$  jest nadmartyngałem.*

*Proof.* Niech  $\Sigma_t = \Sigma(f(X_t), \dots, f(X_0))$ ,  $t \in \mathbb{N}$ . Ponieważ,  $X_t$  jest łańcuchem Markowa, mamy

$$E(f(X_{t+1})|X_t, \dots, X_0) = E(f(X_{t+1})|X_t), \text{ a. s.}$$

oraz, ponieważ  $\Sigma_t \subset \Sigma(X_t, \dots, X_0)$ , z prawa wieży:

$$E[f(X_{t+1})|\Sigma_t] = E[E(f(X_{t+1})|X_t, \dots, X_0)|\Sigma_t] = E[E(f(X_{t+1})|X_t)|\Sigma_t].$$

Z monotoniczności warunkowej wartości pozostaje pokazać  $E(f(X_{t+1})|X_t) \leq f(X_t)$  a.s. co już wynika z równania (10.2). Faktycznie: funkcja  $E(f(X_{t+1})|X_t)$  jest mierzalna względem  $\Sigma(X_t)$  a zatem istnieje funkcja mierzalna  $h: A \rightarrow \mathbb{R}$  spełniająca  $E(f(X_{t+1})|X_t) = h(X_t)$ . Mamy założone w równaniu (10.2):  $E(f(X_{t+1})|X_t = x) = h(x) \leq f(x)$ , a to już implikuje nierówność  $E(f(X_{t+1})|X_t) \leq f(X_t)$  i kończy dowód.  $\square$

Poniższe proste twierdzenie obrazuje równoważności pomiędzy różnymi klasycznymi rodzajami zbieżności algorytmu.

**Twierdzenie 29.** Zakładamy, że funkcja  $f : X \rightarrow \mathbb{R}$  jest ciągła na pewnym zwartym otoczeniu zbioru  $A^*$  oraz że ciąg  $f(X_t)$  jest nadmartyngałem. Następujące warunki są równoważne:

- (1)  $\mathbb{P}_{X_t}$  zbiegają słabo do zbioru  $M^* = \{\mu \in M^1(X) | \mu(A^*) = 1\}$
- (2)  $X_t \rightarrow A^*$  stochastycznie (zbieżność globalna)
- (3)  $d(X_t, A^*) \rightarrow 0$  z prawdopodobieństwem 1,
- (4)  $f(X_t)$  zbiega do 0 słabo,
- (5)  $f(X_t)$  zbiega do 0 stochastycznie,
- (6)  $f(X_t)$  zbiega do 0 z prawdopodobieństwem 1

Jeśli założymy dodatkowo że funkcje mierzalne  $f(X_t)$  oraz  $d(X_t, A^*)$  są ograniczone z góry przez pewną mierzalną funkcję  $Z : \Omega \rightarrow [0, +\infty)$  spełniającą  $E(Z) < \infty$  wtedy warunki (1), (2), (3), (4), (5), (6) są równoważne następującym warunkom:

- (7)  $E(d(X_t, A^*)) \rightarrow 0$ ,
- (8)  $E(f(X_t)) \searrow 0$ .

*Proof.* Zauważmy na początek, że zbieżność globalna w sensie naszej definicji oznacza zbieżność stochastyczną ciągu  $d(X_n, A^*)$  do zera. Ponieważ zbieżność prawie wszędzie jest silniejsza od stochastycznej a ta jest z kolei silniejsza od zbieżności słabej, mamy od razu 6)  $\Rightarrow$  5)  $\Rightarrow$  4) oraz 3)  $\Rightarrow$  2)  $\Rightarrow$  1). Przy naszych założeniach mamy również równoważność 3)  $\Leftrightarrow$  6) (wynikającą z własności (10.1)). Dodatkowo, 4)  $\Leftrightarrow$  5) ponieważ granica jest punktowa oraz 1)  $\Leftrightarrow$  2) na mocy Obserwacji 7. Ponieważ

$$\sup_{t \in \mathbb{N}} E|f(X_t)| = \sup_{t \in \mathbb{N}} Ef(X_t) \leq Ef(X_0) < \infty,$$

ciąg  $f(X_t)$  jest zbieżny prawie wszędzie na mocy twierdzenia Dooba o zbieżności nadmartyngałów. Zatem jeśli jest spełniony warunek (5) oraz wiemy z twierdzenia Dooba że istnieje granica prawie wszędzie, musi ona być równa granicy stochastycznej z warunku (5), co dowodzi implikację 5)  $\Rightarrow$  6). By zakończyć dowód, zauważmy, że z własności funkcji  $f$  pomiędzy (2) i (5) również mamy równoważność - wynika to z faktu, że funkcja  $f$  jest jednostajnie ciągła na pewnym zwartym otoczeniu  $A^*$  skąd, jak można pokazać, mamy:

- $\forall \delta > 0 \exists \varepsilon > 0 \{x \in X : d(x, A^*) < \varepsilon\} \subset \{x \in X : f(x) < \delta\}$
- $\forall \varepsilon > 0 \exists \delta > 0 \{x \in X : d(x, A^*) < \varepsilon\} \supset \{x \in X : f(x) < \delta\}$

Powyższe dwa warunki dowodzą 2)  $\Leftrightarrow$  5) w sposób oczywisty (trzeba skorzystać z definicji zbieżności stochastycznej). Zajmijmy się teraz warunkami 7) oraz 8) - są one silniejsze od zbieżności stochastycznej oraz, skoro mamy założone istnienie całkowalnej majoranty, wynikają ze zbieżności prawie wszędzie na mocy twierdzenia Lebesgue'a o zmajoryzowanym przejściu granicznym. Dowód twierdzenia jest zakończony.  $\square$

Teraz przedstawimy eleganckie warunki dostateczne dla zbieżności globalnej algorytmu  $X_t$  wyrażonego równaniem rekurencyjnym  $X_{t+1} = T(X_t, Y_t)$ .

**Twierdzenie 30.** Zakładamy, że funkcja  $f : X \rightarrow \mathbb{R}^+$  jest ciągła oraz że jej zbiory podpoziomowe  $\{x \in X : f(x) \leq \delta\}$  są zwarte dla każdego  $\delta > 0$ . Zakładamy, że algorytm spełnia następujące dwa warunki:

- (A) Dla każdego  $x \in X$ , funkcja  $f \circ T(x, \cdot)$  jest ciągła
- (C) Dla każdego  $x \in X \setminus A^*$ ,

$$E[f(X_{t+1}) | X_t = x] < f(x)$$

oraz dla każdego  $x \in A^*$ ,  $T(x, Y_t) \in A^*$ . Przy tych założeniach, algorytm  $X_t$  jest zbieżny globalnie.

Kluczowym warunkiem zbieżności jest warunek (C) mówiący, że jeśli algorytm nie jest w stanie optymalnym, jego oczekiwane położenie w kroku kolejnym będzie lepsze niż w kroku obecnym (lepsze w sensie funkcji  $f$ ). Gdybyśmy zastąpili ostrą nierówność warunku (C) nierównością

słabą, łatwo znaleźć kontrprzykład - przykładowo, algorytm mógłby utknąć w pewnym miejscu przestrzeni  $X$  (np. w minimum lokalnym funkcji  $f$ ). Wrócimy do tej sprawy niebawem.

**Słowo o układach dynamicznych.** Zanim przejdziemy do dowodu Twierdzenia 30 przywołamy kilka podstawowych pojęć z teorii układów dynamicznych. Dla przestrzeni metrycznej  $(M, d)$  oraz odwzorowania  $\varphi: M \rightarrow M$  parę  $(M, \varphi)$  będziemy nazywać **układem dynamicznym** (standardowo zakłada się że  $\varphi$  jest ciągle ale my będziemy pracować przy trochę ogólniejszych założeniach). Dla  $x \in M$  zbiór  $o(x) = \{x, \varphi(x), \varphi^2(x), \varphi^3(x), \dots\}$  będziemy nazywać **orbitą** punktu  $X$ . Niech  $K \subset M$  będzie niepustym zbiorem domkniętym który jest **niezmienniczy**, tzn.  $\varphi(K) \subset K$ . Powiemy, że  $K$  jest **globalnym atraktorem** gdy dla dowolnego  $x \in M$  ciąg  $\varphi^n(x)$  zbiega do zbioru  $K$ , czyli  $d(\varphi^n(x), K) \rightarrow 0$  gdy  $n \rightarrow \infty$ . Klasycznym narzędziem badania zbieżności do zbioru jest funkcja Lapunowa  $V: M \rightarrow \mathbb{R}^+$  i poniżej przedstawiam dogodną dla nas wersję twierdzenia Lapunowa z teorii układów dynamicznych.

**Twierdzenie 31.** *Niech  $(M, d)$  będzie przestrzenią metryczną,  $\varphi: M \rightarrow M$  będzie mierzalne oraz niech  $\emptyset \neq K \subset M$  będzie domknięty i niezmienniczy. Niech  $V: M \rightarrow \mathbb{R}^+$  będzie ciągła oraz taka, że  $V \circ \varphi: M \rightarrow \mathbb{R}^+$  jest ciągła oraz niech zachodzi:*

- (1)  $V(x) = 0 \Leftrightarrow x \in K$ ,
- (2)  $V(\varphi(x)) < V(x)$  for any  $x \in M \setminus K$ ,

*Jeśli orbita każdego punktu  $x \in M$  posiada punkt skupienia (czyli istnieje  $x_0 \in M$  spełniający  $\varphi^{t_k}(x) \rightarrow x_0$  dla pewnego podciągu  $t_k \rightarrow \infty$ ), wtedy  $V(\varphi^t(x)) \searrow 0$  as  $t \rightarrow \infty$  dla każdego  $x \in X$ . W szczególności, jeśli dla każdego  $x \in X$ ,  $V(\varphi^t(x)) \searrow 0$  implikuje  $d(\varphi^t(x), K) \rightarrow 0$ , wtedy  $K$  jest globalnym atraktorem*

*Proof.* Dla dowodu nie wprost założymy:

$$(10.3) \quad V(\varphi^t(x)) \searrow \delta \text{ dla pewnego } x \in X \text{ oraz } \delta > 0.$$

Istnieje  $t_k \nearrow \infty$  oraz  $x_0$  spełniający  $\varphi^{t_k}(x) \rightarrow x_0$ . Stąd oraz z (10.3):

$$V(\varphi^{t_k}(x)) \searrow V(x_0) = \delta > 0,$$

a zatem  $x_0 \notin K$  i w efekcie z warunku (2):  $V(\varphi(x_0)) < V(x_0)$ . Dodatkowo,  $V(\varphi^{t_k+1}(x)) \searrow \delta$ . Skoro  $V \circ \varphi$  jest ciągła oraz  $\varphi^{t_k}(x) \rightarrow x_0$ , mamy

$$V(\varphi^{t_k+1}(x)) = V(\varphi(\varphi^{t_k}(x))) \rightarrow V(\varphi(x_0)) < V(x_0) = \lim_{k \rightarrow \infty} V(\varphi^{t_k}(x)) = \delta,$$

sprzeczność. □

**Szkic dowodu Twierdzenia 30.** Sprowadzimy problem zbieżności do Twierdzenia 31 definiując odpowiednie obiekty  $M, \varphi, K, V$  oraz sprawdzając założenia Twierdzenia 31. Rozważmy  $M = \mathcal{M}^1(X)$  z metryką Prochorowa oraz  $\varphi(\mu) = P\mu$ , gdzie  $P$  to operator Markowa, zdefiniowany równaniem  $X_{t+1} = T(X_t, Y_t)$ . Definiujemy teraz zbiór

$$K := M^* = \{\mu \in M: \mu(A^*) = 1\}.$$

Pozostaje zdefiniować funkcję Lapunowa  $V: M \rightarrow \mathbb{R}^+$ . Gdyby funkcja  $f$  była ograniczona z góry, wystarczyłoby rozważyć funkcję  $\mu \rightarrow \int_X f d\mu$ . W ogólnym przypadku, definiujemy:

$$V(\mu) = \int_X \arctan(f(x)) \mu(dx).$$

Powyżej pod całką obłożyliśmy funkcję  $f$  funkcją  $\arctan$  tylko dlatego, funkcja  $\arctan$  jest ograniczona (będziemy pracować ze słabą zbieżnością) oraz ma właściwości nie psujące naszych



założeń o funkcji  $f$ , mianowicie jest ściśle rosnąca, ciągła, dodatnia i wklęsła. W szczególności, funkcja  $\arctan \circ f$  ma taki sam zbiór minimów globalnych jak funkcja  $f$ . Zauważmy teraz, że jeśli  $\mu = \mathbb{P}_{X_0}$ , wtedy  $\varphi^t(\mu) = \mathbb{P}_{X_t}$ . Zatem rozkłady naszego algorytmu stanowią trajektorię punktu  $\mu$ . Pokażemy, że nasz zbiór  $K$  jest globalnym atraktorem co na mocy Twierdzenia 29 jest równoważne globalnej zbieżności. Niezmienniczość zbioru  $K$  wynika bezpośrednio z założenia  $T(A^*, Y_t) \subset A^*$ . Funkcja  $\arctan \circ f$  jest ciągła i ograniczona co gwarantuje ciągłość funkcji  $V$ . Ciągłość funkcji  $V \circ \varphi$  wynika z założenia **(A)** (które jest jednak trochę słabsze niż ciągłość i ograniczoność funkcji  $\arctan \circ f$  i wymagałoby rozpisania (polecam jako ćwiczenie)). Funkcja nieujemna  $\arctan \circ f$  zeruje się tylko na zbiorze  $A^*$  skąd wynika, że  $V$  zeruje się tylko na miarach spełniających  $\mu(A^*) = 1$ . Założenie **(C)** oraz nierówność Jensena gwarantują, że zachodzi nierówność

$$E[\arctan \circ f(T(x, Y_t))] = E[\arctan \circ f(X_{t+1}) | X_t = x] < f(x) \text{ dla } x \notin A^*$$

co z kolei wymusza warunek (2) Twierdzenia 31 - jest to ćwiczenie na zastosowanie twierdzenia Fubiniego ponieważ mamy, pracując na postaci rekurencyjnej łańcucha,

$$V(P\mu) = \int_X \arctan \circ f dP\mu = \int_X \int_B V(T(x, y)) \mathbb{P}_{Y_t}(dy) \mu(dx),$$

gdzie  $B$  to przestrzeń stanów ciągu  $Y_t$ . Pozostaje pokazać, że ciąg  $\mathbb{P}_{X_t}$  posiada punkt skupienia. Na mocy twierdzenia Prochorowa wystarczy pokazać, że rodzina  $P^t\mu = \mathbb{P}_{X_t}$  jest ciasna. Ustalmy  $\varepsilon > 0$ . Z nierówności Czebyszewa oraz z warunku  $E[\arctan \circ f(X_{t+1})] \leq E[\arctan \circ f(X_t)]$  (ciąg  $\arctan \circ f(X_t)$  jest nadmartyngałem na mocy Lematu 4 zastosowanego do funkcji  $\arctan \circ f$ ), dla każdego  $\delta > 0$  mamy

$$\mathbb{P}(\arctan \circ f(X_t) > \delta) \leq \frac{E[\arctan \circ f(X_t)]}{\delta} \leq \frac{E[\arctan \circ f(X_0)]}{\delta}, \quad t \in \mathbb{N}$$

a zatem

$$\mathbb{P}_{X_t}(\{x: \arctan \circ f(x) \leq \delta\}) = \mathbb{P}(\arctan \circ f(X_t) \leq \delta) \geq 1 - \frac{E[V(X_0)]}{\delta}, \quad t \in \mathbb{N}.$$

Funkcja  $\arctan \circ f$  ma zwarte zbiory podpoziomicowe oraz z powyższej równości mamy, że dla dowolnego  $\varepsilon > 0$  znajdzie  $\mathbb{P}_{X_t}(\{x: \arctan \circ f(x) \leq \delta\}) > 1 - \varepsilon$  dla dostatecznie dużego  $\delta$ , a zatem na mocy twierdzenia Prochorowa ciąg  $\mathbb{P}_{X_t} = P^t\mu$  ma punkt skupienia w zbiorze  $M$ . Twierdzenie 31 gwarantuje nam  $V(P^t\mu) = E[\arctan \circ f(X_t)] \searrow 0$  co wymusza  $X_t \xrightarrow{s} A^*$  i kończy dowód ponieważ rozkład początkowy był wybrany dowolnie.  $\square$

Zastosujemy teraz powyższe twierdzenie do Algorytmu 1. W tym celu zdefiniujemy algorytm  $X_t$  bardziej formalnie. Niech  $B$  będzie przestrzenią metryczną ośrodkową, niech ciąg  $Y_n: \Omega \rightarrow B$  będzie i.i.d. oraz niech  $Q: X \times B \rightarrow X$  będzie **ciągłe**. Definiujemy:

$$X_{t+1} = T(X_t, Y_t) := \begin{cases} Q(X_t, Y_t) & \text{jeśli } f(Q(X_t, Y_t)) < f(X_t) \\ X_t & \text{jeśli } f(Q(X_t, Y_t)) \geq f(X_t) \end{cases}.$$

Możemy teraz sformułować i udowodnić:

**Twierdzenie 32.** *Zakładamy, że funkcja  $f: X \rightarrow \mathbb{R}^+$  jest ciągła oraz że jej zbiory podpoziomicowe  $\{x \in X: f(x) \leq \delta\}$  są zwarte dla każdego  $\delta > 0$ . Jeśli dla każdego  $x \in X \setminus A^*$*

$$(10.4) \quad \mathbb{P}[f(Q(x, Y_t)) < f(x)] > 0$$

*wtedy Algorytm 1 jest zbieżny globalnie.*

*Proof.* Wystarczy pokazać, że spełnione są założenia **A)** oraz **C)** Twierdzenia 30. Założenie **A)** jest spełnione natychmiastowo ponieważ

$$f \circ T(x, y) = \min\{f(x), Q(x, y)\}$$

jest funkcją ciągłą jako złożenie funkcji ciągłych. Warunek **C)** jest spełniony natychmiastowo ponieważ

$$f(T(x, Y_t)) \leq f(x) \text{ oraz } \mathbb{P}[f(T(x, Y_t)) < f(x)] > 0 \text{ dla } x \notin A^*$$

dzięki definicji algorytmu oraz warunkowi (10.4), co oznacza:

$$f(X_{t+1}) \leq f(X_t) \text{ oraz } \mathbb{P}(f(X_{t+1}) < f(x) | X_t = x) > 0 \text{ dla } x \notin A^*.$$

□

Tak więc widzimy że warunkiem dostatecznym zbieżności Algorytmu 1 jest możliwość dostania się do regionu o mniejszych wartościach funkcji  $f$  z dowolnego stanu, który nie jest optymalny. W szczególności algorytm ma możliwość z dodatnim prawdopodobieństwem wydostać się z minimum lokalnego, które w pewnym otoczeniu ma najmniejszą wartość funkcji  $f$  i z tego powodu może stanowić pułapkę dla niektórych metod optymalizacji.

A co z tempem zbieżności? Problem jest bardzo trudny do analizy teoretycznej przy założeniach naturalnych i dominują eksperymenty numeryczne. Tempo zbieżności zależy zarówno od doboru konkretnej metody optymalizacji i jej parametrów, jak i od własności optymalizowanej funkcji  $f$  oraz wymiarowości jej dziedziny. Tematyka jest wciąż rozwijana od strony teoretycznej i praktycznej (numerycznej). Znanie nam oprogramowanie R posiada pakiety do optymalizacji funkcji przy użyciu takich metod jak algorytmy ewolucyjne lub algorytm Simulated Annealing. Prześlę coś dodatkowego do poczytania w ramach materiałów uzupełniających.

## 11. IFS ORAZ FRAKTALE

Rozważmy teraz specjalną klasę łańcuchów Markowa - IFS (Iterated Function System). Dla prostoty zakładamy  $X = \mathbb{R}^k$ . Załóżmy, że mamy  $n$  odwzorowań mierzalnych  $S_i: X \rightarrow X$ ,  $i = 1, \dots, n$ . Jeśli łańcuch jest w stanie  $x$ , wtedy losujemy jedno z odwzorowań (zakładamy, że każde ma dodatnie prawdopodobieństwo) i przechodzimy do kolejnego stanu  $S_j(x)$ , gdzie  $S_j$  to wylosowane odwzorowanie. Bardziej formalnie, zakładamy następującą postać rekurencyjną:

$$(11.1) \quad X_{n+1} = S_{Y_n}(X_n),$$

gdzie  $Y_n: \Omega \rightarrow \{1, 2, \dots, n\}$  to mierzalny ciąg i.i.d. Równanie (11.1) definiuje IFS. Rozkład zmiennej  $Y_1$  będziemy utożsamiać z wektorem stochastycznym  $(p_1, \dots, p_n)$ , gdzie  $p_i > 0, i = 1, \dots, n$ . Łatwo wypisać jądro przejścia łańcucha (11.1):

$$P(x, \cdot) = \sum_{i=1}^n p_i \delta_{S_i(x)}.$$

W całej sekcji będziemy zakładać że odwzorowania  $S_i$  są Lipschitzowskie ze stałymi  $L_i$ :

$$|S_i(x) - S_i(z)| \leq L_i |x - z|.$$

**Propozycja 8.** (1) *Jeśli  $\sum_{i=1}^n p_i L_i < 1$ , wtedy operator Markowa  $P$  łańcucha (11.1) jest silną kontrakcją na miarach probabilistycznych z metryką Forter-Mouriera  $\|\cdot\|_F$ , tzn. dla pewnej stałej  $L < 1$  zachodzi*

$$\|P\mu_1 - P\mu_2\|_F \leq L \|\mu_1 - \mu_2\|_F.$$

- (2) Z powyższego, ponieważ przestrzeń  $(\mathcal{M}^1(X), \|\cdot\|_F)$  jest zupełna, z twierdzenia Banacha o punkcie stałym dostajemy, że operator  $P$  jest **ślabo asymptotycznie stabilny**, tzn. istnieje dokładnie jedna miara stacjonarna  $\pi \in \mathcal{M}^1(X)$  spełniająca  $P^n\mu \rightarrow \pi$  w słabej topologii dla dowolnego rozkładu startowego  $\mu \in \mathcal{M}^1(X)$ .

Od teraz skupimy się na IFS-ach **hiperbolicznych**, tzn. spełniających  $L_i < 1$ ,  $i = 1, \dots, n$ . Naszym celem będzie analiza zbioru  $A^* = \text{supp}(\pi)$  będącego nośnikiem miary stacjonarnej  $\pi$ . Przywołajmy definicję: zbiór  $K$  nazywamy **nośnikiem** miary  $\mu \in \mathcal{M}^1(X)$  i oznaczamy  $\text{supp}(\mu)$  gdy jest on najmniejszym zbiorem domkniętym  $H$  spełniającym  $\mu(H) = 1$  (definicja ta jest poprawna i nośnik miary jest wyznaczony jednoznacznie jako przecięcie zbiorów domkniętych na których dana miara jest skupiona).

**Operatorem Barnsley'a**, skojarzonym z operatorem  $P$ , nazywamy funkcję przetwarzającą podzbiory  $X$  następująco:

$$F: \mathcal{P}(X) \ni A \longrightarrow \bigcup_{i=1}^n S_i(A) \in \mathcal{P}(X)$$

Łatwo zauważyć, że jeśli  $A$  jest zbiorem zwartym, zbiór  $F(A)$  jest również zwarty. Zgodnie z poniższym twierdzeniem, odwzorowanie  $F$  mówi nam w jaki sposób zmieniają się nośniki kolejnych rozkładów  $\mu, P\mu, P^2\mu, \dots$  oraz jaką ma postać nośnik miary stacjonarnej. Przypomnijmy, że pracujemy przy założeniu hiperboliczności, czyli odwzorowania  $S_i$  definiujące operator Barnsley'a są silnymi kontrakcjami.

**Twierdzenie 33.** *Jeśli nośnik rozkładu startowego  $\text{supp}(\mu_0)$  jest zwarty, wtedy*

$$\text{supp}(\mu_{n+1}) = F(\text{supp}(\mu_n)).$$

*W szczególności, nośnik  $A^*$  miary stacjonarnej  $\pi$  jest niezmienniczy względem operatora Barnsley'a, tzn.*

$$F(A^*) = A^*.$$

*Dodatkowo, niezależnie od wyboru zwartego nośnika startowego  $A_0$ , zbiór  $A^*$  jest zwarty oraz spełnia*

$$A^* = \lim_{n \rightarrow \infty} F^n(A_0),$$

*gdzie powyższa granica jest granicą w sensie metryki Hausdorffa  $d_H$ , która mierzy odległość pomiędzy zwartymi podzbiorem  $X$  zgodnie z równością:*

$$d_H(C, D) = \inf\{r > 0: \forall x \in C \exists y \in D |x - y| < r \wedge \forall y \in D \exists x \in C |y - x| < r\}.$$

**Przykład 26.** Rozważmy zbiory  $A_n = \{\frac{1}{2^n}, \frac{2}{2^n}, \frac{3}{2^n}, \dots, \frac{2^n-1}{2^n}\}$ . Zachodzi zbieżność  $A_n \rightarrow [0, 1]$  w metryce Hausdorffa (zbiory dwójkowo wymierne  $A_n$  coraz dokładniej pokrywają odcinek  $[0, 1]$ ).

**Przykład 27.** Twierdzenie 33 nie zachodzi bez założenia hiperboliczności. Rozważmy system zadany przez dwa odwzorowania na  $\mathbb{R}$ :  $S_1(x) = x$  oraz  $S_2(x) = 0$ , gdzie  $p_1 > 0, p_2 > 0$ . Na mocy Propozycji 8 operator  $P$  jest ślabo asymptotycznie stabilny (łatwo widać, że  $\pi = \delta_0$ ) ale przykładowo dla zbioru  $A_0 = [0, 1]$  mamy  $F^n(A_0) = A_0$  co nie jest zbieżne do  $A^* = \{0\}$ .

Używając Twierdzenia 33 można łatwo konstruować ciekawe zbiory, mające niecałkowity wymiar Hausdorffa, co jest cechą charakteryzującą **fraktale**. Poniżej zilustruję kilka przykładów, wykorzystując rysunki z pozycji 2 z literatury uzupełniającej, którą podaję na końcu wykładu.

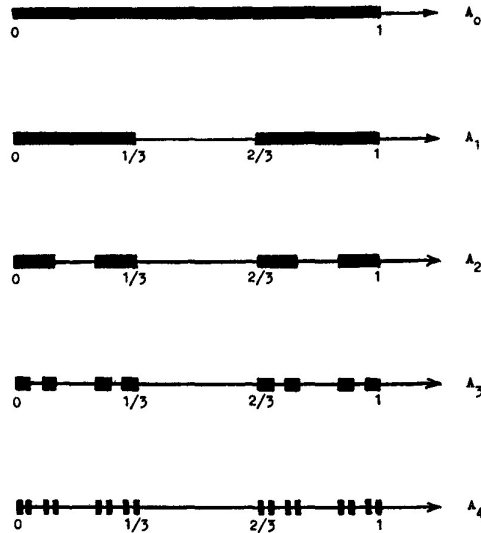
**Przykład 28. Zbiór Cantora.**

Rozważmy  $X = \mathbb{R}$ ,  $A_0 = [0, 1]$  oraz  $S_1(x) = \frac{1}{3} \cdot x$  oraz  $S_2(x) = \frac{1}{3} \cdot x + \frac{2}{3}$ . Łatwo policzyć:

$$A_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1],$$

$$A_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{3}{9}] \cup [\frac{6}{9}, \frac{7}{9}] \cup [\frac{8}{9}, 1],$$

ale przyjemniej obejrzeć na rysunku:



Powyżej oczywiście  $A_n = F^n(X_0)$ . Na mocy dotychczasowych rozważań, zbiory  $A_n$  zbiegają do jedynego zbioru zwartego spełniającego  $F(A^*) = A^*$ , w tym przypadku do zbioru Cantora.

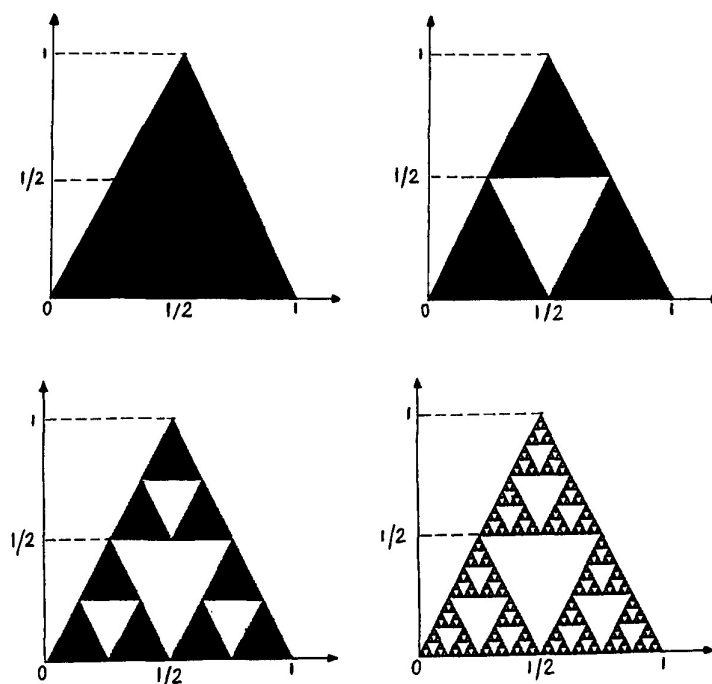
**Przykład 29.** Niech  $X = \mathbb{R}^2$  oraz

$$S_1(x_1, x_2) = \frac{1}{2}(x_1, x_2),$$

$$S_2(x_1, x_2) = \left(\frac{1}{2}x_1 + \frac{1}{2}, \frac{1}{2}x_2\right)$$

$$S_3(x_1, x_2) = \left(\frac{1}{2}x_1 + \frac{1}{4}, \frac{1}{2}x_2 + \frac{1}{2}\right)$$

W tym przykładzie atraktorem spełniającym  $F(A^*) = A^*$  jest **trójkąt Sierpińskiego**. Niech  $A_0$  będzie trójkątem zwanym o wierzchołkach w punktach  $(0, 0)$ ,  $(1, 0)$ ,  $(\frac{1}{2}, 1)$ . Poniżej zbiory  $A_0, A_1, A_2, A_3$ , gdzie  $A^n = F^n(A_0)$

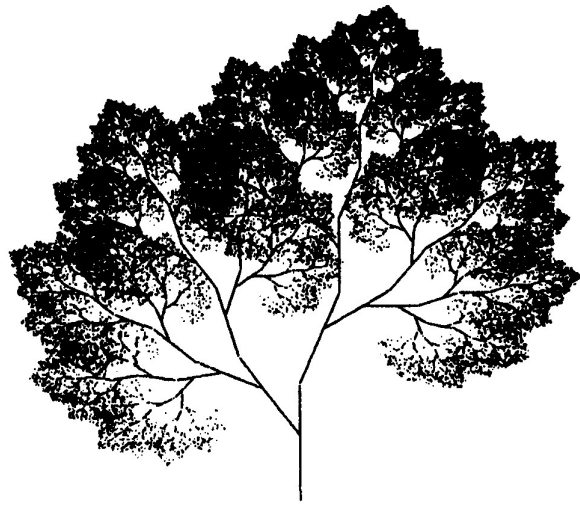


Chcąc otrzymać dokładniejszy obraz zbioru granicznego powinniśmy oczywiście rozważyć większą wartość  $n$ . Powyższe dwie konstrukcje były deterministyczne - operator Barnsley'a nie zależy od prawdopodobieństw  $p_1, \dots, p_n$ . Gdy jednak będziemy używać łańcucha Markowa w iteracyjnej procedurze tworzenia obrazu zbioru  $A^*$ , efekt końcowy nie będzie już jednolicie czarno- biały, ponieważ rozmieszczenie wygenerowanych punktów będzie związane z rozkładem naszej miary granicznej  $\pi$  na zbiorze  $A^*$ .

**Przykład 30.** Rozważmy bardziej skomplikowany przykład, w którym  $X = \mathbb{R}^2$  oraz  $S_1, S_2, S_3$  są dane wzorami:

$$\begin{aligned}
 S_1(x) &= \begin{pmatrix} 0 & 0 \\ 0 & c \end{pmatrix} x + \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} \\
 S_2(x) &= \begin{pmatrix} -r \cos \varphi & -r \sin \varphi \\ -r \sin \varphi & r \cos \varphi \end{pmatrix} x + \begin{pmatrix} \frac{1}{2} + \frac{r}{2} \cos \varphi \\ c + \frac{r}{2} \sin \varphi \end{pmatrix} \\
 S_3(x) &= \begin{pmatrix} q \cos \psi & -r \sin \psi \\ q \sin \psi & r \cos \psi \end{pmatrix} x + \begin{pmatrix} \frac{1}{2} - \frac{q}{2} \cos \psi \\ \frac{3}{5}c - \frac{q}{2} \sin \psi \end{pmatrix},
 \end{aligned}$$

Dla stałych  $c = 0.255$ ,  $r = \frac{3}{4}$ ,  $q = 0.625$ ,  $\varphi = -\frac{\pi}{8}$ ,  $\psi = \frac{\pi}{5}$  oraz prawdopodobieństw  $p_1 = p_2 = p_3 = \frac{1}{3}$  można uzyskać rysunek jak poniżej. Czy coś Ci przypomina?



**Notacja:** Dla zmiennych losowych  $X, Y$  określonych na prz. probabilistycznej  $(\Omega, \Sigma, \mathbb{P})$  piszemy  $X = Y$  gdy  $\mathbb{P}(X = Y) = 1$ .

skrót: c.n. przeliczalne = co najwyżej przeliczalne

skrót: i.i. d. - independent identically distributed

Poniżej różne sposoby zapisu całki z funkcji  $f$  względem miary  $\mu$ :

$$\int_X f d\mu = \int_X f(x)\mu(dx) = \int_X f(x)d\mu(x).$$

$B(X, \mathbb{R})$  - funkcje mierzalne ograniczone

$BC(X, \mathbb{R})$  - funkcje ciągłe ograniczone

$\delta_x$  - delta Diraca w punkcie  $x$

$\delta Z$  - brzeg zbioru  $Z$

$E_\nu[f]$  - wartość oczekiwana z funkcji  $f$  względem miary  $\nu$

$K(X) = \{P: X \times \mathcal{B}(X) \rightarrow [0, 1] | P - \text{jądło przejścia}\}$ .

$M(X, A)$  - funkcje mierzalne  $f: X \rightarrow A$

$\mathcal{M}(X)$  - miary skończone ze znakiem na przestrzeni mierzalnej  $X$

$\mathcal{M}^+(X)$  - miary skończone (nieujemne) na przestrzeni mierzalnej  $X$

$\mathcal{M}^1(X)$  - miary probabilistyczne na przestrzeni mierzalnej  $X$

$\mu$  - miara Lebesgue'a

$\mathbb{P}_X$  - rozkład prawdopodobieństwa zmiennej losowej  $X$

#### **Literatura uzupełniająca:**

##### 1. Literatura Przystępna:

Rozdział 12 w podręczniku: Jakubowski, Sztencel, Wstęp do Rachunku Prawdopodobieństwa

##### 2. Literatura dość przystępna:

Rozdziały 2 oraz 10, 12 w podręczniku: A. Lasota, M. Mackey, Chaos, Fractals and Noise, Springer Verlag, New York, 1994.

##### 3. Literatura dla osób nie znających strachu oraz dla doktorantów:

a) S. Meyn, R. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, London, 1993

b) Randal Douc, Eric Moulines, Pierre Priouret, Philippe Soulier, Markov Chains, Springer, 2018