

# RACHUNEK PRAWDOPODOBIENSTWA 2B

2024

DARIUSZ BURACZEWSKI, PIOTR DYSZEWSKI

---

# Sylabus

## Dane dotyczące przedmiotu

Nazwa przedmiotu: Rachunek prawdopodobieństwa 2b

Jednostka oferująca przedmiot: Instytut Matematyczny

Grupu: Podstawowe przedmioty do wyboru dla spec. teoretycznej (T), Zaawansowane przedmioty do wyboru (M)

Założenia: Teoria prawdopodobieństwa 1 (28-MT-S-tTPrawd1)

Strona www: <https://sites.google.com/site/piotrdyszewski/teaching/rp2r23>

Forma zajęć: wykład + ćwiczenia

Punkty ECTS: 6

## Opis przedmiotu

Celem wykładu jest dokonanie przeglądu najważniejszych procesów stochastycznych w czasie dyskretnym oraz ciągłym wraz z opisem ich podstawowych własności. Zaczniemy od klasy procesów stochastycznych zwanych martyngałami, które modelują wartość uczciwej gry. Jest to klasa procesów szczególnie istotna z punktu widzenia matematyki finansowej. Następnie przeanalizujemy ich związki z łańcuchami Markowa, czyli procesami stochastycznymi które charakteryzują się krótką pamięcią. Są one stosowane w algorytmach probabilistycznych. W ostatniej części wykładu skoncentrujemy się na ruchu Browna. Jest to odpowiednik rozkładu normalnego dla procesów stochastycznych. W skutek czego pojawia się opisach wielu granic spotykanych w zaawansowanej teorii prawdopodobieństwa. Jeżeli pozwoli na to czas, to na ostatnim wykładzie omówimy pokrótce proces Poissona i procesy Lévy'ego. Jednakże aby prawidłowo zbadać wszystkie wymienione wyżej procesy niezbędna jest znajomość warunkowej wartości oczekiwanej, dlatego to właśnie od niej zaczniemy wykład.

Do opanowania podstaw wykładu niezbędny będzie teorio-miarowy rachunek prawdopodobieństwa (Rachunek prawdopodobieństwa 1B). Bardzo pomocne, choć nie niezbędne, będą podstawowe informacje z wykładu z teorii miary.

Podczas zajęć poruszymy zagadnienia takie jak:

- A) **warunkowa wartość oczekiwana:** definicja, podstawowe własności, przykłady i zastosowania;
- B) **martyngały:** definicja i podstawowe własności, twierdzenia o zbieżności, twierdzenia o zatrzymaniu, martyngały całkowalne z kwadratem;
- C) **łańcuchy Markowa:** podstawowe własności, rekurencyjność, twierdzenia graniczne;
- D) **ruch Browna:** konstrukcja, własność Markowa, twierdzenie o zatrzymaniu, własności trajektorii;
- E) **procesy Lévy'ego:** Proces Poissona, rozkłady stabilne, rozkłady nieskończenie podzielne, definicja i własności procesów Lévy'ego

Do wykładu przygotowane są notatki, które będą dostępne na stronie prowadzącego. Podstawową literaturą do wykładu jest:

- R. Durrett, *Probability: theory and examples*. Cambridge university press
- P. Billingsley, *Probability and measure*. John Wiley & Sons

Na literaturę uzupełniającą składają się

- S. Resnick, *Adventures in stochastic processes*, Springer Science & Business Media,
- P. Morters, Y. Peres, *Brownian motion*, Cambridge University Press

## Szczegółowy plan wykładu

1. Definicja warunkowej wartości oczekiwanej
2. Własności Warunkowej wartości oczekiwanej
3. Interpretacja geometryczna i zagadnienie regresji
4. Definicja martyngału
5. Czasy zatrzymania, twierdzenie o zatrzymaniu
6. Twierdzenie o zbieżności martyngau
7. Martyngały całkowalne z kwadratem
8. łańcuchy Markowa
9. Macierz przejścia, tranzytywność
10. rozkłady stacjonarne
11. Twierdzenie graniczne, Markov chain Monte Carlo
12. Definicja i konstrukcja ruchu Browna
13. własności probabilistyczne ruchu Browna,

14. własności trajektorii ruchu Browna, mocna własność Markowa, zasada odbicia
15. Proces Poissona, procesy Lévy'ego

## **Efekty kształcenia**

Po wykładzie student:

1. Formułuje definicję warunkowej wartości oczekiwanej i wyznacza ją w przypadku przeliczalnie generowanego  $\sigma$ -ciała
2. Formułuje definicję i podstawowe własności martyngału
3. Formułuje definicję łańcucha Markowa
4. Formułuje definicję ruchu Browna
5. Weryfikuje własność martyngału w przykładach
6. Formułuje i wykazuje podstawowe własności warunkowej wartości oczekiwanej
7. Weryfikuje własność Markowa w podstawowych przykładach
8. Podaje i dowodzi podstawowe własności probabilistyczne ruchu Browna
9. Wyznacza warunkową wartość oczekiwaną w bardziej złożonych przykładach
10. Stosuje twierdzenia graniczne dla martyngałów
11. Stosuje mocną własność Markowa oraz twierdzenia graniczne dla łańcuchów Markowa.
12. Dowodzi własności trajektorii ruchu Browna

## **Sposób weryfikacji efektów kształcenia**

Na zaliczenie składać się będą dwa sprawdziany pisemne, aktywność na ćwiczeniach oraz zadania dodatkowe. Ocena z wykładu wystawiona będzie na podstawie egzaminu pisemnego.

## **Metody i kryteria oceniania**

Warunkiem dostatecznym zaliczenia przedmiotu jest:

- uzyskanie 50% punktów za zadania stanowiące bieżącą weryfikację efektów kształcenia (sprawdziany pisemne oraz aktywność na zajęciach)
- uzyskanie pozytywnej oceny z egzaminu stanowiącego końcową weryfikację efektów kształcenia

Kryteria ocen:

dst student realizuje punkty 1-5 efektów kształcenia

db student realizuje punkty 1-8 efektów kształcenia

bdb student realizuje punkty 1-12 efektów kształcenia

Wrocław, październik 2023

*Piotr Dyszewski*

---

# Spis treści

<b>1</b>	<b>Warunkowa wartość oczekiwana</b> .....	<b>1</b>
1.1	Definicja warunkowej wartości oczekiwanej .....	2
1.2	Prawdopodobieństwo warunkowe .....	10
1.3	Istnienie warunkowej wartości oczekiwanej .....	11
1.4	Podstawowe własności .....	12
1.5	Interpretacja geometryczna. Zagadnienie regresji .....	16
1.6	Regularne rozkłady warunkowe .....	19
<b>2</b>	<b>Martyngały</b> .....	<b>23</b>
2.1	Definicje i przykłady .....	23
2.2	Czasy zatrzymania .....	28
2.3	Twierdzenia Doob'a .....	31
2.4	Twierdzenia o zbieżności martyngałów .....	36
2.5	Jednostajna całkowalność .....	45
2.6	Zbieżność martyngałów w $L^1$ .....	48
2.7	Nierówności maksymalne .....	49
<b>3</b>	<b>łańcuchy Markowa</b> .....	<b>54</b>
3.1	Przykłady .....	54
3.2	Definicje .....	58
3.3	łańcuchy powracające i chwilowe .....	65
3.3.1	Rekurencyjne spacery losowe na grafach .....	76
3.4	Twierdzenia o zbieżności .....	80
3.4.1	Tempo zbieżności na skończonej przestrzeni stanów ....	84
3.4.2	Przykłady .....	87
3.5	MCMC .....	88
3.5.1	Drzewo rozpinające .....	89
3.5.2	Problem pakowania plecaka (The knapsack problem)....	90
3.5.3	Algorytm Metropolis .....	91
3.6	Definicja ogólna .....	95

<b>4</b>	<b>Ruch Browna</b> .....	98
4.1	Motywacja i konstrukcja .....	98
4.2	Własności probabilistyczne ruchu Browna .....	106
4.3	Trajektorie ruchu Browna .....	109
<b>5</b>	<b>Procesy Lévy'ego</b> .....	115
5.1	Proces Poissona .....	115
5.2	Rozkłady stabilne .....	120
5.3	Procesy Lévy'ego i rozkłady nieskończenie podzielne .....	125



## Warunkowa wartość oczekiwana

**Streszczenie** Wprowadzimy pojęcie prawdopodobieństwa warunkowego względem  $\sigma$ -ciała. Zobaczymy w jaki sposób pojęcie to uogólnia warunkową wartość oczekiwaną względem zdarzenia. Po zbadaniu podstawowych własności zastosujemy warunkową wartość oczekiwaną do zagadnienia regresji.

Przyjmować będziemy, że  $(\Omega, \mathcal{F}, \mathbb{P})$  jest przestrzenią probabilistyczną na której określone są wszystkie rozważane zmienne losowe. Przypomnijmy z podstawowego kursu rachunku prawdopodobieństwa, że dla zdarzenia  $A \in \mathcal{F}$  dla którego  $\mathbb{P}[A] > 0$  prawdopodobieństwo warunkowe  $\mathbb{P}[\cdot | A]$  definiujemy w sposób jawny wzorem

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

dla  $B \in \mathcal{F}$ . Wartość  $\mathbb{P}[B | A]$  jest interpretowana jako prawdopodobieństwo zajścia zdarzenia  $B$  jeżeli *wiemy* że zaszło zdarzenie  $A$ . Co się dzieje w przypadku, gdy nie mamy takiej pewności? W takim wypadku na myśl przychodzi wyrażenie

$$\mathbb{1}_A \mathbb{P}[B | A] + \mathbb{1}_{A^c} \mathbb{P}[B | A^c].$$

Celem niniejszego rozdziału jest nadanie matematycznego rygoru powyższemu wyrażeniu. Zauważmy bowiem, że jest to zmienna losowa. Dodatkowo zbadamy co się dzieje w przypadku, gdy chcemy warunkować większą liczbą zdarzeń co do których nie mamy pewności.

Podobnie dla całkowalnej zmiennej losowej  $X$  (czyli takiej, że  $\mathbb{E}[|X|] < \infty$ ) wartość oczekiwaną  $X$  pod warunkiem zdarzenia  $A$  definiujemy przez

$$\mathbb{E}[X | A] = \int_{\Omega} X(\omega) \mathbb{P}[d\omega | A] = \frac{1}{\mathbb{P}[A]} \mathbb{E}[X \mathbb{1}_A].$$

W dokładnej analogii do prawdopodobieństwa,  $\mathbb{E}[X | A]$  jest średnią wartością  $X$  jeżeli wiemy, że zaszło zdarzenie  $A$ . Rozważmy następujący przykład.

### Przykład 1.1

Założmy, że liczba wypadków jednego dnia w danym mieście ma rozkład Poissona z parametrem 6. Wysokość szkody ma rozkład jednostajny na przedziale  $[2, 10]$ . Niech  $X$  będzie łączną szkodą danego dnia. Zastanówmy się ile wynosi  $\mathbb{E}[X]$ ? Niech  $Y$  oznacza liczbę wypadków danego dnia. Wówczas  $Y$  ma rozkład Poissona z parametrem 6. Zauważmy, że

$$\mathbb{E}[X | Y = k] = 6k.$$

Stąd, stosując wzór na prawdopodobieństwo całkowite,

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}[Y = k] \mathbb{E}[X | Y = k] = 6 \sum_{k=0}^{\infty} k \mathbb{P}[Y = k] = 6\mathbb{E}[Y] = 36.$$

Wróćmy do dowolnej zmiennej losowej  $X$  i dowolnego zdarzenia  $A$ . Co możemy powiedzieć o średniej wartości  $X$  jeżeli nie mamy pewności czy zaszło zdarzenie  $A$ ? Analogicznie jak w przypadku prawdopodobieństwa można zaproponować

$$\mathbb{1}_A \mathbb{E}[X | A] + \mathbb{1}_{A^c} \mathbb{E}[X | A^c].$$

Ponownie mamy do czynienia ze zmienną losową. W tym rozdziale zajmujemy się matematycznym doprecyzowaniem i uzasadnieniem powyższego wyboru. Dokładniej rozważać będziemy warunkową wartość oczekiwaną względem całej kolekcji zdarzeń o zachodzeniu których nie mamy pewności. Zauważmy, że jeżeli  $X = \mathbb{1}_B$ , to  $\mathbb{E}[X | A] = \mathbb{P}[B | A]$ , więc pojęcie wartości oczekiwanej jest ogólniejsze i to na niej skupimy się przez większość tego rozdziału.

## 1.1 Definicja warunkowej wartości oczekiwanej

Przejdziemy teraz do uogólnienia warunkowej wartości oczekiwanej na większą liczbę zdarzeń o zachodzeniu których nie mamy pewności. Zamiast jawnej konstrukcji podamy jednak pewną szczególną cechę, która będzie charakteryzować warunkową wartość oczekiwaną. Opis ten będzie dla nas wygodniejszy przy analizowaniu własności warunkowej wartości oczekiwanej. Niestety jest on na pierwszy rzut oka mało zachęcający. Zaczniemy więc od następującego przykładu.

### Przykład 1.2

Rozważmy zmienną losową  $Z$  przyjmującą przeliczalnie wiele wartości. Niech  $X$  będzie dowolną całkowalną zmienną losową. Wówczas dla każdego  $z \in \mathbb{R}$  takiego, że  $\mathbb{P}[Z = z] > 0$ ,

$$\mathbb{E}[X | Z = z] = \frac{1}{\mathbb{P}[Z = z]} \mathbb{E} \left[ X \mathbb{1}_{\{Z=z\}} \right].$$

Zastanówmy się teraz jak scharakteryzować funkcję

$$h(z) = \begin{cases} \mathbb{E}[X | Z = z], & \mathbb{P}[Z = z] > 0, \\ 0 & \text{poza tym.} \end{cases}$$

Zauważmy, że dla każdego  $B \in \mathcal{B}or(\mathbb{R})$  zachodzi

$$\begin{aligned} \int_{\{Z \in B\}} h(Z(\omega)) \mathbb{P}(d\omega) &= \sum_{z \in B} h(z) \mathbb{P}[Z = z] = \sum_{z \in B} \mathbb{E} \left[ X \mathbb{1}_{\{Z=z\}} \right] \\ &= \mathbb{E} \left[ X \mathbb{1}_{\{Z \in B\}} \right] = \int_{\{Z \in B\}} X(\omega) \mathbb{P}(d\omega). \end{aligned}$$

Pamiętajmy, że wszystkie sumy w powyższym wyliczeniu mają co najwyżej przeliczalnie wiele niezerowych składników. Zmienna losowa  $h(Z)$  jest oczywiście  $\sigma(Z)$ -mierzalna. Dodatkowo całka po każdym zdarzeniu  $F \in \sigma(Z)$  z  $h(Z)$  jest taka sama jak wartość całki z  $X$ . Rzeczywiście, wynika to z powyższego wyliczenia ponieważ  $F$  jest postaci  $F = \{Z \in B\}$  dla pewnego  $B \in \mathcal{B}or(\mathbb{R})$ . Mamy więc

$$\mathbb{E}[h(Z) \mathbb{1}_F] = \mathbb{E}[X \mathbb{1}_F]$$

dla każdego  $F \in \sigma(Z)$ . Jak się niebawem okaże taka własność ma daleko idące konsekwencje. Zauważmy na koniec, że dla  $F = \Omega$  otrzymujemy

$$\mathbb{E}[h(Z)] = \mathbb{E}[X].$$

Należy mieć na względzie, że w kontekście ostatniego przykładu, próba rozpisania  $h(Z)$  w oparciu o definicję funkcji  $h$  może przyprawić o zawrót głowy. Kończymy bowiem z

$$h(Z(\omega)) = \mathbb{E}[X | \{\omega' \in \Omega : Z(\omega') = Z(\omega)\}] = \mathbb{E} \left[ X \mid Z^{-1}[Z(\omega)] \right].$$

Zapis skrótowy stosowany często w probabilistyce jest w tym przypadku nieodpowiedni, ponieważ prowadzi do

$$h(Z) = \mathbb{E}[X | Z = Z].$$

Źródła dysonansu należy szukać w fakcie, że  $Z$  pojawia się tutaj w dwóch charakterach. W napisie  $\mathbb{E}[X | Z = z]$  zmienna  $Z$  ma znaczenie teoretyczne. Innymi słowy gdyby hipotetycznie wiedzieć, że  $Z$  będzie miało wartość  $z$ , to oczekiwana wartość  $X$  wyniesie  $\mathbb{E}[X | Z = z]$ . Natomiast w  $h(Z)$  mamy do czynienia z konkretną realizacją  $Z$ . Inaczej mówiąc z wartością  $Z$ , którą rzeczywiście obserwujemy. Konstrukcje podobne do tej z **Przykładu 1.2** pojawiają się jeszcze wielokrotnie w trakcie wykładu, więc dla pewności pozwolimy sobie zilustrować powyższą sytuację przykładem.

### Przykład 1.3

Powiedzmy, że wykonujemy następujący eksperyment. Ze zbioru  $\{1, \dots, 10\}$  losujemy jednostajnie jedną liczbę  $N$ . Następnie spośród liczb  $\{1, \dots, N\}$  losujemy jednostajnie drugą liczbę  $M$ . Jaka jest wartość oczekiwana  $M$ ? Dla  $n \in \{1, \dots, 10\}$  rozważmy

$$h(n) = \mathbb{E}[M \mid N = n] = (n + 1)/2.$$

Wobec tego  $h(N) = (N + 1)/2$ . Średnią  $M$  możemy teraz wyznaczyć ze znanej nam już własności

$$\mathbb{E}[M] = \mathbb{E}[h(N)] = \mathbb{E}[(N + 1)/2] = 13/4.$$

W powyższych przykładach  $h(Z)$  i  $h(N)$  są warunkowymi wartościami oczekiwanymi odpowiednio  $X$  pod warunkiem  $Z$  i  $M$  pod warunkiem  $N$ . Pokażemy teraz ogólną definicję warunkowej wartości oczekiwanej. Poniższa definicja, w przeciwieństwie do tej z **Przykładu 1.2**, nie podaje wprost konstrukcji. Będzie ona jednak bardzo pomocna przy badaniu własności warunkowej wartości oczekiwanej.

### Definicja 1.4

Niech  $\mathcal{G}$  będzie pod- $\sigma$ -ciałem  $\mathcal{F}$ , a  $X$  całkowalną zmienną losową (tzn.  $\mathbb{E}|X| < \infty$ ). **Warunkową wartością oczekiwaną**  $X$  pod warunkiem  $\mathcal{G}$  nazywamy zmienną losową  $\mathbb{E}[X|\mathcal{G}]$  taką, że

(W<sub>1</sub>)  $\mathbb{E}[X|\mathcal{G}]$  jest mierzalna względem  $\mathcal{G}$ ;

(W<sub>2</sub>) dla każdego  $A \in \mathcal{G}$ ,

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega) \mathbb{P}(d\omega) = \int_A X(\omega) \mathbb{P}(d\omega).$$

Dokładne zrozumienie i przyswojenie powyższej definicji zajmie nam jeszcze trochę czasu. Jak się przekonamy  $\mathbb{E}[X|\mathcal{G}]$  jest aproksymacją  $X$  przy pomocy informacji zawartych w  $\sigma$ -ciele  $\mathcal{G}$ . Na początek sprawdzimy, że warunkowa wartość oczekiwana względem  $\sigma$ -ciała rzeczywiście uogólnia warunkową wartość oczekiwaną względem zdarzenia  $\mathbb{E}[X|Z = z]$ .

### Przykład 1.5

Załóżmy, że  $Z$  jest zmienną losową przyjmującą przeliczalnie wiele wartości. Z **Przykładu 1.2**, zmienna losowa  $h(Z)$ , gdzie  $h(z) = \mathbb{E}[X \mid Z = z]$  spełnia warunki **Definicji 1.4**. Wobec tego

$$\mathbb{E}[X|\sigma(Z)] = h(Z).$$

Podamy teraz dwa konkretne przykłady w których w łatwy sposób wyznaczymy warunkową wartość oczekiwaną powołując się bezpośrednio na

**Definicję 1.2.** Niejednokrotnie będziemy korzystali z poniższych przykładów w zadaniach.

**Przykład 1.6**

Jeżeli  $X \in \mathcal{G}$  (tzn. zmienna losowa  $X$  jest  $\mathcal{G}$  mierzalna), to

$$\mathbb{E}[X|\mathcal{G}] = X.$$

Rzeczywiście, w tym przypadku zmienna losowa  $X$  w oczywisty sposób spełnia oba warunki **Definicji 1.4**. Intuicyjnie oznacza to, że  $\mathcal{G}$  w tym przypadku daje nam więc pełną informację o  $X$ .

**Przykład 1.7**

Jeżeli zmienna losowa  $X$  jest niezależna od  $\mathcal{G}$  (to znaczy  $\sigma$ -ciała  $\sigma(X)$  oraz  $\mathcal{G}$  są niezależne), to wówczas

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X],$$

jest stałą. Istotnie dla  $A \in \mathcal{G}$ ,

$$\int_A X(\omega)\mathbb{P}(d\omega) = \mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A] \cdot \mathbb{E}[X] = \int_A \mathbb{E}[X]\mathbb{P}(d\omega).$$

Ponadto  $\mathbb{E}[X]$  jako stała jest  $\mathcal{G}$ -mierzalna. Zauważmy, że w tym przypadku wynik jest zgodny z oczekiwaniami intuicyjnymi, gdyż 'niezależność' oznacza, iż dodatkowa informacja  $\mathcal{G}$  nie wpływa na naszą wiedzę o  $X$ . W związku z tym nie możemy lepiej przybliżyć  $X$  niż poprzez jej wartość oczekiwaną.

**Przykład 1.8**

Dane są niezależne zmienne losowe  $X, Z$  o rozkładach Poissona z parametrami odpowiednio  $\lambda$  i  $\mu$ . Ile wynosi  $\mathbb{E}[X|\sigma(X+Z)]$ ? Oznaczmy  $Y = X + Z$ . Wówczas  $Y$  ma rozkład Poissona z parametrem  $\lambda + \mu$ :

$$\mathbb{P}[Y = k] = \frac{(\lambda + \mu)^k}{k!} e^{-(\lambda + \mu)}.$$

Z drugiej strony dla  $k > j \geq 0$  mamy

$$\mathbb{P}[X = j, Y = k] = \mathbb{P}[X = j]\mathbb{P}[Z = k - j] = \frac{\lambda^j}{j!} e^{-\lambda} \cdot \frac{\mu^{k-j}}{(k-j)!} e^{-\mu}.$$

Korzystając zatem ze wzoru z **Przykładu 1.5** otrzymujemy

$$\mathbb{E}[X|\sigma(Y)] = h(Y),$$

gdzie

$$h(y) = \mathbb{E}[X|Y = y] = \sum_{0 \leq j \leq y} j \binom{y}{j} \left(\frac{\lambda}{\lambda + \mu}\right)^j \left(1 - \frac{\lambda}{\lambda + \mu}\right)^{y-j} \stackrel{(*)}{=} \frac{\lambda}{\lambda + \mu} \cdot y.$$

Ostatecznie

$$\mathbb{E}[X|\sigma(X + Z)] = \frac{\lambda}{\lambda + \mu}(X + Z).$$

W kroku (\*) skorzystaliśmy ze wzoru

$$\sum_{0 \leq k \leq n} k \binom{n}{k} p^k (1 - p)^{n-k} = np$$

dla  $n = Y$  oraz  $p = \lambda/(\lambda + \mu)$ . Zauważmy, że jest to wzór na wartość oczekiwaną zmiennej losowej o rozkładzie dwumianowym z parametrami  $n$  i  $p$ . Parametr  $p$  określa prawdopodobieństwo sukcesu w jednej próbie, więc liczba oczekiwana sukcesów w  $n$  próbach wynosi  $np$ . Wzór (\*) można też łatwo udowodnić indukcyjnie

$$\begin{aligned} & \sum_{0 \leq k \leq n+1} k \binom{n+1}{k} x^k (1-x)^{n+1-k} \\ &= \sum_{0 \leq k \leq n+1} k \left( \binom{n}{k} + \binom{n}{k-1} \right) x^k (1-x)^{n+1-k} \\ &= (1-x) \sum_{0 \leq k \leq n} k \binom{n}{k} x^k (1-x)^{n-k} + \sum_{1 \leq k \leq n+1} k \binom{n}{k-1} x^k (1-x)^{n+1-k} \\ &\stackrel{ind.}{=} nx(1-x) + x \sum_{0 \leq k \leq n} (k+1) \binom{n}{k} x^k (1-x)^{n-k} \\ &\stackrel{ind.}{=} nx(1-x) + nx^2 + x = (n+1)x. \end{aligned}$$

Definicja warunkowej wartości oczekiwanej wymaga dalszych wyjaśnień. Warto ją zrozumieć zarówno intuicyjnie jak i matematycznie. Należy pokazać, że jest ona poprawna, a więc, że istnieje zmienna losowa  $\mathbb{E}[X|\mathcal{G}]$  i jest ona jednoznacznie wyznaczona. Zaczniemy od wyjaśnienia jednoznaczności.

### Lemat 1.9

Niech  $X$  będzie całkowalną zmienną losową, a  $\mathcal{G}$  pod  $\sigma$ -ciałem  $\mathcal{F}$ . Zmienna losowa  $\mathbb{E}[X | \mathcal{G}]$  jest całkowalna.

*Dowód.* Zauważmy, że dla  $A = \{\mathbb{E}[X | \mathcal{G}] > 0\} \in \mathcal{G}$ . Odwołując się do własności (W2) w Definicji 1.4 mamy

$$0 \leq \int_A \mathbb{E}[X | \mathcal{G}](\omega) \mathbb{P}(d\omega) = \int_A X(\omega) \mathbb{P}(d\omega) \leq \int_A |X(\omega)| \mathbb{P}(d\omega)$$

oraz

$$0 \leq \int_{A^c} -\mathbb{E}[X | \mathcal{G}](\omega) \mathbb{P}(d\omega) = \int_{A^c} -X(\omega) \mathbb{P}(d\omega) \leq \int_{A^c} |X(\omega)| \mathbb{P}(d\omega).$$

Dodając obie nierówności stronami otrzymujemy  $\mathbb{E}[|\mathbb{E}[X | \mathcal{G}]|] \leq \mathbb{E}|X| < \infty$ .  $\square$

**Lemat 1.10**

Jeżeli  $Y$  i  $Y'$  są warunkowymi wartościami oczekiwanymi  $X$  względem  $\mathcal{G}$ , to  $Y = Y'$  p.w.

*Dowód.* Ustalmy dowolny  $\varepsilon > 0$  i rozważmy zdarzenie

$$A_\varepsilon = \{\omega : Y(\omega) - Y'(\omega) \geq \varepsilon\}.$$

Obie zmienne losowe  $Y$  i  $Y'$  są  $\mathcal{G}$  mierzalne (pkt 1.), a więc  $A \in \mathcal{G}$ . Następnie korzystając z definicji zdarzenia  $A_\varepsilon$  a następnie w definicji warunkowej wartości oczekiwanej

$$\begin{aligned} \varepsilon \mathbb{P}[A_\varepsilon] + \int_{A_\varepsilon} Y'(\omega) \mathbb{P}(d\omega) &\leq \int_{A_\varepsilon} Y(\omega) \mathbb{P}(d\omega) = \int_{A_\varepsilon} X(\omega) \mathbb{P}(d\omega) \\ &= \int_{A_\varepsilon} Y'(\omega) \mathbb{P}(d\omega). \end{aligned}$$

Powyższa nierówność może zachodzić jedynie gdy  $\mathbb{P}[A_\varepsilon] = 0$ . Z dowolności  $\varepsilon$  wnioskujemy, że

$$\mathbb{P}[Y \geq Y'] = \mathbb{P}\left[\bigcup_n A_{1/n}\right] = \lim_{n \rightarrow \infty} \mathbb{P}[A_{1/n}] = 0.$$

Należy w tym miejscu zauważyć, że dla  $A_{1/n} \subseteq A_{1/(n+1)}$ . Pokazaliśmy właśnie, że  $Y' \geq Y$  p.w. Powyższy argument jest symetryczny, więc zamieniając rolami  $Y$  i  $Y'$  otrzymujemy odwrotną nierówność  $Y \geq Y'$  p.w., a zatem  $Y = Y'$  p.w.  $\square$

**Przykład 1.11**

Niech  $\Omega = \bigcup_{i=1}^{\infty} A_i$  będzie rozbiem  $\Omega$  na skończoną lub przeliczalną sumę parami rozłącznych zbiorów i niech  $\mathcal{G}$  będzie  $\sigma$ -ciałem generowanym przez zbiory  $A_i$ . Zauważmy, że każdy element  $\mathcal{G}$  jest pewną sumą  $A_i$ , dokładniej jest postaci  $\bigcup_{i \in I} A_i$  dla pewnego  $I \subseteq \mathbb{N}$ . Wówczas warunkowa wartość oczekiwana względem  $\mathcal{G}$  zapisuje się jawnie jako

$$\mathbb{E}[X | \mathcal{G}](\omega) = \sum_{i=1}^{\infty} \mathbb{1}_{A_i}(\omega) \mathbb{E}[X | A_i]. \tag{1.1.1}$$

Powyższa zmienna losowa jest  $\mathcal{G}$  mierzalna (bo jest stała na zbiorach  $A_i$ ), więc wystarczy sprawdzić warunek całkowy 2. Co więcej, ponieważ całkowanie jest operacją liniową, wystarczy sprawdzić go wyłącznie dla zbiorów  $A_i$ :

$$\int_{A_i} \mathbb{E}[X|\mathcal{G}](\omega) \mathbb{P}(d\omega) = \int_{A_i} \frac{\mathbb{E}[X\mathbb{1}_{A_i}]}{\mathbb{P}[A_i]} \mathbb{P}(d\omega) = \mathbb{E}[X\mathbb{1}_{A_i}] = \int_{A_i} X(\omega) \mathbb{P}(d\omega).$$

**Przykład 1.12**

Niech  $\Omega = [0,3]$ ,  $\mathcal{F} = \text{Bor}(\Omega)$  i niech  $\mathbb{P}$  będzie znormalizowaną miarą Lebesguea na odcinku  $[0,3]$ . Rozważmy zmienną losową  $X(\omega) = \omega$ . Zauważmy, że wówczas  $X$  ma rozkład jednostajny na odcinku. Rozważmy  $\sigma$ -ciało  $\mathcal{G} = \sigma(\{[0,1], [1,2], [2,3]\})$ . Korzystając ze wzoru z **Przykładu 1.11** sprawdzamy, że

$$\mathbb{E}[X|\mathcal{G}](\omega) = \lfloor \omega \rfloor + 1/2.$$

**Definicja 1.13**

Załóżmy, że  $X, Y$  są zmiennymi losowymi takimi, że  $X$  jest całkowna. Warunkową wartość oczekiwaną  $X$  pod warunkiem  $Y$  definiujemy jako

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)].$$

**Przykład 1.14**

Załóżmy, że  $X$  i  $Y$  mają rozkład punktowy i niech

$$P_Y(y) = \mathbb{P}[Y = y],$$

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

Wtedy

$$\mathbb{E}[X|Y] = \frac{1}{P_Y(Y)} \sum_x x P_{(X,Y)}(x, Y). \tag{1.1.2}$$

Powyższy wzór wynika z przykładu 1.11, ale powtórzmy rachunki. Zauważmy najpierw, że funkcja występująca po prawej stronie jest funkcją zmiennej losowej  $Y$  więc jest  $\sigma(Y)$  mierzalna. Pozostaje do sprawdzenia warunek całkowy, w tym celu wystarczy pokazać go na zbiorach postaci  $A = \{Y = y\}$  dla  $y$  tż.  $\mathbb{P}[Y = y] > 0$ :

$$\begin{aligned} \int_A \mathbb{E}[X|Y](\omega) \mathbb{P}(d\omega) &= \int_A \sum_x \frac{x P_{(X,Y)}(x, Y(\omega))}{P_Y(Y(\omega))} \mathbb{P}(d\omega) \\ &= \sum_x x P_{(X,Y)}(x, y) = \int_A X(\omega) \mathbb{P}(d\omega). \end{aligned}$$

Również w przypadku, gdy zmienne są absolutnie ciągle można efektywnie obliczyć warunkową wartość oczekiwaną i otrzymać formułę analogiczną do (1.1.2):

**Fakt 1.15**

Niech  $(X, Y)$  będzie wektorem losowym z gęstością  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  i niech  $g_Y(y) = \int_{\mathbb{R}} g(x, y) dx$  będzie gęstością brzegową zmiennej losowej  $Y$ . Wówczas dla dowolnej funkcji borelowskiej  $h: \mathbb{R} \rightarrow \mathbb{R}$  takiej, że  $\mathbb{E}[|h(X)|] < \infty$  zachodzi

$$\mathbb{E}[h(X)|Y] = \int_{\mathbb{R}} h(x) g_{X|Y}(x, Y) dx,$$

gdzie

$$g_{X|Y}(x, y) = \begin{cases} \frac{g(x, y)}{g_Y(y)} & \text{jeżeli } g_Y(y) \neq 0, \\ 0 & \text{jeżeli } g_Y(y) = 0. \end{cases}$$

*Dowód.* Pozostawiamy jako zadanie. □

**Przykład 1.16**

Rozważmy przestrzeń probabilistyczną  $\Omega = [0, 1]$ ,  $\mathcal{F} = \mathcal{Bor}([0, 1])$  gdzie  $\mathbb{P}$  jest miarą Lesbague'a. Dla  $t \in (0, 1)$  rozważmy  $\sigma$ -ciało  $\mathcal{F}_t = \{A \cup B \mid A \in \mathcal{Bor}([0, t]), B \in \{(t, 1], \emptyset\}\}$ . Rozważmy całkowalną zmienną losową  $X: [0, 1] \rightarrow \mathbb{R}$ . Jaka jest warunkowa wartość oczekiwana  $\mathbb{E}[X \mid \mathcal{F}_t]$ ? Widzimy, że charakter **Definicji 1.4** nie wyznacza jawnej konstrukcji  $\mathbb{E}[X \mid \mathcal{F}_t]$ . Będziemy zmuszeni odwołać się do intuicji stojących za warunkową wartością oczekiwaną, zgadnąć postać  $\mathbb{E}[X \mid \mathcal{F}_t]$  na następnie sprawdzić, że spełnia ona żądane postulaty. Zauważmy najpierw, że każda zmienna losowa mierzalna względem  $\mathcal{F}_t$  jest stała na odcinku  $(t, 1]$ . Chcemy zatem funkcję  $X: [0, 1] \rightarrow \mathbb{R}$  przybliżyć funkcją, która jest stała na odcinku  $(t, 1]$ . W sposób naturalny nasuwa się

$$Y(\omega) = X(\omega) \mathbb{1}_{[0, t]}(\omega) + \mathbb{1}_{(t, 1]}(\omega) \frac{1}{1-t} \int_{(t, 1]} X(s) ds.$$

Sprawdźmy teraz, że  $Y = \mathbb{E}[X \mid \mathcal{F}_t]$ .  $Y$  jest  $\mathcal{F}_t$  mierzalna, ponieważ dla  $\omega \in (t, 1]$  przyjmuje ona tę samą wartość  $\int_{(t, 1]} X(s) ds / (1-t)$ . Pozostaje sprawdzić zatem warunek (W2). Niech  $F \in \mathcal{F}_t$  będzie postaci  $F = A \cup B$  dla  $A \in \mathcal{Bor}([0, t])$  oraz  $B \in \{(t, 1], \emptyset\}$ . Zauważmy, że  $X \mathbb{1}_A = Y \mathbb{1}_A$  oraz jeżeli  $B = (t, 1]$  to

$$\mathbb{E}[X \mathbb{1}_B] = \int_{(t, 1]} X(s) ds = \frac{1}{1-t} \int_{(t, 1]} X(s) ds \mathbb{P}[(t, 1]] = \mathbb{E}[Y \mathbb{1}_B].$$

Gdy  $B = \emptyset$ , to oczywiście  $\mathbb{E}[X \mathbb{1}_B] = 0 = \mathbb{E}[Y \mathbb{1}_B]$ . Stąd

$$\mathbb{E}[X \mathbb{1}_F] = \mathbb{E}[X \mathbb{1}_A] + \mathbb{E}[X \mathbb{1}_B] = \mathbb{E}[Y \mathbb{1}_A] + \mathbb{E}[Y \mathbb{1}_B] = \mathbb{E}[Y \mathbb{1}_F].$$

Zaproponowana przez nas funkcja spełnia więc warunki **Definicji 1.4** i stąd

$$\mathbb{E}[X | \mathcal{F}_t](\omega) = X(\omega)\mathbb{1}_{[0,t]}(\omega) + \mathbb{1}_{(t,1]}(\omega)\frac{1}{1-t} \int_{(t,1]} X(s)ds.$$

## 1.2 Prawdopodobieństwo warunkowe

Kolejnym krokiem jest zdefiniowanie **prawdopodobieństwa warunkowego**. Wzór  $\mathbb{P}[A] = \mathbb{E}[\mathbb{1}_A]$  sugeruje jak powinniśmy to uczynić.

### Definicja 1.17

Niech  $\mathcal{G}$  będzie pod  $\sigma$ -ciałem  $\mathcal{F}$ . Prawdopodobieństwo warunkowe pod warunkiem  $\mathcal{G}$  definiujemy jako:

$$\mathbb{P}[A|\mathcal{G}] = \mathbb{E}[\mathbb{1}_A|\mathcal{G}], \quad A \in \mathcal{F}.$$

Zauważmy, że  $\mathbb{P}[A|\mathcal{G}]$  jest zmienną losową taką, że

1.  $\mathbb{P}[A|\mathcal{G}]$  jest  $\mathcal{G}$  mierzalna;
2. dla każdego  $B \in \mathcal{G}$  zachodzi

$$\int_B \mathbb{P}[A|\mathcal{G}](\omega)\mathbb{P}(d\omega) = \mathbb{P}[A \cap B].$$

### Przykład 1.18

Ustalmy zbiór  $B \in \mathcal{F}$  i przyjmijmy  $\mathcal{G} = \sigma(B)$ . Wtedy dla dowolnego  $A \in \mathcal{F}$ , ze wzoru (1.1.1),

$$\mathbb{P}[A|\mathcal{G}] = \mathbb{1}_B \cdot \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} + \mathbb{1}_{B^c} \cdot \frac{\mathbb{P}[A \cap B^c]}{\mathbb{P}[B^c]} = \mathbb{1}_B \cdot \mathbb{P}[A|B] + \mathbb{1}_{B^c} \cdot \mathbb{P}[A|B^c].$$

Przypomnijmy, że poznaliśmy już pojęcie **warunkowego prawdopodobieństwa**:

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

( $\mathbb{P}[\cdot|B]$  jest miarą probabilistyczną). Powyższe rachunki pokazują więc związek pomiędzy obiema definicjami, jeżeli  $\omega \in B$ , to

$$\mathbb{P}[A|\mathcal{G}](\omega) = \mathbb{P}[A|B].$$

### Przykład 1.19

Niech  $E_1$  i  $E_2$  będą niezależnymi zmiennymi losowymi o rozkładzie wykładniczym ze średnią jeden. Niech  $\mathcal{G} = \sigma(E_1)$ . Szukamy reprezentacji

$$\mathbb{P}[E_1 + E_2 > t | \mathcal{G}]$$

dla  $t > 0$ . Dowolne zdarzenie  $F \in \mathcal{G}$  jest postaci  $F = \{E_1 \in C\}$  dla pewnego borelowskiego  $C$ . Mamy

$$\begin{aligned} \mathbb{P}[\{E_1 + E_2 > t\} \cap F] &= \int_{x+y>t, x \in C} e^{-x-y} dx dy \\ &= \int_C e^{-x} e^{-(t-x)_+} dx = \mathbb{E} \left[ e^{-(t-E_1)_+} \mathbb{1}_F \right]. \end{aligned}$$

Stąd

$$\mathbb{P}[E_1 + E_2 > t \mid \mathcal{G}] = e^{-(t-E_1)_+}.$$

### 1.3 Istnienie warunkowej wartości oczekiwanej

Do tej pory zobaczyliśmy kilka podstawowych przykładów warunkowej wartości oczekiwanej. Jednakże z **Definicji 1.4** nie wynika wprost, że zmienna  $\mathbb{E}[X \mid \mathcal{G}]$  w ogóle istnieje.

#### Twierdzenie 1.20

Niech  $(\Omega, \mathcal{F}, \mathbb{P})$  będzie przestrzenią probabilistyczną i niech  $X$  będzie całkowalną zmienną losową, a  $\mathcal{G}$  pod- $\sigma$ -ciałem  $\mathcal{F}$ . Wówczas istnieje warunkowa wartość oczekiwana  $X$  względem  $\mathcal{G}$ :  $\mathbb{E}[X \mid \mathcal{G}]$  i jest ona jednoznacznie wyznaczona.

*Dowód.* Jedyność warunkowej wartości oczekiwanej została wykazana w **Lemacie 1.10**. Dla uzasadnienia, że warunkowa wartość oczekiwana zawsze istnieje przypomnijmy **twierdzenie Radona-Nikodyma**:

#### Twierdzenie 1.21 (Radon, Nikodym 1930 [?])

Jeżeli  $\nu$  i  $\mu$  są  $\sigma$ -skończonymi miarami na  $(\Omega, \mathcal{F}_0)$  oraz  $\nu$  jest absolutnie ciągła względem  $\mu$  (tzn. jeżeli  $\mu(A) = 0$  dla  $A \in \mathcal{F}_0$ , to również  $\nu(A) = 0$ ), to istnieje funkcja  $f$  mierzalna względem  $\mathcal{F}_0$  i taka, że dla każdego  $A \in \mathcal{F}_0$ ,

$$\int_A f(x) \mu(dx) = \nu(A).$$

Funkcja  $f = \frac{d\nu}{d\mu}$  jest nazywana pochodną Radona-Nikodyma.

Zastosujemy powyższe twierdzenie w naszej sytuacji. Załóżmy najpierw, że  $X \geq 0$ . Dla każdego  $A \in \mathcal{G}$  zdefiniujemy

$$\nu(A) = \int_A X(\omega) \mathbb{P}(d\omega).$$

Wówczas  $\nu$  jest miarą  $\sigma$ -addytywną na  $\mathcal{G}$  oraz  $\nu$  jest absolutnie ciągła względem  $\mathbb{P}$ . Z twierdzenia Radona-Nikodyma istnieje zmienna losowa  $Y$ ,  $\mathcal{G}$  mierzalna, taka, że

$$\int_A X(\omega) \mathbb{P}(d\omega) = \nu(A) \stackrel{\text{tw. R.-N.}}{=} \int_A Y(\omega) \mathbb{P}(d\omega).$$

A zatem  $Y = \mathbb{E}[X|\mathcal{G}]$  jest poszukiwaną warunkową wartością oczekiwaną.

Dla dowolnej zmiennej losowej  $X$  zapisujemy ją jako różnicę dwóch nieujemnych zmiennych losowych (części dodatniej i ujemnej):  $X = X^+ - X^-$ . Korzystając z powyższego argumentu wnioskujemy istnienie  $\mathbb{E}[X^+|\mathcal{G}]$  oraz  $\mathbb{E}[X^-|\mathcal{G}]$  i definiujemy

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}].$$

Pokazanie, że jest to szukana warunkowa wartość oczekiwana pozostawiamy jako ćwiczenie.  $\square$

## 1.4 Podstawowe własności

**Twierdzenie 1.22** (Własności warunkowej wartości oczekiwanej)

Niech  $X, X_1, X_2$  będą całkowalnymi zmiennymi losowymi i niech  $\mathcal{G}$  będzie pod- $\sigma$ -ciałem  $\mathcal{F}$ . Wówczas

1. Jeżeli  $X$  jest  $\mathcal{G}$  mierzalna, to  $\mathbb{E}[X|\mathcal{G}] = X$ .
2.  $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}X$ .
3. Jeżeli  $X \geq 0$ , to  $\mathbb{E}[X|\mathcal{G}] \geq 0$  p.w.
4. Dla dowolnych  $a, b \in \mathbb{R}$

$$\mathbb{E}[aX_1 + bX_2|\mathcal{G}] = a\mathbb{E}[X_1|\mathcal{G}] + b\mathbb{E}[X_2|\mathcal{G}].$$

5.  $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$  p.w.
6. Jeżeli  $\mathcal{G}_1, \mathcal{G}_2$  są pod- $\sigma$ -ciałami  $\mathcal{F}$ , takimi, że  $\mathcal{G}_1 \subset \mathcal{G}_2$ , to

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] \stackrel{\text{p.w.}}{=} \mathbb{E}[X|\mathcal{G}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2]$$

7. Dla zmiennej losowej  $Y \in \mathcal{G}$  takiej, że zmienna losowa  $XY$  jest całkowalna

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

*Dowód.* Pkt. 1. został pokazany w **Przykładzie 1.6**. Pkt. 2 wynika z definicji warunkowej wartości oczekiwanej, przyjmując  $A = \Omega$ . W celu pokazania

punktu 3. zdefiniujemy  $A = \{\omega : \mathbb{E}[X|\mathcal{G}](\omega) < 0\}$ . Wtedy  $A \in \mathcal{G}$ . Załóżmy, że  $\mathbb{P}[A] > 0$ . Mamy na mocy definicji warunkowej wartości oczekiwanej

$$0 > \int_A \mathbb{E}[X|\mathcal{G}](\omega)\mathbb{P}(d\omega) = \int_A X(\omega)\mathbb{P}(d\omega) \geq 0,$$

co prowadzi nas do sprzeczności. Wynika stąd, że  $\mathbb{P}[A] = 0$ . Punkt 4. jest oczywisty. Do dowodu pkt. 5. skorzystamy z nierówności

$$-|X| \leq X \leq |X|$$

i dalej na mocy pkt 3. i 4. otrzymujemy

$$-\mathbb{E}[|X||\mathcal{G}] \leq \mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[|X||\mathcal{G}],$$

co implikuje 5. Dowód pkt 6: Weźmy dowolny zbiór  $A \in \mathcal{G}_1 \subset \mathcal{G}_2$ . Wówczas mamy

$$\int_A \mathbb{E}[X|\mathcal{G}_1](\omega)\mathbb{P}(d\omega) \stackrel{A \in \mathcal{G}_1}{=} \int_A X(\omega)\mathbb{P}(d\omega) \stackrel{A \in \mathcal{G}_2}{=} \int_A \mathbb{E}[X|\mathcal{G}_2](\omega)\mathbb{P}(d\omega)$$

co dowodzi pierwszego równania. Drugie wynika z obserwacji  $\mathbb{E}[X|\mathcal{G}_1]$  jest mierzalne względem  $\mathcal{G}_2$  oraz z punktu 1. Pkt. 7 zostanie pokazany podczas ćwiczeń.  $\square$

### Twierdzenie 1.23 (o zbieżności)

Niech  $X$  oraz  $X_n$  będą zmiennymi losowymi.

1. (zbieżność monotoniczna). Jeżeli  $0 \leq X_n \nearrow X$ , to

$$\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}] \quad \text{p.w.}$$

2. (zbieżność ograniczona) Jeżeli  $|X_n| \leq Y$ ,  $\mathbb{E}[Y] < \infty$  oraz  $X_n \rightarrow X$  p.w., to

$$\mathbb{E}[X_n|\mathcal{G}] \rightarrow \mathbb{E}[X|\mathcal{G}] \quad \text{p.w.}$$

3. (warunkowy Lemat Fatout). Jeżeli  $X_n \geq 0$ , to

$$\mathbb{E}[\liminf_{n \rightarrow \infty} X_n|\mathcal{G}] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n|\mathcal{G}] \quad \text{p.w.}$$

*Dowód.* Dowód pkt. 1. Ciąg  $\mathbb{E}[X_n|\mathcal{G}]$  jest monotoniczny (twierdzenie 1.22, pkt 3) i ograniczony przez  $\mathbb{E}[X|\mathcal{G}]$ . Jest więc zbieżny i oznaczmy jego granicę przez  $Y$ . Zmienna losowa  $Y$  jest  $\mathcal{G}$  mierzalna. Niech  $A \in \mathcal{G}$ . Wówczas z twierdzenia o zbieżności monotonicznej,

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

Odwołując się teraz do definicji warunkowej wartości oczekiwanej

$$\lim_{n \rightarrow \infty} \int_A X_n(\omega) \mathbb{P}(d\omega) = \lim_{n \rightarrow \infty} \int_A \mathbb{E}[X_n | \mathcal{G}](\omega) \mathbb{P}(d\omega) = \int_A Y(\omega) \mathbb{P}(d\omega),$$

gdzie ostatnia równość wynika z ponownego zastosowania twierdzenia o zbieżności ograniczonej. Zatem  $Y = \mathbb{E}[X | \mathcal{G}]$ .

Dowód pkt. 3. Zauważmy, że dla każdego ciągu liczb rzeczywistych  $a_n$  prawdą jest, że

$$\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \inf_{k \geq n} a_k.$$

Zauważmy następnie, że dla dowolnej rodziny zmiennych  $\{Z_t\}_{t \in T}$  (tutaj  $T$  jest dowolnym przeliczalnym zbiorem indeksów),

$$\mathbb{E} \left[ \inf_{t \in T} Z_t \middle| \mathcal{G} \right] \leq \inf_{t \in T} \mathbb{E}[Z_t | \mathcal{G}].$$

Rzeczywiście mamy  $X = \inf_{s \in T} Z_s \leq Z_t$  dla każdego  $t \in T$ . Stąd  $\mathbb{E}[X | \mathcal{G}] \leq \mathbb{E}[Z_t | \mathcal{G}]$  dla każdego  $t \in T$  co z kolei pociąga, że  $\mathbb{E}[X | \mathcal{G}] \leq \inf_{t \in T} \mathbb{E}[Z_t | \mathcal{G}]$  co jest postulowaną przez nas nierównością.

Przechodząc do zasadniczej części dowodu punktu 3 napiszmy, powołując się na pkt 1,

$$\begin{aligned} \mathbb{E}[\liminf_{n \rightarrow \infty} X_n | \mathcal{G}] &= \mathbb{E}[\lim_{n \rightarrow \infty} \inf_{k \geq n} X_k | \mathcal{G}] = \lim_{n \rightarrow \infty} \mathbb{E}[\inf_{k \geq n} X_k | \mathcal{G}] \\ &\leq \lim_{n \rightarrow \infty} \inf_{k \geq n} \mathbb{E}[X_k | \mathcal{G}] = \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}]. \end{aligned}$$

Dowód punktu 2. Zauważmy, że zmienne  $X'_n = Y + X_n$  są nieujemne. Zachodzi więc

$$\begin{aligned} \mathbb{E}[Y + \liminf_{n \rightarrow \infty} X_n | \mathcal{G}] &= \mathbb{E}[\liminf_{n \rightarrow \infty} X'_n | \mathcal{G}] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X'_n | \mathcal{G}] \\ &= \mathbb{E}[Y | \mathcal{G}] + \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}]. \end{aligned}$$

Dwukrotnie skorzystaliśmy tutaj z faktu, że dla dowolnego ciągu  $a_n$  i dowolnej stałej  $b$ ,  $\liminf_{n \rightarrow \infty} (b + a_n) = b + \liminf_{n \rightarrow \infty} a_n$ . Pokazaliśmy właśnie, że

$$\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[\liminf_{n \rightarrow \infty} X_n | \mathcal{G}] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}].$$

Zauważmy teraz, że z punktu 3. wynika, że dla całkowalnych zmiennych losowych  $X''_n$  takich, że  $X''_n \leq 0$ ,

$$\mathbb{E}[\limsup_{n \rightarrow \infty} X''_n | \mathcal{G}] \leq \limsup_{n \rightarrow \infty} \mathbb{E}[X''_n | \mathcal{G}] \quad \text{p.w.}$$

Stosując tę obserwację do zmiennych  $X''_n = X_n - Y$  otrzymujemy

$$\begin{aligned}\mathbb{E}[\limsup_{n \rightarrow \infty} (X_n - Y) | \mathcal{G}] &= \mathbb{E}[\limsup_{n \rightarrow \infty} X_n'' | \mathcal{G}] \geq \limsup_{n \rightarrow \infty} \mathbb{E}[X_n'' | \mathcal{G}] \\ &= -\mathbb{E}[Y | \mathcal{G}] + \limsup_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}].\end{aligned}$$

To z kolei dowodzi

$$\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[\limsup_{n \rightarrow \infty} X_n | \mathcal{G}] \leq \limsup_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{G}].$$

Oznacza to, że ciąg  $\mathbb{E}[X_n | \mathcal{G}]$  jest zbieżny i jego granicą jest  $\mathbb{E}[X | \mathcal{G}]$ .  $\square$

#### Twierdzenie 1.24

Niech  $\mathcal{G}$  i  $\mathcal{D}$  będą niezależnymi  $\sigma$ -podciałami  $\mathcal{F}$ . Rozważmy borelowską  $\psi: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  oraz zmienne losowe  $X \in \mathcal{G}$  i  $Y \in \mathcal{D}$  takie, że  $\mathbb{E}[|\psi(X, Y)|] < \infty$ . Wówczas

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

gdzie  $\Psi(x) = \mathbb{E}[\psi(x, Y)]$ .

*Dowód.* Jeżeli funkcja  $\psi$  jest postaci  $\psi(x, y) = \mathbb{1}_A(x)\mathbb{1}_B(y)$  dla borelowskich  $A$  i  $B$ , to

$$\mathbb{E}[\psi(X, Y) | \mathcal{G}] = \mathbb{1}_A(X)\mathbb{P}[Y \in B | \mathcal{G}] = \mathbb{1}_A(X)\mathbb{P}[Y \in B]$$

co okazuje się być zgodne z  $\Psi(X)$ , gdzie

$$\Psi(x) = \mathbb{E}[\mathbb{1}_A(x)\mathbb{1}_B(Y) | \mathcal{G}] = \mathbb{1}_A(x)\mathbb{P}[Y \in B].$$

Pokażemy teraz, że teza zachodzi dla funkcji postaci  $\psi = \mathbb{1}_E$  dla dowolnego  $E \in \mathcal{B}or(\mathbb{R}^2)$ . Przypomnijmy, że dla zbioru  $E \in \mathcal{B}or(\mathbb{R}^2)$  i  $x \in \mathbb{R}$  definiujemy cięcie zbioru  $E$  jako

$$E_x = \{y \in \mathbb{R} : (x, y) \in E\}.$$

Oznaczmy  $p_E(x) = \mathbb{P}[E_x] = \mathbb{P}[(x, Y) \in E]$ . Rozważmy klasę podzbiorów  $\mathbb{R}^2$ ,

$$\mathcal{E} = \left\{ E \in \mathcal{B}or(\mathbb{R}^2) : \mathbb{P}[(X, Y) \in E | \mathcal{G}] = p_E(X) \text{ p.w.} \right\}.$$

Wówczas  $\mathcal{C}$  tworzy  $\lambda$ -układ. Zauważmy też, że  $\mathcal{C}$  to rodzina zbiorów  $E$  dla których funkcja  $\psi = \mathbb{1}_E$  spełnia tezę twierdzenia. Z obserwacji uczynionej na początku dowodu wiemy, że  $A \times B \in \mathcal{E}$  dla dowolnych  $A, B \in \mathcal{B}or(\mathbb{R})$ . Oznacza to, że  $\mathcal{E}$  zawiera  $\pi$ -układ

$$\mathcal{A} = \{A \times B : A, B \in \mathcal{B}or(\mathbb{R})\}.$$

Z twierdzenia o  $\pi$ - $\lambda$  układach  $\mathcal{C} = \mathcal{Bor}(\mathbb{R}^2)$ . Pokazaliśmy właśnie, że teza zadania jest spełniona dla  $\psi(x, y) = \mathbb{1}_E(x, y)$  dla  $E \in \mathcal{Bor}(\mathbb{R}^2)$ .

Teza twierdzenia wynika przez standardową technikę dowodową polegającą na rozważeniu  $\psi$  będącą funkcją prostą, nieujemną i wreszcie spełniającą zadany warunek całkowalności.  $\square$

### Przykład 1.25

Rozważmy niezależne zmienne losowe  $N$  i  $E$  o rozkładach  $\mathcal{N}(0, 1)$  i  $Exp(1)$  odpowiednio. Jaka jest funkcja charakterystyczna zmiennej  $X = EN$ ? Wyznaczmy ją najpierw

$$\mathbb{E}[e^{itX} | E] = \mathbb{E}[e^{itEN} | E].$$

Stosując **Twierdzenie 1.24** do funkcji  $\psi(e, n) = e^{iten}$  znajdujemy funkcję

$$\Psi(e) = \mathbb{E}[\psi(e, N)] = \mathbb{E}[e^{iteN}] = e^{-t^2 e^2 / 2}.$$

Stąd

$$\mathbb{E}[e^{itX} | E] = e^{-t^2 E^2 / 2}.$$

Korzystając teraz w własności warunkowej wartości oczekiwanej

$$\begin{aligned} \mathbb{E}[e^{itX}] &= \mathbb{E}[\mathbb{E}[e^{itX} | E]] = \mathbb{E}[e^{-t^2 E^2 / 2}] = \int_0^\infty e^{-x - x^2 t^2 / 2} dx \\ &= \sqrt{2\pi} t e^{t^2 / 2} \mathbb{P}[N > |t|]. \end{aligned}$$

## 1.5 Interpretacja geometryczna. Zagadnienie regresji

Oznaczmy przez  $(X, Z)$  dwuwymiarowy wektor losowy taki, że  $\mathbb{E}X^2 < \infty$ . Standardowy **problem regresji** polega na przewidzeniu wartości  $X$  na podstawie obserwacji  $Z$ . Innymi słowy chcemy znaleźć taką funkcję  $h$ , aby jak najlepiej przybliżyć  $X$  przez  $h(Z)$  w sensie aproksymacji średniokwadratowej, tzn. tak aby zminimalizować wartość  $\mathbb{E}[(X - h(Z))^2]$ . Zauważmy, że każda zmienna losowa  $h(Z)$  jest mierzalna względem  $\sigma(Z)$ . Okazuje się, że tak wyglądają wszystkie zmienne mierzalne względem  $\sigma(Z)$ .

### Fakt 1.26

Niech  $Z: \Omega \rightarrow \mathbb{R}$  będzie zmienną losową. Dla każdej zmiennej  $Y: \Omega \rightarrow \mathbb{R}$  mierzalnej względem  $\sigma(Z)$  istnieje mierzalna funkcja  $h: \mathbb{R} \rightarrow \mathbb{R}$  taka, że  $Y = h(Z)$ .

*Dowód.* Zadanie.  $\square$

Okazuje się więc, że szukany przez nas kres dolny ma dwie reprezentacje. Mianowicie

$$\inf_{h: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E} \left[ (X - h(Z))^2 \right] = \inf_{Y \in \sigma(Z)} \mathbb{E} \left[ (X - Y)^2 \right]$$

W powyższym wzorze zastosowaliśmy skrócony zapis. Jeżeli zmienna losowa  $Y$  jest mierzalna względem  $\sigma$ -ciała  $\mathcal{G}$ , to piszemy  $Y \in \mathcal{G}$ .

Na potrzeby zagadnienia regresji zakładać będziemy, że rozważana przez nas przestrzeń probabilistyczna  $(\Omega, \mathcal{F}, \mathbb{P})$  jest taka, że  $\sigma$ -ciało  $\mathcal{F}$  zawiera wszystkie zbiory miary zero. Jak pokazuje następujący fakt nie będzie to istotną restrykcją.

**Fakt 1.27**

Niech  $\mathcal{G} \subseteq \mathcal{F}$  będzie  $\sigma$ -ciałem. Rozważmy

$$\mathcal{G}_0 = \sigma(\{A \in \mathcal{F} : \mathbb{P}[A] = 0\} \cup \mathcal{G}).$$

Wówczas dla dowolnej całkowalnej zmiennej losowej  $X$ ,

$$\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X | \mathcal{G}_0] \quad \text{p.w.}$$

*Dowód. Zadanie.*

□

Dla pod  $\sigma$ -ciała  $\mathcal{G} \subseteq \mathcal{F}$  położmy

$$L^2(\mathcal{G}) = \{Y \in \mathcal{G} : \mathbb{E}[Y^2] < \infty\}.$$

Wówczas  $L^2(\mathcal{G})$  jest przestrzenią euklidesową z iloczynem skalarnym

$$\langle X, Y \rangle_{L^2(\mathcal{G})} = \mathbb{E}[XY], \quad X, Y \in L^2(\mathcal{G}).$$

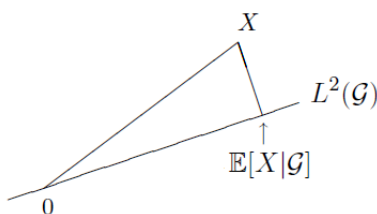
Powyższy iloczyn zadaje normę

$$\|X\|_{L^2(\mathcal{G})} = \sqrt{\mathbb{E}[X^2]}.$$

**Twierdzenie 1.28**

Niech  $\mathcal{G} \subseteq \mathcal{F}$  będzie pod- $\sigma$  ciałem zawierającym wszystkie zbiory miary zero. Wówczas  $L^2(\mathcal{G}) \subset L^2(\mathcal{F})$  jest domkniętą podprzestrzenią. Dla dowolnego  $X$  takiego, że  $\mathbb{E}X^2 < \infty$  ( $X \in L^2(\mathcal{F})$ )  $\mathbb{E}[X|\mathcal{G}]$  jest rzutem  $X$  na  $L^2(\mathcal{G})$ , tzn.  $\mathbb{E}[X|\mathcal{G}]$  minimalizuje wartość

$$\{\mathbb{E}[(X - Y)^2] : Y \in L^2(\mathcal{G})\}$$



**Rysunek 1.1.** Warunkowa wartość oczekiwana jako rzut w  $L^2$

**Uwaga 1.5.1.** Z teorii przestrzeni Hilberta zmienną losową  $X \in L^2(\mathcal{F})$  można jednoznacznie przedstawić w postaci sumy

$$X = Y_0 + Z_0,$$

gdzie  $Y_0 \in L^2(\mathcal{G})$ ,  $(Z_0, Y_0) = \mathbb{E}Y_0Z_0 = 0$  i  $Y_0$  minimalizuje wartość  $\{\mathbb{E}[(X - Y)^2] : Y \in L^2(\mathcal{G})\}$ . Chcemy więc pokazać, że  $Y_0 = \mathbb{E}[X|\mathcal{G}]$ .

**Uwaga 1.5.2.** Z powyższego twierdzenia wynika więc, że rozwiązaniem problemu regresji jest  $\mathbb{E}[X|\sigma(Z)] = h(Z)$  dla pewnej deterministycznej funkcji  $h$ .

### Zadanie 1.1

(Nierówność Jensena) Dana jest funkcja wypukła  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ , przestrzeń probabilistyczna  $(\Omega, \mathcal{F}, \mathbb{P})$  oraz  $\mathcal{G}$  pod- $\sigma$ -ciało  $\mathcal{F}$ . Załóżmy, że zmienne losowe  $X$  i  $\phi(X)$  są całkowalne. Pokaż, że

$$\phi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\phi(X)|\mathcal{G}].$$

### Zadanie 1.2

Niech  $X$  będzie zmienną losową określoną na przestrzeni probabilistycznej  $(\Omega, \mathcal{F}, \mathbb{P})$  taką, że  $\mathbb{E}X^2 < \infty$ . Niech  $\mathcal{G}$  będzie pod- $\sigma$ -ciałem  $\mathcal{F}$ . Pokaż, że  $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]^2] < \infty$ .

*Dowód twierdzenia 1.28.* Weźmy dowolny  $Y \in L^2(\mathcal{G})$ . Wówczas

$$\begin{aligned} \mathbb{E}[(X - Y)^2] &= \mathbb{E}\left[\left((X - \mathbb{E}[X|\mathcal{G}]) - (Y - \mathbb{E}[X|\mathcal{G}])\right)^2\right] \\ &= \mathbb{E}\left[(X - \mathbb{E}[X|\mathcal{G}])^2\right] + \mathbb{E}\left[(Y - \mathbb{E}[X|\mathcal{G}])^2\right] \\ &\quad - 2\mathbb{E}\left[(X - \mathbb{E}[X|\mathcal{G}])(Y - \mathbb{E}[X|\mathcal{G}])\right] \end{aligned}$$

Pokażemy, że ostatni składnik jest równy 0. Oznaczmy  $Y' = Y - \mathbb{E}[X|\mathcal{G}]$ . Wtedy  $Y'$  jest  $\mathcal{G}$  mierzalne, a zatem

$$\begin{aligned} \mathbb{E}\left[(X - \mathbb{E}[X|\mathcal{G}])(Y - \mathbb{E}[X|\mathcal{G}])\right] &= \mathbb{E}[XY'] - \mathbb{E}[Y'\mathbb{E}[X|\mathcal{G}]] \\ &\stackrel{\mathbb{E}[Y'X|\mathcal{G}] = Y'\mathbb{E}[X|\mathcal{G}]}{=} \mathbb{E}[XY'] - \mathbb{E}[\mathbb{E}[Y'X|\mathcal{G}]] \\ &\stackrel{\text{Tw.1.12pkt6}}{=} \mathbb{E}[XY'] - \mathbb{E}[XY'] = 0. \end{aligned}$$

Podsumowując, pokazaliśmy

$$\mathbb{E}[(X - Y)^2] = \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2] + \mathbb{E}[(Y - \mathbb{E}[X|\mathcal{G}])^2]$$

a powyższe wyrażenie jest najmniejsze, gdy drugi składnik znika, czyli  $Y = \mathbb{E}[X|\mathcal{G}]$ .  $\square$

## 1.6 Regularne rozkłady warunkowe

Przypomnijmy, że prawdopodobieństwo warunkowe względem zmiennej losowej jest dane przez  $\mathbb{P}[A|Y] = \mathbb{E}[\mathbb{1}_A|Y]$ . Dla każdego  $y \in \mathbb{R}$  chcielibyśmy móc określić zmienną losową  $\mathbb{P}[A|Y = y]$  taką, że na zdarzeniu  $\{Y = y\}$  zachodzi  $\mathbb{P}[A|Y] = \mathbb{P}[A|Y = y]$ . Dla przykładu może nas interesować opis zjawiska losowego o dwóch losowych krokach. W pierwszym kroku podrabiamy monetę tak aby orzeł wypadł na niej z prawdopodobieństwem  $X$ . Następnie wykonujemy  $n$  niezależnych rzutów  $Y_1, Y_2, \dots, Y_n$  tak spreparowaną monetą. Wówczas zmienne  $(Y_1, \dots, Y_n)$  pod warunkiem  $\{X = x\}$  są iid z rozkładem Bernoulliego z prawdopodobieństwem sukcesu  $x$ .

Przypomnijmy, że dla całkowalnej zmiennej losowej  $X$  istnieje funkcja mierzalna  $h$  taka, że

$$\mathbb{E}[X | \sigma(Y)] = h(Y). \quad (1.6.1)$$

Powyższa równość pozwala zdefiniować  $\mathbb{E}[X | Y = y]$ .

### Definicja 1.29

Niech  $Y$  będzie dowolną zmienną losową. Dla całkowalnej zmiennej losowej  $X$  rozważmy mierzalną funkcję  $h$  spełniającą (1.6.1). Dla  $y \in \mathbb{R}$  wartość oczekiwaną  $X$  pod warunkiem  $\{Y = y\}$  definiujemy jako  $\mathbb{E}[X | Y = y] = h(y)$ . Podobnie definiujemy  $\mathbb{P}[A | Y = y] = \mathbb{E}[\mathbb{1}_A | Y = y]$ .

Jeżeli  $\mathbb{P}[Y = y] > 0$  to prawdopodobieństwo warunkowe z powyższej definicji zgadza się (ćwiczenie) ze znanym nam już wzorem

$$\mathbb{P}[A | Y = y] = \frac{\mathbb{P}[A \cap \{Y = y\}]}{\mathbb{P}[Y = y]}.$$

Powyższa konstrukcja pozwala zdefiniować  $\mathbb{P}[A | Y = y]$  nawet gdy  $\mathbb{P}[Y = y] = 0$ . Czy jednak i w tym przypadku  $\mathbb{P}[\cdot | Y = y]$  jest miarą probabilistyczną? Mamy  $\mathbb{P}[\Omega | Y = y] = 1$  p.w. oraz

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n \mid Y = y\right] = \sum_{n=1}^{\infty} \mathbb{P}[A_n | Y = y] \quad p.w.$$

dla parami rozłącznych zdarzeń  $\{A_n\}_{n=1}^\infty$ . Sugeruje to, że  $\mathbb{P}[\cdot | Y = y]$  spełnia postulaty prawdopodobieństwa. Nie jest jednak zupełnie jasne, czy istnieje taki wybór wersji  $\mathbb{P}[A|\mathcal{G}]$  dla których powyższa równość jest spełniona dla wszystkich możliwych wyborów zdarzeń  $\{A_n\}_{n=1}^\infty$ . Dokładana odpowiedź na to pytanie wymaga argumentów.

### Definicja 1.30

Funkcję  $\kappa: \mathbb{R} \times \mathcal{B}or(\mathbb{R}) \rightarrow [0, 1]$  nazywamy funkcją przejścia, jeżeli

(P1) Dla każdego  $y$ ,  $\kappa(y, \cdot)$  jest rozkładem prawdopodobieństwa na  $\mathbb{R}$ .

(P2) Dla każdego  $A \in \mathcal{B}or(\mathbb{R})$  funkcja  $\kappa(\cdot, A)$  jest borelowska.

### Definicja 1.31

Niech  $X$  i  $Y$  będą zmiennymi losowymi. Funkcję przejścia  $\kappa_{X,Y}$  nazywamy regularnym rozkładem warunkowym  $X$  pod warunkiem  $Y$  jeżeli z prawdopodobieństwem jeden dla każdych  $y \in \mathbb{R}$  oraz  $A \in \mathcal{B}or(\mathbb{R})$ ,

$$\mathbb{P}[X \in A | Y = y] = \kappa_{X,Y}(y, A).$$

### Przykład 1.32

Jeżeli  $X$  i  $Z$  są niezależne, to

$$\mathbb{P}[X + Y \in \cdot | X = x] = (\delta_x * \mu_Y)(\cdot).$$

### Przykład 1.33

Jeżeli zmienne losowe  $(X, Y)$  mają łączną gęstość  $f$  na  $\mathbb{R}^2$ , to

$$\mathbb{P}[X \in dx | Y = y] = f_{X|Y}(x, y)dx = \frac{f(x, y)}{\int f(t, y)dt}dx$$

### Przykład 1.34

Niech  $E_1, E_2$  będą niezależnymi zmiennymi losowymi o standardowym rozkładzie wykładniczym. Pokażemy, że

$$\mathbb{P}[E_1 \in \cdot | E_1 + E_2 = y] = \text{Unif}[0, y](\cdot).$$

Wektor losowy  $(E_1, E_1 + E_2)$  ma rozkład o gęstości  $e^{-y} \mathbb{1}_{\{y \geq x\}}$ . Ze wzoru z poprzedniego przykładu

$$\mathbb{P}[E_1 \in dx | E_1 + E_2 = y] = \mathbb{1}_{\{y \geq x\}}dx.$$

**Twierdzenie 1.35**

Niech  $X$  i  $Y$  będą zmiennymi losowymi. Istnieje funkcja przejścia  $\kappa_{X,Y}$  będąca rozkładem warunkowym  $X$  pod warunkiem  $Y$ .

*Dowód.* Dowód polegał będzie na zdefiniowaniu warunkowej dystrybuanty  $X$  poprzez określenie jej na liczbach wymiernych.

Dla  $r \in \mathbb{Q}$  niech  $F(r, \cdot)$  będzie wersją prawdopodobieństwa warunkowego  $\mathbb{P}[X \leq r | Y]$ . Skoro dla wymiernych  $r \leq s$ ,  $\mathbb{1}_{(-\infty, r]} \leq \mathbb{1}_{(-\infty, s]}$  a co za tym idzie

$$F(r, \cdot) = \mathbb{P}[X \leq r | Y] \leq \mathbb{P}[X \leq s | Y] = F(s, \cdot) \quad p.w.$$

Oznacza to, że istnieje zbiór miary zero  $A_{r,s} \in \mathcal{F}$  taki, że

$$F(r, \omega) \leq F(s, \omega), \quad \omega \in \Omega \setminus A_{r,s}$$

Z warunkowej wersji twierdzenia o zbieżności monotonicznej, dla każdego  $r \in \mathbb{Q}$  istnieje zbiór miary  $B_r$  taki, że

$$\lim_{n \rightarrow \infty} F(r + 1/n, \omega) = F(r, \omega), \quad \omega \in \Omega \setminus B_r.$$

Wreszcie, na mocy tego samego twierdzenia, dla każdego  $r \in \mathbb{Q}$  istnieje zbiór  $C_r$  taki, że

$$\lim_{n \rightarrow \infty} F(n, \omega) = 1, \quad \lim_{n \rightarrow \infty} F(-n, \omega) = 0, \quad \omega \in \Omega \setminus C_r.$$

Rozważmy zbiór

$$N = \bigcup_{r,s \in \mathbb{Q}} A_{r,s} \cup \bigcup_{r \in \mathbb{Q}} (B_r \cup C_r)$$

oraz rozszerzenie  $F$  do całej prostej przez

$$\tilde{F}(x, \omega) = \inf\{F(r, \omega) : r \in \mathbb{Q}, r > x\}, \quad x \in \mathbb{R}, \omega \in \Omega \setminus N.$$

Wówczas  $\tilde{F}(\cdot, \omega)$  jest prawostronnie ciągła i niemalejąca przy czym

$$F(z, \omega) = \tilde{F}(z, \omega), \quad z \in \mathbb{Q}, \omega \in \Omega \setminus N.$$

Stąd  $\tilde{F}(\cdot, \omega)$  jest dystrybuantą dla  $\omega \in \Omega \setminus N$ . Dla  $\omega \in N$  niech  $\tilde{F}(\cdot, \omega) = F_0(\cdot)$  dla dowolnej ustalonej dystrybuanty  $F_0$ .

Niech  $\kappa(\omega, \cdot)$  będzie miarą na  $\mathcal{B}or(\mathbb{R})$  o dystrybuancie  $\tilde{F}(\cdot, \omega)$ . Wówczas, dla  $B = (-\infty, r]$ , zmienna

$$\kappa(\omega, B) = \tilde{F}(r, \omega) \mathbb{1}_{\Omega \setminus N} + F_0(r) \mathbb{1}_N$$

jest  $\sigma(Y)$  mierzalna. Skoro  $\{(-\infty, r] : r \in \mathbb{Q}\}$  jest  $\pi$ -systemem generującym  $\mathcal{Bor}(\mathbb{R})$ , to  $\kappa(\cdot, B)$  jest  $\sigma(Y)$  mierzalna dla wszystkich borelowskich  $B$ .

Pozostaje sprawdzić, że  $\kappa(\omega, B)$  jest wersją prawdopodobieństwa warunkowego  $\mathbb{P}[X \in B | Y]$ . Dla  $A \in \sigma(Y)$ ,  $r \in \mathbb{Q}$  i  $B = (-\infty, r]$  mamy

$$\int_A \kappa(\omega, B) \mathbb{P}(d\omega) = \int_A \mathbb{P}[X \in B | Y](\omega) \mathbb{P}(d\omega) = \mathbb{P}[A \cap \{X \in B\}].$$

Stosując twierdzenie o  $\pi$ - $\lambda$  układach raz jeszcze, powyższa równość zachodzi dla wszystkich borelowskich  $B$ . Skoro  $\kappa(\cdot, B)$  jest  $\sigma(Y)$  mierzalna, to jest mierzalnym obrazem  $Y$ . Innymi słowy, dla pewnej  $\kappa_{X,Y}$  zachodzi  $\kappa(\omega, B) = \kappa_{X,Y}(Y(\omega), B)$ .  $\square$

Użyteczność regularnych rozkładów warunkowych pokazuje następujący rezultat.

**Fakt 1.36**

Dla zmiennych losowych  $X$  i  $Y$  oraz takiej funkcji mierzalnej, że  $\mathbb{E}[|f(X)|] < \infty$  zachodzi

$$\mathbb{E}[f(X)|Y] = \int f(x) \kappa_{X,Y}(Y, dx)$$

*Dowód. Zadanie.*  $\square$

**Fakt 1.37**

Niech  $1/p + 1/q = 1$ . Wówczas

$$\mathbb{E}[|XZ| | Y] \leq \mathbb{E}[|X|^p | Y]^{1/p} \mathbb{E}[|Z|^q | Y]^{1/q}$$

# Martyngały

**Streszczenie** Wprowadzimy pojęcia filtracji i martyngałów. Zaprezentujemy i udowodnimy twierdzenia o zatrzymaniu, zbieżności prawie na pewno oraz w  $L^p$ .

W tym rozdziale zaprezentujemy wyjątkowo przydatną klasę procesów stochastycznych jakimi są martyngały. Są to procesy stochastyczne służące do modelowania przebiegu uczciwej gry. W losowych, mierzalnych momentach wartości takich gier są stałe. Oznacza to, że znalezienie odpowiedniego martyngału pozwala uprościć wiele problemów o złożonej kombinatoryce.

## 2.1 Definicje i przykłady

Martyngały to ciągi zmiennych losowych w których rozwój w czasie dobrze wpisuje się w strukturę posiadanych przez nas informacji. Do modelowania informacji używamy pojęcia filtracji.

### Definicja 2.1

Rodzina  $\sigma$ -ciał  $\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}}$  jest nazywana **filtracją**, jeżeli  $\mathcal{F}_n \subseteq \mathcal{F}$  oraz  $\mathcal{F}_n \subset \mathcal{F}_{n+1}$  dla każdego  $n \in \mathbb{N}$ .

### Przykład 2.2

Niech  $\Omega = \mathbb{R}_+$  i  $\mathbb{F} = \mathcal{B}or(\mathbb{R}_+)$ . Niech  $\mathcal{F}_n$  będzie  $\sigma$ -ciałem podzbiorów  $\mathbb{R}_+$  generowanym przez  $\mathcal{B}[0, n]$ . Wówczas

$$\mathcal{F}_n = \sigma(\mathcal{B}[0, n]) = \{A \cup B : A \in \mathcal{B}or[0, n], B \in \{(n, +\infty), \emptyset\}\}.$$

Rodzina  $\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}}$  tworzy filtrację.

### Przykład 2.3

Wyobraźmy sobie sytuację, w której odczytujemy kolejno cyfry rozwinięcia

dziesiątym liczby z przedziału  $[0, 1]$ . Chcemy opisać stan naszej wiedzy po odczytaniu  $n$  cyfr. Niech  $\Omega = [0, 1]$  i  $\mathcal{F} = \mathcal{Bor}([0, 1])$ . Rozważmy

$$\mathcal{F}_n = \sigma\{[k/10^n, (k+1)/10^n] : 0 \leq k < 10^n\}.$$

Wówczas  $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$  jest filtracją na  $(\Omega, \mathcal{F})$ .

Filtracji  $\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}}$  będziemy używać w kontekście procesów ewoluujących dyskretnych krokach. Wówczas dla  $n \in \mathbb{N}$ ,  $\mathcal{F}_n$  można interpretować jako informacje zgromadzone do chwili  $n$ .

#### Definicja 2.4

Rodzina zmiennych losowych  $\{X_n\}_{n \in \mathbb{N}}$  jest **adaptowalna** do filtracji (zgodna z filtracją)  $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ , jeżeli  $X_n$  jest  $\mathcal{F}_n$  mierzalna dla każdego  $n$ .

W przyszłości dla zmiennej losowej  $X$  i  $\sigma$ -ciała  $\mathcal{G}$  będziemy pisać  $X \in \mathcal{G}$  jeżeli  $X$  jest  $\mathcal{G}$  mierzalna.

Zauważmy, że każda rodzina zmiennych losowych  $X = \{X_n\}_{n \in \mathbb{N}}$  jest adaptowalna do pewnej filtracji. Wystarczy rozważyć tak zwaną naturalną filtrację zadaną przez

$$\mathcal{F}_n^X = \sigma(X_k : k \leq n).$$

Martyngały służą do opisu gier losowych. Zanim wprowadzimy ścisłą definicję rozważmy następujący wprowadzający przykład.

#### Przykład 2.5

Rozważmy grę w rzuty monetą, na której orzeł wypada z prawdopodobieństwem  $p \in (0, 1)$ . Gra toczy się według zasad *double or nothing*: jeżeli wypadnie orzeł, to kapitał gracza jest podwajany. W przeciwnym wypadku gracz traci wszystkie pieniądze. Zastanówmy się, czy taka gra może być sprawiedliwa. Bez zmniejszania ogólności możemy założyć, że gracz zaczyna z jednostkowym kapitałem. Jeżeli przez  $\{\xi_j\}_{j \in \mathbb{N}}$  oznaczymy ciąg zmiennych iid o rozkładzie

$$\mathbb{P}[\xi_j = 2] = 1 - \mathbb{P}[\xi_j = 0] = p,$$

to zmienna losowa zadaną przez

$$X_n = \prod_{j=1}^n \xi_j$$

reprezentuje kapitał gracza po  $n$  rzutach. Rozważmy filtrację  $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$  zadaną przez  $\mathcal{F}_n = \sigma(\xi_j : j \in [n])$ . Sprawdźmy jaka jest oczekiwana wygrana gracza przy każdym kolejnym rzucie monetą. Zauważmy, że proces

$X = (X_n)_{n \in \mathbb{N}}$  spełnia zależność rekurencyjną  $X_{n+1} = \zeta_{n+1} X_n$ . Wykorzystując ją otrzymujemy

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n \mathbb{E}[\zeta_{n+1}] = X_n \cdot 2p.$$

Jeżeli  $2p > 1$ , to gra jest korzystna dla gracza. Jeśli natomiast  $2p < 1$ , to gra jest korzystniejsza dla kasyna. Gra będzie sprawiedliwa tylko dla  $2p = 1$ . Zachodzi wtedy

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n.$$

Pamiętajmy jednak, że nawet w przypadku, gdy  $p \geq 1/2$  ale  $p \neq 1$ , z lematu Borala-Cantelliego  $\lim_{n \rightarrow \infty} X_n = 0$ .

### Definicja 2.6

Niech  $\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}}$  będzie filtracją oraz niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie ciągiem całkowalnych (tzn.  $\mathbb{E}|X_n| < \infty$ ) zmiennych losowych zgodnych z filtracją  $\mathbb{F}$ . Mówimy, że

- $X$  jest  **$\mathbb{F}$ -martyngałem**, jeżeli

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n \quad n \in \mathbb{N};$$

- $X$  jest  **$\mathbb{F}$ -podmartyngałem**, jeżeli

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n \quad n \in \mathbb{N};$$

- $X$  jest  **$\mathbb{F}$ -nadmartyngałem**, jeżeli

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n \quad n \in \mathbb{N};$$

O martyngałach można myśleć jak o portfelu gracza w czasie  $n$ , który bierze udział w uczciwej grze (**Przykład 2.5**). Podmartyngały oznaczają wówczas gry losowe, które faworyzują gracza. Zauważmy, że wartość oczekiwana portfela w dowolnej chwili jest równa wartości początkowej. Rzeczywiście, z **Twierdzenia 1.22** pkt. 6,

$$\mathbb{E}[X_n] = \mathbb{E}[\mathbb{E}[X_n | \mathcal{F}_n]] = \mathbb{E}[X_{n-1}].$$

Stąd przez prostą indukcję

$$\mathbb{E}[X_n] = \mathbb{E}[X_0].$$

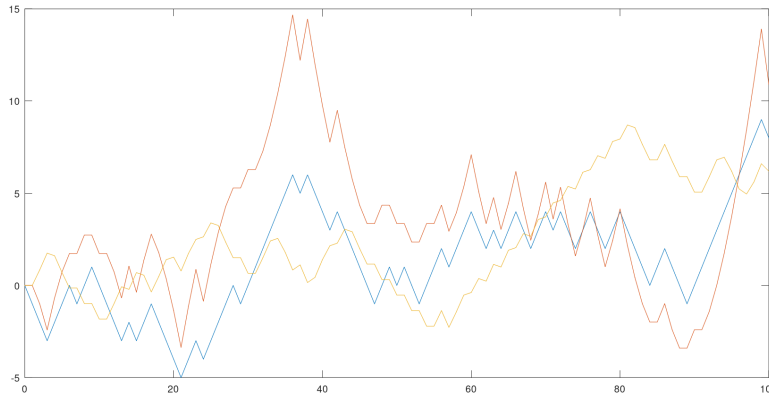
### Przykład 2.7

Rozważmy niezależne, całkowalne zmienne  $\zeta_k$  takie, że  $\mathbb{E}[\zeta_k] = 0$  dla każdego  $k \in \mathbb{N}$ . Wówczas  $X_n = \sum_{k=1}^n \zeta_k$  jest martyngałem. Rzeczywiście, dla naturalnej filtracji  $\mathcal{F}_n = \sigma(X_k : k \leq n) = \sigma(\zeta_k : k \leq n)$  mamy

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_n + \zeta_{n+1} | \mathcal{F}_n] = X_n + \mathbb{E}[\zeta_{n+1}] = X_n.$$

Każdy martynał ma strukturę zbliżoną do tego z powyższego przykładu. Rzeczywiście, dla martynału  $\{X_n\}_{n \in \mathbb{N}}$  rozważmy ciąg różnic martynałowych  $Y_n = X_n - X_{n-1}$  dla  $n \geq 1$ . Wówczas  $X_n = \sum_{k=1}^n Y_k + X_0$ . Zauważmy, że  $\mathbb{E}[Y_n | \mathcal{F}_n] = 0$ .

Nieraz będziemy pomijać w zapisie filtrację i mówić np., że  $\{X_n\}$  jest martynałem, wówczas domyślną filtracją jest  $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ .



**Rysunek 2.1.** Ilustracja procesów w przykładzie 2.8. Trajektoria prostego symetrycznego spaceru losowego  $\{X_n\}$  (niebieska) oraz otrzymana z niej transformata z  $B_j = |X_{j-1}|^{1/2}$  (czerwona) oraz  $B'_j = \cos(X_{j-1})$  (pomarańczowa)

### Przykład 2.8 (Transformata martynałowa)

Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie martynałem. Chcemy modelować sytuację, w której  $X_n$  jest stanem portfela jednego gracza w uczciwej grze. Drugi gracz obserwuje przebieg gry pierwszego i może robić zakłady w oparciu o jej wynik. Tzn. w  $n$ -tym kroku, na podstawie dotychczasowych wyników można obstawić kwotę  $B_n$ . Wówczas wypłata drugiego gracza w  $n$ -tym kroku wynosi  $B_n(X_n - X_{n-1})$ . Tym sposobem, jeżeli  $B_n > 0$  ( $< 0$ ), to drugi gracz obstawia wygraną (przegraną) gracza pierwszego w  $n$ -tym kroku. Drugi gracz dokonuje swych wyborów na podstawie obserwacji gry do danej chwili. Innymi słowy  $B_n \in \mathcal{F}_{n-1}$  jest pewną funkcją  $X_1, \dots, X_{n-1}$ . Wtedy stan portfela drugiego gracza po  $n$  krokach wynosi  $W_0 = X_0$ ,

$$W_n = \sum_{j=1}^n B_j(X_j - X_{j-1}) + X_0.$$

Zauważmy, że  $W_{n+1} = W_n + B_{n+1}(X_{n+1} - X_n)$  a co za tym idzie

$$\begin{aligned}\mathbb{E}[W_{n+1}|\mathcal{F}_n] &= \mathbb{E}[W_n|\mathcal{F}_n] + \mathbb{E}[B_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n] \\ &= W_n + B_{n+1}\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] = W_n.\end{aligned}$$

Zatem  $\{W_n\}_{n \in \mathbb{N}}$  jest martyngałem i oznacza uczciwą grę.

**Przykład 2.9**

Niech  $X$  będzie całkowalną zmienną losową i niech  $\mathcal{F}_n$  będzie dowolną filtracją. Wówczas  $X_n = \mathbb{E}[X|\mathcal{F}_n]$  jest martyngałem, gdyż z **Twierdzenia 1.22** wynika, że zmienne losowe są całkowalne:

$$\mathbb{E}[|X_n|] = \mathbb{E}[|\mathbb{E}[X|\mathcal{F}_n]|] \leq \mathbb{E}[\mathbb{E}[|X|\mathcal{F}_n]] = \mathbb{E}[|X|] < \infty$$

oraz

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_n] = X_n. \quad \square$$

**Przykład 2.10** (Martyngał wykładniczy)

Rozważmy ciąg niezależnych zmiennych losowych  $X_1, X_2, \dots$  o tym samym rozkładzie takim, że

$$\kappa = \log \mathbb{E}[e^{X_1}] < \infty.$$

Wówczas dla  $S_n = X_1 + \dots + X_n$  proces

$$Y_n = \exp\{S_n - \kappa n\}$$

jest martyngałem. Szczegółowe uzasadnienie pozostawiamy jako zadanie.

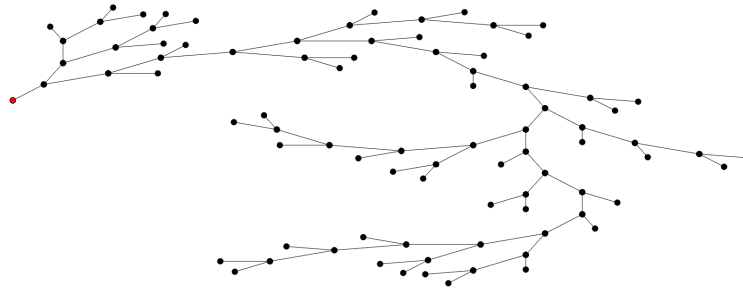
**Przykład 2.11** (Proces Galtona-Watsona)

Rozważmy populację, w której osobnicy (cząsteczki) rozmnażają się niezależnie od siebie z tym samym rozkładem. Chcemy zdefiniować proces  $X = \{X_n\}_{n \in \mathbb{N}}$  w którym  $X_n$  modeluje liczbę osobników (cząsteczek) w populacji w chwili  $n$ .

Ustalmy pewien rozkład  $\mu$  na  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$  o średniej  $m$ . Bez zmniejszania ogólności możemy założyć, że zaczynamy z jedną cząsteczką. Innymi słowy  $X_0 = 1$ . W pierwszym kroku rodzi początkowa cząsteczka rodzi  $Y_1$  cząsteczek, po czym umiera. Zmienna losowa  $Y_1$  ma rozkład  $\mu$ . W kolejnych krokach cząsteczki rozmnażają się niezależnie. Załóżmy, że w  $n$ -tym kroku jest  $X_n$  cząsteczek. Wówczas  $k$ -ta z nich tworzy  $Y_{n+1,k}$  nowych cząsteczek, a sama umiera. Zmienne losowe  $Y_{n,k}$  są niezależne i wszystkie mają rozkład  $\mu$ . Rekurencyjnie ciąg ten można przedstawić następująco

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n+1,k}.$$

Oznaczmy przez  $\mathcal{F}_n$   $\sigma$ -ciało generowane przez zmienne losowe  $\{Y_{j,k}\}_{j \leq n, k \in \mathbb{N}}$ . Wówczas:



**Rysunek 2.2.** Przykładowa realizacja procesu Galtona-Watsona. Kolorem czerwonym zaznaczony jest pierwszy osobnik

$$\begin{aligned} \mathbb{E}[X_{n+1} | \mathcal{F}_n] &= \mathbb{E} \left[ \sum_{k=1}^{X_n} Y_{n+1,k} \middle| \mathcal{F}_n \right] = \mathbb{E} \left[ \sum_{j=0}^{\infty} \mathbf{1}_{\{X_n=j\}} \sum_{k=1}^j Y_{n+1,k} \middle| \mathcal{F}_n \right] \\ &= \sum_{j=0}^{\infty} \mathbf{1}_{\{X_n=j\}} \sum_{k=1}^j \mathbb{E}[Y_{n+1,k} | \mathcal{F}_n] = \sum_{j=0}^{\infty} \mathbf{1}_{\{X_n=j\}} \cdot jm = mX_n. \end{aligned}$$

W powyższych rachunków wynika również, że

$$\mathbb{E}[X_n] = m^n < \infty.$$

Zatem jeżeli

- $m = 1$ , to  $\{X_n\}_{n \in \mathbb{N}}$  jest martyngałem,
- $m < 1$ , to  $\{X_n\}_{n \in \mathbb{N}}$  jest nadmartyngałem,
- $m > 1$ , to  $\{X_n\}_{n \in \mathbb{N}}$  jest podmartyngałem.

Zauważmy też, że  $\{m^{-n} X_n\}_{n \in \mathbb{N}}$  jest martyngałem niezależnie od wartości  $m$ .

Ciekawym pytaniem jest, czy proces  $\{X_n\}$  przeżyje. Pokazuje się, że zależy to wyłącznie od wartości  $m$ . Jeżeli  $m \leq 1$ , to proces wymrze z prawdopodobieństwem 1 (oczywiście należy wykluczyć przypadek  $\mu = \delta_1$ , tzn. gdy każdy osobnik ma zawsze dokładnie jednego potomka). Natomiast jeżeli  $m > 1$ , to proces przeżyje z dodatnim prawdopodobieństwem, a na zbiorze przeżycia będzie rósł wykładniczo szybko.

## 2.2 Czasy zatrzymania

Przypomnijmy, że dla martyngału  $X = \{X_n\}_{n \in \mathbb{N}}$ ,

$$\mathbb{E}[X_n] = \mathbb{E}[X_0].$$

Skoro  $X_n$  możemy interpretować jako stan portfela gracza w uczciwej grze, to powyższa równość oznacza, że nie możemy oszukać systemu (opracować strategii dającej statystycznie zysk) kończąc grę w deterministycznym czasie. Czy jest to możliwe w czasie losowym? Zanim odpowiemy na to pytanie musimy zastanowić się, jakie własności względem filtracji powinien mieć losowy moment zakończenia gry.

### Definicja 2.12

Zmienna losowa  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  nazywana jest **czasem zatrzymania** (momentem stopu), jeżeli dla każdego  $n \in \mathbb{N}$ ,

$$\{T = n\} \in \mathcal{F}_n.$$

Warunek obecny w powyższej definicji oznacza intuicyjnie, że  $T$  jest regułą zatrzymania pewnego procesu, w której decyzję o zatrzymaniu podejmujemy na podstawie zdarzeń z przeszłości i teraźniejszości. Dla przykładu

- trzymamy akcje pewnej spółki, aż ich wartość osiągnie określoną wartość;
- rzucamy kostką tak długo, aż wypadnie 6;
- gramy w kasynie, aż wygramy 1000zł lub zbankrutujemy.

Natomiast czasem zatrzymania **nie jest** zdarzenie: sprzedamy akcje, gdy po raz ostatni ich wartość przekroczy 100 zł (nie znamy przyszłości i nie potrafimy w danej chwili określić kiedy będzie ostatni raz).

Naszym celem jest uzasadnienie, że w przypadku uczciwej gry nie jest możliwe oszukanie systemu (tj.  $\mathbb{E}[X_0] \neq \mathbb{E}[X_T]$ ) nawet z wykorzystaniem najbardziej zmyślnych reguł stopu  $T$ .

### Przykład 2.13

Rzucamy 10 razy monetą. Niech

$$X_n = \begin{cases} 1, & \text{gdy wypadł orzeł w } n\text{-kroku} \\ 0, & \text{gdy wypadła reszka w } n\text{-kroku} \end{cases}$$

Zdefiniujmy naturalną filtrację  $\sigma$ -ciał:  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$  dla  $n = 1, \dots, 10$ . Zdefiniujmy również

- $T$  - moment wypadnięcia pierwszego orła, przyjmujemy  $T = \infty$ , gdy 10 razy nie wypadł orzeł;
- $S$  - moment wypadnięcia ostatniego orła.

$T$  jest czasem zatrzymania, bo

$$\{T = n\} = \{X_1 = 0, \dots, X_{n-1} = 0, X_n = 1\} \in \mathcal{F}_n$$

Natomiast  $S$  nie jest czasem zatrzymania ponieważ

$$\{S = n\} = \{X_n = 1, X_{n+1} = 0, \dots, X_{n+1} = 1, X_{10} = 0\}.$$

Zdarzenie  $\{S = n\}$  zależy od zmiennych  $X_n, X_{n+1}, \dots, X_{10}$ . Nie jest więc elementem  $\mathcal{F}_n$ .

**Lemat 2.14**

Zmienna losowa  $T: \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  jest czasem zatrzymania wtedy i tylko wtedy, gdy  $\{T \leq n\} \in \mathcal{F}_n$  dla każdego  $n \in \mathbb{N}$ .

*Dowód.* Jeżeli  $T$  jest czasem zatrzymania, to

$$\{T \leq n\} = \bigcup_{k=1}^n \{T = k\} \in \mathcal{F}_n.$$

Odwrotnie,

$$\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\} \in \mathcal{F}_n.$$

□

**Przykład 2.15**

Ustalmy  $k$ , wówczas  $T := k$  jest czasem zatrzymania względem dowolnej filtracji  $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ . Istotnie

$$\{T = n\} = \begin{cases} \emptyset & \text{dla } n \neq k \\ \Omega & \text{dla } n = k, \end{cases}$$

a więc  $\{T = n\} \in \mathcal{F}_n$ .

**Przykład 2.16**

Ustalmy  $(\Omega, \mathcal{F}, \mathbb{P})$  z filtracją  $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ . Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie dowolnym procesem stochastycznym (rodziną zmiennych losowych) adoptowalnym do filtracji  $\mathcal{F}_n$ . Ustalmy  $B \in \mathcal{B}(\mathbb{R})$ . Niech

$$T_B = \inf\{n : X_n \in B\}$$

(przyjmujemy, że  $\inf \emptyset = \infty$ ).  $T_B$  jest pierwszym momentem wejścia procesu do zbioru  $B$ . Jest to czas zatrzymania:

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

## 2.3 Twierdzenia Doob'a

Przypomnijmy, że rozważmy stan portfela  $\{X_n\}_{n \in \mathbb{N}}$  gracza, który bierze udział w uczciwej grze. Wróćmy do pytania, czy istnieje strategia, która pozwala na zysk co do wartości średniej? Przez strategię rozumiemy losowy moment  $T$ , w którym gracz przerywa grę. Matematycznie rozważamy martyngał  $\{X_n\}_{n \in \mathbb{N}}$  oraz czas zatrzymania  $T$ . Wówczas stan portfela gracza w chwili zakończenia gry reprezentowany jest przez zmienną losową  $X_T$ , dokładniej na zbiorze  $\{T < \infty\}$  definiujemy  $X_T(\omega) = X_{T(\omega)}(\omega)$ . Wówczas strategia  $T$  przynosi średni zysk, jeżeli  $\mathbb{E}[X_T] > \mathbb{E}[X_0]$ . Czy taka strategia istnieje?

Oznaczmy

$$X_{n \wedge T} = \begin{cases} X_n & \text{jeżeli } T > n \\ X_T & \text{jeżeli } T \leq n. \end{cases}$$

Poniżej prezentujemy uproszczoną wersję twierdzenia Dooba o zatrzymaniu.

### Twierdzenie 2.17 (Doob, Optional Sampling)

Załóżmy, że  $\{X_n\}_{n \in \mathbb{N}}$  jest martyngałem względem  $\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}}$  oraz  $T$  jest  $\mathbb{F}$ -czasem zatrzymania. Wówczas  $\{X_{n \wedge T}\}_{n \in \mathbb{N}}$  jest  $\mathbb{F}$ -martyngałem. W szczególności, dla każdego  $n$ ,

$$\mathbb{E}[X_{T \wedge n}] = \mathbb{E}[X_0].$$

*Dowód.* Zauważmy, że

$$X_{T \wedge n} = \sum_{j=1}^n \mathbb{1}_{\{j \leq T\}} (X_j - X_{j-1}) + X_0,$$

przy czym  $\{T \geq j\} = \{T \leq j-1\}^c \in \mathcal{F}_{j-1}$ . Teza wynika z **Przykładu ??** zastosowanego do zmiennych  $B_j = \mathbb{1}_{\{j \leq T\}}$ .  $\square$

Zauważmy, że jeżeli  $T$  jest skończonym czasem zatrzymania ( $\mathbb{P}[T < \infty] = 1$ ), to  $X_{T \wedge n} \rightarrow X_T$  p.w. W takim przypadku większość naszych zastosowań twierdzenia Dooba będzie polegało na zastosowaniu odpowiedniego twierdzenia granicznego do

$$\mathbb{E}[X_0] = \mathbb{E}[X_{T \wedge n}]$$

i wywnioskowaniu

$$\mathbb{E}[X_0] = \lim_{n \rightarrow \infty} \mathbb{E}[X_{T \wedge n}] = \mathbb{E} \left[ \lim_{n \rightarrow \infty} X_{T \wedge n} \right] = \mathbb{E}[X_T].$$

To które twierdzenie o zamianie wartości oczekiwanej i granicy zastosować zależy od kontekstu.

**Przykład 2.18**

Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie prostym symetrycznym spacerem losowym na  $\mathbb{Z}$ , tzn.  $X_n = Y_1 + \dots + Y_n$ , gdzie  $\{Y_n\}_{n \in \mathbb{N}}$  są niezależnymi zmiennymi losowymi przyjmującymi wartości  $+1$  i  $-1$  z prawdopodobieństwem  $1/2$ . Wówczas  $\{X_n\}_{n \in \mathbb{N}}$  jest martyngałem względem filtracji zadanej przez  $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ . Zdefiniujmy

$$T = \min\{n : X_n = -j \text{ lub } X_n = k\}.$$

Wówczas  $T$  jest czasem zatrzymania. Dodatkowo

$$\mathbb{P}[T > n] = \mathbb{P}[|X_n| \leq j \vee k] \rightarrow 0$$

na podstawie centralnego twierdzenia granicznego. Z twierdzenia Dooba mamy więc

$$\mathbb{E}[X_{T \wedge n}] = \mathbb{E}[X_0] = 0.$$

Aby wykonać przejścia graniczne zauważmy, że  $|X_{n \wedge T}| \leq k \vee j$ . Z twierdzenia o zbieżności ograniczonej mamy

$$\mathbb{E}[X_T] = 0.$$

Zauważmy wreszcie, że  $\mathbb{P}[X_T \in \{-j, k\}] = 1$  i stąd

$$\begin{aligned} 0 &= \mathbb{E}[X_0] = \mathbb{E}[X_T] = -j\mathbb{P}[X_T = -j] + k\mathbb{P}[X_T = k] \\ &= -j(1 - \mathbb{P}[X_T = k]) + k\mathbb{P}[X_T = k] = -j + (j+k)\mathbb{P}[X_T = k], \end{aligned}$$

a stąd

$$\mathbb{P}[X_T = k] = \frac{j}{j+k}.$$

Zauważmy że płynie stąd następujący wniosek

$$\mathbb{P}[\{X_n\}_{n \in \mathbb{N}} \text{ osiągnie } k] = \lim_{j \rightarrow \infty} \mathbb{P}[X_T = k] = \lim_{j \rightarrow \infty} \frac{j}{j+k} = 1.$$

Czasami bywa tak, że odpowiedniego martyngału należy poszukać.

**Przykład 2.19**

Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie niesymetrycznym prostym spacerem losowym na  $\mathbb{Z}$ , tzn.  $X_n = Y_1 + \dots + Y_n$ , gdzie  $\{Y_n\}_{n \in \mathbb{N}}$  są niezależnymi zmiennymi losowymi przyjmującymi wartości  $+1$  i  $-1$  z prawdopodobieństwem odpowiednio  $p$  i  $1-p$ . Załóżmy, że  $p > 1/2$ . Rozważmy ten sam problem jak w powyższym przykładzie, tzn. zdefiniujmy czas zatrzymania

$$T = \min\{n : X_n = -j \text{ lub } X_n = k\}.$$

i chcemy obliczyć  $\mathbb{P}[X_T = -j]$  oraz  $\mathbb{P}[X_T = k]$ . Zauważmy, że w tym przypadku  $\{X_n\}_{n \in \mathbb{N}}$  jest podmartyngałem, bo

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n + \mathbb{E}[Y_{n+1} | \mathcal{F}_n] = X_n + \mathbb{E}[Y_{n+1}] = X_n + (2p - 1) > X_n.$$

Mimo, że same  $\{X_n\}_{n \in \mathbb{N}}$  martyngałem nie jest, to można z nim związać wiele procesów, które martyngałami są. Spróbujmy znaleźć jeden z najprostszych. Będzie on funkcją  $X_n$ . Innymi słowy szukamy funkcji  $f: \mathbb{Z} \rightarrow \mathbb{R}$  takiej, aby ciąg  $\{f(X_n)\}_{n \in \mathbb{N}}$  tworzył martyngał względem naturalnej filtracji  $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ . Wówczas  $f$  powinna spełniać

$$\mathbb{E}[f(X_{n+1}) | \mathcal{F}_n] = f(X_n)$$

dla każdego  $n \in \mathbb{N}$ . Gdy poznamy podstawy teorii łańcuchów Markowa dowiemy się, że takie funkcje nazywamy harmonicznymi. Wracając jednak do warunków na funkcję  $f$ , zauważamy, że

$$\mathbb{E}[f(X_{n+1}) | \mathcal{F}_n] = \mathbb{E}[f(X_n + Y_{n+1}) | \mathcal{F}_n].$$

Skoro  $X_n$  jest  $\mathcal{F}_n$  mierzalna, a  $Y_{n+1}$  niezależna od  $\mathcal{F}_n$ , to

$$\mathbb{E}[f(X_n + Y_{n+1}) | \mathcal{F}_n] = F(X_n),$$

gdzie

$$F(x) = \mathbb{E}[f(x + Y_{n+1})] = pf(x + 1) + (1 - p)f(x - 1).$$

Ostatecznie

$$\mathbb{E}[f(X_{n+1}) | \mathcal{F}_n] = pf(X_n + 1) + (1 - p)f(X_n - 1).$$

Szukamy więc funkcji takiej, że

$$f(x) = pf(x + 1) + (1 - p)f(x - 1), \quad x \in \mathbb{Z}.$$

Znalezienie tego typu funkcji nie jest trudne. Sprawdzając funkcje postaci  $f(x) = \gamma^x$  dla pewnej  $\gamma \in \mathbb{R}$ , otrzymujemy równanie

$$1 = p\gamma + (1 - p)\gamma^{-1}$$

które posiada dwa rozwiązania  $\gamma = 1$  oraz  $\gamma = (1 - p)/p$ . Pierwszy wybór prowadzi do funkcji stałej dzięki której nie wyciągniemy żadnych wniosków. Wybór drugiego rozwiązania prowadzi do funkcji harmonicznej

$$f(x) = \left(\frac{1 - p}{p}\right)^x$$

jest szukaną funkcją harmoniczną. Z naszych rachunków wynika, że

$$f(X_n) = \left(\frac{1-p}{p}\right)^{X_n}$$

jest  $\mathbb{F}$ -martyngałem. Od tej chwili postępujemy jak w poprzednim przykładzie. Oznaczmy  $r = \mathbb{P}[X_T = k]$ . Korzystając z twierdzenia Doob'a i stosując takie samo przejście graniczne otrzymujemy

$$\mathbb{E}[f(X_T)] = \mathbb{E}[f(X_0)] = 1.$$

Pozostaje wykonać jeszcze prosty rachunek

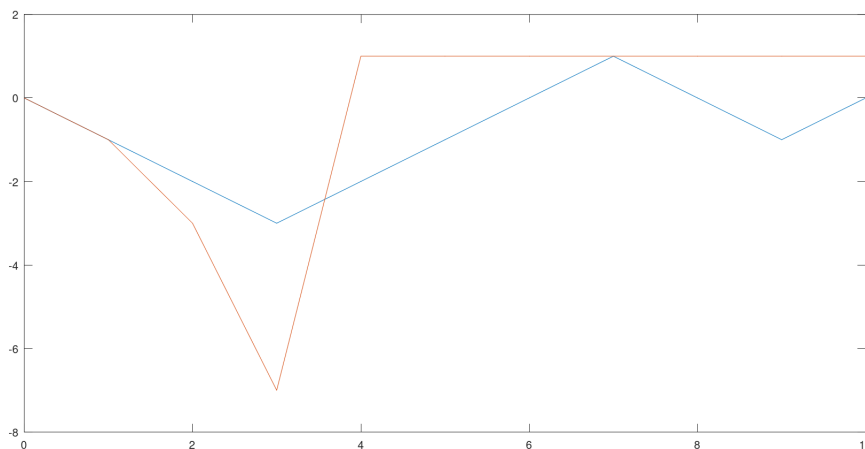
$$\begin{aligned} 1 &= \mathbb{E}[f(X_0)] = \mathbb{E}[f(X_T)] = f(-j)\mathbb{P}[X_T = -j] + f(k)\mathbb{P}[X_T = k] \\ &= (1-r)\gamma^{-j} + r\gamma^k. \end{aligned}$$

Co prowadzi do

$$r = \mathbb{P}[X_T = k] = \frac{1 - \gamma^{-j}}{\gamma^k - \gamma^{-j}} = \frac{1 - \gamma^j}{1 - \gamma^{k+j}}.$$

Przechodząc z  $k \rightarrow \infty$  otrzymujemy również

$$\mathbb{P}[X_n \text{ osiągnie } -j] = \lim_{k \rightarrow \infty} \mathbb{P}[X_T = -j] = \gamma^j.$$



**Rysunek 2.3.** Ilustracja procesów w przykładzie 2.20. Trajektoria prostego symetrycznego spaceru losowego (niebieska) oraz otrzymana z niej transformata  $W$

### Przykład 2.20

Rozważmy, podobnie jak w poprzednich przykładach ciąg niezależnych

zmiennych losowych  $\{Y_n\}_{n \in \mathbb{N}}$  takich, że  $\mathbb{P}[Y_n = 1] = \mathbb{P}[Y_n = -1] = 1/2$ . Przez  $\{W_n\}_{n \in \mathbb{N}}$  oznaczmy przewidywalny ciąg zmiennych losowych, tzn.  $W_n$  jest mierzalne względem  $\mathcal{F}_{n-1}$ . Możemy myśleć, że  $W_n$  jest to wysokość zakładu w  $n$ -tym kroku. Wówczas wygrana po  $n$  krokach wynosi

$$X_n = \sum_{j=1}^n Y_j W_j.$$

Wiemy już, że jest to martyngał. Zastosujmy następującą strategię: gramy do pierwszego momentu, gdy pojawi się 1 i za każdym razem podwajamy stawkę, tzn.  $W_j = 2^{j-1}$  i kończymy grę w chwili  $T = \min_n \{Y_n = 1\}$ .

Wiemy, że  $\mathbb{P}[T < \infty] = 1$  (dlaczego?). Zatem

$$X_T = -1 - 2 \cdot 1 - 4 \cdot 1 - \dots - 2^{T-1} + 2^T = -\frac{2^T - 1}{2 - 1} + 2^T = 1.$$

A więc zawsze, niezależnie od wartości  $T$ , wygramy kwotę 1, tzn. z prawdopodobieństwem 1 zachodzi:  $T < \infty$  oraz  $X_T = 1$ ! Zauważmy też, że

$$\mathbb{E}[X_0] = 0, \quad \mathbb{E}[X_T] = 1.$$

Nie zatem prawdą, że  $\mathbb{E}[X_0] = \mathbb{E}[X_T]$ . Powyższe rachunki pokazują więc, że mamy strategię wygrywającą mimo, iż gra jest sprawiedliwa. Zauważmy jednak, że aby być pewnym wygranej należy mieć nieograniczony kapitał:

$$\mathbb{E}[-X_{T-1}] = \sum_{k=1}^{\infty} (2^k - 1) \cdot \mathbb{P}[T = k] = \sum_{k=1}^{\infty} (2^k - 1) \cdot 2^{-k} = \infty. \quad \square$$

### Przykład 2.21

Rozważmy raz jeszcze prosty symetryczny spacer losowy na  $\mathbb{Z}$ . Proces  $\{X_n\}_{n \in \mathbb{N}}$  zadany jest przez  $X_n = \zeta_1 + \zeta_2 + \dots + \zeta_n$ , gdzie  $\{\zeta_k\}_{k \in \mathbb{N}}$  są iid takie, że  $\mathbb{P}[\zeta_k = \pm 1] = 1/2$ . Dla  $n \in \mathbb{N}$  rozważmy czas wyjścia z przedziału  $(-m, m)$  dany przez

$$T_m = \inf \{n \in \mathbb{N} : |X_n| = m\}.$$

Pokażemy, że  $T_m/m^2$  jest zbieżne według rozkładu. W tym celu sprawdzimy zbieżność transformaty Laplacea  $\mathbb{E} \exp\{-sT_m/m^2\}$  dla  $s > 0$ . Ustalmy w tym celu  $t \neq 0$  i rozważmy martyngał wykładniczy (**Przykład 2.10**)

$$Y_n(t) = e^{tX_n} \cosh(t)^{-n}.$$

Przypomnijmy, że  $\cosh(t) = (e^t + e^{-t})/2 = \mathbb{E}[e^{t\zeta_1}]$ . Z Twierdzenia Dooba, dla dowolnego  $t \neq 0$ ,

$$\mathbb{E}[Y_{T_m}(t)] = \mathbb{E}[Y_0(t)] = 1.$$

Stąd

$$1 = \mathbb{E} \left[ \frac{Y_{T_m}(t) + Y_{T_m}(-t)}{2} \right] = \mathbb{E}[\cosh(tX_{T_m}) \cosh(t)^{-T_m}].$$

Zauważmy, że  $|X_{T_m}| = m$ . Z parzystości funkcji  $\cosh$ ,  $\cosh(tX_{T_m}) = \cosh(tm)$  co prowadzi do tożsamości

$$\mathbb{E}[\cosh(t)^{-T_m}] = 1/\cosh(tm).$$

Dla  $s > 0$  niech  $t = t_{m,s} > 0$  będzie takie, że  $e^{s/m^2} = \cosh(t)$ . Skoro  $s/m^2 \rightarrow 0$ , to  $t = t_{m,s} \rightarrow 0$ . Porównując rozwinięcia funkcji

$$e^{s/m^2} = 1 + s/m^2 + o(s/m^2), \quad \cosh(t) = 1 + t^2/2 + o(t^2)$$

wnioskujemy, że  $m^2 t_{s,m}^2 \rightarrow 2s$ . Podsumowując

$$\mathbb{E} \left[ e^{-sT_m/m^2} \right] = \mathbb{E} \left[ \cosh(t_{m,s})^{-T_m} \right] = 1/\cosh(mt_{s,m}) \rightarrow 1/\cosh(\sqrt{2s}).$$

Stąd  $T_m/m^2$  jest zbieżne słabo do rozkładu, którego transformata Laplacea wynosi  $1/\cosh(\sqrt{2s})$ .

## 2.4 Twierdzenia o zbieżności martyngałów

Zajmiemy się teraz zagadnieniem badania asymptotycznego zachowania martyngałów. Rozważać będziemy martyngał  $X = \{X_n\}_{n \in \mathbb{N}}$ . Wówczas o martyngale  $X$  możemy myśleć jak o losowym ciągu. Przypomnijmy, że z definicji martyngału

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] = 0.$$

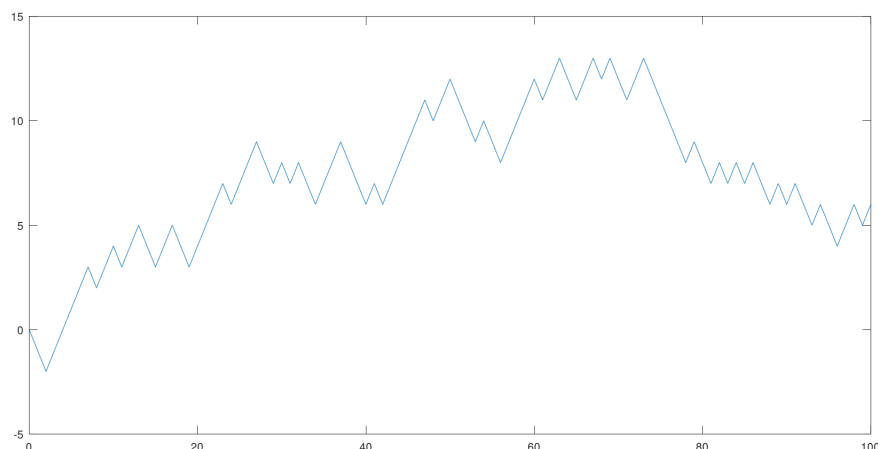
Innymi słowy ciąg  $X$  ma (warunkowo) scentrowane skoki. Oznacza to, że po warunkiem  $\mathcal{F}_n$  zmienna  $X_{n+1} - X_n$  może przyjmować zarówno dodatnie jak i ujemne wartości. W skutek czego trajektoria  $X$  (czyli wykres ciągu) charakteryzować się będzie dużym wahaniami (patrz Rysunek 2.4). Aby nabrać intuicji odnośnie skali tych wahań rozważmy następujące dwa przykłady.

### Przykład 2.22

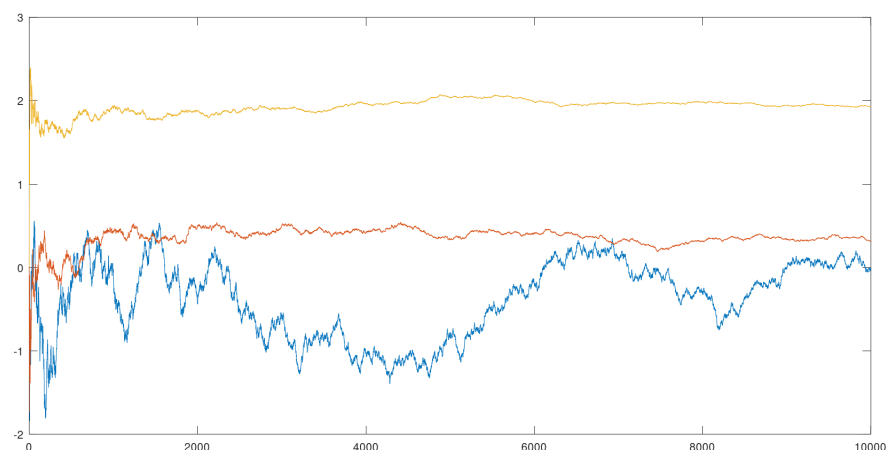
Niech  $\{Y_k\}_{k \in \mathbb{N}}$  będzie ciągiem niezależnych zmiennych takich, że  $Y_k$  ma rozkład  $\mathcal{N}(0, \sigma_k^2)$  dla pewnego ciągu wariancji  $\{\sigma_k^2\}_{k \in \mathbb{N}}$  takiego, że

$$\sum_{k=1}^{\infty} \sigma_k^2 < \infty.$$

Rozważmy ciąg  $M_0 = 0$  i



**Rysunek 2.4.** trajektoria pierwszych stu kroków symetrycznego spaceru losowego



**Rysunek 2.5.** Trajektorie martynałów z przykładu 2.22 z  $\sigma_k^2 = 1/k$  (niebieska),  $\sigma_k^2 = k^{-4/3}$  (czerwona) i  $\sigma_k = k^{-2}$  (pomarańczowa)

$$M_n = \sum_{k=1}^n Y_k.$$

Wówczas  $\{M_n\}_{n \in \mathbb{N}}$  jest  $\mathbb{F}$ -martynałem dla  $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ . Zauważmy, że ciąg  $\{M_n\}_{n \in \mathbb{N}}$  jest zbieżny na podstawie twierdzenia o dwóch szeregach. Istotnie, z nierówności maksymalnej Kołmogorowa

$$\mathbb{P} \left[ \max_{n \leq k \leq N} |M_k - M_n| \geq \epsilon \right] \leq \frac{1}{\epsilon^2} \sum_{k=n}^N \sigma_k^2$$

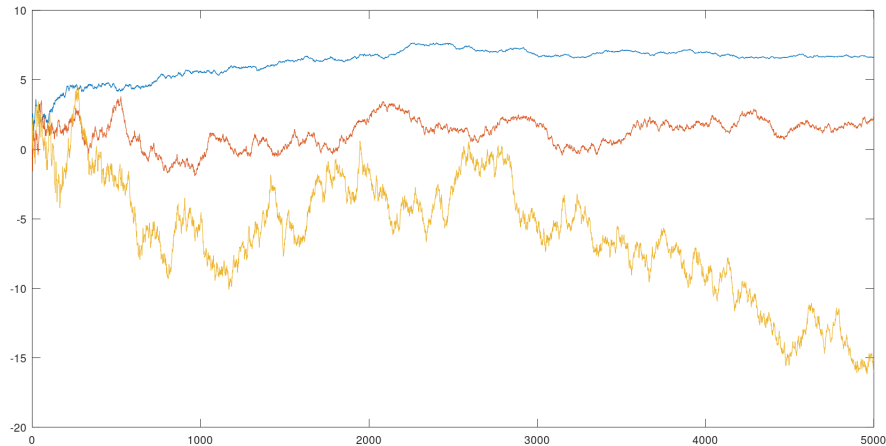
po przejściu z  $N \rightarrow \infty$ ,

$$\mathbb{P} \left[ \max_{n \leq k} |M_k - M_n| \geq \epsilon \right] \leq \frac{1}{\epsilon^2} \sum_{k=n}^{\infty} \sigma_k^2$$

Przy czym prawa strona zbiega do 0 przy  $n \rightarrow \infty$ . Oznacza to, że

$$\max_{n \leq k} |M_k - M_n| \rightarrow 0$$

według prawdopodobieństwa. Z monotoniczności ostatnia zbieżność zachodzi p.w. Oznacza to, że z prawdopodobieństwem jeden  $\{M_n\}_{n \in \mathbb{N}}$  jest ciągiem Cauchy'ego.



**Rysunek 2.6.** Trajektorie martyngałów z przykładu 2.23 z  $\sigma_k^2 = 1/k$  (niebieska),  $\sigma_k^2 = k^{-2/3}$  (czerwona) i  $\sigma_k^2 = k^{-2/5}$  (pomarańczowa)

### Przykład 2.23

W uzupełnieniu do poprzedniego przykładu niech  $\{Y_k\}_{k \in \mathbb{N}}$  będzie teraz ciągiem niezależnych zmiennych takich, że  $Y_k$  ma rozkład  $\mathcal{N}(0, \sigma_k^2)$  dla pewnego ciągu wariancji  $\{\sigma_k^2\}_{k \in \mathbb{N}}$  takiego, że

$$\sum_{k=1}^{\infty} \sigma_k^2 = \infty.$$

Rozważmy ciąg  $M_0 = 0$  i

$$M_n = \sum_{k=1}^n Y_k.$$

Wówczas  $\{M_n\}_{n \in \mathbb{N}}$  jest  $\mathbb{F}$ -martyngałem dla  $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ . Zauważmy, że istnieje ciąg liczb naturalnych  $\{n_k\}_{k \in \mathbb{N}}$  taki, że dla

$$s_k = \sum_{j=n_k+1}^{n_{k+1}} \sigma_j^2 \geq 1.$$

Wówczas, dla zmiennej  $N$  o standardowym rozkładzie normalnym

$$\mathbb{P}[|M_{n_{k+1}} - M_{n_k}| \geq 1] = \mathbb{P}[|N| > 1/\sqrt{s_k}] \geq \mathbb{P}[|N| > 1] > 0.$$

Z lematu Borela-Canteliego z prawdopodobieństwem jeden  $|M_{n_{k+1}} - M_{n_k}| \geq 1$  dla nieskończenie wielu  $k$ . Ciąg  $\{M_n\}_{n \in \mathbb{N}}$  nie może być zatem zbieżny.

**Twierdzenie 2.24** (Twierdzenie Doob'a o zbieżności martyngałów)  
Jeżeli  $\{X_n\}_{n \in \mathbb{N}}$  jest martyngałem takim, że

$$\sup_{n \in \mathbb{N}} \mathbb{E}[X_n^+] < \infty$$

to istnieje całkowalna zmienna losowa  $X$  taka, że  $X_n \rightarrow X$  p.w.

### Przykład 2.25

W nawiązaniu do przykładów 2.22 oraz 2.23 zauważmy, że

$$\mathbb{E}[M_n^+] = \left( \sum_{k=1}^n \sigma_k^2 \right)^{1/2} \mathbb{E}[N_+],$$

gdzie  $N$  ma standardowy rozkład normalny. Zatem kryterium w Twierdzeniu 2.24 jest w tym wypadku równoważne zbieżności martyngału.

### Przykład 2.26

Rozważmy  $\{X_n\}_{n \in \mathbb{N}}$  będący prostym symetrycznym spacerem losowym. Wówczas  $X_n/\sqrt{n}$  zbiega według rozkładu do  $\mathcal{N}(0,1)$ . Oznacza to, że  $X_n$  nie jest zbieżne do żadnej zmiennej losowej. Dzieje się tak ponieważ

$$\mathbb{E}[X_n^+] \geq \sqrt{n} \mathbb{P}[X_n \geq \sqrt{n}] \sim \sqrt{n} \mathbb{P}[\mathcal{N}(0,1) \geq 1].$$

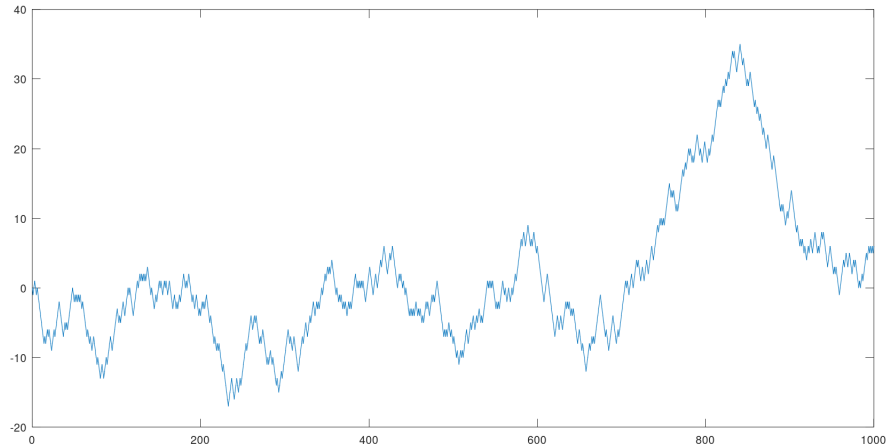
Nie są zatem spełnione założenia powyższego twierdzenia. Zauważmy, że  $\{X_n\}_{n \in \mathbb{N}}$  Warto się jednak zastanowić co się dzieje z typową trajekcją  $X = \{X_n\}_{n \in \mathbb{N}}$ , tj. z ciągiem liczb rzeczywistych  $X(\omega) = \{X_n(\omega)\}_{n \in \mathbb{N}}$  dla typowej  $\omega \in \Omega$  (pochodzącej ze zbioru po prawdopodobieństwie 1). Skoro

$$\mathbb{P}[X \text{ osiągnie } k] = 1$$

dla każdego  $k \in \mathbb{Z}$ , to

$$\mathbb{P}[X \text{ osiągnie każdy punkt } z \mathbb{Z}] = 1.$$

Oznacza to, że typowa trajekcja  $X(\omega)$  musi odwiedzić wszystkie liczby całkowite. Takie ciągi nie mogą być zbieżne.



**Rysunek 2.7.** Trajektoria prostego symetrycznego spaceru losowego

Aby udowodnić Twierdzenie 2.24 będziemy musieli wprowadzić nieco notacji. Możemy założyć bez zmniejszania ogólności, że  $X_0 = 0$  (W przeciwnym razie możemy rozpatrywać  $X'_n = X_n - X_0$ ). Ustalmy  $a < b$ . Policzmy liczbę fluktuacji  $\{X_n\}_{n \in \mathbb{N}}$  wokół przedziału  $(a, b)$ . Zdefiniujmy ciąg czasów zatrzymania

$$\begin{aligned} T_0 &= \inf\{n : X_n < a\}, \\ T_1 &= \inf\{n > T_0 : X_n > b\}, \\ &\dots \\ T_{2k} &= \inf\{n > T_{2k-1} : X_n < a\}, \\ T_{2k+1} &= \inf\{n > T_{2k} : X_n > b\}. \end{aligned}$$

Oznaczmy przez  $U_{a,b}$  liczbę przejść  $X$  w górę przez przedział  $(a, b)$ , tj.

$$U_{a,b} = \begin{cases} \sup\{k : T_{2k-1} < \infty\} & \text{jeżeli } T_1 < \infty \\ 0 & \text{w przeciwnym razie} \end{cases}$$

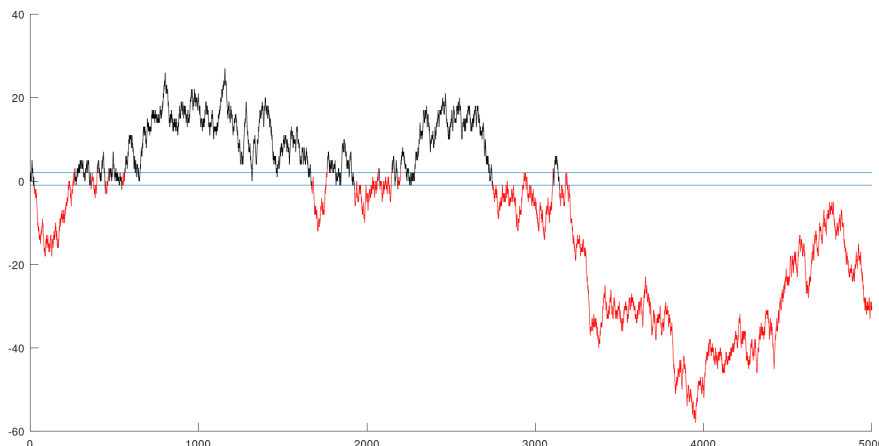
**Lemat 2.27**

Jeżeli  $X = \{X_n\}_{n \in \mathbb{N}}$  jest martyngałem takim, że

$$\sup_{n \in \mathbb{N}} \mathbb{E}[X_n^+] < \infty$$

to dla dowolnych  $a < b$  liczba przejść  $U_{a,b}$  w górę przez przedział  $(a, b)$  spełnia

$$\mathbb{E}[U_{a,b}] \leq \frac{1}{b-a} \sup_{n \in \mathbb{N}} \mathbb{E}[(X_n - a)_+] < \infty.$$



**Rysunek 2.8.** Ilustracja strategii inwestycyjnej z dowodu Twierdzenia 2.24. Czerwone fragmenty trajektorii reprezentują okresy inwestycji.

*Dowód.* Oznaczmy również przez  $U_{a,b}(n)$  liczbę przejść do czasu  $n$ , tj.

$$U_{a,b}(n) = \begin{cases} \sup\{k : T_{2k-1} \leq n\} & \text{jeżeli } T_1 \leq n \\ 0 & \text{w przeciwnym razie} \end{cases}$$

**Krok 1.** Napiszmy

$$X_n = \Delta_1 + \dots + \Delta_n,$$

dla  $\Delta_j = X_j - X_{j-1}$ . Rozważmy transformatę martyngałową

$$W_n = \sum_{j=1}^n B_j \Delta_j, \quad B_j \in \sigma(X_1, \dots, X_{j-1}).$$

Zmienne losowe  $B_j$  będą przyjmować wartości 0 lub 1. Przypomnijmy, że wówczas  $\{W_n\}_{n \in \mathbb{N}}$  jest martyngałem:

$$\begin{aligned} \mathbb{E}[W_n | \mathcal{F}_{n-1}] &= \mathbb{E}\left[\sum_{j=1}^n B_j \Delta_j | \mathcal{F}_{n-1}\right] = W_{n-1} + B_n \mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] \\ &= W_{n-1} + B_n (\mathbb{E}[X_n | \mathcal{F}_{n-1}] - X_{n-1}) = W_{n-1} \end{aligned}$$

oraz ponadto  $\mathbb{E}[W_n] = 0$ .

Chcemy 'ustawić' wartości  $B_j$  zgodnie z zasadą 'tanio kupić, drogo sprzedać'. Jeżeli proces  $\{X_n\}_{n \in \mathbb{N}}$  jest poniżej  $a$ , to spodziewamy się, że w pewnym momencie zacznie rosnąć i zaczynamy obstawiać. Kładziemy

$$B_m = \begin{cases} 1, & \text{jeżeli } T_{2k} < m \leq T_{2k+1} \text{ dla pewnego } k \\ 0 & \text{poza tym} \end{cases}$$

Wówczas  $B_m \in \mathcal{F}_{m-1}$ , bo  $\{T_{2k} < m \leq T_{2k+1}\} = \{T_{2k} \leq m-1\} \cap \{T_{2k+1} \leq m-1\}^c$ . Wartości  $B_m$  wyglądają następująco

- $B_j = 0$  dla  $j < T_0$ ,
- $B_j = 1$  dla  $T_0 \leq j < T_1$ ,
- $B_j = 0$  dla  $T_1 \leq j < T_2$ ,
- ...

Za każdym razem, gdy nastąpi przecięcie z dołu do góry przedziału  $(a, b)$  gracz wygrywa co najmniej kwotę  $b - a$ . Stąd

$$W_n \geq U_{a,b}(n)(b-a) + (X_n - a)\mathbf{1}_{\{X_n \leq a\}}. \quad (2.4.1)$$

Istotnie, zauważmy najpierw  $U_{a,b}(n)$  przejść przez przedział  $(a, b)$  dało nam zysk  $U_{a,b}(n)(b-a)$ . Jednak na zbiorze  $\{X_n \leq a\}$  mogliśmy ponieść stratę. Oznaczmy przez  $T$  ostatni moment, gdy proces znalazł się poniżej  $a$ . Od tego momentu obstawialiśmy 1, więc straciliśmy  $\Delta_T + \dots + \Delta_n = X_n - X_T \geq X_n - a$ .

Ponieważ  $\mathbb{E}W_n = 0$ , korzystając z (2.4.1) mamy

$$\mathbb{E}[U_{a,b}(n)] \leq \frac{\mathbb{E}(X_n - a)_+}{b - a}$$

i przechodząc z  $n \rightarrow \infty$

$$\mathbb{E}[U_{a,b}] \leq \frac{1}{b - a} \sup_{n \in \mathbb{N}} \mathbb{E}[(X_n - a)_+].$$

□

Zatem  $U_{a,b} < \infty$  p.w. Zachodzi to dla wszystkich możliwych wartości  $a, b$ . W dowodzie twierdzenia skorzystamy jeszcze z jednego pomocniczego faktu z analizy.

### Fakt 2.28

Ciąg liczb rzeczywistych  $\{x_n\}_{n \in \mathbb{N}}$  jest zbieżny (być może do  $\pm\infty$ ) wtedy i tylko wtedy, gdy dla wszystkich  $a, b \in \mathbb{Q}$ , liczba  $u_{a,b}$  przejść w górę przez przedział  $(a, b)$  jest skończona.

*Dowód.*  $\implies$  nie wprost. Jeżeli istnieją  $a, b$  takie, że  $U_{a,b} = \infty$ , to istnieją podciągi  $\{x_{n_k}\} < a$  i  $\{x_{n_k}\} > b$ . A zatem ciąg  $\{x_n\}$  nie może być zbieżny.

$\impliedby$  Załóżmy, że  $\liminf x_n < \limsup x_n$ . Wówczas istnieją liczby wymierne  $a, b$  takie, że  $\liminf x_n < a < b < \limsup x_n$ , a zatem  $U_{a,b} = \infty$ .

□

*Dowód Twierdzenia 2.24.* Wracając do martyngału  $\{X_n\}$ , korzystając z powyższej własności możemy napisać

$$\begin{aligned}
\mathbb{P}[\{X_n\} \text{ nie jest zbieżny}] &\stackrel{\text{krok2}}{=} \mathbb{P}[\text{istnieją } a, b \in \mathbb{Q} \text{ takie, że } U_{a,b} = \infty] \\
&= \mathbb{P}\left[\bigcup_{a < b, a, b \in \mathbb{Q}} \{U_{a,b} = \infty\}\right] \\
&\leq \sum_{a < b, a, b \in \mathbb{Q}} \mathbb{P}[U_{a,b} = \infty] \stackrel{\text{krok1}}{=} 0.
\end{aligned}$$

Powyższa własność pokazuje więc, że ciąg  $\{X_n\}$  jest zbieżny p.w. do  $X$ , ale jego granicą może być  $\pm\infty$ . Zauważmy, że z lematu Fatou

$$\mathbb{E}X_+ \leq \mathbb{E} \liminf_n X_n^+ \leq \sup_{n \in \mathbb{N}} \mathbb{E}[X_n^+].$$

Uzasadnimy teraz, że  $\mathbb{E}[X_-] < \infty$ . Mamy

$$\mathbb{E}[X_n^-] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n] = \mathbb{E}[X_n^+] - \mathbb{E}[X_0]$$

co daje

$$\mathbb{E}X_- \leq \mathbb{E} \liminf_n X_n^- \leq \sup_{n \in \mathbb{N}} \mathbb{E}[X_n^+] - \mathbb{E}[X_0].$$

Zatem  $|X| < \infty$  p.w., co kończy dowód twierdzenia.  $\square$

### Uwaga 2.29

Nie każdy martyngał jest zbieżny. Dla przykładu jeżeli  $X_n = Y_1 + \dots + Y_n$ , gdzie  $Y_n$  są niezależne i  $Y_n = \pm 1$  z prawdopodobieństwem  $1/2$  to martyngał nie jest zbieżny. Proces ten oscyluje pomiędzy  $-\infty$  i  $+\infty$ . W szczególności zachodzi  $\sup_n \mathbb{E}|X_n| = \infty$  (można to sprawdzić np. korzystając z centralnego twierdzenia granicznego).

### Wniosek 2.30

Jeżeli  $\{X_n\}$  jest nieujemnym martyngałem, to jest on zbieżny.

*Dowód.* Nieujemny martyngał spełnia założenia poprzedniego twierdzenia, istotnie dla każdego  $n$

$$\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_0 < C.$$

$\square$

Ogólniejsza wersja twierdzenia Doob'a:

**Twierdzenie 2.31** (Twierdzenie Doob'a o zbieżności nadmartyngałów)  
Załóżmy, że proces  $X_n$  jest nadmartyngałem tzn.  $\sup_n \mathbb{E}X_n^- < \infty$ . Wtedy ciąg  $\{X_n\}$  jest zbieżny do pewnej całkowalnej zmiennej losowej  $X$ .

W sformułowaniu twierdzenia  $X^-$  oznacza część ujemną  $X$ , tj.  $X^- = \max\{-X, 0\}$ . Dowód twierdzenia pomijamy. Bazuje on na podobnych technikach jak dowód twierdzenia 2.24.

- Wniosek 2.32**
1. Każdy nieujemny nadmartyngał/martyngał  $X_n$  jest zbieżny p.w.
  2. Jeżeli podmartyngał  $X_n$  spełnia  $\sup_n \mathbb{E}X_n^+ < \infty$ , to  $X_n$  jest zbieżny p.w.
  3. Jeżeli  $X_n$  jest nadmartyngałem, to warunek  $\sup_n \mathbb{E}X_n^- < \infty$  jest równoważny  $\sup_n \mathbb{E}|X_n| < \infty$

*Dowód.* Do dowodu 2. wystarczy rozpatrzyć  $-X_n$ . Punkt 3. wynika z

$$|X_n| = X_n^+ + X_n^- = X_n + 2X_n^-,$$

co implikuje

$$\mathbb{E}|X_n| \leq \mathbb{E}X_0 + 2 \sup_n \mathbb{E}X_n^- < \infty.$$

□

### Przykład 2.33

Niech  $X_n$  będzie procesem gałęzkowym o średniej  $m$ . Wtedy  $X_n/m^n$  jest nieujemnym martyngałem, a więc jest zbieżny do pewnej zmiennej losowej  $X$ . Jeżeli wiemy, że granica jest nietrywialna, to  $X_n$  rośnie jak  $m^n X$ .

### Przykład 2.34

Rozważmy model referendum, w którym każdy z głosujących jest oportunistą. Każdy głosujący będzie przyjmował losowo dwie opinie na ustalony temat, powiedzmy 0 i 1. Niech  $\Lambda \subseteq \mathbb{Z}^d$  będzie zbiorem głosujących. Bez zmniejszania ogólności założymy, że  $\Lambda = \{0, 1, \dots, L-1\}^d$  dla pewnych  $d, L \in \mathbb{N}$ . Każdy punkt  $j \in \Lambda$  utożsamiamy z głosującym. Powiemy, że  $i = (i_1, \dots, i_d)$  oraz  $j = (j_1, \dots, j_d)$  są znajomymi (sąsiadami) jeżeli  $i_k - j_k = 1 \pmod{L}$ . Innymi słowy  $\Lambda$  jest  $d$  wymiarowym torusem. Zauważmy, że jeżeli  $d = 1$ , to  $\Lambda$  traktujemy jako cykl.

Powiemy, że system jest w stanie  $x \in \{0, 1\}^\Lambda$ , jeżeli  $x(j)$  jest opinią  $j \in \Lambda$ . Załóżmy, że zaczynamy od dowolnego ustalonego stanu początkowego  $X_0 = x_0 \in \{0, 1\}^\Lambda$ . W każdej chwili czasu losowo wybrany wierzchołek  $I_n$  przejmuje opinię jednego ze swoich znajomych (sąsiadów). Oznaczmy przez  $X_n$  stan systemu w chwili  $n$ . Wówczas

$$X_n(i) = \begin{cases} X_{n-1}(i) & \text{jeżeli } i \neq I_n \\ X_{n-1}(I_n + N_n) & \text{jeżeli } I_n = i \end{cases}$$

Tutaj  $N_n$  jest losowo wybranym elementem  $2d$  elementowego zbioru

$$\{(\pm 1, 0, \dots, 0), (0, \pm 1, 0, \dots, 0), \dots, (0, \dots, 0, \pm 1)\}.$$

Dodawanie  $I_n + N_n$  w definicji  $X_n(i)$  jest wykonywane modulo  $L$ .

Rozważmy ciąg reprezentujący liczbę głosujących o opinii 1, czyli

$$M_n = \sum_{i \in \Lambda} X_n(i)$$

przez filtrację  $\mathcal{F}_n = \sigma(I_k, N_k : k \leq n)$ . Wówczas  $\{M_n\}_{n \in \mathbb{N}}$  jest  $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$  martyngałem. Mamy

$$\begin{aligned} \mathbb{E}[M_n | \mathcal{F}_{n-1}] &= M_{n-1} - \mathbb{E}[X_{n-1}(I_n) | \mathcal{F}_{n-1}] + \mathbb{E}[X_{n-1}(I_n + N_n) | \mathcal{F}_{n-1}] = \\ &= M_{n-1} - \sum_{i \in \Lambda} \mathbb{P}[I_n = i] X_{n-1}(i) + \sum_{i \in \Lambda} \mathbb{P}[I_n + N_n = i] X_{n-1}(i) = M_{n-1}. \end{aligned}$$

Skoro  $M_n$  jest nieujemny, to jest on zbieżny do pewnego  $M_\infty$ . Skoro  $M_n$  przyjmuje wartości całkowite, to  $M_n = M_\infty$  p.w. dla dostatecznie dużych  $n \in \mathbb{N}$ . Zauważmy, że jedynymi stanami stabilnymi są  $x \equiv 1$  oraz  $x \equiv 0$ . Rzeczywiście, jeżeli  $y$  jest stanem, który nie jest stały, to istnieją sąsiadujące  $i$  oraz  $j$  takie, że  $y(i) \neq y(j)$ . Wówczas

$$\mathbb{P}[X_n \neq X_{n-1} | : X_{n-1} = y] \geq \mathbb{P}[I_{n-1} = i, N_n = j - i] = 1/(2dL^d).$$

Oznacza to, że  $M_\infty \in \{0, L^d\}$  p.w. Z drugiej strony  $M_n$  są ograniczone, więc z twierdzenia o zbieżności ograniczonej

$$\mathbb{E}[M_\infty] = \lim_{n \rightarrow \infty} \mathbb{E}[M_n] = M_0 = \sum_{i \in \Lambda} x(i),$$

gdzie  $x$  jest naszym deterministycznym stanem początkowym. Oznacza to, że

$$\mathbb{P}[M_\infty = L^d] = M_0/L^d, \quad \mathbb{P}[M_\infty = 0] = 1 - M_0/L^d.$$

## 2.5 Jednostajna całkowalność

Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie dowolnym ciągiem zmiennych losowych. Wiemy, że  $X_n \rightarrow X$  p.w. nie wystarcza do tego aby  $\mathbb{E}X_n \rightarrow \mathbb{E}X$ . Najprostrzym przykładem (który nie jest martyngałem) może być ciąg  $X_n = n \mathbb{1}_{[0, 1/n]}$  na  $\Omega = [0, 1]$ . Wówczas  $X_n \rightarrow 0$  p.w. ale  $\mathbb{E}X_n = 1 \not\rightarrow 0$ . Dzieje się tak, ponieważ masa  $X_n$  jako funkcji  $X_n: [0, 1] \rightarrow \mathbb{R}$  są skupione na niemal rozłącznych kawałkach. To nieformalne wyjaśnienie jest doprecyzowane przez koncept jednostajnej całkowalności.

### Definicja 2.35

Proces  $(X_n)_{n \in \mathbb{N}}$  nazywamy *jednostajnie całkowalnym*, jeżeli

$$\lim_{C \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E} \left[ |X_n| \mathbf{1}_{\{|X_n| > C\}} \right] = 0.$$

Zauważmy, że przytoczony proces  $X_n = n \mathbf{1}_{[0, 1/n]}$  nie jest jednostajnie całkowalny ponieważ przy ustalonym  $C$ , dla  $n > C$  mamy  $\mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > C\}}] = 1$ .

### Przykład 2.36

Niech  $Y_n$  będzie ciągiem niezależnych zmiennych losowych takich, że  $Y_n = \pm 1$  z prawdopodobieństwem  $1/2$ . Niech  $W_n = 2^{n-1}$ . Wówczas  $X_n = \sum_{j=1}^n W_j Y_j$  jest martyngałem. Nie jest on jednostajnie całkowalny, gdyż

$$\mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| \geq 2^n - 1\}}] = \frac{1}{2^n} \cdot (2^n - 1) \sim 1.$$

Biorąc więc małe  $\varepsilon > 0$  nie istnieje  $K$  takie, że

$$\mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| \geq K\}}] < \varepsilon.$$

### Przykład 2.37

Proces wspólnie ograniczony przez zmienną całkowalną, tzn. dla wszystkich  $n \in \mathbb{N}$ ,  $|X_n| \leq Y$ , gdzie  $\mathbb{E}Y < \infty$  jest jednostajnie całkowalny. Rzeczywiście

$$\lim_{C \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E} \left[ |X_n| \mathbf{1}_{\{|X_n| > C\}} \right] \leq \lim_{C \rightarrow \infty} \mathbb{E} \left[ |Y| \mathbf{1}_{\{Y > C\}} \right] = 0.$$

### Przykład 2.38

Rodzina stała jest jednostajnie całkowalna. Dla  $X_n = X$  takiego, że  $\mathbb{E}|X| < \infty$  z poprzedniego przykładu widzimy, że  $\{X_n\}_{n \in \mathbb{N}}$  jest jednostajnie całkowalna.

### Przykład 2.39

Niech  $X_n = \mathbb{E}[X | \mathcal{F}_n]$  dla zmiennej  $X \in \mathcal{F}$  takiej, że  $\mathbb{E}|X| < \infty$ . Mówimy wtedy, że  $X_n$  jest martyngałem zamkniętym przez zmienną  $X$ . Wówczas proces  $(X_n)_{n \in \mathbb{N}}$  jest jednostajnie całkowalny. Aby się o tym przekonać zauważmy najpierw, że

$$\mathbb{E}[|X_n|] = \mathbb{E}[|\mathbb{E}[X | \mathcal{F}_n]|] \leq \mathbb{E}[\mathbb{E}[|X| | \mathcal{F}_n]] = \mathbb{E}[|X|].$$

Miara  $\mu(d\omega) = |X(\omega)|\mathbb{P}(d\omega)$  jest absolutnie ciągła względem prawdopodobieństwa  $\mathbb{P}$  ( $\mu \ll \mathbb{P}$ ). Zatem dla dowolnego  $\varepsilon > 0$  istnieje  $\delta > 0$  taka, że

$$\mathbb{P}[A] < \delta \Rightarrow \mathbb{E}[|X| \mathbf{1}_A] = \mu(A) < \varepsilon.$$

Dla  $C \geq \delta^{-1} \mathbb{E}[|X|]$  mamy

$$\mathbb{P}[|X_n| > C] \leq \frac{\mathbb{E}|X_n|}{C} \leq \frac{\mathbb{E}|X|}{C} \leq \delta.$$

Wówczas

$$\mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| > C\}}] \leq \mathbb{E}[\mathbb{E}[|X| \mathbb{1}_{\{|X_n| > C\}} | \mathcal{F}_n]] = \mathbb{E}[|X| \mathbb{1}_{\{|X_n| > C\}}] \leq \varepsilon.$$

**Fakt 2.40**

Proces  $X = (X_n)_{n \in \mathbb{N}}$  jest jednostajnie całkowalny wtedy i tylko wtedy, gdy spełnione są następujące dwa warunki:

- a)  $\sup_{n \in \mathbb{N}} \mathbb{E}|X_n| < \infty$ ;
- b) Dla dowolnego  $\varepsilon > 0$  istnieje  $\delta > 0$  taka, że

$$\mathbb{P}[A] < \delta \Rightarrow \sup_{n \in T} \mathbb{E}[|X_n| \mathbb{1}_A] < \varepsilon.$$

*Dowód.* Załóżmy, że  $X$  jest jednostajnie całkowalny. Dla  $\varepsilon > 0$  możemy dobrać  $C > 0$  tak duże, aby  $\sup_{n \in T} \mathbb{E}|X_n| \mathbb{1}_{\{|X_n| > C\}} \leq \varepsilon/2$ . Wtedy

$$\sup_{n \in \mathbb{N}} \mathbb{E}|X_n| \leq \sup_{n \in \mathbb{N}} \mathbb{E} \left[ |X_n| \mathbb{1}_{\{|X_n| > C\}} \right] + C \leq \frac{\varepsilon}{2} + C < \infty,$$

więc spełniony jest warunek a). Dla  $\mathbb{P}[A] < \delta = \frac{\varepsilon}{2C}$  mamy

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n| \mathbb{1}_A] \leq \sup_{n \in \mathbb{N}} \mathbb{E} \left[ |X_n| \mathbb{1}_A \mathbb{1}_{\{|X_n| > C\}} \right] + C \mathbb{P}[A] < \varepsilon,$$

co pokazuje słuszność warunku b). Aby uzasadnić implikację odwrotną do  $\varepsilon > 0$ , dobieramy  $\delta > 0$  jak w warunku b). Niech  $C > \delta^{-1} \sup_{n \in \mathbb{N}} \mathbb{E}|X_n|$ . Wówczas dla dowolnego  $n \in \mathbb{N}$ ,

$$\mathbb{P}[|X_n| > C] \leq \frac{\mathbb{E}|X_n|}{C} \leq \frac{\sup_{n \in \mathbb{N}} \mathbb{E}|X_n|}{C} < \delta$$

więc

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left[ |X_n| \mathbb{1}_{\{|X_n| > C\}} \right] \leq \varepsilon.$$

Dowodzi to jednostajnej całkowalności  $X$ . □

Dla ciągu  $(X_n)_{n \in \mathbb{N}}$  i  $p \geq 1$  wiemy, że jeżeli  $X_n \rightarrow^{L^p(\Omega)} X_\infty$ , tzn.

$$\|X_n - X_\infty\|_{L^p(\Omega)} = (\mathbb{E}[|X_n - X_\infty|^p])^{1/p} \rightarrow 0,$$

to z nierówności Markowa  $X_n \rightarrow^{\mathbb{P}} X_\infty$ . Implikację tę można odwrócić przy założeniu jednostajnej całkowalności  $(|X_n|^p)_{n \in \mathbb{N}}$ .

**Fakt 2.41**

Niech  $p \geq 1$ . Załóżmy, że proces  $(|X_n|^p)_{n \in \mathbb{N}}$  jest jednostajnie całkowalny. Jeśli  $X_n \xrightarrow{\mathbb{P}} X_\infty$ , to  $X_n \xrightarrow{L^p(\Omega)} X_\infty$ .

*Dowód.* Istnieje podciąg  $n_k$  taki, że  $X_{n_k} \xrightarrow{p.w.} X$ . Z lematu Fatou dla dowolnego  $A \in \mathcal{F}$ ,

$$\mathbb{E}[|X|^p \mathbb{1}_A] = \mathbb{E}\left[\lim_{k \rightarrow \infty} |X_{n_k}|^p \mathbb{1}_A\right] \leq \liminf_{k \rightarrow \infty} \mathbb{E}[|X_{n_k}|^p \mathbb{1}_A] \leq \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p \mathbb{1}_A].$$

W szczególności, sprawdzając powyższe dla  $A = \Omega$ ,  $\mathbb{E}|X| < \infty$ . Dla  $\varepsilon > 0$  możemy dobrać  $\delta > 0$  taką, że jeżeli  $\mathbb{P}[A] < \delta$ , to

$$\mathbb{E}[|X|^p \mathbb{1}_A] \leq \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p \mathbb{1}_A] < \varepsilon.$$

Dla dostatecznie dużych  $n$ ,  $\mathbb{P}[|X - X_n| > \varepsilon] < \delta$ . Wobec tego

$$\begin{aligned} \mathbb{E}|X - X_n|^p &= \mathbb{E}|X - X_n|^p \mathbb{1}_{\{|X - X_n| > \varepsilon\}} + \mathbb{E}|X - X_n|^p \mathbb{1}_{\{|X - X_n| \leq \varepsilon\}} \\ &\leq \mathbb{E}|X - X_n|^p \mathbb{1}_{\{|X - X_n| > \varepsilon\}} + \varepsilon^p \\ &\leq \mathbb{E}2^{p-1}(|X| + |X_n|)^p \mathbb{1}_{\{|X - X_n| > \varepsilon\}} + \varepsilon \leq 2^p \varepsilon + \varepsilon^p. \end{aligned}$$

□

## 2.6 Zbieżność martyngałów w $L^1$

**Twierdzenie 2.42** (Twierdzenie o zbieżności martyngałów w  $L^1$ )

Załóżmy, że  $\{X_n\}$  jest martyngałem. Następujące warunki są równoważne

1. rodzina  $\{X_n\}$  jest jednostajnie całkowalna;
2. ciąg  $\{X_n\}$  jest zbieżny w  $L^1$  do  $X_\infty$  (tzn.  $\mathbb{E}|X_n - X_\infty| \rightarrow 0$ );
3. istnieje  $X_\infty \in L^1$  takie, że  $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ .

**Uwaga 2.43**

Jeżeli zachodzi jeden z powyższych warunków w twierdzeniu, to  $\sup_n \mathbb{E}|X_n| < M$ , a to z kolei implikuje  $X_n \rightarrow X$  p.w. (z twierdzenia 2.24) oraz w  $L^1$ .

*Dowód.* 1.  $\implies$  2. Z **Faktu 2.40**,

$$\sup_{n \in \mathbb{N}} \mathbb{E}X_n^+ \leq \sup_{n \in \mathbb{N}} \mathbb{E}|X_n| < \infty.$$

Stąd  $X_n \rightarrow X_\infty$  dla pewnej całkowalnej zmiennej  $X_\infty$ . Z **Faktu 2.41** zastosowanego dla  $p = 1$  wynika, że zbieżność zachodzi także w  $L^1$ .

2.  $\implies$  3. Niech  $X_n \rightarrow X_\infty$  w  $L^1$ . Dla dowolnych  $n, k \in \mathbb{N}$  oraz  $A \in \mathcal{F}_n$ , poprzez odwołanie się do definicji martyngału oraz warunkowej wartości oczekiwanej,

$$\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[X_{n+k} \mathbb{1}_A]$$

Zauważmy, że ze zbieżności w  $L^1$  wynika, że

$$|\mathbb{E}[X_{n+k} \mathbb{1}_A] - \mathbb{E}[X_\infty \mathbb{1}_A]| \leq \mathbb{E}[|X_{n+k} - X_\infty| \mathbb{1}_A] \leq \mathbb{E}[|X_{n+k} - X_\infty|] \rightarrow 0$$

przy  $k \rightarrow \infty$ . Oznacza to, że dla dowolnego  $A \in \mathcal{F}_n$ ,

$$\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[X_\infty \mathbb{1}_A].$$

Czyli  $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ . 3.  $\implies$  1. Tę implikację pokazaliśmy w **Przykładzie 2.39**. □

## 2.7 Nierówności maksymalne

Jedną z bardziej przydatnych własności martyngałów jest to, że momenty funkcji maksymalnej, tj. dla martyngału  $M = \{M_n\}_{n \in \mathbb{N}}$ ,

$$\sup_{n \in \mathbb{N}} |M_n|$$

są kontrolowane przez momenty zmiennych  $M_n$  dla  $n \in \mathbb{N}$ . Jak się za chwilę przekonamy jest to bezpośrednią konsekwencją twierdzenia o zatrzymaniu. Niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie  $\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}}$ -podmartyngałem. Przypomnijmy, że wówczas dla każdego  $n \in \mathbb{N}$

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n.$$

W odpowiedniku twierdzenia o zatrzymaniu dla martyngałów pojawia się naturalne nierówność.

**Twierdzenie 2.44** (Twierdzenie o zatrzymaniu dla podmartyngałów)  
Jeżeli  $X = \{X_n\}_{n \in \mathbb{N}}$  jest  $\mathbb{F}$ -podmartyngałem a  $T$  jest  $\mathbb{F}$ -czasem zatrzymania, to ciąg  $\{X_{T \wedge n}\}_{n \in \mathbb{N}}$  jest  $\mathbb{F}$ -podmartyngałem. W szczególności

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_{T \wedge n}]$$

*Dowód.* Piszemy

$$X_n - X_{T \wedge n} = \sum_{k=T \wedge n+1}^n X_k - X_{k-1} = \sum_{k=1}^n \mathbb{1}_{\{k-1 \geq T\}} (X_k - X_{k-1})$$

Przy czym każda ze zmiennych pod sumą ma dodatnią wartość oczekiwaną, bo

$$\mathbb{E}[\mathbb{1}_{\{k-1 \geq T\}}(X_k - X_{k-1}) | \mathcal{F}_{k-1}] = \mathbb{1}_{\{k-1 \geq T\}} \mathbb{E}[X_k - X_{k-1} | \mathcal{F}_{k-1}] \geq 0.$$

Stąd

$$\mathbb{E}[\mathbb{1}_{\{k-1 \geq T\}}(X_k - X_{k-1})] \geq 0$$

dla każdego  $k \in \mathbb{N}$  a co za tym idzie  $\mathbb{E}[X_n - X_{n \wedge T}] \geq 0$ .  $\square$

Dla liczby naturalnej  $N$  będziemy stosowali oznaczenie  $[N] = \{1, 2, \dots, N\}$ .

**Fakt 2.45**

Jeżeli  $X = (X_n)_{n \in [N]}$  jest podmartyngałem, to dla każdej  $\lambda > 0$ ,

$$\lambda \mathbb{P} \left[ \sup_{n \in [N]} X_n \geq \lambda \right] \leq \mathbb{E} \left[ X_N \mathbb{1}_{\{\sup_{n \in [N]} X_n \geq \lambda\}} \right].$$

*Dowód.* Zdefiniujmy czas zatrzymania

$$\tau = \min\{n \in [N] : X_n \geq \lambda\}.$$

Wówczas  $X_\tau \geq \lambda$  na zbiorze  $\{\sup_{n \in [N]} X_n \geq \lambda\}$ . Na mocy twierdzenia o zatrzymaniu dla podmartyngałów otrzymujemy

$$\begin{aligned} \mathbb{E}X_N &\geq \mathbb{E}X_{\tau \wedge T} = \mathbb{E} \left[ X_\tau \mathbb{1}_{\{\sup_{n \in [N]} X_n \geq \lambda\}} \right] + \mathbb{E} \left[ X_N \mathbb{1}_{\{\sup_{n \in [N]} X_n < \lambda\}} \right] \\ &\geq \lambda \mathbb{P} \left[ \sup_{n \in [N]} X_n \geq \lambda \right] + \mathbb{E} \left[ X_N \mathbb{1}_{\{\sup_{n \in [N]} X_n < \lambda\}} \right], \end{aligned}$$

co pociąga tezę.  $\square$

Jeżeli  $X = (X_n)_{n \in [N]}$  jest martyngałem, a zmienna  $X_N$  jest w  $L^p$ , to korzystając z nierówności Jensena, w prosty sposób wnioskujemy, że  $(|X_n|^p)_{n \in [N]}$  jest podmartyngałem. Stosując **Fakt 2.45** do nieujemnego podmartyngału  $(|X_n|^p)_{n \in [N]}$  otrzymujemy nasz następny wniosek.

**Wniosek 2.46**

Niech  $X = (X_n)_{n \in [N]}$  będzie martyngałem. Wówczas dla każdego  $p \geq 1$  i każdej  $\lambda > 0$

$$\lambda^p \mathbb{P} [X^* \geq \lambda] \leq \mathbb{E} [|X_N|^p],$$

gdzie  $X^* = \sup_{n \in [N]} |X_n|$ .

Zobaczmy teraz jak szacowanie słabego typu z poprzedniego wniosku implikują oszacowania mocnego typu.

**Fakt 2.47**

Niech  $X = (X_n)_{n \in [N]}$  będzie martyngałem. Wówczas dla każdego  $p > 1$ ,

$$\mathbb{E} \left[ \sup_{n \in [N]} |X_n|^p \right] \leq \left( \frac{p}{p-1} \right)^p \mathbb{E} [|X_N|^p].$$

*Dowód.* Przypomnijmy oznaczenie  $X^* = \sup_{n \in [N]} |X_n|$ . Stosując **Fakt 2.45** otrzymujemy dla dowolnego ustalonego  $K > 0$ ,

$$\begin{aligned} \mathbb{E} [(X^* \wedge K)^p] &= \mathbb{E} \left[ \int_0^{X^* \wedge K} p\lambda^{p-1} d\lambda \right] = \mathbb{E} \left[ \int_0^K p\lambda^{p-1} \mathbb{1}_{\{X^* \geq \lambda\}} d\lambda \right] \\ &= \int_0^K p\lambda^{p-1} \mathbb{P}[X^* \geq \lambda] d\lambda \leq \int_0^K p\lambda^{p-2} \mathbb{E} \left[ |X_N| \mathbb{1}_{\{X^* \geq \lambda\}} \right] d\lambda \\ &= p \mathbb{E} \left[ |X_N| \int_0^{X^* \wedge K} \lambda^{p-2} d\lambda \right] = \frac{p}{p-1} \mathbb{E} \left[ |X_N| (X^* \wedge K)^{p-1} \right] \\ &\leq \frac{p}{p-1} \mathbb{E} [|X_N|^p]^{1/p} \mathbb{E} [(X^* \wedge K)^p]^{(p-1)/p}, \end{aligned}$$

gdzie w ostatnim kroku zastosowaliśmy nierówność Cauchy'ego-Höldera. Skoro  $\mathbb{E} [(X^* \wedge K)^p] < \infty$  możemy uprościć i otrzymać

$$\mathbb{E} [(X^* \wedge K)^p]^{1/p} \leq \frac{p}{p-1} \mathbb{E} [|X_N|^p]^{1/p}.$$

Przechodząc z  $K \rightarrow \infty$  i stosując twierdzenie o zbieżności monotonicznej otrzymujemy tezę.  $\square$

Przechodząc z  $N \rightarrow \infty$  w nierówności z **Faktu 2.47** i korzystając z twierdzenia o zbieżności monotonicznej otrzymujemy nierówność maksymalną z nieskończonym horyzontem czasowym.

**Wniosek 2.48**

Dla martyngału  $\{X_n\}_{n \in \mathbb{N}}$  i  $p > 1$  zachodzi

$$\mathbb{E}[(X^*)^p] \leq \left( \frac{p}{p-1} \right)^p \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p].$$

**Twierdzenie 2.49**

Załóżmy, że  $\{X_n\}$  jest martyngałem. Ustalmy  $p > 1$ . Następujące warunki są równoważne

- 1)  $\sup_n \mathbb{E}|X_n|^p < \infty$ ;
- 2) Istnieje zmienna  $X_\infty$  taka, że  $X_n \rightarrow X_\infty$  p.w. oraz

$$\mathbb{E}[|X_n X_\infty|^p] \rightarrow 0.$$

3) istnieje  $X_\infty \in L^p$  takie, że  $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ .

Dowód Twierdzenia 2.49. 1)  $\implies$  2) Skoro

$$\sup_{n \in \mathbb{N}} \mathbb{E}[X_n^+]^p \leq \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p] < \infty$$

to z twierdzenia o zbieżności istnieje  $X_\infty$  takie, że  $X_n \rightarrow X_\infty$  w  $L^p$ . Pokażemy teraz, że zbieżność zachodzi w  $L^p$ .

Z **Wniosku 2.48** zmienna losowa  $(2X^*)^p$  jest całkowalna i majoryzuje  $\{|X_n - X_\infty|^p\}_{n \in \mathbb{N}}$ . Z twierdzenia o zbieżności ograniczonej

$$\mathbb{E}[|X_n - X_\infty|^p] \rightarrow 0.$$

2)  $\implies$  3) Zauważmy, że dla dowolnych  $m, n \in \mathbb{N}$ ,

$$\mathbb{E}[X_{n+m} | \mathcal{F}_n] = X_n.$$

Zatem dla dowolnego  $A \in \mathcal{F}_n$ ,

$$\mathbb{E}[X_{n+m} \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A]$$

Zauważmy, że  $\mathbb{E}[X_{n+m} \mathbb{1}_A] \rightarrow \mathbb{E}[X_\infty \mathbb{1}_A]$ , bo

$$|\mathbb{E}[X_{n+m} \mathbb{1}_A] - \mathbb{E}[X_\infty \mathbb{1}_A]|^p \leq \mathbb{E}[|X_{n+m} - X_\infty|^p] \rightarrow 0.$$

oznacza to, że

$$\mathbb{E}[X_\infty \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A]$$

Innymi słowy

$$\mathbb{E}[X_\infty | \mathcal{F}_n] = X_n$$

3)  $\implies$  1) Zauważmy, że dla dowolnego  $n \in \mathbb{N}$ ,

$$|X_n|^p = |\mathbb{E}[X_\infty | \mathcal{F}_n]|^p \leq \mathbb{E}[|X_\infty|^p | \mathcal{F}_n]$$

i stąd

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p] \leq \mathbb{E}[|X_\infty|^p] < \infty.$$

□

### Przykład 2.50

Niech  $f: [0, 1] \rightarrow \mathbb{R}$  będzie dowolną borelowską funkcją całkowalną z  $L^p$ , czyli

$$\int_0^1 |f(t)|^p dt < \infty.$$

Dla  $n \in \mathbb{N}$  niech  $I_k^{(n)} = [k2^{-n}, (k+1)2^{-n})$  dla  $0 \leq k < 2^n$ . Dla  $n \in \mathbb{N}$  zdefiniujemy funkcję  $f_n$  wzorem

$$f_n(x) = 2^n \int_{I_k^{(n)}} f(y) dy, \quad x \in I_k^{(n)}.$$

Pokażemy, że  $f_n \rightarrow f$  w  $L^p[0,1]$ , czyli że

$$\int_{[0,1]} |f(x) - f_n(x)|^p dx \rightarrow 0.$$

Rozważmy w tym celu przestrzeń probabilistyczną  $\Omega = [0,1]$ ,  $\mathcal{F} = \text{Bor}([0,1])$  i miarą probabilistyczną  $\mathbb{P}$  będą miarą Lebesguea. Niech  $\mathcal{F}_n = \sigma(I_k^{(n)}, k < 2^n)$  będzie rozważaną przez nas filtracją. Sprawdzamy, że

$$\mathbb{E}[f | \mathcal{F}_n] = \sum_{k < 2^n} \mathbb{1}_{I_k^{(n)}} \mathbb{E}[f | I_k^{(n)}] = \sum_{k < 2^n} \mathbb{1}_{I_k^{(n)}} 2^n \int_{I_k^{(n)}} f(x) dx = f_n(x).$$

Zatem  $f_n$  jest martyngałem. Stosując do niego twierdzenie o zbieżności martyngałów wnioskujemy, że  $f_n \rightarrow f$  w  $L^p$ , czyli

$$\int_0^t |f_n(x) - f(x)|^p dx \rightarrow 0.$$

## łańcuchy Markowa

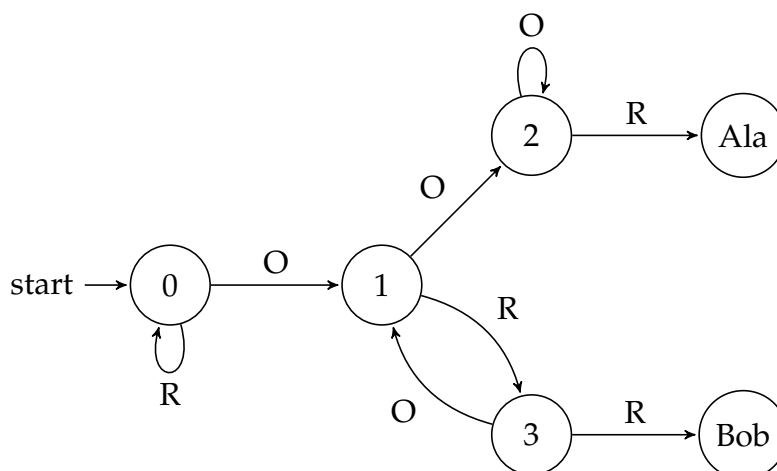
**Streszczenie** Przedstawimy klasę procesów stochastycznych charakteryzującą się krótką pamięcią. Zobaczymy, że mimo restrykcyjnego założenia na strukturę zależności, procesy te z powodzeniem modelują wiele naturalnych sytuacji. Po zaprezentowaniu podstaw ogólnej teorii przedyskutujemy zagadnienie powracalności oraz ergodyczności.

Podczas kolejnych wykładów będziemy badać łańcuchy Markowa. Zanim podamy formalną definicję przytoczymy dwa przykłady, które pozwolą zrozumieć potrzebą badania takich obiektów jak i naszkicują kontekst w którym będziemy o niech dyskutować.

### 3.1 Przykłady

#### Przykład 3.1

Alicja i Bob rzucają symetryczną monetą tak długo, aż wypadnie *OOR* lub *ORR*. Alicja wygrywa, gdy wzorzec *OOR* wypadnie jako pierwszy, natomiast Bob, gdy wypadnie *ORR*. Jakie jest prawdopodobieństwo, że grę wygra Alicja? Powyższe zadanie można opisać przy pomocy spaceru losowego na grafie. Każdy wierzchołek grafu zawiera informacje o stopniu dopasowania końcowych wyników rzutu monetą do jednego z żądanych wzorców. Dla przykładu stan 3 oznacza, że ostatnie dwa wyniki to O i R.



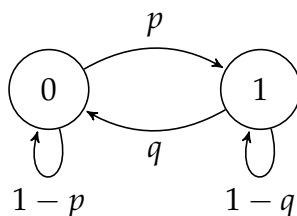
Oznaczmy przez  $p_i$  prawdopodobieństwo, że będąc w wierzchołku  $i$  dojdziemy do wierzchołka Ala. Wówczas, korzystając ze wzoru na prawdopodobieństwo całkowite otrzymujemy równania

$$\begin{aligned}
 p_A &= 1, \quad p_B = 0, \\
 p_2 &= \frac{1}{2}p_2 + \frac{1}{2}p_A, \\
 p_3 &= \frac{1}{2}p_1 + \frac{1}{2}p_B, \\
 p_0 &= p_1 = \frac{1}{2}p_2 + \frac{1}{2}p_3.
 \end{aligned}$$

Stąd otrzymujemy  $p_0 = 2/3$ .

### Przykład 3.2

Rozważmy spacer losowy na zbiorze  $\mathcal{S} = \{0, 1\}$  zdefiniowany następująco. Cząsteczka początkowo znajduje się w punkcie 0. Rzucamy monetą, niekoniecznie symetryczną. Jeżeli wypadnie orzeł, z prawdopodobieństwem  $p$ , to cząsteczka przemieszcza się do 1, w przeciwnym razie pozostaje w 0. W punkcie 1 używamy innej monety, w której orzeł wypada z prawdopodobieństwem  $q$ . Jeżeli wypadnie orzeł, to cząsteczka idzie do 0.



Oznaczmy przez  $X_0 = 0, X_1, \dots$  kolejne pozycje cząsteczki.  $X_0 = 0$ , ale kolejne pozycje są już losowe:

$$\mathbb{P}[X_1 = 0] = 1 - p, \quad \mathbb{P}[X_1 = 1] = p.$$

W drugim kroku:

$$\begin{aligned} \mathbb{P}[X_2 = 0] &= (1 - p)^2 + pq, \\ \mathbb{P}[X_2 = 1] &= (1 - p)p + p(1 - q). \end{aligned}$$

Oznaczmy przez  $\mu_n$  rozkład zmiennej losowej  $X_n$ , tzn.

$$\mu_n = (\mathbb{P}[X_n = 0], \mathbb{P}[X_n = 1]).$$

Ogólnie

$$\begin{aligned} \mu_n(0) &= \mathbb{P}[X_n = 0] = \mathbb{P}[X_n = 0, X_{n-1} = 0] + \mathbb{P}[X_n = 0, X_{n-1} = 1] \\ &= \mathbb{P}[X_n = 0 | X_{n-1} = 0] \mathbb{P}[X_{n-1} = 0] \\ &\quad + \mathbb{P}[X_n = 0 | X_{n-1} = 1] \mathbb{P}[X_{n-1} = 1] \\ &= (1 - p)\mu_{n-1}(0) + q\mu_{n-1}(1), \\ \mu_n(1) &= \mathbb{P}[X_n = 1] = p\mu_{n-1}(0) + (1 - q)\mu_{n-1}(1). \end{aligned}$$

Wygodnie jest używać macierzowej notacji. Reguły rządzące spacerem zapiszmy w postaci macierzy:

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}.$$

Zauważmy

$$\mu_0 = (1, 0), \quad \mu_1 = \mu_0 P, \quad \dots \quad \mu_n = \mu_{n-1} P = \mu_0 P^n$$

(powyżej wektory piszemy jako wiersze).

Chcemy zrozumieć jak wygląda zachowanie  $X_n$  (czy też  $\mu_n$ ) dla dużych wartości  $n$ . W szczególności chcemy zrozumieć, czy rozkłady stabilizują się po pewnym czasie lub dążą do jakiejś granicy. Dzięki powyższym rachunkom te pytania formalnie sprowadzają się do zrozumienia jak wygląda macierz  $P^n$ . W tym prostym przykładzie można bezpośrednio obliczyć tę wartość (np. diagonalizując macierz). Ogólnie, gdy graf ma trudniejszą strukturę, wykorzystywane są również inne metody.

Pokażemy, jak znaleźć odpowiedzi na powyższe pytania przy pomocy łatwych rachunków. Załóżmy najpierw, że istnieje granica  $\mu_n$ , gdy  $n \rightarrow \infty$  i oznaczmy ją przez  $\pi$ . Wektor  $\pi$  musi być miarą probabilistyczną oraz spełniać

$$\pi = \lim \mu_n = \lim \mu_{n-1} P = \pi P,$$

czyli

$$\begin{cases} \pi(0) + \pi(1) = 1 \\ (1-p)\pi(0) + q\pi(1) = \pi(0) \\ p\pi(0) + (1-q)\pi(1) = \pi(1) \end{cases}$$

"Latwo obliczyć, że wówczas  $\pi$  powinno być postaci

$$\pi(0) = \frac{q}{p+q}, \quad \pi(1) = \frac{p}{p+q}.$$

Aby pokazać, że granica  $\lim_n \mu_n$  rzeczywiście istnieje, zdefiniujmy

$$\varepsilon_n = \mu_n(0) - \pi(0).$$

Wówczas

$$\begin{aligned} \varepsilon_{n+1} &= \mu_n(0)(1-p) + (1-\mu_n(0))q - \pi(0) \\ &= \mu_n(0)(1-p-q) + \pi(0)(p+q) - \pi(0) \\ &= (1-p-q)\varepsilon_n. \end{aligned}$$

Zatem, jeżeli  $0 < p+q < 2$ , to

$$\lim_{t \rightarrow \infty} \mu_n(0) = \pi_n(0) \quad \text{oraz analogicznie} \quad \lim_{t \rightarrow \infty} \mu_n(1) = \pi_n(1).$$

Podsumowując, pokazaliśmy że  $\mu_n$  zbiega do pewnej miary  $\pi$ . Co więcej zbieżność ta jest wykładniczo szybka, gdyż  $\varepsilon_n = \rho^n C$  dla  $|\rho| = |1-p-q| < 1$ . A więc dla dużych wartości  $n$  możemy przybliżyć wartości  $\mu_n$  przez  $\pi$ .

Zauważmy również, że jeżeli  $p+q=1$  (np.  $p=q=1/2$ ), to niezależnie od  $\mu_0$ ,  $\mu_1 = \pi$ . Ponadto  $1-p-q$  jest wartością własną macierzy  $P$ .

Naszym celem będzie pokazanie, że ogólne łańcuchy Markowa na skończonych przestrzeniach stanów (czy też spacery losowe na skończonych grafach), przy dosyć naturalnych założeniach, posiadają bardzo podobne własności, a więc zbiegają do pewnego stanu równowagi (ekwilibrium) i zbieżność ta jest wykładniczo szybko.

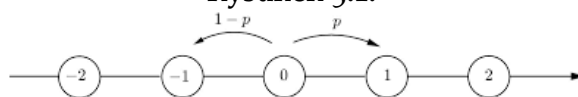
W przypadku, gdy przestrzeń stanów jest nieskończona rozważa się inne pytania, np. czy każdy punkt zostanie odwiedzony nieskończenie wiele razy.

### Przykład 3.3

Rozważmy spacer losowy  $\{X_n\}$  na zbiorze liczb całkowitych, który w każdym kroku przechodzi z prawdopodobieństwem  $p$  w prawo i z prawdopodobieństwem  $1-p$  w lewo. Jego podstawowe własności zostały opisane w zadaniach. Jeżeli spacer losowy jest symetryczny ( $p=1/2$ ), to

- spacer losowy odwiedzi każdy punkt nieskończenie wiele razy;
- $\mathbb{P}[X_n = 0] \sim C/\sqrt{n}$ .

Rysunek 3.1.



- proces oscyluje pomiędzy  $-\infty$  i  $+\infty$ , tzn.

$$-\infty = \liminf X_n \leq \limsup X_n = \infty.$$

Jeżeli spacer losowy jest niesymetryczny (np.  $p > 1/2$ ), to

- spacer losowy odwiedzi każdy punkt jedynie skończenie wiele razy (a niektóre z punktów nie zostaną odwiedzone ani razu);
- $X_n \rightarrow \infty$  p.w.

## 3.2 Definicje

Podczas reszty wykładu skupimy się na łańcuchach Markowa o co najwyżej przeliczalnej przestrzeni stanów  $S$ .

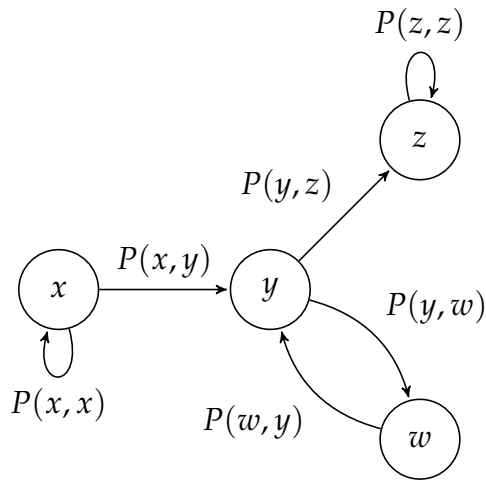
### Definicja 3.4

Niech  $S$  będzie co najwyżej przeliczalną przestrzenią. Funkcję  $P: S \times S \rightarrow [0, 1]$  nazywamy macierzą przejścia (lub macierzą stochastyczną), jeżeli

$$\sum_{y \in S} P(x, y) = 1$$

dla każdego  $x \in S$ .

O funkcji przejścia możemy myśleć na kilka sposobów. Pierwszy z nich polega na utożsamieniu  $P(x, y)$  z wagami na grafie skierowanym.



Inną reprezentacją dla  $P$  jest postać macierzowa. Powiedzmy, że mamy pewne uporządkowanie elementów przestrzeni stanów  $S$  w ciąg. Innymi słowy  $S = \{s_0, s_1, s_2, \dots\}$  dla skończonego lub nieskończonego ciągu  $\{s_n\}_n$ . Wówczas możemy zapisać  $P$  jako

$$P = \begin{pmatrix} P(s_0, s_0) & P(s_0, s_1) & P(s_0, s_2) & \dots \\ P(s_1, s_0) & P(s_1, s_1) & P(s_1, s_2) & \dots \\ P(s_2, s_0) & P(s_2, s_1) & P(s_2, s_2) & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

Należy wspomnieć, że warunek obecny w Definicji 3.4 oznacza, że wiersze w powyższej macierzy sumują się do jedności. Rozważać będziemy  $S$  wyposażone w  $-\sigma$ -ciało wszystkich swoich podzbiorów. Wówczas każda funkcja  $X: \Omega \rightarrow S$  jest zmienną losową (jest mierzalna).

### Definicja 3.5

Niech  $S$  będzie zbiorem co najwyżej przeliczalnym. Proces stochastyczny  $\{X_n\}_{n \in \mathbb{N}}$  jest łańcuchem Markowa z funkcją przejścia  $\mathbf{P}$  wtedy i tylko wtedy dla każdych  $x_0, x_1, \dots, x_{n+1} \in S$  takich, że  $\mathbb{P}[X_n = x_n, \dots, X_0 = x_0] > 0$  zachodzi

$$\begin{aligned} \mathbb{P}[X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n] &= \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n] \\ &= P(x_n, x_{n+1}). \end{aligned}$$

Łańcuch Markowa  $\{X_n\}$  jest to proces bez pamięci, tzn. taki, że  $X_n$  zależy wyłącznie od  $X_{n-1}$ . Dokładniej, jeżeli znamy wartość  $X_{n-1}$ , to  $X_n$  nie zależy od wcześniejszej trajektorii procesu, tj. od stanów  $X_0, X_1, \dots, X_{n-2}$ . Możemy więc myśleć, że jeżeli znamy teraźniejszość (stan  $X_{n-1}$ ), to przyszłość (wartość  $X_n$  i kolejne) nie zależy od przeszłości (zobacz np. pkt. 5

twierdzenie 3.9 poniżej). Kolejne kroki procesu są zdeterminowane przez prawdopodobieństwa określone macierzą  $P$ .

### Przykład 3.6

Spacer losowy  $\{X_n\}$  na liczbach całkowitych opisany w przykładzie 3.3 jest łańcuchem Markowa. Istotnie może on zostać przedstawiony w postaci  $X_n = Y_1 + \dots + Y_n$ , gdzie  $\{Y_n\}$  jest ciągiem niezależnych i jednakowo rozłożonych zmiennych losowych takich, że  $\mathbb{P}[Y_n = 1] = p = 1 - \mathbb{P}[Y_n = -1]$ . Ustalmy dowolny ciąg możliwych trajektorii tego ciągu:  $a_0 = 0, a_1, a_2, \dots$  i zdefiniujmy  $\varepsilon_n = a_n - a_{n-1}$  jako wartość  $n$ -tego kroku. Wówczas  $\varepsilon_n \in \{-1, 1\}$  i mamy

$$\begin{aligned} \mathbb{P}[X_n = a_n | X_0 = a_0, X_1 = a_1, \dots, X_{n-1} = a_{n-1}] \\ &= \mathbb{P}[X_n = X_{n-1} + \varepsilon_n | X_0 = a_0, X_1 = a_1, \dots, X_{n-1} = a_{n-1}] \\ &= \mathbb{P}[X_n = a_n | X_{n-1} = a_{n-1}]. \end{aligned}$$

Powyższy rachunek pokazuje również, że macierz  $P$  (nieskończenie wymiarowa) przyjmuje wartości

$$P(i, i+1) = p, \quad P(i, i-1) = 1-p \quad P(i, j) = 0 \text{ dla } |i-j| \neq 1.$$

Następne twierdzenie wynika ze standardowego zastosowania twierdzenia Kołmogorowa o istnieniu procesu. Można je też udowodnić metodami ad hoc.

### Twierdzenie 3.7

Niech  $P: S \times S \rightarrow [0, 1]$  będzie funkcją przejścia i niech  $\mu: S \rightarrow [0, 1]$  będzie dowolnym rozkładem prawdopodobieństwa. Istnieje przestrzeń probabilistyczna  $(\Omega, \mathcal{F}, \mathbb{P})$  i określone na niej zmienna losowe  $\{X_n\}_{n \in \mathbb{N}}$  o wartościach w  $S$  tworzące łańcuch Markowa z rozkładem początkowym  $\mu$  i funkcją przejścia  $P$ .

Główną konsekwencją powyższego rezultatu jest to, że od tej pory nie będziemy zaprzęтали sobie głowy jawną konstrukcją procesu Markowa. Ograniczać się będziemy jedynie do opisu macierzy przejścia.

### Przykład 3.8

Ważną klasą łańcuchów Markowa są spacer losowe na grafach. Załóżmy, że mamy dany skończony lub nieskończony zbiór wierzchołków (np. pewien zbiór punktów na płaszczyźnie), z których niektóre połączone są krawędziami. Spacer losowy oznacza, że po grafie porusza się pewna cząsteczka (wędrowiec), która będąc w jednym z punktów wybiera losowo jednego z jego sąsiadów (według z góry określonego prawdopodobieństwa i niezależnie od poprzednich kroków) i przechodzi do niego. Następnie

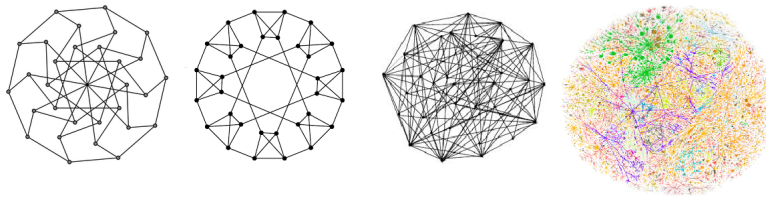
powtarza tę czynność wielokrotnie. Niech  $G = (V, E)$  będzie lokalnie skończonym grafem prostym. Dokładniej niech  $V$  będzie dowolnym, co najwyżej przeliczalnym zbiorem wierzchołków i niech  $E$  będzie zbiorem krawędzi nieskierowanych, czyli dowolna pod kolekcją

$$E \subseteq \{\{x, y\} : x, y \in V, x \neq y\}$$

taką, że dla każdego  $x \in V$ ,

$$\deg(x) = \#\{y \in V : \{x, y\} \in E\} < \infty.$$

Ten prosty proces kryje w sobie wiele pytań:



**Rysunek 3.2.** Graficzne reprezentacje grafów

- Jakie jest prawdopodobieństwo, że po  $n$  krokach cząsteczka jest w danym wierzchołku?
- Czy każdy z wierzchołków zostanie odwiedzony?
- Po jakim czasie odwiedzimy wszystkie wierzchołki (jeżeli graf jest skończony)?
- Jak często będziemy przechodzić przez dany wierzchołek?
- Po jakim czasie wrócimy do danego wierzchołka?
- Jak długo aktualna pozycja zależy od punktu startu?
- Co można powiedzieć o procesie, po upływie długiego czasu?

Grafy mogą mieć dosyć prostą strukturę, być skończone lub nieskończone, mogą też być dosyć skomplikowane (np. graf stworzony przez linki pomiędzy stronami www, sieci społeczne). Powyższe problemy, chociaż czysto teoretyczne znajdują całą gamę praktycznych zastosowań:

- Jak długo należy tasować talię  $n$  kart, aby móc uznać ją za potasowaną? Można ten problem sprowadzić do spaceru losowego na dużym grafie składającym się z  $n!$  elementów, czyli wszystkich możliwych permutacji.
- W informatyce istnieje szereg trudnych obliczeniowo problemów. Jednym z kluczowych wyzwań jest problem czy  $P = NP$ , a więc pytanie czy dla dużej klasy ważnych problemów (np. problem komiwojażera, problem faktoryzacji liczb) istnieją szybkie algorytmy działające w czasie wielomianowym. Przy pomocy spacerów losowych na grafach konstruuje się algorytmy, które w szybkim czasie pozwalają na znalezienie

przybliżonych rozwiązań. Grafy, które się pojawiają podczas analizy, są tak duże (ale ciągle skończone), że nie można wypisać ich wierzchołków.

- Wyszukiwarka Google pozycjonuje strony nadając im pewne wagi. W tym celu używany jest algorytm Pagerank. W dużym uproszczeniu, tworzy się wielki graf, którego wierzchołkami są wszystkie strony internetowe, a krawędzie oznaczają, że jedna strona linkuje do drugiej. Następnie rozważany jest spacer losowy na tym grafie i dla każdej ze stron mierzy się jak często jest ona odwiedzana przez ten spacer. Im częściej, tym większy otrzymuje ona ranking.

Dla macierzy przejścia  $P$  definiujemy jej kolejne potęgi  $P^m$  dla  $m \in \mathbb{N}$  wzorem  $P^1 = P$  oraz dla  $m \in \mathbb{N}$  macierz  $P^{m+1}$  jest zadana przez iloczyn macierzy  $P^m$  i  $P$ , czyli

$$P^{m+1}(x, y) = \sum_{z \in S} P^m(x, z)P(z, y)$$

dla  $x, y \in S$ . Zauważmy, że jeżeli interpretujemy  $P$  jako macierz, to  $P^m$  jest po prostu  $m$ -tą potęgą macierzy  $P$ .

### Twierdzenie 3.9

Niech  $\{X_n\}$  będzie łańcuchem Markowa o macierzy przejścia  $P$ . Wówczas

1. dla dowolnego  $n \in \mathbb{N}$  i dowolnych  $x_0, x_1, \dots, x_n \in S$  zachodzi

$$\begin{aligned} \mathbb{P}[X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0] \\ = \mathbb{P}[X_0 = x_0]P(x_0, x_1) \dots P(x_{n-1}, x_n); \end{aligned}$$

2. Dla każdego  $n, m \in \mathbb{N}$  i dowolnych  $x, y \in S$ ,

$$\mathbb{P}[X_{n+m} = y | X_m = x] = \mathbb{P}[X_n = y | X_0 = x] = P^n(x, y);$$

3. Spełnione są Równania Chapmana-Kołmogorowa: dla każdego naturalnych  $n$  oraz  $m$  oraz  $z, x \in S$  zachodzi

$$\mathbb{P}[X_{n+m} = z | X_0 = x] = \sum_{y \in S} P^n(x, y)P^m(y, z);$$

4. (Własność Markowa) Warunkując na zbiorze  $\{X_m = x\}$ , proces  $\{X_{m+n}\}_{n \geq 0}$  jest łańcuchem Markowa z macierzą przejścia  $P$  i rozkładem początkowym  $\delta_x$  niezależnym od  $\sigma\{X_0, X_1, \dots, X_m\}$ .
5. Jeżeli  $A \in \sigma(X_n, X_{n+1}, \dots)$ , to  $\mathbb{P}[A | X_0, X_1, \dots, X_n] = \mathbb{P}[A | X_n]$

*Dowód. Zadanie.*

□

**Definicja 3.10**

Stan  $y$  nazywamy **osiągalnym** ze stanu  $x$ , jeżeli dla pewnego  $n$  zachodzi  $P^n(x, y) > 0$ . Innymi słowy  $y$  jest osiągalny z  $x$  jeżeli z dodatnim prawdopodobieństwem możemy przejść z  $x$  do  $y$ . Piszemy wówczas  $x \rightarrow y$ . Stany  $x$  i  $y$  nazywamy **wzajemnie komunikującymi się**, jeżeli  $x \rightarrow y$  oraz  $y \rightarrow x$  (a więc można przejść zarówno z  $x$  do  $y$  jak i z  $y$  do  $x$ ). Piszemy wówczas  $x \leftrightarrow y$ .

**Uwaga 3.11**

Relacja osiągalności jest przechodnia, tzn. jeżeli  $x \rightarrow y$  oraz  $y \rightarrow z$ , to  $x \rightarrow z$ . Wynika to natychmiast z równania Chapmana-Kołmogorowa. Weźmy  $n$  i  $m$  takie, że  $P^n(x, y) > 0$  oraz  $P^m(y, z) > 0$ , wówczas

$$\begin{aligned} P^{n+m}(x, z) &= \\ \mathbb{P}[X_{n+m} = z | X_0 = x] &= \sum_{w \in \mathcal{S}} \mathbb{P}[X_n = w | X_0 = x] \mathbb{P}[X_m = z | X_0 = w] \\ &\geq \mathbb{P}[X_n = y | X_0 = x] \mathbb{P}[X_m = z | X_0 = y] = P^n(x, y) P^m(y, z) > 0. \end{aligned}$$

Oczywiście również relacja wzajemnej przechodności jest przechodnia, tzn. jeżeli  $x \leftrightarrow y$  oraz  $y \leftrightarrow z$ , to  $x \leftrightarrow z$ .

**Definicja 3.12**

"Łańcuch Markowa nazywamy **nierozkładalnym** jeżeli każde jego dwa stany komunikują się nawzajem (a więc z dowolnego stanu można osiągnąć każdy inny).

**Przykład 3.13**

Niech teraz  $(G, *)$  będzie grupą z elementem neutralnym  $e_G$ . Rozważmy ciąg  $\{\theta_n\}_{n \in \mathbb{N}}$  niezależnych zmiennych o rozkładzie  $\mu$ , tj.

$$\mu(g) = \mathbb{P}[\theta_0 = g], \quad g \in G.$$

Zdefiniujemy teraz spacer losowy na grupie  $G$ . Niech  $X_0 = e_G$  i niech  $X_{n+1} = \theta_n * X_n$  dla  $n \in \mathbb{N}$ . Wówczas  $\{X_n\}_{n \in \mathbb{N}}$  jest łańcuchem Markowa z funkcją przejścia

$$P(g, h) = \mu(h * g^{-1}).$$

**Przykład 3.14**

Załóżmy, że  $G = \mathbb{Z}_m$  dla pewnego  $m \in \mathbb{N}$  i niech

$$\mu(1) = p, \quad \mu(-1) = 1 - p$$

dla pewnego  $p \in [0, 1]$ . Wówczas łańcuch Markowa skonstruowany w poprzednim przykładzie można zobaczyć jako spacer losowy na cyklu długości  $m$  w którym w każdym kroku poruszamy się o jeden wierzchołek zgodnie z ruchem wskazówek zegara (z prawdopodobieństwem  $p$ ) i przeciwnie z ruchem wskazówek zegara (z prawdopodobieństwem  $1 - p$ ).

### Przykład 3.15

Niech  $(G, *)$  będzie dowolną skończoną grupą. I niech  $\mu$  będzie dowolnym rozkładem skoków spaceru losowego  $\{X_n\}_{n \in \mathbb{N}}$  na grupie  $G$ . Kiedy taki proces jest nierozkładalny? Rozważmy nośnik  $\mu$  dany wzorem

$$H = \{h \in G : \mu(h) > 0\}$$

Innymi słowy  $H$  jest zbiorem wszystkich dozwolonych skoków. Przypomnijmy z podstawowego kursu algebry, że przez  $\langle H \rangle$  oznaczamy najmniejszą podgrupę  $G$  zawierającą  $H$ . Wówczas spacer jest nierozkładalny wtedy i tylko wtedy, gdy  $\langle H \rangle = G$ . Sprawdźmy, że istotnie tak jest. Załóżmy najpierw, że  $\langle H \rangle = G$ . Zauważmy, że wystarczy pokazać, że każdy stan  $g \in G$  komunikuje się z elementem neutralnym  $e_G$  grupy  $G$ . Wobec naszego założenia istnieje  $n \in \mathbb{N}$  oraz ciągi  $h_1, \dots, h_n \in H$  oraz  $\epsilon_1, \epsilon_2, \dots, \epsilon_n \in \{-1, 1\}$  takie, że

$$g * w^{-1} = h_n^{\epsilon_n} * \dots * h_2^{\epsilon_2} * h_1^{\epsilon_1}.$$

Zauważmy, że skoro  $G$  jest skończona, to rzędy elementów  $h_1, h_2, \dots, h_n$  również są skończone, tj  $e_G = h_j^{n_j}$  dla pewnego dodatniego  $n_j$ . Oznacza to, że naszą reprezentację  $g$  możemy przedstawić jako

$$g = h_n^{\epsilon_n + n_n} * \dots * h_2^{\epsilon_2 + n_2} * h_1^{\epsilon_1 + n_1} * w,$$

gdzie  $\epsilon_j + n_j$  jest dodatnie dla każdego  $j$ . Oznacza to, że z  $w$  do  $g$  możemy dojść wykonując  $n_1 + \epsilon_1$  kroków wielkości  $h_1$ , następnie  $n_2 + \epsilon_2$  kroków wielkości  $h_2$  itd. Formalnie

$$\mathbb{P}[X_n = g | X_0 = w] \geq \mu(h_n)^{n_n + \epsilon_n} \cdot \mu(h_1)^{n_1 + \epsilon_1} > 0.$$

Skoro  $g$  i  $w$  były dowolne, pokazuje to, że łańcuch jest nierozkładalny. Jeżeli łańcuch  $\{X_n\}_{n \in \mathbb{N}}$  jest nieredukowalny, to dla każdego  $g \in G$  istnieje ciąg dozwolonych kroków  $h_1, h_2, \dots, h_n \in H$  takich, że  $g = h_n * h_{n-1} * \dots * h_1$ . Oznacza to, że  $g \in \langle H \rangle$ .

W teorii łańcuchów Markowa rozróżnia się kolejne własności strukturalne zbioru stanów. Pominiemy jednak te rozważania i od tej pory będziemy zakładać, że badane łańcuchy Markowa są nieredukowalne, co pozwoli nam się skupić na najważniejszych własnościach procesu.

### 3.3 łańcuchy powracające i chwilowe

Powiedzmy, że interesuje nas pozycja cząsteczki, która porusza się zgodnie z łańcuchem Markowa na pewnej nieskończonej przestrzeni stanów. Szukanie cząsteczki na całej przestrzeni może okazać się czasochłonne. Zauważmy, że jeżeli cząsteczka porusza się bardzo długo, to nie ma potrzeby przeszukiwania stanów, które łańcuch Markowa odwiedza tylko skończenie wiele razy. W najbliższej części wykładu zobaczymy jak scharakteryzować takie stany. Siłą rzeczy opiszemy też stany, które są odwiedzane nieskończenie wiele razy.

#### Definicja 3.16

Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie łańcuchem Markowa. Mówimy, że stan  $x \in S$  jest **powracający**, jeżeli

$$\mathbb{P}[X_n = x \text{ dla nieskończenie wielu } n | X_0 = x] = 1.$$

Mówimy, że stan  $x \in S$  jest **chwilowy**, jeżeli

$$\mathbb{P}[X_n = x \text{ dla nieskończenie wielu } n | X_0 = x] = 0.$$

Stan  $x$  jest powracający, jeżeli proces odwiedzi go nieskończenie wiele razy z prawdopodobieństwem 1, natomiast jest chwilowy, jeżeli zostanie odwiedzony skończenie wiele razy. Pokażemy, że każdy stan jest albo powracający albo chwilowy i przedstawimy odpowiednie kryterium. Co więcej pokażemy, że wszystkie stany muszą być tego samego typu.

#### Przykład 3.17

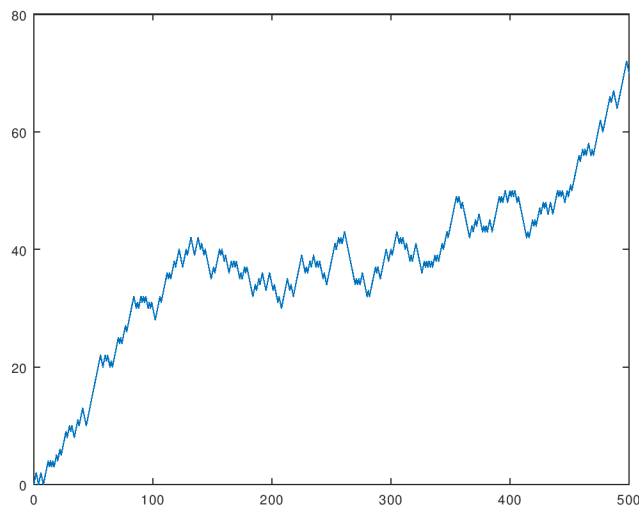
Rozważmy niesymetryczny spacer losowy  $\mathbb{P}[X_{n+1} = j + 1 | X_n = j] = p$ ,  $\mathbb{P}[X_{n+1} = j - 1 | X_n = j] = 1 - p$  dla  $p > 1/2$ . Równoważnie  $X_n = X_0 + \sum_{k=1}^n \xi_k$  gdzie  $\{\xi_k\}_{k \in \mathbb{N}}$  są iid z rozkładem  $\mathbb{P}[\xi_k = 1] = p$  i  $\mathbb{P}[\xi_k = -1] = 1 - p$ . Zauważmy, że z mocnego prawa wielkich liczb z prawdopodobieństwem jeden,

$$\frac{X_n}{n} \rightarrow \mathbb{E}[\xi_1] = 2p - 1 > 0.$$

W szczególności  $X_n \rightarrow \infty$  (patrz rysunek 3.3) oznacza, to że dla  $\mathbb{P}$  prawie wszystkich  $\omega$ , ciąg  $X_n(\omega) \rightarrow \infty$  i w szczególności ciąg  $\{X_n(\omega)\}_{n \in \mathbb{N}}$  zawiera tylko skończenie wiele zer. Oznacza to, że 0 jest stanem chwilowym dla  $\{X_n\}_{n \in \mathbb{N}}$ .

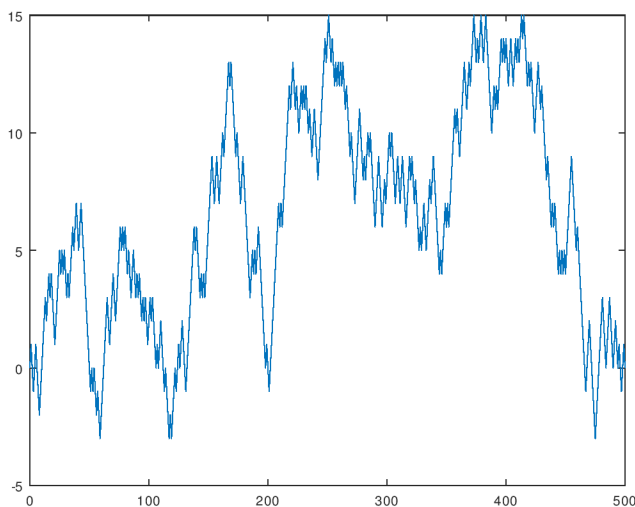
#### Przykład 3.18

Przypomnijmy, że symetryczny prosty spacer na  $\mathbb{Z}$  z prawdopodobieństwem 1 wróci do punktu startu. Oznacza to, że odwiedzi on punkt 0



**Rysunek 3.3.** Przykładowa realizacja niesymetrycznego spaceru losowego

nieskończenie wiele razy (Rysunek 3.4). Formalne uzasadnienie tego faktu wymaga dodatkowych narzędzi.



**Rysunek 3.4.** Przykładowa realizacja symetrycznego spaceru losowego

Przypomnijmy, że  $T : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$  jest czasem zatrzymania jeżeli dla każdego  $n \in \mathbb{N}$  zdarzenie  $\{T = n\}$  jest mierzalne względem  $\sigma$ -ciała  $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ .

**Twierdzenie 3.19** (Silna własność Markowa)

Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie łańcuchem Markowa z macierzą przejścia  $P$  i rozkładem początkowym  $\mu_0$ . Niech  $T$  będzie czasem zatrzymania. Wówczas pod warunkiem  $\{X_T = x, T < \infty\}$ , proces  $\{X_{T+n}\}_{n \in \mathbb{N}}$  jest łańcuchem Markowa z macierzą przejścia  $P$  i rozkładem początkowym  $\delta_x$  niezależnym od  $\sigma\{X_0, X_1, \dots, X_T\}$ .

**Uwaga 3.20**

$\sigma$ -ciało  $\sigma\{X_0, X_1, \dots, X_T\}$  generowane jest przez zbiory postaci  $\{X_0 = s_0, X_1 = s_1, \dots, X_m = s_m, T = m\}$ .

**Uwaga 3.21**

Warto porównać powyższą mocną własność Markowa z własnością Markowa sformułowaną w twierdzeniu 3.9 pkt 4.

*Dowód Twierdzenia 3.19.* W dowodzie dwukrotnie skorzystamy z następującej formuły prawdziwej dla dowolnych zdarzeń  $A, B, C$  takich, że  $C$  jest sumą parami rozłącznych zbiorów  $C_m$ :

$$\mathbb{P}[A|B \cap C] = \sum_m \mathbb{P}[A|B \cap C_m] \cdot \frac{\mathbb{P}[B \cap C_m]}{\mathbb{P}[B \cap C]} \quad (3.3.1)$$

Niech  $B$  będzie dowolnym zdarzeniem zależnym od  $X_0, X_1, \dots, X_T$ . Wówczas

$$\begin{aligned} & \mathbb{P}\left[\{X_{T+1} = x_1, \dots, X_{T+n} = x_n\} \cap B \mid X_T = x, T < \infty\right] \\ & \stackrel{(3.3.1)}{=} \sum_{m \geq 0} \mathbb{P}\left[\{X_{T+1} = x_1, \dots, X_{T+n} = x_n\} \cap B \mid X_m = x, T = m\right] \\ & \quad \cdot \frac{\mathbb{P}[T = m, X_m = x]}{\mathbb{P}[X_T = x, T < \infty]} \\ & = \sum_{m \geq 0} \mathbb{P}\left[\{X_{m+1} = x_1, \dots, X_{m+n} = x_n\} \cap B \mid X_m = x, T = m\right] \\ & \quad \cdot \frac{\mathbb{P}[T = m, X_m = x]}{\mathbb{P}[X_T = x, T < \infty]} \\ & \stackrel{\text{wl. Markowa}}{=} \sum_{m \geq 0} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n \mid X_0 = x] \mathbb{P}[B \mid \{X_m = x, T = m\}] \\ & \quad \cdot \frac{\mathbb{P}[T = m, X_m = x]}{\mathbb{P}[X_T = x, T < \infty]} \\ & \stackrel{(3.3.1)}{=} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n \mid X_0 = x] \cdot \mathbb{P}[B \mid X_T = x, T < \infty]. \end{aligned}$$

□

Zobaczmy teraz jak wykorzystać powyższe twierdzenia do badania rekurencyjności. Zaczniemy od kilku definicji. Zdefiniujemy pierwszy moment trafienia w stan  $x$ :

$$T_x = \inf\{n \geq 1 : X_n = x\}.$$

Następnie oznaczmy przez

$$V_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=x\}}$$

liczbę wizyt procesu w punkcie  $x$ . Zauważmy, że

$$\begin{aligned} \mathbb{E}[V_x | X_0 = x] &= \mathbb{E} \left[ \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=x\}} \middle| X_0 = x \right] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[\mathbf{1}_{\{X_n=x\}} | X_0 = x] = \sum_{n=0}^{\infty} \mathbb{P}[X_n = x | X_0 = x] = \sum_{n=0}^{\infty} P^n(x, x). \end{aligned}$$

Zdefiniujmy również prawdopodobieństwo powrotu do  $x$

$$f_x = \mathbb{P}[T_x < \infty | X_0 = x]$$

**Lemat 3.22**

Dla  $k \geq 0$

$$\mathbb{P}[V_x > k | X_0 = x] = f_x^k$$

*Dowód.* Zdefiniujmy czas  $k$ -tego powrotu do  $x$ :

$$T_x^0 = 0, \quad T_x^1 = T_x, \quad T_x^k = \inf\{n > T_x^{k-1} : X_n = x\}$$

oraz długość  $k$ -tego cyklu

$$S_x^k = \begin{cases} T_x^k - T_x^{k-1} & \text{jeżeli } T_x^k < \infty \\ \infty & \text{w przeciwnym razie} \end{cases}$$

Przeprowadzamy dowód przez indukcję. Dla  $k = 0$  lemat jest oczywisty.

Dla  $k + 1$ :

$$\begin{aligned} \mathbb{P}[V_x > k + 1 | X_0 = x] &= \mathbb{P}[T_x^{k+1} < \infty | X_0 = x] \\ &= \mathbb{P}[T_x^k < \infty \text{ oraz } S_x^{k+1} < \infty | X_0 = x] \\ &= \mathbb{P}[S_x^{k+1} < \infty | T_x^k < \infty, X_0 = x] \mathbb{P}[T_x^k < \infty | X_0 = x] \\ &\stackrel{\text{zał. ind. + mocna wł. M.}}{=} f_x f_x^k = f_x^{k+1} \end{aligned}$$

Wyjaśnijmy dokładniej ostatnią równość. Wynika ona z mocnej własności Markowa. Dla większej przejrzystości odwołajmy się dokładnie do własności, która była sformułowana powyżej w terminach skończenie wielu kroków procesu. Oznaczmy  $T = T_x^k$ , jest to oczywiście czas zatrzymania spełniający  $X_T = x$  na zbiorze  $T < \infty$ . Zatem warunkując na zbiorze  $\{T < \infty\}$ ,

proces  $\{X_{T+n}\}_{n \geq 0}$  jest łańcuchem Markowa z macierzą przejścia  $P$  oraz rozkładem początkowym  $\delta_x$ . Ponadto

$$S := S_x^{k+1} = \inf\{n \geq 1 : X_{T+n} = x\}.$$

Korzystając więc z mocnej własności Markowa otrzymujemy

$$\begin{aligned} \mathbb{P}[S_x^{k+1} < \infty | T_x^k < \infty, X_0 = x] &= \sum_{i=1}^{\infty} \mathbb{P}[S_x^{k+1} = i | T_x^k < \infty, X_0 = x] \\ &= \sum_{i=1}^{\infty} \mathbb{P}[S = i | X_0 = x] = \mathbb{P}[S < \infty | X_0 = x] \\ &= \mathbb{P}[T < \infty | X_0 = x] = f_x. \end{aligned}$$

□

### Twierdzenie 3.23

Zachodzą następujące warunki:

1. jeżeli  $f_x = \mathbb{P}[T_x < \infty | X_0 = x] = 1$ , to stan  $x$  jest powracający oraz

$$\sum_{n \geq 0} P^n(x, x) = \infty;$$

2. jeżeli  $f_x = \mathbb{P}[T_x < \infty | X_0 = x] < 1$ , to stan  $x$  jest chwilowy oraz

$$\sum_{n \geq 0} P^n(x, x) < \infty;$$

W szczególności, każdy stan jest albo powracający, albo chwilowy.

*Dowód.* Jeżeli  $f_x = 1$ , to lematu o ciągłości miary (TP1) oraz z lematu 3.22

$$\mathbb{P}[V_x = \infty | X_0 = x] = \lim_{k \rightarrow \infty} \mathbb{P}[V_x > k | X_0 = x] = 1,$$

a zatem stan  $x$  jest powracający oraz

$$\sum_{n=0}^{\infty} P^n(x, x) = \mathbb{E}[V_x | X_0 = x] = \infty.$$

Z drugiej strony, jeżeli  $f_x < 1$ , wówczas z lematu 3.22

$$\begin{aligned} \sum_{n=0}^{\infty} P^n(x, x) &= \mathbb{E}[V_x | X_0 = x] = \sum_{n=0}^{\infty} \mathbb{E}[V_x > n | X_0 = x] \\ &= \sum_{n=0}^{\infty} f_x^n = \frac{1}{1 - f_x} < \infty. \end{aligned}$$

Zatem  $\mathbb{P}[V_x = \infty | X_0 = x] = 0$  oraz stan  $x$  jest chwilowy.

□

**Wniosek 3.3.2.** Jeżeli  $\{X_n\}_{n \in \mathbb{N}}$  jest nieredukowalnym łańcuchem Markowa, albo wszystkie stany są powracające, albo jest wszystkie stany są chwilowe.

**Definicja 3.24**

Niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie nieredukowalnym łańcuchem Markowa. Powiemy, że  $X$  jest tranzytywny, jeżeli wszystkie jego stany są tranzytywne. Powiemy, że  $X$  jest rekurencyjny, jeżeli wszystkie jego stany są rekurencyjne.

Zauważmy, że powyższe twierdzenie dostarcza nam kryterium, które pozwala efektywnie sprawdzić, czy łańcuch Markowa jest powracający, czy też chwilowy.

**Przykład 3.25**

Niech  $\{S_n\}_{n \in \mathbb{N}}$  będzie prostym (symetrycznym) spacerem losowym na  $\mathbb{Z}$ . Chcemy obliczyć prawdopodobieństwo powrotu do 0 po  $n$  krokach. Zauważmy, że po nieparzystej liczbie kroków nie możemy wrócić do 0:

$$\mathbb{P}[S_{2n+1} = 0] = 0.$$

Dla parzystej liczby kroków mamy

$$\mathbb{P}[S_{2n} = 0] = \binom{2n}{n} \frac{1}{2^{2n}},$$

gdyż z  $2n$  kroków  $n$  powinno być w prawo, a  $n$  w lewo. Przypominając wzór Stirlinga:  $n! \sim \sqrt{2\pi n} (n/e)^n$ , możemy napisać

$$\mathbb{P}[S_{2n} = 0] \sim \frac{\sqrt{2\pi \cdot 2n}}{\sqrt{2\pi n} \sqrt{2\pi n}} \cdot \left(\frac{2n}{e}\right)^{2n} \cdot \left(\frac{e}{n}\right)^{2n} \cdot \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}}.$$

Z powyższego twierdzenia punkt 0 jest więc powracający, a zatem z wniosku, wszystkie pozostałe stany są powracające. Proces ten odwiedzi więc każdy punkt nieskończenie wiele razy.

**Przykład 3.26**

Prosty spacer losowy na  $\mathbb{Z}^d$ . Definiujemy go podobnie jak wcześniej. W każdym kroku cząsteczka wybiera z tym samym prawdopodobieństwem jednego z sąsiadów:

$$X_0 = (0, \dots, 0),$$

$$\mathbb{P}[X_{n+1} = x \pm e_j | X_n = x] = \frac{1}{2d}.$$

Można wówczas pokazać (dla  $d = 2$  jest to zadanie), że

$$\mathbb{P}[X_{2n} = 0] \sim \frac{C_d}{n^{d/2}}$$

co pociąga, że dla  $d \leq 2$  spacer jest rekurencyjny i dla  $d \geq 3$  spacer jest chwilowy.

Przypomnijmy, że dla  $y \in S$  przez  $V_y$  oznaczamy liczbę wizyt w stanie  $y$  dane wzorem

$$V_y = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=y\}}.$$

### Twierdzenie 3.27

Niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie nieredukowalnym łańcuchem Markowa.

a) Jeżeli  $X$  jest rekurencyjny, to dla dowolnych  $x, y \in S$ ,

$$\mathbb{E}[V_y | X_0 = x] = \infty.$$

b) Jeżeli  $X$  jest chwilowy, to dla dowolnych  $x, y \in S$ ,

$$\mathbb{E}[V_y | X_0 = x] < \infty.$$

*Dowód.* Niech  $X$  będzie rekurencyjny. Dla dowolnych  $x, y \in S$  istnieje  $m \in \mathbb{N}$  takie, że  $P^{(m)}(x, y) > 0$ . Z własności Markowa

$$\mathbb{E}[V_y | X_0 = x] \geq P^{(m)}(x, y) \mathbb{E}[V_y | X_0 = y],$$

gdzie ostatnie wartość oczekiwana jest nieskończona z rekurencyjności  $y$ . Załóżmy teraz, że  $X$  jest chwilowy. Wybierzmy  $n \in \mathbb{N}$  takie, że  $P^{(n)}(y, x) > 0$ . Skoro  $x$  jest chwilowy, to

$$\infty > \mathbb{E}[V_y | X_0 = y] \geq P^{(n)}(y, x) \mathbb{E}[V_y | X_0 = x].$$

□

### Definicja 3.28

Niech  $\{X_n\}_{n \in \mathbb{N}}$  będzie łańcuchem Markowa na przestrzeni stanów  $S$  z funkcją przejścia  $P$ . Powiemy, że  $\alpha: S \rightarrow \mathbb{R}_+$  jest miarą niezmenną dla  $P$ , jeżeli  $\alpha$  jest niezerowa oraz  $\alpha = \alpha P$ , czyli dla każdego  $x \in S$ ,

$$\alpha(x) = \sum_{y \in S} \alpha(y) P(y, x).$$

### Przykład 3.29

Niech  $P$  będzie funkcją przejścia prostego, symetrycznego spaceru losowego na  $\mathbb{Z}$ , tj.  $P(k, k \pm 1) = 1/2$ . Wówczas  $\alpha$  jest miarą niezmienniczą wtedy i tylko wtedy, gdy dla dowolnego  $j \in \mathbb{Z}$  spełnione jest

$$\begin{aligned}\alpha(j) &= \sum_{k \in \mathbb{Z}} \alpha(k)P(k, j) = \alpha(j-1)P(j-1, j) + \alpha(j+1)P(j+1, j) \\ &= \frac{\alpha(j+1) + \alpha(j-1)}{2}.\end{aligned}$$

Innymi słowy  $\{\alpha(j)\}_{j \in \mathbb{Z}}$  musi tworzyć ciąg arytmetyczny. Skoro wymagamy  $\alpha(j) \geq 0$  dla każdego  $j \in \mathbb{Z}$ , to jedynym wyborem jest ciąg stały. Zatem przykładowo  $\alpha(j) = 1$  jest miarą niezmienniczą dla prostego, symetrycznego spaceru losowego.

### Twierdzenie 3.30

Niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie nieredukowalnym łańcuchem Markowa. Dla ustalonego  $x_0 \in S$  niech  $T_{x_0}^+$  będzie dane przez

$$T_{x_0}^+ = \inf\{n \geq 1 \mid X_n = x_0\}.$$

Wówczas

$$\alpha(y) = \mathbb{E} \left[ \sum_{n=1}^{T_{x_0}^+} \mathbb{1}_{\{X_n=y\}} \mid X_0 = x_0 \right]$$

jest miarą niezmienniczą.

*Dowód.* Zauważmy, że  $\alpha(x_0) = 1$ , więc  $\alpha$  jest niezerowa. Mamy

$$\alpha(y) = \sum_{n=1}^{\infty} \mathbb{P}[X_n = y, T_{x_0}^+ \geq n \mid X_0 = x_0]$$

Dla  $n = 1$  wyraz szeregu to

$$\mathbb{P}[X_1 = y, T_{x_0}^+ \geq 1 \mid X_0 = x_0] = \mathbb{P}[X_1 = y \mid X_0 = x_0] = P(x_0, y).$$

Dla  $n \geq 2$  mamy

$$\begin{aligned}\mathbb{P}[X_n = y, T_{x_0}^+ \geq n \mid X_0 = x_0] &= \\ &= \sum_{z \in S \setminus \{x_0\}} \mathbb{P}[X_n = y, X_{n-1} = z, T_{x_0}^+ \geq n \mid X_0 = x_0].\end{aligned}$$

Dla każdego  $z \in S \setminus \{x_0\}$  mamy

$$\begin{aligned} \mathbb{P}[X_n = y, X_{n-1} = z, T_{x_0}^+ \geq n | X_0 = x_0] &= \\ \mathbb{P}[X_n = y, X_{n-1} = z, T_{x_0}^+ \geq n - 1 | X_0 = x_0] &= \\ \mathbb{P}[X_n = y | X_{n-1} = z] \mathbb{P}[X_{n-1} = z, T_{x_0}^+ \geq n - 1 | X_0 = x_0]. \end{aligned}$$

Podsumowując

$$\alpha(y) = P(x_0, y) + \sum_{z \in S \setminus \{x_0\}} \alpha(z) P(z, y) = \sum_{z \in S} \alpha(z) P(z, y).$$

Pozostaje pokazać, że  $\alpha(y) < \infty$  dla każdego  $y \in S$ . Zauważmy, że skoro  $\alpha = \alpha P$ , to dla każdego  $n \in \mathbb{N}$ ,

$$\alpha(y) = \sum_{x \in S} \alpha(x) P^{(n)}(x, y).$$

W szczególności

$$1 = \alpha(x_0) = \sum_{x \in S} \alpha(x) P^{(n)}(x, x_0).$$

Załóżmy nie wprost, że istnieje  $x$  którego  $\alpha(x) = \infty$ . Wybierając  $n \in \mathbb{N}$  dla którego  $P^{(n)}(x, x_0) > 0$  otrzymujemy sprzeczność w postaci

$$1 = \alpha(x_0) = \sum_{x \in S} \alpha(x) P^{(n)}(x, x_0) \geq P^{(n)}(x, x_0) \alpha(x) = \infty.$$

□

### Twierdzenie 3.31

Niech  $X$  będzie nieredukowalnym, rekurencyjnym łańcuchem Markowa. Jeżeli  $\alpha, \beta$  są miarami niezmienniczymi, to istnieje stała  $c > 0$  taka, że

$$\alpha(x) = c\beta(x)$$

dla wszystkich  $x \in S$ .

*Dowód.* Coming soon...

□

### Definicja 3.32

Powiemy, że stan  $x \in S$  łańcucha Markowa  $X = \{X_n\}_{n \in \mathbb{N}}$  jest dodatnio rekurencyjny, jeżeli dla

$$T_x^+ = \inf\{n \geq 1 : X_n = x\}$$

spełnia

$$\mathbb{E} [T_x^+ | X_0 = x] < \infty.$$

**Twierdzenie 3.33**

Niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie nieredukowalnym, rekurencyjnym łańcuchem Markowa.

a) Jeżeli istnieje miara niezmiennicza  $\alpha$  dla której

$$\sum_{x \in S} \alpha(x) < \infty,$$

to wszystkie stany są dodatnio rekurencyjne.

b) Jeżeli istnieje stan, który jest dodatnio rekurencyjny, to dla każdej miary niezmienniczej  $\alpha$ ,

$$\sum_{x \in S} \alpha(x) < \infty.$$

Powyższe twierdzenie mówi, że jeżeli łańcuch jest nieredukowalny i rekurencyjny, to jeżeli jeden stan jest dodatnio rekurencyjny, to wszystkie stany są dodatnio rekurencyjne. W takiej sytuacji będziemy mówili, że łańcuch jest dodatnio rekurencyjny.

**Definicja 3.34**

Powiemy, że miara niezmiennicza  $\pi$  jest rozkładem stacjonarnym, jeżeli

$$\sum_{x \in S} \pi(x) = 1.$$

Zauważmy, że jeżeli rozkład  $X_0$  jest rozkładem stacjonarnym, to  $X_1$  również ma rozkład stacjonarny. Rzeczywiście, dla  $x \in S$  mamy

$$\mathbb{P}[X_1 = x] = \sum_{y \in S} \mathbb{P}[X_0 = y] \mathbb{P}[X_1 = x | X_0 = y] = \sum_{y \in S} \pi(y) P(y, x) = \pi(x).$$

Podobnie pokazujemy, że jeżeli  $X_0$  ma rozkład  $\pi$ , to dla każdego  $n \in \mathbb{N}$  ma rozkład  $\pi$ .

**Twierdzenie 3.35**

Niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie nieredukowalnym, dodatnio rekurencyjnym łańcuchem Markowa. Wówczas  $X$  posiada jedyny rozkład stacjonarny zadany wzorem

$$\pi(x) = \frac{1}{\mathbb{E}[T_x^+ | X_0 = x]},$$

gdzie  $T_x^+$  jest czasem powrotu do  $x \in S$ ,

$$T_x^+ = \inf\{n \geq 1 : X_n = x\}.$$

*Dowód.* Załóżmy, że  $\pi$  jest miarą stacjonarną. Ustalmy  $x \in S$ . Wiemy, że

$$\alpha(y) = \mathbb{E} \left[ \sum_{n=1}^{T_x^+} \mathbb{1}_{\{X_n=y\}} \mid X_0 = x \right]$$

zadaje miarę stacjonarną. Wobec jedyności istnieje  $c > 0$  takie, że  $\pi = c\alpha$ , czyli

$$\pi(y) = c\alpha(y)$$

dla każdego  $y$ . Zauważmy, że

$$1 = \sum_{y \in S} \pi(y) = c \sum_{y \in S} \alpha(y) = c \mathbb{E} [T_x^+ | X_0 = x]$$

wobec czego

$$c = \frac{1}{\mathbb{E} [T_x^+ | X_0 = x]}.$$

Z drugiej strony z definicji  $\alpha$  wynika, że  $\alpha(x) = 1$  i stąd

$$\pi(x) = c\alpha(x) = c = \frac{1}{\mathbb{E} [T_x^+ | X_0 = x]}.$$

Z dowodu wynika, że  $c\alpha$  jest miarą stacjonarną, co w szczególności pokazuje, że  $\pi$  zawsze istnieje.  $\square$

### Przykład 3.36

Niech  $X = \{X_n\}_{n \in \mathbb{N}}$  będzie prostym spacerem losowym na grafie  $G = (V, E)$ . Pokażemy, że  $\alpha(x) = \deg(x)$  jest miarą stacjonarną. Przypomnijmy, że funkcja przejścia zadaje się jako  $P(x, y) = 1/\deg(x)$  dla  $x$  i  $y$  będącymi sąsiadami (piszemy wtedy  $x \sim y$ ) i  $P(x, y) = 0$  poza tym. Dla  $x \in S = V$

$$\begin{aligned} \sum_{y \in V} \alpha(y)P(y, x) &= \sum_{y \sim x} \alpha(y)P(y, x) = \sum_{y \sim x} \deg(y) \frac{1}{\deg(y)} \\ &= \sum_{y \sim x} 1 = \deg(x) = \alpha(x). \end{aligned}$$

Zauważmy, że jeżeli  $G$  jest grafem skończonym, to powyższa miara normalizuje się do rozkładu stacjonarnego

$$\pi(x) = \frac{\deg(x)}{2\#E}.$$

W szczególności otrzymujemy jawny wzór na czas powrotu do wierzchołka (stanu)  $x \in S = V$ ,

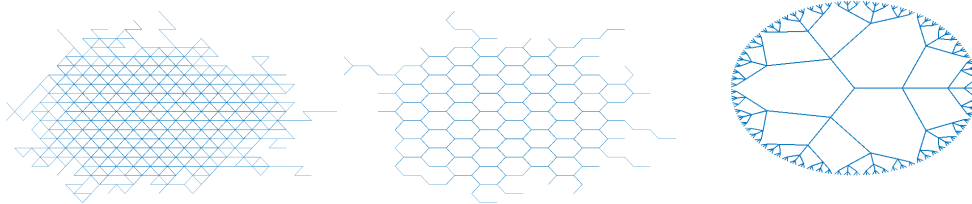
$$\mathbb{E} [T + x^+ | X_0 = x] = \frac{2\#E}{\deg(x)}.$$

### 3.3.1 Rekurencyjne spacery losowe na grafach

Niech  $G = (V, E)$  będzie nieskończonym, spójnym grafem o ograniczonym stopniu, tj.

$$\deg(x) < \infty,$$

gdzie dla wierzchołka  $x \in V$ ,  $\deg(x)$  oznacza liczbę jego sąsiadów. Przykładami takich grafów są kraty  $\mathbb{Z}$ ,  $\mathbb{Z}^2$ ,  $\mathbb{Z}^3 \dots$

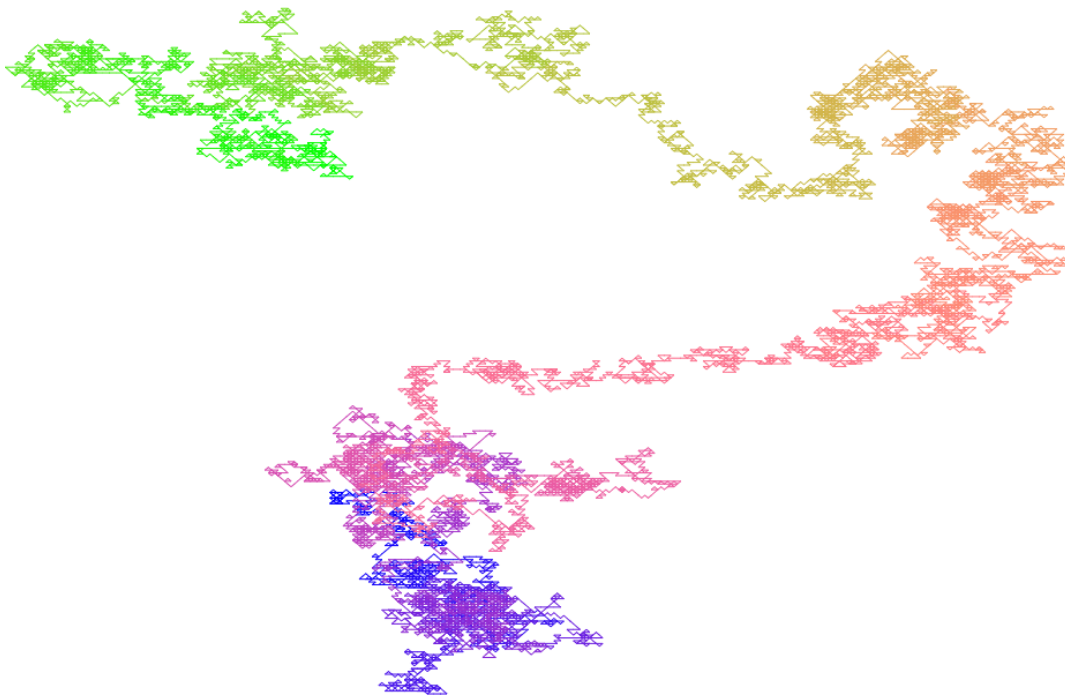


Rysunek 3.5. Przykłady nieskończonych grafów.

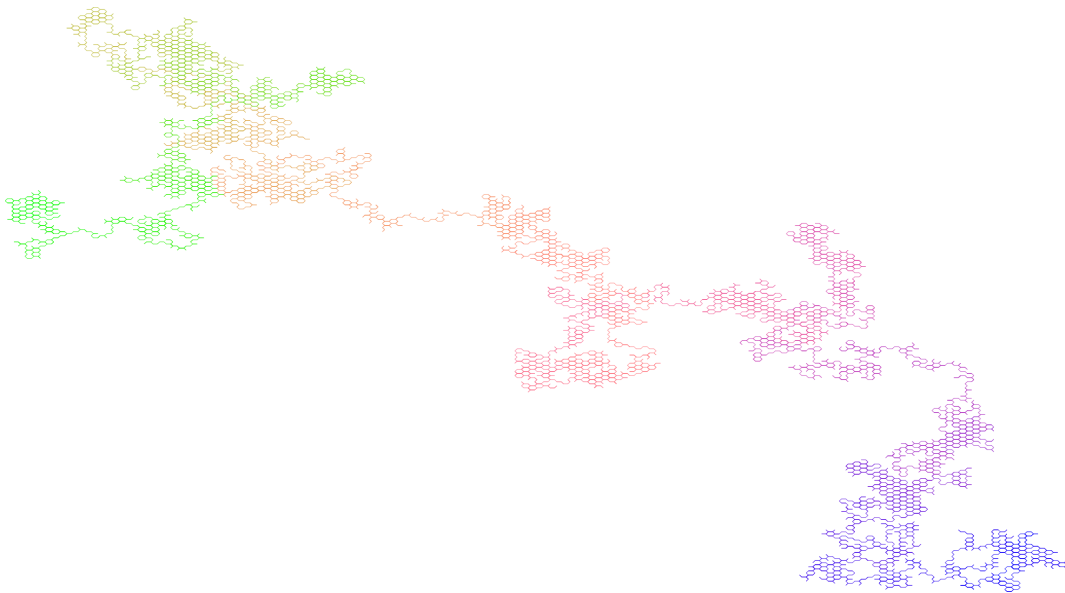
Chcemy sformułować warunki na ogólny graf  $G$  dla których  $\{X_n\}_{n \in \mathbb{N}}$  będący prostym spacerem losowym na  $G$  (w każdym kroku wybieramy losowo jednego z sąsiadów) jest rekurencyjny. Zakładamy, że  $G$  jest grafem spójnym. Wówczas bowiem łańcuch  $\{X_n\}_{n \in \mathbb{N}}$  jest nieredukowalny. Zauważmy, że im mniej krawędzi w grafie, tym łatwiej jest spacerowi wrócić do pozycji początkowej. Rzeczywiście, symulacje pokazują, że jeżeli graf ma więcej krawędzi, to spacer odwiedza więcej stanów (jest on w pewnym sensie szybszy).

#### Definicja 3.37

Kolekcja krawędzi  $C \subset E$  jest nazywana rozcięciem (ang. cutset), jeżeli graf  $(V, E \setminus C)$  nie jest spójny. Powiemy, że ciąg  $\{C_n\}_{n \in \mathbb{N}}$  jest wstępującym ciągiem rozcięć wokół  $x \in V$ , jeżeli dla każdego  $n \in \mathbb{N}$ ,  $C_{n+1}$  jest rozcięciem  $G$  takim, że  $x$  oraz  $C_n$  znajdują się w skończonej składowej  $(V, E \setminus C_{n+1})$ .



Rysunek 3.6. Symulacja 2000 kroków spaceru losowego na kracie trójkątnej



Rysunek 3.7. Symulacja 2000 kroków spaceru losowego na kracie sześciokątnej

**Przykład 3.38**

Jeżeli  $G = \mathbb{Z}$ , to  $C_n = \{(-n-1, -n), (n, n+1)\}$  jest wstępującą rodziną rozcięć 0.

**Przykład 3.39**

Jeżeli  $G = \mathbb{Z}^2$ , to  $C_n$  składający się z krawędzi wychodzących z kwadratu o  $2n \times 2n$  o środku w punkcie  $(0,0)$  jest wstępującą rodziną rozcięć 0.

**Twierdzenie 3.40** (Nash-Williams 1959)

Założmy, że  $\{C_n\}_{n \in \mathbb{N}}$  jest wstępującym ciągiem rozcięć (wokół  $y \in V$ ).

Jeżeli

$$\sum_{n=0}^{\infty} \frac{1}{\#C_n} = \infty,$$

to prosty spacer losowy na  $G$  jest rekurencyjny.

*Dowód.* W każdej krawędzi  $G$  umieszczamy dodatkowy wierzchołek. W ten sposób otrzymujemy większy graf  $G'$ . Zauważmy, że wystarczy uzasadnić powracalność prostego spaceru losowego na  $G'$ . Istotnie, parzyste kroki spaceru na  $G'$  odpowiadają wersji spaceru na  $G$ , w której w każdym kroku możemy pozostać w każdym wierzchołku. Rekurencyjność ostatniego spaceru jest równoważna rekurencyjności prostego spaceru losowego na  $G$ .

Niech  $y \in V$  będzie wierzchołkiem wokół którego  $\{C_n\}_{n \in \mathbb{N}}$  tworzą rozcięcia. Niech  $T_y^+$  będzie czasem powrotu do  $y$  danym przez

$$T_y^+ = \inf\{n \geq 1 \mid X_n = y\}.$$

Jeżeli  $\mathbb{P}[T_y^+ < \infty \mid X_0 = y] = 1$ , to  $y$  jest stanem rekurencyjnym i nie ma czego dowodzić. Założmy więc, że  $\mathbb{P}[T_y^+ < \infty \mid X_0 = y] < 1$ . Oznaczmy przez  $d(x)$  stopień wierzchołka  $x$  w grafie  $G'$ . Wówczas dla funkcji przejścia  $P$  prostego spaceru losowego na  $G'$  mamy

$$d(x)P(x, y) = 1 = d(y)P(y, x)$$

dla dowolnych wierzchołków  $x$  i  $y$  w grafie  $G'$ . Podobnie pokazujemy, że dla dowolnej ścieżki  $\gamma = (x_1, x_2, \dots, x_N)$  prawdopodobieństwo jej przemierzenia

$$P(\gamma) = P(x_1, x_2)P(x_2, x_3) \cdots P(x_{N-1}, x_N)$$

spełnia

$$d(x_1)P(\gamma) = d(x_N)P(\gamma^{-1}),$$

gdzie  $\gamma^{-1} = (x_N, x_{N-1}, \dots, x_2, x_1)$  jest odwróconą ścieżką  $\gamma$ . Niech  $C'_n$  będzie zbiorem wierzchołków dodanych do krawędzi z  $C_n$ . Niech  $\mu_n(x)$ , dla

$x \in C'_n$ , będzie prawdopodobieństwem, że dla prostego spaceru losowego na  $G'$  startującego z  $y$ , zanim wrócimy do  $y$  odwiedzimy  $C'_n$  oraz pierwszym odwiedzionym punktem z  $C'_n$  jest  $x$ . Dokładniej, niech

$$\tau_n = \inf\{k \in \mathbb{N} : X_k \in C'_n\}$$

oraz

$$\mu_n(x) = \mathbb{P}[X_{\tau_n} = x, \tau_n < T_y^+ | X_0 = y].$$

Wówczas

$$\mu_n(x) = \sum_{\gamma \in C(y,x)} P(\gamma),$$

gdzie  $C(y,x)$  jest kolekcją wszystkich skończonych ścieżek  $\gamma = (x_1, \dots, x_N)$  takich, że  $x_1 = y$ ,  $x_N = x$  oraz  $x_2, \dots, x_{N-1} \notin \{y, x\}$ . Rozważmy zdarzenie

$$A_n = \{\text{Spacer losowy odwiedza } C'_n \text{ raz po czym wraca do } y\}.$$

Wówczas rozważając punkt trafienia w  $C'_n$ , i dobierając  $\epsilon > 0$  tak mały aby  $\epsilon d(x) \leq 1$  dla wszystkich  $x$  w  $G'$ ,

$$\begin{aligned} \mathbb{P}[A_n] &= \sum_{x \in C'_n} \sum_{\gamma_1, \gamma_2 \in C(x,y)} P(\gamma_1)P(\gamma_2^{-1}) = \sum_{x \in C'_n} \mu_n(x) \sum_{\gamma_2 \in C(x,y)} P(\gamma_2^{-1}) \\ &\geq \sum_{x \in C'_n} \mu_n(x) \sum_{\gamma_2 \in C(x,y)} \epsilon d(y)P(\gamma_2^{-1}) = \sum_{x \in C'_n} \mu_n(x) \sum_{\gamma_2 \in C(x,y)} \epsilon d(x)P(\gamma_2) \\ &\geq \sum_{x \in C'_n} \epsilon \mu_n(x)^2. \end{aligned}$$

Korzystając teraz z nierówności Jensena

$$\sum_{x \in C'_n} \mu_n(x)^2 \geq \frac{1}{\#C'_n} \left( \sum_{x \in C'_n} \mu_n(x) \right)^2 = \frac{\mathbb{P}[\tau_n < T_y^+]^2}{\#C'_n} \geq \frac{\mathbb{P}[T_y^+ = \infty]^2 | X_0 = y}{\#C'_n},$$

gdzie  $\mathbb{P}[T_y^+ = \infty | X_0 = y] > 0$  wobec naszego założenia. Wobec naszego założenia oraz uzasadnionych nierówności

$$\sum_{n \in \mathbb{N}} \mathbb{P}[A_n] = \infty.$$

Zauważmy, że zdarzenia  $\{A_n\}_{n \in \mathbb{N}}$  są niezależne. Istotnie, dla  $n_1 < n_2 < \dots < n_k$ ,  $k \in \mathbb{N}$ , zdarzenie  $A_{n_1} \cap \dots \cap A_{n_k}$  mówi, że pomiędzy odwiedzinami  $C'_{n_1}$ ,  $C'_{n_2}$ ,  $\dots$ ,  $C'_{n_k}$  spacer za każdym razem wraca do  $y$ . Z mocnej własności Markowa zachodzi zatem

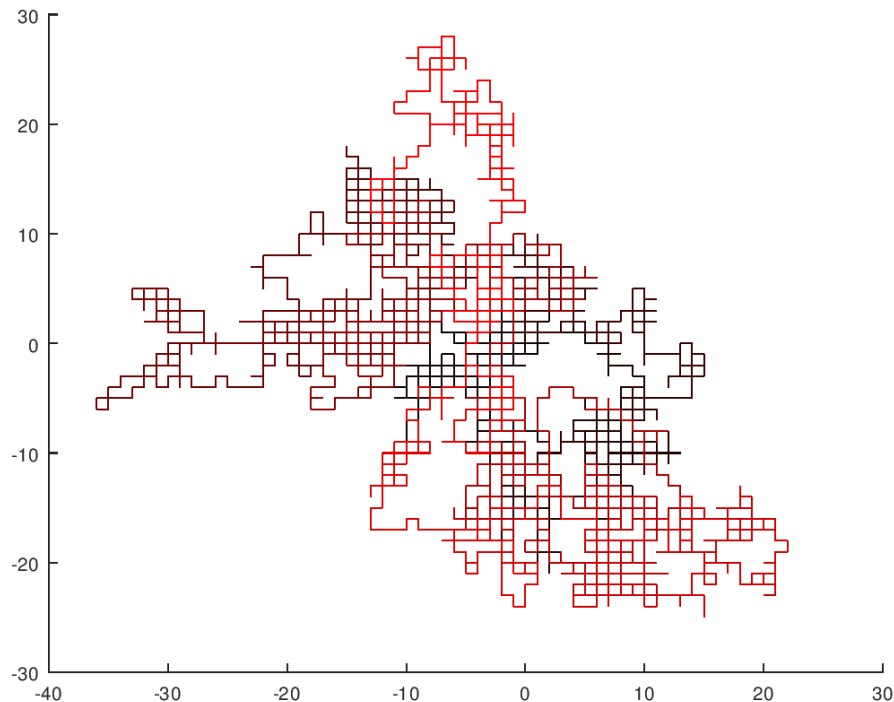
$$\mathbb{P}[A_{n_1} \cap \dots \cap A_{n_k}] = \mathbb{P}[A_{n_1}] \cdots \mathbb{P}[A_{n_k}].$$

Z lematu Borela-Cantelliego z prawdopodobieństwem jeden  $\{A_n\}_{n \in \mathbb{N}}$  zachodzą nieskończenie wiele razy. Oznacza to, że punkt  $y$  zostanie odwiedzony nieskończenie wiele razy.  $\square$

Zastosowanie powyższego twierdzenia pozwala udowodnić część poniższego rezultatu.

**Twierdzenie 3.41 (Pólya)**

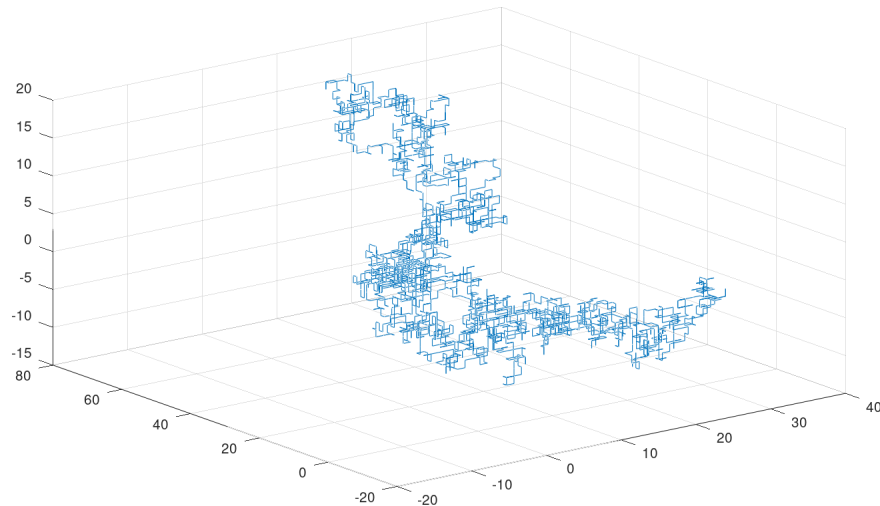
Jeżeli  $d = 1, 2$ , to prosty spacer losowy na  $\mathbb{Z}^d$  jest powracający, natomiast dla  $d \geq 3$ , to prosty spacer losowy na  $\mathbb{Z}^d$  jest chwilowy.



**Rysunek 3.8.** Przykładowa realizacja dwuwymiarowego symetrycznego spaceru losowego. Początek trajektorii zaznaczony jest kolorem czarnym a koniec czerwonym

### 3.4 Twierdzenia o zbieżności

Badać będziemy teraz nierozkładalne, dodatnio rekurencyjne łańcuchy Markowa  $\{X_n\}_{n \in \mathbb{N}}$  na przestrzeni stanów  $S$ . Chcemy odpowiedzieć na pytanie kiedy ciąg  $X_n$  jest zbieżny według rozkładu. Przypomnijmy, że dla przeliczalnej przestrzeni stanów ciąg  $\{X_n\}_{n \in \mathbb{N}}$  jest zbieżny według rozkładu, jeżeli dla każdego  $y \in S$  istnieje liczba  $\mu_\infty(y)$  taka, że



**Rysunek 3.9.** Przykładowa realizacja trójwymiarowego symetrycznego spaceru losowego

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n = y] = \mu_\infty(y)$$

oraz

$$\sum_{y \in S} \mu_\infty(y) = 1.$$

#### Definicja 3.42

Niech  $T(x) = \{n : P^n(x, x) > 0\}$ , wówczas  $\text{NWD}\{T(x)\}_{x \in S}$  nazywane jest **okresem** stanu  $x$ . "Łańcuch Markowa jest **nieokresowy** jeżeli okres każdego stanu wynosi 1.

#### Przykład 3.43

Rozważmy spacer losowy na  $\mathbb{Z}_m = \{0, 1, \dots, m-1\}$  zdefiniowany następująco:  $X_{n+1} = X_n \pm 1 \pmod{m}$ , gdzie  $+1$  i  $-1$  są wybrane z prawdopodobieństwem  $1/2$ . Jeżeli  $m$  jest liczbą parzystą, to spacer ten jest periodyczny. Okres każdego punktu jest wówczas równy 2. Zauważmy, że z uwagi na okresowość rozkład  $X_n$  nie może być zbieżny.

Można uniknąć takiej sytuacji rozważając tzw. 'leniwy spacer losowy', gdzie piechur pozostaje w swojej pozycji z prawdopodobieństwem  $1/2$  i przechodzi sąsiadów z prawdopodobieństwem  $1/4$ .

**Lemat 3.44**

Jeżeli łańcuch Markowa jest nieredukowalny, to wszystkie punkty mają ten sam okres.

**Wniosek 3.4.1.** *Jeżeli łańcuch Markowa jest nieredukowalny i istnieje stan  $x \in S$ , którego okres wynosi 1, to proces ten jest nieokresowy.*

Aby określić w zbieżność dla łańcuchów Markowa posłużymy się odległością całkowitego wahania.

**Definicja 3.45**

Dla miar probabilistycznych  $\nu$  i  $\mu$  na  $S$  ich odległość w sensie całkowitego wahania definiujemy jako

$$d_{TV}(\nu, \mu) = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|.$$

Zauważmy, że jeżeli dla ciągu miar probabilistycznych  $\{\mu_n\}_{n \in \mathbb{N}}$  i  $\mu_\infty$  zachodzi

$$\lim_{n \rightarrow \infty} d_{TV}(\mu_n, \mu_\infty) = 0,$$

to  $\mu_n$  zbiegają słabo do  $\mu_\infty$ , czyli  $\mu_n(x) \rightarrow \mu_\infty(x)$  dla każdego  $x \in S$ .

**Fakt 3.46**

Dla miar probabilistycznych  $\mu$  i  $\nu$

$$d_{TV}(\mu, \nu) = \sup_{A \subseteq S} \mu(A) - \nu(A).$$

*Dowód.* Dla każdego  $A \subseteq S$  mamy

$$\mu(A) - \nu(A) = \sum_{x \in S} \mu(x) - \nu(x).$$

Rozważając zatem zbiór

$$A_0 = \{y \in S : \mu(y) \geq \nu(y)\}.$$

Widzimy, że  $\mu(A_0) - \nu(A_0) \geq \mu(A) - \nu(A)$  dla każdego  $A \subseteq S$ . Innymi słowy

$$\sup_{A \subseteq S} \mu(A) - \nu(A) = \mu(A_0) - \nu(A_0).$$

Mamy też

$$\begin{aligned}
d_{TV}(\mu, \nu) &= \frac{1}{2} \sum_{y \in A_0} \mu(y) - \nu(y) + \frac{1}{2} \sum_{y \in S \setminus A_0} \nu(y) - \mu(y) \\
&= \frac{1}{2} (\mu(A_0) - \nu(A_0)) + \frac{1}{2} (\nu(S \setminus A_0) - \mu(S \setminus A_0)) \\
&= \mu(A_0) - \nu(A_0) = \sup_{A \subseteq S} \mu(A) - \nu(A).
\end{aligned}$$

□

**Fakt 3.47**

Niech  $(M, N)$  będzie wektorem losowym o wartościach w  $S \times S$  taki, że  $M$  ma rozkład  $\mu$  i  $N$  ma rozkład  $\nu$ , tj.

$$\mathbb{P}[M = x] = \mu(x) \quad \text{oraz} \quad \mathbb{P}[N = x] = \nu(x)$$

dla każdego  $x \in S$ . Wówczas

$$\mathbb{P}[M \neq N] \geq d_{TV}(\mu, \nu).$$

*Dowód.* Zauważmy, że dla dowolnego  $A \subseteq S$  mamy

$$\mathbb{P}[M \neq N] \geq \mathbb{P}[M \in A, N \notin A] \geq \mathbb{P}[M \in A] - \mathbb{P}[N \in A]$$

Stąd

$$\mathbb{P}[M \neq N] \geq \sup_{A \subseteq S} \mu(A) - \nu(A) = d_{TV}(\mu, \nu)$$

□

Niech teraz  $\{X_n\}_{n \in \mathbb{N}}$  oraz  $\{Y_n\}_{n \in \mathbb{N}}$  będą łańcuchami markowa z tą samą funkcją przejścia  $P$  oraz rozkładami początkowymi odpowiednio  $\mu$  oraz  $\nu$ . Dla każdego  $n \in \mathbb{N}$  oznaczmy

$$\mu_n(x) = \mathbb{P}[X_n = x] \quad \text{oraz} \quad \nu_n(x) = \mathbb{P}[Y_n = x].$$

**Twierdzenie 3.48** (o zbieżności)

Założmy, że macierz przejścia  $P$  jest nieredukowalna, nieokresowa i dodatnio rekurencyjna (tj. łańcuch  $\{X_n\}_{n \in \mathbb{N}}$  taki jest). Wówczas

$$\lim_{n \rightarrow \infty} d_{TV}(\mu_n, \nu_n) = 0.$$

Zauważmy, że przy założeniach powyższego twierdzenia istnieje jedyna miara stacjonarna  $\pi$ . Zauważmy, że jeżeli  $\nu = \pi$ , to  $\nu_1 = \nu_0 P = \pi P = \pi$ . Indukcyjne pokazujemy, że dla każdego  $n$ ,  $\nu_n = \pi$ . Wobec tego

$$\lim_{n \rightarrow \infty} d_{TV}(\mu_n, \pi) = 0.$$

### Wniosek 3.49

Jeżeli  $\{X_n\}_{n \in \mathbb{N}}$  jest nieredukowalnym, nieokresowym i dodatnio rekurencyjnym łańcuchem Markowa, to dla każdego  $x \in S$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n = x] = \pi(x),$$

gdzie  $\pi$  jest miarą stacjonarną dla  $\{X_n\}_{n \in \mathbb{N}}$ .

### Twierdzenie 3.50

Założmy, że macierz przejścia  $P$  jest nieredukowalna, nieokresowa i dodatnio rekurencyjna (tj. łańcuch  $\{X_n\}_{n \in \mathbb{N}}$  taki jest). Rozważmy czas zatrzymania

$$\tau = \inf\{n \in \mathbb{N} : X_n = Y_n\}. \quad (3.4.2)$$

Wówczas  $\mathbb{P}[\tau < \infty] = 1$ .

*Dowód.* Coming soon.. □

*Dowód twierdzenia o zbieżności.* Niech  $\tau$  będzie czasem zatrzymania danym przez (3.4.2). Rozważmy proces stochastyczny

$$\tilde{Y}_n = \begin{cases} Y_n & n \leq \tau \\ X_n & n \geq \tau \end{cases}$$

Korzystając z mocnej własności Markowa pokazujemy, że  $\{\tilde{Y}_n\}_{n \in \mathbb{N}}$  jest łańcuchem Markowa z funkcją przejścia  $P$  i rozkładem początkowym  $\nu$ . Stąd  $Y_n$  i  $\tilde{Y}_n$  mają ten sam rozkład dla każdego  $n \in \mathbb{N}$ . Wobec tego

$$d_{TV}(\mu_n, \nu_n) \leq \mathbb{P}[X_n \neq \tilde{Y}_n] = \mathbb{P}[\tau \geq n] \rightarrow 0.$$

□

## 3.4.1 Tempo zbieżności na skończonej przestrzeni stanów

### Lemat 3.51

Jeżeli łańcuch Markowa jest nieredukowalny i nieokresowy, to istnieje  $N$  takie, że  $P^n(x, y) > 0$  dla wszystkich  $x, y \in S$  oraz  $n \geq N$ .

### Twierdzenie 3.52

Jeżeli  $\{X_n\}$  jest nieredukowalnym i nieokresowym łańcuchem Markowa o mierze stacjonarnej  $\pi$ , to istnieją  $q < 1$  i  $C$  takie, że dla każdego  $x, y \in S, n \in \mathbb{N}$

$$|P^n(x, y) - \pi(y)| \leq Cq^n.$$

Dowód. Niech

$$m_y^n = \min_x P^n(x, y), \quad M_y^n = \max_x P^n(x, y).$$

Pokażemy, że dla każdego ustalonego  $y \in S$  ciąg  $\{m_y^n\}_n$  jest rosnący, a ciąg  $\{M_y^n\}_n$  jest malejący, a następnie wykażemy, że ich różnica maleje wykładniczo szybko.

**Ciąg  $m_y^n$  jest rosnący:**

$$m_y^{n+1} = \min_x \sum_z P(x, z) P^n(z, y) \geq \min_x \sum_z P(x, z) \min_z P^n(z, y) = m_y^n.$$

Analogicznie dowodzimy  $M_y^{n+1} \leq M_y^n$ .

**Szacowanie różnicy  $M_y^n - m_y^n$ .** Weźmy  $N$  takie jak w lemacie 3.51. Zdefiniujmy

$$\varepsilon = \min_{x, y \in S} P^N(x, y) > 0.$$

Wówczas, z równania Chapmana-Kołmogorowa

$$\begin{aligned} P^{n+N}(x, y) &= \sum_z (P^N(x, z) - \varepsilon P^n(y, z)) P^n(z, y) + \varepsilon \sum_z P^n(y, z) P^n(z, y) \\ &\geq m_y^n \sum_z (P^N(x, z) - \varepsilon P^n(y, z)) + \varepsilon P^{2n}(y, y) \\ &\geq m_y^n (1 - \varepsilon) + \varepsilon P^{2n}(y, y), \end{aligned}$$

Stąd wynika, biorąc minimum po  $x$ , że

$$m_y^{n+N} \geq m_y^n (1 - \varepsilon) + \varepsilon P^{2n}(y, y)$$

i analogicznie

$$M_y^{n+N} \leq M_y^n (1 - \varepsilon) + \varepsilon P^{2n}(y, y).$$

Wnioskujemy więc

$$M_y^{n+N} - m_y^{n+N} \leq (M_y^n - m_y^n)(1 - \varepsilon),$$

a następnie przez indukcję

$$M_y^{n+kN} - m_y^{n+kN} \leq (M_y^n - m_y^n)(1 - \varepsilon)^k$$

Powyższy rachunek pokazuje więc, że podciąg  $M_y^{kN} - m_y^{kN}$  zbiega do zera gdy  $k \rightarrow \infty$ , więc również  $M_y^n - m_y^n$ , jako ciąg monotoniczny zbiega do 0. W szczególności oba ciągi  $m_y^n$  i  $M_y^n$  mają tę samą granicę. Oznaczmy ją przez

$$\bar{\pi}(y) = \lim_{n \rightarrow \infty} m_y^n.$$

Wówczas dla każdego  $y$ ,  $\bar{\pi}(y) \geq m_y(n) \geq \varepsilon > 0$  oraz dla  $n = kN + m$  ( $m < N$ )

$$|P^n(x, y) - \bar{\pi}(y)| \leq M_y^n - m_y^n \leq (M_y^m - m_y^m)(1 - \varepsilon)^k \leq (1 - \varepsilon)^k = [(1 - \varepsilon)^{1/N}]^{kN} \leq C\gamma^n.$$

Pozostaje więc do wykazania, że  $\bar{\pi}$  jest równe mierze stacjonarnej  $\pi$ . Zauważmy najpierw, że  $\bar{\pi}$  jest miarą probabilistyczną:

$$\sum_y \bar{\pi}(y) = \lim_{n \rightarrow \infty} \sum_y P^n(x, y) = 1.$$

Następnie piszemy

$$\pi(y) = \lim_n \sum_{x \in S} \pi(x) P^n(x, y) = \sum_{x \in S} \pi(x) \lim_n P^n(x, y) = \sum_{x \in S} \pi(x) \bar{\pi}(y) = \bar{\pi}(y).$$

□

## Odwracalne łańcuchy Markowa

### Definicja 3.53

Niech  $\{X_n\}$  będzie łańcuchem Markowa na przestrzeni stanów  $\mathcal{S}$  z macierzą przejścia  $P$ . Miara probabilistyczna  $\pi$  na  $\mathcal{S}$  jest nazywana **odwracalną** dla łańcucha Markowa  $\{X_n\}$  jeżeli dla dowolnych  $x, y \in \mathcal{S}$  mamy

$$\pi(x)P(x, y) = \pi(y)P(y, x). \quad (3.4.3)$$

"Łańcuch Markowa jest nazywany **odwracalnym** jeżeli istnieje dla niego odwracalna miara.

### Lemat 3.54

Jeżeli  $\pi$  jest miarą odwracalną dla pewnego łańcucha Markowa, to jest również dla niego miarą stacjonarną

*Dowód.* Korzystając z powyższej definicji oraz stochastyczności macierzy  $P$ :

$$\pi P(x) = \sum_{y \in \mathcal{S}} \pi(y)P(y, x) = \sum_{y \in \mathcal{S}} \pi(x)P(x, y) = \pi(x)$$

□

**Uwaga 3.4.4.** Z powyższego lematu wynika, że jeżeli macierz przejścia  $P$  jest symetryczna, to miara jednostajna jest miarą stacjonarną.

Zauważmy, że jeżeli łańcuch jest odwracalny, to dla każdego  $n$ :

$$\begin{aligned}\mathbb{P}_\pi[X_0 = x_0, \dots, X_n = x_n] &= \pi(x_0)P(x_0, x_1) \dots P(x_{n-1}, x_n) \\ &= \pi(x_n)P(x_n, x_{n-1}) \dots P(x_1, x_0) \\ &= \mathbb{P}_\pi[X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0]\end{aligned}$$

(zapis  $X_\pi$  oznacza, że  $X_0$  ma rozkład  $\pi$ ). Zatem rozkład  $(X_0, X_1, \dots, X_n)$  jest taki sam jak rozkład  $(X_n, X_{n-1}, \dots, X_0)$ , co wyjaśnia pojęcie odwracalności (odwracając czas otrzymujemy ten sam proces).

### 3.4.2 Przykłady

#### Przykład 3.55

Prosty spacer losowy na grafie (przykład ??) jest odwracalny. Przypomnijmy, że miara  $\pi(x) = \deg(x)/2|E|$  jest stacjonarna. Mamy:

$$\pi(x)P(x, y) = \frac{\deg(x)}{2|E|} \cdot \frac{frm[o]--\{x\sim y\}}{\deg(x)} = \frac{frm[o]--\{x\sim y\}}{2|E|} = \pi(y)P(y, x)$$

#### Przykład 3.56

Rozważmy zmodyfikowany spacer losowy na  $\mathbb{Z}_n$ . Cząsteczka porusza się zgodnie ze wskazówkami zegara (+1) z prawdopodobieństwem  $p$  i w przeciwną stronę z prawdopodobieństwem  $1 - p$ . Wówczas miarą stacjonarną jest miara jednostajna:  $\pi(k) = 1/n$ , gdyż

$$\pi P(k) = \sum_{j \in \mathbb{Z}_n} \pi(j)P(j, k) = \pi(k-1)p + \pi(k+1)(1-p) = \frac{1}{n} = \pi(k).$$

Jeżeli jednak  $p \neq 1/2$ , to miara ta nie jest odwracalna:

$$\pi(k)P(k, k+1) = \frac{p}{n} \neq \frac{1-p}{n} = \pi(k+1)P(k+1, k).$$

#### Przykład 3.57

**Spacer losowy na  $m$ -hiperkostce.**  $m$ -hiperkostka jest to graf, którego wierzchołkami są ciągi binarne długości  $m$ , tzn  $\mathcal{S} = \{0, 1\}^m$ . Dwa wierzchołki są połączone ze sobą krawędzią jeżeli różnią się dokładnie na jednej współrzędnej. Tak więc każdy wierzchołek ma dokładnie  $m$  sąsiadów.

Prosty spacer losowy polega na przejściu z wierzchołka  $x = (x_1, \dots, x_m)$  do jednego z jego sąsiadów z prawdopodobieństwem  $1/m$  (jak w przykładzie ??). Dla  $m = 3$  jest to prosty spacer losowy po wierzchołkach sześcianu. Zatem miara jednostajna jest miarą stacjonarną. Zauważmy, że spacer ten jest okresowy. Istotnie, jeżeli spacer losowy startuje w  $(0, \dots, 0)$ , to w  $n$ -tym kroku liczba jedynek w  $X_n$  ma tę samą parzystość, co  $n$ . Zatem do punktu wyjścia  $(0, \dots, 0)$  możemy wrócić jedynie w parzystych krokach.

Często rozważa się **leniwy spacer losowy**, gdy cząsteczka zostaje w danym wierzchołku z prawdopodobieństwem  $1/2$ , a z prawdopodobieństwem  $1/2$  przechodzi jednostajnie do jednego z sąsiadów. Spacer ten jest nieokresowy i miara jednostajna jest jego miarą stacjonarną.

### Przykład 3.58

**Proces urodzin i śmierci.** Jest to proces na przestrzeni stanów  $\mathcal{S} = \{0, 1, \dots, m\}$ . W każdym kroku pozycja może zmienić się maksymalnie o jeden. Wygodnie jest myśleć, że aktualny stan symbolizuje rozmiar populacji. Macierz przejścia zależy od trzech ciągów:  $\{(p_k, r_k, q_k)\}_{k=0}^m$ , gdzie:

- $p_k$  jest prawdopodobieństwem przejścia z  $k$  do  $k + 1$ , dla  $0 \leq k < m$ ;
- $q_k$  jest prawdopodobieństwem przejścia z  $k$  do  $k - 1$ , dla  $0 < k \leq m$ ;
- $r_k$  jest prawdopodobieństwem pozostania w  $k$ , dla  $0 \leq k \leq m$ ;
- $p_k + r_k + q_k = 1$ ;
- $q_0 = p_m = 0$

"Łatwo sprawdzić, że jest to proces odwracalny. Mianowicie miara

$$\tilde{\pi}(k) = \prod_{i=1}^k \frac{p_{i-1}}{q_i}, \quad \tilde{\pi}(0) = 1$$

jest odwracalna, tzn zachodzi

$$p_{k-1} \tilde{\pi}(k-1) = q_k \tilde{\pi}(k),$$

czyli równanie (3.4.3) jest spełnione. Zatem miara

$$\pi(k) = \frac{\tilde{\pi}(k)}{\sum \tilde{\pi}(k)}$$

jest stacjonarna.

## 3.5 MCMC

W tym rozdziale będziemy badać następujący problem: mając dany skończony zbiór  $\mathcal{S}$  oraz miarę probabilistyczną  $\pi$  na  $\mathcal{S}$  chcemy wygenerować zmienną losową o rozkładzie  $\pi$ . Interesować nas będą duże skończone zbiory. Często nie będziemy nawet potrafili określić ich licznosci ani nawet precyzyjnie podać miary  $\pi$ . Przedstawimy kilka przykładów bazujących na metodzie MCMC (Markov Chain Monte Carlo). W tym celu na zbiorze  $\mathcal{S}$  wprowadzimy strukturę grafu i zdefiniujemy odpowiedni spacer losowy z miarą stacjonarną  $\pi$ . Spacery będą nieredukowalne i nieokresowe. Wówczas z twierdzenia 3.52 rozkład spaceru dąży do  $\pi$ . Wyznaczanie elementów o zadanym rozkładzie znajduje zastosowania w szeregu praktycznych problemów, które zostaną omówione poniżej takich jak problemy optymalizacyjne (np. problem pakowania plecaka), deszyfrowanie kodów, ...

### 3.5.1 Drzewo rozpinające

#### Przykład 3.59

**Drzewo rozpinające graf.** Niech  $G = (V, E)$  będzie skończonym grafem. Drzewem rozpinającym  $T$  graf  $G$  nazywamy spójny podgraf  $G$ , który zawiera wszystkie wierzchołki  $V$ , ale nie ma żadnych cykli. Wówczas drzewo spełnia (zadanie)

- $T$  ma  $|V| - 1$  krawędzi
- Każdy podgraf  $G$  posiadający  $|V| - 1$  krawędzi i zawierający wszystkie wierzchołki, bez cykli, jest drzewem rozpinającym.
- Dodanie krawędzi do drzewa rozpinającego  $T$  spowoduje utworzenie cyklu.

Przez  $\mathcal{S}$  oznaczmy zbiór wszystkich drzew rozpinających  $G$ . Chcemy rozważać następujący problem: jak wygenerować losowe (jednostajnie) drzewo rozpinające graf  $G$ ? Podkreślmy, że chcemy, aby wylosowanie każdego drzewa rozpinającego było jednakowo prawdopodobne.

W przypadku małych grafów można próbować wypisać wszystkie drzewa rozpinające. Jednak już dla grafów rozpiętych na 10 wierzchołkach, liczba drzew rozpinających może przekraczać  $10^8$ .

Metoda MCMC polega na wprowadzeniu struktury grafu na zbiorze  $\mathcal{S}$ , a następnie zdefiniowaniu odpowiedniego spaceru losowego. Zbiorem wierzchołków jest zbiór  $\mathcal{S}$ , tak więc wierzchołkiem jest drzewo rozpinające. Następnie definiujemy łańcuch Markowa  $\{X_n\}$  na  $\mathcal{S}$ , w którym określimy też krawędzie w grafie. Startujemy z dowolnego elementu  $\mathcal{S}$ . Załóżmy, że znamy wartość  $X_n$ , wówczas generujemy  $X_{n+1}$  następująco:

1. Wylosujmy jednostajnie krawędź  $e$ , która nie jest w  $X_n$ . Takich krawędzi jest  $|E| - (|V| - 1)$ .
2. Niech  $C$  będzie jedynym cyklem w  $X_n \cup \{e\}$ .
3. Wybierzmy jednostajnie krawędź  $e'$  w  $C$ .
4. Przyjmijmy  $X_{n+1} := (X_n \cup \{e\}) \setminus \{e'\}$ .

Ćwiczenie: Narysuj przykładowy graf i wykonaj kilka pierwszych kroków powyższego algorytmu.

#### Lemat 3.60

Powyższy łańcuch Markowa jest nieokresowy, nieredukowalny i odwracalny.

*Dowód.* Proces jest nieokresowy, bo  $P(T, T) > 0$  dla każdego  $T \in \mathcal{S}$ , gdyż z dodatnim prawdopodobieństwem, w powyższym algorytmie  $e = e'$ . Nieredukowalność zostanie pokazana na ćwiczeniach.

Pokażemy, że proces jest symetryczny, tzn.  $P(T, T') = P(T', T)$  dla dowolnych  $T \neq T'$ , wtedy jest również odwracalny. Załóżmy najpierw, że

$P(T, T') > 0$ . Wtedy  $T' = (T \cup \{e_1\}) \setminus \{e_2\}$  dla pewnych krawędzi  $e_1$  oraz  $e_2$ . Niech  $C$  będzie cyklem zawartym w  $T \cup \{e_1\}$ . Wówczas  $e_2$  musi być też elementem cyklu  $C$ . Mamy więc

$$\begin{aligned} P(T, T') &= \mathbb{P}[e = e_1, e' = e_2 | X_n = T] \\ &= \mathbb{P}[e = e_1 | X_n = T] \cdot \mathbb{P}[e' = e_2 | X_n = T, e = e_1] \\ &= \frac{1}{|E| - (|V| - 1)} \cdot \frac{1}{|C|} \end{aligned}$$

Analogicznie, ponieważ  $T = (T' \cup \{e_2\}) \setminus \{e_1\}$ , mamy

$$\begin{aligned} P(T', T) &= \mathbb{P}[e = e_2, e' = e_1 | X_n = T'] \\ &= \frac{1}{|E| - (|V| - 1)} \cdot \frac{1}{|C|} \end{aligned}$$

Zatem

$$P(T, T') = P(T', T)$$

przy założeniu, że  $P(T, T') > 0$ . Przypadek  $P(T, T') = 0$  pomijamy  $\square$

Z powyższego lematu oraz z lematu 3.54, wynika że miara stacjonarna procesu  $\{X_n\}$  jest jednostajna. Z twierdzenia 3.52 rozkłady  $X_n$  zbiegają do rozkładu jednostajnego i można przyjąć, że  $X_n$  jest bliski rozkładowi jednostajnemu dla dużych wartości  $n$ . Zbieżność jest wykładniczo szybka, ale parametry zależą od struktury grafu i są bardzo trudne do określenia.

### 3.5.2 Problem pakowania plecaka (The knapsack problem)

#### Przykład 3.61

Załóżmy, że mamy  $m$  przedmiotów, każdy o wadze  $w_i$  i wartości  $v_i$ . Dysponujemy plecakiem, który pomieści maksymalnie  $b$  kilogramów. Które przedmioty należy wybrać, aby zmieściły się one do plecaka i miały jak największą wartość? Jest to problem NP-zupełny ...

Formalnie. Mamy zadane:  $w = (w_1, \dots, w_m)$ ,  $v = (v_1, \dots, v_m)$ ,  $b$ . Szukamy wektora decyzyjnego  $z = (z_1, \dots, z_m) \in \{0, 1\}^m$ , tzn.  $z_i = 1$ , jeżeli  $i$ -ty przedmiot jest włożony do plecaka. Przez  $\mathcal{S}$  oznaczamy zbiór dopuszczalnych pakowań plecaka, tzn.

$$\mathcal{S} = \{z \in \{0, 1\}^m : (w, z) \leq b\}$$

Naszym celem jest zoptymalizowanie wartości  $(v, z)$  pod warunkiem  $z \in \mathcal{S}$ .

#### Algorytm 1:

1. Losujemy jednostajnie  $z \in \{0, 1\}^m$ .

2. Jeżeli  $z \in \mathcal{S}$ , to akceptujemy  $z$ , w przeciwnym razie wracamy do punktu 1.

Dla wylosowanego  $z$  sprawdzamy wartość plecaka. Możemy powtórzyć algorytm wielokrotnie i wybrać najlepsze rozwiązanie.

Każdy krok może być szybko obliczony. Jednak jeżeli  $m$  jest duże, to prawdopodobieństwo zaakceptowania wylosowanego  $z$  może być bardzo małe. Dla przykładu jeżeli  $w_i = 1$  i  $b = m/3$ , to można pokazać (zadanie!), że prawdopodobieństwo zaakceptowania jest około  $(0,83)^m$  (dla  $m = 100$  to jest około  $10^{-8}$ ). Zatem tylko mała część losowań zostanie zaakceptowana i trudno przypuszczać, że będą one bliskie optymalnemu rozwiązaniu.

**Algorytm 2 (MCMC):** Rozważmy graf ze zbiorem wierzchołków  $\mathcal{S}$ . Dwie krawędzie są połączone ze sobą, gdy różnią się na dokładnie jednej współrzędnej (to jest podgraf hiperkostki  $\mathbb{Z}_2^m$ ). Konstruujemy spacer losowy na  $\mathcal{S}$ . Mając zadany  $X_n = (z_1, \dots, z_m)$

1. Wybieramy losowo (jednostajnie) jedną ze współrzędnych:  $J \in \{1, \dots, m\}$ .
2. Następnie zamieniamy  $J$ -tą współrzędną:  $z_J = 1 - z_J$  i oznaczmy nowy wektor przez  $z'$ .
3. Jeżeli  $z' \in \mathcal{S}$ , to podstawiamy  $X_{n+1} = z'$ , w przeciwnym razie  $X_{n+1} = X_n = z$ .

Innymi słowy: wylosujmy przedmiot; jeżeli jest w plecaku, to go wyjmijmy; jeżeli nie, to włóżmy go do plecaka, ale pod warunkiem, że się zmieści.

### Lemat 3.62

Jeżeli  $\sum w_i > b$ , to powyższy łańcuch Markowa jest nieokresowy, nieredukowalny i odwracalny.

Dowód pokażemy podczas ćwiczeń.

Uwagi:

- Miarą stacjonarną dla tego łańcucha Markowa jest miara jednostajna, a więc dla dużych  $t$ ,  $\mathbb{P}[X_t = z] \sim 1/|\mathcal{S}|$ . Umiemy zatem aproksymować tę miarę, mimo iż nie znamy dokładnej wartości  $|\mathcal{S}|$
- Próbkowanie z rozkładu jednostajnego, może nie być skuteczne, ponieważ może być niewiele upakowań, które są bliskie optymalnemu, tzn. ich miara może być niezauważalna. Znacznie lepiej jest losować elementy wg rozkładu, który daje większą wagę 'cennym upakowaniom', np.  $\pi(z) = C^{-1}e^{(v,z)}$ , gdzie  $C$  jest (nieznana) stałą normalizującą.

### 3.5.3 Algorytm Metropolis

Algorytm pozwala na przybliżenie zmiennej losowej ze zbioru  $\mathcal{S}$  o rozkładzie  $\pi$ . W tym celu konstruujemy graf  $G$  ze zbiorem wierzchołków  $\mathcal{S}$ . Zbiór

krawędzi może być niemal dowolny z zachowaniem jednak poniższych warunków:

- Graf powinien być spójny, aby otrzymany łańcuch Markowa był nieredukowalny.
- Poszczególne wierzchołki nie powinny mieć zbyt wielu sąsiadów, gdyż spacer stanie się zbyt trudny do symulowania w praktyce.

Dla  $x \in \mathcal{S}$ , przez  $d(x)$  oznaczamy liczbę sąsiadów wierzchołka  $x$ . Definiujemy macierz przejścia

$$P(x, y) = \begin{cases} \frac{1}{d(x)} \min \left\{ \frac{\pi(y)d(x)}{\pi(x)d(y)}, 1 \right\} & \text{jeżeli } x \sim y; \\ 0 & \text{jeżeli } x \neq y \text{ oraz } x \not\sim y; \\ 1 - \sum_{z \sim x} \frac{1}{d(x)} \min \left\{ \frac{\pi(z)d(x)}{\pi(x)d(z)}, 1 \right\} & \text{jeżeli } x = y. \end{cases}$$

Powyższy mechanizm można opisać nieco inaczej. Załóżmy, że  $X_n = x$ . Wybierzmy losowo w sposób jednostajny, tzn. z prawdopodobieństwem  $1/d(x)$  jednego z sąsiadów  $x$  i oznaczmy go przez  $y$ . Następnie podstawmy

$$X_{n+1} = \begin{cases} y \text{ z prawdopodobieństwem } \min \left\{ \frac{\pi(y)d(x)}{\pi(x)d(y)}, 1 \right\} \\ x \text{ z prawdopodobieństwem } 1 - \min \left\{ \frac{\pi(y)d(x)}{\pi(x)d(y)}, 1 \right\}. \end{cases}$$

Zauważmy, że jeżeli graf jest regularny, tzn. każdy wierzchołek ma ten sam stopień, to powyższy algorytm można opisać jeszcze inaczej. Załóżmy, że  $X_n = x$ . Wybierzmy losowo w sposób jednostajny jednego z sąsiadów  $x$  i oznaczmy go przez  $y$ . Następnie

- jeżeli  $\pi(y) \geq \pi(x)$ , to podstawmy  $X_{n+1} = y$
- jeżeli  $\pi(y) < \pi(x)$ , to rzucamy monetę, w której orzeł wypada z prawdopodobieństwem  $\pi(y)/\pi(x)$ . Jeżeli wypadnie orzeł, to  $X_{n+1} = y$ , w przeciwnym razie  $X_{n+1} = x$

Algorytm próbuje więc maksymalizować funkcję  $\pi$ , przechodząc do wierzchołków, w których wartość ta jest większa. Dopuszcza jednak powrót do mniejszych wartości, aby nie utknąć w 'lokalnych maksimach' funkcji  $\pi$ .

### Lemat 3.63

Miarą stacjonarną powyższego łańcucha Markowa jest  $\pi$ .

Dowód przeprowadzimy podczas ćwiczeń

### Przykład 3.64

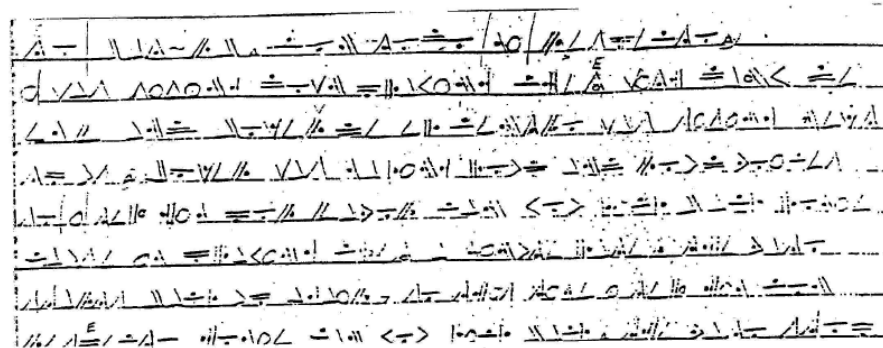
**Problem pakowania plecaka.** Rozważamy strukturę grafu taką samą jak zdefiniowaną powyżej. Definiujemy

$$\pi^\beta(z) = C_\beta^{-1} e^{\beta(v,z)},$$

gdzie  $\beta$  jest pewną ustaloną stałą, a  $C_\beta$  jest stałą normalizującą. Zauważmy, że nie musimy znać jej dokładnej wartości! Definicja macierzy przejścia zależy wyłącznie od ilorazu.

### Przykład 3.65

Persi Diaconis w pracy *The Markov Chain Monte Carlo Revolution* opisuje następujący przykład. Pewnego dnia psycholog ze stanowego więzienia poprosił Departament Statystyki na Uniwersytecie w Stanford o rozszyfrowanie następującego kodu.<sup>1</sup>



Domyślono się natychmiast, że jest to tekst zakodowany przez proste podstawienie, gdzie każdy symbol odpowiada pewnej literze lub znakowi interpunkcyjnemu. Należy więc znaleźć funkcję  $f$  odwzorowującą przestrzeń kodów w zwykły alfabet. Typowa metoda polega na sprawdzeniu statystyk częstości występowania poszczególnych znaków w alfabecie i porównaniu z częstością występowania poszczególnych znaków w zakodowanym tekście. Metoda ta może nie być jednak skuteczna przy krótkich tekstach.

Do odkodowania powyższego kodu użyto skuteczniejszej metody bazującej na sprawdzaniu częstotliwości występowania par znaków. Dokładniej, dla każdych dwóch znaków  $x, y$  można określić częstotliwość występowania  $y$  bezpośrednio po  $x$ . Otrzymujemy w ten sposób macierz  $M(x, y)$ . Definiujemy funkcję wiarygodności

$$G(f) = \prod_i M(f(s_i), f(s_{i+1})),$$

gdzie  $s_i$  przebiega kolejne symbole w zakodowanej wiadomości. Chcemy znaleźć funkcje  $f$  dla których ta wartość jest duża. Te funkcje powinny być dobrymi kandydatami do odkodowania. Przeszukanie wszystkich możliwych funkcji  $f$  nie wchodzi w grę, bo jest ich zbyt wiele (ok. 40!). Zastosowano więc algorytm Metropolis. Należy zdefiniować relację sąsiedztwa, a więc dla zadanej funkcji  $f$  określić funkcje, które możemy otrzymać w

<sup>1</sup> Wszystkie grafiki w tym przykładzie są zaczerpnięte z pracy P. Diaconisa *The Markov Chain Monte Carlo Revolution*

jednym kroku. Mówimy, więc że istnieje krawędź pomiędzy  $f$ , a  $g$ , jeżeli można otrzymać  $g$  z  $f$  używając losowej transpozycji. Algorytm uruchomiono na cytacie z Hamleta, który następnie losowo zakodowano.

ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS  
NOBLER IN THE MIND TO SUFFER THE SLINGS AND ARROWS OF OUTRAGEOUS  
FORTUNE OR TO TAKE ARMS AGAINST A SEA OF TROUBLES AND BY OPPOSING END

Poniżej przedstawione są wyniki działania algorytmu.

```
100 ER ENOHDLAE OHDLO UOZEOUNORU O UOZEO HD OITO HEOQSET IUROFHE HENO ITORUZAEN
200 ES ELOHRNDE OHRNO UOVEOULOSU O UOVEO HR OITO HEOQAET IUSOPHE HELO ITOSUVDL
300 ES ELOHANDE OHANO UOVEOULOSU O UOVEO HA OITO HEOQRET IUSOFHE HELO ITOSUVDL
400 ES ELOHINME OHINO UOVEOULOSU O UOVEO HI OATO HEOQRET AUSOWHE HELO ATOSUVMEL
500 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
600 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
900 ES ELOHANME OHANO UODEOULOSU O UODEO HA OITO HEOQRET IUSOWHE HELO ITOSUDMEL
1000 IS ILOHANMI OHANO RODIORLOSR O RODIO HA OETO HIOQUIT ERSOWHI HILO ETOSRDMIL
1100 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1200 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1300 ISTILOHARMITOHAROT ODIO LOS TOT ODIOTHATOENOTHIOQUINTE SOWHITHILOTENOS DMIL
1400 ISTILOHAMRITOHAMOT OFIO LOS TOT OFIOTHATOENOTHIOQUINTE SOWHITHILOTENOS FRIL
1600 ESTEL HAMRET HAM TO CE OL SOT TO CE THAT IN THE QUENTIOS WHETHEL TIN SOCREL
1700 ESTEL HAMRET HAM TO BE OL SOT TO BE THAT IN THE QUENTIOS WHETHEL TIN SOBREL
1800 ESTER HAMLET HAM TO BE OR SOT TO BE THAT IN THE QUENTIOS WHETHER TIN SOBLER
1900 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
2000 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
```

Jak widać ok. 2000 iteracji wystarczyło, aby odkodować wiadomość. Zaletą tej metody jest fakt, iż działa ona już dla bardzo krótkich tekstów. Dla oryginalnego tekstu więziennego otrzymano następujący wynik

```
to bat-rb. con todo mi respeto. i was sitting down playing chess with
danny de emf and boxer de el centro was sitting next to us. boxer was
making loud and loud voices so i tell him por favor can you kick back
homie cause im playing chess a minute later the vato starts back up again
so this time i tell him con respecto homie can you kick back. the vato
stop for a minute and he starts up again so i tell him check this out shut
the f**k up cause im tired of your voice and if you got a problem with it
we can go to celda and handle it. i really felt disrespected thats why i
told him. anyways after i tell him that the next thing I know that vato
slashes me and leaves. dy the time i figure im hit i try to get away but
the c.o. is walking in my direction and he gets me right dy a celda. so i
go to the hole. when im in the hole my home boys hit doxer so now "b" is
also in the hole. while im in the hole im getting schoold wrong and
```

## 3.6 Definicja ogólna

Naszym celem jest teraz przedstawienie formalnej definicji łańcuchów Markowa i pokazanie ich fundamentalnych własności.

### Definicja 3.66

Funkcję  $\mathbf{P}(x, A)$ , dla  $x \in \mathbb{R}$ ,  $A \in \mathcal{B}or(\mathbb{R})$  nazywamy (dyskretną) funkcją przejścia jeżeli

- dla każdego  $x \in \mathbb{R}$ ,  $\mathbf{P}(x, \cdot)$  jest miarą probabilistyczną na  $\mathbb{R}$ ;
- dla każdego  $A \in \mathcal{B}or(\mathbb{R})$ ,  $\mathbf{P}(\cdot, A)$  jest mierzalna.

Funkcje przejścia posłużą nam do zdefiniowania i badania procesów Markowa. O funkcji przejścia należy myśleć jak o

$$\mathbf{P}(x, A) = \mathbb{P} \left[ \begin{array}{l} \text{w chwili } n+1 \\ \text{proces trafia w } A \end{array} \middle| \begin{array}{l} \text{w chwili } n \text{ proces} \\ \text{jest w punkcie } x \end{array} \right].$$

Należy mieć na względzie, że w zupełnej ogólności zdarzenie polegające na tym, że w chwili  $n$  proces jest w punkcie  $x$  może mieć prawdopodobieństwo zero. Stąd konieczność wprowadzenia funkcji przejścia.

### Przykład 3.67

Niech  $\mu$  będzie dowolnym rozkładem prawdopodobieństwa na  $\mathbb{R}$ . Wówczas  $\mathbf{P}(x, A) = \mu(A)$  jest funkcją przejścia na  $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}or(\mathbb{R}))$ .

### Przykład 3.68

Niech  $Y$  będzie dowolną zmienną losową. Wówczas

$$\mathbf{P}(x, A) = \mathbb{P}[x + Y \in A]$$

zadaje funkcje przejścia na  $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}or(\mathbb{R}))$ .

### Przykład 3.69

Niech  $(Q, T)$  będzie dowolnym dwuwymiarowym wektorem losowym. Wówczas  $\mathbf{P}(x, A) = \mathbb{P}[Qx + T \in A]$  jest funkcją przejścia na  $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}or(\mathbb{R}))$ .

Przypomnijmy, że dla ciągu zmiennych losowych  $X = \{X_n\}$  określamy jego naturalną filtrację przez  $\mathcal{F}_n^X = \sigma(X_k : k \leq n)$ .

### Definicja 3.70

Ciąg zmiennych losowych  $\{X_n\}$  o wartościach w  $(S, \mathcal{S})$  nazywamy (jednorodnym) **łańcuchem Markowa** z funkcją przejścia  $\mathbf{P}$  jeżeli każdego

$A \in \mathcal{S}$ ,

$$\mathbb{P}[X_{n+1} \in A \mid \mathcal{F}_n^X] = \mathbf{P}(X_n, A).$$

### Przykład 3.71

Spacer losowy jest łańcuchem Markowa. Rzeczywiście, ciągu iid  $\{Y_k\}$  o rozkładzie  $\mu$  definiujemy spacer przez  $S_0 = 0$ ,  $S_n = Y_1 + Y_2 + \dots + Y_n$ ,  $n \geq 1$ . Mamy  $S_{n+1} = S_n + Y_{n+1}$ . Skoro  $S_n$  i  $Y_{n+1}$  są niezależne, to

$$\mathbb{P}[S_{n+1} \in A \mid \mathcal{F}_n^X] = \mathbb{P}[S_n + X \in A \mid \mathcal{F}_n^X] = \mathbb{E}[\psi(S_n, Y_{n+1}) \mid \mathcal{F}_n^X],$$

gdzie  $\psi(s, y) = \mathbb{1}_{\{s+y \in A\}}$ . Stąd

$$\mathbb{P}[S_{n+1} \in A \mid \mathcal{F}_n^X] = \Psi(S_n),$$

gdzie

$$\Psi(s) = \mathbb{P}[s + Y_{n+1} \in A] = \mathbf{P}(s, A).$$

Istotnie więc  $\{S_n\}$  jest łańcuchem Markowa z funkcją przejścia  $\mathbf{P}$  zdefiniowaną powyżej.

Procesy Markowa charakteryzują się krótką pamięcią. Zauważmy, że

$$\begin{aligned} \mathbb{P}[X_{n+1} \in A \mid X_n] &= \mathbb{E}[\mathbb{P}[X_{n+1} \in A \mid \mathcal{F}_n^X] \mid X_n] = \mathbb{E}[\mathbf{P}(X_n, A) \mid X_n] \\ &= \mathbf{P}(X_n, A) = \mathbb{P}[X_{n+1} \in A \mid \mathcal{F}_n^X]. \end{aligned}$$

Oznacza to, że

$$\mathbb{P}[X_{n+1} \in A \mid X_n] = \mathbb{P}[X_{n+1} \in A \mid \mathcal{F}_n^X].$$

Innymi słowy do aproksymacji rozkładu zmiennej  $X_{n+1}$  znając  $X_1, X_2, \dots, X_n$  wystarcza nam jedynie zmienna  $X_n$ .

### Definicja 3.72

Rozkładem początkowym łańcucha Markowa  $\{X_n\}_{n \in \mathbb{N}}$  nazywamy rozkład zmiennej  $X_0$ .

Definicja łańcucha Markowa zakłada istnienie jednej przestrzeni probabilistycznej, na której określony jest proces  $\{X_n\}$  co wymaga dodatkowej argumentacji. Następujące twierdzenie pozostawimy bez dowodu

### Twierdzenie 3.73

Załóżmy, że dana jest funkcja przejścia  $\mathbf{P}$  przestrzeni stanów  $(S, \mathcal{S})$  oraz niech  $\mu_0$  będzie zadaną miarą probabilistyczną na  $\mathcal{S}$ . Wówczas istnieje przestrzeń probabilistyczna  $(\Omega, \mathcal{F}, \mathbb{P})$  na której można zdefiniować łań-

cuch Markowa  $\{X_n\}$  o funkcji przejścia  $\mathbf{P}$  oraz rozkładzie początkowym  $\mu_0$ .

## Ruch Browna

**Streszczenie** Wprowadzimy matematyczny model ruchów Browna. Jest to proces stochastyczny szczególnie istotny w teorii prawdopodobieństwa. Jak się przekonamy stanowi on zarówno martyngał jak i proces Markowa. Szeroki wachlarz własności probabilistycznych ruchu Browna pozwoli nam bardzo dokładnie przeanalizować jego trajektorie.

### 4.1 Motywacja i konstrukcja

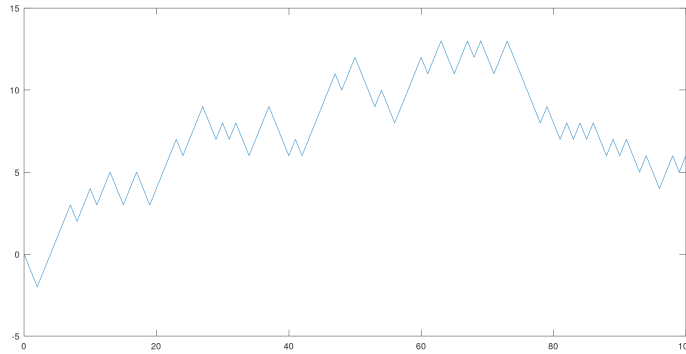
Szkocki botanik Robert Brown w 1827 roku zaobserwował ruchy wykonywane przez pyłki kwiatowe zawieszona w cieczy. Uwagę botanika zwrócił fakt, że ruchy:

- nie są spowodowane przez ruchy cieczy
- nie ustają
- są bardzo nieregularne

Po publikacji R. Browna powstało kilka teorii odnośnie pochodzenia ruchu pyłków, ale dopiero w 1905 roku niezależnie Albert Einstein oraz Marian Smoluchowski powiązali ruchy ze zderzeniami w mikroskali z innymi cząsteczkami. Pierwszy matematyczny opis tego zjawiska pojawił się w 1900 r. w pracy Bacheliera, który badał wykresy notowań giełdowych. Jego definicja nie była jednak zupełnie ścisła. Pierwsza formalna matematyczna konstrukcja została zaprezentowana przez Norberta Wienera w 1926 roku. Stąd też ruch Browna nazywany jest niekiedy procesem Wienera.

Aby zrozumieć filozofię stojącą za matematyczną definicją ruchu Browna rozważmy następujące przybliżenie. Należy mieć na względzie, że nasze rozważania do momentu pojawienia się **Definicji 4.2** nie będą całkowicie formalne. Dla uproszczenia będziemy badać ruch odbywający się w jednym wymiarze. Ustalmy  $T > 0$  i załóżmy, że interesuje nas pozycja  $B_t \in \mathbb{R}$  cząsteczki w chwili  $t \in [0, T]$ . W tym celu dokonujemy obserwacji cząsteczki w dyskretnych chwilach i zapisujemy odczytaną przez nas pozycję  $X_t$ . Zakładamy dodatkowo, że

- Obserwujemy cząsteczkę w równych odstępach  $\Delta t > 0$ ,
- Każdorazowo cząsteczka może przesunąć się w lewo bądź prawo o  $\Delta x > 0$  z prawdopodobieństwem  $1/2$ ,
- cząsteczka zaczyna swój ruch w chwili 0 w punkcie  $X_0 = 0$ .



**Rysunek 4.1.** Kolejne przybliża  $B_t$

Powyżej  $\Delta t$  i  $\Delta x$  są parametrami, których dobór sprecyzujemy niebawem. Będziemy chcieli zbadać probabilistyczne własności  $X_t$  i zbadać zachowanie przy  $\Delta t \rightarrow 0$ . Będzie to prowadziło do zrozumienia  $B_t$  ponieważ

$$B_t = \lim_{\Delta t \rightarrow 0} X_t.$$

Powyższej granicy nie należy traktować formalnie, ponieważ nie wprowadziliśmy odpowiednich narzędzi do jej prawidłowej interpretacji. Brak formalizmu w tym miejscu nie przeszkodzi nam jednak wywnioskować odpowiednich postulatów dla procesu  $(B_t)$ . Jeżeli dokonujemy obserwacji w chwilach  $k\Delta t$ , dla naturalnych  $k$ , to w przedziale czasowym dokonamy  $n_T = \lceil T/\Delta t \rceil$  obserwacji. Niech  $\epsilon_1, \epsilon_2, \dots, \epsilon_{n_T}$  będą niezależnymi zmiennymi losowymi o rozkładzie

$$\mathbb{P}[\epsilon_k = 0] = \mathbb{P}[\epsilon_k = 1] = 1/2$$

dla  $k \leq n_T$ . Wówczas  $S_n = \epsilon_1 + \dots + \epsilon_n$  oznacza liczbę kroków w prawo podczas pierwszych  $n$  obserwacji. Do chwili  $t \in [0, T]$  dokonujemy  $n_t = \lceil t/\Delta t \rceil$  obserwacji i odnotowana przez nas pozycja zapisuje się jako

$$X_t = \Delta x S_{n_t} - \Delta x (n_t - S_{n_t}) = \Delta x \sum_{k=1}^{n_t} (2\epsilon_k - 1).$$

Stąd

$$\mathbb{E}[X_t] = 0, \quad \text{Var}(X_t) = (\Delta x)^2 n_t.$$

Zauważmy, że przy  $\Delta t \rightarrow 0$  wariancja  $X_t$  zachowuje się jak  $t(\Delta x)^2/\Delta t$ . Aby zapobiec eksplozji zakładamy, że  $\Delta x$  jest rzędu  $\sqrt{\Delta t}$ . Dla uproszczenia przyjmijmy

$$(\Delta x)^2 = \Delta t.$$

Wówczas

$$X_t = \sqrt{\frac{\Delta t}{[t/\Delta t]}} \cdot \frac{\sum_{k=1}^{n_t} (2\epsilon_k - 1)}{\sqrt{n_t}}$$

Pierwszy składnik, przy  $\Delta t \rightarrow 0$ , zbiega do  $t$  a dla drugiego zachodzi centralne twierdzenie graniczne

$$\frac{\sum_{k=1}^{n_t} (2\epsilon_k - 1)}{\sqrt{n_t}} \Rightarrow \mathcal{N}(0, t).$$

Oznacza to, że zmienna  $B_t$  powinna mieć rozkład normalny o średniej zero i wariancji  $t$ . Zauważmy dodatkowo, że dla  $s < r < t$  oraz  $n_s = [s/\Delta t]$ ,  $n_r = [r/\Delta t]$  i  $n_t = [t/\Delta t]$  zmienne

$$X_t - X_r = \Delta x \sum_{k=n_r+1}^{n_t} (2\epsilon_k - 1) \quad \text{oraz} \quad X_r - X_s = \Delta x \sum_{k=n_s+1}^{n_r} (2\epsilon_k - 1).$$

są niezależne. To z kolei prowadzi do niezależności  $B_t - B_s$  oraz  $B_s - B_r$ .

Po tym nieformalnym wstępie możemy wprowadzić ścisłą definicję ruchu Browna.

#### Definicja 4.1

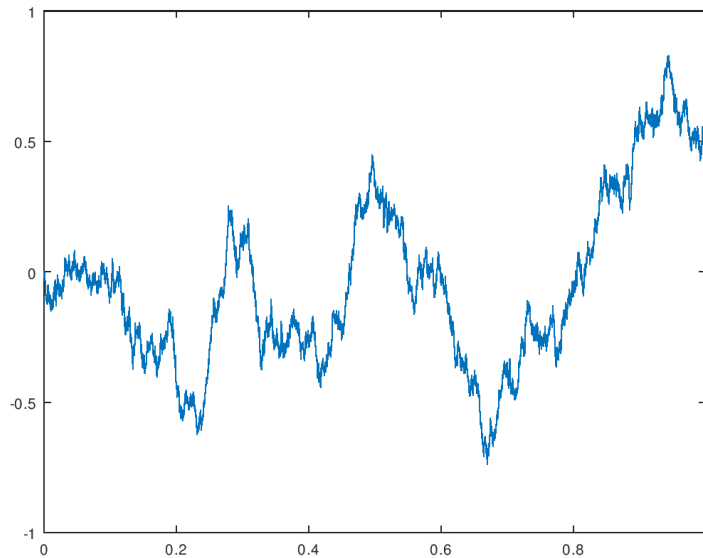
Procesem stochastycznym  $(X_t)_{t \geq 0}$  nazywamy dowolną rodzinę zmiennych losowych.

#### Definicja 4.2

Proces stochastyczny  $(B_t)_{t \geq 0}$  nazywamy i **ruchem Browna** (procesem Wienera) jeżeli

- (B1)  $B_0 = 0$ ;
- (B2) ma stacjonarne przyrosty: dla każdego  $s < t$ ,  $B_t - B_s$  ma rozkład  $\mathcal{N}(0, t - s)$ ;
- (B3) ma niezależne przyrosty: dla dowolnego  $n \in \mathbb{N}$  i dowolnych liczb  $0 \leq t_0 < t_1 < \dots < t_n$  zmienne losowe

$$B_{t_0}, B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$$



**Rysunek 4.2.** Trajektoria procesu  $(B_t)_{t \geq 0}$ , czyli wykres położenia cząsteczki  $B_t$  w zależności od czasu  $t$

są niezależne.  
**(B4)** ma ciągle trajektorie: dla każdej  $\omega$ , funkcja  $t \mapsto B_t(\omega)$  jest ciągła.

W zależności od czytelności będziemy stosowali zamiennie notacje  $B_t = B(t)$ . Przez proces stochastyczny w powyższej definicji rozumiemy rodzinę zmiennych losowych  $B_t$  indeksowaną  $t \in [0, +\infty)$ . W różnych kontekstach warto jest korzystać z innych interpretacji. Zauważmy, że dla każdej  $\omega \in \Omega$  kolekcja liczb  $(B_t(\omega))_{t \geq 0}$  to po prostu funkcja, która każdej liczbie  $t \in [0, +\infty)$  przyporządkowuje wartość  $B_t(\omega)$  (patrz punkty **(B4)** **Definicji 4.2**). Oznacza to, że o procesie stochastycznym  $(B_t)_{t \geq 0}$  możemy myśleć ja o losowej funkcji. Zauważmy wreszcie, że traktując  $t$  i  $\omega$  równocześnie otrzymujemy funkcję dwóch zmiennych  $(t, \omega) \mapsto B_t(\omega)$ .

**Twierdzenie 4.3** (Wiener 1926)

Istnieje przestrzeń probabilistyczna  $(\Omega, \mathcal{F}, \mathbb{P})$  oraz określone na niej zmienne losowe  $B_t$  dla  $t \geq 0$  takie, że  $(B_t)_{t \geq 0}$  jest ruchem Browna.

Znane są rozmaite konstrukcje ruchu Browna, np. w [JS] jest opisana elegancka konstrukcja Ciesielskiego (1961) bazująca na falkach (funkcja Hara). Poniżej przedstawimy konstrukcję pochodzącą od P. Lévy'ego (1948).



Rysunek 4.3. Konstrukcja Lévy'ego ruchu Browna

*Dowód.* Konstrukcję przeprowadzimy w kilku krokach według następującego planu:

1. zdefiniujemy  $B_t$  dla liczb diadycznych postaci  $k/2^n \in [0, 1]$ ;
2. korzystając z gęstości liczb diadycznych w  $[0, 1]$  rozszerzymy definicję do  $B_t$  dla  $t \in [0, 1]$ ;
3. rozszerzamy  $(B_t)$  z odcinka  $[0, 1]$  do  $[0, \infty)$ .

**Krok 1. Definicja  $W_t$  dla liczb diadycznych**

Niech

$$D_n = \left\{ \frac{k}{2^n} : k = 0, 1, \dots, 2^n \right\}$$

będzie zbiorem liczb wymiernych w  $[0, 1]$  o mianowniku  $2^n$ . Wówczas  $D_0 \subset D_1 \subset D_2 \subset \dots$  oraz  $D = \bigcup_n D_n$  jest zbiorem wszystkich liczb diadycznych w  $[0, 1]$ . Zbiór  $D$  jest przeliczalny i gęsty w  $[0, 1]$ . Możemy więc założyć, że zadana jest rodzina niezależnych zmiennych losowych  $\{Z_q\}_{q \in D}$  o rozkładzie  $N(0, 1)$ . Rodzinę taką możemy zrealizować na przestrzeni  $\Omega = \mathbb{R}^{\mathbb{N}}$ . Używając tej rodziny zmiennych losowych zdefiniujemy ruch Browna na liczbach diadycznych  $\{B_t\}_{t \in D}$ .

Podczas konstrukcji wielokrotnie będziemy odwoływać się do następującego lematu (pojawił się on jako jedno z zadań podczas kursu TP1), umożliwi on nam konstrukcję  $W_t$  dla  $t$  z kolejnych zbiorów  $D_0, D_1, \dots$

**Lemat 4.4**

Niech  $t \geq 0$ . Jeżeli  $X, Y$  są niezależnymi zmiennymi losowymi o rozkładzie  $N(0, t)$ , to zmienne losowe

$$\frac{X+Y}{\sqrt{2}}, \quad \frac{X-Y}{\sqrt{2}}$$

są niezależne i obie mają rozkład  $N(0, t)$ .

**Konstrukcja dla  $q \in D_1 = \{0, 1/2, 1\}$**  Zdefiniujemy  $B_0 = 0$ ,  $B_1 = Z_1$ , a następnie

$$B_{\frac{1}{2}} = \frac{B_1}{2} + \frac{Z_1}{2}.$$

Zauważmy, że zmienne losowe  $B_{\frac{1}{2}}$  i  $B_1 - B_{\frac{1}{2}}$  są niezależne, gdyż można je przedstawić w postaci

$$B_{\frac{1}{2}} = \frac{B_1 + Z_1}{2}, \quad B_1 - B_{\frac{1}{2}} = \frac{B_1 - Z_1}{2},$$

a następnie skorzystać z lematu dla zmiennych losowych  $B_1/\sqrt{2}$  oraz  $Z_1/\sqrt{2}$ . Wynika stąd również, że  $B_{\frac{1}{2}} \stackrel{d}{=} N(0, 1/2)$ .

**Konstrukcja dla  $q \in D_n$ .** Załóżmy, że proces został skonstruowany dla parametrów ze zbioru  $D_{n-1}$ . Przedstawimy teraz definicję dla  $q \in D_n \setminus D_{n-1}$ . Zauważmy, że wówczas  $q$  musi być postaci

$$q = \frac{2k+1}{2^n}.$$

Dodatkowo

$$q \pm 2^{-n} \in D_{n-1}.$$

kładziemy

$$B(q) = \frac{1}{2} (B(q + 2^{-n}) + B(q - 2^{-n})) + \frac{Z_q}{2^{(n+1)/2}}.$$

Wówczas należy pokazać, że zmienne losowe

$$\{B(q) - B(q - 2^{-n})\}_{q \in D_n} = \left\{ B_{\frac{j}{2^n}} - B_{\frac{j-1}{2^n}} \right\}_{j=1, \dots, 2^n}$$

są niezależne i każda z nich ma rozkład  $\mathcal{N}(0, 2^{-n})$ . Odwołując się do własności rozkładu normalnego wystarczy pokazać niezależność parametrów. Postępujemy indukcyjnie. Ustalmy  $q = \frac{2k+1}{2^n}$ . Wówczas

$$B_q - B_{\frac{2k}{2^n}} = \frac{\frac{1}{\sqrt{2}}(B_{\frac{k+1}{2^{n-1}}} - B_{\frac{k}{2^{n-1}}}) + \frac{Z_q}{2^{\frac{n}{2}}}}{\sqrt{2}}$$

$$B_{\frac{2k+2}{2^n}} - B_q = \frac{\frac{1}{\sqrt{2}}(B_{\frac{k+1}{2^{n-1}}} - B_{\frac{k}{2^{n-1}}}) - \frac{Z_q}{2^{\frac{n}{2}}}}{\sqrt{2}}.$$

Zmienne losowe  $\frac{1}{\sqrt{2}}(B_{\frac{k+1}{2^{n-1}}} - B_{\frac{k}{2^{n-1}}})$  oraz  $\frac{Z_q}{2^{\frac{n}{2}}}$  są niezależne i obie mają rozkład  $\mathcal{N}(0, 2^{-n})$ . Z lematu wynika więc, że również  $B_q - B_{\frac{2k}{2^n}}$  i  $B_{\frac{2k+2}{2^n}} - B_q$  są niezależne i obie mają rozkład  $\mathcal{N}(0, 2^{-n})$ .

Trzeba jeszcze pokazać, że dla dowolnych  $q, d \in D_n$ ,  $q > d$  zmienna losowa  $B(q) - B(d)$  ma rozkład  $\mathcal{N}(0, q - d)$  i jest niezależna od zmiennych  $\{B(w)\}_{w \in D_n, w \leq q}$ . Własność ta wynika z niezależności  $\{B(q) - B(q - 2^{-n})\}_{q \in D_n}$ .

**Krok 2. Rozszerzenie na  $[0, 1]$ .**

Definiujemy ciąg losowych funkcji  $F_n: [0, 1] \rightarrow \mathbb{R}$  wzorem

$$F_0(t) = \begin{cases} Z_1 & q = 1 \\ \text{liniowa poza tym} \end{cases}$$

oraz dla  $n \geq 1$ ,

$$F_n(t) = \begin{cases} 2^{-(n+1)/2} Z_q, & q \in D_n \setminus D_{n-1} \\ 0 & q \in D_{n-1} \\ \text{liniowa} & \text{poza tym} \end{cases}$$

Wówczas dla każdego  $d \in D$ ,

$$B(d) = \sum_{k=0}^{\infty} F_k(d). \quad (4.1.1)$$

Aby się o tym przekonać zauważmy, że jeżeli  $d \in D_n$ , to z definicji  $F_k(d) = 0$  dla wszystkich  $k > n$ . Wystarczy zatem pokazać indukcyjnie po  $n$ , że dla każdego  $q \in D_n$ ,

$$B(d) = \sum_{k=0}^n F_k(d).$$

Dalej, skoro dla  $j \leq n - 1$  funkcja  $F_j$  jest liniowa na przedziale  $[d - 2^{-n}, d + 2^{-n}]$ ,

$$F_j(d) = \frac{1}{2}(F_j(d - 2^{-n}) + F_j(d + 2^{-n})), \quad q \in D_n, j < n.$$

Stąd, z założenia indukcyjnego

$$\sum_{j=0}^{n-1} F_j(d) = \sum_{j=0}^{n-1} \frac{1}{2}(F_j(d - 2^{-n}) + F_j(d + 2^{-n})) = \frac{1}{2}(B(d + 2^{-n}) + B(d - 2^{-n}))$$

Skoro  $F_n(d) = 2^{-(n+1)/2}Z_d$ , to po odwołaniu się do definicji  $B(d)$  dla  $d \in D_n \setminus D_{n-1}$  otrzymujemy (4.1.1).

**Szereg jest zbieżny jednostajnie.** Dla  $t \in [0, 1]$  definiujemy

$$B(t) = \sum_{k=0}^{\infty} F_k(t).$$

Wiemy już, że dla  $t = d \in D$  powyższy suma powyższego szeregu jest zgodna z przeprowadzoną do tej pory konstrukcją. Pokażemy teraz, że szereg ten jest zbieżny jednostajnie. Wówczas, skoro ciąg sumy częściowe  $\sum_{k=0}^n F_k(t)$  są ciągłymi funkcjami zmiennej  $t$ , otrzymamy ciągłość  $B(t)$  (przypomnijmy, że jeżeli ciąg funkcji ciągłych jest zbieżny jednostajnie, to granica musi być funkcją ciągłą). Niech

$$\|F_n\|_{\infty} = \sup_{t \in [0,1]} |F_n(t)| = 2^{-(n+1)/2} \max_{d \in D_n \setminus D_{n-1}} |Z_d|.$$

Wówczas dla  $c > \sqrt{2 \log(2)}$ ,

$$\sum_{n=1}^{\infty} \mathbb{P}[\|F_n\| > c\sqrt{n}2^{-(n+1)/2}] \leq \sum_{n=1}^{\infty} 2^n \mathbb{P}[|Z_1| > c\sqrt{n}] \leq \sum_{n=1}^{\infty} 2^n \exp\{c^2 n/2\}.$$

Z lematu Borela-Cantellego z prawdopodobieństwem jeden

$$\|F_n\|_{\infty} \leq c\sqrt{n}2^{-(n+1)/2}$$

dla dostatecznie dużych  $n$ . Stąd

$$\left\| \sum_{k=n}^{\infty} F_k(t) \right\|_{\infty} \leq \sum_{k=n}^{\infty} \|F_k(t)\|_{\infty} \leq \sum_{k=n}^{\infty} c\sqrt{k}2^{-(k+1)/2} \leq C2^{-n}$$

dla pewnej uniwersalnej stałej  $C > 0$ .

**Własności probabilistyczne na  $[0, 1]$ .**

Należy jeszcze pokazać, że dla  $t, s \in [0, 1]$ ,  $t > s$  zmienna  $B(t) - B(s)$  ma rozkład normalny  $\mathcal{N}(0, t - s)$ . Zauważmy, że z konstrukcji powyższa własność zachodzi jeżeli  $t, s \in D$ . W tym celu korzystamy z gęstości  $D$  w  $[0, 1]$ , ciągłości  $B$ .

**Krok 4. Rozszerzenie na  $[0, +\infty)$**  Niech  $(B_k(t))_{t \in [0,1]}$  będą niezależnymi ruchami Browna na  $[0, 1]$ . Ruch Browna na  $[0, +\infty)$  definiujemy przez

$$B(t) = B_{[t]}(t - [t]) + \sum_{k=1}^{[t]-1} B_k(1).$$

□

## 4.2 Własności probabilistyczne ruchu Browna

Dla  $s > 0$  definiujemy ruch Browna po czasie  $s$  wzorem

$$B_t^{(s)} = B_{t+s} - B_s, \quad t \geq 0.$$

### Twierdzenie 4.5

Dla każdego  $s > 0$  proces stochastyczny  $(B_t^{(s)})_{t \geq 0}$  jest ruchem Browna.

*Dowód.* Należy sprawdzić, że rodzina zmiennych losowych  $(B_t^{(s)})_{t \geq 0}$  spełnia warunki **Definicji 4.2**. Pozostawiamy to jako zadanie dla czytelnika.  $\square$

Zastanowimy się teraz jak proces  $(B_t^{(s)})_{t \geq 0}$  ma się do ruchu Browna do chwili  $s$ , czyli  $(B_t)_{0 \leq t \leq s}$ . W tym celu posłużymy się filtracją w czasie ciągłym. Dla każdego  $t > 0$  rozważmy

$$\mathcal{F}_t^B = \sigma(B_r : r \leq t).$$

Zauważmy, że z ciągłości ruchu Browna wartości  $(B_t)_{t \geq 0}$  są jednoznacznie wyznaczone przez wartości  $(B_t)_{0 \leq t \in \mathbb{Q}}$ . Stąd

$$\mathcal{F}_t^B = \sigma(B_r : r \leq t, r \in \mathbb{Q}).$$

### Fakt 4.6

Dla dowolnych  $t > s > 0$  zmienna losowa  $B_t - B_s$  jest niezależna od  $\sigma$ -ciała  $\mathcal{F}_s^B$ .

*Dowód.* Z niezależności przyrostów ruchu Browna otrzymujemy, że dla dowolnego  $n \in \mathbb{N}$  i dowolnych  $t_1 < t_2 < \dots < t_n \leq s$  zmienne losowe

$$B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}, B_s - B_{t_n}, B_t - B_s$$

są niezależne. Stąd wektor losowy

$$(B_{t_1}, B_{t_2}, \dots, B_{t_n})$$

jest niezależny od  $B_t - B_s$ . Oznacza to, że ostatnie zmienna jest niezależna od  $\pi$ -układu

$$\mathcal{C} = \{ \{ B_{t_1} \in A_1, B_{t_2} \in A_2, \dots, B_{t_n} \in A_n \} : n \in \mathbb{N}, t_k \leq s, A_k \in \mathcal{B}or(\mathbb{R}) \}.$$

Dokładniej dla każdego  $C \in \mathcal{C}$  i  $A \in \mathcal{B}or(\mathbb{R})$ ,

$$\mathbb{P}[C, B_t - B_s \in A] = \mathbb{P}[C] \mathbb{P}[B_t - B_s \in A].$$

Rozważając zatem  $\lambda$ -układ

$$\mathcal{A} = \{ C \in \mathcal{F} : \forall A \mathbb{P}[C, B_t - B_s \in A] = \mathbb{P}[C] \mathbb{P}[B_t - B_s \in A] \}$$

i stosując lemat o  $\pi$  i  $\lambda$  układach otrzymujemy  $\mathcal{F}_s^B = \sigma(\mathcal{C}) \subseteq \mathcal{A}$  to z kolei pociąga tezę.  $\square$

**Definicja 4.7**

Powiemy, że proces stochastyczny  $(X_t)_{t \geq 0}$  jest niezależny od  $\sigma$ -ciała  $\mathcal{G}$ , jeżeli  $\sigma$ -ciała  $\sigma(X_t : t \geq 0)$  oraz  $\mathcal{G}$  są niezależne.

**Fakt 4.8**

Dla dowolnego  $s > 0$  proces stochastyczny  $B_t^{(s)} = B_{t+s} - B_s$  jest niezależny od  $\sigma$ -ciała  $\mathcal{F}_s^B$ .

*Dowód.* Rozumowanie jest analogiczne do dowodu **Faktu 4.6**. Szczegóły są pozostawione jako zadanie.  $\square$

Niech  $t > s$ . Wykorzystując zgromadzone do tej pory informacje możemy rozważyć warunkową wartość oczekiwaną

$$\mathbb{E} \left[ B_t \mid \mathcal{F}_s^B \right] = \mathbb{E} \left[ B_t - B_s \mid \mathcal{F}_s^B \right] + \mathbb{E} \left[ B_s \mid \mathcal{F}_s^B \right] = B_s.$$

Zatem najlepszą aproksymacją  $B_t$  zrobioną na podstawie historii procesu do chwili  $s$  jest  $B_s$ . Procesy stochastyczne, w analogii do czasów dyskretnych, nazywane są martyngałami w czasie ciągłym.

**Definicja 4.9**

Filtracją w czasie ciągłym  $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$  nazywamy rodzinę  $\sigma$ -ciał taką, że dla  $s < t$ ,  $\mathcal{F}_s \subseteq \mathcal{F}_t$ .

Zdefiniowana przez nas rodzina  $\sigma$ -ciał  $\mathbb{F}^B = (\mathcal{F}_t^B)_{t \geq 0}$  tworzy filtrację.

**Definicja 4.10**

Niech  $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$  będzie filtracją. Proces stochastyczny  $(M_t)_{t \geq 0}$  jest  $\mathbb{F}$ -martyngąłem jeżeli

- jest adoptowalny do tej filtracji: dla każdego  $t \geq 0$  zmienna losowa  $M_t$  jest  $\mathcal{F}_t$  mierzalna
- $\mathbb{E}|M_t| < \infty$ , dla każdego  $t \geq 0$ ;
- dla każdego  $s \leq t$  zachodzi

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s.$$

**Przykład 4.11**

Jak już wyliczyliśmy wcześniej,  $\mathbb{E}[B_t \mid \mathcal{F}_s^B] = B_s$ , więc ruch Browna jest martyngąłem.

**Przykład 4.12**

Jeżeli  $(B_t)_{t \geq 0}$  jest ruchem Browna, to proces  $M_t = B_t^2 - t$  jest martyngałem. Rzeczywiście, wykorzystując niezależność przyrostów

$$\mathbb{E} \left[ B_t^2 \mid \mathcal{F}_s^B \right] = B_s^2 + t - s.$$

**Definicja 4.13**

Niech  $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$  będzie filtracją. Mówimy, że zmienna losowa  $T : \Omega \rightarrow [0, \infty]$  jest  **$\mathbb{F}$ -czasem zatrzymania** jeżeli dla każdego  $t \geq 0$ ,

$$\{T \leq t\} \in \mathcal{F}_t.$$

**Przykład 4.14**

Ustalmy  $y \in \mathbb{R}$  i niech

$$T_y = \inf\{t \geq 0 : B_t = y\}.$$

Wówczas  $T_y$  jest czasem zatrzymania względem filtracji  $\mathcal{F}_t^B$ , gdyż

$$\{T_y \leq t\} = F = \bigcap_{n=1}^{\infty} \bigcup_{s \in \mathbb{Q} \cap (0, t)} \{|B_s - y| \leq 1/n\}.$$

Zauważmy najpierw, że  $F \in \mathcal{F}_t^B$  jako wynik przeliczalnych operacji mnogościowych na zdarzeniach z  $\mathcal{F}_t^B$ . Aby uzasadnić, że powyższa równość zachodzi pokażemy dwa zawierania. Niech  $\omega \in \{T_y \leq t\}$ . Istnieje malejący ciąg  $s_n$  zbieżny do

**Twierdzenie 4.15**

Niech  $\{X_t\}$  będzie  $\mathbb{F}$ -martyngałem oraz niech  $\tau$  będzie czasem zatrzymania takim, że dla pewnego  $t_0$ ,  $\tau(\omega) < t_0$  dla wszystkich  $\omega \in \Omega$ . Wówczas

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

Dowód powyższego twierdzenia przedstawimy na końcu tego rozdziału. Skupimy się teraz na przykładach.

**Przykład 4.16**

Niech  $a > b > 0$ . Rozważmy

$$T = \inf\{s > 0 : B_s \in \{-a, b\}\}.$$

Wówczas  $T = T_b \wedge T_{-a}$  jest czasem zatrzymania. Sprawdźmy najpierw, że  $T$  jest skończone prawie wszędzie. Dla  $t > 0$  mamy

$$\mathbb{P}[T > t] \leq \mathbb{P}[B_t \in (-a, b)] = \mathbb{P}[B_1 \in (-a/\sqrt{t}, b/\sqrt{t})].$$

Po przejściu  $t \rightarrow \infty$  widzimy, że

$$\mathbb{P}[T = \infty] = \lim_{t \rightarrow \infty} \mathbb{P}[T > t] = \mathbb{P}[B_1 = 0] = 0.$$

Rozważmy martyngał  $X_t = (a + B_t)(b - B_t) + t$ . Stosując twierdzenie o zatrzymaniu do czasu zatrzymania  $\tau = T \wedge t_0$  otrzymujemy  $\mathbb{E}[X_\tau] = ab$ , co z kolei pociąga

$$\mathbb{E}(a + B_{T \wedge t_0})(b - B_{T \wedge t_0}) + \mathbb{E}[T \wedge t_0] = ab.$$

Przechodząc teraz z  $t_0$  do nieskończoności i powołując się na twierdzenia o zbieżności ograniczonej i monotonicznej

$$\mathbb{E}(a + B_T)(b - B_T) + \mathbb{E}[T] = ab.$$

Zauważmy, że z ciągłości trajektorii ruchu Browna  $B_T \in \{-a, b\}$  i stąd  $(a + B_T)(b - B_T) = 0$ . Ostatecznie

$$\mathbb{E}[T] = ab.$$

## 4.3 Trajektorie ruchu Browna

### Definicja 4.17

Trajektoria procesu stochastycznego  $(X_t)_{t \geq 0}$  nazywamy losową funkcję  $t \mapsto X_t$ .

Skupimy się teraz na opisie własności trajektorii ruchu Browna. Z konstrukcji i definicji wiemy, że są one ciągłe. Okazuje się, że można powiedzieć nieco więcej. Zaczniemy jednak od  $\sigma$ -ciała

$$\mathcal{F}_{0+}^B = \bigcap_{s>0} \mathcal{F}_s^B.$$

$\sigma$ -ciało  $\mathcal{F}_{0+}^B$  składa się ze zdarzeń, które są determinowane przez trajektorie ruchu Browna na przedziałach  $[0, h]$  dla dowolnie małego  $h$ .

### Przykład 4.18

Rozważmy zdarzenie

$$A = \{(B)_{t \geq 0} \text{ ma w } 0 \text{ ekstremum lokalne}\}$$

Wówczas dla dowolnego  $h > 0$ ,

$$A = \{(B)_{t \in [0, h]} \text{ ma w } 0 \text{ ekstremum lokalne}\} \in \mathcal{F}_h^B.$$

Skoro  $A \in \mathcal{F}_h^B$  dla dowolnego  $h$ , to należy też do przekroju

$$A \in \bigcap_{h > 0} \mathcal{F}_h^B = \mathcal{F}_{0+}^B.$$

**Fakt 4.19**

Ruch Browna  $(B_t)_{t \geq 0}$  jest niezależny do  $\mathcal{F}_{s+}^B$ .

*Dowód.* Dla ustalonego ciągu  $\epsilon_n \downarrow 0$ , z ciągłości trajektorii  $B$ ,

$$B(t) = \lim_{n \rightarrow \infty} B(t + \epsilon_n) - B(\epsilon_n).$$

Dla każdego ustalonego  $n \in \mathbb{N}$  proces  $B_t^{(\epsilon)} = B(t + \epsilon_n) - B(\epsilon_n)$  jest niezależny od  $\sigma$ -ciała  $\mathcal{F}_{\epsilon_n} \supseteq \mathcal{F}_{0+}$ . Skoro wszystkie zmienne procesy  $B(t + \epsilon_n) - B(\epsilon_n)$  są niezależne od  $\mathcal{F}_{0+}$ , to to samo tyczy się granicy  $B(t + s) - B(s)$ .  $\square$

**Twierdzenie 4.20** (Prawo 0-1 Blumenthala)

Dla każdego  $A \in \mathcal{F}_{0+}^B$ ,  $\mathbb{P}[A] \in \{0, 1\}$ .

*Dowód.* Zauważmy, że  $\mathcal{F}_{0+}^B \subseteq \sigma(B_t : t \geq 0)$ . Dodatkowo, z **Faktu 4.19** dla  $s = 0$  każde zdarzenie  $A \in \sigma(B_t : t \geq 0)$  jest niezależne od  $\mathcal{F}_{0+}^B$ . Oznacza to, że  $A \in \mathcal{F}_{0+}^B$  jest niezależne od  $\mathcal{F}_{0+}^B$ , więc w szczególności  $A$  jest niezależne od samego siebie. Stąd  $\mathbb{P}[A] \in \{0, 1\}$ .  $\square$

**Przykład 4.21**

Niech  $\tau = \inf\{t > 0 : B_t > 0\}$ . Pokażemy, że  $\mathbb{P}[\tau = 0] = 1$ . Mamy

$$\{\tau = 0\} = \bigcap_{n=1}^{\infty} \left\{ \sup_{0 < s < 1/n} B_s > 0 \right\} \in \mathcal{F}_{0+}.$$

Wystarczy zatem pokazać, że  $\mathbb{P}[\tau = 0] > 0$ . Zauważmy, że

$$\mathbb{P}[\tau \leq t] = \mathbb{P} \left[ \sup_{s \in [0, t]} B_s > 0 \right] \geq \mathbb{P}[B_t > 0] = \frac{1}{2}.$$

Przejście  $t \rightarrow 0$  pociąga  $\mathbb{P}[\tau = 0] \geq \frac{1}{2}$ . Z 0-1 Blumenthala wnioskujemy, że  $\mathbb{P}[\tau = 0] = 1$ . W szczególności pokazaliśmy, że dla każdego  $t > 0$ ,

$$\sup_{s \in [0, t]} B_s > 0 \quad \text{p.w.}$$

**Twierdzenie 4.22** (Mocna własność Markowa dla ruchu Browna)

Niech  $\tau$  będzie skończonym czasem zatrzymania. Wówczas proces  $(B_{t+\tau} - B_\tau : t \geq 0)$  jest ruchem Browna niezależnym do procesu  $(B_{\tau \wedge t})_{t \geq 0}$ .

*Dowód.* Rozważmy ciąg czasów zatrzymania  $\tau_n$  przybliżających  $\tau$  od góry

$$\tau_n = (m+1)2^{-n} \quad \text{na zbiorze } \tau \in [m2^{-n}, (m+1)2^{-n}).$$

Ustalmy  $n \in \mathbb{N}$ . Niech  $B_k$  będzie ruchem Browna zadany przez  $B_k(t) = B(t + k2^{-n}) - B(k2^{-n})$ . Niech  $X$  będzie procesem danym przez  $X(t) = B(t + \tau_n) - B(\tau_n)$ . Dla zbioru  $A \in \mathcal{B}(\mathbb{R}^{\mathbb{R}^+})$ , tj zbioru  $A$  postaci

$$A = \{x \in \mathbb{R}^{\mathbb{R}^+} : x_{t_k} \in A_k, k = 1, 2, \dots, n\}, \quad t_k \in \mathbb{R}_+, A_k \in \mathcal{B}(\mathbb{R}),$$

oraz zbioru  $E \in \mathcal{F}_{\tau_n}^B$ , napiszmy

$$\begin{aligned} \mathbb{P}[\{X \in A\} \cap E] &= \sum_{k=0}^{\infty} \mathbb{P}[\{X \in A\} \cap E \cap \{\tau_n = k2^{-n}\}] \\ &= \sum_{k=0}^{\infty} \mathbb{P}[\{B_k \in A\} \cap E \cap \{\tau_n = k2^{-n}\}]. \end{aligned}$$

Zauważmy, że  $E \cap \{\tau_n = k2^{-n}\} \in \mathcal{F}_{k2^{-n}}^B$ . Wobec tego z własności Markowa dla ruchu Browna

$$\mathbb{P}[\{X \in A\} \cap E] = \sum_{k=0}^{\infty} \mathbb{P}[\{B_k \in A\}] \mathbb{P}[E \cap \{\tau_n = k2^{-n}\}] = \mathbb{P}[B \in A] \mathbb{P}[E].$$

Pokazaliśmy właśnie, że  $(B(t + \tau_n) - B(\tau_n))_{t \in [0, \infty)}$  jest ruchem Browna niezależnym od  $\mathcal{F}_{\tau_n}^B$ . Skoro  $\mathcal{F}_{\tau_n}^B \supseteq \mathcal{F}_\tau^B$  to dla każdego  $n$ ,  $(B(t + \tau_n) - B(\tau_n))_{t \in [0, \infty)}$  jest niezależny od  $\mathcal{F}_\tau^B$ . Wobec tego granica

$$\lim_{n \rightarrow \infty} B(t + \tau_n) - B(\tau_n) = B(t + \tau) - B(\tau)$$

również jest niezależna od  $\mathcal{F}_\tau^B$ . Dla każdego  $n$ ,  $(B(t + \tau_n) - B(\tau_n))_{t \in [0, \infty)}$  jest ruchem Browna, więc jego rozkłady skończone wymiarowe spełniają warunki (B1)-(B3) **Definicji 4.2**. Z powyższej granicy rozkłady skończone wymiarowe  $(B(t + \tau) - B(\tau))_{t \in [0, \infty)}$  są takie same jak rozkłady  $(B(t + \tau_n) - B(\tau_n))_{t \in [0, \infty)}$ . W szczególności  $(B(t + \tau) - B(\tau))_{t \in [0, \infty)}$  spełnia warunki (B1)-(B3) **Definicji 4.2**. Ciągłość  $t \mapsto B(t + \tau) - B(\tau)$  jest z kolei oczywista. Pokazuje to, że  $(B(t + \tau) - B(\tau))_{t \in [0, \infty)}$  jest ruchem Browna.  $\square$

### Przykład 4.23

Aby zilustrować **Twierdzenie 4.22** uzasadnimy zasadę odbicia dla ruchu Browna. Dla  $a > 0$  niech

$$\tau_a = \inf\{t \geq 0 : B_t \geq a\}.$$

Rozważmy proces  $X = (X_t)_{t \in \mathbb{R}_+}$  będący ruchem Browna odbitym w  $a$ . Wówczas  $X$  jest zadany przez

$$X_t = \begin{cases} B_t & t \leq \tau_a \\ 2a - B_t & t > \tau_a \end{cases} \quad (4.3.1)$$

Korzystając z mocnej własności Markowa dla  $B$ , procesy

$$(B(t + \tau_a) - B(\tau_a))_{t \in \mathbb{R}_+} \quad \text{oraz} \quad (-B(t + \tau_a) + B(\tau_a))_{t \in \mathbb{R}_+} \quad (4.3.2)$$

są ruchami Browna niezależnymi od  $(B_{\tau_a \wedge t})_{t \in \mathbb{R}_+}$ . Wówczas sklejenie ostatniego procesu w punkcie  $\tau_a$  z każdym z procesów (4.3.2) da proces o takim samym rozkładzie. Sklejenie z pierwszym daje  $(B_t)_{t \in \mathbb{R}_+}$  a z drugim  $X$ .

### Przykład 4.24

Rozważmy  $M_t = \sup_{s \in [0, t]} B_s$ . Dla  $a > 0$  mamy

$$\{M_t > a\} = \{B_t > a\} \cup \{B_t \leq a, M(t) > a\}.$$

Zauważmy, że

$$\{B_t \leq a, M_t > a\} = \{X_t > a\},$$

gdzie  $X$  jest procesem zadany przez (4.3.1). Stąd

$$\mathbb{P} \left[ \sup_{s \in [0, t]} B_s > a \right] = \mathbb{P}[M_t > a] = \mathbb{P}[B_t > a] + \mathbb{P}[X_t > a] = 2\mathbb{P}[B_t > a].$$

### Nieróżniczkowalność ruchu Browna

Trajektorie ruchu Browna są ciągłe, ale bardzo nieregularne. Poniżej opiszemy jego lokalne zachowanie. Zaczniemy od ogólnej obserwacji. Zgodnie z definicją

$$W_{t+\Delta t} - W_t \stackrel{d}{=} W_{\Delta t}$$

(interesują nas bardzo małe wartości  $\Delta t$ ) ma rozkład  $N(0, \Delta t)$ , a więc, oznaczając przez  $Y$  zmienną o rozkładzie  $N(0, 1)$ , otrzymujemy

$$W_{t+\Delta t} - W_t = \sqrt{\Delta t} \cdot Y \approx \sqrt{\Delta t}$$

gdy  $\Delta t \rightarrow 0$ . Powyższa formuła pokazuje ciągłość, ale nie pociąga różniczkowalności, bo

$$\frac{W_{t+\Delta t} - W_t}{\Delta t} \approx \frac{1}{\sqrt{\Delta t}} \rightarrow \infty, \quad \Delta t \rightarrow 0.$$

Powyższy argument jest nieformalny. Zachodzi jednak

**Twierdzenie 4.25**

Z prawdopodobieństwem 1, dla każdego  $s \in [0, 1]$  funkcja  $t \mapsto W_t$  nie jest różniczkowalna w punkcie  $s$ .

**Uwaga 4.3.3.** Powyższe twierdzenie jest mocniejsze niż warunek: dla każdego  $t$ , z prawdopodobieństwem 1, pochodna  $W_t$  w punkcie  $t$  nie istnieje.

*Dowód.* Jeżeli funkcja  $f$  jest różniczkowalna w  $t_0$ , to istnieje  $\delta > 0$  taka, że dla każdego  $s$  spełniającego  $|s - t_0| < \delta$  zachodzi

$$|f(s) - f(t_0)| \leq 2|f'(t_0)||s - t_0|.$$

Gdyby  $W_t$  było różniczkowalne w pewnym punkcie, to moglibyśmy znaleźć duże  $M$  (zależne od trajektorii  $W_t$  oraz wartości pochodnej) takie, że dla każdego  $n > N$  istnieje  $k \leq n$  tż.

$$Y_{k,n} \leq \frac{M}{n}$$

dla

$$Y_{k,n} = \max \left\{ \left| W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right|, \left| W_{\frac{k+2}{n}} - W_{\frac{k+1}{n}} \right|, \left| W_{\frac{k+3}{n}} - W_{\frac{k+2}{n}} \right| \right\}.$$

Wtedy również dla każdego  $n > N$

$$Y_n \leq \frac{M}{n} \quad \text{dla} \quad Y_n = \min_k Y_{k,n}.$$

Niech

$$A_M = \{Y_n \leq M/n \text{ dla odpowiednio dużych } n\} = \{\omega : \exists N \forall n > N Y_n(\omega) \leq M/n\}.$$

Napiszmy również

$$A_M = \bigcup_N A_{M,N}, \quad \text{dla} \quad A_{M,N} = \{Y_n \leq M/n \forall n \geq N\}$$

co pociąga

$$A_M \subset \{Y_n \leq M/n\} \quad \text{dla odpowiednio dużych } n \quad (4.3.4)$$

Z powyższych obserwacji wynika, że

$$\{W_t \text{ jest różniczkowalne dla pewnego } t\} \subset \bigcup_{M=1}^{\infty} A_M$$

Dla dowolnego  $n$  mamy

$$\begin{aligned}\mathbb{P}[Y_{k,n} \leq M/n] &= \left(\mathbb{P}[|W_{\frac{1}{n}}| \leq M/n]\right)^3 = \mathbb{P}[|W_1|/\sqrt{n} \leq M/n]^3 \\ &= \left[\int_{-\frac{M}{\sqrt{n}}}^{\frac{M}{\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy\right]^3 \leq \left(\frac{2M}{\sqrt{n}} \cdot \frac{1}{\sqrt{2\pi}}\right)^3 \leq \frac{CM^3}{n^{3/2}}.\end{aligned}$$

Stąd

$$\mathbb{P}[Y_n \leq M/n] \leq \sum_{k=0}^{n-1} \mathbb{P}[Y_{k,n} \leq M/n] \leq \frac{M^3}{\sqrt{n}} \rightarrow 0$$

i z (4.3.4) dla każdego  $M$ ,  $\mathbb{P}[A_M] = 0$ . Implikuje to

$$\mathbb{P}\left[\bigcup_{M=1}^{\infty} A_M\right] = 0.$$

Ostatecznie

$$\mathbb{P}[W_t \text{ jest różniczkowalne dla pewnego } t] = 0.$$

□

## Procesy Levy'ego

**Streszczenie** W tym rozdziale opiszemy szeroką klasę procesów stochastycznych o niezależnych przyrostach. Zanim jednak przedstawimy ogólną definicję zbadamy dokładniej kilka konkretnych przykładów.

Skonstruowany przez nas w poprzednim rozdziale ruch Browna  $(B_t)_{t \geq 0}$  jest procesem stochastycznym o niezależnych przyrostach. Dokładniej, dla dowolnych  $0 < s < t$  przyrost ruchu Browna  $B_t - B_s$  jest niezależny od historii procesu do chwili  $s$ . Okazuje się, że to dokładnie ta własność spotykana jest w wielu zastosowaniach i modelowaniu. W tym rozdziale przyjrzemy się dokładniej procesom o tej własności. Najpierw jednak zbadamy dokładniej kilka konkretnych przykładów.

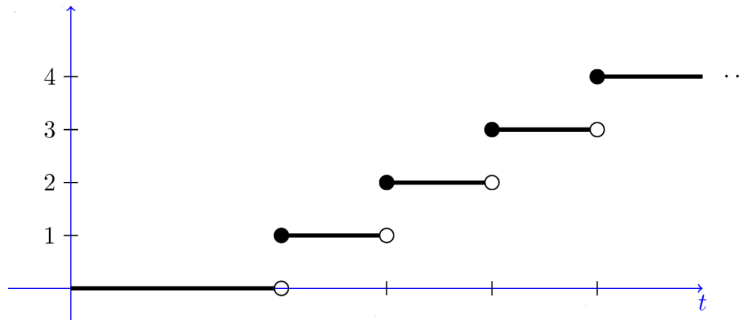
### 5.1 Proces Poissona

Przykładem procesu o niezależnych przyrostach jest proces liczący  $\{N_t\}_{t \geq 0}$ , który zlicza zdarzenia pojawiające się w czasie  $[0, t)$ . Ma on cały szereg praktycznych zastosowań (np. przychodzące rozmowy telefoniczne, czasy przyjścia klientów, czasy uszkodzenia maszyn, momenty pojawiania się błędów w tekście, rozkład materiałów radioaktywnych). Zanim podamy formalną definicję pokażemy, że zmienne losowe pojawiające się w tym procesie powinny mieć rozkład Poissona. Zakładać będziemy, że dla  $t \geq 0$ , zmienna  $N_t$  liczy liczbę zgłoszeń do chwili  $t$ . Oznaczmy przez  $N(s, t] = N_t - N_s$  liczbę zdarzeń w czasie  $(s, t]$ . Naszym punktem wyjścia będą cztery postulaty:

1. liczby zdarzeń w rozłącznych przedziałach są niezależne: dla  $n \in \mathbb{N}$ ,  $0 \leq t_0 < t_1 < \dots < t_n$  zmienne

$$N(t_0, t_1], N(t_1, t_2], \dots, N(t_{n-1}, t_n]$$

są niezależne.



Rysunek 5.1. Przykładowa trajektoria procesu  $N_t$

2. Dla każdych  $s < t$ ,

$$N(s, t] \stackrel{d}{=} N(0, t - s].$$

3. prawdopodobieństwo, że w małym przedziale zajdzie jedno zdarzenie jest liniowe, tzn. dla każdego ustalonego  $t > 0$  oraz dla małych wartości  $h$

$$\mathbb{P}[N(t, t + h] = 1] = \lambda h + o(h),$$

gdzie  $\lambda$  jest parametrem intensywności, a symbol  $o(h)$  oznacza  $o(h)/h \rightarrow 0$ , gdy  $h \rightarrow 0$ ;

4. prawdopodobieństwo, że w małym przedziale zajdą co najmniej dwa zdarzenia jest zaniedbywane, tzn. dla małych wartości  $h$

$$\mathbb{P}[N(t, t + h] \geq 2] = o(h).$$

Okazuje się, że wychodząc od powyższych postulatów można wywnioskować szereg dalszych własności. Dla  $n > 0$  oraz  $t > 0$  rozważmy

$$N_t = \sum_{j=0}^{n-1} N(tj/n, t(j+1)/n).$$

Dla dużego  $n$ , z dużym prawdopodobieństwem zmienne pod sumą przyjmują wartości 0 lub 1. Rzeczywiście, dla każdego  $j$  mamy

$$\begin{aligned} \mathbb{P} \left[ N(tj/n, t(j+1)/n] \neq \mathbb{1}_{\{N(tj/n, t(j+1)/n]=1\}} \right] = \\ \mathbb{P} [N(tj/n, t(j+1)/n] \geq 2] = r_n = o(1/n). \end{aligned}$$

Rozważmy więc

$$N_t^{(n)} = \sum_{j=0}^{n-1} \mathbb{1}_{\{N(tj/n, t(j+1)/n]=1\}}.$$

Wówczas  $\mathbb{P}[N_t \neq N_t^{(n)}] \leq nr_n$ . Teraz łatwo możemy przybliżyć rozkład  $N_t$  przez  $N_t^{(n)}$ , ponieważ

$$\left| \mathbb{E} \left[ e^{isN_t} \right] - \mathbb{E} \left[ e^{isN_t^{(n)}} \right] \right| \leq 2\mathbb{P} \left[ N_t \neq N_t^{(n)} \right].$$

Oznaczmy  $p_n = \mathbb{P}[N(0, t/n] = 1]$ . Widzimy, że  $N_t^{(n)}$  ma rozkład dwumianowy z parametrami  $n, p_n$ . Skoro  $np_n \rightarrow \lambda t$ , to  $N_t^{(n)}$  zbiega według słabo do rozkładu Poissona z parametrem  $\lambda t$ . Rzeczywiście, mamy

$$\mathbb{E} \left[ e^{isN_t^{(n)}} \right] = \left( 1 + (e^{is} - 1)p_n \right)^n \rightarrow e^{\lambda t(e^{is} - 1)}.$$

Oznacza to, że

$$\mathbb{E} \left[ e^{iN_t} \right] = e^{\lambda(e^{is} - 1)}.$$

Widzimy stąd, że zmienna  $N_t$  ma rozkład Poissona z parametrem  $\lambda t$ . Prowadzi nas to do następującej definicji formalnej procesu Poissona.

### Definicja 5.1

Proces stochastyczny  $N = (N_t)_{t \geq 0}$  nazywamy jednorodnym procesem Poissona z parametrem  $\lambda > 0$ , jeżeli

- (N1)  $N_0 = 0$
- (N2)  $N$  ma niezależne przyrosty
- (N3) Dla każdego  $s < t$ ,  $N_t - N_s$  ma rozkład Poissona z parametrem  $\lambda(t - s)$
- (N4) Trajektorie  $N$  są prawostronnie ciągłe.

Naszym pierwszym celem będzie podanie jawnej konstrukcji procesu Poissona. W tym celu zastanówmy się jak powyższe postulaty wpływają na czas pierwszego zgłoszenia, które w terminach zmiennych  $(N_t)_{t \geq 0}$  zadaje się wzorem

$$T_1 = \inf\{t \geq 0 : N_t = 1\}.$$

Zauważmy, że dla każdego  $t > 0$  mamy

$$\mathbb{P}[T_1 > t] = \mathbb{P}[N_t = 0] = e^{-\lambda t}.$$

Czas zatrzymania  $T_1$  ma więc rozkład wykładniczy z parametrem  $\lambda$  (przypomnijmy, że wówczas gęstość tego rozkładu jest zadana wzorem  $\lambda e^{-\lambda t}$ ). Powyższa obserwacja nasuwa pewną jawną konstrukcję procesu Poissona.

**Twierdzenie 5.2**

Niech  $E_1, E_2, \dots$  będą niezależnymi zmiennymi losowymi z rozkładem wykładniczym o parametrze  $\lambda > 0$ . Dla  $n \geq 1$  połóżmy  $S_n = \sum_{k=1}^n E_k$ . Wówczas rodzina zmiennych losowych  $N = (N_t)_{t \geq 0}$  dana przez

$$N_t = \# \{n \geq 1 : S_n \leq t\} \quad (5.1.1)$$

jest jednorodnym procesem Poissona z parametrem  $\lambda$ .

*Dowód.* W pierwszym kroku sprawdzamy indukcyjnie, że dla każdego  $n \in \mathbb{N}$  zmienna  $S_n$  ma rozkład o gęstości

$$f_{S_n}(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \mathbb{1}_{\{x > 0\}}.$$

Następnie sprawdzamy, że dla  $u, t \geq 0$  mamy

$$\begin{aligned} \mathbb{P}[S_n \leq t, S_{n+1} > u] &= \int_0^t \mathbb{P}[E_{n+1} > u - s] f_{S_n}(s) ds \\ &= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} e^{-\lambda(u-s)} ds = e^{-\lambda u} \frac{(\lambda t)^n}{n!}. \end{aligned} \quad (5.1.2)$$

Sprawdzimy teraz, że dla każdego  $t \geq 0$  zmienna  $N_t$  ma rozkład Poissona z parametrem  $\lambda t$ . Mamy

$$\mathbb{P}[N_t = n] = \mathbb{P}[S_n > t] = e^{-\lambda t}$$

oraz dla  $n \geq 1$ , podstawiając  $u = t$  w (5.1.2)

$$\mathbb{P}[N_t = n] = \mathbb{P}[S_n \leq t < S_{n+1}] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Stąd  $N_t$  ma istotnie rozkład Poissona z parametrem  $\lambda t$ . Pokażemy teraz, że  $N$  ma niezależne przyrosty. Dla każdego  $n \in \mathbb{N}$  i  $u \geq t$ , korzystając z (5.1.2) mamy

$$\mathbb{P}[S_{n+1} \geq u \mid N_t = n] = \mathbb{P}[S_{n+1} \geq u, S_n \leq t] / \mathbb{P}[N_t = n] = e^{-\lambda(u-t)}.$$

Biorąc  $u = t + s$  wnioskujemy, że

$$\mathbb{P}[S_{n+1} - t \geq s \mid N_t = n] = e^{-\lambda s}.$$

Oznaczmy  $E'_1 = S_{N(t)+1} - t$ ,  $E'_k = E_{N(t)+k}$  dla  $k \geq 2$ . Zauważmy, że dla  $m \in \mathbb{N}$  i dowolnych  $s, t, v_2, \dots, v_m > 0$  mamy

$$\begin{aligned}
\mathbb{P}[N(t) = n, E'_1 > s, E'_2 \geq v_2, \dots, E'_m \geq v_m] &= \\
&\mathbb{P}[S_n \leq t, S_{n+1} > t + s, E_{n+2} \geq v_2, \dots, E_{n+m} \geq v_m] = \\
&\mathbb{P}[S_n \leq t, S_{n+1} > t + s \mid N_t = n] \mathbb{P}[N_t = n] \prod_{j=2}^m \mathbb{P}[E_{n+j} > v_j] \\
&= e^{-\lambda s} \mathbb{P}[N_t = n] \prod_{j=2}^m \mathbb{P}[E_{n+j} \geq v_j].
\end{aligned}$$

Oznacza to, że zmienne  $E'_1, E'_2, \dots$  są niezależne od  $N_t$  i mają rozkład wykładniczy z parametrem  $\lambda$ . Aby skończyć dowód musimy uzasadnić, że dla każdego  $n \in \mathbb{N}$  i dowolnych  $0 = t_0 < t < t_2 < \dots < t_n$  zmienne losowe

$$N(t), N(t_2) - N(t), \dots, N(t_n) - N(t_{n-1})$$

są niezależne i mają rozkład Poissona z odpowiednimi parametrami. Dla każdego  $k \leq n$  mamy

$$\begin{aligned}
N(t_k) - N(t) &= \#\{j \geq 1 : S_j \in (t, t_k]\} \\
&= \#\{j \geq N(t) + 1 : S_j - t \in (0, t_k - t]\} \\
&= \#\{j \geq 1 : S_{N(t)+j} - t \in (0, t_k - t]\} \\
&= \#\{j \geq 1 : S'_j \in (0, t_k - t]\},
\end{aligned}$$

gdzie  $S'_j = E'_1 + \dots + E'_j$ . To pokazuje, że wektor losowy

$$(N(t_2) - N(t), N(t_3) - N(t_2), \dots, N(t_n) - N(t_{n-1}))$$

jest mierzalny względem  $\sigma(E'_k : k \in \mathbb{N})$ , więc jest niezależny od  $N(t)$  oraz, że ma on taki sam rozkład co

$$(N(t_1 - t), N(t_2 - t) - N(t_1 - t), \dots, N(t_n - t) - N(t_{n-1} - t)).$$

W szczególności pokazuje to, że  $N(t_1) - N(t)$  i  $N(t_1 - t)$  mają rozkład Poissona z parametrem  $\lambda(t_1 - t)$ . Iterując to rozumowanie pokazujemy, że zmienne

$$N(t), N(t_2) - N(t), \dots, N(t_n) - N(t_{n-1})$$

są niezależne. □

### Wniosek 5.3

Niech  $N = (N_t)_{t \geq 0}$  będzie jednorodnym procesem Poissona z parametrem  $\lambda > 0$ . Wówczas

$$\lim_{t \rightarrow \infty} N_t/t = 1/\lambda$$

prawie wszędzie.

*Dowód.* Możemy założyć, że  $N_t$  jest zadane przez (5.1.1). Z mocnego prawa wielkich liczb

$$\lim_{n \rightarrow \infty} S_n/n = \mathbb{E}[S_1] = \lambda.$$

Skoro  $t \mapsto N_t$  jest niemalejąca oraz  $n = N(S_n)$ , to  $N_t \rightarrow \infty$  przy  $t \rightarrow \infty$ . Zauważmy też, że

$$S_{N(t)} \leq t < S_{N(t)+1}$$

Stąd

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)} \quad (5.1.3)$$

Odwołując się do prawa wielkich liczb wzdłuż wartości  $N(t) \rightarrow \infty$  otrzymujemy

$$\lim_{t \rightarrow \infty} \frac{S_{N(t)}}{N(t)} = \lim_{t \rightarrow \infty} \frac{S_{N(t)+1}}{N(t)+1} = \lambda.$$

Odwołując się do twierdzenia o trzech ciągach w (5.1.3) otrzymujemy tezę.  $\square$

## 5.2 Rozkłady stabilne

W naszych dotychczasowych rozważaniach wyjątkową rolę pełnił rozkład normalny, który pojawił się zarówno w centralnym twierdzeniu granicznym jak i w opisie ruchu Browna. Będziemy chcieli teraz zrozumieć, które z własności tego rozkładu są istotne oraz czy istnieją inne rozkłady, które mogą pojawiać się jako granice w twierdzeniach granicznych, czy też jako rozkłady w procesach o niezależnych, stacjonarnych przyrostach (takich jak ruch Browna, czy proces Poissona).

Kluczowa własność jest opisana w następującym lemacie:

### Lemat 5.4

Niech  $X$  będzie zmienną losową taką, że  $eX^2 < \infty$ . Wówczas  $X$  można przedstawić w postaci

$$X \stackrel{d}{=} \frac{X' + X''}{\sqrt{2}}, \quad (5.2.1)$$

gdzie  $X'$  i  $X''$  są niezależnymi zmiennymi losowymi o takim samym rozkładzie jak  $X$  wtedy i tylko wtedy, gdy  $X$  ma rozkład  $N(0, \sigma^2)$ .

*Dowód.* Załóżmy najpierw, że  $X \sim N(0, \sigma^2)$ , wówczas oznaczając przez  $\{X_n\}$  ciąg niezależnych kopii  $X$  i korzystając dwukrotnie z centralnego twierdzenia granicznego otrzymujemy

$$X \stackrel{d}{\leftarrow} \frac{X_1 + \dots + X_{2n}}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \left( \frac{X_1 + \dots + X_n}{\sqrt{n}} + \frac{X_{n+1} + \dots + X_{2n}}{\sqrt{n}} \right) \stackrel{d}{\rightarrow} \frac{1}{\sqrt{2}} (X' + X''),$$

gdzie  $X', X''$  są niezależnymi zmiennymi losowymi o rozkładzie  $N(0, 1)$ . Powyżej skorzystaliśmy też z faktu, że jeżeli ciągi zmiennych losowych  $\{Y_n\}, \{Z_n\}$  są niezależne i  $Y_n \xrightarrow{d} Y, Z_n \xrightarrow{d} Z$ , to  $Y_n + Z_n \xrightarrow{d} Y + Z$ .<sup>1</sup>

Założmy teraz, że  $X \stackrel{d}{=} \frac{X' + X''}{\sqrt{2}}$ . Wówczas z równości

$$eX = \frac{1}{\sqrt{2}}(eX' + eX'') = \sqrt{2}eX$$

wynika, że  $eX = 0$ . Oznaczmy jak powyżej przez  $\{X_n\}$  ciąg niezależnych kopii  $X$ , wówczas korzystając wielokrotnie z równania rekurencyjnego otrzymujemy

$$\begin{aligned} X &\stackrel{d}{=} \frac{X_1 + X_2}{\sqrt{2}} \stackrel{d}{=} \frac{\frac{X_3 + X_4}{\sqrt{2}} + \frac{X_5 + X_6}{\sqrt{2}}}{\sqrt{2}} \stackrel{d}{=} \frac{X_1 + X_2 + X_3 + X_4}{\sqrt{4}} \\ &\stackrel{d}{=} \dots \stackrel{d}{=} \frac{X_1 + \dots + X_{2^n}}{2^{n/2}} \xrightarrow{d} N(0, \sigma^2). \end{aligned}$$

Zauważmy, że w powyższym rozumowaniu nie jesteśmy w stanie wyznaczyć jednoznacznie wariancji i to wynika z prostej obserwacji, że jeżeli zmienna losowa  $X$  spełnia równanie 5.2.1, to również  $cX$  dla dowolnej stałej  $c$  ją spełnia. Natomiast centralne twierdzenie graniczne gwarantuje, że  $X$  musi mieć rozkład normalny.  $\square$

Rozkład normalny ma więc własność 'samorozkładalności'. Będziemy poszukiwać rozkładów o podobnych własnościach.

### Definicja 5.5

Zmienna losowa  $X$  jest **stabilna**, jeżeli dla każdego  $n$  istnieją stałe  $a_n$  i  $b_n$  takie, że

$$a_n X + b_n \stackrel{d}{=} S_n,$$

gdzie  $S_n = X_1 + \dots + X_n$  i  $\{X_n\}$  są niezależnymi zmiennymi losowymi o takim samym rozkładzie jak  $X$ .

**Przykład 5.2.2.** • Rozkład normalny  $N(0, \sigma^2)$ : jeżeli  $X_i \sim N(0, \sigma^2)$ , to  $S_n \sim N(0, n\sigma^2)$ , a więc definicja zachodzi dla  $a_n = \frac{1}{\sqrt{n}}, b_n = 0$ . Przypomnijmy wzór na funkcję charakterystyczną rozkładu normalnego:

<sup>1</sup> Ta własność nie jest prawdziwa bez założenia niezależności. Natomiast, jeżeli oba ciągi są niezależne, to można ją udowodnić wykorzystując funkcje charakterystyczne:

$$\phi_{Y_n + Z_n}(t) = \phi_{Y_n}(t) \cdot \phi_{Z_n}(t) \rightarrow \phi_Y(t) \cdot \phi_Z(t) = \phi_{Y+Z}(t).$$

$$\phi(t) = e[e^{itX}] = e^{-\sigma^2 t^2 / 2}.$$

- Rozkład Cauchy'ego z parametrami  $(m, \sigma)$ : wówczas gęstość  $X$  jest zadana wzorem

$$f(x) = \frac{1}{\pi} \frac{\sigma}{(x - m)^2 + \sigma^2}.$$

Przypomnijmy, że wówczas  $eX = \infty$ . Niech  $X \sim \text{Cauchy}(0, \sigma)$ , wówczas

$$e[e^{itX}] = e[e^{-\sigma|t|}].$$

Zauważmy, że  $S_n \sim \text{Cauchy}(0, n\sigma)$ , gdyż korzystając z funkcji charakterystycznych

$$e[e^{itS_n}] = (e[e^{itX}])^n = e[e^{-n\sigma|t|}] = e[e^{itnX}],$$

stąd

$$e[e^{itS_n/n}] = e[e^{-\sigma|t|}].$$

Zatem definicja jest spełniona dla  $a_n = n$  i  $b_n = 0$ .

- Ogólniej: zmienna losowa  $X$  ma symetryczny rozkład  $\alpha$ -stabilny jeżeli

$$e[e^{itX}] = e^{-c|t|^\alpha}$$

dla pewnych stałych  $\alpha \in (0, 2]$  oraz  $c > 0$ . Wówczas  $a_n = n^{1/\alpha}$ ,  $b_n = 0$ .

Istnienie takiej zmiennej losowej nie jest w tej chwili oczywiste, ale będzie wynikało z Twierdzenia 5.7. Wiadomo, że gęstość  $g_\alpha$  jest symetryczna oraz dla  $\alpha < 2$

$$g_\alpha(x) \sim 1/|x|^{1+\alpha}, \quad \text{gdy } |x| \rightarrow \infty.$$

Okazuje się, że wzór na gęstość takiej symetrycznej zmiennej stabilnej nie jest znany poza szczególnymi przypadkami (np. rozkład normalny dla  $\alpha = 2$ , Cauchy'ego dla  $\alpha = 1$ )

Okazuje się, że jeżeli badamy twierdzenia graniczne dla sum niezależnych zmiennych losowych o takim samym rozkładzie, to w granicy pojawia się rozkład stabilny.

### Definicja 5.6

Jeżeli  $\{X_n\}$  jest ciągiem i.i.d. oraz  $S_n = X_1 + \dots + X_n$ , to mówimy, że  $X_1$  (również dystrybuanta  $F_{X_1}$ ) należy do obszaru przyciągania dystrybuanty  $G$ , jeżeli istnieją ciągi normalizujące  $a_n, b_n$  takie, że

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} G.$$

Piszemy wówczas  $F_{X_1} \in \mathcal{D}(G)$  ( $X_1 \in \mathcal{D}(Z)$  dla  $G = G_Z$ ).

**Przykład 5.2.3.** • Jeżeli  $eX^2 < \infty$ , to  $X \in \mathcal{D}(N(0,1))$  z parametrami  $a_n = \sqrt{n\text{var}X}$ ,  $b_n = nE X$ .

- Jeżeli  $X$  jest zmienną stabilną, to z definicji  $X \in \mathcal{D}(X)$ , bo wtedy  $X \stackrel{d}{=} \frac{S_n - b_n}{a_n}$  dla każdego  $n$ .

Pokazuje się, że tylko zmienne stabilne posiadają niepuste obszary przyciągania, a więc mogą występować jako granice w twierdzeniach granicznych. Ponadto można podać precyzyjne warunki zbieżności (jednak nie w terminach momentów jak w CTG, a w terminach regularnego zachowania zmiennych losowych z nieskończoności). Twierdzenie to wykracza poza ramy naszego wykładu (choć dowód bazuje na znanych nam metodach wykorzystujących funkcje charakterystyczne). Pokażemy szczególny przypadek tego twierdzenia.

### Twierdzenie 5.7

Załóżmy, że  $\{X_n\}$  jest ciągiem i.i.d. i zmienne losowe mają ograniczoną gęstość  $f$  spełniającą

$$f(x) = f(-x) \quad \text{oraz} \quad f(x) \sim \frac{c}{x^{1+\alpha}}, \text{ gdy } x \rightarrow \infty$$

dla  $c > \alpha > 0$  (symbol  $\sim$  oznacza, że iloraz obu funkcji zbiega do 1).

1. Jeżeli  $\alpha \in (0, 2)$ , to ciąg

$$Z_n = \frac{X_1 + \dots + X_n}{n^{1/\alpha}}$$

zbiega według rozkładu do symetrycznej  $\alpha$ -stabilnej zmiennej losowej o funkcji charakterystycznej  $e^{-r|t|^\alpha}$ .

2. Jeżeli  $\alpha > 2$ , to ciąg

$$Z_n = \frac{X_1 + \dots + X_n}{\sqrt{n}}$$

zbiega według rozkładu do  $N(0, \sigma^2)$ .

3. Jeżeli  $\alpha = 2$ , to ciąg

$$Z_n = \frac{X_1 + \dots + X_n}{\sqrt{n \log n}}$$

zbiega według rozkładu do  $N(0, \sigma^2)$ .

*Dowód.* Przedstawimy dowód jedynie części 1. (pkt. 2 to CTG, gdyż wówczas  $eX^2 < \infty$ , a pkt. 3 pozostawiamy czytelnikowi).

W dowodzie użyjemy funkcji charakterystycznych. Pokażemy, że

$$e[e^{itX}] = e^{-r|t|^\alpha + o(|t|^\alpha)}, \quad t \rightarrow 0. \quad (5.2.4)$$

dla pewnej stałej  $r$ .

Wówczas dla każdego  $t$

$$e[e^{itZ_n}] = e[e^{n \frac{it}{n^{1/\alpha}} \cdot (X_1 + \dots + X_n)}] = e^{n(r|t/n^{1/\alpha}|^\alpha + o(|t/n^{1/\alpha}|^\alpha))} = e^{-r|t|^\alpha + no(|t|^\alpha/n)} \rightarrow e^{-r|t|^\alpha}, \quad n \rightarrow \infty$$

Z twierdzenia Lévy'ego o ciągłości powyższa zbieżność funkcji charakterystycznych implikuje zbieżność według rozkładu, ale pociąga również istnienie symetrycznej stabilnej zmiennej losowej.

Do dowodu (5.2.4) ustalmy  $t > 0$  oraz oznaczmy  $\phi(t) = e[e^{itX}]$ . Wówczas

$$\begin{aligned} \phi(t) - 1 &= \int_{\mathbb{R}} (e^{itx} - 1)f(x)dx = 2 \int_0^\infty (\cos(tx) - 1)f(x)dx \\ &\stackrel{y:=tx}{=} \frac{2}{t} \int_0^\infty (\cos y - 1)f(y/t)dy = 2t^\alpha \int_0^\infty \frac{\cos y - 1}{t^{1+\alpha}} f(y/t)dy. \end{aligned}$$

Pokażemy teraz

$$\lim_{t \searrow 0} \int_0^\infty \frac{\cos y - 1}{t^{1+\alpha}} f(y/t)dy = cI, \quad \text{dla } I = \int_0^\infty \frac{\cos y - 1}{y^{1+\alpha}} dy. \quad (5.2.5)$$

Zauważmy, że powyższa całka jest zbieżna. Ustalmy  $\varepsilon < (2 - \alpha)/3$  (czyli  $3\varepsilon + \alpha < 2$ ) i rozbijmy powyższą całkę na dwie części. Dla  $y < t^{1-\varepsilon}$  piszemy

$$\int_0^{t^{1-\varepsilon}} \frac{\cos y - 1}{t^{1+\alpha}} f(y/t)dy \leq \frac{C}{t^{1+\alpha}} \int_0^{t^{1-\varepsilon}} 1 \wedge y^2 dy \leq Ct^{3(1-\varepsilon)}/t^{1+\alpha} = Ct^{2-(\alpha+3\varepsilon)} \rightarrow 0,$$

gdy  $t \rightarrow 0$ .

Jeżeli  $y \geq t^{1-\varepsilon}$ , to  $y/t \rightarrow \infty$ , co pociąga

$$f(y/t) = c(y/t)^{-(1+\alpha)}(1 + o(1)).$$

Stąd

$$\lim_{t \rightarrow 0} \int_{t^{1-\varepsilon}}^\infty \frac{\cos y - 1}{t^{1+\alpha}} f(y/t)dy = c \lim_{t \rightarrow 0} \int_{t^{1-\varepsilon}}^\infty \frac{\cos y - 1}{y^{1+\alpha}} dy = cI.$$

Pozostaje do pokazania, że (5.2.4) wynika z (5.2.5). Zauważmy, że (5.2.5) możemy zapisać w postaci

$$\frac{\phi(t) - 1}{t^\alpha} \rightarrow 2cI, \quad \text{gdy } t \rightarrow 0,$$

a więc również

$$\frac{\log \phi(t)}{t^\alpha} \rightarrow 2cI, \quad \text{gdy } t \rightarrow 0,$$

co jest równoważne z (5.2.4) dla  $r = 2cI$ . □

## 5.3 Procesy Lévy'ego i rozkłady nieskończenie podzielne

2

### Definicja 5.8

Proces stochastyczny  $\{X_t\}_{t \geq 0}$  jest nazywany procesem Lévy'ego jeżeli

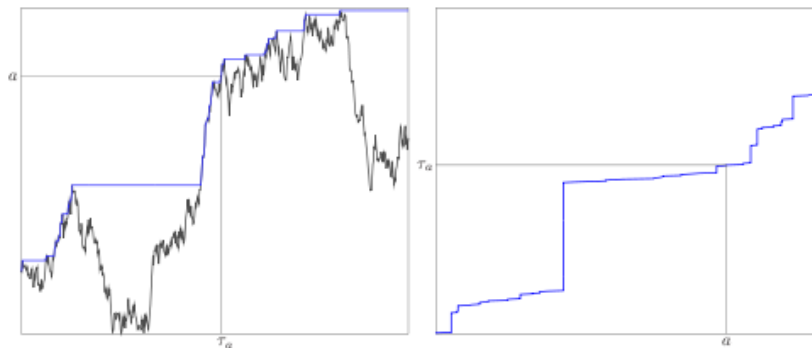
1.  $X_0 = 0$ ;
2. (stacjonarne przyrosty) dla każdych  $s < t$ ,  $X_t - X_s$  ma taki sam rozkład jak  $X_{t-s}$ ;
3. (niezależne przyrosty) dla każdych  $s < t$ , zmienna losowa  $X_t - X_s$  jest niezależna od  $\sigma\{X_r\}_{r < s}$ ;
4. (trajektorie cadlag) z prawdopodobieństwem 1, funkcja  $t \mapsto X_t$  jest prawostronnie ciągła i posiada lewe granice.

**Przykład 5.3.1. Ruch Browna** (jest to jedyny proces Lévy'ego o ciągłych trajektoriach).

**Przykład 5.3.2. Proces Poissona** jest procesem Lévy'ego.

**Przykład 5.3.3.** Niech  $W_t$  będzie standardowym ruchem Browna. Zdefiniujemy

$$T_s = \inf\{t : W_t = s\}, \quad s \geq 0.$$



Wówczas zachodzi

### Lemat 5.9

$\{T_s\}$  jest ma stacjonarne i niezależne przyrosty.

+

<sup>2</sup> Rozdział ma charakter pogładowy i większość faktów jest przedstawiona bez kompletnych dowodów

*Dowód.* Istotnie,  $T_0 = 0$ . Ustalmy  $s \leq p < q$ . Z mocnej własności Markowa proces  $W'_t = W_{T_s+t} - W_{T_s}$  jest ruchem Browna niezależnym od  $W_{T_s}$ . Wtedy

$$T_p = T_s + \inf\{t : W'_t = p - s\}, \quad T_q = T_s + \inf\{t : W'_t = q - s\}.$$

Zatem  $T_q - T_p$  nie zależy od  $T_s$ , co daje warunek 3. Ponadto biorąc  $s = p$  otrzymujemy

$$T_q - T_p = \inf\{t : W'_t = q - p\} \stackrel{d}{=} \inf\{t : W_t = q - p\} = T_{q-p},$$

a więc przyrosty są stacjonarne.  $\square$

Ostatecznie wyliczymy gęstość  $T_s$ . Oznaczmy przez  $\Phi$  dystrybuantę rozkładu normalnego, a przez  $\phi$  jego gęstość. Korzystając z zasady odbicia, piszemy

$$\mathbb{P}[T_s < t] = \mathbb{P}\left[\sup_{u \in [0,t]} W_u > s\right] = 2\mathbb{P}[W_t > s] = 2\mathbb{P}[W_1 > s/\sqrt{t}] = 2(1 - \Phi(s/\sqrt{t})).$$

Różniczkując obie strony otrzymujemy gęstość:

$$f_s(t) = \phi(s/\sqrt{t}) \cdot \frac{s}{t^{3/2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2t}} \cdot \frac{s}{t^{3/2}}.$$

Zauważmy jednak, że  $T_s$  nie jest procesem Lévy'ego, bo trajektorie nie są cadlag (ale są lewostronnie ciągłe i mają prawe granice). Można ten proces jednak trochę poprawić. Zdefiniujemy

$$U_s = \inf\{t : W_t > s\}, \quad s \geq 0.$$

Wówczas  $\{U_s\}_s$  jest procesem Lévy'ego.

**Przykład 5.3.4. Złożony proces Poissona** (*compound Poisson process*). Niech  $\{T_i\}$  będą niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z parametrem  $\lambda$ . Oznaczmy

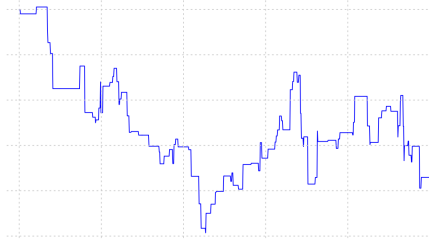
$$\tau_n = T_1 + \dots + T_n.$$

Niech  $\{Y_i\}_{i \in \mathbb{N}}$  będą niezależnymi zmiennymi losowymi o takim samym rozkładzie  $\mu_0$ . Będą to skoki procesu. Zakładamy, że oba ciągi  $\{T_i\}$  i  $\{Y_i\}$  są od siebie niezależne. Zakładamy, że  $\mathbb{P}[Y_1 = 0] = 0$  (skoki o długości 0 nie wpłyną na proces). Oznaczmy  $S_n = Y_1 + \dots + Y_n$ . Niech  $N_t$  będzie procesem Poissona, tj

$$N_t = n, \quad \text{dla } \tau_n \leq t < \tau_{n+1}.$$

Zdefiniujemy proces

$$X_t = S_{N_t}.$$



Złożony proces Poissona  $X_t$  jest procesem Lévy'ego. Zauważmy, że gdy,  $\mu_0 = \delta_1$ , to  $X_t = N_t$  jest procesem Poissona.

Policzmy funkcję charakterystyczną  $X_t$ . Oznaczmy  $\mu = \lambda\mu_0$ , wtedy  $\mu(\mathbb{R}) = \lambda$ . Miara  $\mu$  jest nazywana miarą Lévy'ego. Zachodzi

**Lemat 5.10**

$$\phi_{X_t}(s) = e^{t \int_{\mathbb{R}} (e^{isx} - 1) \mu(dx)}.$$

*Dowód.* Zauważmy najpierw, że

$$\phi_Y(s) = e[e^{isY}] = \int_{\mathbb{R}} e^{isx} \mu_0(dx).$$

Dalej

$$\begin{aligned} \phi_{X_t}(s) &= e[e^{isX_t}] = \sum_{n=0}^{\infty} \mathbb{P}[N_t = n] e[e^{isX_t} | N_t = n] \\ &= \sum_{n=0}^{\infty} \mathbb{P}[N_t = n] \phi_Y(s)^n \stackrel{N_t \sim \text{Pois}(\lambda t)}{=} \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(t\lambda)^n}{n!} \phi_Y(s)^n \\ &= \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(t\lambda \phi_Y(s))^n}{n!} = e^{t\lambda(\phi_Y(s) - 1)} = e^{t \int_{\mathbb{R}} (e^{isx} - 1) \mu(dx)} \end{aligned}$$

□

Jakie rozkłady mogą się pojawić jako rozkłady procesów Lévy'ego  $X_t$ ?  
Niech

$$Y_{j,n} = X_{\frac{j}{n}} - X_{\frac{j-1}{n}},$$

wówczas

$$X_1 = \sum_{j=1}^n Y_{j,n}$$

ponadto zmienne losowe  $Y_{j,n}$  są niezależne i wszystkie mają ten sam rozkład (wynika to z definicji procesu Lévy'ego). Prowadzi nas to do definicji

**Definicja 5.11**

Zmienna losowa  $X$  ma **nieskończenie podzielny rozkład** jeżeli dla każdego  $n$  istnieją zmienne losowe  $Y_{1,n}, \dots, Y_{n,n}$ , które są i.i.d. oraz ponadto  $X$  ma taki sam rozkład jak  $Y_{1,n} + \dots + Y_{n,n}$ , tj.

$$X \stackrel{d}{=} Y_{1,n} + \dots + Y_{n,n}.$$

**Przykład 5.3.5.** • *Rozkład normalny jest nieskończenie podzielny:*

$$X \sim N(m, \sigma^2), Y_{j,n} \sim N(m/n, \sigma^2/n).$$

• *Rozkład Poissona jest nieskończenie podzielny:*

$$X \sim \text{Poiss}(\lambda), Y_{j,n} \sim \text{Poiss}(\lambda/n).$$

Jeżeli  $X_t$  jest procesem Lévy'ego, to  $X_1$  (oraz każde  $X_t$ ) ma rozkład nieskończenie podzielny. Również odwrotna implikacja jest prawdziwa, tzn. dla każdego rozkładu nieskończenie podzielnego istnieje odpowiedni proces Lévy'ego.

Chcemy zrozumieć jak mogą wyglądać rozkłady nieskończenie podzielne oraz procesy Lévy'ego. Na zakończenie wykładu podamy dwa różne sformułowania charakteryzacji Lévy'ego-Chinczyna. Mówi ona, że rozkład nieskończenie podzielny jest sumą niezależnych zmiennych losowych o rozkładzie normalnym i uogólnionego rozkładu Poissona. Podobnie każdy proces Lévy'ego jest sumą ruchu Browna i 'uogólnionego' procesu Poissona

**Twierdzenie 5.12** (Charakteryzacja Lévy'ego-Chinczyna)

Niech  $\{X_t\}_{t \geq 0}$  będzie procesem Lévy'ego. Wówczas  $X_t$  może być przedstawiony jako suma niezależnych procesów

$$X_t = mt + \sigma W_t + C_t + Y_t,$$

gdzie

- $W_t$  jest standardowym ruchem Browna;
- $C_t$  jest złożonym procesem Poissona z miarą Lévy'ego  $\mu_C$  spełniającą

$$\mu_C\{|x| \leq 1\} = 0, \quad \mu_C(\mathbb{R}) = \mu_C\{|x| > 1\} < \infty;$$

- $Y_t$  jest uogólnionym procesem Poissona z miarą Lévy'ego  $\mu_Y$  spełniającą

$$\mu_Y\{|x| > 1\} = 0, \quad \mu_Y\{0\} = 0, \quad \int_{-1}^1 x^2 \mu_Y(dx) < \infty;$$

W powyższym rozkładzie mamy więc liniowy dryft, ruch Browna oraz część skokową procesu  $C_t + Y_t$ . Miara  $\mu = \mu_C + \mu_Y$  nazywa się miarą Lévy'ego. Proces  $C_t$  zawiera w sobie duże skoki, powyżej 1. Natomiast proces  $Y_t$  może posiadać nieskończenie wiele małych skoków (taką własność ma np. proces  $\{T_t\}$  opisujący czasy trafienia ruchu Browna). Powyższy warunek całkowy pozwala jednak kontrolować całkowitą sumę tych skoków. Każdy proces Lévy'ego jest jednoznacznie wyznaczony przez trójkę  $(m, \sigma, \mu)$ .

### Twierdzenie 5.13

Jeżeli zmienna losowa  $X$  ma nieskończenie podzielny rozkład, to istnieją:  $m \in \mathbb{R}$ ,  $\sigma > 0$  oraz miara  $\mu$  spełniająca

$$\mu\{0\} = 0, \quad \int_{\mathbb{R}} (1 \wedge x^2) \mu(dx) < \infty,$$

takie, że  $\phi_X$ , funkcja charakterystyczna  $X$ , może być zapisana w postaci

$$\begin{aligned} \phi_X(t) &= \exp \left( imt - \frac{\sigma^2}{2} \cdot t^2 + \int_{\mathbb{R}} (e^{itx} - 1 - itx \mathbf{1}_{\{|x| \leq 1\}}) \mu(dx) \right) \\ &= \exp \left( imt - \frac{\sigma^2}{2} \cdot t^2 + \int_{\{|x| > 1\}} (e^{itx} - 1) \mu_C(dx) + \int_{\{|x| \leq 1\}} (e^{itx} - 1 - itx) \mu_Y(dx) \right), \end{aligned}$$

gdzie  $\mu_C(A) = \mu(A \cap [-1, 1])$ , a  $\mu_T(A) = \mu(A \cap [-1, 1]^c)$ .