

# SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH<sup>1</sup>

DARIUSZ BURACZEWSKI

## 1. ŁAŃCUCHY MARKOWA I MIARY STACJONARNE

**Przykład 1.1.** Rozważmy spacer losowy na  $\Omega = \{0, 1\}$  zdefiniowany następująco. Cząsteczka początkowo znajduje się w punkcie 0. Rzucamy monetą  $m_0$ , niekoniecznie symetryczną. Jeżeli wypadnie orzełek, z prawdopodobieństwem  $p$ , to cząsteczka przemieszcza się do 1, w przeciwnym razie pozostaje w 0. W punkcie 1 używamy monety  $m_1$ , w której orzełek wypada z prawdopodobieństwem  $q$ . Jeżeli wypadnie orzeł, to cząsteczka idzie do 0.

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

$$\mathbb{P}[X_1 = 0|X_0 = 0] = 1 - p, \quad \mathbb{P}[X_1 = 1|X_0 = 0] = p.$$

W drugim kroku:

$$\begin{aligned} \mathbb{P}[X_2 = 0|X_0 = 0] &= (1 - p)^2 + pq, \\ \mathbb{P}[X_2 = 1|X_0 = 0] &= (1 - p)p + p(1 - q). \end{aligned}$$

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

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

Oznaczmy przez  $\mu_t$  rozkład zmiennej losowej  $X_t$ ,<sup>2</sup> tzn.

$$\mu_t = (\mathbb{P}[X_t = 1|X_0 = 0], \mathbb{P}[X_t = 0|X_0 = 0])$$

Zauważmy

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

(powyżej wektory piszemy jako wiersze).

Chcemy zrozumieć jak wygląda zachowanie  $X_t$  (czy też  $\mu_t$ ) dla dużych wartości  $t$ . W szczególności chcemy zrozumieć, czy rozkłady stabilizują się po pewnym czasie lub dążą do jakiejś granicy. Jeżeli tak, to oznaczmy tę granicę przez  $\pi$ . Powinna ona spełniać

$$\pi = \pi P.$$

Łatwo obliczyć, że wówczas  $\pi$  powinno być postaci

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

Zdefiniujmy

$$\varepsilon_t = \mu_t(0) - \pi(0).$$

Wówczas

$$\begin{aligned} \varepsilon_{t+1} &= \mu_t(0)(1 - p) + (1 - \mu_t(0))q - \pi(0) \\ &= \mu_t(0)(1 - p - q) + \pi(0)(p + q) - \pi(0) \\ &= (1 - p - q)\varepsilon_t. \end{aligned}$$

Zatem, jeżeli  $0 < p + q < 2$ , to

$$\lim_{t \rightarrow \infty} \mu_t(0) = \pi(0) \quad \text{oraz} \quad \lim_{t \rightarrow \infty} \mu_t(1) = \pi(1).$$

<sup>1</sup>Skrypt powstał na bazie wykładu *Spacery losowe na skończonych grafach*, który odbywał się wiosną 2015 roku na Wydziale Matematyki i Informatyki Uniwersytetu Wrocławskiego.

<sup>2</sup>przez  $t$  będziemy oznaczać czas

Zauważmy, 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  $P$  (wrócimy do tej własności później)

**1.1. Łańcuchy Markowa.** Powyższy przykład jest łańcuchem Markowa. Będziemy używać tej samej notacji.

- $\{X_t\}_{t \in \mathbb{N}}$  będzie oznaczać proces losowy na skończonej przestrzeni  $\Omega$ ;
- $P$  - macierz przejścia;  $P$  będzie macierzą stochastyczną o wymiarach  $|\Omega| \times |\Omega|$ , tzn. taką, że jej wyrazy są nieujemne oraz

$$\sum_{y \in \Omega} P(x, y) = 1 \quad \forall x \in \Omega.$$

- $\mu_t$  - rozkład  $X_t$ ;
- $\pi$  - miara stacjonarna procesu  $\{X_t\}$ , tzn. miara spełniająca  $\pi = \pi P$ .

Łańcuch Markowa  $\{X_t\}$  jest to process bez pamięci, tzn. taki, że  $X_{t+1}$  zależy wyłącznie od  $X_t$ . Dokładniej

**Definicja 1.1.** Ciąg zmiennych losowych  $\{X_t\}$  o wartościach w  $\Omega$  nazywamy (jednorodnym) **łańcuchem Markowa** z macierzą przejścia  $P$  jeżeli dla każdych  $x, y \in \Omega$  oraz  $x_0, x_1, \dots, x_{t-1}$  mamy

$$\begin{aligned} \mathbb{P}[X_{t+1} = y | X_0 = x_0, X_1 = x_1, \dots, X_{t-1} = x_{t-1}, X_t = x] &= \mathbb{P}[X_{t+1} = y | X_t = x] \\ &= P(x, y). \end{aligned}$$

Naszym celem jest otrzymanie podobnych wyników jak powyżej, tj. opisanie miar stacjonarnych dla konkretnych przykładów oraz tempa zbieżności  $\mu_t$  to  $\pi$ .

Podobnie jak w przykładzie mamy:

$$\mu_t = \mu_0 P^t.$$

Troche oznaczeń: Często będziemy oznaczać zależność  $X_t$  od początkowego rozkładu  $\mu_0 = \mu$  i będziemy pisać  $\mathbb{P}_\mu$  i  $\mathbb{E}_\mu$  dla odpowiedniego prawdopodobieństwa i wartości oczekiwanej. Zazwyczaj  $\mu = \delta_x$  jest miarą skoncentrowaną w pewnym stanie  $x \in \Omega$ . Wówczas piszemy po prostu:  $\mathbb{P}_x$  i  $\mathbb{E}_x$ .

Wtedy

$$\mathbb{P}_x[X_t = y] = \mathbb{P}_{\delta_x}[X_t = y] = \mathbb{P}[X_t = y | X_0 = x] = P^t(x, y).$$

**Definicja 1.2.** Łańcuch Markowa jest **nieredukowalny** jeżeli dla każdych dwóch stanów  $x, y \in \Omega$  istnieje  $t \in \mathbb{N}$  takie, że  $P^t(x, y) > 0$

**Definicja 1.3.** Niech  $T(x) = \{t : P^t(x, x) > 0\}$ , wówczas  $\text{NWDT}(x)$  nazywane jest okresem stanu  $x$ . Łańcuch Markowa jest **aperiodyczny** jeżeli okres każdego stanu wynosi 1.

Zazwyczaj będziemy zakładać, że badane łańcuchy Markowa są nieredukowalne i aperiodyczne.

**Przykład 1.2.** Rozważmy spacer losowy na  $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$  zdefiniowany następująco:  $X_{t+1} = X_t \pm 1 \pmod n$ , gdzie  $+1$  i  $-1$  są wybrane z prawdopodobieństwem  $1/2$ . Jeżeli  $n$  jest liczbą parzystą, to spacer ten jest periodyczny. Okres każdego punktu jest wówczas równy 2. Będziemy unikać 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 1.1.** *Jeżeli łańcuch Markowa jest nieredukowalny, to wszystkie punkty mają ten sam okres.*

**Lemat 1.2.** *Jeżeli łańcuch Markowa jest nieredukowalny i aperiodyczny, to istnieje  $M$  takie, że  $P^t(x, y) > 0$  dla wszystkich  $x, y \in \Omega$  oraz  $t \geq M$ .*

## 1.2. Miara stacjonarna.

**Definicja 1.4.** Miarę probabilistyczną nazywamy **miarą stacjonarną** łańcucha Markowa, jeżeli  $\pi = \pi P$ , tzn.

$$\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y) \quad \forall x, y \in \Omega.$$

Jeżeli  $X_0$  ma rozkład  $\pi$ , to również  $X_t$  ma rozkład  $\pi$ . Jak zobaczymy poniżej dla dowolnego łańcucha nieredukowalnego i aperiodycznego rozkłady  $X_t$  zbiegają do miary stacjonarnej.

**Przykład 1.3. Proste spacery losowe na grafach.** Graf  $G = (V, E)$  składa się ze zbioru wierzchołków  $V$  oraz zbioru krawędzi  $E$ , gdzie  $E$  zbiorem nieuporządkowanych par wierzchołków:

$$E \subset \{(x, y) : x, y \in V, x \neq y\}.$$

Jeżeli  $(x, y) \in E$ , to wierzchołki  $x, y$  nazywamy sąsiadami i oznaczamy  $x \sim y$ . Stopniem wierzchołka  $x \in V$  nazywamy liczbę jego sąsiadów i oznaczamy  $\deg(x)$ .

Na danym grafie  $G = (V, E)$  definiujemy prosty spacer losowy. Jest to łańcuch Markowa na przestrzeni stanów  $V$  z macierzą przejścia

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)} & \text{jeżeli } y \sim x \\ 0 & \text{w przeciwnym razie} \end{cases}$$

Gdy łańcuch znajduje się w wierzchołku  $x$ , to wybiera losowo (jednostajnie) jednego z jego sąsiadów i przechodzi do niego. Jeżeli graf jest spójny, to spacer jest nieredukowalny.

Łatwo jest wówczas wyznaczyć miarę stacjonarną spaceru losowego. Mianowicie

$$\pi(y) = \frac{\deg(y)}{2|E|}$$

jest miarą stacjonarną, gdyż

$$\pi P(y) = \sum_{x \in \Omega} \pi(x) P(x, y) = \sum_{x \sim y} \frac{\deg(x)}{2|E|} \cdot \frac{1}{\deg(x)} = \frac{\deg(y)}{2|E|} = \pi(y).$$

Zauważmy, że jeżeli graf  $G$  jest  $d$ -regularny (każdy wierzchołek ma ten sam stopień równy  $d$ ), to  $2|E| = d|V|$  i miara jednostajna  $\pi(y) = 1/|V|$  jest stacjonarna.

**Twierdzenie 1.3.** *Każdy nieredukowalny i aperiodyczny łańcuch Markowa posiada jedyną stacjonarną miarę.*

*Miara ta zadana jest wzorem*

$$\pi(x) = \frac{1}{\mathbb{E}_x \tau_x^+}.$$

*Ponadto  $\pi(x) > 0$  dla każdego  $x \in \Omega$*

Teraz pokażemy istnienie miary stacjonarnej. Później, w twierdzeniu (2.4), udowodnimy że rozkłady  $\mu_t$  zbiegają szybko do  $\pi$  oraz, że miara stacjonarna musi być jedyna.

**Definicja 1.5.** Dla każdego  $x \in \Omega$  definiujemy **pierwszy czas powrotu do  $x$** :

$$\tau_x := \min\{t \geq 0 : X_t = x\},$$

jest to pierwszy moment, gdy łańcuch znajduje się w  $x$ . Będziemy chcieli wyróżnić sytuację, gdy proces startuje z  $x$  i po raz pierwszy wraca do  $x$ . W tym celu definiujemy **pierwszy dodatni czas uderzenia**

$$\tau_x^+ := \min\{t \geq 1 : X_t = x\}.$$

Obie definicje różnią się jedynie dla punktu startu.

**Lemat 1.4.** Dla każdych dwóch stanów  $x, y \in \Omega$  nieredukowalnego i aperiodycznego łańcucha Markowa zachodzi

$$\mathbb{E}_x \tau_y^+ < \infty.$$

*Dowód.* Z lematu 1.2 istnieje  $M$  takie, że dla każdych dwóch stanów  $x, y \in \Omega$ ,  $P^M(x, y) > 0$ . Oznaczmy

$$\varepsilon = \min_{x, y} P^M(x, y) > 0.$$

Wtedy dla dowolnego  $k > 0$

$$\begin{aligned} \mathbb{P}_x[\tau_y^+ > kM] &\leq \mathbb{P}_x[X_{kM} \neq y \text{ i } \tau_y^+ > (k-1)M] \\ &= \sum_{z \neq y} \mathbb{P}_x[X_{kM} \neq y, X_{(k-1)M} = z \text{ i } \tau_y^+ > (k-1)M] \\ &= \sum_{z \neq y} \mathbb{P}_x[X_{kM} \neq y \text{ i } \tau_y^+ > (k-1)M | X_{(k-1)M} = z] \mathbb{P}[X_{(k-1)M} = z] \\ &= \sum_{z \neq y} \mathbb{P}_x[X_{kM} \neq y | X_{(k-1)M} = z] \mathbb{P}_x[\tau_y^+ > (k-1)M | X_{(k-1)M} = z] \mathbb{P}[X_{(k-1)M} = z] \\ &= \sum_{z \neq y} \mathbb{P}_z[X_M \neq y] \mathbb{P}_x[\tau_y^+ > (k-1)M \text{ i } X_{(k-1)M} = z] \\ &\leq (1 - \varepsilon) \mathbb{P}_x[\tau_y^+ > (k-1)M] \\ &\leq \dots \leq (1 - \varepsilon)^k. \end{aligned}$$

Dalej mamy

$$\mathbb{E}_x \tau_y^+ = \sum_{t \geq 0} \mathbb{P}_x[\tau_y^+ > t] \leq \sum_{k \geq 0} M \mathbb{P}_x[\tau_y^+ > kM] \leq M \sum_{k \geq 0} (1 - \varepsilon)^k < \infty.$$

□

*Dowód Twierdzenia 1.3.* Ustalmy dowolny  $z \in \Omega$ . Zdefiniujemy

$$\begin{aligned} \tilde{\pi}(y) &:= \mathbb{E}_z[\text{liczba wizyt w } y \text{ przed pierwszym powrotem do } z] \\ &= \sum_{t=0}^{\infty} \mathbb{P}_z[X_t = y, \tau_z^+ > t]. \end{aligned}$$

Powyższa wartość jest oczekiwaną liczbą wizyt w  $y$  przed pierwszym powrotem do  $z$ . Jeżeli  $y = z$ , to  $\tilde{\pi}(z) = 1$ . Wartość ta jest skończona, gdyż  $\tilde{\pi}(y) \leq \mathbb{E}_z \tau_z^+ < \infty$  (Lemat 1.4).

Pokażemy, że  $\tilde{\pi}$  jest miarą niezmienniczą (tzn.  $\tilde{\pi}P = \tilde{\pi}$ ):

$$\begin{aligned}
\tilde{\pi}P(y) &= \sum_{x \in \Omega} \tilde{\pi}(x)P(x, y) \\
&= \sum_{x \in \Omega} \sum_{t=0}^{\infty} \mathbb{P}_z[X_t = x, \tau_z^+ > t]P(x, y) \\
&= \sum_{t=0}^{\infty} \sum_{x \in \Omega} \mathbb{P}_z[X_t = x, \tau_z^+ > t]P(x, y) \\
&= \sum_{t=0}^{\infty} \mathbb{P}_z[X_{t+1} = y, \tau_z^+ \geq t+1] \\
&= \sum_{t=1}^{\infty} \mathbb{P}_z[X_t = y, \tau_z^+ \geq t] \\
&= \tilde{\pi}(y) - \mathbb{P}_z[X_0 = y, \tau_z^+ > 0] + \sum_{t=1}^{\infty} \mathbb{P}_z[X_t = y, \tau_z^+ = t] \\
&= \tilde{\pi}(y) - \mathbb{P}_z[X_0 = y] + \mathbb{P}_z[X_{\tau_z^+} = y] \\
&= \tilde{\pi}(y).
\end{aligned}$$

$\tilde{\pi}$  jest więc miarą niezmienniczą. Należy ją jednak znormalizować, aby otrzymać miarę stacjonarną (probabilistyczną). Zauważmy, że

$$\sum_{x \in \Omega} \tilde{\pi}(x) = \mathbb{E}_z \tau_z^+.$$

Zatem

$$\pi(x) = \frac{\tilde{\pi}(x)}{\mathbb{E}_z \tau_z^+}$$

jest miarą stacjonarną i probabilistyczną. W szczególności przyjmując  $z = x$  otrzymujemy

$$\pi(x) = \frac{1}{\mathbb{E}_x \tau_x^+}.$$

□

### 1.3. Odwracalne łańcuchy Markowa.

**Definicja 1.6.** Niech  $\{X_t\}$  będzie łańcuchem Markowa na przestrzeni stanów  $\Omega$  z macierzą przejścia  $P$ . Miara probabilistyczna  $\pi$  na  $\Omega$  jest nazywana **odwracalną** dla łańcucha jeżeli dla dowolnych  $x, y \in \Omega$  mamy

$$(1.5) \quad \pi(x)P(x, y) = \pi(y)P(y, x).$$

Łańcuch Markowa jest nazywany **odwracalnym** jeżeli istnieje dla niego odwracalna miara.

**Lemat 1.6.** Jeżeli  $\pi$  jest miarą odwracalną dla pewnego łańcucha Markowa, to jest również dla niego miarą stacjonarną

*Dowód.* Korzystając z powyższej definicji oraz stochastyczności macierzy  $P$ :

$$\pi P(x) = \sum_{y \in \Omega} \pi(y)P(y, x) = \sum_{y \in \Omega} \pi(x)P(x, y) = \pi(x)$$

□

Z powyższego lematu wynika, że jeżeli macierz przejścia  $P$  jest symetryczna, to miara jednostajna jest miarą symetryczną.

Zauważmy, że jeżeli łańcuch jest odwracalny, to dla każdego  $n$ :

$$\begin{aligned}\mathbb{P}_\pi[X_0 = x_0, \dots, X_n = x_n] &= \pi(x_0)P(x_0, x_1) \dots P(x_{n-1}, x_n) \\ &= \pi(x_n)P(x_n, x_{n-1}) \dots P(x_1, x_0) \\ &= \mathbb{P}_\pi[X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0]\end{aligned}$$

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).

#### 1.4. Przykłady.

**Przykład 1.4.** Prosty spacer losowy na grafie (przykład 1.3) jest odwracalny. Przypomnijmy, że miara  $\pi(x) = \deg(x)/2|E|$  jest stacjonarna. Mamy:

$$\pi(x)P(x, y) = \frac{\deg(x)}{2|E|} \cdot \frac{\mathbf{1}_{\{x \sim y\}}}{\deg(x)} = \frac{\mathbf{1}_{\{x \sim y\}}}{2|E|} = \pi(y)P(y, x)$$

**Przykład 1.5.** 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

$$\pi P(y) = \sum_{j \in \mathbb{Z}_n} \pi(j)P(j, k) = \pi(k-1)p + \pi(k+1)(1-p) = \frac{1}{n} = \pi(k).$$

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 1.6. Spacer losowy na  $n$ -hiperkostce.**  $n$ -hiperkostka jest to graf, którego wierzchołkami są ciągi binarne długości  $n$ , tzn  $\Omega = \{0, 1\}^n$ . Dwa wierzchołki są połączone ze sobą krawędzią jeżeli różnią się dokładnie na jednej współrzędnej. Tak więc każdy wierzchołek ma dokładnie  $n$  sąsiadów. Prosty spacer losowy polega na przejściu z wierzchołka  $x = (x_1, \dots, x_n)$  do jednego z jego sąsiadów z prawdopodobieństwem  $1/n$  (jak w przykładzie 1.3). Dla  $n = 3$  jest to prosty spacer losowy po wierzchołkach sześcianu. Zatem miara jednostajna jest miarą stacjonarną. Zauważmy, że spacer ten jest periodyczny. Istotnie, jeżeli spacer losowy startuje w  $(0, \dots, 0)$ , to w  $t$ -tym kroku liczba jedynek w  $X_t$  ma tę samą parzystość, co  $t$ . 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 w prawdopodobieństwie  $1/2$ , a z prawdopodobieństwem  $1/2$  przechodzi jednostajnie do jednego z sąsiadów. Spacer ten jest aperiodyczny i miara jednostajna jest jego miarą stacjonarną.

**Przykład 1.7. Proces urodzin i śmierci.** Jest to proces na przestrzeni stanów  $\Omega = \{0, 1, \dots, n\}$ . 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:  $\{(p_k, r_k, q_k)\}_{k=0}^n$ , gdzie:

- $p_k$  jest prawdopodobieństwem przejścia z  $k$  do  $k+1$ , dla  $0 \leq k < n$ ;
- $q_k$  jest prawdopodobieństwem przejścia z  $k$  do  $k-1$ , dla  $0 < k \leq n$ ;
- $r_k$  jest prawdopodobieństwem pozostania w  $k$ , dla  $0 \leq k \leq n$ ;

- $p_k + r_k + q_k = 1$ ;
- $q_0 = p_n = 0$

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 (1.5) jest spełnione. Zatem miara

$$\pi(k) = \frac{\tilde{\pi}(k)}{\sum \tilde{\pi}(k)}$$

jest stacjonarna.

**1.5. Spacery losowe na grupach.** Niech  $\mu$  będzie miarą probabilistyczną na skończonej grupie  $G$ . Będziemy rozważać **lewe** spacery losowe na grupie  $G$  zadane przez macierz przejścia

$$P(g, hg) = \mu(h).$$

Wówczas jeżeli  $X_0 = e$  oraz  $g_1, \dots, g_t$  jest ciągiem i.i.d., to

$$X_t = g_t \dots g_1.$$

Powyżej poznaliśmy przykłady

- Prosty spacer losowy na  $\mathbb{Z}_n$ . Wówczas:  $G = \mathbb{Z}_n$ ,  $\mu = 1/2(\delta_{-1} + \delta_{+1})$ .
- Prosty spacer losowy na hiperkostce. Należy utożsamić hiperkostkę z  $\mathbb{Z}_2^n$  i miarę  $\mu$  położyć jednostajnie na elementach  $\mathbf{e}_j$ , które mają 1 na  $j$  współrzędnej i 0 na pozostałych.
- Podobnie można opisać leniwe spacery losowe na  $\mathbb{Z}_n$  i  $\mathbb{Z}_2^n$ .

**Lemat 1.7.** Niech  $U$  będzie jednostajną miarą probabilistyczną na grupie  $G$ , tzn.  $U(g) = 1/|G|$ . Wówczas dla dowolnego spaceru losowego na  $G$ ,  $U$  jest miarą stacjonarną.

*Dowód.* Niech  $\mu$  będzie miarą opisującą spacer losowy na  $G$ . Wówczas:

$$UP(g) = \sum_{h \in G} U(h)P(h, g) = \frac{1}{|G|} \sum_{k \in G} P(k^{-1}g, g) = \frac{1}{|G|} \sum_{k \in G} \mu(k) = \frac{1}{|G|} = U(g).$$

□

## 2. PAROWANIE I MIESZANIE

**2.1. Norma całkowitego wahanía.** Wiemy już, że spacery losowe posiadają miarę stacjonarną. Do tej pory nie pokazaliśmy jednak bezpośredniego związku (poza formułą definiującą stacjonarność) pomiędzy spacerem, a miarą stacjonarną. W tym rozdziale pokażemy, że rozkład spaceru zbiega do miary stacjonarnej. Głównym celem wykładu będzie opis tempa zbieżności dla konkretnych przykładów.

W tym celu musimy umieć mierzyć odległość pomiędzy miarami. Najbardziej naturalna jest tzw. norma całkowitego wahanía (jak zobaczymy poniżej pochodząca od normy  $l^1$ ).

**Definicja 2.1.** Normę całkowitego wahanía (ang. total variation distance) pomiędzy dwoma miarami probabilistycznymi  $\mu$  i  $\nu$  definiujemy następująco

$$\|\mu - \nu\|_{\text{TV}} = \max_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

Zatem odległość pomiędzy dwoma miarami probabilistycznymi jest równa maksymalnej różnicy ich wartości przyporządkowanej pojedynczemu zdarzeniu. Zauważmy, że

- $\|\mu - \nu\|_{\text{TV}} = 0$  wtedy i tylko wtedy gdy  $\mu = \nu$ .
- $\|\mu - \nu\|_{\text{TV}} \leq 1$ , a równość zachodzi, gdy miary mają rozłączne nośniki.

**Przykład 2.1.** Powróćmy do przykładu 1.1. Przypomnijmy

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}, \quad \pi = \left( \frac{q}{p+q}, \frac{p}{p+q} \right)$$

Załóżmy, że  $\mu_0 = (1, 0)$ . Wówczas są jedynie 4 możliwe zdarzenia:  $\emptyset, \{0\}, \{1\}, \{0, 1\}$  i jak łatwo widać

$$\|\mu_t - \pi\|_{\text{TV}} = \varepsilon_t = P^t(0, 0) - \pi(0) = \pi(1) - P^t(0, 1),$$

dla  $\varepsilon_t = \mu_t(0) - \pi(0)$ . Jak wiemy  $\varepsilon_t = (1 - p - q)^t \varepsilon_0$ . Zatem odległość  $\mu_t$  od  $\pi$  maleje wykładniczo szybko.

**Lemat 2.1.** Dla dowolnych miar probabilistycznych na  $\Omega$  zachodzi

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \sum_{\{x: \mu(x) > \nu(x)\}} (\mu(x) - \nu(x))$$

*Dowód.* Niech

$$B = \{x : \mu(x) \geq \nu(x)\}$$

i niech  $A \subset \Omega$  będzie dowolnym zdarzeniem. Wówczas

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B),$$

Zauważmy również, że

$$\mu(B) - \nu(B) = \nu(B^c) - \mu(B^c).$$

Rozumując analogicznie

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \nu(B^c) - \mu(B^c) = \mu(B) - \nu(B).$$

Zatem

$$\max_{A \subset \Omega} |\mu(A) - \nu(A)| \leq \mu(B) - \nu(B) = \nu(B^c) - \mu(B^c)$$

i jednocześnie maksimum to jest przyjęte dla  $A = B$ . stąd

$$\|\mu - \nu\|_{\text{TV}} = \mu(B) - \nu(B) = \sum_{x \in B} (\mu(x) - \nu(x)).$$

Ostatecznie

$$\begin{aligned} \|\mu - \nu\|_{\text{TV}} &= \frac{1}{2} (\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)) \\ &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \end{aligned}$$

□

**Wniosek 2.2.** Norma całkowitego wahanía spełnia nierówność trójkąta, tzn dla trzech dowolnych miar probabilistycznych  $\nu, \mu, \eta$  zachodzi

$$\|\mu - \nu\|_{\text{TV}} \leq \|\mu - \eta\|_{\text{TV}} + \|\eta - \nu\|_{\text{TV}}.$$



## 2.2. Parowanie.

**Definicja 2.2. Parowaniem** (ang. coupling) dwóch miar probabilistycznych  $\mu$  i  $\nu$  nazywamy parę zmiennych losowych  $(X, Y)$  zdefiniowanych na tej samej przestrzeni probabilistycznej  $\Omega$  taką, że  $X$  ma rozkład  $\mu$ , a  $Y$  rozkład  $\nu$ . Tzn. parowanie  $(X, Y)$  spełnia  $\mathbb{P}[X = x] = \mu(x)$  i  $\mathbb{P}[Y = y] = \nu(y)$ .

**Przykład 2.2.** Niech  $\mu = \nu = 1/2(\delta_0 + \delta_1)$ . Przykładami parowania są

- para  $(X, Y)$  niezależnych zmiennych losowych taka, że  $\mathbb{P}[X = x, Y = y] = 1/4$  dla wszystkich  $x, y \in \{0, 1\}$ ;
- para  $(X, Y)$  taka, że  $X = Y$ , tzn.  $\mathbb{P}[X = Y = x] = 1/2$  dla  $x \in \{0, 1\}$ .

Przykład pokazuje, że parowanie jest opisane przez wspólny rozkład  $(X, Y)$  na  $\Omega \times \Omega$ , czyli przez wartości

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

Jeżeli miary  $\mu$  i  $\nu$  są równe, to można zdefiniować parowanie takie, że  $X = Y$  prawie wszędzie.

Istnieje związek pomiędzy parowaniem, a normą całkowitego wahania:

**Twierdzenie 2.3.** Niech  $\mu$  i  $\nu$  będą miarami probabilistycznymi na  $\Omega$ . Wówczas

$$\|\mu - \nu\|_{\text{TV}} = \inf\{\mathbb{P}[X \neq Y] : (X, Y) \text{ jest parowaniem } \mu \text{ i } \nu\}.$$

Ponadto istnieje parowanie, dla którego powyższe infimum jest przyjęte.

*Dowód.* W terminach powyższej macierzy  $q$ , chcemy znaleźć parowanie, które ma jak największą masę na diagonalu. Łatwo powyżej pokazać nierówność. Niech  $(X, Y)$  będzie dowolnym parowaniem miar  $\mu$  i  $\nu$  i niech  $A \subset \Omega$  będzie dowolnym zdarzeniem. Wtedy

$$\mu(A) - \nu(A) = \mathbb{P}[X \in A] - \mathbb{P}[Y \in A] \leq \mathbb{P}[X \in A, Y \notin A] \leq \mathbb{P}[X \neq Y].$$

Zatem

$$\|\mu - \nu\|_{\text{TV}} \leq \inf\{\mathbb{P}[X \neq Y] : (X, Y) \text{ jest parowaniem } \mu \text{ i } \nu\}.$$

Pozostaje skonstruować parowanie realizujące infimum. Zatem chcemy, aby  $X$  było równe  $Y$  na jak największym zbiorze. W tym celu zapiszemy  $\|\mu - \nu\|_{\text{TV}}$  w nieco inny sposób. Niech

$$p = \sum_{x \in \Omega} \min\{\mu(x), \nu(x)\}.$$

Wtedy

$$\begin{aligned} p &= \sum_{x \in \Omega} \min\{\mu(x), \nu(x)\} = \sum_{\{x: \mu(x) \leq \nu(x)\}} \mu(x) + \sum_{\{x: \mu(x) > \nu(x)\}} \nu(x) \\ &= \left( \sum_{\{x: \mu(x) \leq \nu(x)\}} \mu(x) + \sum_{\{x: \mu(x) > \nu(x)\}} \mu(x) \right) + \left( \sum_{\{x: \mu(x) > \nu(x)\}} \nu(x) - \sum_{\{x: \mu(x) > \nu(x)\}} \mu(x) \right) \\ &= 1 - \sum_{\{x: \mu(x) > \nu(x)\}} (\mu(x) - \nu(x)) = 1 - \|\mu - \nu\|_{\text{TV}}. \end{aligned}$$

Ostatnia równość wynika z lematu 2.1. Zatem

$$\|\mu - \nu\|_{\text{TV}} = 1 - p.$$

Musimy więc znaleźć parowanie takie, że

$$\mathbb{P}[X = Y] = p.$$

Znalezienie parowania polega na zdefiniowaniu funkcji  $q(x, y) = \mathbb{P}[X = x, Y = y]$ . Zauważmy, że  $p = \mathbb{P}[X = Y] = \sum_x q(x, x)$ . Jednocześnie dla każdych  $x, y$ :  $q(x, y) \leq \mu(x)$  oraz  $q(x, y) \leq \nu(y)$ . Zatem na diagonalu możemy położyć co najwyżej  $\min\{\mu(x), \nu(x)\}$ .

W tym celu postępujemy następująco. Definiujemy trzy miary probabilistyczne (sprawdź, że są to istotnie nieujemne miary probabilistyczne):

$$\begin{aligned}\gamma_D(x) &= \frac{\min\{\mu(x), \nu(x)\}}{p}, \\ \gamma_X(x) &= \frac{\mu(x) - \min\{\mu(x), \nu(x)\}}{1 - p}, \\ \gamma_Y(y) &= \frac{\nu(y) - \min\{\mu(y), \nu(y)\}}{1 - p}\end{aligned}$$

a następnie definiujemy parę  $(X, Y)$ . Rzucamy monetą, w której orzeł wypada z prawdopodobieństwem  $p$ :

- jeżeli wypadnie orzeł, to kładziemy  $X=Y=Z$ , gdzie  $Z$  jest zmienną wylosowaną względem miary  $\gamma_D$ ;
- jeżeli wypadnie reszka, to losujemy niezależnie  $X$  względem miary  $\gamma_X$  i  $Y$  względem miary  $\gamma_Y$ .

Oczywiście

$$\mathbb{P}[X = Y] = \mathbb{P}[\text{wypadnie orzeł}] = p.$$

Pozostaje do sprawdzenia, że jest to rzeczywiście parowanie. Z definicji  $\gamma_X$  wynika

$$\mu = p\gamma_D + (1 - p)\gamma_X.$$

Ponadto  $\mathbb{P}[X = x, Y = x \text{ oraz wypadła reszka}] = 0$  dla każdego  $x \in \Omega$ . Stąd

$$\begin{aligned}\mathbb{P}[X = x] &= \mathbb{P}[X = Y = x \text{ oraz wypadł orzeł}] + \sum_{y \neq x} \mathbb{P}[X = x, Y = y \text{ oraz wypadła reszka}] \\ &= p\gamma_D(x) + \sum_y \mathbb{P}[X = x, Y = y \text{ oraz wypadła reszka}] \\ &= p\gamma_D(x) + (1 - p)\gamma_X(x) \\ &= \mu(x).\end{aligned}$$

Analogicznie pokazujemy

$$\mathbb{P}[Y = y] = \nu(y).$$

Zauważmy, że macierz  $q$  możemy napisać wzorem:

$$\begin{aligned}q(x, x) &= p\gamma_D(x), \\ q(x, y) &= (1 - p)\gamma_X(x)\gamma_Y(y), \quad x \neq y.\end{aligned}$$

□

Chcemy zdefiniować parowanie nie tylko dwóch zmiennych losowych, ale całych łańcuchów Markowa.

**Przykład 2.3.** Rozważamy prosty spacer losowy na  $\Omega = \{0, 1, \dots, n\}$ , przy założeniu, że spacer nie przechodzi przez punkty brzegowe, tzn.  $P(0, 0) = P(0, 1) = 1/2$ . Intuicyjnie jest oczywiste, że dla  $x \leq y$

$$P^t(x, n) \leq P^t(y, n)$$

Jak to udowodnić?

Pokażemy dowód używający metody parowania. Niech  $\Delta_i$  będzie ciągiem iid. takim, że  $\mathbb{P}[\Delta_i = \pm 1] = 1/2$ . Zdefiniujemy dwa spacery losowe na  $\Omega$ :

- proces  $X_t$  startuje z punktu  $x$  i w każdym kroku (jeżeli to możliwe) dodaje  $\Delta_t$ ;
- proces  $Y_t$  startujący z  $y \geq x$  i zdefiniowany jak wyżej.

Łatwo pokazać podstawowe własności obu procesów:

- rozkładem  $X_t$  jest  $P^t(x, \cdot)$ , a rozkładem  $Y_t$  jest  $P^t(y, \cdot)$ ;
- jeżeli oba procesy znajdują się w tym samym stanie, to od tego momentu będą miały te same trajektorie;
- $Y_t \geq X_t$  dla każdego  $t$ , w szczególności jeżeli  $X_t = n$ , to również  $Y_t = n$ . Zatem

$$P^t(x, n) = \mathbb{P}[X_t = n] \leq \mathbb{P}[Y_t = n] = P^t(y, n). \quad \square$$

Będziemy używać parowania łańcuchów Markowa.

**Definicja 2.3. Parowaniem łańcuchów Markowa** z macierzą przejścia  $P$  nazywamy proces  $(X_t, Y_t)$  taki, że zarówno  $X_t$  jak i  $Y_t$  jest łańcuchem Markowa z macierzą przejścia  $P$ . Oba procesy mogą mieć różne rozkłady początkowe.

Zauważmy, że parowanie łańcuchów Markowa może być zmodyfikowane, w taki sposób, że jeżeli oba procesy spotkają się w tym samym stanie, to zostaną połączone. Dokładniej, niech

$$Z_t = \begin{cases} X_t & \text{dla } t \leq T \\ Y_t & \text{dla } t > T \end{cases}$$

dla

$$T = \min\{n : X_n = Y_n\}.$$

Wówczas  $(Z_t, Y_t)$  też jest parowaniem.

**Twierdzenie 2.4** (Twierdzenie o zbieżności). *Załóżmy, że łańcuch Markowa jest nieredukowalny oraz aperiodyczny z miarą stacjonarną  $\pi$ . Istnieją stałe  $\beta \in (0, 1)$  oraz  $C > 0$  takie, że*

$$\|\mu_t - \pi\|_{\text{TV}} \leq C\beta^t$$

*Dowód.* Niech  $\{X_t\}_{t \in \mathbb{N}}$  i  $\{Y_t\}_{t \in \mathbb{N}}$  będą dwoma **niezależnymi** łańcuchami Markowa z macierzą przejścia zdefiniowanymi na wspólnej przestrzeni probabilistycznej (parowanie). Załóżmy ponadto, że  $Y_0$  ma rozkład  $\pi$  (zatem każdy z  $Y_t$  ma rozkład  $\pi$ ), a  $X_t$  ma rozkład  $\mu_t$ .

Pokażemy najpierw, że oba procesy w pewnym momencie spotkają się, tzn. jeżeli

$$\tau_{\text{par}} = \min\{n : X_n = Y_n\},$$

to  $\mathbb{P}[\tau_{\text{par}} < \infty] = 1$ .

Ponieważ łańcuch jest nieredukowalny i aperiodyczny, to z lematu 1.2 istnieją  $M$  i  $\delta$  takie, że

$$P^M(x, y) > \delta > 0$$

dla wszystkich  $x, y \in \Omega$ . Ustalmy  $x$ , wówczas

$$\begin{aligned}
\mathbb{P}[\tau_{\text{par}} \leq M] &\geq \mathbb{P}[X_M = Y_M] \\
&\geq \mathbb{P}[X_M = x, Y_M = x] \\
&= \mathbb{P}[X_M = x] \mathbb{P}[Y_M = x] \\
&= \left( \sum_{y \in \Omega} \mathbb{P}[X_0 = y, X_M = x] \right) \cdot \left( \sum_{y \in \Omega} \mathbb{P}[Y_0 = y, Y_M = x] \right) \\
&= \left( \sum_{y \in \Omega} \mathbb{P}[X_M = x | X_0 = y] \mathbb{P}[X_0 = y] \right) \cdot \left( \sum_{y \in \Omega} \mathbb{P}[Y_M = x | Y_0 = y] \mathbb{P}[Y_0 = y] \right) \\
&\geq \left( \delta \sum_{y \in \Omega} \mathbb{P}[X_0 = y] \right) \cdot \left( \delta \sum_{y \in \Omega} \mathbb{P}[Y_0 = y] \right) \\
&= \delta^2.
\end{aligned}$$

Zatem

$$\mathbb{P}[\tau_{\text{par}} > M] < 1 - \delta^2.$$

Podobne argumenty dają:

$$\begin{aligned}
\mathbb{P}[X_{2M} = Y_{2M} | \tau_{\text{par}} > M] &\geq \frac{1}{\mathbb{P}[\tau_{\text{par}} > M]} \sum_{x, y \neq z} \mathbb{P}[X_{2M} = x, Y_{2M} = x, X_M = y, Y_M = z, \tau_{\text{par}} > M] \\
&\geq \sum_{x, y \neq z} \mathbb{P}[X_{2M} = x, Y_{2M} = x | X_M = y, Y_M = z] \mathbb{P}[X_M = y, Y_M = z] \\
&\geq \dots \geq \delta^2
\end{aligned}$$

i dalej

$$\begin{aligned}
\mathbb{P}[\tau_{\text{par}} > 2M] &= \mathbb{P}[\tau_{\text{par}} > M] \mathbb{P}[\tau_{\text{par}} > 2M | \tau_{\text{par}} > M] \\
&\leq (1 - \delta^2) \mathbb{P}[\tau_{\text{par}} > 2M | \tau_{\text{par}} > M] \\
&\leq (1 - \delta^2) \mathbb{P}[X_{2M} \neq Y_{2M} | \tau_{\text{par}} > M] \\
&\leq (1 - \delta^2) (1 - \mathbb{P}[X_{2M} = Y_{2M} | \tau_{\text{par}} > M]) \\
&\leq (1 - \delta^2)^2
\end{aligned}$$

i przez indukcję

$$\mathbb{P}[\tau_{\text{par}} > kM] \leq (1 - \delta^2)^k.$$

To pokazuje, że  $\lim_{n \rightarrow \infty} \mathbb{P}[\tau_{\text{par}} > n] = 0$ , zatem  $\tau_{\text{par}}$  jest skończone p.w.

Następnie konstruujemy nowy łańcuch Markowa

$$Z_t = \begin{cases} X_t & \text{dla } t \leq \tau_{\text{par}} \\ Y_t & \text{dla } t > \tau_{\text{par}} \end{cases}$$

Nowy proces zachowuje się najpierw jak  $X_t$ , a w momencie spotkania obu procesów 'skleja się' z  $Y_t$ . Ponieważ  $X_0 = Z_0$ , to  $\mu_t$  jest rozkładem  $Z_t$  dla każdego  $t$ . Dalej mamy dla każdego  $x \in \Omega$

$$\mu_t(x) - \pi(x) = \mathbb{P}[Z_t = x] - \mathbb{P}[Y_t = x] \leq \mathbb{P}[Z_t = x, Y_t \neq x] \leq \mathbb{P}[Z_t \neq Y_t] = \mathbb{P}[\tau_{\text{par}} > t].$$

Analogicznie

$$\pi(x) - \mu_t(x) \leq \mathbb{P}[\tau_{\text{par}} > t].$$

Z lematu 2.1 mamy więc

$$\|\mu_t - \pi\|_{\text{TV}} \leq C \mathbb{P}[\tau_{\text{par}} > t].$$

Powyższe wyrażenie zbiega wykładniczo do 0.  $\square$

**Wniosek 2.5.** *Każdy nieredukowalny i aperiodyczny łańcuch Markowa ma dokładnie jedną miarę stacjonarną*

*Dowód.* Załóżmy, że łańcuch Markowa posiada dwie stacjonarne miary:  $\pi$  i  $\pi'$ . Niech  $\mu_0 = \pi'$ , wówczas również  $\mu_t = \pi'$ , ale z twierdzenia

$$\|\pi' - \pi\|_{\text{TV}} = \|\mu_t - \pi\|_{\text{TV}} \rightarrow 0.$$

Więc  $\pi' = \pi$ .  $\square$

**Definicja 2.4.** Naszym celem jest badanie odległości pomiędzy  $P^t(x, \cdot)$  i  $\pi$  dlatego też definiujemy

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}}$$

Będziemy również rozważać:

$$\bar{d}(t) = \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}}$$

**Lemat 2.6.** *Mamy*

$$d(t) \leq \bar{d}(t) \leq 2d(t)$$

*Dowód.* Prawa nierówność wynika bezpośrednio z nierówności trójkąta. Do pokazania lewej nierówności skorzystamy ze stacjonarności miary  $\pi$  ( $\pi = \pi P^t$ ), która implikuje

$$\pi(A) = \sum_{y \in \Omega} \pi(y) P^t(y, A)$$

dla każdego zbioru  $A$  (przypomnijmy  $\pi(A) = \sum_{x \in A} \pi(x)$ ). Stąd

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{\text{TV}} &= \max_{A \subset \Omega} |P^t(x, A) - \pi(A)| \\ &= \max_{A \subset \Omega} \left| \sum_{y \in \Omega} \pi(y) [P^t(x, A) - P^t(y, A)] \right| \\ &\leq \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \\ &= \bar{d}(t). \end{aligned}$$

$\square$

**Lemat 2.7.** *Funkcja  $\bar{d}$  jest podmultiplikatywna, tzn*

$$\bar{d}(t + s) \leq \bar{d}(t)\bar{d}(s).$$

Dowód powyższego lematu pomijamy.

**Definicja 2.5.** Definiujemy czas mieszania (ang. mixing time)

$$t_{\text{miks}}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}.$$

Jest to moment, gdy odległość łańcucha Markowa od miary stacjonarnej jest mała. To jest kluczowy parametr, który będziemy badać podczas wykładu.

Dla ustalenia uwagi definiuje się

$$t_{\text{miks}} = t_{\text{miks}}(1/4)$$

Wybór  $1/4$  jest dowolny i pokazuje się (zadanie), że dla małych  $\varepsilon$

$$t_{\text{miks}}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{\text{miks}}.$$

We wszystkich przykładach interesuje nas głównie rząd wielkości. Wartość stałej jest drugorzędna.

**Twierdzenie 2.8.** *Niech  $(X_t, Y_t)$  będzie parowaniem takim, że  $X_0 = x, Y_0 = y$  oraz*

$$X_t = Y_t \quad \text{dla } t \geq \tau_{\text{par}}.$$

*Wtedy*

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq \mathbb{P}_{x,y}[\tau_{\text{par}} > t].$$

( $\mathbb{P}_{x,y}$  jest miarą probabilistyczną na przestrzeni, na której jest zdefiniowane parowanie  $(X_t, Y_t)$  z warunkiem początkowym  $X_0 = x, Y_0 = y$ ).

*Dowód.* Ustalmy  $t$ . Mamy

$$P^t(x, z) = \mathbb{P}_{x,y}[X_t = z] \quad \text{oraz} \quad P^t(y, z) = \mathbb{P}_{x,y}[Y_t = z].$$

Zatem  $(X_t, Y_t)$  jest parowaniem  $P^t(x, \cdot)$  i  $P^t(y, \cdot)$ , więc z twierdzenia 2.3

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq \mathbb{P}_{x,y}[X_t \neq Y_t] \leq \mathbb{P}_{x,y}[\tau_{\text{par}} > t].$$

□

**Przykład 2.4. Leniwy spacer losowy na  $\mathbb{Z}_n$ .** Konstruujemy parowanie  $(X_t, Y_t)$  dwóch cząsteczek realizujących leniwy spacer na  $\mathbb{Z}_n$ . Jedna startuje z  $x$ , a druga z  $y$ . Obie nie poruszają się jednocześnie. W ten sposób nie mogą przez siebie przeskoczyć (zamienić się miejscami).

Reguły są następujące. W każdym kroku, aż do pierwszego momentu spotkania, rzucamy monetą:

- jeżeli wypadnie orzeł, to w tym kroku porusza się pierwsza cząsteczką, a kierunek jest wskazany przez kolejny rzut monetą;
- jeżeli wypadnie reszka, to w tym kroku porusza się druga cząsteczką, a kierunek jest wskazany przez kolejny rzut monetą.

Z punktu widzenia cząsteczek, obie wykonują leniwy spacer losowy (tzn. z prawdopodobieństwem  $1/2$  pozostają w miejscu, a z prawdopodobieństwem  $1/2$  przechodzą do jednego z sąsiadów). Jeżeli obie cząsteczki znajdują się w tym samym punkcie, to od tej pory poruszają się wspólnie. Jest to parowanie (sprawdź!).

Niech  $D_t$  oznacza odległość pomiędzy cząsteczkami w chwili  $t$ . Zauważmy, że  $D_t$  jest prostym spacerem losowym na zbiorze  $\{0, 1, \dots, n\}$ , z absorbującymi punktami  $0$  i  $n$  (porównaj z problemem ruiny gracza<sup>3</sup>). Dla

$$\tau = \min\{t \geq 0 : D_t \in \{0, n\}\},$$

mamy  $\tau = \tau_{\text{par}}$ . Ponadto z zadania 9 (lista 1) wiemy, że  $\mathbb{E}_{x,y}\tau = k(n - k)$ , gdzie  $k$  jest odległością na  $\mathbb{Z}_n$  pomiędzy  $x$  i  $y$ . Zatem z lematu 2.6, twierdzenia 2.8 oraz nierówności Markowa<sup>4</sup>

$$d(t) \leq \max_{x,y \in \mathbb{Z}_n} \mathbb{P}_{x,y}[\tau > t] \leq \frac{\max_{x,y} \mathbb{E}_{x,y}(\tau)}{t} \leq \frac{n^2}{4t}.$$

Prawa strona jest równa  $1/4$  dla  $t = n^2$ , stąd  $t_{\text{miks}} \leq n^2$ . Pokażemy później, że  $t_{\text{miks}} \geq c_1 n^2$ .

<sup>3</sup>Lista 1, zadanie 9

<sup>4</sup>Nierówność Markowa: dla każdej całkowalnej zmiennej losowej zachodzi  $\mathbb{P}[X > t] \leq \mathbb{E}X/t$ .

**Przykład 2.5. Leniwy spacer losowy na torusie.** Rozważamy  $d$ -wymiarowy torus. Jest to graf, którego zbiorem wierzchołków jest  $\mathbb{Z}_n^d = \mathbb{Z}_n \times \cdots \times \mathbb{Z}_n$ . Dwa wierzchołki są połączone ze sobą krawędzią, gdy różnią się od siebie na dokładnie jednej współrzędnej o  $(1 \bmod n)$ . Jeżeli  $n$  jest parzyste, to prosty spacer losowy jest periodyczny, będziemy więc badać leniwy spacer losowy na  $\mathbb{Z}_n^d$ .

Chcemy oszacować czas mieszania  $t_{\text{miks}}$ , w tym celu skonstruujemy parowanie na torusie. Ustalmy  $x, y \in \mathbb{Z}_n^d$ . Definiujemy dwa spacery losowe:  $\{X_t\}$  oraz  $\{Y_t\}$  startujące odpowiednio w  $x$  i  $y$ . Losujemy jedną ze współrzędnych

- jeżeli na tej współrzędnej oba spacery się zgadzają, to oba przesuwamy o  $1, -1, 0$  z prawdopodobieństwem  $1/4, 1/4, 1/2$ ;
- jeżeli na wylosowanej współrzędnej spacery się różnią, to losujemy jeden z nich i przesuwamy o  $1$  lub  $-1$ .

Wówczas

$$\tau_{\text{par}} = \max_{1 \leq i \leq d} \tau_i,$$

gdzie

$$\tau_i = \min\{t \geq 0 : X_t^i = Y_t^i\}$$

Zauważmy, że jeżeli będziemy patrzeć wyłącznie na momenty, gdy wylosowano  $i$ -tą współrzędną, to  $\{X_t^i\}$  oraz  $\{Y_t^i\}$  -  $i$ -te współrzędne  $\{X_t\}$  oraz  $\{Y_t\}$ ; zachowują się jak leniwy spacer losowy na  $\mathbb{Z}_n$ .

Obliczmy  $\mathbb{E}_{x,y} \tau_i$ . Oznaczmy przez  $\{Z_j\}$  ciąg niezależnych zmiennych losowych takich, że  $Z_1$  oznacza pierwszy moment, w którym wylosowano  $i$ -tą współrzędną,  $Z_1 + Z_2$  - drugi moment itd. Wówczas  $Z_j$  mają rozkład geometryczny z prawdopodobieństwem sukcesu  $1/d$ , zatem  $\mathbb{E} Z_j = d$ . Niech  $\tau$  będzie zdefiniowane jak w poprzednim przykładzie, tzn. oznacza numer przemieszczenia jednej z cząsteczek na  $i$ -tej współrzędnej, po którym obie mają identyczną  $i$ -tą współrzędną.  $\tau$  jest niezależne od wartości  $Z_j$ . Wtedy, z przykładu 2.4 oraz tożsamości Walda (zadanie ..., lista 2), mamy

$$\mathbb{E}_{x,y} \tau_i = \mathbb{E}_{x,y} \left[ \sum_{j=1}^{\tau} Z_j \right] = \mathbb{E}_{x,y} \tau \cdot \mathbb{E}_{x,y} Z_j \leq \frac{dn^2}{4},$$

a stąd (korzystamy z lematu 2.6, twierdzenia 2.8 oraz nierówności Markowa)

$$\begin{aligned} d(t) &\leq \max_{x,y \in \mathbb{Z}_n^d} \mathbb{P}_{x,y}[\tau_{\text{par}} > t] \\ &\leq \frac{\max_{x,y} \mathbb{E}_{x,y}(\tau_{\text{par}})}{t} \\ &\leq \frac{\max_{x,y} \mathbb{E}_{x,y}(\sum \tau_i)}{t} \\ &\leq \frac{d^2 n^2}{4t}. \end{aligned}$$

Przyjmując  $t = d^2 n^2$  otrzymujemy

$$t_{\text{miks}} \leq d^2 n^2$$

**Przykład 2.6. Leniwy spacer losowy na hiperkostce.** Popatrzmy na ten proces nieco inaczej.  $\Omega = \mathbb{Z}_2^n$ . Przejście do sąsiada powoduje zmianę wartości na jednej ze współrzędnych (z prawdopodobieństwem  $1/(2n)$ ). Wylosujemy jednostajnie jedną ze współrzędnych, a następnie zmieńmy ją z prawdopodobieństwem  $1/2$ . Zobaczmy, że proces nie zostanie zmieniony z prawdopodobieństwem  $1/2$ , a jednostajnie przejdzie do jednego z sąsiadów. Wykorzystamy tę obserwację do zdefiniowania parowania.

Rozważamy dwa procesy startujące z różnych pozycji. W każdym kroku losujemy jedną ze współrzędnych  $i$ , a następnie losujemy jej wartość, którą przyporządkowujemy obu procesom. Jest to parowanie. Chcemy oszacować  $\tau_{\text{par}}$ . Ta wartość jest nie większa niż pierwszy moment, w którym wylosowano ostatnią współrzędną. Zatem z zadania 10 lista 1 (przykład kolekcjonera kuponów) mamy

$$d(n \log n + cn) \leq \mathbb{P}[\tau_{\text{par}} > n \log n + cn] \leq e^{-c}$$

co implikuje ( $c = \log(1/\varepsilon)$ )

$$t_{\text{miks}}(\varepsilon) \leq n \log n + \log(1/\varepsilon)n$$

Współczynnik powyżej nie jest optymalny. W dalszej części wykładu pokażemy, że czas mieszania jest rzędu  $(1/2)n \log n$ .

### 3. SILNE CZASY STACJONARNE.

**Przykład 3.1.** Tasowanie kart metodą TOP-TO-RANDOM. Rozważmy następującą metodę tasowania talii  $n$  kart: bierzemy kartę z góry i wkładamy ją losowo (w sposób jednostajny) do talii, możliwych jest więc  $n$  pozycji. Powtarzając tę czynność odpowiednio wiele razy talia będzie potasowana. Naszym celem jest sprecyzowanie pojęcia 'potasowania' oraz obliczenie ile kroków należy wykonać. Możemy rozważać tę operację jako spacer losowy na grupie permutacji  $S_n$ , wtedy każdy krok polega na wylosowaniu odpowiedniej permutacji  $\sigma = \begin{pmatrix} 1 & 2 & 3 & \dots & k & k+1 & \dots & n \\ k & 1 & 2 & \dots & k-1 & k+1 & \dots & n \end{pmatrix}$ . Przypomnijmy, że grupa  $S_n$  składa się z  $n!$  elementów. Zatem z lematu 1.7 miarą stacjonarną jest miara jednostajna. Idealnie byłoby uzyskać w wyniku tasowania rozkład jednostajny. Będziemy chcieli jednak zbliżyć się do niego dowolnie blisko (względem normy  $\|\cdot\|_{\text{TV}}$ ).

Kiedy możemy uznać, że talia jest potasowana? Odpowiedź ukryta jest w czasie losowym  $\tau_{\text{top}}$  - jest to pierwszy moment, gdy karta, która początkowo była na spodzie talii, oznaczmy ją przez  $A$ , znalazła się na jej górze i jest wkładana losowo do talii. Jest to szczególny przypadek następującego wyniku:

**Lemat 3.1.** *Niech  $X_t$  będzie spacerem losowym na  $S_n$  odpowiadającym tasowaniu  $n$  kart metodą TOP-TO-RANDOM. Jeżeli w chwili  $t$  jest  $k$  kart pod  $A$ , kartą która początkowo była na dole, to wszystkie  $k!$  uporządkowań jest jednakowo prawdopodobne. Zatem rozkład  $X_{\tau_{\text{top}}}$  jest jednostajny*

*Dowód.* Dla  $t = 0$  nie ma kart pod  $A$  więc teza jest oczywista. Załóżmy, że teza zachodzi dla  $t$ . W chwili  $t + 1$  możliwe są 2 przypadki:

- karta z wierzchołka umieszczana jest nad  $A$ . Wówczas ułożenie kart pod  $A$  się nie zmienia.
- karta w wierzchołka umieszczana jest pod  $A$ . Wtedy każda z  $k + 1$  lokalizacji pod  $A$  jest jednakowo prawdopodobna. Zatem wszystkie ułożenia  $k + 1$  kart, które są pod  $A$  są jednakowo prawdopodobne.

□

**Lemat 3.2.** *Mamy*

$$\mathbb{E}\tau_{\text{top}} \sim n \log n$$

oraz

$$\mathbb{P}[\tau_{\text{top}} > n \log n + cn] \leq e^{-c}$$



*Dowód.* Popatrzmy ponownie na kartę  $A$  znajdującą się początkowo na dole talii i opiszmy jej kolejne ruchy. Zauważmy, że jeżeli karta ta znajduje się na samym dole, to prawdopodobieństwo, że aktualnie wkładana karta znajdzie się pod  $A$  wynosi  $1/n$ . Niech  $\tau_1$  oznacza więc moment, gdy pod  $A$  pojawi się pierwsza karta. Wówczas  $\tau_1$  ma rozkład geometryczny z prawdopodobieństwem sukcesu  $1/n$ . Podobnie, gdy pod  $A$  znajduje się już  $k - 1$  kart, niech  $\tau_k$  oznacza czas przebywania karty  $A$  na tej pozycji. Ponownie rozkład tej zmiennej losowej jest geometryczny, tym razem z prawdopodobieństwem sukcesu  $k/n$ . Zauważmy, że  $\tau_{\text{top}} = \tau_1 + \dots + \tau_n$  i  $\tau_{\text{top}}$  ma taki sam rozkład jak zmienna  $\tau$  w zadaniu o kolekcjonerze kuponów (lista 1). Lemat wynika więc z odpowiedniego zadania.  $\square$

W momencie  $\tau_{\text{top}}$  karty są już dobrze potasowane i jak łatwo widać w kolejnych krokach ich rozkład pozostaje jednostajny. Jest to moment losowy, ale jak pokażemy poniżej można go użyć do oszacowania deterministycznej wartości  $t_{\text{miks}}$ .

### 3.1. Czasy zatrzymania.

**Definicja 3.1.** Niech  $\{X_t\}$  będzie ciągiem zmiennych losowych o wartościach w  $\Omega$ . **Czasem zatrzymania** nazywamy zmienną losową  $\tau$  o wartościach w  $\{0, 1, \dots, \infty\}$ , taką, że zbiór  $\{\tau = t\}$  jest mierzalny względem  $\sigma$ -ciała generowanego przez  $X_0, \dots, X_t$ .

To jest formalna definicja. Wygodnie jest myśleć, że  $\tau$  zależy od wartości  $X_1, \dots, X_\tau$ , ale nie zależy już od kolejnych wartości spaceru.

Przykłady:

- Akcje firmy Bogdanka przekroczą poziom 95zł.
- Niech  $U \subset \Omega$ , wtedy pierwszy moment trafienia w zbiór  $U$

$$\tau_U = \min\{t \geq 0 : X_t \in U\}$$

jest czasem zatrzymania.

- Niech  $U \subset \Omega$ , wtedy ostatni moment trafienia w zbiór  $U$

$$\tau_U = \max\{t \geq 0 : X_t \in U\}$$

**nie** jest czasem zatrzymania.

- zmienna losowa  $\tau_{\text{top}}$  zdefiniowana w przykładzie 3.1 jest czasem zatrzymania. Wtedy  $\tau_{\text{top}} = \tau_U + 1$ , gdzie  $U$  jest zbiorem uporządkowań talii takim, że  $A$  jest na górze;

**Przykład 3.2. Leniwy spacer losowy na hiperkostce.** Niech  $X_t$  będzie leniwym spacerem losowym. Przypomnijmy (jak w przykładzie 2.6), że spacer ten może być skonstruowany następująco: wybieramy losowo element  $(j, B) \in \{1, 2, \dots, n\} \times \{0, 1\}$  i współrzędną  $j$  zastępujemy bitem  $B$ .

W tej konstrukcji spacer jest zdeterminowany przez ciąg  $Z_t = (j_t, B_t)$  iid. Zdefiniujmy

$$\tau_0 = \min\{t \geq 0 : \{j_1, \dots, j_t\} = \{1, \dots, n\}\},$$

jest to pierwszy czas, gdy wybrano wszystkie współrzędne ( $\tau_0 = \tau_{\text{par}}$  z przykładu 2.6). W tym momencie w sposób losowy zostały zamienione wszystkie współrzędne, więc  $X_{\tau_0}$  ma już rozkład jednostajny (to odpowiada zmiennej  $\tau_{\text{top}}$  z przykładu 3.1 oraz  $\tau$  z problemu kolekcjonera kuponów, lista 1).

Zauważmy, że  $\tau_0$  jest czasem zatrzymania dla  $\{Z_t\}$ , ale nie jest czasem zatrzymania dla  $\{X_t\}$  (!). Tzn. obserwacja wyłącznie wartości  $X_t$  nie pozwala na wyznaczenie  $\tau_0$ . Jest to funkcja ciągu  $Z_n$ .  $\tau_0$  jest nazywane zrandomizowanym czasem zatrzymania dla ciągu  $X_t$ .

Formalna definicja jest następująca:

**Definicja 3.2.** Niech  $P$  będzie macierzą przejścia. Wówczas  $P$  ma **reprezentację w postaci odwzorowań losowych** (ang. random mapping representation), tzn. istnieje ciąg  $\{Z_t\}$  iid (nie specyfikujemy tutaj zbioru wartości) oraz funkcja deterministyczna  $f$  takie, że

$$X_0 = x, \quad X_t = f(X_{t-1}, Z_t)$$

jest łańcuchem Markowa z macierzą przejścia  $P$  startującym w  $x$  ([LPW], Proposition 1.5). Czas losowy  $\tau$  jest nazywany **zrandomizowanym czasem zatrzymania** dla łańcucha  $\{X_t\}$ , jeżeli jest czasem zatrzymania dla ciągu  $\{Z_t\}$ .

**Definicja 3.3. Czasem stacjonarnym** (ang. stationary time) dla łańcucha Markowa  $X_n$  z miarą stacjonarną  $\pi$  nazywamy zrandomizowany czas zatrzymania  $\tau$  taki, że

$$(3.3) \quad \mathbb{P}_x[X_\tau = y] = \pi(y).$$

Zatem  $X_\tau$  ma rozkład  $\pi$  i w kolejnych krokach proces będzie miał ten sam rozkład.

Potrzebujemy mocniejszej definicji:

**Definicja 3.4. Silnym czasem stacjonarnym** (ang. strong stationary time) dla łańcucha Markowa  $X_n$  z miarą stacjonarną  $\pi$  nazywamy zrandomizowany czas zatrzymania  $\tau$  taki, że

$$(3.4) \quad \mathbb{P}_x[\tau = t, X_\tau = y] = \mathbb{P}_x[\tau = t]\pi(y).$$

Zatem  $X_\tau$  ma rozkład  $\pi$  i w kolejnych krokach proces będzie miał ten sam rozkład.

Czasy losowe zdefiniowane powyżej dla TOP-TO-RANDOM i spaceru na hiperkostce są silnymi czasami stacjonarnymi.

Kluczowy dla nas jest następujący wynik:

**Twierdzenie 3.5.** *Jeżeli  $\tau$  jest silnym czasem stacjonarnym, to*

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq \max_{x \in \Omega} \mathbb{P}_x[\tau > t].$$

*Dowód.* Piszemy

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{\text{TV}} &= \sum_{\{y: P^t(x, y) < \pi(y)\}} (\pi(y) - P^t(x, y)) \\ &= \sum_{\{y: P^t(x, y) < \pi(y)\}} \pi(y) \left(1 - \frac{P^t(x, y)}{\pi(y)}\right) \\ &\leq \max_y \left(1 - \frac{P^t(x, y)}{\pi(y)}\right) \end{aligned}$$

Oznaczmy powyższą wartość przez  $s_x(t)$  (jest ona nazywana w literaturze 'separation distance').

Z (3.4) wynika (to wymaga pewnych rachunków, które pominiemy)

$$\mathbb{P}_x[\tau \leq t, X_t = y] = \mathbb{P}_x[\tau \leq t]\pi(y).$$

Z powyższej równości dla dowolnego  $y \in \Omega$  mamy

$$\begin{aligned} 1 - \frac{P^t(x, y)}{\pi(y)} &= 1 - \frac{\mathbb{P}_x[X_t = y]}{\pi(y)} \\ &\leq 1 - \frac{\mathbb{P}_x[X_t = y, \tau \leq t]}{\pi(y)} \\ &= 1 - \frac{\pi(y)\mathbb{P}_x[\tau \leq t]}{\pi(y)} \\ &= \mathbb{P}_x[\tau > t]. \end{aligned}$$

Zatem dla każdego  $x \in \Omega$

$$\|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq s_x(t) \leq \mathbb{P}_x[\tau > t].$$

□

**Wniosek 3.6.** *Dla leniwego spaceru na hiperkostce oraz problemu TOP-TO-RANDOM, mamy*

$$t_{\text{miks}}(\varepsilon) \leq n \log n + \log(\varepsilon^{-1})n.$$

#### 4. RIFFLE SHUFFLE

RIFFLE SHUFFLE jest najbardziej popularną metodą tasowania kart. Metoda zazwyczaj używana jest do potasowania talii 52 kart. Osoba tasująca dzieli stos kart na dwa (w sposób losowy, ale stara się aby oba zbiory miały w przybliżeniu podobną liczbę kart). Następnie oba stosy kart są wspólnie 'przekartkowane'. Trzymając je w obu rękach osoba tasująca opuszcza karty (liczba kart jest zmienna i też losowa) z obu stosów tak, aby karty 'wchodziły' pomiędzy siebie. Celem tego rozdziału jest odpowiedź na pytanie: ile razy należy wykonać tę procedurę, aby uznać talię kart za potasowaną?

Zanim przejdziemy do analizy warto zrozumieć jaki jest sens dokładnego tasowania kart. Odpowiedź jest zawarta w poniższej tabelce.<sup>5</sup> Przedstawia ona doświadczenie wykonane przez Bergera po koniec lat sześćdziesiątych. Przeanalizował on rozkłady kart, które otrzymał jeden z graczy (na pozycji S) podczas szeregu turniejów brydżowych. W pierwszej kolumnie są oczekiwane rozkłady, w drugiej rozkłady po tasowaniu przez komputer, a w trzeciej po tasowaniu przez człowieka. Jak widać trzecia kolumna wyraźnie odbiega od oczekiwań (nie przechodzi też typowych testów statystycznych). Zbyt często pojawiają się zrównoważone rozkłady rąk (z punktu widzenia graczy, mniej ciekawe).

<sup>5</sup>Tabela pochodzi z książki Diaconisa [D], strona 78.

Table 1  
Frequency of Computer-dealt Hands Versus Theoretical  
Expected Frequencies from Berger (1973)

Distribution of the 4 suits	Expected Frequencies	Actual Frequencies of Computer-dealt Hands	Actual Frequencies of Man-dealt Hands
4,4,3,2*	216	198	241
5,3,3,2	155	160	172
5,4,3,1	129	116	124
5,4,2,2	106	92	105
4,3,3,3	105	103	129
6,3,2,2	56	64	46
6,4,2,1	47	53	36
6,3,3,1	34	40	41
5,5,2,1	32	40	19
4,4,4,1	30	35	25
7,3,2,1 and others	90	99	62
	1,000	1,000	1,000

\* by "4,4,3,2" we mean that the thirteen cards contained 4 cards in one suit, 4 cards in another suit, 3 cards in another suit, and 2 cards in the remaining suit.

4.1. **Model GSR (Gilbert, Shannon, Reeds)**. Badanie RIFFLE SHUFFLE wymaga ścisłego matematycznego modelu, którym można opisać tę metodę tasowania. Bada się następujący model tasowania talii  $n$  kart, opisany na trzy równoważne sposoby:

- (1) Niech  $k$  będzie zmienną losową o rozkładzie dwumianowym z parametrami  $n, 1/2$  (można myśleć, że  $k$  jest liczbą orłów w  $n$ -krotnym rzucie monetą). Dzielimy karty na dwa stosy: górne  $k$  kart tworzy pierwszy z nich, a pozostałe  $n - k$  drugi. Następnie wykonujemy tasowanie, musimy więc zmieszać oba zbiory zachowując jednak uporządkowanie wewnątrz nich. Możemy to zrobić na  $\binom{n}{k}$  sposobów, tzn. na  $n$  pozycjach wybieramy  $k$  miejsc, gdzie umieszczamy pierwszy stos (możemy to zrobić tylko na jeden sposób, gdyż musimy zachować jego uporządkowanie), a na pozostałe  $n - k$  miejsc wkładamy pozostałe karty. Każde takie tasowanie wybieramy losowo w sposób jednostajny.
- (2) Niech  $k$  będzie zmienną losową o rozkładzie dwumianowym z parametrami  $n, 1/2$ . Dzielimy karty na dwa stosy: górne  $k$  kart tworzy pierwszy z nich, a pozostałe  $n - k$  drugi. Następnie wykonujemy tasowanie. Bierzymy jeden stos do lewej ręki, drugi do prawej. Upuszczamy kolejno karty, losowo, z lewej lub prawej ręki z prawdopodobieństwem proporcjonalnym do rozmiaru zbioru. To znaczy, jeżeli w lewej ręce pozostało nam  $L$  kart, a w prawej  $P$  kart, to kartę z lewej ręki upuszczamy z prawdopodobieństwem  $L/(L + P)$ , a z prawej z prawdopodobieństwem  $P/(L + P)$ . Powtarzamy tę czynność tak długo, aż upuścimy wszystkie karty.
- (3) (Tasowanie odwrotne). Opiszemy jak wykonać czynność odwrotną do RIFFLE SHUFFLE. Oznaczmy tył każdej karty przez 0 lub 1 w zależności od wyniku rzutu monetą (dla każdej karty wykonujemy niezależnie rzut). Wyjmijmy z talii wszystkie karty oznaczone zerem i połóżmy je na górze talii, zachowując ich uporządkowanie

**Lemat 4.1.** *Powyższe opisy (1), (2) i (3) są równoważne, tzn. wygenerowana permutacja ma ten sam rozkład.*

*Dowód.* Modele (1) i (3) są równoważne. Istotnie, zauważmy, że liczba zer ma rozkład dwumianowy z parametrami  $n, 1/2$ . Ponadto każdy rozkład zer jest jednakowo prawdopodobny.

Modele (1) i (2) są równoważne. Oznaczmy przez  $L_1, \dots, L_k$  karty z pierwszego stosu, a przez  $P_1, \dots, P_{n-k}$  karty z drugiego stosu. Obliczmy prawdopodobieństwo wylosowania konkretnej permutacji, np.  $L_1 L_2 P_1 L_3 P_2 \dots P_{n-k} L_k$ :

$$\frac{k}{n} \cdot \frac{k-1}{n-k-1} \cdot \frac{n}{n-k-2} \cdot \frac{k-2}{n-k-3} \cdot \frac{n-1}{n-k-4} \cdot \dots \cdot \frac{1}{2} \cdot 1 = \frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}}.$$

Dla każdej permutacji otrzymujemy ten sam wynik, więc każde ułożenie kart jest jednakowo prawdopodobne.  $\square$

**Lemat 4.2.** *Niech  $X_t$  będzie spacerem losowym na grupie  $G$  generowanym przez miarę probabilistyczną  $\mu$ . Zdefiniujmy miarę  $\tilde{\mu}(g) = \mu(g^{-1})$  na  $G$  i niech  $\tilde{X}_t$  będzie spacerem losowym generowanym przez  $\tilde{\mu}$ . Oznaczmy przez  $\mu_t$  i  $\tilde{\mu}_t$  rozkłady  $X_t$  i  $\tilde{X}_t$ . Wówczas*

$$\|\mu_t - U\|_{\text{TV}} = \|\tilde{\mu}_t - U\|_{\text{TV}},$$

gdzie  $U$  jest miarą jednostajną na grupie  $G$ .

Z lematu wynika, że wystarczy badać spacer losowy generowany przez odwrotne permutacje. Skonstruujemy silny czas stacjonarny. W tym celu budujemy losową macierz o  $n$  wierszach. Wiersze numerowane są kolejnymi kartami (każda karta ma na stałe przyporządkowany jeden z wierszy). Elementami macierzy są 0 i 1. W pierwszej kolumnie zapisujemy liczbę przyporządkowaną danej karcie w pierwszym odwrotnym tasowaniu. Czynność powtarzamy, tzn. w drugiej kolumnie zapisujemy wynik przyporządkowany w drugim tasowaniu itd.

**Lemat 4.3.** *Niech  $\tau_{\text{RS}}$  oznacza pierwszy czas, w którym macierz binarna opisana powyżej zawiera parami różne rzędy. Wówczas  $\tau_{\text{RS}}$  jest silnym czasem stacjonarnym.*

*Dowód.* Po pierwszym tasowaniu karty, które mają 0 są przeniesione na górę, a karty z 1 na dół. Po drugim losowaniu na górę przenoszone są karty, które mają 0 w drugiej kolumnie, ale zachowując ich wcześniejsze uporządkowanie. Tak więc na samej górze są karty, które w dwóch pierwszych wierszach mają 00, potem 10, 01, 11. Po trzecim losowaniu kolejno: 000, 100, 010, 110, 001, 101, 011, 111, itd. W poszczególnych grupach karty zachowują jednak swoje początkowe uporządkowanie. Zauważmy, że odwrotne tasowanie sortuje wg. odwrotnego porządku leksykograficznego.

W momencie  $\tau_{\text{RS}}$  wszystkim kartom zostały przyporządkowane **różne** wektory o długości  $\tau_{\text{RS}}$ . Więc zostały one posortowane w sposób losowy niezależnie od ich początkowego ułożenia. Jednocześnie, ponieważ przyporządkowanie wektora karcie jest losowe, każda ich permutacja jest jednakowo prawdopodobna.  $\square$

Zgodnie z twierdzeniem 3.5 musimy więc oszacować  $\mathbb{P}[\tau_{\text{RS}} > t]$ :

**Twierdzenie 4.4.** *Dla modelu RIFFLE SHUFFLE mamy*

$$\|\mu^t - U\|_{\text{TV}} \leq \mathbb{P}[\tau_{\text{RS}} > t] = 1 - \prod_{i=1}^{n-1} \left(1 - \frac{i}{2^t}\right).$$

*Ponadto*

$$t_{\text{miks}} \leq 2 \log_2(4n/3).$$

*Dowód.* Obliczymy wartość  $\mathbb{P}[\tau_{\text{RS}} \leq t]$ . W czasie  $t$  każdej z kart przyporządkowano losowy wektor 0-1 o długości  $t$  i chcemy policzyć prawdopodobieństwo, że wektory te są parami różne. Zauważmy, że jest to dokładnie problem urodzin (oblicz prawdopodobieństwo, że w grupie  $n$  osób nie ma dwóch osób obchodzących urodziny tego samego dnia). Zatem

$$\mathbb{P}[\tau_{\text{RS}} \leq t] = \prod_{i=1}^{n-1} \left(1 - \frac{i}{2^t}\right).$$

Powyższa formułka jest niewygodna w obliczeniach. Można ją uprościć. Niech  $t = 2 \log_2(n/c)$  dla pewnej stałej  $c$ . Wtedy  $2^t = n^2/c^2$  i mamy (przypomnijmy  $\log(1+x) = x + O(x^2)$ )

$$\begin{aligned} \log \prod_{i=1}^{n-1} \left(1 - \frac{i}{2^t}\right) &= - \sum_{i=1}^{n-1} \left(\frac{c^2 i}{n^2} + O\left(\frac{i}{n^2}\right)^2\right) \\ &= - \frac{c^2 n(n-1)}{2n^2} + O\left(\frac{n^3}{n^4}\right) = -\frac{c^2}{2} + O(1/n) \end{aligned}$$

Zatem

$$\mathbb{P}[\tau_{\text{RS}} \leq t] \sim e^{-c^2/2}, \quad \text{gdy } n \rightarrow \infty.$$

Przyjmując  $c = 3/4$  otrzymujemy  $1 - e^{-c^2/2} \approx 0,245$ . Zatem z Twierdzenia 3.5 otrzymujemy dowód.  $\square$

Korzystając z powyższych wzorów można oszacować  $\|\mu^t - U\|_{\text{TV}}$  dla  $n = 52$

$t$	10	11	12	13	14
górne szacowanie	0,73	0,48	0,28	0,15	0,08

Z powyższych danych wynika więc,  $t_{\text{miks}} \leq 13$ . Dostyc łatwo jest dostać dolne szacowania. Mianowicie zauważmy, że są 32 wektory 0-1 o długości 5 i 64 o długości 6. Dla  $n = 52$ , 5 tasowań to zdecydowanie za mało, podobnie 6 wydaje się być niewystarczające. Ogólnie

**Lemat 4.5.** *Ustalmy  $0 < \varepsilon, \delta < 1$ . Rozważmy RIFFLE SHUFFLE talii  $n$  kart. Wówczas dla dużych  $n$*

$$t_{\text{miks}}(\varepsilon) \geq (1 - \delta) \log_2 n$$

*Dowód.* Z zadania 1, lista 3 wiemy, że

$$t_{\text{miks}}(\varepsilon) \geq \frac{\log(|\Omega|(1 - \varepsilon))}{\log \Delta},$$

gdzie  $\Delta = \max_{x \in \Omega} |\{y : P(x, y) > 0\}|$ . W naszym przypadku  $|\Omega| = n!$ ,  $\Delta = 2^n$  (w jednym kroku generujemy wektor 0-1 o długości  $n$ , który jednoznacznie wyznacza permutację). Przypomnijmy formułę Stirlinga

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}.$$

Stąd

$$t_{\text{miks}}(\varepsilon) \geq \frac{\log(|\Omega|(1 - \varepsilon))}{\log \Delta} \sim \frac{\log n!}{\log 2^n} \sim \log_2 n$$

$\square$

Policzmy:  $\log_2 52 \approx 5,7$ .

**4.2. Twierdzenie Bayera i Diaconisa.** Bayer i Diaconis [BD] otrzymali znacznie dokładniejszy wynik

**Twierdzenie 4.6.** *Jeżeli  $n$  kart jest tasowane  $t$  razy metodą RIFFLE SHUFFLE i  $t = 3/2 \log_2 n + c$ , to dla dużych  $n$*

$$\|\mu_t - U\|_{\text{TV}} = 1 - 2\Phi\left(\frac{-2^{-c}}{4\sqrt{3}}\right) + O\left(\frac{1}{n^{1/4}}\right),$$

gdzie  $\Phi$  jest dystrubuantą rozkładu normalnego

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

Powyższy wynik pozwala dokładnie obliczyć odległość dla  $n = 52$

$t$	4	5	6	7	8	8	10
$\ \mu_t - U\ _{\text{TV}}$	1,00	0,92	0,61	0,33	0,16	0,08	0,04

Zauważmy, że dla wartości 4,5,6 odległość jest duża, a potem zaczyna gwałtownie spadać. Z każdym kolejnym krokiem zmniejsza się dwukrotnie. Przyjmuje się, że **7 tasowań wystarczy** do potasowania talii 52 kart.

**4.3. Szkic dowodu twierdzenia 4.6.** Przypatrzmy się wyglądają permutacje losowane podczas jednego wykonania tasowania RIFFLE SHUFFLE. Dla przykładu, załóżmy, że karty ułożone są rosnąco 12345678. Dzielimy je na dwa zbiory 123 i 45678, a następnie łączymy otrzymując np. 4<sub>12</sub>567<sub>3</sub>8. Zauważmy, że permutacja zawiera 2 ciągi rosnące (przez ciąg rosnący rozumiemy ciąg, w którym kolejne elementy składają się z kolejnych liczb naturalnych). Tasując metodą RIFFLE SHUFFLE, możemy otrzymać jedynie identyczność, lub permutację zawierającą dokładnie 2 ciągi rosnące.

Ta obserwacja została poczyniona już na początku XX wieku, przez magików. Wykonywali oni następującą sztuczkę. Wręczali talię kart komuś z publiczności i prosili, aby ta osoba potasowała trzykrotnie karty metodą RIFFLE SHUFFLE. Następnie osoba tasująca brała kartę z wierzchu oglądała ją i wkładała w losowe miejsce w talii. Wprawny magik pozwalał na kolejne potasowanie kart. Następnie talia kart wracała do magika, który z dużym prawdopodobieństwem potrafił określić która karta została włożona do wnętrza talii. Wyjaśnienie sztuczki polega na tym, że po trzech tasowaniach w talii powstaje zazwyczaj 8 ciągów rosnących. Włożenie karty do wnętrza talii powoduje zazwyczaj powstanie kolejnego ciągu złożonego z tej jednej karty. Sprawne oko potrafi ja szybko wyłapać.

Zdefiniujmy  $a$ -RIFFLE SHUFFLE. Najwygodniej myśleć, że osoba tasująca używa  $a$ -rąk. Dzielimy talię  $n$  kart na  $a$  zbiorów w sposób losowy, składające się odpowiednio z  $k_1, k_2, \dots, k_a$  kart ( $k_1 + \dots + k_a = n$ ) z prawdopodobieństwem  $\binom{n}{k_1, k_2, \dots, k_a}$ . Następnie łączymy je w jeden zbiór tak, aby każde połączenie, zachowujące wewnętrzne uporządkowanie zbiorów było jednakowo prawdopodobne. Równoważnie ... (zadanie!).

Pokazuje się, że (zadanie)

- wykonanie  $a$ -RIFFLE SHUFFLE, a następnie  $b$ -RIFFLE SHUFFLE jest równoważne tasowaniu  $ab$ -RIFFLE SHUFFLE.
- $Q_a(\pi) = \binom{n+a-r(\pi)}{n} / a^n$ , gdzie  $r(\pi)$  jest liczbą rosnących ciągów w  $\pi$ , a  $Q_a$  jest rozkładem  $S_n$  po jednym tasowaniu  $a$ -RIFFLE SHUFFLE.

Zatem

$$\|\mu_t - U\|_{\text{TV}} = \frac{1}{2} \sum_{j=1}^n k_n(j) \left| \frac{\binom{n+2^t-j}{n}}{2^{tn}} - \frac{1}{n!} \right|,$$

gdzie  $k_n(j)$  jest liczbą permutacji z  $j$  rosnącymi ciągami. Okazuje się, że te liczby są dosyć dobrze zbadane i są współczynnikami wielomianów Euleriana...

**4.4. 7 tasowań nie wystarczy.** Peter Doyle zaproponował następujące doświadczenie. Załóżmy, że karty są ponumerowane  $R1, \dots, R26, B1, \dots, B26$  oraz ułożone są w talii w następującej kolejności  $R1, R2, \dots, R26, B26, B25, \dots, B1$ . Wykonajmy 7 razy tasowanie metodą RIFFLE SHUFFLE. Następnie po kolei przeglądamy karty. Jeżeli napotkamy na kartę  $R1$  kładziemy ją na stos  $R$ , jeżeli  $B1$ , na stos  $B$ . Jeżeli wyciągnięta karta ma numer o jeden większy niż odkryta karta w jej kolorze, to kładziemy ją na odpowiedni stos, w przeciwnym razie odkładamy ją na dół talii. Kończymy grę, gdy jeden ze stosów jest pełny, tzn. składa się z 26 kart. Powiedzmy, że wygramy grę, jeżeli pełny jest stos  $R$ . Gdyby karty zostały dobrze potasowane, to prawdopodobieństwo wygrania wynosiłoby  $1/2$ . Tymczasem okazuje się (poprzez komputerowe symulacje), że wygrywamy w 81%. Wyjaśnienie jest następujące. Czerwone karty były ułożone zgodnie z ich kolejnością i ta struktura nie została rozbita w ciągu 7 tasowań.

Ta gra realizuje w pewnym sensie najgorszy możliwy scenariusz. Niech  $S$  oznacza zbiór tych permutacji, po których wylosowaniu wygrywamy. Wtedy wiemy, że

$$0,33 \approx \|\mu_7 - U\|_{\text{TV}} \geq |\mu_7(S) - U(S)| \approx |0,81 - 0,5| = 0,31.$$

## 5. MCMC

W tym rozdziale będziemy badać następujący problem: mając dany skończony zbiór  $S$  oraz miarę probabilistyczną  $\pi$  na  $S$  chcemy wygenerować zmienną losową o rozkładzie  $\pi$ . Interesować nas będą duże skończone zbiory. Często nie będziemy nawet potrafili określić ich liczności ani nawet precyzyjnie podać miary  $\pi$ . Przedstawimy kilka przykładów bazujących na metodzie MCMC (Markov Chain Monte Carlo). W tym celu na zbiorze  $S$  wprowadzimy strukturę grafu i zdefiniujemy odpowiedni spacer losowy z miarą stacjonarną  $\pi$ . Spacery będą nieredukowalne i aperiodyczne. Wówczas z twierdzenia 2.4 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, ...

### 5.1. Drzewo rozpinające.

**Przykład 5.1. 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  $S$  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  $S$ , a następnie zdefiniowaniu odpowiedniego spaceru losowego. Zbiorem wierzchołków jest zbiór  $S$ , tak więc wierzchołkiem jest drzewo rozpinające. Następnie definiujemy łańcuch Markowa  $\{X_t\}$  na  $S$ ,



w którym określimy też krawędzie w grafie. Startujemy z dowolnego elementu  $S$ . Załóżmy, że znamy wartość  $X_t$ , wówczas generujemy  $X_{t+1}$  następująco:

- (1) Wylosujemy jednostajnie krawędź  $e$ , która nie jest w  $X_t$ . Takich krawędzi jest  $|E| - (|V| - 1)$ .
- (2) Niech  $C$  będzie jedynym cyklem w  $X_t \cup \{e\}$ .
- (3) Wybierzmy jednostajnie krawędź  $e'$  w  $C$ .
- (4) Przyjmijmy  $X_{t+1} := (X_t \cup \{e\}) \setminus \{e'\}$ .

Ćwiczenie: Narysuj przykładowy graf i wykonaj kilka pierwszych kroków powyższego algorytmu.

**Lemat 5.1.** *Powyższy łańcuch Markowa jest aperiodyczny, nieredukowalny i odwracalny.*

*Dowód.* Proces jest aperiodyczny, bo  $P(T, T) > 0$  dla każdego  $T \in S$ , 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

$$\begin{aligned} P(T, T') &= \mathbb{P}[e = e_1, e' = e_2 | X_t = T] \\ &= \mathbb{P}[e = e_1 | X_t = T] \cdot \mathbb{P}[e' = e_2 | X_t = T, e = e_1] \\ &= \frac{1}{|E| - (|V| - 1)} \cdot \frac{1}{|C|} \end{aligned}$$

Analogicznie, ponieważ  $T = (T' \cup \{e_2\}) \setminus \{e_1\}$ , mamy

$$\begin{aligned} P(T', T) &= \mathbb{P}[e = e_2, e' = e_1 | X_t = T'] \\ &= \frac{1}{|E| - (|V| - 1)} \cdot \frac{1}{|C|} \end{aligned}$$

Zatem

$$P(T, T') = P(T', T)$$

przy założeniu, że  $P(T, T') > 0$ . Przypadek  $P(T, T') = 0$  pomijamy □

Z powyższego lematu oraz z lematu 1.6, wynika że miara stacjonarna  $\{X_t\}$  jest jednostajna. Z twierdzenia 2.4 rozkłady  $X_t$  zbiegają do jednostajnego i można przyjąć, że  $X_t$  jest bliski rozkładowi jednostajnemu dla dużych wartości  $t$ .

## 5.2. Problem pakowania plecaka (The knapsack problem).

**Przykład 5.2.** 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  $S$  oznaczamy zbiór dopuszczalnych pakowań plecaka, tzn.

$$S = \{z \in \{0, 1\}^m : (w, z) \leq b\}$$

Naszym celem jest zoptymalizowanie wartości  $(v, z)$  pod warunkiem  $z \in S$ .

### Algorytm 1:

- (1) Losujemy jednostajnie  $z \in \{0, 1\}^m$ .
- (2) Jeżeli  $z \in S$ , 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  $S$ . 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  $S$ . Mając zadany  $X_t = (z_1, \dots, z_m)$

- (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  $z' \in S$ , to podstawiamy  $X_{t+1} = z'$ , w przeciwnym razie  $X_{t+1} = X_t = z$ .

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 5.2.** *Jeżeli  $\sum w_i > b$ , to powyższy łańcuch Markowa jest aperiodyczny, nieredukowalny i odwracalny.*

*Dowód.* Zauważmy, że z dowolnego stanu możemy dojść do  $(0, \dots, 0)$ , wyjmując po kolei wszystkie przedmioty. Podobnie możemy dojść do każdego elementu  $S$ , to pokazuje nieredukowalność. Aperiodyczność wynika z obserwacji, że istnieje stan, w którym nie możemy dołożyć już kolejnego przedmiotu. Zgodnie z powyższym algorytmem z dodatnim prawdopodobieństwem więc w nim pozostaniemy. Zatem okres tego stanu wynosi 1.

Odwracalność wynika, podobnie jak wcześniej, z faktu, że łańcuch jest symetryczny, a to z kolei jest konsekwencją, że jeżeli  $x, y \in S$  są sąsiadami, to  $P(x, y) = P(y, x) = 1/m$ .  $\square$

Uwagi:

- Miarą stacjonarną dla tego łańcucha Markowa jest miara jednostajna, a więc dla dużych  $t$ ,  $\mathbb{P}[X_t = z] \sim 1/|S|$ . Umiemy zatem aproksymować tę miarę, mimo iż nie znamy dokładnej wartości  $|S|$
- 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.

**5.3. Algorytm Metropolis.** Algorytm pozwala na przybliżenie zmiennej losowej ze zbioru  $S$  o rozkładzie  $\pi$ . W tym celu konstruujemy graf  $G$  ze zbiorem wierzchołków  $S$ . 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  $x \in S$ , 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 5.3.** *Miarą stacjonarną powyższego łańcucha Markowa jest  $\pi$ .*

*Dowód.* Wystarczy pokazać, że spacer losowy jest odwracalny względem  $\pi$ , tzn. spełniony jest warunek

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \forall x, y.$$

Oczywiście wystarczy sprawdzić wyłącznie dla  $x \neq y$  połączonych krawędzią. Załóżmy, że  $\frac{\pi(y)d(x)}{\pi(x)d(y)} \geq 1$ . Wówczas

$$\pi(x)P(x, y) = \frac{\pi(x)}{d(x)}$$

oraz

$$\pi(y)P(y, x) = \pi(y) \cdot \frac{1}{d(y)} \frac{\pi(x)d(y)}{\pi(y)d(x)} = \frac{\pi(x)}{d(x)}$$

Jeżeli  $\frac{\pi(y)d(x)}{\pi(x)d(y)} < 1$ , analogicznie pokazuje się, że

$$\pi(x)P(x, y) = \frac{\pi(x)}{d(x)} = \frac{\pi(y)}{d(y)}.$$

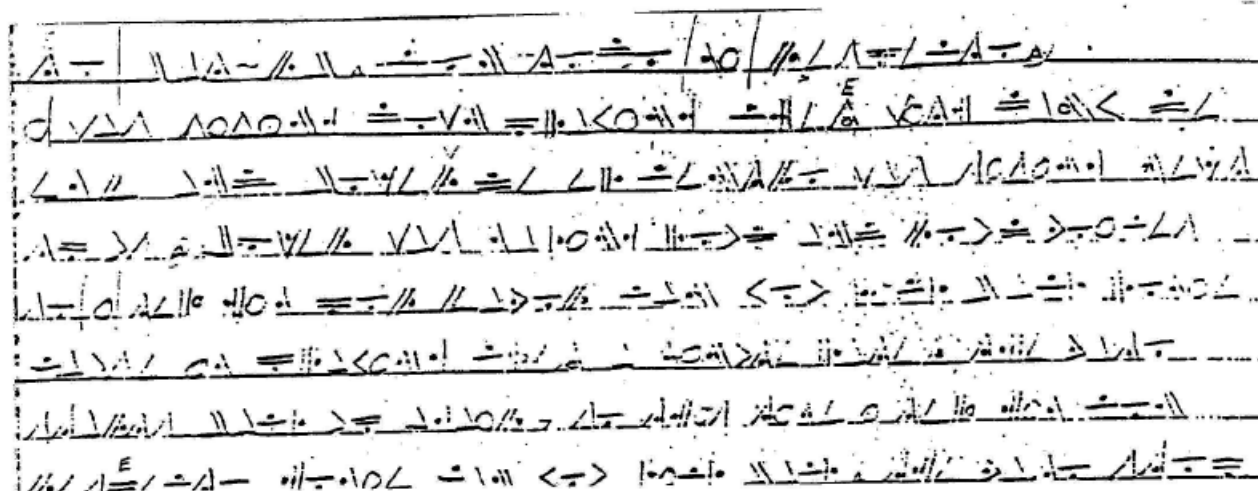
□

**Przykład 5.3. Problem pakowania plecaka.** Rozważmy 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 5.4.** Persi Diaconis w pracy *The Markov Chain Monte Carlo Revolution* opisuje następujący przykład. Pewnego dnia psycholog ze stanowego więzienia poprosił Department Statystyki na Uniwersytecie w Stanford o rozszyfrowanie następującego kodu.<sup>6</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 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**

<sup>6</sup>Wszystkie grafiki w tym przykładzie są zaczerpnięte z pracy P. Diaconisa *The Markov Chain Monte Carlo Revolution*

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 ITOSUVDEL
300 ES ELOHANDE OHANO UOVEOULOSU O UOVEO HA OITO HEOQRET IUSOFHE HELO ITOSUVDEL
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

```

5.4. Symulowane wyżarzanie (Simulated Annealing). Problemy pakowania plecaka, odkodowania szyfrów, komiwojażera są szczególnymi przypadkami następującego problemu optymalizacyjnego: Mając zadany zbiór  $S = \{s_1, \dots, s_k\}$  oraz funkcję  $f : S \mapsto \mathbb{R}$  należy znaleźć  $s_i \in S$ , maksymalizujący (lub minimalizujący)  $f(s)$ .

Rozważa się modyfikację algorytmu Metropolis nazywaną *symulowanym wyżarzaniem* (ang. simulated annealing). Obserwując wykonanie algorytmu Metroplis można znaleźć wartości bliskie optymalnym. Jeżeli interesuje nas jednak nie losowy element o zadanym prawdopodobieństwie, a deterministyczna liczba. Można zwiększyć wagę elementów o dużej

wartości  $f(s)$  i rozważać  $\pi(s) = e^{\beta f(s)}$ . Ponadto  $\beta$  może zmieniać się w czasie. Dokładniej zamiast  $\beta$  rozważa się  $\beta(t)$ , funkcję która bardzo wolno rośnie w czasie, np.  $\beta(t) = \log t$ ,  $\beta(t) = (1,001)^t$  i wtedy rozważamy niejednorodny łańcuch Markowa zdefiniowany następująco

$$\mathbb{P}[X_t = x, X_{t+1} = y] = \begin{cases} \frac{1}{d(x)} \min \left\{ \frac{e^{\beta(t)f(y)d(x)}}{e^{\beta(t)f(x)d(y)}}, 1 \right\} & \text{jeżeli } x \sim y; \\ 0 & \text{jeżeli } x \neq y \text{ oraz } x \not\sim y; \\ 1 - \sum_{z \sim x} \frac{1}{d(x)} \min \left\{ \frac{e^{\beta(t)f(z)d(x)}}{e^{\beta(t)f(x)d(z)}}, 1 \right\} & \text{jeżeli } x = y. \end{cases}$$

Przypomnijmy, że dla stałego  $\beta$  miarą stacjonarną procesu jest

$$\pi^\beta(x) = C_\beta^{-1} e^{\beta f(x)},$$

dla pewnej (zazwyczaj nieznannej) normalizującej stałej  $C_\beta$ . Jeżeli  $\beta$  zbiega do  $\infty$ , to miara koncentruje się na optymalnych rozwiązaniach. Dokładniej, niech  $S_{\max} = \{s : s = M\}$  dla  $M = \max_{r \in S} f(r)$ . Wtedy

$$\lim_{\beta \rightarrow \infty} e^{-\beta M} \sum e^{\beta f(s)} = |S_{\max}|$$

i stąd

$$\lim_{\beta \rightarrow \infty} \pi(\beta)(s) = \begin{cases} |S_{\max}|^{-1} & \text{dla } s \in S_{\max} \\ 0 & \text{dla } s \notin S_{\max} \end{cases}$$

Powyższe rachunki tłumaczą ideę algorytmu. Zwiększając  $\beta$  proces powinien spędzać większość czasu blisko optymalnych rozwiązań. Istnieją jednak potencjalne zagrożenia:

- Jeżeli  $\beta$  jest duże, to łańcuch Markowa może bardzo wolno zbiegać do miary stacjonarnej. Wówczas większość stanów ma bardzo małe wagi i proces może nie rozumieć, w którą stronę powinien iść, aby zmierzać do optymalnego rozwiązania.
- Jeżeli  $\beta$  rośnie za szybko, to proces może zablokować się w lokalnych maksimach.

Powyższy algorytm dosyć dobrze radzi sobie w praktyce, jednak jego istnieje niewiele teoretycznych wyników na jego temat.

**5.5. MCMC - dynamika Glaubera.** W tym rozdziale przedstawimy inną metodę generowania zmiennej losowej o zadanym rozkładzie, w sytuacji gdy przestrzeń stanów jest postaci  $S^V$ , gdzie  $V$  jest zbiorem wierzchołków, a  $S$  skończonym zbiorem. Elementy  $S^V$  są nazywane konfiguracjami i są to funkcje na  $V$  o wartościach w  $S$ .

**Przykład 5.5. Hardcore model.** Dla zadanego grafu  $G = (V, E)$  przyporządkowujemy jego wierzchołkom wartości 0 i 1 w sposób losowy, tak aby dwa sąsiednie wierzchołki nie przyjmowały wartości 1. Takie przyporządkowanie nazywamy dopuszczalną konfiguracją. Ten model pojawia się np. w fizyce statystycznej, gdzie na trójwymiarowej kratce modeluje się dyskretny ruch cząsteczek gazu. Wówczas 1 oznacza pozycję cząsteczki, a 0 puste miejsce. Zakłada się tutaj, że cząsteczki są kulami o pewnej objętości i dwie cząstki nie mieszczą się na sąsiednich polach.

Przez  $\pi_G$  oznaczmy miarę jednostajną na wszystkich dopuszczalnych konfiguracjach. Chcemy wygenerować zmienną losową o rozkładzie  $\pi_G$ . Jako motywację można podać naturalne pytanie o oczekiwaną liczbę jedynek w losowej dopuszczalnej konfiguracji, względem miary  $\pi_G$ . Można przybliżyć tę liczbę symulując wielokrotnie losowe dopuszczalne konfiguracje, a następnie licząc liczbę ich jedynek i uśredniając. Mocne prawo wielkich liczb gwarantuje, że średnia dąży do szukanej wartości oczekiwanej.

Rozważmy następujący algorytm. W czasie  $t + 1$

- Losujemy jednostajnie jeden z wierzchołków  $v \in V$ .

- Rzucamy monetą.
- Jeżeli wypadnie orzeł oraz wszyscy sąsiedzi  $v$  mają przyporządkowane 0 w momencie  $t$ , to  $X_{t+1}(v) = 1$ ; w przeciwnym razie  $X_{t+1}(v) = 0$ .
- Nie zmieniamy wartości pozostałych wierzchołków, tzn.  $X_{t+1}(w) = X_t(w)$  dla  $w \neq v$ .

**Lemat 5.4.** *Zdefiniowany powyżej łańcuch Markowa jest aperiodyczny, nieredukowalny i jego miara stacjonarna jest  $\pi_G$ .*

**Przykład 5.6.** Ustalmy zbiór kolorów  $S = \{1, \dots, q\}$ . **Poprawnym pokolorowaniem** grafu  $G = (V, E)$  jest przyporządkowanie koloru do każdego wierzchołka, w taki sposób, aby każde dwa sąsiadujące ze sobą wierzchołki otrzymały różne kolory. Chcemy wylosować jednostajnie poprawne pokolorowanie ze zbioru wszystkich poprawnych pokolorowań. Oznaczmy odpowiednią miarę przez  $\pi_{G,q}$ .

Rozważmy następujący algorytm. Niech  $X_t = x$  ( $x$  jest poprawnym kolorowaniem). W czasie  $t + 1$

- Losujemy jednostajnie jeden z wierzchołków  $v \in V$ .
- Wybieramy losowo kolor  $j$  - jednostajnie spośród dopuszczalnych kolorów w  $v$  (tzn. jednostajnie ze zbioru  $S \setminus \{x(w)\}_{w \sim v}$ ).
- Podstawiamy  $X_{t+1}(v) = j$  oraz  $X_{t+1}(w) = X_t(w)$  dla  $w \neq v$ .

**Lemat 5.5.** *Miarą stacjonarną zdefiniowanego powyżej łańcucha Markowa jest  $\pi_{G,q}$ .*

Ogólnie, niech  $V$  i  $S$  będą skończonymi zbiorami i załóżmy, że  $\Omega$  jest podzbiorem  $S^V$  (oba powyższe przykłady są tej postaci). Niech  $\pi$  będzie miarą probabilistyczną na  $\Omega$ . **Dynamiką Glaubera** dla  $\pi$  (ang. Glauber dynamics, nieraz używa się również nazwy Gibbs sampler) nazywamy odwracalny łańcuch Markowa na przestrzeni stanów  $\Omega$ , z miarą probabilistyczną  $\pi$  i następującą funkcją przejścia. Załóżmy, że spacer znajduje się w stanie  $x$ , tzn  $X_t = x$ . Wówczas

- Losujemy jednostajnie jeden z wierzchołków  $v \in V$ .
- Niech

$$\Omega(x, v) = \{y \in \Omega : y(w) = x(w) \forall w \neq v\}.$$

Definiujemy miarę  $\pi^{x,v}$  warunkując  $\pi$  na zbiorze  $\Omega(x, v)$ :

$$\pi^{x,v}(y) = \pi\Omega(x, v) = \begin{cases} \frac{\pi(y)}{\pi(\Omega(x,v))} & \text{dla } y \in \Omega(x, v) \\ 0 & \text{dla } y \notin \Omega(x, v) \end{cases}$$

- Losujemy  $X_{t+1}$  względem miary  $\pi^{x,v}$ .

Zauważmy, że oba powyższe przykłady są tej postaci.

**Lemat 5.6.** *Miarą stacjonarną zdefiniowanego powyżej łańcucha Markowa jest  $\pi$ .*

**5.6. Tempo zbieżności dla  $q$ -kolorowania.** Poniżej znajdziemy czas mieszania dla nieco zmodyfikowanego spaceru losowego na zbiorze poprawnych kolorowań. Różnica polega na tym, że zamiast losować jednostajnie spośród dopuszczalnych kolorów dla konfiguracji  $x$  w wierzchołku  $x$ , losujemy dowolny kolor i jeżeli jest on dopuszczalny, to zmieniamy odpowiednio wartość  $x(v)$ . Zwiększa to więc prawdopodobieństwo pozostania w konfiguracji  $x$ .

**Twierdzenie 5.7.** *Niech  $G$  będzie grafem o  $n$  wierzchołkach i niech  $\Delta$  będzie maksymalnym stopniem wierzchołka. Jeżeli  $q > 3\Delta$ , to dla opisanego powyżej spaceru losowego na*

poprawnych kolorowaniach mamy

$$t_{\text{miks}}(\varepsilon) \leq (1 - (3\Delta/q))^{-1} n [\log n + \log(1/\varepsilon)]$$

*Dowód.* Chcemy skorzystać z twierdzenia 2.8, które mówi, że

$$d(t) \leq \sup_{x,y} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq \sup_{x,y} \mathbb{P}_{x,y}[\tau_{\text{par}} > t].$$

Zamiast zwykłego parowania dwóch procesów będziemy jednak rozważać 'wielkie parowanie' (grand coupling). Rozszerzymy też przestrzeń stanów i zamiast poprawnych kolorowań  $\Omega$ , będziemy badać procesy na zbiorze wszystkich kolorowań  $\tilde{\Omega}$ .

Definiujemy rodzinę zmiennych losowych  $\{X_t^x : x \in \tilde{\Omega}, t = 0, 1, \dots\}$  następująco

- $X_0^x = x$ .
- W kroku  $t + 1$  losujemy jednostajnie jeden z wierzchołków  $v \in V$  oraz jeden z kolorów  $k$ . Dla każdego  $x$ , jeżeli sąsiedzi wierzchołki  $v$  mają inny kolor niż  $k$ , tzn.  $X_t^x(w) \neq k$  dla  $w \sim v$ , to zmieniamy wartość  $w$  na  $k$  i jest to  $X_{t+1}^x$ . W przeciwnym razie  $X_{t+1}^x = X_t^x$ . Zauważmy, że używamy tych samych  $v$  i  $k$  dla każdego procesu  $X_t^x$ !

Powyższą rodzinę procesów nazywamy wielkim parowaniem. Dla  $x, y \in \tilde{\Omega}$  chcemy znaleźć moment, w którym oba procesy się spotkają.

Dla dwóch kolorowań  $x, y \in \tilde{\Omega}$  zdefiniujmy ich odległość

$$\rho(x, y) = \sum_{v \in V} \mathbf{1}_{\{x(v) \neq y(v)\}}$$

jako liczbę wierzchołków, na których się one różnią. Zauważmy, że  $\rho$  jest metryką na  $\tilde{\Omega}$ .

Załóżmy, że  $\rho(x, y) = 1$ , tzn.  $x$  i  $y$  różnią się jedynie na jednym wierzchołku  $v_0$ , a poza tym są identyczne. Chcemy zrozumieć co stanie się wykonaniu jednego kroku w algorytmie, tzn. jak wygląda  $\rho(X_1^x, X_1^y)$ . Oznaczmy przez  $N$  zbiór kolorów przyporządkowanych w  $x$  sąsiadom  $v_0$ . Ten zbiór jest identyczny dla  $y$ .

Po pierwszym kroku  $x$  i  $y$  pokrywają się, tzn.  $\rho(X_1^x, X_1^y) = 0$  tylko wówczas, gdy wylosowany zostanie wierzchołek  $v_0$  oraz kolor  $k$  nie należący do  $N$ . To zachodzi z prawdopodobieństwem

$$(5.8) \quad \mathbb{P}[\rho(X_1^x, X_1^y) = 0] = \frac{1}{n} \cdot \frac{q - |N|}{q} \geq \frac{q - \Delta}{nq}.$$

Może się również zdarzyć, że  $\rho(X_1^x, X_1^y) = 2$ . Taka sytuacja może zajść wyłącznie wtedy, gdy wylosowany zostanie w jeden z sąsiadów  $v_0$  oraz jeden z kolorów  $x(v_0)$  lub  $y(v_0)$ . Istotnie załóżmy, że wylosowano  $w$  oraz  $x(v_0)$ . Wtedy  $x(w)$  nie zostanie zmienione, ale może się zmienić  $y(w)$ . Stąd

$$(5.9) \quad \mathbb{P}[\rho(X_1^x, X_1^y) = 2] \leq \frac{\Delta}{n} \cdot \frac{2}{q}.$$

Korzystając z (5.8) oraz (5.9) otrzymujemy

$$\begin{aligned} \mathbb{E}[\rho(X_1^x, X_1^y)] &= 2\mathbb{P}[\rho(X_1^x, X_1^y) = 2] + \mathbb{P}[\rho(X_1^x, X_1^y) = 1] \\ &= 2\mathbb{P}[\rho(X_1^x, X_1^y) = 2] + (1 - \mathbb{P}[\rho(X_1^x, X_1^y) = 2] - \mathbb{P}[\rho(X_1^x, X_1^y) = 0]) \\ &\leq 1 - \frac{q - 3\Delta}{nq} < 1. \end{aligned}$$



Załóżmy teraz, że  $\rho(x, y) = r$ . Wtedy istnieją kolorowania  $x_0 = x, x_1, \dots, x_r = y$  takie, że  $\rho(x_k, x_{k-1}) = 1$ . Korzystając z nierówności trójkąta oraz powyższych oszacowań

$$\begin{aligned} \mathbb{E}[\rho(X_1^x, X_1^y)] &\leq \mathbb{E}[\rho(X_1^x, X_1^{x_1})] + \mathbb{E}[\rho(X_1^{x_1}, X_1^{x_2})] + \dots + \mathbb{E}[\rho(X_1^{x_{r-1}}, X_1^y)] \\ &\leq r \left(1 - \frac{q - 3\Delta}{nq}\right) = \rho(x, y) \left(1 - \frac{q - 3\Delta}{nq}\right). \end{aligned}$$

Chcemy teraz zobaczyć co się dzieje dla dużych czasów.

$$\begin{aligned} \mathbb{E}[\rho(X_t^x, X_t^y)] &= \mathbb{E}\left[\mathbb{E}[\rho(X_t^x, X_t^y) | X_{t-1}^x, X_{t-1}^y]\right] \\ &\leq \mathbb{E}\left[\rho(X_{t-1}^x, X_{t-1}^y) \left(1 - \frac{q - 3\Delta}{nq}\right)\right] \\ &\leq \dots \leq \rho(x, y) \left(1 - \frac{q - 3\Delta}{nq}\right)^t \end{aligned}$$

Dalej, z nierówności Markowa dla  $x \neq y$  mamy

$$\mathbb{P}[X_t^x \neq X_t^y] = \mathbb{P}[\rho(X_t^x, X_t^y) \geq 1] \leq \rho(x, y) \left(1 - \frac{q - 3\Delta}{nq}\right)^t \leq ne^{-t(1-3\Delta/q)/n}.$$

Powyższe oszacowania zachodzą dla wszystkich kolorowań  $x, y \in \tilde{\Omega}$ , w szczególności dla  $x, y \in \Omega$ . Z Twierdzenia 2.8

$$d(t) \leq ne^{-t(1-3\Delta/q)/n}.$$

Zatem jeżeli

$$t > (1 - 3\Delta/q)^{-1} n [\log n + \log(1/\varepsilon)],$$

to  $d(t) \leq \varepsilon$ , co kończy dowód.  $\square$

**5.7. Algorytm Proppa - Wilsona.** Algorytmy MCMC przedstawione powyżej mają dwie wady

- Zmienne losowe generowane przez podane algorytmy nie mają dokładnie rozkładu  $\pi$ , który chcieliśmy osiągnąć. Rozkład tych zmiennych jest bliski pożądanemu, jednak jest obciążony pewnym błędem. Zwiększenie liczby iteracji jedynie zmniejsza ten błąd.
- W celu zminimalizowania błędu należy dokładnie oszacować różnicę pomiędzy rozkładem  $X_t$ , a  $\pi$ . W wielu przypadkach jest to trudne do zrobienia.

W 1996 roku Jim Propp i David Wilson zaproponowali alternatywny algorytm, bazujący również na technikach MCMC, który rozwiązuje jednocześnie oba problemy:

- Generuje zmienną losową, której rozkład jest dokładnie  $\pi$ .
- Algorytm sam decyduje kiedy powinien się zatrzymać. W ten sposób nie ma potrzeby wyznaczania np. czasu mieszania.

Algorytm ten, pomimo prostoty, wymaga jednak dużej przestrzeni obliczeniowej (wymaga uruchomienia kopii  $|S|$  łańcuchów Markowa) co ogranicza jego praktyczne zastosowania. Istnieją jednak metody pozwalające istotnie zredukować problem. Polecamy rozdziały 10-12 książki Haggstroma [H] oraz notatki Berestyckiego [B], rozdział 9.

## 6. WŁASNOŚCI SPEKTRALNE MACIERZY PRZEJŚCIA

Chcemy badać zachowanie  $P^t$  dla dużych wartości  $t$ , gdzie  $P$  jest macierzą przejścia pewnego łańcucha Markowa. Naturalna jest więc próba opisu spektrum  $P$ . Będziemy badać przypadek gdy łańcuchy Markowa generowane przez  $P$  są odwracalne, a więc spełniony jest warunek (1.5), tzn

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \forall x, y \in \Omega.$$

Przypomnijmy, że wówczas  $\pi$  jest miarą stacjonarną.

**Lemat 6.1.** *Niech  $P$  będzie macierzą przejścia skończonego łańcucha Markowa. Wówczas*

- (1) *Jeżeli  $\lambda$  jest wartością własną  $P$ , to  $|\lambda| \leq 1$ .*
- (2) *Jeżeli  $P$  jest nieredukowalna, to przestrzeń własna odpowiadająca 1 jest jednowymiarowa.*
- (3) *Jeżeli  $P$  jest nieredukowalna i aperiodyczna, to  $-1$  nie jest wartością własną  $P$ .*

Niech  $\pi$  będzie miarą stacjonarną. Definiujemy iloczyn skalarny na  $\Omega$ :

$$\langle f, g \rangle_\pi = \sum_{x \in \Omega} f(x)g(x)\pi(x).$$

Przez  $l^2(\pi)$  będziemy oznaczać przestrzeń funkcji o wartościach rzeczywistych na  $\Omega$  wyposażonych w powyższy iloczyn skalarny. Będziemy też rozważać standardowy iloczyn skalarny  $\langle \cdot, \cdot \rangle$

**Lemat 6.2.** *Niech  $P$  będzie odwracalne względem  $\pi$ . Wówczas*

- (1) *Przestrzeń  $l^2(\pi)$  ma bazę ortonormalną złożoną z rzeczywistych funkcji własnych  $P$ :  $\{f_j\}$  i  $f_1 = (1, \dots, 1)$ . Odpowiednie wartości własne będziemy oznaczać przez  $\{\lambda_j\}$ .*
- (2) *Macierz  $P$  można rozłożyć*

$$\frac{P^t(x, y)}{\pi(y)} = \sum_{j=1}^{|\Omega|} f_j(x)f_j(y)\lambda_j^t.$$

*Dowód.* Zdefiniujemy

$$A(x, y) = \frac{\sqrt{\pi(x)}}{\sqrt{\pi(y)}}P(x, y).$$

Odwracalność  $P$  implikuje, że  $A$  jest macierzą symetryczną:

$$\begin{aligned} A(x, y) &= \frac{\sqrt{\pi(x)}}{\sqrt{\pi(y)}}P(x, y) = \frac{1}{\sqrt{\pi(x)\pi(y)}}\pi(x)P(x, y) \\ &= \frac{1}{\sqrt{\pi(x)\pi(y)}}\pi(y)P(y, x) = \frac{\sqrt{\pi(y)}}{\sqrt{\pi(x)}}P(y, x) = A(y, x). \end{aligned}$$

Z twierdzenia spektralnego dla macierzy symetrycznych wynika, że istnieje baza ortonormalna  $\mathbb{R}^{|\Omega|}$  względem  $\langle \cdot, \cdot \rangle$ :  $\{\phi_j\}$  taka, że  $\phi_j$  jest funkcją własną rzeczywistej wartości własnej  $\lambda_j$ . Łatwo sprawdzić, że  $\phi_1(x) = \sqrt{\pi(x)}$  jest wartością własną odpowiadającą  $\lambda_1 = 1$ . Istotnie

$$A\phi_1(x) = \sum_y A(x, y)\sqrt{\pi(y)} = \sum_y \sqrt{\pi(x)}P(x, y) = \sqrt{\pi(x)} = \phi_1(x).$$

Niech  $D_\pi$  oznacza macierz diagonalną z wektorem  $\pi$  na przekątnej. Wtedy  $A = D_\pi^{-\frac{1}{2}} P D_\pi^{-\frac{1}{2}}$ . Definiujemy  $f_j = D_\pi^{-\frac{1}{2}} \phi_j$  i mamy

$$P f_j = P D_\pi^{-\frac{1}{2}} \phi_j = D_\pi^{-\frac{1}{2}} (D_\pi^{\frac{1}{2}} P D_\pi^{-\frac{1}{2}}) \phi_j = D_\pi^{-\frac{1}{2}} A \phi_j = D_\pi^{-\frac{1}{2}} \lambda_j \phi_j = \lambda_j f_j.$$

Funkcje  $f_j$  są więc wartościami własnymi  $P$ . Ponadto

$$\delta_{ij} = \langle \phi_i, \phi_j \rangle = \langle D_\pi^{\frac{1}{2}} f_j, D_\pi^{\frac{1}{2}} f_j \rangle = \langle f_i, f_j \rangle_\pi,$$

co dowodzi (1).

Następnie zdefiniujemy funkcję

$$\delta_y(x) = \begin{cases} 1 & \text{dla } y = x, \\ 0 & \text{dla } y \neq x, \end{cases}$$

i rozkładamy ją

$$\delta_y(x) = \sum_j \langle \delta_y, f_j \rangle_\pi f_j(x) = \sum_j f_j(y) \pi(y) f_j(x).$$

Ponieważ  $P^t(x, y) = (P^t \delta_y)(x)$  i  $P^t f_j = \lambda_j^t f_j$ ,

$$P^t(x, y) = \sum_j f_j(y) \pi(y) \lambda_j^t f_j(x).$$

□

**Wniosek 6.3.** *Mamy*

$$P^t f = \sum_{j=1}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t.$$

*Dowód.* Korzystając z powyższego twierdzenia

$$\begin{aligned} P^t f(x) &= P^t \left( \sum_y f(y) \delta_y \right) (x) = \sum_y f(y) P^t \delta_y(x) \\ &= \sum_{y,j} f(y) f_j(y) \pi(y) \lambda_j^t f_j(x) = \sum_j \langle f, f_j \rangle_\pi \lambda_j^t f_j(x). \end{aligned}$$

□

Sortujemy kolejne wartości własne:

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} \geq -1.$$

**Definicja 6.1.** Niech

$$\lambda_* = \max_{i>1} \{|\lambda_i|\}.$$

Wartość  $\gamma_* = 1 - \lambda_*$  nazywamy **absolutną dziurą spektralną** (ang. absolute spectral gap). Wartość  $\gamma = 1 - \lambda_2$  nazywana jest **dziurą spektralną** (ang. spectral gap). Definiujemy **czas relaksacji**  $t_{\text{rel}} = 1/\gamma_*$

Często  $\gamma_* = \gamma$ . Np. dla leniwych spacerów losowych ich macierz przejścia  $\tilde{P} = (1/2)(I + P)$  ma wszystkie wartości własne nieujemne.

Istnieje związek pomiędzy czasem mieszania, a czasem relaksacji:

**Twierdzenie 6.4.** Niech macierz przejścia  $P$  będzie nieredukowalna, aperiodyczna i odwracalna względem  $\pi$ . Zdefiniujmy  $\pi_{\min} = \min_x \pi(x)$ . Wówczas

$$(6.5) \quad (t_{\text{rel}} - 1) \log \left( \frac{1}{2\varepsilon} \right) \leq t_{\text{miks}}(\varepsilon) \leq \log \left( \frac{1}{\varepsilon \pi_{\min}} \right) t_{\text{rel}}.$$

*Dowód. Krok 1.* Przypomnijmy, z dowodu twierdzenia 3.5:

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{\text{TV}} &= \sum_{\{y: P^t(x, y) < \pi(y)\}} (\pi(y) - P^t(x, y)) \\ &= \sum_{\{y: P^t(x, y) < \pi(y)\}} \pi(y) \left( 1 - \frac{P^t(x, y)}{\pi(y)} \right) \\ &\leq \max_y \left( 1 - \frac{P^t(x, y)}{\pi(y)} \right) \end{aligned}$$

Oznaczmy powyższą wartość przez  $s_x(t)$  (jest ona nazywana w literaturze 'separation distance'). Stąd

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq \max_x s_x(t) = \max_x \max_y \left( 1 - \frac{P^t(x, y)}{\pi(y)} \right)$$

Dalej z lematu 6.2 oraz nierówności Schwarz'a mamy

$$\left| 1 - \frac{P^t(x, y)}{\pi(y)} \right| \leq \sum_{j=2}^{|\Omega|} |f_j(x) f_j(y)| |\lambda_j|^t \leq \lambda_*^t \left( \sum_{j=2}^{|\Omega|} f_j^2(x) \sum_{j=2}^{|\Omega|} f_j^2(y) \right)^{\frac{1}{2}}.$$

Zauważmy

$$\pi(x) = \langle \delta_x, \delta_x \rangle_{\pi} = \left\langle \sum_{j=1}^{|\Omega|} f_j(x) \pi(x) f_j, \sum_{j=1}^{|\Omega|} f_j(x) \pi(x) f_j \right\rangle_{\pi} = \pi^2(x) \sum_{j=1}^{|\Omega|} f_j^2(x),$$

więc  $\sum_{j=2}^{|\Omega|} f_j^2(x) \leq \pi(x)^{-1}$ . Stąd

$$\left| 1 - \frac{P^t(x, y)}{\pi(y)} \right| \leq \frac{\lambda_*^t}{\sqrt{\pi(x)} \sqrt{\pi(y)}} \leq \frac{\lambda_*^t}{\pi_{\min}} = \frac{(1 - \gamma_*)^t}{\pi_{\min}} \leq \frac{e^{-\gamma_* t}}{\pi_{\min}}.$$

Z definicji  $t_{\text{miks}}$  wynika teraz prawa nierówność w (6.5). (To szacowanie nie wymagało aperiodyczności)

W szczególności otrzymaliśmy powyżej nierówność, która przyda nam się później:

$$(6.6) \quad d(t) \leq \frac{\lambda_*^t}{\pi_{\min}}$$

**Krok 2.** Niech  $f_*$  będzie funkcją własną  $P$  z wartością własną odpowiadającą wartości własnej, której moduł jest równy  $\lambda_*$ . Ponieważ wartości własne  $P$  są prostopadłe względem iloczynu skalarnego  $\langle \cdot, \cdot \rangle_{\pi}$ , to

$$0 = \langle f_*, \mathbf{1} \rangle_{\pi} = \sum_y f_*(y) \pi(y)$$

Wtedy

$$\lambda_*^t |f_*(x)| = |P^t f_*(x)| = \left| \sum_y (P^t(x, y) f_*(y) - \pi(y) f_*(y)) \right| \leq \|f_*\|_{\infty} 2d(t),$$

bo  $2\|P^t(x, \cdot) - \pi\|_{\text{TV}} = \sum_y |P^t(x, y) - \pi(y)|$ . Biorąc  $x_0$  takie, że  $f_*(x_0) = \|f_*\|_\infty$  otrzymujemy

$$(6.7) \quad \lambda_*^t \leq 2d(t).$$

Zatem  $\lambda_*^{t_{\text{miks}}(\varepsilon)} \leq 2\varepsilon$ . Korzystając z nierówności  $\log(1+x) < x$  dla  $x > 0$ , mamy

$$t_{\text{miks}}(\varepsilon) \left( \frac{1}{\lambda_*} - 1 \right) \geq t_{\text{miks}}(\varepsilon) \log \left( \frac{1}{\lambda_*} \right) \geq \log \left( \frac{1}{2\varepsilon} \right).$$

Używając równania  $t_{\text{rel}} - 1 = \frac{\lambda_*}{1-\lambda_*}$  otrzymujemy lewą stronę nierówności (6.5).  $\square$

**Wniosek 6.8.** *Dla odwracalnego, nieredukowalnego i aperiodycznego łańcucha Markowa mamy*

$$\lim_{t \rightarrow \infty} d(t)^{\frac{1}{t}} = \lambda_*$$

*Dowód.* Z (6.7) mamy

$$\lambda_* \leq \liminf_{t \rightarrow \infty} d(t)^{\frac{1}{t}}.$$

Z drugiej strony, z (6.6) wynika, że

$$\limsup_{t \rightarrow \infty} d(t)^{\frac{1}{t}} = \lambda_*.$$

$\square$

**Przykład 6.1. Prosty spacer losowy na  $\mathbb{Z}_n$ .** Możemy myśleć o tym przykładzie jako spacerze losowym na pierwiastkach  $n$  stopnia z jedności:  $\Omega = \{\omega, \omega^2, \dots, \omega^{n-1}, 1\}$  dla  $\omega = e^{2\pi i/n}$ . Wówczas dla  $\phi_j(\omega^k) = \omega^{kj}$ ,  $0 \leq j \leq n-1$ , mamy

$$P\phi_j(\omega^k) = \frac{\phi_j(\omega^{k-1}) + \phi_j(\omega^{k+1})}{2} = \frac{\omega^{jk-1} + \omega^{jk+1}}{2} = \omega^{jk} \left( \frac{\omega^{-1} + \omega}{2} \right) = \cos(2\pi j/n) \phi_j(\omega^k).$$

Zatem  $\phi_j$  jest funkcją własną  $P$  z wartością własną  $\cos(2\pi j/n)$ . Ponieważ macierz  $P$  i wartości własne są rzeczywiste, to również funkcje

$$f_j(\omega^k) = \Re(\phi_j(\omega^k)) = \cos(2\pi jk/n)$$

są funkcjami własnymi.

Jeżeli  $n$  jest parzyste, to  $-1$  jest wartością własną i spacer jest periodyczny, wtedy  $\gamma_* = 0$ . Jeżeli  $n$  jest nieparzyste, to spacer jest aperiodyczny i absolutna dziura spektralna wynosi (dziura spektralna jest tego samego rzędu)

$$\lambda_* = 1 - \cos \left( \frac{2\pi(n-1)}{2n} \right) = 1 - \cos(\pi/n) = \frac{\pi^2}{2n^2} + O(n^{-4}),$$

więc dziura spektralna jest rzędu  $n^{-2}$ , a czas relaksacji rzędu  $n^2$ .

**Definicja 6.2.** Mówimy, że łańcuch Markowa jest **tranzytywny** jeżeli dla każdej pary  $(x, y) \in \Omega \times \Omega$  istnieje bijekcja  $\phi = \phi_{(x,y)} : \Omega \mapsto \Omega$  taka, że

$$\phi(x) = y \quad \text{oraz} \quad P(z, w) = P(\phi(z), \phi(w)) \quad \forall z, w \in \Omega.$$

Zatem w każdym punkcie spacer losowy wygląda tak samo.

Tę własność łatwo zauważyć dla prostych spacerów losowych na  $\mathbb{Z}_n$  lub  $\mathbb{Z}_2^d$ . Ogólniej, np. spacery losowe na grupach są tranzytywne:  $\phi_{(x,y)}(g) = gx^{-1}y$ .

$$\text{Definiujemy } \|f\|_2^2 = \langle f, f \rangle_\pi$$

**Lemat 6.9.** Niech  $P$  będzie odwracalną macierzą przejścia. Wówczas

$$4\|P^t(x, \cdot) - \pi\|_{\text{TV}}^2 \leq \sum_{j=2}^{|\Omega|} f_j(x)^2 \lambda_j^{2t}$$

Ponadto jeżeli spacer losowy jest tranzytywny, to

$$4\|P^t(x, \cdot) - \pi\|_{\text{TV}}^2 \leq \sum_{j=2}^{|\Omega|} \lambda_j^{2t}$$

*Dowód.* Korzystając z nierówności  $\|\cdot\|_1 \leq \|\cdot\|_2$ , mamy

$$\begin{aligned} 4\|P^t(x, \cdot) - \pi\|_{\text{TV}}^2 &\leq \left( \sum_y \left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \pi(y) \right)^2 \\ &= \left\| \frac{P^t(x, \cdot)}{\pi(\cdot)} - 1 \right\|_1^2 \\ &\leq \left\| \frac{P^t(x, \cdot)}{\pi(\cdot)} - 1 \right\|_2^2 \\ &= \left\| \sum_{j=2}^{|\Omega|} \lambda_j^t f_j(x) f_j \right\|_2^2 \\ &= \sum_{j=2}^{|\Omega|} \lambda_j^{2t} f_j(x)^2. \end{aligned}$$

to dowodzi pierwszej nierówności.

Do dowodu drugiej użyjemy obserwacji, że jeżeli spacer losowy jest tranzytywny, to jego miarą stacjonarną jest miara jednostajna. Do dowodu tego faktu oznaczmy przez  $U$  miarę jednostajną na  $\Omega$ . Wtedy dla dowolnych  $x, y \in \Omega$  weźmy odpowiednią bijekcję  $\phi = \phi_{x,y}$ . Wówczas

$$UP(x) = \sum_{z \in \Omega} U(z)P(z, x) = \sum_{z \in \Omega} U(\phi(z))P(\phi(z), y) = \sum_{w \in \Omega} U(w)P(w, y) = UP(y).$$

Wartość  $UP(x)$  nie zależy więc od  $x$  i musi być stała, z uwagi na to, że  $UP$  jest miarą probabilistyczną

$$UP(x) = 1/|\Omega| = U(x).$$

Zauważmy, że w powyższych szacowaniach lewa strona nie zależy od  $x$  (wynika to z tranzytywności), zatem dla każdych  $x, y$  mamy

$$4\|P^t(y, \cdot) - \pi\|_{\text{TV}}^2 \leq \sum_{j=2}^{|\Omega|} \lambda_j^{2t} f_j(x)^2.$$

Sumując obie strony po  $x$  i korzystając z faktu, że miara jednostajna jest stacjonarna

$$|\Omega|4\|P^t(y, \cdot) - \pi\|_{\text{TV}}^2 \leq |\Omega| \sum_{j=2}^{|\Omega|} \left( \sum_{x \in \Omega} f_j(x)^2 U(x) \right) \lambda_j^{2t} = |\Omega| \sum_{j=2}^{|\Omega|} \lambda_j^{2t},$$

co kończy dowód. □

**Wniosek 6.10.** Rozważamy prosty spacer losowy na  $\mathbb{Z}_n$  dla nieparzystego  $n \geq 7$ . Jeżeli  $t \geq n^2$ , to

$$d(t) \leq e^{-\alpha \frac{t}{n^2}}$$

gdzie  $\alpha = \pi^2/2$ . Na odwrót, dla każdego  $t \geq 0$

$$d(t) \geq \frac{1}{2} e^{-\alpha \frac{t}{n^2} - \beta \frac{t}{n^4}}$$

dla  $\beta = \pi^4/11$ .

Zauważmy, że dla  $t \leq n^2$  wartość  $d(t)$  jest duża, a po przekroczeniu  $n^2$  zaczyna gwałtownie maleć.

*Dowód.* Z poprzedniego twierdzenia wynika, że

$$d(t)^2 \leq \frac{1}{4} \sum_{j=1}^{n-1} \cos\left(\frac{2\pi j}{n}\right)^{2t} \leq \frac{1}{2} \sum_{j=1}^{(n-1)/2} \cos\left(\frac{\pi j}{n}\right)^{2t}.$$

(uwaga na przeniebrowanie wartości własnych!). Używając nierówności  $\cos x \leq e^{-x^2/2}$  dla  $x \in [0, \pi/2]$  otrzymujemy

$$\begin{aligned} d(t)^2 &\leq \frac{1}{2} \sum_{j=1}^{(n-1)/2} \exp\left(-\frac{\pi^2 j^2 t}{n^2}\right) \\ &\leq \frac{1}{2} \exp\left(-\frac{\pi^2 t}{n^2}\right) \sum_{j=0}^{\infty} \exp\left(-\frac{3j\pi^2 t}{n^2}\right) \\ &= \frac{1}{2} \frac{\exp\left(-\frac{\pi^2 t}{n^2}\right)}{1 - \exp\left(-\frac{3\pi^2 t}{n^2}\right)} \end{aligned}$$

i to daje pierwszą nierówności. Dolne szacowania wynikają z nierówności 6.7:  $\lambda_*^t \leq 2d(t)$  oraz  $\cos x \geq \exp\left(-\frac{x^2}{2} - \frac{x^4}{11}\right)$  dla  $x \in [0, 1/2]$  (sprawdź!). Dają one

$$d(t) \geq \frac{1}{2} \cos\left(\frac{\pi}{n}\right)^t \geq \frac{1}{2} e^{-\alpha \frac{t}{n^2} - \beta \frac{t}{n^4}}$$

□

**6.1. Leniwy spacer losowy na hiperkostce  $\mathbb{Z}_2^n$ .** Przypomnijmy, że spacer ten można opisać na dwa sposoby:

- z prawdopodobieństwem  $1/2$  zostajemy w aktualnym wierzchołku, a z prawdopodobieństwem  $(1/2n)$  przechodzimy do jednego z sąsiadów (czyli wektora różniącego tylko na jednej współrzędnej od aktualnego stanu);
- losujemy jednostajnie jedną ze współrzędnych i tam kładziemy losowo 0 lub 1.

Do tej pory używając parowania oraz silnych czasów stacjonarnych pokazaliśmy, że

$$t_{\text{miks}}(\varepsilon) \leq n \log n + \log(\varepsilon^{-1})n.$$

Chcemy użyć teraz metody spektralnej. Bezpośrednie zdiagonalizowanie macierzy jest trudne. Popatrzmy jednak na spacer losowy jak na produkt spacerów losowych.

Mianowicie zdefiniujemy  $\Omega_1 = \{-1, 1\}$  (w tym rozdziale zamiast hiperkostki  $\{0, 1\}^n$ , będziemy patrzeć na zbiór  $\{-1, 1\}^n$ ) i rozważmy spacer losowy zadany macierzą przejścia

$$P_1(x, y) = \frac{1}{2} \quad \text{dla wszystkich } x, y \in \{-1, 1\}.$$

Łatwo zdiagonalizować macierz  $P_1$  ma ona dwie wartości własne

- 1, z wartością własną  $\mathbf{1}$  (funkcja stała);
- 0, z wartością własną  $\mathbf{I}$  ( $\mathbf{I}(x) = x$ , tzn.  $\mathbf{I} = (-1, 1)$ ).

Rozważmy kopie  $\{(\Omega_i, P_i)\}_{i=1}^n$ . Wtedy leniwy spacer losowy na  $\{-1, 1\}^n$  może być zapisany jako

$$P(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{j=1}^n P_j(x_j, y_j) \prod_{i \neq j} \mathbf{1}_{\{x_i = y_i\}},$$

gdzie  $\mathbf{x} = (x_1, \dots, x_n)$ . Istotnie, losujemy jedną ze współrzędnych, a następnie randomizujemy tę współrzędną, ale tylko pod warunkiem, że  $\mathbf{x}$  i  $\mathbf{y}$  są sąsiadami, czyli zgadzają się na pozostałych współrzędnych.

Dla każdego  $j = 1, \dots, n$  umiemy zdiagonalizować  $P_j$  o napisać funkcje własne. Wykorzystując tę wiedzę możemy również zdiagonalizować  $P$ . Istotnie, dla każdego podzbioru  $J \subset \{1, 2, \dots, n\}$  zdefiniujemy

$$f_J(x_1, \dots, x_n) = \prod_{j \in J} \mathbf{I}(x_j) \prod_{j \in J} \mathbf{1}(x_j) = \prod_{j \in J} x_j.$$

Oznaczmy teraz

$$\tilde{P}_j(\mathbf{x}, \mathbf{y}) = P_j(x_j, y_j) \prod_{i \neq j} \mathbf{1}_{\{x_i = y_i\}}$$

i zauważmy, że

$$\begin{aligned} \tilde{P}_j f_J(\mathbf{x}) &= \sum \tilde{P}_j(\mathbf{x}, \mathbf{y}) f_J(\mathbf{y}) \\ &= \tilde{P}_j(\mathbf{x}, (x_1, \dots, x_j, \dots, x_n)) f_J(x_1, \dots, x_j, \dots, x_n) \\ &\quad + \tilde{P}_j(\mathbf{x}, (x_1, \dots, -x_j, \dots, x_n)) f_J(x_1, \dots, -x_j, \dots, x_n) \\ &= \frac{1}{2} (f_J(x_1, \dots, x_j, \dots, x_n) + f_J(x_1, \dots, -x_j, \dots, x_n)) \end{aligned}$$

i stąd

$$\tilde{P}_j f_J(\mathbf{x}) = \begin{cases} f_J(\mathbf{x}) & \text{gdzie } j \notin J, \\ 0 & \text{gdzie } j \in J. \end{cases}$$

Zatem

$$P f_J(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^n \tilde{P}_j f_J(\mathbf{x}) = \frac{n - |J|}{n} f_J(\mathbf{x}),$$

więc  $\frac{n-|J|}{n}$  są wartościami własnymi. Wskazaliśmy  $2^n$  różnych funkcji własnych, więc nie ma innych. To pokazuje, że wartości własne  $P$  wynoszą:  $1, 1 - 1/n, 1 - 2/n, \dots, 0$ . Więc  $\lambda_* = 1 - 1/n$  oraz  $\gamma_* = 1/n$ ,  $t_{\text{rel}} = n$ . Przypomnijmy, że miarą stacjonarną jest miara jednostajna, stąd  $\pi_{\min} = 1/2^n$ . Ostatecznie z twierdzenia 6.4 mamy

$$t_{\text{miks}}(\varepsilon) \leq n(\log 2^n - \log \varepsilon) = O(n^2).$$



Szacowania te są znacznie słabsze od dotychczas otrzymanych, ale pokażemy jak pełna wiedza o strukturze spektrum i lemat 6.9 dadzą optymalne szacowania. Mamy

$$\begin{aligned} 4d(t)^2 &\leq \sum_{k=1}^n \left(1 - \frac{k}{n}\right)^{2t} \binom{n}{k} \\ &\leq \sum_{k=1}^n e^{-2tk/n} \binom{n}{k} \\ &\leq (1 + e^{-2t/n})^n - 1. \end{aligned}$$

Biorąc  $t = (1/2)n \log n + cn$  dostajemy

$$4d(t)^2 \leq (1 + 1/n e^{-2c})^n - 1 \leq e^{e^{-2c}} - 1 \leq 2e^{-2c}$$

stąd

$$d((1/2)n \log n + cn) \leq \frac{1}{2} e^{-c}$$

Można pokazać ([LPW], p.95), że

$$d((1/2)n \log n - cn) \geq 1 - \frac{8}{e^{2c}}(1 + o(1))$$

To pokazuje, że tuż przed czasem  $(1/2)n \log n$  odległość rozkładu spaceru od miary stacjonarnej jest bliska 1, ale dla czasów nieco większych od  $(1/2)n \log n$  bardzo szybko zaczyna zbiegać do zera. Takie zachowanie nazywane jest fenomenem *cutoff* (patrz [LPW], rozdział 18)

## 7. FORMA DIRICHLETA I NIERÓWNOŚĆ POINCARÉ

W poprzednim rozdziale pokazaliśmy związek pomiędzy dziurą spektralną, a czasem mieszania. Kluczowa była jednak znajomość drugiej wartości własnej. W wielu przypadkach nie można jej obliczyć. Poniżej pokażemy jednak techniki bazujące na geometrycznych własnościach grafu pozwalające przybliżyć wartość dziury spektralnej.

**7.1. Forma Dirichleta, a dziura spektralna.** Przypomnijmy działanie operatora  $P$  na funkcjach:

$$Pf(x) = \sum_y P(x, y)f(y)$$

Przez  $\pi$  będziemy oznaczać miarę stacjonarną  $P$ . Poniżej będziemy badać jedynie odwracalne spacery losowe.

**Definicja 7.1.** Niech  $f, h$  będą funkcjami na  $\Omega$ . **Formę Dirichleta** definiujemy następująco

$$\mathcal{E}(f, h) = \langle (I - P)f, h \rangle_\pi$$

Można ją zapisać nieco inaczej. Korzystając z odwracalności

$$\begin{aligned}
\mathcal{E}(f, h) &= \sum_x \pi(x)(f(x) - Pf(x))h(x) \\
&= \sum_x \pi(x) \left( \sum_y P(x, y)(f(x) - f(y)) \right) h(x) \\
&= \sum_{x, y} \pi(x) P(x, y) h(x) (f(x) - f(y)) \\
&= \frac{1}{2} \sum_{x, y} \pi(x) P(x, y) (f(y) - f(x))(h(y) - h(x))
\end{aligned}$$

W szczególności

$$\mathcal{E}(f) := \mathcal{E}(f, f) = \frac{1}{2} \sum_{x, y} \pi(x) P(x, y) (f(y) - f(x))^2$$

Dla uproszczenia notacji (a przede wszystkim uwzględniając obecność formy Dirichleta w znacznie ogólniejszym kontekście) dla ustalonej krawędzi  $e = (x, y)$  definiujemy

$$Q(e) = \frac{1}{2} \pi(x) P(x, y)$$

oraz *gradient*

$$\nabla f(e) = f(y) - f(x).$$

Wówczas

$$\mathcal{E}(f, g) = \sum_e Q(e) \nabla f(e) \nabla g(e).$$

(w tym miejscu warto powiązać formę Dirichleta z ciągłą sytuacją, laplasjanem i gradientem).

**Twierdzenie 7.1.** *Załóżmy, że łańcuch Markowa  $(P, \pi)$  jest odwracalny. Wtedy dziura spektralna  $\gamma$  spełnia*

$$\gamma = \min_{\{f: \mathbb{E}_\pi(f)=0, \|f\|_2=1\}} \mathcal{E}(f).$$

*Ponadto równość jest przyjęta dla  $f = f_2$ , gdzie  $f_2$  jest drugą funkcją własną  $P$ .*

*Dowód.* Warunek  $\mathbb{E}_\pi(f) = 0$  oznacza, że  $\langle f, f_1 \rangle_\pi = 0$ . Stąd

$$f = \sum_{j=1}^n \langle f, f_j \rangle_\pi f_j = \sum_{j=2}^n \langle f, f_j \rangle_\pi f_j.$$

Dalej mamy

$$\begin{aligned}
\mathcal{E}(f) &= \langle (I - P)f, f \rangle_\pi \\
&= \langle f, f \rangle_\pi - \langle Pf, f \rangle_\pi \\
&= \sum_{j=2}^n \langle f, f_j \rangle_\pi^2 - \sum_{j=2}^n \langle \langle f, f_j \rangle_\pi Pf_j, f \rangle_\pi \\
&= \sum_{j=2}^n \langle f, f_j \rangle_\pi^2 (1 - \lambda_j) \\
&\geq (1 - \lambda_2) \sum_{j=2}^n \langle f, f_j \rangle_\pi^2 \\
&= (1 - \lambda_2) \|f\|_2^2 = \gamma.
\end{aligned}$$

Widać, że dla  $f = f_2$  otrzymujemy równość.  $\square$

**7.2. Nierówność Poincare i metoda ścieżek.** Powyższe twierdzenie mówi, że dla każdej funkcji  $f$ , takiej, że  $\mathbb{E}_\pi(f) = 0$  mamy

$$\gamma \|f\|_2^2 \leq \mathcal{E}(f),$$

innymi słowy, w terminach probabilistycznych

$$\text{var}_\pi(f) \leq \frac{1}{\gamma} \mathcal{E}(f).$$

**Definicja 7.2.** Mówimy, że funkcja przejścia  $P$  spełnia nierówność Poincare ze stałą  $C$  jeżeli dla wszystkich funkcji  $f$

$$(7.2) \quad \text{var}_\pi(f) \leq C \mathcal{E}(f).$$

Powyżej, w twierdzenie 7.1 pokazaliśmy nierówność Poincare ze stałą  $1/\gamma$ . Jeżeli umiemy więc, innymi metodami, pokazać nierówność (7.2) dla pewnej stałej  $C$ , to  $t_{\text{rel}} \leq C$ .<sup>7</sup>

Poniżej prezentujemy technikę 'ścieżek' wprowadzoną przez Diaconisa i Saloff-Coste. Dla każdych  $x, y \in S$ , ustalamy raz na zawsze ścieżkę  $\gamma_{x,y}$ , tzn. ciąg wierzchołków  $x_0 = x, x_1, \dots, x_n = y$  takich, że  $P(x_i, x_{i+1}) > 0$ . Przez  $|\gamma_{x,y}| = n$  oznaczamy długość tej ścieżki.

**Twierdzenie 7.3.** *Nierówność Poincare (7.2) zachodzi ze stałą*

$$C = \max_{e \in E} \left\{ \frac{1}{Q(e)} \sum_{\{x,y: e \in \gamma_{x,y}\}} |\gamma_{x,y}| \pi(x) \pi(y) \right\}.$$

W szczególności  $\gamma \geq 1/C$ .

<sup>7</sup>Uwaga: czas relaksacji został zdefiniowany jako  $t_{\text{rel}} = 1/\gamma^*$ . W tym rozdziale przyjmujemy  $t_{\text{rel}} = 1/\gamma$ . Będzie nas interesować jedynie rząd poszczególnych parametrów, w tym czasu mieszania, a nie konkretne stałe. Jak wiemy każdy spacer można spowolnić, rozważając leniwy spacer losowy, a ten, jak już wiemy, ma wszystkie wartości własne nieujemne

*Dowód.* W dowodzie dwukrotnie używamy nierówności Schwartza

$$\begin{aligned}
\text{var}_\pi(f) &= \sum_x \left| f(x) - \sum_y \pi(y) f(y) \right|^2 \pi(x) \\
&= \sum_x \left( \sum_y (f(x) - f(y)) \pi(y) \right)^2 \pi(x) \\
&\leq \sum_x \left( \sum_y (f(x) - f(y))^2 \pi(y) \right) \left( \sum_y 1 \pi(y) \right) \pi(x) \\
&= \sum_{x,y} (f(x) - f(y))^2 \pi(y) \pi(x) \\
&= \sum_{x,y} \left( \sum_{e \in \gamma_{x,y}} \nabla f(e) \right)^2 \pi(y) \pi(x) \\
&\leq \sum_{x,y} |\gamma_{x,y}| \sum_{e \in \gamma_{x,y}} |\nabla f(e)|^2 \pi(y) \pi(x) \\
&= \sum_e \left( \frac{1}{Q(e)} \sum_{\{x,y: e \in \gamma_{x,y}\}} |\gamma_{x,y}| \pi(x) \pi(y) \right) Q(e) |\nabla f(e)|^2 \pi(y) \pi(x) \\
&\leq C \mathcal{E}(f).
\end{aligned}$$

□

Nieraz wygodnie jest inaczej mierzyć odległość, uwzględniając geometrię grafu i kładąc różne wagi na krawędziach. Niech  $\omega : \mathbb{E} \mapsto \mathbb{R}_+$  będzie wagą na krawędziach. Definiujemy wagę ścieżki

$$|\gamma|_\omega = \sum_{e \in \gamma} \omega(e).$$

Ten sam dowód, co powyżej daje

**Wniosek 7.4.** *Nierówność Poincare (7.2) zachodzi ze stałą*

$$C = \max_{e \in \mathcal{A}} \left\{ \frac{1}{\omega(e)Q(e)} \sum_{\{x,y: e \in \gamma_{x,y}\}} |\gamma_{x,y}|_\omega \pi(x) \pi(y) \right\}.$$

Poniżej prezentujemy kilka przykładów

**Przykład 7.1. Kostka**  $V = \{1, \dots, n\}^d$ . Krawędzie łączą dwa wierzchołki różniące się na dokładnie jednej współrzędnej o 1. Pokażemy, jak używając metody ścieżek oszacować dziurę spektralną. Zastosujemy metodę ścieżek. Dla każdych  $\mathbf{x}, \mathbf{y} \in V$  definiujemy ścieżkę łączącą te punkty w taki sposób, że najpierw idziemy wzdłuż pierwszej współrzędnej, potem drugiej, itd.  $\gamma_{\mathbf{x}, \mathbf{y}}$  jest więc sumą co najwyżej  $d$  prostych odcinków. W tym przypadku można napisać bezpośrednio miarę stacjonarną, ale nas interesuje jedynie rząd i wiemy, że  $\pi(\mathbf{x}) \leq C/n^d$  oraz  $Q(e) \geq c/n^d$ . Ponadto maksymalna długość ścieżki  $\gamma_{\mathbf{x}, \mathbf{y}}$  wynosi  $dn$ . Stąd

$$\begin{aligned}
C &= \max_e \left\{ \frac{1}{Q(e)} \sum_{\{x,y: e \in \gamma_{x,y}\}} |\gamma_{x,y}| \pi(x) \pi(y) \right\} \\
&\leq \max_r \left\{ \frac{1}{n^{d-1}} \#\{x, y : e \in \gamma_{x,y}\} \right\} \\
&\leq n^2
\end{aligned}$$

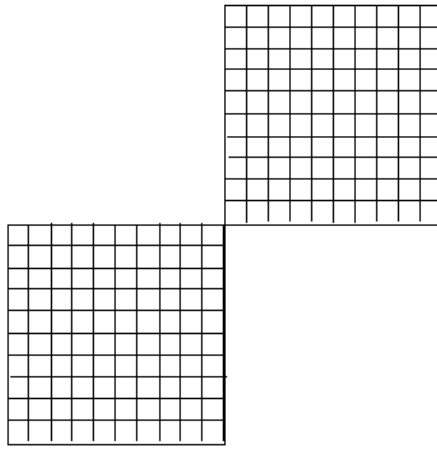
(uzasadnij ostatni krok)

**Przykład 7.2. Drzewo.** Rozważmy skończone drzewo o  $n$  wierzchołkach i maksymalnym stopniu  $d$  oraz maksymalnej wysokości  $H$ . Wówczas

$$\gamma \geq \frac{c}{d^2 n H}.$$

Dla każdego wyboru ścieżek  $\gamma_{x,y}$  wymuszamy, aby ścieżka przechodziła przez korzeń. Z definicji maksymalna długość ścieżki jest co najwyżej  $2H$ . Dalej,  $\#\{x, y \in V : e \in \gamma_{x,y}\}$  jest mniejsza od  $n^2$ . To daje  $C \leq cd^2 n H$ .

**Przykład 7.3. Zwężenia.** Rozważmy dwie siatki kwadratowe połączone ze sobą w jednym wierzchołku.



Czas mieszania dla pojedynczego kawałka jest rzędu  $n^2$ . Pokażemy, że z uwagi na zwężenie czas mieszania na całym grafie jest wyższego rzędu. Dla uproszczenia rachunków możemy założyć, że graf jest regularny. Wtedy miara jednostajna jest miarą stacjonarną. W tym celu wystarczy dodać pętle w niektórych wierzchołkach (jedną dla wierzchołków na brzegu i dwie na rogach). Wówczas

**Twierdzenie 7.5.** Dla każdego  $n \geq 1$

$$t_{\text{rel}} \leq 64(n+1)^2 \log(2n+1).$$

Ponadto dla dużych  $n$

$$t_{\text{rel}} \geq 2n^2 \log n.$$

*Dowód.* Pokażemy najpierw górne szacowanie. Zauważmy, że

- liczba wierzchołków:  $|V| = 2(n+1)^2 - 1$ ;
- miara stacjonarna jest jednostajna:  $\pi(x) = 1/|V|$ ;
- $Q(e) = 1/(4|V|)$

Chcemy użyć wniosku 7.4. W tym celu musimy zdefiniować ścieżki i wagi. Każda ścieżka  $\gamma_{x,y}$  jest zdefiniowana jako suma dwóch ścieżek:  $\gamma_{x,y} = \gamma_{x,o} \cup \gamma_{o,y}$ , gdzie  $o$  jest punktem łączącym obie siatki. Natomiast  $\gamma_{x,o}$  definiujemy jako ścieżkę, która jest najbliższa euklidesowemu odcinkowi łączącemu  $x$  z  $o$ .

Jeżeli  $e = (u, v)$  jest krawędzią taką, że  $d(u, o) = i+1$  oraz  $d(v, o) = i$  ( $d$  oznacza odległość na grafie), to

$$\#\{x \in V : e \in \gamma_{x,o}\} \leq \frac{4(n+1)^2}{i+1}.$$

Ponadto (Uzasadnij!!!)

$$\#\{(x, y) \in V^2 : e \in \gamma_{x,y}\} \leq \frac{8(n+1)^2|V|}{i+1}.$$

Następnie definiujemy wagi: jeżeli  $d(o, e) = i$ , to  $\omega(e) = 1/(i+1)$ . Zatem dla  $x, y \in S$

$$|\gamma_{x,y}|_\omega \leq 2 \sum_{i=0}^{2n-1} \frac{1}{i+1} \leq 2 \log(2n+1).$$

Z wniosku 7.4 zachodzi nierówność Poincare ze stałą

$$\begin{aligned} C &= \max_e \left\{ \frac{1}{\omega(e)Q(e)} \sum_{\{x,y: e \in \gamma_{x,y}\}} |\gamma_{x,y}|_\omega \pi(x)\pi(y) \right\} \\ &\leq \max_{0 \leq i \leq n} \left\{ (i+1)4|V|2 \log(2n+1) \frac{\#\{x,y \in V : e \in \gamma_{x,y}\}}{|V|^2} \right\} \\ &\leq 64 \log(2n+1)n^2. \end{aligned}$$

Z drugiej strony weźmy  $f(x) = \text{sgn}(x) \log(1+d(0, x))$ , gdzie funkcja  $\text{sgn}(x)$  jest  $-1, 0$  lub  $1$ , zależnie od tego, czy  $x$  jest w lewym kwadracie,  $x = o$ , czy w prawym. Z symetrii  $\mathbb{E}_\pi(f) = 0$ .

Istnieje  $i+2$  punktów w odległości  $i$  od  $o$  w każdym kwadracie, stąd

$$\begin{aligned} \text{var}_\pi(f) &= \mathbb{E}_\pi(f^2) \geq \frac{1}{|V|} \sum_{i=0}^n (i+2) \log(i+1)^2 \\ &\geq \frac{n^2(\log n)^2}{2|V|} \end{aligned}$$

dla dużych  $n$ . Ponadto

$$\begin{aligned} \mathcal{E}(f) &= \frac{1}{4|V|} \sum_{i=0}^{2n-1} [(i+1) \wedge (2n-i+1)] (\log(i+2) - \log(i+1))^2 \\ &\leq \frac{1}{4|V|} \sum_{i=1}^{2n-1} \frac{1}{i+1} \\ &\leq \frac{\log(2n+1)}{4|V|}. \end{aligned}$$

Stąd

$$\gamma \leq \frac{\mathcal{E}(f)}{\text{var}_\pi(f)} \leq \frac{1}{2n^2 \log n}$$

dla dostatecznie dużych  $n$ . □

**7.3. Współczynnik zwężenia (ang. bottleneck ratio).** W tym rozdziale rozważamy nieredukowalny i aperiodyczny spacer losowy na grafie  $(E, V)$ . Chcemy opisać czas mieszania poprzez geometryczne własności grafu/przestrzeni stanów. Wówczas 'zwężenia' fragmentów grafu powodują, że niektóre z wierzchołków są ciężko osiągalne, co zwiększa czas mieszania.

Zacniemy od kilku definicji.

**Definicja 7.3.** Niech  $Q(x, y)$  będzie miarą krawędziową:

$$Q(x, y) = \pi(x)P(x, y).$$

Wtedy

$$Q(A, B) = \sum_{x \in A, y \in B} Q(x, y)$$

jest prawdopodobieństwem przejścia ze zbioru  $A$  do  $B$  w jednym kroku, startując z miary stacjonarnej.

**Współczynnikiem zwężenia** grafu nazywamy

$$\Phi_* = \min_{\{S: \pi(S) \leq 1/2\}} \Phi(S),$$

dla

$$\Phi(S) = \frac{Q(S, S^c)}{\pi(S)}.$$

Ponadto definiujemy **brzeg zbioru**

$$\partial S = \{(x, y) \in E : x \in S \text{ oraz } y \in S^c\}.$$

### Przykłady

- Dla prostego spaceru losowego na grafie

$$Q(x, y) = \begin{cases} \frac{\deg(x)}{2|E|} \frac{1}{\deg(x)} = \frac{1}{2|E|} & \text{jeżeli } (x, y) \in E \\ 0 & \text{jeżeli } (x, y) \notin E \end{cases}$$

Wtedy

$$\Phi(S) = \frac{\frac{|\partial S|}{2|E|}}{\sum_{x \in S} \frac{\deg(x)}{2|E|}} = \frac{|\partial S|}{\sum_{x \in S} \deg(x)}.$$

- Jeżeli spacer jest leniwy, to

$$Q(x, y) = \frac{1}{4|E|} \mathbf{1}_{\{(x, y) \in E\}} \quad \text{and} \quad \Phi(S) = \frac{2|\partial S|}{\sum_{x \in S} \deg(x)}.$$

- Jeżeli graf jest  $d$ -regularny,

$$\Phi(S) = \frac{|\partial S|}{d|S|}.$$

- Dla ostatniego przykładu z poprzedniego wykładu (dwie siatki kwadratowe połączone w jednym wierzchołku),  $\Phi_*$  jest rzędu  $1/n^2$  (przelicz)

Istnieją związki pomiędzy współczynnikiem zwężenia, czasami mieszania i dziurą spektralną

**Twierdzenie 7.6.** *Mamy*

$$t_{\text{miks}} = t_{\text{miks}}(1/4) \geq \frac{1}{4\Phi_*}$$

**Twierdzenie 7.7.**

$$\frac{\Phi_*^2}{2} \leq \gamma \leq 2\Phi_*$$

*Dowód.* Pokażemy jedynie prawą nierówność. W tym celu korzystamy z twierdzenia 7.1:

$$\gamma = \min_{\{f: \mathbb{E}_\pi(f)=0, f \neq 0\}} \frac{\mathcal{E}(f)}{\|f\|_2^2} = \min_{\{f: \mathbb{E}_\pi(f)=0, f \neq 0\}} \frac{\sum_{x, y \in \Omega} \pi(x)P(x, y)(f(x) - f(y))^2}{\sum_{x, y \in \Omega} \pi(x)\pi(y)(f(x) - f(y))^2}.$$

Dla każdego zbioru  $S$  takiego, że  $\pi(S) \leq 1/2$  zdefiniujemy funkcję

$$f_S(x) = \begin{cases} -\pi(S^c) & \text{for } x \in S, \\ \pi(S) & \text{for } x \notin S. \end{cases}$$

Dalej mamy

$$\gamma \leq \frac{2Q(S, S^c)}{2\pi(S)\pi(S^c)} \leq \frac{2Q(S, S^c)}{\pi(S)} = 2\Phi(S).$$

Biorąc infimum po wszystkich zbiorach  $S$  otrzymujemy prawą nierówność.  $\square$

Kilka dalszych przykładów

- **Leniwy spacer losowy na hiperkostce  $\mathbb{Z}_2^n$ .** Weźmy zbiór  $S = \{\mathbf{x} : x_1 = 0\}$ . Wówczas

$$\Phi(S) = 2 \sum_{x \in S, y \in S^c} 2^{-n} P(x, y) = 2^{-n+1} 2^{n-1} \cdot 1/(2n) = \frac{1}{2n}.$$

Zatem  $\Phi_* \leq 1/(2n)$ . Z drugiej strony wiemy, że  $\gamma = n^{-1}$ , więc z powyższego twierdzenia

$$\frac{1}{n} \leq 2\Phi_* \leq \frac{1}{n}.$$

Zatem  $2\Phi_* = n^{-1} = \gamma$ . Zatem górne szacowanie w powyższym twierdzeniu jest ostre.

- **Leniwy spacer losowy na  $\mathbb{Z}_{2n}$ .** Wiemy, że  $\gamma = \pi^2/(4n^2) + O(n^{-4})$ . Dla każdego zbioru  $S$ ,

$$\Phi(S) = \frac{|\partial S| \frac{1}{4} \frac{1}{2n}}{\frac{|S|}{2n}} = \frac{|\partial S|}{4|S|}.$$

Powyższa wartość przyjmuje minimum (na zbiorze  $\{S : \pi(S) \leq 1/2\}$ ) gdy  $S$  zawiera kolejne punkty i brzeg jest dwupunktowy. Wtedy, w dolnego szacowania w powyższym twierdzeniu  $\gamma \geq 1/(8n^2)$ . Zatem rząd jest poprawny.

## 8. EKSPANDERY

Jeżeli graf ma 'duże zwężenie', to czasy mieszania są duże. Teraz zajmiemy się sytuacją odwrotną, gdy zwężenie nie ma, a czasy mieszania są najszybsze z możliwych. Będziemy badać grafy, które są mocno połączone.

### 8.1. Definicja i podstawowe własności.

**Definicja 8.1.** Rodzina grafów  $\{G_n\}$  jest nazywana rodziną  $(d, \alpha)$ -ekspanderów jeżeli

- $\lim_{n \rightarrow \infty} |V(G_n)| = \infty$

oraz spełnione są następujące warunki dla każdego  $n$ :

- $G_n$  jest  $d$ -regularny,
- współczynnik zwężenia prostego spaceru losowego na  $G_n$  spełnia  $\Phi_*(G_n) \geq \alpha$ .

Przypomnijmy, że w tym przypadku

$$\Phi_*(G_n) = \min_{\{S : |S| \leq |V(G_n)|/2\}} \frac{|\partial S|}{d|S|}.$$

Poniższy lemat mówi, że ekspandery mają najszybsze możliwe czasy mieszania spośród grafów o ograniczonych stopniach wierzchołków

**Lemat 8.1.** *Jeżeli  $G_n$  jest  $(d, \alpha)$ -ekspanderem, to prosty spacer losowy na  $G_n$  spełnia*

$$t_{\text{miks}}(G_n) = O(\log |V(G_n)|).$$



*Dowód.* Przypomnijmy, że z twierdzenia 6.4

$$t_{\text{miks}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon\pi_{\min}}\right)t_{\text{rel}},$$

dla  $\pi_{\min} = \min_x \pi(x)$ . W przypadku ekspanderów  $\pi_{\min}$  jest rzędu  $1/|V(G_n)|$ . Ponadto w twierdzeniu 7.7 pokazaliśmy  $\gamma \leq 2\Phi_*$ . Stąd

$$t_{\text{miks}}(\varepsilon) \leq C \log(V(G_n)).$$

Z drugiej strony średnica grafu  $G_n$  jest rzędu  $\log(V(G_n))$ , więc

$$t_{\text{miks}}(\varepsilon) \geq \delta \log(V(G_n)).$$

□

**8.2. Istnienie.** Z powyższej definicji nie jest oczywiste istnienie ekspanderów. Bezpośrednie konstrukcje są dosyć trudne. Poniżej pokażemy probabilistyczny dowód istnienia.

**Twierdzenie 8.2.** *Istnieje rodzina  $(3; 0, 001)$  ekspanderów.*

*Dowód. Krok 1.* Zaczniemy od zdefiniowania rodziny losowych 3-regularnych grafów, które mogą posiadać wielokrotne krawędzie. Oznaczmy

$$A = \{a_1, \dots, a_n\} \quad B = \{b_1, \dots, b_n\}.$$

Niech

$$V(G_n) = A \cup B.$$

Wybermy losowe permutacje  $\sigma_1, \sigma_2 \in S_n$  (niezależnie, wg rozkładu jednostajnego) i definiujemy zbiór krawędzi

$$E(G_n) = \{(a_i, b_i), (a_i, b_{\sigma_1(i)}), (a_i, b_{\sigma_2(i)}) : 1 \leq i \leq n\}.$$

Zgodnie z definicją są to grafy dwudzielne, tzn. wszystkie krawędzie mają jeden koniec w zbiorze  $A$ , a drugi w zbiorze  $B$ .

**Krok 2.** Pokażemy, że

$$(8.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}[\Phi_*(G_n) > 0, 01] = 1.$$

Ustalmy  $\delta < 0, 03$ . Dla dowolnego zbioru wierzchołków  $S$ , przez  $N(S)$  oznaczmy zbiór sąsiadów  $S$ . Udowodnimy najpierw, że

$$(8.4) \quad \lim_{n \rightarrow \infty} \mathbb{P}[|N(S)| \leq (1 + \delta)|S| \text{ dla pewnego } S \subset A \text{ takiego, że } |S| \leq n/2] = 0.$$

Niech  $S$  będzie dowolnym podzbiorem  $A$  takim, że  $k = |S| \leq n/2$ . Oznaczmy przez  $I_S$  zbiór indeksów elementów zbioru  $S$ , tzn  $S = \{a_i\}_{i \in I(S)}$ . Wówczas  $S_B = \{b_i\}_{i \in I(S)}$  - podzbiór  $B$  o tych samych indeksach co  $S$ , zgodnie z konstrukcją grafu, jest podzbiorem  $N(S)$ . Jeżeli  $|N(S)| \leq (1 + \delta)k$ , to do zbioru  $S_B$  może zostać dołączony zbiór o rozmiarze  $\delta k$  taki, że  $N(S)$  jest zawarty w ich sumie. W szczególności obie losowe permutacje  $\sigma_1, \sigma_2$  przekształcają zbiór  $I_S$  w zbiór większy o co najwyżej  $\delta k$ . Zatem

$$\mathbb{P}[|N(S)| \leq (1 + \delta)k] = \frac{\binom{n-k}{\delta k} \binom{(1+\delta)k}{k}^2}{\binom{n}{k}^2} \leq \frac{\binom{n}{\delta k} \binom{(1+\delta)k}{k}^2}{\binom{n}{k}^2}$$

i stąd

$$\lim_{n \rightarrow \infty} \mathbb{P}[|N(S)| \leq (1 + \delta)|S| \text{ dla pewnego } S \subset A \text{ takiego, że } |S| \leq n/2] \leq \sum_{k=1}^{n/2} \binom{n}{k} \frac{\binom{n}{\delta k} \left(\frac{(1+\delta)k}{k}\right)^2}{\binom{n}{k}^2}.$$

Pokazuje się, że powyższe wyrażenie zbiega do 0 (zadanie, w dowodzie wystarczy użyć nierówności  $\binom{n}{\delta k} \leq n^{\delta k}/(\delta k)!$ ,  $\binom{n}{k} \geq n^k/k^k$  oraz  $j! > (j/e)^j$ ). To pokazuje (8.4), co z kolei jest równoważne

$$(8.5) \quad \lim_{n \rightarrow \infty} \mathbb{P}[|N(S)| \geq (1 + \delta)|S| \text{ dla każdego } S \subset A \text{ takiego, że } |S| \leq n/2] = 1.$$

Aby pokazać (8.3) musimy uzasadnić, że jeżeli każdy podzbiór  $A$  o rozmiarze  $k \leq n/2$  ma więcej niż  $(1 + \delta)k$  sąsiadów, to  $\Phi_* > \delta/2$ . Załóżmy, że powyższa własność jest spełniona i weźmy dowolny podzbiór  $S \subset V$  taki, że  $|S| \leq n$ . Niech  $A' = S \cap A$ ,  $B' = S \cap B$ . Możemy przyjąć, że  $|A'| \geq |B'|$ . Jeżeli  $|A'| \leq n/2$ , to z (8.5) zbiór  $A'$  ma co najmniej  $(\delta/2)|S|$  sąsiadów w  $B \setminus B'$  i te wszystkie krawędzie łączą  $S$  z  $S^c$ . Stąd

$$\Phi(S) = \frac{|\partial S|}{3|S|} \geq \frac{\delta}{6}.$$

Jeżeli  $|A'| \geq n/2$ , to bierzemy  $A'' \subset A'$ , którym składa się z  $\lceil n/2 \rceil$  elementów i powtarzamy rozumowanie. To pokazuje  $\Phi_* > \delta/6$  i dla  $\delta = 0,06$  daje (8.3).

**Krok 3. Konkluzja.** Z (8.3) dla każdego  $n$  możemy wybrać odpowiedni graf. Chcemy jednak skonstruować rodzinę bez wielokrotnych krawędzi. Najpierw pokażemy, że istnieje rodzina (deterministyczna) grafów  $\{G_n\}$  takich, że każdy z nich ma co najwyżej 100 podwójnych krawędzi, żadnej potrójnej i współczynnik zwężenia co najmniej 0,01.

W tym celu postępujemy następująco. Formuła (8.3) gwarantuje, że asymptotycznie prawie każdy graf, z rodziny skonstruowanej powyżej, ma współczynnik zwężenia co najmniej 0,01. Zauważmy, że istnienie potrójnej krawędzi rozspaja graf, a wtedy współczynnik zwężenia jest 0. Dalej, wartość oczekiwana i wariancja liczby podwójnych krawędzi dla każdej pary z trzech permutacji  $\{\text{id}, \sigma_1, \sigma_2\}$  wynosi 1 (to jest łatwe do policzenia). Zatem prawdopodobieństwo, że wszystkich podwójnych krawędzi jest powyżej 100 jest bardzo małe (nierówność Czebyszewa). Dla dużych  $n$  ten zbiór musi mieć więc niepusty przekrój ze zbiorem gdzie współczynnik zwężenia jest większy niż 0,01. Wybieramy więc graf z tego przekroju.



Pozostają do naprawienia podwójne krawędzie. Postępujemy jak na rysunku. Do pary wierzchołków  $a_i, b_j$  połączonych dwoma krawędziami dokładamy dwa nowe wierzchołki  $c_{ij}^1, c_{ij}^2$  oraz nowe krawędzie  $(a_i, c_{ij}^1), (a_i, c_{ij}^2), (b_i, c_{ij}^1), (b_i, c_{ij}^2), (c_{ij}^1, c_{ij}^2)$  (rysunek!). Nowy graf oznaczymy przez  $\tilde{G}_n$ . Tak zdefiniowane grafy tworzą rodzinę ekspanderów, a modyfikacja ma zaniedbywalny wpływ na współczynnik zwężenia.  $\square$

Bezpośrednie konstrukcje są bardzo trudne. Przytoczymy dwa niedawne wyniki (Kassabov, 2007):

- Istnieje  $k \in \mathbb{N}$  oraz  $\varepsilon > 0$  takie, że dla każdej nieabelowej skończonej grupy prostej<sup>8</sup> istnieje symetryczny zbiór jej generatorów o liczności co najwyżej  $k$ , taki, że odpowiedni graf Cayleya jest  $\varepsilon$ -ekspanderem.
- Istnieje  $k \in \mathbb{N}$  oraz  $\varepsilon > 0$  takie, że dla każdego  $n \geq 5$ , grupa symetryczna  $S_n$  posiada symetryczny zbiór generatorów o liczności co najwyżej  $k$ , taki, że odpowiedni graf Cayleya jest  $\varepsilon$ -ekspanderem.

**8.3. Zastosowania.** Informatyka: teoria informacji (automatyczna korekcja błędów w przesyłanej informacji); redukcja prawdopodobieństwa błędu w algorytmach probabilistycznych; algorytmy aproksymacyjne, ...

Matematyka: teoria liczb, teoria grup, ...

#### LITERATURA

- [BD] D. BAYER, P. DIACONIS. Trailing the Dovetail Shuffle To Its Lair. *Annals of Applied Probability*, 1992.
- [B] N. BERESTYCKI. *Lectures on Mixing Times*
- [D] P. DIACONIS. *Group Representations in Probability and Statistics*.
- [D1] P. DIACONIS. *The Markov Chain Monte Carlo Revolution*.
- [H] O. HAGGSTROM. *Finite Markov Chains and Algorithmic Applications*.
- [LPW] D. LEVIN, Y. PERES, E. WILMER. *Markov Chains and Mixing Times*.
- [M] N. MADRAS. *Lectures on Monte Carlo methods*.
- [Ma] B. MANN. How many times should you shuffle a deck of cards.
- [P] Y. PERES. *Mixing for Markov Chains and Spin Systems*.
- [SC] L. SALOFF-COSTE. *Random walks on finite groups*.

#### DODATEK A. PRZYKŁADY

Poniżej znajduje się lista przykładów omówionych podczas wykładu wraz z otrzymanymi szacowaniami

- Leniwy spacer losowy na  $Z_n$ .  
 $t_{\text{miks}} \leq n^2$  (przykład 2.4, strona 14).  
 $t_{\text{miks}} \geq \varepsilon n^2$  ([LPW], strona 96, przykład 7.4.1).
- Leniwy spacer losowy na hiperkostce  $\mathbb{Z}_2^n$ .  
 $t_{\text{miks}}(\varepsilon) \leq n \log n + \log(1/\varepsilon)n$  (przykład 2.6, strona 16).  
 $t_{\text{miks}}(\varepsilon) \leq n \log n + \log(1/\varepsilon)n$  (przykład 3.2 oraz wniosek 3.6, strona 19).  
 $t_{\text{miks}}(\varepsilon) \geq 1/2n \log n - Cn$  ([LPW], strona 95, prop. 7.13)  
 Optymalny współczynnik, to  $1/2$ .
- Leniwy spacer losowy na torusie  $\mathbb{Z}_n^d$   
 $t_{\text{miks}} \leq d^2 n^2$  (przykład 2.5, strona 15; stała może być poprawiona - lista nr 2)
- Tasowanie TOP-TO-RANDOM  
 $t_{\text{miks}}(\varepsilon) \leq n \log n + \log(1/\varepsilon)n$  (przykład 3.1 oraz wniosek 3.6, strona 19)  
 $t_{\text{miks}}(\varepsilon) \geq n \log n - \alpha n$  ([LPW], strona 96, przykład 7.4.2, oraz [D], strona 70)

---

<sup>8</sup>Grupa jest prosta, jeżeli nie zawiera nietrywialnej podgrupy normalnej, np.  $\mathbb{Z}_p$  dla liczby pierwszej  $p$ ,  $A_n$  dla  $n \geq 5$

## SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH

### LISTA NR 1

1. Niech  $\{X_0\}$  będzie łańcuchem Markowa o macierzy przejścia  $P$  i niech  $\mu_k$  będzie rozkładem  $X_k$ . Pokaż, że  $\mu_t = \mu_0 P^t$ .
2. (Funkcja łańcucha Markowa nie musi być łańcuchem Markowa.) Niech  $\{X_t\}$  będzie łańcuchem Markowa na przestrzeni stanów  $\Omega = \{s_1, s_2, s_3\}$ , z macierzą przejścia

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

i rozkładem początkowym  $\mu_0 = (1/3, 1/3, 1/3)$ . Dla każdego  $n$ , zdefiniujmy

$$Y_n = \begin{cases} 0 & \text{jeżeli } X_n = s_1 \\ 1 & \text{w przeciwnym razie} \end{cases}$$

Pokaż, że  $Y_n$  nie jest łańcuchem Markowa.

3. Pokaż, że jeżeli łańcuch Markowa jest nieredukowalny, to wszystkie punkty mają ten sam okres.
4. Jeżeli łańcuch Markowa jest nieredukowalny i aperiodyczny, to istnieje  $M$  takie, że  $P^t(x, y) > 0$  dla wszystkich  $x, y \in \Omega$  oraz  $t \geq M$ .
5. Pokaż, że nierozkładalny łańcuch Markowa posiada co najwyżej jedną miarę stacjonarną. **Wskazówka:** Niech  $P$  będzie macierzą przejścia, wówczas jedynymi funkcjami harmonicznymi (tzn. spełniającymi  $Ph = h$ ) są funkcje stałe ...
6. Uzasadnij, że jeżeli łańcuch Markowa ma dwie miary stacjonarne, to ma ich nieskończenie wiele.
7. (Inny dowód istnienia miary stacjonarnej) Niech  $P$  będzie macierzą przejścia łańcucha Markowa na skończonej przestrzeni stanów  $\Omega$ . Dla dowolnego rozkładu początkowego  $\mu$  na  $\Omega$  i  $n > 0$  definiujemy

$$\nu_n = \frac{1}{n}(\mu + \mu P + \dots + \mu P^{n-1}).$$

- Pokaż, że dla każdego  $x \in \Omega$  i  $n > 0$ ,

$$|\nu_n P(x) - \nu_n(x)| \leq \frac{2}{n}.$$

- Pokaż, że istnieje podciąg  $\{\nu_{n_k}\}$  taki, że  $\lim_{k \rightarrow \infty} \nu_{n_k}(x)$  istnieje dla każdego  $x \in \Omega$ .
  - Dla  $x \in \Omega$  zdefiniujmy  $\nu(x) = \lim_{k \rightarrow \infty} \nu_{n_k}(x)$ . Wykaż, że  $\nu$  jest miarą stacjonarną  $P$ .
8. (Ruina gracza) Gracz startuje z kapitałem  $k$ . W każdej rundzie rzuca monetą, jeżeli wypadnie orzeł wygrywa 1, w przeciwnym razie traci 1. Gracz kończy grę, gdy jego kapitał osiągnie poziom  $n$  lub zbankrutuje. Przedstaw odpowiedni spacer losowy na grafie.
    - Czy jest on nieredukowalny? Czy posiada miarę stacjonarną?
    - Oblicz prawdopodobieństwo zakończenia gry przez gracza z kapitałem  $n$ , tzn.  $\mathbb{P}_k[X_\tau = n]$ , gdzie  $\tau$  oznacza moment zakończenia gry.
    - Oblicz wartość oczekiwaną długości gry, tzn.  $\mathbb{E}_k \tau$ . Wskazówka: ułóż odpowiednie równanie rekurencyjne, a następnie je rozwiąż.
  9. (Kolekcjonowanie kuponów) Kolekcjoner zbiera kupony  $n$  różnych typów. Za każdym razem, gdy kupuje nowy kupon, rozkład ich typów jest jednostajny. Niech  $\tau$  oznacza moment, gdy zbierze pełną kolekcję.
    - Oblicz  $\mathbb{E} \tau$ . **Wskazówka:** niech  $X_t$  oznacza liczbę różnych typów zgromadzonych przez kolekcjonera w chwili  $t$

- Pokaż, że dla  $c > 0$

$$\mathbb{P}[\tau > n \log n + cn] \leq e^{-c}$$

**Wskazówka:** Niech  $A_i$  oznacza zdarzenie, że wśród pierwszych  $n \log n + cn$  kuponów nie było kuponu  $i$ -tego typu...

**10.** Niech  $\mu$  będzie miarą probabilistyczną na skończonej grupie  $G$ . Pokaż, że spacer losowy generowany przez  $\mu$  jest nieredukowalny wtedy i tylko wtedy, gdy  $S = \{g \in G : \mu(g) > 0\}$  generuje  $G$  (tzn. najmniejszą grupą zawierającą  $S$  jest  $G$ ).

**11.** Pokaż, że spacer losowy na  $G$  generowany przez  $\mu$  jest odwracalny wtedy i tylko wtedy, gdy miara  $\mu$  jest symetryczna, tzn.  $\mu(g) = \mu(g^{-1})$  dla każdego  $g$

**12.** Spacer losowy na torusie. Niech  $m, n \in \mathbb{N}$ . Rozważmy łańcuch Markowa na przestrzeni stanów

$$\Omega = \{(x, y) : x \in \{0, \dots, m-1\}, y \in \{0, \dots, n-1\}\}.$$

Jeżeli proces jest w stanie  $(x, y)$ , to przechodzi do stanów  $((x+1) \bmod m, y)$  lub  $(x, (y+1) \bmod n)$  z prawdopodobieństwem  $1/2$ .

- Pokaż, że ten łańcuch Markowa jest nieredukowalny.
- Pokaż, że proces jest aperiodyczny wtedy i tylko wtedy gdy  $\text{NWD}(m, n) = 1$ .

**13.** Skoczek znajduje się w rogu szachownicy. Zaczyna się poruszać w sposób losowy wybierając za każdym razem w sposób jednostajny jeden z dozwolonych ruchów. Jaka jest wartość oczekiwana liczby ruchów po których skoczek wróci do punktu startowego?

**14.** Rozważmy spacer losowy na  $\Omega = \{0, 1, \dots, n\}$ , w którym cząsteczka porusza się w lewo lub w prawo z prawdopodobieństwem  $1/2$ , za wyjątkiem punktów  $0$  i  $n$ . W punkcie  $n$  spacer pozostaje lub przechodzi do  $n-1$  z prawdopodobieństwem  $1/2$ . Stan  $0$  jest stanem absorbującym, a więc po trafieniu w niego cząsteczka pozostanie w nim na zawsze. Oblicz wartość oczekiwaną czasu trafienia w  $0$  cząsteczki, która startuje w punkcie  $n$ .

## SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH

### LISTA NR 2

1. Rozważmy prosty spacer losowy na  $\mathbb{Z}_n$ . Oznaczmy przez  $\tau$  pierwszy moment w którym zostały odwiedzone wszystkie stany. Oblicz  $\mathbb{E}\tau$ . **Wskazówka:** skorzystaj z zadania 9 z listy 1 o ruinie gracza.

2. W dwóch urnach znajduje się w sumie  $k$  kul ( $k$ -ustalona liczba). Definiujemy w następujący sposób łańcuch Markowa. W każdym kroku losujemy jednostajnie jedną z kul i przekładamy ją do sąsiedniej urny. Niech  $X_t$  oznacza liczbę kul znajdujących się w czasie  $t$  w pierwszej urnie. Wówczas  $X_t$  jest łańcuchem Markowa. Znajdź jego miarę stacjonarną. Pokaż, że proces jest odwracalny.

3. Oblicz  $\|\mu - \nu\|_{\text{TV}}$  dla następujących przykładów

- $\mu \sim \text{Bin}(4, 1/2)$ ,  $\nu \sim \text{Bin}(4, 1/4)$
- $\mu$  jest miarą jednostajną na  $S_n$  (grupie permutacji), a  $\nu$  jest miarą jednostajną na wszystkich permutacjach zachowujących 1.

4. Pokaż, że

$$d(t) = \sup_{\mu} \|\mu P^t - \pi\|_{\text{TV}},$$

$$\bar{d}(t) = \sup_{\mu, \nu} \|\mu P^t - \nu P^t\|_{\text{TV}},$$

gdzie  $\mu$  i  $\nu$  przebiegają po wszystkich miarach probabilistycznych na  $\Omega$ .

5. Pokaż, że

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}}$$

jest ciągiem malejącym.

6. Podczas wykładu pokazaliśmy, że dla leniwego spaceru na torusie  $\mathbb{Z}_n^d$  zachodzi  $t_{\text{miks}} \leq d^2 n^2$ . Pokaż jak otrzymać lepsze szacowanie  $t_{\text{miks}} \leq O(d \log d) n^2$ . W tym celu należy przeanalizować dowód podany na wykładzie. Niech  $t > k d n^2$ .

- Oszacuj prawdopodobieństwo zdarzenia, że w czasie  $t$  pierwsze współrzędne obu spacerów są różne, przez  $(1/4)^k$
- Wybierz odpowiednie  $k$  i rozważ wszystkie współrzędne.

7. (Tożsamość Walda) Niech  $\{X_n\}$  będzie ciągiem niezależnych i całkowalnych zmiennych losowych o tym samym rozkładzie i niech  $\tau$  będzie zmienną losową niezależną od tego ciągu o wartościach w  $\mathbb{N}$ . Wtedy

$$\mathbb{E} \left[ \sum_{i=1}^{\tau} X_i \right] = \mathbb{E} X_1 \cdot \mathbb{E} \tau$$

8. Niech  $\{X_t\}$  będzie nieredukowalnym i aperiodycznym łańcuchem Markowa z miarą stacjonarną  $\pi$ . Oznaczmy przez  $N(x, t)$  liczbę wizyt w punkcie  $x$  w pierwszych  $t$  krokach. Pokaż, że

$$\lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}[N(x, t)] = \pi(x)$$

oraz  $\frac{1}{t} N(x, t)$  zbiega do  $\pi(x)$  według prawdopodobieństwa.

9. Rozważmy aperiodyczny i nieredukowalny łańcuch Markowa z miarą stacjonarną  $\pi$ . Załóżmy, że istnieje stan  $x$  oraz zbiór  $A \subset \Omega$  taki, że  $\pi(A) = \sum_{y \in A} \pi(y) > 1/4$  oraz odległość  $x$  od zbioru  $A$  jest większa niż  $M$ . Pokaż, że  $t_{\text{miks}} \geq M$ .

10. Korzystając z poprzedniego zadania znajdź jak najlepsze dolne oszacowania na  $t_{\text{miks}}$  dla leniwego spaceru losowego na hiperkostce  $\mathbb{Z}_2^d$  oraz torusie  $\mathbb{Z}_n^d$

SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH

LISTA NR 3

1. Udowodnij

$$t_{\text{miks}}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{\text{miks}}.$$

2. Pokaż, że

$$t_{\text{miks}}(\varepsilon) \geq \frac{\log(|\Omega|(1-\varepsilon))}{\log \Delta},$$

gdzie  $\Delta = \max_{x \in \Omega} |\{y : P(x, y) > 0\}|$

3. Pokaż, że dla leniwego spaceru losowego na  $\mathbb{Z}_n$  zachodzi  $t_{\text{miks}} \geq \varepsilon n^2$ . **Wskazówka:** oszacuj  $\mathbb{P}[\sup_{t \leq \delta n^2} |S_t| > n/4]$ , gdzie  $S_t$  jest leniwym spacerem losowym na  $\mathbb{Z}$ .

4. Niech

$$\Omega = \{x \in \{0, 1\}^{n+1} : x(n+1) = 1\}.$$

Spacer losowy zdefiniowany jest następująco. W kroku  $t+1$  wybieramy losowo jedną ze współrzędnych  $k$  ze zbioru  $\{1, \dots, k\}$  i jeżeli  $x_t(k+1) = 1$ , to zmieniamy wartość  $k$ -tej współrzędnej, definiując w ten sposób  $X_{t+1}$ . Znajdź miarę stacjonarną. Pokaż, że  $t_{\text{miks}} \geq n^2 - 2n^{3/2}$ .

5. Znajdź silny czas jednostajny dla zmodyfikowanego leniwego spaceru losowego na  $\mathbb{Z}_2^n$  takiego, że będąc w stanie  $X_t$  pozostajemy w nim z prawdopodobieństwem  $1/(n+1)$ , a z prawdopodobieństwem  $1/(n+1)$  przechodzi do jednego z sąsiadów.

6. Niech  $G$  będzie skończoną grupą i  $\mu$  miarą probabilistyczną na  $G$ . Załóżmy, że istnieje  $\varepsilon > 0$  takie, że

$$\mathbb{P}[X_1 \in A] \geq \varepsilon U(A)$$

dla każdego  $A \subset G$  (przypomnijmy  $X_0 = e$ ), gdzie  $U$  jest miarą jednostajną na  $G$ . Znajdując odpowiedni silny czas stacjonarny, pokaż że

$$\|\mu_t - U\| \leq (1-\varepsilon)^t,$$

gdzie  $\mu_t$  jest rozkładem  $X_t$ . **Wskazówka:** skorzystaj z rozkładu  $\mu = \varepsilon U + (1-\varepsilon)\tilde{\mu}$ .

7. Rozważmy graf powstały przez sklejenie w jednym wierzchołku dwóch pełnych grafów. Dla każdego wierzchołka dodajmy pętle, tak, aby nowy graf był regularny. Oszacuj z dołu i z góry czas mieszania  $t_{\text{miks}}$ .

8. Rozważmy następujący sposób tasowania kart (odwrotny do TOP-TO-RANDOM). Wyciągamy (losowo) jedną z kart z talii i kładziemy ją na górze talii. Oszacuj czas mieszania  $t_{\text{miks}}$ .

9. Niech  $Y$  oznacza ostatni stan oznaczony przez prosty spacer losowy na  $\mathbb{Z}_n$  startujący z 0. Pokaż, że  $Y$  ma rozkład jednostajny na zbiorze  $\{1, \dots, n-1\}$ .

10. Skonstruuj silny czas zatrzymania dla leniwego spaceru na  $\mathbb{Z}_{2^k}$ , a następnie oblicz jego wartość oczekiwaną. **Wskazówka:** postępuj przez indukcję.

**Referat.** 11. ([LPW], strony 56 i 69) Opowiedz o przykładzie *winning streak*.

**Referat.** 12. ([LPW], strona 96) Pokaż, że w tasowaniu TOP-TO-RANDOM

$$t_{\text{miks}}(\varepsilon) \geq n \log n - \alpha n.$$

**Referat.** 13. ([LPW], strona 95) Przedstaw dolne oszacowania czasu mieszania leniwego spaceru losowego na hiperkostce.

## SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH

### LISTA NR 4

**1.** Niech  $P$  będzie macierzą przejścia. Pokaż, że  $P$  ma reprezentację w postaci odwzorowań losowych, tzn. istnieje ciąg  $\{Z_t\}$  iid oraz funkcja deterministyczna  $f$  takie, że

$$X_0 = x, \quad X_t = f(X_{t-1}, Z_t)$$

jest łańcuchem Markowa z macierzą przejścia  $P$  startującym w  $x$ .

**2.** Niech  $X_t$  będzie spacerem losowym na grupie  $G$  generowanym przez miarę probabilistyczną  $\mu$ . Zdefiniujmy miarę  $\tilde{\mu}(g) = \mu(g^{-1})$  na  $G$  i niech  $\tilde{X}_t$  będzie spacerem losowym generowanym przez  $\tilde{\mu}$ . Oznaczmy przez  $\mu_t$  i  $\tilde{\mu}_t$  rozkłady  $X_t$  i  $\tilde{X}_t$ . Pokaż, że

$$\|\mu_t - U\|_{\text{TV}} = \|\tilde{\mu}_t - U\|_{\text{TV}},$$

gdzie  $U$  jest miarą jednostajną na grupie  $G$ .

**3.** Chcemy wygenerować losową permutację  $\sigma \in S_n$ , tzn funkcję różnowartościową  $\sigma : \{1, \dots, n\} \mapsto \{1, \dots, n\}$ . Stosujemy następujący algorytm. Dla każdego  $i$  podstawiamy za  $\sigma(i)$  losową liczbę ze zbioru  $\{1, \dots, n\}$ . Kolejne losowania są niezależne i jednostajne. Jeżeli otrzymana funkcja jest permutacją, to kończymy. W przeciwnym razie powtarzamy procedurę. Oszacuj oczekiwaną liczbę losowań niezbędnych do wygenerowania permutacji.

**4.** Rozważmy następujący algorytm generujący losową permutację  $\sigma \in S_n$ . Przyjmujemy  $\sigma_0 = Id$ . Następnie dla  $k = 1, \dots, n-1$  postępujemy indukcyjnie. Mając zadane  $\sigma_{k-1}$  losujemy liczbę  $J_k$  jednostajnie spośród  $\{k, \dots, n\}$ . Definiujemy  $\sigma_k$ , zamieniając wartości  $\sigma_{k-1}(k)$  z  $\sigma_{k-1}(J_k)$ , tzn.  $\sigma_k(k) = \sigma_{k-1}(J_k)$ ,  $\sigma_k(J_k) = \sigma_{k-1}(k)$ , a dla pozostałych wartości  $\sigma_k(i) = \sigma_{k-1}(i)$ . Pokaż, że  $\sigma_{n-1}$  ma rozkład jednostajny na  $S_n$ .

**5.** Załóżmy, że w poprzednim zadaniu liczby  $J_k$  są losowane ze zbioru  $\{1, \dots, n\}$ . Czy  $\sigma_{n-1}$  ma rozkład jednostajny na  $S_n$ ?

**6.** Czy prawdziwe jest następujące zdanie: niech  $Q$  będzie rozkładem na  $S_n$  i niech  $\sigma$  będzie losową permutacją o rozkładzie  $Q$ , wówczas jeżeli

$$\mathbb{P}[\sigma(i) > \sigma(j)] = 1/2,$$

dla każdych  $i, j$ , to  $Q$  jest rozkładem jednostajnym na  $S_n$ .

**7.** Czy prawdziwe jest następujące zdanie: niech  $Q$  będzie rozkładem na  $S_n$  i niech  $\sigma$  będzie losową permutacją o rozkładzie  $Q$ , wówczas jeżeli

$$\mathbb{P}[\sigma(i) = j] = 1/n,$$

dla każdych  $i, j$ , to  $Q$  jest rozkładem jednostajnym na  $S_n$ .

**8.** Niech  $Q$  będzie rozkładem na grupie  $S_n$  odpowiadającym jednemu tasowaniu RIFFLE SHUFFLE. Pokaż, że

$$Q(\pi) = \begin{cases} (n+1)/2^n & \text{jeżeli } \pi = id \\ 1/2^n & \text{jeżeli } \pi \text{ ma dwa ciągi rosnące} \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

**9.** Zdefiniujmy  $a$ -RIFFLE SHUFFLE, jako tasowanie gdzie talię kart dzielimy na  $a$ -zbiorów o zbliżonym rozmiarze, a następnie ... tasujemy (możemy myśleć o tasowaniu RIFFLE SHUFFLE, takim, że dysponujemy  $a$  rękoma). Podobnie jak na wykładzie zdefiniuj 3 równoważne modele, a następnie spróbuj oszacować niezbędną liczbę tasowań dla  $a = 4$  i  $a = 8$

**10.** (Geometryczna interpretacja RIFFLE SHUFFLE). Wylosujmy (niezależnie i jednostajnie)  $n$  punktów z odcinka  $[0, 1]$ . Oznaczmy je rosnąco  $x_1, \dots, x_n$ . Następnie zastosujmy odwzorowanie piekarza  $x \mapsto 2x \pmod{1}$  do tych punktów. Pokaż, że indukowana permutacja punktów ma dokładnie taki sam rozkład jak w tasowaniu RIFFLE SHUFFLE.



**11.** Pokaż, że wykonanie  $a$ -RIFFLE SHUFFLE, a następnie  $b$ -RIFFLE SHUFFLE jest równoważne tasowaniu  $ab$ -RIFFLE SHUFFLE. **Wskazówka:** możesz skorzystać z poprzedniego zadania (ale nie jest to konieczne).

**12.** Niech  $Q_a$  będzie rozkładem  $S_n$  po jednym tasowaniu  $a$ -RIFFLE SHUFFLE. Pokaż, że  $Q_a(\pi) = \binom{n+a-r(\pi)}{n} / a^n$ , gdzie  $r(\pi)$  jest liczbą rosnących ciągów w  $\pi$ .

**Referat. 13.** ([LPW], strony 101-106) Opowiedz o tasowaniu przez losowe permutacje.

## SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH

### LISTA NR 5

1. Uzasadnij, że spacer losowy podany na wykładzie generujący losowe drzewo rozpinające jest nieredukowalny.
2. Rozważmy problem pakowania plecaka dla  $m$  przedmiotów z wagami 1 i plecaka o rozmiarze  $b = m/3$ . Przez  $S$  oznaczamy zbiór dopuszczalnych pakowań plecaka, tzn.  $S = \{z \in \{0, 1\}^m : (w, z) \leq b\}$ . Dla  $m = 100$  oszacuj prawdopodobieństwo, że losowo wygenerowany wektor  $z \in \{0, 1\}^m$  jest elementem  $S$ .
3. Rozważmy graf  $G = (V, E)$ , gdzie  $V = \{1, \dots, N\}$  i  $E = \{(i, i + 1)\}$ . Oblicz liczbę poprawnych konfiguracji w modelu hardcore
4. Dla grafu z poprzedniego zadania oblicz liczbę poprawnych  $q$ -kolorowań.
5. Oblicz liczbę poprawnych  $q$ -kolorowań dla pełnego drzewa binarnego o głębokości  $n$ .
6. Niech  $G = (V, E)$  będzie spójnym grafem i niech  $X$  będzie losowym  $q$ -kolorowaniem grafu (niekoniecznie poprawnym). Pokaż, że prawdopodobieństwo, że  $X$  jest właściwym  $q$ -kolorowaniem grafu jest mniejsze niż  $(\frac{q-1}{q})^{|V|-1}$ .
7. Niech  $G = (V, E)$  będzie spójnym grafem. Oznaczmy przez  $\Delta$  maksymalny stopień wierzchołka. Niech  $q = \Delta + 1$ . Pokaż, że wówczas istnieje przynajmniej jedno poprawne  $q$ -kolorowanie grafu. Uzasadnij również, że spacer losowy podany na wykładzie nie musi być w tym przypadku nieredukowalny.
8. Niech  $q = \Delta + 2$  jak w poprzednim zadaniu. Uzasadnij, że spacer losowy podany na wykładzie musi być w tym przypadku nieredukowalny.
9. Zdefiniuj problem komiwojażera i wyjaśnij jak można znaleźć jego przybliżone rozwiązanie przy pomocy algorytmu Metropolis.
10. Pokaż, że łańcuch Markowa zdefiniowany na wykładzie dla modelu hardcore jest aperiodyczny, nieredukowalny i jego miara stacjonarna jest  $\pi_G$ .
11. Pokaż, że miarę stacjonarną łańcuch Markowa zdefiniowanego na wykładzie dla  $q$ -kolorowania jest  $\pi_{G,q}$ .
12. Pokaż, że miarę stacjonarną dynamiki Glaubera jest  $\pi$ .
13. Zapoznaj się z 'hardcore model with fugacity' ([LPW], rozdział 3.3.4), a następnie pokaż, że

$$t_{\text{miks}}(\varepsilon) \leq cn(\log n + \log(1/\varepsilon))$$

14. Dla  $G = (V, E)$  podzbiór wierzchołków  $I \subset V$  nazywa się niezależnym w  $G$  jeśli żadne dwa wierzchołki z  $I$  nie są połączone krawędzią w  $G$ . Niech  $\mathcal{I}(G)$  oznacza wszystkie zbiory niezależne w  $G$ . Podaj przykład łańcucha, którego rozkład stacjonarny jest rozkładem jednostajnym na  $\mathcal{I}(G)$ .

## SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH

### LISTA NR 6

**1.** Niech  $\pi$  będzie dowolną miarą probabilistyczną na hiperkostce  $\mathbb{Z}_2^d$  taką, że  $\pi(x) > 0$  dla każdego  $x \in \mathbb{Z}_2^d$ . Zdefiniujemy następujący łańcuch Markowa na  $\mathbb{Z}_2^d$ . Niech  $X_t = x$ . W chwili  $t + 1$  losujemy jednostajnie jedną ze współrzędnych  $k$ . Niech  $x'$  będzie elementem  $\mathbb{Z}_2^d$  różniącym się od  $x$  tylko na  $k$ -tej pozycji. Definiujemy  $X_{t+1} = x'$  z prawdopodobieństwem  $\min\{1, \frac{\pi(x')}{\pi(x)}\}$  i  $X_{t+1} = x$  z prawdopodobieństwem  $1 - \min\{1, \frac{\pi(x')}{\pi(x)}\}$ . Pokaż, że  $\pi$  jest miarą stacjonarną tego łańcucha Markowa.

**2. Model Isinga** jest to losowe uporządkowanie spinów na zadanym grafie  $G = (V, E)$ , tzn. losowe przyporządkowanie wierzchołkom grafu 1 lub  $-1$ . Wybieramy więc losowo element  $\{-1, 1\}^V$ , który będziemy oznaczać przez  $\sigma$ . Fizycznie interpretuje się ten model jako rozmieszczenie w wierzchołkach grafu magnesów o orientacjach  $+1$  lub  $-1$ .

Energią systemu jest funkcja

$$H(\sigma) = - \sum_{\substack{v, w \in V \\ v \sim w}} \sigma(v)\sigma(w).$$

Rozkładem Gibbsa odpowiadającym energii  $H$  jest rozkład

$$\pi(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}$$

( $Z(\beta)$  jest stałą normalizującą).

Wyjaśnij jak wygenerować zmienną o rozkładzie  $\pi$ . Oblicz macierz przejścia.

**3.** Niech  $P$  będzie macierzą przejścia skończonego łańcucha Markowa. Jeżeli  $\lambda$  jest wartością własną  $P$ , to  $|\lambda| \leq 1$ .

**4.** Jeżeli  $P$  jest nieredukowalna, to przestrzeń własna odpowiadająca 1 jest jednowymiarowa.

**5.** Jeżeli  $P$  jest nieredukowalna i aperiodyczna, to  $-1$  nie jest wartością własną  $P$ .

**6.** Załóżmy, że łańcuch Markowa jest nieredukowalny. Pokaż, że  $\{t : P^t(x, x) > 0\} \subset 2\mathbb{Z}$  wtedy i tylko wtedy gdy  $-1$  jest wartością własną  $P$ .

**7.** Załóżmy, że łańcuch Markowa jest nieredukowalny. Pokaż, że  $\{t : P^t(x, x) > 0\} \subset k\mathbb{Z}$  wtedy i tylko wtedy gdy pierwiastek  $k$ -tego stopnia z 1 jest wartością własną  $P$ .

**8.** Niech  $P$  będzie nieredukowalna. Załóżmy, że macierz  $A$  spełnia  $0 \leq A(i, j) \leq P(i, j)$  oraz  $A \neq P$ . Pokaż, że każda wartość własna  $\lambda$  macierzy  $A$  spełnia  $|\lambda| < 1$ .

**9.** Niech  $\tilde{P} = \frac{1}{2}P + \frac{1}{2}I$  będzie macierzą przejścia leniwej wersji spaceru losowego generowanego przez macierz przejścia  $P$ . Pokaż, że wszystkie wartości własne  $\tilde{P}$  są nieujemne.

## SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH

### LISTA NR 7

1. Pokaż, że jeżeli  $P$  jest macierzą przejścia odwracalną względem pewnej miary  $\pi$ , to

$$P^{2t+2}(x, x) \leq P^t(x, x).$$

2. W dwóch urnach znajduje się w sumie  $n$  kul. W każdym kroku losujemy jednostajnie jedną z kul i wkładamy ją do losowej urny (z prawdopodobieństwem  $1/2$ ). Wyznacz miarę stacjonarną tego procesu. Oszacuj czas mieszania  $t_{\text{miks}}$ .

3. Rozważmy spacer losowy na  $\{0, 1, \dots, n\}$  taki, że

$$P(k, k-1) = P(k, k+1) = \frac{1}{2} \quad \text{dla } k = 1, \dots, n-1$$

oraz  $P(0, 1) = P(n, n-1) = 1$ . Znajdź wartości własne i funkcje własne  $P$ .

4. Rozważmy spacer losowy na  $0, 1, \dots, n$  taki, że

$$P(k, k-1) = P(k, k+1) = \frac{1}{2} \quad \text{dla } k = 1, \dots, n-1$$

oraz

$$P(0, 1) = P(0, 0) = P(n, n-1) = P(n, n) = \frac{1}{2}.$$

Znajdź wartości własne i funkcje własne  $P$ .

5. Używając metody spektralnej oszacuj czas mieszania dla prostego spaceru losowego na torusie  $\mathbb{Z}_n \times \mathbb{Z}_m$

6. Używając metody spektralnej oszacuj czas mieszania dla leniwego spaceru losowego na  $\mathbb{Z}_3^n$  (zostajemy w aktualnym wierzchołku z prawdopodobieństwem  $1/3$ ).

7. Niech  $\Omega$  będzie przestrzenią metryczną z metryką  $\rho$  i niech  $P$  będzie macierzą przejścia pewnego łańcucha Markowa na  $\Omega$ . Załóżmy, że istnieje stała  $\theta < 1$  taka, że dla każdego  $x, y \in \Omega$  istnieje parowanie  $(X_1, Y_1)$  miar  $P(x, \cdot)$ ,  $P(y, \cdot)$  spełniające

$$\mathbb{E}_{x,y}[\rho(X_1, Y_1)] \leq \theta \rho(x, y).$$

Pokaż, że jeżeli  $\lambda \neq 1$  jest wartością własną  $P$ , to  $|\lambda| \leq \theta$ . W szczególności  $\gamma_* \geq 1 - \theta$ .

8. Wywnioskuj z poprzedniego zadania, że w algorytmie generującym losowe poprawne  $q$ -kolorowanie

$$\gamma_* \geq \frac{1}{3n\Delta}.$$

9. Rozważmy na hiperkostce  $\mathbb{Z}_2^n$  odległość Hamminga  $\rho(\mathbf{x}, \mathbf{y}) = \sum |x_i - y_i|$ . Uzasadnij, że dla leniwego spaceru losowego  $\gamma_* \geq 1/n$ .

**Referat.** 10. ([LPW], strona 96) Pokaż, że w tasowaniu TOP-TO-RANDOM

$$t_{\text{miks}}(\varepsilon) \geq n \log n - \alpha n.$$

**Referat.** 11. ([LPW], strona 95) Przedstaw dolne oszacowania czasu mieszania leniwego spaceru losowego na hiperkostce.

SPACERY LOSOWE NA SKOŃCZONYCH GRAFACH  
KLASÓWKA

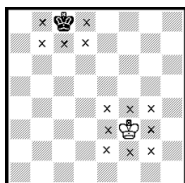
**Rozgrzewka:**

**R1.** Zdefiniuj pojęcie miary stacjonarnej.

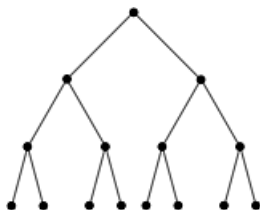
**R2.** Zdefiniuj normę całkowitego wahania. Podaj 2 równoważne definicje.

1. Rozważmy leniwy spacer losowy na  $\mathbb{Z}_3$ , taki, że w każdym kroku pozostajemy z prawdopodobieństwem  $1/2$  w aktualnym wierzchołku i z prawdopodobieństwem  $1/4$  przechodzimy do jednego z sąsiadów. Podaj miarę stacjonarną. Uzasadnij odpowiedź.

2. Król znajduje się w rogu szachownicy. Zaczyna się poruszać w sposób losowy wybierając za każdym razem w sposób jednostajny jeden z dozwolonych ruchów. Jaka jest wartość oczekiwana liczby ruchów po których król wróci do punktu startowego? Dla przypomnienia: król może przemieścić się na jedno sąsiadujących pól



3. Rozważamy prosty spacer losowy na pełnym drzewie binarnym o głębokości  $n$ . Jest to drzewo z korzeniem, w którym każdy wierzchołek, do głębokości  $n$ , ma 2 potomków. Poniżej przykład pełnego drzewa binarnego o głębokości 3.



- Czy ten spacer jest odwracalny?
- Podaj miarę stacjonarną
- Załóżmy, że spacer losowy startuje w korzeniu. Niech  $\tau$  będzie pierwszym momentem, w którym odwiedzony zostanie jeden z liści (tzn. wierzchołek bez potomków). Oblicz  $\mathbb{E}\tau$ .

4. Niech  $X_t$  będzie spacerem losowym na grupie  $G$  generowanym przez miarę probabilistyczną  $\mu$ . Zdefiniujmy miarę  $\tilde{\mu}(g) = \mu(g^{-1})$  na  $G$  i niech  $\tilde{X}_t$  będzie spacerem losowym generowanym przez  $\tilde{\mu}$ . Oznaczmy przez  $\mu_t$  i  $\tilde{\mu}_t$  rozkłady  $X_t$  i  $\tilde{X}_t$ . Pokaż, że

$$\|\mu_t - U\|_{\text{TV}} = \|\tilde{\mu}_t - U\|_{\text{TV}},$$

gdzie  $U$  jest miarą jednostajną na grupie  $G$ .

5. Oblicz  $\|\mu - \nu\|_{\text{TV}}$  dla następujących przykładów

- $\mu \sim \text{Bin}(4, 1/2)$ ,  $\nu \sim \text{Bin}(4, 1/4)$
- $\mu$  jest miarą jednostajną na  $S_n$  (grupie permutacji), a  $\nu$  jest miarą jednostajną na wszystkich permutacjach zachowujących 1.