Uniwersytet Wrocławski
Wydział Matematyki i Informatyki
Instytut Matematyczny
specjalność: Analiza Danych

Tamara Frączek

Czasy silnie stacjonarne i łańcuchy Markowa

Praca licencjacka
napisana pod kierunkiem
prof. dr. hab. Dariusza Buraczewskiego

Wrocław 2021

Wprowadzenie

Niniejsza praca podejmuje temat znajdowania czasów mieszania łańcuchów Markowa. Została zainspirowana pytaniem: ile razy należy przetasować talię kart sposobem „riffle shuffle", aby można było ją uznać za potasowaną? Problem ten został opisany przez Persiego Diaconisa w czwartym rozdziale książki „Group representations in probability and statistics" [3] i rozwiązany za pomocą *metody czasów silnie stacjonarnych*. Jednak dowody tam zamieszczone są raczej pobieżne i nieformalne. Celem pracy jest sformalizowanie rozważań P. Diaconisa oraz precyzyjne opisanie kilku innych przykładów zastosowania metody czasów silnie stacjonarnych.

Łańcuch Markowa jest to ciąg zmiennych losowych $\{X_t\}$, taki że X_{t+1} nie zależy od $X_{t-1}, \ldots, X_0$. Pierwszy element tego ciągu X_0 nazywamy *punktem początkowym*. Przy pewnych warunkach, dla łańcucha Markowa istnieje jedyny rozkład π, do którego zbiegają rozkłady X_t. Nazywany go *rozkładem stacjonarnym*. Dla danego łańcucha Markowa określa się *czas mieszania* t_{mix}, czyli moment, po którym (czyli dla $t \geqslant t_{mix}$) rozkład X_t jest odpowiednio blisko rozkładu stacjonarnego niezależnie od punktu początkowego X_0.

Najważniejszym teoretycznym wynikiem pracy jest twierdzenie 2.16, pozwalające oszacować odległość między rozkładami X_t oraz π. Wykorzystuje się w nim tzw. *czas silnie stacjonarny*, którego zdefiniowaniu poświęcony jest podrozdział 2.1. Twierdzenie to mówi, że odległość między rozkładami X_t oraz π jest nie większa niż prawdopodobieństwo zdarzenia, że czas silnie stacjonarny dla łańcucha jest większy niż t. Rezultat ten jest podstawą wspomnianej na wstępie metody czasów silnie stacjonarnych. Główną część pracy stanowią zastosowania tej metody w konkretnych przykładach. W każdym z nich, określając odpowiedni czas silnie stacjonarny oraz stosując twierdzenie 2.16, otrzymujemy oszacowanie na czas mieszania.

Strukturę pracy podporządkowano jej celowi. W rozdziale 1 definiujemy łańcuchy Markowa oraz wybrane ich proste własności. Następnie podajemy (bez dowodów) kilka twierdzeń przydatnych w późniejszych rozważaniach. Podane w tym rozdziale definicje i przykłady są zaczerpnięte z [2] oraz [4].

W rozdziale 2 wyjaśniamy na czym polega metoda czasów silnie stacjonarnych. W tym celu wprowadzamy potrzebne pojęcia oraz udowodniamy serię rezultatów prowadzących do dowodu twierdzenia 2.16. Ta część rozważań jest przygotowana na podstawie rozdziału 6 z [4], w szczególności podrozdziałów 6.2, 6.3 i 6.4.

W rozdziale 3 prezentujemy cztery zastosowania metody czasów silnie stacjonarnych. W podrozdziale 3.1 dotyczącym tasowania „top-to-random" korzystamy z wyników podrozdziału 6.1 z [4], lematu 4.4.2 z [3] oraz twierdzenia 1 z rozdziału 4 z [5]. W podrozdziale 3.2 dotyczącym „leniwego" spaceru po n-kostce korzystamy z przykładów 6.3, 6.4 oraz 6.5.2 z [4]. W podrozdziale 3.3 dotyczącym „leniwego" spaceru po 2^k-cyklu korzystamy z przykładu 6.5.5 oraz zadań 6.6 i 6.8 z [4]. W podrozdziale 3.4 dotyczącym tasowania „riffle shuffle" korzystamy z wyników lematów 2 i 3 oraz twierdzenia 7 z rozdziału 4 z [3] oraz z wyników podrozdziału 8.3 z [4].

Spis treści

1 Podstawowe definicje i twierdzenia

W tym rozdziale zdefiniujemy łańcuchy Markowa oraz przytoczymy bez dowodów kilka podstawowych faktów i definicji potrzebnych w dalszych rozważaniach.

Definicja 1.1. Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie przestrzenią probabilistyczną. Ciąg zmiennych losowych $\{X_t\}$ określonych na Ω o wartościach w skończonym lub przeliczalnym zbiorze $\mathcal{S}$ (przestrzeni stanów) nazywamy **łańcuchem Markowa** z macierzą przejścia P, jeśli dla każdych $x_0, \dots, x_t \in \mathcal{S}$ mamy

$$\mathbb{P}[X_t = x_t | X_0 = x_0, \dots, X_{t-1} = x_{t-1}] = \mathbb{P}[X_t = x_t | X_{t-1} = x_{t-1}] = P(x_{t-1}, x_t).$$

Łańcuch Markowa jest to proces bez pamięci, to znaczy taki, że X_{t+1} zależy wyłącznie od X_t. Kolejne kroki procesu są determinowane przez macierz P. Macierz ta jest przykładem macierzy stochastycznej.

Definicja 1.2. Macierz $P = \{P(x, y)\}_{x,y \in \mathcal{S}}$ nazywamy **macierzą stochastyczną**, jeżeli jej wszystkie wyrazy są nieujemne oraz

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

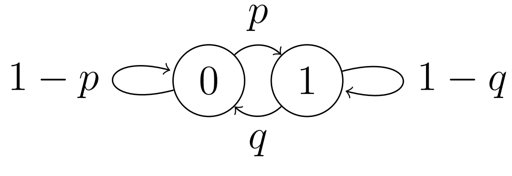
W niniejszej pracy będziemy rozważać wyłącznie łańcuchy takie, dla których $\mathcal{S}$ jest zbiorem skończonym.

Rozkład początkowy X_0 oznaczmy przez μ. Będziemy rozważać zależność X_t od początkowego rozkładu. Standardowo zapisuje się to jako $\mathbb{P}_\mu$, $\mathbb{E}_\mu$. My jednak rozważać będziemy wyłącznie przypadki takie, że $\mu = \delta_x$, czyli miara początkowa to miara skupiona w danym punkcie $x \in \mathcal{S}$. Zastosujemy więc oznaczenia $\mathbb{P}_x$, $\mathbb{E}_x$. Wtedy

$$\mathbb{P}_x(X_t = y) = \mathbb{P}_{\delta_x}(X_t = y) = \mathbb{P}(X_t = y | X_0 = x).$$

Rozważmy przykład łańcucha Markowa. Zaprezentujemy w ten sposób wyniki, do których dążymy w tym rozdziale.

Przykład 1.3. Rozważmy spacer losowy na $\mathcal{S} = \{0, 1\}$ zilustrowany poniższym grafem (Rysunek 1).



Rysunek 1: Graf, po którym porusza się spacer z prawdopodobieństwami przejścia

Mamy dwie monety, pierwszą o prawdopodobieństwie wyrzucenia orła p i drugą o prawdopodobieństwie wyrzucenia orła q. Cząsteczka startuje w punkcie 0. W każdym ruchu spacer działa następująco.

- Jeżeli cząsteczka znajduje się w punkcie 0, to rzucamy pierwszą monetą. Jeśli wypadnie orzeł, to cząsteczka przemieszcza się do punktu 1, w przeciwnym wypadku pozostaje w 0.
- Jeżeli cząsteczka znajduje się w punkcie 1, to rzucamy drugą monetą. Jeśli wypadnie orzeł, to cząsteczka przemieszcza się do punktu 0, w przeciwnym wypadku pozostaje w 1.

Dla rozważanego spaceru

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

W jakim celu definiujemy dla łańcuchów macierz P? Oznaczmy jako μ_t rozkład zmiennej X_t, tzn. w naszym przypadku

$$\mu_t = (\mathbb{P}(X_t = 0), \mathbb{P}(X_t = 1)),$$

wtedy

$$\mu_{t+1} = \begin{pmatrix} \mu_{t+1}(0) \\ \mu_{t+1}(1) \end{pmatrix} = \begin{pmatrix} (1-p)\mu_t(0) + q\mu_t(1) \\ p\mu_t(0) + (1-q)\mu_t(1) \end{pmatrix} = \mu_t P. \tag{2}$$

Co pociąga

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

W naszym przypadku $\mu_0 = (1, 0)$. Jak się okazuje (1) i (2) są to ogólne własności łańcuchów Markowa.

Jednym z ważniejszych pytań, dotyczących łańcuchów Markowa jest: jak wygląda zachowanie X_t (czyli μ_t) dla dużych wartości t? Pokażemy, że, dla rozważanego łańcucha μ_t zbiega do pewnej miary π, co więcej, zbieżność ta jest wykładniczo szybka.

Zdefiniujmy miarę π na $\mathcal{S}$ jako

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

Jest to taka miara, że

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

Łatwo zauważyć, że jeżeli granica μ_t istnieje, to musi spełniać powyższą własność. W związku z tym, miary spełniające tę własność będą dla nas szczególnie ważne.

Aby pokazać, że granica $\lim_t \mu_t = \pi$, zdefiniujmy

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

Wówczas

$$\varepsilon_{t+1} = \mu_t(0)(1-p) + (1-\mu_t(0))q - \pi(0) = (1-p-q)\varepsilon_t,$$

więc, jeśli $0 < p+q < 2$, to

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

oraz zbieżności te są wykładnicze.

Łańcuchy Markowa na skończonych przestrzeniach generalnie posiadają podobne do opisanych w przykładzie 1.3 własności, a więc zbiegają do pewnej miary π, nazywanej miarą stacjonarną, a zbieżność ta jest wykładniczo szybka. Takie wyniki wymagają założenia, że łańcuch jest nieredukowalny oraz nieokresowy. Te pojęcia teraz zdefiniujemy.

Definicja 1.4. Stan y nazywamy **osiągalnym** ze stanu x, jeżeli dla pewnego n zachodzi $P^n(x, y) > 0$. Inaczej mówiąc, y jest osiągalny z x, jeżeli z dodatnim prawdopodobieństwem możemy przejść z x do y. Oznaczamy to $x \to y$.

Definicja 1.5. Stany x i y nazywamy **wzajemnie się komunikującymi**, jeżeli $x \to y$ i $y \to x$ (a wieć można przejść zarówno z x do y, jak i z y do x).

Uwaga 1.6. Obie relacje osiągalności i wzajemnej przechodniości są przechodnie.

Definicja 1.7. Łańcuch Markowa nazywamy **nierozkładalnym (nieredukowalnym)**, jeżeli każde jego dwa stany komunikują się nawzajem (a więc z dowolnego stanu można osiągnąć każdy inny).

Definicja 1.8. Miarę probabilistyczną nazywamy **miarą stacjonarną (rozkładem stacjonarnym)** łańcucha Markowa (ekwilibrum, stanem równowagi), jeżeli $\pi = \pi P$, czyli

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

Oznacza to, że jeśli X_t ma rozkład π, to również wszystkie kolejne X_s $s \geqslant t$ mają rozkład π.

Dla danego procesu może nie istnieć żaden, istnieć jeden lub więcej niż jeden rozkład stacjonarny. Jak zobaczymy za chwilę, dla dowolnego łańcucha nieredukowalnego miara ta jest jedyna.

Twierdzenie 1.9. *Każdy nieredukowalny łańcuch Markowa na skończonej przestrzeni stanów posiada jedyną miarę stacjonarną. Ponadto $\pi(x) > 0$ dla każdego $x \in S$.*

Przykład 1.10. **(Spacery losowe na grupach)** Niech ν oznacza dowolną miarę probabilistyczną na skończonej grupie G. Rozważać będziemy lewe spacery losowe na grupie G, które są zadane przez macierz przejścia

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

Zatem, jest to taki łańcuch Markowa, w którym prawdopodobieństwo przejścia ze stanu g do h zależy wyłącznie od wartości miary na elemencie $h \circ g^{-1}$. W związku z tym, taki łańcuch wygodnie jest przedstawić w postaci

$$X_t = g_t \dots g_1 X_0,$$

gdzie $g_1, \ldots, g_t$ to niezależne zmienne losowe o rozkładzie ν. $g_1, \ldots, g_t$ nazywamy **przyrostami**.

Jeżeli U jest jednostajną miarą probabilistyczną na G, czyli $U(g) = \frac{1}{|G|}$, to dla dowolnego spaceru losowego na G, U jest miarą stacjonarną:

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

Naszym kolejnym celem jest pokazanie, że rozkład X_t zbiega do miary stacjonarnej. Do tego potrzebujemy jednak dodatkowego założenia nieokresowości, co możemy zilustrować przykładem.

Przykład 1.11. Niech n będzie liczbą parzystą. Rozważamy spacer losowy na $\mathcal{S} = \mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ zdefiniowany następująco: $X_{t+1} = X_t \pm 1 \pmod{n}$ z prawdopodobieństwem $\frac{1}{2}$. Można zauważyć, że okres każdego punktu jest równy 2, ponieważ jedynie w parzystej liczbie ruchów jesteśmy w stanie wrócić do każdego punktu. Co więcej, dla tego łańcucha prawdopodobieństwo przejścia z danego punktu do niego samego będzie wynosiło 0. Stąd, skoro miarą stacjonarną tego łańcucha jest miara jednostajna, to rozkład X_t nie może być zbieżny.

Można uniknąć takiej sytuacji rozważając tak zwany „leniwy spacer", gdzie z prawdopodobieństwem $\frac{1}{2}$ spacer pozostaje w tym samym miejscu, a z prawdopodobieństwami $\frac{1}{4}$ przechodzi do jednego z punktów sąsiednich.

Definicja 1.12. Niech $T(x) = \{t : P^t(x,x) > 0\}$, wówczas $\mathrm{NWD}\{T(x)\}_{x \in \mathcal{S}}$ nazywane jest **okresem** stanu x. Łańcuch Markowa jest **nieokresowy**, jeżeli okres każdego stanu wynosi 1.

Lemat 1.13. *Jeżeli łańcuch Markowa jest nieredukowalny, to wszystkie punkty mają ten sam okres.*

Wniosek 1.14. *Jeżeli łańcuch Markowa jest nieredukowalny i istnieje stan $x \in S$, którego okres wynosi 1, to proces jest nieokresowy.*

Twierdzenie 1.15. *Jeżeli $\{X_t\}_{t=0}^{\infty}$ jest nieredukowalnym i nieokresowym łańcuchem Markowa o mierze stacjonarnej π, to istnieją $q < 1$ i C takie, że dla każdych $x, y \in S$, $n \in \mathbb{N}$*

$$|P^t(x,y) - \pi(y)| \leqslant Cq^t.$$

Od tego momentu będziemy zatem rozważać wyłącznie łańcuchy nieredukowalne oraz nieokresowe.

2 Metoda czasów silnie stacjonarnych

2.1 Czas silnie stacjonarny

W niniejszym podrozdziale zdefiniujemy czas silnie stacjonarny. W tym celu potrzebne jednak będą pomocnicze definicje. Wprowadzimy je od najsłabszych do najmocniejszych.

Definicja 2.1. Dla ciągu zmiennych losowych $\{X_t\}$ określonych na $(\Omega, \mathcal{F}, \mathbb{P})$, zmienna losowa $\tau : \Omega \to \{0, 1, \ldots, \infty\}$ jest nazywana **momentem zatrzymania (momentem stopu)**, jeśli dla każdego $t \in \{0, 1, \ldots\}$ istnieje $B_t \subset \Omega^{t+1}$, taki że

$$\{\tau = t\} = \{(X_0, \ldots, X_t) \in B_t\}.$$

To znaczy, że zmienna τ jest momentem zatrzymania wtedy i tylko wtedy, gdy funkcja $\mathbf{1}_{\{\tau=t\}}$ jest funkcją wektora $(X_0, \ldots, X_t)$.

Intuicyjnie, τ jest „rozsądną” regułą zatrzymania pewnego procesu, ponieważ decyzję podejmujemy na podstawie wiedzy z przeszłości i teraźniejszości, a nie przyszłości.

Przykładem momentu zatrzymania może być instrukcja „trzymamy akcje spółki tak długo, aż ich wartość osiągnie 100 złotych”. Natomiast momentem zatrzymania nie będzie polecenie „sprzedajemy akcje, gdy po raz ostatni ich wartość przekroczy 100 złotych”, ponieważ nie znamy przyszłości i nie potrafimy w danej chwili określić kiedy będzie ostatni raz.

Aby zdefiniować kolejne pojęcie, silniejsze niż moment stopu, potrzebne będzie pojęcie reprezentacji przez losowe odwzorowania.

Definicja 2.2. **Reprezentacja przez losowe odwzorowania** macierzy stochastycznej P na przestrzeni stanów $\mathcal{S}$ to para (f, Z), gdzie $f : \mathcal{S} \times \Lambda \to \mathcal{S}$, a Z jest zmienną losową przyjmującą wartości w Λ, taką że

$$\mathbb{P}(f(x, Z) = y) = P(x, y) \quad \forall x, y \in \mathcal{S}.$$

Jaki jest związek definicji 2.2 z łańcuchem Markowa? Mając rozkład X_0 oraz reprezentację przez losowe odwzorowania można skonstruować łańcuch.

Uwaga 2.3. Jeśli $Z_1, Z_2, \ldots$ to ciąg niezależnych zmiennych losowych o rozkładzie takim jak Z, a X_0 ma rozkład μ, to ciąg $(X_0, X_1, \ldots)$ zdefiniowany jako

$$X_n = f(X_{n-1}, Z_n) \quad \text{dla } n \geqslant 1$$

jest łańcuchem Markowa z macierzą przejścia P z rozkładem początkowym μ.

Można się spotkać z definicjami, które właśnie ciąg $(Z_1, Z_2, \ldots)$ nazywają reprezentacją przez losowe odwzorowania.

Reprezentacja przez losowe odwzorowania często pomaga zrozumieć, jak wygląda krok spaceru. Ponadto ułatwia formalny zapis. Spróbujemy to zilustrować w kolejnym przykładzie.

Przykład 2.4. Wróćmy na chwilę do przykładu 1.11.

Klasyczny spacer po n-cyklu można opisać następująco. W każdym kroku rzucamy monetą. Jeśli wypadnie orzeł, to idziemy do kolejnego wierzchołka zgodnie ze wskazówkami zegara. Jeśli wypadnie reszka - przeciwnie. Dzięki takiemu opisowi łatwo stworzymy reprezentację przez losowe odwzorowania takiego spaceru:

$$\Lambda = \{-1, 1\}, \quad Z \sim U(\Lambda), \quad f(x, z) = x + z \pmod n.$$

Jak można zdefiniować reprezentację przez losowe odwzorowania dla leniwego spaceru po n-cyklu? Przykładową modyfikacją poprzedniej reprezentacji może być

$$\Lambda = \{-1, 1\} \times \{0, 1\}, \quad Z \sim U(\Lambda), \quad f(x, (z_1, z_2)) = x + z_1 \cdot z_2 \pmod n.$$

Tym razem dwa razy rzucamy monetą - raz aby zdecydować do którego wierzchołka chcemy się ruszyć, drugi raz - aby zdecydować, czy chcemy się do niego ruszyć.

Twierdzenie 2.5. *Dla każdej macierzy stochastycznej P na skończonej przestrzeni stanów istnieje reprezentacja przez losowe odwzorowania.*

Uwaga 2.6. Dla danego łańcucha może istnieć wiele reprezentacji przez losowe odwzorowania.

Definicja 2.7. Zmienną losową τ nazywamy **zrandomizowanym momentem zatrzymania** dla łańcucha Markowa $\{X_t\}$, jeżeli jest ona momentem zatrzymania dla ciągu $\{Z_t\}$ z reprezentacji przez losowe odwzorowania.

Uwaga 2.8. Ciąg $\{Z_t\}$ zawiera więcej informacji niż $\{X_t\}$, jako że $\{X_t\}$ jest funkcją $\{Z_t\}$. W związku z tym moment zatrzymania $\{X_t\}$ jest zrandomizowanym momentem zatrzymania. Natomiast zrandomizowany moment zatrzymania dla $\{X_t\}$ niekoniecznie jest momentem zatrzymania.

Z powyższej uwagi wynika, że aby udowodnić, że dana zmienna losowa jest zrandomizowanym momentem zatrzymania wystarczy udowodnić, że jest momentem zatrzymania (o ile nim jest). W niektórych przykładach rozdziału 3 można by z tego ułatwienia skorzystać. My będziemy jednak udowadniać bycie zrandomizowanym momentem zatrzymania wprost z definicji, ponieważ definiowanie reprezentacji przez losowe odwzorowania ułatwi formalizację dowodów.

Definicja 2.9. Niech $\{X_t\}$ będzie nieredukowalnym łańcuchem Markowa z miarą stacjonarną π. Zmienną losową τ nazywamy **czasem stacjonarnym** dla $\{X_t\}$, jeżeli jest zrandomizowanym momentem zatrzymania (możliwe, że zależnym od pozycji początkowej x), takim że X_τ ma rozkład π, tzn. $\mathbb{P}_x(X_\tau = y) = \pi(y)$.

Definicja 2.10. Niech $\{X_t\}$ będzie nieredukowalnym łańcuchem Markowa z miarą stacjonarną π. Zmienną losową τ nazywamy **czasem silnie stacjonarnym** dla $\{X_t\}$, jeżeli jest zrandomizowanym momentem zatrzymania (możliwe, że zależnym od pozycji początkowej x), takim że

$$\mathbb{P}_x(\tau = t, X_t = y) = \mathbb{P}_x(\tau = t)\pi(y). \tag{3}$$

Inaczej mówiąc, X_τ ma rozkład π, czyli dla $t \geqslant \tau$ X_t mają rozkład π oraz X_τ jest niezależny od τ.

Lemat 2.11. *Niech $\{X_t\}$ będzie łańcuchem Markowa z miarą stacjonarną π. Jeżeli τ jest czasem silnie stacjonarnym dla $\{X_t\}$, to dla każdego $t \geqslant 0$ zachodzi*

$$\mathbb{P}_x(\tau \leqslant t, X_t = y) = \mathbb{P}_x(\tau \leqslant t)\pi(y) \tag{4}$$

Dowód. Udowodnimy, że dla każdego $s \leqslant t$

$$\mathbb{P}_x(\tau = s, X_t = y) = \mathbb{P}_x(\tau = s)\pi(y),$$

co zsumowane stronami po wszystkich całkowitych $0 \leqslant s \leqslant t$ da tezę.

Zacznijmy od rozpisania lewej strony ($s \leqslant t$).

$$\begin{aligned}
\mathbb{P}_x(\tau = s, X_t = y) &= \sum_{z\in\mathcal{S}} \mathbb{P}_x(\tau = s, X_t = y, X_s = z) \\
&= \sum_{z\in\mathcal{S}} \mathbb{P}_x(X_t = y|\tau = s, X_s = z)\mathbb{P}_x(\tau = s, X_s = z) \\
&= \sum_{z\in\mathcal{S}} \mathbb{P}_x(X_t = y|\tau = s, X_s = z)\mathbb{P}_x(\tau = s)\pi(z).
\end{aligned}$$

Skupimy się teraz na wyrazie $\mathbb{P}_x(X_t = y|\tau = s, X_s = z)$.

Niech $\{Z_t\}$ oznacza ciąg zmiennych losowych z definicji zrandomizowanego momentu zatrzymania dla łańcucha $\{X_t\}$, a f jest funkcją z tej definicji.

Ponieważ τ jest czasem zatrzymania dla $\{Z_t\}$, to istnieje takie $B \subseteq \Lambda^s$, że $\{\tau = s\} = \{(Z_1, \ldots, Z_s) \in B\}$.

Co więcej, dla całkowitych $r, s \geqslant 0$ istnieje funkcja $\tilde{f}_r : \mathcal{S} \times \Lambda^r \to \mathcal{S}$, taka że

$$X_{s+r} = \tilde{f}_r(X_s, Z_{s+1}, \ldots, Z_{s+r}).$$

Można ją indukcyjnie zdefiniować jako $\tilde{f}_1(X_s, Z_{s+1}) = f(X_s, Z_{s+1}) = X_{s+1}$ oraz

$$\tilde{f}_r(X_s, Z_1, \ldots, Z_{s+r}) = f(\tilde{f}_{r-1}(X_s, Z_1, \ldots, Z_{s+r-1}), Z_{s+r}).$$

Korzystając z tych faktów (dla $r = t - s \geqslant 0$) oraz niezależności wektorów $(Z_1, \ldots, Z_s)$ i $(Z_{s+1}, \ldots, Z_t)$:

$$\begin{aligned}
&\mathbb{P}_x(X_t = y|\tau = s, X_s = z) = \\
&\qquad = \mathbb{P}_x(\tilde{f}_{t-s}(X_s, Z_{s+1}, \ldots, Z_t) = y|(Z_1, \ldots, Z_s) \in B, X_s = z) \\
&= \mathbb{P}_x(\tilde{f}_{t-s}(z, Z_{s+1}, \ldots, Z_t) = y|(Z_1, \ldots, Z_s) \in B, X_s = z) = P^{t-s}(z, y). \qquad (5)
\end{aligned}$$

Czyli

$$\mathbb{P}_x(\tau = s, X_t = y) = \mathbb{P}_x(\tau = s)\left(\sum_{z\in\mathcal{S}} P^{t-s}(z,y)\pi(z)\right) = \mathbb{P}_x(\tau = s)\pi(y),$$

co kończy dowód. $\square$

2.2 Metoda czasów silnie stacjonarnych

Celem tego podrozdziału jest udowodnienie nierówności do szacowania odległości między rozkładami X_t oraz π. Na koniec zdefiniujemy formalnie czas mieszania.

Zacznijmy od wprowadzenia miary odległości dwóch rozkładów, dzięki której będziemy mogli oceniać, jak blisko miary stacjonarnej jest rozkład X_t.

Definicja 2.12. Dla dwóch miar probabilistycznych μ i ν na Ω definiujemy **normę całkowitego wahania**

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

Uwaga 2.13.
- Oznaczenie TV bierze się od słów „total variation” (po ang. całkowita zmienność).
- Odległość między dwoma miarami probabilistycznymi jest równa maksymalnej różnicy ich wartości przyporządkowanej pojedynczemu wydarzeniu.
- Wielkość ta przyjmuje wartości między 0 a 1. Ponadto $\|\mu-\nu\|_{TV} = 0$ wtedy i tylko wtedy, gdy $\mu = \nu$.

Fakt 2.14. Dla dowolnych miar probabilistycznych na Ω zachodzi

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)| = \sum_{\substack{x\in\Omega,\\ \mu(x)\geqslant\nu(x)}} [\mu(x) - \nu(x)]$$

Dowód. Ustalmy $A \subseteq \Omega$ oraz $B = \{x \in \Omega : \mu(x) \geqslant \nu(x)\}$. Wtedy

$$\mu(A) - \nu(A) \leqslant \mu(A\cap B) - \nu(A\cap B) \leqslant \mu(B) - \nu(B).$$

Pierwsza nierówność wynika z własności miar

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

oraz z definicji B

$$\mu(A\cap B^c) - \nu(A\cap B^c) = \sum_{x\in A\cap B^c} (\mu(x) - \nu(x)) \leqslant 0.$$

Aby uzyskać drugą nierówność przyprowadzamy analogiczne rozumowanie korzystając z tego, że $B = (A \cap B) \cup (A^c \cap B)$ i

$$\mu(A^c \cap B) - \nu(A^c \cap B) = \sum_{x \in A^c \cap B} (\mu(x) - \nu(x)) \geqslant 0.$$

Zamieniając w powyższym rozumowaniu μ i ν kolejnością dostajemy (wtedy musimy zamienić też B na B^c)

$$\nu(A) - \mu(A) \leqslant \nu(B^c) - \mu(B^c) = 1 - \nu(B) - (1 - \mu(B)) = \mu(B) - \nu(B).$$

Podsumowując

$$|\mu(A) - \nu(A)| \leqslant \mu(B) - \nu(B),$$

czyli (bo równość dostajemy dla $A = B$)

$$\|\mu - \nu\|_{TV} = \mu(B) - \nu(B) = \sum_{\substack{x \in \Omega, \\ \mu(x) \geqslant \nu(x)}} [\mu(x) - \nu(x)].$$

Otrzymaliśmy w ten sposób drugą równość z tezy. Pierwszą równość możemy udowodnić korzystając w ostatniej równości z faktu, że

$$\mu(B) - \nu(B) = \frac{1}{2}(\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)) = \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

□

Definicja 2.15. Odległość pomiędzy rozkładami X_t i π zdefiniujemy jako

$$d(t) := \max_{x \in \mathcal{S}} \|P^t(x, \cdot) - \pi\|_{TV}.$$

Twierdzenie 2.16. *Jeżeli τ jest czasem silnie stacjonarnym dla łańcucha Markowa X_t z macierzą przejścia P i miarą stacjonarną π, to*

$$d(t) = \max_{x \in \mathcal{S}} \|P^t(x, \cdot) - \pi\|_{TV} \leqslant \max_{x \in \mathcal{S}} \mathbb{P}_x(\tau > t).$$

Wygodnie będzie rozbić dowód tego twierdzenia na dwa lematy. Pomocniczo zdefiniujmy wielkość $s_x(t)$.

Definicja 2.17.

$$s_x(t) := \max_{y \in \mathcal{S}} \left[1 - \frac{P^t(x,y)}{\pi(y)}\right] \qquad s(t) := \max_{x \in \mathcal{S}} s_x(t)$$

Uwaga 2.18. $s_x(t)$ to inna możliwa definicja odległości pomiędzy $P^t(x,\cdot)$ i π. W literaturze można ją znaleźć pod nazwą „separation distance". Lemat 2.21 pokazuje zależność między dwiema definicjami dystansu.

Lemat 2.19. *Przy założeniach takich jak w twierdzeniu 2.16 zachodzi*

$$s_x(t) \leqslant \mathbb{P}_x(\tau > t). \tag{6}$$

Dowód. Ustalamy $x \in \mathcal{S}$. Wtedy dla dowolnego $y \in \mathcal{S}$

$$1 - \frac{P^t(x,y)}{\pi(y)} = 1 - \frac{\mathbb{P}_x(X_t = y)}{\pi(y)} \leqslant 1 - \frac{\mathbb{P}_x(X_t = y, \tau \leqslant t)}{\pi(y)}. \tag{7}$$

Z lematu 2.11 (bo τ - czas silnie stacjonarny) prawa strona nierówności wynosi

$$1 - \frac{\pi(y)\mathbb{P}_x(\tau \leqslant t)}{\pi(y)} = 1 - \mathbb{P}_x(\tau \leqslant t) = \mathbb{P}_x(\tau > t).$$

□

Uwaga 2.20. Dla łańcucha Markowa $\{X_t\}$ z ustalonym punktem początkowym $X_0 = x$ oraz dla jego momentu zatrzymania τ możemy zdefiniować **punkt postojowy** y jako taki, że jeśli $X_t = y$, to $\tau \leqslant t$. Przykład punktu postojowego zaprezentujemy w przykładzie o „leniwym" spacerze po n - kostce w uwadze 3.19.

Łatwo zauważyć, że w nierówności w (7) (czyli również w (6)) równość zachodzi wtedy i tylko wtedy, gdy y jest punktem postojowym dla punktu początkowego x.

Można udowodnić, że w powyższym twierdzeniu dla każdego łańcucha da się zdefiniować taki czas silnie stacjonarny, że zachodzi równość.

Lemat 2.21. *Przy założeniach takich jak w twierdzeniu 2.16 zachodzi*

$$\|P^t(x,\cdot) - \pi\|_{TV} \leqslant s_x(t),$$

czyli $d(t) \leqslant s(t)$.

Dowód. Korzystając z faktu 2.14 oraz z tego, że π jest miarą probabilistyczną otrzymujemy

$$\begin{aligned}\|P^t(x,\cdot) - \pi\|_{TV} &= \sum_{\substack{y\in\mathcal{S},\\ P^t(x,y)<\pi(y)}} [\pi(y) - P^t(x,y)] = \sum_{\substack{y\in\mathcal{S},\\ P^t(x,y)<\pi(y)}} \pi(y)\left[1 - \frac{P^t(x,y)}{\pi(y)}\right] \\ &\leqslant \max_{y\in\mathcal{S}}\left[1 - \frac{P^t(x,y)}{\pi(y)}\right] = s_x(t).\end{aligned}$$

Drugą nierówność z tezy dowodzimy poprzez nałożenie na obie strony pierwszej nierówności $\max_{x\in\mathcal{S}}$. □

Oczywiście nierówności otrzymane w lematach 2.19 i 2.21 składają się na dowód twierdzenia 2.16.

Na koniec tego podrozdziału zdefiniujmy czas mieszania.

Definicja 2.22. **Czasem mieszania** (ang. mixing time) nazywamy taki moment, gdy odległość łańcucha Markowa od miary stacjonarnej jest odpowiednio mała, czyli

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leqslant \varepsilon\}.$$

Dla ustalenia uwagi definiuje się

$$t_{mix} := t_{mix}\left(\frac{1}{4}\right).$$

Czas mieszania to kluczowy parametr, który będziemy badać, ponieważ daje on odpowiedź na pytanie, ile kroków dany spacer musi wykonać, aby można było uznać, że ma rozkład stacjonarny. Przykładowo, czas mieszania informuje ile razy musimy potasować talię, aby można było uznać ją za potasowaną.

Zauważmy, że twierdzenie 2.16 daje nam metodę szacowania czasu mieszania z góry. Ograniczając prawą stronę nierówności w twierdzeniu przez ε będziemy próbowali obliczyć optymalne t. Nie mamy narzędzi, aby wyznaczyć dokładną wartość $t_{mix}(\varepsilon)$, jedynie jej szacowanie z góry. W naszych rozważaniach jest to jednak informacja wystarczająca.

3 Przykłady

3.1 Tasowanie kart metodą „top-to-random"

Definicja 3.1. Rozważymy następującą (powolną) metodę tasowania talii złożonej z n kart. Bierzemy kartę na górze i wkładamy ją w dowolne miejsce w talii (na n możliwych sposobów z takim samym prawdopodobieństwem). Czynność powtarzamy. Proces taki nazwiemy **tasowaniem „top-to-random"**.

Lemat 3.2. *Takie tasowanie możemy utożsamić z łańcuchem Markowa takim, że $\mathcal{S} = \mathcal{S}_n$ (zbiór wszystkich permutacji zbioru $\{1, \ldots, n\}$). Jest to łańcuch nieredukowalny i nieokresowy. Miarą stacjonarną takiego łańcucha jest miara jednostajna.*

Dowód. Ustawienia n kart możemy utożsamić z permutacjami n elementowego zbioru, więc przestrzenią stanów naszego procesu będzie $\mathcal{S}_n$. Ustalmy, że numerujemy karty od szczytu talii.

Łatwo zauważyć, że ustawienie kart w czasie t zależy wyłącznie ustawienia w czasie $t-1$. Aby formalnie zdefiniować macierz P tego łańcucha, opiszemy go jako składanie permutacji. Zdefiniujmy

$$\Lambda = \left\{ \begin{pmatrix} 1 & 2 & \ldots & k \\ k & 1 & \ldots & k-1 \end{pmatrix} : k \in \{1, \ldots, n\} \right\}.$$

Jest to podzbiór $\mathcal{S}_n$ takich permutacji, które odpowiadają możliwym tasowaniom. k oznacza numer miejsca, w które wkładamy kartę z góry.

Teraz już łatwo zdefiniujemy macierz przejścia P. Dla wszystkich $x, y \in \mathcal{S}_n$

$$P(x,y) = \begin{cases} \frac{1}{n} & \text{jeżeli istnieje } \lambda \in \Lambda \text{ taka, że } \lambda \circ x = y, \\ 0 & \text{w przeciwnym wypadku} \end{cases}.$$

Z ustalonego układu kart możemy przejść do dokładnie n różnych układów (bo Λ zawiera n różnych permutacji).

Łańcuch jest nieredukowalny, ponieważ Λ jest generatorem grupy $\mathcal{S}_n$, czyli w skończenie wielu krokach można otrzymać każdą permutację początkowego układu. Λ jest generatorem, ponieważ

$$\begin{pmatrix} 1 & 2 & \dots & k \\ k & 1 & \dots & k-1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & \dots & k+1 \\ k+1 & 1 & \dots & k \end{pmatrix}^k = \begin{pmatrix} k & k+1 \\ k+1 & k \end{pmatrix},$$

które już w oczywisty sposób są generatorami.

Jest to łańcuch nieokresowy, bo dla każdego $x \in \mathcal{S}_n$ $P(x,x) = \frac{1}{n}$, czyli okres każdego stanu jest równy 1. Jego miarą stacjonarną jest $\pi(y) = \frac{1}{n!}$, co wynika z przykładu 1.10. $\square$

Definicja 3.3. Dla $\{X_t\}$ opisanego w lemacie 3.2 definiujemy zmienną losową $\tau_{top} : \mathcal{S}_n \to \{1, 2, \dots, \infty\}$ jako moment pierwszego ruchu po tym jak karta K znajdzie się na szczycie talii.

Twierdzenie 3.4. *τ_{top} jest zrandomizowanym momentem zatrzymania dla łańcucha zdefiniowanego w lemacie 3.2.*

Dowód. Najpierw znajdujemy reprezentację przez losowe odwzorowania dla wspomnianego łańcucha. Naturalnym wyborem jest $Z \sim U(\Lambda)$ (gdzie Λ jest zdefiniowane jak w lemacie 3.2) oraz $f(x,z) = z \circ x$.

Wtedy

$$\tau_{top} = \min\{t : Z_{t-1} \circ \dots \circ Z_1 \text{ jest taką permutacją, że } n \text{ przechodzi na } 1\}.$$

Ponieważ $\mathbf{1}_{\{\tau_{top}=t\}}$ jest funkcją wektora $(Z_1, \dots, Z_t)$, dlatego τ_{top} jest zrandomizowanym momentem zatrzymania. $\square$

Udowodnimy, że τ_{top} jest czasem silnie stacjonarnym dla rozważanego łańcucha. Będziemy potrzebować do tego poniższego lematu pomocniczego.

Lemat 3.5. *Niech $\{X_t\}$ będzie łańcuchem Markowa zdefiniowanym w lemacie 3.2. Niech K oznacza kartę, która przed rozpoczęciem tasowania jest na samym dole talii. Jeżeli w danym momencie t oznaczymy liczbę kart pod K jako k, to wszystkie $k!$ układów tych kart są jednakowo prawdopodobne.*

Dowód. Intuicyjnie można myśleć o tym następująco. Gdy pod kartą K znajduje się dokładnie jedna karta stwierdzenie jest oczywiste. Następnie postępujemy indukcyjnie. Zakładamy, że pod kartą K znajduje się k kart i wszystkie ich układy są równie prawdopodobne. W takiej sytuacji układ kart pod K zmienia jedynie włożenie nowej karty pod K. Jeżeli nową kartę wkładamy pod K, to w każde z $k+1$ miejsc z jednakowym prawdopodobieństwem. Czyli również po włożeniu nowej karty wszystkie układy kart pod K są jednakowo prawdopodobne.

Sformalizujmy tę intuicję. Najpierw wprowadźmy dodatkowe oznaczenia:

- T_i - czas (liczba ruchów) dopóki i kart znajdzie się pod K.
- X_t^k - ciąg oznaczający k dolnych kart w danym momencie.
- $\widehat{X_t^k}$ - zbiór kart z X_t^k.

Chcemy udowodnić, że dla każdego k i każdego t zachodzi

$$\mathbb{P}_x(X_t^k = s | T_k \leqslant t < T_{k+1}, \widehat{X_t^k} = \widehat{s}) = \frac{1}{k!}.$$

To znaczy, że jeżeli w danym momencie t mamy k kart pod K (czyli $T_k \leqslant t < T_{k+1}$), to wszystkie ustawienia tych kart są równie prawdopodobne.

Krok pierwszy Zauważmy najpierw, że dla $k = 1$ teza jest oczywista. Pod kartą K znajduje się dokładnie jedna karta, czyli ma ona możliwe tylko jedno ułożenie, więc

$$\mathbb{P}_x(X_t^1 = s | T_1 \leqslant t < T_2, \widehat{X_t^1} = \widehat{s}) = 1.$$

Krok indukcyjny Zakładamy, że dla t teza jest prawdziwa i chcemy udowodnić ją dla $t+1$. Problem rozkładamy na dwa przypadki. Albo w momencie $t+1$ włożyliśmy kartę nad K, czyli liczba kart pod K się nie zmieniła, albo włożyliśmy kartę pod K, czyli liczba kart pod K się zmieniła.

Najpierw rozważmy przypadek, że liczba kart pod K się nie zmieniła. To znaczy, że zakładamy, że $t+1 \in [T_k, T_{k+1})$ i $t \in [T_k, T_{k+1})$. Wtedy

$$\begin{aligned}
\mathbb{P}_x(X_{t+1}^k = & s | T_k \leqslant t < t+1 < T_{k+1}, \widehat{X_{t+1}^k} = \widehat{s}) \\
= & \mathbb{P}_x(X_{t+1}^k = s | X_t^k = s, T_k \leqslant t < t+1 < T_{k+1}, \widehat{X_{t+1}^k} = \widehat{s}) \\
& \cdot \mathbb{P}_x(X_t^k = s | T_k \leqslant t < t+1 < T_{k+1}, \widehat{X_t^k} = \widehat{s}) \\
= & 1 \cdot \mathbb{P}_x(X_t^k = s | T_k \leqslant t < t+1 < T_{k+1}, \widehat{X_t^k} = \widehat{s}) \\
\overset{(\star)}{=} & \mathbb{P}_x(X_t^k = s | T_k \leqslant t < T_{k+1}, \widehat{X_t^k} = \widehat{s}) = \frac{1}{k!}.
\end{aligned}$$

Przejście $(\star)$ wynika z faktu, że

$$\{T_k \leqslant t < t+1 < T_{k+1}\} = \{T_k \leqslant t < T_{k+1}\} \cap \{Z_{t+1} \in B\},$$

gdzie Z_{t+1} to zmienna zdefiniowana w reprezentacji przez losowe odwzrowania, a B to taki podzbiór permutacji z Λ, że karta z góry przechodzi nad K, czyli 1 przechodzi na miejsce o numerze mniejszym niż $n-k$. Zdarzenia $\{T_k \leqslant t < T_{k+1}\}$,

$\{X_t^k = s\}$ i $\{\widehat{X_t^k} = \widehat{s}\}$ zależą wyłącznie od $Z_1, \ldots, Z_t$ i X_0, czyli, z niezależności $\{Z_t\}$, nie zależą od Z_{t+1}.

Rozważmy przypadek, że w momencie $t+1$ liczba kart pod K się zmieniła. Tym razem zakładamy więc, że $t+1 \in [T_k, T_{k+1})$ i $t < T_k$, czyli $T_k = t+1$. Poniżej przeprowadzamy rachunek w tej sytuacji. W pierwszym przejściu wykorzystujemy wzór na prawdopodobieństwo całkowite. Rozbijamy prawdopodobieństwo ze względu na to jaką kartę z $\hat{s}$ wkładamy pod K w ruchu $t+1$. Kartę tą oznaczamy u. W kolejnym przejściu wykorzystujemy fakt, że jeżeli ustalimy u, to istnieje dokładnie jedna permutacja (oznaczona $s^{'}$) zbioru $\widehat{s} \setminus \{u\}$, taka że prawdopodobieństwo przejścia do s jest niezerowe.

$$
\begin{aligned}
&\mathbb{P}_x(X_{t+1}^k = s | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}) \\
&\quad = \sum_{u \in \widehat{s}} \mathbb{P}_x(X_{t+1}^k = s, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\} | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}) \\
&\quad = \sum_{u \in \widehat{s}} \mathbb{P}_x(X_{t+1}^k = s, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}, X_t^{k-1} = s^{'} | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}) \\
&\quad = \sum_{u \in \widehat{s}} \mathbb{P}_x(X_{t+1}^k = s | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}, X_t^{k-1} = s^{'}, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}) \\
&\qquad \cdot \mathbb{P}_x(X_t^{k-1} = s^{'} | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}) \\
&\qquad \cdot \mathbb{P}_x(\widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\} | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}) \\
&\quad \overset{(\diamond)}{=} \sum_{u \in \widehat{s}} \frac{1}{k} \cdot \frac{1}{(k-1)!} \cdot \mathbb{P}_x(\widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\} | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}) = \frac{1}{k!}.
\end{aligned}
$$

Przejście $(\diamond)$ wykorzystuje podobną obserwację jak rachunek $(\star)$, czyli

$$\{T_k = t+1\} = \{T_{k-1} \leqslant t < T_k\} \cap \{Z_{t+1} \in B^c\}.$$

Zdarzenia $\{T_{k-1} \leqslant t < T_k\}$, $\{X_t^{k-1} = s^{'}\}$, $\{\widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}\}$, $\{\widehat{X_{t+1}^k} = \widehat{s}\}$ zależą wyłącznie od $X_0, Z_1, \ldots, Z_t$. Zatem z niezależności $\{Z_t\}$ otrzymujemy niezależność wspomnianych zdarzeń od Z_{t+1}, czyli

$$
\begin{aligned}
&\mathbb{P}_x(X_t^{k-1} = s^{'} | T_k = t+1, \widehat{X_{t+1}^k} = \widehat{s}, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}) = \\
&\quad = \mathbb{P}_x(X_t^{k-1} = s^{'} | T_{k-1} \leqslant t < T_k, Z_{t+1} \in B^c, \widehat{X_{t+1}^k} = \widehat{s}, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}) \\
&\quad = \mathbb{P}_x(X_t^{k-1} = s^{'} | T_{k-1} \leqslant t < T_k, \widehat{X_{t+1}^k} = \widehat{s}, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}).
\end{aligned}
$$

Następnie zauważmy, że znając s oraz $\hat{s} \setminus \{u\}$ wiemy jakie będzie $\widehat{X_{t+1}^k}$. Zatem

$$
\begin{aligned}
&\mathbb{P}_x(X_t^{k-1} = s^{'} | T_{k-1} \leqslant t < T_k, \widehat{X_{t+1}^k} = \widehat{s}, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}) = \\
&\quad = \mathbb{P}_x(X_t^{k-1} = s^{'} | T_{k-1} \leqslant t < T_k, \widehat{X_t^{k-1}} = \widehat{s} \setminus \{u\}) = \frac{1}{(k-1)!}.
\end{aligned}
$$

Tezę otrzymujemy korzystając ze wzoru na prawdopodobieństwo całkowite

$$
\begin{aligned}
\mathbb{P}_x(X^k_{t+1} = s | T_k \leqslant t+1 < T_{k+1}, \widehat{X^k_{t+1}} = \widehat{s}) =& \\
&= \mathbb{P}_x(X^k_{t+1} = s | T_k \leqslant t < t+1 < T_{k+1}, \widehat{X^k_{t+1}} = \widehat{s}) \\
&\quad \cdot \mathbb{P}_x(T_k \leqslant t | T_k \leqslant t+1 < T_{k+1}, \widehat{X^k_{t+1}} = \widehat{s}) \\
&\quad + \mathbb{P}_x(X^k_{t+1} = s | T_k = t+1, \widehat{X^k_{t+1}} = \widehat{s}) \\
&\quad \cdot \mathbb{P}_x(t < T_k | T_k \leqslant t+1 < T_{k+1}, \widehat{X^k_{t+1}} = \widehat{s}) \\
&= \frac{1}{k!}\Big(\mathbb{P}_x(T_k \leqslant t | T_k \leqslant t+1 < T_{k+1}, \widehat{X^k_{t+1}} = \widehat{s}) \\
&\quad + \mathbb{P}_x(t < T_k | T_k \leqslant t+1 < T_{k+1}, \widehat{X^k_{t+1}} = \widehat{s})\Big) = \frac{1}{k!}.
\end{aligned}
$$

□

Twierdzenie 3.6. *τ_{top} jest czasem silnie stacjonarnym dla łańcucha zdefiniowanego w lemacie 3.2.*

Dowód. Ponieważ $n-1$ dolnych kart definiuje cały układ talii oraz $T_{n-1} = \tau_{top} - 1$, to teza lematu 3.5 dla $k = n-1$ daje nam

$$
\mathbb{P}_x(X_{t-1} = s | t = \tau_{top} - 1, U_{t-1} = u) = \frac{1}{(n-1)!},
$$

gdzie U_t to zmienna losowa oznaczająca górną kartę w momencie t (w naszym przypadku kartę K). Czyli, w momencie $\tau - 1$, gdy K jest na szczycie talii, wszystkie układy pozostałych kart są równie prawdopodobne.

Pozostaje zatem pokazać, że skoro kartę K wkładamy w każde miejsce z równym prawdopodobieństwem, to wszystkie układy talii są jednakowo prawdopodobne. Przeprowadzimy w tym celu podobne obliczenia jak w dowodzie lematu 3.5. Niech $s^{'}$ oznacza jedyną permutację $\widehat{s} \setminus \{u\}$ taką, że prawdopodobieństwo przejścia z s do $s^{'}$ jest niezerowe.

$$
\begin{aligned}
\mathbb{P}_x(X_t =& s | t = \tau) \\
&= \sum_{u \in \{1,\dots,n\}} \mathbb{P}_x(X_t = s | t = \tau, X_{t-1} = s^{'}, U_{t-1} = u) \\
&\quad \cdot \mathbb{P}_x(X_{t-1} = s^{'} | t = \tau, U_{t-1} = u) \cdot \mathbb{P}_x(U_{t-1} = u | t = \tau) \\
&= \sum_{u \in \{1,\dots,n\}} \frac{1}{n} \cdot \frac{1}{(n-1)!} \cdot \mathbb{P}_x(U_{t-1} = u | t = \tau) = \frac{1}{n!} = \pi(s).
\end{aligned}
$$

Mnożąc stronami przez $\mathbb{P}_x(t = \tau)$ dostajemy warunek z definicji czasu silnie stacjonarnego. To, połączone z poprzednim lematem, kończy dowód. □

Twierdzenie 3.7. *Dla łańcucha zdefiniowanego w lemacie 3.2 zachodzi*

$$
d(n \log n + cn) \leqslant e^{-c} \quad \text{dla } c \geqslant 0, n \geqslant 2, \tag{8}
$$

$$
d(n \log n + c_n n) \xrightarrow{n \to \infty} 1 \quad \text{dla } c_n \to -\infty. \tag{9}
$$

Dowód. Aby udowodnić (8), będziemy chcieli pokazać, że τ_{top} (dalej nazywane τ) ma taki sam rozkład jak inna zmienna losowa.

Na początek zauważmy, że

$$\tau = T_1 + (T_2 - T_1) + \cdots + (T_{n-1} - T_{n-2}) + (\tau - T_{n-1}),$$

gdzie T_i oznacza czas (liczbę ruchów) dopóki i kart znajdzie się pod K. Wtedy $T_i - T_{i-1}$ ma rozkład geometryczny z prawdopodobieństwem sukcesu $\frac{i}{n}$. Sukcesem w tym rozkładzie jest włożenie i-tej karty pod K w momencie, w którym pod nią znajduje się $i-1$ kart. Czyli możemy ją włożyć na i z n sposobów. Z definicji τ mamy $\tau - T_{n-1} = 1$. Co więcej, wszystkie różnice w sumie są niezależne. Stąd

$$\tau_{top} \sim Geo\left(\frac{1}{n}\right) + Geo\left(\frac{2}{n}\right) + \cdots + Geo\left(\frac{n-1}{n}\right) + 1.$$

Teraz rozważmy znany problem kolekcjonera kuponów. Przypuśćmy, że w urnie znajduje się n różnych kul. Losujemy je ze zwracaniem. Jako V oznaczymy liczbę losowań dopóki każdą kulę wylosujemy co najmniej raz. Pokażmy, że zmienna V ma taki sam rozkład jak zmienna τ. V może być zapisane jako

$$V = V_n = (V_n - V_{n-1}) + (V_{n-1} - V_{n-2}) + \cdots + (V_2 - V_1) + V_1,$$

gdzie V_i oznacza liczbę losowań dopóki wylosujemy co najmniej raz dokładnie i kul. Wtedy $V_i - V_{i-1}$ jest czasem oczekiwania na wylosowanie i-tej nowej kuli, po tym jak wylosowaliśmy $i-1$ różnych kul. Zmienna taka ma rozkład geometryczny z prawdopodobieństwem sukcesu $\frac{n-(i-1)}{n}$, gdyż każda z $n-(i-1)$ kul jeszcze nie była wylosowana, więc może być nową kulą. Znowu różnice są niezależne. Stąd dostajemy, że

$$V \sim Geo\left(\frac{n-1}{n}\right) + Geo\left(\frac{n-2}{n}\right) + \cdots + Geo\left(\frac{1}{n}\right) + 1,$$

czyli τ i V mają taki sam rozkład.

W kolejnym kroku, dla każdej kuli b zdefiniujmy zdarzenie A_b „b nie została wylosowana w pierwszych m losowaniach". Wtedy

$$\mathbb{P}(V > m) = \mathbb{P}\left(\bigcup_b A_b\right) \leqslant \sum_b \mathbb{P}(A_b) = \sum_b \left(1 - \frac{1}{n}\right)^m = n\left(1 - \frac{1}{n}\right)^m \leqslant ne^{-m/n}.$$

Stosując powyższą nierówność dla $m = n\log n + cn$, korzystając z twierdzenia 2.16 oraz faktu, że τ nie zależy od X_0 (z tego jak jest zdefiniowane) otrzymujemy

$$d(n\log n + cn) \leqslant \max_{x\in\mathcal{S}} \mathbb{P}_x(\tau > n\log n + cn) = \mathbb{P}(\tau > n\log n + cn) \leqslant e^{-c}.$$

Przechodzimy do dowodu (9). Ponieważ $d(m) \leqslant 1$, więc wystarczy znaleźć odpowiednie oszacowanie z dołu. Ustalmy łańcuch $\{X_t\}$, taki że $X_0 = x$. Dla dowolnego j rozważmy j dolnych kart z x. Zdefiniujmy zbiór A_j ustawień talii kart takich, dla których porządek ustalonych j kart jest taki sam jak w x. To znaczy mogą być jakieś karty między nimi, ale ich kolejność względem siebie pozostaje taka sama. Oczywiście $|A_j| = \frac{n!}{j!}$. Pokażemy, że dla $m = n \log n + c_n n$, gdzie $c_n \xrightarrow{n\to\infty} -\infty$ zachodzi

$$d(m) \geqslant \max_{1\leqslant j\leqslant n} \left(\max_{x\in A_j} \sum_{y\in A_j} P^m(x,y) - \frac{1}{j!} \right) \xrightarrow{n\to\infty} 1. \tag{10}$$

Nierówność w (10) wynika z prostego rachunku. Dla każdego ustalonego j oraz dowolnego t zachodzi

$$\begin{aligned} d(t) &= \max_{x\in\mathcal{S}} \|P^t(x,\cdot) - \pi\|_{TV} = \max_{x\in\mathcal{S}} \sum_{\substack{y\in\mathcal{S},\\ P^t(x,y)\geqslant\pi(y)}} [P^t(x,y) - \pi(y)] \\ &\geqslant \max_{x\in A_j} \sum_{\substack{y\in A_j,\\ P^t(x,y)\geqslant\pi(y)}} [P^t(x,y) - \pi(y)] \geqslant \max_{x\in A_j} \sum_{y\in A_j} \left[P^t(x,y) - \frac{1}{n!}\right] \\ &= \max_{x\in A_j} \sum_{y\in A_j} P^t(x,y) - \frac{1}{n!}|A_j| = \max_{x\in A_j} \sum_{y\in A_j} P^t(x,y) - \frac{1}{j!}. \end{aligned}$$

Aby udowodnić zbieżność w (10) wprowadźmy pomocnicze oznaczenie

$$Y_{j,n} = \max_{x\in A_j} \sum_{y\in A_j} P^m(x,y).$$

Teraz udowodnimy, że dla ustalonego j

$$Y_{j,n} \geqslant \mathbb{P}_x(\tau - T_{j-1} > m) \xrightarrow{n\to\infty} 1. \tag{11}$$

To zakończy dowód zbieżności, ponieważ będzie stąd wynikało, że $Y_{j,n} \to 1$, jako że $Y_{j,n} \leqslant 1$. Możemy wtedy dla ustalonego $\varepsilon > 0$ wziąć takie j_0, że $\frac{1}{j_0!} \leqslant \varepsilon/2$, a następnie tak duże $n_0 > j_0$, że

$$\forall n > n_0 \quad \left|Y_{j_0,n_0} - 1\right| < \varepsilon/2.$$

Wtedy dla każdego $n > n_0$

$$\left| \max_{1\leqslant j\leqslant n} \left(Y_{j,n} - \frac{1}{j!} \right) - 1 \right| = \min_{1\leqslant j\leqslant n} \left(1 - Y_{j,n} + \frac{1}{j!} \right) < 1 - Y_{j_0,n_0} + \frac{1}{j_0!} < \varepsilon.$$

Aby wykazać nierówność z (11) rozważmy rozkład zmiennej $\tau - T_{j-1}$. Jest to czas liczony od momentu, w którym pod K znajduje się $j-1$ kart do momentu włożenia K w talię. Czyli jest to czas, dopóki j-ta karta od dołu talii przejdzie na górę i zostanie włożona w talię. A zatem, jeżeli nie minął moment $\tau - T_{j-1}$, to pierwotne j dolnych kart nadal jest w tej samej kolejności (i nigdy nie przestało być). Czyli, jeśli $\tau - T_{j-1} > m$, to spacer co najmniej do momentu m jest cały czas w A_j. Korzystając z tego, że

$$Y_{j,n} = \max_{x\in A_j} \sum_{y\in A_j} P^m(x,y) = \max_{x\in A_j} \mathbb{P}_x(X_m \in A_j)$$

otrzymujemy nierówność z (11).

Pozostaje udowodnić zbieżność z (11). Udowodnimy, że $\mathbb{P}(\tau - T_{j-1} \leqslant m) \to 0$. Posłuży nam do tego osłabiona wersja nierówności Czebyszewa

$$\mathbb{P}(Z \leqslant \mathbb{E}Z - a) \leqslant \frac{Var(Z)}{a^2} \quad \text{dla } a \geqslant 0. \tag{12}$$

Znamy rozkłady $T_i - T_{i-1}$, stąd

$$\mathbb{E}(T_i - T_{i-1}) = \frac{1}{p} = \frac{n}{i}, \qquad Var(T_i - T_{i-1}) = \frac{1-p}{p^2} = \frac{n^2}{i^2} - \frac{n}{i},$$

czyli

$$\mathbb{E}(\tau - T_{j-1}) = \sum_{i=j}^{n} \frac{n}{i} = n \log n + O(n),$$
$$Var(\tau - T_{j-1}) = \sum_{i=j}^{n} \frac{n^2}{i^2} - \frac{n}{i} = O(n^2),$$

gdzie w obu równościach skorzystaliśmy z tego, że

$$\sum_{i=1}^{n} \frac{1}{i} = \log n + \gamma + O\left(\frac{1}{n}\right)$$

oraz w drugiej z tego, że $\sum \frac{1}{n^2} < \infty$.

Biorąc zatem (12) dla

$$Z = \tau - T_{j-1} \quad \text{oraz} \quad a = \mathbb{E}Z - m = O(n) - c_n n = b_n n \quad (b_n \to +\infty)$$

otrzymujemy

$$\mathbb{P}(\tau - T_{j-1} \leqslant m) \leqslant \frac{O(n^2)}{b_n^2 n^2} = O\left(\frac{1}{b_n^2}\right) \xrightarrow{n\to\infty} 0.$$

Co kończy dowód drugiej części. □

Uwaga 3.8. Problem rozważony w pierwszej części dowodu nazywa się „problemem kolekcjonera kuponów”. Urnę z kulami możemy potraktować jako loterię z nieskończoną liczbą kuponów n rodzajów. Celem kolekcjonera kuponów będzie zebranie chociaż jednego kuponu z każdego rodzaju. W takim myśleniu zmienna V będzie oznaczała, ile kuponów kolekcjoner musi kupić, żeby tego dokonać.

Wniosek 3.9. *Wstawiając $c = \log(\varepsilon^{-1})$ do (8) z twierdzenia 3.7 otrzymujemy*

$$d(n \log n + \log(\varepsilon^{-1})n) \leqslant \varepsilon,$$

czyli

$$t_{mix}(\varepsilon) \leqslant n \log n + \log(\varepsilon^{-1})n.$$

Więc dla standardowej talii kart ($n = 52$)

$$t_{mix} < 278.$$

Wniosek 3.10. *Z* (9) *wynika w szczególności, że dla każdego* ε *istnieje tak duże* α, *że dla dużych* n

$$d(n \log n - \alpha n) \geqslant 1 - \varepsilon,$$

czyli

$$t_{mix}(\varepsilon) \geqslant n \log n - \alpha n.$$

3.2 „Leniwy" spacer po hiperkostce

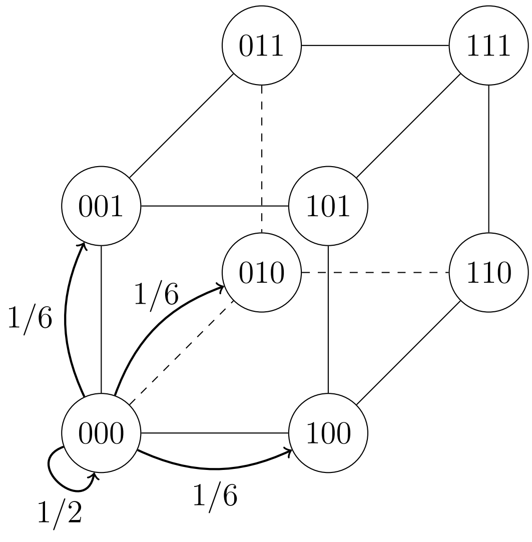
Definicja 3.11. n**-wymiarową hiperkostką** nazywamy graf, którego wierzchołkami są ciągi binarne długości n, czyli elementy zbioru $\{0,1\}^n$. Dwa wierzchołki są połączone krawędzią jeżeli różnią się na dokładnie jednej współrzędnej.

Definicja 3.12. **„Leniwym" spacerem po** n**-wymiarowej hiperkostce** będziemy nazywać taki łańcuch Markowa $\{X_t\}$, że $\mathcal{S} = \{0,1\}^n$ oraz

$$P(x,y) = \begin{cases} \frac{1}{2} & \text{jeżeli} \quad y = x, \\ \frac{1}{2n} & \text{jeżeli } \exists! j \in \{1,\ldots,n\} \;\; x(j) \neq y(j), \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

Co oznacza, że w danym kroku z prawdopodobieństwem $\frac{1}{2}$ spacer pozostaje w punkcie, w którym się znajduje, a z prawdopodobieństwem $\frac{1}{2n}$ przechodzi do któregoś z n sąsiadujących wierzchołków.

„Leniwy" spacer dla $n = 3$ ilustruje poniższy graf (Rysunek 2).



Rysunek 2: 3-wymiarowa hiperkostka i przykładowe prawdopodobieństwa przejścia

Uwaga 3.13. „Leniwy" spacer po kostce jest nieokresowy, bo $P(x,x) = \frac{1}{2}$. Jest też nieredukowalny, ponieważ z każdego wierzchołka da się dojść do każdego innego. Jego miarą stacjonarną jest miara jednostajna.

Uwaga 3.14. Można się zastanawiać, dlaczego rozważamy „leniwy” spacer po n-kostce, a nie klasyczny, czyli taki, że z prawdopodobieństwem $\frac{1}{n}$ przechodzimy do jednego z sąsiednich wierzchołków. Podobnie jak w przypadku spaceru po n-cyklu, klasyczny spacer po hiperkostce jest okresowy z okresem każdego wierzchołka równym 2. Wynika to z faktu, że w t-tym kroku liczba jedynek w X_t ma tę samą parzystość co t, czyli do punktu wyjściowego da się wrócić jedynie w parzystej liczbie kroków. Okres równy 2 wyklucza zbieżność łańcucha do miary stacjonarnej.

Uwaga 3.15. Popatrzmy na ten proces trochę inaczej. Przejście do sąsiada (z prawdopodobieństwem $\frac{1}{2n}$) powoduje zmianę wartości na jednej ze współrzędnych. Można zatem inaczej skonstruować ten sam proces. Najpierw wylosować jedną ze współrzędnych, a następnie rzucić monetą czy tę współrzędną zmienić czy nie. Czyli proces można zdefiniować przy pomocą zmiennej

$$Z_t = (j_t, B_t) \sim U(\{1, 2, \dots, n\} \times \{0, 1\}).$$

j mówi, którą współrzędną rozważamy, a B czy ją zmieniamy.

Tą obserwację wykorzystamy do zdefiniowania czasu silnie stacjonarnego dla spaceru po hiperkostce.

Twierdzenie 3.16. *Dla „leniwego” spaceru po n-wymiarowej hiperkostce czasem silnie stacjonarnym jest zmienna losowa*

$$\tau := \min\{t \geqslant 0 : \{j_1, \dots, j_t\} = \{1, \dots, n\}\},$$

czyli jest to pierwszy moment, w którym wylosowaliśmy każdą współrzędną.

Dowód. Uwaga 3.15 sugeruje reprezentację przez losowe odwzorowania dla tego łańcucha Markowa

$$Z_t = (j_t, B_t) \quad Z_t \sim U(\{1, 2, \dots, n\} \times \{0, 1\})$$

$$f((x_1, x_2, \dots, x_n), (j, b)) = (x_1, \dots, x_j +_2 b, \dots, x_n).$$

Z definicji τ zdarzenie $\{\tau = t\}$ zależy wyłącznie od $\{j_1, \dots, j_t\}$, czyli τ jest czasem zatrzymania dla Z_t. Zatem jest zrandomizowanym czasem zatrzymania wyjściowego łańcucha.

Aby udowodnić, że τ jest czasem silnie stacjonarnym, zauważmy, że skoro w momencie τ wylosowaliśmy już każdą współrzędną, to X_τ przyjmuje z równym prawdopodobieństwem wartość 0 lub 1 na każdej współrzędnej. Wynika to z faktu, że po wylosowaniu danej współrzędnej przyjmuje ona wartość 0 lub 1 z równym prawdopodobieństwem. A τ jest momentem późniejszym niż momenty wylosowania każdej współrzędnej po raz pierwszy.

Przeprowadźmy formalne uzasadnienie. Niech τ_i oznacza moment, w którym po raz pierwszy wylosowaliśmy i-tą współrzędną. Dodatkowo, niech X_t^i oznacza i-tą współrzędną punktu, w którym jesteśmy w momencie t. Wtedy

$$\mathbb{P}_x(X_t^i = y | t = \tau_i) = \frac{1}{2}.$$

Co więcej, równość ta zachodzi również dla wszystkich $t \geqslant \tau_i$, co można udowodnić indukcyjnie.

- Dla $t = \tau_i$ zachodzi.
- Jeżeli teza zachodzi dla jakiegoś $t = T \geqslant \tau_i$, to zachodzi również dla $t+1$, ponieważ

$$
\begin{aligned}
&\mathbb{P}_x(X^i_{t+1} = y | t = \tau_i) = \\
&\quad = \mathbb{P}_x(X^i_{t+1} = y | X^i_t = y, t = \tau_i) \cdot \mathbb{P}_x(X^i_t = y | t = \tau_i) \\
&\qquad + \mathbb{P}_x(X^i_{t+1} = y | X^i_t = y +_2 1, t = \tau_i) \cdot \mathbb{P}_x(X^i_t = y +_2 1 | t = \tau_i) \\
&\quad = \frac{1}{2}\Big(\mathbb{P}_x(X^i_{t+1} = y | j_{t+1} = i, X^i_t = y, t = \tau_i) \cdot \mathbb{P}_x(j_{t+1} = i | X^i_t = y, t = \tau_i) \\
&\qquad + \mathbb{P}_x(X^i_{t+1} = y | j_{t+1} \neq i, X^i_t = y, t = \tau_i) \cdot \mathbb{P}_x(j_{t+1} \neq i | X^i_t = y, t = \tau_i) \\
&\qquad + \mathbb{P}_x(X^i_{t+1} = y | j_{t+1} = i, X^i_t = y +_2 1, t = \tau_i) \cdot \mathbb{P}_x(j_{t+1} = i | X^i_t = y +_2 1, t = \tau_i) \\
&\qquad + \mathbb{P}_x(X^i_{t+1} = y | j_{t+1} \neq i, X^i_t = y +_2 1, t = \tau_i) \cdot \mathbb{P}_x(j_{t+1} \neq i | X^i_t = y +_2 1, t = \tau_i)\Big) \\
&\quad = \frac{1}{2}\left(\frac{1}{2} \cdot \frac{1}{n} + 1 \cdot \frac{n-1}{n} + \frac{1}{2} \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n}\right) = \frac{1}{2}.
\end{aligned}
$$

Do zakończenia dowodu musimy jeszcze skorzystać z niezależności wszystkich współrzędnych od siebie. Wtedy

$$
\mathbb{P}_x(X_t = y | t = \tau) = \mathbb{P}_x(X^1_t = y^1 | t = \tau) \ldots \mathbb{P}_x(X^n_t = y^n | t = \tau) = \frac{1}{2^n} = \pi(y).
$$

□

Uwaga 3.17. Zmienna τ, zdefiniowana w 3.16, jest przykładem zrandomizowanego momentu zatrzymania, który nie jest momentem zatrzymania. Spostrzeżenie to jest intuicyjne. Zauważmy, że obserwując jedynie zachowanie spaceru nie jesteśmy w stanie powiedzieć kiedy dochodzimy do τ. Dzieje się tak, ponieważ w danym momencie, jeżeli spacer się nie rusza, to nie wiemy, którą współrzędną wylosował. Hipotetycznie mogło by się zdarzyć nawet tak, że spacer do momentu τ stałby w tym samym miejscu co na początku.

Twierdzenie 3.18. *Dla „leniwego” spaceru po hiperkostce*

$$
d(n \log n + cn) \leqslant e^{-c},
$$

czyli

$$
t_{mix}(\varepsilon) \leqslant n \log n + \log(\varepsilon^{-1}) n.
$$

Dowód. Treść twierdzenia jest podobna do pierwszej części twierdzenia 3.7, zatem wykorzystamy jej dowód. Przypomnijmy, że rozważaliśmy tam problem kolekcjonera kuponów (losujemy ze zwracaniem n różnych kul) oraz zmienną V, czyli liczbę losowań, dopóki każdą kulę wylosujemy co najmniej raz. Utożsamiając $\{j_1, \ldots, j_n\}$ z definicji τ z kulami z problemu kolekcjonera otrzymujemy, że τ i V mają taki sam rozkład. Obydwa są pierwszymi momentami, w których wylosujemy wszystkie elementy, a losowania te są niezależne i prawdopodobieństwo wylosowania każdego elementu jest takie samo. Co więcej, τ jest tak zdefiniowane, że nie zależy od X_0. Czyli

$$d(n \log n + cn) \leqslant \max_{x \in \mathcal{S}} \mathbb{P}_x(\tau > n \log n + cn) = \mathbb{P}(\tau > n \log n + cn) \leqslant e^{-c}.$$

Druga część tezy wynika z pierwszej, co zostało uzasadnione we wniosku 3.9. □

Uwaga 3.19. Wróćmy do uwagi 2.20, w której definiowaliśmy punkt postojowy. Jest to taki punkt dla spaceru $\{X_t\}$ z punktem początkowym $X_0 = x$, że jeżeli znajdujemy się w tym punkcie, to na pewno minął już moment τ. W przypadku spaceru po kostce takim punktem jest ten naprzeciwko x, tzn. taki, że ma wszystkie współrzędne inne niż x. Na przykład, jeżeli $x = (0, \ldots, 0)$, to $(1, \ldots, 1)$ jest punktem postojowym.

3.3 „Leniwy" spacer po cyklu o 2^k wierzchołkach

Teraz rozważymy łańcuch zdefiniowany w przykładzie 1.11. Jednak ograniczymy się jedynie do przypadków, gdy $n = 2^k$. Przypomnijmy definicję.

Definicja 3.20. Łańcuch Markowa, taki że $\mathcal{S} = \mathbb{Z}_n = \{0, 1, \ldots, n-1\}$, a macierz przejścia to

$$P(j,k) = \begin{cases} \frac{1}{4} & \text{gdy } k \equiv j+1 \pmod{n}, \\ \frac{1}{2} & \text{gdy } k \equiv j \pmod{n}, \\ \frac{1}{4} & \text{gdy } k \equiv j-1 \pmod{n}, \\ 0 & \text{w przeciwnym wypadku,} \end{cases}$$

będziemy nazywać **„leniwym" spacerem po cyklu $\mathbb{Z}_n$ (n-cyklu)**.

Definicja 3.21. Określamy rekurencyjnie zmienną losową $\tau_k : \mathbb{Z}_{2^k} \to \{0, 1, \ldots \infty\}$ dla „leniwego" spaceru $\{X_t\}$ po 2^k-cyklu w następujący sposób

- $\tau_1 = 1$,
- $\tau_{i+1} = T_{\tau_i} + 1$, gdzie zmienne $T_0, T_1, \ldots$ definiujemy również rekurencyjnie
  - $T_0 = 0$
  - $T_1 = t_1$, gdzie t_1 jest czasem, po którym spacer zrobił dwa kroki, takie że w każdym zmienił punkt,
  - $T_{j+1} = T_j + t_{j+1}$, gdzie t_{j+1} jest czasem, po którym spacer zrobił dwa kroki, takie że w każdym zmienił punkt po czasie T_j.

Fakt 3.22. τ_k jest zrandomizowanym momentem zatrzymania dla „leniwego" spaceru po 2^k-cyklu.

Dowód. Przypomnijmy przykład 2.4, gdzie znaleźliśmy przykładową reprezentację przez odwzorowania losowe dla „leniwego" spaceru po n-cyklu. Jest to

$$Z = (j, B) \sim U(\{-1, 1\} \times \{0, 1\}) \quad \text{i} \quad f(x, z) = x + j \cdot B \mod 2^k.$$

Zauważmy, że T_{τ_i} to czas, po którym spacer zrobił $2 \cdot \tau_i$ kroków, takich że zmienił miejsce. O tym, czy w kroku t spacer zmienił miejsce, informuje zmienna B_t. Czyli τ_k można zdefiniować indukcyjnie przy pomocy Z_t w następujący sposób

- $\tau_1 = 1$,
- $\tau_{i+1} = \min\{t : \text{w ciągów } B_1, \ldots, B_{t-1} \text{ znajduje się } 2 \cdot \tau_i \text{ wartości } 1\}$.

Ponieważ $\mathbf{1}_{\{\tau = t\}}$ zależy wyłącznie od wektora $(Z_1, \ldots, Z_t)$, więc τ_k jest zrandomizowanym momentem zatrzymania. □

Twierdzenie 3.23. *τ_k jest czasem silnie stacjonarnym dla „leniwego" spaceru po 2^k-cyklu.*

Dowód. Dla ułatwienia rachunków będziemy rozważać wyłącznie spacery takie, że $X_0 = 0$. Każdy spacer możemy utożsamić ze spacerem zaczynającym się w 0, obracając odpowiednio cykl.

Skorzystamy z poprzedniego faktu. Żeby udowodnić, że τ jest czasem silnie stacjonarnym potrzebujemy zatem jeszcze wykazać, że

$$\mathbb{P}(\tau_k = t, X_{\tau_k} = y) = \mathbb{P}(\tau_k = t)\mathbb{P}(X_{\tau_k} = y).$$

Posłużymy się metodą indukcji matematycznej względem k.

Krok pierwszy Rozpatrujemy sytuację, gdy $k = 1$, czyli spacer odbywa się po $\mathbb{Z}_2$. Wiemy, że $\tau_1 = 1$. Stąd

$$\mathbb{P}(\tau_1 = 1, X_1 = y) = \begin{cases} \frac{1}{2} & \text{gdy } y = 0 \\ \frac{1}{2} & \text{gdy } y = 1 \end{cases} = \frac{1}{2} = \mathbb{P}(\tau_1 = 1)\pi(y),$$

a dla $t \neq 1$

$$\mathbb{P}(\tau_1 = t, X_t = y) = 0 = \mathbb{P}(\tau_1 = t)\pi(y).$$

Krok indukcyjny Zakładamy, że zmienna τ_k jest czasem silnie stacjonarnym dla spaceru po $\mathbb{Z}_{2^k}$, a chcemy pokazać, że zmienna τ_{k+1} jest czasem silnie stacjonarnym dla spaceru po $\mathbb{Z}_{2^{k+1}}$.

Zauważmy, że $\{X_{T_j}\}_{j \geqslant 0}$ jest „leniwym" spacerem po wierzchołkach parzystych grafu $\mathbb{Z}_{2^{k+1}}$. Po pierwsze, spacer ten zawsze znajduje się w wierzchołku o indeksie parzystym (zaczynamy z 0, a z T_i do T_{i+1} zmieniając wierzchołek 2 razy możemy przejść jedynie do parzystego). Po drugie, dlatego że z prawdopodobieństwem $\frac{1}{2}$ ten spacer zostaje w miejscu, a z prawdopodobieństwem $\frac{1}{4}$ przechodzi do któregoś z parzystych sąsiadów. Działania $y \pm 1$ rozumiemy jako modulo 2^k.

$$\begin{aligned}
&\mathbb{P}(X_{T_i} = y | X_{T_{i-1}} = y) = \\
&\quad = \mathbb{P}(X_{T_i} = y, X_{T_i - 1} = y + 1 | X_{T_{i-1}} = y) \\
&\qquad + \mathbb{P}(X_{T_i} = y, X_{T_i - 1} = y - 1 | X_{T_{i-1}} = y) \\
&\quad = \mathbb{P}(X_{T_i} = y | X_{T_i - 1} = y + 1, X_{T_{i-1}} = y) \cdot \mathbb{P}(X_{T_i - 1} = y + 1 | X_{T_{i-1}} = y) \\
&\qquad + \mathbb{P}(X_{T_i} = y | X_{T_i - 1} = y - 1, X_{T_{i-1}} = y) \cdot \mathbb{P}(X_{T_i - 1} = y - 1 | X_{T_{i-1}} = y) \\
&\quad = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}
\end{aligned}$$

W rachunku korzystamy z tego, że ze względu na definicję T_k spacer między momentami T_i i $T_{i+1}-1$ oraz momentami $T_{i+1}-1$ i T_{i+1} musi się gdzieś ruszyć. W związku z tym, na przykład jeżeli w momencie T_i byliśmy w y, to w momencie $T_{i+1}-1$ możemy być w $y+1$ albo w $y-1$. Ze względu na symetrię otrzymujemy

$$\mathbb{P}(X_{T_1-1} = y-1 | X_{T_{i-1}} = y) = \mathbb{P}(X_{T_1-1} = y+1 | X_{T_{i-1}} = y) = \frac{1}{2}.$$

Analogiczny rachunek przeprowadzamy dla $\mathbb{P}(X_{T_i} = y+2 | X_{T_{i-1}} = y)$ oraz $\mathbb{P}(X_{T_i} = y-2 | X_{T_{i-1}} = y)$. Przykładowo przeprowadźmy pierwszy z nich.

$$\begin{aligned}
&\mathbb{P}(X_{T_i} = y+2 | X_{T_{i-1}} = y) = \\
&\quad = \mathbb{P}(X_{T_i} = y+2, X_{T_i-1} = y+1 | X_{T_{i-1}} = y) \\
&\quad = \mathbb{P}(X_{T_i} = y+2 | X_{T_i-1} = y+1, X_{T_{i-1}} = y) \cdot \mathbb{P}(X_{T_i-1} = y+1 | X_{T_{i-1}} = y) \\
&\quad = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.
\end{aligned}$$

A zatem, skoro $\{X_{T_j}\}$ jest „leniwym" spacerem po cyklu o 2^k wierzchołkach, to τ_k jest jego czasem silnie stacjonarnym. Czyli T_{τ_k} jest czasem silnie stacjonarnym dla spaceru $\{X_t\}$. Stąd, w momencie T_{τ_k} rozkład X_{T_τ} jest jednostajny na parzystych wierzchołkach $\mathbb{Z}_{2^{k+1}}$ oraz

$$\mathbb{P}(X_t = y | t = T_{\tau_k}) = \frac{1}{2^k}.$$

Policzmy $\mathbb{P}(X_t = y | t = \tau_{k+1})$. Dla y parzystego X_{T_τ} musiało być w y. Stąd

$$\begin{aligned}
&\mathbb{P}(X_t = y | t = \tau_{k+1}) = \\
&\qquad = \mathbb{P}(X_t = y, X_{t-1} = y | t = \tau_{k+1}) \\
&\qquad = \mathbb{P}(X_t = y | X_{t-1} = y, t-1 = T_{\tau_k}) \mathbb{P}(X_{t-1} = y | t-1 = T_{\tau_k}) \\
&\qquad \overset{(\star)}{=} \mathbb{P}(X_t = y | X_{t-1} = y) \mathbb{P}(X_{t-1} = y | t-1 = T_{\tau_k}) \\
&\qquad = \frac{1}{2} \cdot \frac{1}{2^k} = \frac{1}{2^{k+1}} = \pi(y),
\end{aligned}$$

gdzie przejście $(\star)$ wynika z rachunków (5) w dowodzie lematu 2.11 dla $s = t-1$. Dla y nieparzystego X_{T_τ} musiało być w $y \pm 1$. Czyli

$$\begin{aligned}
&\mathbb{P}(X_t = y | t = \tau_{k+1}) = \\
&\quad = \mathbb{P}(X_t = y, X_{t-1} = y+1 | t = \tau_{k+1}) + \mathbb{P}(X_t = y, X_{t-1} = y-1 | t = \tau_{k+1}) \\
&\quad = \mathbb{P}(X_t = y | X_{t-1} = y+1, t = \tau_{k+1}) \mathbb{P}(X_{t-1} = y+1 | t = \tau_{k+1}) \\
&\qquad + \mathbb{P}(X_t = y | X_{t-1} = y-1, t = \tau_{k+1}) \mathbb{P}(X_{t-1} = y-1 | t = \tau_{k+1}) \\
&\quad = \mathbb{P}(X_t = y | X_{t-1} = y+1) \mathbb{P}(X_{t-1} = y+1 | t-1 = T_{\tau_k}) \\
&\qquad + \mathbb{P}(X_t = y | X_{t-1} = y-1) \mathbb{P}(X_{t-1} = y-1 | t-1 = T_{\tau_k}) \\
&\quad = \frac{1}{4} \cdot \frac{1}{2^k} + \frac{1}{4} \cdot \frac{1}{2^k} = \frac{1}{2^{k+1}} = \pi(y).
\end{aligned}$$

Co kończy dowód. □

Do znalezienia czasu mieszania dla „leniwego” spaceru po n-cyklu użyjemy technicznej własności, nazywanej tożsamością Walda.

Lemat 3.24. *(tożsamość Walda) Niech $\{Y_t\}$ będzie ciągiem niezależnych zmiennych losowych o takich samych rozkładach takich, że $\mathbb{E}(|Y_t|) < \infty$. Niech τ oznacza moment zatrzymania dla łańcucha $\{Y_t\}$. Wtedy*

$$\mathbb{E}\left(\sum_{t=1}^{\tau} Y_t\right) = \mathbb{E}(\tau)\mathbb{E}Y_1.$$

Dowód. Zauważmy, że jeżeli τ jest momentem zatrzymania dla ciągu $\{Y_t\}$, to zmienne $\mathbf{1}_{\{\tau\geqslant t\}}$ i Y_t są niezależne. Ponieważ τ jest momentem zatrzymania dla $\{Y_t\}$ to, istnieje takie $B_{t-1} \subset \Omega^{t-1}$, że

$$\{\tau \geqslant t\} = \{\tau \leqslant t-1\}^c = \{(Y_1, \dots, Y_{t-1}) \in B_{t-1}\}^c.$$

Ponieważ Y_t jest niezależne od wektora $(Y_1, \dots, Y_{t-1})$, to powyższy rachunek pokazuje, że $\{\tau \geqslant t\}$ i Y_t są niezależne.

Z tej niezależności oraz twierdzenia o zbieżności monotonicznej otrzymujemy

$$\begin{aligned}\mathbb{E}\left(\sum_{t=1}^{\infty} |Y_t|\mathbf{1}_{\{\tau\geqslant t\}}\right) &= \sum_{t=1}^{\infty} \mathbb{E}\left(|Y_t|\mathbf{1}_{\{\tau\geqslant t\}}\right) = \sum_{t=1}^{\infty} \mathbb{E}\left(|Y_t|\right)\mathbb{P}\left(\tau \geqslant t\right) \\ &= \mathbb{E}(|Y_1|)\sum_{t=1}^{\infty} \mathbb{P}\left(\tau \geqslant t\right) = \mathbb{E}(|Y_1|)\mathbb{E}(\tau) < \infty. \qquad (13)\end{aligned}$$

Skąd dzięki (13) oraz

$$\Big|\sum_{t=1}^{\infty} Y_t\mathbf{1}_{\{\tau\geqslant t\}}\Big| \leqslant \sum_{t=1}^{\infty} |Y_t|\mathbf{1}_{\{\tau\geqslant t\}},$$

możemy zastosować twierdzenie o zbieżności zmajoryzowanej

$$\mathbb{E}\left(\sum_{t=1}^{\infty} Y_t\mathbf{1}_{\{\tau\geqslant t\}}\right) = \sum_{t=1}^{\infty} \mathbb{E}\left(Y_t\mathbf{1}_{\{\tau\geqslant t\}}\right) = \sum_{t=1}^{\infty} \mathbb{E}\left(Y_t\right)\mathbb{P}(\tau \geqslant t) = \mathbb{E}(T_1)\mathbb{E}(\tau).$$

Na koniec zauważmy, że

$$\sum_{t=1}^{\tau} Y_t = \sum_{t=1}^{\infty} Y_t\mathbf{1}_{\{\tau\geqslant t\}},$$

skąd otrzymujemy

$$\mathbb{E}\left(\sum_{t=1}^{\tau} Y_t\right) = \mathbb{E}\left(\sum_{t=1}^{\infty} Y_t\mathbf{1}_{\{\tau\geqslant t\}}\right) = \mathbb{E}(T_1)\mathbb{E}(\tau).$$

□

Twierdzenie 3.25. *Dla τ_k zdefiniowanego jak w 3.21 zachodzi*

$$\mathbb{E}(\tau_k) = \sum_{i=0}^{k-1} 4^i = \frac{4^k - 1}{3}.$$

Dowód. Aby to udowodnić pokażemy, że

$$\mathbb{E}(\tau_k) = 4 \cdot \mathbb{E}(\tau_{k-1}) + 1 \tag{14}$$

skąd już łatwym rachunkiem otrzymamy

$$\mathbb{E}(\tau_k) = 4(\ldots(4(4\tau_1 + 1) + 1)\ldots) + 1 = \sum_{i=0}^{k-1} 4^i.$$

Zauważmy, że (stosujemy podobną metodę jak w pierwszej części dowodu twierdzenia 3.7)

$$t_1 = Y_1 + (t_1 - Y_1),$$

gdzie Y_1 jest czasem dopóki spacer zrobił pierwszy krok, taki że zmienił punkt. Zmienne Y_1 i $t_1 - Y_1$ mają taki sam rozkład, czyli geometryczny z prawdopodobieństwem sukcesu równym $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ (sukcesem jest ruszenie się do któregokolwiek sąsiedniego punktu). W związku z tym $\mathbb{E}(Y_1) = \mathbb{E}(t_1 - Y_1) = 2$, czyli

$$\mathbb{E}(t_1) = \mathbb{E}(Y_1) + \mathbb{E}(t_1 - Y_1) = 4.$$

Ponieważ $\{t_i\}$ są zmiennymi niezależnymi o jednakowych rozkładach, to $\mathbb{E}(t_i) = 4$. Z konstrukcji τ_{k-1} jest momentem zatrzymania dla $\{t_i\}$. Możemy zatem skorzystać z tożsamości Walda

$$\begin{aligned}\mathbb{E}(\tau_k) &= \mathbb{E}(T_{\tau_{k-1}} + 1) = \mathbb{E}(T_{\tau_{k-1}}) + 1 = \mathbb{E}(t_{\tau_{k-1}} + t_{\tau_{k-1}-1} + \cdots + t_1) + 1 \\ &= \mathbb{E}(\tau_{k-1})\mathbb{E}(t_1) + 1 = 4 \cdot \mathbb{E}(\tau_{k-1}) + 1.\end{aligned}$$

□

Wniosek 3.26. *Dla „leniwego” spaceru po cyklu* $\mathbb{Z}_{2^k}$

$$t_{mix}(\varepsilon) \leqslant \frac{4^k - 1}{\varepsilon}.$$

Dowód. Stosując twierdzenie 2.16, niezależność τ od X_0, nierówność Markowa oraz poprzednie twierdzenie otrzymujemy

$$d(t) \leqslant \max_{x \in \mathcal{S}} \mathbb{P}_x(\tau > t) = \mathbb{P}(\tau > t) \leqslant \frac{\mathbb{E}(\tau)}{t} = \frac{4^k - 1}{3t},$$

czyli

$$d\left(\frac{4^k - 1}{3\varepsilon}\right) \leqslant \varepsilon.$$

□

3.4 Tasowanie „riffle shuffle"

Na koniec zajmiemy się tasowaniem kart o nazwie „riffle shuffle". Metoda ta jest popularna w kasynach ze względu na małe prawdopodobieństwo ujawnienia kart podczas tasowania. Przebiega ono w następujący sposób. Dzielimy karty na dwie mniej więcej równe części. Następnie bierzemy w każdą rękę jedną część i uwalniamy karty po kolei, tak aby spadały na stół mniej więcej naprzemiennie z każdej ręki. Na koniec karty łączymy w jedną talię.

Uwaga 3.27. Podobnie jak tasowanie „top-to-random" tasowanie „riffle shuffle" będziemy utożsamiać ze spacerem po grupie permutacji n kart, czyli $\mathcal{S}_n$. Znowu karty ponumerujemy od góry talii.

Aby zdefiniować model matematyczny tasowania „riffle shuffle", posłużymy się następującym lematem.

Lemat 3.28 (model Gilberta-Shannona-Reedsa)**.** *Następujące 3 opisy dają ten sam rozkład prawdopodobieństwa na grupie permutacji kart:*

1. *Wylosuj k z rozkładu dwumianowego $Bin(n, \frac{1}{2})$ i podziel talię n kart na 2 części: k i $n-k$ elementową (biorąc górne k kart do pierwszej części). Następnie wybierz jednostajnie jedno z $\binom{n}{k}$ możliwych ustawień kart z pierwszej grupy we wszystkich kartach (ważne, że nie zmieniamy ustawienia kart w częściach względem siebie).*

2. *Podziel talię na dwie części tak jak w opisie powyżej. Przypuśćmy, że pierwsza część kart jest w lewej ręce, druga w prawej. Układaj z nich jedną talię w następujący sposób. Jeżeli w danym ruchu masz L kart w lewej ręce, a P w prawej, to wykładasz kartę z góry stosu z lewej ręki z prawdopodobieństwem $\frac{L}{L+P}$, a kartę z góry stosu z prawej z prawdopodobieństwem $\frac{P}{L+P}$.*

3. *(Tasowanie odwrotne) Opiszemy jak wykonać czynność odwrotną do tasowania „riffle shuffle". Połóż karty koło siebie grzbietami do góry. Każdej karcie (niezależnie) przypisz liczbę 0 lub 1 (możesz w tym celu rzucić monetą dla każdej z nich). Weź wszystkie karty zaetykietowane numerem 0 (nie zmieniając ich ułożenia względem siebie) i ustaw je na górze talii.*

Uwaga 3.29. Odpowiedzmy na pytanie, dlaczego metoda (3) jest czynnością odwrotną do tasowania „riffle shuffle". Rozważmy permutację kart otrzymaną pojedynczym tasowaniem

$$\begin{pmatrix} 1 & 2 & \dots & n \\ a_1 & a_2 & \dots & a_n \end{pmatrix}.$$

Wtedy ciąg $a_1, \dots, a_n$ ma co najwyżej dwa podciągi rosnące. Każdy z podciągów otrzymujemy z jednej z części, na które wcześniej podzieliliśmy talię. Jak przeprowadzić czynność odwrotną? Należy przypisać etykiety „0" elementom podciągu zaczynającego się od 1, a etykiety „1" pozostałym elementom. A następnie elementy z numerem 0 wyjąć i dać na początek talii.

Formalnie zdefiniujemy, co to znaczy być spacerem odwrotnym. W dowodzie pokażemy, że opis (3) zadaje spacer odwrotny do spaceru zadanego w opisach (1) i (2).

Definicja 3.30. Niech $\{X_t\}$ będzie spacerem na grupie G generowanym przez miarę probabilistyczną ν. **Miarą odwrotną** do tego łańcucha będziemy nazywać miarę $\tilde{\nu}$, taką że

$$\tilde{\nu}(g) := \nu(g^{-1}) \qquad (\forall g \in G).$$

Łańcuch generowany przez miarę odwrotną oznaczać będziemy $\tilde{X}_t$ i nazywać **spacerem odwrotnym**.

Dowód. Najpierw pokażemy, że opisy (1) i (2) są równoważne. Talię dzielimy w ten sam sposób. Niech K będzie zmienną losową mówiącą jak dzielimy talię. Przypuśćmy zatem, że ją podzieliliśmy, czyli ustaliliśmy $K = k$. Policzmy rozkład ustawień kart w modelu (2). Po pierwsze, każdemu możliwemu do osiągnięcia ustawieniu możemy jednoznacznie przypisać ciąg $D = \{D_1, \ldots, D_n\}$, taki że $D_i \in \{L, P\}$ mówi, z której ręki wzięliśmy kartę w i-tym ruchu. Oczywiście L występuje w tym ciągu k razy, a P występuje $n - k$ razy. Ustalmy teraz konkretne ustawienie i jego ciąg $\{d_1, \ldots, d_n\}$. Niech $\{l_1, \ldots, l_k\}$ oznaczają uporządkowane rosnąco indeksy, takie że $d_{l_i} = L$ (i jest to i-te L w D). Analogicznie, $\{p_1, \ldots, p_{n-k}\}$ to, uporządkowane rosnąco, pozostałe indeksy. Wtedy

$$\begin{aligned}
&\mathbb{P}(D_{l_i} = d_{l_i} | D_1 = d_1, \ldots, D_{l_i-1} = d_{l_i-1}, K = k) \\
&\qquad = \mathbb{P}(D_{l_i} = L | D_1 = d_1, \ldots, D_{l_i-1} = d_{l_i-1}, K = k) = \frac{k - (i-1)}{n - (l_i - 1)} \\
&\mathbb{P}(D_{p_i} = d_{p_i} | D_1 = d_1, \ldots, D_{p_i-1} = d_{p_i-1}, K = k) \\
&\qquad = \mathbb{P}(D_{p_i} = P | D_1 = d_1, \ldots, D_{p_i-1} = d_{p_i-1}, K = k) = \frac{(n-k) - (i-1)}{n - (p_i - 1)}.
\end{aligned}$$

Mnożąc te równości dla wszystkich $i \in \{1, \ldots, k\}$ i $j \in \{1, \ldots, n-k\}$ stronami (i korzystając z tego, że $\{l_1, \ldots l_k\} \cup \{p_1, \ldots, p_{n-k}\} = \{1, \ldots, n\}$) otrzymujemy

$$\begin{aligned}
&\mathbb{P}(D = \{d_1, \ldots, d_n\} | K = k) \\
&= \mathbb{P}(D_1 = d_1 | K = k) \ldots \mathbb{P}(D_n = d_n | D_1 = d_1, \ldots, D_{n-1} = d_{n-1}, K = k) \\
&= \frac{(n-k)!k!}{n!} = \frac{1}{\binom{n}{k}}.
\end{aligned}$$

W następnym kroku wykażemy, że równoważne są opisy (1) i (3). Ustalmy najpierw jak wygląda miara generująca łańcuch z (1) i (2) opisu. Wiemy, że $\mathbb{P}(K = k) = \binom{n}{k}\left(\frac{1}{2}\right)^n$. Co więcej, zbiory permutacji możliwych do otrzymania dla różnych k mają jedyny punkt wspólny, czyli permutację identycznościową. Stąd wartość miary na permutacjach możliwych do otrzymania (nie identycznościowej) to

$$\mathbb{P}(D = \{d_1, \ldots, d_n\}) = \mathbb{P}(D = \{d_1, \ldots, d_n\} | K = k)\mathbb{P}(K = k) = \frac{\binom{n}{k}\left(\frac{1}{2}\right)^n}{\binom{n}{k}} = \frac{1}{2^n}.$$

Permutację identycznościową możemy otrzymać na $(n+1)$ sposobów. Zatem miarą generującą łańcuch z (1) i (2) opisu jest

$$Q(\sigma) = \begin{cases} \frac{1}{2^n} & \text{jeżeli } \sigma \text{ zawiera dwa podciągi rosnące,} \\ \frac{n+1}{2^n} & \text{jeżeli } \sigma = \mathrm{id}, \\ 0 & \text{w przeciwnym wypadku,} \end{cases}.$$

Szukamy teraz miary generującej proces tasowania odwrotnego. Permutacje, które można uzyskać poprzez tasowanie odwrotne, są generowane przez ciągi z $\{0,1\}^n$. Weźmy ciąg $a_1, \ldots, a_n$ i oznaczmy jako k liczbę jego wyrazów zerowych. Takiemu ciągowi przypisujemy permutację

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}$$

taką, że $i_1, \ldots, i_k$ to ustawione rosnąco l, takie że $a_l = 0$, natomiast $i_{k+1}, \ldots, i_n$ to ustawione rosnąco l, takie że $a_l = 1$. Ciągów z $\{0,1\}^n$ jest 2^n, a ponieważ wyrazy ich są losowane niezależnie, to prawdopodobieństwo otrzymania każdego z nich wynosi $\frac{1}{2^n}$. Ciągi postaci $(0, \ldots, 0, 1, \ldots, 1)$ (jest ich $n+1$) dają permutację identycznościową. Natomiast pozostałe ciągi generują różne permutacje. W związku z tym, miara generująca proces tasowania odwrotnego to

$$\widetilde{Q}(\sigma) = \begin{cases} \frac{1}{2^n} & \text{jeżeli } \sigma \in \Lambda, \\ \frac{n+1}{2^n} & \text{jeżeli } \sigma = \mathrm{id}, \\ 0 & \text{w przeciwnym wypadku,} \end{cases}$$

gdzie Λ oznacza zbiór permutacji generowanych przez ciągi $\{0,1\}^n$ bez identyczności. W uwadze 3.29 uzasadniliśmy, że takie permutacje są odwrotnymi do permutacji uzyskiwanych za pomocą tasowania „riffle shuffle”. □

Do dalszych rozważań, będziemy potrzebowali więcej informacji na temat spacerów odwrotnych.

Uwaga 3.31. Niech $\widetilde{P}$ oznacza macierz przejścia dla łańcucha $\tilde{X}_t$. Ponieważ $\tilde{X}_t$ również jest spacerem losowym na grupie G, to

$$\widetilde{P}(g, hg) = \tilde{\nu}(h) = \nu(h^{-1}) = P(g, h^{-1}) = P(hg, g).$$

Co więcej, jego miarą stacjonarną również jest U, ponieważ

$$\begin{aligned} U\widetilde{P}(g) &= \sum_{h \in G} U(h)\widetilde{P}(h, g) = \frac{1}{|G|} \sum_{k \in G} \widetilde{P}(k^{-1}g, g) = \frac{1}{|G|} \sum_{k \in G} P(g, k^{-1}g) \\ &= \frac{1}{|G|} \sum_{k \in G} \nu(k^{-1}) = \frac{1}{|G|} = U(g). \end{aligned}$$

Lemat 3.32. *Niech X_t będzie spacerem losowym na grupie G generowanym przez miarę probabilistyczną ν. Natomiast $\tilde{X}_t$ będzie oznaczać spacer losowy generowany przez $\tilde{\nu}$. Oznaczmy przez ν_t i $\tilde{\nu}_t$ rozkłady X_t i $\tilde{X}_t$. Wówczas*

$$\|\nu_t - U\|_{TV} = \|\tilde{\nu}_t - U\|_{TV},$$

gdzie U jest miarą jednostajną na grupie G.

Dowód. Niech $\{X_t\} = (e, X_1, \dots)$ oznacza dowolny spacer losowy na grupie G generowanym przez miarę probabilistyczną ν, taki że zaczynamy w elemencie neutralnym. Wtedy możemy napisać, że $X_k = g_k \dots g_1$, gdzie $g_1, \dots, g_k \in G$ to jego przyrosty. natomiast niech $\{Y_t\}$ oznacza dowolny spacer losowy na grupie G generowany przez miarę probabilistyczną $\tilde{\nu}$, a jego przyrosty oznaczymy h_t.

Dla ustalonych $a_1, \dots, a_t \in G$ zachodzi

$$\mathbb{P}(g_1 = a_1, \dots, g_t = a_t) = \mathbb{P}(g_1 = a_1) \dots \mathbb{P}(g_t = a_t) = \nu(a_1) \dots \nu(a_t) = \\ = \widetilde{\nu}(a_1^{-1}) \dots \widetilde{\nu}(a_t^{-1}) = \mathbb{P}(h_1 = a_t^{-1}, \dots, h_t = a_1^{-1}).$$

Co, zsumowane stronami po $a_1, \dots, a_t$, takich że $a_t \dots a_1 = a$, daje

$$P^t(e, a) = \widetilde{P}^t(e, a^{-1}).$$

Stąd

$$\sum_{a \in G} \left| P^t(e, a) - \frac{1}{|G|} \right| = \sum_{a \in G} \left| \widetilde{P}^t(e, a^{-1}) - \frac{1}{|G|} \right| = \sum_{a \in G} \left| \widetilde{P}^t(e, a) - \frac{1}{|G|} \right|,$$

co w połączeniu z faktem 2.14 daje

$$\|P^t(e, \cdot) - U\|_{TV} = \|\widetilde{P}^t(e, \cdot) - U\|_{TV},$$

czyli tezę, bo oczywiście $P^t(e, \cdot) = \nu_t$. □

Z lematu wynika, że wystarczy badać spacer odwrotny do tasowania „riffle shuffle", czyli tasowanie odwrotne. Zdefiniujemy dla niego czas silnie stacjonarny, a następnie zastosujemy metodę czasów silnie stacjonarnych.

Uwaga 3.33. Tasowanie odwrotne n kart można przedstawić za pomocą macierzy binarnej o n wierszach i tylu kolumnach ile tasowań wykonujemy. i-ty wiersz macierzy odpowiada i-tej karcie w pierwotnym ułożeniu kart, a k-ta kolumna odpowiada k-temu tasowaniu. Zatem w polu (i, k) wpisujemy etykietę (0 lub 1), którą dostała i-ta karta w k-tym tasowaniu.

Lemat 3.34 (J. Reeds). *Niech $\{X_t\}$ będzie łańcuchem Markowa na $\mathcal{S}_n$ odpowiadającym modelowi tasowania odwrotnego. Definiujemy zmienną losową $\tau : \mathcal{S}_n \to \{1, 2, \dots, \infty\}$, jako najmniejszą liczbę tasowań taką, że macierz binarna opisana w uwadze 3.33 ma wszystkie wiersze parami różne. τ jest silnym czasem stacjonarnym dla tak zdefiniowanego łańcucha Markowa.*

Dowód. Opiszmy, jak wygląda tasowanie odwrotne. W pierwszym ruchu każdą z kart etykietujemy 0 lub 1. Następnie te z etykietami 0 dajemy na górę talii, pozostałe na dół (nie zmieniając kolejności kart w tych grupach względem siebie!). Następnie czynność tę powtarzamy. Po dwóch tasowaniach na górze talii są karty z wierszem macierzy zaczynającym się od $(0, 0)$, następnie $(1, 0)$, $(0, 1)$, a na końcu $(1, 1)$. Po k tasowaniach, każda karta będzie miała etykietę o długości k. Karty o różnych etykietach będą posortowane leksykograficznie, natomiast karty o tych samych etykietach będą w wyjściowym ułożeniu względem siebie. W momencie τ kończymy tasowanie.

τ definiujemy jako moment w którym wszystkie wiersze macierzy są różne, ponieważ od tego momentu układ żadnej karty nie zależy od układu żadnej innej. Dzieje się tak, bo każde dwie karty się od siebie „odkleiły" w momencie, w którym wiersze im odpowiadające pierwszy raz się różniły.

Przeprowadzimy dowód formalny. Najpierw udowadniamy, że τ jest zrandomizowanym momentem zatrzymania dla spaceru $\{X_t\}$. W tym celu użyjemy następującej reprezentacji przez losowe odwzorowania

$$\Lambda = \{0,1\}^n, \quad Z \sim U(\Lambda),$$

gdzie $Z_i \sim Z$ będzie oznaczało etykiety nadane w i-tym ruchu. W tym języku łatwo zauważyć, że

$$\tau = \min\{t : \text{wiersze macierzy } (Z_1, \ldots, Z_t) \text{ są różne}\},$$

czyli funkcja $\mathbf{1}_{\{\tau=t\}}$ zależy wyłącznie od wektora $(Z_1, \ldots, Z_t)$. Stąd τ jest zrandomizowanym momentem zatrzymania $\{X_t\}$.

Teraz dążymy do pokazania, że

$$\mathbb{P}_x(X_t = y | t = \tau) = \frac{1}{n!}.$$

Co zakończy dowód.

W dalszej części rozważać będziemy wiersze macierzy. Niech zmienna losowa V_i oznacza i-ty wiersz macierzy.

Zauważmy, że dla macierzy binarnych reprezentujących tasowanie odwrotne

$$\mathbb{P}(V_1 = v_1, \ldots, V_n = v_n | \{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau) = \frac{1}{n!}.$$

Oznacza to, że w momencie τ, jeżeli ustalimy jakie wiersze ma mieć otrzymana macierz, to otrzymanie każdej (z możliwych do otrzymania) macierzy jest równie prawdopodobne. Dzieje się tak, ponieważ wszystkie wiersze macierzy tworzą się niezależnie, czyli

$$\begin{aligned}
&\mathbb{P}(V_1 = v_1, \ldots, V_n = v_n | \{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau) \\
&\quad = \frac{\mathbb{P}(V_1 = v_1, \ldots, V_n = v_n, \{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau)}{\mathbb{P}(\{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau)} \\
&\quad = \frac{\mathbb{P}(V_1 = v_1, t = \tau) \cdot \cdots \cdot \mathbb{P}(V_n = v_n, t = \tau)}{\mathbb{P}(\{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau)} = \frac{\left(\frac{1}{2^\tau}\right)^n}{\left(\frac{1}{2^\tau}\right)^n \cdot n!} = \frac{1}{n!}.
\end{aligned}$$

Ponieważ macierz binarna oraz X_0 wyznaczają cały spacer (czyli w szczególności X_τ), a macierz binarna nie zależy od X_0, to możemy napisać, że

$$\begin{aligned}
&\mathbb{P}_x(X_t = y | \{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau) \\
&\quad = \mathbb{P}_x(V_1 = v_{i_1}, \ldots, V_n = v_{i_n} | \{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau) \\
&\quad = \mathbb{P}(V_1 = v_{i_1}, \ldots, V_n = v_{i_n} | \{V_1, \ldots, V_n\} = \{v_1 \ldots, v_n\}, t = \tau) = \frac{1}{n!}.
\end{aligned}$$

Stąd dostajemy tezę

$$
\begin{aligned}
&\mathbb{P}_x(X_t = y|t=\tau)\\
&\quad = \sum_{\{v_1,\ldots,v_n\}} \mathbb{P}_x\left(X_t = y|\{V_1,\ldots,V_n\} = \{v_1\ldots,v_n\}, t=\tau\right)\\
&\qquad \cdot \mathbb{P}_x\left(\{V_1,\ldots,V_n\} = \{v_1\ldots,v_n\}|t=\tau\right)\\
&\quad =\frac{1}{n!}\sum_{\{v_1,\ldots,v_n\}} \mathbb{P}_x(\{V_1,\ldots,V_n\} = \{v_1\ldots,v_n\}|t=\tau) = \frac{1}{n!}.
\end{aligned}
$$

□

Aby oszacować do ilu tasowań możemy się ograniczyć, tak by móc powiedzieć, że potasowaliśmy karty, pozostaje policzyć prawdopodobieństwo zdarzenia, że $\tau > t$ oraz skorzystać z twierdzenia 2.16. $\mathbb{P}(\tau \leqslant t)$ to prawdopodobieństwo, że jeżeli wrzucimy n kul do 2^t urn, to w żadnej urnie nie będzie dwóch kul. Inaczej, jeśli pomyślimy o kulach jako o ludziach, dostajemy popularny problem urodzin. Stąd następujące twierdzenie.

Twierdzenie 3.35. *Dla spaceru odpowiadającego tasowaniu odwrotnemu zachodzi*

$$d(t) \leqslant \mathbb{P}(\tau > t) = 1 - \prod_{i=1}^{n-1}\left(1 - \frac{i}{2^t}\right)$$

Dowód. Z konstrukcji τ wiemy, że nie zależy ono od X_0. Pozostaje więc udowodnić, że

$$\mathbb{P}(\tau \leqslant t) = \prod_{i=1}^{n-1}\left(1 - \frac{i}{2^t}\right),$$

co w połączeniu z twierdzeniem 2.16 da tezę.

Zauważmy, że zdarzenie $\tau \leqslant t$ jest równoważne temu, że macierz $n \times t$ ma wszystkie wiersze różne. Czyli, prawdopodobieństwo liczymy następująco. Wybieramy n wierszy z 2^t możliwych i je ustawiamy, a potem dzielimy przez liczbę wszystkich macierzy binarnych $n \times t$. Stąd

$$\mathbb{P}(\tau \leqslant t) = \frac{\binom{2^t}{n}n!}{(2^t)^n} = \frac{\frac{(2^t)!}{(2^t-n)!}}{(2^t)^n} = \frac{2^t\cdot(2^t-1)\ldots(2^t-n+1)}{(2^t)^n} = \prod_{i=1}^{n-1}\left(1 - \frac{i}{2^t}\right).$$

□

Wniosek 3.36. *Wstawiając $n = 52$ możemy obliczyć otrzymaną wielkość. Odpowie ona na pytanie, ile razy należy wykonać tasowanie „riffle shuffle", tak aby talię można było uznać za potasowaną.*

t	10	11	12	13	14
górne szacowanie	0.73	0.48	0.28	0.15	0.08

Z tabeli powyżej widzimy, że z pewnością w takim problemie $t_{mix} \leqslant 13$.

Jednak w ogólności wielkość otrzymana po prawej stronie nierówności nie jest łatwa do obliczania. Możemy ją trochę uprościć zakładając pewne zbieżności.

Wniosek 3.37. *Jeżeli* $t = 2\log_2(n/c)$, *to*

$$\mathbb{P}(\tau > t) \underset{\infty}{\overset{n}{\sim}} 1 - e^{-\frac{c^2}{2}} \underset{0}{\overset{c}{\sim}} \frac{c^2}{2},$$

gdzie $a_n \underset{\infty}{\overset{n}{\sim}} b_n$ *oznacza* $a_n/b_n \xrightarrow{n\to\infty} 1$.

Dowód. Pierwsza część dowodu to udowodnienie zbieżności

$$\mathbb{P}(\tau \leqslant t) = \prod_{i=1}^{n-1}\left(1 - \frac{i}{2^t}\right) = \prod_{i=1}^{n-1}\left(1 - \frac{ic^2}{n^2}\right) \xrightarrow{n\to\infty} e^{-\frac{c^2}{2}}.$$

Aby oszacować iloczyn z góry korzystamy z nierówności $1 - x \leqslant e^{-x}$:

$$\prod_{i=1}^{n-1}\left(1 - \frac{ic^2}{n^2}\right) \leqslant \exp\left(-\left(\frac{c}{n}\right)^2 \cdot \sum_{i=1}^{n-1} i\right) = \exp\left(-\left(\frac{c}{n}\right)^2 \cdot \frac{(n-1)n}{2}\right) \xrightarrow{n\to\infty} e^{-c^2}.$$

Aby oszacować iloczyn z dołu zastosujemy nierówność $e^{-x^2-x} \leqslant 1 - x$, która jest prawdziwa dla dla $x \in \left[0, \frac{1}{2}\right]$ (poniżej podamy jej dowód). Dla ustalonego c bierzemy tak duże n, że $\frac{(n-1)c^2}{n^2} \leqslant \frac{1}{2}$, wtedy

$$\begin{aligned}\prod_{i=1}^{n-1}\left(1 - \frac{ic^2}{n^2}\right) &\geqslant \exp\left(-\left(\frac{c}{n}\right)^2 \cdot \sum_{i=1}^{n-1} i - \left(\frac{c}{n}\right)^4 \cdot \sum_{i=1}^{n-1} i^2\right) \\ &= \exp\left(-\left(\frac{c}{n}\right)^2 \cdot \frac{(n-1)n}{2} - \left(\frac{c}{n}\right)^4 \cdot \sum_{i=1}^{n-1} i^2\right) \xrightarrow{n\to\infty} e^{-c^2},\end{aligned}$$

bo

$$0 \leqslant \left(\frac{c}{n}\right)^4 \cdot \sum_{i=1}^{n-1} i^2 \leqslant \left(\frac{c}{n}\right)^4 \cdot \sum_{i=1}^{n-1} (n-1)^2 = \frac{c^4(n-1)^3}{n^4} \leqslant \frac{c^4}{n} \to 0.$$

Przeprowadźmy dowód nierówności $e^{-x-x^2} \leqslant 1 - x$ dla $x \in [0, \frac{1}{2}]$. Logarytmujemy ją stronami i dostajemy równoważną nierówność $-x - x^2 \leqslant \log(1-x)$ Nowa nierówność jest spełniona dla 0, więc wystarczy pokazać, że na $[0, \frac{1}{2}]$ pochodna $\log(1-x) + x + x^2$ jest dodatnia

$$1 + 2x + \frac{-1}{1-x} \geqslant 0 \quad \Leftrightarrow \quad (1-x)(1+2x) \geqslant 1 \quad \Leftrightarrow \quad 2x \leqslant 1$$

bo rozważamy liczby dodatnie.

Druga część dowodu sprowadza się do policzenia granicy

$$\lim_{c\to 0} \frac{1 - e^{-c^2}}{\frac{c^2}{2}} \overset{(H)}{=} \lim_{c\to 0} \frac{-e^{-c^2} \cdot (-c)}{c} = \lim_{c\to 0} e^{-c^2} = 1.$$

$\square$

Wniosek 3.38. *Dla dużych n*

$$t_{mix} \leqslant 2\log_2(4n/3).$$

Wynik ten otrzymujemy wstawiając $c = \frac{3}{4}$ do lewej zbieżności w poprzednim wniosku. Wtedy

$$d(2\log_2(4n/3)) \leqslant 1 - e^{-\frac{c^2}{2}} \approx 0.245.$$

Uwaga 3.39. W ogólności można udowodnić mocniejszy wynik. Przyjmuje się, że 7 tasowań wystarczy, aby potasować talię. Do tego natomiast potrzebne są znacznie trudniejsze metody. Wyniki te można znaleźć w artykule [1] z 1992.

Literatura

[1] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992.

[2] D. Buraczewski. Spacery losowe na skończonych grafach. Skrypt do wykładu `http://www.math.uni.wroc.pl/~dbura/dydaktyka/random_walks.pdf`.

[3] P. Diaconis. *Group representations in probability and statistics*, volume 11. Hayward, CA: Institute of Mathematical Statistics, 1988.

[4] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times. With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.* Providence, RI: American Mathematical Society (AMS), 2017.

[5] P. Lorek. *Speed of convergence to stationarity for stochastically monotone Markov chains.* PhD thesis, Uniwersytet Wrocławski, 2007. `http://www.math.uni.wroc.pl/~lorek/files/lorek_thesis.pdf`.