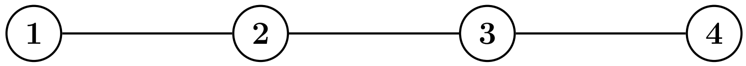
Wstęp do rachunku prawdopodobieństwa
13. Łańcuchy Markowa. Twierdzenie ergodyczne.

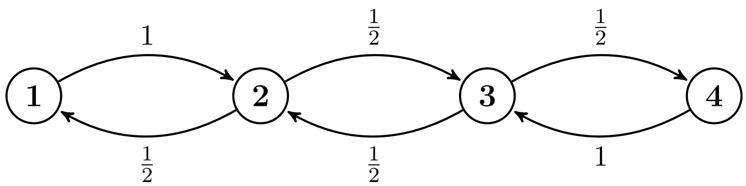
Zadanie 1. *Przedstaw za pomocą łańcucha Markowa następujące procesy:*

(a) *Cząstka błądzi klasycznie na ścieżce o 4 wierzchołkach.*

Przypomnijmy, że ścieżka na czterech wierzchołkach to poniższy graf:



Graf skierowany łańcucha Markowa dla błądzenie na tej ścieżce ma te same wierzchołki co ścieżka, natomiast każda krawędź $\{i, j\}$ jest zastąpiona dwiema krawędziami skierowanymi (i, j) oraz (j, i) z wagami równymi odpowiednio $\deg(i)^{-1}$ i $\deg(j)^{-1}$, jak na poniższym rysunku:



Macierzą przejścia dla tego łańcucha jest macierz:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Zbadaj, czy ten łańcuch jest nieprzywiedlny i nieokresowy, a następnie wyznacz jego rozkład stacjonarny.

Wiemy, że błądzenie na grafie G jest łańcuchem:

- nieprzywiedlnym wtedy i tylko wtedy, gdy G jest spójny,
- nieokresowym wtedy i tylko wtedy, gdy G nie jest dwudzielny.

Ponieważ ścieżka na 4 wierzchołkach jest grafem spójnym i dwudzielnym, błądzenie klasyczne na tym grafie jest łańcuchem nieprzywiedlnym, ale nie jest łańcuchem nieokresowym.

Ponadto, zgodnie z twierdzeniem wektor

$$\left(\frac{\deg(1)}{2e(G)}, \frac{\deg(2)}{2e(G)}, \ldots, \frac{\deg(v(G))}{2e(G)}\right)$$

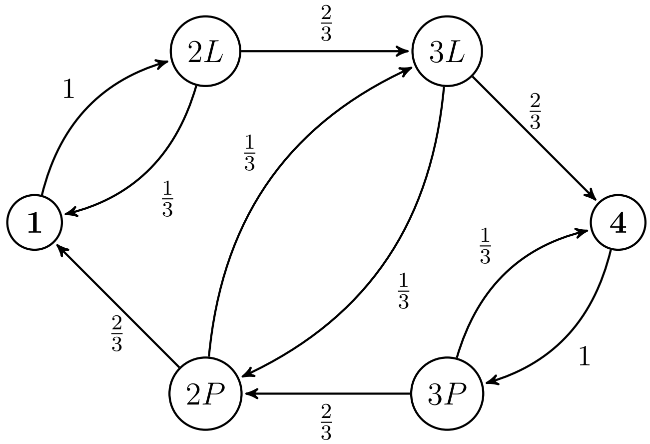
jest rozkładem stacjonarnym dla błądzenia na grafie G.

W naszym przypadku $e(G) = 3$, natomiast ciągiem stopni jest ciąg $(1, 2, 2, 1)$. Zatem rozkładem stacjonarnym dla błądzenia ma ścieżce o 4 wierzchołkach jest wektor:

$$\left(\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\right).$$

(b) *Cząstka błądzi (nieklasycznie) na ścieżce o 4 wierzchołkach w następujący sposób: jeśli przybyła do wierzchołka wewnętrznego ścieżki z lewej strony, to w następnym kroku szansa pójścia w lewo jest dwa razy mniejsza od szansy pójścia w prawo. Analogiczna reguła dotyczy przybycia z prawej strony.*

W tym przypadku prawdopodobieństwo przejścia ze środkowych wierzchołków, czyli wierzchołków numer 2 i 3, będzie zależało od tego, z której strony weszliśmy do takiego wierzchołka. Aby zareprezentować to błądzenie za pomocą łańcucha Markowa, musimy w takim razie dla każdego ze środkowych wierzchołków wprowadzić dwa stany, np. $2P$ będzie oznaczać, że jesteśmy w wierzchołku numer 2 i weszliśmy do niego z prawej strony, a $2L$ że jesteśmy w wierzchołku numer 2, ale weszliśmy do niego z lewej strony. Analogicznie definiujemy stany $3P$ i $3L$. Stanów numer 1 i 4 nie musimy dublować, bo tam sytuacja jest jasna. Szukany łańcuch Markowa wygląda następująco:

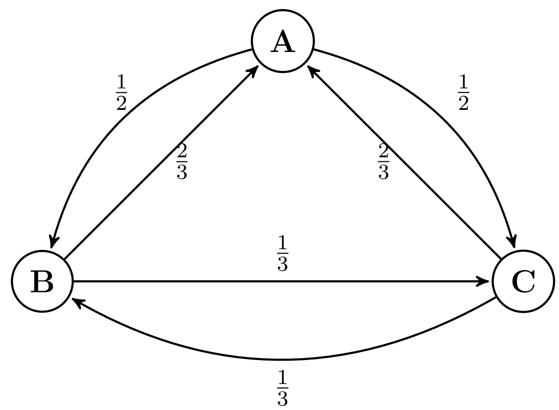


Macierzą przejścia dla tego łańcucha jest poniższa macierz (wiersze indeksujemy kolejno stanami $1, 2L, 3L, 4, 3P, 2P$).

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \end{bmatrix}$$

Zadanie 2. *Komputer generuje tekst złożony z liter A, B, C według następujących reguł: Zaczyna od losowo wybranej litery. Jeżeli ostatnio napisana litera jest samogłoską, następna będzie spółgłoską wybraną z jednakowym prawdopodobieństwem. Jeżeli ostatnio napisana litera jest spółgłoską, następna będzie inna, przy czym samogłoska jest wybierana z prawdopodobieństwem dwa razy większym niż spółgłoska.*

(a) *Przedstaw ten proces w postaci łańcucha Markowa.*



Macierzą przejścia tego łańcucha jest macierz

$$\Pi = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}.$$

(b) *Wyznacz jego rozkład stacjonarny.*

Szukamy wektora $\bar{\pi} = (x, y, 1-x-y)$ spełniającego zależność $\bar{\pi}\Pi = \bar{\pi}$. A zatem mamy:

$$\begin{aligned}(x, y, 1-x-y) \cdot \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix} &= \left(\frac{2}{3}y + \frac{2}{3}(1-x-y), \frac{1}{2}x + \frac{1}{3}(1-x-y), \frac{1}{2}x + \frac{1}{3}y\right) \\ &= \left(\frac{2}{3} - \frac{2}{3}x, \frac{1}{3} + \frac{1}{6}x - \frac{1}{3}y, \frac{1}{2}x + \frac{1}{3}y\right),\end{aligned}$$

skąd otrzymujemy następujący układ równań:

$$\begin{cases} x = \frac{2}{3} - \frac{2}{3}x, \\ y = \frac{1}{3} + \frac{1}{6}x - \frac{1}{3}y, \\ 1 - x - y = \frac{1}{2}x + \frac{1}{3}y. \end{cases}$$

Rozwiązaniem tego układu jest:

$$\begin{cases} x = \frac{2}{5}, \\ y = \frac{3}{10}, \end{cases}$$

zatem szukany rozkład stacjonarny to wektor $\bar{\pi} = \left(\frac{2}{5}, \frac{3}{10}, \frac{3}{10}\right)$.

(c) *Zbadaj, czy łańcuch ten jest nieprzywiedlny i nieokresowy.*

Nietrudno zauważyć, że łańcuch jest nieprzywiedlny, bo z każdego stanu możemy dojść do każdego innego w skończonej liczbie kroków, szczególnie widać to na przykładzie skierowanego cyklu A–B–C–A. Z kolei skoro mamy do czynienia z łańcuchem nieprzywiedlnym, wszystkie stany mają ten sam okres i wystarczy obliczyć okres jednego z nich. Na przykład dla stanu A mamy:

$$p_{AA}^{(2)} > 0 \ \text{ oraz } \ p_{AA}^{(3)} > 0,$$

zatem

$$d(A) = NWD\{t \geqslant 1 : p_{AA}^{(t)} > 0\} = NWD\{2, 3, \ldots\} = 1.$$

Czyli okres każdego stanu wynosi 1 i łańcuch ten jest nieokresowy.

Zadanie 3. *Komputer generuje tekst złożony z liter A, B, C według następujących reguł: Zaczyna od losowo wybranego słowa długości trzy składającego się z różnych liter, wybranego spośród tych, które nie zawierają podsłowa CA ani BAC. Każda następna litera jest wybrana z jednakowym prawdopodobieństwem spośród spełniających dwa warunki:*

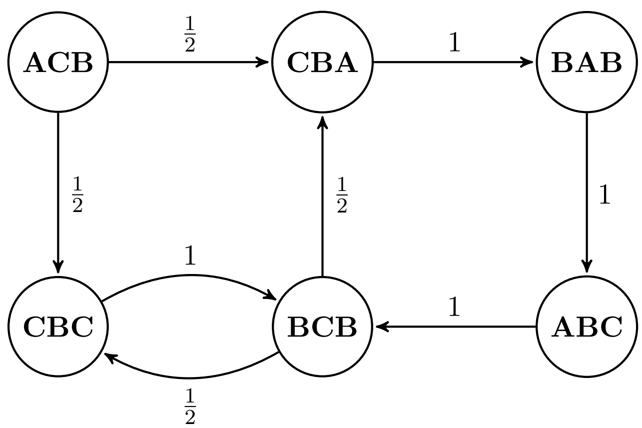
- *litera jest inna od poprzedniej,*
- *w tekście nie mogą pojawić się wyrazy: BABA, BAC, CA.*

(a) *Przedstaw ten proces w postaci łańcucha Markowa.*

Zacznijmy od **stanów startowych**. Spośród słów długości trzy składających się z różnych liter, które nie zawierają podsłowa CA ani BAC do dyspozycji mamy: **ABC**, **ACB**, **CBA**.

Przechodząc do kolejnych stanów będziemy przechowywać w pamięci trzy ostatnio wybrane litery, zatem stanami będą trzyliterowe słowa, które mogą pojawić się w tekście.

- Będąc w stanie **ABC** nie możemy wybrać litery C, bo nowa litera ma być inna od poprzedniej. Nie możemy też wybrać litery A, bo w tekście pojawi się wyraz CA. Zatem z prawdopodobieństwem 1 wybieramy literę B i przechodzimy do stanu **BCB**.
- W stanie **BCB** jedyna litera jakiej nie możemy wybrać to B. Zatem z równym prawdopodobieństwem wybierzemy literę C lub A i przejdziemy, odpowiednio, do stanu **CBC** lub **CBA**.
- Będąc w stanie **CBC** nie możemy wybrać ani litery A, bo utworzymy słowo CA, ani litery C, bo utworzymy powtórzenie, zatem z prawdopodobieństwem 1 przechodzimy do stanu **BCB**.
- Z kolei w stanie **CBA** nie możemy wybrać litery A, bo utworzymy powtórzenie, ani litery C, bo utworzymy słowo BAC. Zatem z prawdopodobieństwem 1 wybieramy literę B i przesuwamy się do stanu **BAB**.
- W stanie **BAB** nie możemy wybrać litery B, bo utworzymy powtórzenie, ani litery A, bo utworzymy słowo BABA, zatem z prawdopodobieństwem 1 wybieramy literę C i przechodzimy do stanu **ABC**.
- Jeśli natomiast zaczniemy od stanu **ACB**, to możemy wybrać zarówno literę A, jak i C, zatem z równym prawdopodobieństwem przechodzimy do jednego ze stanów **CBA** lub **CBC**.



(b) *Zbadaj, czy łańcuch jest nieprzywiedlny.*

Łańcuch ten nie jest nieprzywiedlny, bo np. do stanu **ACB** nie da się dojść z żadnego innego stanu.

(c) *Zbadaj, czy łańcuch jest nieokresowy.*

Łańcuch nie jest nieokresowy. Rozważmy na przykład stan **CBC**. Wszystkie spacery, które zaczynają się i kończą w tym stanie mają parzystą długość. Ponadto jest wśród nich spacer długości 2, zatem okres tego stanu jest też równy 2.

Zadanie 4. *Ala i Franek rzucają na zmianę standardową monetą (zaczyna Ala). Franek wygrywa, gdy po raz pierwszy w ciągu rzutów pojawi się ciąg ROR (reszka, orzeł, reszka), a Ala gdy pojawi się ciąg OOR (orzeł, orzeł, reszka). Oblicz prawdopodobieństwo wygranej Ali.*

Stworzymy łańcuch Markowa opisujący powyższą grę. Na początek wprowadzimy stan startowy **S** oznaczający początek gry. Oprócz tego będziemy rozważali tylko takie sekwencje ostatnich rzutów, które mają wpływ na potencjalną wygraną któregoś z graczy. Zatem dopóki nie pojawiła się sekwencja wygrywająca wystarczy, że będziemy pamiętali wyniki co najwyżej dwóch ostatnich rzutów. Oprócz tego wprowadzimy dwa stany **ROR** i **OOR**, które kończą grę. Są to stany odpowiadające sekwencjom wygrywającym. Aby zasymulować koniec gry możemy przyjąć, że są to stany pochłaniające.

- Zanim wykonamy pierwszy rzut znajdujemy się w stanie startowym **S**. Ze stanu **S** możemy przejść z równym prawdopodobieństwem albo do stanu **R** – po wyrzuceniu reszki, albo do stanu **O** – po wyrzuceniu orła. Po czym nigdy nie powrócimy do tego stanu.

- Załóżmy, że jesteśmy teraz w stanie **R**. Z punktu widzenia gry oznacza to, że Franek ma szansę wygrać, jeśli w dwóch kolejnych rzutach wypadnie odpowiednio orzeł, a potem reszka. Zatem po wyrzucenia orła przesuwamy się do stanu **RO**, dzięki czemu wiemy, że w dwóch ostatnich rzutach wypadła właśnie taka sekwencja. Jeśli natomiast wypadnie reszka, otrzymamy dwie reszki pod rząd w dwóch ostatnich rzutach. Na dalszy przebieg gry wpływa jedynie fakt, że wynikiem ostatniego rzutu była reszka, zatem możemy przyjąć, że pozostajemy w stanie **R**. Podsumowując, ze stanu **R** z równym prawdopodobieństwem przechodzimy albo do stanu **RO**, albo z powrotem do stanu **R**.

- Jeżeli jesteśmy w stanie **RO** i w kolejnym rzucie wypadnie reszka, to otrzymamy sekwencję ROR, która oznacza wygraną Franka. Zatem po wyrzuceniu reszki przesuwamy się do stanu **ROR**. Natomiast jeśli wypadnie orzeł, w dwóch ostatnich rzutach będziemy mieli dwa orły, co daje szansę na wygraną Ali, jeśli w kolejnym rzucie wypadnie reszka. W związku z tym po wyrzuceniu orła przesuwamy się do stanu **OO**. Czyli ze stanu **RO** z równym prawdopodobieństwem przesuwamy się do stanu **ROR** lub **OO**.

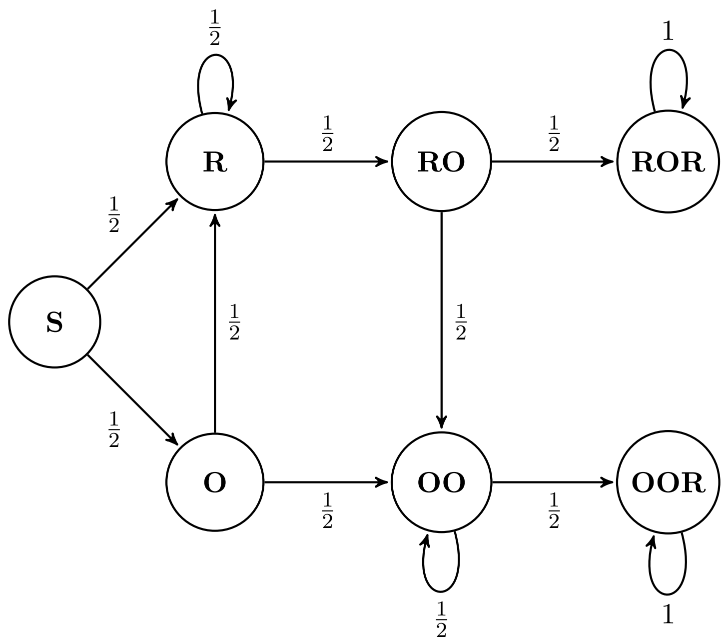
- Stan **ROR** oznacza wygraną Franka i koniec gry. W związku z tym po dojściu do tego stanu zostajemy w nim z prawdopodobieństwem 1.

- Stan **OO** oznacza, że w dwóch ostatnich rzutach wypadł orzeł. Jeśli teraz wypadnie reszka, to otrzymamy sekwencję OOR, która oznacza wygraną Ali. W tej sytuacji przesuwamy się do stanu **OOR**. Natomiast jeśli wypadnie orzeł, to zapamiętujemy jedynie wynik dwóch ostatnich rzutów, czyli dwa orły i pozostajemy w stanie **OO**. Podsumowując, ze stanu **OO** z równym prawdopodobieństwem przechodzimy albo do stanu **OOR**, albo pozostajemy w stanie **OO**.

- Stan **OOR** jest też stanem kończącym grę, bo oznacza wygraną Ali. Po dojściu do tego stanu już z niego nie wyjdziemy.

- Zostaje nam jeszcze stan **O**, czyli orzeł wyrzucony w pierwszym rzucie gry. Jeżeli w kolejnym rzucie wypadnie znów orzeł, to otrzymamy sekwencję dwóch orłów pod rząd, zatem przesuwamy się do stanu **OO**. Jeśli natomiast wypadnie reszka, daje to sekwencję OR, która z punktu widzenia gry jest tym samym, co start z reszki. Zatem po wyrzuceniu reszki przesuwamy się do stanu **R**. Innymi słowy będąc w stanie **O** z równym prawdopodobieństwem przechodzimy do jednego ze stanów **OO** lub **R**.

Powyższe rozważania możemy przedstawić na grafie skierowanym, co da nam pełny obraz gry.



Chcemy teraz obliczyć prawdopodobieństwo wygranej Ali, czyli dojście do stanu **OOR**. Musimy w tym celu rozważyć wszystkie skończone ścieżki prowadzące ze stanu startowego **S** do stanu końcowego **OOR.** Mamy tutaj trzy możliwe scenariusze:

1. **S** → **O** → **OO** → **OOR**,
2. **S** → **R** → **RO** → **OO** → **OOR**,
3. **S** → **O** → **R** → **RO** → **OO** → **OOR**

przy czym jeśli wejdziemy do stanu **R** lub do stanu **OO** możemy wykonać dowolną skończoną liczbę pętli w tym stanie. Prawdopodobieństwa ścieżek dla każdego z tych scenariuszy wynoszą:

1. $$\frac{1}{2}\cdot\frac{1}{2}\cdot\left(\left(\frac{1}{2}\right)^0+\left(\frac{1}{2}\right)^1+\left(\frac{1}{2}\right)^2+\dots\right)\cdot\frac{1}{2}=\frac{1}{2}\cdot\frac{1}{2}\cdot 2\cdot\frac{1}{2}=\frac{1}{4}$$

2. $$\frac{1}{2}\cdot\left(\left(\frac{1}{2}\right)^0+\left(\frac{1}{2}\right)^1+\left(\frac{1}{2}\right)^2+\dots\right)\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\left(\left(\frac{1}{2}\right)^0+\left(\frac{1}{2}\right)^1+\left(\frac{1}{2}\right)^2+\dots\right)\cdot\frac{1}{2}=\frac{1}{4}$$

3. $$\frac{1}{2}\cdot\frac{1}{2}\cdot\left(\left(\frac{1}{2}\right)^0+\left(\frac{1}{2}\right)^1+\left(\frac{1}{2}\right)^2+\dots\right)\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\left(\left(\frac{1}{2}\right)^0+\left(\frac{1}{2}\right)^1+\left(\frac{1}{2}\right)^2+\dots\right)\cdot\frac{1}{2}=\frac{1}{8}$$

A zatem prawdopodobieństwo wygranej Ali jest równe

$$\frac{1}{4}+\frac{1}{4}+\frac{1}{8}=\frac{5}{8}.$$