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

Twierdzenie. *Jeśli łańcuch Markowa $\Pi = [p_{ij}]$ jest łańcuchem z* **symetryczną** *macierzą przejścia, to znaczy $p_{ij} = p_{ji}$ dla każdych i oraz j, to łańcuch ten posiada co najmniej jeden rozkład stacjonarny, mianowicie rozkład jednostajny na stanach.*

Przypomnijmy, że przez $p_{ij}^{(t)}$ oznaczamy prawdopodobieństwo przejścia ze stanu i do stanu j w dokładnie t krokach.

Definicja. *Łańcuch Markowa nazywamy* **nieprzywiedlnym**, *gdy dla dowolnych jego stanów i oraz j istnieje takie $t \geqslant 1$, że $p_{ij}^{(t)} > 0$.*

Definicja. **Okresem** *stanu j nazywamy liczbę $d(j) = NWD\{t \geqslant 1 : p_{jj}^{(t)} > 0\}$. Dla stanu j nie definiujemy okresu gdy $p_{jj}^{(t)} = 0$ dla wszystkich $t \geqslant 1$.*

Gdy $d(j) > 1$, wtedy mówimy, że stan j jest **okresowy**, *w przeciwnym wypadku mówimy, że stan ten jest* **nieokresowy**. *Przyjmujemy, że stany nieposiadające okresu też są nieokresowe.*

Podobnie łańcuch, w którym wszystkie stany są okresowe nazywamy łańcuchem **okresowym**, *w przeciwnym wypadku mówimy, że łańcuch ten jest* **nieokresowy**.

Fakt. W nieprzywiedlnym łańcuchu Markowa wszystkie stany mają ten sam okres.

Definicja. *Mówimy, że łańcuch Markowa jest* **nieokresowy**, *gdy wszystkie jego stany są nieokresowe. Natomiast łańcuch Markowa, który jest nieprzywiedlny i nieokresowy nazywamy łańcuchem* **ergodycznym**.

Twierdzenie (ergodyczne). *Każdy ergodyczny łańcuch Markowa ma dokładnie jeden rozkład stacjonarny $\bar{\pi} = (\pi_1, \pi_2, \ldots, \pi_s)$. Ponadto dla dowolnych stanów i, j zachodzi $\lim_{t\to\infty} p_{ij}^{(t)} = \pi_j$. Innymi słowy, niezależnie od początkowego rozkładu $\bar{\rho}^0$ zachodzi $\lim_{t\to\infty} \bar{\rho}^t = \bar{\pi}$.*

Przypomnijmy, że przez V oznaczamy zbiór wierzchołków, a przez E zbiór krawędzi grafu $G = (V, E)$. Ponadto przyjmujemy oznaczenia $v(G) = |V|$ oraz $e(G) = |E|$, natomiast $\deg(i)$ to stopień wierzchołka i w grafie G, czyli liczba krawędzi, które zawierają ten wierzchołek.

Błądzeniem klasycznym (spacerem losowym) na grafie $G = (V, E)$ bez wierzchołków izolowanych nazywamy proces, w którym cząstka znajdująca się w jakimś wierzchołku grafu w każdym kroku przemieszcza się do jednego z sąsiadów tego wierzchołka, wybranego z jednakowym prawdopodobieństwem. Procesowi temu odpowiada łańcuch Markowa o zbiorze stanów V, gdzie dla dowolnej krawędzi $ij \in E$ mamy

$$p_{ij} = \frac{1}{\deg(i)},$$

natomiast gdy $ij \notin E$ to $p_{ij} = 0$. Gdy mowa o błądzeniu na grafie, zawsze zakładamy, że graf ten nie ma wierzchołków izolowanych. W szczególności musi on mieć co najmniej dwa wierzchołki.

Fakt. Błądzenie na grafie G jest łańcuchem nieprzywiedlnym wtedy i tylko wtedy, gdy G jest spójny. Błądzenie na grafie G jest łańcuchem nieokresowym wtedy i tylko wtedy, gdy G nie jest dwudzielny.

Twierdzenie. *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 $G = (V, E)$.

Dodatek A. Zadania na ćwiczenia

Zadanie 1. Przedstaw za pomocą łańcucha Markowa następujące procesy:

(a) Cząstka błądzi klasycznie na ścieżce o 4 wierzchołkach. Zbadaj, czy ten łańcuch jest nieprzywiedlny i nieokresowy, a następnie wyznacz jego rozkład stacjonarny.

(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.

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.

(b) Wyznacz jego rozkład stacjonarny.

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

(b) Zbadaj, czy łańcuch jest nieprzywiedlny.

(c) Zbadaj, czy łańcuch jest nieokresowy.

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.

Dodatek B. Zadania domowe

Zadanie 1. Macierz przejścia pewnego łańcucha Markowa dana jest następującym wzorem:

$$(a)\begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}, \qquad (b)\begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}, \qquad (c)\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(a) Sprawdź, czy łańcuch jest nieprzywiedlny.

(b) Sprawdź, czy istnieją w nim stany okresowe.

(c) Wyznacz liczbę rozkładów stacjonarnych.

Zadanie 2. Rozpatrzmy klasyczne błądzenie po cyklu o 5 wierzchołkach w_1, w_2, w_3, w_4, w_5, zaczynające się w wierzchołku w_1.

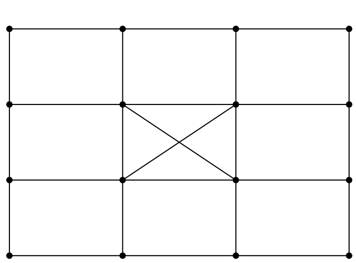
(a) Czy odpowiadający temu błądzeniu łańcuch Markowa jest nieprzywiedlny? Czy jest okresowy?

(b) Wyznacz rozkład stacjonarny tego łańcucha.

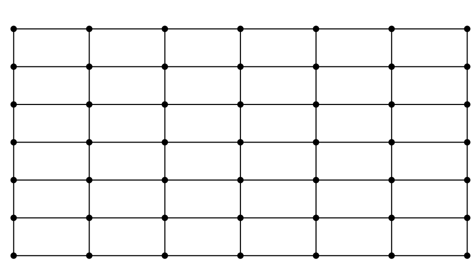
(c) Wyznacz $\lim_{t\to\infty} p^{(t)}_{w_1 w_3}$.

Zadanie 3. Uzasadnij, że błądzenie klasyczne na pierwszym grafie jest nieokresowym łańcuchem Markowa, a błądzenie na drugim grafie jest łańcuchem okresowym.

(a)



(b)



Zadanie 4. Rozważmy błądzenie na pierwszym grafie z poprzedniego zadania. Załóżmy, że startujemy z lewego górnego rogu. Wyznacz rozkład stacjonarny dla tego błądzenia.

Zadanie 5. Rozważamy błądzenie na grafie G o wierzchołkach $w_1, w_2, \ldots, w_{10}$, który jest spójny i nie jest dwudzielny. Niech $\bar{\pi} = (\pi_1, \pi_2, \ldots, \pi_{10})$ będzie rozkładem stacjonarnym dla tego błądzenia. Oceń poprawność poniższych zdań. Odpowiedź należy poprzeć albo uzasadnieniem ogólnym, albo przykładem, albo kontrprzykładem.

(a) $p^{(t)}_{w_1 w_2} \to \frac{\deg(w_1)}{2e(G)}$ przy $t \to \infty$

(b) Jeżeli wierzchołek 2 jest stopnia 5, a $p^{(t)}_{w_1 w_2} \to \frac{1}{8}$ przy $t \to \infty$, to G ma dokładnie 20 krawędzi.

(c) Jeżeli $\bar{\pi} = \left(\frac{1}{10}, \frac{1}{10}, \ldots, \frac{1}{10}\right)$, to G jest grafem regularnym.

Dodatek C. Odpowiedzi do zadań domowych

B.1 (a) Nieprzywiedlny, wszystkie stany mają okres 2, jedyny rozkład stacjonarny to $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$.
(b) Nieprzywiedlny, nieokresowy, jeden rozkład stacjonarny.
(c) Nie jest nieprzywiedlny, nie ma stanów okresowych, ma nieskończenie wiele rozkładów stacjonarnych.

B.2 (a) Łańcuch jest nieprzywiedlny i nieokresowy. (b) $\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$, (c) $\frac{1}{5}$

B.3 (a) nie jest grafem dwudzielnym, zawiera cykl C_3, (b) jest grafem dwudzielnym

B.4 Numerując wierzchołki od lewej strony do prawej, z góry na dół, otrzymujemy jedyny rozkład stacjonarny (bo łańcuch jest ergodyczny) zadany przez ciąg stopni:

$$\left(\frac{2}{52}, \frac{3}{52}, \frac{3}{52}, \frac{2}{52}, \frac{3}{52}, \frac{5}{52}, \frac{5}{52}, \frac{3}{52}, \frac{3}{52}, \frac{5}{52}, \frac{5}{52}, \frac{3}{52}, \frac{2}{52}, \frac{3}{52}, \frac{3}{52}, \frac{2}{52}\right)$$

B.5 (a) Nie. Łańcuch jest ergodyczny, więc jedynym rozkładem stacjonarnym jest wektor $\bar{\pi} = \left(\frac{\deg(w_1)}{2e(G)}, \ldots, \frac{\deg(w_{10})}{2e(G)}\right)$, zatem $\lim_{t\to\infty} p^{(t)}_{w_1 w_2} = \pi_2 = \frac{\deg(w_2)}{2e(G)}$ i wystarczy wziąć graf, w którym $\deg(w_1) \neq \deg(w_2)$
(b) Tak. Zauważmy, że $p^{(t)}_{w_1 w_2} \to \frac{\deg(w_2)}{2e(G)} = \frac{5}{2e(G)} = \frac{1}{8}$, skąd $e(G) = 20$.
(c) Tak, podobnie jak powyżej.