Sandsynlighedsteori og Lineær algebra

Workshop 2 - PageRank

Sebastian Livoni Larsen

June 6, 2022

Delopgave 1

Opgave 2 - Sandsynlighederne $p(X = y)$ for de ikke-tomme urbilleder af y.

$$p(X = 0) = p(w_i w_i w_i) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = 0.125$$

$$p(X = 1) = p(w_1 w_i w_i) + p(w_i w_1 w_i) + p(w_i w_i w_1) = \frac{1}{2} \cdot 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{8} = 0.625$$

$$p(X = 2) = p(w_1 w_1 w_i) + p(w_i w_1 w_1) + p(w_1 w_i w_1) = \frac{1}{2} \cdot 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 \cdot \frac{1}{2} = \frac{2}{8} = 0.25$$

$$p(X = 3) = p(w_1 w_1 w_1) = \frac{1}{2} \cdot 0 \cdot 0 = 0$$

Delopgave 1

Opgave 3 - Middelværdien og variansen for X.

Middelværdi: $E(X) = \sum_{s \in S} p(s)X(s) = 0.125 \cdot 0 + 0.625 \cdot 1 + 0.25 \cdot 2 = \mathbf{1.125}$

$$\begin{aligned} V(X) &= \sum_{s \in S} (X(s) - E(X))^2 p(s) \\ &= (0 - 1.125)^2 \cdot 0.125 + (1 - 1.125)^2 \cdot 0.625 + (2 - 1.125)^2 \cdot 0.25 = \mathbf{0.359375} \end{aligned}$$

Delopgave 1

Opgave 4 - Bernoullifordeling med sandsynlighed $\frac{3}{8}$ for success.

Bernoulli trial $= b(k, n, p)$ where k=successes, n=independent trials, p=probability of success

$$b(k, n, p) = \mathrm{C}(n, k)p^k q^{n-k}$$

$$b\left(0; 3; \frac{3}{8}\right) = \frac{3!}{0! \cdot (3-0)!} \cdot \frac{3}{8}^0 \cdot \left(1 - \frac{3}{8}\right)^{3-0} = 0.244140625$$

$$b\left(1; 3; \frac{3}{8}\right) = \frac{3!}{1! \cdot (3-1)!} \cdot \frac{3}{8}^1 \cdot \left(1 - \frac{3}{8}\right)^{3-1} = 0.439453125$$

$$b\left(2; 3; \frac{3}{8}\right) = \frac{3!}{2! \cdot (3-2)!} \cdot \frac{3}{8}^2 \cdot \left(1 - \frac{3}{8}\right)^{3-2} = 0.263671875$$

$$b\left(3; 3; \frac{3}{8}\right) = \frac{3!}{3! \cdot (3-3)!} \cdot \frac{3}{8}^3 \cdot \left(1 - \frac{3}{8}\right)^{3-3} = 0.052734375$$

Middelværdi: $E(X) = \sum_{s \in S} p(s)X(s) = 0.244140625 \cdot 0 + 0.439453125 \cdot 1 + 0.263671875 \cdot 2 + 0.052734375 \cdot 3 = \mathbf{1.125}$

Varians: $V(X) = \sum_{s \in S} (X(s) - E(X))^2 p(s)$

$$= (0 - 1.125)^2 \cdot 0.244 + (1 - 1.125)^2 \cdot 0.439 + (2 - 1.125)^2 \cdot 0.263 + (3 - 1.125)^2 \cdot 0.052 = \mathbf{0.703125}$$

Delopgave 2

Opgave 1 - Sandsynligheden for at være i tilstand w_1 til tiden $t = 5$.

$$S = \begin{bmatrix} \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}^T$$

$$P^{(5)} = P^5 P^{(0)} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}^5 \begin{bmatrix} \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}^T = \begin{bmatrix} \mathbf{0.328125} \\ 0.16796875 \\ 0.16796875 \\ 0.16796875 \\ 0.16796875 \end{bmatrix}$$

Delopgave 2

Opgave 2 - Bestem Markov-kædens stationære fordeling hvis den har en.

- Ja, den har en stationær fordeling pga. $N \times N$ og ingen negative værdier.

$$(P - I)x = 0 \iff \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & -\frac{7}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{4} & \frac{1}{8} & -\frac{7}{8} & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & -\frac{7}{8} & \frac{1}{8} & 0 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & -\frac{7}{8} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{matrix} x_1 = 2x_5 \\ x_2 = 1x_5 \\ x_3 = 1x_5 \\ x_4 = 1x_5 \\ x_5 \text{ is free} \end{matrix}$$

$$x_5 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$q = \begin{bmatrix} 2/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{bmatrix}$$

- Større sandsynlighed for at befinde sig på side 1.

Delopgave 3

Opgave 1 - Find stationær fordeling for Markov-kæden beskrevet ud for den stokastiske matrix P_1

$$(P_1 - I)x = 0 \iff \begin{bmatrix} -1 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & -1 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & -1 & 0 & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$w = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, q = \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 0 \\ 0.5 \\ 0 \\ 0 \end{bmatrix}$$

Delopgave 3

Opgave 3 - Er der entydig stationær fordeling i de to tilfælde?

- Det kræves at den stokastiske matrix er **regular**
- Den er kun regular hvis alle elementer i en a $P^{(k)}$ potens kun indeholder non-zero entries.
- Ingen af de to tilfælde har en entydig stationær fordeling.

Delopgave 4

Opgave 1 - Hvorfor vil en Markov-kæde med stokastisk matrix P altid have en entydig stationær fordeling hvis $0 < \alpha < 1$?

$$P = \alpha \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \end{bmatrix} + (1-\alpha) \begin{bmatrix} \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \end{bmatrix}$$

- Når $0 < \alpha < 1$ så vil alle indgange/entries være strengt/strictly positive og dermed er P en regulær stokastisk matrix. P^1 indeholder nemlig kun positive tal hvilket ved definition i Theorem 18 derfor har en entydig stationær fordeling.

Delopgave 4

Opgave 2 - Forklar hvad et lille/sort α betyder for vores vurdering af surferens adfærd?

$$P = \alpha \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \end{bmatrix} + (1-\alpha) \begin{bmatrix} \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \end{bmatrix}$$

- Et lille α vil betyde at der er stor sandsynlighed for at brugeren blot indtaster en ny internetadresse i stedet for at følge links på siden.
- Et stort α må så betyde at der er stor sandsynlighed for at brugeren følger links på siden.

Delopgave 4

Opgave 3 - Beregning af den entydige stationære fordeling for de to eksempler i Delopgave 3. Hvordan vil jeg "ranke" internetsiderne?

$$P = \alpha P_1 + (1-\alpha) P_2$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1.7083 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -4.8769 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -2.9570 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -4.8003 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{Inlusiv de røde pile: } \mathbf{q} = \begin{bmatrix} 1.7083/17.3425 \\ 1.000/17.3425 \\ 4.8796/17.3425 \\ 2.9570/17.3425 \\ 4.8003/17.3425 \\ 1/17.3425 \\ 1/17.3425 \end{bmatrix} = \begin{bmatrix} 0.0985048 \\ 0.05766135 \\ 0.28120974 \\ 0.17050718 \\ 0.27679424 \\ 0.05766135 \\ 0.05766135 \end{bmatrix}$$

Delopgave 4

Opgave 3 - Beregning af den entydige stationære fordeling for de to eksempler i Delopgave 3. Hvordan vil jeg "ranke" internetsiderne? (fortsættelse)

- Uden de røde pile og ligeligt fordelt sandsynlighed:

$$\mathbf{q} = \begin{bmatrix} 7.1058/46.6667 \\ 4.6883/46.6667 \\ 7.9685/46.6667 \\ 13.0176/46.6667 \\ 8.1982/46.6667 \\ 4.6883/46.6667 \\ 1/46.6667 \end{bmatrix} = \begin{bmatrix} 0.15 \\ 0.10 \\ 0.17 \\ 0.28 \\ 0.18 \\ 0.10 \\ 0.02 \end{bmatrix} \tag{1}$$