MARKOVKJEDER (DEL 1) (SEC. 5.9)

Hva er en Markovkjede? En Markovkjede er en stokastisk prosess hvor et system utvikler seg probabilistisk sådan at fremtiden kun avhenger på tilstanden systemet er i lige nu men ikke på hele historikken av tilstander som den har haft før. Det lyder lidt abstrakt men egentlig er det ret naturligt og bliver klart når man studerer nogen eksempler. Vi starter med en første definisjon som vi kommer til at erstatte med en lidt nemmer definisjon senere:

Definisjon (Markovkjeder i Sannsynlighetsteori)**.** *En* Markovkjede *er beskrivet med en endlig mængde* $\mathcal{S}$ *av tilstander og sannsynligheter* $\{p(i \to j)\}_{i,j\in\mathcal{S}}$ *for at skifte en tilstand* i *til* j*. At være sannsynligheter betyder:*

(1) For alle tillstander $i, j \in \mathcal{S}$ *har vi* $p(i \to j) \geq 0$ *og* $p(i \to j) \leq 1$.

(2) For hvert tillstand $i \in \mathcal{S}$ *har vi*

$$\sum_{j\in\mathcal{S}} p(i \to j) = 1.$$

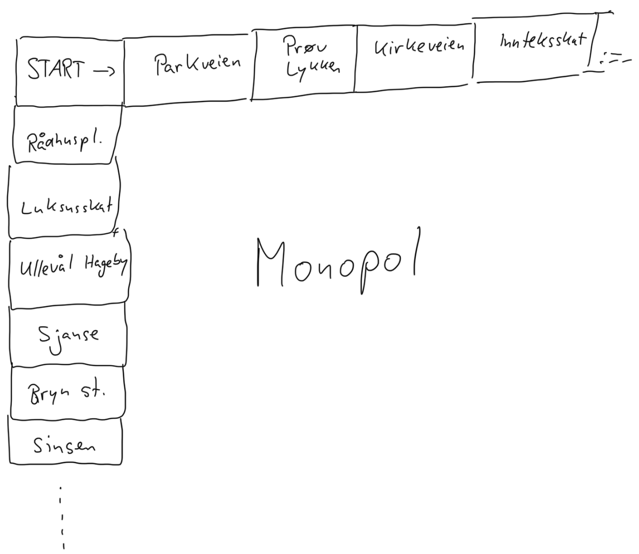
Prosessen som starter i et tillstand og utvikler sig probabilistisk med at skifte tilstand med de givende sannsynligheder kalles en Markov kjede.

Markovkjeder finnes mange steder og vi kan se nogen eksempler i kenner:

Eksempel.

(1) En eksempel er brettspiller som Monopol. Tillstander i Monopol er de forskjellige felder på spillebrett hvor figuren kan står:

$$\mathcal{S} = \Big\{\text{START}, \text{Parkveien}, \text{Prøv Lykken}, \text{Kirkeveien}, \dots\Big\}.$$



FIGUR 1. Spillebrett av Monopol (delvist).

Sannsynligheter for at skifte tillstander er sannsynligheterne for at kaste to terninger sådan at figuren bevæger sig fra den ene til den annen feld (og muligvis kombinert med nogen særlige spilmekaniker som aktionskort i Monopol). For eksempel har vi

$$p\,(\text{Parkveien} \to \text{Kirkeveien}) = \frac{1}{36},$$

og

$$p\,(\text{Parkveien} \to \text{Prinsensgate}) = \frac{1}{6}.$$

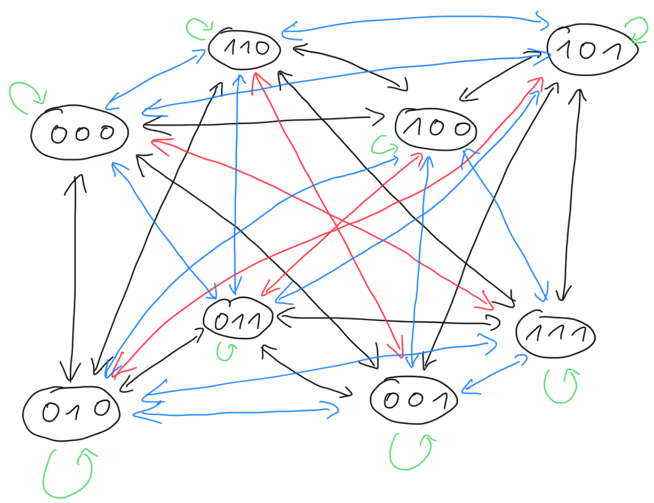
Vi kan sjekke at det kommer til at oppfylle definisjonen av en Markovkjede. Alle $p(i \to j)$ er mellom 0 og 1 (de er jo sannsynligheter) og vi har

$$\sum_{j \in \mathcal{S}} p(i \to j) = 1,$$

fordi vi har inkludert alle spillfelder i $\mathcal{S}$.

(2) Et annen eksempel er et modell for støj av et datalager i en datamaskine. La oss anta at datalageren har kun 3 bits kapasitet. De mulige tilstander er derfor

$$\mathcal{S} = \{000, 100, 010, 001, 110, 101, 011, 111\}.$$



Figur 2. Modell for støj som virker på en 3-bit datalager. Hvert transisjon går i begge retninger med den sammen sannsynlighed. Grønne kanter har sannsynlighed $(1-\lambda)^3$, hvide kanter har sannsynlighed $\lambda(1-\lambda)^2$, blåe kanter har sannsynlighed $\lambda^2(1-\lambda)$ og røde kanter har sannsynlighed λ^3.

Datalager i virkligheden er ikke perfekt. For eksempel kan der være en fysikalisk prosess som virker på datalageren (måske bliver datalageren ramt av elekromagnetistisk stråling) og hvert sekund kan et bit i datalageren vendes fra 0 til 1 eller fra 1 til 0 med en sannsynlighed λ. Denne prosess kan vi modellere med en Markovkjede! Spesifisk kan vi regne ut sannsynligheden for at skifte fra en tilstand til en annen, og vi finner for eksempel

$$\begin{aligned}
p(000 \to 001) &= \lambda(1-\lambda)^2 \\
p(000 \to 101) &= \lambda^2(1-\lambda) \\
p(110 \to 101) &= \lambda^2(1-\lambda) \\
p(101 \to 010) &= \lambda^3 \\
p(100 \to 100) &= (1-\lambda)^3.
\end{aligned}$$

Se Fig. 2 for en diagram av alle muligheder for disse sannsynligheder.

(3) For et tredje eksempel kan vi modellere årlig migrasjon av mensker, for eksempel mellom Oslo og resten av Norge. La oss anta for simplisitet at mennesker i Norge ikke migrer til andre stiller i verden, og at mensker ikke migrerer til Norge, og at der er ingen dødsfall eller fødsler i Norge (Ja, det er et meget merkligt model, men la oss tag det alligvel). I denne modell har vi to tillstander

$$\mathcal{S} = \{\text{Oslo}, \text{ikke Oslo}\}.$$

For at modellere migrasjon kan vi definere sannsynligheder for et tilfaldig person at migrere til eller fra Oslo (for eksempel med at se i sentralregister). Fiktivt kunne vi har

$$p(\text{Oslo} \to \text{ikke Oslo}) = 0.05,$$

og

$$p(\text{ikke Oslo} \to \text{Oslo}) = 0.03.$$

Hva er sannsynligheden for et person som er i Oslo at blive i Oslo? Husk at vi har

$$\sum_{j \in \mathcal{S}} p(i \to j) = 1,$$

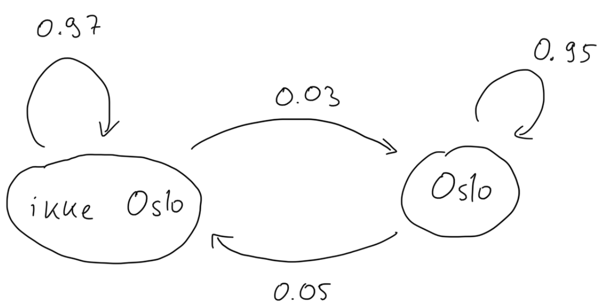
for $i = \text{Oslo}$ og $i = \text{ikke Oslo}$. Derfor finner vi

$$p(\text{Oslo} \to \text{Oslo}) = 1 - 0.05 = 0.95,$$

og

$$p(\text{ikke Oslo} \to \text{ikke Oslo}) = 1 - 0.03 = 0.97.$$

Se Fig. 3 for en diagram av alle transisjonssannsynligheder.



FIGUR 3. Modell for migrasjon med en Markovkjede.

Hvordan beskriver vi Markovkjeder med lineær algebra? Først, er det måske en god ide at organisere dataen lidt bedre med at bruke matrise-terminologi. Det er nemest at gøre hvis vi identifisere tilstander med tall sådan at vi kan snakke om den 1. tillstand, den 2. tillstand, ... Hvis der finnes n tilstander i alt kan vi derfor anta at tilstandsmængden er

$$\mathcal{S} = \{1, 2, \ldots, n\}.$$

Nu kan vi organisere sannsynligheter i en $n \times n$ matrise P med

$$P_{ij} = p(i \to j).$$

Fordi vi har konstruert matrisen P fra sannsynligheter $p(i \to j)$ er indganger P_{ij} mellom 0 og 1 og vi har at

$$\sum_{j=1}^{n} P_{ij} = 1,$$

for hvert i. Disse matriser har et navn:

Definisjon. *En $n \times n$ matrise P kalles en* stokastiske matrise *hvis*

(1) Indganger P_{ij} er mellom 0 og 1.
(2) For hvert i har vi

$$\sum_{j=1}^{n} P_{ij} = 1.$$

La oss se på migrasjonseksemplen fra før:

Eksempel. Med tallene fra den sidste eksempel finner vi den stokastiske matrise

$$P = \begin{pmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{pmatrix}.$$

Her er 'Oslo' vores første tilstand og 'ikke Oslo' den annen.

At skrive ned denne matrise gør det nemt at regne ut hvordan Markovkjeden utvikler sig i tiden. Der er forskjellige muligheder for at tænke om denne tidsevolusjon. For eksempel kan vi forstille oss at har et abstrakt startpopulasjon i hvert tillstand i $\mathcal{S}$ og vi kan bruke Markovkjeden til at regne ut hvordan populasjonen utvikler sig i tid. La oss se på migrasjonseksemplen.

Eksempel. Anta at vi har et år hvor 600000 mensker bor i Oslo og 4400000 mensker bor i resten av Norge. Hvordan utvikler sig populasjonen i den næste år. Vi kan bar regne matrise-vektor produkt og får

$$\begin{pmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{pmatrix} \begin{pmatrix} 600000 \\ 4400000 \end{pmatrix} = \begin{pmatrix} 702000 \\ 4298000 \end{pmatrix}.$$

I den næste år bor der 702000 mensker i Oslo. Observere at antall av mensker i alt har ikke ændret sig! Der var

$$600000 + 4400000 = 5000000$$

mensker i forvejen og efter et år er der stedigvæk

$$702000 + 4298000 = 5000000.$$

Grunnen til det er at matrisen P er stokastisk og vi kan sjekke at

$$\begin{aligned} y_1 + y_2 &= (0.95x_1 + 0.03x_2) + (0.05x_1 + 0.97x_2) \\ &= (0.95 + 0.05)x_1 + (0.03 + 0.97)x_2 = x_1 + x_2 \end{aligned}$$

når

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Totalt antall av mennesker bliver det sammen i vores denne model.

Et annen mulighed at tænke på Markovkjeder er med at tænke om relative andeler i stedet for absolute antall.

Eksempel. For eksempel kan vi dividere startvektoren med antal av alle mensker som bor i Oslo og resten av Norge sammen. I eksemplen er der 5000000 mensker i alt. Derfor kan vi starte med den relativt populasjonsvektor

$$\frac{1}{5000000}\begin{pmatrix} 600000 \\ 4400000 \end{pmatrix} = \begin{pmatrix} 0.12 \\ 0.88 \end{pmatrix}.$$

og den utvikler sig til

$$\begin{pmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{pmatrix}\begin{pmatrix} 0.12 \\ 0.88 \end{pmatrix} = \begin{pmatrix} 0.1404 \\ 0.8596 \end{pmatrix}$$

i den næste år. Denne vektor beskriver igen en relativ andel fordi $0.1404 + 0.8596 = 1$. Grunnen at det virker sådan er igen at matrisen P er en stokastiske matrise.

Den sidste eksempel viser at vi kan beskrive evolusjonen av en Markovkjede med vektorer som inneholder relative andeler av populasjoner i de forskjellige tilstander. En annen interpretasjon av relative andeler er sannsynligheder! I stedet for at tænke på relativt andeler av et populasjon med en fikset størrelse kan vi tænke på sannsynligheder at være i et bestemt tillstand og hvordan denne sannsynlighed utvikler sig i tiden. Hvis vi for eksempel tænker på den gennemsnitlige menneske som bor i Modell-Norge fra før så bor den med sannsynlighed 0.12 i Oslo og med sannsynlighed 0.88 ikke i Oslo. Efter et år bor den med sannsynlighed 0.1404 i Oslo og med sannsynlighed 0.8596 ikke i Oslo.

For den beskrivelse har vi bruk for en definisjon mere:

Definisjon. *En vektor* $\vec{x}$ *kalles en* sannsynlighetsvektor *hvis*

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

med indganger $x_1, \ldots, x_n$ *mellom* 0 *og* 1 *slik at*

$$x_1 + x_2 + \cdots + x_n = 1.$$

Det betyder at indgangerne av vektoren $\vec{x}$ *kan interpreteres som sannsynligheder for at være i en ut av* n *tilstander.*

Eksempler for sannsynlighetsvektorer er

$$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/4 \\ 1/4 \end{pmatrix}.$$

Eksempler for vektorer som ikke er sannsynlighetsvektorer er

$$\begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} 1/3 \\ 5 \end{pmatrix}.$$

La oss nu sige hvordan lineær algebraists definerer Markovkjeder:

Definisjon (Markovkjeder i lineær algebra)**.** *En Markovkjede er en følge av sannsynlighetsvektorer* $\vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$ *sammen med en stokastiske matrise* P *slik at*

$$\vec{x}_{k+1} = P\vec{x}_k$$

for alle $k \in \mathbb{N}$. *Vektorerne* $\vec{x}_k$ *kalles også tillstandsvektorer.*

Indgang i i tillstandsvektoren $\overrightarrow{x}_0$ beskriver sannsynligheden for at være i tillstand i i starten av tidsevolusjon (man kan godt anta en deterministisk startvektor hvor der kun finnes et enkelt 1 bland indgangerne). Efter k tidstrinn har vi tillstandsvektoren $\overrightarrow{x}_k$ og indgangen i i denne vektor beskriver sannsynligheden for at være i denne tillstand hvis vi har started Markovkjeden med startvektor $\overrightarrow{x}_0$.

Eksempel. La oss se på migrasjonseksemplen med startvektoren

$$\overrightarrow{x}_0 = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix},$$

og husk at vi har stokastiske matrisen

$$P = \begin{pmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{pmatrix}.$$

i denne eksempel. Startvektoren beskriver en situasjon hvor sannsynligheden for at bor i Oslo eller i resten av Norge er det sammen og like med $1/2$ for et tillfældig menneske. Hvis vi las Markovkjeden udvikle sig hvordan forændrer sig sannsynlighederne? Vi finner

$$\overrightarrow{x}_1 = P\overrightarrow{x}_0 = \begin{pmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{pmatrix} \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 0.49 \\ 0.51 \end{pmatrix}.$$

Vi kan regne videre og finne

$$\overrightarrow{x}_2 = P\overrightarrow{x}_1 = \begin{pmatrix} 0.4808 \\ 0.5192 \end{pmatrix}$$

$$\overrightarrow{x}_3 = P\overrightarrow{x}_2 = \begin{pmatrix} 0.472336 \\ 0.527664 \end{pmatrix}$$

$$\overrightarrow{x}_4 = P\overrightarrow{x}_3 = \begin{pmatrix} 0.4808 \\ 0.5192 \end{pmatrix}$$

$$\overrightarrow{x}_5 = P\overrightarrow{x}_4 = \begin{pmatrix} 0.4645 \\ 0.5355 \end{pmatrix}$$

$$\vdots$$

Hvordan utvikler sig systemet for meget lange tider? Hvis vi regner meget langt så finner vi at systemet når et likevekt. For stor k finner vi

$$\overrightarrow{x}_k \approx \begin{pmatrix} 0.375 \\ 0.625 \end{pmatrix},$$

og vektoren ændrer sig ikke mere, dvs. vi har

$$P \begin{pmatrix} 0.375 \\ 0.625 \end{pmatrix} = \begin{pmatrix} 0.375 \\ 0.625 \end{pmatrix}.$$

Hvis vi ser på populasjonen av 5000000 mennesker så svarer det til situasjonen hvor 1875000 mennesker bor i Oslo og 3125000 mennesker bor i resten av Norge. For disse tall har vi lige mange mennesker som flytter fra Oslo som vi har mennesker som flytter til Oslo. Det er likevekten!

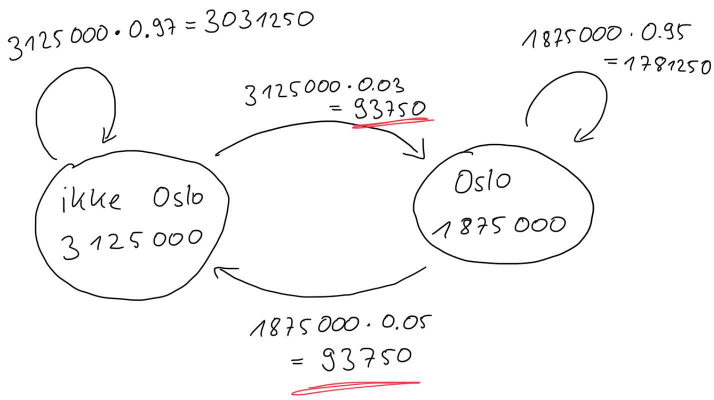
Med vores terminologi som vi har lært i kurset kan vi formalisere hva et likevektsvektor er. Det er bar en særlig egenvektor for egenverdien 1.

Definisjon. *En sannsynlighetsvektor $\overrightarrow{q}$ som oppfyller*

$$P\overrightarrow{q} = \overrightarrow{q},$$

kalles en likevektsvektor*.*

Fra eksemplen er det meget klart at der er to viktige spørgesmåler i teorien av Markovkjeder:



FIGUR 4. Likevekten i migrasjon.

(1) Finnes der altid en likevektsvektor? For at svar denne spørgesmål skal vi vise at 1 er en egenverdi av hvert stokastiske matrise P og at der finnes en egenvektor til egenverdien 1 med positive indganger.
(2) Hvis den finnes, konvergerer så tillstandsvektorer $\overrightarrow{x}_k$ altid mod likevektsvektoren?

La oss se på en eksempel for hva der kan ske med Markovkjeder:

Eksempel.

(1) La oss komme tilbake til migrasjonseksemplen. Men nu vil vi modellere migrasjon mellom Oslo, resten av Norge og Tyskland. Det betyder at vi har tillstander

$$\mathcal{S} = \{\text{resten av Norge}, \text{Oslo}, \text{Tyskland}\}.$$

La oss anta at tysker er meget mærklig og de vil ikke flytte til Norge eller andre steder hen. De vil bar blive i Tyskland hele tiden. Fordi tysker er så merklig vil Nordmenn heller ikke flytte til tyskland og vi har sannsynligheder som i Figure 5.

Den tilsvarende stokastiske matrise er

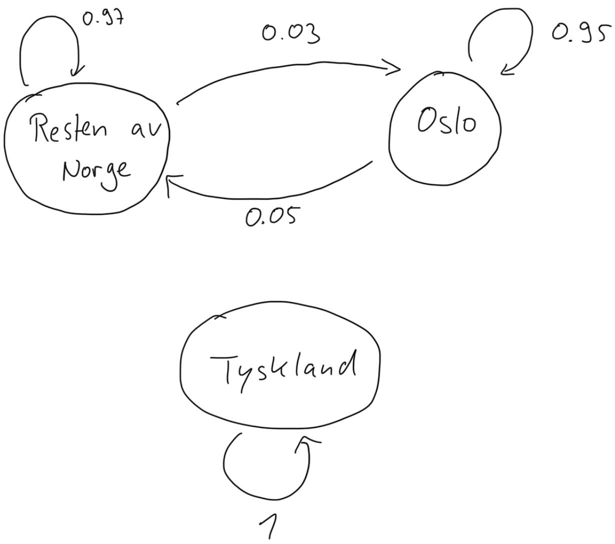
$$P = \begin{pmatrix} 0.95 & 0.03 & 0 \\ 0.05 & 0.97 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

som inneholder den stokastiske matrise fra sidste eksempel som blok. Har Markovkjeden som er definert med P et likevektsvektor? Selvføglig har det det. Hvis der ikke finnes nogen som bor i tyskland så kan vi bare bruke den gamle likevektsvektor og finne en likevektsvektor

$$\overrightarrow{q} = \begin{pmatrix} 0.375 \\ 0.625 \\ 0 \end{pmatrix}.$$

Men vi kan også anta at der bor nogen mennesker i tyskland. For eksempel kunne halvdelen av mennesker i modellen bor i tyskland. Det ville svar til likevektsvektoren

$$\overrightarrow{q} = \frac{1}{2}\begin{pmatrix} 0.375 \\ 0.625 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.1875 \\ 0.3125 \\ 0.5 \end{pmatrix}$$



FIGUR 5. Fiktivt migrasjon mellom Norge og Tyskland.

I den her eksempel konvergerer følgen $\overrightarrow{x}_{k+1} = P\overrightarrow{x}_k$ ikke for alle sannsynlighedsvektorer $\overrightarrow{x}_0$. For eksempel ville $\overrightarrow{x}_0 = \overrightarrow{q}$ give $\overrightarrow{x}_k = \overrightarrow{q}$ for alle k, men $\overrightarrow{x}_0 = \overrightarrow{q'}$ ville give $\overrightarrow{x}_k = \overrightarrow{q'}$ for alle k.