MARKOV KETENS, OF: 'WAT IS DE KANS DAT MEVROUW DE VRIES NAT ZAL WORDEN?'

KARMA DAJANI

In deze lezing gaan we over een bijzonder model in kansrekening spreken. Maar eerst een paar woorden vooraf.
Wat doen we eigenlijk in kansrekening? In het algemeen sprekend bestuderen wij onvoorspelbare ('random') verschijnselen. Bijvoorbeeld: als we met een dobbelsteen werpen, dan weten we van te voren zeker dat één van de zes zijden zal verschijnen, maar we kunnen niet met absolute zekerheid vooraf zeggen welke zijde zal verschijnen (behalve als de dobbelsteen vals is). Als de dobbelsteen eerlijk is dan is elke zijde gelijkwaardig, en wegens symmetrie heeft iedere zijde dus kans 1/6 om boven te komen. Een ander argument dat in het algemeen werkt is: werp n keer met de dobbelsteen (bijvoorbeeld $n = 1000$), en tel het aantal keer dat een bepaalde zijde (bijvoorbeeld 1) voorkomt. Noem dit aantal k_n. Dan kan men bewijzen dat k_n/n naar 1/6 zal convergeren als n naar oneindig gaat. Trouwens, als de dobbelsteen vals is, zal k_n/n naar de kans dat 1 bovenkomt convergeren als n naar oneindig gaat.

Neem aan dat we de dobbelsteen nu elke minuut één keer werpen. In de loop van de tijd zullen de uitkomsten verschillend zijn, maar in de loop van de tijd zal de kans dat 1 (of 2, of 3, 4, 5, of 6) altijd hetzelfde zijn: 1/6. We noemen deze eigenschap *stationariteit*. Merk op, dat de uitkomsten van de huidige worp, en die van eerdere worpen geen invloed hebben op wat er in de toekomst zal gebeuren. We noemen dit *onafhankelijkheid*. Wiskundig gezegd, als

$$X_n = \text{ uitkomst van de } n\text{de worp}$$

en als we met de notatie $P(X_{n+1} = j|X_n = i)$ bedoelen: de kans dat we op 'tijdstip' $n + 1$ de uitkomst j hebben, *gegeven* dat we op tijdstip n als uitkomst i hadden, dan is onafhankelijkheid equivalent met

$$P(X_{n+1} = j|X_n = i) = P(X_{n+1} = j) = \frac{1}{6}.$$

We noemen $X_1, X_2, \ldots$ een onafhankelijk stochastisch process.

Er zijn in de praktijk heel veel stochastische processen die niet onafhankelijk zijn. Het eenvoudigste afhankelijke model is dat van de zogenaamde *Markov ketens*. Ruwweg gesproken hangt bij een Markov keten de toekomst alleen van het heden en niet van het verleden af. Of in wiskundige formulering:

$$P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1) = P(X_{n+1} = j|X_n = i).$$

We noemen de Markov keten *homogeen* als

$$P(X_{n+m} = j|X_n = i) = P(X_{m+1} = j|X_1 = i).$$

Aan de hand van simpele voorbeelden wil ik nu een aantal eigenschappen van Markov ketens bestuderen.

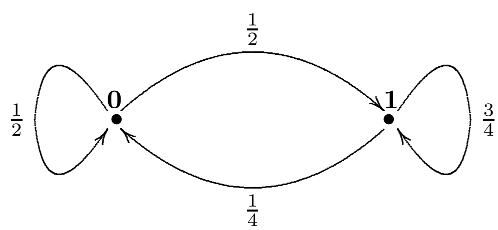
Voorbeeld 1. Zeg dat een machine in twee mogelijke toestanden kan zijn: 'aan' en 'uit'. Op elk uur kijken we of de machine aan of uit is. Laat

$$X_n = \begin{cases} 1 & \text{als de machine op tijdstip } n \text{ aan is,} \\ 0 & \text{als de machine op tijdstip } n \text{ uit is.} \end{cases}$$

We modeleren het stochastische proces $X_1, X_2, \dots$ als een Markov keten. Zeg dat de volgende kansen bekend zijn:

$$\begin{aligned} p_{00}(1) &:= P(X_{n+1} = 0 | X_n = 0) = \frac{1}{2} \\ p_{01}(1) &:= P(X_{n+1} = 1 | X_n = 0) = \frac{1}{2} \\ p_{10}(1) &:= P(X_{n+1} = 0 | X_n = 1) = \frac{1}{4} \\ p_{11}(1) &:= P(X_{n+1} = 1 | X_n = 1) = \frac{3}{4} \end{aligned}$$

We kunnen deze informatie in een graaf weergeven:



We kunnen ons nu afvragen wat de fractie van de tijd is dat de machine aan zal staan? Om deze vraag te kunnen beantwoorden gaan we eerst de informatie in de graaf op een andere manier—met behulp van een matrix—interpreteren. Definieer

$$P = P_1 = \begin{pmatrix} p_{00}(1) = \frac{1}{2} & p_{01}(1) = \frac{1}{2} \\ \\ p_{10}(1) = \frac{1}{4} & p_{11}(1) = \frac{3}{4} \end{pmatrix}$$

We noemen P de *eenstapovergangs matrix*. Merk op dat alle rijen van P tot 1 optellen (waarom?). Laten we de *tweestaps overgangskansen* berekenen:

$$p_{ij}(2) := P(x_{n+2} = j | X_n = i).$$

In een matrix:

$$P_2 = \begin{pmatrix} p_{00}(2) = \frac{3}{8} & p_{01}(2) = \frac{5}{8} \\ \\ p_{10}(2) = \frac{5}{16} & p_{11}(2) = \frac{11}{16} \end{pmatrix}.$$

Het is eenvoudig om te zien dat $P_2 = P^2$, waar $P^2 = P \cdot P$, en waar de vermenigvuldiging matrix vermenigvuldiging is. In het algemeen geldt er dat de $n$de stapovergangsmatrix P_n gelijk is aan:

$$P_n = \begin{pmatrix} p_{00}(n) & p_{01}(n) \\ \\ p_{10}(n) & p_{11}(n) \end{pmatrix} = P^n.$$

Met behulp van `Mathematica` (en later in het vak *stochastische processen*) kunnen we zien dat

$$\lim_{n\to\infty} P_n = \lim_{n\to\infty} P^n = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

We noemen de kansvector

$$v = \left(\frac{1}{3}, \frac{2}{3}\right)$$

de *stationaire verdeling*, en we interpreteren de coordinaten alsvolgt:

$$\frac{1}{3} = \text{de fractie van de tijd dat de machine uit is}$$

$$\frac{2}{3} = \text{de fractie van de tijd dat de machine aan is.}$$

Deze stationaire verdeling heeft als eigenschap dat

$$vP^n = v, \quad \text{voor elke } n.$$

Voorbeeld 2. Mevrouw de Vries, werkzaam bij de Nederlandse Bank, bezit 3 paraplu's. Elke ochtend en middag wandelt ze van huis naar de bank, respektijvelijk van de bank naar huis. Indien het op het moment van vertrek regent dan neemt zij een paraplu (als er een is) anders niet. Veronderstel dat op elk vertrekstip, en onafhankelijk van alle andere vertrekstrippen, de kans op regen 1/3 is.

Vraag: bepaal de fractie van de tijd dat Mw. de Vries in de ochtend een nat pak oploopt = P(nat pak in de ochtend)?

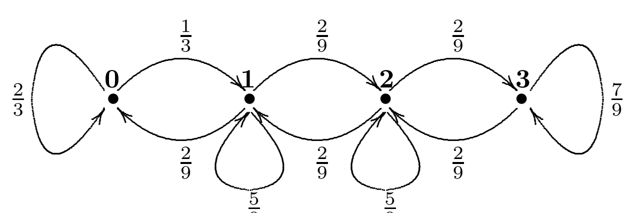
Merk eerst op dat,

$$\begin{aligned} P(\text{nat pak in de ochtend}) &= P(\text{regen en geen paraplu thuis in de ochtend}) \\ &= P(\text{regent}) \cdot P(\text{geen paraplu thuis in de ochtend}) \\ &= \frac{1}{3} \cdot P(\text{geen paraplu thuis in de ochtend}). \end{aligned}$$

We zijn dus geintereseerd in het aantal paraplu's dat bij ieder vertrek s'ochtends thuis aanwezig is, en we modeleren dit als een Markov keten met

$$X_n = \text{het aantal paraplu's thuis op de } n\text{de ochtend}$$

Als wij de overgangskansen uitrekenen vinden we de volgende graaf:



De corresponderende overgangs matrix $P_1 = P$ is dan

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ \frac{2}{9} & \frac{5}{9} & \frac{2}{9} & 0 \\ 0 & \frac{2}{9} & \frac{5}{9} & \frac{2}{9} \\ 0 & 0 & \frac{2}{9} & \frac{7}{9} \end{pmatrix}.$$

Met wat theorie (of voorlopig met `Mathematica`) zie je dat

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{2}{11} & \frac{3}{11} & \frac{3}{11} & \frac{3}{11} \\ \frac{2}{11} & \frac{3}{11} & \frac{3}{11} & \frac{3}{11} \\ \frac{2}{11} & \frac{3}{11} & \frac{3}{11} & \frac{3}{11} \\ \frac{2}{11} & \frac{3}{11} & \frac{3}{11} & \frac{3}{11} \end{pmatrix}.$$

We zien dus dat de stationaire verdeling gegeven wordt door

$$v = \left(\frac{2}{11}, \frac{3}{11}, \frac{3}{11}, \frac{3}{11}\right).$$

Merk op dat $vP = v$. De $i$de coordinaat van v geeft ons de fractie van de tijd dat we 's ochtends i paraplu's hebben. De kans op een nat pak in de morgen is dus

$$\frac{1}{3} \cdot \frac{2}{11} = \frac{2}{33},$$

oftewel, ongeveer 2 keer per maand! (Mw. de Vries is een workaholic! Ze gaat iedere dag naar het werk—sneeuw, regen, of zoiets triviaals als een week-end doen er niet toe!).

In het algemeen, gegeven een $n \times n$ overgangsmatrix P, een stationaire verdeling is een vector $v = (v_1, v_2, \ldots, v_n)$ met de eigenschappen: (i) $\sum_{i=1}^{n} v_i = 1$, en (ii) $vP = v$. We interpreteren de $i$de coordinaat v_i als de fractie of frequentie van de tijd dat de Markovketen in toestand i is. In de bovenstaande voorbeelden, de markovketen is *aperiodiek* d.w.z. er is een n waarvoor $p_{ij}(n) > 0$ for alle toestanden i, j. In dit geval is de stationaire verdeling v uniek, en

$$\lim_{n\to\infty} P^n = \begin{pmatrix} v \\ \vdots \\ v \end{pmatrix}.$$

Voorbeeld 3. In dit voorbeeld laten we zien dat een stationaire verdeling v bestaat (d.w.z. $vP = v$ en $\sum_i v_i = 1$) maar $\lim_{n\to\infty} P^n$ bestaat niet. Beshouw de markovketen met overgangsmatrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Het is makkelijk te zien dat

$$P^{3n+1} = P,$$

$$P^{3n+2} = P^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

en

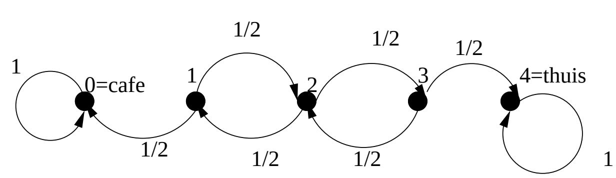
$$P^{3n} = P^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Er volgt dus dat $\lim_{n\to\infty} P^n$ bestaat niet. Aan de andere kant de stelsel $vP = v$ en $v_1 + v_2 + v_3 = 1$ heeft een unieke oplossing $v = (1/3, 1/3, 1/3)$ zoals verwacht. Merk op dat de bovenstaande matrix is *irreducibele* d.w.z. voor ieder i, j er bestaat een n zodanig dat $p_{ij}(n) > 0$. Verder,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P^i = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

Voorbeeld 4. (Absorberend Markovketens): Hier hebben we te maken met een andere soort markovketens, namelijk *absorberende markovketens*:
(1) er bestaat tenminste een toestand i die absorberend is, d.w.z. $p_{ii} = 1$,
(2) vanuit iedere toestand is het mogelijk om naar een absorberende toestand te gaan in een of meer stappen.
In dit geval kan men bewijzen dat vanuit elk beginpunt, met kans 1 de markovketen in een absorberende toestand terecht zal komen.

Beschouw een dronken wandelaar die op de punten $0, 1, 2, 3, 4$ loopt. Stel dat $0 =$cafe en $4 =$thuis. Hij begint op een willekeurige punt en elke seconde gaat hij met kans $1/2$ of een stap naar rechts of een stap naar links totdat hij in toestand 0 of 4 is, dan stopt hij.



In de overgangsmatrix, schrijven we eerst de absorberende toestanden en dan de niet absorberende toestanden, d.w.z. we gebruiken de volgorde $0, 4, 1, 2, 3$. We krijgen

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Men kan bewijzen dat

$$\lim_{n\to\infty} P^n = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ r_{10} = \frac{3}{4} & r_{14} = \frac{1}{4} & 0 & 0 & 0 \\ r_{20} = \frac{1}{2} & r_{24} = \frac{1}{2} & 0 & 0 & 0 \\ r_{30} = \frac{1}{4} & r_{34} = \frac{3}{4} & 0 & 0 & 0. \end{pmatrix}.$$

We interpreteren bijvoorbeeld r_{10} als de kans dat de Markovketen geabsorbeerd is in toestand 0, gegeven dat het beginpunt 1 is. We zien dus dat als het beginpunt 1 is dan met kans 3/4 zal de markovketen geabsorbeerd worden in toestand 0, en met kans 1/4 zal de markovketen gebasorbeerd worden in toestand 4. Merk op dat, hoe dichter een niet absorberende toestand i in de beurt is van een absorberende toestand j, hoe groter de kans is om vanuit i in j geabsorbeerd te worden.

Opgave: Elke ochtend gaat uw docent een stukje hardlopen in de omgeving van haar huis. Het is even waarschijnlijk dat zij haar huis via de voordeur of via de achterdeur verlaat. Bij het verlaten van het huis trekt zij een paar hardloopschoenen aan (indien aanwezig, anders gaat ze blootsvoets de deur uit). Bij terugkeer laat ze haar schoenen achter bij de voordeur of achterdeur, al naar gelang ze binnenkomt (waarbij voordeur of achterdeur opnieuw willekeurig gekozen worden). Neem aan dat uw docent 3 paar schoenen heeft, en laat X_n het aantal paar schoenen dat is aan het begin van de $n$de dag bij de deur waaruit zij vertrekt.

(a) Bepaal de overgangsmatrix van de Markovketen $\{X_n : n \geq 0\}$.
(b) Met welke frequentie zal uw docent blootsvoets rondhollen?