Markovkæder og samfundsforskning

Af Ernst Lykke Jensen[1])

Inden for sandsynlighedsregningen har en væsentlig del af bestræbelserne været rettet mod udforskningen af egenskaberne ved hændelsesforløb, i hvilke hændelserne er indbyrdes uafhængige. Som et simpelt eksempel kan man tænke sig et forsøg, der kun kan give to udfald, A_1 og A_2, der gensidigt udelukker hinanden, f. eks. krone og plat ved kast med en mønt. Ved n gentagelser af forsøget realiseres et eller andet hændelsesforløb, f. eks. følgende for $n=5$

$$H=A_1A_2A_2A_1A_2 \ . \tag{1}$$

Sandsynligheden $S(H)$ for netop dette hændelsesforløb er

$$S(H)=S_1(A_1)\cdot S_2(A_2|A_1)\cdot S_3(A_2|A_1A_2)\cdot S_4(A_1|A_1A_2A_2)\cdot S_5(A_2|A_1A_2A_2A_1),\tag{2}$$

hvor $S_1(A_1)$ er sandsynligheden for hændelsen A_1 i første forsøg, og hvor f. eks. $S_3(A_2|A_1A_2)$ er sandsynligheden for forekomst i tredie forsøg af hændelsen A_2 betinget af, at A_1 forekom i første og A_2 i andet forsøg. Opstiller man nu den

Forudsætning 1, at udfaldet af et vilkårligt forsøg er uafhængig af udfaldene i de foregående forsøg,

forenkles formel (2) til

$$S(H)=S_1(A_1)\cdot S_2(A_2)\cdot S_3(A_2)\cdot S_4(A_1)\cdot S_5(A_2) \quad . \tag{3}$$

Forudsætningen indebærer nemlig, at f. eks. A_2 i tredie forsøg forekommer lige hyppigt, hvadenten A_2 efterfølger A_1A_1, A_1A_2, A_2A_1 eller A_2A_2, således at de betingede sandsynligheder $S_3(A_2|A_1A_1)$, $S_3(A_2|A_1A_2)$,

[1]) cand. polit., Københavns Universitets Statistiske Institut.

Markovkæder og samfundsforskning

Af ERNST LYKKE JENSEN[1])

Inden for sandsynlighedsregningen har en væsentlig del af bestræbelserne været rettet mod udforskningen af egenskaberne ved hændelsesforløb, i hvilke hændelserne er indbyrdes uafhængige. Som et simpelt eksempel kan man tænke sig et forsøg, der kun kan give to udfald, A_1 og A_2, der gensidigt udelukker hinanden, f. eks. krone og plat ved kast med en mønt. Ved n gentagelser af forsøget realiseres et eller andet hændelsesforløb, f. eks. følgende for $n=5$

$$H = A_1A_2A_2A_1A_2 \ . \tag{1}$$

Sandsynligheden S(H) for netop dette hændelsesforløb er

$$S(H) = S_1(A_1) \cdot S_2(A_2|A_1) \cdot S_3(A_2|A_1A_2) \cdot S_4(A_1|A_1A_2A_2) \cdot S_5(A_2|A_1A_2A_2A_1), \tag{2}$$

hvor $S_1(A_1)$ er sandsynligheden for hændelsen A_1 i første forsøg, og hvor f. eks. $S_3(A_2|A_1A_2)$ er sandsynligheden for forekomst i tredie forsøg af hændelsen A_2 betinget af, at A_1 forekom i første og A_2 i andet forsøg. Opstiller man nu den

Forudsætning 1, at udfaldet af et vilkårligt forsøg er uafhængig af udfaldene i de foregående forsøg,

forenkles formel (2) til

$$S(H) = S_1(A_1) \cdot S_2(A_2) \cdot S_3(A_2) \cdot S_4(A_1) \cdot S_5(A_2) \quad . \tag{3}$$

Forudsætningen indebærer nemlig, at f. eks. A_2 i tredie forsøg forekommer lige hyppigt, hvadenten A_2 efterfølger A_1A_1, A_1A_2, A_2A_1 eller A_2A_2, således at de betingede sandsynligheder $S_3(A_2|A_1A_1)$, $S_3(A_2|A_1A_2)$,

[1]) cand. polit., Københavns Universitets Statistiske Institut.

$S_3(A_2|A_2A_1)$ og $S_3(A_2|A_2A_2)$ er lige store og altså netop lig med sandsynligheden $S_3(A_2)$ for, at A_2 indtræffer i tredie forsøg.

I formel (3) har man ikke udelukket den mulighed, at sandsynligheden for A_1 henholdsvis A_2 kan variere fra forsøg til forsøg. Gør man imidlertid følgende

Forudsætning 2, at sandsynligheden for et bestemt udfald af forsøget er konstant,

forenkles formel (3) til følgende udtryk

$$S(H)=S(A_1)\cdot S(A_2)\cdot S(A_2)\cdot S(A_1)\cdot S(A_2) \quad . \tag{4}$$

Idet udfaldene A_1=krone og A_2=plat ved møntkast antages at tilfredsstille de opstillede forudsætninger, er $S(H)=\frac{1}{32}$ sandsynligheden for rækkefølgen krone, plat, plat, krone, plat ved fem kast med en „ægte" mønt, d.v.s. en mønt for hvilken $S(A_1)=S(A_2)=\frac{1}{2}$.

De foranstående bemærkninger tjener til at vise vej mod en generalisation af den model, der hviler på forudsætningerne 1 og 2. Det er nærliggende som et første skridt på vejen at lade udfaldet af et bestemt forsøg afhænge af udfaldet i det umiddelbart foregående forsøg, men derimod ikke af det tidligere hændelsesforløb. Dermed erstattes forudsætning 1 af

Forudsætning 1'. Udfaldet af et vilkårligt forsøg er afhængig af, og kun af, udfaldet i det nærmest foregående forsøg.

Hændelsesforløbet H i (1) har da, jfr. (2), sandsynligheden

$$S(H)=S_1(A_1)\cdot S_2(A_2|A_1)\cdot S_3(A_2|A_2)\cdot S_4(A_1|A_2)\cdot S_5(A_2|A_1) \quad . \tag{5}$$

Om de i (5) indgående betingede sandsynligheder gør vi nu

Forudsætning 2'. Sandsynligheden for et bestemt udfald i et forsøg betinget af et bestemt udfald i det foregående forsøg er konstant.

Dette betyder, at $S_2(A_2|A_1)$ og $S_5(A_2|A_1)$ er lige store, og vi kan derfor skrive (5) på formen

$$S(H)=S_1(A_1)\cdot S(A_2|A_1)\cdot S(A_2|A_2)\cdot S(A_1|A_2)\cdot S(A_2|A_1) \quad . \tag{6}$$

En følge af forsøg, for hvilke sandsynligheden for ethvert endeligt hændelsesforløb af typen (1) kan beregnes på den i (5) angivne måde, kaldes en Markovkæde [2]). Er forudsætning 2' opfyldt, gælder (6), og Markovkæden siges at have konstante overgangssandsynligheder.

Som sagt, de betingede sandsynligheder, der indgår i (6), kaldes overgangssandsynligheder; f. eks. er $S(A_2|A_1)$, som herefter betegnes p_{12}, sandsynligheden for overgang fra udfaldet A_1 i et vilkårligt forsøg til udfaldet A_2 i det efterfølgende forsøg. Da vi betragter et forsøg med to udfald, må der specificeres fire overgangssandsynligheder p_{11}, p_{12}, p_{21}, p_{22}, som det viser sig at være bekvemt at ordne i rækker og søjler i følgende kvadratformede, såkaldt stokastiske, matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \quad . \tag{7}$$

Det er klart, at rækkesummerne i P er lig med 1; f. eks. er $p_{11}+p_{12}=1$, fordi det er sikkert, at A_1 i et forsøg enten efterfølges af A_1 eller A_2 i det næste forsøg.

Af hensyn til anvendelserne indføres nu en anden terminologi. I stedet for at sige, at et givet forsøg resulterede i udfaldet A_2, vil vi tale om, at systemet på det pågældende tidspunkt befinder sig i tilstanden A_2. Vi vil endvidere nummerere tidspunkterne fra 0 og opefter, således at hændelsesforløbet (1) angiver, at systemet er i tilstanden A_1 på tidspunktet $t = 0$, i tilstanden A_2 på tidspunkt $t = 1$ o.s.v. Den i (1) angivne udvikling af systemet fra $t = 0$ til $t = 4$ kan da tilknyttes sandsynligheden

$$S(H) = u_1 \cdot p_{12} \cdot p_{22} \cdot p_{21} \cdot p_{12} \quad , \tag{8}$$

hvor u_1 er sandsynligheden for, at systemet til tiden $t = 0$ befinder sig i tilstanden A_1. Da vi betragter et system med kun to tilstande, angiver talparret (u_1, u_2) systemets sandsynlighedsfordeling på udgangstidspunktet, d.v.s. $u_1+u_2=1$. Den kaldes for udgangs- eller initialfordelingen.

Det står nu klart, at med en given udgangsfordeling og en given stokastisk matrix P kan systemets sandsynlighedsfordeling beregnes for alle værdier af t. Med andre ord, systemets fremtidige udvikling kan forudsiges i sandsynlighedsmæssig forstand. Speciel interesse knytter der sig til fordelingens opførsel for store værdier af t; grænsefordelingen, d.v.s. fordelingen for $t \to \infty$, kaldes stationær, hvis dens udseende ikke beror på t.

Det siger sig selv, at modellen uden videre generaliseres til et vilkår-

[2]) Efter den russiske sandsynlighedsteoretiker A. A. Markov (1856–1922). En indgående behandling af teorien findes i W. Feller: An Introduction to Probability Theory and its Applications, Vol. I, Second Edition. Wiley, New York, 1957. Kapitlerne 15 og 16.

ligt endeligt[3]) antal tilstande $A_1, A_2, \cdots, A_k$. Med konstante overgangssandsynligheder kan systemets fremtidige udvikling forudsiges ved hjælp af udgangsfordelingen

$$u_1, u_2, \cdots, u_k \tag{9}$$

og den stokastiske matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}, \tag{10}$$

i hvilken diagonalelementet p_{jj} angiver sandsynligheden for, at systemet forbliver i tilstanden A_j fra et tidspunkt til det næste, medens p_{ij} er sandsynligheden for overgang fra tilstanden A_i til tilstanden A_j. Som tidligere nævnt er rækkesummerne i P lig med 1, d.v.s.

$$p_{j1} + p_{j2} + \cdots + p_{jj} + \cdots + \cdots + p_{jk} + = 1. \tag{11}$$

Lad os på dette sted standse et øjeblik og se på de eventuelle anvendelsesmuligheder for den betragtede model. Det bedste udgangspunkt for en sådan drøftelse får man, når man erkender modellens fundamentale kendetegn: Systemets fremtidige udvikling er betinget udelukkende af dets nutidige tilstand, hvorimod den historiske udvikling, der har realiseret denne tilstand, er uden betydning. Dette forhold er naturligvis i flere forbindelser en svær hindring for anvendelse af Markov-modellen inden for økonomisk og social forskning. På den anden side er den dynamiske betragtningsmåde modellens fornemste kendemærke, og dermed er der i hvert fald knyttet et stærkt bånd mellem adskillig samfundsforskning på den ene side og sandsynlighedsregning på den anden side. Modellen er med sit dynamiske oplæg et skridt i en rigtig retning. På de områder, hvor modellen er åbenbart urimelig, har man i det generelle oplæg i indledningen, jfr. (2), i princippet mulighed for at ændre den på en sådan måde, at den bliver mere virkelighedsbetonet. Noget andet er så, om modellen i matematisk henseende bliver håndterlig. Hvad angår den beregningstekniske side af sagen kommer man vel ikke i nød i elektronregnemaskinens tidsalder.

På denne baggrund skal der i det følgende nævnes en række områder for mulig anvendelse af det fremførte begrebsapparat; den faktiske an-

[3]) Teorien for uendelige Markovkæder skal ikke her gøres til genstand for omtale.

vendelse af det forbliver i hvert enkelt tilfælde, i større eller mindre grad, et „quaestio facti".

1. Ved *opinionsundersøgelser* [4]) er det af interesse at kende ikke alene befolkningens indstilling på et bestemt tidspunkt, men også at kunne foretage ekstrapolationer vedrørende udviklingen i denne indstilling. Det er naturligvis en mulighed at foretage en undersøgelse på to eller flere hinanden efterfølgende tidspunkter, hvorefter man på grundlag af de indsamlede oplysninger skønsmæssigt vurderer den fremtidige udvikling. Markov-modellen har imidlertid den fordel, at den objektivt giver anvisning på, hvorledes ekstrapolationen skal udføres.

   Det kan f. eks. dreje sig om en valgforudsigelse, således at systemets tilstande er de partier, der kan stemmes på; elementet p_{ij} i den stokastiske matrix er sandsynligheden for, at et individ på ét tidspunkt vil afgive sin stemme på partiet A_j, når det på det umiddelbart foregående tidspunkt havde til hensigt at stemme på A_i. Det er klart, at man ved analysen nødsages til at tage hensyn til en række forhold; der rejser sig således spørgsmålet om inddeling af vælgerne efter køn, alder, erhverv o.s.v., ligesom der må tages hensyn til, at vælgerskaren består af en hård kerne og en mere eller mindre flydende vælgermasse.

2. Som et andet muligt anvendelsesområde kan man nævne de såkaldte *konjunkturtest* [5]), hvor man klassificerer de adspurgte virksomheder efter forskellige kendetegn, f. eks. lagerstørrelse, i f. eks. tre grupper (tilstande): fremgang, stilstand og tilbagegang.

3. I hvert fald som et udgangspunkt kunne man forsøge at angribe problemet om *arbejdskraftens bevægelser* [6]) sandsynlighedsteoretisk efter de her beskrevne retningslinier. Drejer det sig f. eks. om den faglige bevægelighed, repræsenterer fag eller industrigrupper de forskellige tilstande i modellen. Også her vil man naturligvis stå overfor vanskelige spørgsmål af samme karakter som nævnt foran i forbindelse med opinionsundersøgelserne, hvortil kommer, at mobiliteten er afhængig af situationen på boligmarkedet, sæsonvariationer i beskæftigelsen, det økonomiske udviklingstrin og tempo i forskellige egne

[4]) Se eksempel 2, p. 275.

[5]) Helge Munksgaard: Konjunkturtesten. – Et moderne bidrag til bedømmelse af afsætningssituationen. Nationaløkonomisk Tidsskrift, 1958, 96. bind, 1.–2. hefte.

[6]) F. Blumen, M. Kogan, P. J. McCarthy: The Industrial Mobility of Labor as a Probability Process. Cornell Studies in Industrial and Labor Relations, Vol. 6, 1955.

af landet og konjunktursituationen i det hele taget; også hele spørgsmålet om uddannelsen kommer ind i billedet som en vigtig faktor.

4. En *markedsundersøgelse* har f. eks. til formål at belyse en virksomheds markedsandel for en vare. Ved hjælp af en repræsentativ undersøgelse kan markedsandelen fastslås på undersøgelsestidspunktet [7]); det kan imidlertid også have interesse at forudsige udviklingen i markedsandelen; en Markovkæde med to tilstande kan eventuelt benyttes hertil. En husstand er i tilstanden A_1, hvis den forbruger pågældende varemærke, ellers i A_2, og markedsandelen er antallet af husstande, der forbruger varemærket i procent af antal husstande, der forbruger varearten.

5. Hidtil har vi karakteriseret de forskellige tilstande kvalitativt. Man kan imidlertid knytte en talværdi, x_i, til hver tilstand, A_i. Derved fremkommer den særlige type af stokastiske processer, som kaldes Markovprocesser. Herved forstås altså en følge af stokastiske variable X_0, X_1, $X_2, \cdots, X_t, \cdots$, hvor den variable til tiden t, X_t, kan antage de mulige værdier $x_1, x_2, \cdots, x_k$, med den egenskab, at f. eks.

$$S(X_0 = x_i,\ X_1 = x_j,\ X_2 = x_m,\ X_3 = x_k)$$

$$= S(X_0 = x_i) \cdot S(X_1 = x_j | X_0 = x_i) \cdot S(X_2 = x_m | X_1 = x_j) \cdot S(X_3 = x_k | X_2 = x_m),$$

jfr. (6). Den variabel, hvis tidslige udvikling betragtes, kan f. eks. være prisen på en vare, virksomhedernes lagerhold o.s.v. Det vil ses, at denne betragtning åbner mulighed for anvendelse af teorien for Markovprocesser inden for dele af økonomisk teori, f. eks. konjunkturforskningens periodeanalyse og pristeorien [8]).

Efter denne omtale af nogle mulige anvendelser af Markovmodellen vil vi forsøge på simpel måde at redegøre for den teknik, der muliggør beregning af systemets sandsynlighedsfordeling på et vilkårligt fremtidigt tidspunkt på grundlag af udgangsfordelingen (9) og den stokastiske matrix (10). Af hensyn til overskueligheden indskrænker vi os til at betragte en Markovkæde med tre tilstande. Idet altså fordelingen på tidspunkt $t = 0$ er

$$u_1,\ u_2,\ u_3 \quad , \tag{12}$$

[7]) Finn Madsen: Nye metoder i markeds analysen. Nordisk Tidsskrift for industriel Statistik, Bind 3, Hæfte 2, 1958.

[8]) F. Zeuthen: Economic Theory and Method, 1955. Kapitel 35 og 36.

hvor $u_1 + u_2 + u_3 = 1$, og de konstante overgangssandsynligheder er givet ved den stokastiske matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} , \tag{13}$$

ønsker vi at beregne systemets sandsynlighedsfordeling for $t = 1, 2, 3, \cdots$. Betragt f. eks. tilfældet $t = 2$. Lad det være givet, at systemet på tidspunktet $t = 0$ befinder sig i tilstanden A_3; under denne forudsætning er sandsynligheden for, at det på tidspunktet $t = 2$ befinder sig i tilstanden A_2, lig med

$$p_{32}^{(2)} = p_{31}p_{12} + p_{32}p_{22} + p_{33}p_{32} \quad , \tag{14}$$

idet f. eks. $p_{13}p_{12}$ er sandsynligheden for, at systemet fra $t = 0$ til $t = 1$ bevæger sig fra tilstanden A_3 til tilstanden A_1 og dernæst fra $t = 1$ til $t = 2$ overgår fra A_1 til A_2. Det bemærkes, at $p_{32}^{(2)}$ er lig med produktsummen af elementerne i tredie række og anden søjle i den stokastiske matrix (13). Beregnes på samme måde som angivet ved (14) de øvrige sandsynligheder, kan disse opstilles i følgende matrix

$$P^2 = P \cdot P = \begin{pmatrix} p_{11}^{(2)} & p_{12}^{(2)} & p_{13}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} & p_{23}^{(2)} \\ p_{31}^{(2)} & p_{32}^{(2)} & p_{33}^{(2)} \end{pmatrix} . \tag{15}$$

Ved række-søjle multiplikation af (13) med (15) fås dernæst

$$P^3 = P \cdot P^2 = \begin{pmatrix} p_{11}^{(3)} & p_{12}^{(3)} & p_{13}^{(3)} \\ p_{21}^{(3)} & p_{22}^{(3)} & p_{23}^{(3)} \\ p_{31}^{(3)} & p_{32}^{(3)} & p_{33}^{(3)} \end{pmatrix} , \tag{16}$$

hvor f. eks.

$$p_{21}^{(3)} = p_{21} \cdot p_{11}^{(2)} + p_{22}p_{21}^{(2)} + p_{23}p_{31}^{(2)} \tag{17}$$

er sandsynligheden for, at systemet til tidspunkt $t = 3$ befinder sig i tilstanden A_1, når det til tidspunkt $t = 0$ befandt sig i tilstanden A_2. Man kan nu fortsætte på den angivne måde og har generelt

$$P^t = P \cdot P^{t-1} = \begin{pmatrix} p_{11}^{(t)} & p_{12}^{(t)} & p_{13}^{(t)} \\ p_{21}^{(t)} & p_{22}^{(t)} & p_{23}^{(t)} \\ p_{31}^{(t)} & p_{32}^{(t)} & p_{33}^{(t)} \end{pmatrix} \tag{18}$$

Vi er nu ved hjælp af (12) og (18) i stand til at finde de søgte sandsynligheder $u_1^{(t)}$, $u_2^{(t)}$ og $u_3^{(t)}$ for, at systemet befinder sig i de tre tilstande på et vilkårligt tidspunkt t:

$$\begin{aligned} u_1^{(t)} &= u_1 p_{11}^{(t)} + u_2 p_{21}^{(t)} + u_3 p_{31}^{(t)} \\ u_2^{(t)} &= u_1 p_{12}^{(t)} + u_2 p_{22}^{(t)} + u_3 p_{32}^{(t)} \\ u_3^{(t)} &= u_1 p_{13}^{(t)} + u_2 p_{23}^{(t)} + u_3 p_{33}^{(t)} \end{aligned} \quad . \tag{19}$$

Sandsynligheden fremkommer altså ved multiplikation af udgangssandsynlighederne (12) med søjleelementerne i (18). (Læsere, der er bekendt med matrixregning, vil kunne udtrykke det foregående mere kortfattet og elegant ved indførelse af matrixnotation og anvendelse af reglerne for regning med matricer).

Hvis p^t for voksende t konvergerer mod en matrix med ens rækker, d.v.s.

$$\begin{pmatrix} p_{11}^{(t)} & p_{12}^{(t)} & p_{13}^{(t)} \\ p_{21}^{(t)} & p_{22}^{(t)} & p_{23}^{(t)} \\ p_{31}^{(t)} & p_{32}^{(t)} & p_{33}^{(t)} \end{pmatrix} \rightarrow \begin{pmatrix} p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \end{pmatrix} \quad , \tag{20}$$

ses det af (19), da $u_1 + u_2 + u_3 = 1$, at systemets sandsynlighedsfordeling konvergerer mod en stationær fordeling:

$$\begin{aligned} u_1^{(t)} &\rightarrow p_1 \\ u_2^{(t)} &\rightarrow p_2 \\ u_3^{(t)} &\rightarrow p_3 \end{aligned} \quad . \tag{21}$$

Grænsefordelingen er altså uafhængig af udgangsfordelingen [9]).

Vi vil herefter gennemgå et par eksempler, der kan tjene til at belyse teorien.

Eksempel 1.

Bl. a. på baggrund af hvad der er omtalt under punkt 4 p. 16 vedrørende markedsundersøgelser, har det en vis interesse at betragte en Markovkæde med to tilstande med følgende overgangssandsynligheder

$$P = \begin{pmatrix} 1-\beta & \beta \\ \alpha & 1-\alpha \end{pmatrix} \quad .$$

[9]) Se herom Feller, op. cit. p. 356 ff. eller B. W. Gnedenko: Lehrbuch der Wahrscheinlichkeitsrechnung. Akademie-Verlag. Berlin, 1957, p. 105.

Man kan tænke sig, at man på ethvert tidspunkt inddeler husstandene i to grupper: husstande i gruppe 1, der er forbrugere af varemærket, og husstande i gruppe 2, der ikke er forbrugere af varemærket. α er da den brøkdel af husstandene i gruppe 2, som fra et tidspunkt til det næste flytter over i gruppe 1. Tilsvarende er β sandsynligheden for, at en husstand, der på et tidspunkt er forbruger af varemærket, på det nærmest efterfølgende tidspunkt ikke er forbruger af varemærket.

Det kan vises [10]), at

$$p_{11}^{(t)} = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^t$$

$$p_{21}^{(t)} = \frac{\alpha}{\alpha+\beta} \div \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^t$$

$$p_{12}^{(t)} = \frac{\beta}{\alpha+\beta} \div \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^t$$

$$p_{22}^{(t)} = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^t.$$

Da der kun knytter sig interesse til tilfældene $0<\alpha<1$ og $0<\beta<1$, ses det, at $(1-\alpha-\beta)^t$ går mod nul for voksende t. Fordelingen på de to tilstande er altså for store værdier af t stationær med

$$p_1 = \frac{\alpha}{\alpha+\beta} \text{ og } p_2 = \frac{\beta}{\alpha+\beta}$$

Eksempel 2.

Talmaterialet til dette eksempel stammer fra T. W. Anderson: „Probability Models for Analyzing Time Changes in Attitudes“ i P. F. Lazarsfeld (red.): „Mathematical Thinking in the Social Sciences“. The Free Press, Glencoe, Illinois. Second Edition, 1955.

En gruppe på 600 personer i Erie County, Ohio, blev i forbindelse med en undersøgelse, foretaget af „Bureau of Applied Social Research“, interviewet i hver af månederne maj–oktober 1940. Et af spørgsmålene angik hvilken kandidat den adspurgte havde til hensigt at stemme på ved valget i november. Svarene blev grupperet i tre grupper: Republikaner (R), Demokrat (D) og ved ikke (O); den sidste gruppe omfatter også svarnægtere og svaret „Andre kandidater“. Der var 445 personer, som besvarede alle seks interviews.

[10]) Feller, op. cit. p. 384, se også S. Goldberg: Introduction to Difference Equations. Wiley, New York, 1958, p.p. 217–33.

I nedenstående fem tabeller er de 445 personer grupperet efter deres hensigt med hensyn til stemmeafgivning i to på hinanden følgende måneder.

		Juni R	Juni D	Juni O	Ialt
Maj	R	125	5	16	146
	D	7	106	15	128
	O	11	18	142	171
	Ialt	143	129	173	445

		Juli R	Juli D	Juli O	Ialt
Juni	R	124	3	16	143
	D	6	109	14	129
	O	22	9	142	173
	Ialt	152	121	172	445

		August R	August D	August O	Ialt
Juli	R	146	2	4	152
	D	6	111	4	121
	O	40	36	96	172
	Ialt	192	149	104	445

		September R	September D	September O	Ialt
August	R	184	1	7	192
	D	4	140	5	149
	O	10	12	82	104
	Ialt	198	153	94	445

		Oktober R	Oktober D	Oktober O	Ialt
September	R	192	1	5	198
	D	2	146	5	153
	O	11	12	71	94
	Ialt	205	159	81	445

Ved rækkevis at dividere ind i en tabel med tallene i tabellens højre margin fås skøn over overgangssandsynlighederne; resultatet af beregningerne fremgår af nedenstående tabeller.

		Juni R	Juni D	Juni O
Maj	R	0.856	0.034	0.110
	D	0.055	0.828	0.117
	O	0.064	0.105	0.831

		Juli R	Juli D	Juli O
Juni	R	0.867	0.021	0.112
	D	0.047	0.845	0.108
	O	0.127	0.052	0.821

		August R	August D	August O
Juli	R	0.961	0.013	0.026
	D	0.050	0.917	0.033
	O	0.233	0.209	0.558

		September R	September D	September O
August	R	0.958	0.005	0.037
	D	0.027	0.940	0.033
	O	0.096	0.115	0.789

		Oktober R	Oktober D	Oktober O
September	R	0.970	0.005	0.025
	D	0.013	0.954	0.033
	O	0.117	0.128	0.755

Lad os et øjeblik antage, at der er tale om en markovkæde med konstante overgangssandsynligheder, d.v.s. at de fem stokastiske matrixer kun afviger tilfældigt fra hinanden. Efter juni-interviewet opstilles maj-juni-matricen; dernæst beregnes P^5, jfr. (18), hvorefter de beregninger, der er vist i (19), udføres; u'erne repræsenterer fordelingen af de 445 personer på de tre grupper i juni, d.v.s. $u_1 = 143$, $u_2 = 129$ og $u_3 = 173$. Dermed har man en forudsigelse for november måned af, hvorledes de 445 personer vil fordele sig på grupperne R, D og O. Efter juli-interviewet foretages en sammenfatning af maj-juni og juni-juli matricerne; hyppigheden, der f. eks. skal stå på pladsen (DR), bliver $\frac{7+6}{128+129}$. Denne matrix multipliceres med sig selv 4 gange, og man har så ved hjælp af julifordelingen en forudsigelse for november måned. Tallene i tabellerne viser imidlertid klart, at der ikke kan være tale om en for hele perioden fælles matrix af overgangssandsynligheder. Det oplyses i den citerede afhandling, at det republikanske konvent afholdtes mellem juni og juli-interviewet, medens det demokratiske afholdtes mellem juli og august interviewene; og „it may be expected that the probabilities of one changing his mind are different before the conventions and after the conventions. This is observed casualy in the tables above. We see that before the conventions the probability is about ·85 of retaining one's opinion for a month. However, after the conventions the probability is about ·95 of retaining vote intention for one of the major parties" (op.cit.p.49). Ved hjælp af et statistisk test påvises det, at der ikke er noget til hinder for at antage, at de to første matricer kun afviger tilfældigt fra hinanden, og det samme gælder for de to sidste matricer: „The results of this analysis is that the changes in the first two periods are explained by one model, and the changes in the third period are explained by a second model, and the changes in the fourth and fifth periods are explained by a third model" (op.cit.p.49–50). Selv om konklusionen lyder besnærende, kan man måske være lidt skeptisk overfor, om observationerne virkelig støtter den. Betragter man grupperne R og D, vil man se, at diagonalelementerne *gennem hele perioden* systematisk forøges, medens elementerne uden for diagonalen systematisk aftager. Det er derfor måske mere rimeligt at betragte perioden under et og anlægge det synspunkt, at hele perioden kan beskrives ved een model; der bliver i så fald tale om en model, hvor overgangssandsynlighederne ændrer sig efter et eller andet bestemt mønster, jfr. (5).

Akcepterer man, at perioden for valgkampagnen med hensyn til vælgernes stemmehensigter kan opdeles i tre faser, kan de to sidste matricer på den tidligere angivne måde sammenfattes til følgende matrix med gyldighed for valgkampens slutfase

$$\hat{P} = \begin{pmatrix} 0.964 & 0.005 & 0.031 \\ 0.020 & 0.947 & 0.033 \\ 0.106 & 0.121 & 0.773 \end{pmatrix} .$$

Vi kan på grundlag af $\hat{P}$ forudsige valget i november. I tabellen neden for er forudsigelsen sammenlignet med oktoberhensigterne og den faktiske stemmeafgivning.

	R	D	O
Oktober-hensigten	205	159	81
Forudsigelse ved $\hat{P}$	210	161	74
Faktiske stemmetal	204	146	95

Tallene, der angiver forudsigelsen ved $\hat{P}$, fremkommer ved at beregne produktsummerne af søjleelementerne i $\hat{P}$ med tallene, der angiver fordelingen i oktober, se september-oktober matricen. F. eks. er

$$205 \cdot 0.964 + 159 \cdot 0.020 + 81 \cdot 0.106 = 210 \quad .$$

Der er forskelle mellem forudsigelsen og den faktiske stemmeafgivning. „The most reasonable explanation of this discrepancy is the difference between verbalizing a vote intention and action. It is well known that in general not everyone who intends to vote actually does so. Furthermore, other studies show that Democrats are more vote-delinquent. Our analysis bears this out" (op.cit.p.52). –

Vi har i det foregående kun betragtet Markovkæder med konstante overgangssandsynligheder. Man kan i dette tilfælde beregne sandsynligheden for systemets overgang fra én tilstand til en anden tilstand t perioder senere ved at multiplicere overgangsmatricen P med sig selv t gange, jfr. (18). Disse sandsynligheder er de samme, uanset hvor tidsintervallet af længden t placeres. Betegner man med $p_{ij}(m,n)$ sandsynligheden for systemets overgang fra tilstand i på tidspunkt m til tilstand j på tidspunkt n, gælder det altså, at

$$p_{ij}(m+s,\ n+s) = p_{ij}(m,n) = p_{ij}^{(n-m)} \quad , \tag{22}$$

d.v.s., at overgangssandsynligheden er uafhængig af udgangstidspunktet m og sluttidspunktet n; den afhænger kun af tidsafstanden n-m. Det fremgår endvidere af det foregående, at overgangssandsynlighederne er knyttet sammen ved relationen

$$p_{ij}^{(t_1+t_2)} = \sum_{\nu=1}^{k} p_{i\nu}^{(t_1)} \cdot p_{\nu j}^{(t_2)} \quad . \tag{23}$$

Når overgangsmatricen ændrer sig med tiden, sådan som det f. eks. var tilfældet i eksempel 2, har vi at gøre med en Markovkæde med variable overgangssandsynligheder, og teorien må baseres på udtrykket (5) i stedet for udtrykket (6). I så fald gælder relationen (22) ikke længere idet overgangssandsynligheden $p_{ij}(m,n)$ afhænger af såvel begyndelsestidspunktet m som sluttidspunktet n, og relationen (23) må erstattes af den mere generelle relation mellem overgangssandsynlighederne

$$p_{ij}(t_1,t_2) = \sum_{\nu=1}^{k} p_{i\nu}(t_1,r) \cdot p_{\nu j}(r,t_2) \quad , \tag{24}$$

der kaldes Chapman-Kolmogorovs ligning. Det bemærkes, at r er et mellem t_1 og t_2 vilkårligt indskudt tidspunkt.

Det vil ses, at de to modeller, Markovkæder med konstante og variable overgangssandsynligheder, i matematisk og beregningsmæssig henseende blot adskiller sig ved, at sidstnævnte model er lidt vanskeligere at håndtere end den førstnævnte. I stedet for at operere med een matrix P af overgangssandsynligheder nødsages man til at arbejde med flere overgangsmatricer $P_1, P_2, P_3, \cdots$ svarende til forskellige tidspunkter. Medens sandsynligheden for overgang fra én tilstand til en anden i løbet af f. eks. tre perioder fås ved beregning af P^3 i den første model, fås den i den anden model ved beregning af $P_4 \cdot P_5 \cdot P_6$, hvis begyndelsestidspunktet er $t=4$, men af $P_5 \cdot P_6 \cdot P_7$, hvis begyndelsestidspunktet er $t=5$.

Der er i den her givne fremstilling lagt afgørende vægt på at fremhæve modellens muligheder i prognostisk henseende. Inden man kommer så langt, må der naturligvis udføres et stort statistisk og faglig-teoretisk arbejde med henblik på en simpel beskrivelse af det pågældende procesforløb. I denne forbindelse er det vigtigt, at det foreliggende statistiske materiale inddeles i homogene undergrupper, f. eks. efter kriterier som alder, køn, geografisk beliggenhed o.s.v. Det drejer sig da om for hver undergruppe at finde frem til karakteristiske egenskaber ved overgangsmatricen og de forhold af f. eks. sæsonmæssig karakter, der betinger et bestemt mønster for udviklingen i overgangssandsynlighederne. Det næste skridt består i en sammenligning af de forskellige undergrupper for derved at finde frem til kriterier, der bringer de forskellige delområder ind under en fælles synsvinkel. Ved en sådan analyse rejser der sig en række statistiske problemer vedrørende estimation og hypoteseprøvning, som falder udenfor rammerne af nærværende artikel.