INTRODUKTION TILL MARKOVKEDJOR

av Göran Rundqvist, KTH

Läs detta först:

Det här kompendiet är avsett som en introduktion till kompendiet av Enger och Grandell. Det är absolut inget fel på ”det officiella kompendiet” annat än att det kan vara lite svårt att ta till sig det som står där utan ”viss förberedelse”, kanske i form av en mindre stringent och därför mer lättillgänglig framställning. (Observera dock att kursen fortfarande ”definieras” enligt det officiella kompendiet.)

Jag har bara kunnat få med det mest grundläggande och har till stora delar byggt upp teorin på basis av exempel. Fall nu inte för frestelsen att bara lära dig exemplen utantill! Den metoden för att klara en tenta kan förstås fungera men misslyckas för det mesta. Ett lite orginellt men bra sätt att se om man tillgodogjort sig materialet på rätt sätt är att **själv** försöka konstruera ett tentatal. Principerna för talets lösning ska vara desamma som de i exemplen och gärna hämtade från flera olika exempel. I övrigt ska ditt tal (jämfört med exemplen) ”se så olikt ut som möjligt.”

I kompendiet finns många ”. Ibland markerar de ett ”hemmagjort” begrepp, ibland står de kring ett vedertaget begrepp när det införs första gången och slutligen används de när jag vill säga något på ett sätt som inte är helt stringent men ändå fullt begripligt.

Ha dock inte för bråttom när du läser. Kör du fast kan det visserligen ordna sig om du bara läser vidare, men lika ofta är det enda rätta att gå tillbaka i texten. För att få maximalt utbyte av kompendiet är det vidare viktigt att du verkligen försöker lösa de uppgifter som finns i texten. Har du några frågor går det dock bra att skriva till mig. (goranr@kth.se) Några särskilda förkunskaper behövs inte men repetera gärna ”totala sannolikhetslagen” (se t ex”tärningskompendiet” sid 7 eller kursboken sats 2.9) och matrismultiplikation. Med ”kursboken” avses (här och i fortsättningen): ”Sannolikhetsteori och statistikteori med tillämpningar” av Gunnar Blom m. fl.

Kapitel 1

Markovkedjor i diskret tid

1.1 Ändliga kedjor

Adam, Bertil och Cecilia kastar boll enligt sannolikheter som framgår av matrisen nedan:

$$\begin{array}{cccc} & A & B & C \\ A & 0 & 0.4 & 0.6 \\ B & 0.3 & 0.7 & 0 \\ C & 1 & 0 & 0 \end{array}$$

Om vi betecknar ”Bertil kastar till Adam” med ”$B \to A$” så ser vi t ex att $P(B \to A) = 0.3$ etc. Vidare är Bertil den ende som kan kasta till själv och $P(B \to B) = 0.7$.

Kastandet inleds med ett ”inkast” som också är slumpmässigt enligt ”startvektorn”: $(0.1 \;\; 0.5 \;\; 0.4)$ d v s inkastet går t ex till Adam med sannolikhet 0.1 etc. Lägg märke till att radsummorna både i matrisen och i startvektorn måste vara 1.

Ovanstående är ett exempel på något vars fullständiga namn är: ”Ändlig tidshomogen markovkedja i diskret tid.” Matrisen ovan kallas för övergångsmatrisen medan startvektorn redan fått sitt rätta namn. Ändlig syftar på antalet ”tillstånd” som här motsvaras av vem som har bollen d v s vi har 3 tillstånd. En markovkedja är det eftersom sannolikheterna varmed bollen kastas vidare ej påverkas av varifrån bollen kom. Detta uttrycks ibland som att ”en markovkedja saknar minne.” Om t ex Adam vore ”mer benägen” att kasta tillbaka bollen till Bertil om han just fått bollen från honom så skulle det inte längre vara en markovkedja.

”Tidshomogen” betyder att övergångssannolikheterna inte ändras med tiden vilket inte gäller om t ex Bertil blir trött och ”mer benägen” att kasta till sig själv när man kastat länge. ”Diskret tid” slutligen betyder att det inte finns någon tid i egentlig mening; ”tidsenheten” är ju kast. (I den allmänna modellen heter tidsenheten ”steg”.)

Låt oss nu med A_1 (B_1, C_1) beteckna händelserna att Adam (Bertil,Cecilia) har bollen efter ett kast inkastet **oräknat** och låt oss kalla ”inkastaren” för David. Väsentligen genom ”totala sannolikhetlagen” får man:

$$
\begin{aligned}
P(A_1) &= P((D \to A \to A) \text{ eller } (D \to B \to A) \text{ eller } (D \to C \to A)) \\
&= (\text{disjunkta händelser}) \\
&= P(D \to A \to A) + P(D \to B \to A) + P(D \to C \to A) \\
&= (\text{Markovegenskapen!}) \\
&= 0.1 \cdot 0 + 0.5 \cdot 0.3 + 0.4 \cdot 1 = 0.55 \\
P(B_1) &= 0.1 \cdot 0.4 + 0.5 \cdot 0.7 + 0.4 \cdot 0 = 0.39 \\
P(C1) &= 0.1 \cdot 0.6 + 0.5 \cdot 0 + 0.4 \cdot 0 = 0.06.
\end{aligned}
$$

Kallar man matrisen ovan för $\boldsymbol{P}$, startvektorn för $\boldsymbol{s}(0)$ och radvektorn $(P(A_1), P(B_1), P(C_1))$ för $\boldsymbol{s}(1)$, kan dessa räkningar sammanfattas i matrismultiplikationen: $\boldsymbol{s}(1) = \boldsymbol{s}(0) \cdot \boldsymbol{P}$. Genom att se $\boldsymbol{s}(1)$ som "ny startvektor" får man p s s (=på samma sätt) att $\boldsymbol{s}(2) = \boldsymbol{s}(1) \cdot \boldsymbol{P}$ eller mera allmänt att:

$$
\boldsymbol{s}(i+1) = \boldsymbol{s}(i) \cdot \boldsymbol{P}, \quad \text{för } i = 0, 1, 2, 3, 4 \ldots \tag{1.1}
$$

En del räknande ger

$$
\begin{aligned}
\boldsymbol{s}(2) &= (0.177,\ 0.493,\ 0.330), \\
\boldsymbol{s}(3) &= (0.478,\ 0.416,\ 0.106), \\
&\vdots \\
\boldsymbol{s}(14) &= (0.327,\ 0.458,\ 0.215), \\
&\vdots \\
\boldsymbol{s}(19) &= (0.346,\ 0.453,\ 0.201), \\
&\vdots \\
\boldsymbol{s}(24) &= (0.339,\ 0.455,\ 0.206)
\end{aligned}
$$

och här verkar det onekligen som om $\boldsymbol{s}(n)$ går mot en "gränsvektor" ($\boldsymbol{g}$) då $n \to \infty$.

Om det finns en sådan gränsvektor så får man av (1.1) ovan att den i så fall måste uppfylla $\boldsymbol{g} = \boldsymbol{g} \cdot \boldsymbol{P}$. Sätter vi $\boldsymbol{g} = (g_1, g_2, g_3)$ och utför multiplikationen fås

$$
\begin{aligned}
g_1 &= g_2 \cdot 0.3 + g_3 \\
g_2 &= g_1 \cdot 0.4 + g_2 \cdot 0.7 \\
g_3 &= g_1 \cdot 0.6.
\end{aligned}
$$

Detta är ett ekvationssystem med 3 ekvationer och 3 obekanta, men en av ekvationerna följer av de bägge andra så vi behöver en ekvation till. En sådan har vi i $g_1+g_2+g_3 = 1$ som självklart måste gälla. Det är praktiskt med ett namn för denna typ av ekvationssystem så låt oss kalla det för "G-systemet". Vi löser G-systemet och får $g_1 = 15/44$, $g_2 = 20/44$, $g_3 = 9/44$. Avrundat ger detta: $\boldsymbol{g} = (0.3410,\ 0.4545,\ 0.2045)$ och vi ser nu vart $\boldsymbol{s}(n)$ "eventuellt var på väg" när n gick mot oändligheten. Lägg märke till att $\boldsymbol{g}$ ej beror på startvektorn.

Hur ser då de villkor ut som garanterar att gränsvektorn existerar? Antag t ex att inkastet alltid sker till Adam d v s $\boldsymbol{s}(0) = (1, 0, 0)$ och att alla kast sker med sannolikhet 1 enligt $A \to B \to C \to A \to B$ osv. I så fall får vi: $\boldsymbol{s}(1) = (0, 1, 0)$, $\boldsymbol{s}(2) = (0, 0, 1)$, $\boldsymbol{s}(3) = (1, 0, 0)$

etc. d v s i det fallet finns ingen gränsvektor. Detta är ett exempel på en periodisk kedja (med period 3). Ett nödvändigt villkor för att $\boldsymbol{g}$ ska existera borde alltså vara att kedjan är aperiodisk (=ej periodisk).

För att komma vidare behöver vi nu begreppen "irreducibel" och "ergodisk". En kedja kallas för irreducibel om alla tillstånd i den "kommunicerar" d v s givet två tillstånd A och B vilka som helst så finns det alltid minst ett sätt att (i ett eller flera steg) komma från A till B och vice versa. (Låt oss beteckna detta med $A< - \cdot\cdot - >B$.) När en gränsfördelning (som inte beror på hur startvektorn ser ut!) existerar kallas kedjan för "ergodisk".

Vi kan nu formulera vår **huvudsats** för **ändliga** markovkedjor: "Så snart kedjan är irreducibel och aperiodisk så är den ergodisk". Gränsfördelningen $\boldsymbol{g}$ erhålles som (den entydiga) lösningen till G-systemet.

Låt oss med "utflykt" mena en väg som börjar och slutar i samma tillstånd. Ett praktiskt sätt att visa att en (ändlig) kedja är irreducibel är att visa att det finns en "sightseeing" varmed menas en utflykt som "besöker" alla andra tillstånd minst en gång. För "bollkedjan" ovan finns t ex $B \to A \to C \to A \to B$ vilket visar att kedjan är irreducibel.

Hur undersöker man sedan om kedjan är aperiodisk? Ja, för en irreducibel kedja räcker det att visa att **ett** tillstånd är aperiodiskt eftersom det därav följer att alla andra också är det. Detta är lätt att bevisa:

Låt L och H vara två tillstånd vilka som helst i en irreducibel kedja d v s $L< - \cdot\cdot - >H$ gäller. Antag nu att L är periodiskt med period $= d$. Detta betyder att stegantalet för varje utflykt från L är en multipel av ett heltal $d > 1$.

Säg nu att vi har en **godtycklig** utflykt från H d v s $H - \cdot\cdot - >H$. Stegantalet för $H - \cdot\cdot - >H$ är = stegantalet för $L - \cdot\cdot - >H - \cdot\cdot - >H - \cdot\cdot - >L$ minskat med stegantalet för $L - \cdot\cdot - >L$. Eftersom de båda sista vägarna är utflykter från L måste deras stegantal vara multipler av d. Detta måste då också gälla för stegantalet för $H - \cdot\cdot - >H$ vilket i sin tur betyder att H är periodiskt (med samma period som L vilket man också kan visa.) Sammanfattningsvis är alltså alla tillstånd i en irreducibel kedja antingen aperiodiska eller periodiska (med samma period).

Hur avgör man om ett tillstånd är aperiodiskt? Det räcker om man kan hitta utflykter med stegantal som är "relativt prima" d v s de har ingen gemensam faktor annan än 1.

Allra enklast blir detta förstås om det finns en utflykt med stegantal = 1 d v s tillståndet kan gå till sig själv i ett steg. Detta svarar mot att det finns ett element > 0 på huvuddiagonalen i övergångsmatrisen. Är detta inte fallet får man göra en närmare undersökning. Exempel: Antag att vi hittar 2 utflykter med stegantal 6 och 15. Detta visar ännu ingenting eftersom 3 är en gemensam faktor. Men skulle vi nu också hitta en utflykt med stegantalet 10 är det klart att tillståndet är aperiodiskt eftersom den enda gemensamma faktorn till 6, 15 och 10 är 1. (Fast såhär krångligt är det sällan; det brukar räcka med två utflykter.)

Vi har visat att "bollkedjan" som vi inledde med är irreducibel och eftersom Bertil kan kasta till sig själv är den också aperiodisk. Den är alltså ergodisk och gränsfördelningen bestämde vi ju ovan. En fråga som återstår är vad gränsfördelningen säger praktiskt. Det finns två tolkningar. Att gränssannolikheten för t ex Cecilia är 0.205 betyder att sannolikheten är 0.205 att Cecilia har bollen "efter lång tid" (=efter ett stort antal kast) men 0.205 uttrycker också att Cecilia har fått bollen i ungefär 20.5% av kasten.

Villkoret aperiodisk är nödvändigt för att en kedja ska vara ergodisk. Hur är det med irreducibel d v s kan en kedja som inte är irreducibel vara ergodisk? Ja det kan den, men i så fall finns där tillstånd vilkas gränssannolikheter är triviala (d v s = 0 eller 1). Enklast förstår man detta genom ett exempel som vi (för att inte skymma det väsentliga) ger som en figur utan siffervärden på övergångssannolikheterna.

$$\leftrightarrow (1) \leftrightarrow (2) \rightarrow (3) \leftrightarrow (4) \leftrightarrow$$

De möjliga övergångarna (i ett steg) mellan tillstånden ovan är markerade med pilar d v s en sådan pil betyder att det finns en sannolikhet > 0 att gå som pilen visar. $\leftrightarrow$ till vänster om (1) betyder att (1) kan gå till sig själv i ett steg vilket även gäller för (4). Att kedjan ej är irreducibel syns direkt; det går t ex inte att komma från (3) till (2). Antag nu att kedjan startar i (1). Den kommer då att vara i (1) eller (2) en viss ändlig tid innan den går till (3). Därefter återvänder den aldrig vare sig till (1) eller (2). (1) och (2) är s k "genomgångstillstånd" eller som de också kallas "obeständiga tillstånd". Gränssannolikheterna för (1) och (2) är därför noll.

Om man **reducerar** kedjan genom att ta bort (1) och (2) bildar de tillstånd som är kvar en irreducibel (ty (3)→(4)→(3) är möjligt) och aperiodisk kedja (ty (4)→(4) i ett steg är möjligt). Löser man G-systemet för denna (ergodiska) delkedja får man g_3 och g_4. Den ursprungliga kedjan är också ergodisk och dess gränsfördelning är $\boldsymbol{g} = (0, 0, g_3, g_4)$.

Låt oss slutligen göra kedjan "ännu trivialare" genom att ta bort vägen från 4 till 3 så att 4 blir ett s k "absorberande tillstånd". Kedjan är då fortfarande ergodisk nu med $\boldsymbol{g} = (0, 0, 0, 1)$.

Det är praktiskt att beteckna tillstånden med 1,2,3 etc. I så fall är det element som finns på rad k i den j:te kolonnen i övergångsmatrisen $\boldsymbol{P}$ lika med sannolikheten att i ett steg gå från k till j. "På samma ställe" fast i matrisen $\boldsymbol{P} \cdot \boldsymbol{P}$ istället finns sannolikheten att gå från k till j i **två** steg och p s s ger matrisen $\boldsymbol{P} \cdot \boldsymbol{P} \cdot \boldsymbol{P}$ sannolikheterna för övergångar i 3 steg etc.

Praktiskt är det också att införa stokastiska variabler för att ange i vilket tillstånd kedjan är efter ett visst antal steg. Om vi numrerar tillstånden i "bollkedjan" $(A = 1, B = 2, C = 3)$ och inför $X(t)$: Tillståndsnummer efter t kast (inkastet oräknat) kan vi t ex skriva $X(1) = 3$ för att beteckna att Cecilia har bollen efter 1 kast (inkastet oräknat) och $X(0) = 2$ betyder p s s att Bertil har bollen direkt efter inkastet. Vi kan också uttrycka att sannolikheten är $9/44$ att Cecilia har bollen "efter lång tid" genom att skriva: $\lim_{t\to\infty} P(X(t) = 3) = 9/44$.

Det som vi har kallat gränsfördelningen ovan kallas ofta för den stationära fördelningen. Med "stationär fördelning" avses egentligen **varje** lösning till G-systemet, men när kedjan är ergodisk är G-systemet entydigt lösbart och då sammanfaller stationär fördelning och gränsfördelning.

1.2 Oändliga kedjor

För att en kedja med oändligt antal tillstånd ska vara ergodisk räcker det inte med att den är irreducibel och aperiodisk utan vi måste även som **villkor** ha med att G-systemet går att lösa.

Vad som kan hända är att kedjan "sticker iväg till oändligheten". I så fall får man från de ekvationer i G-systemet som kommer från matrismultiplikationen att $\boldsymbol{g} = (0, 0, 0, 0, 0, \ldots)$ vilket inte satisfierar den sista ekvationen att radsumman ska vara $= 1$. **Alla** kedjans tillstånd blir då obeständiga och någon gränsfördelning existerar inte.

Vi ger nu ett exempel (som kan hoppas över vid en första genomläsning.)

$$\leftrightarrow (1) \leftrightarrow (2) \leftrightarrow (3) \leftrightarrow (4) \leftrightarrow (5) \leftrightarrow \cdots$$

Någon "sightseeing" går inte att genomföra men vi ser direkt av figuren att alla tillstånd kommunicerar vilket betyder att kedjan är irreducibel. Tillstånd (1) kan gå till sig själv i ett steg vilket betyder att kedjan också är aperiodisk. Vore kedjan ändlig skulle vi nu kunna sluta att den var ergodisk och lösa G-systemet. Säg nu att sannolikheterna att "gå åt höger f o m (1)" är p med $0 < p < 1$. (Övriga sannolikheter $= 1 - p$ förstås.) Kalla övergångsmatrisen $\boldsymbol{P}$.

Med $\boldsymbol{g} = (g_1, g_2, g_3, \ldots)$ blir resultatet av $\boldsymbol{g} = \boldsymbol{g} \cdot \boldsymbol{P}$

$$\begin{aligned} g_1 &= g_1 \cdot (1-p) + g_2 \cdot (1-p) \\ g_{k+1} &= g_k \cdot p + g_{k+2} \cdot (1-p), \quad (k = 1, 2, 3, \ldots) \end{aligned}$$

Det är inte särskilt svårt att lösa detta system trots att där finns oändligt många ekvationer och oändligt många obekanta. Precis som i det ändliga fallet han en ekvation strykas. Låt oss stryka den första. De andra ekvationerna kan lösas med hjälp av teorin för differensekvationer (Se t ex Mathematical Handbook). Med $u = p/(1-p)$ där p **ej** är lika med 0.5 får vi:

$$g_k = a + bu^k \quad \text{där } a \text{ och } b \text{ är konstanter.}$$

För att detta ska uppfylla $g_1+g_2+g_3+\cdots = 1$ måste gälla $a = 0$ och $b = 1/(u+u^2+u^3+\cdots)$. Om $p < 0.5$ gäller så medför detta att $u < 1$ och serien (en geometrisk serie) konvergerar med summan $u/(1-u) = p/(1-2p)$. Kedjan är då ergodisk och med $b = (1-2p)/p$ och $u = p/(1-p)$ blir gränsfördelningen: $g_k = bu^k$ för $k = 1, 2, 3, \ldots$

Skulle å andra sidan $p > 0.5$ gälla divergerar serien mot oändligheten och vi får $g_k = 0$ för $k = 1, 2, 3, \ldots$ d v s alla tillstånden blir obeständiga och kedjan "sticker iväg mot oändligheten".

För $p = 0.5$ blir lösningen till differensekvationen: $g_k = a + bk$. $g_1 + g_2 + g_3 + \cdots = 1$ ger nu $a = 0$ och $b = 1/(1+2+3+4+\cdots)$ d v s även nu blir $g_k = 0$ för $k = 1, 2, 3, 4, \ldots$

Kapitel 2

Markovkedjor i kontinuerlig tid

(Vi begränsar oss till s k reguljära kedjor. Eftersom alla kedjor som man träffar på i tillämpningarna är reguljära spelar denna inskränkning ingen roll i praktiken.)

2.1 Ändliga kedjor

Om våra bollspelande vänner i exemplet ovan istället spelar enligt reglerna för en markovkedja i kontinuerlig tid blir mycket annorlunda. Detta beror på att tiden nu verkligen finns med. I det diskreta fallet fanns visserligen också tid men inte i egentlig mening eftersom ”tidsenheten” där var **steg**. Antag att inkastet går till Bertil. Spelar man efter ”de nya reglerna” behåller Bertil nu bollen en viss slumpmässig tid innan han kastar den vidare till någon av de bägge andra som p s s behåller bollen en viss slumpmässig tid innan den kastas vidare. Observera att den tid själva kasten tar nu antas vara försumbar.

Vad Bertil eller någon annan gör med bollen när han har den är ointressant d v s möjligheten för ett tillstånd att gå till sig själv i ett steg ”finns inte längre”. Vidare kan en kedja i kontinuerlig tid aldrig vara periodisk. Att det begreppet kom in i bilden i det diskreta fallet berodde ju helt på att vi där mätte tiden i steg.

Vilken sannolikhetsfördelning får nu den tid som kedjan ligger kvar i ett tillstånd? Man kan visa att den blir exponentialfördelad, vilket är en fördelning som passar bra ihop med den ”minneslöshet” som karakteriserar en markovkedja. För exponentialfördelningen gäller nämligen att fördelningen för den återstående tiden ej beror på hur lång tid som redan gått! (Detta framgår av anmärkning 3.2 på sid 61 i kursboken.)

Precis som i det diskreta fallet kan kedjans beteende efter start sammanfattas i en matris ($\boldsymbol{Q}$) som dock nu ser ganska annorlunda ut:

$$\begin{array}{cccc} & A & B & C \\ A & -8 & 4 & 4 \\ B & 3 & -5 & 2 \\ C & 0 & 2 & -2 \end{array}$$

Matrisen $\boldsymbol{Q}$ ovan är en s k intensitetsmatris som bl a berättar om den ”intensitet” med vilken bollen kastas till de olika spelarna. För att utvinna den information som matrisen innehåller behövs vissa manipulationer. Genom att invertera elementen i huvuddiagonalen och ta bort minustecknen får man t ex väntevärdena för de exponentialfördelningar som bestämmer hur länge kedjan stannar i de olika tillstånden.

Säg att inkastet går till Bertil. Av det mittersta elementet i matrisen ser vi att han kommer att behålla bollen i en tid som är exponentialfördelad med väntevärdet 1/5. Om vi räknar tiden i minuter betyder detta att Bertil i genomsnitt behåller bollen i 12 sekunder.

Hur ser nu den matris ut som bestämmer sannolikheterna för övergångarna mellan de olika tillstånden? Huvuddiagonalen i den består förstås av nollor eftersom ett tillstånd "inte kunde gå till sig själv". De andra elementen får vi om vi multiplicerar varje rad med den parameter som vi (p s s som ovan) får från diagonalelementet i raden. I den första raden ska vi alltså multiplicera med 1/8 i den andra med 1/5 och i den tredje med 1/2. Detta ger den så kallade "uthoppsmatrisen":

$$\begin{array}{cccc} & A & B & C \\ A & 0 & 1/2 & 1/2 \\ B & 3/5 & 0 & 2/5 \\ C & 0 & 1 & 0 \end{array}$$

Observera att radsummorna blir ett som de ska, vilket är en följd av att radsummorna i $\boldsymbol{Q}$-matrisen var noll. (Vi kunde inte behålla sannolikheterna för Bertil och har ändrat även för Adam och Cecilia.)

Hur är det nu med ergodiciteten? Ja det enda villkor som behövs när kedjan är ändlig, är att kedjan är irreducibel eftersom kedjan "automatiskt är aperiodisk". (Observera dock att på samma sätt som vid diskret tid kan en kedja i kontinuerlig tid vara ergodisk utan att vara irreducibel.)

När man ska bestämma gränsfördelningen ($\boldsymbol{g}$) blir det dock annorlunda. Den ges av $\boldsymbol{g} \cdot \boldsymbol{Q} =$ nollvektorn och **inte** av $\boldsymbol{g}\cdot$uthoppsmatrisen $= \boldsymbol{g}$. Dessutom behövs villkoret att radsumman i $\boldsymbol{g}$ ska vara ett eftersom en av ekvationerna i $\boldsymbol{g} \cdot \boldsymbol{Q} = \boldsymbol{0}$ följer av de andra p s s som i det diskreta fallet. Genomför gärna dessa beräkningar här. Du ska få $\boldsymbol{g} = (0.12,\ 0.32,\ 0.56)$. Observera att $\boldsymbol{g}$ kallas för π i F.S. (Jfr. F.S. 14.2.3)

I köteori och därmed besläktade problem används ofta s k "födelse och dödsprocesser". Dessa är markovkedjor i kontinuerlig tid där en (direkt) övergång bara är möjlig om skillnaden mellan tillståndsnumren är ett. Numrerar vi tillstånden i "bollkedjan" ovan ($A = 1, B = 2, C = 3$) ser vi att den inte uppfyller detta villkor eftersom $A \to C$ är möjligt.

Vi går nu till några exempel som vi formulerar på det sätt som är vanligt på tentor. (De nya begreppen förklaras i lösningarna.)

Ex 1. Till en korvkiosk anländer kunder enligt en poissonprocess med intensitet 20 kunder/timme, men när 4 kunder (= de som står i kö plus den som betjänas) finns "i systemet" går (potentiella) kunder någon annanstans för att handla korv. Betjäningstiderna är oberoende och exponentialfördelade med väntevärde 3 min. Bejäningstiderna är också oberoende av ankomstprocessen. Stationärt tillstånd förutsätts.

a. Bestäm sannolikhetsfördelningen för antal kunder i systemet.

b. Beräkna väntevärdet för antal kunder i systemet.

c. Beräkna väntevärdet för antal kunder i kön.

d. Beräkna hur många kunder som i genomsnitt bortfaller under en timme.

LÖSNING: Att stationärt tillstånd förutsätts betyder att det som söks kan bestämmas genom gränsfördelningen. Inför nu antal kunder i systemet som tillståndsnummer. (I köteori är det naturligt att ha noll som tillståndsnummer.) Att ankomsterna sker enligt en poissonprocess betyder (bland annat) att tiderna mellan ankomsterna är oberoende och exponentialfördelade. Den "intensitet" varmed ankomsterna sker är $20/60 = 1/3$ kund per min och eftersom 1 kund betjänas på i genomsnitt 3 min så är "betjäningsintensiteten" också $1/3$ kund per min. "Intensitetsgrafen" blir:

$$(0) \leftrightarrow (1) \leftrightarrow (2) \leftrightarrow (3) \leftrightarrow (4)$$

Intensiteterna för övergångarna åt höger de s k "födelseintensiteterna" betecknar vi λ_k, medan intensiteterna för övergångarna åt vänster de s k "dödsintensiteterna"betecknas μ_k. Här är k det tillstånd man **kommer från** så vi har att $\lambda_k = 1/3$ för $k = 0, 1, 2, 3$ och $\mu_k = 1/3$ för $k = 1, 2, 3, 4$. För att lösa G-systemet kan vi nu använda formelsamlingen 14.2.8 som visserligen är avsedd för oändliga kedjor men som går att använda också för ändliga. Vi har $\rho_0 = 1$, $\rho_1 = \lambda_0/\mu_1 = 1$, $\rho_2 = (\lambda_0/\mu_1) \cdot (\lambda_1/\mu_2) = 1$ etc. Vi får $\rho_k = 1$ för $k = 0, 1, 2, 3, 4$ och sätter $\rho_k = 0$ för $k > 4$.

a. Enligt F.S. har vi $g_k = \rho_k/(\rho_0 + \cdots + \rho_4)$ d v s $g_k = 1/5$ för $k = 0, 1, 2, 3, 4$. Eftersom kedjan är ändlig och irreducibel är detta gränsfördelningen och svaret på uppgift a.

b. I b. söks $E(X)$ där X är antal kunder i systemet. Vi får $E(X) = (0 \cdot 1/5) + \cdots + (4 \cdot 1/5) = 2$.

c. Svaret här är **inte** $E(X) - 1$ eftersom sambandet kölängd $= X - 1$ inte gäller då $X = 0$. Sökt är istället: $(0 \cdot 2/5) + (1 \cdot 1/5) + \cdots + (3 \cdot 1/5) = 1.2$. (Kölängden är ju noll både då $X = 0$ och $X = 1$ d v s med sannolikhet $1/5 + 1/5$.)

d. Bortfall av kunder sker då $X = 4$ och $P(X = 4) = 1/5$. På en timme kommer i genomsnitt 20 potentiella kunder. Av dessa bortfaller "alltså" $1/5$ d v s 4 stycken (i genomsnitt). Detta resonemang kan göras matematiskt bättre om man utnyttjar att antalet kunder som anländer på en timme är poissonfördelat med väntevärde 20. (Jfr tal 48a i kompendiet av Enger–Grandell.)

KOMMENTAR: Eftersom exponentialfördelningen är en kontinuerlig fördelning kan det inte komma två kunder **precis** samtidigt. Inte heller kan en kund komma till systemet **precis** samtidigt som en annan kund lämnar det. När en övergång sker blir det alltså alltid till ett "bredvidliggande" tillstånd d v s vår kedja är en födelse-döds-process. Hur är det nu med den tid som kedjan ligger kvar i ett tillstånd? Eftersom vi har en kontinuerlig markovkedja ska den vara exponentialfördelad. Ligger kedjan i tillstånd (0) är detta omedelbart klart eftersom tiden till dess en förändring sker bestäms av tiden till dess en kund kommer. Men hur är det när vi är exempelvis i tillstånd (1)? Tiden till dess en förändring sker bestäms då av vad som händer först. Antingen kommer en ny kund och kedjan går till tillstånd (2) eller så blir den kund som betjänas färdigbetjänad och kedjan går till tillstånd (0). Tiden i (1) blir alltså den minsta av dessa tider. Att denna tid också blir exponentialfördelad följer av följande viktiga sats: $X_1, X_2, \ldots, X_n$ oberoende och exponentialfördelade $\Rightarrow Z = \min(X_1, X_2, \ldots, X_n)$ är exponentialfördelad. För bevis se lösningen till tal 4.16 i kursboken. Av lösningen (som finns på hemsidan) framgår även att intensiteten för Z blir lika med summan av intensiteterna för X_i:na. (Jfr ex. 2 nedan.) Även för de exempel som följer är ovanstående kommentarer relevanta.

Ex 2. Tre maskiner "servas" av två reparatörer. Maskinerna "havererar" **var och en** enligt en poissonprocess med intensitet = 5 haverier/år och oberoende av varandra. Lagningstiderna för var maskin är oberoende och exponentialfördelade med väntevärde 0.25 år, men då **en** maskin är sönder kan bara **en** reparatör användas. Lagningstiderna är dessutom oberoende av "haveriprocessen". Hur många dygn (i genomsnitt) under ett år fungerar alla tre maskinerna samtidigt? (Stationärt tillstånd förutsätts.)

LÖSNING: Också detta är en födelse och dödsprocess. Som tillståndsnummer inför vi antal hela maskiner Med samma beteckningar som i ex 1. får vi: $\lambda_0 = \lambda_1 = 8$ och $\lambda_2 = 4$. "Lagningsintensiteten" är nämligen $1/0.25 = 4$ för **en** maskin och när t ex alla tre är trasiga kan vilken som helst av de två som då lagas bli färdiglagad, vilket betyder att intensiten för att gå från tillstånd 0 till tillstånd 1 **fördubblas**. P s s får vi $\mu_1 = 5$, $\mu_2 = 10$ och $\mu_3 = 15$ ty om t ex alla tre maskinerna är hela kan vilken som helst av dem gå sönder vilket ger tre gånger så stor intensitet som då bara en maskin är hel. FS 14.2.8 ger nu: $\rho_0 = 1$, $\rho_1 = 8/5$, $\rho_2 = (8/5)\cdot(8/10)$, $\rho_3 = (8/5)\cdot(8/10)\cdot(4/15)$ och $\rho_k = 0$ för $k > 3$. Vi får $\rho_0 + \rho_1 + \rho_2 + \rho_3 = 4.221$ och $g_3 = \rho_3/4.221 = 0.081$. Sökt är $0.081 \cdot 365 = 29.5$ dygn.

2.2 Oändliga kedjor

För en reguljär markovkedja i kontinuerlig tid med oändligt många tillstånd gäller att den är ergodisk om den är irreducibel och G-systemet är lösbart.

Ex 3. Vår korvhandlare flyttar ut på landet. Här finns det gott om plats så antalet kunder i systemet kan vara i princip hur stort som helst. Något kundbortfall blir det inte eftersom det inte finns någon annanstans att handla. Nu kommer dock bara i genomsnitt 10 kunder per timme. För övrigt gör vi samma antaganden som i ex 1. och räknar a, b och c.

a. Vi får $\lambda_k = 10/60 = 1/6$ kund per minut och $\mu_k = 1/3$ kund per minut. (Obs. att λ_k och μ_k måste ha samma sort!) Vi följer FS 14.2.8 och får: $\rho_0 = 1$, $\rho_1 = \lambda_0/\mu_1 = 1/2$, $\rho_2 = (1/2)\cdot(1/2)$ etc. och allmänt $\rho_k = (1/2)^k$ för $k = 0, 1, 2, 3, \ldots$ Genom formeln för summan av en oändlig geometrisk serie får vi $g_0 + g_1 + g_2 + g_3 + \cdots = 1/(1-0.5) = 2$ och $g_k = (1/2)^{k+1}$ för $k = 0, 1, 2, 3, \ldots$ Att detta är den sökta gränsfördelningen följer av att kedjan är irreducibel och reguljär.

b. Vi ska nu bestämma

$$E(X) = 0 \cdot g_0 + 1 \cdot g_1 + 2 \cdot g_2 + 3 \cdot g_3 + \cdots = \frac{1}{4} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{16} + \cdots .$$

Detta kan man göra med samma knep som man använder för att härleda formeln för summan av en geometrisk serie. Multiplikation med 1/2 ger:

$$\frac{1}{2}E(X) = \frac{1}{8} + 2 \cdot \frac{1}{16} + 3 \cdot \frac{1}{32} + \cdots .$$

Subtraktion ger:

$$E(X) - \frac{1}{2}E(X) = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \cdots = \text{(Geometrisk serie!)} = \frac{1/4}{1-(1/2)} = \frac{1}{2}.$$

Detta ger $E(X) = 1$. (Matematiskt korrektare hade varit att använda ett gränsvärdesresonemang.)

c. Precis som i ex. 1 blir kölängden noll både då $X = 0$ och då $X = 1$ d v s med sannolikhet $g_0 + g_1 = 1/2 + 1/4 = 3/4$. Det sökta blir därför: $0 \cdot (3/4) + 1 \cdot g_2 + 2 \cdot g_3 + 3 \cdot g_4 + \cdots = 1/8 + 2/16 + 3/32 + \cdots$. Samma knep som i b. ger $E(\text{kölängden}) = 0.5$. (Genomför räkningarna själv!)

KOMMENTAR: Det vi har i ex. 3 är en s k (M/M/1) kö där det första M:et berättar att ankomstprocessen är en poissonprocess och det andra visar att betjäningstiderna är oberoende och exponentialfördelade. Ettan slutligen anger antalet betjäningsstationer. Exempel 3 kan också lösas med hjälp av formelsamlingen. Först bestämmer man den s k "trafikintensiteten" som finns definierad på sid 11. Sedan går man till "System med $c = 1$" på sidan 12. Där står först "p_n" som ger svaret i ex.3a och omedelbart därefter "l_q" som ger svaret i ex.3c. Svaret i ex.3b. kallas i FS för "l" och hittas långt ner på sid 11, där man också hittar formler för att bestämma medelväntetiden i kön respektive medelväntetiden i systemet. Observera att alla dessa formler förutsätter att trafikintensiteten är < 1. Korvhandlaren får tacka sin lyckliga stjärna att inte kundintensiteten var som i ex.1 ty då går antal personer i systemet mot oändligheten!

Ex.4 För att göra om modellen i ex.2 till ett (M/M/2)-system får vi dels förutsätta en "oändlig" maskinpark och dels ändra intensiteten 5 haverier per år till att gälla maskinparken **som helhet**. Inför nu antalet **trasiga** maskiner som tillståndsnummer och beräkna hur lång tid det i genomsnitt tar från det en maskin havererat till dess den blivit färdiglagad. Beräkna sedan också hur många trasiga maskiner som i genomsnitt väntar på att bli lagade. Den första uppgiften löser man med hjälp av FS medan den andra också kan lösas p s s som i ex.3. Eftersom detta är bra uppgifter att öva på ger vi bara svaren: 16/39 år resp. 125/156 stycken.

Kapitel 3

Rekursiva metoder

Med hjälp av s k ”rekursiv teknik” kan man lösa problem som i förstone verkar helt omöjliga. Vi ger ett antal exempel där tekniken tillämpas på lite olika sätt även om det i grunden rör sig om samma teknik.

3.1 Ändliga kedjor i diskret tid

Kedjan

$$\leftrightarrow (1) \leftarrow (2) \leftrightarrow (3) \leftrightarrow (4) \rightarrow (5) \leftrightarrow$$

är given och vi vet att $P(2 \to 3) = 0.1$, $P(3 \to 4) = 0.4$ och $P(4 \to 5) = 0.2$ gäller.

Tillstånden (1) och (5) är som synes absorberande d v s $P(1 \to 1)$ och $P(5 \to 5)$ är lika med ett. Komplementräkning ger att $P(2 \to 1) = 0.9$, $P(3 \to 2) = 0.6$ och $P(4 \to 3) = 0.8$. Den här kedjan är inte ergodisk eftersom gränsfördelningen beror på startvektorn; om vi startar i (1) blir den $(1, 0, 0, 0, 0)$ och startar vi i (5) blir den $(0, 0, 0, 0, 1)$.

Säg att kedjan startar i (2) och vår uppgift är att bestämma sannolikheten för att den absorberas i (1). Uppgiften kan klaras om vi inför:

$a_k = P(\text{Kedjan absorberas i (1) vid start i tillstånd } k)$ för $k = 1, 2, 3, 4, 5$. Vi söker a_2 och ser direkt att $a_1 = 1$ och $a_5 = 0$. Totala sannolikhetslagen ger:

$$a_2 = P(2 \to 1) + P(2 \to 3) \cdot P(\text{abs. i (1) vid start i tillstånd 3})$$

d v s vi har $a_2 = 0.9 + 0.1a_3$. P s s: $a_3 = 0.6a_2 + 0.4a_4$ och $a_4 = 0.8a_3 + 0.2a_5$. Eftersom $a_5 = 0$ gällde är detta ett ekvationssystem med 3 ekvationer och 3 obekanta som vi kan lösa. Vi får det sökta $a_2 = 0.987$. (På köpet får vi också $a_3 = 0.871$ och $a_4 = 0.697$.)

Låt oss nu istället fråga efter **väntevärdet** för antal steg till dess kedjan absorberas. Absorbtionen sker ju i antingen (1) eller (5), men att bara fråga efter väntevärdet för antal steg till absorbtion i (1) vore meningslöst eftersom detta blir oändligt. (Det finns ju en sannolikhet > 0 att kedjan istället absorberas i (5).)

För att bestämma det sökta väntevärdet behöver vi ett ”utvidgat” sätt att beräkna väntevärde. Vi repeterar först det vanliga beräkningssättet. genom ett exempel. Låt Z vara resultatet av ett kast med en vanlig (symmetrisk) tärning. Vi får väntevärdet för Z genom att multiplicera sannolikheterna för de olika utfallen med dess värden d v s vi har:

$$E(Z) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6$$

Det nya är att denna typ av formel gäller även då (alla eller en del av) värdena **själva** innehåller väntevärden. (Läs bara vidare så förstår du snart hur detta fungerar.)

Vi inför: $b_k = E$(antalet steg till absorbtion vid start i tillstånd k). Det nya utvidgade beräkningssättet av väntevärde ger nu:

$$b_2 = P(2 \to 1) \cdot 1 + P(2 \to 3) \cdot (1 + b_3)$$

som förklaras på följande sätt:

Om kedjan absorberas direkt sker detta med sannolikhet $P(2 \to 1)$ och tar ett steg. Detta förklarar den första termen i högerledet ovan. Kedjan kan också gå till (3) vilket sker med sannolikhet $P(2 \to 3)$. Det ”värde” som motsvarar detta är $1 + b_3$ där 1 är steget från (2) till (3) och b_3 är väntevärdet för det antal steg som sedan är kvar till absorbtion.

Med insatta värden får vi: $b_2 = 0.9 + 0.1(1 + b_3)$ och p s s även $b_3 = 0.6(1 + b_2) + 0.4(1 + b_4)$ och $b_4 = 0.2 \cdot 1 + 0.8(1 + b_3)$. Vi får det sökta $b_2 = 1.323$ (Och på köpet $b_3 = 3.226$ och $b_4 = 3.581$.)

3.2 Ändliga kedjor i kontinuerlig tid

Den rekursiva tekniken för att beräkna sannolikheter är precis densamma som i diskret tid, men vill vi istället beräkna väntevärden blir det annorlunda. Som illustration väljer vi att fortsätta på ex 2. Säg att vi ”kommer in i handlingen” när bara en maskin är hel och att reparatörerna har bestämt sig för att ta semester tillsammans så snart alla tre maskinerna fungerar. Hur länge får de i genomsnitt vänta tills detta händer? För beräkningen behövs intensitetsmatrisen. Elementen i huvuddiagonalen får man genom att radsummorna ska vara noll och de övriga elementen får man direkt från de födelse- och dödsintensiteter som gällde. Som tillståndsnummer väljer vi antal **hela** maskiner. Eftersom man inte är intresserad av vad som händer efter det att kedjan nått tillstånd (3) kan man se (3) som ett absorberande tillstånd och får:

$$\begin{array}{cccc} -8 & 8 & 0 & 0 \\ 5 & -13 & 8 & 0 \\ 0 & 10 & -14 & 4 \\ 0 & 0 & 0 & 0 \end{array}$$

Ett absorberande tillstånd motsvaras som synes av en rad med bara nollor. Genom att ta bort minustecknen och invertera elementen i huvuddiagonalen får vi ju väntevärdet för tiderna i de olika tillstånden och ser då speciellt att vad det gäller ett absorberande tillstånd så blir väntevärdet oändligt vilket är som det ska vara.

Vad vi söker är väntevärdet för tiden till absorbtion givet att man befinner sig i (1). På grund av exponentialfördelningens minneslöshet behöver man inte veta hur länge man varit i (1) så låt oss anta att kedjan just kommit dit.

Antag att vi befinner oss i tillstånd (k) och just har kommit dit. Om vi med U_k betecknar tiden som nu är kvar till absorbtion har vi: $U_k = W_k + R_k$ där W_k är den tid som vi är kvar i tillstånd (k) och R_k är resten av tiden till absorbtion. För väntevärdena gäller:

$$E(U_k) = E(W_k) + E(R_k) \tag{3.1}$$

$E(W_k)$ får vi från intensitetsmatrisen och genom det utvidgade sättet att beräkna väntevärde får vi här också:

$$E(R_k) = P(k \to k-1)E(U_{k-1}) + P(k \to k+1)E(U_{k+1}).$$

Om vi sätter in detta i (3.1) och inför $b_k = E(U_k)$ får vi:

$$b_k = E(W_k) + P(k \to k-1)b_{k-1} + P(k \to k+1)b_{k+1}.$$

Sannolikheterna ovan får vi genom uthoppsmatrisen. (Denna får läsaren själv skriva upp!)

För $k = 1$ får vi: $b_1 = 1/13 + (5/13)b_0 + (8/13)b_2$ och för $k = 2$: $b_2 = 1/14 + (5/7)b_1 + (2/7)b_3$ men b_3 är förstås $= 0$ så vi behöver bara en ekvation till. Tar vi $k = 0$ så kommer bara de termer med som är definierade och vi får: $b_0 = 1/8 + 1b_1$. Löser vi detta ekvationssystem får vi $b_1 = 123/128$ (år) så nästan ett år får reparatörerna alltså i genomsnitt vänta på sin semester!

Alla dessa rekursiva problem går att lösa med hjälp av formelsamlingen fast då kommer själva metoden i bakgrunden vilket är mindre bra. Den formel för väntevärdet i det diskreta fallet som finns i FS (sid 9 högst upp) är för övrigt "hopdragen" vilket gör den svårare att förstå.

I vissa andra situationer kan dock formelsamlingen vara till stor hjälp, men man måste vara försiktig så att man tolkar det som står där rätt. Som exempel tar vi först "bollkedjan" i diskret tid. Säg att Cecilia har bollen och låt T_3 beteckna det antal kast det tar innan Cecilia åter har bollen sedan hon kastat den vidare. I detta antal inkluderar vi **både** "utflyktens" första och sista kast. Om vi numrerar tillstånden som förut kan det t ex se ut såhär: (3)→(1)→(2)→(2)→(2)row(1)→(3) där T_3 har värdet 6.

FS 14.1.3 ger $E(T_3) = 1/g_3 = 4.89$. FS 14.1.3 besvarar också frågan om hur många **gånger** Adam repektive Bertil i genomsnitt har bollen under dess utflykt från Cecilia. Vi får svaren $g_1/g_3 = 1.67$ resp. $g_2/g_3 = 2.22$. **Observera** att "tiden" som det talas om i 14.1.3 mäts i **steg**, så att "tiden mellan" betyder "antal steg mellan".

Också för "bollkedjan" i kontinuerlig tid kan FS användas för att direkt besvara liknande frågor. Antag att Cecilia har bollen och just kastar den ifrån sig. Hur länge dröjer det nu i genomsnitt innan Cecilia återfår bollen? Säg att vi räknar tiden i minuter. FS 14.2.3 ger väntevärdet för tiden mellan två **inträden** i tillstånd (3). Den blir $-1/(\boldsymbol{Q}(3,3) \cdot g_3)$. Här betecknar $\boldsymbol{Q}(3,3)$ elementet i $\boldsymbol{Q}$-matrisens tredje rad och tredje kolumn d v s $\boldsymbol{Q}(3,3) = -2$. Eftersom $g_3 = 0.56$ enligt tidigare beräkningar får vi $1/1.12 = 0.893$ minuter. För att få det sökta måste vi från detta dra bort den tid som Cecilia i genomsnitt har bollen d v s det sökta blir: $0.893 - 0.5 = 0.393$ minuter.

FS 14.2.3 ger också att Adam i genomsnitt har bollen i $-g_1/(\boldsymbol{Q}(3,3) \cdot g_3 = 0.107$ minuter under dess utflykt från Cecilia, medan samma genomsnitt för Bertil blir$= 0.286$ minuter. (Kontroll: $0.107 + 0.286 = 0.393$.)

Lägg märke till att vi **inte** behöver förutsätta att Cecilia har bollen "efter lång tid" för att få använda FS 14.1.3 och FS 14.2.3.

3.3 Oändliga kedjor i diskret tid

$$\leftrightarrow (0) \leftrightarrow (1) \leftrightarrow (2) \leftrightarrow (3) \leftrightarrow (4) \leftrightarrow (5) \leftrightarrow \cdots$$

Kedjan ovan är samma som vi hade tidigare men vi har infört 0 som tillståndsnummer. Detta är naturligt eftersom den praktiska bakgrunden kan vara ett "spel" där noll svarar mot "ruin". Sannolikheterna att "gå åt höger f o m (0)" är p med $0 < p < 1$ och övriga sannolikheter $= 1 - p$.

Antag att kedjan är i tillstånd (1). Vad är sannolikheten att den förr eller senare hamnar i (0)? Vi har sett tidigare att kedjan är ergodisk om $p < 0.5$ och om detta gäller så är den sökta sannolikheten självklart = ett. Låt oss anta att $p > 0.5$.

Nu finns det en sannolikhet > 0 att kedjan "försvinner ut i oändligheten". Den sannolikhet vi söker är därför < 1. (Slutsatsen är dock inte så självklar som den kan verka, men det skulle föra för långt att ge en korrekt matematisk motivering.)

Låt $(k)\Longrightarrow(j)$ beteckna en väg sådan att kedjan startar i (k) och går till (j) utan att ha passerat (j) tidigare d v s den kommer till (j) **för första gången**. Det vi söker är summan av sannolikheterna för alla de olika vägar som uppfyller $(1)\Longrightarrow(0)$. Låt oss beteckna det sökta med $P((1)\Longrightarrow(0))$. På grund av symmetrin är detta är lika med $P((2)\Longrightarrow(1))$ eller mera allmänt $= P((k+1)\Longrightarrow(k))$ för $k = 2, 3, 4, \ldots$

Inför nu $a_k = P((k)\Longrightarrow(0))$. Totala sannolikhetslagen ger:

$$a_1 = (1-p) + pa_2 \tag{3.2}$$

En godtycklig väg från (2) till (0) kan delas upp i två vägar, först vägen till dess man når (1) **för första gången** sedan resten. Detta ger

$$\begin{aligned} a_2 &= P((2)\Longrightarrow(0)) = P((2)\Longrightarrow(1)\Longrightarrow(0)) = (\text{Markov-egenskapen!}) \\ &= P((2)\Longrightarrow(1)) \cdot P((1)\Longrightarrow(0)) = (\text{symmetrin!}) = a_1 \cdot a_1 \end{aligned}$$

Sätter vi in $a_2 = a_1^2$ i (3.2) ovan får vi en andragradsekvation för a_1. Den har två lösningar: $a_1 = 1$ och $a_1 = p/(1-p)$. Det är den andra lösningen som gäller eftersom vi visste att $a_1 < 1$.

På analogt sätt kan man bestämma a_k för $k > 1$. Med samma beteckningar:

$$\begin{aligned} a_k &= P((k)\Longrightarrow(k-1)\Longrightarrow(k-2)\Longrightarrow\cdots\Longrightarrow(1)\Longrightarrow(0)) = (\text{Markov!}) \\ &= P((k)\Longrightarrow(k-1)) \cdot P((k-1)\Longrightarrow(k-2)) \cdots P((1)\Longrightarrow(0)) = (\text{symmetrin!}) \\ &= a_1 \cdot a_1 \cdots a_1 = a_1^k. \end{aligned}$$

Observera att det var viktigt att det fanns en symmetri. Visserligen gäller $a_2 = (1-p)a_1 + pa_3$ etc, men för varje ny ekvation vi lägger till, tillkommer också en ny obekant. Vi behövde alltså en "fristående" relation och en sådan fick vi genom symmetrin.

Antag återigen att kedjan är i tillstånd (1) men att vi nu istället söker **väntevärdet** för antal steg till dess vi når tillstånd (0) för första gången. För att väntevärdet ska bli ändligt krävs att kedjan är ergodisk d v s vi måste anta att $p < 0.5$ gäller.

Med "naturliga beteckningar" blir $E((1)\Longrightarrow(0))$ det som söks och på grund av symmetrin ser man att detta är lika med $E((2)\Longrightarrow(1))$ eller mera allmänt lika med $E((k+1)\Longrightarrow(k))$ för $k = 2, 3, 4, \ldots$

Inför nu $b_k = E((k)\Longrightarrow(0))$. Det som söks är alltså b_1. Samma väguppdelning som förut ger att

$$\begin{aligned} b_2 &= E((2)\Longrightarrow(0)) = E((2)\Longrightarrow(1)\Longrightarrow(0)) = (\text{enligt en väntevärdeslag}) \\ &= E((2)\Longrightarrow(1)) + E((1)\Longrightarrow(0)) = (\text{symmetrin!}) = b_1 + b_1 = 2b_1. \end{aligned}$$

Enligt det utvidgade beräkningssättet för väntevärde har vi $b_1 = (1-p)\cdot 1 + p(1+b_2)$ och sätter vi in $b_2 = 2b_1$ och löser ut b_1 får vi: $b_1 = 1/(1-2p)$. Mera allmänt kan man genom att dela upp vägen visa att $b_k = kb_1$ vilket rekommenderas som övning.

Vad händer nu då $p = 0.5$? Att kedjan inte är ergodisk det vet vi, men beräkningen av a_k ger nu $a_k = 1$ för $k = 1, 2, 3, \ldots$ d v s med sannolikhet ett kommer vi alltid förr eller senare till (0) var vi än startar. Å andra sidan ger beräkningen av b_k ett oändligt värde vilket betyder att väntevärdet för den tid det tar att komma till (0) alltid är oändligt var vi än startar. Som så ofta när oändligheten är inblandad kan man få resultat som ser motsägelsefulla ut.

3.4 Oändliga kedjor i kontinuerlig tid

Som exempel tar vi vår korvgubbe sedan han flyttat ut på landet. Vid ett tillfälle var det bara en person i systemet. Hur länge dröjer det nu i genomsnitt till dess systemet är tomt?

Med b_k definierad som i 3.2 och med samma beteckningar som när vi beräknade väntevärdet i 3.3 får vi:

$$\begin{aligned} b_2 &= E((2) \Longrightarrow (0)) = E((2) \Longrightarrow (1) \Longrightarrow (0)) = E((2) \Longrightarrow (1)) + E((1) \Longrightarrow (0)) \\ &= (\text{symmetrin!}) = b_1 + b_1 = 2b_1. \end{aligned}$$

D v s vi har $b_2 = 2b_1$.

Sambandet som vi använde då vi beräknade väntevärdet i 3.2 ger nu: $b_1 = E(W_1) + P(1 \to 2)b_2$ där $E(W_1) = 2$ enligt intensitetsmatrisen och $P(1 \to 2) = 1/3$ enligt uthoppsmatrisen. Se till att du kontrollerar dessa värden genom att sätta upp matriserna (i alla fall en del av dem, de är ju oändliga!) Vi får det sökta $= b_1 = 6$ minuter. Mera allmänt kan man genom att dela upp vägen visa att $b_k = kb_1$ för $k = 1, 2, 3, \ldots$