Förord

Detta kompendium behandlar grunderna för diskreta Markovprocesser i diskret och kontinuerlig tid. Ambitionen har varit att, åtminstone då tillståndsrummet är ändligt, ge en så matematisk fullständig beskrivning som möjligt, och ge bevis för de satser och påståenden som görs. Bevisen bygger oftast på enkla sannolikhetsteoretiska resonemang som gör att man skall kunna förstå och inse vad som händer i en Markovprocess. Ingen avancerad matematik behövs för att förstå bevisen, även om några av dem kan synas långa och litet komplicerade. Många av de satser och resultat som visas kan man förstå intuitivt, och det är värdefullt att få en sådan förståelse.

De grafteoretiska egenskaperna hos Markovprocesser, som har så stor betydelse i konvergensteorin, har fått ett eget kapitel.

Markovprocesser tillämpas i ett otal sammanhang. En viktig tillämpning är köteorin, även om köteori också omfattar modeller som inte kan behandlas med Markovteknik. Kompendiets köteorikapitel omfattar sådana icke-markovska modeller även om tyngdpunkten ligger på modeller, som kan studeras med Markovteknik.

Finstilt skrivna stycken behandlar ”speciella” frågor, ofta av mera matematisk natur. Dessa stycken är främst skrivna med tanke på de särskilt matematiskt intresserade och på de som i ”framtiden” kommer att tillämpa Markovteori och köteori i situationer som går utanför vad vi kan behandla inom ramen för kursen. Den som så önskar kan ögna igenom dessa kursivt utan att därmed förlora sammanhanget.

Problemsamlingen är till viss del baserad på en problemsamling sammanställd av Mikael Möller och Gunnar Englund.

Författarna vill tacka Torgny Lindvall, Chalmers tekniska högskola, för många värdefulla kommentarer som bakats in i kompendiet.

F.o.m. utgåvan Höstterminen 2006 använder vi beteckningen Exp(λ) för en exponentialfördelning med *intensitet* λ. I tidigare utgåvor var parametern väntevärdet.

Jan Enger   Jan Grandell

Innehåll

Kapitel 1

Inledning

I många sammanhang vill man beskriva ett slumpmässigt skeende i tid eller rum, t.ex. vinstutveckling vid successiva spelomgångar eller antalet inkommande samtal till en växel från tiden 0 till tiden t. Vid varje tidpunkt t beskriver en stokastisk variabel $X(t)$ det fenomen man intresserar sig för. Oftast är man då inte bara intresserad av $X(t)$ vid en fix tidpunkt utan framförallt av uppträdandet under en hel tidsperiod. Detta uppträdande beskrivs då av familjen av stokastiska variabler, $\{X(t); t \in T\}$, där T är ett ändligt eller oändligt tidsintervall. Man säger att $\{X(t); t \in T\}$ är en *stokastisk process.*

Definition 1.1 *En familj av stokastiska variabler* $\{X(t);\, t \in T\}$ *kallas en* stokastisk process.

För det mesta svarar parametern t mot ”tid”. Mängden T kallas *parameterrummet.* Typiska exempel är:

$T = \{0, 1, \ldots\}$, d.v.s. $T = \boldsymbol{N}$, då vi talar om *diskret tid*;

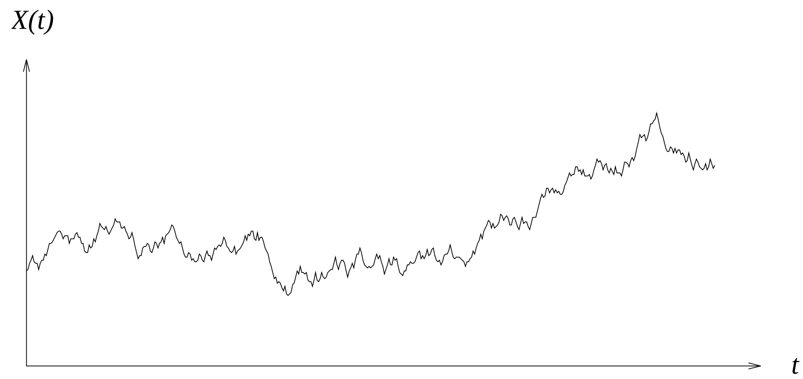
$T = [0, \infty)$, d.v.s. $T = \boldsymbol{R}_+$, då vi talar om *kontinuerlig tid.*

Teorin för stokastiska processer är en betydelsefull del av den matematiska statistiken och har många praktiska tillämpningar inom bl.a. naturvetenskap och teknik. Den används för att t.ex. studera populationers tillväxt, hur radioaktivt sönderfall sker, hur trafik- och kommunikationssystem uppför sig, eller för att studera olika slag av ekonomiska tidsserier. Andra viktiga tillämpningsområden är tillförlitlighetsteknik, signalbehandling, statistisk mekanik och köteori.

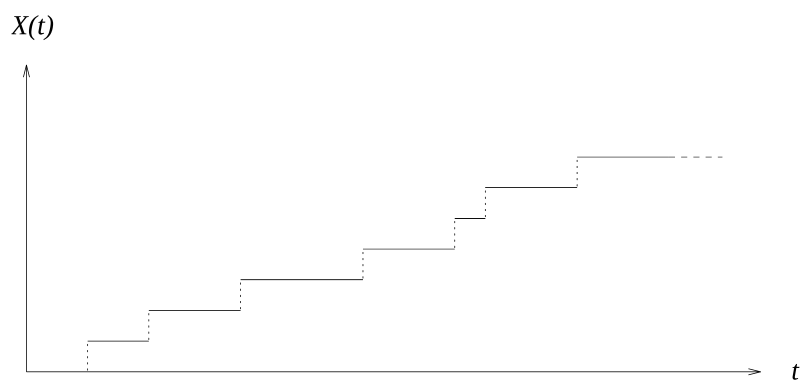
En stokastisk process i diskret tid kallas ibland en *stokastisk följd*, *kedja* eller *tidsserie.*

Mängden av värden, $\boldsymbol{E}$, som $X(t)$ kan anta kallas *tillståndsrummet.* Liksom för ”tiden” skiljer vi ofta på diskret (ändligt eller uppräkneligt) och kontinuerligt tillståndsrum.

Exempel 1.1 Tjockleken hos papper i en pappersmaskin, som studeras under ett tidsintervall $[0, 24)$. Låt $X(t)$ vara tjockleken vid tiden t. Exemplet illustreras i figur 1.1. $\square$



Figur 1.1: Tjockleken hos papper i en pappersmaskin.

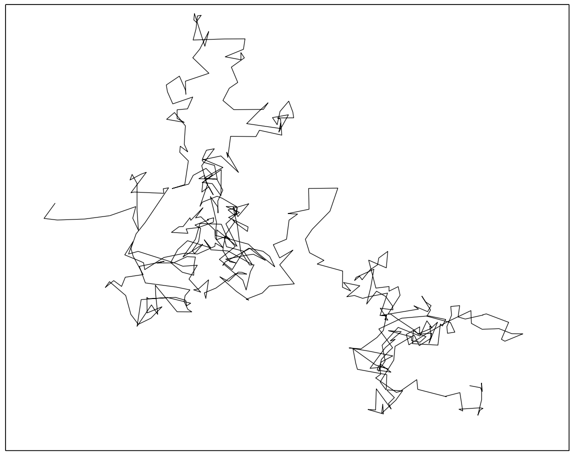


Figur 1.2: Antal kunder som besöker en affär.

Exempel 1.2 Antal kunder som besöker en affär. Låt $X(t)$ vara antalet kunder som besöker en affär under tidsintervallet $[0, t)$. Exemplet illustreras i figur 1.2. □

Exempel 1.3 (Brownsk rörelse) En liten partikels rörelse i ett prov studeras. Partikeln tycks utföra slumpmässiga och oregelbundna rörelser. Den engelske botanisten Brown studerade (1827) pollenkorn som rörde sig på detta slumpmässiga sätt. Rörelsen fick senare sin förklaring av Einstein. Rörelsen beskrivs av en s.k. Wienerprocess. I två dimensioner beskrivs läget vid tiden t av $\{(X(t), Y(t)); t \geq 0\}$ där $\{X(t); t \geq 0\}$ och $\{Y(t); t \geq 0\}$ är oberoende stokastiska processer för vilka gäller att ökningarna i de disjunkta tidsintervallen $[t_1, t_2]$ och $[t_3, t_4]$, $X(t_2) - X(t_1)$ respektive $X(t_4) - X(t_3)$ är normalfördelade och oberoende och motsvarande för Y-processen. □

Det som gör studiet av processer intressant, är beroendet mellan $X(t)$ och $X(s)$ för $t, s \in T$. Typiska sådana "beroendetyper" är stationära processer och Markovprocesser.



Figur 1.3: Brownsk rörelse

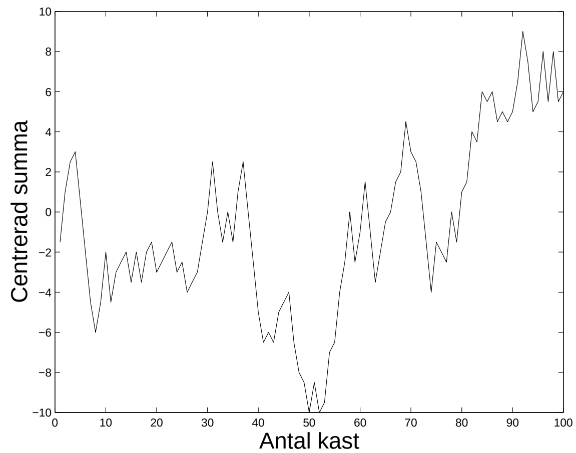
Markovberoendet innebär, populärt uttryckt, att processens framtida utseende, beror av nuet men ej av den tidigare historien; Markovprocessen saknar minne. ”Givet nuet är det förgångna och det framtida oberoende” är ett annat sätt att uttrycka Markovegenskapen.

Exempel 1.4 (Slumpvandring) Antag att en tärning kastas upprepade gånger och att kastens utfall är oberoende av varandra. Låt S_n vara ögonsumman efter n kast. Fördelningen för ögonsumman efter $n+1$ kast, S_{n+1} kan lätt beräknas om vi vet vad S_n är, och värdena av S_k för kastomgångar k, ”tider”, $k < n$ är irrelevanta i sammanhanget; $\{S_n; n \in N\}$ är en Markovkedja. Givet nuet S_n, är de framtida värdena på S_k helt oberoende av de tidigare ögonsummorna upp till omgång n. Om vi centrerar S_n och bildar $Y_n = S_n - 3.5n$ så har Y_n väntevärde 0 för alla n. Figur 1.4 är en simulering av processen $\{Y_n\}$ för $n = 1, 2, \ldots, 100$. □

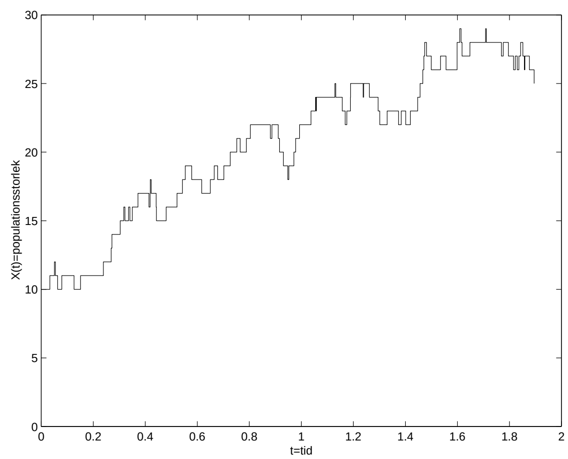
Exempel 1.5 (Populationstillväxt) I en population föder varje individ, oberoende av varandra, efter en stokastisk tid nya individer. Individernas livslängder är slumpmässiga, dvs en individ dör efter en stokastisk tid. Låt $X(t)$ vara antalet invider i populationen vid tid t. $\{X(t); t \geq 0\}$ är då en stokastisk process i kontinuerlig tid. En realisering av processen kan se ut som i figur 1.5. □

Vi kommer att studera sådana *födelse-dödsprocesser* senare i kompendiet.

Exempel 1.6 (Kortblandning) En kortlek med 52 kort kan ordnas på 52! olika sätt. En kortblandning avser att göra ordningen mer slumpmässig. Blandningsförfarandena kan utföras på många olika sätt, t.ex. medelst följande enkla och primitiva metod. Det översta kortet tas och sätts in slumpmässgt i leken.



Figur 1.4: Centrerad följd av tärningskast



Figur 1.5: Populationstillväxt

En ny ordning, permutation, av korten har då åstadkommits. Låt X_n vara ordningen av korten efter n sådana blandningar. Man inser att ordningen efter n blandningar beror på ordningen efter $n-1$ blandningar, men att de tidigare ordningarna då inte spelar någon roll; $\{X_n; n \geq 1\}$ är en Markovkedja. Tillstånden är inte tal utan permutationer. Antalet tillstånd är visserligen ändligt, men astronomiskt stort, $52! \approx 8 \cdot 10^{67}$. Vi skall studera Markovkedjor med ändligt antal tillstånd i detta kompendium, men för så stora tillståndsrum som i detta exempel, blir de metoder som beskrivs oftast omöjliga att använda.

Den formella definitionen av Markovkedja ger vi i kapitel 3.

Kapitel 2

Betingning

Markovprocesser definieras i termer av betingade sannolikheter, på ett sätt som leder till att processerna ”saknar minne”. Betingade sannolikheter spelar därför en viktig roll i Markovteorin. Vi påminner om definitionen.

Definition 2.1 *Låt A och B vara två händelser och antag $P(B) > 0$. Då är den* betingade sannolikheten *för A givet B*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Betingade sannolikheter kan användas för att beräkna sannolikheten för snittet av ett antal händelser.

Sats 2.1 *Låt $A_1, A_2, \ldots, A_n$ vara n händelser. Då gäller*

$$P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \quad (2.1)$$

Bevis. Lämnas till läsaren. □

En fördelning med ”glömskeegenskap” är exponentialfördelningen, vilken är fundamental vid behandling av Markovprocesser i kontinuerlig tid.

Den stokastiska variabeln X är *exponentialfördelad*, $\mathrm{Exp}(\lambda)$, om dess täthetsfunktion är

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Fördelningsfunktionen är $F(x) = 1 - e^{-\lambda x}$. Vi säger också att X är exponentialfördelad med *intensitet* λ. Väntevärdet för X är $1/\lambda$ och variansen är $1/\lambda^2$.

Med en terminologi hämtad från tillförlitlighetsteorin kallas funktionen $R(x)$ definierad

$$R(x) = P(X > x)$$

överlevnadsfunktionen till X. Överlevnadsfunktionen är helt enkelt ”komplementet” till fördelningsfunktionen, $R(x) = 1 - F(x)$. Om X är exponentialfördelad med intensitet λ är överlevnadsfunktionen $R(x) = P(X > x) = 1 - F(x) = e^{-\lambda x}$.

Vi har följande karaktäriseringssats för exponentialfördelningen.

Sats 2.2 *Låt X vara en icke-negativ stokastisk variabel X sådan att $P(X > x) > 0$ för alla x.Då är X exponentialfördelad om och endast om det för alla $x \geq 0$ och $y \geq 0$ gäller att*

$$P(X > x + y \mid X > y) = P(X > x) \tag{2.2}$$

eller ekvivalent

$$R(x + y) = R(x)R(y) \tag{2.3}$$

där R är överlevnadsfunktionen till X.

Glömskeegenskapen ges av det första villkoret (2.2) i satsen. Tolka X som en livslängd av en komponent. Likheten (2.2) utsäger då att en komponent som fungerar vid tidpunkt y har samma sannolikhet att fungera ytterligare x tidsenheter som en ny komponent. I tillförlitlighetstermer uttrycks detta som att en gammal komponent är lika bra som en ny, de har samma sannolikheter att leva ytterligare ett givet antal tidsenheter.

Bevis. Ekvivalensen av villkoren följer från definitionen av betingad sannolikhet.

Antag först att X är exponentialfördelad. Då gäller $R(x + y) = e^{-\lambda(x+y)} = e^{-\lambda x}e^{-\lambda y} = R(x)R(y)$.

För att bevisa omvändningen antar vi att $R(x + y) = R(x)R(y)$ för alla x och $y > 0$. Antag n är ett godtyckligt heltal. Vi får då, för varje reellt tal a, $R(na) = R((n - 1)a + a) = R((n - 1)a)R(a) = R((n - 2)a)R(a)R(a)$
$=$ (upprepad användning) $= R(a)^n$

För $a = 1/n$ erhålls $R(1) = R(1/n)^n$ eller $R(1/n) = R(1)^{1/n}$. Härav inses att om m och n är heltal, är $R(m/n) = R(1/n)^m = R(1)^{m/n}$. Eftersom $R(1) \leq 1$ är $R(1) = e^{-\lambda}$ för något $\lambda \geq 0$, och $R(m/n) = e^{-\lambda m/n}$, dvs $R(x) = e^{-\lambda x}$ om x är rationellt.

Om x är godtycklig utnyttjar vi att fördelningsfunktionen, och således även överlevnadsfunktionen, är högerkontinuerlig. Låt $\{x_n\}$ vara en följd rationella tal sådana att $x_n \to x$ från höger då $n \to \infty$. Vi erhåller då $R(x) = \lim_{n\to\infty} R(x_n) = \lim_{n\to\infty} e^{-\lambda x_n} = e^{-\lambda x}$, dvs X är exponentialfördelad med intensitet λ.

Fallet $\lambda = 0$ kan uteslutas, ty i så fall skulle vi ha $R(x) = 1$ alla x, vilket är orimligt. $\square$

En annan viktig egenskap hos exponentialfördelningen ges i följande sats.

Sats 2.3 *Låt $X_1, X_2, \ldots, X_n$ vara oberoende stokastiska variabler där X_i är exponentialfördelad med intensitet λ_i och sätt $Y = \min\{X_1, X_2, \ldots X_n\}$. Då är Y exponentialfördelad med intensitet $\lambda = \sum_i \lambda_i$.*

Bevis. Vi har

$$R_Y(x) = P(Y > x) = P(\min\{X_1, X_2, \ldots, X_n\} > x)$$
$$= P(X_1 > x, X_2 > x, \ldots, X_n > x) = \prod_{i=1}^{n} P(X_i > x) = \prod_{i=1}^{n} e^{-\lambda_i x} = e^{-x \sum_{i=1}^{n} \lambda_i},$$

d.v.s. Y är exponentialfördelad med intensitet $\sum_{i=1}^{n} \lambda_i$. $\square$

För första gången (ffg) fördelningen och den geometriska (Ge) fördelningen är diskreta motsvarigheter till exponentialfördelningen. Om är $X \in \text{ffg}(p)$ så är $X - 1 \in \text{Ge}(p)$, så dessa fördelningar är i princip ekvivalenta. Om X är antalet gånger ett visst försök upprepas tills en händelse A inträffar (oberoende försök) är X ffg(p), där p är sannolikheten att A inträffar i ett enskilt försök. Sannolikhetsfunktionen för $X \in \text{ffg}(p)$ ges av

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

Väntevärdet och variansen är

$$E(X) = \frac{1}{p} \qquad V(X) = \frac{1-p}{p^2}. \tag{2.4}$$

Vi påminner också om *satsen om total sannolikhet.*

Sats 2.4 *Låt $H_1, H_2, \ldots$ vara en följd av ändligt många eller uppräkneligt många parvis disjunkta händelser vars union är hela utfallsrummet, d.v.s. vid ett försök inträffar alltid en och endast en av händelserna H_i. Låt A vara en godtycklig händelse. Då gäller*

$$P(A) = \sum_{i=1}^{\infty} P(A \mid H_i)P(H_i). \tag{2.5}$$

Om H_i är händelsen $\{Y = i\}$, där Y är en diskret stokastisk variabel och A händelsen $\{X = k\}$ kan likheten ovan skrivas

$$P(X = k) = \sum_i P(X = k \mid Y = i)P(Y = i), \tag{2.6}$$

där summationen sker över alla i som Y kan anta.

En viktig generalisering av denna sats är satsen om total förväntan. Vi definierar först begreppet betingat väntevärde.

Definition 2.2 *Låt X vara en diskret stokastisk variabel och B en händelse. Med det* betingade väntevärdet *för X givet B menas*

$$E(X \mid B) = \sum_k kP(X = k \mid B) \tag{2.7}$$

där summationen sker över alla värden k som X kan anta.

Om X skulle vara oberoende av B, d.v.s. om händelserna $\{X = k\}$ är oberoende av B för alla k, ser man lätt att $E(X \mid B) = E(X)$.

Man kan definiera betingat väntevärde för en kontinuerlig stokastisk variabel på liknande sätt.

Vi kan nu formulera *satsen om total förväntan.*

Sats 2.5 *Låt $H_1, H_2, \dots$ vara en följd av ändligt eller uppräkneligt många parvis disjunkta händelser vars union är hela utfallsrummet, d.v.s. vid ett försök inträffar alltid en och endast en av händelserna H_i. Låt X vara en stokastisk variabel med ändligt väntevärde. Då gäller*

$$E(X) = \sum_{i=1}^{\infty} E(X \mid H_i)P(H_i). \tag{2.8}$$

Bevis. Vi har

$$\begin{aligned} E(X) &= \sum_k kP(X = k) = (\text{satsen om total sannolikhet}) \\ &= \sum_k k \sum_i P(X = k \mid H_i)P(H_i) = (\text{omkastning av summationsordning}) \\ &= \sum_i P(H_i) \sum_k kP(X = k \mid H_i) = \sum_i E(X \mid H_i)P(H_i) \end{aligned} \tag{2.9}$$

vilket skulle visas. □

Sats 2.6 *Låt $X_1, X_2, \dots$ vara en följd av stokastiska variabler alla med väntevärde m. Låt N vara stokastisk variabel som endast antar icke-negativa heltalsvärden och antag att för alla heltal n, de stokastiska variablerna $X_1, X_2, \dots, X_n$ är oberoende av händelsen $\{N = n\}$ och att $E(N) < \infty$. Sätt $S_N = X_1 + X_2 + \cdots + X_N$, ($S_0 = 0$ definitionsmässigt). Då gäller att*

$$E(S_N) = E(N)m.$$

Bevis. Satsen om total förväntan ger att

$$E(S_N) = \sum_n E(S_N \mid N = n)P(N = n).$$

Men

$$\begin{aligned} E(S_n \mid N = n) &= E(X_1 + X_2 + \cdots + X_n \mid N = n) \\ &= (\text{oberoendet!}) = E(X_1 + X_2 + \cdots + X_n) = nm. \end{aligned} \tag{2.10}$$

Alltså fås $E(S_N) = \sum_n mnP(N = n) = m \sum_n nP(N = n) = mE(N)$. □

Ofta är N oberoende av de stokastiska variablerna $X_1, X_2, \dots$, men notera att detta inte är ett nödvändigt villkor i satsen.

Exempel 2.1 I en population är antalet avkommor till en individ Ge(0.2). Antalet avkommor från de olika inviderna är oberoende. Om vi låter S_N vara antalet barnbarn till en individ kan S_N skrivas $S_N = X_1 + X_2 + \cdots + X_N$ där N är antalet barn till individen, och $X_1, X_2, \cdots X_N$ dessa barns antal avkommor. I detta fall är N och X_i-variablerna oberoende, alla Ge(0.2), och vi har att $E(S_N) = E(N)E(X_i) = 4 \cdot 4 = 16$. □

Kapitel 3

Diskreta Markovkedjor, grundläggande egenskaper

3.1 Grundläggande begrepp

Vi skall betrakta processer $\{X_n; n = 0, 1, \dots\}$ i diskret tid, d.v.s. som i tidpunkterna $0, 1, 2, \dots$ förflyttar sig inom ett ändligt tillståndsrum $\boldsymbol{E} = \{i_k, k = 1, 2, \dots, N\}$ eller uppräkneligt tillståndsrum $\boldsymbol{E} = \{i_k, k = 1, 2, \dots\}$. Tillstånden i_k behöver inte vara tal, men för enkelhetens skull kommer vi ofta att anta (och detta är egentligen ingen inskränkning) att $\boldsymbol{E}$ är en heltalsmängd, t.ex. $\boldsymbol{E} = \{1, 2, \dots, N\}$ eller $\boldsymbol{E} = \{1, 2, \dots\}$.

Definition 3.1 $\{X_n; n \geq 0\}$ *är en* Markovkedja *om*

$$P(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, ..., X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

för alla n *och tillstånd* $i_0, i_1, \dots, i_{n+1}$.

Definitionen innebär att en Markovkedja saknar minne i den meningen att fördelningen för ett tidssteg framåt beror av kedjans läge ”nu”, men inte av dess tidigare historia.

Anm: I definitionen ovan kan vänsterledet vara odefinierat på grund av att den givna händelsen har sannolikhet 0, medan högerledet kan vara väldefinierat. I sådana fall definierar vi vänsterledet att vara lika med högerledet.

Genom att använda sats 2.1 på sidan 5 och Markovegenskapen erhåller vi

$$\begin{aligned}
P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) \\
&= P(X_0 = i_0)P(X_1 = i_1 \mid X_0 = i_0)P(X_2 = i_2 \mid X_0 = i_0, X_1 = i_1) \\
&\qquad \cdots P(X_n = i_n \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) \\
&= P(X_0 = i_0)P(X_1 = i_1 \mid X_0 = i_0)P(X_2 = i_2 \mid X_1 = i_1) \cdots \\
&\qquad \cdots P(X_n = i_n, \mid X_{n-1} = i_{n-1}). \qquad (3.1)
\end{aligned}$$

I definitionen förutsätter vi att X_k är givet för alla tidpunkter upp till $n - 1$, och ger den betingade sannolikheten ett steg framåt. Man kan kanske intuitivt inse att detta medför att sannolikhetsfördelningen för kedjan under ett godtyckligt antal steg framåt, bestäms av värdet på X_k för den sist kända tidpunkten.

Formellt kan man visa detta påstående på följande sätt. För alla $i_1, i_2, \ldots i_k$ i $\boldsymbol{E}$, och $n_1 < n_2 < \cdots < n_k$ gäller att

$$\begin{aligned} &P(X_{n_k} = i_k \mid X_{n_1} = i_1, X_{n_2} = i_2, \ldots, X_{n_{k-1}} = i_{k-1}) \\ &\qquad = P(X_{n_k} = i_k \mid X_{n_{k-1}} = i_{k-1}) \end{aligned} \quad (3.2)$$

och för alla mängder $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots \boldsymbol{A}_m$ i $\boldsymbol{E}$, och $n_1 < n_2 < \cdots < n_{k-1} < n_k < \cdots < n_m$ gäller

$$\begin{aligned} &P(X_{n_m} \in \boldsymbol{A}_m, \ldots, X_{n_k} \in \boldsymbol{A}_k \mid X_{n_1} \in \boldsymbol{A}_1, X_{n_2} \in \boldsymbol{A}_2, \ldots, X_{n_{k-2}} \in \boldsymbol{A}_{k-2}, X_{n_{k-1}} = i_{k-1}) \\ &\qquad = P(X_{n_m} \in \boldsymbol{A}_m, \ldots, X_{n_k} \in \boldsymbol{A}_k \mid X_{n_{k-1}} = i_{k-1}). \end{aligned} \quad (3.3)$$

Händelsen $X_{n_m} \in \boldsymbol{A}_m$ kan till exempel vara händelsen $X_{n_m} \neq i_m$, $\boldsymbol{A}_m = \boldsymbol{E} \setminus \{i_m\}$, d.v.s. mängden $\boldsymbol{E}$ utom elementet $\{i_m\}$. Notera att man inte kan ersätta den sist givna händelsen $X_{n_{k-1}} = i_{k-1}$ med $X_{n_{k-1}} \in \boldsymbol{A}_{k-1}$.

Vi inför nu övergångssannolikheterna p_{ij} och kommer därvid endast att betrakta så kallade tidshomogena Markovkedjor.

Definition 3.2 *Övergångssannolikheterna p_{ij} i en tidshomogen Markovkedja definieras av*

$$p_{ij} = P(X_n = j \mid X_{n-1} = i) \quad i, j \in \boldsymbol{E}$$

d.v.s. p_{ij} är sannolikheten att gå från i till j i ett tidssteg.

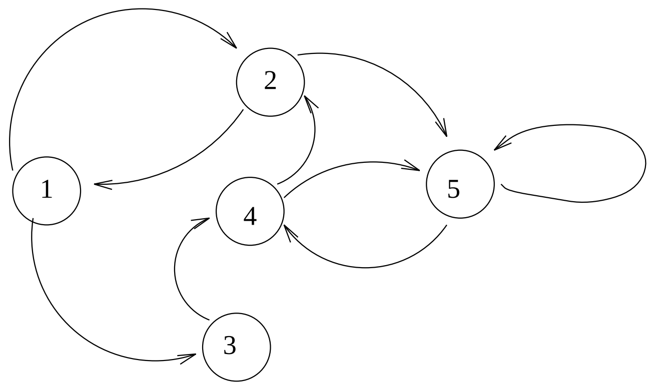
I det allmänna fallet tillåter man dessa övergångssannolikheter att bero på tidpunkten n. Vi kommer kommer dock endast att studera tidshomogena Markovkedjor.

En tidshomogen Markovkedja kan illustreras av en övergångsgraf. Det är en riktad graf där tillstånden finns representerade som noder och där riktade pilar anger om det är möjligt att gå från ett tillstånd direkt till ett annat. Vid pilarna är det vanligt att ange övergångssannolikheterna.

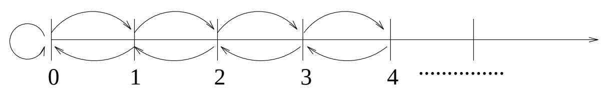
Figuren 3.1 illustrerar att det är möjligt att gå direkt från 1 till 2 eller 3, från 2 till 1 eller 5, från 3 till 4, från 4 till 2 eller 5 samt från 5 till 4 eller 5.

Exempel 3.1 (Slumpvandring på de naturliga talen) Betrakta en partikel som vandrar mellan heltalspunkterna $0, 1, 2 \ldots$. Partikeln går antingen till höger eller till vänster, utom om den befinner sig i punkten 0, då den i nästa tidpunkt antingen är kvar i 0 eller går till höger. Denna slumpvandring kan illustreras av figuren 3.2

□



Figur 3.1: Övergångsgraf



Figur 3.2: Slumpvandring på positiva heltalen

Definition 3.3 *Med* övergångsmatrisen $\boldsymbol{P}$ *menas matrisen* $(p_{ij})_{i,j\in\boldsymbol{E}}$ *av övergångssannolikheter*

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{3.4}$$

Elementet i i:te raden, j:te kolumnen i en övergångsmatris är alltså sannolikheten att gå från tillstånd i till tillstånd j. Är tillståndsrummet oändligt, blir matrisen oändlig.

Notera också att summan av alla elementen i en rad är 1, eftersom man med sannolikhet 1 går från ett tillstånd till något annat, $\sum_j p_{ij} = 1$.

Sannolikheten att gå från i till j i n steg betecknar vi $p_{ij}^{(n)}$ och övergångsmatrisen av ordning n definierar vi naturligtvis som

$$\boldsymbol{P}^{(n)} = \begin{pmatrix} p_{11}^{(n)} & p_{12}^{(n)} & p_{13}^{(n)} & \cdots \\ p_{21}^{(n)} & p_{22}^{(n)} & p_{23}^{(n)} & \cdots \\ p_{31}^{(n)} & p_{32}^{(n)} & p_{33}^{(n)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{3.5}$$

På grund av tidshomogeniteten är $P(X_{m+n} = j \mid X_m = i) = p_{ij}^{(n)}$ för alla m.

För $n = 0$ definierar vi $p_{ij}^{(0)} = 1$ om $i = j$ och $p_{ij}^{(0)} = 0$ om $i \neq j$. Vi erhåller då $\boldsymbol{P} = \boldsymbol{I}$, enhetsmatrisen.

Definition 3.4 Startfördelningen *är fördelningen för* X_0 *och den anges av vektorn*

$$\boldsymbol{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, p_3^{(0)}, \dots)$$

där $p_k^{(0)} = P(X_0 = k)$*. Fördelningen för* X_n *anges av vektorn*

$$\boldsymbol{p}^{(n)} = (p_1^{(n)}, p_2^{(n)}, p_3^{(n)}, \dots)$$

där $p_k^{(n)} = P(X_n = k)$*.*

Exempel 3.2 Andrei Andreevich Markov var en rysk matematiker som levde 1857-1918. Han använde Pusjkins versroman Eugene Onegin till att uppfinna Markovkedjan. I denna versroman följs på ryska en vokal av en konsonant i 87 % av alla fall och en konsonant följs av en konsonant i 34 % av alla fall. Betraktar man "vokal" och "konsonant" som två olika tillstånd i en Markovkedja kan man ställa upp följande övergångsmatris för följd av vokaler och konsonanter i ryskan.

$$\boldsymbol{P} = \begin{pmatrix} 0.13 & 0.87 \\ 0.66 & 0.34 \end{pmatrix}$$

Eftersom sannolikheten är 0.87 att en vokal följs av en konsonant är sannolikheten 0.13 att en vokal följs av en vokal. Det ger oss övergångssannolikheten för övergång vokal till vokal som är det första elementet i matrisen. På samma sätt erhåller vi de övriga elementen. Konsonanter och vokaler blandas alltså inte slumpmässigt. Å andra sidan utgörs inte heller en skriven text av en Markovkedja, nästa tecken beror också i hög grad av bokstäver före den som ligger till vänster. Vissa aspekter i språket kan dock studeras, bl.a. kan språkigenkänning göras med hjälp av sådana övergångsmatriser.

Exempel 3.3 (Ehrenfests urnmodell) Antag att det finns sammanlagt N molekyler i två kommunicerande kärl A och B. Vid varje tidpunkt tas en molekyl på måfå och flyttas till andra kärlet. Låt X_n vara antalet partiklar i kärl A efter n dragningar. Om nu i partiklar befinner sig i A kommer med sannolikhet i/N någon av dessa att väljas och flyttas över till B, dvs med sannolikhet i/N är $X_{n+1} = i-1$. Med sannolikhet $1-i/N$ kommer en partikel ur B att väljas och vid nästa tidpunkt finns då $i+1$ partiklar i A. Om $i = 0$ eller $i = N$ befinner sig med sannolikhet 1 vid nästa tidpunkt 1 respektive $N-1$ partiklar i A. Därför är det lätt att inse att $\{X_n; n \geq 0\}$ är en Markovkedja med övergångsmatris

$$\boldsymbol{P} = \begin{array}{c} \\ \mathit{0} \\ \mathit{1} \\ \mathit{2} \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{ccccccc} \mathit{0} & \mathit{1} & \mathit{2} & \mathit{3} & \dots & N-\mathit{1} & N \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1/N & 0 & 1-1/N & 0 & \dots & 0 & 0 \\ 0 & 2/N & 0 & 1-2/N & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \end{array}.$$

Vi har markerat tillstånden vid övergångsmatrisen. □

Exempel 3.4 (Slumpvandring med absorberande barriärer) En partikel vandrar mellan heltalspunkterna $0, 1, 2 \dots, N$ och hoppar till höger med sannolikhet p och till vänster med sannolikhet $q = 1 - p$, utom i ändpunkterna 0 och N, där den förblir med sannolikhet 1. Låt X_n vara partikelns läge vid tidpunkt n. Övergångsmatrisen är

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & \dots & 0 & 0 & 0 \\ 0 & q & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

där första raden och kolonnen representerar punkten 0, den andra raden och kolonnen punkten 1 o.s.v.

Slumpvandringen kan representera en spelares totalvinst vid upprepade spel, vid vilket hon i en spelomgång vinner en krona med sannolikhet p och förlorar en krona med sannolikhet q. Vi skall senare beräkna sannolikheten för ruin, d.v.s. att hamna i tillstånd 0, givet att man startar i godtyckligt tillstånd. □

Vi skall ägna oss åt och mer eller mindre fullständigt besvara följande problem:

1. Hur skall fördelningen för X_n beräknas?
2. Ange villkor för att fördelningen för X_n skall konvergera då tiden n går mot oändligheten, och beräkna i så fall gränsfördelningen.
3. Om kedjan har absorberande tillstånd, vad är sannolikheten att hamna i ett sådant, och hur lång tid tar det?

3.2 Fördelningen för X_n

Vi visar först

Sats 3.1 (Chapman-Kolmogorov)

a) $p_{ij}^{(m+n)} = \sum_{k \in \boldsymbol{E}} p_{ik}^{(m)} p_{kj}^{(n)}$

b) $\boldsymbol{P}^{(m+n)} = \boldsymbol{P}^{(m)} \boldsymbol{P}^{(n)}$

c) $\boldsymbol{P}^{(n)} = \boldsymbol{P}^n$

d) $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)} \boldsymbol{P}^{(n)} = \boldsymbol{p}^{(0)} \boldsymbol{P}^n$.

Bevis. a) Vi gör ett typiskt Markovresonemang. Vi skall beräkna sannolikheten att kedjan går från i till j i $m+n$ steg. Det kan ske genom att den går från i till tillståndet k i m steg och sedan därifrån till j i n steg. Om vi summerar sannolikheterna för alla dessa möjligheter, får vi den totala sannolikheten

$p_{ij}^{(m+n)}$. Vi erhåller

$$
\begin{aligned}
p_{ij}^{(m+n)} &= P(X_{m+n} = j \mid X_0 = i) = \sum_{k \in \boldsymbol{E}} P(X_m = k, X_{m+n} = j \mid X_0 = i) \\
&= \sum_{k \in \boldsymbol{E}} P(X_m = k \mid X_0 = i) P(X_{m+n} = j \mid X_0 = i, X_m = k) \\
&= (\text{Markovegenskapen}) = \sum_{k \in \boldsymbol{E}} P(X_m = k \mid X_0 = i) P(X_{m+n} = j \mid X_m = k) \\
&= \sum_{k \in \boldsymbol{E}} p_{ik}^{(m)} p_{kj}^{(n)}. \qquad (3.6)
\end{aligned}
$$

b) Högerledet i a) är ingenting annat än resultatet vid en matrismultiplikation! Det är i:te raden i $\boldsymbol{P}^{(m)}$ multiplicerat med j:te kolumnen i $\boldsymbol{P}^{(n)}$, varav fås likheten i b).

c) Vi har

$$
\begin{aligned}
\boldsymbol{P}^{(n)} &= \boldsymbol{P}^{(n-1+1)} = \boldsymbol{P}^{(n-1)} \boldsymbol{P}^{(1)} = \boldsymbol{P}^{(n-1)} \boldsymbol{P} = \boldsymbol{P}^{(n-2)} \boldsymbol{P} \boldsymbol{P} \\
&= \boldsymbol{P}^{(n-2)} \boldsymbol{P}^2 = \cdots = \boldsymbol{P}^n. \qquad (3.7)
\end{aligned}
$$

d) Vi har

$$
\begin{aligned}
p_j^{(n)} &= P(X_n = j) = (\text{satsen om total sannolikhet}) \\
&= \sum_{i \in \boldsymbol{E}} P(X_0 = i) P(X_n = j \mid X_0 = i) = \sum_{i \in \boldsymbol{E}} p_i^{(0)} p_{ij}^{(n)}. \qquad (3.8)
\end{aligned}
$$

Detta sista uttryck är multiplikation av vektorn $\boldsymbol{p}^{(0)}$ med j:te kolumnen i $\boldsymbol{P}^{(n)}$. Det ger den första likheten i d). Den andra följer av c). □

Exempel 3.5 En Markovkedja har följande övergångsmatris

$$
\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.4 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}.
$$

Vi får

$$
\boldsymbol{P}^2 = \begin{pmatrix} 0.70 & 0.12 & 0.18 \\ 0.45 & 0.16 & 0.39 \\ 0.45 & 0.12 & 0.43 \end{pmatrix}.
$$

Om kedjan startar i tillstånd 2 är alltså sannolikheten att efter två steg befinna sig i tillstånd 3 lika med 0.39. Om kedjan startar i de tre tillstånden med samma sannolikhet, ges sannolikheterna att efter två steg befinna sig i de olika tillstånden av vektorn

$$
(1/3, 1/3, 1/3)\boldsymbol{P}^2 = (1.6, 0.4, 1.0)/3 = (0.533, 0.133, 0.333)
$$

□

Exempel 3.6 (Slumpmässig beläggning) Vid tidpunkter $n = 1, 2, \ldots$ kastas bollar så att en boll hamnar slumpmässigt och med lika sannolikhet i ett av totalt N fack. Vid tidpunkt $n = 0$ är alla facken tomma. Sätt X_n = antal icke-tomma (belagda) fack vid tidpunkt n. Man inser att $\{X_n; n \geq 0\}$ är en Markovkedja med $X_0 = 0$ och $\boldsymbol{P} = (p_{ij})_{(N+1)\times(N+1)}$ övergångsmatris med övergångssannolikheter

$$p_{ij} = \begin{cases} \frac{N-i}{N} & \text{för } j = i+1, i = 0, 1, \ldots, N-1 \\ \frac{i}{N} & \text{för } j = i, i = 0, 1, \ldots, N \\ 0 & \text{annars.} \end{cases}$$

Startfördelning är $(1, 0, 0, \ldots, 0)$. Vektorn $(1, 0, \ldots, 0)\boldsymbol{P}^n$ ger då fördelningen för antalet icke-tomma fack efter n kast. För N=5 och $n = 6$ erhåller man till exempel

$$(1,0,0,0,0,0)\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 & 0 \\ 0 & 0 & 2/5 & 3/5 & 0 & 0 \\ 0 & 0 & 0 & 3/5 & 2/5 & 0 \\ 0 & 0 & 0 & 0 & 4/5 & 1/5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}^6 = \left(0, \frac{1}{3125}, \frac{124}{3125}, \frac{216}{625}, \frac{312}{625}, \frac{72}{625}\right)$$

Sannolikheten att inget fack är tomt efter 6 kast är således $\frac{72}{625} = 0.1152$. □

Sannolikheten att en kedja hoppar ur ett tillstånd i är $1 - p_{ii}$ och antalet gånger, tiden, uthoppsförsök görs tills det lyckas är då ffg$(1 - p_{ii})$. Den förväntade tiden i ett tillstånd i är därför

$$\frac{1}{1 - p_{ii}}. \tag{3.9}$$

Låt $\tilde{p}_{ij}$ vara sannolikheten att uthopp görs till tillstånd j från tillstånd $i, i \neq j$. Sannolikheten $\tilde{p}_{ij}$ kallas uthoppssannolikhet. Vi har

$$\begin{aligned} \tilde{p}_{ij} &= P(\text{ hopp från } i \text{ till } j \mid \text{ hopp från } i) \\ &= \frac{P(\text{ hopp från } i \text{ till } j \text{ och hopp från } i)}{P(\text{ hopp från } i)} \\ &= \frac{P(\text{ hopp från } i \text{ till } j)}{P(\text{ hopp från } i)} = \frac{p_{ij}}{1 - p_{ii}}. \end{aligned} \tag{3.10}$$

En Markovkedja uppträder alltså på följande sätt. Om kedjan går in i ett tillstånd i, befinner den sig där en ffg$(1 - p_{ii})$-fördelad tid. Efter denna tid hoppar den med sannolikheten $\tilde{p}_{ij}$ till tillstånd j, där den sedan befinner sig en ffg-fördelad tid, osv. Vi skall senare se att Markovkedjor i kontinuerlig tid uppför sig på liknande sätt.

3.3 Absorption

Vi säger att ett tillstånd i *leder till* tillstånd j om det är möjligt att i ett ändligt antal steg komma från i till j. Ett tillstånd är *absorberande* om kedjan alltid förblir i tillståndet, givet att den kommit dit. Det betyder att i är ett absorberande tillstånd om och endast om $p_{ii} = 1$.

Definition 3.5 *En Markovkedja kallas* A-kedja *om varje tillstånd* i *antingen är absorberande eller leder till ett absorberande tillstånd.*

Definition 3.6 *Ett tillstånd i som leder till ett tillstånd från vilket kedjan ej kan återvända till i kallas ett* genomgångstillstånd.

I en A-kedja är således ett tillstånd antingen ett absorberande tillstånd eller ett genomgångstillstånd.

Tillståndsrummet $\boldsymbol{E}$ i en A-kedja kan alltså delas upp i en mängd av absorberande tillstånd $\boldsymbol{A}$ och en mängd genomgångstillstånd $\boldsymbol{G}$, $\boldsymbol{E} = \boldsymbol{A} \cup \boldsymbol{G}$.

Vi sätter a_{ij}= sannolikheten att absorberas i j givet start i tillstånd i, $i \in \boldsymbol{G}, j \in \boldsymbol{A}$. Man inser att $a_{ij} = \lim_{n\to\infty} P(X_n = j \mid X_0 = i)$.

Vi låter T_i vara tiden tills kedjan absorberas, givet att den startar i tillstånd i. Om $i \in \boldsymbol{A}$ är givetvis $T_i = 0$ och det intressanta fallet är då $i \in \boldsymbol{G}$. Det är à priori inte säkert att T_i är ändlig med sannolikhet 1. Väntevärdet $E(T_i)$ betecknar vi t_i.

Sats 3.2 *Låt $\{X_n; n \geq 0\}$ vara en ändlig Markovsk A-kedja. Då gäller att*

a) t_i är ändlig för alla i

b) $P(T_i < \infty) = 1$, d.v.s. T_i är ändlig med sannolikhet 1.

Bevis. Om g är antalet genomgångstillstånd kan vi utan inskränkning anta att genomgångstillstånden är $1, 2, \ldots, g$. Vi betraktar $i = 1$. Övriga tillstånd behandlas på samma sätt. Vi kan skriva $T_1 = U_1 + U_2 + \cdots + U_g$, där U_k är tiden som tillbringas i tillstånd k, innan kedjan absorberas.

Betrakta först U_1. Vid varje besök i 1 finns en möjlighet att kedjan inte återvänder till 1, eftersom 1 leder till ett absorberande tillstånd. Vid varje besök i tillstånd 1 inträffar händelsen ”inget nytt besök” med sannolikhet $\gamma_1 > 0$, oberoende av tidigare besök (Markovegenskapen!). Antalet besök i tillstånd 1 kommer därför att vara ffg(γ_1). Härav fås att $E(U_1) = 1/\gamma_1 < \infty$, se (2.4) på sidan 7.

På samma sätt inses att om tillstånd 2 överhuvudtaget besöks, kommer väntevärdet för U_2 att vara ändligt. Låt H_1 vara händelsen att tillstånd 2 besöks givet start i 1 och H_2 händelsen att så ej sker. Satsen om total förväntan 2.5 på sidan 8 ger då

$$E(U_2) = E(U_2 \mid H_1)P(H_1) + E(U_2 \mid H_2)P(H_2) = E(U_2 \mid H_1)P(H_1) + 0 < \infty.$$

På samma sätt för övriga genomgångstillstånd. Vi erhåller

$$E(T_1) = E(U_1) + E(U_2) + \cdots + E(U_g) < \infty.$$

Eftersom väntevärdet är ändligt är T_1 självt ändlig med sannolikhet 1. Därmed är satsen visad. □

Att $P(T_i < \infty) = 1$ innebär att Markovkedjan förr eller senare kommer att absorberas.

Vi skall nu visa hur man kan beräkna absorptionssannolikheterna och de förväntade tiderna tills absorption sker. Bevisen är exempel på Markovresonemang.

Sats 3.3 *Absorptionssannolikheterna a_{ij}, där a_{ij} är sannolikheten att absorberas i tillstånd* j *vid start i tillstånd* i *i en ändlig Markovkedja, uppfyller för alla $j \in \boldsymbol{A}$ följande ekvationssystem,*

$$a_{ij} = p_{ij} + \sum_{k \in \boldsymbol{G}} p_{ik} a_{kj}, \quad i \in \boldsymbol{G}. \tag{3.11}$$

$\sum_{j \in \boldsymbol{A}} a_{ij} = 1$, *dvs sannolikheten är 1 att kedjan absorberas, vilket tillstånd man än startar vid.*

Bevis. Vi betraktar kedjans tillstånd ett tidssteg framåt. Kedjan kan gå direkt till det absorberande tillståndet j. Sannolikheten för detta är p_{ij}. I annat fall går kedjan till ett annat genomgångstillstånd k för att senare hamna i j. Sannolikheten att först gå till k och sedan därifrån absorberas i j är p.g.a. Markovegenskapen $p_{ik}a_{kj}$. Summerar vi över alla dessa möjligheter får vi likheten (3.11). Att summan av absorptionssannolikheterna är 1 följer av att tiden tills absorption sker är ändlig, $P(T_i < \infty) = 1$. □

Följdsats 3.1 *Låt $\boldsymbol{P}_{AG} = (p_{ij})_{i \in \boldsymbol{G}, j \in \boldsymbol{A}}$ vara den matris som fås ur $\boldsymbol{P}$ då rader respektive kolonner som representerar absorptions- respektive genomgångstillstånd stryks och $\boldsymbol{P}_G = (p_{ij})_{i,j \in \boldsymbol{G}}$ den matris som fås då både rader och kolonner som representerar absorptionstillstånd stryks. Låt vidare $\boldsymbol{A} = (a_{ij})_{i \in \boldsymbol{G}, j \in \boldsymbol{A}}$ vara matrisen av absorptionssannolikheter. Då är*

$$\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{P}_G)^{-1} \boldsymbol{P}_{AG}$$

där $\boldsymbol{I}$ är enhetsmatrisen.

Bevis. Fås genom att skriva om (3.11) på matrisform och lämnas som övning till läsaren. □

Sats 3.4 *Låt i en Markovkedja $t_i = E(T_i)$ $i \in \boldsymbol{G}$, vara de förväntade tiderna tills absorption vid start i tillstånd i. Då uppfyller dessa följande ekvationssystem,*

$$t_i = 1 + \sum_{k \in \boldsymbol{G}} p_{ik} t_k \quad i \in \boldsymbol{G}. \tag{3.12}$$

Bevis. Betrakta än en gång vad som sker ett tidssteg framåt. Sätt T_k'=tid efter första hoppet tills absorption sker, givet hopp till k. Naturligtvis är $T_k' = 0$ om $k \in \boldsymbol{A}$.

Man inser då lätt att

$$T_i = 1 + T_k^{'} \text{ om kedjan går till tillstånd } k. \tag{3.13}$$

Men på grund av Markoviteten har T_k och $T_k^{'}$ samma fördelning och samma väntevärde. Satsen om total förväntan 2.5 på sidan 8 ger

$$t_i = E(T_i) = \sum_k p_{ik} E(T_i \mid \text{hopp till } k) = \sum_k p_{ik} E(1 + T_k^{'})$$
$$= \sum_k p_{ik} + \sum_{k \in \boldsymbol{A}} p_{ik} E(T_k) + \sum_{k \in \boldsymbol{G}} p_{ik} E(T_k) = 1 + 0 + \sum_{k \in \boldsymbol{G}} p_{ik} t_k = 1 + \sum_{k \in \boldsymbol{G}} p_{ik} t_k. \tag{3.14}$$

□

Följdsats 3.2 *Låt $\boldsymbol{P}_G = (p_{ij})_{i,j \in \boldsymbol{G}}$ vara den matris som fås ur $\boldsymbol{P}$ då rader och kolonner som representerar absorptionstillstånd stryks och $\boldsymbol{t}$ kolumnvektorn av förväntade tider. Då är*

$$\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{P}_G)^{-1} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

där $\boldsymbol{I}$ är enhetsmatrisen.

Bevis. Lämnas som övning. □

Vi kan resonera oss fram till ekvationssystemen för absorptionssannolikheter och förväntade absorptionstider på ett annat sätt. Vi gör en liten omskrivning av (3.11) och (3.12) och finner att ($\boldsymbol{G} \setminus \{i\}$ är mängden $\boldsymbol{G}$ utom elementet i)

$$a_{ij}(1 - p_{ii}) = p_{ij} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} p_{ik} a_{kj}, \quad i \in \boldsymbol{G}, j \in \boldsymbol{A} \tag{3.15}$$

eller

$$a_{ij} = \frac{p_{ij}}{(1 - p_{ii})} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \frac{p_{ik}}{(1 - p_{ii})} a_{kj}, \quad i \in \boldsymbol{G}, j \in \boldsymbol{A} \tag{3.16}$$

och på samma sätt

$$t_i = \frac{1}{(1 - p_{ii})} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \frac{p_{ik}}{(1 - p_{ii})} t_k, \quad i \in \boldsymbol{G}. \tag{3.17}$$

Den sista av dessa ekvationer kan man få fram med följande resonemang.

Givet start i tillstånd i är förväntad tid till uthopp lika med $\frac{1}{1-p_{ii}}$ och sannolikheten att hopp då sker till tillstånd k är $\tilde{p}_{ik} = \frac{p_{ik}}{1-p_{ii}}$, se (3.9) på sidan 15. Förväntad tid tills absorption är förväntad tid till uthopp från tillstånd i plus förväntad tid därefter tills absorption, vilket enligt satsen om total förväntan är summan i (3.17). På samma sätt kan man härleda ekvationssystemet (3.16).

Dessa sista ekvationsystem pekar på ett allmännare fall än de vi nu mött. Beviset av formlerna håller för en process i vilken uppehållstiderna i de olika tillstånden är oberoende av varandra och av det tillstånd till vilket uthopp sker. Om uthoppssannolikheterna är $\tilde{p}_{ij}, i, j \in \boldsymbol{E}$ och den förväntade tiden i tillstånd i är u_i erhåller vi

$$t_i = u_i + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \tilde{p}_{ik} t_k, \quad i \in \boldsymbol{G} \tag{3.18}$$

och

$$a_{ij} = \tilde{p}_{ij} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \tilde{p}_{ik} a_{kj}, \quad i \in \boldsymbol{G}, j \in \boldsymbol{A}. \tag{3.19}$$

Uppehållstiderna behöver inte ens vara heltalsvariabler, utan kan vara kontinuerliga. Vi återkommer till detta i avsnitt 6.6. Kedjan behöver inte vara markovsk, men däremot är den inbäddade hoppkedja man får om man endast betraktar uthoppstiderna en Markovkedja. Den ursprungliga processen kallas då en *semi-Markovkedja.*

Vi tillämpar dessa satser på ett ruinproblem.

Exempel 3.7 (fortsättning exempel 3.4 på sidan 13). En partikel vandrar mellan heltalspunkterna $0, 1, 2 \ldots, N$ och hoppar till höger med sannolikhet p och till vänster med sannolikhet $q = 1 - p$, utom i ändpunkterna 0 och N, där den förblir med sannolikhet 1. Vi sätter a_i lika med sannolikheten att partikeln absorberas i tillstånd 0, givet start i tillstånd i. Ekvationssystemet (3.11) ger då

$$a_i = qa_{i-1} + pa_{i+1} \quad i = 1, 2, \ldots, N-1 \tag{3.20}$$

med randvillkoren $a_0 = 1$ och $a_N = 0$.

Med hjälp av teorin för differensekvationer kan detta ekvationssystem lösas enkelt, men vi använder här ett annat knep. Om vi sätter $b_i = a_i - a_{i-1}$ erhåller vi

$$\begin{aligned}
(p+q)a_i =& qa_{i-1} + pa_{i+1} \quad i = 1, 2, \ldots, N-1 \text{ d.v.s.} \\
q(a_i - a_{i-1}) =& p(a_{i+1} - a_i) \text{ d.v.s.} \\
qb_i =& pb_{i+1} \quad i = 1, 2, \ldots, N-1
\end{aligned}$$

vilket ger $b_{i+1} = \frac{q}{p} b_i =$(upprepad användning)$=(\frac{q}{p})^2 b_{i-1} = \cdots = (\frac{q}{p})^i b_1$.

Om vi summerar b_i erhåller vi $\sum_{i=0}^{N-1} b_{i+1} = \sum_{i=0}^{N-1} a_{i+1} - a_i = a_2 - a_1 + a_3 - a_2 + \cdots + a_N - a_{N-1} = -a_1 + a_N = -1$. Om $p \neq q$ kan denna summa också beräknas som

$$\sum_{i=0}^{N-1} b_{i+1} = \sum_{i=0}^{N-1} (\frac{q}{p})^i b_1 = (\text{geometrisk serie}) = b_1 \frac{1-(q/p)^N}{1-q/p}$$

vilket ger $b_1 = -\frac{1-q/p}{1-(q/p)^N}$. Om $p = q = \frac{1}{2}$ erhåller vi istället $b_1 = -1/N$. Till sist beräknar vi a_i

$$a_i = a_0 + \sum_{j=0}^{i-1} (a_{j+1} - a_j) = 1 + \sum_{j=0}^{i-1} b_j =$$
$$1 + \sum_{j=0}^{i-1} (\frac{q}{p})^j b_1 = 1 - \frac{1-(q/p)^i}{1-(q/p)^N} = \frac{(q/p)^i - (q/p)^N}{1-(q/p)^N}$$

om $p \neq q$. Om $p = q = \frac{1}{2}$ är $a_i = 1 - i/N$.

Låt oss nu se på förväntad tid tills absorption. Ekvationssystemet (3.12) övergår till

$$t_i = 1 + pt_{i+1} + qt_{i-1} \quad i = 1, 2, \ldots, N-1. \tag{3.21}$$

Randvillkor är naturligtvis $t_0 = t_N = 0$.

På liknande sätt som ovan kan man beräkna t_i och erhåller

$$t_i = \frac{N\big(1-(q/p)^i\big) - i\big(1-(q/p)^N\big)}{(p-q)\big(1-(q/p)^N\big)} \tag{3.22}$$

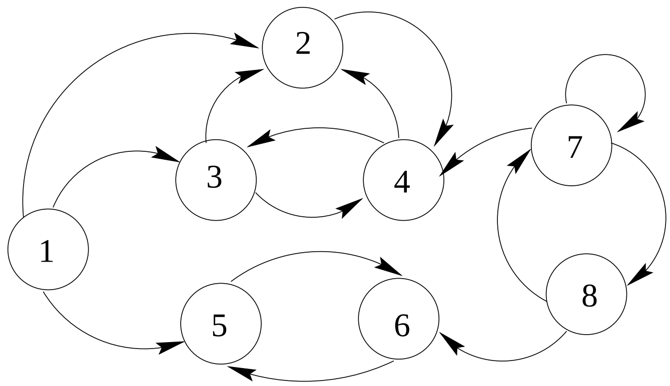
om $p \neq q$. Om $p = q = \frac{1}{2}$ fås $t_i = i(N-i)$.

Läsaren kan verifiera detta påstående genom insättning av formlerna för t_i i (3.21) på denna sida.

Antag att två spelare A och B har a resp b kronor och spelar ett spel där var och en har lika chans att vinna respektive förlora en krona. Spelet slutar då en av dem ruinerats. Om X_n är A:s kassa efter n spel slutar spelet om X_n är 0 eller $N = a+b$. Markovkedjan $\{X_n; n \geq 0\}$ beskriver då en slumpvandring med absorberande barriärer. Sannolikheten att A vinner är enligt ovan $\dfrac{a}{a+b}$. Den förväntade tiden tills spelet slutar är ab. Man ser att den som från början är förmögnast har störst chans att ta hem spelet, sannolikheten att ruinera motspelaren är proportionell mot den egna insatsen. □

I ett senare kapitel skall vi studera konvergens av Markovkedjor. Vi kommer då att stöta på problem som illustreras i följande exempel.

Exempel 3.8 Låt $\{X_n; n \geq 0\}$ vara en Markovkedja enligt figur 3.3. Vi tänker oss att $n \to \infty$. Vi ser av övergångsgrafen att om kedjan hamnat i tillstånd 2,3 eller 4 kommer den alltid att leva kvar bland dessa tillstånd. Likaså kommer den alltid att vandra runt bland tillstånden 5 och 6 om den hamnat i något av dessa tillstånd. Det asymptotiska förloppet beror alltså i hög grad i vilken av dessa ”delkedjor” Markovkedjan hamnat i. Vi ser dessutom att om den hamnat i ”delkedjan” {5,6} kommer den varannan gång att vara i tillstånd 5 och



Figur 3.3: Övergångsgraf till exempel 3.8

varannan gång i tillstånd 6, dvs vad som asymptotiskt händer beror på om vi betraktar jämna eller udda tidpunkter. Tillstånd 1 är ett genomgångstillstånd, eftersom kedjan kan gå till ett annat tillstånd från vilket ingen återvändo till tillstånd 1 kan ske. □

I exemplet är det de rena grafteoretiska egenskaperna som styr det asymptotiska förloppet och som komplicerar kedjans uppträdande. Det visar sig att för Markovkedjor med ändligt tillståndsrum, är villkoren för existens av asymptotisk fördelning rent grafteoretiska och därför skall vi härnäst studera dessa.

Kapitel 4

Markovkedjors grafteoretiska egenskaper

Vi betraktar nu en ”kedja” som vandrar i ett tillståndsrum $\boldsymbol{E}$, men där kedjan inte har några speciella sannolikheter att gå mellan tillstånden. Som tidigare kan vi illustrera en sådan kedja med en övergångsgraf, som anger om det är möjligt att i ett tidssteg gå från ett tillstånd till ett annat.

Ett annat sätt att illustrera övergångsmöjligheterna är med hjälp av *grannmatrisen* (”adjacency matrix”) $\boldsymbol{G}$. Elementet i i:te raden, j:te kolumnen i $\boldsymbol{G}$, g_{ij} är 1 om det är möjligt att gå från i direkt till j, och 0 annars.

Grannmatrisen för kedjan i figur 3.1 på sidan 11 är således

$$\boldsymbol{G} = \begin{array}{c} \\ \mathit{1} \\ \mathit{2} \\ \mathit{3} \\ \mathit{4} \\ \mathit{5} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathit{1} & \mathit{2} & \mathit{3} & \mathit{4} & \mathit{5} \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

och i slumpvandringsexemplet 3.1 på sidan 10

$$\boldsymbol{G} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots \\ 1 & 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & 1 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

där första raden och kolonnen motsvarar punkten 0, andra raden och kolonnen punkten 1 o.s.v. Observera att en punkt kan vara granne med sig själv.

Grannmatrisen är en motsvarighet till övergångsmatrisen och man kan visa att elementet $g_{ij}^{(n)}$ i $\boldsymbol{G}^n$ är lika med antalet olika sätt att gå från i till j i n tidssteg. Vi överlåter åt den intresserade läsaren att visa detta. Vi kommer inte att använda detta resultat som dock är en motsvarighet till Chapman-Kolmogorovs ekvationer, se sats 3.1 på sidan 13.

Vi är nu beredda att definiera den första grafegenskapen.

Definition 4.1 *Ett tillstånd i sägs* leda till *tillstånd j om det är möjligt att gå från i till j i noll, ett eller flera tidssteg, skrivs $i \to j$.*

Två tillstånd i och j sägs kommunicera *om $i \to j$ och $j \to i$, skrivs $i \leftrightarrow j$.*

Med $\boldsymbol{E}_i$ kommer vi att mena mängden av alla tillstånd j som kommunicerar med tillståndet i, d.v.s. $\boldsymbol{E}_i$ är en ekvivalensklass relativt $\leftrightarrow$.

Med den formulering vi valt kommunicerar i alltid med sig självt eftersom man definitionsenligt kan gå från i till i i noll tidssteg. Det innebär att i tillhör mängden $\boldsymbol{E}_i$.

Sats 4.1 *Egenskapen $\leftrightarrow$ definierar en ekvivalensrelation, d.v.s.*

a) symmetrisk, dvs om $i \leftrightarrow j$ så gäller $j \leftrightarrow i$,

b) transitiv, dvs om $i \leftrightarrow j$ och $j \leftrightarrow k$ så gäller $i \leftrightarrow k$,

c) reflexiv, d.v.s $i \leftrightarrow i$.

Bevis. Lämnas till läsaren. □

Sats 4.2 *Det gäller att*

a) $\boldsymbol{E}_i$ innehåller tillstånd i.

b) $\boldsymbol{E}_i$ och $\boldsymbol{E}_j$ är antingen disjunkta eller sammanfallande.

Bevis. a) Följer av c) i sats 4.1

b) Antag att k tillhör såväl $\boldsymbol{E}_i$ som $\boldsymbol{E}_j$, och låt l vara ett tillstånd som tillhör $\boldsymbol{E}_i$. Vi har att $l \leftrightarrow k, k \leftrightarrow j$ och således $l \leftrightarrow j$. Därför tillhör l också $\boldsymbol{E}_j$ och $\boldsymbol{E}_i \subset \boldsymbol{E}_j$. På samma sätt gäller $\boldsymbol{E}_j \subset \boldsymbol{E}_i$ och mängderna är identiska. Satsen följer för övrigt direkt av att $\boldsymbol{E}_i$ är en ekvivalensklass relativt $i \leftrightarrow j$.□

Definition 4.2 *En delmängd sägs vara* sluten *om inget tillstånd i den leder ut från delmängden.*

Det innebär att en kedja som hamnat i en sluten delmängd, alltid kommer att förbli där.

Definition 4.3 *En mängd av tillstånd vilka kommunicerar med varandra kallas en* irreducibel tillståndsmängd*. En kedja vars tillståndsrum är irreducibelt kallas en* irreducibel kedja*.*

Tillståndsmängderna $\boldsymbol{E}_i$ är, som lätt inses, irreducibla.

Enligt definition 3.6 är i ett genomgångstillstånd om det leder till ett tillstånd från vilket kedjan inte kan återvända till i. Man inser lätt att detta är ekvivalent med att $\boldsymbol{E}_i$ är icke-sluten.

Från ett genomgångstillstånd finns det således någon väg till ett annat tillstånd från vilket man ej kan komma tillbaka.

Genom att betrakta tillståndsrummen $\boldsymbol{E}_i$ inser man att tillståndsmängden $\boldsymbol{E}$ entydigt kan delas upp i delmängder $\boldsymbol{F}_0, \boldsymbol{F}_1, \ldots$ med följande egenskaper.

Sats 4.3 *Låt $\boldsymbol{F}_0$ vara mängden av alla genomgångstillstånd, dvs föreningen av alla icke-slutna $\boldsymbol{E}_i$. Välj sedan successivt $\boldsymbol{F}_i, i = 1, 2, \ldots$ så att $\boldsymbol{F}_i$ är en av de ej tidigare valda $\boldsymbol{E}_j$-mängderna och disjunkt från $\boldsymbol{F}_0, \boldsymbol{F}_1, \ldots, \boldsymbol{F}_{i-1}$. Det gäller då att*

a) $\boldsymbol{F}_0, \boldsymbol{F}_1, \ldots$ *är parvis disjunkta,*

b) $\boldsymbol{F}_i$ *är irreducibla och slutna,* $i = 1, 2, \ldots$

c) $\boldsymbol{E} = \cup_i \boldsymbol{F}_i$

Bevis. Lämnas till läsaren. □

Konstruktionen i satsen innebär helt enkelt att $\boldsymbol{F}_0$ är mängden av genomgångstillstånd och de övriga $\boldsymbol{F}_j$-mängderna är de *slutna* $\boldsymbol{E}_i$-mängderna. Vitsen med satsen är att den visar att man kan dela upp hela tillståndsrummet $\boldsymbol{E}$ i en genomgångstillståndmängd och ett antal disjunkta deltillståndsrum som utgör irreducibla slutna delklasser.

Exempel 4.1 Betrakta kedjan i exempel 3.8 på sidan 20. Vi ser att $\boldsymbol{E}_1 = \{1\}$, $\boldsymbol{E}_2 = \boldsymbol{E}_3 = \boldsymbol{E}_4 = \{2, 3, 4\}$, $\boldsymbol{E}_5 = \boldsymbol{E}_6 = \{5, 6\}$ och $\boldsymbol{E}_7 = \boldsymbol{E}_8 = \{7, 8\}$. 1,7 och 8 är genomgångstillstånd och vi får att $\boldsymbol{F}_0 = \boldsymbol{E}_1 \cup \boldsymbol{E}_7 = \{1, 7, 8\}$, $\boldsymbol{F}_1 = \{2, 3, 4\}$ och $\boldsymbol{F}_2 = \{5, 6\}$. □

Sats 4.4 *En ändlig kedja har minst ett slutet irreducibelt deltillståndsrum.*

Bevis. Antag motsatsen, d.v.s. att alla N tillstånd är genomgångstillstånd. Starta kedjan i tillstånd i_1. Det är då möjligt att efter ett visst antal steg gå till ett tillstånd som inte leder tillbaka till i_1. Kalla detta tillstånd i_2. Från i_2 är det möjligt att i ett visst antal steg gå till ett tillstånd som inte leder tillbaka till i_2, och inte heller till i_1, efter vad som tidigare antagits (ty då kan man gå från i_2 till i_1), o.s.v. Efter $N{+}1$ sådana operationer har $N{+}1$ olika tillstånd besökts, vilket naturligtvis är omöjligt om antal möjliga tillstånd är N. Motsägelsen bevisar satsen. □

Om kedjan har oändligt antal tillstånd, gäller inte satsen vilket följande triviala exempel visar: Låt kedjan vara en vandring på de icke-negativa heltalen och antag att man bara kan gå till höger ett steg. Kedjan driver då mot oändligheten då antal tidssteg växer och alla tillstånd är genomgångstillstånd.

Vi har ju tidigare studerat absorption. Att ett tillstånd är absorberande är ju en rent grafteoretisk egenskap: Ett tillstånd i är *absorberande* om man inte kan gå från i till något annat tillstånd.

En kedja som hamnat i ett absorberande tillstånd kommer därför alltid att förbli där.

I grannmatrisen $\boldsymbol{G}$ är således i absorberande om

$$g_{ij} = \begin{cases} 1 \text{ om } j = i \\ 0 \text{ om } j \neq i. \end{cases}$$

Den sista grafteoretiska egenskap vi skall ägna oss åt är periodicitet.

Definition 4.4 *Låt D_i vara mängden av heltal n sådan att det är möjligt att från tillståndet* i *återvända till detta tillstånd i n tidssteg. Med* perioden d_i *till* i *menas den största gemensamma delaren till talen i D_i. Om $d_i = 1$ kallas* i *för* aperiodiskt *tillstånd.*

Exempel 4.2 Tillståndet 0 i slumpvandringen i exempel 3.1 på sidan 10 är aperiodiskt, eftersom man kan gå från 0 till 0 i ett steg. Tillstånd 1 är också aperiodiskt eftersom man kan går från 1 till 1 i 2 steg (t.e.x. $1 \to 0 \to 1$) men också i 3 steg (t.e.x. $1 \to 0 \to 0 \to 1$). Eftersom 2 och 3 är relativt prima är perioden d_1=1. □

Att 0 och 1 båda är aperiodiska i exemplet är ingen tillfällighet, utan följer av

Sats 4.5 *Antag att $i \leftrightarrow j$. Då är $d_i = d_j$, d.v.s. kommunicerande tillstånd har samma period.*

Bevis. Från i tillbaka till i kan man gå via j, eftersom de två tillstånden kommunicerar.

Antag att man kan gå från i till j i n_1 steg. Låt n_2 vara ett godtyckligt tal sådant att man kan gå från j tillbaka till j i n_2 steg, och antag att man kan gå från j till i i n_3 steg. Från i tillbaka till i kan man alltså gå på $n_1 + n_2 + n_3$ steg.

Men kedjan kan naturligtvis naturligtvis gå från i till i i $n_1 + n_3$ steg, d.v.s. utan att göra mellanloopen från j till j. Eftersom i har period d_i är både $n_1 + n_3$ och $n_1 + n_2 + n_3$ multipler av d_i, och härav följer att n_2 är multipel av d_i.

Men av konstruktionen följer att n_2 alltid är multipel av d_j eftersom tillstånd j har period d_j. Dessutom är d_j det största talet som delar alla sådana n_2. Talet d_j måste därför vara minst lika stort som d_i, $d_i \leq d_j$. Av symmetriskäl gäller också $d_j \leq d_i$, varför perioderna är lika. □

Eftersom alla tillstånd i en irreducibel delklass således har samma period, kan man tala om aperiodisk delklass eller delklass med period d.

Tillståndsrummet till en irreducibel kedja med period d kan delas upp i d disjunkta delmängder på följande sätt.

Låt $\boldsymbol{D}_r$ ={i; kedjan kan gå från tillstånd 1 till tillstånd i i $nd + r$ steg för något n}, $r = 0, 1, \ldots, d-1$. Det är lätt att se att $1 \in \boldsymbol{D}_0$ eftersom 1 har period d. Vi kan nu visa

Sats 4.6

a) $\boldsymbol{D}_0, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_{d-1}$ *är parvis disjunkta.*

b) $\cup_{i=0}^{d-1} \boldsymbol{D}_i = \boldsymbol{E}$

c) Kedjan förflyttar sig cykliskt mellan tillstånden enligt följande mönster

$$\boldsymbol{D}_0 \to \boldsymbol{D}_1 \to \cdots \to \boldsymbol{D}_{d-1} \to \boldsymbol{D}_0 \cdots$$

Bevis. a) Antag att h tillhör både $\boldsymbol{D}_i$ och $\boldsymbol{D}_j, i \neq j$. Då kan man gå från 1 till h i $n_1 d + i$ steg för något n_1 och även i $n_2 d + j$ steg för något n_2. Antag att man kan gå från h till 1 i k steg. Då inses lätt att man kan gå från 1 tillbaka till 1 i både $n_1 d + i + k$ och $n_2 d + j + k$ steg. Eftersom tillstånd 1 har period d är skillnaden $(n_1 - n_2)d + i - j$ och därmed även $i - j$ delbart med d. Eftersom $|i - j| < d$ är då $i = j$ vilket är en motsägelse.

b) Låt i vara godtyckligt tillstånd. 1 leder till i i n_i steg för något $n_i = nd + r$ där $0 < r < d$ och i tillhör således $\boldsymbol{D}_r$. Av detta följer b).

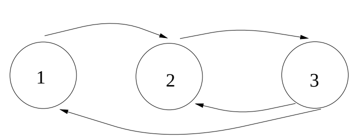
c) Antag att h tillhör $\boldsymbol{D}_i$. Om kedjan befinner sig i h kommer den vid nästa tidpunkt att befinna sig i $\boldsymbol{D}_j$. Då kan kedjan gå från 1 till ett tillstånd i $\boldsymbol{D}_j$ i $nd + i + 1$ steg. Men av detta följer att tillståndet h tillhör $\boldsymbol{D}_{i+1}$. Eftersom D-delmängderna enligt a) är disjunkta är $D_j = D_{i+1}$. □

Exempel 4.3 En kedja har följande grannmatris.

$$\boldsymbol{G} = \begin{array}{cc} & \begin{array}{cccccccc} \mathit{1} & \mathit{2} & \mathit{3} & \mathit{4} & \mathit{5} & \mathit{6} & \mathit{7} & \mathit{8} \end{array} \\ \begin{array}{c} \mathit{1} \\ \mathit{2} \\ \mathit{3} \\ \mathit{4} \\ \mathit{5} \\ \mathit{6} \\ \mathit{7} \\ \mathit{8} \end{array} & \begin{pmatrix} 0 & * & * & 0 & * & * & 0 & * \\ 0 & * & * & * & 0 & * & 0 & 0 \\ 0 & * & * & 0 & * & 0 & * & * \\ 0 & 0 & 0 & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \end{pmatrix} \end{array}$$

där vi för åskådlighetens skull bytt ettor mot stjärna $*$. Vi ser att 1, 2 och 3 är genomgångstillstånd. 1 kommunicerar inte med något annat tillstånd, medan 2 och 3 kommunicerar och utgör en irreducibel delmängd dock ej sluten. Tillstånden 4 och 5 bildar en sluten irreducibel delklass som är aperiodisk. För att visa att delklassen är aperiodisk räcker det att observera att man kan komma från 4 till 4 i ett tidssteg. Tillstånden 6 och 7 bildar också en sluten irreducibel delklass, men denna har period 2. Slutligen är tillstånd 8 absorberande. □

Exempel 4.4 Tillstånd *1* i kedjan som beskrivs av följande övergångsgraf är aperiodiskt.



Man kan gå från tillstånd *1* tillbaka till *1* i t.ex. 3 och 5 steg. Det är också lätt att se att man inte kan återvända i 1,2 eller 4 steg, men däremot i 6,7,8, osv steg, dvs vi kan gå från *1* tillbaka till *1* i n steg för alla $n \geq 5$. Alla aperiodiska kedjor uppför sig på samma sätt vilket följande sats visar. □

De återstående resultaten i detta kapitel är hjälpsatser, varav de två första egentligen talteoretiska.

Sats 4.7 *Betrakta ett aperiodisk tillstånd i. Det existerar ett n_0 sådant att för varje $n \geq n_0$ kan kedjan från tillstånd* i *återvända till detta tillstånd i n steg.*

Bevis. Vi noterar först rent allmänt, att en kedja som kan gå från i till j i både n och m steg, kan också göra detta i $an + bm$ steg om a och b är heltal ≥ 0 (definitionsenligt går man från i till i i 0 steg, ty då sker ingen övergång alls).

Sätt nu

$$u = \min\{n_1 - n_2; n_2 < n_1, \text{ och kedjan kan återvända till } i \text{ i både } n_2 \text{ och } n_1 \text{ steg}\} \tag{4.1}$$

Vi påstår att $u = 1$, ty antag motsatsen, $u > 1$. Kedjan kan då gå från i till i i såväl n_1 steg som $n_2 = n_1 - u$ och $ku + r$ steg för något n_1 samt något $k \geq 0$ och $0 < r \leq u - 1$. Om det det inte fanns något sådant $r > 0$ skulle antalet steg till återkomst alltid vara ku för något k och perioden vara $\geq u$.

Kedjan kan då gå från i till i i $kn_2 + ku + r = k(n_1 - u) + ku + r = kn_1 + r$ steg men även i kn_1 steg. Men skillnaden mellan dessa möjliga antal steg är $0 < r \leq u - 1$ vilket motsäger definitionen av u.

Låt nu $n \geq n_0 = n_2^2$. Låt r vara resten då n divideras med n_2. Det innebär att $n = kn_2 + r$ för något $k \geq n_2$ och $r, 0 \leq r < n_2$,

$$n = kn_2 + r = (\text{ty } n_1 - n_2 = 1) = kn_2 + r(n_1 - n_2) = (k - r)n_2 + rn_1.$$

Eftersom $k - r > 0$ kan kedjan alltså gå från i till i i n steg. □

Följdsats 4.1 *Låt i och j vara två tillstånd i en irreducibel aperiodisk kedja. Då existerar ett tal n_{ij} sådant att kedjan kan gå från i till j i n steg för alla $n \geq n_{ij}$.*

Bevis. Kedjan kan gå från i till j i n_1 steg, säg. Låt n_0 vara som i satsen och låt $n \geq n_{ij} = n_0 + n_1$. Genom att först gå från i tillbaka till i i $n - n_1 \geq n_0$ steg och sedan gå från i till j i n_1 steg går kedjan från i till j i n steg. □

Som en tillämpning på de grafteoretiska begreppen visar vi följande sats som vi kommer att använda senare.

Sats 4.8 *Låt $\boldsymbol{G}$ vara en ändlig, irreducibel, aperiodisk riktad graf med tillståndsrum $\boldsymbol{E} = \{1, 2, \ldots, N\}$. Låt $\boldsymbol{G} \times \boldsymbol{G}$ vara produktgrafen med tillståndsrum $\boldsymbol{E} \times \boldsymbol{E} = \{(i, j), i, j \in \boldsymbol{E}\}$ och för vilken tillståndet (i_1, i_2) leder till (j_1, j_2) om och endast om i_1 leder till j_1 och i_2 leder till j_2 i $\boldsymbol{G}$, d.v.s. grannmatrisen $(g_{(i_1,i_2),(j_1,j_2)})_{E \times E}$ ges av*

$$g_{(i_1,i_2),(j_1,j_2)} = g_{i_1,j_1} g_{i_2,j_2}.$$

Då är $\boldsymbol{G} \times \boldsymbol{G}$ ändlig, irreducibel och aperiodisk.

Bevis. Att $\boldsymbol{G} \times \boldsymbol{G}$ är ändlig är trivialt.

Varje vandring från tillstånd 1 tillbaka till tillstånd 1, ger en motsvarande väg att längs ”diagonalen” $\big((1,1),(2,2),\ldots,(N,N)\big)$ i produktgrafen gå från (1,1) tillbaka till (1,1). Eftersom 1 är ett aperiodiskt tillstånd i $\boldsymbol{G}$, är (1,1) ett aperiodiskt tillstånd i produktgrafen. Visar vi sedan att produktgrafen är irreducibel, följer att alla tillstånd är aperiodiska. Återstår alltså att visa irreducibiliteten.

Enligt följdsatsen ovan kan den ursprungliga kedjan gå från i_1 till j_1 i n steg om $n \geq n_1$ för något n_1. Från i_2 till j_2 kan kedjan på liknande sätt gå i n steg om $n \geq n_2$. Låt $n_{12} = \max(n_1, n_2)$ och låt $n \geq n_{12}$. Den ursprungliga kedjan kan då gå både från i_1 till j_1 och i_2 till j_2 i n steg, dvs i produktgrafen leder (i_1, j_1) till (i_2, j_2). På samma sätt leder (i_2, j_2) till (i_1, j_1) och tillstånden kommunicerar. $\square$

Kapitel 5

Konvergens av Markovkedjor

5.1 Stationaritet och ergodicitet

Vi skall nu studera uppträdandet av en Markovkedja $\{X_n; n \geq 0\}$ då n går mot ∞. Speciellt skall vi undersöka under vilka förutsättningar fördelningen för X_n konvergerar och gränsfördelningen är oberoende av starttillståndet. De så kallade stationära fördelningarna är fundamentala i detta sammanhang.

Definition 5.1 *En fördelning* $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots)$ *är en* stationär *fördelning till en Markovkedja med övergångsmatris* $\boldsymbol{P}$ *om*

$$\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}. \tag{5.1}$$

Om en Markovkedja har startfördelningen $\boldsymbol{\pi}$, så erhålls fördelningen för X_n som

$$\boldsymbol{p}^{(n)} = \boldsymbol{\pi}\boldsymbol{P}^n = \boldsymbol{\pi}\boldsymbol{P}\boldsymbol{P}^{n-1} = \boldsymbol{\pi}\boldsymbol{P}^{n-1} = \cdots = \boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi} \tag{5.2}$$

dvs X_n har fördelning $\boldsymbol{\pi}$ för alla n. Därav termen stationär.

Relationen (5.1) kan skrivas

$$\boldsymbol{\pi}(\boldsymbol{P} - \boldsymbol{I}) = \boldsymbol{0} \text{ där } \boldsymbol{I} \text{ är enhetsmatrisen och } \boldsymbol{0} \text{ nollvektorn.} \tag{5.3}$$

Denna relation definierar ett linjärt ekvationssystem i $\pi_1, \pi_2, \dots$. Koefficientmatrisen $\boldsymbol{P} - \boldsymbol{I}$ är singulär, ty radsummorna är $\sum_j p_{ij} - 1 = 1 - 1 = 0$ för alla i. Alltså finns en lösning som är skild från 0-lösningen (ändligt tillståndsrum). Man kan dock inte säkert påstå om det finns en lösning som är en sannolikhetsfördelning, dvs en lösning med $\sum_i \pi_i = 1$ och $\pi_i \geq 0$. Om tillståndsrummet är ändligt är emellertid så fallet.

Sats 5.1 *Låt* $\{X_n; n \geq 0\}$ *vara en Markovkedja med ett ändligt tillståndsrum och med övergångsmatris* $\boldsymbol{P}$*. Då existerar en stationär fördelning.*

Bevis. Eftersom kedjan är ändlig finns ett slutet irreducibelt deltillståndsrum. Vi kan anta att tillstånd 1 tillhör detta. Antag att kedjan startar i tillstånd 1, d.v.s. $X_0 = 1$. Låt T_1 vara antalet steg tills kedjan återkommer till tillstånd 1. Att $E(T_1)$ är ändlig följer av att vi kan se T_1 som en absorptionstid och att förväntade tider tills absorption är ändliga; på samma sätt som vid härledningen av (3.17) på sidan 18 inser man att

$$t_1 = \frac{1}{(1-p_{11})} + \sum_{k \in \boldsymbol{G} \setminus \{1\}} \frac{p_{1k}}{(1-p_{11})} t_{k1},$$

där $t_1 = E(T_1)$ och t_{k1} är förväntad tid tills absorption i tillstånd 1 givet start i tillstånd k. Låt U_j vara antalet tidpunkter kedjan är i tillstånd j innan den återvänt till tillstånd 1 och sätt u_j lika med väntevärdet av U_j, $u_j = E(U_j)$. Väntevärdena för U_j-variablerna är också ändligt eftersom $U_j \leq T_1$. Vi skall visa följande likhet:

$$u_j = \sum_{i \in \boldsymbol{E}} u_i p_{ij}, \quad j = 1, 2, \ldots \tag{5.4}$$

I ord innebär (5.4) att det ”förväntade antalet tidpunkter kedjan är i tillstånd j är lika med förväntat antal tidpunkter kedjan är i tillstånd i och därefter går till tillstånd j, summerat över alla i”, något som kanske verkar intuitivt plausibelt.

Låt nu $\pi_i = \frac{u_i}{\sum_{k \in \boldsymbol{E}} u_k}$. Då är $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ en sannolikhetsfördelning och från (5.4) fås

$$\pi_j = \frac{u_j}{\sum_{k \in \boldsymbol{E}} u_k} = \sum_{i \in \boldsymbol{E}} \frac{u_i}{\sum_{k \in \boldsymbol{E}} u_k} p_{ij} = \sum_{i \in \boldsymbol{E}} \pi_i p_{ij} \tag{5.5}$$

eller i vektorform

$$\boldsymbol{\pi} = \boldsymbol{\pi} \boldsymbol{P} \tag{5.6}$$

$\boldsymbol{\pi}$ är alltså en stationär fördelning. Det kan vara nyttigt att även tolka den första likheten i (5.5) heuristiskt. Den säger i ord att den stationära sannolikheten att vara i tillstånd j är förhållandet mellan den förväntade tiden som kedjan ligger i tillstånd j mellan två besök i tillstånd 1 och den totala förväntade tiden mellan två sådana besök.

Vi skall alltså visa (5.4) och inför ett antal indikatorvariabler. Sätt för $n \geq 1$

$$I_n(i) = \begin{cases} 1 \text{ om } X_1 \neq 1, X_2 \neq 1, \ldots, X_{n-1} \neq 1, X_n = i \\ 0 \text{ om annars} \end{cases}$$

d.v.s. $I_n(i)$ är 1 om och endast om processen är i tillstånd i vid tidpunkt n och inte har återvänt till tillstånd 1 före denna tidpunkt. Man inser lätt att $E\big(I_n(i)\big) = 0 \cdot P(I_n(i) = 0) + 1 \cdot P(I_n(i) = 1) = P(I_n(i) = 1)$. Vi definierar $I_0(i)$ som 1 om $i = 1$ och 0 annars.

Om som ovan $U_j =$ tid i tillstånd j innan tillstånd 1 besöks igen, kan vi skriva

$$U_j = \sum_{n=1}^{\infty} I_n(j) = \sum_{n=1}^{\infty} \sum_{i \in \boldsymbol{E}} I_{n-1}(i) I_n(j) \tag{5.7}$$

Den första likheten kan inses av att indikatorerna i högerledet är 1 precis när kedjan är i tillstånd j vid tidpunkt n och inte har återkommit till tillstånd 1 före denna tidpunkt. Summan räknar alltså antalet gånger kedjan är i tillstånd j före återkomst till tillstånd 1. Den andra likheten inses av att $\sum_{i\in\boldsymbol{E}} I_{n-1}(i) = 1$ eftersom processen är i ett och endast ett tillstånd vid en given tidpunkt. Den sista summan kan också ses som en räknare av antalet tidspar $(n-1, n)$ sådana att processen är i tillstånd j vid den vid tid n, innan den återvänt till tillstånd 1. Detta parantal måste vara detsamma som antalet gånger processen är i tillstånd j före återkomst till 1. Notera att likheterna även gäller för j=1 i vilket fall alla leden blir lika med 1.

Vi har nu

$$
\begin{aligned}
u_j = E(U_j) = E\big(\sum_{n=1}^{\infty} I_n(j)\big) = E\big(\sum_{n=1}^{\infty}\sum_{i\in\boldsymbol{E}} I_{n-1}(i)I_n(j)\big) = \sum_{n=1}^{\infty}\sum_{i\in\boldsymbol{E}} E\big(I_{n-1}(i)I_n(j)\big)\\
= \sum_{n=1}^{\infty}\sum_{i\in\boldsymbol{E}} 0\cdot P(I_{n-1}(i)I_n(j)=0) + 1\cdot P(I_{n-1}(i)I_n(j)=1)\\
= \sum_{n=1}^{\infty}\sum_{i\in\boldsymbol{E}} P(I_{n-1}(i)I_n(j)=1) = \sum_{n=1}^{\infty}\sum_{i\in\boldsymbol{E}} P(I_{n-1}(i)=1)P(I_n(j)=1 \mid I_{n-1}(i)=1)
\end{aligned}
\tag{5.8}
$$

Den sista betingade sannolikheten kan skrivas

$$
\begin{aligned}
&P(I_n(j)=1 \mid I_{n-1}(i)=1)\\
&= P(X_n = j, X_1\neq 1,\ldots,X_{n-1}\neq 1 \mid X_1\neq 1,\ldots,X_{n-2}\neq 1, X_{n-1}=i)
\end{aligned}
\tag{5.9}
$$

Om $i \neq 1$ följer av markoviteten, att ovanstående sannolikhet är

$$
P(X_n = j \mid X_1\neq 1,\ldots,X_{n-2}\neq 1, X_{n-1}=i) = P(X_n = j \mid X_{n-1}=i) = p_{ij}.
$$

(observera att händelsen $\{X_1 \neq 1,\ldots,X_{n-1}\neq 1\}$ i (5.9) följer av den givna händelsen).

Om $i = 1$ och $n > 1$ kan inte den givna och den betingade händelsen i (5.9) inträffa samtidigt; om $I_{n-1}(1) = 1$ har processen återvänt till tillstånd 1 och $I_n(j) = 0$. För $n = 1$ är sannolikheten i (5.9) $P(I_1(j) = 1 \mid I_0(1) = 1) = P(X_1 = j \mid X_0 = 1) = p_{1j}$.

Sätter vi in detta i (5.8) erhålls

$$
\begin{aligned}
u_j = \sum_{n=1}^{\infty}\sum_{i\in\boldsymbol{E}\setminus\{1\}} P(I_{n-1}(i)=1)p_{ij} + P(X_0=1)p_{1j}\\
= \sum_{n=1}^{\infty}\sum_{i\in\boldsymbol{E}\setminus\{1\}} E\big(I_{n-1}(i)\big)p_{ij} + p_{1j}\\
= \sum_{i\in\boldsymbol{E}\setminus\{1\}} p_{ij}\sum_{n=1}^{\infty} E\big(I_{n-1}(i)\big) + p_{1j} = \sum_{i\in\boldsymbol{E}\setminus\{1\}} p_{ij}u_i + p_{1j}u_1 = \sum_{i\in\boldsymbol{E}} p_{ij}u_i
\end{aligned}
\tag{5.10}
$$

eftersom

$$
\sum_{n=1}^{\infty} E\big(I_{n-1}(i)\big) = \sum_{n=0}^{\infty} E\big(I_n(i)\big) = \sum_{n=1}^{\infty} E\big(I_n(i)\big) = E\big(\sum_{n=1}^{\infty} I_n(i)\big) = E(U_i) = u_i
$$

för $i \neq 1$ och $u_1 = 1$.

Därmed har vi visat (5.4). Vi har också visat att

$$\pi_1 = \frac{u_1}{\sum_{j \in \boldsymbol{E}} u_j} = \frac{1}{E(T_1)}$$

där T_1 är tid mellan två besök i tillstånd 1. Summan $\sum_{j \in \boldsymbol{E}} u_j$ är ju den totala förväntade tiden mellan två besök i tillstånd 1. För framtida bruk noterar vi också att $\pi_j/\pi_1 = \frac{u_j/\sum_k u_k}{u_1/\sum_k u_k} = u_j/u_1 = u_j$, d.v.s.

$$\begin{aligned} \pi_j/\pi_1 &= E(U_j) \\ &= \text{förväntat antal besök i tillstånd j mellan två besök i tillstånd 1.} \end{aligned} \tag{5.11}$$

□

Beviset av satsen håller även för oändligt tillståndsrum om vi istället kräver att den förväntade tiden mellan två besök i tillstånd 1 är ändligt. Ett sådant tillstånd kallas *positivt rekurrent*. Vi återkommer till detta i slutet av kapitlet.

I beviset kan vi naturligtvis låta vilket tillstånd k som helst i ett irreducibelt deltillståndsrum ersätta tillstånd 1. Om k är ett genomgångstillstånd är sannolikheten positiv att man inte återvänder till tillståndet, d.v.s. T_k, tiden tills man återvänt, är oändlig med positiv sannolikhet. Därför är $E(T_k) = \infty$ och $1/E(T_k) = 0$. Av beviset ovan följer att det finns en stationär fördelning med $\pi_k = 0$ i detta fall. Därför har vi följande sats.

Följdsats 5.1 *Låt k vara ett godtyckligt fixt tillstånd i en ändlig Markovkedja. Då existerar en stationär fördelning $\boldsymbol{\pi}$ sådan att $\pi_k = \frac{1}{E(T_k)}$ där $E(T_k)$ är förväntad tid mellan två besök i tillstånd k.*

Intresset för stationära fördelningar kommer från att det är dessa som är de möjliga gränsfördelningarna.

Definition 5.2 *En Markovkedja $\{X_n; n \geq 0\}$ sägs vara* ergodisk *om en gränsfördelning $\boldsymbol{p}$ existerar och om den är oberoende av startfördelning, d.v.s. om $\lim_{n\to\infty} \boldsymbol{p}^{(0)}\boldsymbol{P}^{(n)}$ existerar och inte beror av $\boldsymbol{p}^{(0)}$.*

Följande sats är en av huvudsatserna i teorin för Markovkedjor.

Sats 5.2 *En ändlig, irreducibel, aperiodisk Markovkedja är ergodisk och dess gränsfördelning är den entydiga, stationära fördelningen.*

Bevis. Vi använder oss av så kallad kopplingsteknik. Låt $\{X_n; n \geq 0\}$ och $\{Y_n; n \geq 0\}$ vara två oberoende Markovkedjor med övergångsmatrisen $\boldsymbol{P}$. Låt vidare $\{X_n; n \geq 0\}$ ha startfördelning $\boldsymbol{p}^{(0)}$ och $\{Y_n; n \geq 0\}$ en stationär startfördelning $\boldsymbol{\pi}$. Vi sätter $Z_n = (X_n, Y_n)$. Då är $\{Z_n; n \geq 0\}$ definerad på produkttillståndsrummet $\boldsymbol{E} \times \boldsymbol{E}$ och är

(i) en Markovkedja (bevis överlämnas till läsaren)

(ii) irreducibel och aperiodisk. Detta följer av sats 4.8 på sidan 28.

Låt T vara tiden tills Z_n för första gången hamnar i $(1,1)$. På samma sätt som i sats 3.2 på sidan 16 fås att T är ändlig: $P(T < \infty) = 1$. Man inser att denna tid är densamma som tiden tills absorption i den kedja man får om man gör om $(1,1)$ till ett absorberande tillstånd. När Z_n når $(1,1)$ är $Y_n = X_n$ och dess framtida fördelningar blir lika. Eftersom fördelningen för Y_n är $\boldsymbol{\pi}$ för alla n, kommer sålunda gränsfördelningen för X_n att vara densamma. Formellt visar vi detta på följande sätt.

$$
\begin{aligned}
|P(X_n = i) - \pi_i| &= |P(X_n = i) - P(Y_n = i)| = \text{(satsen om total sannolikhet)} \\
&= \left| \sum_k P(X_n = i \mid T = k)P(T = k) - \sum_k P(Y_n = i \mid T = k)P(T = k) \right| \\
&= \left| \sum_k \big(P(X_n = i \mid T = k) - P(Y_n = i | T = k)\big) P(T = k) \right| \qquad (5.12)
\end{aligned}
$$

Den sista summan delar vi upp i två delar och använder triangelolikheten.

$$
\begin{aligned}
&\left| \sum_k \big(P(X_n = i \mid T = k) - P(Y_n = i \mid T = k)\big) P(T = k) \right| \\
&\quad \leq \left| \sum_{k=1}^{n} \big((P(X_n = i \mid T = k) - P(Y_n = i \mid T = k)\big) P(T = k) \right| \\
&\quad + \left| \sum_{k=n+1}^{\infty} \big(P(X_n = i \mid T = k) - P(Y_n = i \mid T = k)\big) P(T = k) \right|. \qquad (5.13)
\end{aligned}
$$

Betrakta först den sista summan ovan.

$$
\begin{aligned}
&\left| \sum_{k=n+1}^{\infty} \big(P(X_n = i \mid T = k) - P(Y_n = i \mid T = k)\big) P(T = k) \right| \leq \\
&\quad \sum_{k=n+1}^{\infty} \big|(P(X_n = i \mid T = k) - P(Y_n = i \mid T = k)\big| P(T = k) \leq \\
&\quad\quad \sum_{k=n+1}^{\infty} P(T = k) = P(T \geq n+1).
\end{aligned}
$$

Betrakta sedan den första summan i högerledet av (5.13). Denna summa är 0 ty

$$P(X_n = i \mid T = k) = P(Y_n = i \mid T = k)$$

eftersom X_k och Y_k båda är 1 vid tid $T = k$ och X- och Y-processerna efter denna tid är processer med samma övergångsmatris $\boldsymbol{P}$ och med samma starttillstånd 1. När $n - k$ tidsenheter efter tid $T = k$ förflutit har X_n och Y_n då samma fördelning, d.v.s. sannolikheten att båda antar värdet i är lika.

Insatt i (5.12) erhåller vi $|P(X_n = i) - \pi_i| \leq 0 + P(T \geq n+1)$. Men då $n \to \infty$ går $P(T \geq n+1)$ mot 0 och således är $\lim_{n\to\infty} P(X_n = i) = \pi_i$.

Av beviset framgår att den stationära fördelningen är entydig; låt nämligen startfördelningen för X_n vara en stationär fördelning $\boldsymbol{\pi}'$. Då följer av beviset att $\boldsymbol{\pi}'$ och $\boldsymbol{\pi}$ sammanfaller. □

Vi har alltså visat att alla ändliga, irreducibla, aperiodiska kedjor har en entydig stationär fördelning. Genom ett litet knep kan vi utvidga detta resultat även till periodiska kedjor.

Sats 5.3 *En ändlig, irreducibel kedja har en entydig stationär fördelning* $\boldsymbol{\pi}$ *och det gäller att* $\pi_i = \dfrac{1}{E(T_i)}$ *där* T_i *är tiden mellan två besök i tillstånd* i. *Kvoten* $\dfrac{\pi_j}{\pi_i}$ *är förväntat tid i tillstånd* j *mellan två besök i tillstånd* i.

Bevis. Antag $\boldsymbol{P}$ är övergångsmatrisen och sätt $\boldsymbol{P}_1 = \frac{1}{2}(\boldsymbol{I} + \boldsymbol{P})$. Man visar lätt att $\boldsymbol{P}_1$ är en övergångsmatris och eftersom diagonalelementen är $\geq \frac{1}{2} > 0$ är den aperiodisk. Att den är ändlig är trivialt och eftersom $\boldsymbol{P}$ är irreducibel är även $\boldsymbol{P}_1$ irreducibel.

Låt nu $\boldsymbol{\gamma}$ vara en fördelning. Vi har

$$\boldsymbol{\gamma}\boldsymbol{P}_1 = \boldsymbol{\gamma} \iff \boldsymbol{\gamma}\frac{1}{2}(\boldsymbol{I} + \boldsymbol{P}) = \boldsymbol{\gamma} \iff \frac{1}{2}\boldsymbol{\gamma}\boldsymbol{P} = \frac{1}{2}\gamma \iff \boldsymbol{\gamma}\boldsymbol{P} = \boldsymbol{\gamma}$$

d.v.s. $\boldsymbol{\gamma}$ är en stationär fördelning till $\boldsymbol{P}$ om och endast om $\boldsymbol{\gamma}$ är stationär till $\boldsymbol{P}_1$. Men $\boldsymbol{P}_1$ har en entydig stationär fördelning, således även $\boldsymbol{P}$. Om $\boldsymbol{\pi}$ är den stationära fördelningen följer från följdsatsen 5.1 att $\pi_i = \dfrac{1}{E(T_i)}$, där T_i tiden mellan två besök i tillstånd i. Det sista påståendet i satsen ges för $i = 1$ av ekvationen (5.11) på sidan 34 och gäller p.g.a. symmetrin för ett godtyckligt tillstånd i. □

Låt $\{X_n\}$ vara en ergodisk Markovkedja med gränsfördelning $\boldsymbol{p}$ och X en stokastisk variabel med $\boldsymbol{p}$ som fördelning. För varje funktion g sådan att $E(g(X))$ existerar gäller då att

$$\frac{1}{n}\sum_{k=1}^{n} g(X_k) \to E(g(X)) \text{ då } n \to \infty. \tag{5.14}$$

Då tillståndsrummet består av reella tal ger speciellt valet $g(x) = x$ att

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \to E(X),$$

vilket är en generalisering av stora talens lag från en följd av oberoende stokastiska variabler till Markovkedjor och innebär att väntevärdet i fördelningen $\boldsymbol{p}$ kan skattas på vanligt sätt. Genom att välja g som en indikatorfunktion kan även motsvarande komponent i $\boldsymbol{p}$ själv skattas.

Resultatet i (5.14) återspeglar egentligen bättre det allmänna begreppet ergodicitet än definition 5.2.

Exempel 5.1 En Markovkedja med tillstånd 1, 2 och 3 har följande övergångsmatris

$$\boldsymbol{P} = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.6 & 0 & 0.4 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}.$$

Denna ändliga Markovkedja är irreducibel och aperiodisk, läsaren får själv visa detta. Den har alltså en gränsfördelning som är oberoende av var kedjan startar. Denna fördelning fås ur $\boldsymbol{\pi P} = \boldsymbol{\pi}$ och $\pi_1 + \pi_2 + \pi_3 = 1$, vilket är ekvivalent med ekvationssystemet

$$\begin{aligned} 0.2\pi_1 + 0.6\pi_2 + 0.1\pi_3 &= \pi_1 \\ 0.3\pi_1 + \phantom{0.6\pi_2 +{}} 0.3\pi_3 &= \pi_2 \\ 0.5\pi_1 + 0.4\pi_2 + 0.6\pi_3 &= \pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Ekvationssystemet har lösningen $\pi_1 = 0.239, \pi_2 = 0.231, \pi_3 = 0.530$. □

Exempel 5.2 Betrakta en partikels slumpvandring på talen $\{0, 1, \ldots, N\}$, och antag att slumpvandringen sker ett steg till höger med sannolikhet p eller ett steg till vänster med sannolikhet $q = 1 - p$. I punkten 0 stannar partikeln kvar med sannolikhet q och går ett steg till höger med sannolikhet p. I punkten N går den ej till höger utan ligger kvar med sannolikhet p. Övergångsmatrisen är

$$\boldsymbol{P} = \begin{pmatrix} q & p & 0 & \ldots & 0 & 0 & 0 \\ q & 0 & p & \ldots & 0 & 0 & 0 \\ 0 & q & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & q & 0 & p \\ 0 & 0 & 0 & \ldots & 0 & q & p \end{pmatrix}$$

För att erhålla den stationära fördelningen har vi att lösa ekvationssystemet $\boldsymbol{\pi} = \boldsymbol{\pi P}$, vilket ger

$$\begin{aligned} &\pi_0 = q\pi_0 + q\pi_1 && \text{vilket ger } \pi_1 = \pi_0 \frac{p}{q} \\ &\pi_1 = p\pi_0 + q\pi_2 && \text{vilket ger } \pi_2 = \frac{1}{q}(\pi_1 - p\pi_0) = \pi_0 \frac{p}{q}(\frac{1}{q} - 1) = \pi_0 (\frac{p}{q})^2 \\ &\pi_2 = p\pi_1 + q\pi_3 && \text{vilket ger } \pi_3 = \frac{1}{q}(\pi_2 - p\pi_1) = \pi_0 (\frac{p}{q})^3 \\ &\quad\vdots && \\ &\pi_N = p\pi_{N-1} + p\pi_N && \text{vilket ger } \pi_N = \frac{p}{q}\pi_{N-1} = (\frac{p}{q})^N \pi_0 \end{aligned}$$

Summan av sannolikheterna är ett, $1 = \sum_{i=0}^{N} \pi_i = \sum_{i=0}^{N} (\frac{p}{q})^i \pi_0$. Denna summa av en ändlig geometrisk serie är $\pi_0 \frac{1-(p/q)^{N+1}}{1-p/q}$ varför vi får

$$\pi_0 = \frac{1-p/q}{1-(p/q)^{N+1}} \text{ och således } \pi_i = \left(\frac{p}{q}\right)^i \frac{1-p/q}{1-(p/q)^{N+1}}, i = 0, 1, \ldots, N$$

Det förväntade antalet steg slumpvandringen varit i tillstånd i mellan två besök i tillstånd 0 är $\frac{\pi_i}{\pi_0} = (\frac{p}{q})^i$. □

5.2 Något om periodiska kedjor

Antag att $\boldsymbol{P}$ är övergångsmatrisen till en irreducibel Markovkedja med period d. Det finns då d deltillståndsrum $\boldsymbol{D}_0, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_{d-1}$ mellan vilka kedjan förflyttar sig, se sats 4.6 på sidan 26. Vidare vet vi att det finns en entydig stationär fördelning. Vi skall kort studera det asymptotiska uppträdandet och lämnar åt läsaren att verifiera nedanstående påståenden.

a) $\boldsymbol{P}^d$ har d irreducibla, aperiodiska, slutna delklasser, $\boldsymbol{D}_0, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_{d-1}$. Varje delklass $\boldsymbol{D}_r$ definierar en stationär fördelning med avseende på $\boldsymbol{P}^d$, $\boldsymbol{\pi}^{(r)} = (\pi_1^{(r)}, \pi_2^{(r)}, \ldots, \pi_N^{(r)})$ där $\pi_i^{(r)} = 0$ om $i \notin \boldsymbol{D}_r$.

b) $\boldsymbol{\pi}^{(0)}\boldsymbol{P}^r$ är en stationär fördelning med avseende på $\boldsymbol{P}^d$ och har hela sin sannolikhetsmassa på $\boldsymbol{D}_r$. Alltså är $\boldsymbol{\pi}^{(r)} = \boldsymbol{\pi}^{(0)}\boldsymbol{P}^r$.

c) Sätt $\boldsymbol{\pi} = \frac{\boldsymbol{\pi}^{(1)}+\boldsymbol{\pi}^{(2)}+\cdots+\boldsymbol{\pi}^{(d)}}{d}$. Då är $\boldsymbol{\pi}$ stationär med avseende på $\boldsymbol{P}$, d.v.s. $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ och är alltså den entydiga stationära fördelningen $(\pi_1, \pi_2, \ldots, \pi_N)$ med avseende på $\boldsymbol{P}$.

d) Om $i \in \boldsymbol{D}_r$ fås från ovan att $\pi_i^{(r)} = d\pi_i$.

e) Om $i, j \in \boldsymbol{D}_r$, ligger i samma delklass, gäller således att

$$p_{ij}^{(n)} \begin{cases} \to d\pi_j, \text{ om } n \to \infty \text{ genom } d, 2d, \ldots \\ = 0 \text{ om } n \neq d, 2d, \ldots \end{cases}$$

Exempel 5.3 (forts Ehrenfest urnmodell) Modellen beskrivs i exempel 3.3 på sidan 12. Kedjan är periodisk med period 2, irreducibel och ändlig. En stationär fördelning existerar och den bestäms av ekvationssystemet

$$\begin{aligned} \pi_0 &= \pi_1/N \\ &\vdots \\ \pi_i &= \pi_{i-1}(1 - \frac{i-1}{N}) + \pi_{i+1}\frac{i+1}{N} \\ &\vdots \\ \pi_N &= \pi_{N-1}/N. \end{aligned}$$

Genom insättning kan man verifiera att $\pi_i = \binom{n}{i}\pi_0$ uppfyller ekvationssystemet och summerar vi dessa värden erhålls

$$1 = \sum_{i=0}^{N} \pi_i = \sum_{i=0}^{N} \binom{N}{i} \pi_0 = \pi_0 2^N$$

Det betyder att $\pi_i = \binom{N}{i}(\frac{1}{2})^N$, d.v.s. att antalet partiklar i en urna är $\mathrm{Bin}(N, \frac{1}{2})$ vid stationära förhållanden. Kedjan är inte ergodisk men om man startar med 0 partiklar i urnan kommer $p_i^{(2n)}$ att gå mot $2\binom{N}{i}(\frac{1}{2})^N$ för jämna värden på i. Sannolikheten $p_i^{(2n+1)}$ kommer att gå mot $2\binom{N}{i}(\frac{1}{2})^N$ för udda värden på i då $n \to \infty$. □

5.3 Det asymptotiska uppträdandet hos ändliga Markovkedjor

Vi har hittills studerat det asymptotiska uppträdandet hos irreducibla kedjor. Vad kan man då säga i det allmänna fallet?

Sats 5.4 *En ändlig Markovkedja är ergodisk om och endast om det existerar en enda sluten irreducibel delklass och denna är aperiodisk. Den asymptotiska fördelningen är lika med den stationära.*

Bevis Om det finns två eller fler slutna irreducibla delklasser, kommer var och en av dem att ge upphov till en stationär fördelning. Om det finns exakt en irreducibel delklass kommer Markovkedjan förr eller senare att hamna i denna. Detta följer av de resultat som visats i avsnittet om absorption. Den asymptotiska fördelningen ges sedan av den irreducibla delklassens stationära fördelning. □

Låt nu $\boldsymbol{P}$ vara en godtycklig övergångsmatris till en ändlig Markovkedja och antag att $\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_\nu$ är dess slutna delklasser och $\boldsymbol{F}_0$ klassen av dess genomgångstillstånd, se sats 4.3 på sidan 25. Med sannolikhet 1 kommer kedjan, var den än börjar, att hamna i någon av de slutna delklasserna och sannolikheten att hamna i de olika delklasserna kan beräknas som absorptionssannolikheter. När den väl hamnat i en delklass, kan den asymptotiska fördelningen beräknas enligt metoder ovan. Är delklassen aperiodisk fås en asymptotisk fördelning, är kedjan periodisk fås det asymptotiska uppträdandet enligt avsnitt 5.2.

Sats 5.5 *Låt $\{X_n; n \geq 0\}$ vara en ändlig Markovkedja. Då existerar*

$$\lim_{n\to\infty} P(X_n = j \mid X_0 = i)$$

om och endast om j ej tillhör en periodisk delklass. Gränsvärdet är då

$$\lim_{n\to\infty} P(X_n = j \mid X_0 = i) = a_{ik}\pi_j \ \text{ om } j \in \boldsymbol{F}_k \tag{5.15}$$

där a_{ik} är sannolikheten att hamna i delklassen $\boldsymbol{F}_k$ vid start i tillstånd i och där π_j fås från den stationära sannolikhetsfördelningen med avseende på den aperiodiska delklassen $\boldsymbol{F}_k$. Speciellt är gränsvärdet lika med 0 om j är ett genomgångstillstånd.

Bevis. Satsen följer av att

$$\begin{aligned}
&P(X_n = j \mid X_0 = i)\\
&= P(\text{kedjan hamnar i } \boldsymbol{F}_k \mid X_0 = i)P(X_n = j \mid X_0 = i, \text{kedjan hamnar i } \boldsymbol{F}_k)\\
&= P(\text{kedjan hamnar i } \boldsymbol{F}_k \mid X_0 = i)P(X_n = j \mid \text{kedjan hamnar i } \boldsymbol{F}_k) \to a_{ik}\pi_j
\end{aligned}$$

då $n \to \infty$. □

Observera att vi i satsen inte förutsätter att gränssannolikheten är oberoende av startfördelning. Kedjan behöver alltså ej vara ergodisk, och är det endast om det finns en enda sluten delklass.

Exempel 5.4 En Markovkedja i diskret tid har tillstånden $E = \{1, 2, 3, 4, 5, 6\}$ och övergångsmatrisen

$$\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0\\
0.1 & 0 & 0.7 & 0.2 & 0 & 0\\
0.2 & 0.8 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0.5 & 0 & 0.5\\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}$$

Klassificera tillstånden och avgör om $\lim_{n\to\infty} P\left(X_n = j \mid X_0 = i\right)$ för $i, j = 1, 2, 3, 4, 5, 6$ existerar samt beräkna i så fall dessa gränsvärden.

Lösning. Vi noterar att 1 är ett absorberande tillstånd. Vidare är $\{4, 5, 6\}$ en sluten irreducibel delklass, medan 2 och 3 är genomgångstillstånd. Om vi klumpar ihop 4, 5 och 6 i ett tillstånd får vi en A-kedja med övergångsmatrisen

$$\boldsymbol{P}_1 = \begin{pmatrix}
1 & 0 & 0 & 0\\
0.1 & 0 & 0.7 & 0.2\\
0.2 & 0.8 & 0 & 0\\
0 & 0 & 0 & 1
\end{pmatrix}$$

Om vi låter $a_{ij} = P(\text{absorption i tillstånd } j \mid \text{start i tillstånd } i)$ så erhåller vi $a_{21} = 0.1 + 0.7a_{31}$ och $a_{31} = 0.2 + 0.8a_{21} = 0.2 + 0.8(0.1 + 0.7a_{31}) = 0.2 + 0.08 + 0.56a_{31}$ (se (3.11) på sidan 17) som ger

$$a_{31} = \frac{0.28}{(1 - 0.56)} = \frac{28}{44} = \frac{7}{11}$$

och alltså $a_{3,\{4,5,6\}} = 4/11$. Vidare är enligt ovan $a_{21} = 0.1 + 0.7a_{31} = 0.1 + 0.7 \cdot 7/11 = 6/11$ dvs $a_{2,\{4,5,6\}} = 5/11$. Om vi betraktar delkedjan $\{4, 5, 6\}$ så har den övergångsmatrisen

$$\begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 1 & 0 & 0 \end{pmatrix}$$

och denna är irreducibel ty $4 \to 5 \to 6 \to 4$ och aperiodisk ty $4 \to 5 \to 4$ (två steg) och $4 \to 5 \to 6 \to 4$ (tre steg) och 2 och 3 är relativt prima tal. Alltså är denna delkedja ergodisk. Sker absorption i delkedjan $\{4,5,6\}$ så är den stationära fördelningen på denna delkedja den asymptotiska fördelningen. För att bestämma denna stationära fördelning (π_4, π_5, π_6) löser vi ekvationssystemet $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ dvs

$$\begin{cases} \pi_4 &= 0.5\pi_5 + \pi_6 \\ \pi_5 &= \pi_4 \\ \pi_6 &= 0.5\pi_5 \\ 1 &= \pi_4 + \pi_5 + \pi_6 \end{cases}$$

som ger $\pi_4(1+1+0.5) = 1$ dvs $\pi_4 = \pi_5 = 2/5$ och $\pi_6 = 1/5$.
Vi får alltså att gränsvärdena existerar för alla i och j. T.ex. erhåller vi $\lim_{n\to\infty} P(X_n = 4 \mid X_0 = 2) = a_{2,\{4,5,6\}} \cdot \pi_4 = 5/11 \cdot 2/5 = 2/11$. Övriga gränsvärden ges på samma sätt av

1) $\lim_{n\to\infty} P(X_n = 1 \mid X_0 = 1) = 1$

2) $\lim_{n\to\infty} P(X_n = j \mid X_0 = i) = 2/5$ för $j = 4, 5$
samt $\lim_{n\to\infty} P(X_n = 6 \mid X_0 = i) = 1/5$ för $i = 4, 5, 6$.

3) $\lim_{n\to\infty} P(X_n = 1 \mid X_0 = 2) = 6/11$

4) $\lim_{n\to\infty} P(X_n = 1 \mid X_0 = 3) = 7/11$

5) $\lim_{n\to\infty} P(X_n = j \mid X_0 = 2) = 5/11 \cdot 2/5 = 2/11$ för $j = 4, 5$
samt $\lim_{n\to\infty} P(X_n = 6 \mid X_0 = 2) = 5/11 \cdot 1/5 = 1/11$

6) $\lim_{n\to\infty} P(X_n = j \mid X_0 = 3) = 4/11 \cdot 2/5 = 8/55$ för $j = 4, 5$
samt $\lim_{n\to\infty} P(X_n = 6 \mid X_0 = 3) = 4/11 \cdot 1/5 = 4/55$

Övriga gränsvärden är 0. □

5.4 Det asymptotiska uppträdandet hos Markovkedjor med oändligt antal tillstånd

Villkoren för konvergens av Markovkedjor med ändligt antal tillstånd är, som framgått av tidigare avsnitt, grafteoretiska. Karaktären av det asymptotiska uppträdandet bestäms av sådana egenskaper som irreducibilitet, periodicitet och absorption. Så är inte fallet om antalet tillstånd är oändligt, då möjligheten att driva mot oändligheten tillkommer. Denna ”drift” bestäms inte enbart av de grafteoretiska egenskaperna.

Exempel 5.5 Betrakta en partikels slumpvandring på de naturliga talen, som i exempel 3.1 på sidan 10, och antag att slumpvandringen sker ett steg till höger med sannolikhet p och ett steg till vänster med sannolikhet $q = 1 - p$. I punkten 0 stannar partikeln kvar med sannolikhet q och går ett steg till höger med sannolikhet p. Om $0 < p < 1$ är kedjan aperiodisk och irreducibel, men dess asymptotiska uppträdande beror på p-värdet. Om $p < \frac{1}{2}$ kommer en asymptotisk fördelning att existera och vi skall senare beräkna denna. Om $p > \frac{1}{2}$ kommer kedjan att driva mot oändligheten och om $p = \frac{1}{2}$ kommer kedjan att oscillera mer och mer och sannolikheten att finna den i en given punkt kommer att gå mot 0. Någon gränsfördelning kommer inte att existera i dessa fall. □

För ändliga, aperiodiska, irreducibla Markovkedjor är den asymptotiska fördelningen lika med den stationära. För oändliga kedjor gäller samma sak, och dessutom är existensen av en stationär fördelning ett både nödvändigt och tillräckligt villkor för att kedjan skall konvergera. I det ändliga fallet finns alltid en stationär fördelning, men så behöver inte vara fallet för en oändlig Markovkedja.

Sats 5.6 *Låt P vara övergångsmatrisen till en irreducibel, aperiodisk Markovkedja. Betrakta följande ekvationssystem, $\boldsymbol{x} = (x_1, x_2, \ldots)$.*

$$\begin{aligned} &\boldsymbol{x} = \boldsymbol{xP} \\ &x_i \geq 0 \quad i = 1, 2, \ldots \end{aligned} \tag{5.16}$$

Då gäller

a) Om (5.16) har en lösning sådan att

$$0 < \sum_{i=1}^{\infty} x_i < \infty \tag{5.17}$$

så har Markovkedjan en entydig stationär fördelning som ges av $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ där

$$\pi_i = \frac{x_i}{\sum_{j=1}^{\infty} x_j} \tag{5.18}$$

Markovkedjan är ergodisk med gränsfördelning som ges av $\boldsymbol{\pi}$.

b) Om (5.16) har en lösning som satisfierar

$$\sum_{i=1}^{\infty} x_i = \infty \tag{5.19}$$

så har kedjan ingen stationär fördelning och är inte ergodisk.

Bevis. Visas ej. □

Forts av exempel 5.5 Ekvationsystemet (5.16) leder till

$$x_0 = qx_0 + qx_1 \tag{5.20}$$

$$x_i = px_{i-1} + qx_{i+1} \quad i = 1, 2, \ldots \tag{5.21}$$

Vi kan skriva om denna

$$x_1 = \frac{p}{q} x_0 \tag{5.22}$$

$$x_{i+1} = \frac{1}{q} x_i - \frac{p}{q} x_{i-1} \quad i = 1, 2, \dots \tag{5.23}$$

Man kan lösa detta system iterativt och erhåller (jfr. exempel 5.2 på sidan 37)

$$x_i = \left(\frac{p}{q}\right)^i x_0 \tag{5.24}$$

Om $p < q$, dvs $p < \frac{1}{2}$, konvergerar serien $\sum_{i=0}^{\infty} x_i = \sum_{i=0}^{\infty} (\frac{p}{q})^i x_0$ och summan är $x_0 \frac{1}{1-p/q}$. En stationär fördelning existerar och denna ges av $\boldsymbol{\pi}$ med

$$\pi_i = \frac{x_i}{\sum_j x_j} = \left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^i$$

en geometrisk fördelning med parameter $1 - \frac{p}{q}$. Denna är också den asymptotiska fördelningen.

Om $p \geq q$, dvs $p \geq \frac{1}{2}$, konvergerar inte summan $\sum_{i=0}^{n} x_i$, vilket lätt inses. Kedjan har ingen stationär fördelning och är inte ergodisk. □

Även om kedjan i detta exempel inte har någon asymptotisk fördelning om $p \geq \frac{1}{2}$ är det en kvalitativ skillnad i det asymptotiska uppträdandet mellan fallen $p = \frac{1}{2}$ och $p > \frac{1}{2}$. I det sista fallet är det en drift mot oändligheten och sannolikheten att komma tillbaka till ett givet tillstånd i är strikt mindre än 1 (sannolikheten att återkomma är q/p). I fallet $p = \frac{1}{2}$ är däremot sannolikheten att komma tillbaka till i lika med 1, dvs man kommer säkert tillbaka till tillståndet och därav följer att man återkommer ett oändligt antal gånger. Fallet $p = \frac{1}{2}$ är i detta avseende kvalitativt detsamma som fallet $p < \frac{1}{2}$. Tillstånd till vilka man säkert kommer tillbaka kallas *rekurrenta* eller *beständiga* tillstånd. I fallet $p = \frac{1}{2}$ är dock den förväntade tiden tills kedjan återkommer till ett tillstånd oändlig, men i fallet $p < \frac{1}{2}$ är denna förväntade tid ändlig. Rekurrenta tillstånd för vilka förväntad återkomsttid är oändlig kallas *nollrekurrenta* eller *nollbeständiga*. Om förväntad tid är ändlig kallas tillståndet för *positivt rekurrent* eller *positivt beständigt*. Tillstånd som man med sannolikhet mindre än 1 återvänder till kallas *transienta* (kallas även *obeständiga*). Genomgångstillstånd är transienta tillstånd och om kedjan är ändlig gäller även motsatsen. I en Markovkedja med oändligt antal tillstånd kan tillstånd vara transienta utan att vara genomgångstillstånd, t.e.x. är alla tillstånd i slumpvandringsexemplet ovan transienta om $\frac{1}{2} < p < 1$ men inget är genomgångstillstånd.

Två kommunicerande tillstånd är antingen båda nollrekurrenta, eller båda positivt rekurrenta eller båda transienta. Det innebär att alla tillstånd i en irreducibel kedja har samma karaktär. Av anmärkningen efter beviset av sats 5.1, se sidan 34, följer att en Markovkedja med ett positivt rekurrent tillstånd har en stationär fördelning. Satserna 5.1 på sidan 31 och 5.3 på sidan 36 gäller om villkoret ”ändlig kedja” ersätts med villkoret ”positivt rekurrent kedja”.

I själva verket är villkoret ”positivt rekurrent kedja” också nödvändigt för existens av en stationär fördelning i en irreducibel kedja, d.v.s. i en irreducibel Markovkedja som har en stationär fördelning är alla tillstånd positivt rekurrenta.

Kapitel 6

Markovprocesser i kontinuerlig tid

6.1 Inledning

Vi ska nu diskutera fallet med $T = [0, \infty)$ och börjar med några begrepp som på det stora hela är naturliga modifieringar från fallet med diskret tid. Tillståndsmängden $\boldsymbol{E}$ låter vi i allmänhet vara $\boldsymbol{E} = \{0, 1, \ldots, N\}$ eller $\boldsymbol{E} = \{0, 1, \ldots\}$.

Definition 6.1 *En stokastisk process $X(t)$ kallas en (diskret)* Markovprocess *om*

$$P(X(t_{n+1}) = i_{n+1} \mid X(t_n) = i_n, X(t_{n-1}) = i_{n-1}, \ldots, X(t_0) = i_0) \tag{6.1}$$
$$= P(X(t_{n+1}) = i_{n+1} \mid X(t_n) = i_n) \tag{6.2}$$

för alla n, alla $0 \leq t_0 < t_1 < \ldots < t_{n+1}$ och alla tillstånd $i_0, \ldots, i_{n+1} \in \boldsymbol{E}$.

Om $P(X(t+s) = j \mid X(s) = i)$, $i, j \in \boldsymbol{E}$, inte beror av s, så är processen *tidshomogen*. Liksom för Markovkedjor i diskret tid begränsar vi oss till detta fall.

Vi sätter $p_{ij}(t) = P(X(t+s) = j \mid X(s) = i)$ och

$$\boldsymbol{P}(t) = \big(p_{ij}(t)\big)_{i,j \in \boldsymbol{E}} = \begin{pmatrix} p_{00}(t) & p_{01}(t) & \ldots & p_{0j}(t) & \ldots \\ p_{10}(t) & p_{11}(t) & \ldots & p_{1j}(t) & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ p_{i0}(t) & p_{i1}(t) & \ldots & p_{ij}(t) & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix} \tag{6.3}$$

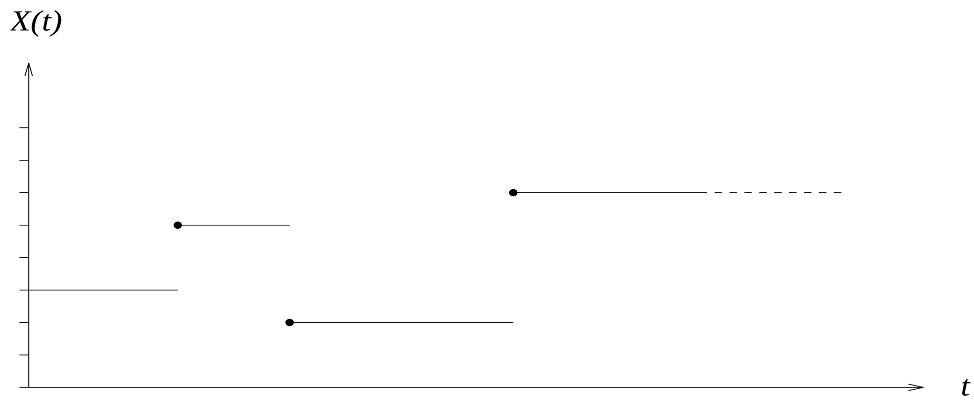
där $p_{ij}(t) = P(X(t) = j \mid X(0) = i)$ eftersom processen är tidshomogen. Matrisen $\boldsymbol{P}(t)$ är övergångsmatris för tid t.

Vi sätter också

$$\boldsymbol{p}(t) = (p_0(t), p_1(t), \ldots) \text{ där } p_i(t) = P(X(t) = i)$$

d.v.s. $\boldsymbol{p}(t)$ är fördelningen för $X(t)$.

På samma sätt som för Markovkedjor visar man



Figur 6.1: Illustration av en realisering av en reguljär Markovprocess.

Sats 6.1 (Chapman-Kolmogorov) *För alla* $s, t \geq 0$ *gäller*

a) $p_{ij}(s+t) = \sum_{k \in \boldsymbol{E}} p_{ik}(s) p_{kj}(t)$

b) $\boldsymbol{P}(s+t) = \boldsymbol{P}(s)\boldsymbol{P}(t)$

c) $\boldsymbol{p}(s+t) = \boldsymbol{p}(s)\boldsymbol{P}(t)$.

Det enklaste fallet av Markovprocesser kan tyckas vara då familjen av $X(t)$:n är oberoende, d.v.s. då $X(t)$ och $X(s)$ är oberoende så snart $t \neq s$. Antag att $X(t)$-variablerna är oberoende och 0 med sannolikhet $\frac{1}{2}$ och 1 med sannolikhet $\frac{1}{2}$. I så fall skulle i varje ändligt tidsintervall, oändligt många $X(t)$ vara 0 och oändligt många vara 1. Processen $\{X(t); t \geq 0\}$ är en synnerligen irreguljär, av ett slag som vi vill undvika. Intressantare är processer som med sannolikhet 1 ligger kvar i ett tillstånd en positiv tid.

Definition 6.2 *En Markovprocess kallas* reguljär *om den med sannolikhet 1 gör högst ändligt många hopp under ändliga tidsintervall och med sannolikhet 1 ligger kvar en positiv tid i ett tillstånd den gått in i.*

Vi kan anta att en reguljär Markovprocess har högerkontinuerliga realiseringar. En realisering av en reguljär Markovprocess kan således se ut som i figur 6.1.

Låt oss se på några exempel på Markovprocesser.

Exempel 6.1 Vi betraktar följande enkla process. Vi låter $\{X(t); t \geq 0\}$ starta i tillstånd 0, d.v.s. $X(0) = 0$, och efter en stokastisk tid T hoppar den till tillstånd 1 där den stannar för evigt.

För att processen skall vara tidshomogen måste

$$P(X(s+t) = 0 \mid X(s) = 0) = P(X(t) = 0 \mid X(0) = 0)$$

vilket är ekvivalent med att

$$P(T > s+t \mid T > s) = P(T > t)$$

Men enligt karaktäriseringssatsen 2.2 på sidan 6 är då T exponentialfördelad med någon intensitet λ_0. Vi får

$$\begin{aligned} P(X(t+h)=0 \mid X(t)=0) &= P(T>t+h \mid T>t) = P(T>h) = e^{-\lambda_0 h} \\ &= \text{(Taylorutveckla)} = 1-\lambda_0 h + \frac{(\lambda_0 h)^2}{2} + \cdots = 1-\lambda_0 h + o(h) \end{aligned}$$

och

$$\begin{aligned} P(X(t+h)=1 \mid X(t)=0) &= P(T\le t+h \mid T>t) = P(T\le h) \\ &= 1-e^{-\lambda_0 h} = \lambda_0 h + o(h). \end{aligned}$$

Naturligtvis är $P(X(t+h)=1 \mid X(t)=1)=1$. Vi ser att $\{X(t); t\ge 0\}$ är en Markovprocess som uppfyller

$$\begin{aligned} p_{00}(h) &= 1-\lambda_0 h + o(h) \qquad & p_{01}(h) &= \lambda_0 h + o(h) \\ p_{10}(h) &= 0 & p_{11}(h) &= 1. \end{aligned}$$

Med $o(h)$, *litet ordo* av h, menas en funktion som går mot 0 fortare än h själv; $\lim_{h\to 0} \dfrac{o(h)}{h} = 0$. Observera *ordon* som står på olika sidor om ett likhetstecken i allmänhet är olika funktioner och inte ”tar ut” varandra. □

Vi modifierar exemplet något.

Exempel 6.2 Antag att processen i exemplet ovan inte för evigt förblir i tillstånd 1, utan stannar i detta tillstånd en exponentialfördelad tid med intensitet λ_1 varefter den hoppar till tillstånd 0, och allt börjar om. De successiva tiderna i tillstånden antas vara oberoende.

Sätt A händelsen ”inget hopp sker i intervallet (0,h)” och B händelsen ”minst två hopp sker i intervallet (0,h)”. Då har vi

$$\begin{aligned} P(X(t+h)=0 \mid X(t)=0) &= P(\{X(t+h)=0\}\cap A \mid X(t)=0) \\ &\quad + P(\{X(t+h)=0\}\cap B) \mid X(t)=0). \qquad (6.4) \end{aligned}$$

Observera att vi efter <u>ett</u> hopp inte kan befinna oss i tillstånd 0, om vi startar där. Sannolikheten för minst två hopp i tidsintervallet (0,h) är $P(T_0+T_1<h)$ där T_0 och T_1 är de två första uppehållstiderna. De är oberoende och Exp(λ_0) respektive Exp(λ_1). Vi får

$$P(T_0+T_1<h) \le P(T_0<h, T_1<h) = P(T_0<h)P(T_1<h) \le \lambda_0\lambda_1 h^2 = o(h) \qquad (6.5)$$

eftersom $P(T_i<h) = \int_0^h \lambda_i e^{-\lambda_i u} du \le \lambda_i h$

Den sista termen i (6.4) är alltså $o(h)$. Den första sannolikheten i högerledet kan skrivas

$$P(T_0>h) = e^{-\lambda_0 h} = 1-\lambda_0 h + o(h)$$

och sätter vi samman detta erhålls

$$p_{00}(h) = P(X(t+h) = 0 \mid X(t) = 0) = 1 - \lambda_0 h + o(h)$$

På liknande sätt erhålls

$$\begin{aligned} p_{01}(h) &= P(X(t+h) = 1 \mid X(t) = 0) = \lambda_0 h + o(h) \\ p_{11}(h) &= P(X(t+h) = 1 \mid X(t) = 1) = 1 - \lambda_1 h + o(h) \\ p_{10}(h) &= P(X(t+h) = 0 \mid X(t) = 1) = \lambda_1 h + o(h). \end{aligned}$$

Derivatorna $p'_{ij}(0)$ är således

$$p'_{00}(0) = -\lambda_0 \quad p'_{01}(0) = \lambda_0 \quad p'_{10}(0) = \lambda_1 \quad p'_{11}(0) = -\lambda_1. \tag{6.6}$$

□

Derivatorna av övergångssannolikheterna $p_{ij}(t)$ vid tiden 0 är alltså lika med intensiteterna av uppehållstiderna om $i \neq j$ och negativa intensiteterna om $i = j$. Det leder till följande definition.

Definition 6.3 *En Markovprocess har* intensitetsmatris *(*generator*)*

$$\boldsymbol{Q} = (q_{ij})_{i,j \in \boldsymbol{E}} = \begin{pmatrix} q_{00} & q_{01} & \cdots & q_{0j} & \cdots \\ q_{10} & q_{11} & \cdots & q_{1j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ q_{i0} & q_{i1} & \cdots & q_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{6.7}$$

om högerderivatorna av övergångssannolikheterna $p_{ij}(t)$ *existerar i* $t = 0$ *och*

$$p'_{ij}(0) = q_{ij}. \tag{6.8}$$

Ekvationerna (6.8) är ekvivalenta med

$$p_{ij}(h) = \begin{cases} 1 - q_i h + o(h) & \text{om } i = j \\ q_{ij} h + o(h) & \text{om } i \neq j \end{cases} \tag{6.9}$$

där

$$q_i = -q_{ii}.$$

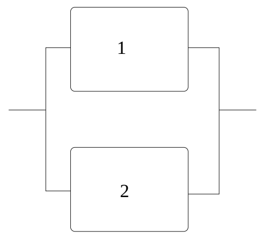
Det framgår av framställningen (6.9) att $q_{ij}, i \neq j$ och q_i är icke-negativa tal. Sannolikheten att under ett kort tidsintervall hoppa från ett tillstånd i till tillståndet $j, j \neq i$ är alltså approximativt q_{ij} gånger tidsintervallets längd. Sannolikheten att hoppa är approximativt q_i gånger intervallets längd.

Eftersom $\sum_{j \in \boldsymbol{E}} p_{ij}(t) \equiv 1$ finner vi efter derivering och insättning av $t = 0$, att $\sum_{j \in \boldsymbol{E}} q_{ij} = 0$ eller efter omskrivning

$$q_i = \sum_{j \in \boldsymbol{E} \setminus \{i\}} q_{ij} \tag{6.10}$$

d.v.s. radsummorna i $\boldsymbol{Q}$ är 0. En kontroll visar att exemplen ovan uppfyller dessa ekvationer.

Exempel 6.3 Markovprocesser används ofta inom tillförlitlighetsteorin. Betrakta parallellsystemet i figur 6.2 som brister om båda komponenterna felar.



Figur 6.2: Parallellsystem

Vi antar att komponenterna har livslängder som är oberoende och Exp(λ_1)- respektive Exp(λ_2)-fördelade. När en enhet går sönder repareras den och reparationstiden är exponentialfördelad med väntevärde $1/\mu$. Två reparatörer är tillgängliga, så båda komponenterna kan repareras samtidigt om systemet är trasigt. Parametrarna λ_1, λ_2 och μ är fel- respektive reparationsintensiteterna för komponenterna i detta fall. Inför nu följande fyra tillstånd.

$$\begin{aligned}
S_0 &= \text{ båda komponenterna hela} \\
S_1 &= \text{ komponent 1 hel, komponent 2 felar} \\
S_2 &= \text{ komponent 1 felar, komponent 2 hel} \\
S_3 &= \text{ båda komponenterna felar}
\end{aligned}$$

Låt $X(t)$ vara processens tillstånd vid tid t. Om processen går in i till exempel tillstånd S_1, i vilken komponent 1 är hel, har den återstående livslängden samma fördelning som den ursprungliga livslängden på grund av exponentialfördelningens glömskeegenskap. Processen väntar alltså på att någon av två händelser, att komponenten under reparation blir hel eller att den hela komponenten brister, skall inträffa och tiderna till dessa händelser är exponentialfördelade, Exp(μ) respektive Exp(λ_1). Kalla dessa två händelser A_{10} respektive A_{13}. Om A_{10} inträffar före A_{13} går processen in i tillstånd S_0, om A_{13} inträffar före A_{10} går den in i tillstånd S_3. På samma sätt kan vi analysera processen då den befinner sig i övriga tillstånd.

Som vi snart skall visa, är då $\{X(t); t \geq 0\}$ är Markovprocess. □

Exemplet kan generaliseras och följande sats ger en allmän beskrivning hur en Markovprocess kan uppkomma och hur den beter sig.

Sats 6.2 *Antag att* $\{X(t); t \geq 0\}$ *är en stokastisk process med tillståndsrummet* $\{0, 1, 2, 3 \dots\}$, *ändligt eller uppräkneligt. Om processen befinner sig i tillstånd* i, *antas händelser* $A_{i0}, A_{i1}, \dots$ *kunna inträffa. Tiden tills* $A_{ij}, i \neq j$ *inträffar,*

T_{ij}, antas vara exponentialfördelad med intensitet q_{ij}, och alla tider antas vara oberoende av varandra. Om A_{ij} inträffar först hoppar processen till tillstånd j, där den väntar att nya händelser på samma sätt inträffar. Sätt $q_{ii} = -q_i = -\sum_{j \neq i} q_{ij}$.

Då är $\{X(t); t \geq 0\}$ en Markovprocess med intensitetsmatris $\boldsymbol{Q} = (q_{ij})_{i,j \in \boldsymbol{E}}$. Sannolikheten att hopp från tillstånd i *sker till tillstånd* j *är*

$$\tilde{p}_{ij} = \frac{q_{ij}}{q_i}, \quad i \neq j. \tag{6.11}$$

Tiden som processen befinner sig i tillstånd i är exponentialfördelad med intensitet q_i och oberoende av processens uppförande före inträdet i tillståndet i och oberoende av vart man hoppar.

Definiera $\tilde{p}_{ii} = 0$. Matrisen

$$\widetilde{\boldsymbol{P}} = (\tilde{p}_{ij})_{i,j \in \boldsymbol{E}} \tag{6.12}$$

kallas uthoppsmatrisen *och är övergångsmatris till den* inbäddade hoppkedjan $\{\widetilde{X}_n; n \geq 0\}$ *där $\widetilde{X}_n$ är processens tillstånd efter* n:*te hoppet. Hoppkedjan är markovsk.*

Bevis. Att processen är tidshomogen markovsk följer av att tiderna är exponentialfördelade på samma sätt som tidigare.

Låt T_i vara tiden som processen befinner sig i tillstånd i då den gått in i detta tillstånd. Enligt sats 2.3 på sidan 6 är

$$T_i = \min\{T_{ik}; k \neq i\} \text{ och } U_{ij} = \min\{T_{ik}; k \neq i, j\}$$

exponentialfördelade med intensiteter $\sum_{k \neq i} q_{ik} = q_i$ respektive $q_i - q_{ij}$. Sannolikheten att processen hoppar till tillstånd j är sannolikheten att T_{ij} är minst av alla $T_{ik}, k \in \boldsymbol{E}$. Tiden T_{ij} är oberoende av U_{ij} och därför är

$$\begin{aligned}
\tilde{p}_{ij} = P(T_{ij} < U_{ij}) &= \iint_{t<u} f_{T_{ij}}(t) f_{U_{ij}}(u)\,du\,dt = \int_0^\infty \int_t^\infty f_{T_{ij}}(t) f_{U_{ij}}(u)\,du\,dt \\
&= \int_0^\infty \int_t^\infty q_{ij} e^{-q_{ij}t} (q_i - q_{ij}) e^{-(q_i - q_{ij})u}\,du\,dt = \int_0^\infty q_{ij} e^{-q_{ij}t} e^{-(q_i - q_{ij})t}\,dt \\
&= \int_0^\infty q_{ij} e^{-q_i t}\,dt = \frac{q_{ij}}{q_i}.
\end{aligned} \tag{6.13}$$

Vi skall nu beräkna $P(X(t+h) = j \mid X(t) = i)$. På grund av att tiden till hopp är exponentialfördelad, är processen tidshomogen, d.v.s.

$$P(X(t+h) = j \mid X(t) = i) = P(X(h) = j \mid X(0) = i).$$

Antag först $i \neq j$. Man kan visa att sannolikheten för två eller flera hopp i tidsintervallet $(0, h]$ är $o(h)$. Genom eftertanke inser man därför att

$$\begin{aligned}
p_{ij}(h) &= P(X(h) = j \mid X(0) = i) = P(T_{ij} < h, h < U_{ij} \mid X(0) = i) + o(h) \\
&= (\text{ober}) = P(T_{ij} < h \mid X(0) = i) P(h < U_{ij} \mid X(0) = i) + o(h) \\
&= (1 - e^{-q_{ij}h}) e^{-(q_i - q_{ij})h} + o(h) = e^{-(q_i - q_{ij})h} - e^{-q_i h} + o(h) = (\text{Taylorutveckla}) \\
&= 1 - (q_i - q_{ij})h - (1 - q_i h) + o(h) = q_{ij} h + o(h) \qquad (6.14)
\end{aligned}$$

och att

$$\begin{aligned}
p_{ii}(h) &= P(X(h) = i \mid X(0) = i) = P(T_i > h) + o(h) = e^{-q_i h} + o(h) \\
&= 1 - q_i h + o(h). \qquad (6.15)
\end{aligned}$$

Den inbäddade hoppkedjan har övergångssannolikheterna $\tilde{p}_{ij}$ och då hoppen sker oberoende av den ursprungliga processens tidigare historia, är hoppkedjan en Markovkedja. Från (6.14), (6.15) och (6.9) på sidan 48 ser vi att Markovprocessens övergångsintensiteter ges av q_{ij}. □

Man måste göra en modifiering av definitionen av hoppmatris om tillståndet i är absorberande, d.v.s. då $q_{ij} = q_i = 0$ alla j. I det fall definierar vi $\tilde{p}_{ij} = 0$ om $i \neq j$ och $\tilde{p}_{ii} = 1$, vilket innebär att hoppkedjan absorberas i tillstånd i.

Vi säger att en Markovprocess är irreducibel om hoppkedjan är irreducibel.

Fortsättning exempel 6.3 Tillförlitlighetsexemplet 6.3 beskriver en situation som vi har i satsen. I varje tillstånd väntar processen på att komponent 1 går sönder eller blir färdigreparerad beroende i vilket tillstånd processen är, och likadant för komponent 2. Tiderna tills dessa händelser är exponentialfördelade och processen $\{X(t); t \geq 0\}$ är en Markovprocess med intensitetsmatris

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_1 - \lambda_2 & \lambda_2 & \lambda_1 & 0 \\ \mu & -\mu - \lambda_1 & 0 & \lambda_1 \\ \mu & 0 & -\mu - \lambda_2 & \lambda_2 \\ 0 & \mu & \mu & -2\mu \end{pmatrix}$$

□

För reguljära Markovprocesser gäller omvändningen av satsen. En reguljär Markovprocess ligger en exponentialfördelad tid i ett tillstånd den trätt in i, innan den hoppar till ett nytt tillstånd. Markovprocesserna i kontinuerlig tid uppför sig alltså på liknande sätt som de i diskret tid, se resonemanget på sidan 15.

Nedanstående följdsats ger en mer generell beskrivning som dock kan återföras till det föregående fallet. Först ett exempel.

Exempel 6.4 (Populationstillväxt) I en population föds och dör individer oberoende av varandra. Tiden tills en individ föder en ny individ är exponentialfördelad med intensitet λ och tiden tills en individ dör är exponentialfördelad med intensitet μ. En individ kan således dö innan den föder en ny individ och efter att ha fött en individ kan den föda en ny efter en exponentialfördelad tid o.s.v. Alla tider är oberoende.

Låt $X(t)$ vara antalet individer i populationen vid tid t. Om populationen har i individer vid tid t kommer för varje individ händelserna ”föder ny individ” och ”dör” att ske efter exponentialfördelade tider, på grund av glömskeegenskapen hos exponentialfördelningen. Så snart en individ föds hoppar processen till tillstånd $i+1$ och så snart en dör hoppar processen till tillstånd $i-1$. Om populationen har i individer låter vi A_{ik} vara händelsen att den k:te individen föder en ny individ, och $A_{i,i+k}$ låter vi vara händelsen att k:te individen dör, $k = 1, 2, \ldots, i$. Tiden tills A_{ik} inträffar är Exp(λ) för $k = 1, 2, \ldots, i$ och processen hoppar då till tillståndet $i + 1$. Tiden tills $A_{i,i+k}$ inträffar är Exp(μ) och processen hoppar då till tillståndet $i - 1$ om $i > 0$. Detta är en generellare beskrivning än den i satsen ovan eftersom två händelser kan leda till samma tillstånd. Som vi dock skall se kan fallet överföras till grundfallet och $\{X(t); t \geq 0\}$ är en Markovprocess. □

Följdsats 6.1 *Antag att processen $\{X(t); t \geq 0\}$ och händelserna A_{ij} uppfyller villkoren i sats 6.2, men att då A_{ij} inträffar, processen hoppar till tillståndet $S(A_{ij})$. Två händelser A_{ij_1} och A_{ij_2} kan således leda till samma nya tillstånd.*

Då är $\{X(t); t \geq 0\}$ en Markovprocess med övergångsintensiteter

$$q'_{ij} = \sum_{k;S(A_{ik})=j} q_{ik}, i \neq j. \tag{6.16}$$

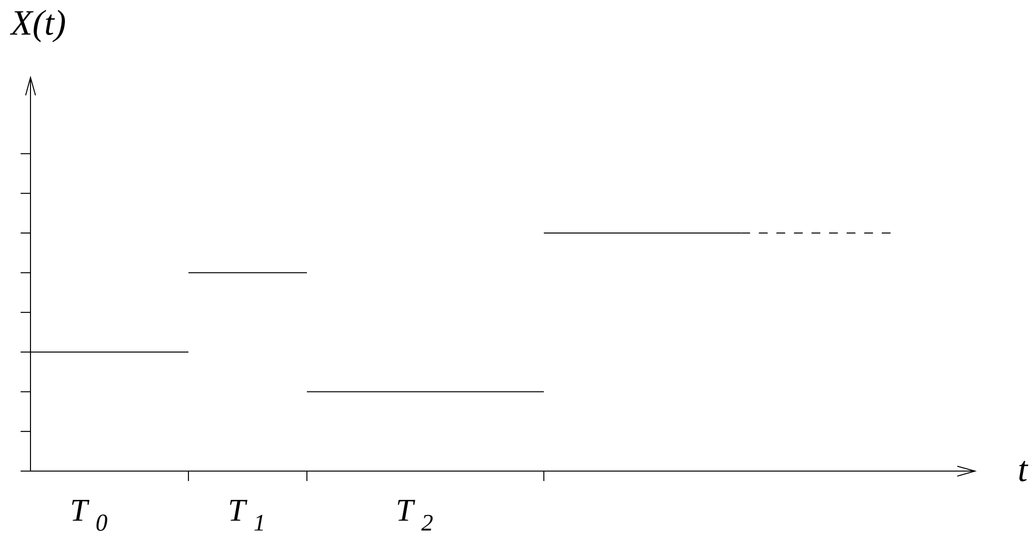
Summeringen i (6.16) sker alltså över händelser som medför att hopp sker till tillstånd j.

Bevis. Låt B_{ij} vara händelsen att någon av de A_{ik} som leder till tillstånd j inträffar, d.v.s. $B_{ij} = \cup_{k;S(A_{ik}=j)} A_{ik}$. Men tiden till att B_{ij} inträffar är då minimum av oberoende exponentialfördelade stokastiska variabler och blir, som vi tidigare sett också, exponentialfördelad med intensitet $q'_{ij} = \sum_{k;S(A_{ik})=j} q_{ik}$. Men vi har då exakt samma situation som i huvudsatsen och följsatsen är bevisad. □

Figur 6.3 illustrerar det allmänna uppträdandet av en Markovprocess.

Fortsättning exempel 6.4 I populationsexemplet 6.4 har vi en situation som beskrivs i följdsatsen. Vi kan direkt använda följdsatsen och ser att $\{X(t); t \geq 0\}$ är en Markovprocess med övergångsintensiteter

$$q_{ij} = \begin{cases} \sum_{k=1}^{i} \lambda = i\lambda & \text{om } j = i+1,\ i = 0, 1, \ldots \\ \sum_{k=1}^{i} \mu = i\mu & \text{om } j = i-1,\ i = 1, 2, \ldots \\ -i(\lambda + \mu) & \text{om } j = i,\ i = 1, 2, \ldots \\ 0 & \text{för övrigt} \end{cases} \tag{6.17}$$



Figur 6.3: Uppträdande av Markovprocesser

Detta är ett exempel på födelse-dödsprocesser som vi skall studera i avsnitt 6.8 □

6.2 Fördelningen för X(t)

I det tidsdiskreta fallet är fördelningen för X_n bestämd av övergångsmatrisen $\boldsymbol{P}$. I kontinuerlig tid finns ingen "första" matris $\boldsymbol{P}(t)$ men vi skall se att intensitetsmatrisen $\boldsymbol{Q}$ i viss utsträckning har övertagit denna roll. Av definitionen av intensiteter följer att

$$q_{ij} = \lim_{h \to 0+} \frac{p_{ij}(h)}{h} \text{ om } i \neq j$$
$$q_{ii} = \lim_{h \to 0+} \frac{p_{ii}(h) - 1}{h}$$

vilket i matrisform kan skrivas

$$\boldsymbol{Q} = \lim_{h \to 0+} \frac{\boldsymbol{P}(h) - \boldsymbol{I}}{h}, \tag{6.18}$$

där $\boldsymbol{I}$ är identitetsmatrisen. Vi kommer senare att se att $\boldsymbol{Q}$ är motsvarigheten till matrisen $\boldsymbol{P} - \boldsymbol{I}$ i diskret tid.

Av Chapman-Kolmogorovs sats fås

$$\boldsymbol{P}(t+h) = \boldsymbol{P}(t)\boldsymbol{P}(h) = \boldsymbol{P}(h)\boldsymbol{P}(t)$$
$$\Leftrightarrow$$
$$\boldsymbol{P}(t+h) - \boldsymbol{P}(t) = \boldsymbol{P}(t)(\boldsymbol{P}(h) - \boldsymbol{I}) = (\boldsymbol{P}(h) - \boldsymbol{I})\boldsymbol{P}(t)$$

$h \to 0+$ ger

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{P}(t), \tag{6.19}$$

vilket är Kolmogorovs *framåt-* resp. *bakåt-ekvationer*. Dessa differentialekvationer är synnerligen grundläggande inom Markovteorin, även om vi inte kommer att utnyttja dem så mycket. Bakåtekvationerna gäller alltid men kan, om processen är irreguljär, ha flera lösningar.

Framåtekvationerna kan konkret skrivas som ekvationssystemet

$$\begin{aligned}
p'_{00}(t) &= p_{00}(t)q_{00} + p_{01}(t)q_{10} + p_{02}(t)q_{20} + \cdots \\
p'_{10}(t) &= p_{10}(t)q_{00} + p_{11}(t)q_{10} + p_{12}(t)q_{20} + \cdots \\
&\vdots \\
p'_{ij}(t) &= p_{i0}(t)q_{0j} + p_{i1}(t)q_{1j} + p_{i2}(t)q_{2j} + \cdots \\
&\vdots
\end{aligned} \tag{6.20}$$

En irreguljär process där antalet hopp under ändlig tid är oändligt med positiv sannolikhet har – med positiv sannolikhet – en ändlig hopningspunkt av hopp och vad som händer efter denna hopningspunkt är ej bestämt av $\boldsymbol{Q}$. Framåtekvationerna kan sägas bestämda av vad som händer efter en tid, då en sådan hopningspunkt av hopp kan ha skett, och har därför inte alltid en lösning. Att bakåtekvationerna kan ha flera lösningar beror på att processen inte är tillräckligt specificerad efter en sådan hopningspunkt och att det finns flera sätt för definiera processen efter hopningspunkten så att bakåtekvationerna blir uppfyllda.

Framåt- och bakåtekvationerna definierar ett system av linjära differentialekvationer som kan lösas, åtminstone om tillståndsrummet ändligt. En vanligt hjälpmedel för att lösa ekvationerna är Laplacetransformationer, se nästa avsnitt.

En mycket formell lösning till ekvationssystemet

$$\boldsymbol{P}'(t) = \boldsymbol{Q}\boldsymbol{P}(t)$$

är

$$\boldsymbol{P}(t) = e^{\boldsymbol{Q}t}. \tag{6.21}$$

Högerledet i (6.21) skall tolkas som taylorutvecklingen

$$e^{\boldsymbol{Q}t} = \sum_{n=0}^{\infty} \frac{t^n \boldsymbol{Q}^n}{n!} \tag{6.22}$$

I konkreta fall är inte (6.22) till så stor nytta, eftersom det kräver beräkning av $\boldsymbol{Q}^n$ för alla n.

Man kan också sätta upp ett differentialekvationssystem för den absoluta sannolikhetsfördelningen $\boldsymbol{p}(t)$:

$$\boldsymbol{p}(t) = \boldsymbol{p}(0)\boldsymbol{P}(t).$$

Derivera med avseende på t,

$$\boldsymbol{p}'(t) = \boldsymbol{p}(0)\boldsymbol{P}'(t) = \boldsymbol{p}(0)\boldsymbol{P}(t)\boldsymbol{Q} = \boldsymbol{p}(t)\boldsymbol{Q}. \tag{6.23}$$

Vi har utnyttjat framåtekvationen, vilket är tillåtet om processen är reguljär.

Dessa ekvationer får vi också om vi låter h gå mot 0 i

$$P(X(t+h) = j) = \sum_{i \in \boldsymbol{E}} P(X(t) = i)P(X(t+h) = j \mid X(t) = i)$$

$$\Leftrightarrow$$

$$p_j(t+h) = \sum_i p_i(t)p_{ij}(h).$$

Men $p_{ij}(h) = q_{ij}h + o(h), i \neq j$ och $p_{jj}(h) = 1 + q_{jj}h + o(h)$ varför vi erhåller

$$p_j(t+h) - p_j(t) = \sum_i p_i(t)q_{ij}h + o(h). \tag{6.24}$$

Dividera med h och låt h gå mot 0

$$p_j'(t) = \sum_i p_i(t)q_{ij} \tag{6.25}$$

vilket ger j:te komponenten i (6.23).

6.3 Laplacetransformer

Ekvationssystemet (6.23) kan lösas, åtminstone teoretiskt, med hjälp av Laplacetransformer.

Definition 6.4 *Låt $p(t)$ vara en begränsad, kontinuerlig funktion definierad för ≥ 0. Med* Laplacetransformen *till $p(t)$ menas funktionen*

$$p^*(s) = \int_0^\infty p(t)e^{-st}dt, \quad s \geq 0.$$

Laplacetransformen är entydig, d.v.s två kontinuerliga funktioner som har samma Laplacetransform är identiska.

Exempel 6.5 *Om $p(t) = at^k e^{-\lambda t}$, k heltal ≥ 0 och $\lambda \geq 0$ erhåller vi* $p^*(s) = \int_0^\infty at^k e^{-\lambda t} e^{-st} dt = \big((\lambda + s)t = u\big) = \frac{a}{(\lambda+s)^{k+1}} \int_0^\infty u^k e^{-u} du = \frac{ak!}{(\lambda+s)^{k+1}}$ □

Om funktionen $p(t)$ är deriverbar erhåller vi Laplacetransformen av derivatan som

$$\begin{aligned}(p')^*(s) = \int_0^\infty p'(t)e^{-st}dt &= \text{(partiell integration)} \\ &= [p(t)e^{-st}]_0^\infty + s\int_0^\infty p(t)e^{-st}dt = -p(0) + sp^*(s)\end{aligned}$$

Om vi nu multiplicerar vänster- och högerleden i ekvationssystemet (6.20) och integrerar från 0 till oändligheten får vi således systemet

$$\begin{aligned} sp^*_{00}(s) - p_{00}(0) &= p^*_{00}(s)q_{00} + p^*_{01}(s)q_{10} + p^*_{02}(s)q_{20} + \cdots \\ sp^*_{10}(s) - p_{10}(0) &= p^*_{10}(s)q_{00} + p^*_{11}(s)q_{10} + p^*_{12}(s)q_{20} + \cdots \\ &\vdots \\ sp^*_{ij}(s) - p_{ij}(0) &= p^*_{i0}(s)q_{0j} + p^*_{i1}(s)q_{1j} + p^*_{i2}(s)q_{2j} + \cdots \\ &\vdots \end{aligned} \tag{6.26}$$

eller i matrisform

$$s\boldsymbol{P}^*(s) - \boldsymbol{I} = \boldsymbol{P}^*(s)\boldsymbol{Q} \text{ dvs } \boldsymbol{P}^*(s)(s\boldsymbol{I} - \boldsymbol{Q}) = \boldsymbol{I} \tag{6.27}$$

där $\boldsymbol{P}^*(s)$ är matrisen av Laplacetransformer, $\boldsymbol{P}^*(s) = (p^*_{ij}(s))_{i,j\in\boldsymbol{E}}$. Notera att $p_{ij}(0) = 0$ för $i \neq j$ och $p_{ii}(0) = 1$ alla i.

(6.27) kan fås genom att formellt bilda laplactransformen av vänster- och högerled i (6.23).

Ovanstående definierar ett linjärt ekvationssystem för Laplacetransformerna, som alltså i princip kan lösas. Genom att invertera Laplacelösningen kan övergångssannolikheterna $p_{ij}(t)$ fås.

Lösningen till 6.27 kan skrivas

$$\boldsymbol{P}^*(s) = (s\boldsymbol{I} - \boldsymbol{Q})^{-1}.$$

Högerledet kan beräknas och de olika elementen i den resulterande matrisen ger Laplacetransformerna till $p_{ij}(t)$. Genom att invertera Laplacetransformen beräknar vi sedan $p_{ij}(t)$.

Exempel 6.6 Betrakta exempel 6.2 på sidan 47. Intensitetsmatrisen är

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{pmatrix}$$

se formel (6.6). Laplacetransformen kan alltså beräknas ur

$$\begin{aligned} p^*(s) = (s\boldsymbol{I} - \boldsymbol{Q})^{-1} &= \begin{pmatrix} s+\lambda_0 & -\lambda_0 \\ -\lambda_1 & s+\lambda_1 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{\lambda_1}{(\lambda_0+\lambda_1)s} + \frac{\lambda_0}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_1+s)} & \frac{\lambda_0}{(\lambda_0+\lambda_1)s} - \frac{\lambda_0}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_1+s)} \\ \frac{\lambda_1}{(\lambda_0+\lambda_1)(\lambda_1+\lambda_1+s)} & \frac{\lambda_0}{(\lambda_0+\lambda_1)s} + \frac{\lambda_1}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_1+s)} \end{pmatrix}. \end{aligned}$$

Funktionen $1/s$ är Laplacetransformen av funktionen $f(t) = 1$ och $\frac{1}{\lambda_0+\lambda_1+s}$ är Laplacetransformen av funktionen $e^{-(\lambda_0+\lambda_1)t}$, se exempel 6.5. Härav fås att

$$\begin{aligned} p_{00}(t) &= \frac{\lambda_1}{\lambda_0+\lambda_1} + \frac{\lambda_0}{\lambda_0+\lambda_1}e^{-(\lambda_0+\lambda_1)t} & p_{01}(t) &= \frac{\lambda_0}{\lambda_0+\lambda_1} - \frac{\lambda_0}{\lambda_0+\lambda_1}e^{-(\lambda_0+\lambda_1)t} \\ p_{10}(t) &= \frac{\lambda_1}{\lambda_0+\lambda_1} - \frac{\lambda_1}{\lambda_0+\lambda_1}e^{-(\lambda_0+\lambda_1)t} & p_{11}(t) &= \frac{\lambda_0}{\lambda_0+\lambda_1} + \frac{\lambda_1}{\lambda_0+\lambda_1}e^{-(\lambda_0+\lambda_1)t} \end{aligned}$$

I detta fall kunde vi också ha utnyttjat att $p_{01}(t) = 1 - p_{00}(t)$ och $p_{11}(t) = 1 - p_{10}(t)$ för att beräkna två av de okända sannolikheterna.

Istället för att använda matrisapparaten kan man naturligtvis sätta upp ekvationssystemet

$$\begin{aligned} sp^*_{00}(s) - 1 &= -\lambda_0 p^*_{00}(s) + \lambda_1 p^*_{01}(s) \\ sp^*_{01}(s) &= \lambda_0 p^*_{00}(s) - \lambda_1 p^*_{01}(s) \\ sp^*_{10}(s) &= -\lambda_0 p^*_{10}(s) + \lambda_1 p^*_{11}(s) \\ sp^*_{11}(s) - 1 &= \lambda_0 p^*_{10}(s) - \lambda_1 p^*_{11}(s) \end{aligned}$$

och lösa ut transformerna. □

6.4 Poissonprocessen

Poissonprocessen kan definieras på flera olika sätt. Vi väljer en ”markovsk” definition.

Definition 6.5 $\{N(t),\, t \geq 0\}$ *är en* Poissonprocess *med intensitet* λ, Po(λ)-*process om* $N(t)$ *är en Markovprocess med* $N(0) = 0$ *och med övergångsintensiteter*

$$q_{ij} = \begin{cases} \lambda, & j = i+1, \\ -\lambda, & j = i, \\ 0, & \textit{annars.} \end{cases}$$

Betrakta följande situation. Händelser inträffar oberoende av varandra och med intensitet λ, d.v.s. tiden från en händelse till nästa är Exp(λ). Låt $N(t)$ vara antalet händelser i tidsintervallet $(0, t)$. Från definitionen och karaktäriseringssatsen 6.2 på sidan 49 kan man se att $\{N(t); t \geq 0\}$ är en Poissonprocess med intensitet λ. Vi skall visa att $N(t)$ är Poissonfördelad.

Processen illustrerad i figur 6.4 är ett exempel på hur en realisering av en Poissonprocess kan se ut.

Sats 6.3 *Låt* $\{N(t),\, t \geq 0\}$ *vara en stokastisk process med tillståndsrum*

$$E = \{0, 1, \ldots\}.$$

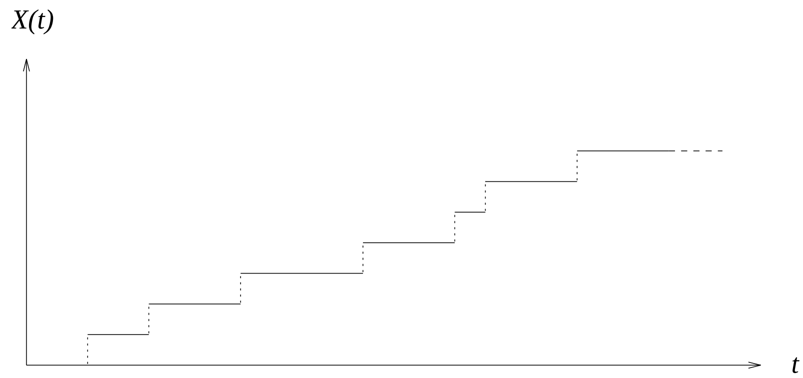
Följande uttalanden är ekvivalenta.

a) $\{N(t); t \geq 0\}$ *är en* Po(λ)-*process, enl. definitionen ovan.*

b) $\{N(t); t \geq 0\}$ *är en Markovprocess med* $N(0) = 0$ *och med övergångssannolikheter*

$$p_{ij}(t) = \begin{cases} \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t}, & j \geq i, \\ 0, & j < i. \end{cases}$$

c) $\{N(t); t \geq 0\}$ *är en stokastisk process sådan att*



Figur 6.4: Realisering av en Poissonprocess.

1) *För alla* $0 \le s_1 < t_1 \le s_2 < t_2 \le \ldots \le s_n < t_n$ *gäller att* $N(t_1) - N(s_1), N(t_2) - N(s_2), \ldots, N(t_n) - N(s_n)$ *är oberoende stokastiska variabler.*
2) $N(t) - N(s)$ *är* $\mathrm{Po}(\lambda(t-s))$*-fördelad.*
3) $N(0) = 0$.

Bevis.

a) $\Rightarrow$ *b)*

Från Kolmogorovs framåtekvationer (6.20) på sidan 54 erhåller vi

$$p'_{ij}(t) = \lambda p_{i,j-1}(t) - \lambda p_{ij}(t). \tag{6.28}$$

För $j < i$ är $p_{ij}(t) = 0$ då processen aldrig förflyttar sig nedåt. Det ger att

$$p'_{ii}(t) = -\lambda p_{ii}(t).$$

Eftersom $p_{ii}(0) = 1$ fås

$$p_{ii}(t) = e^{-\lambda t} = \frac{(\lambda t)^0}{0!}\, e^{-\lambda t}.$$

Ekvationerna (6.28) kan sedan lösas rekursivt för $j = i+1, i+2, \ldots$ Vi nöjer oss här med att göra första steget. För $j = i+1$ fås

$$p'_{i,i+1}(t) = -\lambda(p_{i,i+1}(t) - p_{ii}(t)) = -\lambda(p_{i,i+1}(t) - e^{-\lambda t})$$

som har lösningen

$$p_{i,i+1}(t) = \lambda t e^{-\lambda t} + C e^{-\lambda t}.$$

Eftersom $p_{i,i+1}(0) = 0$ fås

$$p_{i,i+1}(t) = \lambda t e^{-\lambda t} = \frac{(\lambda t)^1}{1!}\, e^{-\lambda t}.$$

b) $\Rightarrow$ *c*) Oberoende ökningar följer av att $p_{ij}(t)$ beror av i och j endast genom $j-i$.

Det följer av tidshomogeniteten att $N(t)-N(s)$ har samma fördelning som $N(t-s)-N(0) = N(t-s)$. Men $P(N(t-s)=j) = p_{0j}(t-s) = \frac{(\lambda(t-s))^j}{j!}\,e^{-\lambda(t-s)}, j \geq 0$.

c) $\Rightarrow$ *a*)

Antag att $0 \leq t_1 < t_2 < \ldots < t_n$. Då har vi att

$$
\begin{aligned}
&P(N(t_n)=i_n \mid N(t_1)=i_1, N(t_2)=i_2, \ldots, N(t_{n-1})=i_{n-1}) \\
&= P(N(t_n)-N(t_{n-1})=i_n-i_{n-1} \mid N(t_1)-N(0)=i_1, \ldots, N(t_{n-1})-N(t_{n-2})=i_{n-1}-i_{n-2})
\end{aligned}
$$

På grund av oberoende ökningar är den sista sannolikheten lika med $P(N(t_n)-N(t_{n-1})=i_n-i_{n-1})$, d.v.s. oberoende av processens värden i tidpunkter $t_1, t_2, \ldots, t_{n-2}$, vilket betyder att processen är markovsk.

Vi har dessutom att

$$
\begin{aligned}
&P(N(t+h)=j \mid N(t)=i) = P(N(t+h)-N(t)=j-i \mid N(t)-N(0)=i) \\
&= (\text{oberoende ökningar}) = P(N(t+h)-N(t)=j-i) = \frac{(\lambda h)^{j-i}}{(j-i)!}\,e^{-\lambda h}
\end{aligned}
$$

om $j \geq i$. Härav får vi att processen är tidshomogen och genom direkt derivering av ovanstående att

$$
q_{ij} = p'_{ij}(0) = \begin{cases} -\lambda & \text{om } j = i \\ \lambda & \text{om } j = i+1 \\ 0 & \text{annars} \end{cases}
$$

□

Eftersom Poissonprocessen är en reguljär Markovprocess följer att tiderna mellan hopp är oberoende och Exp(λ)-fördelade. Man kan också inse detta på följande sätt. Låt T vara tiden tills tills processen når tillstånd 1 d.v.s. tid till första hoppet. Då är $P(T>t) = P(N(t)=0) = e^{-\lambda t}$ vilket betyder att T är Exp(λ). Men intensiteten för hopp är alltid λ så i vilket tillstånd processen än befinner sig i är tiden tills hopp sker Exp(λ).

Tiden T_n tills en Poissonprocess når nivån n är summan av n oberoende exponetialfördelade variabler med intensitet λ, eftersom tiderna mellan hopp är exponentialfördelade. Följande sats visar att T_n är gammafördelad.

Sats 6.4

a) Låt T_n vara tiden tills en Poissonprocess med intensitetet λ når nivån n. Då är T gammafördelad, $\Gamma(n,\lambda)$, d.v.s. har täthetsfunktion

$$
f_{T_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!}\,e^{-\lambda t} \ \text{om } t \geq 0
$$

Det innebär att summan av n oberoende Exp(λ)-fördelade stokastiska variabler är $\Gamma(n,\lambda)$-fördelad.

b) Låt N vara ffg(p)-fördelad, d.v.s.

$$P(N = n) = (1-p)^{n-1}p, \quad \text{för } n = 1, 2, \ldots,$$

och oberoende av $U_1, U_2, \ldots$, där $U_1, U_2, \ldots$ är oberoende, alla Exp(λ). Då gäller att den stokastiska summan $T_N = U_1 + \ldots + U_N$ *är Exp($p\lambda$)-fördelad.*

Bevis. a) Att T_n är större än t är detsamma som att Poissonprocessen vid tid t ej har nått upp till värdet n, d.v.s. $P(T_n > t) = P(N(t) < n)$, där $\{N(t); t \geq 0\}$ är Poissonprocessen ifråga. Men $N(t)$ är Po(λt). Alltså har vi att

$$\begin{aligned} F_{T_n}(t) = P(T_n \leq t) = 1 - P(T_n > t) = 1 - P(N(t) \leq n-1) \\ = 1 - \sum_{k=0}^{n-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!} \end{aligned}$$

Vi deriverar för att få täthetsfunktionen.

$$\begin{aligned} f_{T_n}(t) = F'_{T_n}(t) &= -\sum_{k=0}^{n-1}\left(-\lambda e^{-\lambda t}\frac{(\lambda t)^k}{k!} + e^{-\lambda t}k\frac{\lambda^k t^{k-1}}{k!}\right) \\ &= \lambda e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} - \lambda e^{-\lambda t}\sum_{k=1}^{n-1}\frac{(\lambda t)^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda t}\left(\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} - \sum_{j=0}^{n-2}\frac{(\lambda t)^j}{j!}\right) \\ &= \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} = e^{-\lambda t}\frac{\lambda^n t^{n-1}}{(n-1)!} \end{aligned}$$

b) Låt T_N ha tätheten $f_{T_N}(t)$. Vi betingar med $N = n$. Med en viss generalisering av lagen om total sannolikhet fås

$$f_{T_N}(t) = \sum_{n=1}^{\infty} P(N = n) f_{T_n}(t) = \sum_{n=1}^{\infty}(1-p)^{n-1}p \cdot \frac{\lambda^n t^{n-1}}{(n-1)!}e^{-\lambda t}$$

$$\begin{aligned} = p\lambda e^{-\lambda t}\sum_{n=1}^{\infty}(1-p)^{n-1}\frac{\lambda^{n-1}t^{n-1}}{(n-1)!} = p\lambda e^{-\lambda t}\sum_{k=0}^{\infty}(1-p)^k\frac{\big((1-p)\lambda t\big)^k}{k!} \\ = p\lambda e^{-\lambda t}e^{(1-p)\lambda t} = p\lambda e^{-p\lambda t}, \end{aligned}$$

d.v.s. T_N är Exp($p\lambda$)-fördelad. □

Summan av oberoende Poissonprocesser är en Poissonprocess med motsvarande summa av intensiteter. Detta följer eftersom samma resultat gäller för Poissonfördelningen och eftersom ”oberoende ökningar” lever kvar.

Poissonprocessen kan generaliseras på olika sätt. En generalisering är att låta intensiteten vara beroende av tiden istället för att vara konstant. Om vi låter $N(t)$ vara antalet inkomna telefonsamtal till en växel, är det naturligt att tänka sig att teletrafikens intensitet varierar med tidpunkten på dygnet, d.v.s. den konstanta intensiteten är ersatt av en intenstitetsfunktion $\lambda(t)$. En sådan process blir dock inte tidshomogen och kallas *inhomogen Poissonprocess*.

En annan generalisering är att låta intensiteten vara stokastisk. Först slumpas då en intensitsfunktion $\lambda(t)$ ut och sedan bildas en Poissonprocess med denna intensitet. Den process som detta förfarande mynnar ut kallas en *dubbelt stokastisk Poissonprocess* eller *Coxprocess*. Vi återkommer till denna generalisering i exempel 7.5 på sidan 75.

Ytterligare en generalisering är att låta tiderna mellan hopp vara generell och ej nödvändigtvis exponentialfördelade. Om tiderna är oberoende och likafördelade kallas processen en *förnyelseprocess*. Denna process blir ej markovsk om tiderna ej är exponentialfördelade. Om tiderna är exponentialfördelade blir processen enligt vad vi tidigare visat, en Poissonprocess.

Poissonprocesser och förnyelseprocesser är exempel på *punktprocesser*. För en punktprocess finns en slumpmekanism som generar en ändlig eller uppräknelig punktmängd på tidsaxeln. Dessa genererade punkter är processens hopptider.

I nästa avsnitt skall vi så betrakta en generalisering där markoviteten bibehålles.

6.5 Födelseprocesser

En födelseprocess är en generalisering av en Poissonprocess. I en födelseprocess tillåter man att hoppintensiteterna är olika.

Definition 6.6 *En Markovprocess $\{X(t); t \geq 0\}$ med $\boldsymbol{E} = \{0, 1, 2, \dots\}$ som tillståndsrum är en födelseprocess med födelseintensiteter $\lambda_i,\ i = 0, 1, 2, \dots$ om dess övergångsintensiteter är*

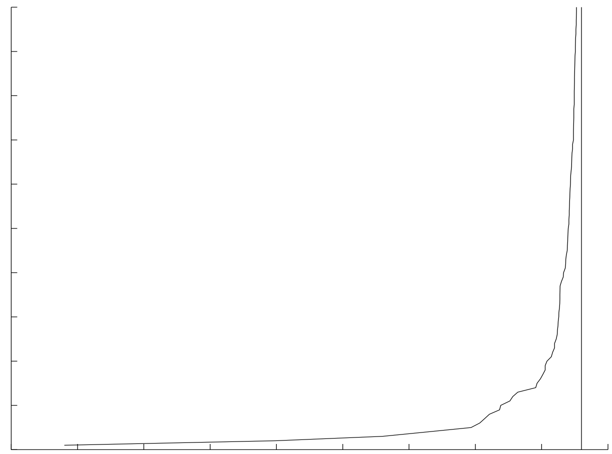
$$q_{ij} = \begin{cases} \lambda_i & \text{om } j = i+1 \\ -\lambda_i & \text{om } j = i \\ 0 & \text{annars.} \end{cases}$$

En födelseprocess går alltså ett steg uppåt vid varje hopp, men intensiteten beror på vilket tillstånd processen befinner sig i.

Om processen startar i tillstånd 0 fås fördelningen för $X(t)$, d.v.s $\boldsymbol{p}(t)$, genom att lösa ekvationssystemet (se ekvation (6.25) på sidan 55)

$$p_i'(t) = -\lambda_i p_i(t) + \lambda_{i-1} p_{i-1}(t),\ i = 0, 1, 2, \dots \tag{6.29}$$

och $p_0(0) = 1$ och $p_i(0) = 0, i \geq 1$.



Figur 6.5: Exploderande födelseprocess.

Förväntad tid i tillstånd i är λ_i, och om processen startar i tillstånd 0, blir förväntad tid att nå tillstånd n

$$\sum_{i=0}^{n-1} \frac{1}{\lambda_i}. \tag{6.30}$$

Om nu $\sum_{i=0}^{\infty} \frac{1}{\lambda_i}$ är ändlig, når kedjan oändligheten inom ändlig förväntad tid. Detta är ett exempel på ”explosion” och en irreguljär process. Man kan också visa omvändningen, nämligen att processen är reguljär om serien (6.30) divergerar då $n \to \infty$. Figur 6.5 är ett realisering av en exploderande födelseprocess.

Exempel 6.7 (Yule-processen) $\{X(t); t \geq 0\}$ där $X(t)$ är antalet individer vid tidpunkt t i population där var och en av individerna, med intensitet λ och oberoende av varandra, ger upphov till en ny individ är en födelseprocess med födelseintensiteter $\lambda_i = i\lambda$, se exempel 6.4 på sidan 52. Vi antar att populationen startar med 1 individ. Exploderar en sådan population? Svaret är nej, ty

$$\sum_{i=1}^{\infty} \frac{1}{i\lambda} = \frac{1}{\lambda} \sum_{i=1}^{\infty} \frac{1}{i}$$

är divergent. □

6.6 Absorption

Ett absorptionstillstånd i är ett tillstånd från vilket man ej kan gå till något annat, vilket innebär att $q_{ij} = 0$ för alla j. Begreppet A-kedja definieras precis som i fallet diskret tid. Låt som tidigare a_{ij} vara sannolikheten att absorberas i tillstånd j givet start i tillstånd i och t_i vara förväntad tid tills kedjan når ett absorberande tillstånd. Vi visade redan i avsnittet 3.3 hur man i det generella fallet finner ekvationssystem för att beräkna absorptionssannolikheterna och de förväntade tiderna. De förväntade uppehållstiden i tillstånd i är $\dfrac{1}{q_i}$ och uthoppssannolikheterna är $\tilde{p}_{ij} = \dfrac{q_{ij}}{q_i}$. Från ekvationssystemen (3.18) och (3.17) på sidan 18 erhåller vi alltså följande sats.

Sats 6.5 *Låt a_{ij} vara sannolikheten att absorberas i tillstånd j vid start i tillstånd i och t_i förväntade tider tills absorption i en A-kedja med intensitetsmatris $\boldsymbol{Q}$. Då gäller för alla $j \in \boldsymbol{A}$*

$$a_{ij} = \frac{q_{ij}}{q_i} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \frac{q_{ik}}{q_i} a_{kj}, \quad i \in \boldsymbol{G} \tag{6.31}$$

och

$$t_i = \frac{1}{q_i} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \frac{q_{ik}}{q_i} t_k, \quad i \in \boldsymbol{G} \tag{6.32}$$

Exempel 6.8 Låt oss betrakta tillförlitlighetsexemplet 6.1 på sidan 51 och söka förväntad tid tills systemet brister givet att det startar med två hela komponenter. Vi gör om tillstånd E_3 till ett absorberande tillstånd och med beteckningar som i satsen erhåller vi ekvationssystemet

$$\begin{aligned}
t_0 &= \frac{1}{\lambda_1 + \lambda_2} + \frac{\lambda_2}{\lambda_1 + \lambda_2} t_1 + \frac{\lambda_1}{\lambda_1 + \lambda_2} t_2 \\
t_1 &= \frac{1}{\mu + \lambda_1} + \frac{\mu}{\mu + \lambda_1} t_0 \\
t_2 &= \frac{1}{\mu + \lambda_2} + \frac{\lambda_1}{\mu + \lambda_2} t_0
\end{aligned}$$

vilket efter en del räknande ger

$$t_0 = \frac{\mu^2 + 2\,\mu\,\lambda_1 + {\lambda_1}^2 + 2\,\lambda_2\,\mu + \lambda_1\,\lambda_2 + {\lambda_2}^2}{\lambda_1\,\lambda_2\,(2\,\mu + \lambda_1 + \lambda_2)}$$

□

6.7 Stationaritet och konvergens

Vi skall nu studera det asymptotiska uppträdandet av Markovprocesser. Som i det tidsdiskreta fallet visar det sig att de stationära fördelningarna är fundamentala.

Definition 6.7 *En Markovprocess med övergångsmatris $\boldsymbol{P}(t)$ säges ha* stationär fördelning $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ *om $\boldsymbol{\pi}$ är en fördelning, d.v.s. om $\pi_0 + \pi_1 + \ldots = 1$, $\pi_n \geq 0$, och om*

$$\boldsymbol{\pi} = \boldsymbol{\pi} \boldsymbol{P}(t), \quad t \geq 0. \tag{6.33}$$

Om vi deriverar (6.33) får vi $\mathbf{0} = \boldsymbol{\pi}\boldsymbol{P}'(t) = \boldsymbol{\pi}\boldsymbol{P}(t)\boldsymbol{Q} = \boldsymbol{\pi}\boldsymbol{Q}$ åtminstone formellt. Om tillståndsrummet är ändligt uppstår inga problem, men om tillståndsrummet är oändligt får vi kräva att processen är reguljär.

Sats 6.6 $\boldsymbol{\pi}$ *är en stationär fördelning till en reguljär Markovprocess om och endast om*

$$\boldsymbol{\pi}\boldsymbol{Q} = \mathbf{0} \tag{6.34}$$

Bevis. Att $\boldsymbol{\pi}$ stationär medför $\boldsymbol{\pi}\boldsymbol{Q} = \mathbf{0}$ visade vi ovan. Omvändningen följer av att om $\boldsymbol{\pi}$ är startfördelningen och $\boldsymbol{p}(t) = \boldsymbol{\pi}\boldsymbol{P}(t)$ så är $\boldsymbol{p}'(t) = \boldsymbol{\pi}\boldsymbol{P}'(t) = \boldsymbol{\pi}\boldsymbol{Q}\boldsymbol{P}(t) = \mathbf{0}$. Således är $\boldsymbol{p}(t)$ konstant och därmed lika med $\boldsymbol{\pi}$. □

Ekvationen (6.34) kan efter matrisutveckling skrivas

$$q_i\pi_i = q_{0i}\pi_0 + q_{1i}\pi_1 + q_{2i}\pi_2 + \cdots + q_{i-1,i}\pi_{i-1} + q_{i+1,i}\pi_{i+1} + \cdots, \quad i = 0, 1, 2, \ldots \tag{6.35}$$

Ekvationerna (6.35) kallas de *globala balansekvationerna*. Dessa har följande flödestolkning. Vänsterledet är "flödet" ut ur tillståndet i; sannolikheten att vara i tillståndet gånger intensiteten ut. Termerna i högerledet anger "flödet" in till tillstånd i från tillstånden $j = 1, 2, \ldots, j \neq i$. Summan är totala "flödet" in; sannolikheterna att vara i tillstånden gånger intensiteten in till tillstånd i. För att processen skall vara i balans måste ut- och inflödet vara desamma.

Sats 6.7 *En ändlig Markovprocess har en stationär fördelning.*

Bevis. Betrakta den Markovkedja i diskret tid som har övergångssannolikheter lika med uthoppssannolikheterna i Markovprocessen, d.v.s. som har övergångsmatris $\widetilde{\boldsymbol{P}} = (\tilde{p}_{ij})_{i,j\in E}$ där $\tilde{p}_{ij} = \frac{q_{ij}}{q_i}$ och $\tilde{p}_{ii} = 0$. Denna Markovkedja är ändlig och har enligt sats 5.1 på sidan 31 en stationär fördelning $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \tilde{\pi}_2, \ldots, \tilde{\pi}_N)$.

Sätt $\pi_i = \frac{\tilde{\pi}_i/q_i}{\tilde{\pi}_{sum}}$ där $\tilde{\pi}_{sum} = \sum_j \frac{\tilde{\pi}_j}{q_j}$. Då är $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$ är en sannolikhetsfördelning eftersom $\pi_i \geq 0$ och $\sum_i \pi_i = 1$.

Eftersom $\tilde{\boldsymbol{\pi}}$ är en stationär fördelning till $\widetilde{\boldsymbol{P}}$ och $\tilde{p}_{jj} = 0$ gäller

$$\tilde{\pi}_j = \tilde{p}_{0j}\tilde{\pi}_0 + \tilde{p}_{1j}\tilde{\pi}_1 + \cdots + \tilde{p}_{Nj}\tilde{\pi}_N = \sum_i \tilde{p}_{ij}\tilde{\pi}_i = \sum_{i\neq j} \tilde{p}_{ij}\tilde{\pi}_i = \sum_{i\neq j} \frac{q_{ij}}{q_i}\tilde{\pi}_i \tag{6.36}$$

Eftersom $q_{ii}/q_i = -1$ ser man lätt att (6.36) kan skrivas

$$\sum_i \frac{q_{ij}}{q_i}\tilde{\pi}_i = 0 \tag{6.37}$$

eller efter division med $\tilde{\pi}_{sum}$

$$0 = \sum_i \frac{q_{ij}}{q_i}\frac{\tilde{\pi}_i}{\tilde{\pi}_{sum}} = \sum_i q_{ij}\frac{\tilde{\pi}_i/q_i}{\tilde{\pi}_{sum}} = \sum_i q_{ij}\pi_i \tag{6.38}$$

vilket innebär att $\boldsymbol{\pi}$ är en stationär fördelning till processen. □

Sats 6.8 *En ändlig, irreducibel Markovprocess* $\{X(t); t \geq 0\}$ *är ergodisk, d.v.s. har en gränsfördelning som är oberoende av startfördelningen, och gränsfördelningen ges av den entydiga stationära fördelningen* $\boldsymbol{\pi}$. *Kvoten* $\pi_i = \dfrac{1/q_i}{E(T_i)}$ *där* T_i *är tiden från det processen gått in i tillstånd i tills det den nästa gång går in i tillstånd i. Förväntad tid i tillstånd* j *mellan två inträden i tillstånd* i *är* $\pi_j/(q_i\pi_i)$.

Bevis. Vi bevisar satsen på samma sätt som den diskreta motsvarigheten, sats 5.2 och skisserar därför bara beviset här.

Låt således $Z(t) = \big(X(t), Y(t)\big)$ där $\{Y(t), t \geq 0\}$ är en av $\{X(t); t \geq 0\}$ oberoende process med samma intensitetsmatris och med startvektor $\boldsymbol{\pi}$. $\{Z(t), t \geq 0\}$ är en ändlig irreducibel Markovprocess och om vi låter T vara tiden tills denna process hamnar på "diagonalen", t.ex. tillstånd $(1, 1)$, kommer T att vara ändlig med sannolikhet 1. Efter den stokastiska tiden T kommer Y- och X-processen att stokastiskt uppföra sig på samma sätt och därför har de två processerna samma gränsfördelning, som således är den stationära eftersom fördelningen för $Y(t)$ är $\boldsymbol{\pi}$ för alla t.

Betrakta nu den inbäddade hoppkedjan som har stationär fördelning $\tilde{\boldsymbol{\pi}}$. Vi betraktar tillstånd 1, de andra tillstånden behandlas på samma sätt.

Låt U_i vara antal gånger hoppkedjan är i tillstånd i mellan två besök i tillstånd 1 och låt $u_i = E(U_i)$. Förväntade antalet gånger som hopprocessen är i tillstånd i under denna tidscykel kan enligt ekvation (5.11) på sidan 34 skrivas $u_i = \tilde{\pi}_i/\tilde{\pi}_1$. Låt $V_1, V_2, \ldots, V_{U_i}$ vara de tider som den ursprungliga processen är i tillstånd i efter de U_i olika hoppen in till i. Låt Z_i vara den totala tiden i tillstånd i mellan två inträden i 1. Enligt sats 2.6 på sidan 8 är $E(Z_i) = E(\sum_{k=1}^{U_i} V_k) = E(U_i)\frac{1}{q_i} = u_i\frac{1}{q_i} = (\tilde{\pi}_i/\tilde{\pi}_1)\frac{1}{q_i}$, eftersom $V_1, V_2, \ldots, V_{U_i}$ alla är $\mathrm{Exp}(q_i)$ och oberoende av U_i (följer av Markovegenskapen). Totala förväntade tiden före återkomst till tillstånd 1 är därför $E(T_1) = \sum_k E(Z_k) = \sum_k \dfrac{\tilde{\pi}_k}{\tilde{\pi}_1 q_k} = \dfrac{\tilde{\pi}_{sum}}{\tilde{\pi}_1} = \dfrac{\tilde{\pi}_{sum}}{\tilde{\pi}_1/q_1}/q_1 = \dfrac{1}{q_1\pi_1}$, vilket ger $\pi_1 = \dfrac{1/q_1}{E(T_1)}$.

Totala förväntade tiden i tillstånd j mellan två inträden i tillstånd 1 är enligt ovan $E(Z_j) = \dfrac{\tilde{\pi}_j}{\tilde{\pi}_1 q_j} = \dfrac{\tilde{\pi}_j/q_j}{\tilde{\pi}_{sum}} \cdot \dfrac{\tilde{\pi}_{sum}}{\tilde{\pi}_1/q_1} \cdot \dfrac{1}{q_1} = \dfrac{\pi_j}{\pi_1 q_1}$ vilket skulle visas. □

Frågan om periodicitet dyker aldrig upp i det kontinuerliga fallet. Det förenklar studiet av en Markovprocess asymptotiska uppträdande.

Om antalet tillstånd är oändligt, är en irreducibel Markovkedja inte säkert ergodisk, utan vi har följande motsvarighet till sats 5.6 på sidan 42.

Sats 6.9 *Låt* $\boldsymbol{Q}$ *vara intensitetsmatrisen till en irreducibel, reguljär Markovprocess. Betrakta följande ekvationssystem där* $\boldsymbol{x} = (x_0, x_1, \ldots)$.

$$\begin{aligned} \boldsymbol{x}\boldsymbol{Q} &= \boldsymbol{0} \\ x_i &\geq 0 \quad i = 0, 1, \ldots \end{aligned} \tag{6.39}$$

Då gäller

a) Om (6.39) har en lösning sådan att

$$0 < \sum_{i=0}^{\infty} x_i < \infty \tag{6.40}$$

så har Markovprocessen en entydig stationär fördelning som ges av $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ *där*

$$\pi_i = \frac{x_i}{\sum_{j=0}^{\infty} x_j} \tag{6.41}$$

Markovkedjan är ergodisk med gränsfördelning som ges av $\boldsymbol{\pi}$*.*

b) Om (6.39) har en lösning som satisfierar

$$\sum_{i=0}^{\infty} x_i = \infty \tag{6.42}$$

så finns ingen stationär fördelning och kedjan är inte ergodisk.

Bevis. Visas ej, men notera att satsen är nästan ordagrant lika med motsvarande sats i diskret tid; ersätt matrisen $\boldsymbol{P} - \boldsymbol{I}$ med $\boldsymbol{Q}$.

Precis som i fallet diskret tid, kan man i satserna 6.7 på sidan 64 och 6.8 på föregående sida ersätta villkoret ”ändlig Markovprocess” med ”positivt rekurrent Markovprocess”, d.v.s. att förväntad återkomst till ett tillstånd är ändligt. Dessutom gäller att en process är positivt rekurrent om den är reguljär och har en stationär fördelning.

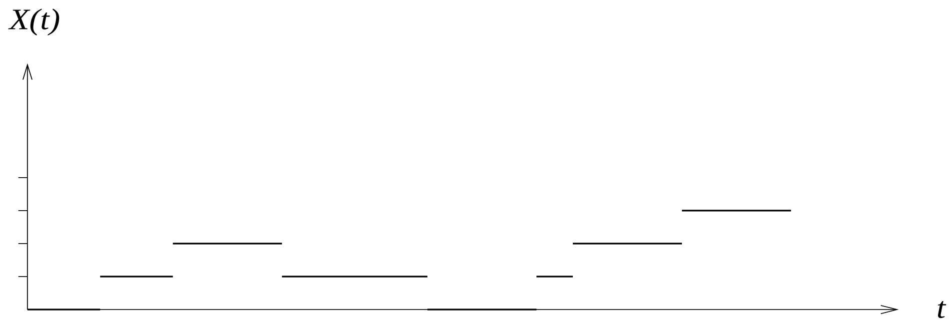
6.8 Födelse-döds-processer

Definition 6.8 *En* födelse-döds-process $\{X(t); t \geq 0\}$ *med* födelseintensiteter λ_0, $\lambda_1, \ldots$ *och* dödsintensiteter $\mu_1, \mu_2, \ldots$ *är en Markovprocess med övergångsintensiteter*

$$q_{ij} = \begin{cases} \lambda_i, & j = i+1, i = 0, 1, 2, \ldots \\ \mu_i, & j = i-1, i = 1, 2, \ldots \\ -\lambda_i - \mu_i, & j = i, i = 1, 2, \ldots \\ -\lambda_0, & i = j = 0 \\ 0, & \textit{annars,} \end{cases}$$

d.v.s. dess intensitetsmatris $\boldsymbol{Q}$ *är given av*

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & & \end{pmatrix}.$$



Figur 6.6: Illustration av en realisering av en födelse-döds-process.

En födelse-döds-process är således en Markovprocess som hoppar precis ett steg upp eller ned. Processen illustrerad i figur 6.6 är ett exempel på hur en realisering av en födelse-döds-process kan se ut. Namnet födelse-döds-process kommer av att processen har används för att beskriva populationers utveckling. Ett hopp uppåt svarar mot att en individ föds och ett hopp nedåt mot att en individ dör. Vi kommer ibland att kalla uppåthoppen för *"födelser"* och nedåthoppen för *"dödslar"*.

Processens uppförande kommer naturligt nog att bero på förhållandet mellan födelse- och döds-intensiteterna. Det kommer att visa sig lämpligt att definiera

$$\rho_0 = 1 \quad \text{och} \quad \rho_n = \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n}.$$

Vi såg tidigare att en födelseprocess kan explodera. En födelse-döds-process kan också vara irreguljär och explodera. Antag att processen startar i tillstånd 0 och låt t_n vara förväntad tid tills processen för första gången når värdet n. Om vi sätter $t_\infty = \lim_{n\to\infty} t_n$ är ett nödvändigt och tillräckligt villkor för att processen skall vara reguljär, att $t_\infty = \infty$, jämför födelseprocessen sidan 62.

Sats 6.10 *Låt t_n vara förväntad tid tills en födelse-dödsprocess når tillstånd n, givet start i tillstånd 0. Då är*

$$t_n = \sum_{k=0}^{n-1} \frac{1}{\lambda_k \rho_k} \sum_{j=0}^{k} \rho_j. \tag{6.43}$$

Bevis. Vi gör om tillstånd n till ett absorberande tillstånd och låter t_{in} vara förväntad tid tills absorption, givet start i tillstånd i, $i \le n$. Absortionstidsekvationerna (6.32) ger i detta fall

$$\begin{aligned} t_{0n} &= \frac{1}{\lambda_0} + t_{1n} \\ t_{in} &= \frac{1}{\lambda_i + \mu_i} + \frac{\lambda_i}{\lambda_i + \mu_i} t_{i+1,n} + \frac{\mu_i}{\lambda_i + \mu_i} t_{i-1,n}, i = 1, 2, \ldots, n-1 \\ t_{nn} &= 0. \end{aligned} \tag{6.44}$$

Genom insättning kan man efter en del räknande verifiera att

$$t_{in} = \sum_{k=i}^{n-1} \frac{1}{\lambda_k \rho_k} \sum_{j=0}^{k} \rho_j, \ \ i = 0, 1, \ldots, n-1 \tag{6.45}$$

är lösning till (6.44). Sätt i=0 och beviset är klart. □

Om nu summan $\sum_{k=0}^{\infty} \frac{1}{\lambda_k \rho_k} \sum_{j=0}^{k} \rho_j$ är ändlig når man under ändlig förväntad tid oändligheten, och vi har följande sats.

Sats 6.11 *En födelse-döds-process är reguljär om och endast om*

$$\sum_{k=0}^{\infty} \frac{1}{\lambda_k \rho_k} \sum_{j=0}^{k} \rho_j \tag{6.46}$$

divergerar.

Exempel 6.9 Antag att alla födelseintenstiteter är λ och alla dödsintensiteter är μ. Då är $\rho_k = (\frac{\lambda}{\mu})^k$. Om $x = \frac{\lambda}{\mu}$ så erhåller vi för $x \neq 1$

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{1}{\lambda_k \rho_k} \sum_{j=0}^{k} \rho_j &= \frac{1}{\lambda} \sum_{k=0}^{\infty} x^{-k} \sum_{j=0}^{k} x^j \\ &= \frac{1}{\lambda(1-x)} \sum_{k=0}^{\infty} x^{-k}(1 - x^{k+1}) = \frac{1}{\lambda|1-x|} \sum_{k=0}^{\infty} |x^{-k} - x| \end{aligned} \tag{6.47}$$

vilken divergerar eftersom termerna inte går mot 0. Om $x = 1$ finner man lätt att summan blir $\sum_{k=0}^{\infty} \frac{k+1}{\lambda}$ vilken naturligtvis är divergent. För alla λ och μ är således processen reguljär. □

Vi ska nu betrakta gränsfördelningar i födelse-döds-processer.

Sats 6.12 *En födelse-döds-process där alla födelse- och döds-intensiteter är positiva är reguljär och ergodisk om och endast om*

$$\sum_{i=0}^{\infty} \rho_i < \infty \tag{6.48}$$

och

$$\sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} = \infty. \tag{6.49}$$

Den asymptotiska fördelningen är lika med den stationära som är

$$\pi_i = \frac{\rho_i}{\sum_{j=0}^{\infty} \rho_j}, \tag{6.50}$$

Om tillståndsrummet är ändligt, {0,1,...,N}, är processen reguljär och med en asymptotisk fördelning som ges av (6.50) där nämnaren ersatts med en ändlig summa $\sum_{j=0}^{N} \rho_j$.

Anm: Det är villkoret (6.48) som är det väsentliga. I alla realistiska fall är reguljaritetsvillkoret (6.49) uppfyllt.

Bevis. Den asymptotiska fördelningen ges av den stationära. Balansekvationerna $\boldsymbol{\pi Q} = \mathbf{0}$ blir, explicit utskrivna,

$$\begin{aligned} 0 &= -\lambda_0\pi_0 + \mu_1\pi_1, \\ 0 &= \lambda_{i-1}\pi_{i-1} - (\lambda_i + \mu_i)\pi_i + \mu_{i+1}\pi_{i+1} \quad \text{för } i = 1, 2, \ldots \end{aligned}$$

För $i = 1$ fås

$$\begin{aligned} 0 &= \lambda_0\pi_0 - (\lambda_1 + \mu_1)\pi_1 + \mu_2\pi_2 \\ &= \lambda_0\pi_0 - \mu_1\pi_1 - \lambda_1\pi_1 + \mu_2\pi_2 = -\lambda_1\pi_1 + \mu_2\pi_2 \end{aligned}$$

Betraktar vi nu $i = 2, 3, \ldots$, så får vi följande förenkling:

$$\lambda_i\pi_i = \mu_{i+1}\pi_{i+1}, \quad i = 0, 1, \ldots. \tag{6.51}$$

Vi får av detta

$$\begin{aligned} \pi_i &= \frac{\lambda_{i-1}}{\mu_i}\pi_{i-1} \\ &= \frac{\lambda_{i-1}}{\mu_i}\frac{\lambda_{i-2}}{\mu_{i-1}}\pi_{i-2} \\ &= \frac{\lambda_{i-1}}{\mu_i}\frac{\lambda_{i-2}}{\mu_{i-1}}\frac{\lambda_{i-3}}{\mu_{i-2}}\pi_{i-3} \\ &\vdots \\ &= \frac{\lambda_{i-1}}{\mu_i}\frac{\lambda_{i-2}}{\mu_{i-1}}\frac{\lambda_{i-3}}{\mu_{i-2}}\cdots\frac{\lambda_0}{\mu_1}\pi_0 = \rho_i\pi_0 \end{aligned}$$

Detta ger, då $\rho_0 = 1$, att

$$1 = \sum_{i=0}^{\infty}\pi_i = \pi_0\sum_{i=0}^{\infty}\rho_i.$$

Sätter vi samman detta med $\pi_i = \rho_i\pi_0$ så följer 6.50.

Om nu $\sum_{i=0}^{\infty}\rho_i < A < \infty$ så får vi att

$$\sum_{k=0}^{\infty}\frac{1}{\lambda_k\rho_k} \le \sum_{k=0}^{\infty}\frac{1}{\lambda_k\rho_k}\sum_{j=0}^{k}\rho_j \le \sum_{k=0}^{\infty}\frac{1}{\lambda_k\rho_k}A$$

av vilket man finner att mellanledet divergerar om och endast om $\sum_{k=0}^{\infty}\frac{1}{\lambda_k\rho_k}$ divergerar, vilket var villkoret för att processen skulle vara reguljär enligt sats 6.11.

Om tillståndsrummet är ändligt, $\boldsymbol{E} = \{0, 1, \ldots, N\}$ är $\rho_i = 0$ om $i > N$. Processen alltid reguljär och har en asymptotisk fördelning, vilket framgår av den allmänna satsen 6.8. □

Vi har tidigare något litet diskuterat de villkor som man behöver lägga på en Markov process för att undvika (onödiga) komplikationer, d.v.s. för att det ska finnas en unik lösning till bakåt- och framåtekvationerna eller för att processen ska varar reguljär. Vi ska här något mer diskutera dessa villkor för en födelse-och-dödsprocess där alla födelse- och dödsintensiteter är positiva så att processen är irreducibel.

Antag att $\boldsymbol{Q}$ är given. Då finns en entydigt bestämd övergångsmatris $\boldsymbol{P}(t)$ som uppfyller bakåt- och framåtekvationerna om och endast om

$$\sum_{i=0}^{\infty} \rho_i + \sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} = \infty.$$

I fortsättningen förutsätter vi att detta villkor är uppfyllt.

En födelse-och-dödsprocess är rekurrent, d.v.s. återkommer med sannolikhet 1 till ett tillstånd, om och endast om

$$\sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} = \infty.$$

En födelse-och-dödsprocess är positivt rekurrent, d.v.s. ergodisk, om och endast om

$$\sum_{i=0}^{\infty} \rho_i < \infty \text{ och } \sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} = \infty.$$

Detta visste vi ju redan från Sats 6.12 på sidan 68. De förväntade tiderna mellan återbesök till ett tillstånd är ändliga.

En födelse-och-dödsprocess är nollrekurrent, d.v.s. återkommer till varje tillstånd med sannolikhet 1 men efter förväntad oändlig tid, om och endast om

$$\sum_{i=0}^{\infty} \rho_i = \infty \text{ och } \sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} = \infty.$$

En födelse-och-dödsprocess är transient om och endast om

$$\sum_{i=0}^{\infty} \rho_i = \infty \text{ och } \sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} < \infty.$$

Om en födelse-och-dödsprocess är transient så ”försvinner den ut i oändligheten”. Detta kan den göra på två olika sätt. Antingen ”exploderar” den vilket innebär att processen blir oändlig redan inom ändlig tid eller också ”glider” den ut i oändligheten. I det första fallet är processen ej reguljär.

Om slutligen

$$\sum_{i=0}^{\infty} \rho_i < \infty \text{ och } \sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} < \infty.$$

så är processen irreguljär och det finns ingen entydig lösning till framåt- och bakåtekvationerna.

Exempel 6.10 Antag som i exempel 6.9 på sidan 68 att födelseintensiteterna respektive dödsintensiterna är konstanta, $\lambda_i = \lambda$ och $\mu_i = \mu$ alla i. Då är $\rho_i = (\frac{\lambda}{\mu})^i$ och vi erhåller

$$p_i = \frac{\rho_i}{\sum_{j=0}^{\infty} \rho_j} = \frac{(\frac{\lambda}{\mu})^i}{\sum_{j=0}^{\infty} (\frac{\lambda}{\mu})^j} = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right)$$

□

om $\lambda < \mu$. Gränsfördelningen är $\mathrm{Ge}(1 - \frac{\lambda}{\mu})$. Om $\lambda \geq \mu$, konvergerar inte summan $\sum_{j=0}^{\infty} \rho_j$ och processen är inte ergodisk. Om $\lambda > \mu$ driver processen mot oändligheten om $\lambda = \mu$ fluktuerar den allt starkare.

Kapitel 7

Köteori

7.1 Inledning

Köteori, eller ”teorin för masservice”, är en livaktig gren av den tillämpade matematiska statistiken. Dess grundidé kan kortfattat beskrivas på följande sätt:

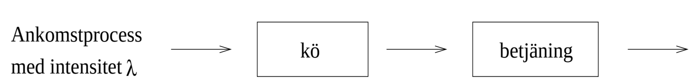
- Kunder kommer till ett betjäningsställe med en *intensitet* λ, d.v.s. i genomsnitt kommer λ kunder/tidsenhet.
- Kunderna kan tvingas vänta, och bildar då en kö eller lämnar systemet.
- Kunderna betjänas, och detta tar i genomsnitt $b = 1/\mu$ tidsenheter.
- Kunderna lämnar betjäningsstället.

Grundidén illustreras i figur 7.1.

Innan vi börjar diskutera de olika delarna av grundidén så ger vi några exempel, vilka vi ska återkomma till.

Exempel 7.1 (Kunder i en kiosk) Antag att n personer per tidsenhet går förbi en kiosk. Det är inte viktigt för den fortsatta diskussionen att antalet personer per tidsenhet är deterministiskt, men vi gör detta antagande för att inte komplicera resonemanget.

Antag nu att var och en av dessa personer bestämmer sig för att handla i kiosken oberoende av varandra och med sannolikheten p. Oberoendet verkar vara ett rimligt antagande, åtminstone om vi bortser från att personer kan komma i sällskap och därför påverka varandra. En annan invändning kan vara att om många personer under kort tid, av slumpmässiga skäl, har bestämt sig



Figur 7.1: Illustration av ett kösystem.

för att handla så har det bildats en kö utanför kiosken. En person som kommer, och som egentligen vill handla, kanske avstår från detta då han ser kön. Vi bortser för tillfället även från dessa invändningar.

Det slumpmässiga i ankomsterna är således huruvida en person – en ”potentiell kund” – bestämmer sig för att bli en kund. Under ett tidsintervall av längden t har vi nt potentiella kunder. Låt N_t vara antalet av dessa som blir (riktiga) kunder. Under våra antaganden gäller att N_t är $\mathrm{Bin}(nt, p)$-fördelad. (Vi förutsätter att t är valt så att nt är ett heltal.) Om nu n är stort och p är litet och om $np = \lambda$ så gäller det att N_t är approximativt $\mathrm{Po}(\lambda t)$-fördelad. Ska vi vara lite matematiskt strikta så antar vi att $n \to \infty$ och $p \to 0$ på sådant sätt att $np \to \lambda$. Antalet kunder som kommer under olika (disjunkta) tidsintervall är oberoende stokastiska variabler, eftersom olika potentiella kunder blir kunder oberoende av varandra.

Låt nu $N(t)$ vara antalet kunder som kommer till kiosken under tidsintervallet $(0, t]$. Enligt definition 6.5 på sidan 57 är $N = \{N(t),\, t \geq 0\}$ en Poissonprocess med intensiteten λ. Vi har under våra antaganden, som verkar ganska rimliga, en situation där *ankomstprocessen* är en Poissonprocess.

Om en kund har tur så finns det inga andra kunder i kiosken och kunden blir genast betjänad. Befinner sig däremot en kund i kiosken så får vår kund snällt vänta tills denna kund är betjänad. Finns det flera kunder så förutsätter vi att dessa har bildat en kö och att vår kund ställer sig sist i kön. Någon annan *köordning* i denna situation torde ställa till bråk.

Tiden det tar för en kund att bli betjänad beror på en mängd faktorer: vad kunden vill köpa, hur snabb kunden och expediten är osv. Vi beskriver därför kundens *betjäningstid* med en stokastisk variabel U med fördelningsfunktion $B(u) = P(U \leq u)$. Vidare antar vi att olika kunders betjäningstider beskrivs av oberoende stokastiska variabler som alla har samma fördelning B. Vad som är ett rimligt antagande om fördelningen B är inte alls lätt att säga. Vi kommer dock ofta – nästan alltid – att anta att U är exponentialfördelad. Skälet till detta antagande är att den matematiska analysen blir kraftigt förenklad.

Frågor som är intressanta kan t.ex. vara:

- Behöver en kund köa?
- Hur länge behöver en kund köa?
- Hur lång tid tar hela kioskexpeditionen?
- Hur lång är kön?

□

Exempel 7.2 (Kunder i en stor affär) I exempel 7.1 underförstod vi att det bara fanns en expedit eller *betjäningsstation*. In en större affär finns det förhoppningsvis flera. Vi tänker oss att ankomstprocessen, precis som i exempel 7.1, är en Poissonprocess med intensiteten λ. Det vanliga i en affär är att det finns nummerlappar, vilket gör att även köordningen blir som i exempel 7.1. Skillnaden är att vi nu har flera betjäningsstationer.

I vissa fall kan man ha en kö till varje expedit, t.ex. i äldre postkontor och systembolag. Om en kund väljer kö slumpmässigt och aldrig byter kö så är vi tillbaka i exempel 7.1 för varje betjäningsstation. Om vi har c betjäningsstationer så är ankomstprocessen till varje betjäningsstation en Poissonprocess med intensiteten λ/c. Man kan vidare visa – detta är en av Poissonprocessens många speciella egenskaper – att de c ankomstprocesserna är oberoende.

I verkligheten väljer nog inte kunderna kö slumpmässigt utan försöker bedöma vilken kö som är snabbast. Vi har väl alla erfarenhet av att vi försökt välja kö och haft otur! □

Exempel 7.3 (Akutmottagning på sjukhus) Vid en akutmottagning på ett sjukhus finns det i regel flera läkare och vi kan betrakta mottagningen som ett betjäningssystem med c betjäningsstationer. Om vi bortser från större olyckor, då flera patienter kan komma samtidigt, är det nog rimligt att även här anta att ankomstprocessen är en Poissonprocess med intensiteten λ. I detta sammanhang är det lämpligt att uppfatta patienter som kunder; ett synsätt som ju dessvärre för närvarande även är politiskt gångbart. Köordningen är dock lyckligtvis annorlunda. När en patient kommer görs en första preliminär bedömning av patientens tillstånd, och denna bedömning bestämmer den vidare behandlingen av patienten. Patienter i behov av mycket akut vård får ”gå före” i kön.

Man brukar i sådana fall tala om *prioriterad kö*. □

Exempel 7.4 (Telefonsamtal) Sverige har en lång tradition av forskning och utveckling i samband med teletrafik. En privatpersons telefon kan betraktas som ett betjäningssystem med en betjäningsstation. Ankomstprocessen består av de personer som ringer och betjäningstiden svarar mot längden av ett telefonsamtal. När en person, kund, ringer så kan tre saker inträffa: ingen svarar, samtalet kommer till stånd eller det är upptaget. Vi bortser tills vidare från fallet att ingen svarar. Man vet att längden av telefonsamtal väl beskrivs av exponentialfördelningen. Skillnaden mot exempel 7.1 är främst att ingen kö kan bildas. En person som får upptagetsignal lägger på och ”försvinner”. Vi talar i detta fall om ett *förlustsystem* eller ett system utan kö.

Situationen som vi har beskrivit kan ses som den klassiska för teletrafik, och vi kan numera tänka oss flera modifieringar. En första modifiering, som dock inte har med modern teknik att göra, är att abonnenten är en familj. Förutom möjligheterna ingen svarar, samtal kommer till stånd eller att det är upptaget finns nu möjligheten att den som söks inte är hemma. Den typen av samtal är ur teletrafiksynvinkel ”lyckade” samtal, men de kan ses som samtal ”som inte kommit till stånd” eller korta samtal. Man vet att uttalandet om att längden av telefonsamtal är exponentialfördelade egentligen gäller samtal som kommit till stånd eller, annorlunda uttryckt, att korta samtal är ”överrepresenterade”. En annan komplikation är att olika familjemedlemmar kan ha olika telefonkultur. Även om varje familjemedlem talar exponentialfördelat länge, så kan medlemmarna ha högst individuella väntevärden för samtalslängden.

Det utbredda (miss)bruket av telefonsvarare medför nog att korta ”samtal” blir ytterligare överrepresenterade. Möjligheten till automatisk återuppringning som AXE-växlarna ger innebär – matematiskt sett – en komplikation. Denna möjlighet innebär att det bildas ett beroende mellan avslutandet av ett samtal och ankomsten av ett nytt samtal.

En mera långtgående modifiering är att betrakta en företagsväxel som har flera ingående linjer. Först om alla linjer är upptagna så kommer samtalet inte till stånd. Vissa växlar, ett typexempel är tidsbeställningen till bilprovningen, har ett visst antal aktiva linjer men dessutom ett ”väntrum”. I praktiken innebär detta att man antingen direkt får tala med en person eller att man får svar av en telefonsvarare som meddelar att det finns samtal före och att man står i kö. □

7.1.1 Kendalls beteckningssystem

Ofta beskriver man kortfattat ett kösystem med Kendalls beteckningssystem:

$$A/B/c,$$

där

A anger ankomstprocessen;

B anger betjäningstidsfördelningen;

c är antalet betjäningsstationer.

Ankomstprocessen:

Ankomstprocessen är en *punktprocess*, d.v.s. en stokastisk beskrivning av hur kunder kommer till systemet. I en Poissonprocess med intensitet λ är avstånden mellan händelserna – kunderna – oberoende och Exp(λ)-fördelade. Vi kommer att beteckna fördelningsfunktionen för händelseavstånden i en förnyelseprocess med $A(t)$.

M anger att ankomstprocessen är en Poissonprocess, d.v.s. tiderna mellan ankomster är oberoende, med samma exponentialfördelning. Bokstaven M betyder ”Markov”. Poissonprocessen är ju en Markovprocess!

D anger att ankomstprocessen är deterministisk, d.v.s. det finns ingen slump i hur kunderna kommer,

GI står för ”general independent” och innebär att ankomstprocessen är en förnyelseprocess, d.v.s. tiderna mellan ankomster är likafördelade och oberoende,

G står för ”general” och innebär att ankomstprocessen är en godtycklig punktprocess.

I våra inledande exempel var det rimligt att anta att ankomstprocessen är en Poissonprocess. En naturlig invändning mot detta antagande i exempel 7.1 på sidan 71 är att antalet personer som går förbi kiosken per tidsenhet varierar med tiden. I verkligheten har vi säkert en högst påtaglig dygnsvariation – det är ju mycket vanligt att man passar på att handla i en kiosk på väg till eller från arbetet. Tar vi hänsyn till detta så borde vi låta ankomstprocessen vara en inhomogen Poissonprocess, se slutet avsnittet om Poissonprocesser. Å andra sidan kan man tycka att denna variation är förhållandevis långsam i förhållande till slumpvariationen. Vi kommer alltid att förutsätta att vi betraktar kösystemen under så pass begränsad tid att vi kan bortse från dygnsvariation och liknade typer av säsongsvariationer.

Förnyelseprocessen är en mycket populär generalisering av Poissonprocessen. Detta gäller inte bara inom köteori utan inom en mängd olika modeller där Poissonprocessen kan ses som den ”första” ansatsen. Skälet är nog till stor del att man ofta kan genomföra en sådan generalisering utan att på ett fundamentalt sätt förändra den matematiska analysen. Det finns dock lömska fallgropar eftersom Poissonprocessen har många speciella egenskaper. Det är inte ovanligt i tillämpade uppsatser att något – mer eller mindre explicit – antagande görs som innebär att förnyelseprocessen i själva verket är en Poissonprocess. I kösammanhang kan man nog säga att förnyelseprocessen kan vara en högst relevant generalisering i ”tekniska” sammanhang men kanske inte i lika hög grad för ”mänskliga” köer. Ett sådant tekniskt fall kan vara enheter på ett löpande band som i vissa fall kommer (nästan) deterministiskt. Detta svarar då mot att händelseavstånden är enpunktsfördelade, d.v.s. att

$$A(t) = \begin{cases} 0 & \text{om } 0 \le t < d \\ 1 & \text{om } t \ge d, \end{cases}$$

där d är avståndet mellan enheterna.

Vi kan inom ramen för denna kurs egentligen inte alls gå in på fallet med allmän ankomstprocess, d.v.s. då ankomstprocessen är en godtycklig punktprocess. Vi ska dock i exempel 7.5 betrakta en, kanske inte helt seriös, situation där ankomstprocessen nog borde uppfattas som en punktprocess av annan typ än en förnyelseprocess.

Exempel 7.5 (Kunder i paraplyaffär) Låt oss betrakta en affär på en turistort som mest säljer paraplyer. Skälet till att vi väljer en turistort är att vi nog i stort sett kan bortse från dygnsvariationen. Intresset för att köpa paraplyer är nog däremot högst beroende av om det regnar eller ej. Under en dag med omväxlande skurar och solsken är det därför rimligt att tänka sig att ankomstprocessens intensitet kraftigt varierar med vädret. En enkel och naturlig modell kan vara att låta ankomstprocessen vara en Poissonprocess med intensitet λ_s då solen skiner och med intensitet λ_r då det regnar där λ_r är väsentligt större än λ_s. Detta innebär således att ankomstprocessen är en inhomogen Poissonprocess där intensitetsfunktionen $\lambda(t)$ hoppar mellan λ_s och λ_r. Problemet är nu att beskriva intensitetsfunktionen. Eftersom variationen mellan sol och regn svårligen beskrivs deterministiskt är det naturligt att uppfatta $\lambda(t)$ som utfall av en stokastisk process. Detta innebär att vi först genererar ett utfall $\lambda(t)$ av en intensitetsprocess och att detta utfall sedan används som intensitetsfunktion i en inhomogen Poissonprocess. En punktprocess uppkommen på detta sätt kallas en *Coxprocess* eller en *dubbelt stokastisk Poissonprocess.* □

Coxprocesser en mycket viktig klass av punktprocesser. Denna klass av processer används t.ex. inom teorin för optiska signaler. Den underliggande intensitetsprocessen är då av helt annan typ än i exemplet. Vi ska dock inte fördjupa oss i dessa processer.

Betjäningstidsfördelning:

Inom ramen för detta beteckningssystem förutsätter vi att olika kunders betjäningstider beskrivs av oberoende stokastiska variabler som alla har samma fördelning B och att dessa tider även är oberoende av ankomstprocessen.

M anger att betjäningstiderna är exponentialfördelade. Bokstaven M betyder, liksom för ankomstprocessen, ”Markov”. Det markovska med exponentialfördelningen är ju att den saknar minne.

G står för ”general” och innebär att betjäningstiderna har en godtycklig fördelning B. Ett specialfall är

D som innebär att betjäningstiderna är deterministiska, d.v.s. att B är en enpunktsfördelning.

Vi påminner om diskussionen i exempel 7.1 på sidan 71 angående valet av betjäningstidsfördelning.

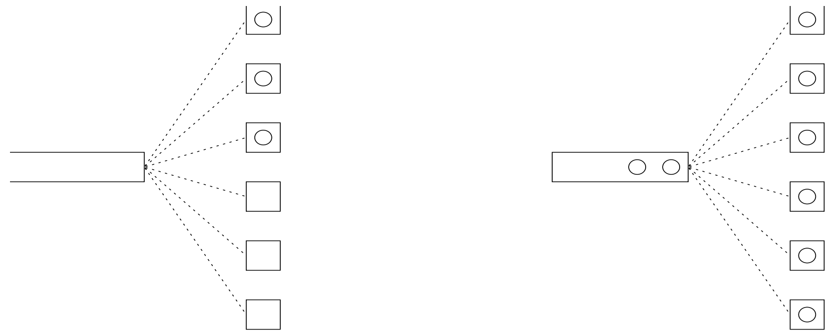
Antal betjäningsstationer:

Vi kommer bara att betrakta fallet med en eller flera betjäningsstationer, som arbetar parallellt och med samma fördelning för betjäningstiden. Antalet stationer betecknas med c. Figur 7.2 illustrerar detta för $c = 6$.

Det finns ytterligare några egenskaper hos ett kösystem, som ibland bakas in in Kendalls beteckningar.

Köordning:

Vanligast är FIFO (first-in, first-out), eller ordnad kö, och det är det som vi skall betrakta här. Varianter finns, tex parallella köer (post), prioritering (akutmottagning på sjukhus). I datorsammanhang förekommer FILO (first-in, last-out) och slumpmässig ordning.



Figur 7.2: Illustration av ett betjäningssystem med 6 stationer, för 3 resp. 8 kunder i systemet. En kund är markerad med $\circ$.

Ändligt väntrum:

Det finns vidare möjligheten att kunder lämnar systemet, utan att bli betjänade: Väntrummet är fullt! Vi ska endast betrakta *förlustsystem* som innebär att en kund lämnar systemet om inte betjäning kan ske omedelbart. Ett typiskt exempel på detta är telefonsamtal till privatpersoner som diskuterades i exempel 7.4 på sidan 73.

7.1.2 Beteckningar

Den långa lista av beteckningar som kommer kan nog vid första ögonkastet verka lite avskräckande. En förstagångsläsare rekommenderas att ögna igenom listan och sedan använda den som referens när beteckningarna dyker upp.

Vi förutsätter att inblandade gränsvärden existerar.

$X(t)$ = antalet kunder i systemet vid tid t.

En realisering av $X(t)$ kan illustreras av figur 6.6 på sidan 67. Tolkningen av en ”födelse” är att en kund kommer och av en ”död” att en kund lämnar systemet. Vi har därmed inte sagt att $X(t)$ är en födelse-döds-process, d.v.s. en Markovprocess, även om det är detta fall som vi främst ska betrakta.

$\ell(t) = E[X(t)]$.
$\ell = \lim_{t\to\infty} \ell(t)$.
$p_i = \lim_{t\to\infty} P(X(t) = i)$.
$X_q(t)$ = antalet kunder i kön vid tid t.
$\ell_q(t) = E[X_q(t)]$.
$\ell_q = \lim_{t\to\infty} \ell_q(t)$.
U_n = n:te kundens betjäningstid.
$B(t) = P(U_n \le t)$ = betjäningstidens fördelningsfunktion.
Q_n = n:te kundens kötid.
$S_n = Q_n + U_n$ = n:te kundens tid i systemet.

$$G_q(\tau) = \lim_{n\to\infty} P(Q_n \le \tau).$$
$$G(\tau) = \lim_{n\to\infty} P(S_n \le \tau).$$

Låt Q vara en s.v. med fördelningsfunktion $G_q(\tau)$, U en stokastisk variabel med fördelningsfunktion $B(t)$, oberoende av Q, och $S = Q + U$.

$w_q = E[Q]$ = förväntad kötid.
$b = E[U]$ = förväntad betjäningstid.
$\mu = 1/b$ = betjäningsintensitet.
$w = E[Q + U]$ = förväntad tid i systemet.

7.2 Littles formel

Vi antar nu att de olika kundernas betjäningstider är oberoende och likafördelade s.v., och dessutom oberoende av ankomstprocessen. Vi betraktar nu ett system med c parallella betjäningsstationer, och definierar

$$\textit{trafikintensiteten (betjäningsfaktorn) } \rho = \frac{\lambda}{c\mu}. \tag{7.1}$$

Trafikintensiteten är ett mått på belastningen av systemet, och vi förutsätter att $\rho < 1$. Detta innebär att det inte kommer flera kunder än systemet ”tål”, och är ett nödvändigt villkor för att kön inte ska ”explodera”.

Vi kommer att tala om ”mycket allmänna villkor”. Detta innebär, på det stora hela, att resultaten gäller för ett $G/G/c$-system.

Det är egentligen två krav som behövs.

Dels kommer vi att förutsätta att systemet befinner sig i stationärt tillstånd. För att uppnå detta måste vi anta, som vi gjort, att $\rho < 1$. Vidare måste inprocessen vara, mer eller mindre, stationär. Matematiskt betyder det att inprocessens stokastiska egenskaper inte får bero ”för mycket” av hur den startar. Enklast är ju att förutsätta att inprocessens egenskaper inte alls beror av startsättet, vilket vi gör i resonemanget, men då ”straffar vi ut” $GI/G/c$-systemet som är tillåtet. I praktiken innebär detta att vi inte tillåter trender eller säsongsvariationer i inprocessen.

Det andra kravet är ett ”ergodicitetsvillkor”. Detta innebär (nästan) att inprocessens stokastiska egenskaper på intervall som tidsmässigt ligger långt ifrån varandra är oberoende. Mera precist innebär ergodicitet att ”tidsmedelvärden” konvergerar mot ”rumsmedelvärden”.

Vi ska längre fram något lite återkomma till dessa två krav.

Antag nu att en kund kommer till systemet, som befinner sig i stationärt tillstånd. När denna kund, efter i genomsnitt tiden w_q, köat färdigt och börjar betjänas, så har alla kunder som står i kön kommit under denna kunds kötid. I genomsnitt har då $w_q\lambda$ kunder kommit. Detta leder till Littles formel:

Under mycket allmänna villkor gäller: $\ell_q = w_q\lambda$.

Antag att $c = 1$, d.v.s. vi har bara en betjäningsstation. Eftersom alla kunder som kommer till systemet även lämnar det, så måste utprocessen ha intensitet λ. Betjäningsstationen har dock, när den arbetar, ”utintensiteten” μ. Detta ger

$$\lambda = p_0 \cdot 0 + (1 - p_0) \cdot \mu$$

eller

$$\text{sannolikheten att betjäning pågår} = 1 - p_0 = \frac{\lambda}{\mu} = \rho. \tag{7.2}$$

Vi kan nu formulera följande sats, vars vits är att om vi lyckats beräkna något av de grundläggande väntevärdena ℓ_q, ℓ, w_q eller w, så kan vi lätt bestämma de andra.

Sats 7.1 *Under mycket allmänna villkor gäller:*

$$\ell = \ell_q + c\rho \tag{7.3}$$

$$w_q = \frac{\ell_q}{\lambda} \tag{7.4}$$

$$w = \frac{\ell}{\lambda}\,. \tag{7.5}$$

Heuristiskt bevis. (7.4) är Littles formel. För att visa (7.3) generaliserar vi resonemanget som ledde till (7.2) i fallet $c = 1$. Detta ger

$$\lambda = \text{ förväntat antal kunder som betjänas } \cdot \mu,$$

eller

$$\text{förväntat antal kunder som betjänas } = \frac{\lambda}{\mu} = c\rho.$$

Detta ger

$$\ell = \ell_q + \text{ förväntat antal kunder som betjänas } = \ell_q + c\rho,$$

vilket är (7.3). Vidare följer från (7.3) och (7.4) att

$$\ell = \ell_q + c\rho = \lambda w_q + \frac{\lambda}{\mu} = \lambda \cdot \left(w_q + \frac{1}{\mu}\right) = \lambda \cdot w,$$

d.v.s. (7.5). □

I strikt matematisk mening har vi inte *bevisat* sats 7.1, vilket ju bland annat hade krävt att vi ordentligt hade angett vad vi menar med ”mycket allmänna villkor”. Vi ska dock, med direkta räkningar, i avsnitt 7.3.1 på sidan 80 *visa* att sats 7.1 gäller i ett $M/M/1$-system.

Man bör observera att sats 7.1 behandlar köns stokastiska egenskaper då den befinner sig i stationärt tillstånd med avseende på tiden. Dessa är, i denna allmänna formulering, inte nödvändigtvis desamma som de egenskaper som möter kunderna då de kommer. För att vara mera precisa, så betraktar vi $X(t)$ vid de tidpunkter då kunder kommer till systemet. Låt $T_1, T_2, \ldots$ vara dessa slumpmässiga tidpunkter, och sätt

$$\widetilde{Y}_n = X(T_n - 0),$$

d.v.s. antalet kunder i systemet då den n-te kunden kommer. Istället för att betrakta kön efter lång tid kan vi betrakta den efter att många kunder har kommit, d.v.s. vi betraktar $\widetilde{Y}_n$ i dess stationära tillstånd. Medelantalet kunder i systemet som en ankommande kund möter är då $\lim_{n\to\infty} E[\widetilde{Y}_n]$. Man talar ibland om *tidsstationärt* tillstånd respektive *kundstationärt* tillstånd beroende på om man betraktar $X(t)$ eller $\widetilde{Y}_n$. Sats 7.1 behandlar således det tidsstationära fallet. Då ankomstprocessen är en Poissonprocess så överensstämmer dessa två fall. Detta beror av att Poissonprocessen har oberoende ökningar. I ett system med en annan ankomstprocess behöver man dock skilja på systemets stokastiska egenskaper vid godtyckliga tidpunkter och dess egenskaper då kunder ankommer. Vi ska något återkomma till denna distinktion i avsnitten 7.4.1 på sidan 99 och 7.5 på sidan 101.

Efter definitionen (7.1) av trafikintensiteten så nämnde vi att ρ är ett mått på belastningen av systemet. Belastningen kan i fallet $c = 1$ tolkas som sannolikheten för att betjäningsstationen arbetar och för $c \geq 1$ som den förväntade proportionen av betjäningsstationerna som arbetar. Eftersom "antalet kunder som betjänas" och "antalet betjäningsstationer som arbetar" är samma sak så följer det av (7.2) och (7.3) att denna tolkning, eller definition, är likvärdig med (7.1).

7.3 Markovska köer

Ett kösystem kallas en *markovsk kö* om $X(t)$, d.v.s. antalet kunder i systemet vid tid t, är en Markovprocess. I alla våra tillämpningar kommer i detta fall $X(t)$ att speciellt vara en födelse-döds-process. Som vi tidigare nämnt svarar en "födelse" mot att en kund kommer och en "död" mot att en kund lämnar systemet.

I en markovsk kö är tiderna mellan ändringar i tillståndet, d.v.s. mellan hopp i $X(t)$, exponentialfördelade. Detta gäller ju för alla Markovprocesser, se sats 6.2 på sidan 49.

7.3.1 $M/M/1$-system

Detta är det enklaste, och kanske mest grundläggande kösystemet. Vi påminner om att ett $M/M/1$-system innebär att kunder kommer enligt en Poissonprocess med intensitet λ och att betjäningstiderna är oberoende Exp(μ)-fördelade stokastiska variabler som dessutom är oberoende av ankomstprocessen. Följande sats är grundläggande för vår analys av $M/M/1$-system.

Sats 7.2 *Antalet kunder i ett $M/M/1$-system, $X(t)$, är en födelse-döds-process med födelseintensiteter*

$$\lambda_0 = \lambda_1 = \ldots = \lambda \text{ (ankomstintensiteten)}$$

och dödsintensiteter

$$\mu_1 = \mu_2 = \ldots = \mu \text{ (betjäningsintensiteten).}$$

Bevis. Satsen följer av sats 6.2 på sidan 49 som beskriver allmänt hur en Markovkedja kan uppkomma. Vid varje tidpunkt t väntar processen $\{X(t); t \geq 0\}$ på att en kund anländer eller att en försvinner. I det första fallet ökar processen med värdet ett, i det andra minskar processen med värdet ett. Intensiteterna för dessa händelser är λ respektive μ. □

Vi påminner om att trafikintensiteten ρ i detta fall med en betjäningsstation är definierad av

$$\rho = \frac{\lambda}{\mu}.$$

Sats 7.3 *Antalet kunder i ett $M/M/1$-system, $X(t)$, har en gränsfördelning om och endast om*

$$\rho < 1.$$

I detta fall gäller

$$p_n = \rho^n(1-\rho), \quad \text{för } n = 0, 1, \ldots, \tag{7.6}$$

d.v.s. antalet kunder i systemet, när detta befinner sig i stationärt tillstånd, är geometriskt fördelat.

Bevis. Detta följer av exempel 6.10 på sidan 70. □

I resten av detta avsnitt låter vi X vara en stokastisk variabel vars fördelning är given av (7.6).

Betrakta nu $X_q(t)$ = antalet kunder i kön vid tid t. Det är lätt att inse att

$$X_q(t) = \begin{cases} 0 & \text{om } X(t) = 0 \text{ eller } 1 \\ X(t) - 1 & \text{annars.} \end{cases}$$

På motsvarande sätt bildar vi den stokastiska variabeln

$$X_q = \begin{cases} 0 & \text{om } X = 0 \text{ eller } 1 \\ X - 1 & \text{annars.} \end{cases}$$

Sats 7.4 *I ett $M/M/1$-system gäller*

$$\ell = E[X] = \frac{\rho}{1-\rho} \qquad V[X] = \frac{\rho}{(1-\rho)^2}$$

och

$$\ell_q = E[X_q] = \frac{\rho^2}{1-\rho} \qquad V[X_q] = \frac{\rho^2(1+\rho-\rho^2)}{(1-\rho)^2}.$$

Bevis. Eftersom X är Ge$(1-\rho)$ är $X+1$ ffg$(1-\rho)$ och $E(X) = \frac{1}{1-\rho} - 1 = \frac{\rho}{1-\rho}$ och $V(X) = V(X+1) = \frac{\rho}{(1-\rho)^2}$, se (2.4) på sidan 7.

Vidare gäller

$$P(X_q = 0) = P(X = 0) + P(X = 1) = p_0 + p_1 = 1 - \rho^2 \quad \text{och}$$
$$P(X_q = i) = P(X = i+1) = p_{i+1} \quad \text{för } i \geq 1 \tag{7.7}$$

vilket ger

$$E[X_q] = \sum_{i=1}^{\infty} i p_{i+1} = \rho E[X] = \frac{\rho^2}{1-\rho}$$

och

$$E[X_q^2] = \rho E[X^2] = \frac{\rho^2(1+\rho)}{(1-\rho)^2}$$

$$\Rightarrow V[X_q] = \frac{\rho^2(1+\rho)}{(1-\rho)^2} - \frac{\rho^4}{(1-\rho)^2} = \frac{\rho^2(1+\rho-\rho^2)}{(1-\rho)^2}\,.$$

□

Vi kan observera att

$$\ell = \frac{\rho}{1-\rho} = \frac{\rho-\rho^2+\rho^2}{1-\rho} = \frac{\rho-\rho^2}{1-\rho} + \frac{\rho^2}{1-\rho} = \rho + \ell_q$$

vilket är (7.3) i sats 7.1 på sidan 79.

Låt oss nu betrakta väntetiden för en kund som kommer efter lång tid, d.v.s. då systemet befinner sig i stationärt tillstånd. Som förut använder vi följande beteckningar:

$Q =$ kundens kötid.
$U =$ kundens betjäningstid som är Exp(μ)-fördelad och oberoende av Q.
$S = Q + U =$ kundens tid i systemet.
$G_q(\tau) = P(Q \le \tau)$.
$G(\tau) = P(S \le \tau)$.

Sats 7.5 *I ett $M/M/1$-system gäller*

$$w = E[S] = \frac{1}{\mu(1-\rho)} \qquad V[S] = w^2 \qquad 1 - G(\tau) = e^{-\tau/w}, \quad \tau \ge 0,$$

d.v.s. S är Exp$(1/w)$*-fördelad, och*

$$w_q = E[Q] = \frac{\rho}{\mu(1-\rho)} \qquad V[Q] = \frac{(2-\rho)\rho}{\mu^2(1-\rho)^2} \qquad 1 - G_q(\tau) = \rho e^{-\tau/w}, \quad \tau \ge 0.$$

Bevis. Antag att det finns n kunder i systemet, d.v.s. att $X = n$. Vi påminner om (den finstilta) diskussionen efter sats 7.1 på sidan 79.

Betrakta nu en ankommande kund. Då gäller att kundens kötid Q är summan av den återstående betjäningstiden för kunden som betjänas plus de $(n-1)$ betjäningstiderna för kunderna i kön. Alla dessa tider är oberoende och Exp(μ)-fördelade. Den totala tiden i systemet S är därför summan av $n+1$ oberoende Exp(μ)-fördelade stokastiska variabler. Eftersom det följer av (7.6) att $X+1$ är ffg$(1-\rho)$-fördelad så ger sats 6.4 på sidan 59 utsagorna angående S.

Vi betraktar nu Q. Om $X = 0$ så blir kunden direkt betjänad, d.v.s. även $Q = 0$. Detta sker med sannolikheten $p_0 = 1-\rho$. Om $X > 0$ så är Q en summa av exponentialfördelade variabler. Således har Q en blandad fördelning. X betingat av $X > 0$ är ffg$(1-\rho)$-fördelad. Lagen om den totala sannolikheten ger således

$$1 - G_q(\tau) = (1-\rho)\cdot 0 + \rho\cdot(1 - G(\tau)), \quad \text{för } \tau > 0.$$

För att beräkna w_q och $V(Q)$ kan man räkna direkt och utnyttja att $E[S] = \rho E[Q]$ och $E[S^2] = \rho E[Q^2]$. En lite roligare härledning – nåja, smaken kan ju vara delad – är att utnyttja $S = Q + U$ vilket ger

$$w_q = w - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}$$

och

$$V[S] = V[Q] + V[U] \Rightarrow V[Q] = V[S] - V[U] = \frac{1}{\mu^2(1-\rho)^2} - \frac{1}{\mu^2} = \frac{(2-\rho)\rho}{\mu^2(1-\rho)^2}.$$

□

Jämför vi satserna 7.4 och 7.5 så ser vi att

$$\ell = \frac{\rho}{(1-\rho)} = \frac{\lambda}{\mu(1-\rho)} = \lambda w,$$

vilket är (7.5) i sats 7.1. På samma sätt kan vi visa att (7.4) i sats 7.1 gäller.

Den fördelningsmässigt enklaste och nog även mest intressanta av variablerna X, X_q, S och Q är S, så låt oss titta lite närmare på den.

Exempel 7.6 (Lunch på KTH) Vi har nog alla erfarenhet av att det då och då är väldigt långa köer vid KTHs lunchställen. Det vi ska koncentrera oss på här är inte om köerna i allmänhet är för långa utan på att de *då och då* är väldigt långa. Detta kan naturligtvis bero av flera saker, matsedeln som kan påverka både λ och μ, tidpunkten både under dagen och under året, osv. Förutom variation i parametrarna så finns det en påtaglig slumpmässig variation. Detta är särskilt påtagligt för oss som äter på Drottning Kristina på torsdagar. Under vintern serveras där alltid ärtor och pannkaka, som uppenbarligen är en mycket populär kombination. Vi på matematisk statistik har med viss irritation insett att vi inte alls kan förutse hur lång kön ska vara. Vi försöker naturligtvis välja tidpunkten lämpligt, speciellt under läsperioderna. Trots detta misslyckas vi. Förklaring torde – förhoppningsvis – vara den stora slumpmässiga variationen. Låt oss anta att kön beskrivs av ett $M/M/1$-system. Detta stämmer naturligtvis inte riktigt, gäster kommer ofta i sällskap vilket motsäger Poissonankomster och säkert är inte tiden det tar att lägga upp maten exponentialfördelad. Vi borde trots detta kunna få en viss uppfattning med hjälp av $M/M/1$-systemet.

Låt oss säga att vi tror att en person i genomsnitt får vänta $w = E(S) = 10$ minuter. Detta medför i så fall att $P(S > 20) = e^{-2} \approx 0.135$ och $P(S > 30) = e^{-3} \approx 0.05$. Detta skulle innebära att man ungefär 2 gånger av 15, d.v.s. ungefär tre gånger per månad, får vänta mer än 20 minuter på maten. På motsvarande sätt skulle man 1 gång på 20, d.v.s. i ungefär en gång per månad behöva vänta mer än 30 minuter. Var och en får själv avgöra om dessa värden är någorlunda relevanta.

Klart är dock att de slumpmässiga variationerna i kölängd och väntetid är stora i ett kösystem där ρ är nära ett. □

Låt $\{U(t),\, t \geq 0\}$ vara *utprocessen*, d.v.s. $U(t)$ är antalet färdigbetjänade kunder som lämnar systemet under tidsintervallet $(0, t]$. Följande sats är huvudsatsen för utprocessen från ett $M/M/1$-system.

Sats 7.6 *I ett $M/M/1$-system som befinner sig i stationärt tillstånd gäller att utprocessen $U(t)$ är en Poissonprocess med intensitet λ.*

Sats 7.6 är på intet sätt uppenbar och beviset av den bygger på att ”vända tiden”. Vi kan här inte ge ett strikt bevis av satsen, utan måste nöja oss med en mera lös diskussion.

Låt nu $\{X(t),\, t \geq 0\}$ vara en Markovprocess som har stationär fördelning $\boldsymbol{\pi}$ och välj $\boldsymbol{p}(0) = \boldsymbol{\pi}$. Då vet vi att $\boldsymbol{p}(t) = \boldsymbol{\pi}$ för alla $t \geq 0$. Det är givetvis inget speciellt med tidpunkten 0, så vi kan precis lika väl studera processen $\{X(t),\, t \geq t_0\}$. I princip kan vi låta $t_0 \to -\infty$ och betrakta $X(t)$ som en process på hela reella linjen. Formellt är detta inga problem inom ramen för den allmänna teorin för stokastiska processer. Vill vi dock ha en mera konstruktiv beskrivning av processens egenskaper, d.v.s. hur och när den hoppar, så är det inte fullt så lätt att säga hur man bör bete sig. Ett naturligt sätt är dock att starta i, låt oss säga, tidpunkten $t = 0$ och där kräva att $X(0)$ har fördelningen $\boldsymbol{p}$. Utgående från denna tidpunkt kan vi låta processen utveckla sig åt ”båda hållen”. Utvecklingen åt positiva hållet är då vad vi har studerat, d.v.s. för en Markovprocess vet vi, se sats (6.2) på sidan 49 att processen kommer att ligga still exponentialfördelat länge och därefter hoppa på ett sätt som beskrivs av $\boldsymbol{Q}$ eller, om man så vill, av den inbäddade Markovkedjan $\widetilde{X}$. Frågan är därför:
”Hur ska man göra åt det negativa hållet?”

Denna fråga är detsamma som att ställa frågan om hur den tidsreverserade processen $Y(t) = X(-t)$ uppför sig. Första frågan är om $Y(t)$ är en Markovprocess. Det är den. Sätt nämligen $X(t_i) = X_i, t_1 \leq t_2 \ldots$. Då får vi att

$$P(X_1 = i_1 \mid X_2 = i_2, X_3 = i_3, \ldots, X_n = i_n) = \frac{P(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)}{P(X_2 = i_2, \ldots, X_n = i_n)} =$$

$$(\text{Markovegenskapen}) = \frac{P(X_1 = i_1)\prod_{i=2}^{n} P(X_i = i_i \mid X_{i-1} = i_{i-1})}{P(X_2 = i_2)\prod_{i=3}^{n} P(X_i = i_i \mid X_{i-1} = i_{i-1})}$$

$$= \frac{P(X_1 = i_1)P(X_2 = i_2 \mid X_1 = i_1)}{P(X_2 = i_2)} = \frac{P(X_1 = i_1, X_2 = i_2)}{P(X_2 = i_2)} = P(X_1 = i_1 \mid X_2 = i_2)$$

vilket betyder att den reverserade processen är markovsk.

Nästa fråga är då hur övergångsintensiteterna uppför sig. Pga. stationariteten räcker det att betrakta $t = 0$. Då gäller, jmf. Bayes' sats,

$$P(Y(h) = i \mid Y(0) = j) = \frac{P(Y(0) = j, Y(h) = i)}{P(Y(0) = j)}$$

$$= \frac{P(Y(h) = i)P(Y(0) = j \mid Y(h) = i)}{P(Y(0) = j)} = \frac{P(X(0) = i)P(X(h) = j \mid X(0) = i)}{P(X(h) = j)}.$$

Låter vi nu $h \to 0$ och betecknar övergångsintensiteterna för $Y(t)$ med q_{ji}^{Y} så fås

$$q_{ji}^{Y} = \frac{\pi_j}{\pi_i} q_{ij}.$$

Det enklaste är naturligvis om $Y(t)$ och $X(t)$ har samma stokastiska egenskaper. Om så är fallet säger vi att processen är *tidsreversibel*. I det tidskontinuerliga fallet gäller att en Markovprocess är tidsreversibel om och endast om $\boldsymbol{\pi}$ och $\boldsymbol{Q}$ satisfierar de *lokala balansekvationerna*

$$\pi_i q_{ij} = \pi_j q_{ji} \quad \text{för alla } i \text{ och } j.$$

För en födelse-döds-process reduceras dessa, se avsnitt 6.8

$$\pi_i \lambda_i = \pi_{i+1} \mu_{i+1} \quad \text{för alla } i \geq 0,$$

vilket är (6.51) på sidan 69. En födelse-döds-process som har en gränsfördelning är således alltid tidsreversibel.

Av ovanstående, något abstrakta, resonemang följer sats 7.7.

Sats 7.7 *Låt $X(t)$ vara en födelse-döds-process med alla födelseintensiteter $\lambda_i = \lambda$ och sådan att en gränsfördelning existerar. Antag att processen befinner sig i stationärt tillstånd, d.v.s. att $\boldsymbol{p}(0) = \boldsymbol{\pi}$. Då bildar nedåthoppen en Poissonprocess med intensitet λ.*

Bevis. Då alla födelseintensiteter $\lambda_i = \lambda$ så bildar uppåthoppen en Poissonprocess med intensitet λ. Uppåthoppen svarar mot nedåthopp i den tidsreverserade processen. Då en födelse-döds-process som har en gränsfördelning är tidsreversibel så följer satsen. □

Sats 7.6 följer nu av sats 7.7. Observera dock att vi i sats 7.7 inte kräver att dödsintensiteterna är konstanta.

En följd av sats 7.6 är följdsats 7.1, som vi dock kan visa inom ramen för denna framställning.

Följdsats 7.1 *Tiden från en godtycklig tidpunkt tills en kund lämnar ett $M/M/1$-system som befinner sig i stationärt tillstånd är* Exp(λ)*-fördelad.*

Bevis. Betrakta systemet vid en godtycklig tidpunkt. Vid denna tidpunkt kan systemet antingen vara tomt eller ”ej tomt”. Låt T vara tiden tills en kund lämnar systemet. Vi har

$$P(T \le t) = P(T \le t \mid \text{systemet tomt}) \cdot P(\text{systemet tomt})$$

$$+P(T \le t \mid \text{systemet ej tomt}) \cdot P(\text{systemet ej tomt})$$

$$= P(T \le t \mid \text{systemet tomt}) \cdot (1-\rho) + P(T \le t \mid \text{systemet ej tomt}) \cdot \rho. \quad (7.8)$$

Om systemet ej är tomt så är T Exp(μ)-fördelad. Om systemet är tomt så måste vi först invänta en kund, och den tiden är Exp(λ)-fördelad, och sedan vänta tills denna kund är betjänad. Då är T summan av två oberoende exponentialfördelade stokastiska variabler med intensiteter λ resp. μ. Detta innebär enligt faltningsformeln för summan av två oberoende stokastiska variabler att

$$P(T \le t \mid \text{systemet tomt}) = \int_0^t \lambda e^{-\lambda x}\big(1 - e^{-\mu(t-x)}\big)\ dx$$

$$= \lambda \int_0^t e^{-\lambda x}\ dx - \lambda e^{-\mu t} \int_0^t e^{(\mu-\lambda)x}\ dx$$

$$= 1 - e^{-\lambda t} - \frac{\lambda}{\mu-\lambda} e^{-\mu t}\left(e^{(\mu-\lambda)t} - 1\right) = 1 - \left(1 + \frac{\lambda}{\mu-\lambda}\right) e^{-\lambda t} + \frac{\lambda}{\mu-\lambda} e^{-\mu t}$$

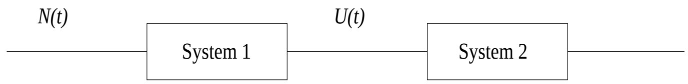
$$= 1 - \frac{\mu}{\mu-\lambda} e^{-\lambda t} + \frac{\lambda}{\mu-\lambda} e^{-\mu t}.$$

Från (7.8) fås nu

$$P(T \le t) = \left(1 - \frac{\mu}{\mu-\lambda} e^{-\lambda t} + \frac{\lambda}{\mu-\lambda} e^{-\mu t}\right)(1-\rho) + (1 - e^{-\mu t})\rho$$

$$= 1 - \frac{\mu-\lambda}{\mu}\left(\frac{\mu}{\mu-\lambda} e^{-\lambda t} - \frac{\lambda}{\mu-\lambda} e^{-\mu t}\right) - \frac{\lambda}{\mu} e^{-\mu t} = 1 - e^{-\lambda t} \quad \text{för } t \ge 0,$$

d.v.s. T är Exp(λ)-fördelad. □



Figur 7.3: Illustration av två betjäningssystem i serie.

Antag att vi står utanför en $M/M/1$-kö och betraktar kunderna som lämnar systemet. Vi förutsätter att vi inte kan se hur många kunder som befinner sig i systemet och att vi inte heller kan se de kunder som ankommer. Mer precist uttryckt så antar vi att vi observerar utprocessen men att vi inte kan observera varken antalet kunder i systemet $X(t)$ eller inprocessen. Kan vi då dra några slutsatser om $X(t)$? Svaret, som kanske är lite förvånande, är att det kan vi inte göra, vilket följer av följande sats.

Sats 7.8 *I ett $M/M/1$-system gäller att $X(t)$ och $U(s)$ för $s \leq t$ är oberoende stokastiska variabler.*

Vi kan inte här ge ett bevis av sats 7.8 men vi ska försöka göra den trolig. Av de $N(t)$ kunder som har kommit till systemet under tidsintervallet $(0, t]$ har $U(t)$ stycken redan lämnat det medan resten, d.v.s. $X(t) = N(t) - U(t)$ fortfarande befinner sig i systemet. De som lämnat systemet kom tidigare än de som ännu befinner sig där, och således kom dessa två grupper under disjunkta tidsintervall. Eftersom inprocessen är en Poissonprocess så är antalen kunder i dessa bägge grupper oberoende.

Satserna 7.6 och 7.8 är användbara då man har betjäning i serie, d.v.s. då en kund först köar för en betjäning och sedan direkt ställer sig i kö för nästa betjäning.

Exempel 7.7 (Lunch på KTH, forts.) När vi besöker något lunchställe på KTH så måste vi ju faktiskt stå i två köer. Vi har först betjäningsstationen där maten delas ut, och det är ju där som den ”stora” kön är. När vi äntligen fått maten så ska vi ju betala den och vi får då ställa oss i kön till kassan. Våra lunchställen är således egentligen exempel på seriekopplade betjäningssystem. □

Låt oss betrakta fallet med två betjäningssystem i serie, som illustreras i Figur 7.3. Som vanligt förutsätter vi att systemen befinner sig i stationärt tillstånd.

Betjäningstiderna i de bägge systemen är Exp(μ_1)- resp. Exp(μ_2)-fördelade. Låt $X_k(t)$ vara antalet kunder i system k. Utprocessen $U(t)$ från system 1, som enligt sats 7.6 är en Poissonprocess med intensitet λ, är inprocess till system 2. Enligt sats 7.8 följer att $X_1(t)$ och $X_2(t)$ är oberoende. Vi kan därför tillämpa (7.6) på sidan 81 för vart och ett av systemen, vilket ger

$$P(X_1 = i_1,\, X_2 = i_2) = P(X_1 = i_1)P(X_2 = i_2)$$

$$= \rho_1^{i_1}(1-\rho_1) \cdot \rho_2^{i_2}(1-\rho_2), \quad \text{för } i_1, i_2 = 0, 1, \ldots, \tag{7.9}$$

där $\rho_k = \lambda/\mu_k$.

Vi har hittills betraktat kösystemen främst ur kundernas synvinkel. Låt oss nu se på det hela ur expeditens synvinkel. Så länge det finns kunder så arbetar expediten och först då systemet är tomt så får expediten en – förmodligen välbehövlig – paus. Bildar vi nu processen $Y(t)$, definierad av

$$Y(t) = \begin{cases} 0 & \text{om } X(t) = 0 \\ 1 & \text{om } X(t) > 0, \end{cases}$$

så innebär $Y(t) = 0$ vila och $Y(t) = 1$ arbete för expediten vid tiden t. Processen $Y(t)$ är inte en Markovprocess men viloperioderna och sysselsättningsperioderna är oberoende. Det följer av Poissonprocessens egenskaper att viloperiodernas längder är Exp(λ)-fördelade. Fördelningen för sysselsättningsperiodernas längd kan beräknas, men det är svårt och fördelningen är komplicerad så vi lämnar det därhän. (Hade $Y(t)$ varit en Markovprocess så hade av nödvändighet även denna fördelning varit en exponentialfördelning, men det är den inte.) Man kan dock visa att

$$E[T] = \frac{1}{\mu - \lambda} \quad \text{och} \quad V[T] = \frac{\mu + \lambda}{(\mu - \lambda)^3},$$

där T är längden för en sysselsättningsperiod. Uttrycket för $E[T]$ är inte svårt att motivera genom att beräkna $P(Y(t) = 0)$ på två sätt. Enligt sats 7.6 på sidan 81 gäller att

$$P(Y(t) = 0) = P(X(t) = 0) = 1 - \rho = \frac{\mu - \lambda}{\mu}.$$

Å andra sidan gäller

$$P(Y(t) = 0) = \frac{\text{förväntad viloperiod}}{\text{förväntad viloperiod} + \text{förväntad sysselsättningsperiod}}$$

$$= \frac{1/\lambda}{1/\lambda + E[T]} = \frac{1}{1 + \lambda E[T]}.$$

Sätter vi ihop dessa två uttryck för $P(Y(t) = 0)$ så fås

$$\frac{\mu - \lambda}{\mu} = \frac{1}{1 + \lambda E[T]} \Rightarrow E[T] = \frac{1}{\lambda} \cdot \left(\frac{\mu}{\mu - \lambda} - 1 \right) = \frac{1}{\mu - \lambda}.$$

(Uttrycket för $E[T]$ kan även beräknas som i problem 58 på sidan 125.)

I exempel 7.6 på sidan 83 påpekade vi att de slumpmässiga variationerna i kölängd och väntetid är stora i ett kösystem där ρ är nära ett. Detta gäller i ännu högre grad för sysselsättningsperioderna. Som ett exempel väljer vi $\mu = 1$, vilket innebär att vi väljer den förväntade betjäningstiden som tidsenhet, och $\lambda = 0.9$. Vi får då $E[T] = w = 10$, d.v.s. den förväntade sysselsättningsperiodens längd överensstämmer med den förväntade tiden i systemet för en kund. Ur expeditens synvinkel verkar ju inte detta så ”farligt” utan kanske ”inte mer än rättvist”. Å andra sidan har vi $D[S] = 10$, vilket är mycket nog, och $D[T] = 43.6$. Ur expeditens synvinkel är sysselsättningsperiodernas ganska rimliga förväntade längd rätt ointressant i förhållande till de slumpmässiga variationerna. Ur arbetsmiljösynvinkel räcker det således inte att betrakta förväntade arbetsförhållanden i hårt belastade kösystem.

7.3.2 $M/M/c$-system

Vi generaliserar nu $M/M/1$-kön till fallet med c parallella betjäningsstationer.

Sats 7.9 *$X(t)$ är en födelse-döds-process med födelseintensiteter*

$$\lambda_0 = \lambda_1 = \ldots = \lambda \text{ (ankomstintensiteten)}$$

och dödsintensiteter

$$\mu_i = \begin{cases} i\mu, & i = 1, 2, \ldots, c, \\ c\mu, & i = c+1, c+2, \ldots \end{cases}$$

Bevis. Eftersom minimum av Exp-fördelade variabler är Exp-fördelad, se sats 2.3 på sidan 6, så bevisas detta väsentligen på samma sätt som sats 7.2 på sidan 80. □

Sats 7.10 *För ett $M/M/c$-system med $\rho = \frac{\lambda}{c\mu} < 1$ gäller*

$$p_n = \begin{cases} \left(\sum_{i=0}^{c-1} \frac{(c\rho)^i}{i!} + \frac{(c\rho)^c}{c!(1-\rho)}\right)^{-1} & \text{om } n = 0, \\ p_0 \cdot \frac{(c\rho)^n}{n!} & \text{om } n = 1, 2, \ldots, c, \\ p_c \cdot \rho^{n-c} & \text{om } n = c+1, c+2, \ldots \end{cases} \tag{7.10}$$

$$\ell_q = \frac{p_c\rho}{(1-\rho)^2} \tag{7.11}$$

$$1 - G_q(\tau) = \frac{p_c}{1-\rho} e^{-c\mu(1-\rho)\tau}, \quad \tau > 0. \tag{7.12}$$

Bevis. Ekvation (7.10) följer av sats 6.12 på sidan 68:

Vi har

$$\rho_i = \begin{cases} \frac{\lambda^i}{i!\mu^i} = \frac{(c\rho)^i}{i!} & \text{om } i < c \\ \frac{\lambda^i}{c!c^{i-c}\mu^i} = \frac{(c\rho)^i}{c!c^{i-c}} = \frac{(c\rho)^c}{c!}\rho^{i-c} & \text{om } i \geq c \end{cases}$$

vilket ger

$$\sum_{i=0}^{\infty} \rho_i = \sum_{i=0}^{c-1} \frac{(c\rho)^i}{i!} + \frac{(c\rho)^c}{c!} \sum_{i=c}^{\infty} \rho^{i-c} = \sum_{i=0}^{c-1} \frac{(c\rho)^i}{i!} + \frac{(c\rho)^c}{c!(1-\rho)}.$$

Formel (7.10) följer nu som en ren insättningsövning.

Ekvation (7.11) fås av

$$\ell_q = 0 \cdot \sum_{\nu=0}^{c} p_\nu + \sum_{\nu=1}^{\infty} \nu p_{c+\nu} = p_c \sum_{\nu=1}^{\infty} \nu \rho^\nu = p_c \frac{\rho}{(1-\rho)^2}.$$

För att härleda (7.12) krävs en hel del räkningar. Idén är, precis som i beviset av sats 7.5 på sidan 82, att betrakta en stokastisk summa av exponentialfördelade variabler. För att det skall bli positiv kötid krävs att alla betjäningsstationer är upptagna. Tiden till dess att första kunden lämnar systemet är Exp($c\mu$). Tiderna mellan kunder som lämnar systemet, så länge det är fullt, har samma fördelning. □

Faktorn $p_c/(1-\rho)$ i (7.12) är en intressant storhet i sig själv eftersom den är sannolikheten för att en kund måste köa:

$$P(\text{en kund måste köa}) = \sum_{i=c}^{\infty} p_i = p_c \sum_{i=c}^{\infty} \rho^{i-c} = \frac{p_c}{1-\rho}.$$

Denna sannolikhet kallas *Erlangs fördröjningsformel* eller Erlangs C-formel:

$$C_c = \frac{p_c}{1-\rho} = \frac{\frac{(c\rho)^c}{c!}}{(1-\rho)\left(\sum_{i=0}^{c-1} \frac{(c\rho)^i}{i!} + \frac{(c\rho)^c}{c!(1-\rho)}\right)}.$$

Erlangs fördröjningsformel finns i och för sig tabulerad. Har man inte tillgång till sådana tabeller kan en vanlig tabell över Poissonfördelningen användas. Låt Z vara en Po($c\rho$)-fördelad stokastisk variabel. Förlänger vi uttrycket för C_c med $e^{-c\rho}$ så fås

$$C_c = \frac{P(Z=c)}{(1-\rho)P(Z \le c-1) + P(Z=c)}.$$

Dessa beräkningsaspekter har nog främst historiskt intresse från en tid då datorer inte var så lättillgängliga som nu.

Utnyttjar vi sats 7.1 på sidan 79 och Erlangs fördröjningsformel fås

$$\ell_q = \frac{C_c \rho}{1-\rho} \tag{7.13}$$

$$\ell = \frac{\lambda}{\mu} + \frac{C_c \rho}{1-\rho} \tag{7.14}$$

$$w_q = \frac{C_c \rho}{\lambda(1-\rho)} \tag{7.15}$$

$$w = \frac{1}{\mu} + \frac{C_c \rho}{\lambda(1-\rho)}. \tag{7.16}$$

Vi har sett, även om formlerna blir lite mera komplicerade, att olika storheter i $M/M/c$-systemet i stort sett är naturliga generaliseringar av motsvarande storheter i $M/M/1$-systemet. Ett lite otrevligt undantag från detta är fördelningen G för S. I $M/M/1$-systemet är G enklare än G_q, och vi gick ju via G då vi i sats 7.5 på sidan 82 beräknade G_q. Problemet är att S inte längre är en stokastisk summa av oberoende och *likafördelade* exponentialfördelade stokastiska variabler. Man kan dock gå den omvända vägen och utgå från G_q

och beräkna G eftersom $S = Q + U$, där Q och U fortfarande är oberoende. Genomför vi detta fås

$$1 - G(\tau) = \begin{cases} \dfrac{(c(1-\rho) - 1 + C_c)e^{-\mu\tau} - C_c e^{-c\mu(1-\rho)\tau}}{c(1-\rho) - 1} & \text{om } c(1-\rho) \neq 1 \\ (1 + C_c\mu\tau)e^{-\mu\tau} & \text{om } c(1-\rho) = 1, \end{cases} \qquad \tau \geq 0.$$

Det kan tyckas förvånande att vi får två olika fall. Detta händer dock även för ”vanliga” exponentialfördelade variabler. Summan av två exponentialfördelade variabler med samma intensitet är ju Γ-fördelad, som vi visade i sats 6.4 på sidan 59, vilket svarar mot fallet $c(1-\rho) = 1$. Detta gäller dock ej om intensiteterna är olika, jmf. följdsats 7.1 på sidan 85.

Ett viktigt specialfall, åtminstone för att förenkla formlerna, är naturligtvis $M/M/2$-systemet.

Följdsats 7.2 *För ett $M/M/2$-system med $\rho = \frac{\lambda}{2\mu} < 1$ gäller*

$$p_n = \begin{cases} \frac{1-\rho}{1+\rho} & \text{om } n = 0, \\ \frac{2(1-\rho)\rho^n}{1+\rho} & \text{om } n = 1, 2, \ldots \end{cases} \tag{7.17}$$

$$C_c = \frac{2\rho^2}{1+\rho} \tag{7.18}$$

$$\ell_q = \frac{2\rho^3}{1-\rho^2} \tag{7.19}$$

$$1 - G_q(\tau) = \frac{2\rho^2}{1+\rho} e^{-2\mu(1-\rho)\tau}, \quad \tau > 0 \tag{7.20}$$

$$\ell = \frac{2\rho}{1-\rho^2} \tag{7.21}$$

$$w_q = \frac{\rho^2}{\mu(1-\rho^2)} \tag{7.22}$$

$$w = \frac{1}{\mu(1-\rho^2)}\,. \tag{7.23}$$

Bevis. Följdsatsen följer som en insättningsövning. □

Exempel 7.8 (Effektivisering av ett $M/M/1$-system) Låt oss betrakta ett $M/M/1$-system där belastningen börjar bli lite väl stor, d.v.s. köerna börjar bli väl långa och expediten börjar bli utsliten (eller för att använda ett modeord ”utbränd”). För att råda bot på detta så bestämmer man sig för att öka effektiviteten hos systemet genom att anställa ytterligare en expedit. Man kan nu genomföra effektiviseringen på minst tre olika sätt:

Fall 1 Man låter expediterna samarbeta om betjäningen av varje kund. Detta låter vi svara emot att vi har ett $M/M/1$-system med ingångsparametrar λ och 2μ. För mänskliga expediter är nog detta inte helt realistiskt, men om ”expediten” i själva verket är en dator så bör det vara fullt möjligt.

Fall 2 Man bildar två parallella $M/M/1$-system. Under förutsättning att kunderna väljer kö slumpmässigt, jmf. exempel 7.2 på sidan 72, så har vardera systemet ingångsparametrarna $\lambda/2$ och μ.

	Fall 1	Fall 2	Fall 3
ℓ	$\frac{\rho}{1-\rho}$	$\frac{2\rho}{1-\rho}$	$\frac{2\rho}{1-\rho^2}$
ℓ_q	$\frac{\rho^2}{1-\rho}$	$\frac{2\rho^2}{1-\rho}$	$\frac{2\rho^3}{1-\rho^2}$
w	$\frac{1}{2\mu(1-\rho)}$	$\frac{1}{\mu(1-\rho)}$	$\frac{1}{\mu(1-\rho^2)}$
w_q	$\frac{\rho}{2\mu(1-\rho)}$	$\frac{\rho}{\mu(1-\rho)}$	$\frac{\rho^2}{\mu(1-\rho^2)}$
$\ell \cdot \frac{1-\rho}{\rho}$	1	2	$\frac{2}{1+\rho}$
$\ell_q \cdot \frac{1-\rho}{\rho^2}$	1	2	$\frac{2\rho}{1+\rho}$
$w \cdot 2\mu(1-\rho)$	1	2	$\frac{2}{1+\rho}$
$w_q \cdot \frac{2\mu(1-\rho)}{\rho}$	1	2	$\frac{2\rho}{1+\rho}$

Tabell 7.1: Sammanställning av formler i exempel 7.8.

Fall 3 Man bildar ett $M/M/2$-system med ingångsparametrarna λ och μ.

Alla ovanstående kösystem har $\rho = \lambda/(2\mu)$. Frågan är nu vilket av alternativen som är bäst.

Jämför vi först de två $M/M/1$-systemen med parametrar λ och 2μ respektive $\lambda/2$ och μ så kan man säga att tiden går dubbelt så fort i det första fallet. Detta påverkar inte den stationära fördelningen för systemen och bägge systemen har därför samma ℓ_q och ℓ. Väntetiderna påverkas dock, och w_q och w är dubbelt så stora i det andra fallet som i det första. Den som inte utan vidare sväljer detta resonemang kan lätt kontrollera, se satserna 7.4 och 7.5 på sidorna 81 och 82, att det stämmer. Nu har vi ju dock två parallella system i fall 2 så det totala förväntade antalet kunder i kön eller i systemet är i själva verket även det dubbelt så stort i fall 2 som i fall 1. Sammanfattningsvis kan man säga att effektiviseringen i fall 1 är dubbelt så stor som i fall 2. Vi ska dock hålla i minnet att fall 1 långt ifrån alltid går att genomföra.

Mera intressant är att jämföra med fall 3. För att förenkla jämförelsen gör vi i tabell 7.1 en sammanställning, där uttrycken svarande mot fall 1 och fall 2 kommer från resonemanget ovan och uttrycken svarande mot fall 3 är tagna från följdsats 7.2. Formlerna i undre delen av tabell 7.1 är normerade för att ytterligare förenkla jämförelsen.

Tittar vi först på de normerade storheterna som hör till systemet, d.v.s. $\ell \cdot \frac{1-\rho}{\rho}$ och $w \cdot 2\mu(1-\rho)$, så ser vi att fall 3 uppför sig som fall 2 då $\rho \to 0$ och som fall 1 då $\rho \to 1$. Detta är ganska naturligt. Då ρ är liten så är systemet i regel tomt och då är betjäningstiden det viktigaste. Denna är densamma i fallen 2 och 3. Då ρ är nära 1 så finns det i regel många kunder i systemet. Då uppför sig systemen i fallen 1 och 3 lika, så när som att betjäningen av en kund, när den kommit igenom kön, är snabbare i fall 1 än i fall fall 3. Den skillnaden betyder dock ”i limes” inget.

För de normerade storheterna som hör till kön, d.v.s. $\ell_q \cdot \frac{1-\rho}{\rho^2}$ och $w_q \cdot \frac{2\mu(1-\rho)}{\rho}$

så ser vi att fall 3 mera effektivt än fall 1. Förklaringen är att det i fall 3 är större chans att bli ögonblickligen betjänad än i fall 1. Dock uppför sig fall 3 som fall 1 då $\rho \to 1$. Detta är naturligt med samma motivering som för systemet. □

7.3.3 $M/M/\infty$-system

Vi ska nu betrakta ett system med oändligt många parallella betjäningsstationer. Den matematiska analysen av $M/M/\infty$-systemet är mycket enklare än i $M/M/c$-systemet, eftersom det inte finns någon kö – alla kunder blir ju direkt betjänade då det alltid finns lediga betjäningsstationer. En användning av $M/M/\infty$-systemet är därför som en approximation av $M/M/c$-systemet då c är stort, jmf. problem 71 på sidan 129.

I sats 7.11 nedan ger vi en uppsättning resultat som svarar mot dem som vi betraktade för $M/M/1$- och $M/M/2$-systemen.

Sats 7.11 *För ett $M/M/\infty$-system med $r = \lambda/\mu$ gäller*

$$
\begin{aligned}
p_n &= \frac{r^n}{n!}\, e^{-r}, \quad n = 0, 1, \dots && (7.24)\\
\ell &= r && (7.25)\\
w &= \frac{1}{\mu} && (7.26)\\
1 - G(\tau) &= e^{-\mu\tau}, \quad \tau > 0 && (7.27)\\
\ell_q &= 0 && (7.28)\\
w_q &= 0. && (7.29)
\end{aligned}
$$

Bevis. Eftersom inga kunder behöver köa så gäller (7.28) och (7.29). En kunds tid i systemet är således detsamma som dess betjäningstid, vilket ger (7.26) och (7.27). Om vi kan visa (7.24), d.v.s. att antalet kunder i systemet är Po(r)-fördelat, så följer (7.25). Återstår således att visa (7.24).

Man inser, jmf. sats 7.9 på sidan 88, att $X(t)$ är en födelse-döds-process med födelseintensiteter $\lambda_i = \lambda$ och dödsintensiteter $\mu_i = i\mu$. Vi utnyttjar sats 6.12 på sidan 68. Eftersom

$$\rho_n = \frac{\lambda^n}{n!\mu^n} = \frac{r^n}{n!}$$

och

$$\sum_{i=0}^{\infty} \rho_i = e^r$$

så följer

$$p_n = \frac{\rho_n}{\sum_{i=0}^{\infty} \rho_i} = \frac{r^n}{n!}\, e^{-r},$$

vilket är (7.24). □

Vi ska nu roa oss med att se vad som händer om vi utnyttjar kända resultat för $M/M/c$-systemet och låter $c \to \infty$. I detta fall gäller då att $\rho \to 0$. I formlerna förekommer ofta $c\rho$ och vi ersätter då $c\rho$ med r eftersom $c\rho = r$ för alla $c < \infty$.

Om vi utnyttjar att $\ell_q = 0$ så fås av sats 7.1 på sidan 79 att

$$\ell = 0 + r = r, \quad w_q = 0 \quad \text{och} \quad w = \frac{r}{\lambda} = \frac{\lambda/\mu}{\lambda} = \frac{1}{\mu}.$$

Låt oss nu utnyttja (7.10) på sidan 88. Vi observerar att

$$\frac{(c\rho)^c}{c!} = \frac{r^c}{c!} \to 0 \quad \text{då } c \to \infty.$$

Då fås

$$p_0 = \Big(\sum_{i=0}^{\infty} \frac{r^i}{i!}\Big)^{-1} = \big(e^r\big)^{-1} = e^{-r} \quad \text{och} \quad p_n = \frac{r^n}{n!}\, e^{-r}.$$

Även (7.27) går, med viss möda, att få fram genom en gränsövergång. Ingen skulle nog, på fullt allvar, härleda (7.27) på detta sätt.

Vi ska nu ge ett ett exempel där $M/M/\infty$-systemet kan ses som en modell mera än som en approximation av ett $M/M/c$-system.

Exempel 7.9 (Bilar längs en väg) Vid en viss plats A längs en väg, med gles trafik, betraktar vi då bilar passerar i en viss riktning. Ett naturligt antagande är att bilarna kommer enligt en Poissonprocess $N(t)$ med någon intensitet λ. Antagandet om "gles trafik", d.v.s. att λ inte är så stort, är viktigt för Poisson-antagandet. I "tät trafik" påverkar bilarna varandra genom att lokala bilköer bildas, vilket strider mot antagandet om en Poissonprocess.

Vi betraktar nu en annan plats B längre fram längs vägen. Vi förutsätter att inga bilar försvinner eller tillkommer mellan dessa två platser. Tiden U det tar för en bil att förflytta sig mellan dessa platser är

$$U = \frac{\text{Avståndet mellan } A \text{ och } B}{\text{Medelhastigheten för bilen ifråga.}}$$

Låt nu $U_1, U_2, \ldots$ vara tiderna det tar för successiva bilar att förflytta sig mellan platserna A och B. Pga antagandet om "gles trafik" är det rimligt att anta att $U_1, U_2, \ldots$ är oberoende stokastiska variabler. Det är även naturligt att anta att de stokastiska variablerna är likafördelade med något väntevärde $1/\mu$. Vi kan uppfatta $N(t)$ som ankomstprocessen och $U_1, U_2, \ldots$ som betjäningstiderna i ett kösystem med oändligt många betjäningsstationer. För att få ett $M/M/\infty$-system måste vi dessutom anta att U_i-variablerna är exponentialfördelade. Alla som har kört bil, och som vet vad en exponentialfördelning innebär, inser att detta antagande är fullständigt orealistiskt. Det kommer dock att visa sig, något förvånande kanske, att antagandet om exponentialfördelning "inte gör så mycket".

Det följer nu av sats 7.11 att antalet som befinner sig mellan punkterna A och B är $\text{Po}(\lambda/\mu)$-fördelat, åtminstone då trafiken pågått en tid så att systemet befinner sig i stationärt tillstånd. $\square$

Som vi nämnde i exempel 7.9 ovan så ”gör det inte så mycket” om antagandet om exponentialfördelade betjäningstider är helt orealistiskt. Lite mera precist betyder detta att sats 7.11 gäller – med en trivial modifiering av (7.27) – även för ett $M/G/\infty$-system, d.v.s. då betjäningstiderna har en godtycklig fördelning B med väntevärde $1/\mu$. Vi återkommer till detta i avsnitt 7.4.2 på sidan 100.

7.3.4 Förlustsystem med $c \geq 1$

Ett förlustsystem med c betjäningsstationer, eller ett $M/M/c$-system utan kö är ett system där en kund som inte får betjäning omedelbart lämnar detta. Exempel på förlustsystem är telefonsamtal till en privatperson, $c = 1$, eller till en telefonväxel, $c \geq 1$, vilket vi diskuterade i exempel 7.4 på sidan 73. Naturligtvis kan man tänka sig helt andra typer av situationer som t.ex. en kiosk, jmf. exempel 7.1 på sidan 71, under ett ösregn. Detta är nog dock inte ett renodlat förlustsystem; en riktigt röksugen kund lär nog köa trots regnet.

På liknande sätt som i satserna 7.2 och 7.9 på sidorna 80 och 88 följer att $X(t)$ är en födelse-döds-process med tillståndsrummet $\{0, 1, \ldots, c\}$, födelse-intensiteter

$$\lambda_0 = \lambda_1 = \ldots = \lambda_{c-1} = \lambda$$

och dödsintensiteter

$$\mu_i = i\mu, \quad i = 1, 2, \ldots, c.$$

Sats 7.12 *För ett förlustsystem med $c \geq 1$ gäller att*

$$p_n = \begin{cases} \dfrac{\frac{r^n}{n!}}{1 + \frac{r}{1!} + \frac{r^2}{2!} + \ldots + \frac{r^c}{c!}} & \text{om } n = 0, 1, \ldots, c \\ 0 & \text{om } n = c + 1, c + 2, \ldots \end{cases}$$

där $r = \lambda/\mu$.

Bevis. Satsen är (tämligen) direkt följd av sats 6.12 på sidan 68. □

På liknande sätt som för Erlangs fördröjningsformel kan sannolikheterna p_n beräknas med hjälp av en tabell över Poissonfördelningen. Låt Z vara en Po(r)-fördelad stokastisk variabel. Förlänger vi uttrycket för p_n med e^{-r} så fås

$$p_n = \frac{P(Z = n)}{P(Z \leq c)} = P(Z = n \mid Z \leq c), \quad \text{för } n = 0, 1, \ldots, c.$$

Denna fördelning kallas för en *trunkerad Poissonfördelning.*

Sannolikheten p_c är av särskilt intresse eftersom det är sannolikheten för att en kund ska bli ”förlorad”, eller annorlunda uttryckt sannolikheten för att en kund omedelbart ska lämna systemet utan att bli betjänad. Sannolikheten p_c kallas ofta *spärrsannolikheten.* Andra namn på p_c är *Erlangs förlustformel* eller Erlangs första formel. Beteckningen Erlangs B-formel förekommer även – kärt barn har många namn. Explicit utskriven är spärrsannolikheten given av

$$B_c(r) = p_c = \frac{\frac{r^c}{c!}}{1 + \frac{r}{1!} + \frac{r^2}{2!} + \ldots + \frac{r^c}{c!}}. \tag{7.30}$$

Vi har i formuleringen av sats 7.12 valt att använda beteckningen r i stället för ρ eller $c\rho$ eftersom kvoten mellan λ och μ här inte direkt är ett mått på belastningen av betjäningsstationerna.

För att en kund ska komma in i systemet under tidsintervallet $(t, t+dt)$, där vi förutsätter att t är så stort att systemet befinner sig i stationärt tillstånd, krävs två händelser:

- Det ska komma en potentiell kund, d.v.s. en kund som vill in i systemet, under tidsintervallet $(t, t+dt)$. Sannolikheten för detta är λdt.
- Den potentiella kunden ska bli en verklig kund. För detta krävs att systemet inte är fullt vilket är fallet med sannolikheten $1 - p_c$.

Eftersom processen av potentiella kunder är en Poissonprocess så är dessa två händelser oberoende. Sannolikheten för en kund ska komma in i systemet under tidsintervallet $(t, t+dt)$ är således Λdt där

$$\Lambda = \lambda \cdot (1 - p_c).$$

Inprocessen av kunder in i systemet är därför en punktprocess med intensiteten Λ. Ett naturligt mått på belastningen av systemet verkar vara

$$\tilde{\rho} = \frac{\Lambda}{c\mu}.$$

Det är uppenbart av systemets konstruktion att $\ell_q = w_q = 0$, eftersom det inte finns någon kö, och att $w = 1/\mu$ eftersom tiden i systemet överensstämmer med betjäningstiden. Enligt (7.5) på sidan 79 fås

$$\ell = \Lambda w = \lambda \cdot (1 - p_c)/\mu. \tag{7.31}$$

Ska nu (7.3) stämma så är $\tilde{\rho}$ det ”riktiga” måttet på belastningen, jmf. diskussionen i slutet på avsnitt 7.2.

En kritisk läsare kan, högst motiverat, ställa sig frågan om man verkligen får använda sats 7.1 på sidan 79 som vi gjort. Det är emellertid lätt att visa att (7.31) stämmer. Vi har

$$\ell = \sum_{n=0}^{c} np_n = p_0 \sum_{n=0}^{c} n\frac{r^n}{n!} = p_0 r \sum_{n=1}^{c} \frac{r^{n-1}}{(n-1)!} = r \cdot (1 - p_c) = \lambda \cdot (1 - p_c)/\mu$$

vilket är (7.31).

7.3.5 Modifierade $M/M/1$-system

Man kan givetvis modifiera ett $M/M/1$-system på en mängd olika sätt. Vi ska här bara betrakta sådana modiferingar där markoviteten hos $X(t)$ bibehålles. Ett modifierat $M/M/1$-system beskrivs således av födelseintensiteterna λ_k och dödsintensiteterna μ_k. Detta innebär att endast antalet kunder i systemet får styra kundernas och expeditens beteende.

Vi ska begränsa våra exempel till situationer där antingen kundernas eller expeditens uppträdande beror av $X(t)$.

Vi börjar med att betrakta fallet då expeditens beteende beror av kön. Allmänt verkar det rimligt att tänka sig att expediten försöker ”snabba på” betjäningen om det finns många i kön. En naturlig modifiering kan vara att låta μ_n vara växande i n. I verkligheten kan naturligtvis önskemålet om att snabba upp betjäningen då kön är lång vara ett gemensamt önskemål hos kunderna och expediten; vem vågar börja diskutera vinsorter på systemet då en törstig kö står och stampar.

Exempel 7.10 (Social expedit) Låt oss betrakta butiken i ett litet samhälle där alla känner alla. Handlandet i en sådan by innebär ofta inte bara att man får de varor som man behöver, utan är även en viktig del av umgänget. Kanske passerar minst lika mycket information som varor disken. En hel del av denna information kan vara av den karaktären att både kunden och expediten vill ge intryck av att informationen är ”exklusiv”. Skvaller är ett mindre elegant uttryck för detta. En rimlig modell kan då vara

$$\mu_1 = \mu \quad \text{och} \quad \mu_2 = \mu_3 = \ldots = c\mu \quad \text{där } c > 1.$$

I fallet $c = 2$ uppför sig $X(t)$ som om vi hade haft ett $M/M/2$-system och vi begränsar oss till detta fall. Det följer således av (7.17) att

$$p_0 = \frac{2\mu - \lambda}{2\mu + \lambda}$$

och av (7.21) att

$$\ell = \frac{\lambda/\mu}{1 - (\lambda/(2\mu))^2} = \frac{4\mu\lambda}{(2\mu + \lambda)(2\mu - \lambda)}\,.$$

Skälet till att vi har uttryckt (7.17) och (7.21) i λ och μ är att den naturliga definitionen av ρ, jmf. (7.2), är

$$\rho = 1 - p_0 = \frac{2\lambda}{2\mu + \lambda}\,.$$

Eftersom betjäningstiderna inte är oberoende av ankomsterna så kan sats 7.1 på sidan 79 inte utan vidare tillämpas. Med vår definition av ρ gäller dock (7.3) vilket ger

$$\ell_q = \ell - \rho = \frac{2\lambda^2}{(2\mu + \lambda)(2\mu - \lambda)}\,.$$

Resonemanget som ledde till Littles formel, eller (7.4), fungerar oförändrat, vilket ger

$$w_q = \frac{2\lambda}{(2\mu + \lambda)(2\mu - \lambda)}\,.$$

Eftersom vi bara har en betjäningsstation fungerar detta resonemang även på hela systemet, och inte bara på kön, vilket medför att även (7.5) gäller, d.v.s.

$$w = \frac{4\mu}{(2\mu + \lambda)(2\mu - \lambda)}\,.$$

Trots beroende gäller således sats 7.1 i detta fall. Vi har vidare, vilket inte alls är något uppenbart resultat,

$$\text{medelbetjäningstid} = w - w_q = \frac{2}{2\mu + \lambda} .$$

$\square$

Vi går nu över till modifieringar av kundernas beteende, d.v.s. vi låter födelseintensiteterna variera.

Exempel 7.11 (Reparationssystem) Vi ska nu betrakta ett $M/M/1$-system där populationen av kunder är ändlig. Om byn i exempel 7.10 är mycket liten och inga andra än byborna skulle komma på idén att handla i affären så kan man ju säga att dessa utgör populationen av kunder. Det är nog i detta fall naturligare att inte tänka på mänskliga köer. Vi kommer att låta kunderna vara maskiner och expediten vara en reparatör, vilken förvisso fortfarande är en människa.

Vi antar att vi har m stycken likvärdiga maskiner, vilka servas av en reparatör. En maskin kan befinna sig i tre olika tillstånd:

- Maskinen fungerar. Vi antar att en maskin fungerar $\text{Exp}(\alpha)$-fördelat länge innan den behöver repareras.
- Maskinen är trasig och väntar på reparation.
- Maskinen repareras. En reparation tar $\text{Exp}(\mu)$-fördelat lång tid. Efter reparationen återgår maskinen till att fungera.

Vi förutsätter vidare att alla ingående stokastiska variabler i modellen är oberoende. Låt $X(t)$ vara antalet trasiga maskiner vid tid t. På samma sätt som i exempel 6.3 på sidan 49 följer att $X(t)$ är en födelse-döds-process med tillståndsrummet $\{0, 1, \ldots, m\}$, födelseintensiteter

$$\lambda_0 = m\alpha,\ \lambda_1 = (m-1)\alpha,\ \lambda_2 = (m-2)\alpha,\ \ldots,\ \lambda_{m-1} = \alpha$$

och dödsintensiteter

$$\mu_1 = \ldots = \mu_m = \mu.$$

Detta ger

$$\rho_0 = 1,\ \rho_1 = \frac{m\alpha}{\mu} = \frac{m!\alpha}{(m-1)!\mu},$$

$$\rho_2 = \frac{m(m-1)\alpha^2}{\mu^2} = \frac{m!\alpha^2}{(m-2)!\mu^2},\ \ldots,\ \rho_m = \frac{m!\alpha^m}{\mu^m},$$

eller, på en mera kompakt form,

$$\rho_n = \frac{m!}{(m-n)!}\left(\frac{\alpha}{\mu}\right)^n \quad \text{för } n = 0, 1, \ldots m.$$

Sats 7.13 *För ett reparationssystem med $m \geq 1$ maskiner och en reparatör gäller att*

$$p_n = \frac{\frac{m!}{(m-n)!}\left(\frac{\alpha}{\mu}\right)^n}{\sum_{i=0}^{m} \frac{m!}{(m-i)!}\left(\frac{\alpha}{\mu}\right)^i} \quad \text{för } n = 0, 1, \ldots, m.$$

Bevis. Satsen följer av sats 6.12 på sidan 68. □

Speciellt följer av sats 7.13 att

$$p_0 = \frac{1}{\sum_{i=0}^{m} \frac{m!}{(m-i)!}\left(\frac{\alpha}{\mu}\right)^i} = \frac{\left(\frac{\mu}{\alpha}\right)^m \frac{1}{m!}}{\sum_{i=0}^{m} \left(\frac{\mu}{\alpha}\right)^m \frac{1}{m!} \frac{m!}{(m-i)!}\left(\frac{\alpha}{\mu}\right)^i} = \frac{\left(\frac{\mu}{\alpha}\right)^m \frac{1}{m!}}{\sum_{j=0}^{m} \left(\frac{\mu}{\alpha}\right)^j \frac{1}{j!}} = B_m\left(\frac{\mu}{\alpha}\right),$$

så Erlangs förlustformel dyker upp även här.

I motsats till mänskliga köer så är nog ℓ det intressantaste förväntade värdet – det talar ju om hur många maskiner som i genomsnitt är oanvändbara i produktionen – och det kan direkt beräknas med hjälp av sats 7.13. □

Exempel 7.12 (Köaversion) Det är rimligt att förmoda att ett vanligt beteende hos personer som egentligen önskar besöka ett betjäningsställe ”drar öronen åt sig” om de ser en lång kö. Matematiskt kan vi beskriva detta som att ankomstintensiteterna λ_k avtar med växande k och att λ_0 är intensiteten av potentiella kunder. Det mest extrema exemplet på köaversion är ett förlustsystem där de potentiella kunderna vägrar stå i kö. Naturligvis kan man även tänka sig situationer då en lång kö är ”lockande”: reor, krogar en fredag- eller lördagskväll.

En enkel modell för köaversion är

$$\lambda_n = \frac{\lambda}{n+1}.$$

Sätt, som vanligt,

$$\rho_0 = 1 \quad \text{och} \quad \rho_n = \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n} = \frac{(\lambda/\mu)^n}{n!}.$$

Detta ger $\sum_{i=0}^{\infty} \rho_i = e^{\lambda/\mu}$ och det följer av sats 6.12 på sidan 68 att

$$p_n = \frac{(\lambda/\mu)^n}{n!} e^{-\lambda/\mu},$$

d.v.s. X är Po(λ/μ)-fördelad. Av detta följer att

$$\ell = \frac{\lambda}{\mu}.$$

Med ungefär samma resonemang som vi gjorde för förlustsystem inser man att inprocessen i systemet är en punktprocess med intensiteten

$$\Lambda = \sum_{n=0}^{\infty} \lambda_n p_n = \lambda \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{(n+1)!} e^{-\lambda/\mu} = \mu \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^{(n+1)}}{(n+1)!} e^{-\lambda/\mu} = \mu(1 - e^{-\lambda/\mu}).$$

Sätter vi $\rho = \Lambda/\mu$ fås

$$\rho = 1 - e^{-\lambda/\mu} = 1 - p_0.$$

Sats 7.1 på sidan 79 ger nu

$$\ell_q = \frac{\lambda}{\mu} - 1 + e^{-\lambda/\mu}, \quad w = \frac{\ell}{\Lambda} = \frac{\lambda}{\mu^2(1 - e^{-\lambda/\mu})} \text{ och } w_q = \frac{\ell}{\Lambda} = w - \frac{1}{\mu}.$$

□

7.4 $M/G/c$-system

Vi ska bara betrakta fallen $c = 1$ och ∞.

7.4.1 $M/G/1$-system

Vi släpper nu antagandet att betjäningstiderna är exponentialfördelade och underförstår att dessa har någon annan fördelning men behåller antagandet om Poissonankomster. Denna generalisering nämndes i avsnitt 6.2, och vi ska nöja oss med att ge några kompletterande kommentarer. Som vi tidigare påpekat är nog antagandet om Poissonankomster ofta ganska realistiskt medan antagandet om exponentialfördelade betjäningstider är betydligt mera tveksamt. Generaliseringen av $M/M/1$ till $M/G/1$ är därför ur tillämpningssynpunkt högst relevant.

Knepet är att betrakta $X(t)$ vid de tidpunkter då kunder lämnar systemet. Låt $T_1, T_2, \ldots$ vara dessa slumpmässiga tidpunkter, och sätt

$$\widetilde{Y}_n = X(T_n + 0),$$

d.v.s. antalet kunder i systemet omedelbart efter att den n-te kunden lämnat detta. Man "inser" att $\widetilde{Y}_n$ är en Markovkedja med övergångsmatris

$$\boldsymbol{P} = \begin{pmatrix} k_0 & k_1 & k_2 & k_3 & \cdots \\ k_0 & k_1 & k_2 & k_3 & \cdots \\ 0 & k_0 & k_1 & k_2 & \cdots \\ 0 & 0 & k_0 & k_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

där

$$k_r = \int_0^\infty \frac{(\lambda t)^r}{r!} e^{-\lambda t} dB(t).$$

I integralen ovan tolkas $dB(t)$ som $b(t)\,dt$ om $B(t)$ är kontinuerlig med täthet $b(t)$.

Skälet till att vi genomgående skiljer på diskreta och kontinuerliga fördelningar är just för att kunna använda "vanliga" integraler. Om B är en kontinuerlig fördelning så fås genom partialintegration

$$b = E[U] = \int_0^\infty (1 - B(u))\,du \text{ och } E[U^2] = 2\int_0^\infty u(1 - B(u))\,du.$$

Dessa formler gäller i själva verket för godtyckliga fördelningar på positiva halvaxeln, så vi kan alltid beräkna väntevärdet och variansen med vanliga integraler.

Att ”inse” att matrisen $\boldsymbol{P}$ är övergångsmatrisen för den inbäddade Markovkedjan då $\widetilde{Y}_n \geq 1$ är nog inte så svårt. Om $\widetilde{Y}_n = 0$ så ”händer ingenting” förrän en kund kommer, och då är vi i samma situation som om $\widetilde{Y}_n = 1$.

Med hjälp av transformer kan man uttrycka den stationära fördelningen för $\widetilde{Y}_n$. Detta är även den stationära fördelningen för $X(t)$. Att detta är fallet är inte alls uppenbart, utan hänger, på ett ganska djupt sätt, ihop med att Poissonprocessen har oberoende inkrement. Detta faktum brukar kallas *PASTA* vilket ska utläsas ”Poisson Arrivals See Time Averages”.

Den naturliga definitionen av ρ är nu

$$\rho = \lambda b$$

Sats 7.14 *För ett $M/G/1$-system gäller*

$$\ell_q = \frac{\rho^2}{2(1-\rho)}\left(1 + \frac{V[U]}{b^2}\right). \tag{7.32}$$

Med hjälp av sats 7.1 på sidan 79 kan ℓ, w_q och w beräknas.

Vi lämnar sats 7.14 utan bevis. Formeln (7.32) kallas Pollaczeks formel, eller för att vara exakt, är en av de formler som kallas Pollaczeks eller Pollaczek-Khintchines formel.

Sett ur tillämpningssynvinkel så ökar inte svårighetsgraden särskilt mycket om vi tillåter allmänna betjäningstidsfördelningar, så länge vi nöjer oss med förväntade värden.

7.4.2 $M/G/\infty$-system

Vi ska nu betrakta ett system med oändligt många parallella betjäningsstationer. Låt B beteckna fördelningen för en godtycklig betjäningstid U och sätt $b = E[U]$, jmf. ovan.

Som vi nämnde i avsnittet 7.3.3 om $M/M/\infty$-system så gäller sats 7.11 på sidan 92 med en trivial modifiering. Detta innebär att följande sats gäller.

Sats 7.15 *För ett $M/G/\infty$-system med $r = \lambda b$ gäller*

$$p_n = \frac{r^n}{n!}e^{-r}, \quad n = 0, 1, \ldots \tag{7.33}$$

$$\ell = r \tag{7.34}$$

$$w = b \tag{7.35}$$

$$G(\tau) = B(\tau) \tag{7.36}$$

$$\ell_q = 0 \tag{7.37}$$

$$w_q = 0. \tag{7.38}$$

Bevisskiss. Eftersom inga kunder behöver köa så gäller (7.37) och (7.38). En kunds tid i systemet är således detsamma som dess betjäningstid, vilket ger (7.35) och (7.36). Om vi kan visa (7.33), d.v.s. att antalet kunder i systemet är Po(r)-fördelat, så följer (7.34). Återstår således att visa (7.33). Så här långt har vi ordagrant följt beviset av sats 7.11.

För att åtminstone antyda hur (7.33) kan inses antar vi att systemet har fungerat oändligt länge och är definierat på $(-\infty, \infty)$. Vi kan därför betrakta antalet kunder X i systemet vid tiden $t = 0$. (Detta är bara ett annat sätt att uttrycka att X är antalet kunder i systemet efter lång tid eller då systemet befinner sig i stationärt tillstånd.) Välj nu ett $t > 0$ och betrakta tidsintervallet $(-t, -t+h)$, där h är ett litet positivt tal. Ankomsterna i intervallet $(-t, -t+h)$ bidrar till X om dels

- det kommer en kund i intervallet $(-t, -t + h)$, vilket sker med sannolikheten $\lambda h + o(h)$,

och dels

- kunden ifråga är kvar i systemet vid tid 0, vilket sker om kunden ännu inte är färdigbetjänad och har sannolikheten $P(U > t) = 1 - B(t)$.

Av detta följer att intervallet $(-t, -t + h)$ ”bidrar” till X med sannolikhet $\lambda(1 - B(t))h$. Antalet kunder som kommer från detta intervall är således approximativt $\text{Bin}(1, \lambda(1 - B(t))h)$-fördelat och därmed även approximativt $\text{Po}(\lambda(1 - B(t))h)$-fördelat eftersom h är litet. Låt oss nu indela hela negativa tids-axeln i små intervall. Bidragen från disjunkta intervall är, pga. Poisson-processens egenskaper, oberoende. Eftersom summor av oberoende Poisson-fördelade stokastiska variabler är Poissonfördelade så ”följer” att X är Poissonfördelad. (Här återstår en del för att resonemanget ska bli strikt.) Väntevärdet för X fås om vi summerar över alla intervall. Om vi nu låter $h \to 0$, eller om man så vill ersätter h med dt, så ”fås”

$$E[X] = \int_0^\infty \lambda(1 - B(t))\, dt = \lambda b = r,$$

vilket antyder att (7.33) är sann. □

7.5 $GI/M/1$-system

Som vi tidigare har påpekat så är är denna generalisering populär, förmodligen främst för att metodiken för $M/G/1$-systemet på det stora hela fungerar även här. I detta fall betraktar man $X(t)$ vid de tidpunkter då kunder kommer till systemet. Låt $T_1, T_2, \ldots$ vara dessa slumpmässiga tidpunkter, och sätt

$$\widetilde{Y}_n = X(T_n - 0),$$

d.v.s. antalet kunder i systemet omedelbart innan den n-te kunden ankommer. I detta fall är ankomstprocessen en förnyelseprocess vilket innebär att

$$T_1, T_2 - T_1, T_3 - T_2, \ldots$$

är oberoende stokastiska variabler med fördelningsfunktion $A(t)$ och väntevärde $a = 1/\lambda$. Den naturliga definitionen av ρ är i detta fall

$$\rho = 1/(a\mu) = \lambda/\mu.$$

Man "inser" även här att $\widetilde{Y}_n$ är en Markovprocess. Då $\rho < 1$ existerar en gränsfördelning $\tilde{\boldsymbol{p}}$ för $\widetilde{Y}_n$, där speciellt

$$\tilde{p}_0 = P(\text{en ankommande kund blir genast betjänad}).$$

En väsentlig skillnad mot $M/G/1$-systemet är att PASTA-egenskapen saknas, d.v.s. $\tilde{\boldsymbol{p}}$ är inte den stationära fördelningen för $X(t)$. Detta innebär, med andra ord, att köns stokastiska egenskaper vid de tidpunkter då kunder ankommer skiljer sig från dess egenskaper vid en godtycklig tidpunkt. Man talar ofta om en *virtuell kund*, vilket är en tänkt kund som kommer vid en godtycklig tidpunkt då kön är i stationärt tillstånd. Det intressanta bör ju vara riktiga kunder och inte virtuella kunder, så allt kan ju tyckas vara gott och väl. Man måste dock hålla i minnet att sats 7.1 på sidan 79 gäller för virtuella kunder och inte för riktiga kunder, något som vi kortfattad diskuterade efter sats 7.1.

Man kan visa att

$$\tilde{\ell} = \lim_{n\to\infty} E[\widetilde{Y}_n] = \frac{1-\tilde{p}_0}{\tilde{p}_0}$$

och att $\tilde{p}_0$ är den entydiga lösningen till ekvationen

$$1 - \tilde{p}_0 = E\Big[e^{-\mu\tilde{p}_0 T_1}\Big] \left(= \int_0^\infty e^{-\mu\tilde{p}_0 t} dA(t)\right).$$

En ankommande kund måste köa under den tid det tar för kunden som betjänas att bli färdigbetjänad och ytterligare den tid det tar för kunderna i kön att betjänas. Enligt 2.2 på sidan 6 är den tid det tar för kunden som betjänas att bli färdigbetjänad Exp(μ) liksom betjäningstiderna för de köande kunderna. Den förväntade tiden tills alla i systemet blivit färdigbetjänade kan man visa är förväntat antal kunder gånger betjäningstiden för de enskilda kunderna (jämför 6.4 på sidan 59), d.v.s.

$$\tilde{w_q} = \tilde{\ell}/\mu,$$

där $\tilde{w_q}$ är kundens förväntade kötid. Observera att vi utnyttjat att betjäningstiderna är exponentialfördelade och att $\tilde{w_q} = \tilde{\ell}/\mu$ inte "har med sats 7.1 att göra".

Man kan vidare visa att

$$\ell = \lim_{t\to\infty} E[X(t)] = \frac{\rho}{\tilde{p}_0},$$

vilket är det förväntade antalet kunder i systemet då en virtuell kund ankommer. Här kan sats 7.1 användas på vanligt sätt.

För resten av dessa kommentarer skulle vi egentligen vilja använda ännu mindre stil, men då blir dessa nog även typografiskt oläsbara. När vi diskuterade villkoren för sats 7.1 så nämnde vi att inprocessen skulle vara mer eller mindre stationär. Den enda förnyelseprocessen som verkligen är stationär är Poissonprocessen. Detta hänger ihop med att exponentialfördelning är den enda fördelning som saknar minne. Man kan dock visa att om man låter T_1 ha fördelningsfunktionen

$$P(T_1 \le t) = \frac{1}{a}\int_0^t (1 - A(s))\, ds,$$

men bibehåller oberoendet och fördelningen A för $T_2 - T_1, T_3 - T_2, \ldots$, så blir inprocessen stationär. Man är nog villig att tro på att denna ganska obetydliga modifiering inte påverkar köns egenskaper i dess stationära tillstånd.

7.6 $G/M/1$-system

Vi ska bara betrakta ett exempel som visar att sats 7.1 på sidan 79, som ju gäller "under mycket allmänna villkor", trots allt inte gäller hur allmänt som helst.

Vi utgår från $M/M/1$-systemet och låt oss ha exempel 7.1 på sidan 71 i minnet. Beroende på var kiosken ligger, så är kundtillströmningen mer eller mindre stabil dag från dag. En naturlig fråga för expediten vid arbetets början är då: Jag undrar hur jobbigt det blir i dag?

Mera precist uttryckt så kan det vara rimligt att förutsätta att kunderna visserligen kommer enligt en Poissonprocess varje dag, men att dess intensitet kan variera dag från dag. Denna variation kan man beskriva genom att uppfatta "dagens ankomstintensitet" som utfall av en stokastisk variabel L. En expedit med en god utbildning i matematisk statistik hade då kunnat formulera sin fråga som: Jag undrar vilket utfallet av L blir i dag?

En rimlig modell för ankomstprocessen är att först generera ett utfall L och sedan använda detta utfall som intensitet i Poissonprocessen. En ankomstprocess uppkommen på detta sätt kallas en *vägd Poissonprocess* och är en Coxprocess, se exempel 7.5 på sidan 75.

Antag nu att $P(L < \mu) = 1$ så att systemet för alla utfall av L kan förutsättas befinna sig i stationärt tillstånd. För att analysera kön så betingar vi av utfallet av L och räknar som för ett $M/M/1$-system. De storheter som vi beräknar blir därför funktioner av L, d.v.s. stokastiska variabler. Vi kan nu tillämpa t.ex. (7.5) för varje utfall av av L, vilket ger

$$\ell(L) = Lw(L).$$

Viktar vi över utfallen av L så ger detta, då L och $w(L)$ är positivt korrelerade, att

$$\ell = E[\ell(L)] = E[Lw(L)] > E[L]E[w(L)] = \lambda w,$$

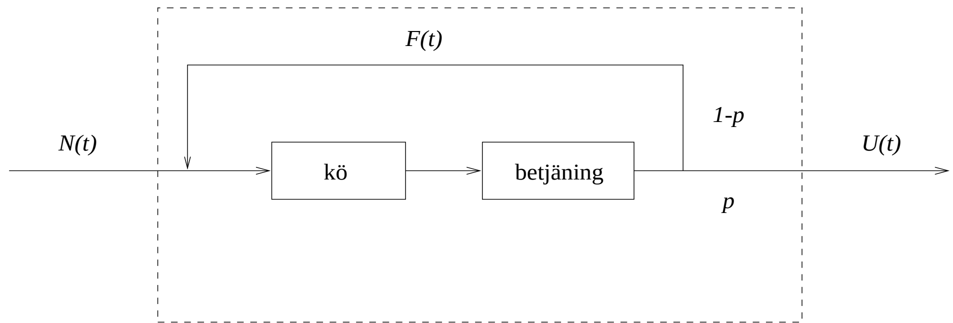
och således gäller (7.5) inte då ankomstprocessen är en vägd Poissonprocess. Det villkor för sats 7.1 som den vägda Poissonprocessen bryter mot är ergodicitetskravet, som (nästan) innebar att inprocessens stokastiska egenskaper på intervall som tidsmässigt ligger långt ifrån varandra skulle vara oberoende. I det vägda Poissonfallet beror de stokastiska egenskaperna på L, oavsett hur långt ifrån varandra intervallen än ligger.

7.7 Köer med återkoppling

Vi har hittills betraktat "klassisk" köteori där – i stort sett – kunder kommer, ställer sig i kö och går nöjda(?) därifrån när de har blivit betjänade. Betjäning i serie, d.v.s. då en kund först köar för en betjäning och sedan direkt ställer sig i kö för nästa betjäning, kan dock ses som en avvikelse från den "klassiska" situationen.

En modern tillämpning av köteori är modellering av datasystem, där situationen kan vara annorlunda. I ett fleranvändarsystem kommer då och då en ny kund, vilket svarar mot att någon vill köra ett nytt program. Programmet får köa tills det får plats i CPUn. Där "betjänas" programmet tills beräkningarna antingen är färdiga eller tills ytterligare information behövs. I det senare fallet sänds programmet till I/O (in-out) enheten där det eventuellt igen måste köa tills det betjänas. Efter I/O betjäningen får progammet åter köa tills det får plats i CPUn osv. tills beräkningarna är färdiga. Med viss reservation för beskrivningen av datasystemets arbetssätt – den datakunnige som läst ända hit må förtjäna ett gott skratt – så leder detta till en etablerad modell som vi ska återkomma till i exempel 7.16 på sidan 108.

Låt oss nu betrakta det enklast tänkbara systemet med återkoppling.



Figur 7.4: Illustration av ett kundströmmar vid ”enkel återkoppling”.

Exempel 7.13 (Enkel återkoppling) I ett $M/M/1$-system tänker vi oss att betjäningen i själva verket är någon typ av reparation. När en enhet (kund) är reparerad (betjänad) så görs en kontroll av reparationen. Om allt är gott och väl, d.v.s. att reparationen har lyckats, så lämnar enheten systemet. Vi förutsätter att detta sker med sannolikheten p. Reparatören är dock inte perfekt utan ibland fungerar inte enheten trots reparationen och i så fall återförs enheten till kön. Detta sker med sannolikheten $1-p$. Vi förutsätter vidare att huruvida en reparation lyckas eller ej är oberoende av hur många reparationsförsök som gjorts. Den läsare som tycker att reparatören verkar slarvig och således borde ersättas av en noggrannare kan byta ordet ”reparation” mot ”test av ett program” och blir kanske då mera överseende! Systemet och dess olika kundströmmar är illustrerat i figur 7.4.

Vi antar, som vanligt, att $N(t)$ är en Poissonprocess med intensitet λ och att reparationstiderna (betjäningstiderna) är Exp(μ)-fördelade. Av vårt antagande att ”huruvida en reparation lyckas eller ej är oberoende av hur många reparationsförsök som gjorts” följer att det totala antalet reparationer, eller snarare reparationsförsök, som behövs är ffg(p)-fördelat. Den totala reparationstiden är då enligt sats 6.4 b) på sidan 59 Exp($p\mu$)-fördelad. Detta innebär att vi i detta fall kan ”smita ifrån” problemet med återkoppling och betrakta hela systemet, d.v.s. det som i figur 7.4 är inramat med de streckade linjerna, som ett $M/M/1$-system med parametrar λ och $p\mu$. Under förutsättning att

$$\rho = \frac{\lambda}{p\mu} < 1$$

så följer det av satserna 7.3 och 7.6 på sidorna 81 och 83 att antalet kunder $X(t)$ har en gränsfördelning

$$p_n = \rho^n(1-\rho) = \left(\frac{\lambda}{p\mu}\right)^n \left(1 - \frac{\lambda}{p\mu}\right), \quad \text{för } n = 0, 1, \ldots \tag{7.39}$$

och att utprocessen $U(t)$ är en Poissonprocess med intensitet λ.

Låt oss nu betrakta systemet ur reparatörens synvinkel, som ser de enheter som kommer, d.v.s. uppfattar $N(t) + F(t)$ som inprocessen. Detta är en punktprocess, och låt oss beteckna dess intensitet med Λ. Kontrollanten uppfattar utprocessen som en punktprocess, även den med intensiteten Λ. Denna utprocess uppdelas nu i processerna $U(t)$, som vi vet är en Poissonprocess med intensitet λ, och $F(t)$. Beteckningen $F(t)$ är vald efter engelskans uttryck "feedback". Man inser att $U(t)$ har intensitet $p\Lambda = \lambda$, vilket ger

$$\Lambda = \frac{\lambda}{p},$$

och att $F(t)$ har intensitet

$$(1-p)\Lambda = \frac{(1-p)\lambda}{p}.$$

I motsats till processen $U(t)$ är $F(t)$ *inte* en Poissonprocess. Detta kan vid första, och kanske även andra, anblicken verka förvånande. Om vi speciellt har $p = 1/2$ så kan man ju tycka att bägge processerna av symmetriskäl borde vara av samma typ. Så är det dock inte, vilket man kan inse om man väljer λ litet, p litet och μ så stor att ρ är litet. Med detta val av parametrar så gäller i stort sett att då en enhet kommer till reparatören så är han ledig och gör ett stort antal mycket snabba reparationsförsök. Slutligen lyckas han och får åter vila sig tills nästa enhet kommer. Detta innebär att $F(t)$ då och då innehåller ett stort antal ihopklumpade punkter, och detta motsäger att $F(t)$ är en Poissonprocess. Processen $U(t)$, å andra sidan, innehåller endast de punkter som svarar mot att reparationerna har lyckats.

Eftersom $F(t)$ inte är en Poissonprocess så är inte heller $N(t) + F(t)$ det. Den naturliga definitionen av ρ är nu

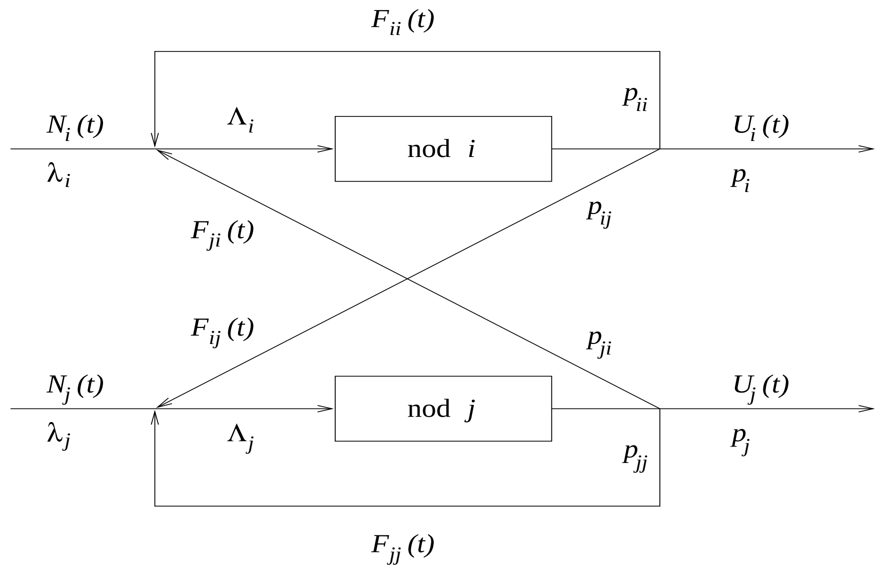
$$\rho = \frac{\Lambda}{\mu} = \frac{\lambda/p}{\mu} = \frac{\lambda}{p\mu},$$

d.v.s. vi får, vilket väl inte är förvånande, samma ρ oavsett om vi betraktar det systemet som ett $M/M/1$-system med parametrarna λ och $p\mu$ eller som ett kösystem med parametrarna Λ och μ. Antalet kunder är givetvis detsamma oavsett vilken tolkning vi ger. Detta innebär att (7.39) gäller trots att inprocessen $N(t) + F(t)$ inte är en Poissonprocess. Annorlunda uttryckt *kan man räkna som om inprocessen vore en Poissonprocess.*

Till sist en liten varning. Så länge som vi betraktar "antal kunder", d.v.s. p_n, ℓ eller ℓ_q spelar det ingen roll om vi uppfattar systemet som ett $M/M/1$-system med parametrarna λ och $p\mu$ eller med parametrarna Λ och μ. Detta gäller dock *inte* om vi betraktar tiden i systemet, eftersom det är skillnad mellan tiden som behövs för ett "varv" i systemet (detta svarar mot parametrarna Λ och μ) och tiden för det antal "varv" det tar tills reparationen har lyckats (vilket svarar mot parametrarna λ och $p\mu$). □

7.7.1 Jacksonnätverk

"Hemligheten" med de återkopplade system som vi ska studera är just sista meningen i exempel 7.13, d.v.s. man räkna som om inprocessen vore en Poisson-



Figur 7.5: Illustration av ett kundströmmar i ett Jacksonnätverk.

process. Naturligtvis ställer detta mycket speciella krav på hur återkopplingen får se ut. Klassen av de system som vi ska betrakta är Jacksonnätverken.

Definition 7.1 *Ett könätverk som har m noder kallas ett* Jacksonnätverk *om följande villkor är uppfyllda:*

- *Varje nod har identiska betjäningsstationer med exponentialfördelade betjäningstider. Nod i har c_i betjäningsstationer med förväntad betjäningstid $1/\mu_i$.*
- *Kunder som kommer till nod i utifrån anländer enligt en Poissonprocess med intensitet λ_i. (Kunder kan även komma till nod i från andra noder i systemet.)*
- *Så snart en kund är betjänad i nod i så går kunden till nod j med sannolikheten p_{ij} för $j = 1, \ldots, m$ eller lämnar systemet med sannolikheten $p_i = 1 - \sum_{j=1}^{m} p_{ij}$. Alla förflyttningar sker ögonblickligen.*
- *Alla ankomstprocesser, betjäningstider och förflyttningar är oberoende av varandra och av systemet i övrigt.*

Jacksonnätverket och dess olika kundströmmar illustreras i figur 7.5.

Ankomstintensiteten till nod i ges av

$$\Lambda_i = \lambda_i + \sum_{j=1}^{m} p_{ji}\Lambda_j, \quad \text{för } i = 1, \ldots, m. \tag{7.40}$$

Låt $X_i(t)$ vara antalet kunder som befinner sig i nod i.

Vi kan nu formulera följande allmänna sats för Jacksonnätverk, som vi ger utan bevis.

Sats 7.16 (Jacksons sats) *För ett Jacksonnätverk med* $\Lambda_i/(c_i\mu_i) < 1$ *för* $i = 1, \ldots, m$ *gäller*

$$\lim_{t\to\infty} P(X_1(t) = n_1, X_2(t) = n_2, \ldots, X_m(t) = n_m) = p_{n_1}^{(1)} p_{n_2}^{(2)} \cdots p_{n_m}^{(m)},$$

där $(p_0^{(i)}, p_1^{(i)}, p_2^{(i)}, \ldots)$ *är sannolikhetsfördelningen för ett* $M/M/c_i$*-system i stationärt tillstånd med ankomstintensitet* Λ_i *och förväntad betjäningstid* $1/\mu_i$*.*

Sats 7.16 säger att i ett Jacksonnätverk i stationärt tillstånd så kan de m noderna behandlas som oberoende $M/M/c_i$-system, trots att ankomstprocesserna inte är Poissonprocesser. Vidare gäller att utprocesserna $U_1(t), \ldots, U_m(t)$ är oberoende Poissonprocesser med intensiteter $p_1\Lambda_1, \ldots, p_m\Lambda_m$.

Vi ska nu i två exempel illustrera hur sats 7.16 kan användas i två enkla fall där vi redan gett resultatet.

Exempel 7.14 (Betjäningssystem i serie) Ett betjäningssystem i serie, se sid. 87, är ett Jacksonsystem med parametrar

$$m = 2 \quad c_1 = c_2 = 1$$

$$\lambda_1 = \lambda \quad \lambda_2 = 0$$

$$p_{12} = 1 \quad p_2 = 1.$$

Från (7.40) får vi, vilket i och för sig inses direkt av konstruktionen,

$$\Lambda_1 = \lambda_1 + 0 = \lambda \quad \text{och} \quad \Lambda_2 = 0 + 1 \cdot \Lambda_1 = \lambda$$

och vi ser att sats 7.16 reduceras till (7.9) på sidan 87. □

Exempel 7.15 (Enkel återkoppling, forts.) Systemet med enkel återkoppling, som vi betraktade i exempel 7.13 och illustrerade i figur 7.4, är ett Jacksonsystem med parametrar

$$m = 1 \quad c_1 = 1$$

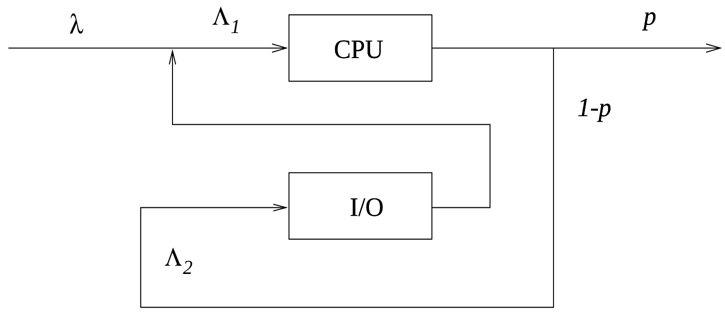
$$\lambda_1 = \lambda$$

$$p_{11} = p \quad p_1 = 1 - p.$$

Från (7.40) får vi

$$\Lambda_1 = \lambda + (1-p)\Lambda_1 \quad \text{eller} \quad \Lambda_1 = \frac{\lambda}{p}$$

och vi ser att sats 7.16 reduceras till (7.39). □



Figur 7.6: Illustration av ett kundströmmar i ett fleranvändarsystem.

Exempel 7.16 (Fleranvändarsystem) Vi ska nu återvända till det fleranvändarsystem som vi inledde avsnitt 7.7 med. Systemet illustreras i figur 7.6. För enkelhets skull förutsätter vi att det endast finns en CPUenhet och en I/O-enhet och att den förväntade "betjäningstiden" i CPUenheten är Exp(μ_1)-fördelad och i I/O-enheten är Exp(μ_2)-fördelad. Systemet är då ett Jacksonnätverk med parametrar

$$m = 2 \quad c_1 = c_2 = 1$$

$$\lambda_1 = \lambda \quad \lambda_2 = 0$$

$$p_{12} = 1 - p \quad p_1 = p \quad p_{21} = 1.$$

Från (7.40) får vi

$$\Lambda_1 = \lambda + \Lambda_2 \quad \text{och} \quad \Lambda_2 = (1-p)\Lambda_1$$

vilket ger

$$\Lambda_1 = \frac{\lambda}{p} \quad \text{och} \quad \Lambda_2 = \frac{(1-p)\lambda}{p}.$$

Av sats 7.4 på sidan 81 följer att

$$\ell = \ell_1 + \ell_2 = \frac{\Lambda_1}{\mu_1 - \Lambda_1} + \frac{\Lambda_2}{\mu_2 - \Lambda_2} = \frac{\lambda}{p\mu_1 - \lambda} + \frac{(1-p)\lambda}{p\mu_2 - (1-p)\lambda},$$

där ℓ_1 är antalet program i CPUn – i kö eller under behandling – och ℓ_2 antalet program i I/O-enheten.

Den förväntade totala körtiden w, inräknat köande, är enl. sats 7.1 på sidan 79

$$w = \frac{\ell}{\lambda} = \frac{1}{p\mu_1 - \lambda} + \frac{1-p}{p\mu_2 - (1-p)\lambda}. \tag{7.41}$$

Vi ska visa att sats 7.1 är tillämpbar i denna situation, vilket inte alls är självklart med det heuristiska bevis som vi gett. Den totala körtiden per CPUomgång – där sats 7.5 på sidan 82 kan tillämpas – är $1/(\mu_1 - \lambda/p)$ och ett program gör i genomsnitt $1/p$ CPUomgångar. På motsvarande sätt är tiden per I/O-omgång $1/(\mu_2 - (1-p)\lambda/p)$. Antalet I/O-omgångar är en mindre än antalet CPUomgångar, d.v.s. i genomsnitt $(1-p)/p$. Sätter vi ihop detta fås

$$w = \frac{1}{p}\frac{1}{\mu_1 - \lambda/p} + \frac{1-p}{p}\frac{1}{\mu_2 - (1-p)\lambda/p} = \frac{\ell}{\lambda} = \frac{1}{p\mu_1 - \lambda} + \frac{1-p}{p\mu_2 - (1-p)\lambda},$$

d.v.s. (7.41) gäller.

Den förväntade effektiva körtiden, d.v.s. tiden då datorn verkligen arbetar med programmet, är

$$\frac{1}{p}\frac{1}{\mu_1} + \frac{1-p}{p}\frac{1}{\mu_2} = \frac{1}{p\mu_1} + \frac{1-p}{p\mu_2}.$$

För att ge någon känsla av vad detta innebär betraktar vi ett datasystem som arbetar 10 timmar/dygn och i genomsnitt mottar 8000 ”jobb” per dygn. Detta ger $\lambda = 8000/36000 = 2/9$ jobb/sekund. Vi antar vidare att CPUtiden i genomsnitt är 4 sekunder med ett I/O ”avbrott” efter (i genomsnitt) var 0.25 sekund. Slutligen antar vi att I/O-tiden i genomsnitt är 0.2 sekunder. Detta ger, naturligtvis i genomsnitt, $4/0.25 = 16$ I/O avbrott/jobb. Med dessa siffror, som vi har vi tagit från [1], är parametrarna

$$\lambda = 2/9 \quad \mu_1 = 1/0.25 = 4 \quad \mu_2 = 1/0.2 = 5 \quad p = 1/16$$

vilket ger

$$\Lambda_1 = \lambda/p = 32/9 \quad \text{och } \Lambda_2 = (1-p)\lambda/p = 10/3.$$

Således får vi

$$w = \frac{1}{p\mu_1 - \lambda} + \frac{1}{\frac{p}{1-p}\mu_2 - \lambda} = \frac{1}{\frac{1}{4} - \frac{2}{9}} + \frac{1}{\frac{1}{3} - \frac{2}{9}} = 36 + 9 = 45 \text{ sekunder}$$

vilket skall jämföras med den förväntade effektiva körtiden

$$\frac{1}{p\mu_1} + \frac{1-p}{p\mu_2} = 4 + \frac{15}{5} = 7 \text{ sekunder.}$$

Det kan tyckas att en total (förväntad) körtid på 45 sekunder är mycket då den effektiva körtiden endast är 7 sekunder. Med dessa siffror gäller dock att $\lambda_1/\mu_1 = 8/9$ och $\lambda_2/\mu_2 = 2/3$ så datorn, och i synnerhet CPUn, är hårt belastad. □

Kapitel 8

Problem

8.1 Markovkedjor i diskret tid

1 En Markovkedja med tillståndsrum $\{0, 1, 2\}$ har övergångsmatrisen

$$\boldsymbol{P} = \begin{pmatrix} 1/3 & & 1/3 \\ 1/2 & 0 & \\ & 3/4 & 0 \end{pmatrix}$$

a) Fyll i de tomma platserna i övergångsmatrisen ovan.

b) Beräkna övergångsmatrisen av andra ordningen, dvs. $\boldsymbol{P}^{(2)}$.

c) Antag att Markovkedjan har startvektor $\boldsymbol{p}^{(0)} = (1/2, 1/2, 0)$. Bestäm de absoluta sannolikheterna $\boldsymbol{p}^{(2)} = (p_0^{(2)}, p_1^{(2)}, p_2^{(2)})$.

2 Låt $(X_n : n \geq 0)$ vara en Markovkedja med tillståndsrum $E = \{r, w, b, y\}$ och övergångsmatris

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0.4 & 0.6 & 0 \\ 0.8 & 0 & 0.2 & 0 \\ 0.2 & 0.3 & 0 & 0.5 \end{pmatrix}$$

a) Beräkna

$$P(X_5 = b,\ X_6 = r,\ X_7 = b,\ X_8 = b \mid X_4 = w)$$

b) Beräkna

$$E(f(X_5)f(X_6) \mid X_4 = y)$$

där f definieras enligt

$$f(x) = \begin{cases} 2 & x = r \\ 4 & x = w \\ 7 & x = b \\ 3 & x = y \end{cases}$$

3 En Markovkedja $\{X_n; n \geq 0\}$ har tillståndsrummet $\{1, 2, 3\}$ och följande övergångsmatris

$$\boldsymbol{P} = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.2 & 0.4 \\ 0.3 & 0.6 & 0.1 \end{pmatrix}$$

a) Beräkna $P(X_5 = 3 \mid X_3 = 1, X_2 = 1)$.

b) Beräkna $P(X_8 = 3, X_7 = 1, X_5 = 2 \mid X_3 = 2, X_2 = 1)$.

4 En Markovkedja $\{X_n; n \geq 0\}$ med tillståndrum $\{1, 2, 3\}$ startar i tillstånd 3, $X_0 = 3$, och har följande övergångsmatris.

$$\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0 & 0.6 & 0.4 \\ 0.8 & 0.1 & 0.1 \end{pmatrix}$$

a) Beräkna $P(X_2 = i), i = 1, 2, 3$.

b) Beräkna förväntad antal steg innan den för första gången når tillstånd 1.

c) Beräkna sannolikheten att den når tillstånd 1 före tillstånd 2.

5 Bestäm alla stationära fördelningar $\boldsymbol{\pi}$ till Markovkedjorna med övergångsmatriser enligt nedan.

$$\text{a)} \quad \boldsymbol{P} = \begin{pmatrix} 0.4 & 0 & 0.6 \\ 0 & 0.5 & 0.5 \\ 0.25 & 0.75 & 0 \end{pmatrix} \qquad \text{b)} \quad \boldsymbol{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 3/4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

6 Klassificera Markovkedjorna med övergångsmatrisen nedan, d.v.s. undersök om de är irreducibla, tillståndens perioder och om tillstånden är genomgångstillstånd eller ej.

$$\text{a)} \quad \boldsymbol{P} = \begin{pmatrix} 0 & 3/4 & 1/4 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/5 & 0 & 0 & 4/5 \\ 0 & 1/3 & 2/3 & 0 \end{pmatrix} \qquad \text{b)} \quad \boldsymbol{P} = \begin{pmatrix} 0.6 & 0 & 0.4 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.8 & 0 & 0.2 \end{pmatrix}$$

7 I två urnor A och B finns det tre röda respektive två gröna bollar. Man drar på måfå en boll ur den urna som innehåller tre bollar och lägger den i den andra urnan. Definiera

$$X_n = \text{antalet gröna bollar i den urna som efter } n \text{ dragningar har två bollar} \qquad n = 1, 2, \ldots$$

$$X_0 = 2$$

samt

$$Y_n = \begin{cases} 1 & \text{om den boll som drogs i } n\text{:te dragningen är röd,} \\ 2 & \text{om den boll som drogs i } n\text{:te dragningen är grön} \end{cases}$$

Y_0 är godtycklig.

a) Visa att $(X_n : n \geq 0)$ är en Markovkedja med tillståndsrum $\{0, 1, 2\}$.

b) Bestäm begynnelsefördelningen och övergångsmatrisen $\boldsymbol{P}$.

c) Är $(Y_n : n \geq 0)$ en Markovkedja?

d) Beräkna $P(X_n = j)$ numeriskt för $j = 0, 1, 2$, $n = 0, 1, 2, 3$.

e) Beräkna den stationära fördelningen till $\boldsymbol{P}$.

8 En Markovkedja har övergångsmatrisen

$$\boldsymbol{P} = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$$

a) Undersök om kedjan är irreducibel.

b) Undersök om Markovkedjan är ergodisk och bestäm i så fall dess gränsfördelning.

9 Som bekant studerade Markov följden av vokaler och konsonanter i ryska diktverk. Liknande studier kan naturligtvis göras av svensk litteratur. I Selma Lagerlöfs En Herrgårdssägen följs en vokal av en konsonant i 78.7 % av fallen och av ordavskiljningstecken (blanksteg, punkt, komma osv) i 20.9 % av fallen. Om vi betraktar ”vokal”, ”konsonant” och ”ordavskiljningssträng” som tre tillstånd i en Markovkedja och antar att de har samma övergångssannolikheter som motsvarande frekvenser i ovannämnda roman erhålls övergångsmatrisen

$$\boldsymbol{P} = \begin{pmatrix} 0.004 & 0.787 & 0.209 \\ 0.483 & 0.271 & 0.246 \\ 0.238 & 0.762 & 0 \end{pmatrix}$$

Naturligtvis kan inte det skrivna språket modelleras exakt som en Markovkedja, men vissa aspekter i språket kan ändå studeras med en sådan modell och antag därför att en ”roman” är skriven som en följd av vokaler, konsonanter och ordavskiljningssträngar $\{X_n; n \geq 1\}$, med ovanstående övergångssannolikheter. Första tecknet, X_1, väljs slumpmässigt enligt den stationära fördelningen.

a) Motivera, utan att utföra några beräkningar, att en entydig, stationär fördelning existerar. Beräkna sedan denna.

b) Beräkna sannolikheten att ett ord börjar med en konsonant.

c) Beräkna sannolikheten att ett ord slutar med en vokal.

d) Beräkna genomsnittlig ordlängd.

e) Beräkna genomsnittligt antal vokaler i ett ord och genomsnittligt antal konsonanter i ett ord.

10 I ett signalsystem sänds 0:or och 1:or ut. En 1:a sänds ut med sannolikhet p och en 0:a med sannolikhet $q = 1 - p$. Signalerna är oberoende av varandra.

Låt X_n vara den n:te utsända signalen, $n = 1, 2, \ldots$. Den första signalen är en 0:a, $X_1 = 0$.

Inför följande tillstånd:

S_1={de två sista signalerna är 0:or},
S_2={de två sista signalerna är 0:a och 1:a},
S_3={de två sista signalerna är 1:a och 0:a} och
S_4={de två sista signalerna är 1:or}.

Låt vidare $\{Y_n; n \geq 2\}$ vara den stokastiska process som anger i vilket av dessa tillstånd signalprocessen hamnat i.

a) Motivera att $\{Y_n; n \geq 2\}$ är en Markovkedja och ange övergångsmatrisen.

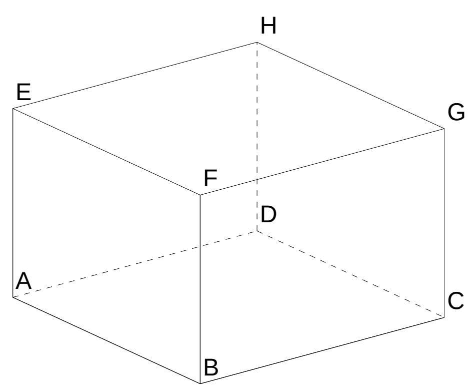
b) Beräkna förväntat antal utsända signaler (inklusive den första 0:an) tills två 1:or i rad har sänts ut.

11 För en irreducibel Markovkedja med ändligt många tillstånd $1, 2, \ldots, n$ gäller att övergångsmatrisen är s.k. dubbelt stokastisk, dvs att

$$\sum_{j=1}^{n} p_{ij} = 1, \quad \text{alla } i,$$
$$\sum_{i=1}^{n} p_{ij} = 1, \quad \text{alla } j.$$

Bestäm den stationära sannolikhetsfördelningen $\{\pi_i\}$ $i = 1, 2, \ldots, n$.

12 En myra rör sig efter kanterna på en kub A B C D E F G H. På varje tidsenhet förflyttar hon sig från ett hörn till något av de tre närmsta belägna hörnen och hon väljer respektive kant med samma sannolikhet 1/3. Befinner hon sig t.ex. i A är sannolikheten för att hon efter en tidsenhet är i B, D eller E resp. 1/3. Hörnen G och H är dock försedda med lim, så att myran fastnar om hon når något av dessa hörn och kan inte gå till något av de andra.



Om myran startar i A, vad är sannolikheten att hon fastnar i hörnet G, resp. H?

13 En Markovkedja med tillstånden {1,2,3} i diskret tid har övergångsmatrisen

$$\begin{pmatrix} 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}$$

a) Kedjan startar i tillstånd 1. Beräkna förväntad tid tills den hamnar i tillstånd 3.

b) Beräkna sannolikheten att kedjan inte någon gång har nått tillstånd 3 efter 4 tidsenheter.

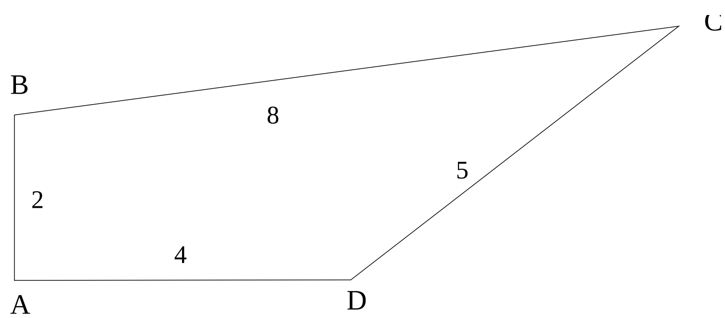
c) Visa att en asymptotisk fördelning existerar samt beräkna denna.

14 Till glädje för alla barn (men kanske till förtvivlan för föräldrar) förekommer samlarbilder i vissa livsmedelspaket. Betrakta följande variant. Varje paket innehåller 2 olika bilder. Dessa kan betraktas som slumpmässigt valda ur en serie om fyra. Efter första paketet har man sålunda 2 olika bilder, efter ytterligare ett paket 2, 3 eller 4 olika bilder. Sätt X_n = antal olika bilder efter n paket. Följden $\{X_n\}$ är en Markovkedja.

a) Beräkna övergångsmatrisen.

b) Om varje paket kostar 15 kr, beräkna väntevärdet av kostnaden tills full serie erhållits.

15 En person skall gå från punkten A till punkten C i figuren.



Vandringen är synnerligen slumpmässig. När han befinner sig i en av punkterna går han till en av de närliggande punkterna med sannolikheter som anges av övergångsmatrisen

$$\boldsymbol{P} = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 0.4 & 0 & 0.6 \\ 0.7 & 0 & 0.3 & 0 \\ 0 & 0 & 1 & 0 \\ 0.8 & 0 & 0.2 & 0 \end{pmatrix} \end{array}$$

Avstånden, i km, mellan punkterna är angivna i figuren. Beräkna den förväntade totalsträcka som personen har gått innan han anländer till C.

Ledning. Jämför resonemanget för beräkning av absorptionstider.

16 A och B deltar i ett tärningsspel, som tillgår på följande sätt: Om tärningen visar 1 eller 2 vinner kastaren, om den visar 3 får han göra ytterligare ett kast. Visar tärningen 4 eller 5 får den andre kasta, och kommer 6 upp förlorar kastaren. Antag att A börjar kasta. Beräkna

a) sannolikheten för att B vinner

b) väntevärdet för det antal spelomgångar som krävs för att avsluta spelet.

17 En Markovkedja med tillstånden $1, 2, 3, \ldots, n$ har övergångsmatrisen

$$\begin{pmatrix} p & q & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & p & q & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & p & q \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \end{pmatrix} \qquad p + q = 1.$$

Beräkna förväntade värdet av den tid det tar innan kedjan absorberas, givet att den startar i tillstånd 1.

18 Antag att en Markovkedja har N tillstånd och låt d_i vara perioden till ett tillstånd i till vilket kedjan kan återvända. Visa att $d_i \leq N$.

19 Låt $\boldsymbol{P}$ vara övergångsmatrisen till en Markovkedja och låt λ vara ett tal sådant att $|\lambda| < 1$.

a) Visa att $\sum_{n=0}^{\infty} \lambda^n \boldsymbol{P}^n$ konvergerar. Med konvergens av en följd av matriser menas komponentvis konvergens.

b) Visa att $\sum_{n=0}^{\infty} \lambda^n \boldsymbol{P}^n = (\boldsymbol{I} - \lambda \boldsymbol{P})^{-1}$.

c) Sätt $\boldsymbol{R} = (\boldsymbol{I} - \lambda \boldsymbol{P})^{-1}$, $(r_{ij}) = \boldsymbol{R}$, med $0 < \lambda < 1$. Visa att
1) $i \rightarrow j$ om och endast om $r_{ij} > 0$
2) $i \leftrightarrow j$ om och endast om $r_{ij} r_{ji} > 0$
3) i är ett genomgångstillstånd om och endast om det existerar ett tillstånd j sådant att $r_{ij} > 0$ och $r_{ji} = 0$.

Ovanstående resultat kan användas för analytisk beräkning av $\boldsymbol{P}^n$ och för automatisk klassificering av tillstånden i en Markovkedja.

20 $X_0, X_1, \ldots$ är en Markovkedja med tillståndsmängd $E = \{1, 2, 3, 4, 5, 6, 7\}$ och övergångsmatris

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

a) Beräkna $\boldsymbol{P}^2$ och $\boldsymbol{P}^3$.

b) Visa att kedjan är irreducibel.

c) Vilken period har kedjan?

21 En maskin tillverkar detaljer av ett visst slag. Detaljerna klassificeras som korrekta, acceptabla eller defekta. Om en tillverkad detalj är korrekt är sannolikheten att nästa tillverkad enhet är korrekt 0.9 och sannolikheten att den är acceptabel 0.1. Om däremot en tillverkad enhet är acceptabel är sannolikheten att nästa är korrekt 0.1, acceptabel 0.7 och defekt 0.2. Om en tillverkad enhet är defekt justeras maskinen varefter nästa tillverkad enhet blir korrekt. Beräkna väntevärdet för antalet tillverkade icke-defekta enheter mellan två justeringar.

22 Klassificera tillstånden i nedanstående övergångsmatriser samt dela upp tillståndsmängden i en union av slutna irreducibla tillståndsmängder och en mängd av genomgångstillstånd.

a) $$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} & \frac{1}{8} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

b) $$\begin{pmatrix} 0.6 & 0 & 0.4 & 0 & 0 \\ 0.2 & 0.6 & 0.2 & 0 & 0 \\ 0.2 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

c) $$\begin{pmatrix} 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 & 0 & 2/3 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 3/4 & 0 \\ 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 \end{pmatrix}$$

23 En Markovkedja $\{X_n; n \geq 0\}$ med tillståndsrum $\{0,1,2\}$ har övergångsmatrisen

$$\begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/4 & 0 & 3/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

Motivera att kedjan är ergodisk och beräkna gränsfördelningen $\boldsymbol{p}$.

24 $X_0, X_1, X_2, \ldots$ är en Markovkedja med tillståndsmängden $\{1, 2, 3, 4\}$. Kedjan har stationära övergångssannolikheter med nedanstående övergångsmatris

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/3 & 1/6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Kedjans initialfördelning är $\left(\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}\right)$.
Redogör för existensen av nedanstående gränsvärden, samt beräkna dem i den

mån de existerar.

$$\lim_{n\to\infty} P(X_n = i), \quad i = 1, 2, 3, 4.$$

25 En Markovkedja $\{X_n; n \geq 0\}$ med tillstånd 0,1 och 2 har följande övergångsmatris.

$$\boldsymbol{P} = \begin{pmatrix} 3/8 & 2/8 & 3/8 \\ 1/3 & 0 & 2/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

a) Visa att en asymptotisk fördelning existerar samt beräkna denna.

b) Sätt $Y_0 = X_0$ och $Y_n = X_n + X_{n-1}$ för $n \geq 1$ och betrakta $\{Y_n; n \geq 0\}$. Undersök om $\{Y_n; n \geq 0\}$ är en Markovkedja. Ange den asymptotiska fördelningen om sådan existerar som är oberoende av startfördelningen, till Y-processen, dvs beräkna om de existerar, $\lim_{n\to\infty} P(Y_n = i)$ för de värden i som Y-processen kan anta.

26 Betrakta en Markovkedja med följande övergångsmatris

$$\begin{pmatrix} \frac{1}{3} & 0 & 0 & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

Klassificera de olika tillstånden och beräkna

$$\lim_{n\to\infty} p_{ik}^{(n)} \quad \text{för alla } i \text{ och } k.$$

27 En Markovkedja med tillstånden 1, 2, 3, 4, 5 har övergångsmatrisen

$$\boldsymbol{P} = \begin{pmatrix} 1/10 & 2/10 & 3/10 & 4/10 & 0 \\ 0 & 1/3 & 0 & 0 & 2/3 \\ 0 & 0 & 1/5 & 4/5 & 0 \\ 0 & 0 & 3/4 & 1/4 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

Kedjan startar i tillståndet 1. Beräkna $\lim\limits_{n\to\infty} P(X_n = k), k = 1, 2, 3, 4, 5.$

28 En Markovkedja med tillståndsrum {0,1,2,3,4} startar i tillstånd 0 och har följande övergångsmatris

$$\boldsymbol{P} = \begin{pmatrix} 1/4 & 1/4 & 1/6 & 1/6 & 1/6 \\ 0 & 2/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/6 & 0 & 5/6 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/4 & 0 & 3/4 \end{pmatrix}$$

Beräkna $\lim\limits_{n\to\infty} P(X_n = k)$, $k = 0, 1, 2, 3, 4.$

29 På en bokhylla står N böcker, av vilka en är röd och resten gröna. Den röda boken står längst till höger i hyllan. Man väljer slumpmässigt en av böckerna på sådant sätt att var och en av de gröna böckerna har sannolikheten $1/(N+1)$ att bli vald, medan den röda väljs med sannolikheten $2/(N+1)$. Den valda boken ställs tillbaka längst till vänster i bokraden. Ovanstående urvals- och återställningsförfarande upprepas gång på gång och de olika urvalen utförs oberoende av varandra. Låt

$$
\begin{aligned}
X_n = {} & \text{den röda bokens position i hyllan (från vänster räknat)} \\
& \text{efter } n\text{:te urvals- och återställningsomgången}, \quad n = 1, 2, \ldots
\end{aligned}
$$

Redogör för existensen av följande gränsvärden samt beräkna dem när de existerar,

$$\lim_{n\to\infty} P(X_n = i), \quad i = 1, 2, \ldots, N.$$

30 En Markovkedja med oändligt många tillstånd, $0, 1, 2, \ldots$ har övergångsmatrisen

$$
\begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & \cdots \\
\frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \cdots \\
\frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & \cdots \\
\frac{4}{5} & 0 & 0 & 0 & \frac{1}{5} & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
$$

Beteckna som vanligt med $p_{jk}^{(n)}$ övergångssannolikheten från tillståndet j till tillståndet k i n steg. Beräkna $\lim_{n\to\infty} p_{jk}^{(n)}$.

31 En partikel utför en slumpvandring på $\boldsymbol{E} = \{0, 1, 2, \ldots\}$, och går till höger med sannolikhet p och till vänster med sannolikhet $q = 1 - p$. I punkten 0 hoppar den dock med sannolikhet 1 ett steg åt höger. Vi antar $p > q$. Låt a_1 vara sannolikheten att man överhuvudtaget når tillstånd 0 vid start i tillstånd 1. Denna sannolikhet är strikt mindre än 1.

a) Beräkna a_1.
Ledning. Om kedjan hoppar ett steg åt höger vid start i 1, vad är sannolikheten att den kommer att återvända till 1 och sedan nå 0?

b) Beräkna sannolikheten a_n att nå tillstånd 0, vid start i tillstånd n.

32 Betrakta en slumpvandring på de icke-negativa heltalen. En partikel som befinner sig i punkten i, hoppar ett steg till höger med sannolikhet p_i och ett steg till vänster med sannolikhet $q_i = 1 - p_i, i = 1, 2, 3, \ldots$, oberoende av tidigare hopp. I punkten 0 ligger den kvar med sannolikhet q_0 istället för att hoppa till vänster. Vi antar $0 < p_i < 1$ alla i.

Sätt

$$
\begin{aligned}
r_0 &= 1 \\
r_i &= \frac{p_0 p_1 \cdots p_{i-1}}{q_1 q_2 \cdots q_i} \; i = 1, 2, \ldots.
\end{aligned}
$$

Visa att ett nödvändigt och tillräckligt villkor för att en asymptotisk fördelning skall existera är att $\sum_{i=0}^{\infty} r_i < \infty$ och ange den asymptotiska fördelningen i detta fall.

33 En person önskar tillverka n st felfria likadana verktyg. Händelsen att ett verktyg blir felaktigt är oberoende av utfallet beträffande övriga verktyg. Sannolikheten för ett felaktigt verktyg är p. Han börjar med att tillverka n st verktyg. Om a av dessa är felaktiga tillverkas en ny sats om a st verktyg. Om b av dessa är felaktiga tillverkas en ny sats om b st osv till dess han fått totalt n st felfria verktyg. Hur många satser, den första inräknad, måste i genomsnitt tillverkas? (OBS! Problemet gäller antalet *satser*, ej antalet tillverkade verktyg.) Detta problem kan bl.a. behandlas med Markovkedjeteknik.

a) Låt tillståndet k bestå i att k felfria verktyg totalt har tillverkats, $k = 0, 1, 2, \ldots, n$. Uppskriv övergångsmatrisen och klassificera kedjans tillstånd.

b) Beteckna med x_{n-k}, $k = 0, 1, 2, \ldots, n-1$, medelantalet övergångar med start i tillståndet k till dess tillståndet n uppnåtts. Sålunda betyder x_n medelantalet satser (med start i tillståndet 0) till dess tillståndet n, dvs n felfria verktyg, uppnåtts. Det förutsättes och behöver ej visas, att alla x_{n-k} är ändliga. Uppskriv ett ekv.system för x_{n-k}.

c) Visa att ekv.systemet i b) satisfieras av

$$x_i = \sum_{r=0}^{\infty} \left[1 - (1 - p^r)^i\right], \quad i = 1, \ldots, n.$$

Speciellt är alltså $x_n = \sum_{r=0}^{\infty} \left[1 - (1 - p^r)^n\right]$.

d) Visa att

$$x_n \sim \frac{\log n}{\log 1/p} \quad \text{för stora } n.$$

e) Beräkna förväntat antal verktyg som har tillverkats tills n stycken blivit felfria.
Denna deluppgift löses lämpligen ej med markovteknik.

8.2 Markovkedjor i kontinuerlig tid

34 En Markovprocess $X(t); t \geq 0$ har tillståndsrum $\{0, 1, 2\}$ och följande intensitetsmatris. Processen startar i 2, $X(0) = 2$.

$$\begin{pmatrix} -8 & & 4 \\ & -5 & 2 \\ 0 & 2 & \end{pmatrix}$$

a) Fyll i de tomma platserna i matrisen.

b) Beräkna sannolikheten att processen ligger kvar i tillstånd 2 under hela tidsintervallet $[0, 1]$.

c) Beräkna sannolikheten att kedjans andra hopp sker till tillstånd 0.
d) Ställ upp det system av differentialekvationer ur vilket man kan lösa ut $p_i(t) = P(X(t) = i), i = 0, 1, 2$.

e) Beräkna förväntad tid tills processen för första gången går in i tillstånd 0.

f) Motivera att en gränsfördelning existerar samt beräkna denna.

35 En Markovprocess har generatorn (intensitetsmatrisen)

$$\boldsymbol{Q} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -3 & 2 \\ 0 & 2 & -2 \end{pmatrix}.$$

a) Bestäm uthoppsmatrisen $\widetilde{\boldsymbol{P}}$.

b) Undersök om kedjan är ergodisk och bestäm i så fall gränsfördelningen $\boldsymbol{p}$.

36 En Markovprocess med tillståndsrum $\{1, 2, 3, 4\}$ har följande intensitetsmatris

$$\begin{pmatrix} -8 & 2 & 2 & 4 \\ 3 & -6 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Tillstånd 3 och 4 är alltså absorberande.

a) Beräkna sannolikheten att kedjan absorberas i tillstånd 3 respektive 4 vid start i tillstånd 1.

b) Beräkna förväntad tid tills absorption vid start i tillstånd 1.

37 $\{X(t); t \geq 0\}$ är en Markovprocess med tillståndsrum $\{0,1,2\}$ och $X(0) = 0$. Intensitetsmatrisen är

$$\boldsymbol{Q} = \begin{pmatrix} -3 & 1 & 2 \\ 4 & -10 & 6 \\ 1 & 4 & -5 \end{pmatrix}$$

a) Beräkna förväntad tid tills processen för tredje gången återkommer till tillstånd 0.

b) Beräkna den förväntade tid processen varit i tillstånd 1 under denna tid.

38 Ett seriesystem med två komponenter brister så snart en av komponenterna brister. Komponenternas livslängder är exponentialfördelade, Exp(1/400) (timmar) oberoende av om den andra komponenten fungerar eller ej. En komponent som brister repareras och reparationstiden för en komponent är Exp(1/20). Man har tillgång till två reparatörer, så båda komponenterna kan repareras samtidigt om båda är sönder. Alla tider är oberoende av varandra.

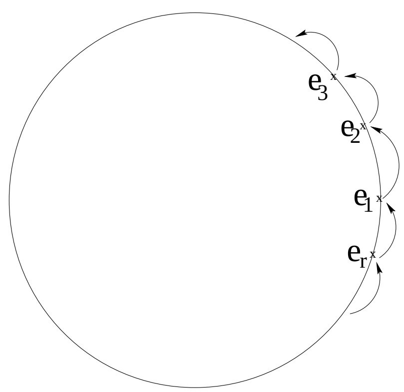
Beräkna asymptotisk tillgänglighet av systemet, d.v.s. sannolikheten att systemet efter lång tid fungerar.

39 På en plats kan vädret klassificeras på tre sätt: uppehåll, duggregn och regnskur. Längden av en regnskur (i timmar) är Exp(5/2) och övergår med sannolikhet 0.4 till ett duggregn. Ett duggregn varar Exp(5/3)-fördelad tid och övergår till uppehåll med sannolikhet 0.8. Uppehåll varar en tid som är Exp(2/5) och övergår sedan till duggregn med sannolikhet 0.75. Alla tider är oberoende av varandra.

a) Beräkna sannolikheten att det vid en (asymptotisk) tidpunkt t är uppehåll.

b) Beräkna förväntade längden av totala nederbördstiden mellan två uppehållsperioder.

40 En partikel rör sig slumpmässigt på punkterna $e_1, e_2, \ldots, e_r$ som ligger på enhetscirkeln. Slumpmekanismen är följande. Om partikeln befinner sig i e_k vid tiden t, så är, oavsett vad som hänt tidigare sannolikheten att den vid tiden $t+h$ flyttad sig ett steg moturs lika med $a_k h + o(h)$ då $h \to 0, k = 1, 2, \ldots, r$. Sannolikheten att den flyttat sig någon annanstans är $o(h)$. Låt $p_k(t)$ beteckna sannolikheten att partikeln befinner sig i e_k vid tiden t givet att den startade i tillstånd e_1 vid tiden $t = 0$.



Motivera varför $\lim_{t\to\infty} p_k(t)$ existerar samt beräkna detta gränsvärde för $k = 1, 2, \ldots, r$.

41 En maskin kan befinna sig i tre tillstånd, {1,2,3}. I tillståndet 1 är maskinen perfekt och genererar en intäkt av 100 000 kr/år. I tillstånd 2 är maskinen delvis felaktig och i "halvfart". Den genererar då en intäkt av 40 000 kr/år. I tillstånd 3 är den helt felaktig och genererar ingen intäkt. I detta tillstånd byts maskinen ut mot en ny. I tillstånd 2 försöker man reparera maskinen och övergångarna mellan tillstånden sker enligt Markovprocess med intensitetsmatris (enhet: per år)

$$\begin{pmatrix} -8 & 7 & 1 \\ 32 & -36 & 4 \\ 100 & 0 & -100 \end{pmatrix}$$

a) Beräkna en maskins förväntade livslängd.

b) Beräkna förväntad intäkt av en maskin under dess livslängd.

42 Låt $\{X(t); t \geq 0\}$ vara en Markovprocess med tillstånd 1,2,3 och intensitetsmatris

$$\boldsymbol{Q} = \begin{pmatrix} -4 & 1 & 3 \\ 3 & -7 & 4 \\ 3 & 0 & -3 \end{pmatrix}$$

Processen startar i tillstånd 3, $X(0) = 3$. Beräkna $P(X(t) = i), i = 1, 2, 3$.

43 Till en biltunnel anländer bilar enligt en Poissonprocess med intensitet 2 bilar per minut. Bilarna kör 60 km/timme. Tunneln är 1 km lång.

Beräkna sannolikheten att det vid en tidpunkt t finns högst 3 bilar i tunneln.

44 Låt $\{X(t); t \geq 0\}$ och $\{Y(t); t \geq 0\}$ vara två oberoende Poissonprocesser med intensiteter λ_1 respektive λ_2. Sätt $Z(t) = X(t) + Y(t)$.

Visa att $\{Z(t); t \geq 0\}$ är en Poissonprocess.

45 Låt $X(t); t \geq 0$ vara en Poissonprocess med intensitet λ.

Beräkna $P(X(2) = 3 \mid X(1) = 2, X(5) = 5)$.

46 Låt $X(t)$, $t \geq 0$, vara en Poissonprocess med intensiteten λ. Sätt $Y(t) = [X(t)/2]$, där $[x]$ betecknar heltalsdelen av x, dvs det heltal n för vilket $n \leq x < n + 1$.

a) Bestäm $P(Y(3) = 1 \mid Y(1) = 1, Y(2) = 1)$.

b) Visa att $Y(t)$, $t \geq 0$, ej är en Markovprocess. (Beräkningen i a) kan härvid vara till hjälp.)

47 Kunder anländer till en betjäningsstation enligt en Poissonprocess med intensiteten λ. Visa att sannolikheten för att ett jämnt antal kunder anländer under tidsintervallet $(s, s + t]$ är $\frac{1}{2}(1 + e^{-2\lambda t})$ och att sannolikheten för att ett udda antal kunder anländer är $\frac{1}{2}(1 - e^{-2\lambda t})$.

48 [Uttunning av Poissonprocess]Alfa-partiklar utsänds från ett radiumpreparat enligt en Poissonprocess med intensiteten λ. Antag att en Geigerräknare registrerar en utsänd partikel med sannolikheten p oberoende av tidigare registreringar.

a) Visa att $\{X(t); t \geq 0\}$ är en Poissonprocess med intensitet $p\lambda$.

b) Låt $U_1, U_2, \ldots$ vara tiderna mellan de tidpunkter då alfa-partiklarna utsänds och T_1 var tidpunkten då första alfa-partiklar räknas. Låt vidare N var antalet utsända partiklar tills en räknas. Ange fördelningarna för U_i, T_1 och N.

c) Beskriv T_1 med hjälp av $U_1, U_2, \ldots$ och N samt visa sats 6.4 b) på sidan 59.

49 I en källa A ”föds” A-partiklar enligt en Poissonprocess med intensiteten λ_A partiklar/tidsenhet, och i en källa B ”föds” B-partiklar enligt en Poissonprocess med intensiteten λ_B partiklar/tidsenhet. A- och B-källorna fungerar oberoende av varandra.

En ”nyfödd” partikel vandrar med hastigheten 1 längdenhet/tidsenhet rakt mot punkten C där den absorberas. Avståndet mellan A och C är 1 längdenhet och avståndet mellan B och C är 2 längdenheter. Man börjar observera systemet vid en tidpunkt t_0. Då finns inga partiklar på vandring.

Beräkna sannolikheten för att den första (efter t_0) partikeln som når C är en B-partikel.

50 I ett fysikaliskt experiment vill man fotografera ett visst elementarpartikelfenomen. Man har följande matematiska modell för experimentsituationen.

De händelser man är intresserad av kallar vi A-händelser. Vidare förekommer vissa störningshändelser, vilka vi kallar S-händelser. A-händelser och S-händelser inträffar enligt Poissonprocesser med intensiteter λ_A respektive λ_S, och dessa två processer är oberoende av varandra. Man startar ett försök vid tid $t = 0$. Sätt

$$T = \text{ tidpunkten då första A-händelsen inträffar.}$$

Försöket ifråga räknas som lyckat om nedanstående betingelser är uppfyllda.

1) $T \geq 1$

2) Ingen S-händelse inträffar på tidsintervallet $[0, T + 2]$.

Bestäm sannolikheten att försöket lyckas.

51 (Den trunkerade Poissonprocessen.) I en ren födelseprocess är födelseintensiteterna

$$\lambda_n = \begin{cases} \lambda & \text{för } 0 \leq n \leq M - 1, \\ 0 & \text{för } n \geq M. \end{cases}.$$

Vidare befinner sig processen i tillståndet 0 vid $t = 0$. Bestäm sannolikheten $P_M(t)$, att processen vid tidpunkten $t > 0$ befinner sig i tillståndet M.

52 I en födelseprocess är födelseintensiteterna

$$\lambda_i = a^i, \quad i = 0, 1, 2, \ldots$$

där a är en positiv konstant. $X(0) = 0$.

a) Beräkna förväntad tid till processen når upp till värdet n.

b) För vilka värden på a är processen reguljär, d.v.s. exploderar inte?

53 Tillväxten i en bakteriekultur kan beskrivas med en födelseprocess $\{X(t); t \geq 0\}$ med födelseintensiteter, λ_i, som är proportionella mot det redan uppnådda värdet, dvs $\lambda_i = \lambda \cdot i$ och med $X(0) = 1$. Visa att

$$P(X(t) = i) = g(t)\,(1 - g(t))^{i-1}, \quad i = 1, 2, 3, \ldots$$

med en lämpligt vald funktion g samt beräkna denna.

54 En population har födelseintensitet λ oberoende av populationens storlek, d.v.s. populationen ökar med en individ med denna intensitet. Populationen kan råka ut för en katastrof, varvid alla individer dör. Denna katastrof inträffar med intensitet μ, oberoende av populationens storlek. Populationen består vid tidpunkt 0 av en individ.

a) Låt N vara populationens storlek ögonblicket före katastrof. Bestäm fördelningen för N.

Ledning: Betrakta inbäddade hoppkedjan.

b) Bestäm fördelningen för den tidpunkt då katastrof inträffar.

c) Använd resultatet i a) och b) för att visa sats 6.4 b) på sidan 59.

55 (Ehrenfests urnmodell, kontinuerlig tid). I två behållare, A och B, finns sammanlagt N partiklar. Var och en av partiklarna överförs, oberoende av varandra och med intensitet λ, till den andra behållaren. Låt $X(t)$ vara antalet partiklar i behållare A vid tid t.

Motivera att $\{X(t); t \geq 0\}$ är en ergodisk Markovprocess och beräkna gränsfördelningen. Om behållare A innehåller en enda partikel, hur lång är den förväntade tiden tills behållare A blir tom?

56 Antalet individer i en population förändras dels genom födslar och dödsfall, dels genom inflyttning till och utflyttning ur populationen. Varje individ i populationen föder en ny individ med intensiteten λ och varje individ dör med intensiteten μ. Om k individer finns i populationen inflyttar en ny individ med intensiteten $\lambda/(k+2)$ och utflyttar en ny individ med intensitet $\nu \cdot k$. (Födelse- och dödsintensiteter för inflyttade individer densamma som för övriga individer.)

Under vilka villkor på λ, μ och ν finns en gränsfördelning? Beräkna denna i detta fall.
Ledning. $-\ln(1-x) = +x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots = \sum_{k=1}^{\infty} \frac{x^k}{k}$ (Taylorutveckling).

57 En population tillväxer dels genom att varje individ i populationen föder en ny individ med intensitet λ och dels genom immigration. Immigrationsintensiteten är också λ, oavsett populationens storlek. Om populationen har i individer, är dödsintensiteten $i \cdot \mu$ för var och en av de i individerna.

a) Visa att en asymptotisk fördelning existerar samt beräkna denna.

b) Antag att $\lambda = 100$ och $\mu = 2$. Beräkna approximativt $P(X(t) \leq 60)$, där $X(t)$ är populationens storlek vid tid t.

58 Betrakta en födelse-döds-process med födelseintensiteter i vilken alla födelseintensiteter är lika och alla dödsintensiteter är lika, d.v.s. $\lambda_i = \lambda$, alla i, och dödsintensiteter $\mu_i = \mu$, alla i. Vi antar dessutom att $\mu > \lambda$. Låt t_1 vara den förväntade tiden tills processen för första gången befinner sig i tillstånd 0 vid start i tillstånd 1. Denna förväntade tid är ändlig.

a) Beräkna t_1.
Ledning: Om processen hoppar till höger från starttillståndet 1, hur stor är förväntade tiden tills den återkommer till 1 och sedan 0?

b) Låt T_i vara tiden tills processen hamnar i tillstånd 0, givet start i tillstånd i. Beräkna $E(T_i)$.

59 En kontorsslav har en ärendekorg som han betar av genom att undan för undan ta det översta ärendet, behandla det och ta nästa. Nya ärenden kommer hela tiden in och ett nytt ärende läggs överst i korgen. Ärenden kommer in enligt en Poissonprocess med intensitet λ och behandlingstiderna för ärendena är oberoende och Exp(μ). Vid ett tillfälle finns n ärenden i korgen och ett under behandling.

a) Antag $\lambda < \mu$. Hur lång är den förväntade tiden innan det understa ärendet börjar behandlas?

Ledning. Jämför uppgift 58 b)

b) Antag $\lambda > \mu$. Då finns en positiv sannolikhet att det understa ärendet aldrig kommer att behandlas. Beräkna denna.

Ledning. Betrakta inbäddade hoppkedjan och använd resultatet i uppgift 31 b)

8.3 Köteori

60 Kunder anländer till ett $M/M/1$-system med ankomstintensiteten 2 och betjäningsintensiteten 3. Detta innebär att kunder anländer till en betjäningsstation enligt en Poissonprocess med intensitet 2 och att betjäningstiderna är exponentialfördelade med intensitet 3 och oberoende av varandra och av ankomstprocessen. Vi betraktar systemet efter lång tid, d.v.s. vi förutsätter att det befinner sig i stationärt tillstånd. Låt S vara en kunds tid i systemet och låt X vara antalet kunder i systemet.

a) Bestäm $w = E[S]$ och $P(S > 2)$.

b) Bestäm $\ell = E[X]$ och $P(X > 6)$.

61 Kunder anländer till en betjäningsstation enligt en Poissonprocess med intensitet λ. Betjäningstiderna är oberoende och Exp(1/2)-fördelade. Man är angelägen om att hålla ned köbildningen och önskar att den förväntade tiden i kön för en kund ej ska överstiga 0.3, när systemet befinner sig i stationärt tillstånd.

Vilket är det högsta värde på λ som är förenligt med ovanstående önskemål?

62 Kunder anländer till en betjäningsstation enligt en Poissonprocess med intensitet λ. Betjäningstiderna är oberoende och Exp(μ)-fördelade. Inkommande potentiella kunder ansluter sig inte säkert till kön, utan tar hänsyn till köns längd. I exempel 7.12 på sidan 98 betraktade vi en enkel modell för köaversion där

$$\lambda_n = \frac{\lambda}{n+1},$$

som förutsätter att det finns oändligt väntrum. Ett sätt att ta hänsyn till väntrummets storlek är att sätta

$$\lambda_n = \begin{cases} \frac{\lambda}{n+1}\left(1 - \frac{n}{c}\right) & \text{om } n = 0, 1, \ldots, c \\ 0 & \text{om } n = c+1, c+2, \ldots. \end{cases}$$

c är således det maximala antalet kunder i systemet.

Beräkna gränsfördelningen $\boldsymbol{p}$ för antalet kunder i systemet samt visa att denna gränsfördelning svarar mot en Bin($c, \tilde{p}$)-fördelning och bestäm $\tilde{p}$.

63 Betrakta samma situation som i problem 62, men antag att en kund ansluter sig till kön med

$$\text{sannolikheten} = \begin{cases} \frac{c-i}{c} & \text{om } i = 0, 1, \ldots, c-1 \\ 0 & \text{om } i \geq c, \end{cases}$$

där c är ett heltal ≥ 1 och i är antalet kunder i systemet.

Beräkna gränsfördelningen $\boldsymbol{p}$ för antalet kunder i systemet.

64 Ett rederi äger en hamn för lossning och lastning av egna fartyg. Hamnen kan vid kaj mottaga endast ett fartyg åt gången och fartygen anländer till hamnen enligt en Poissonprocess med intensitet λ. Lossnings- och lastningsintensiteten är μ fartyg/dag, d.v.s. betjäningstiderna är Exp(μ)-fördelade, och μ ($\mu > \lambda$) kan väljas fritt. Den fasta kostnaden för hamnanläggningen är $a + b\mu$ kr/dag. Avlöningen till stuveriarbetarna jämte andra rörliga kostnader för själva arbetet uppgår till $c\mu$ kr/arbetsdag, d.v.s. per dag som det ligger ett fartyg vid kajen. Omkostnaderna för ett fartyg som ligger i hamn, d.v.s. vid kajen eller på redden, är d kr/dag. Storheterna a, b, c och d är givna konstanter.

Hur skall μ väljas för att den förväntade totala kostnaden ska bli så liten som möjligt?

65 Ett visst kösystem arbetar under följande förutsättningar. Kunderna anländer enligt en Poissonprocess med intensitet 6 kunder/timme. Antalet betjäningsstationer är 3. Dessa arbetar parallellt, oberoende av varandra och av ankomstprocessen. Betjäningstiderna i varje station är exponentialfördelade och i genomsnitt expedieras 3 kunder/timme och arbetande station. Ködisciplinen är ordnad kö, d.v.s. FIFO. Utrymme för godtyckligt lång kö står till förfogande.

Beräkna medeltiden i kön för en kund, innan denne börjar betjänas.

66 Till en betjäningssystem med 2 parallella betjäningsställen anländer kunder enligt en Poissonprocess med intensitet λ. Betjäningstiderna är oberoende och Exp(μ)-fördelade. Låt $w_{q,1}$ vara medelväntetiden i kön för en person som anländer efter lång tid. Man planerar att dela upp betjäningssystemet i två betjäningssystem med en betjäningstation vardera och antar då att kundankomsterna till vardera systemet utgör en Poissonprocess med intensitet $\lambda/2$. Låt $w_{q,2}$ vara medelväntetiden i kön för en person som anländer till ett av dessa system efter lång tid.

Beräkna

$$\frac{w_{q,2}}{w_{q,1}} \quad \text{och visa att} \quad \frac{w_{q,2}}{w_{q,1}} > 2$$

då $\rho = \lambda/(2\mu) < 1$.

67 Betrakta ett kösystem med Poisson-ankomster med intensiteten λ och exponentialfördelade betjäningstider. Definiera

$$\text{förlustkvoten } R = \frac{\text{medelväntetiden i kön}}{\text{medelbetjäningstiden}}.$$

a) Bestäm förlustkvoten R_1 för det fall att systemet består av en betjäningsstation med intensiteten 2μ.

b) Bestäm förlustkvoten R_2 för det fall att systemet består av två parallella betjäningsstationer med intensiteten μ.

c) Jämför förlustkvoten i de två fallen och undersök vilken som är störst, under förutsättningen att μ är samma i a) och b).

68 Vid ett postkontor finns en kassa A för insättning på postgiro och en annan kassa B för inbetalning av postanvisningar. Till kassa A anländer i genomsnitt 30 kunder/timme, till B i genomsnitt 20 kunder/timme (Poisson-ankomst). Betjäningstiderna i både A och B antas vara är exponentialfördelade med väntevärdet 1.5 minuter. Ordnad kö.

a) Hur lång är medelväntetiden i kön framför kassa A resp. B?

b) Hur lång blir medelväntetiden i kön om A och B båda mottar både postanvisningar och postgiro men arbetar oberoende av varandra, så att A och B har varsin kö och så att ingen kund som står i en kö får gå över till den andra? Kunderna väljer dessutom resp. kassa slumpmässigt med sannolikheten 1/2.

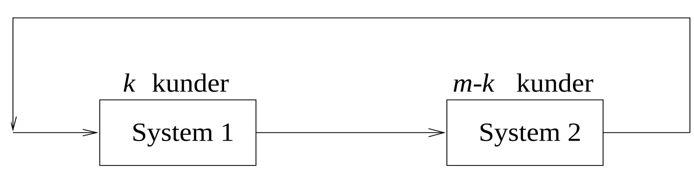
c) Hur lång blir medelväntetiden in kön om A och B samarbetar och har gemensam kö?

69 Ett betjäningsställe ska bestå av ett antal parallella betjäningsstationer, vilka arbetar oberoende av varandra och till vilka kunderna köar i en gemensam kö med ordnad ködiciplin. Kunderna antages komma enligt en en Poissonprocess. Betjäningstiderna är oberoende och Exp(1)-fördelade.

Följande önskemål ställs på systemet. Vid ankomstintensiteten 1 skall den förväntade tiden i kön för en kund ej överstiga 0.2 när systemet befinner sig i stationärt tillstånd.

Hur många betjäningsstationer måste systemet minst innehålla för att uppfylla ovanstående önskemål?

70 Betrakta en ”cyklisk kö” efter lång tid i vilken m kunder cirkulerar mellan två betjäningssystem med Exp($\tilde{\mu}_1$)- resp. Exp($\tilde{\mu}_2$)-fördelade betjäningstider, enl. figur 8.1.



Figur 8.1: Illustration av en cyklisk kö.

Sätt

$$p_k = P(k \text{ kunder i system 1 och } m-k \text{ kunder i system 2}).$$

Bestäm p_k.

71 Antag att en telefonväxel kan samtidigt mottaga obegränsat många samtal och att samtal ankommer till växeln enligt en Poissonprocess med intensiteten λ. Telefonsamtalens längder antas vidare vara Exp(μ)-fördelade och oberoende av såväl varandra som av ankomstprocessen. Ett betjäningssystem av denna typ kallas ett $M/M/\infty$-system, se avnitt 7.3.3 på sidan 92.

Bestäm gränsfördelningen $\boldsymbol{p}$ för antalet pågående samtal.

72 Ett betjäningssystem innehåller 5 parallella betjäningsstationer. Betjäningstiderna är exponentialfördelade med väntevärdet 8 minuter och olika betjäningstider är oberoende.

Systemet "stängs" (d.v.s. inga nya kunder tillåts komma in) i ett läge då det innehåller totalt 12 kunder. Låt T vara tiden det tar att avveckla systemet, d.v.s.

$$T = \text{tiden från stängningsdags tills dess den sista kunden fått betjäning.}$$

Beräkna väntevärdet och standardavvikelsen för T.

73 Till en taxistation anländer taxibilar och kunder enligt oberoende Poissonprocesser med intensiteterna 1 resp. 1.25 per minut. En taxibil väntar vid stationen oavsett hur många bilar som redan finns inne. En kund väntar endast om två eller färre kunder finns före honom.

Bestäm

a) medelantalet taxibilar som väntar på kunder;

b) medelantalet kunder som väntar på taxi;

c) medelantalet kunder som bortfaller under en timme på grund av att det redan finns tre väntande kunder vid stationen.

74 Till en snabbkassa i ett varuhus kommer kunder enligt en Poissonprocess med intensiteten 0.5 kunder/minut. Betjäningen av en kund kan ses som bestående av två delar. Först ska priserna registreras i kassan och sedan ska betalningen ske. I en snabbkassa där kunden får ha högst 5 olika varor kan man mer eller mindre bortse från slumpvariationen i tiden för prisregistreringen, och vi antar att den delen av betjäningen tar (exakt) 20 sekunder oavsett antalet varor. Tiden som betalningen tar kan dock variera högst väsentligt, bl.a. beroende på om kunden betalar kontant, med kort eller med check. Vi antar att denna del av betjäningstiden är exponentialfördelad med väntevärdet 1 minut. Systemet förutsätts befinna sig i stationärt tillstånd. Som vanligt förutsätts alla betjäningstider vara oberoende av varandra och av ankomstprocessen.

Beräkna det förväntade antalet kunder som befinner sig i systemet, dvs som antingen köar i snabbkassan eller betjänas.

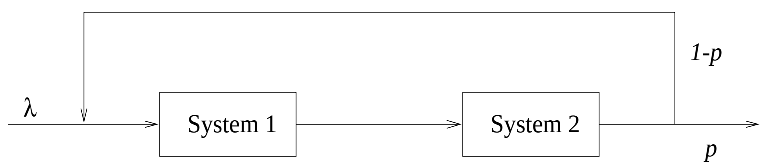
75 Kunder anländer till ett betjäningsställe enligt en Poissonprocess med intensitet λ. En kund har med sannolikhet 1/2 ett ärende och med sannolikhet 1/2 två ärenden som ska expedieras. Varje *ärende* tar en Exp(μ)-fördelad tid att expediera. Alla betjäningstider är oberoende av varandra och av ankomstprocessen.

a) Beräkna väntevärde och varians för betjäningstiden för en *kund*.

b) Bestäm ℓ_q.

c) Bestäm ℓ, w och w_q.

76 Låt oss betrakta en modifiering av den cykliska kön i Problem 70 där kunder kommer "utifrån" enligt en Poissonprocess med intensitet λ. Kunderna kommer först till system 1, där betjäningstiden är Exp($\tilde{\mu}_1$)-fördelad och går sedan till system 2 där betjäningstiden är Exp($\tilde{\mu}_2$)-fördelad. Efter att en kund lämnat system 2 "försvinner" kunden ur hela kösystemet, detta sker med sannolikhet p, eller går till system 1, vilket sker med sannolikhet $1-p$. Samtliga inblandade stokastiska variabler förutsätts oberoende, vilket bl.a. innebär att huruvida en kund försvinner ur systemet efter att passerat system 2 är oberoende av allt annat. Systemet illustreras i figur 8.2.



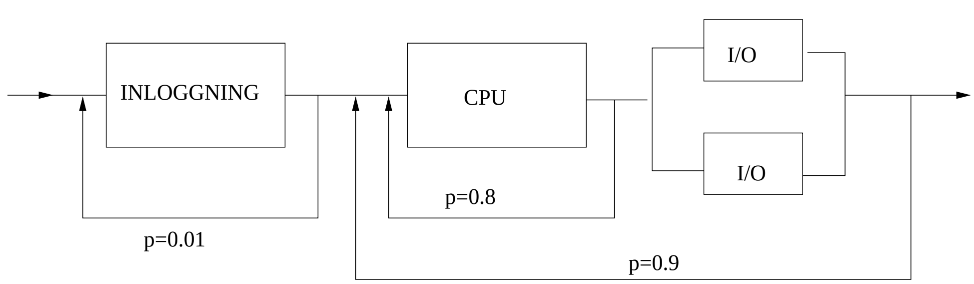
Figur 8.2: Illustration av en "modifierad" cyklisk kö.

Sätt

$$p_{k,n} = P(k \text{ kunder i system 1 och } n \text{ kunder i system 2}).$$

Bestäm $p_{k,n}$.

77 En enkel modell för ett datorsystem är följande (se också figur 8.3).



Figur 8.3: Illustration av en modell för ett datorsystem.

Vi har ett inloggingsförfarande som vid varje försök misslyckas med sannolikhet 0.01 och försöken tar 0.1 tidsenheter i genomsnitt. Man har vidare en CPU-enhet som det tar 0.1 tidsenheter (i genomsnitt) att passera och man behöver sedan med sannolikheten 0.8 en ny CPU-betjäning och med sannolikheten 0.2 vill man sedan använda I/O-enheten där man har två parallella enheter (gemensam kö) där betjäningstiden för vardera eneheten är i genomsnitt 0.2 tidsenheter. Efter I/O-användning behöver man med sannolikheten 0.9 en ny CPU-betjäning och med sannolikheten 0.1 är exekveringen klar. Alla betjäningstider är oberoende och exponentialfördelade. Nya jobb anländer till systemet enligt en poissonprocess med intensiteten 0.15 jobb/tidsenhet. Systemet är i sitt stationära skede.

a) Beräkna förväntad antal jobb i systemet.

b) Beräkna sannolikheten att CPU-enheten är upptagen samt förväntad tid ett jobb tar att exekveras (inklusive inloggning och väntetider).

Kapitel 9

Lösningsförslag

1 a) Eftersom radsummorna i övergångsmatris är 1 har vi att

$$\boldsymbol{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/4 & 3/4 & 0 \end{pmatrix}$$

b) Vi har

$$\boldsymbol{P}^{(2)} = \boldsymbol{P}^2 = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/4 & 3/4 & 0 \end{pmatrix}^2 = \begin{pmatrix} 13/36 & 13/36 & 5/18 \\ 7/24 & 13/24 & 1/6 \\ 11/24 & 1/12 & 11/24 \end{pmatrix}$$

c) Det gäller att

$$\boldsymbol{p}^{(2)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^{(2)} = (1/2, 1/2, 0) \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/4 & 3/4 & 0 \end{pmatrix}^2 = (47/144, 65/144, 2/9)$$

2 a)

$$\begin{aligned} &P(X_5 = b, X_6 = r, X_7 = b, X_8 = b \mid X_4 = w) \\ &\quad = P(X_5 = b \mid X_4 = w) \cdot P(X_6 = r \mid X_4 = w, X_5 = b) \\ &\cdot P(X_7 = b \mid X_4 = w, X_5 = b, X_6 = r) \cdot P(X_8 = b \mid X_4 = w, X_5 = b, X_6 = r, X_7 = b) \\ &\quad = (\text{Markovegenskapen}) \\ &= P(X_5 = b \mid X_4 = w)P(X_6 = r \mid X_5 = b)P(X_7 = b \mid X_6 = r)P(X_8 = b \mid X_7 = b) \\ &\quad = (\text{enligt matrisen}) = 0.6 \cdot 0.8 \cdot 1 \cdot 0.2 = 0.096. \end{aligned}$$

b) Vi får sannolikheten för de olika vägarna vid tidpunkterna 4,5 och 6 enligt tabellen.

Tidpunkt	4	5	6	Sannolikhet
Väg	$y \to$	$r \to$	b	$0.2 \cdot 1 \quad = 0.2$
	$y \to$	$w \to$	w	$0.3 \cdot 0.4 = 0.12$
	$y \to$	$w \to$	b	$0.3 \cdot 0.6 = 0.18$
	$y \to$	$y \to$	r	$0.5 \cdot 0.2 = 0.1$
	$y \to$	$y \to$	w	$0.5 \cdot 0.3 = 0.15$
	$y \to$	$y \to$	y	$0.5 \cdot 0.5 = 0.25$

som ger

$$\begin{aligned}E(f(X_5)f(X_6) \mid X_4 = y) &= f(r)f(b) \cdot 0.2 + f(w)f(w) \cdot 0.12 + f(w)f(b) \cdot 0.18 \\ &+ f(w)f(r) \cdot 0.1 + f(y)f(w) \cdot 0.15 + f(y)f(y) \cdot 0.25 = 14.41\end{aligned}$$

3 a) $P(X_5 = 3 \mid X_3 = 1, X_2 = 1) = P(X_5 = 3 \mid X_3 = 1) = p_{13}^{(2)}$ på grund av markovegenskapen. $p_{13}^{(2)}$ ges av motsvarande element i $P^{(2)} = P^2$. Vi erhåller

$$\boldsymbol{P}^2 = \begin{pmatrix} 0.31 & 0.42 & 0.27 \\ 0.28 & 0.40 & 0.32 \\ 0.33 & 0.27 & 0.4 \end{pmatrix}$$

Den sökta sannolikheten är alltså 0.27.

b) Vi erhåller på grund av markovegenskapen

$$P(X_8 = 3, X_7 = 1, X_5 = 2 \mid X_3 = 2, X_2 = 1) = P(X_8 = 3, X_7 = 1, X_5 = 2 \mid X_3 = 2)$$

Men

$$\begin{aligned}&P(X_8 = 3, X_7 = 1, X_5 = 2 \mid X_3 = 2) \\ &\quad = P(X_5 = 2 \mid X_3 = 2)P(X_7 = 1 \mid X_5 = 2)P(X_8 = 3 \mid X_7 = 1) \\ &\quad = p_{22}^{(2)} p_{21}^{(2)} p_{13} = 0.40 \cdot 0.28 \cdot 0.5 = 0.056\end{aligned}$$

4 a) Fördelningen för X_2 fås genom

$$\begin{aligned}\boldsymbol{p}^{(2)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^{(2)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^2 &= (0\ 0\ 1) \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0 & 0.6 & 0.4 \\ 0.8 & 0.1 & 0.1 \end{pmatrix}^2 \\ &= (0\ 0\ 1) \begin{pmatrix} 0.41 & 0.35 & 0.24 \\ 0.32 & 0.40 & 0.28 \\ 0.48 & 0.31 & 0.21 \end{pmatrix} = (0.48\ 0.31\ 0.21)\end{aligned}$$

b) Gör om 1 till ett absorberande tillstånd och låt t_i vara förväntat antal steg till absorbtion vid start i tillstånd $i,\ i = 2, 3$. Vi får

$$\begin{aligned}t_2 &= 1 + 0.6t_2 + 0.4t_3 \\ t_3 &= 1 + 0.1t_2 + 0.1t_3\end{aligned}$$

vilket har lösningen $t_2 = 65/16$ och $t_3 = 25/16$. Det sökta antalet är alltså 25/16=1.5625 eftersom kedjan startar i tillstånd 3.

c) Gör om tillstånden 1 och 2 till absorberande tillstånd och låt a_i vara sannolikheten att kedjan absorberas i tillstånd i, $i = 1, 2$. Vi får att då att $a_1 = 0.8 + 0.1a_1$, dvs $a_1 = 8/9 = 0.8889$. Detta är då sannolikheten att nå tillstånd 1 före tillstånd 2 i den ursprungliga kedjan. Denna sannolikhet kan också fås som hoppsannolikheten ur tillstånd 3. Sannolikheten för hopp till tillstånd 1 är ju $\frac{p_{31}}{1-p_{33}} = \frac{0.8}{1-0.1} = \frac{8}{9}$.

5 Den stationära fördelningen $\boldsymbol{\pi}$ ges av ekvationssystemet $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}, \pi_1+\pi_2+\pi_3 = 1$.

Vi får

$$\begin{aligned} \pi_1 &= 0.4\pi_1 + 0.25\pi_3 \\ \pi_2 &= 0.5\pi_2 + 0.75\pi_3 \\ \pi_3 &= 0.6\pi_1 + 0.5\pi_2 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

vilket har lösningen $\pi_1 = 1/7, \pi_2 = 18/35, \pi_3 = 12/35$.

b) I detta fall erhåller vi

$$\begin{aligned} \pi_1 &= 0.5\pi_1 + 0.25\pi_2 \\ \pi_2 &= 0.5\pi_1 + 0.75\pi_2 \\ \pi_3 &= \pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Detta system har oändligt många lösningar, $\pi_2 = 2\pi_1, \pi_3 = 1 - 3\pi_1$. Eftersom alla sannolikheter måste ligga mellan 0 och 1 gäller att $0 \le \pi_1 \le 1/3$.

6 a) Man kan göra vandringen $1 \to 2 \to 4 \to 3 \to 1$ varför alla tillstånden kommunicerar, kedjan är irreducibel. Perioden är 2.

b) Kedjan har två irreducibla delkedjor. {1,3} respektive {2,4}. Tillstånden är aperiodiska eftersom man kan går från ett tillstånd tillbaka i ett tidssteg.

7 a) Möjliga värden på X_n är 0, 1, 2. Vi vill visa att $\{X_n; n \ge 0\}$ är en Markovkedja. Vi studerar därför $P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1, X_0 = 2)$. Om vi vet att $X_n = i$ betyder det att $2-i$ st kulor i den ena urnan är röda (och i st är gröna). I den andra urnan är då $2 - i$ st gröna och $3 - (2 - i) = i + 1$ st är röda. Vi vet alltså precis fördelningen av kulor inför den n:te dragningen. Vi får alltså ingen ytterligare relevant information av $X_{n-1}, X_{n-2}, \ldots, X_0$:s värden, dvs $P(X_{n+1} = j \mid X_{n-1} = i_{n-1} = i_{n-1}, \ldots, X_0 = 2) = P(X_{n+1} = j \mid X_n = i)$. Detta betyder att $\{X_n; n \ge 0\}$ är en Markovkedja.

b) Om X_n är 0 finns 2 gröna och en röd i urnan med 3 bollar. Efter $n + 1$:a dragningen har denna urna 2 gröna med sannolikhet 1/3 och en röd och en grön boll med sannolikheten 2/3. Om X_n är 1 finns 1 grön och 2 röda bollar i urnan med 3 bollar och efter $n+1$:a sannolikheten är 1/3 att den har 2 röda bollar och 2/3 att den har 1 röd och 1 grön boll. Om slutligen X_n är 2 finns 3 röda bollar i urnan med 3 bollar och efter $n+1$:a dragningen är sannolikheten är 1 att den har 0 gröna bollar. Övergångsmatrisen ges därför av

$$\boldsymbol{P} = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 2/3 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

Startfördelning $\boldsymbol{p}^{(0)} = (0,0,1)$. c) $\{Y_n; n \geq 0$ är ej en Markovkedja. Vi har till

exempel att $P(Y_4 = 2 \mid Y_3 = 1, Y_2 = 1, Y_1 = 1) = 2/3$, ty genom eftertanke inser man att om händelsen $\{Y_3 = 1, Y_2 = 1, Y_1 = 1\}$ inträffat finns 1 röd och 2 gröna bollar i den urna som innehåller tre bollar. Sannolikheten att man drar en grön boll är således 2/3. Däremot är $P(Y_4 = 2 \mid Y_3 = 1, Y_2 = 2, Y_1 = 1) = 1/3$ eftersom man då efter tredje dragningen har 2 röda och 1 grön boll i urnan med 3 bollar. d) Fördelningen för X_n fås som

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^{(n)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^n = (0,0,1)\begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 2/3 & 0 \\ 1 & 0 & 0 \end{pmatrix}^n$$

Man erhåller $\boldsymbol{p}^{(1)} = (1,0,0), \boldsymbol{p}^{(2)} = (0,2/3,1/3)$ och $\boldsymbol{p}^{(3)} = (5/9,4/9,0)$.

e) Den stationära fördelningen uppfyller ekvationssystemet

$$\begin{aligned} \pi_0 &= \frac{1}{3}\pi_1 + \pi_2 \\ \pi_1 &= \frac{2}{3}\pi_0 + \frac{2}{3}\pi_1 \\ \pi_2 &= \frac{1}{3}\pi_0 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

vilket har lösningen $\pi_0 = 0.3, \pi_1 = 0.6, \pi_2 = 0.1$.

8 a) Man kan göra vandringen $1 \to 3 \to 4 \to 2 \to 1$ varför alla tillstånden kommunicerar, kedjan är irreducibel. Perioden är 1, ty övergången $1 \to 1$ är möjlig.

b) Ändlig, irreducibel och aperiodisk kedja. Alltså är den ergodisk. Den asymptotiska fördelningen ges av den stationära, $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}, \pi_1+\pi_2+\pi_3+\pi_4 = 1$ vilken har lösningen $\pi_1 = 0.5, \pi_2 = 0.125, \pi_3 = 0.25, \pi_4 = 0.125$.

9 a) Vi kallar de olika tillstånden 1,2 och 3. Markovedjan $\{X_n; n \geq 1\}$ är ändlig och irreducibel eftersom man från varje tillstånd kan gå till de övriga. Den stationära fördelningen fås ur $\boldsymbol{\pi} = \boldsymbol{\pi P}$, $\pi_1 + \pi_2 + \pi_3 = 1$ vilket ger ekvationssystemet

$$\begin{aligned}
\pi_1 &= 0.004\pi_1 + 0.483\pi_2 + 0.238\pi_3 \\
\pi_2 &= 0.787\pi_1 + 0.271\pi_2 + 0.762\pi_3 \\
\pi_3 &= 0.209\pi_1 + 0.246\pi_2 \\
\pi_1 &+ \pi_2 + \pi_3 = 1
\end{aligned}$$

vilket har lösningen $\pi_1 = 0.2953, \pi_2 = 0.5160, \pi_3 = 0.1887$.

b) Givet att ett tecken är en ordavskiljningssträng (tillstånd 3) skall vi beräkna sannolikheten att nästa tecken är en konsonant (tillstånd 2). Men denna sannolikhet kan skrivas $P(X_{n+1} = 2 \mid X_n = 3) = p_{32} = 0.762$.

c) Givet att ett tecken är en ordavskiljningssträng skall vi beräkna sannolikheten att föregående tecken är en vokal. Men denna sannolikheten kan skrivas $P(X_n = 1 \mid X_{n+1} = 3)$. Den får vi inte direkt ur övergångsmatrisen men kan beräknas som

$$\begin{aligned}
P(X_n = 1 \mid X_{n+1} = 3) &= \frac{P(\{X_n = 1\} \cap \{X_{n+1} = 3\})}{P(X_{n+1} = 3)} \\
&= \frac{P(X_n = 1)P(X_{n+1} = 3 \mid X_n = 1)}{P(X_{n+1} = 3)} = \frac{\pi_1 p_{13}}{\pi_3} = 0.3272
\end{aligned}$$

eftersom både X_n och X_{n+1} har fördelning $\boldsymbol{\pi}$. Det följer av att startfördelningen är den stationära fördelningen.

Notera att övergångssannolikheten p_{13} inte är sannolikheten att ett ord slutar med vokal, utan sannolikheten att en vokal är slutbokstav i ett ord, d.v.s. andelen vokaler som är slutbokstav i ett ord. Tydligast syns denna skillnad om p_{13} vore 1. Alla vokaler är då slutbokstav i ett ord, men det betyder inte att alla ord slutar med vokal.

d) $1/\pi_3 = 5.30$ är förväntat antal steg mellan två ordavskiljningssträngar, en av dessa inkluderad. Den förväntade ordlängden är därför $5.30 - 1 = 4.30$.

e) Det förväntade antal gånger Markovkedjan är i tillstånd 1 mellan två besök i tillstånd 3 är $\pi_1/\pi_3 = 1.565$. Men detta antal är ju just förväntat antal vokaler i ett ord. På samma sätt är förväntat antal konsonanter $\pi_2/\pi_3 = 2.735$. Notera att summan av dessa är förväntade ordlängden i d).

10 a) Givet $Y_2, Y_3, \ldots, Y_{n-1}$, bestäms Y_n av den sist givna signalen och av den nya, och bestäms sålunda av Y_{n-1} och den nya signalen. Kedjan är därför markovsk. Från S_1 (två sista signalerna 00) kan kedjan gå till tillstånd S_1 (00) eller S_2 (01) med sannolikheter q respektive p. På samma sätt för övriga övergångar. Vi får övergångsmatrisen

$$\boldsymbol{P} = \begin{pmatrix} q & p & 0 & 0 \\ 0 & 0 & q & p \\ q & p & 0 & 0 \\ 0 & 0 & q & p \end{pmatrix}$$

b) Gör om S_4 till ett absorberande tillstånd. Låt t_i vara förväntat antal steg till Markovprocessen hamnar i tillstånd S_4 vid start i S_i. Vi kan betrakta kedjan som att den startar i S_1. Vi får då ekvationssystemet

$$\begin{aligned} t_1 &= 1 + qt_1 + pt_2 \\ t_2 &= 1 + qt_3 \\ t_3 &= 1 + qt_1 + pt_2 \end{aligned}$$

Vi ser att $t_1 = t_3$ och således $t_1 = 1 + qt_1 + p(1 + qt_1)$ vilket ger, efter insättning av, $q = 1 - p$, $t_1 = 1/p + 1/p^2$ och $t_2 = 1/p^2$. Om vi räknar med det första 0-signalen är förväntat antalet utsända signaler tills två ettor i rad erhålls lika med $1 + t_1 = 1 + 1/p + 1/p^2$.

11 Vi gissar $\pi_k = \frac{1}{n}$, $k = 1, 2, \ldots, n$, dvs likformig stationär fördelning.

$$\begin{aligned} \boldsymbol{\pi P} &= \frac{1}{n}(1, 1, \ldots, 1)\boldsymbol{P} = \frac{1}{n}\left(\sum_{i=1}^{n} p_{i1}, \sum_{i=1}^{n} p_{i2}, \ldots, \sum_{i=1}^{n} p_{in}\right) \\ &= \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = \boldsymbol{\pi}, \end{aligned}$$

dvs $\boldsymbol{\pi} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ löser $\sum_{j=1}^{n} \pi_j = 1$, $\boldsymbol{\pi} = \boldsymbol{\pi P}$. Gissningen var alltså korrekt. En ändlig, irreducibel och aperiodisk Markovkedja har en unik stationär fördelning och det är alltså denna vi funnit.

12 Sätt $X_n =$ myrans position efter n steg. Tillståndsrum $=\{A, B, C, D, E, F, G, H\}$. Låt $a_{jk} = P$ (myran fastnar i k |start i tillstånd j). Vi har då att $\{X_n; n \geq 0\}$ är en A-kedja med absorbtionstillstånd $= \{G, H\}$ och genomgångstillstånd $\boldsymbol{\Gamma} = \{A, B, C, D, E, F\}$. Vi har

$$a_{ij} = p_{ij} + \sum_{k \in \Gamma} p_{ik} a_{kj}, \quad i = A, B, C, D, E, F, \quad j = G, H.$$

Vi har enligt texten övergångsmatrisen

$$\begin{pmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Vi får ekvationssystemet

$$\begin{aligned}
a_{AG} &= \frac{1}{3}a_{BG} + \frac{1}{3}a_{DG} + \frac{1}{3}a_{EG} \\
a_{BG} &= \frac{1}{3}a_{AG} + \frac{1}{3}a_{CG} + \frac{1}{3}a_{FG} \\
a_{CG} &= \frac{1}{3} + \frac{1}{3}a_{BG} + \frac{1}{3}a_{DG} \\
a_{DG} &= \frac{1}{3}a_{AG} + \frac{1}{3}a_{CG} \\
a_{EG} &= \frac{1}{3}a_{AG} + \frac{1}{3}a_{FG} \\
a_{FG} &= \frac{1}{3} + \frac{1}{3}a_{BG} + \frac{1}{3}a_{EG}
\end{aligned}$$

Man inser ur ekvationssystements symmetri (eller ur figurens) att $a_{FG} = a_{CG}$ och $a_{EG} = a_{DG}$. Ekvationssystemet ger $a_{AG} = \frac{3}{7}$, vilket ger $a_{AH} = 1 - \frac{3}{7} = \frac{4}{7}$ ty $a_{AG} + a_{AH} = 1$ (förr eller senare måste vi absorberas i antingen G eller H).

13 a) Gör om 3 till ett absorberande tillstånd och låt t_i vara förväntat antal steg till absorbtion vid start i tillstånd i, $i = 1, 2$. Vi får

$$\begin{aligned}
t_1 &= 1 + 0.2t_1 + 0.5t_2 \\
t_2 &= 1 + 0.4t_1 + 0.3t_2
\end{aligned}$$

vilket har lösningen $t_1 = t_2 = 10/3 = 3.333$.

b) Betrakta kedjan som fås då tillstånd 3 gjorts om till ett absorberande tillstånd. Den sökta sannolikheten är då sannolikheten att denna kedja inte är i tillstånd 3 efter fyra tidsenheter, ty om den hamnat där tidigare är den där även i tidpunkt 4. Vi har att

$$\boldsymbol{P}^{(4)} = \boldsymbol{P}^4 = \begin{pmatrix} 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0 & 0 & 1 \end{pmatrix}^4 = \begin{pmatrix} 0.1076 & 0.1325 & 0.7599 \\ 0.1060 & 0.1341 & 0.7599 \\ 0 & 0 & 1 \end{pmatrix}$$

Sannolikheten att efter 4 tidsenheter ha nått tillstånd 3 vid start i tillstånd 1 är således $p_{13}^{(4)} = 0.7599$ och sannolikheten att inte någon gång varit i tillstånd 3 är $1 - 0.7599 = 0.2401$.

c) Kedjan är ändlig, irreducibel (man kan gå från ett tillstånd till alla andra) och aperiodisk (diagonalelement > 0). Alltså är kedjan ergodisk. Den asymptotiska fördelningen fås som den stationära, $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ vilket ger ekvationssystemet

$$\begin{aligned}
\pi_1 &= 0.2\pi_1 + 0.4\pi_2 + 0.3\pi_3 \\
\pi_2 &= 0.5\pi_1 + 0.3\pi_2 + 0.2\pi_3 \\
\pi_3 &= 0.3\pi_1 + 0.3\pi_2 + 0.5\pi_3 \\
\pi_1 &+ \pi_2 + \pi_3 = 1
\end{aligned}$$

vilket har lösningen $\pi_1 = 0.3021,\ \pi_2 = 0.3229$ och $\pi_3 = 0.3750$.

14 a) Låt tillståndet k vara tillståndet att k olika bilder erhållits, $k = 2, 3, 4$. Om 2 olika bilder erhållits är sannolikheten att 2 olika bilder erhållits efter nästa köp lika med sannolikheten att det sist köpta paketet innehåller samma bilder som de man har. Eftersom det finns $\binom{4}{2} = 6$ sätt att välja ut 2 bilder ur 4 givna är denna sannolikhet 1/6. På samma sätt är sannolikheten att efter nästa köp ha 4 bilder 1/6 och således är sannolikheten att ha 3 bilder 2/3.

Om man har 3 olika bilder är sannolikheten att efter nästa köp fortfarande ha 3 bilder 1/2, eftersom de två nya bilderna skall vara två av de redan erhållna och antalet sätt att ta ut två bilder ur dessa tre givna är $\binom{3}{2} = 3$. Sannolikheten att efter nästa köp ha 4 bilder är således också 3/6=1/2. Tillstånd 4 är givetvis absorberande. Vi har övergångsmatrisen

$$\boldsymbol{P} = \begin{pmatrix} 1/6 & 2/3 & 1/6 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}$$

$X_1 = 2$, d.v.s. startfördelningen är $(1, 0, 0)$.

b) Antalet köp, efter första paketet, tills full serie erhållits är antalet hopp som kedjan tar tills den absorberas. Med sedvanliga beteckningar har vi

$$\begin{aligned} t_2 &= 1 + \frac{1}{6}t_2 + \frac{2}{3}t_3 \\ t_3 &= 1 + \frac{1}{2}t_3 \end{aligned}$$

vilket har lösningen $t_2 = 2.8, t_3 = 2$. Tillsammans med det första paketet blir alltså den förväntade kostnaden $3.8 \cdot 15 = 57$ kronor.

15 Vi låter t_i vara förväntad tid tills absorbtion i C givet start i tillstånd $i,\ i = A, B, D$. Om han står i en viss punkt väljer han en av de möjliga vägarna och går i figuren angiven sträcka plus den som återstår då han kommit till nästa punkt. Man inser därför att

$$\begin{aligned} t_A &= 0.4(2 + t_B) + 0.6(4 + t_D) \\ t_B &= 0.7(2 + t_A) + 0.3 \cdot 8 \\ t_D &= 0.8(4 + t_A) + 0.2 \cdot 5 \end{aligned}$$

Detta ekvationssystem är lätt att lösa och man får $t_A = 181/6 = 30\frac{1}{6}$.

16 Identifiera tillståndet enligt
B skall kasta $\Leftrightarrow$ 1, A skall kasta $\Leftrightarrow$ 2, B vinner $\Leftrightarrow$ 3, A vinner $\Leftrightarrow$ 4. Sätt $X_n =$

"det tillstånd spelet befinner sig i efter n kast". Då är $\{X_n\}$ en Markovkedja med tillståndsmängden $E = \{1, 2, 3, 4\}$ och övergångsmatrisen

$$\boldsymbol{P} = \begin{pmatrix} 1/6 & 2/6 & 2/6 & 1/6 \\ 2/6 & 1/6 & 1/6 & 2/6 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

a) Låt a_{i3} vara sannolikheten att kedjan absorberas i tillstånd 3 givet start i tillstånd $i, i = 1, 2$. Vi söker a_{23} och får

$$\begin{aligned} a_{13} &= \frac{2}{6} + \frac{1}{6}a_{13} + \frac{2}{6}a_{23} \\ a_{23} &= \frac{1}{6} + \frac{2}{6}a_{13} + \frac{1}{6}a_{23} \end{aligned}$$

vilket har lösningen $a_{13} = 4/7$ och $a_{23} = 3/7$. Sannolikheten att B vinner är således $3/7$.

b) Sätt t_i lika med förväntad tid tills spelet avslutas vi har

$$\begin{aligned} t_1 &= 1 + \frac{1}{6}t_1 + \frac{2}{6}t_2 \\ t_2 &= 1 + \frac{2}{6}t_1 + \frac{1}{6}t_2 \end{aligned}$$

vilket har lösningen $t_1 = t_2 = 2$.

17 Kedjan hoppar ett steg åt höger vid en övergång. Tiden som kedjan ligger kvar i ett tillstånd är ffg(q) eftersom sannolikheten för uthopp är q. Den förväntade tiden i ett tillstånd är $\frac{1}{q}$. Efter $n-1$ uthopp absorberas kedjan och den förväntade tiden tills detta inträffar är således $\frac{n-1}{q}$.

18 Antag att $d_i > N$. Kedjan kan då inte återvända till tillstånd i annat än i $d_i, 2d_i, 3d_i, \ldots$ steg. Eftersom kedjan kan återvändas till i finns en möjlig väg $i \to i_1 \to i_2 \to \cdots \to i_k \to i$. Här kan vi anta att $i, i_1, i_2, \ldots, i_k$ är olika tillstånd ty i annat fall kan vi stryka tillstånden mellan två lika tillstånd i vägen ovan och fortfarande ha en möjlig väg från i till i. Men det innebär att vägen har högst N tillstånd och kedjan kan återkomma i högst N steg till i. Denna motsägelse visar att $d_i \le N$.

19 a) Vi har att $\sum_{n=0}^{\infty} \lambda^n p_{ij}^{(n)}$ konvergerar eftersom $0 \le p_{ij}^{(n)} \le 1$ och $\sum_{n=0}^{\infty} \lambda^n < \infty$.

b) $(\boldsymbol{I} - \lambda\boldsymbol{P}) \sum_{n=0}^{\infty} \lambda^n \boldsymbol{P}^n = \sum_{n=0}^{\infty} \lambda^n \boldsymbol{P}^n - \sum_{n=0}^{\infty} \lambda^{n+1} \boldsymbol{P}^{n+1} = \sum_{n=0}^{\infty} \lambda^n \boldsymbol{P}^n - \sum_{n=1}^{\infty} \lambda^n \boldsymbol{P}^n = I$, vilket skulle visas.

c) $r_{ij} = \sum_{n=0}^{\infty} \lambda^n p_{ij}^{(n)}$ varför $r_{ij} > 0$ om och endast om $p_{ij}^{(n)} > 0$ för något n, d.v.s. om och endast om $i \to j$.

På samma sätt erhåller vi att $i \leftrightarrow j$ om och endast om $r_{ij} > 0$ och $r_{ji} > 0$, d.v.s. om och endast om $r_{ij} r_{ji} > 0$.

Tillståndet i är ett genomgångstillstånd om och endast om det finns ett tillstånd j tills vilket i leder, men som ej leder tillbaka till i. Men det är detsamma som att det finns ett tillstånd j sådant att $r_{ij} > 0$ och $r_{ji} = 0$.

20 a)

$$\boldsymbol{P}^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 23/48 & 25/48 \\ 0 & 0 & 0 & 0 & 0 & 11/18 & 7/18 \\ 1/3 & 2/3 & 0 & 0 & 0 & 0 & 0 \\ 3/8 & 5/8 & 0 & 0 & 0 & 0 & 0 \\ 7/16 & 9/16 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5/12 & 1/8 & 11/24 & 0 & 0 \\ 0 & 0 & 3/8 & 1/16 & 9/16 & 0 & 0 \end{pmatrix}$$

och

$$\boldsymbol{P}^3 = \begin{pmatrix} 71/192 & 121/192 & 0 & 0 & 0 & 0 & 0 \\ 29/72 & 43/72 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 7/18 & 1/12 & 19/36 & 0 & 0 \\ 0 & 0 & 19/48 & 3/32 & 49/96 & 0 & 0 \\ 0 & 0 & 13/32 & 7/64 & 31/64 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 157/288 & 131/288 \\ 0 & 0 & 0 & 0 & 0 & 37/64 & 27/64 \end{pmatrix}$$

b) Matriserna $\boldsymbol{P}, \boldsymbol{P}^2$ och $\boldsymbol{P}^3$ visar att man från tillstånd 1 kan komma till till var och en av de andra i 1,2 eller 3 steg, och att man från de övriga kan komma till tillstånd 1 i 1, 2 eller 3 steg. Alla tillstånd kommuicerar således och kedjan är irreducibel

c) Perioden är 3. Om $\boldsymbol{D}_0 = \{1, 2\}$, $\boldsymbol{D}_1 = \{3, 4, 5\}$ och $\boldsymbol{D}_2 = \{6, 7\}$ gör kedjan vandringen $\boldsymbol{D}_0 \to \boldsymbol{D}_1 \to \boldsymbol{D}_2 \to \boldsymbol{D}_0 \to \cdots$.

21 Låt tillstånden 1, 2 och 3 representera att en enhet blivit korrekt, acceptabel respektive defekt och låt X_n vara tillståndet för den n:te tillverkade enheten. $\{X_n; n \geq 1\}$ är en Markovkedja med övergångsmatris

$$\begin{pmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.7 & 0.2 \\ 1 & 0 & 0 \end{pmatrix}$$

Den stationära sannolikhetsfördelningen ges av $\boldsymbol{\pi} = \boldsymbol{\pi P}, \pi_1 + \pi_2 + \pi_3 = 1$ vilken har lösningen $\pi_1 = 5/7, \pi_2 = 5/21$ och $\pi_3 = 1/21$. Förväntat antal tillverkade enheter mellan två defekta (en av dessa är inkluderad) är $1/\pi_3 = 21$ och förväntade antalet icke defekta enheter mellan två justeringar är således 20.

Detta förväntade antal kan också beräknas med absorptionsteknik.

22 a) 1 och 5 är absorberande tillstånd, de övriga genomgångstillstånd.

b) tillståndsrummet kan delas upp i två irreducibla slutna deltillståndsmängder, {1,3} och {4,5} och genomgångstillstånd {2}

c) {1,3} är en irreducibel sluten deltillståndsmängd.
{2,7,9} är en irreducibel sluten deltillståndsmängd.
4,5, 8 och 10 är genomgångstillstånd.
6 är ett absorberande tillstånd.

23 Kedjan är ändlig. Den kan göra vandringen $0 \to 1 \to 2 \to 0$ varför tillstånden kommunicerar, kedjan är irreducibel. Kedjan kan också göra vandringen $0 \to 1 \to 0$, varför den kan återkomma till tillstånd 0 i såväl 2 som 3 steg. Kedjan är alltså aperiodisk och vi har också visat att den är ergodisk. Den stationära fördelningen ges av den stationära som vi får ur $\boldsymbol{\pi} = \boldsymbol{\pi P}$. Vi får ekvationssystemet

$$\begin{aligned}
\pi_0 &= \pi_1 \frac{1}{4} + \pi_2 \frac{1}{2} \\
\pi_1 &= \pi_0 \frac{2}{3} + \pi_2 \frac{1}{2} \\
\pi_2 &= \pi_0 \frac{1}{3} + \pi_1 \frac{3}{4} \\
\pi_0 + \pi_1 + \pi_2 &= 1
\end{aligned}$$

Löses ekvationssystemet erhålles $\pi_0 = 3/11, \pi_1 = 4/11, \pi_2 = 4/11$. Denna fördelning är också den asymptotiska $\boldsymbol{p}$.

24 Markovkedjan är en A-kedja med tillstånden 1 och 3 som genomgångstillstånd och 2 och 4 absorbtionstillstånd. Härav följer direkt att

$$\lim_{n\to\infty} P(X_n = i) = 0, \quad i = 1, 3$$

Låt nu a_{i2} vara sannolikheten att absorberas i tillstånd 2 vid start i tillstånd $i, i = 1, 3$. Vi får då att

$$\begin{aligned}
a_{12} &= \frac{1}{3} + \frac{1}{3} a_{12} + \frac{1}{3} a_{32} \\
a_{32} &= \frac{1}{2} a_{12} + \frac{1}{3} a_{32}
\end{aligned}$$

som har lösningen $a_{12} = 0.8$ och $a_{32} = 0.6$. Härav erhåller vi

$$\begin{aligned}
P(X_n = 2) &= \sum_{i=1}^{4} P(X_0 = i) P(X_n = 2 \mid X_0 = i) \\
&\to \frac{1}{10} \cdot 0.8 + \frac{2}{10} \cdot 1 + \frac{3}{10} \cdot 0.6 + 0 = 0.46
\end{aligned}$$

då $n \to \infty$. Av ovanstående följer till slut att $\lim_{n\to\infty} P(X_n = 4) = 0.54$.

25 a) Man kan gå från vilket tillstånd som helst till de andra tillstånden, så kedjan är irredicibel. Den är aperiodisk eftersom t.ex. $p_{11} > 0$. Eftersom kedjan är ändlig är den då ergodisk. Den asymptotiska fördelningen ges av ekvationssystemet

$$\begin{aligned}
\pi_0 &= \frac{3}{8}\pi_0 + \frac{1}{3}\pi_1 + \frac{1}{4}\pi_2 \\
\pi_1 &= \frac{2}{8}\pi_0 + \frac{1}{4}\pi_2 \\
\pi_2 &= \frac{3}{8}\pi_0 + \frac{2}{3}\pi_1 + \frac{1}{2}\pi_2 \\
\pi_0 + \pi_1 + \pi_2 &= 1
\end{aligned}$$

som har lösningen $\pi_0 = \frac{32}{105}, \pi_1 = \frac{1}{5}$ och $\pi_2 = \frac{52}{105}$.

b) Vi har $P(Y_2 = 2 \mid Y_0 = 0, Y_1 = 1) = P(X_1 + X_2 = 2 \mid X_0 = 0, X_0 + X_1 = 1) = P(X_1 + X_2 = 2 \mid X_0 = 0, X_1 = 1) = P(X_2 = 1 \mid X_1 = 1) = 0$ och $P(Y_2 = 2 \mid Y_0 = 1, Y_1 = 1) = P(X_1 + X_2 = 2 \mid X_0 = 1, X_0 + X_1 = 1) = P(X_1 + X_2 = 2 \mid X_0 = 1, X_1 = 0) = P(X_2 = 2 \mid X_1 = 0) = 3/8$. De är inte lika och Y-processen är ej markovsk. Däremot kan ju eventuellt en gränsfördelning existera, och låt oss se på detta. Vi har

$$\begin{aligned}
P(Y_n = i) = P(X_n + X_{n-1} = i) &= \sum_{k=0}^{i} P(X_{n-1} = k) P(X_{n-1} + X_n = i) \mid X_{n-1} = k) \\
&= \sum_{k=0}^{i} P(X_{n-1} = k) P(X_n = i - k \mid X_{n-1} = k) = \sum_{k=0}^{i} P(X_{n-1} = k) p_{k,i-k}
\end{aligned}$$

Men vi har ju att $P(X_{n-1} = k) \to \pi_k$ då $n \to \infty$ och vi erhåller $\lim_{n\to\infty} P(Y_n = i) = \sum_{k=0}^{i} \pi_k p_{k,i-k}$. Med siffror insatta erhåller vi

$$\lim_{n\to\infty} P(Y_n = i) = \begin{cases} 4/35, & i = 0 \\ 1/7, & i = 1 \\ 5/21, & i = 2 \\ 9/35, & i = 3 \\ 26/105, & i = 4 \end{cases}$$

26 Man ser att 2 och 3 är genomgångstillstånd medan $\{1,4,5\}$ är en sluten irreducibel deltillståndsmängd. Det finns alltså endast en irreducibel deltillståndsmängd och denna är aperiodisk. Kedjan är ändlig och således ergodisk. Den asymptotiska fördelningen ges av den stationära $\boldsymbol{\pi} = \boldsymbol{\pi P}, \sum_k \pi_k = 1$ vilken har lösningen $\pi_1 = 3/13, \pi_2 = \pi_3 = 0, \pi_4 = 4/13$ och $\pi_5 = 6/13$.

27 1 är ett genomgångstillstånd, $\{2,5\}$ och $\{3,4\}$ slutna irreducibla deltillståndsmängder. Låt A_k vara händelsen att kedjan hamnar i den irredicibla deltillståndsmängd i vilken k ingår. $\lim\limits_{n\to\infty} P(X_n = k) = P(\lim\limits_{n\to\infty} X_n = k \mid A_k) P(A_k)$.

Vi beräknar först sannolikheterna att kedjan hamnar i de två deltillståndsmängderna. Uthoppssannolikheten att hopp från 1 till {2,5} är $\frac{2/10}{1-1/10} = \frac{2}{9}$ och från 1 till {3,4} således $\frac{7}{9}$. När kedjan hamnat i ett av de två deltillståndsmängderna får den asymptotiska fördelningen som den stationära eftersom deltillståndsmängderna är irredicibla, ändliga och aperiodiska. Då kedjan hamnat i {2,5} är dess övergångsmatris

$$\begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}$$

vars stationära fördelning är $\pi_2 = 3/7, \pi_5 = 4/7$. På samma sätt fås den stationära fördelningen för kedjan om den hamnat i {3,4}, $\pi_3 = 15/31, \pi_4 = 16/31$. Vi erhåller således

$$\lim_{n\to\infty} P(X_n = k) = \begin{cases} 0, & k = 1 \\ \frac{2}{9}\cdot\frac{3}{7} = \frac{2}{21}, & k = 2 \\ \frac{7}{9}\cdot\frac{15}{31} = \frac{35}{93}, & k = 3 \\ \frac{7}{9}\cdot\frac{16}{31} = \frac{112}{279}, & k = 4 \\ \frac{2}{9}\cdot\frac{4}{7} = \frac{8}{63}, & k = 5 \end{cases}$$

28 $\{1,3\}$ och $\{2,4\}$ är slutna irreducibla underkedjor. $\{0\}$ är ett genomgångstillstånd. Gör om dessa underkedjor till absorberande tillstånd. Låt a vara sannolikheten att absorberas i delkedjan $\{1,3\}$ vi start i tillstånd 0. Med absorptionssannolikhetsteknik erhåller vi

$$a = \frac{1}{4} + \frac{1}{6} + \frac{1}{4}a$$

vilket har lösningen $a = 5/9$. Således är sannolikheten att hamna i deltillståndsmängden $\{2,4\}$ lika med 4/9.

Den stationära fördelningen för $\{1,3\}$ och $\{2,4\}$ är $(3/5, 2/5)$ respektive $(3/13, 10/13)$. Härav fås

$$\lim_{n\to\infty} P(X_n = k) = \begin{cases} 0 & k = 0 \\ 5/9\cdot 3/5 = 1/3 & k = 1 \\ 4/9\cdot 3/13 = 4/39 & k = 2 \\ 5/9\cdot 2/5 = 2/9 & k = 3 \\ 4/9\cdot 10/13 = 40/117 & k = 4 \end{cases}$$

29 Om den röda boken står i position i kan den flyttas till position 1 (längst till vänster), till position $i+1$ eller stå kvar i samma position efter urval och återställning. Sannolikheten att den flyttas till position $i+1$ är sannolikheten att en av de böcker som står till höger om den röda dras ($N-i$ stycken) och sannolikheten för detta är $\frac{N-i}{N+1}$. På samma sätt kan man analysera övriga fall. Vi får att $\{X_n; n \geq 1\}$ är en Markovkedja med övergångssannolikheter

$$p_{ij} = \begin{cases} \frac{2}{N+1} & \text{om } i = 1, 2, \ldots, N, j = 1 \\ \frac{i-1}{N+1} & \text{om } i = j > 1 \\ \frac{N-i}{N+1} & \text{om } i = 1, 2, \ldots, N-1, j = i+1 \\ 0 & \text{för övrigt} \end{cases}$$

Kedjan är ändlig, irreducibel och aperiodisk och således ergodisk. Den asymptotiska fördelningen ges av den stationära och vi erhåller ekvationssystemet

$$\pi_1 = \frac{2}{N+1}\pi_1 + \frac{2}{N+1}\pi_2 + \cdots + \frac{2}{N+1}\pi_N = \frac{2}{N+1}\sum_{j=1}^{N}\pi_j = \frac{2}{N+1}$$
$$\pi_i = \pi_{i-1}\frac{N+1-i}{N+1} + \pi_i\frac{i-1}{N+1},\ i = 2, 3, \ldots, N$$

Av den sista ekvationen erhålls successivt

$$\begin{aligned}\pi_i &= \pi_{i-1}\frac{N+1-i}{N+1-(i-1)} = \pi_{i-2}\frac{N+1-(i-1)}{N+1-(i-2)}\frac{N+1-i}{N+1-(i-1)} \\ &= \pi_{i-2}\frac{N+1-i}{N+1-(i-2)} = \cdots = \pi_1\frac{N+1-i}{N+1-1} = \frac{2(N+1-i)}{(N+1)N}.\end{aligned}$$

Det gäller alltså att $\lim\limits_{n\to\infty} P(X_n = i) = \frac{2(N+1-i)}{(N+1)N}$.

30 Markovkedjan är oändlig och irreducibel eftersom den från ett tillstånd kan,via tillstånd 0, kan komma till vilket annat tillstånd som helst. Kedjan är aperiodisk eftersom den kan gå från 0 till 0 i ett steg. Betrakta ekvationssystemet $\boldsymbol{x} = \boldsymbol{xP}$ där $\boldsymbol{P}$ är övergångsmatrisen. Det ger ekvationssystemet

$$\begin{aligned}x_0 &= \frac{1}{2}x_0 + \frac{2}{3}x_1 + \cdots + \frac{i+1}{i+2}x_i + \cdots \\ x_1 &= \frac{1}{2}x_0 \\ x_2 &= \frac{1}{3}x_1 = \frac{1}{2\cdot 3}x_0 \\ x_3 &= \frac{1}{4}x_2 = \frac{1}{4!}x_0 \\ \vdots &= \vdots \\ x_i &= \frac{1}{i+1}x_{i-1} = \frac{1}{(i+1)!}x_0 \\ \vdots &= \vdots\end{aligned}$$

Insättes $x_i = \frac{1}{(i+1)!}x_0$ i högerledet av den första ekvationen erhålles

$$\begin{aligned}x_0\sum_{i=0}^{\infty}\frac{1}{(i+1)!}\frac{i+1}{i+2} &= x_0\sum_{i=0}^{\infty}\frac{1}{(i+1)!}\frac{i+2-1}{i+2} \\ &= x_0\sum_{i=0}^{\infty}\left(\frac{1}{(i+1)!} - \frac{1}{(i+2)!}\right) = x_0\left(1 - \frac{1}{2!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{3!} - \cdots +\right) = x_0\end{aligned}$$

d.v.s. första ekvationen är uppfylld. En stationär fördelning existerar och den ges av $\pi_i = \frac{x_i}{\sum_{j=0}^{\infty} x_j}$. Summan kan beräknas,

$$\sum_{j=0}^{\infty} x_j = x_0 \sum_{j=0}^{\infty} \frac{1}{(j+1)!} = x_0 \sum_{j=1}^{\infty} \frac{1}{j!} = x_0 \Big(\sum_{j=0}^{\infty} \frac{1}{j!} - 1 \Big) = x_0(e-1)$$

Vi erhåller alltså $\pi_i = \frac{1}{(e-1)(i+1)!}$. Kedjan är ergodisk och den asymptotiska fördelningen är lika med den stationära. Oavsett starttillstånd j konvergerar således $p_{jk}^{(n)}$ mot π_k.

31 a) Betrakta kedjan efter första hoppet. Antingen har den gått till tillstånd 0. Sannolikheten för detta är q. I annat fall hoppar den till tillstånd 2. För att komma till tillstånd 0 måste processen först besöka tillstånd 1 igen. På grund av symmetrin är denna sannolikhet densamma som att komma från 1 till 0, d.v.s. a_1. Därefter har vi samma situation som från början, sannolikheten att nu nå 0:an är a_1. På grund av Markovkedjeegenskapen är sannolikheten att nå tillstånd 0, $q + p \cdot a_1 \cdot a_1$, d.v.s. vi har ekvationen $a_1 = q + pa_1^2$. Denna andragradsekvation har lösningarna $a_1 = 1$ och $a_1 = q/p$. Eftersom vi har att $a_1 < 1$ är den sökta sannolikheten $a_1 = q/p$.

b) För att vid start i tillstånd n nå 0, skall kedjan nå värdet till vänster n gånger. Enligt a) är denna sannolikhet varje gång q/p. På grund av att händelsen att nå värdet till vänster är oberoende av vad som tidigare inträffat, är $a_n = a_1^n = (q/p)^n$.

32 Låt X_n vara partikelns läge vid tidpunkt n. Då är $\{X_n; n \geq 0\}$ är en irreducibel aperiodisk Markovkedja, eftersom partikeln från ett tillstånd kan komma till vilket annat tillstånd som helst och eftersom man kan gå från 0 till 0 i ett steg. Enligt sats är kedjan ergodisk om och endast om

$$\boldsymbol{x} = \boldsymbol{x}\boldsymbol{P} \text{ där } \boldsymbol{x} = (x_0, x_1, \dots)$$

har en lösning med $\sum_{i=0}^{\infty} x_i < \infty$.

Övergångsmatrisen är

$$\begin{pmatrix} q_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & 0 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & 0 & p_2 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Ekvationssystemet blir utskrivet

$$\begin{aligned} x_0 &= q_0 x_0 + q_1 x_1 \\ x_1 &= p_0 x_0 + q_2 x_2 \\ &\vdots \\ x_i &= p_{i-1} x_{i-1} + q_{i+1} x_{i+1} \\ &\vdots \end{aligned}$$

Vi skall visa att $x_i = r_i$ uppfyller ekvationssystemet. Vi observerar först att $r_{i+1} = r_i \frac{p_i}{q_{i+1}}$ och $r_{i-1} = r_i \frac{q_i}{p_{i-1}}$. För $i = 1, 2, \ldots$ har vi att högerledet i ekvationerna ovan är $p_{i-1}r_{i-1} + q_{i+1}r_{i+1} = p_{i-1}r_i \frac{q_i}{p_{i-1}} + q_{i+1}r_i \frac{p_i}{q_{i+1}} = r_i(q_i + p_i) = r_i$ vilket är vänsterledet i ekvationerna. För $i = 0$ får vi att $q_0 r_0 + q_1 r_1 = q_0 + q_1 \frac{p_0}{q_1} = q_0 + p_0 = 1 = r_0$, så att $x_i = r_i$ uppfyller ekvationssystemet. Ett nödvändigt och tillräckligt villkor för att kedjan skall vara ergodisk är då att $\sum_{i=0}^{\infty} r_i < \infty$ och den asymptotiska fördelningen ges av den stationära, med $\pi_i = \frac{r_i}{\sum_{k=0}^{\infty} r_k}$.

33 Från tillstånd k kan endast övergång ske till tillstånden $k, k+1, \ldots, n$. I tillstånd k görs $n-k$ nya verktyg och antalet, X, av dessa är som är felfria är $\text{Bin}(p, n-k)$. Sannolikheten att r av dessa är felfria är således $\binom{n-k}{r} q^r p^{n-k-r}$ för $k = 0, 1, \ldots, n,\ r = 0, 1, \ldots, n-k$. Denna sannolikhet är alltså övergångssannolikheten $p_{k,k+r}$.

Alla tillstånd är genomgångstillstånd utom tillstånd n som är absorberande.

b) Vi har att

$$x_i = 1 + \sum_{r=0}^{i-1} p_{n-i,n-i+r} x_{i-r}, \quad i = 1, 2, \ldots, n-1$$
$$x_n = 0$$

c) $p_{n-i,n-i+r} = \binom{i}{r} q^r p^{i-r}$. Insättes detta oc h $x_{i-r} = \sum_{\nu=1}^{\infty} \left[1 - (1-p^\nu)^{i-r}\right]$ i högra ledet i b) erhålles

$$1 + \sum_{r=0}^{i-1} \binom{i}{r} q^r p^{i-r} \sum_{\nu=1}^{\infty} \left[1 - (1-p^\nu)^{i-r}\right] = 1 + \sum_{\nu=1}^{\infty} \sum_{r=0}^{i-1} \binom{i}{r} q^r p^{i-r} \left[1 - (1-p^\nu)^{i-r}\right]$$
$$= 1 + \sum_{\nu=1}^{\infty} \sum_{r=0}^{i-1} \left[\binom{i}{r} q^r p^{i-r} - \binom{i}{r} q^r (p(1-p^\nu))^{i-r}\right]$$
$$= 1 + \sum_{\nu=1}^{\infty} \sum_{r=0}^{i} \left[\binom{i}{r} q^r p^{i-r} - \binom{i}{r} q^r (p(1-p^\nu))^{i-r}\right] = (\text{binomialteoremet})$$
$$= 1 + \sum_{\nu=0}^{\infty} \left(1 - (1-p^\nu)^i\right) = \sum_{\nu=0}^{\infty} \left(1 - (1-p^\nu)^i\right) = x_i$$

vilket skulle visas.

d) Vi gör en integraljämförelse med summan.

$$x_n = \sum_{\nu=0}^{\infty} \left(1 - (1-p^\nu)^n\right) \approx \int_0^{\infty} \left(1 - (1-p^x)^n\right) dx = (\text{gör substitutionen}$$
$$1 - p^x = y) = \frac{1}{\ln(1/p)} \int_0^1 \frac{1-y^n}{1-y} dy = \frac{1}{\ln(1/p)} \int_0^1 (1 + y + y^2 + \cdots + y^{n-1}) dy$$
$$= \frac{1}{\ln(1/p)} (1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}) \approx \frac{1}{\ln(1/p)} \int_1^n \frac{1}{x} dx = \frac{\ln(n)}{\ln(1/p)}$$

e) Personen tillverkar verktyg efter verktyg tills n felfria erhållits, även om han indelar förfarandet i satser. Låt Y_i vara antalet tillverkade verktyg mellan det $i-1$:a och $i:te$ felfria, det sista felfria medräknat. Y_i är då för första gången fördelad, ffg$(1-p)$ och $E(Y_i)=\frac{1}{1-p}$, $i=1,2,\ldots,n$. Totala antalet tillverkade verktyg är $Y=Y_1+Y_2+\cdots+Y_n$ och således är $E(Y)=\frac{n}{1-p}$.

34 a) Eftersom radsummorna i en intensitetsmatris är 0 är den sökta matrisen

$$\begin{pmatrix} -8 & 4 & 4 \\ 3 & -5 & 2 \\ 0 & 2 & -2 \end{pmatrix}$$

b) Tiden, T_2, som processen ligger i tillstånd 2 är Exp(q_2)=Exp(2). Alltså är $P(T_2 \geq 1) = e^{-2\cdot 1} = e^{-2} = 0.1353$.

c) Processens första hopp sker från tillstånd 2 till tillstånd 1 eftersom intesiteten att hoppa till tillstånd 0 är 0. Sannolikheten att nästa hopp sker till tillstånd 0 är $q_{10}/q_1 = 3/5$.

d) Om $\boldsymbol{p}(t) = (p_0(t), p_1(t), p_2(t))$ ges ekvationssystemet av $\boldsymbol{p}'(t) = \boldsymbol{p}(t)\boldsymbol{Q}$ eller utskrivet

$$\begin{aligned} p_0'(t) &= -8p_0(t) + 3p_1(t) \\ p_1'(t) &= 4p_0(t) - 5p_1(t) + 2p_2(t) \\ p_2'(t) &= 4p_0(t) + 2p_1(t) - 2p_2(t) \end{aligned}$$

Det kan vara illustrativt att härleda detta ekvationssystem enligt markovresonemang. I ett tidsintervall av längd h, är sannolikheten att processen flyttar sig från i till j, $q_{ij}h + o(h)$. Sannolikheten att ligga kvar i tillstånd i är $1 - q_i h + o(h)$. Sannolikheten för mer än en förflyttning är $o(h)$. Vi får därför

$$\begin{aligned} p_0(t+h) &= P(X(t+h) = 0) \\ &= P(X(t) = 0)(1 - q_i h) + P(X(t) = 1)q_{10}h + P(X(t) = 2)q_{20}h + o(h) \\ &= p_0(t)(1 - q_i h) + p_1(t)q_{10}h + p_2(t)q_{20}h + o(h) \end{aligned}$$

vilket ger

$$\frac{p_0(t+h) - p_0(t)}{h} = -p_0(t)q_i + p_1(t)q_{10} + p_2(t)q_{20} + \frac{o(h)}{h}$$

Om vi låter h gå mot 0 erhålles

$$p_0'(t) = -p_0(t)q_i + p_1(t)q_{10} + p_2(t)q_{20} = -8p_0(t) + 3p_1(t)$$

med siffror insatta. På samma sätt fås de övriga ekvationerna.

e) Gör om 0 till ett absorberande tillstånd. Låt t_i vara förväntad tid tills processen hamnat i tillstånd 0, givet start i tillstånd i, $i = 1, 2$. Vi får då ekvationssystemet

$$\begin{aligned} t_1 &= \frac{1}{5} + 0.4t_2 \\ t_2 &= \frac{1}{2} + t_1 \end{aligned}$$

vilket har lösningen $t_1 = 2/3$ och $t_2 = 7/6$. Den sökta tiden är alltså 7/6.

f) Kedjan är ändlig och irreducibel, man kan gå från vilket tillstånd som helst till de andra. Alltså är Markovprocessen ergodisk och den asymptotiska fördelningen fås ur den stationära, vilken fås ur ekvationssystemet $\boldsymbol{\pi Q} = \mathbf{0}$, $\pi_0 + \pi_1 + \pi_2 = 1$ eller utskrivet

$$\begin{aligned} -8\pi_0 + 3\pi_1 &= 0 \\ 4\pi_0 - 5\pi_1 + 2\pi_2 &= 0 \\ 4\pi_0 + 2\pi_1 - 2\pi_2 &= 0 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

Löses detta ekvationssystem erhåller man $\pi_0 = 3/25$, $\pi_1 = 8/25$ och $\pi_2 = 14/25$.

35 a)

$$\widetilde{\boldsymbol{P}} = \begin{pmatrix} 0 & 1 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 1 & 0 \end{pmatrix}$$

b) Markovprocessen har ändlig tillståndsrum och är irreducibel. Den är således ergodisk och den asymptotiska fördelningen ges av den stationära, $\boldsymbol{\pi Q} = 0, \pi_1 + \pi_2 + \pi_3 = 1$ dvs

$$\begin{aligned} -\pi_1 + \pi_2 &= 0 \\ \pi_1 - 3\pi_2 + 2\pi_3 &= 0 \\ 2\pi_2 - 2\pi_3 &= 0 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

vilket har lösningen $\pi_1 = \pi_2 = \pi_3 = 1/3$.

36 Om vi sätter a_{ij} lika med sannolikheten att absorberas i tillstånd j givet start i tillstånd i, $i = 1, 2$ $j = 3, 4$, erhåller vi

$$\begin{aligned} a_{13} &= \frac{2}{8} + \frac{2}{8}a_{23} \\ a_{23} &= \frac{2}{6} + \frac{3}{6}a_{13} \end{aligned}$$

Löses detta ekvationssystem erhåller man $a_{13} = 8/21$ och $a_{23} = 11/21$. Sannolikheten att absorberas i tillstånd 4 vid start i tillstånd 1 är $a_{14} = 1 - a_{13} = 13/21$.

b) Om vi låter t_i vara förväntad tid tills absorbtion vid start i tillstånd i, erhåller vi

$$\begin{aligned} t_1 &= \frac{1}{8} + \frac{1}{4}t_2 \\ t_2 &= \frac{1}{6} + \frac{1}{2}t_1 \end{aligned}$$

Sätt in den andra ekvationen i den första och man erhåller lätt $t_1 = 4/21$.

37 Vi beräknar först den stationära fördelningen. Den fås ur ekvationssystemet $\boldsymbol{\pi Q} = \mathbf{0}$,

$$\begin{aligned} -3\pi_0 + 4\pi_1 + \pi_2 &= 0 \\ \pi_0 - 10\pi_1 + 4\pi_2 &= 0 \\ 2\pi_0 + 6\pi_1 - 5\pi_2 &= 0 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

Ekvationssystemet har lösningen $\pi_0 = \pi_2 = 0.4$ och $\pi_1 = 0.2$. Den förväntade tiden mellan två ingångar i tillstånd 0 är $\frac{1}{q_0\pi_0} = \frac{1}{3\cdot 0.4}$. Det betyder att den förväntade tiden tills processen för tredje gången besöker tillstånd 0 är $\frac{3}{3\cdot 0.4} = 2.5$.

b) Den förväntade tiden som processen ligger i tillstånd 1 mellan två inträden i tillstånd 0 är $\frac{\pi_1}{q_0\pi_0} = \frac{0.2}{3\cdot 0.4}$. Den förväntade tiden som processen ligger i tillstånd 1 under den tid tills kedjan för tredje gången återkommer till tillstånd 0 är därför 3 gånger detta värde, d.v.s. 0.5.

38 Inför tillstånden

$$\begin{aligned} S_0 &= \text{ båda komponenterna hela} \\ S_1 &= \text{ en komponent hel, en komponent felar} \\ S_2 &= \text{ båda komponenterna felar} \end{aligned}$$

Låt $X(t)$ vara processens tillstånd vid tid t. $\{X(t); t \geq 0\}$ är en Markov-process med intensitetsmatris

$$\boldsymbol{Q} = \begin{pmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -\lambda - \mu & \lambda \\ 0 & 2\mu & -2\mu \end{pmatrix}$$

med $\lambda = 1/400$ och $\mu = 1/20$. Om till exempel processen är i tillstånd S_0, kan två händelser inträffa, komponent 1 kan brista och komponent två kan brista. Båda inträffar med intensitet λ och processen hoppar till tillstånd S_1. Intensiteten för detta hopp är då $\lambda + \lambda = 2\lambda$. Den asymptotiska tillgängligheten är sannolikheten att processen är i tillstånd S_0. Eftersom processen är ändlig och irreducibel är den asymptotiska sannolikheten lika med den stationära som fås ur ekvationssystemet $\boldsymbol{\pi Q} = \mathbf{0}$,

$$\begin{aligned} -2\lambda\pi_0 + \mu\pi_1 &= 0 \\ 2\lambda\pi_0 - (\lambda+\mu)\pi_1 + 2\mu\pi_2 &= 0 \\ \lambda\pi_1 - 2\mu\pi_2 &= 0 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

vilket har lösningen $\pi_0 = \mu^2/(\lambda+\mu)^2$, $\pi_1 = 2\lambda\mu/(\lambda+\mu)^2$ och $\pi_0 = \lambda^2/(\lambda+\mu)^2$. Med siffror insatta fås $\pi_0 = 400/441$, $\pi_1 = 40/441$ och $\pi_2 = 1/441$. Den asymptotiska tillgängligheten är alltså $400/441 \approx 90.7\%$.

39 a) Inför tillstånden S_1=uppehåll, S_2=regnskur och S_3=duggregn och låt $X(t)$ vara tillståndet vid tid t. $\{X(t); t \geq 0\}$ är en Markovprocess med intensitetsmatris

$$\boldsymbol{Q} = \begin{pmatrix} -0.4 & 0.25\cdot 0.4 & 0.75\cdot 0.4 \\ 0.6\cdot 2.5 & -2.5 & 0.4\cdot 2.5 \\ 0.8\cdot 5/3 & 0.2\cdot 5/3 & -5/3 \end{pmatrix}$$

Betrakta t.ex. tillstånd S_1. Om intensiteterna för övergång till tillstånden S_2 och S_3 är q_{12} respektive q_{13} och $q_1 = q_{12}+q_{13}$ gäller ju att $1/q_1 = 2.5$, förväntad tid i tillstånd S_1. Hoppsannolikheten till tillstånd S_3 är $\tilde{p}_{13} = \frac{q_{13}}{q_1}$ som alltså är 0.75, vilket ger $q_{13} = q_1\tilde{p}_{13} = \frac{0.75}{2.5} = 0.3$. På samma sätt för övriga övergångar, $q_{ij} = \tilde{p}_{ij}q_j$. Markovprocessen är ändlig och irreducibel och därför existerar en asymptotisk fördelning. Denna ges av den stationära fördelningen som fås ur ekvationssystemet

$$\begin{aligned} -0.4\pi_1 + 1.5\pi_2 + 4\pi_3/3 &= 0 \\ 0.1\pi_1 - 2.5\pi_2 + \pi_3/3 &= 0 \\ 0.3\pi_1 + \pi_2 - 5\pi_3/3 &= 0 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

vilket har lösningen $\pi_1 = 230/297 \approx 0.7744$, $\pi_2 = 16/297 \approx 0.0539$ och $\pi_3 = 17/99 \approx 0.1717$. Sannolikheten för uppehåll är alltså 0.7744.

b) Den förväntade duggregnstiden plus förväntade regnskurstiden mellan två uppehållsperioder är $\frac{\pi_2}{q_1\pi_1} + \frac{\pi_3}{q_1\pi_1} = \frac{\pi_2+\pi_3}{q_1\pi_1} = \frac{67/297}{0.4\cdot 230/297} = 67/92 \approx 0.728$

40 Låt $X(t)$ vara partikeln läge vid tidpunkt t. Från beskrivningen av partikelns vandring inser man att $\{X(t); t \geq 0\}$ är en Markovprocess med övergångsintensiteter

$$q_{ij} = \begin{cases} a_i \text{ om } j = i+1,\ i = 1,2,\ldots,r-1 \\ a_r \text{ om } j = 1,\ i = r \\ -a_i \text{ om } j = i,\ i = 1,2,\ldots,r \\ 0 \text{ för övrigt} \end{cases}$$

dvs intensitetsmatrisen är

$$\boldsymbol{Q} = \begin{pmatrix} -a_1 & a_1 & 0 & \dots & 0 & 0 \\ 0 & -a_2 & a_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -a_{r-1} & a_{r-1} \\ a_r & 0 & 0 & \dots & 0 & -a_r \end{pmatrix}$$

Markovprocessen är ändlig, alla tillstånd kommunicerar med varandra, d.v.s. processen är irreducubel och därmed ergodisk. Den asymptotiska fördelningen är lika med den stationära vilken fås ur $\boldsymbol{\pi Q} = \mathbf{0}$. Detta ger ekvationssystemet

$$\begin{aligned} -a_1\pi_1 + a_r\pi_r &= 0 \\ a_1\pi_1 - a_2\pi_2 &= 0 \\ \vdots \quad &= \vdots \\ a_{i-1}\pi_{i-1} - a_i\pi_i &= 0 \\ \vdots \quad &= \vdots \\ a_{r-1}\pi_{r-1} - a_r\pi_r &= 0 \end{aligned}$$

vilket ger $a_r\pi_r = a_{r-1}\pi_{r-1} = a_{r-2}\pi_{r-2} = \cdots = a_1\pi_1$ d.v.s. $\pi_i = \frac{a_1\pi_1}{a_i}$. Från $\sum_{k=1}^r \pi_k = 1$ erhåller vi så $\pi_1 = \frac{1/a_1}{\sum_{k=1}^r 1/a_k}$ och således $\pi_i = \frac{1/a_i}{\sum_{k=1}^r 1/a_k}$. Det ger alltså den asymptotiska fördelningen. Man kan notera att $1/a_i$ är den förväntade tid som processen ligger i tillstånd i innan den hoppar ur, så resultatet säger att den asymptotiska sannolikheten att vara i ett visst tillstånd är proportionell mot den förväntade tiden i detta tillstånd, vilket kanske inte är så förvånande.

41 a) Låt $X(t)$ vara maskinens tillstånd vid tid t och sätt t_i lika med förväntad tid tills processen når tillstånd 3 givet start i tillstånd i, $i = 1, 2$. Vi får

$$\begin{aligned} t_1 &= \frac{1}{8} + \frac{7}{8}t_2 \\ t_2 &= \frac{1}{36} + \frac{32}{36}t_1 \end{aligned}$$

vilket har lösningen $t_1 = 43/64$ och $t_2 = 5/8$. En maskins förväntade livslängd är alltså $43/64$ år.

b) Låt nu u_i, $i = 1, 2$, vara den förväntade tiden i tillstånd i under maskinens livslängd. Den kan beräknas ur den stationära fördelningen som $\frac{\pi_i}{100\pi_3}$ (förväntad tid i tillstånd i mellan två inträden i tillstånd 3, byte av maskin), men också genom följande resonemang. Efter start är maskinen i tillstånd 1 i förväntad tid $1/8$. Antingen går den sedan till tillstånd 3 utan att återkomma till tillstånd 1. I annat fall går den till tillstånd 2 sedan till tillstånd 1 och i så fall återstår igen den förväntade tiden u_1 innan maskinen hamnar i tillstånd 3. Vi har alltså

$$u_1 = \frac{1}{8} + \frac{7}{8} \cdot \frac{8}{9} u_1$$

vilket ger $u_1 = 9/16$. Härav följer att $u_2 = t_1 - u_1 = 7/64$. Den förväntade intäkten är alltså $100000\frac{9}{16} + 40000\frac{7}{64} = 60625$ kronor.

42 Sätt $p_i(t) = P(X(t) = i)$. Vi erhåller följande diffekvationssystem.

$$\begin{aligned} p_1'(t) &= -4p_1(t) + 3p_2(t) + 3p_3(t) \\ p_2'(t) &= p_1(t) - 7p_2(t) \\ p_3'(t) &= 3p_1(t) + 4p_2(t) - 3p_3(t) \end{aligned}$$

Bilda nu Laplacetransformen av vänster och högerled. vi får då

$$\begin{aligned} sp_1^*(s) &= -4p_1^*(s) + 3p_2^*(s) + 3p_3^*(s) \\ sp_2^*(s) &= p_1^*(s) - 7p_2^*(s) \\ sp_3^*(s) - 1 &= 3p_1^*(s) + 4p_2^*(s) - 3p_3^*(s) \end{aligned}$$

eftersom Laplacetransformen av $p_i'(t)$ är $sp_i^*(s) - p_i(0)$. Ekvationssystemet har lösningen

$$\begin{aligned} p_1^*(s) &= \frac{3}{s(s+7)} = \frac{3/7}{s} - \frac{3/7}{s+7} \\ p_2^*(s) &= \frac{3}{s(s+7)^2} = \frac{3/49}{s} - \frac{3/7}{(s+7)^2} - \frac{3/49}{s+7} \\ p_3^*(s) &= \frac{s^2+11s+25}{s(s+7)^2} = \frac{25/49}{s} + \frac{3/7}{(s+7)^2} + \frac{23/49}{s+7} \end{aligned}$$

Genom invertering av Laplacetransformen har vi till slut

$$\begin{aligned} p_1(t) &= \frac{3}{7} - \frac{3}{7}e^{-7t} \\ p_2(t) &= \frac{3}{49} - \frac{3}{7}te^{-7t} - \frac{3}{49}e^{-7t} \\ p_3(t) &= \frac{25}{49} + \frac{3}{7}te^{-7t} + \frac{24}{49}e^{-7t} \end{aligned}$$

43 Eftersom tunneln är 1 km lång och bilarna kör med en hastighet av 60 km/timme har de bilar som finns i tunneln vid en tidpunkt t, anlänt under tidsintervallet $[t-1, t]$. Det tar ju en minut att köra genom tunneln.

Antalet bilar,X, som anländer i ett tidsintervall av längd 1 är Po$(2 \cdot 1)$. Vi får att $P(X \leq 3) = 0.8571$. Använd tabell eller beräkna detta via sannolikhetsfunktionen.

44 Vi ser att

1) $Z(0) = X(0) + Y(0) = 0$
2) $Z(t) - Z(s) = (X(t) - X(s)) + (Y(t) - Y(s))$ är Po$(\lambda_1(t-s) + \lambda_2(t-s))=$ Po$((\lambda_1 + \lambda_2)(t-s))$

eftersom summan av två oberoende Poissonfördelade variabler är Poissonfördelad.

3) $Z(t_2) - Z(s_2) = (X(t_2) - X(s_2)) + (Y(t_2) - Y(s_2))$ och $Z(t_1) - Z(s_1) = (X(t_1) - X(s_1)) + (Y(t_1) - Y(s_1))$ är oberoende om $s_1 < t_1 < s_2 < t_2$ eftersom X- och Y-processerna är oberoende och var och en har oberoende inkrement.

Sålunda är $\{Z(t); t \geq 0\}$ en Poissonprocess.

45 Vi har

$$P(X(2) = 3 \mid X(1) = 2, X(5) = 5) = \frac{P(X(2) = 3, X(1) = 2, X(5) = 5)}{P(X(1) = 2, X(5) = 5)} =$$

$$\frac{P(X(1) = 2, X(2) - X(1) = 1, X(5) - X(2) = 2)}{P(X(1) = 2, X(5) - X(1) = 3)} = \text{(oberoende inkrement)} =$$

$$\frac{P(X(1) = 2)P(X(2) - X(1) = 1)P(X(5) - X(2) = 2)}{P(X(1) = 2)P(X(5) - X(1) = 3)} =$$

$$\frac{P(X(2) - X(1) = 1)P(X(5) - X(2) = 2)}{P(X(5) - X(1) = 3)} =$$

$$\left(e^{-\lambda}\lambda e^{-3\lambda}\frac{(3\lambda)^2}{2!}\right)\Big/\left(e^{-4\lambda}\frac{(4\lambda)^3}{3!}\right) = 27/64$$

46 Inför beteckningen $Y_t = Y(t)$ och $X_t = X(t)$. $Y_k = 1$ innebär att X_k är 2 eller 3. Y_k är icke-avtagande i k och vi har

$$\begin{aligned}\{Y_1 = 1 \cap Y_2 = 1 \cap Y_3 = 1\} &= \{(X_1 \text{ är 2 eller 3}) \cap (X_2 \text{ är 2 eller 3}) \cap (X_3 \text{ är 2 eller 3})\} \\ &= \{(X_1 = 2, X_2 = 2, X_3 = 2) \cup (X_1 = 2, X_2 = 2, X_3 = 3) \\ &\qquad \cup (X_1 = 2, X_2 = 3, X_3 = 3) \cup (X_1 = 3, X_2 = 3, X_3 = 3)\}\end{aligned}$$

Vi skall ha sannolikheten för ovanstående händelse som är unionen av fyra disjunkta händelser. Om vi noterar att Poissonprocessen har oberoende inkrement får vi sannolikheten för den första delhändelsen till

$$P(X_1 = 2, X_2 = 2, X_3 = 2) = P(X_1 = 2, X_2 - X_1 = 0, X_3 - X_2 = 0)$$

$$= P(X_1 = 2)P(X_2 - X_1 = 0)P(X_3 - X_2 = 0) = e^{-\lambda}\frac{\lambda^2}{2!} \cdot e^{-\lambda} \cdot e^{-\lambda} = e^{-3\lambda}\frac{\lambda^2}{2}$$

eftersom X_1, $X_2 - X_1$ och $X_3 - X_2$ är oberoende och alla $\text{Po}(1 \cdot \lambda)$.

På liknande sätt erhålles sannolikheterna för de övriga delhändelserna och summan blir den totala sannolikheten. Man erhåller så småningom

$$P(Y_1 = 1, Y_2 = 1, Y_2 = 1) = e^{-3\lambda}(\lambda^2/2 + 7\lambda^3/6)$$

På samma sätt fås att $P(Y_2 = 1 \cap Y_1 = 1) = e^{-2\lambda}(\lambda^2/2 + 4\lambda^3/6)$ och således

$$\begin{aligned}P(Y_3 = 1 \mid Y_1 = 1, Y_2 = 1) &= \frac{P(Y_3 = 1, Y_1 = 1, Y_2 = 1)}{Y_1 = 1, Y_2 = 1)} \\ &= \frac{e^{-3\lambda}(\lambda^2/2 + 7\lambda^3/6)}{e^{-2\lambda}(\lambda^2/2 + 4\lambda^3/6)} = e^{-\lambda}\frac{1/2 + 7\lambda/6}{1/2 + 4\lambda/6}\end{aligned}$$

b) På samma sätt som i a) finner man

$$P(Y_3 = 1 \mid Y_1 = 0, Y_2 = 1) = e^{-\lambda}\frac{3/2 + 13\lambda/6}{3/2 + 2\lambda/6}$$

som ej är lika med sannolikheten i a), och således är $Y(t)$ ej en Markovprocess.

47 Antalet kunder som kommer i tidsintervallet (av längd t) är Po(λt). Därför är

$$P(X \text{ jämnt}) = \sum_{j \text{ jämnt}} e^{-\lambda t}\frac{(\lambda t)^j}{j!} = e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^{2k}}{(2k)!}$$

Men $\sum_{k=0}^{\infty}\frac{(\lambda t)^{2k}}{(2k)!} = \frac{1}{2}\sum_{k=0}^{\infty}(\frac{(\lambda t)^k}{k!} + \frac{(-\lambda t)^k}{k!}) = \frac{1}{2}(e^{\lambda t} + e^{-\lambda t})$ eftersom de termer i mellanledets summa som har udda k tar ut varandra. Insättes det sista uttrycket i ekvationen ovan fås $P(X \text{ jämnt}) = \frac{1}{2}(1 + e^{-2\lambda t})$. Av $P(X \text{ udda}) = 1 - P(X \text{ jämnt})$ fås den andra likheten.

48 a) Låt $Z = Y(s+t) - Y(s)$ vara antalet utsända signaler i tidsintervallet $(s, s+t)$. Givet att $Z = n$ är $X(t+s) - X(s)$, antalet registrerade signaler i intervallet, Bin(n, p) eftersom varje signal registreras med sannolikhet p oberoende av de andra. Från satsen om total sannolikhet fås

$$\begin{aligned}
P(X(t+s) - X(s) = k) &= \sum_{n=k}^{\infty} P(X(t+s) - X(s) = k \mid Z = n)P(Z = n) \\
&= \sum_{n=k}^{\infty}\binom{n}{k}p^k(1-p)^{n-k}e^{-\lambda t}\frac{(\lambda t)^n}{n!} = \sum_{n=k}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= \frac{p^k e^{-\lambda t}}{k!}e^{-\lambda t}\sum_{j=0}^{\infty}\frac{(\lambda t)^{j+k}}{j!}(1-p)^j = \frac{e^{-\lambda t}(p\lambda t)^k}{k!}e^{-\lambda t}\sum_{j=0}^{\infty}\frac{((1-p)\lambda t)^j}{j!} \\
&= \frac{e^{-\lambda t + (1-p)\lambda t}(p\lambda t)^k}{k!} = \frac{e^{-p\lambda t}(p\lambda t)^k}{k!}
\end{aligned}$$

dvs $X(s+t) - X(s)$ är Po($p\lambda t$). X-processen har oberoende inkrement eftersom Y-processen har det, och vidare är $X(0) = 0$, varför $\{X(t); t \geq 0\}$ är en Poissonprocess.

Ett annat och enklare sätt att härleda detta är att betrakta intensiteter. X-processen hoppar ett steg uppåt vid varje övergång och man inser att den är markovsk eftersom Y-processen är det. Sannolikheten att en partikel registeras i ett intervall $(t, t+h)$ är sannolikheten att en partikel sänds ut och att denna räknas. Men sannolikheten för detta är $(\lambda h + o(h))p$, d.v.s. $p\lambda h + o(h)$. Sannolikheten för två eller fler förflyttningar är $o(h)$ (minst två måste ha sänts ut) och sannolikheten för att ingen förflyttning skett således $1 - p\lambda h + o(h)$. Men detta definierar ju en födelseprocess med konstant intensitet $p\lambda$, d.v.s. en Poissonprocess.

b) N är ffg(p), T_1 är Exp($p\lambda$) och U_i är Exp(λ) för alla i.

c) Man har att $T_1 = U_1+U_2+\cdots+U_N$. U-variablerna är oberoende av varandra och av N. Om man från b) noterar vilka fördelningar variablerna har, inses att sats 6.4 b) är visad.

49 Låt T_A vara tiden då den första A-partikeln sänds ut (efter tid t_0) och T_B då B-partikeln sänds ut. T_A och T_B är Exp(λ_A) respektive Exp(λ_B). Tiderna när de når punkten C är T_A+1 resektive T_B+2. Vi söker $P(T_B+2 < T_A+1)$. Vi får eftersom T_A och T_B är oberoende

$$\begin{aligned} P(T_B+2<T_A+1) &= P(T_B+1<T_A) = \iint_{t+1<u} f_{T_A}(u) f_{T_B}(t)\,du\,dt \\ &= \int_0^\infty \lambda_B e^{-\lambda_B t}dt \int_{t+1}^\infty \lambda_A e^{-\lambda_A u}du = \int_0^\infty \lambda_B e^{-\lambda_B t} e^{-\lambda_A(t+1)}dt \\ &= e^{-\lambda_A}\int_0^\infty \lambda_B e^{-(\lambda_A+\lambda_B)t}dt = e^{-\lambda_A}\frac{\lambda_B}{\lambda_A+\lambda_B} \end{aligned}$$

50 Sätt U tiden till första S-händelse. Då söker vi $P(T \geq 1 \cap U > T+2)$. De stokastiska variablerna T och U är oberoende och exponentialfördelade med intensiteter λ_A respektive λ_S. Vi får därför att

$$\begin{aligned} \iint_{t\geq 1\cap u>t+2} f_T(t) f_U(u)\,du\,dt &= \int_1^\infty \lambda_A e^{-\lambda_A t}dt \int_{t+2}^\infty \lambda_S e^{-\lambda_S u}du \\ &= \int_1^\infty \lambda_A e^{-\lambda_A t} e^{-\lambda_S(t+2)}dt = e^{-2\lambda_S}\lambda_A \int_1^\infty e^{-(\lambda_A+\lambda_S)t}dt \\ &= e^{-2\lambda_S}\lambda_A\Big[-\frac{e^{-(\lambda_A+\lambda_S)t}}{\lambda_A+\lambda_S}\Big]_1^\infty = \frac{e^{\lambda_A-3\lambda_S}\lambda_A}{\lambda_A+\lambda_S} \end{aligned}$$

51 Den trunkerade Poissonprocessen uppför sig som en otrunkerad Poissonprocess tills den når upp till värdet M. Eftersom processerna bara kan gå ett steg uppåt och den trunkerade processen absorberas i M är sannolikheten att den trunkerade processen nått upp till värden M, densamma som att den otrunkerade nått upp till värden M eller högre. Denna sannolikhet är $P(X(t) \geq M)$, där $X(t)$ är Po(λt), d.v.s. $P_M(t) = 1 - \sum_{i=0}^{M-1} e^{-\lambda t}\frac{(\lambda t)^i}{i!}$.

52 a) Låt t_i vara förväntad tid i tillstånd i, $i = 0, 1, \ldots$. Eftersom födelseprocessen enbart går ett steg uppåt vid varje övergång är förväntad tid till tillstånd n nås, $t_0 + t_1 + \cdots + t_{n-1} = 1 + 1/a + 1/a^2 + \cdots + 1/a^{n-1}$. Detta är en ändlig geometrisk serie och summan är $\frac{1-(1/a)^n}{1-1/a}$ om $a \neq 1$. Om $a = 1$ är summan n.

b) Processen är reguljär om och endast om summan $1 + 1/a + 1/a^2 + \cdots = \sum_{i=0}^\infty 1/a^i$ divergerar, vilket den gör om och endast om $a \leq 1$.

53 Födelseprocessen har intensitetsmatris

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -2\lambda & 2\lambda & 0 & \dots \\ 0 & 0 & -3\lambda & 3\lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Sätt $p_i(t) = P(X(t) = i)$. Diffekvationssystemet $\boldsymbol{p}'(t) = \boldsymbol{Q}\boldsymbol{p}(t)$ ger

$$\begin{aligned} p_1'(t) &= -\lambda p_1(t) \\ p_2'(t) &= -2\lambda p_2(t) + \lambda p_1(t) \\ &\vdots \\ p_i'(t) &= -i\lambda p_i(t) + (i-1)\lambda p_{i-1}(t) \\ &\vdots \end{aligned}$$

Vi ansätter $p_i(t) = g(t)(1 - g(t))^{i-1},\ i = 1, 2, \dots$. Då fås

$$p_i'(t) = g'(t)(1-g(t))^{i-1} - (i-1)g(t)g'(t)(1-g(t))^{i-2} = g'(t)(1-g(t))^{i-2}(1-ig(t))$$

Sätter vi in detta uttryck i diffekvationssystemet får vi

$$\begin{aligned} g'(t)(1-g(t))^{i-2}(1-ig(t)) &= -i\lambda g(t)(1-g(t))^{i-1} + (i-1)\lambda g(t)(1-g(t))^{i-2} \\ &= \lambda g(t)(1-g(t))^{i-2}(ig(t) - 1) \end{aligned}$$

d.v.s. $g'(t) = -\lambda g(t)$, vilket har lösningen $g(t) = Ce^{-\lambda t}$. Eftersom $g(0) = P(X(0) = 1) = 1$ är $C = 1$ varför $g(t) = e^{-\lambda t}$.

54 a) Låt $X(t)$ vara populationens storlek vid tid t. Då är $\{X(t); t \geq 0\}$ är en Markovprocess som vid varje hopp hoppar antingen ett steg uppåt eller till tillståndet 0. Sannolikheten att ett hopp sker till 0 är $\frac{\mu}{\lambda+\mu}$, och hoppen sker oberoende av varandra och av den tid som processen varit i de olika tillstånden. Antalet hopp som görs tills händelsen hopp till 0:an är då ffg$(\frac{\mu}{\lambda+\mu})$. Men man inser med litet eftertanke att antalet hopp som görs tills hopp sker till 0:an, är lika med populationens storlek N innan katastrofen. N är alltså ffg$(\frac{\mu}{\lambda+\mu})$.

b) Eftersom katastrof inträffar med intensitet μ, oavsett populationens storlek, är tidpunkten T, då katastrof sker Exp(μ).

c) Tiden T tills katastrof kan också skrivas $T = T_1 + T_2 + \cdots + T_N$, där T_i är tiden i tillstånd i. T_i är Exp$(\lambda + \mu)$ och är oberoende, $i = 1, 2, \dots$. De är också oberoende av N. Sätter vi nu $p = \frac{\mu}{\lambda+\mu}$ och $\nu = \lambda + \mu$ så har vi visat, att om $T_1, T_2, \dots \in$ Exp(ν) och $N \in$ ffg(p) är oberoende, så är $T = T_1 + T_2 + \cdots + T_N \in$Exp$(\mu)$=Exp$(p\nu)$, vilket ger sats 6.4.

55 Man inser att $\{X(t); t \geq 0\}$ är en födelse-dödsprocess med ändligt tillståndsrum. Den är alltså ergodisk. Födelse- och dödsintensiteter ges av

$$\lambda_i = (N-i)\lambda, \qquad i = 0, 1, 2 \ldots, N-1$$
$$\mu_i = i\lambda, \qquad i = 1, 2, \ldots, N$$

Detta kan inses av att i tillstånd i var och en partiklarna efter exponentialfördelad tid förflyttas, varvid tillståndet ökar eller minskar med 1. Om vi sätter

$$\rho_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} = \frac{N(N-1)\cdots(N-i+1)}{i!} = \binom{N}{i}, \ \rho_0 = 1$$

ges den asymptotiska fördelningen av $\pi_i = \frac{\rho_i}{\sum_{j=0}^{N}\rho_j}$. Men enligt binomialteoremet är $\sum_{j=0}^{N}\rho_j = \sum_{j=0}^{N}\binom{N}{j} = (1+1)^N = 2^N$. Detta ger $\pi_i = \binom{N}{i}(\frac{1}{2})^N$ dvs den asymptotiska fördelningen är $\mathrm{Bin}(N, \frac{1}{2})$. Om behållare A inte innehåller någon partikel är den förväntade tiden tills den blir tom densamma som förväntade tiden som processen ligger i tillstånden $1, 2, \ldots, N$ mellan två inträden i det tomma tillståndet. Denna förväntade tid blir $\frac{\sum_{i=1}^{N}\pi_i}{q_0\pi_0} = \frac{1-\pi_0}{N\lambda\pi_0} = \frac{2^N-1}{N\lambda}$.

56 Vi har en födelse-dödsprocess med

$$\lambda_i = i\lambda + \frac{\lambda}{i+2} = \lambda\frac{(i+1)^2}{i+2} \qquad i = 0, 1, \ldots$$
$$\mu_i = i(\mu+\nu) \qquad i = 1, 2, \ldots$$

Sätt $\rho_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} = (\frac{\lambda}{\mu+\nu})^i\frac{(i!)^2/(i+1)!}{i!} = \frac{\rho^i}{i+1}$ där $\rho = \frac{\lambda}{\mu+\nu}$.

En stationär fördelning existerar om $\sum_{i=0}^{\infty}\rho_i = \sum_{i=0}^{\infty}\frac{\rho^i}{i+1} < \infty$. Serien konvergerar om $0 \leq \rho < 1$. För $0 < \rho < 1$ är $\sum_{i=0}^{\infty}\rho_i = \sum_{i=0}^{\infty}\frac{\rho^i}{i+1} = \rho^{-1}\sum_{i=0}^{\infty}\frac{\rho^{i+1}}{i+1} = \rho^{-1}\ln(\frac{1}{1-\rho})$. För dessa ρ är $\sum_{i=0}^{\infty}\frac{1}{\lambda_i\rho_i} = \sum_{i=0}^{\infty}\frac{(i+2)}{\lambda(i+1)\rho^i}$, vilken divergerar. En asymptotisk fördelning existerar alltså och den ges av den stationära; $\pi_i = \frac{\rho_i}{\sum_{j=0}^{\infty}\rho_j} = \frac{\rho^i/(i+1)}{\ln(1/(1-\rho))}$

Om $\rho \geq 1$ divergerar serien $\sum_{i=0}^{\infty}\frac{\rho^i}{i+1}$ och ingen asymptotisk fördelning existerar.

57 Låt $X(t)$ vara populationens storlek vid tid t. $\{X(t); t \geq 0\}$ är en födelsedödsprocess med födelse- och dödsintensiteter

$$\lambda_i = i\cdot\lambda + \lambda = (i+1)\lambda, \ i = 0, 1, 2, \ldots$$
$$\mu_i = i\cdot i\cdot\mu = i^2\mu, \ i = 1, 2, \ldots$$

Detta kan inses av att processen vid ett tillstånd i väntar på att någon av individerna föder en ny individ, eller någon av dem dör eller att en individ immigrerar. Så snart någon av dessa händelser inträffar hoppar processen antingen upp ett steg eller ner ett steg. Vi sätter nu

$$\rho_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} = \frac{i!(\lambda)^i}{(i!)^2(\mu)^i} = \frac{(\lambda/\mu)^i}{i!}.$$

Vi har att $\sum_{i=0}^{\infty}\rho_i = \sum_{i=0}^{\infty}\frac{(\lambda/\mu)^i}{i!} = e^{\lambda/\mu} < \infty$, varför en asymptotisk fördelning existerar. Denna ges av den stationära, $\boldsymbol{\pi}$ med $\pi_i = \rho_i/\sum_{i=0}^{\infty}\rho_i = e^{-\lambda/\mu}\frac{(\lambda/\mu)^i}{i!}$, d.v.s. den asymptotiska fördelningen är Po(λ/μ).

b) I detta fall är $\lambda/\mu = 50$. Resultatet i a) ger då att för stora t är $X(t)$ så gott som Po(50) som är approximativt N$(50,\sqrt{50})$. Vi får således att $P(X(t) \geq 60) \approx \Phi(\frac{60-50}{\sqrt{50}}) = \Phi(1.4142) = 0.9214$.

58 a) Om processen hoppar till höger från tillstånd 1, skall den för att överhuvudtaget komma till tillstånd 0, först komma tillbaka till tillstånd 1, för att därefter någon gång komma till 0:an. På grund av symmetrin är förväntad tid att nå tillstånd 1 från tillstånd 2, densamma som att den förväntade tiden att från tillstånd 1 nå tillstånd 0, d.v.s. t_1. Sedvanligt absorbtionstidsresonemang ger oss därför relationen $t_1 = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}(t_1 + t_1)$ vilket har lösningen $t_1 = \frac{1}{\mu-\lambda}$.

b) Om man startar i tillstånd i skall man först nå tillstånd $i-1$, sedan tillstånd $i-2$ o.s.v. De förväntade tiderna för att från tillstånd i komma till $i-1$, är enligt a) $\frac{1}{\mu-\lambda}$, varför $E(T_i) = \frac{i}{\mu-\lambda}$.

59 a) Låt $X(t)$ vara antalet ärenden som är före det understa i korgen vid tidpunkt t (d.v.s. de ovanför i korgen plus det under behandling). $\{X(t); t \geq 0\}$ är en födelse-dödsprocess med födelseintensiteter λ och dödsintensiteter μ. Processen startar i tillsånd n och enligt lösningen på uppgift 58 är en förväntade tiden tills sista ärendet tas $\frac{n}{\mu-\lambda}$.

b) Betrakta inbäddade hoppkedjan. Hopp sker ett steg uppåt med sannolikhet $p = \frac{\lambda}{\lambda+\mu}$ och ett steg neråt med sannolikhet $q = \frac{\mu}{\lambda+\mu}$. Sannolikheten att tillstånd 0 överhuvudtaget nås, då processen startar i tillstånd n, är enligt lösningen av uppgift 31, $(q/p)^n = (\mu/\lambda)^n$. Sannolikheten att det sista ärendet inte kommer att behandlas är därför $1 - (\mu/\lambda)^n$.

60 a) Vi har $\rho = \lambda/\mu = 2/3$. Enligt sats 7.5 på sidan 82 fås

$$w = E[S] = \frac{1}{\mu(1-\rho)} = \frac{1}{3\cdot(1-\frac{2}{3})} = 1$$

och

$$P(S > 2) = 1 - G(2) = e^{-2} \approx 0.135.$$

b) Enligt sats 7.4 på sidan 81 fås

$$\ell = E[X] = \frac{\rho}{1-\rho} = \frac{\frac{2}{3}}{1-\frac{2}{3}} = 2$$

och, se sats 7.3 på sidan 81,

$$P(X > 6) = \sum_{n=7}^{\infty} p_n = \sum_{n=7}^{\infty} \rho^n (1-\rho) = (1-\rho)\rho^7 \sum_{n=7}^{\infty} \rho^{n-7}$$

$$= (1-\rho)\rho^7 \sum_{k=0}^{\infty} \rho^k = (1-\rho)\rho^7 \frac{1}{1-\rho} = \rho^7 = (2/3)^7 \approx 0.0585.$$

61 Vi har ett $M/M/1$-system med $\mu = 1/2$ och $\rho = \lambda/\mu = 2\lambda$. Sats 7.5 på sidan 82 ger

$$w_q = \frac{\rho}{\mu(1-\rho)} = \frac{2\rho}{(1-\rho)} = \frac{4\lambda}{(1-2\lambda)}.$$

Villkoret $w_q \leq 0.3$ ger

$$\frac{4\lambda}{(1-2\lambda)} \leq 0.3 \Rightarrow 4\lambda \leq 0.3 - 0.6\lambda \Rightarrow \lambda \leq \frac{0.3}{4.6} = 0.065.$$

62 Sätt

$$\rho_0 = 1 \quad \text{och} \quad \rho_n = \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n} = \frac{(\lambda/\mu)^n}{n!} \prod_{k=0}^{n-1} \frac{c-k}{c} = \left(\frac{\lambda}{\mu c}\right)^n \binom{c}{n}$$

Detta ger

$$\sum_{i=0}^{c} \rho_i = \left(1 + \frac{\lambda}{\mu c}\right)^c$$

och det följer av sats (6.12) på sidan 68 att

$$p_n = \binom{c}{n} \left(\frac{\lambda}{\mu c}\right)^n \left(1 + \frac{\lambda}{\mu c}\right)^{-c}$$

Speciellt gäller nu $p_0 = \left(1 + \frac{\lambda}{\mu c}\right)^{-c}$. Ska detta svara mot en Bin$(c, \tilde{p})$-fördelning så måste vi ha

$$1 - \tilde{p} = \left(1 + \frac{\lambda}{\mu c}\right)^{-1} \Rightarrow \tilde{p} = \frac{\frac{\lambda}{\mu c}}{1 + \frac{\lambda}{\mu c}}.$$

Detta $\tilde{p}$-värde svarar mot $\frac{\lambda}{\mu c} = \frac{\tilde{p}}{1-\tilde{p}}$, vilket ger

$$p_n = \binom{c}{n} \left(\frac{\tilde{p}}{1-\tilde{p}}\right)^n \left(1 + \frac{\tilde{p}}{1-\tilde{p}}\right)^{-c}$$

$$= \binom{c}{n} \left(\frac{\tilde{p}}{1-\tilde{p}}\right)^n \left(\frac{1}{1-\tilde{p}}\right)^{-c} = \binom{c}{n} \tilde{p}^n (1-\tilde{p})^{c-n}.$$

63 Vi har

$$\lambda_i = \begin{cases} \lambda \frac{c-i}{c} & \text{om } i = 0, 1, \ldots, c-1 \\ 0 & \text{om } i \geq c. \end{cases}$$

På i stort sett samma sätt som i problem 62 fås

$$p_i = \frac{(c)_i \left(\frac{\lambda}{\mu c}\right)^i}{\sum_{k=0}^{c} (c)_k \left(\frac{\lambda}{\mu c}\right)^k}, \quad i = 0, 1, \ldots, c,$$

där $(c)_k = c(c-1)\cdots(c-k+1) = c!/k!$.

Det torde inte gå att associera denna fördelning med någon mera ”välkänd” fördelning.

64 Låt X vara antalet fartyg som ligger i hamnen. Enligt satserna 7.3 och 7.4 på sidan 81 gäller att

$$P(X > 0) = \rho \quad \text{och} \quad \ell = E[X] = \frac{\rho}{1-\rho},$$

där $\rho = \lambda/\mu$.

Den totala genomsnittliga kostnaden $k(\mu)$, som funktion av μ, är

$$k(\mu) = a + b\mu + c\mu\rho + \frac{d\rho}{1-\rho} = a + b\mu + c\lambda + \frac{d\lambda}{\mu - \lambda}.$$

Derivation ger

$$k'(\mu) = b - \frac{d\lambda}{(\mu-\lambda)^2}.$$

Ekvationen $k'(\mu) = 0$ har lösningarna $\mu = \lambda \pm \sqrt{d\lambda/b}$, men då $\mu > \lambda$ fås $\mu = \lambda + \sqrt{d\lambda/b}$ som enda giltiga lösningen. Denna svarar mot ett minimum!

65 Vi har ett $M/M/3$-system med (tidsenhet minuter) $\lambda = 1/10$ och $\mu = 1/20$. Enligt (7.15) på sidan 89 fås

$$w_q = \frac{C_3 \rho}{\lambda(1-\rho)},$$

där

$$\rho = \frac{\lambda}{3\mu} = \frac{2}{3}$$

$$C_3 = \frac{\frac{2^3}{6}}{\frac{1}{3}\left(1 + 2 + \frac{2^2}{2} + \frac{2^3}{2}\right)} = \frac{4}{9}.$$

Detta ger

$$w_q = \frac{\frac{4}{9}\frac{2}{3}}{\frac{1}{10}\frac{1}{3}} = \frac{80}{9} \approx 9 \text{ minuter.}$$

66 Situationen i uppgiften diskuteras i exempel 7.8 på sidan 90, men vi upprepar lösningen här. Enligt följdsats 7.2 (7.22) på sidan 90 resp. sats 7.5 på sidan 82 följer

$$w_{q,1} = \frac{\rho^2}{\mu(1-\rho^2)} \quad \text{resp.} \quad w_{q,2} = \frac{\rho}{\mu(1-\rho)}.$$

Detta ger

$$\frac{w_{q,2}}{w_{q,1}} = \frac{\frac{\rho}{\mu(1-\rho)}}{\frac{\rho^2}{\mu(1-\rho^2)}} = \frac{1+\rho}{\rho} = 1 + \frac{1}{\rho} > 2.$$

67 a) Enligt sats 7.5 på sidan 82 följer

$$R_1 = \frac{\rho}{1-\rho}.$$

b) Enligt följdsats 7.2 (7.22) på sidan 90 följer

$$R_2 = \frac{\rho^2}{1-\rho^2}.$$

c) Vi har

$$\frac{R_1}{R_2} = \frac{\rho}{1-\rho} \cdot \frac{1-\rho^2}{\rho^2} = \frac{1+\rho}{\rho} = 1 + \frac{1}{\rho} > 2,$$

d.v.s. $R_1 > R_2$.

68 Enligt sats 7.5 på sidan 82 så gäller det för ett $M/M/1$-system att

$$w_q = \frac{\rho}{\mu(1-\rho)}.$$

a) För kassa A gäller $\rho = \lambda/\mu = (30/60) \cdot 1.5 = 0.75$ vilket ger $w_q = 0.75 \cdot 1.5/0.25 = 4.5$ min. På motsvarande sätt fås för kassa B att $\rho = \lambda/\mu = (20/60) \cdot 1.5 = 0.5$ vilket ger $w_q = 0.5 \cdot 1.5/0.5 = 1.5$ min.

b) För både A och B gäller att $\lambda = \frac{1}{2}(20+30)/60 = 5/12$ vilket ger $\rho = (5/12) \cdot 1.5 = 5/8$ och således fås $w_q = (5/8) \cdot 1.5/(3/8) = 5/2 = 2.5$ min.

c) Kassorna utgör nu ett $M/M/2$-system med $\lambda = (20+30)/60 = 5/6$ och $\rho = (5/6) \cdot 1.5/2 = 5/8$, jmf. b)-delen. Av följdsats 7.2 (7.22) på sidan 90 fås

$$w_q = \frac{\rho^2}{\mu(1-\rho^2)} = \frac{(5/8)^2 \cdot 1.5}{(1-(5/8)^2} = \frac{25 \cdot 3}{13 \cdot 3 \cdot 2} = 25/26\text{min}.$$

69 För ett $M/M/c$-system med $\lambda = \mu = 1$ har vi $\rho = 1/c$. Vi betraktar $c = 1, 2, \ldots$

$c = 1$
$\rho = 1$ medför att $w_q = \infty$, d.v.s. att kön "exploderar".

$c = 2$

Vi har nu $\rho = 1/2$. Av följdsats 7.2 (7.22) på sidan 90 fås $w_q = \frac{\rho^2}{\mu(1-\rho^2)} = \frac{1/4}{3/4} = 1/3 > 0.2$.

$c = 3$

Av (7.15) fås

$$w_q = \frac{C_3\rho}{\lambda(1-\rho)} = \frac{C_3}{2}$$

där

$$C_3 = \frac{\frac{(3\rho)^3}{3!}}{(1-\rho)\Big(\sum_{i=0}^{2}\frac{(3\rho)^i}{i!} + \frac{(3\rho)^3}{3!(1-\rho)}\Big)} = \frac{\frac{1}{6}}{\frac{2}{3}\big(1+1+\frac{1}{2}+\frac{1}{6\cdot(2/3)}\big)} = \frac{1}{4\cdot\frac{11}{4}} = \frac{1}{11},$$

vilket ger $w_q = 1/22 < 0.2$. Således räcker 3 kassor.

70 Låt $X(t)$ vara antalet kunder i system 1 vid tid t. Om $X(t) = 1, 2, \ldots, m-1$ väntar system 1 på att en kund lämnar systemet eller att ny anländer. Om $X(t) = 0$ väntar systemet bara på att en kund anländer, d.v.s. blir färdigexpedierad i system 2. Om $X(t) = m$ väntar system 1 bara på att en kund blir färdigexpedierad. Av sats (6.2) på sidan 49 följer då att $X(t)$ är en födelsedöds-process med tillståndsrummet $\{0, 1, \ldots, m\}$, födelseintensiteter

$$\lambda_i = \tilde{\mu}_2, \quad i = 0, \ldots, m-1,$$

och dödsintensiteter

$$\mu_i = \tilde{\mu}_1, \quad i = 1, \ldots, m.$$

Detta ger

$$\rho_i = \left(\frac{\tilde{\mu}_2}{\tilde{\mu}_1}\right)^i, \quad i = 0, \ldots, m.$$

Vi skiljer nu på fallen $\tilde{\mu}_1 \neq \tilde{\mu}_2$ och $\tilde{\mu}_1 = \tilde{\mu}_2$. För fallet $\tilde{\mu}_1 \neq \tilde{\mu}_2$ fås

$$\sum_{i=0}^{m}\rho_i = \frac{1-(\tilde{\mu}_2/\tilde{\mu}_1)^{m+1}}{1-(\tilde{\mu}_2/\tilde{\mu}_1)},$$

enligt räknereglerna för geometriska summor. Det följer nu av sats (6.12) på sidan 68 att

$$p_i = \frac{\rho_i}{\sum_{i=0}^{m}\rho_i} = \frac{(\tilde{\mu}_2/\tilde{\mu}_1)^i(1-(\tilde{\mu}_2/\tilde{\mu}_1))}{1-(\tilde{\mu}_2/\tilde{\mu}_1)^{m+1}}, \quad i = 0, \ldots, m.$$

För fallet $\tilde{\mu}_1 = \tilde{\mu}_2$ fås $\rho_i = 1$ för $i = 0, \ldots, m$, vilket ger

$$\sum_{i=0}^{m}\rho_i = m+1,$$

och

$$p_i = \frac{1}{m+1}, \quad i = 0, \ldots, m.$$

71 Låt $X(t)$ vara antalet pågående samtal vid tiden t. Då är $X(t)$ en födelse-döds-process med alla födelseintensiteter $\lambda_i = \lambda$ och dödsintensiteter $\mu_i = i\mu$. Sätt, som vanligt,

$$\rho_0 = 1 \quad \text{och} \quad \rho_n = \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n} = \frac{(\lambda/\mu)^n}{n!} \,.$$

Detta ger $\sum_{i=0}^{\infty} \rho_i = e^{\lambda/\mu}$ och det följer av sats 6.12 på sidan 68 att

$$p_n = \frac{(\lambda/\mu)^n}{n!}\, e^{-\lambda/\mu},$$

d.v.s. X är Po(λ/μ)-fördelad.

72 Låt T_i vara tidpunkten (räknat från stängningsdags) då den i:te kunden fått betjäning, och sätt

$$S_i = T_i - T_{i-1}, \quad \text{där} \quad T_0 = 0.$$

Således gäller det att

$$T = T_{12} = \sum_{i=1}^{12} S_i.$$

Så länge det finns minst 5 kunder i systemet så arbetar alla 5 kassorna för fullt, vilket innebär att $T_1, \ldots, T_8$ är fördelade som minimum av 5 st. oberoende Exp(1/8)-fördelade stokastiska variabler, d.v.s. $T_1, \ldots, T_8$ är Exp(5/8)-fördelad, se sats (2.3) på sidan 6. På motsvarande sätt är T_9 Exp(4/8)-fördelad osv. Vidare är $T_1, \ldots, T_8$ är oberoende. Detta ger

$$E[T] = 8 \cdot \frac{8}{5} + \frac{8}{4} + \frac{8}{3} + \frac{8}{2} + 8 = \frac{442}{15} = 29.5,$$

$$V[T] = 8 \cdot \left(\frac{8}{5}\right)^2 + \left(\frac{8}{4}\right)^2 + \left(\frac{8}{3}\right)^2 + \left(\frac{8}{2}\right)^2 + 8^2 = \frac{25108}{225}$$

och

$$D[T] = \sqrt{\frac{25108}{225}} = 10.6.$$

73 Vi betecknar taxibilarnas och kundernas ankomstintensiteter med λ_T resp. λ_K. Låt $(X_T(t), X_K(t))$ vara antalet taxibilar resp. kunder vid stationen vid tid t. De möjliga tillstånden är

$$(0,3),\ (0,2),\ (0,1),\ (0,0),\ (1,0),\ (2,0),\ (3,0), \ldots$$

där en övergång åt höger sker då en bil kommer, d.v.s. med intensitet λ_T, och en övergång åt vänster då en kund kommer, d.v.s. med intensitet λ_K. Bildar vi nu

$$X(t) = X_T(t) - X_K(t) + 3,$$

så är $X(t)$ en födelse-döds-process med födelseintensiteter λ_T och dödsintensiteter λ_K. (Det extra "+3" är till för att ge $X(t)$ det vanliga tillståndsrummet.)

Bortsett från beteckningarna för intensiteterna så är $X(t)$ den födelse-dödsprocess som beskriver antalet kunder i ett $M/M/1$-system. Det följer således från sats 7.3 på sidan 81 att

$$p_n = \lim_{t\to\infty} P(X(t) = n) = \rho^n(1-\rho), \quad \text{för } n = 0, 1, \ldots,$$

under förutsättning att $\rho = \lambda_T/\lambda_K < 1$. I vårt fall är $\rho = 1/1.25 = 0.8$.

a) Vi får, jmf. sats 7.4 på sidan 81,

$$E[X_T] = \sum_{n=0}^{\infty} n\rho^{n+3}(1-\rho) = \rho^3 \sum_{n=0}^{\infty} n\rho^n(1-\rho)$$

$$= \rho^3\ell = \frac{\rho^4}{1-\rho} = 0.8^4/0.2 = 2.048.$$

b) Vi får

$$E[X_K] = 3p_0 + 2p_1 + 1p_2 = (3 + 2\rho + \rho^2)(1-\rho) = 1.048.$$

En alternativ lösning är följande:

$$E[X_K] = E[-X + X_T + 3] = -\frac{\rho}{1-\rho} + \frac{\rho^4}{1-\rho} + 3 = -4 + 2.048 + 3 = 1.048.$$

c) En kund bortfaller med sannolikheten $p_0 = (1-\rho) = 0.2$. I medeltal kommer $60\lambda_K = 75$ kunder/timme, vilket medför att i medeltal bortfaller $p_0 \cdot 60\lambda_K = 0.2 \cdot 75 = 15$ kunder/timme.

74 Vi kan skriva $U = U_1 + U_2$ där $U_1 = 1/3$ minut och U_2 är Exp(1)-fördelad. Observera att U inte är exponentialfördelad. Med sedvanliga beteckningar fås

$$b = E(U_1) + E(U_2) = \frac{1}{3} + 1 = \frac{4}{3}$$
$$V(U) = V(U_1) + V(U_2) = 0 + 1 = 1$$
$$\rho = \lambda \cdot b = \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}.$$

Detta ger, se sats 7.14 på sidan 100,

$$\ell_q = \frac{4/9}{2 \cdot 1/3} \cdot \left(1 + \frac{1}{16/9}\right) = \frac{25}{24}.$$

Vidare fås med sats 7.1 på sidan 79 att

$$\ell = \rho + \ell_q = \frac{2}{3} + \frac{25}{24} = \frac{41}{24} = 1.71.$$

75 Betrakta en kund och låt N vara antalet ärenden som kunden har, d.v.s.

$$P(N = 1) = P(N = 2) = 1/2.$$

Låt U vara kundens betjäningstid och U_i, $i = 1, 2$, vara ärendenas betjäningstider. Detta innebär att

$$U = \sum_{i=1}^{N} U_i,$$

d.v.s. U är en stokastisk summa. Eftersom U_i är Exp(μ)-fördelad så gäller det att

$$E[U_i] = 1/\mu \quad \text{och} \quad E[U_i^2] = V[U_i] + E[U_i]^2 = 1/\mu^2 + 1/\mu^2 = 2/\mu^2.$$

Av detta fås

$$\begin{aligned} E[U] &= \frac{1}{2}E[U \mid N = 1] + \frac{1}{2}E[U \mid N = 2] = \frac{1}{2}E[U_1] + \frac{1}{2}E[U_1 + U_2] \\ &= \frac{1}{2}\frac{1}{\mu} + \frac{1}{2}\frac{2}{\mu} = \frac{3}{2\mu} \end{aligned}$$

och

$$\begin{aligned} E[U^2] &= \frac{1}{2}E[U^2 \mid N = 1] + \frac{1}{2}E[U^2 \mid N = 2] \\ &= \frac{1}{2}E[U_1^2] + \frac{1}{2}E[(U_1 + U_2)^2] = \frac{1}{2}\frac{2}{\mu^2} + \frac{1}{2}E[U_1^2 + U_2^2 + 2U_1U_2] \\ &= \frac{1}{2}\frac{2}{\mu^2} + \frac{1}{2}\left(\frac{2}{\mu^2} + \frac{2}{\mu^2} + 2\frac{1}{\mu}\frac{1}{\mu}\right) = \frac{1}{2}\frac{2}{\mu^2} + \frac{1}{2}\frac{6}{\mu^2} = \frac{4}{\mu^2}. \end{aligned}$$

Detta ger slutligen

$$V[U] = E[U^2] - E[U]^2 = \frac{4}{\mu^2} - \left(\frac{3}{2\mu}\right)^2 = \frac{4}{\mu^2} - \frac{9}{4\mu^2} = \frac{7}{4\mu^2}.$$

b) Systemet är ett $M/G/1$-system, och det följer av avsnitt 7.4.1 att

$$\rho = \lambda E[U] = \frac{3\lambda}{2\mu}.$$

Det följer nu av sats 7.14 på sidan 100 att

$$\ell_q = \frac{\rho^2}{2(1-\rho)}\left(1 + \frac{V[U]}{E[U]^2}\right) = \frac{8\rho^2}{9(1-\rho)},$$

under förutsättning att $\rho < 1$.

c) Av sats 7.1 på sidan 79 fås

$$\begin{aligned} \ell &= \rho + \ell_q = \frac{\rho(9-\rho)}{9(1-\rho)} \\ w_q &= \frac{\ell_q}{\lambda} = \frac{8\rho^2}{9\lambda(1-\rho)} \\ w &= \frac{\ell}{\lambda} = \frac{\rho(9-\rho)}{9\lambda(1-\rho)} \end{aligned}$$

76 Det beskrivna systemet är ett Jacksonnätverk med parametrar

$$m = 2 \quad c_1 = c_2 = 1$$

$$\lambda_1 = \lambda \quad \lambda_2 = 0$$

$$p_{12} = 1 \quad p_2 = p \quad p_{21} = 1 - p.$$

Från (7.40) får vi

$$\Lambda_1 = \lambda + (1-p)\Lambda_2 \quad \text{och} \quad \Lambda_2 = \Lambda_1$$

vilket ger

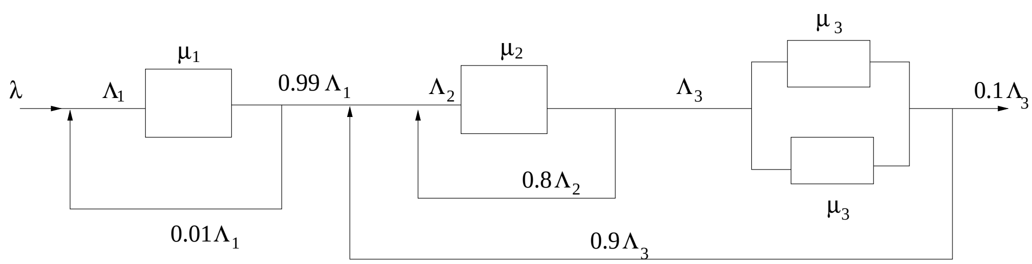
$$\Lambda_1 = \Lambda_2 = \frac{\lambda}{p}$$

Av satserna 7.16 och 7.3 på sidorna 107 och 81 följer att

$$p_{k,n} = p_n = \rho_1^k(1-\rho_1) \cdot \rho_2^n(1-\rho_2), \quad \text{för } k, n = 0, 1, \ldots,$$

där $\rho_1 = \lambda/(p\tilde{\mu}_1)$ och $\rho_2 = \lambda/(p\tilde{\mu}_2)$. Man kan observera att detta system är likt fleranvändarsystemet som diskuterades i exempel 7.16 på sidan 108. Skillnaden är från vilket delsystem som kunder kan försvinna.

77 Jackson-nätverk!



Vi får

$$\Lambda_1 = \lambda + 0.01\Lambda_1, \text{ d.v.s. } \Lambda_1 = \frac{\lambda}{0.99} = \frac{0.15}{0.99} = \frac{5}{33}.$$

$$\Lambda_2 = 0.99\Lambda_1 + 0.9\Lambda_3 + 0.8\Lambda_2 \text{ och } \Lambda_3 = 0.2\Lambda_2 \text{ vilket ger}$$
$$\Lambda_2 = 0.99\frac{\lambda}{0.99} + 0.9 \cdot 0.2\Lambda_2 + 0.8\Lambda_2, \text{ d.v.s.}$$
$$\Lambda_2 = \frac{\lambda}{0.02} = \frac{0.15}{0.02} = 7.5 \text{ och } \Lambda_3 = 0.2\Lambda_2 = 1.5$$

Vi får vidare (ρ_i nodernas betjäningsfaktorer)

$$\mu_1 = \frac{1}{0.1} = 10, \qquad \rho_1 = \frac{\Lambda_1}{\mu_1} = \frac{5/33}{10} = \frac{1}{66} < 1,$$
$$\mu_2 = \frac{1}{0.1} = 10, \qquad \rho_2 = \frac{\Lambda_2}{\mu_2} = \frac{7.5}{10} = 0.75 < 1,$$
$$\mu_3 = \frac{1}{0.2} = 5, \qquad \rho_3 = \frac{\Lambda_3}{2\mu_3} = \frac{1.5}{2 \cdot 5} = 0.15 < 1.$$

Genomsnitten antalet ”kunder” (d.v.s. jobb) i de tre delsystemen blir alltså

$$\ell_1 = \frac{\rho_1}{1-\rho_1} = \frac{1/66}{1-1/66} = \frac{1}{65} = 0.0154,$$
$$\ell_2 = \frac{\rho_2}{1-\rho_2} = \frac{0.75}{1-0.75} = 3,$$
$$\ell_3 = \frac{2\rho_3}{1-\rho_3^2} = \frac{2 \cdot 0.15}{1-0.15^2} = 0.3069,$$

d.v.s. totalt finns i systemet $\ell = \ell_1 + \ell_2 + \ell_3 = 3.3213$ jobb i genomsnitt.

b) $P(N_2 \geq 1) = 1 - P(N_2 = 0) = 1 - (1-\rho_2)\rho_2^0 = \rho_2 = 0.75$
och
$W = \frac{\ell}{\lambda} = \frac{3.307}{0.15} = 22.65$ tidsenheter.

Litteraturförteckning

Litteraturen inom markovprocessteorin är mycket omfattande. En modern och trevlig bok, som kräver ungefär samma förkunskaper som detta kompendium är [11]. Den är mer omfattande än detta kompendium och behandlar också martingaler, potentialteori, elektriska nätverk och brownsk rörelse. Även andra tillämpningar tas upp, såsom biologiska tillämpningar av Markovkedjor.

[6] omfattar förutom Markovprocesser även allmän sannolikhetsteori och konvergens av fördelningar. Den innehåller också ett stort antal problem att lösa. Även [12] omfattar allmän sannolikhetsteori, men behandlar även Markov- och Poissonprocesser. [13] behandlar enbart stokastiska processer mer ingående än [12]. Båda ger många tillämpningar på processteori.

[4] är en klassisk bok. Den behandlar sannolikhetsteori, diskreta stokastiska variabler, slumpvandring (random walk) och Markovkedjor. Boken innehåller många intresseväckande exempel och varje kapitel avslutas med flera problem. Feller har också skrivit en Volume II, som är matematiskt sett svårare och behandlar kontinuerliga stokastiska variabler och processer i kontinuerlig tid.

Beviset av ergodicitet i detta kompendium, använder kopplingsteknik. [10] behandlar kopplingsmetoder med flera exempel från Markovkedjor.

[7] ger en omfattande framställning av Markovkedjor i diskret tid. Den är lämplig för högre studier i markovteori.

En trevlig framställning av köteorin, på ungefär samma nivå och med liknande innehåll som detta kompendium, återfinns i [1]. Denna bok kan varmt rekomenderas som brevidläsningslitteratur för den som vill ha mera kött på benen. Den som — senare — behöver en mera omfattande framställning hänvisas t.ex. till [8] och [9].

[14] är den klassiska framställningen av avancerad köteori då ankomstprocessen är en förnyelseprocess.

En modern behandling av återkopplade system finner man i [3], som dock kräver omfattande kunskaper om stokastiska processer. Boken kan rekommenderas för doktorander i matematisk statistik eller i optimeringslära och systemteori med en inriktning åt stokastisk analys. Standardverket om $G/G/c$-system är [5]. Denna bok är dock mycket svårläst och och rekomenderas i första hand för doktor(and)er i matematisk statistik med en inriktning åt punktprocesser.

[1] Allen, A. (1990) *Probability, Statistics and Queueing Theory with Computer Science Applications.* Academic Press, New York.

[2] Anderson, W.J. (1991) *Continuous-Time Markov Chains.* Springer-Verlag, New York.

[3] Brémaud, P. (1981) *Point Processes and Queues. Martingale Dynamics.* Springer-Verlag, New York.

[4] Feller, W. (1968) *An Introduction to Probability Theory ans Its Application. Volume I.* John Wiley & Sons, New York.

[5] Franken, P., König, D., Arndt, U., and Schmidt, V. (1981) *Queues and Point Processes.* Akademie-Verlag, Berlin and John Wiley & Sons, New York.

[6] Grimmett, G.R., Stirzaker, D.R. (2001) *Probability and Random Processes.* Oxford University Press.

[7] Kemeny, J.G., Snell, J.L., Knapp, A.W. (1976) *Denumerable Markov Chains.* Springer-Verlag.

[8] Kleinrock, L. (1975) *Queueing Systems, Volume I: Theory.* John Wiley & Sons, New York.

[9] Kleinrock, L. (1976) *Queueing Systems, Volume II: Computer Applications.* John Wiley & Sons, New York.

[10] Lindvall, T. (2002) *Lectures on the Coupling Method.* Dover Publications.

[11] Norris, J.R. (1997) *Markov Chains.* Cambridge University Press.

[12] Ross, S.M. (1997) *Introduction to Probability Models.* Academic Press.

[13] Ross, S.M. (1995) *Stochastic Processes.* John Wiley & Sons, New York.

[14] Takács, L. (1962) *Introduction to the Theory of Queues.* Oxford University Press, New York.

Sakregister