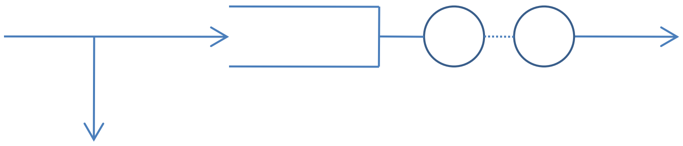
Föreläsning 2, Kösystem

Vi ska börja titta på enskilda kösystem som ser ut på följande sätt:



Det kan finnas en eller fler betjänare och bufferten kan vara ändlig eller oändlig. Om bufferten är oändligt så kommer aldrig kunder att spärras. Så snart en betjänare är färdig med en kund så hämtar den en ny kund från kön om det finns någon där. Kunder kan hämtas ur bufferten på flera olika sätt:

FIFO – First In First Out (det vanligaste i köer)

LIFO – Last In First Out (brukar kallas stack i programmering)

SIRO – Service In Random Order (en kund plockas helt slumpmässigt)

Vi inför några beteckningar:

$A(t) =$ fördelningsfunktionen för tiderna mellan ankomster

$a(t) = A'(t) =$ frekvensfunktionen för tiden mellan ankomster

$B(t) =$ fördelningsfunktionen för betjäningstiderna

$b(t) = B'(t) =$ frekvensfunktionen för betjäningstiderna

Om $A(t)$ och $B(t)$ är kända och vi vet hur många betjänare och buffertplatser det finns så vill vi beräkna sannolikheten för spärr och de statistiska egenskaperna hos tiden i systemet för en kund som inte spärras.

En beteckning som ofta används: A/B/n*kommentar. A anger fördelningen mellan ankomster, B fördelningen för betjäningar, n är antalet betjänare. I kommentaren kan man lägga till ytterligare information. Om inget annat sägs så antar vi att antalet buffertplatser är oändligt. Ofta används följande förkortningar för fördelningar:

M – exponentialfördelning

D – deterministisk (det vill säga konstant)

G – en godtycklig (men känd) fördelning

För att det ska finnas enkla sätt att beräkna det vi vill för kösystem så måste man ställa vissa krav på fördelningarna för tiderna mellan ankomsterna och betjäningstiderna. I denna kurs kommer vi att räkna på så kallade markovska köer och på M/G/1-köer. Vi kommer också att studera simulering som man kan använda till mycket allmännare kösystem.

Markovska köer

En markovsk kö kan beskrivas med en markovkedja. Det finns markovkedjor i diskret tid och markovkedjor i kontinuerlig tid. Vi ska använda Markovkedjor i kontinuerlig tid. En markovkedja befinner sig i ett tillstånd. Tillstånden kan numreras med heltal och det kan finnas ändligt många eller oändligt många tillstånd. I denna kurs så representerar för det mesta tillståndets nummer antal kunder i ett kösystem. Det trevliga med markovkedjor är att det finns ett matematiskt maskineri som kan användas för att utföra beräkningar. Dock måste man ställa vissa krav på $A(t)$ och $B(t)$ för att man ska kunna beskriva ett kösystem med en markovkedja. Först kommer en definition på markovkedjor i kontinuerlig tid (i fortsättningen skriver jag bara markovkedja) och sedan reder vi ut vilka villkor som ett kösystem måste uppfylla för att man ska få beskriva det med en markovkedja.

Vad är en markovkedja?

Vi sätter

$X(t) =$ tillståndet för ett system vid tiden t

Tillståndet är ett heltal och tiden är kontinuerlig det vill säga kan anta vilket reellt värde som helst. Antag att $t_1 < t_2 < \cdots < t_n < t_{n+1}$ och att följande gäller för godtyckliga t_i och n:

$$P(X(t_{n+1}) = k_{n+1} | X(t_n) = k_n, X(t_{n-1}) = k_{n-1} \dots X(t_1) = k_1) = P(X(t_{n+1}) = k_{n+1} | X(t_n) = k_n)$$

Då säger vi att $X(t)$ är en markovkedja. Formeln ovan innebär att vi kan kasta bort all historia före tiden t_n när vi vill förutsäga vad som händer efter t_n. Det enda som har betydelse för vad som händer i framtiden är i vilket tillstånd som man befinner sig, hur man har tagit sig dit spelar ingen roll. Man kan visa att om Δt är ett mycket litet tidsintervall så är om $i \neq j$:

$$\lim_{\Delta t \to 0} \frac{P(X(t + \Delta t) = j | X(t) = i)}{\Delta t} = \lambda_{ij}$$

Detta innebär att för små Δt så är

$$P(X(t + \Delta t) = j | X(t) = i) \approx \lambda_{ij} \cdot \Delta t$$

Hur räknar man på en markovkedja?

För att slippa skriva så mycket inför vi en ny beteckning:

$$p_i(t) = P(X(t) = i)$$

Låt oss anta att vi har en markovkedja med tre tillstånd, 1, 2 och 3. Då får vi följande ekvationer:

$$\begin{aligned}
p_1(t + \Delta t) &= (1 - (\lambda_{12} + \lambda_{13})\Delta t) \cdot p_1(t) + \lambda_{21}\Delta t \cdot p_2(t) + \lambda_{31}\Delta t \cdot p_3(t) \\
p_2(t + \Delta t) &= \lambda_{12}\Delta t \cdot p_1(t) + (1 - (\lambda_{21} + \lambda_{23})\Delta t) \cdot p_2(t) + \lambda_{32}\Delta t \cdot p_3(t) \\
p_3(t + \Delta t) &= \lambda_{13}\Delta t \cdot p_1(t) + \lambda_{23}\Delta t \cdot p_2(t) + (1 - (\lambda_{31} + \lambda_{32})\Delta t) \cdot p_3(t)
\end{aligned}$$

Dessa ekvationer kan skrivas om på följande sätt

$$\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t} = -(\lambda_{12}+\lambda_{13})p_1(t)+\lambda_{21}p_2(t)+\lambda_{31}p_3(t)$$
$$\frac{p_2(t+\Delta t)-p_2(t)}{\Delta t} = \lambda_{12}p_1(t)-(\lambda_{21}+\lambda_{23})p_2(t)+\lambda_{32}p_3(t)$$
$$\frac{p_3(t+\Delta t)-p_3(t)}{\Delta t} = \lambda_{13}p_1(t)+\lambda_{23}p_2(t)-(\lambda_{31}+\lambda_{32})p_3(t)$$

Låter vi nu $\Delta t \to 0$ så får vi

$$p_1'(t) = -(\lambda_{12}+\lambda_{13})p_1(t)+\lambda_{21}p_2(t)+\lambda_{31}p_3(t)$$
$$p_2'(t) = \lambda_{12}p_1(t)-(\lambda_{21}+\lambda_{23})p_2(t)+\lambda_{32}p_3(t)$$
$$p_3'(t) = \lambda_{13}p_1(t)+\lambda_{23}p_2(t)-(\lambda_{31}+\lambda_{32})p_3(t)$$

Om vi inför Q-matrisen

$$Q = \begin{bmatrix} -\lambda_{12}-\lambda_{13} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -\lambda_{21}-\lambda_{23} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -\lambda_{31}-\lambda_{32} \end{bmatrix}$$

och radmatrisen

$$p(t) = [p_1(t) \quad p_2(t) \quad p_3(t)]$$

så kan vi skriva

$$p'(t) = p(t) \cdot Q$$

Det innebär att om vi känner till $p(0)$ så kan vi beräkna sannolikheten att befinna sig i ett visst tillstånd vid en godtycklig tidpunkt $t > 0$. Naturligtvis går detta att göra för ett godtyckligt antal tillstånd, att det visas för tre tillstånd här är bara för att slippa skriva så mycket. Oftast är man inte intresserad av $p(t)$ utan vill beräkna den så kallade ***jämviktsfördelningen*** för markovkedjan. Man kan visa att för de flesta markovkedjor så existerar följande gränsvärde

$$\lim_{t\to\infty} p_i(t) = p_i$$

Vektorn $p = [p_1 \quad p_2 \quad p_3]$ kallar vi oftast för ***tillståndsfördelningen*** för markovkedjan. Det gäller att

$\lim_{t\to\infty} p_i(t)$ existerar $\Longrightarrow \lim_{t\to\infty} p_i'(t) = 0$

Således kan vi bestämma p med hjälp av ekvationssystemet

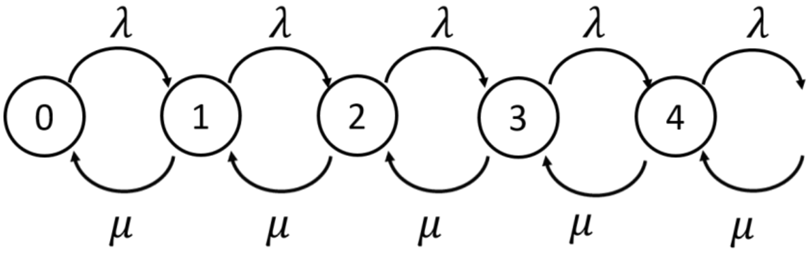
$$\begin{cases} 0 = p \cdot Q \\ 1 = \sum_{\forall k} p_i \end{cases}$$

Den första ekvationen får vi genom att låta $t \to \infty$ i $p'(t) = p(t) \cdot Q$ och den andra behövs för att lösningen ska bli entydig. Den andra säger ju helt enkelt att man måste befinna sig i något tillstånd i markovkedjan, summan av alla sannolikheter måste då vara =1.

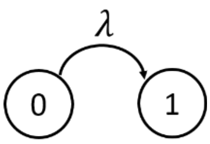
Ofta ritas markovkedjor som tillståndsdiagram där intensiteterna λ_{ij} skrivs på bågarna, se nedan.

Markovkedjor och kösystem

Eftersom det finns ett matematiskt maskineri för markovkedjor vore det ju trevligt att få använda det för kösystem. Till exempel kunde man låta tillstånd i i markovkedjan betyda att det finns i kunder i systemet och sedan beskriva kösystemet med följande markovkedja:



Detta är bara möjligt om fördelningen för tiden mellan ankomster och fördelningen för betjäningstiderna uppfyller vissa villkor. Låt oss ta reda på vilka de är! Om vi börjar med tiden mellan ankomsterna så måste det gälla att den är lika med tiden som man tillbringar i tillstånd 0 i markovkedjan nedan om man befinner sig i 0 vid tiden 0.



Q-matrisen för denna markovkedja är

$$\begin{bmatrix} -\lambda & \lambda \\ 0 & 0 \end{bmatrix}$$

Det ger följande system av differentialekvationer

$$p'(t) = p(t) \cdot Q \Leftrightarrow \begin{cases} p_0'(t) = -\lambda p_0(t) \\ p_1'(t) = \lambda p_0(t) \end{cases}$$

Dessutom har vi begynnelsevillkoret

$$p_0(0) = 1$$
$$p_1(0) = 0$$

Vi löser ekvationen för $p_0(t)$:

$$p_0'(t) = \lambda p_0(t) \Longrightarrow p_0(t) = Ce^{-\lambda t}$$

Eftersom $p_0(0) = 1$ så måste $C = 1$ vilket ger att $p_0(t) = e^{-\lambda t}$. Nu observerar vi att om X är tiden i tillstånd 0 så gäller

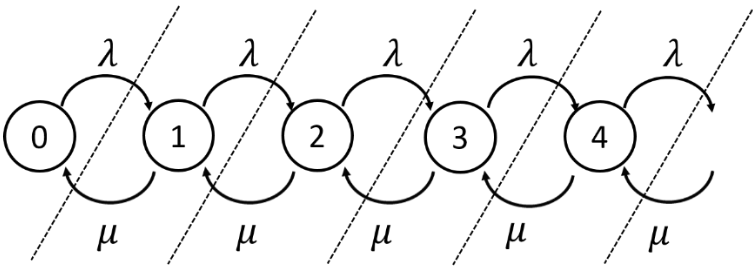
$$p_0(t) = P(X > t) = 1 - P(X \leq t) = 1 - F_X(t) \Longrightarrow F_X(t) = 1 - p_0(t) = 1 - e^{-\lambda t}$$

Detta innebär att tiden mellan ankomster måste vara exponentialfördelade. På samma sätt visas att betjäningstiderna är exponentialfördelade.

Vi kan beskriva ett kösystem med en markovkedja om alla tiden mellan ankomster och betjäningstiden är exponentialfördelade.

M/M/1

Detta kösystem har exponentialfördelade betjäningstider och tiden mellan ankomster är också expoentialfördelade. Vi får följande markovkedja:



Observera att vi har ritat linjer som ”skär av” markovkedjan mellan tillstånden. Kedjan fortsätter mot oändligheten åt höger. Det enklaste sättet att hitta sannolikheten att man i jämvikt befinner sig i ett visst tillstånd för denna markovkedja är att konstatera att antalet hopp per tidsenhet från höger till vänster över en av linjerna i kedjan måste vara mycket nära antalet hopp från vänster till höger, åtminstone på lång sikt. Detta leder till den så kallade ***snittmetoden***:

Antal hopp åt vänster per tidsenhet genom snittet mellan 0 och 1: λp_0

Antal hopp åt höger per tidsenhet genom snittet mellan 0 och 1: μp_1

Om dessa ska vara lika och dessutom gälla för alla snitten i figuren så får vi:

$$\lambda p_0 = \mu p_1 \Rightarrow p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0$$

$$\lambda p_1 = \mu p_2 \Rightarrow p_2 = \frac{\lambda}{\mu} p_1 = \rho^2 p_0$$

$$\lambda p_2 = \mu p_3 \Rightarrow p_3 = \frac{\lambda}{\mu} p_2 = \rho^3 p_0$$

Man inser att i det allmänna fallet gäller

$$p_k = \rho^k p_0$$

Nu återstår att bestämma p_0. Det kan vi göra genom att utnyttja att summan av alla sannolikheter ska vara = 1. Det ger:

$$1 = \sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \rho^k p_0 = p_0 \sum_{k=0}^{\infty} \rho^k = p_0 \frac{1}{1-\rho} \Rightarrow p_0 = 1 - \rho$$

Således gäller att

$$p_k = (1-\rho)\rho^k$$

Observera också att detta går bara att göra om $\rho < 1$. Om $\rho \geq 1$ så finns det inte någon jämviktsfördelning eftersom antalet kunder i kön kommer att växa hela tiden.

Man kan också lösa den här typen av problem genom att utnyttja att antalet hopp in i ett tillstånd måste vara lika med antalet hopp ut ur tillståndet vilket ger att

$$\lambda p_0 = \mu p_1$$

$$(\lambda + \mu)p_1 = \lambda p_0 + \mu p_1$$

$$(\lambda + \mu)p_2 = \lambda p_1 + \mu p_3$$

osv. Denna metod bruka kallas ***flöde-in-flöde-ut***.

Man kan också utnyttja att

$$\begin{cases} p = p \cdot Q \\ \sum_k p_1 = 1 \end{cases}$$

När väl alla p_k är kända så kan man beräkna medelantal kunder i systemet med definitionen

$$E(N) = \sum_{k=0}^{\infty} k p_k = \frac{\rho}{1-\rho}$$

Därefter ger Littles sats medeltiden i systemet:

$$E(T) = \frac{E(N)}{\lambda} = \frac{1}{\lambda} \cdot \frac{\rho}{1-\rho} = \frac{1}{\lambda} \cdot \frac{\lambda/\mu}{1-\rho} = \frac{1}{\mu} \cdot \frac{\rho}{1-\rho}$$

När man räknar på ett markovskt kösystem så följer man nästan alltid detta schema:

1. Rita tillståndsdiagrammet
2. Bestäm tillsåndssannolikheterna (kallas också tillståndsfördelningen) med någon av metoderna
    a. Snittmetoden
    b. Flöde in, flöde ut
    c. $0 = p \cdot Q$ kombinerat med $\sum p_i = 1$
3. Beräkna medelantalet kunder i kön med definitionen på medelvärde eller med z-transform
4. Beräkna medeltiden i systemet (mer om detta på föreläsning 3)

Man bör observera att ett M/M/1-system är stabilt om $\rho < 1$, annars instabilt.