TAMS15: SS1
Markovprocesser

Johan Thim (johan.thim@liu.se)

21 november 2018

Vad händer om vi i en Markovkedja har kontinuerlig tid istället för diskreta steg? Detta är ett specialfall av en kategori stokastiska processer som kallas för **Markovprocesser**. Dessa definieras av följande villkor.

Markovprocess

Definition. En stokastisk process $\{X_t\}_{t\in I}$, som tar värden i $\mathbf{R}$, med kontinuerlig tid, kallas en **Markovprocess** om

$$P\big(X_{t_n} \leq x \mid X_{t_1} = x_1, \ldots, X_{t_{n-1}} = x_{n-1}\big) = P\big(X_{t_n} \leq x \mid X_{t_{n-1}} = x_{n-1}\big)$$

för alla $x, x_1, x_2, \ldots, x_{n-1}$ och alla växande följder $t_1 < t_2 < \cdots < t_n$ av tider.

Slarvigt uttryckt innebär detta att det inte spelar någon roll *hur* processen har kommit till ett visst tillstånd, utan *endast* det nuvarande tillståndet bestämmer vad som händer härnäst; precis som för Markovkedjor.

Markovegenskapen och exponentialfördelningen

Sats. Om en icke-negativ stokastisk variabel T har Markovegenskapen, d v s

$$P\big(T > t + a \mid T > a\big) = P(T > t), \quad a, t \in [0, \infty[,$$

så är T exponentialfördelad.

Bevis: Markovegenskapen implicerar att

$$P(T > t) = P\big(T > t + a \mid T > a\big) = \frac{P(T > t + a)}{P(T > a)},$$

så $P(T > t + a) = P(T > a)P(T > t)$ för alla $a, t \geq 0$. Låt $g(s) = \ln P(T > s)$ för $s \geq 0$. Då kommer g att vara högerkontinuerlig (varför?) och

$$g(x + y) = g(x) + g(y), \quad x, y \geq 0.$$

Låt $m, n \in \mathbf{N}$ med $n \neq 0$. Då måste

$$g(1) = g\left(\sum_{k=1}^{n} \frac{1}{n}\right) = \sum_{k=1}^{n} g\left(\frac{1}{n}\right) = ng\left(\frac{1}{n}\right) \quad \Rightarrow \quad g\left(\frac{1}{n}\right) = \frac{g(1)}{n}.$$

Vidare ser vi då att

$$g\left(\frac{m}{n}\right) = \sum_{k=1}^{m} g\left(\frac{1}{n}\right) = \frac{m}{n} g(1).$$

Alltså har vi visat att $g(s) = sg(1)$ för alla positiva rationella tal s. Eftersom alla reella tal kan approximeras godtyckligt nära med rationella tal (vi kan välja en sekvens r_n så att $r_n \in \mathbf{Q}$ och $r_n \to s$ med $r_n \geq s$ för alla n) och g är högerkontinuerlig, följer det att $g(s) = sg(1)$ för alla $s > 0$. Detta implicerar direkt att

$$F_T(s) = 1 - P(T > s) = 1 - e^{g(s)} = 1 - e^{-\lambda s}, \quad s \geq 0,$$

där $\lambda = -\ln P(T > 1)$. Detta är fördelningsfunktionen för en exponentialfördelad variabel T.

11.1 Markovkedjor med kontinuerlig tid

Vi betraktar den så kallade **födelse-döds-processen**. Typexemplet är att låta $X(t)$ vara antalet kunder vid ett betjäningsställe vid tiden t. Det följer att tillståndsrummet E lämpligen väljs som $E = \{0, 1, 2, \ldots\}$ (ändligt eller oändligt). Tanken är att processen nu endast kan ändra sitt tillstånd med ett steg i vardera riktning, alternativt inte ändra sig alls. Vi preciserar i följande definition.

Födelse-döds-process

Definition. Den stokastiska processen $\{X(t)\}_{t\in[0,\infty[}$ är en **födelse-döds-process** om och endast om, i alla små tidsintervall $]t, t+h[$, följande gäller:

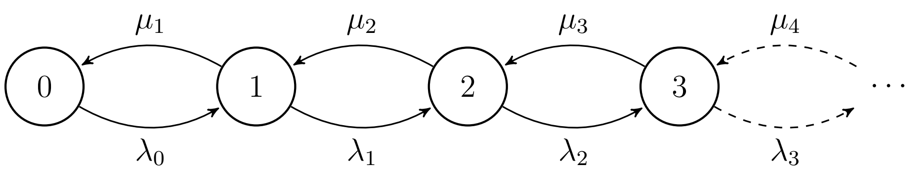
(i) tillståndet ökar från k till $k+1$ med sannolikhet $\lambda_k h + o(h)$,

(ii) tillståndet minskar från k till $k-1$ med sannolikhet $\mu_k h + o(h)$,

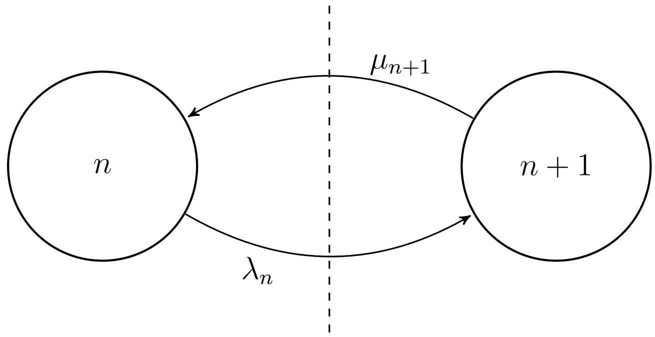
(iii) tillståndet är oförändrat k med sannolikhet $1 - \lambda_k h - \mu_k h + o(h)$.

Vi kräver att $\mu_0 = 0$.

Vi kallar λ_k för **födelseintensiteterna** och μ_k för **dödsintensiteterna**.



Detta är en Markovkedja med kontinuerlig tid. Tyvärr kan vi inte lösa ut sannolikheter lika enkelt som för Poissonprocessen (det går att ställa upp liknande differentialekvationer, men förutom i specialfall blir dessa svårlösta). Vad vi kan göra är att med lite argumentation ta fram stationära sannolikheter. Om kedjan befinner sig i ett stationärt tillstånd måste det rimligen vara samma "flöde" in och ut ur ett tvärsnitt av grafen ovan. Vi kikar mellan tillstånd n och $n+1$:



Antag att kedjan befinner sig i jämnvikt (har uppnått ett stationärt tillstånd). Låt p_n vara sannolikheten att kedjan befinner sig i tillstånd n. Då måste $p_n\lambda_n = p_{n+1}\mu_{n+1}$ för att vi ska ha jämnvikt. Således blir

$$p_{n+1} = \frac{\lambda_n}{\mu_{n+1}}p_n = \frac{\lambda_n}{\mu_{n+1}}\frac{\lambda_{n-1}}{\mu_n}p_{n-1} = \cdots = \frac{\lambda_n\lambda_{n-1}\cdots\lambda_0}{\mu_{n+1}\mu_n\cdots\mu_1}p_0.$$

Vidare vet vi att alla p_n måste summera till ett (kedjan måste befinna sig i något tillstånd), så

$$1 = p_0 + p_1 + \cdots = p_0\left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots\right),$$

förutsatt att

$$1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots < \infty.$$

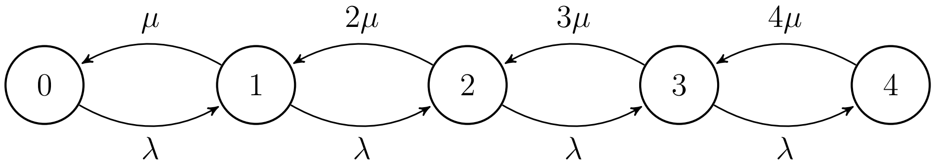
Vi finner alltså att

$$p_0 = \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots\right)^{-1} \qquad \text{och} \qquad p_{n+1} = \frac{\lambda_n\lambda_{n-1}\cdots\lambda_0}{\mu_{n+1}\mu_n\cdots\mu_1}p_0.$$

Exempel

Biluthyraren Billy har fyra stycken likadana bilar och hyr ut i linjär taxa (så har du bilen i 28 timmar betalar du för $28/24 \approx 1.17$ dygn). Antag att kunder anländer som en Poissonprocess med intensitet $\lambda = 1.8$ per dygn. Varje uthyrning har exponentialfördelad utlåningstid med väntevärde 1.5 dygn och vi antar att olika uthyrningar är oberoende av varandra. Vad är sannolikheten att en kund förloras (på grund av att alla bilar är uthyrda)? Skulle man tjäna på att ha en extra bil för samma kostnad man betalar för att leasa övriga bilar?

Lösning: Vi tänker oss en födelse-döds process $X(t)$ med fem tillstånd (noll till alla fyra bilar uthyrda), där $\lambda_i = 1.8$, $\mu_i = i\mu$ för $i = 0, 1, 2, 3, 4$ och $\mu = \dfrac{1}{1.5}$. Vi skissar processen:



Vi ställer upp uttrycken för den stationära fördelningen:

$$p_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} + \frac{\lambda^3}{3!\mu^3} + \frac{\lambda^4}{4!\mu^4}\right)^{-1} \qquad \text{och} \qquad p_n = \frac{\lambda^n}{n!\mu^n} p_0.$$

Med siffror erhåller vi vektorn

$$p = \begin{pmatrix} 0.0779 & 0.2103 & 0.2839 & 0.2555 & 0.1725 \end{pmatrix}.$$

Det är alltså 17.3% risk att uthyraren förlorar en kund (detta inträffar då processen befinner sig i tillstånd 4 och alla bilar är uthyrda). Det förväntade antalet uthyrda bilar är

$$E(X(t)) = \sum_{k=0}^{4} k p_k = 2.234$$

så förväntad vinst per tidsenhet blir $2.234\alpha - 4\beta$ där α är uthyrningspriset och β är kostnad för en bil.
Vi gör om samma beräkning med fem bilar istället. Det som ändras är att vi har ett extra tillstånd för fem bilar uthyrda. Vi finner nu att

$$p = \begin{pmatrix} 0.0712 & 0.1924 & 0.2597 & 0.2337 & 0.1578 & 0.0852 \end{pmatrix}.$$

Alltså är det nu 8.5% chans att vi förlorar en kund. Bättre, men är det värt priset? Det förväntade antalet uthyrda bilar är

$$E(X(t)) = \sum_{k=0}^{5} k p_k = 2.47$$

så förväntad vinst per tidsenhet blir $2.47\alpha - 5\beta$ med α och β enligt ovan. Svaret på om det är värt priset att skaffa en bil till beror alltså på uthyrningspris och kostnad för bilarna, men vi har nu en uppskattning av på vilket sätt!

11.2 Mer om Exponentialfördelningen

Om vi summerar två oberoende variabler X och Y ges täthetsfunktionen för $Z = X + Y$ av

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx;$$

se föreläsning 4. Antag att $X, Y \sim \mathrm{Exp}(\mu)$ är oberoende. Då blir

$$f_Z(z) = \int_0^z f_X(x) f_Y(z-x)\, dx = \mu^2 e^{-\mu z} \int_0^z e^{-\mu x + \mu x}\, dx = z\mu^2 e^{-\mu z}, \quad z \geq 0.$$

Genom induktion kan man visa att, om $W_1, W_2, \ldots, W_{n+1} \sim \mathrm{Exp}(\mu)$ är oberoende, så har $W = W_1 + W_2 + \cdots W_{n+1}$ täthetsfunktionen

$$f_W(w) = \frac{(\mu w)^n}{n!} \mu e^{-\mu x}, \quad w \geq 0.$$

Denna fördelning brukar kallas **Gammafördelningen**. Ofta ser man $W \sim \Gamma(n+1, \mu)$. Genom partialintegration (i n steg) kan vi räkna ut att

$$F_W(w) = \int_0^w f_W(t)\, dt = 1 - e^{-\mu w} \sum_{k=0}^{n} \frac{(\mu w)^k}{k!}, \quad w \geq 0.$$

Gammafördelning

Sats. Vi skriver att $X \sim \Gamma(\alpha, \mu)$, $\alpha, \mu > 0$, om

$$f_X(x) = \frac{\mu^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\mu x}, \quad x > 0.$$

Variabeln X uppfyller $E(X) = \dfrac{\alpha}{\mu}$ och $V(X) = \dfrac{\alpha}{\mu^2}$.

Här är $\Gamma(\alpha)$ gammafunktionen, och om $\alpha = n \geq 1$ är ett heltal kan den beräknas enligt $\Gamma(n) = (n-1)!$. Observera även att $\Gamma(1, \mu)$ är $\text{Exp}(\mu)$.

Summa av oberoende Exp(μ)-variabler

Sats. Om $W_i \sim \text{Exp}(\mu)$, $i = 1, 2, \ldots, n$, är oberoende så är

$$W = W_1 + W_2 + \cdots + W_n \sim \Gamma(n, \mu).$$

Ibland vill man även ta minimum av exponentialfördelade variabler, och då gäller följande.

Minimum av oberoende Exp(μ_k)-variabler

Sats. Antag att $X_i \sim \text{Exp}(\mu_i)$, $i = 1, 2, \ldots, n$, är oberoende. Då gäller att

$$X = \min\{\, X_1, X_2, \ldots, X_n \,\} \sim \text{Exp}(\mu_1 + \mu_2 + \cdots + \mu_n).$$

Vi låter

$$X = \min\{\, X_1, X_2, \ldots, X_n \,\}.$$

Eftersom variablerna X_i är oberoende erhåller vi

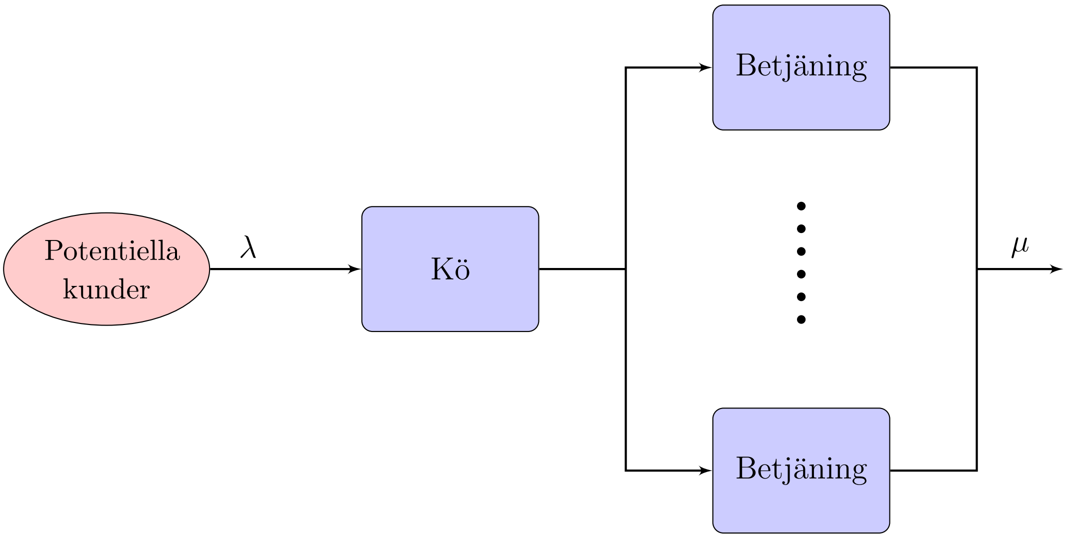
$$\begin{aligned}
F_X(x) &= P(X \leq x) = 1 - P(X > x) = 1 - P(\min\{\, X_1, X_2, \ldots, X_n \,\} > x) \\
&= 1 - P(X_1 > x)P(X_2 > x) \cdots P(X_n > x) = 1 - e^{-\mu_1 x} e^{-\mu_2 x} \cdots e^{-\mu_n x} \\
&= 1 - \exp\left(-x \sum_{k=1}^{n} \mu_k \right),
\end{aligned}$$

vilket är fördelningsfunktionen för en exponentialfördelad varibel.

11.3 Köteori: terminologi och notation

Det naturligaste exemplet på ett kösystem är kanske en affär med ett visst antal kassor. Kunder kommer, ställer sig i kö, betjänas, och går därifrån. Men kömodeller existerar i väldigt många mer abstrakta tillämpningar. Till exempel använder en router för datapaket en kö för att ta mot paket samtidigt som paket vidarebefodras. Vi kommer mest att studera köer i termer av stationära fördelningar $\underline{\pi}$.

Följande illustration beskriver situationen.



Kunder ankommer till ett kösystem med viss intensitet λ, väntar, betjänas och lämnar systemet med intensitet μ.

Vi kommer att beteckna olika kösystem med notationen **A/B/c/K/m/O**, där dessa bokstäver har följande betydelse:

A: Fördelning för tiden mellan kundankomster; typiskt **M** (Markov).
B: Fördelning för betjäningstiden; typiskt **M** (Markov).
c: Antal betjäningsställen.
K: Maximala antalet tillåtna kunder i systemet; typiskt ∞.
m: Maximala antalet kunder i populationen; typiskt ∞.
O: Betjäningsordning; typiskt FIFO (first-in-first-out).

Till exempel **M/M/1** är en kö med "Markovsk" ankomst- och betjäningstid (vilket innebär exponentialfördelade tider). Det innebär även att ankomsterna drivs av en Poissonprocess. Vidare har vi bara ett betjäningsställe. Om bokstäver är utelämnade i slutet antar vi de typiska värderna, så i vårt exempel är $K = m = \infty$ och ordningen FIFO.
För att beskriva trafiken genom systemet introducerar vi följande stokastiska variabler:

$$\begin{aligned}
N_q(t) &= \text{antalet kunder i kön vid tiden } t,\\
N_s(t) &= \text{antalet kunder i betjäning vid tiden } t,\\
N(t) &= \text{totala antalet kunder i systemet vid tiden } t,\\
W &= \text{kötiden för en kund,}\\
S &= \text{betjäningstiden för en kund,}\\
T &= \text{totala tiden i systemet för en kund.}
\end{aligned}$$

Observera att $T = W + S$ samt $N(t) = N_q(t) + N_s(t)$. När vi pratar om kösystem använder vi vissa parametrar för att beskriva trafiken. Intensiteten λ anger **ankomstintensiteten** och $\mu = \dfrac{1}{E(S)}$ anger **betjäningsintensiteten**. Vi använder också **utnyttjandegraden**

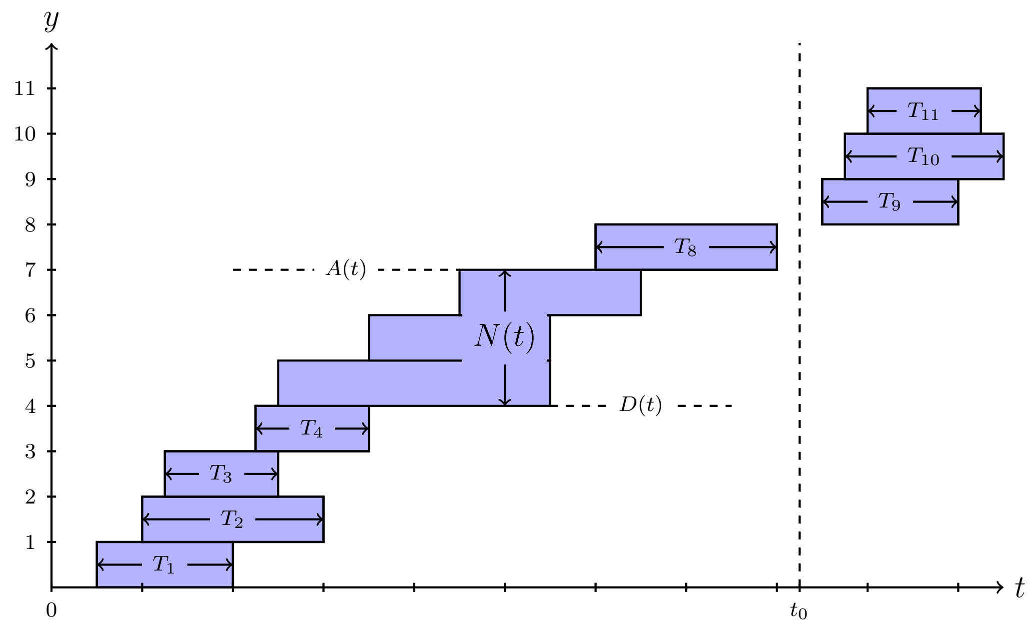
$$\rho = \frac{\lambda E(S)}{c} = \frac{\lambda}{c\mu}.$$

Littles formler

Sats. Med beteckningarna ovan gäller att

$$E(N(t)) = \lambda E(T) \qquad \text{och} \qquad E(N_q(t)) = \lambda E(W).$$

Båda dessa känns inte helt orimliga intuitivt. Det förväntade antalet kunder i systemet borde vara lika med ankomstintensiteten gånger den förväntade tiden varje kund befinner sig i systemet; Alltså $E(N(t)) = \lambda E(T)$. På liknande sätt förefaller Littles andra ekvation rimlig. Lite mer noggrant: vi inför beteckningarna T_i för tiden kund i är i systemet, och funktionerna $A(t)$ och $D(t)$ är antalet ankomna kunder respektive avgågna kunder vi tiden t. Det följer därmed att $N(t) = A(t) - D(t)$ är antalet kunder i systemet vid tiden t. Vi antar att systemet är tomt vid tiden $t = 0$ ($N(0) = 0$) och fixerar en tidpunkt t_0 då vi för enkelhetens skull antar att systemet är tomt igen: $N(t_0) = 0$.



En realisering av den stokastiska processen $\{N(t)\}$.

Den skuggade arean kan beräknas på flera sätt:

$$\int_0^{t_0} N(t)\,dt = \sum_{k=1}^{D(t_0)} T_i = \sum_{k=1}^{A(t_0)} T_i, \tag{1}$$

eftersom $N(t_0) = 0$ så $A(t_0) = D(t_0)$.
Det förefaller rimligt (behöver bevisas) att

$$\frac{1}{t_0} \int_0^{t_0} N(t)\,dt \to E(N(t)), \quad t_0 \to \infty.$$

På samma sätt borde också vara så att, om T är tiden för en kund i systemet,

$$E(T) = \lim_{t_0 \to \infty} \frac{1}{A(t_0)} \sum_{k=1}^{A(t_0)} T_i \qquad \text{och} \qquad \lambda = \lim_{t_0 \to \infty} \frac{A(t_0)}{t_0}.$$

Men detta innebär enligt (1) ovan att

$$E(N(t)) = \lambda E(T),$$

vilket är precis vad vi ville visa. Saken är biff! På liknande sätt kan Littles andra formel "bevisas."

Observera att vi inte gjort några antaganden kring fördelningar och turordning i systemet, så dessa formler gäller i väldigt generella fall. Vi kräver egentligen bara att systemet befinner sig i ett stationärt tillstånd. Kom bara ihåg att λ är den verkliga intensiteten in i systemet, så om man har ett system där kunder kan avvisas måste λ skalas om för att ta hänsyn till detta.

Exempel

Tre inköpare fyller på ett förråd med varor med intensiteterna 24, 48 respektive 36 varor per vecka. Förrådet är stort och blir aldrig fullt. Genom bokföring vet man att det finns i genomsnitt 2800 varor i förrådet. Chefen blir en dag orolig för att varorna ska bli för gamla och frågar statistikern i företaget hur lång tid i medel en vara ligger i förrådet.

Lösning: Frågan verkar vid första anblick svår att svara på, vi har ju nästan ingen information! Hur plockas varor ut (kö-ordning)? Hur ofta? Men faktum är att så länge vi antar att förrådet befinner sig i ett stationärt tillstånd, kan vi använda Littles formel. Total intensitet in i systemet är $\lambda = 24 + 48 + 36 = 108$ varor per vecka och $E(N(t)) = 2800$ varor. Vi erhåller alltså förväntad tid i lagret $E(T) = 2800/108 \approx 28$ veckor.