

Kap 5 Kösystem

F08b

Kap 5.1 Begrepp

Kendalls notation

A / S / N / C / D

OH 14

Här betecknar

A ankomstprocessen ("Arrival process")

S betjäningsprocessen ("Service process")

N antalet betjäningsstationer ("Number of servers")

C kapacitet ("system Capacity")

D ködisciplin ("queue Discipline")

När vi studerar köer är tillståndsrummet alltid hela $E = \{0, 1, 2, \dots\}$ eller (om kapaciteten är begränsad) av typen $E = \{0, 1, 2, \dots, K\}$. Köprocessen betecknas oftast $\{N_t; t \geq 0\}$, där $N_t = j$ betyder att vid tidpunkten t är det j kunder i systemet (inkl de som ev betjänas).

Det som vi ska rikta in oss på i fallet med oändligt tillståndsrum E , villkor för existens av stationär fördelning och hur den isåfall är ser ut. Då tillståndsrummet är ändligt existerar alltid en stationär fördelning och vi ska beräkna dess utseende.

När stationär fördelning $\mathbf{p}$ existerar, gäller

$$p_t(i, j) = \Pr\{N_t = j | N_0 = i\} \rightarrow p(j) \quad \text{då } t \rightarrow \infty$$

Vi ska tänka oss en stokastisk variabel N som representerar antalet kunder i systemet i jämvikt, och som därför är fördelad enligt stationära fördelning $\mathbf{p}$. Vi har alltså

$$p(j) = \Pr\{N = j\}$$

Om antalet betjäningsstationer är c , låter vi

$$N_q = \begin{cases} 0 & \text{då } N \leq c \\ N - c & \text{då } N > c \end{cases}$$

representera antalet kunder i kön i jämvikt.

I det som följer ska vi studera system av typen M/M/ c / K , där $1 \leq c \leq \infty$ är antalet betjäningsstationer och $1 \leq K \leq \infty$ är maximala antalet kunder i kön (obs inkl de som är under betjäning). Ködisciplinen är hela tiden FIFO.

Kap 5.2 En betjäningsstation

M/M/1

M/M/1 är en förkortning av M/M/1/ ∞ /FIFO.

Vi har alltså $c = 1$ betjäningsstation och $K = \infty$ betyder att det inte finns någon begränsning av antalet kunder i systemet.

Första M:et i M/M/1 betyder att ankomstprocessen är Poisson med intensitet λ , vilket betyder att tiderna mellan ankomster är oberoende och $\text{Exp}(1/\lambda)$ -fördelade. Ankomster sker alltså enl en Poissonprocess med intensiteten λ .

Andra M:et i M/M/1 betyder att betjädningsprocessen definieras av att alla servicetider är oberoende $\text{Exp}(1/\mu)$ -fördelade stokastiska variabler.

Rita tillståndsdigram.

Om tillståndet $j = 0$ så pågår bara ankomstprocessen. Uppehållsintensiteten då $j = 0$ är därför λ . Raden i generatoren $\mathbf{G}$ svarande mot tillståndet $j = 0$ är därför

$$[-\lambda \quad \lambda \quad 0 \quad \dots]$$

Och om tillståndet $j > 0$ så pågår både ankomstprocessen och en betjädningsprocess (vi har ju bara en betjädningsstation). Uppehållsintensiteten då $j > 0$ är därför $\mu + \lambda$. Tillståndsförändringen $j \rightarrow j - 1$ sker med intensiteten μ , och tillståndsförändringen $j \rightarrow j + 1$ sker med intensiteten λ . Raden i $\mathbf{G}$ svarande mot tillståndet $j > 0$ är därför

$$[\dots \quad 0 \quad \mu \quad -(\mu + \lambda) \quad \lambda \quad 0 \quad \dots]$$

Vi har alltså

$$\mathbf{G} = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \mu & -(\mu + \lambda) & \lambda & 0 & 0 & \\ 0 & \mu & -(\mu + \lambda) & \lambda & 0 & \\ \vdots & & & & & \end{bmatrix}$$

och $\mathbf{pG} = \mathbf{0}$ ger oss ekvationssystemet

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0 \\ \lambda p_0 - (\mu + \lambda)p_1 + \mu p_2 = 0 \\ \lambda p_1 - (\mu + \lambda)p_2 + \mu p_3 = 0 \\ \vdots \end{cases}$$

som vi ska lösa rekursivt.

Men innan dess ska vi skriva om ekvationssystemet och relatera det till tillståndsdigrammet. Ekvationssystemet kan skrivas

$$\begin{cases} \mu p_1 &= \lambda p_0 \\ \lambda p_0 + \mu p_2 &= (\mu + \lambda)p_1 \\ \lambda p_1 + \mu p_3 &= (\mu + \lambda)p_2 \\ &\vdots \end{cases}$$

och vi ska tänka på termerna på vänster sida som flödet in i ett tillstånd och på termerna på höger sida som flödet ut ur tillståndet. Jämför med tillståndsdigrammet för M/M/1-systemet. Skrivna på detta sätt, brukar ekvationerna kallas för *balansekvationerna*.

Den första balansekvationen ger

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

Den andra ger därför

$$p_2 = \frac{\mu + \lambda}{\mu} p_1 - \frac{\lambda}{\mu} p_0 = \frac{\mu + \lambda}{\mu} \cdot \frac{\lambda}{\mu} p_0 - \frac{\lambda}{\mu} \cdot \frac{\mu}{\mu} p_0 = \frac{\lambda^2}{\mu^2} p_0$$

Analogt ger den tredje att

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

etc. Vi sluter oss till att

$$p_n = \frac{\lambda^n}{\mu^n} p_0$$

Ur $\sum_n p_n = 1$, får vi nu

$$1 = \sum_n \frac{\lambda^n}{\mu^n} p_0 = p_0 \sum_n \left(\frac{\lambda}{\mu} \right)^n = p_0 \sum_n \rho^n$$

där $\rho = \lambda/\mu$. Serien till höger är geometrisk och en konvergerar om, och endast om, dess kvot $\rho < 1$. Om så är fallet följer

$$1 = p_0 \frac{1}{1 - \rho} \Rightarrow p_0 = 1 - \rho$$

Vi har visat att stationär fördelning finns om, och endast om, $\rho < 1$ och då gäller

$$p_n = \Pr\{N = n\} = \rho^n (1 - \rho), \quad n = 0, 1, \dots$$

Observera att

F09

$$P\{N + 1 = k\} = P\{N = k - 1\} = \rho^{k-1} (1 - \rho)$$

för $k = 1, 2, \dots$. Alltså, $N + 1 \sim \text{Geo}(1 - \rho)$.

Ur formelsamlingen får vi att $E[N + 1] = 1/(1 - \rho)$, vilket medför att i genomsnitt befinner sig

$$L = E[N] = \frac{1}{1 - \rho} - 1 = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda}$$

kunder i systemet (i jämnvikt).

Innan vi studerar tiden en kund tillbringar i M/M/1-kön, ska vi lära oss lite om betingade väntevärden. Vi nöjer oss med det diskreta fallet. (De exakta definitionerna är lite svårare att skriva ned i det kontinuerliga fallet, fast intuitivt analoga.)

Vi definierar det betingade väntevärdet av X givet A enl,

$$E[X|A] = \sum_k k \Pr\{X = k|A\}$$

Notera att om X och A är oberoende, så gäller att $E[X|A] = E[X]$.

Låt nu N vara en annan diskret stokastisk variabel. Enl lagen om total sannolikhet gäller att

$$\Pr\{X = k\} = \sum_n \Pr\{X = k|N = n\} \Pr\{N = n\}$$

Härur följer att

$$\begin{aligned} E[X] &= \sum_k k \Pr\{X = k\} \\ &= \sum_k k \sum_n \Pr\{X = k|N = n\} \Pr\{N = n\} \\ &= \sum_n \left(\sum_k k \Pr\{X = k|N = n\} \right) \Pr\{N = n\} \\ &= \sum_n E[X|N = n] \Pr\{N = n\} \end{aligned}$$

Man kan alltså räkna ut väntevärdet av en stokastisk genom att betinga map resultatet av en annan stokastisk variabel och väga ihop de betingade väntevärdena. Samma räkneregler gäller då X är kontinuerlig (fast vi avstår ifrån att skriva ned den exakta matematiska definitionen av det betingade väntevärdet för sådana X) och då även N är kontinuerlig gäller

$$E[X] = \int_{-\infty}^{\infty} E[X|N = n] f(n) dn$$

där vi skrivit $f(n)$ för den kontinuerliga tätheten för N . Vi ska nu använda oss av detta trick.

Låt W vara väntevärdet av den tid en godtycklig kund tillbringar i systemet. Då kunden anländer till systemet finns i jämvikt N st kunder före i systemet, varav en är under betjäning om $N > 0$. På grund av Markovegenskapen blir kundens totala tid i systemet därför en summa av $N + 1$ oberoende $\text{Exp}(1/\mu)$ -variabler S_i . Väntevärdet av denna summa är

$$\begin{aligned} W &= E \left[\sum_{i=1}^{N+1} S_i \right] = \sum_n E \left[\sum_{i=1}^{N+1} S_i | N = n \right] \Pr\{N = n\} \\ &= \sum_n E \left[\sum_{i=1}^{n+1} S_i | N = n \right] \Pr\{N = n\} = \sum_n \sum_{i=1}^{n+1} E[S_i | N = n] \Pr\{N = n\} \end{aligned}$$

Av oberoendet mellan S_i och N följer nu att

$$\begin{aligned} W &= \sum_n \left(\sum_{i=1}^{n+1} E[S_i] \right) \Pr\{N = n\} = \sum_n (n+1) E[S] \Pr\{N = n\} \\ &= E[N+1] E[S] = \frac{1}{1-\rho} \cdot \frac{1}{\mu} = \frac{1}{\mu-\lambda} = \frac{L}{\lambda} \end{aligned}$$

Vi har härlett formeln

$$L = \lambda W$$

Detta är ett berömt köteoriresultat, som kallas *Littles formel* efter J. D. C. Little som visade att det gäller för väldigt allmänna kösystem. Det brukar formuleras så här:

$$L = \lambda_e W$$

där λ_e är den effektiva ankomstintensiteten. Det är nämligen så att i kösystem med begränsad kapacitet avvisas alla kunder som anländer då systemet är fullt, och det som påverkar tiden inuti systemet måste ju vara den ankomsterna av kunder som tas emot. För M/M/1-kön är $\lambda_e = \lambda$.

Låt W_q vara den genomsnittliga kötiden. Eftersom totala tiden i systemet T kan delas i tiden i kö T_q och tiden under betjäning, som har väntevärdet $1/\mu$, följer att

$$W = W_q + \frac{1}{\mu}$$

och vi ser att

$$W_q = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \rho \cdot \frac{1}{\mu - \lambda} = \rho W$$

Littles formel gäller även för paret L_q, W_q . Vi får därför

$$L_q = E[N_q] = \lambda W_q = \lambda \cdot \rho \cdot \frac{1}{\mu - \lambda} = \frac{\rho^2}{1 - \rho} = \rho L$$

Sammanfattningsvis konstaterar vi att stationär fördelning existerar precis då $\rho = \lambda/\mu < 1$. Man brukar kalla ρ för *trafikintensiteten*. Sannolikheten $1 - p_0 = \Pr\{N > 0\}$ brukar kallas för *betjäningsfaktorn*. För M/M/1-kön är $1 - p_0 = \rho$.

Vi har vidare sett att i jämvikt finns i genomsnitt

$$L = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda}$$

kunder i systemet, och deras genomsnittliga väntetid är

$$W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}$$

De i genomsnitt $L = E[N]$ kunderna i systemet delas upp i $L_q = E[N_q] = \rho L$ st i kön, och $(1 - \rho)L$ st under betjäning. Totala förväntade tiden i systemet, W , delas upp i $W_q = \rho W$ tidsenheter i kön, och $(1 - \rho)W = 1/\mu$ tidsenheter i betjäning.

Exempel 5.2 Ett företag som säljer och reparerar bilar, lånar ut en reservbil till OH 15 de kunder som har problem som behöver åtgärdas med sina bilar. Man uppskattar kostnaden för detta till $c = 10$ USD per bil och dag. I snitt får man in 1 bil som behöver åtgärdas varannan dag (så $\lambda = 1/2$). Tiden som mekanikern använder på en bil har medelvärde $1/\mu = 1.6$ dagar (så $\mu = 1/1.6 = 5/8$).

Antag Markovska ankomster och betjäningar.

Hur mycket kostar bilutlåningen per tidsenhet?

Hur mycket kostar bilutlåningen per bil?

Vi såg i exempel 5.2 att om man har kostnaden c USD per kund och tidsenhet, så blir genomsnittskostnaden i jämnvikt

$$E[C] = cL \text{ USD per tidsenhet}$$

och att detta svarar mot genomsnittskostnaden

$$d = cW \text{ USD per kund}$$

M/M/1/K

Rita tillståndsdigram. Tillståndsrummet är

$$E = \{0, 1, \dots, K\}$$

Balansekvationerna är

$$\left\{ \begin{array}{rcl} \mu p_1 & = & \lambda p_0 \\ \lambda p_0 + \mu p_2 & = & (\mu + \lambda)p_1 \\ \vdots & & \\ \lambda p_{j-1} + \mu p_{j+1} & = & (\mu + \lambda)p_j \\ \vdots & & \\ \lambda p_{K-2} + \mu p_K & = & (\mu + \lambda)p_{K-1} \\ \lambda p_{K-1} & = & \mu p_K \end{array} \right.$$

Sätt $\rho = \lambda/\mu$. Vi erhåller rekursivt, att

$$\left\{ \begin{array}{l} p_1 = \rho p_0 \\ p_2 = \rho^2 p_0 \\ p_3 = \rho^3 p_0 \\ \vdots \\ p_K = \rho^K p_0 \end{array} \right.$$

(jämför med M/M/1-fallet). Villkoret $\sum_n p_n = 1$ ger nu

$$1 = \sum_{n=0}^K \rho^n p_0 = p_0 \frac{1 - \rho^{K+1}}{1 - \rho}$$

om $\rho \neq 1$, och om $\rho = 1$,

$$1 = \sum_{n=0}^K \rho^n p_0 = (K+1)p_0$$

Stationära fördelningen är alltså

$$p_n = \left\{ \begin{array}{ll} \frac{\rho^n(1 - \rho)}{1 - \rho^{K+1}} & \text{då } \rho \neq 1 \\ \frac{1}{K+1} & \text{då } \rho = 1 \end{array} \right.$$

Sannolikheten

F10

$$p_K = \Pr\{N = K\} = \begin{cases} \frac{\rho^K(1-\rho)}{1-\rho^{K+1}} & \text{då } \rho \neq 1 \\ \frac{1}{K+1} & \text{då } \rho = 1 \end{cases}$$

talas om hur stor del av den totala tiden som systemet är blockerat. Den effektiva ankomstintensiteten för M/M/1/K-kön blir därför

$$\lambda_e = \lambda(1 - p_K)$$

Det är den effektiva ankomstintensiteten som ska in i Littles formel:

$$L = \lambda_e W$$

eller

$$W = L/\lambda_e$$

Notera att

$$\rho \neq 1 \Rightarrow 1 - p_K = \frac{1 - \rho^K}{1 - \rho^{K+1}} \Rightarrow \lambda_e = \lambda \frac{1 - \rho^K}{1 - \rho^{K+1}}$$

medan

$$\rho = 1 \Rightarrow 1 - p_K = 1 - \frac{1}{K+1} = \frac{K}{K+1} \Rightarrow \lambda_e = \lambda \frac{K}{K+1}$$

Då $\rho = 1$ har N den likformiga fördelningen på $\{0, 1, \dots, K\}$, så

$$L = E[N] = \frac{K}{2}$$

i detta fall. Littles formel ger

$$W = \frac{L}{\lambda_e} = \frac{K+1}{2\lambda}$$

Då $\rho \neq 1$ följer

$$L = E[N] = \sum_{n=0}^K np_n = \dots = \frac{\rho[1 + K\rho^{K+1} - (K+1)\rho^K]}{(1-\rho)(1-\rho^{K+1})}$$

och medelst division med

$$\lambda_e = \lambda(1 - p_K) = \lambda \frac{1 - \rho^K}{1 - \rho^{K+1}}$$

får vi värdet av W .

Sedan fås W_q ur

$$W = W_q + \frac{1}{\mu}$$

och L_q ur Littles formel

$$L_q = \lambda_e W_q$$

Ofta genererar de missade kunderna en kostnad eller missad intäkt. Vi vet att i jämnvikt är systemet stängd i p_K tidsenheter och öppet för ankomster i $1 - p_K$. Ur detta följer att det genomsnittliga antalet missade kunder är λp_K och om man har en kostnad om USD d per missad kund, så kommer man att ha utgiften USD

$$E[D] = \lambda p_K d$$

per tidsenhet i snitt i jämnvikt.

Exempel 5.3 Reparation av traktorer. I snitt ankommer 3 st/vecka, så $\lambda = 3$. I snitt OH 16 är reparationstiden $1/2$ vecka, så $\mu = 2$. Just nu kan man ha högst 2 st i garaget, varav en är under reparation, så $K = 2$. Varje traktor i reparationsgaraget kostar USD $c = 100$ per vecka. Varje traktor som man ej kan ta emot kostar USD $d = 500$. Antag Markovska ankomster och betjäningar.

Vad har man för genomsnittskostnad?

Om man ökar K , minskar då genomsnittskostnaden?

Kap 5.3 Flera betjäningsstationer

Den allmänna födelse- och dödsprocessen

Gjorde ej allmän födelse- och dödsprocess på föreläsning. Ritade istället upp tillståndsdigram för M/M/2/4-kön, skrev upp och löste balansekvationerna.

Rita upp tillståndsdigrammet för en allmän födelse- och dödsprocess.

Födelseintensiteterna betecknas $\lambda_0, \lambda_1, \lambda_2, \dots$ och dödsintensiteterna betecknas $\mu_1, \mu_2, \dots$ (observera att $\mu_0 = 0$, ty det kan ju inte försvinna någon ur ett tomt system).

Balansekvationerna är

$$\begin{cases} \mu_1 p_1 &= \lambda_0 p_0 \\ \lambda_0 p_0 + \mu_2 p_2 &= (\mu_1 + \lambda_1) p_1 \\ \lambda_1 p_1 + \mu_3 p_3 &= (\mu_2 + \lambda_2) p_2 \\ &\vdots \end{cases}$$

Medelst successiv substitution fås

$$\begin{cases} p_1 &= p_0 \frac{\lambda_0}{\mu_1} \\ p_2 &= p_0 \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \\ &\vdots \end{cases}$$

Normeringen $\sum_n p_n = 1$ ger nu

$$1 = p_0 + p_0 \frac{\lambda_0}{\mu_1} + p_0 \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \dots = p_0 \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \right)$$

Stationär fördelning existerar alltså om, och endast om,

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} < \infty$$

och sannolikheten p_n fås ur

$$p_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} p_0$$

där

$$p_0 = \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \right)^{-1}$$

Exempel 5.4 Referera till exempel 5.3. Reparation av traktorer.

OH 17

Det ursprungliga kösystemet var ett M/M/1/2 med genomsnittskostnaden USD 836.84 per vecka och räknade fram att ett M/M/1/3 skulle få den lägre genomsnittskostnaden USD 821.54 per vecka.

Att anställa ytterligare en mekaniker kostar USD 400 per vecka¹. Kommer det att ge en lägre genomsnittskostnad?

Exempel 5.5 Man producerar något med tre maskiner som går parallellt. Då och då

OH 18

går någon ned och detta sker för varje maskin med intensiteten $\lambda = 1$ gång per vecka.

Man har en reparatör som i snitt klarar av att reparera $\mu = 2$ maskiner per vecka. Etc.

M/M/c för $c < \infty$

Gjorde ej M/M/c-kön på föreläsning. Refererade till formler i läroboken.

För M/M/c-kön gäller att

$$\lambda_n = \lambda, \quad n = 0, 1, \dots$$

$$\mu_n = \begin{cases} n\mu & \text{då } n = 1, \dots, c \\ c\mu & \text{då } n = c, c+1, \dots \end{cases}$$

Vi får genom att relatera till en födelse- och dödsprocess med dessa intensiteter, att

$$p_n = \begin{cases} p_0 \frac{\lambda^n}{n! \mu^n} & = p_0 \frac{r^n}{n!} \quad \text{då } n = 1, \dots, c \\ p_0 \frac{\lambda^n}{c! c^{n-c} \mu^n} & = p_0 \frac{c^c}{c!} \rho^n \quad \text{då } n = c, c+1, \dots \end{cases}$$

där $r = \lambda/\mu$ och $\rho = r/c$. Sannolikheten att systemet är tomt fås ur

$$1 = p_0 \left(1 + \sum_{n=1}^{c-1} \frac{r^n}{n!} + \sum_{n=c}^{\infty} \frac{c^c}{c!} \rho^n \right) = p_0 \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!(1-\rho)} \right)$$

¹Vi ändrar från USD 600 per vecka i F & V-F

Den sista likheten gäller bara om $\rho < 1$. Så stationär fördelning existerar precis då $\rho < 1$ och då gäller att

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!(1-\rho)} \right)^{-1}$$

Man kan visa att

$$L_q = \sum_{n=c}^{\infty} (n-c)p_n = \frac{p_0 r^c \rho}{c!(1-\rho)^2}$$

Sedan följer av Littles formel att

$$W_q = \frac{L_q}{\lambda}$$

(Observera att $\lambda_e = \lambda$.) Sedan kan man beräkna

$$W = W_q + \frac{1}{\mu}$$

och

$$L = \lambda W = L_q + r$$