Basic queues

COMPSCI 655
Lecture 7
Seen last lecture: CTMC

\{X(t), t \geq 0\}, X(t) \in S, CTMC \text{ if}

\[ P(X(t)=j | X(s_1)=i_1, \ldots, X(s_k)=i_k, X(s)=i) \]

\[ = P(X(t)=j | X(s)=i) \quad \text{(Markov property)} \]

for all \(0 \leq s_1 < \ldots < s_k < s < t, i_1, \ldots, i_k, i, j \in S, k=1,2,\ldots\)

\[ = F_{i,j} (t-s) \quad \text{(homogeneity)} \]

with infinitesimal generator \(Q = [q_{i,j}]\)

\[ Q = \lim_{u \to 0} \frac{(F(u)-I)}{h}, \quad F(u) = [F_{i,j}(u)] \]

\((\Sigma_{i,i} q_{i,i} = 0)\)
Seen last lecture: CTMC (cont')

If \( \{X(t), \, t \geq 0\} \) irreducible CTMC on \( S \) with infinitesimal generator \( Q = [q_{i,j}] \) and if eqns

\[
x \cdot Q = 0, \quad x \cdot 1 = 1
\]

have unique, strictly positive, solution \( x = (x(i), \, i \in S) \), then \( x(i) = \lim_{t \to \infty} P(X(t) = i), \quad i \in S \), regardless of \( X(0) \)

\( q_{i,j} \) = transition rate from state \( i \) to state \( j \neq i \)

Interpretation of \( x \cdot Q = 0 \) (equilibrium eqns):

\[
\text{Prob. flow out a state} = \text{Prob. flow into a state}
\]
Construction rule for checking if \( \{X(t)\}_t \) is CTMC

When process enters state \( i \in S \) at time \( t \) then

- for each \( j \neq i \) an exponential rv \( Y_{i,j} \) with rate \( \eta_{i,j} \) is generated. These rvs are mutually independent and independent of what happened before \( t \).

One may have \( \eta_{i,j} = 0 \) in which case \( Y_{i,j} = \infty \) (i.e. state \( j \) cannot be reached in one transition from state \( i \))

- at time \( t + Y_{i,k} \) with \( Y_{i,k} = \min\{Y_{i,j}\}_j \) process instantaneously jumps into state \( k \)

If true, then \( \{X(t)\}_t \) is CTMC with \( Q = [q_{i,j}] \) given by

\[
q_{i,j} = \begin{cases} 
\eta_{i,j} & \text{if } i \neq j \\
-\sum_{l \neq i} \eta_{i,l} & \text{if } i = j 
\end{cases}
\]
Queues

Single server queue

Multiple servers and queues
Queues (cont’)

Typical performance metrics of interest

• Condition of stability
• Expected number of customers/jobs/etc. in system
• Expected waiting time in queue
• Expected system response time
• System throughput
• Etc.
Queue represented by Kendall’s notation: A/S/c/K/D

- **A** = nature of arrival process (e.g. A=M if Poisson)
- **S** = nature of service time process (S=M if expon.)
- **c** = number of servers
- **K** = size of queue including # servers (default=infinite)
- **D** = service discipline (default = FIFO)

Unless otherwise mentioned interarrival times and service times are independent iid sequences
Classic queues

- **M/M/1**
Poisson arrivals, exponential service times, 1 server, infinite waiting room

- **M/M/1/K**
Poisson arrivals, exponential service times, 1 server, waiting room of size K-1

- **M/M/c**
Poisson arrivals, exponential service times, c servers, infinite waiting room

- **M/M/c/c**
Poisson arrivals, generally distributed service times, c servers, no waiting room
Classic queues (con't)

- **M/M/∞**
  Poisson arrivals, exponential service times, infinitely many servers, no waiting room

- **M/G/1**
  Poisson arrivals, generally distributed service times, 1 server, infinite waiting room

- **M/D/1**
  Poisson arrivals, constant service time all equal to D, 1 server, infinite waiting room

- **G/M/1**
  Arbitrary interarrival times, exponential service times, 1 server, infinite waiting room
Classic queues (con’t)

Most common service disciplines:
FIFO = First In First Out
LIFO = Last In First Out (stack)
PS = Processor Sharing
**M/M/1 queue**

- Arrival times \( \{t_n\}_n \) are Poisson with rate \( \lambda \)
- Service times exponentially distributed with rate \( \mu \)

Namely, with \( \tau_n := t_{n+1} - t_n \) and \( \sigma_n = \text{time needed to serve nth customer} \), then

\[
P(\tau_n < x) = 1 - e^{-\lambda x} \quad P(\sigma_n < x) = 1 - e^{-\mu x}
\]

with \( \{\tau_n\}_n \) and \( \{\mu_n\}_n \) mutually ind. iid sequences of rvs
M/M/1 queue

\( X(t) = \# \) customers in system at time \( t \), including the one in service, if any

- When does system is stable? (= queue does not build up)
- \( \lim_{t \to \infty} P(X(t) = n) = ? \)

- System throughput?

\[
\begin{array}{c}
\text{Poisson} \\
\text{Exp.}
\end{array}
\]

Infinite waiting room
Claim: \( \{X(t), t \geq 0\} \) is a CTMC with state-space \( S = \{0, 1, 2, \ldots\} \)

Let us see if construction rule applies

- Assume first that \( X(t) = 0 \)

From state 0 one can only go to state 1 (arrival). Memoryless feature of Poisson implies that next customer will arrive after a time exponentially distributed with rate \( \lambda \)

\( \Rightarrow \) Construction rule applies to state 0 with \( \eta_{0,1} = \lambda \) and \( \eta_{0,j} = 0 \) for \( j > 1 \)
Assume $X(t) = i > 0$

From state $i > 0$ one can either go to state $i+1$ (arrival) or to state $i-1$ (departure)

Time before next arrival is exponentially distributed with rate $\lambda$ (memoryless property of interarrival times)

Time before next departure is exponentially distributed with rate $\mu$ (memoryless property of service times)

Whichever event occurs first, will trigger the next event. Hence, time before next event is min of two independent exponential rvs with rate $\lambda$ and $\mu$

$\Rightarrow$ Construction rule applies to state $i > 0$ with $\eta_{i,i+1} = \lambda$, $\eta_{i,i-1} = \mu$, $\eta_{i,j} = 0$ for $j \neq i-1, i$
Rate transition diagram for $M/M/1$ queue

Irreducible as one can go from any state $i$ to any other state $j \neq i$ with positive prob.

E.g. if $j = i+2$ then we can go from $i$ to $i+2$ if there are no departure and two arrivals
**M/M/1 queue (cont')**

Infinitesimal generator

\[
Q = \begin{pmatrix}
-\lambda & \lambda & 0 & 0 & \cdots \\
\mu & -(\lambda + \mu) & \lambda & 0 & \cdots \\
0 & \mu & -(\lambda + \mu) & \lambda & 0 \\
0 & 0 & \mu & -(\lambda + \mu) & \lambda \\
\end{pmatrix}
\]
Balance eqns for M/M/1 queue

$$\lambda x_0 = \mu x_1$$

$$(\lambda + \mu) x_i = \lambda x_{i-1} + \mu x_{i+1}, \ i > 0$$

Easy to find that $x_i = x_0 \rho^i$ for $i > 0$

$x.1 = 1$ gives $x_0 \sum_{i \geq 0} \rho^i = 1$ with $\rho := \lambda/\mu$

- If $\rho < 1$ then $\sum_{i \geq 0} \rho^i = 1/(1-\rho)$, $x_0 = 1-\rho$ and $x_i = (1-\rho)\rho^i > 0$, $i=0,1,...$

**Unique strictly positive sol. $\Rightarrow$ CMTQ Thm applies**

- $\pi_i = \lim_t P(X(t) = i) = (1-\rho)\rho^i$, $i \geq 0$ - Geometric distr.
  $\rho < 1$ called stability condition (makes sense)

- If $\rho \geq 1$ only solution is $\pi_i = 0$ for all $i$
  Thm does not apply. Can be shown syst. unstable
Alternative approach for \(M/M/1\) queue

\(\{X(t)\}_t\) is a birth and death process with birth rate \(\lambda\) in any state and death rate \(\mu\) in any state \(i \neq 0\)
Assume $\rho < 1$ (queue stable)

- Queue length distribution: $\pi_i = (1-\rho)\rho^i$, $i \geq 0$

- System throughput: $\mu(1 - \pi_0) = \mu \rho = \lambda$ \hspace{1em} ($\rho = \lambda/\mu$)

Note: makes sense that when system stable input rate ($\lambda$) = output rate

- Expected queue-length: $E[X] = (1-\rho)\sum_{i \geq 1} i\rho^i$

\[ F(z) = \sum_{i \geq 0} z^i = 1/(1-z), \quad F'(z) = \sum_{i \geq 1} iz^{i-1} = 1/(1-z)^2 \]

$\Rightarrow \sum_{i \geq 1} i\rho^i = \rho/(1-\rho)^2$ \hspace{1em} so that $E[X] = \rho/(1-\rho)$
$E_{M/M/1}[X]$
$P(X \geq K)$?

$P(X < K) = \sum_{i=0}^{K-1} \pi_i = (1 - \rho) \sum_{i=0}^{K-1} \rho^i = (1 - \rho) \frac{1 - \rho^K}{1 - \rho} = 1 - \rho^K$

$P(X \geq K) = \rho^K$
M/M/1/K queue

Poisson

Lost when full

Waiting room of size K-1

Exp.
$X(t) \in S = \{0, 1, \ldots, K\}$ # customers in system at time $t$, including the one in service, if any

{\{X(t)\}_t}$ CMTT as construction rule applies with

- if $X(t) = 0$: $\eta_{0,1} = \lambda$, $\eta_{0,j} = 0$ for $j \neq 1$

- if $0 \leq X(t) = i \leq K-1$: $\eta_{i,i+1} = \lambda$, $\eta_{i,i-1} = \mu$, $\eta_{i,j} = 0$ for $j \neq i-1, i+1$

- if $X(t) = K$: $\eta_{K,K-1} = \mu$, $\eta_{K,j} = 0$ for $j \neq K-1$

Balance eqns:

\[
\begin{align*}
\lambda \pi_0 &= \mu \pi_1 \\
(\lambda + \mu) \pi_i &= \lambda \pi_{i-1} + \mu \pi_{i+1}, \quad i = 1, \ldots, K-1 \\
\mu \pi_K &= \lambda \pi_{K-1}
\end{align*}
\]
Balance eqns:  
\[ \lambda x_0 = \mu x_1 \]
\[ (\lambda + \mu) x_i = \lambda x_i + \mu x_i, \quad i = 1, \ldots, K-1 \]
\[ \lambda x_{K-1} = \mu x_K \]

Unique normalized (i.e. \(x.1 = 1\)) solution
\[ x_i = \rho^i(1-\rho)/(1-\rho^{K+1}), \quad i = 0,1,\ldots,K \]

Strictly positive (for all \(\rho\)) + \{X(t)\}_t irreducible yields by CMTQ theorem:

\[ \pi_i = \lim_{t \to \infty} P(X(t)=i) = \rho^i (1-\rho)/(1-\rho^{K+1}), \quad i = 0,1,\ldots,K \]

Always stable (makes sense, as system finite)

Note: \(K \to \infty\) and \(\rho < 1\) one retrieves \(M/M/1\) (as \(\rho^{K+1} \to 0\))
\[ \pi_i = \rho^i \frac{(1-\rho)}{(1-\rho^{K+1})}, \ i = 0,1,\ldots,K \]

\[ E[X] = \sum_{i=0}^{K} i\pi_i = \frac{1-\rho}{1-\rho^{K+1}} \sum_{i=0}^{K} i\rho^i = \rho \frac{1-(K+1)\rho^K + K\rho^{K+1}}{(1-\rho)(1-\rho^{K+1})} \]

Make use of
\[ \frac{d}{dz} \left( \sum_{i=0}^{K} z^i \right) = \frac{d}{dz} \left( \frac{1-z^{K+1}}{1-z} \right) = \sum_{i=1}^{K} iz^{i-1} \]
$E_{M/M/1/K}[X]$
$\pi_i = \rho^i (1-\rho)/(1-\rho^{K+1})$, $i = 0,1,...,K$

- Prob. arriving customer lost = $\pi_K$
  
  $$= \frac{\rho^K(1-\rho)}{(1-\rho^{K+1})}$$

- Throughput = $\mu(1-\pi_0) = \mu \left[1 - \frac{(1-\rho)}{(1-\rho^{K+1})}\right]$

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

  $$= \lambda (1-\pi_K)$$
M/M/c queue

Poisson → Infinite waiting room

Exp

1 → 2 → c
\( M/M/c \) queue (cont’)

\( X(t) \in S = \{0,1,...\} \) # customers in system at time \( t \), including the one in service, if any

\[ \{X(t)\}_t \text{ CMTC as construction rule applies with} \]

- if \( X(t) = 0 \): \( \eta_{0,1} = \lambda, \eta_{0,j} = 0 \) for \( j \neq 1 \)
- if \( 1 \leq X(t) = i \leq c \): \( \eta_{i,i+1} = \lambda, \eta_{i,i-1} = i \mu, \eta_{i,j} = 0 \) for \( j \neq i-1, i+1 \)
- if \( X(t) = i > c \): \( \eta_{i,i+1} = \lambda, \eta_{i,i-1} = c \mu, \eta_{i,j} = 0 \) for \( j \neq i-1, i+1 \)

Balance eqns:

\[
\begin{align*}
\lambda \pi_0 &= \mu \pi_1 \\
(\lambda + i \mu) \pi_i &= \lambda \pi_{i-1} + (i+1) \mu \pi_{i+1}, \quad i = 1, \ldots, c-1 \\
(\lambda + c \mu) \pi_i &= \lambda \pi_{i-1} + c \mu \pi_{i+1}, \quad i \geq c
\end{align*}
\]
Balance eqns: \[ \lambda x_0 = \mu x_1 \]
\[ (\lambda + i\mu) x_i = \lambda x_{i-1} + (i+1)\mu x_{i+1}, \quad i = 1, \ldots, c-1 \]
\[ (\lambda + c\mu) x_i = \lambda x_{i-1} + c\mu x_{i+1}, \quad i \geq c \]

Unique solution
\[ x_i = x_0 \rho^i / i! \quad \text{if } i = 0, 1, \ldots, c \]
\[ = x_0 (\rho/c)^i c^c / i! \quad \text{if } i > c \]

x.1=1 gives \[ x_0 = \begin{cases} 0 & \text{if } \rho \geq c \\ \left[ \sum_{i=0}^{c-1} \frac{\rho^i}{i!} + \frac{\rho^c}{c!} \left( \frac{1}{1 - \rho / c} \right) \right]^{-1} & \text{if } \rho < c \end{cases} \]

Strictly positive solution only if \( \rho < c \)

CMTC thm: \( \pi_i = \lim_{t \to \infty} P(X(t) = i) = x_0 \rho^i / i! \quad \text{if } i = 0, 1, \ldots, c \]
\[ = x_0 (\rho/c)^i c^c / i! \quad \text{if } i > c \]
when \( \rho < c \) -- Stability condition of M/M/c queue