Queuing Theory

Richard Lockhart

Simon Fraser University

STAT 870 — Summer 2011

Purposes of Today's Lecture

- Describe general queuing theory jargon.
- Establish relation of some queues to Markov Chains.

Queuing Theory

- Ingredients of Queuing Problem:
  1. Queue input process.
  2. Number of servers
  3. Queue discipline: first come first serve? last in first out? pre-emptive priorities?
  4. Service time distribution.
- Example: Imagine customers arriving at a facility at times of a Poisson Process N with rate λ.
- This is the input process, denoted M (for Markov) in queuing literature.

Single server case

- Service distribution: exponential service times, rate μ.
- Queue discipline: first come first serve.
- $X(t)$ = number of customers in line at time t.
- X is a Markov process called $M/M/1$ queue:

$$v_i = \lambda + \mu 1(i > 0)$$

$$\mathbf{P}_{ij} = \begin{cases} \frac{\mu}{\mu+\lambda} & j = i - 1 \geq 0 \\ \frac{\lambda}{\mu+\lambda} & j = i + 1, i > 0 \\ 1 & j = 1, i = 0 \\ 0 & \text{otherwise} \end{cases}$$

Example: $M/M/\infty$ queue

- Customers arrive according to PP rate λ.
- Each customer begins service immediately.
- $X(t)$ is number being served at time t.
- X is a birth and death process with

$$v_n = \lambda + n\mu$$

and

$$\mathbf{P}_{ij} = \begin{cases} \frac{i\mu}{i\mu+\lambda} & j = i-1 \geq 0 \\ \frac{\lambda}{i\mu+\lambda} & j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Stationary distributions

- For $M/M/1$ queue:
- Solve

$$\{\lambda + \mu 1(n>0)\}\pi_n = \mu\pi_{n+1} + \lambda 1(n>0)\pi_{n-1}$$

- Just use general birth and death process formulation:

$$\lambda_n = \lambda \quad \mu_n = \mu 1(n>0)$$

so

$$\frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = (\lambda/\mu)^n$$

and

$$\sum_{n=0}^{\infty} (\lambda/\mu)^n = 1/(1-\lambda/\mu)$$

so

$$\pi_n = \frac{(\lambda/\mu)^n}{1+1/(1-\lambda/\mu)}$$

- Exists only if $\lambda < \mu$.

For $M/M/\infty$ queue

- $$\pi_n \propto \frac{\lambda^n}{\mu^n n!}$$

  and

  $$\sum_{n=0}^{\infty} \frac{\lambda^n}{\mu^n n!} = \exp(\lambda/\mu)$$

  so

  $$\pi_n = \exp(-\lambda/\mu)\frac{\lambda^n}{\mu^n n!}$$

- Notice this exists for all $\lambda > 0$ and all $\mu > 0$.

Scope of Queuing Theory

- $M/M/k$ queues.
  - $X(t)$ is number queued or in service.
  - Birth and Death process; death rate maxes out at $k\mu$.
  - Stationary distribution exists if $\lambda < k\mu$.
- Same input / service processes as $M/M/k$ but customers not served leave.
- Question of interest: customers lost per time unit?
- Take X to be number in service. $(0 \le X(t) \le k)$.
- Find stationary distribution.
- Fraction of time spent in state k is π_k.
- During time in state k lose customers at rate λ.
- So lost $\pi_k\lambda$ customers per unit time.

More Queues

4. $G/M/1$ queue. General distribution of interarrival times for input. Input is a **renewal process**. Not Markov.
5. $M/G/1$ and others.
6. Networks: output of 1 queue is input of next; feedback ...
7. Quantities of potential interest:
   - Average fraction of time server idle.
   - Average time in system for customer.
   - Average wait to see server.

One example calculation: in $G/M/1$ queue

- Compute long run fraction time system is idle.
- Idea: interarrival times are iid with cdf G.
- Service rate μ.
- Let X_n be number of customers in service / in line when nth customer arrives.
- Claim X_n is a Markov chain.
- Example of general tactic: find simple process buried within process of interest.

Example Continued

- Notation: $T_1, T_2, \cdots$ iid interarrival times.
- Given $X_n = i$ and $T_{n+1} = t$ number served between nth arrival and $n+1$st arrival is
$$\min\{\text{Poisson}(\mu t), i+1\}$$
- So: if $X_n = i$ and the Poisson variable above is j then
$$X_{n+1} = i + 1 - \min\{j, i+1\}$$
- Now to compute prob of j served must average over T_{n+1}:
$$P(j \text{ served}) = \int e^{-\mu t}\frac{(\mu t)^j}{j!} dG(t) \equiv a_j$$
for $j \le i+1$.
- This gives:
$$P_{ik} = \begin{cases} a_{i+1-k} & 1 \le k \le i+1 \\ 1 - \sum_0^i a_j & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

Computing stationary distribution

- No particularly trivial way to compute this.
- Solve equations. For $k \geq 1$:

$$\begin{aligned}\pi_k &= \sum_j \pi_j P_{jk} \\ &= \sum_j \pi_j a_{j+1-k} 1(k \leq i+1) \\ &= \int_0^\infty e^{-\mu t} \left\{ \sum_{j=k-1}^{\infty} \frac{\pi_j (\mu t)^{j-(k-1)}}{(j-(k-1))!} \right\} dG(t)\end{aligned}$$

- Note that if π_j is a jth power the infinite sum has a closed form.

Stationary Initial Distributions Continued

- So try $\pi_j = c\beta^j$.
- Inside sum is

  $$c\beta^{k-1} \times \exp\{\beta\mu t\}$$

  so the RHS is

  $$c\beta^{k-1} \int_0^\infty e^{-\mu t} e^{\mu\beta t} dG(t)$$

  while the LHS is

  $$c\beta^k$$

- These two are equal if

  $$\beta = \int_0^\infty e^{\mu t(\beta-1)} dG(t)$$

- The LHS is a function of β which is increasing and runs from 0 to 1 as β runs from 0 to 1.

Stationary Initial Distributions Continued

- The RHS is a convex function of β and runs from
$$\int_0^\infty e^{-\mu t} dG(t)$$
at $\beta = 0$ to 1 at $\beta = 1$.
- $\mathrm{RHS}(\beta)$ is positive at $\beta = 0$ (so above the line β) and 1 at $\beta = 1$.
- If slope of RHS at 1 is more than 1 there is unique root $\beta \in (0, 1)$.
- The slope at 1 is
$$\mu \int_0^\infty t dG(t)$$
which is more than 1 if the mean interarrival time
$$\int_0^\infty t dG(t)$$
is more than $1/\mu$ which is the mean service time.
- In this case there is a unique β solving the equation and we get $c = 1 - \beta$.

Busy and idle periods

- **Renewals** at times when customer arrives to find no-one in line or in service.
- Time between successive renewals called a **cycle**.
- Cycle composed of busy period followed by idle period.
- Want to compute fraction of time system idle.
- Want to compute fraction of time system is in state k.
- Use renewal theory ideas.