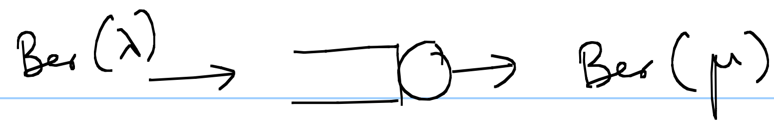
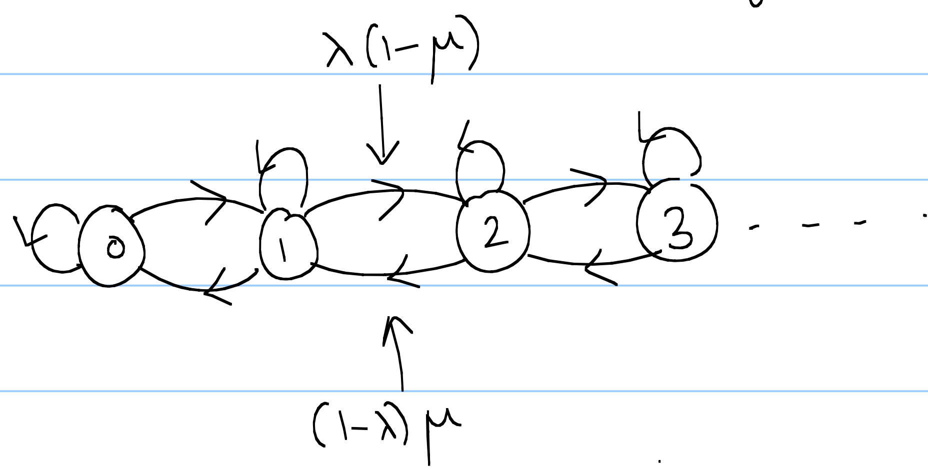
Note Title 1/20/2015

05. Queueing Theory



(Also called Geo | Geo | 1 queue)



$\boxed{\text{Assumption: Arrival occurs before any departure in each time slot.}}$

- If $0 < \lambda, \mu < 1$, the probability of staying in a state > 0, hence DTMC is aperiodic.

Easy to check irreducibility.

- To find π, we need to solve

$$\pi = \pi P.$$

For any state i, this looks like

$$\pi_i = \pi_{i-1} P_{i-1,i} + \pi_i P_{i,i} + \pi_{i+1} P_{i+1,i}$$

$$\Leftrightarrow \quad \pi_i (1 - P_{ii}) = \pi_{i-1} P_{i-1,i} + \pi_{i+1} P_{i+1,i}$$

$$\Leftrightarrow \quad \pi_i \left(P_{i,i-1} + P_{i,i+1} \right) = \pi_{i-1} P_{i-1,i} + \pi_{i+1} P_{i+1,i}$$

using row sum of P = 1

So a sufficient condition to solve for $\pi = \pi P$ is

$$\boxed{\pi_i P_{i,i+1} = \pi_{i+1} P_{i+1,i}}$$

$$\lambda(1-\mu)\,\pi_i = \mu(1-\lambda)\,\pi_{i+1}$$

$$\Longleftrightarrow \quad \pi_{i+1} = \rho\,\pi_i\,, \qquad \rho \triangleq \frac{\lambda(1-\mu)}{\mu(1-\mu)}$$

$$\pi_i = \rho^i \pi_0$$

$\Rightarrow$ if $\rho < 1$,

$$1 = \sum_{i=0}^{\infty} \rho^i \pi_0 \quad \text{or} \quad \boxed{\pi_0 = 1-\rho}$$

$$\Rightarrow \quad \boxed{\pi_i = \rho^i (1-\rho)}\,, \quad i \geq 0$$

If $\rho \geq 1$, $\sum_{i=0}^{\infty} \rho^i = \infty$, so stationary distribution does not exist.

If the stationary distribution, we say that the DTMC is stable. The reason is that in this case,

$$\lim_{k \to \infty} P(X_k \geq B) = \sum_{i=B}^{\infty} \pi_i \longrightarrow 0$$

as $B \to \infty$. Thus, the queue is stable in the sense that the probability it takes very large value is small.

Mean queue length in steady-state

$$=: E(q_\infty) = \sum_{i=0}^{\infty} i \pi_i$$

$$= \sum_{i=0}^{\infty} i \rho^i (1-\rho)$$

$$= (1-\rho)\rho \sum_{i=0}^{\infty} i \rho^{i-1}$$

$$= \rho(1-\rho) \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)$$

$$= \boxed{\frac{\rho}{1-\rho}}$$

Little's law: Consider any queueing system (not just the Geo|Geo|1 queue). Let the delay or waiting time of a packet be the amount of time it spends in the system, with the convention that the waiting time = 0 if a packet arrives and departs in the same time slot.

Let $I_i(t)$ indicate if the i^{th} arrival arrived before time slot t and departed in a time slot $\geq t$.

Then,

$$q(t) = \sum_{i=1}^{|A(t-1)|} I_i(t),$$

where

$A(t)$ = set of arrivals upto and incl. time t.

The waiting time of the i^{th} packet

$$w_i = \sum_{t=1}^{\infty} I_i(t)$$

The time-average of the queue length

$$\frac{1}{T}\sum_{t=1}^{T} q(t) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{|A(t-1)|} I_i(t)$$

$$= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{|A(T-1)|} I_i(t)$$

$$= \frac{1}{T}\sum_{i=1}^{|A(T-1)|}\sum_{t=1}^{T} I_i(t)$$

$$\leq \frac{1}{T}\sum_{i=1}^{|A(T)|} w_i$$

$$= \frac{|A(T)|}{T}\frac{1}{|A(T)|}\sum_{i=1}^{|A(T)|} w_i$$

As $T \to \infty$, $L \triangleq \lim_{T\to\infty} \frac{q(T)}{T}$

and $W = \lim_{T \to \infty} \frac{\sum_{i=1}^{|A(T)|} w_i}{|A(T)|}$

are related by

$$L \le \lambda W, \quad \text{where } \lambda$$

is the arrival rate $\lim_{T \to \infty} \frac{|A(T-1)|}{T}$

Also

$$\frac{1}{T} \sum_{t=1}^{T} q(t) \geq \frac{1}{T} \sum_{i=1}^{D(T-1)} \sum_{t=1}^{T} I_i(t),$$

where $D(t) = \#$ departures in $[1, t]$

Since $I_i(t) = 0 \quad \forall\, t \geq T$.

for $i \in D(T-1)$, we have

$$\frac{1}{T} \sum_{t=1}^{T} q(t) \geq \frac{1}{T} \sum_{i=1}^{D(T-1)} w_i$$

$$= \frac{|A(T-1)|}{T} \frac{1}{|D(T-1)|} \sum_{i=1}^{D(T-1)} w_i$$

$$\to \lambda W.$$

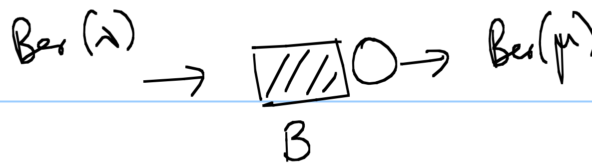
Thus, $\boxed{L = \lambda W}$ → Little's law

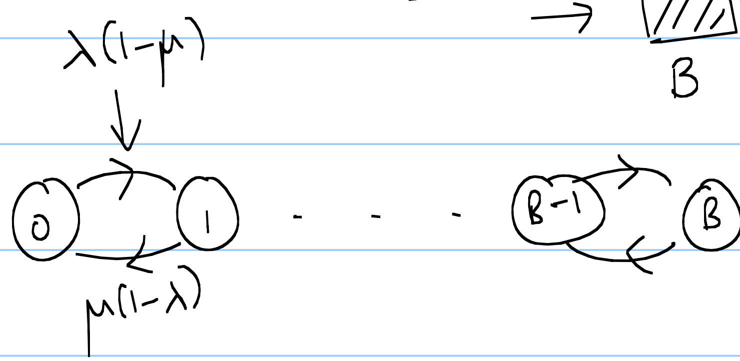
Holds very generally.

Back to Geo | Geo | 1 queue.

$$\boxed{W = \frac{\rho}{(1-\rho)\lambda}}$$

Geo | Geo | 1 | B queue:





$$\pi_i = \rho^i \pi_0$$

$$\sum_{i=0}^{B} \rho^i \pi_0 = 1$$

$$\pi_i = \frac{\rho^i}{\sum_{i=0}^{B} \rho^i} = \frac{\rho^i}{\frac{1-\rho^{B+1}}{1-\rho}}$$

$$\boxed{\pi_i = \frac{\rho^i (1-\rho)}{1-\rho^{B+1}}}$$

$\Rightarrow$ π exists $\forall \rho$.

$\pi_B = P($ queue length $= B)$ in steady-state

$\overset{?}{=}$ $P($ arriving packets sees full buffer$)$ in steady state.

i.e,

Is $\pi_B =$ blocking probability?

Blocking probability

$= P(\, q_i(k) = B \mid a_i(k) = 1 \,)$

$$= \frac{P(a_i(k)=1 \mid q_i(k)=B)\, P(q_i(k)=B)}{P(a_i(k)=1)}$$

Bernoulli:

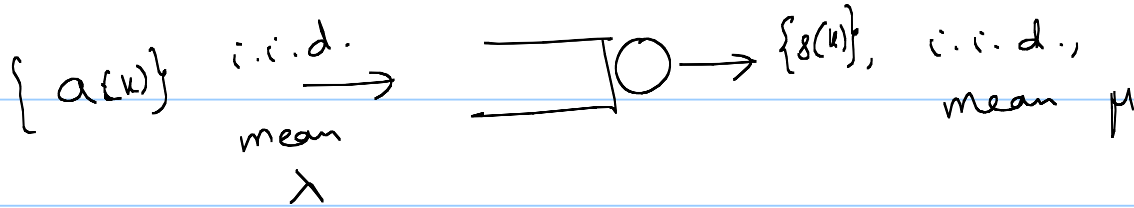
$$= \underbrace{\cancel{\frac{P(a_i(k)=1)}{P(a_i(k)=1)}}}_{=1} \cdot P(q_i(k)=B)$$

$$= \pi_B$$

This property is called BASTA

or **Bernoulli Arrivals See Time Averages**

Discrete-time G|G|1 queues



G: general distribution

$$q(k+1) = \left(q(k) + a(k) - s(k)\right)^+$$

$$E\left(q^2(k+1) - q^2 \mid q(k) = q\right)$$

$$= E\left(\left(q + a - s\right)^{+2} - q^2\right)$$

$$\leq E\left(\left(q + a - s\right)^2 - q^2\right)$$

$$= E\left(\left(a - s\right)^2\right) + 2q\,E(a - s)$$

$$\leq K + 2q(\lambda - \mu),$$

where we have assumed

$$E(a^2) < \infty \quad \text{and} \quad E(s^2) < \infty$$

If $\lambda < \mu$, the drift $\leq -\varepsilon$ for sufficiently large q. Thus, by Foster-Lyapunov, the queue is stable if it is irreducible. We will assume the arrivals and service are such that the DTMC is irreducible and aperiodic. For example,

$$P(a(k)=0),\ P(a(k)=1) > 0,$$

$$P(s(k)=0),\ P(s(k)=1) > 0.$$

Now that we know that the system is stable, let us calculate $E(q)$ in steady-state.

In steady state,

$$0 = E\left(q^2(k+1) - q^2(k)\right)$$

$$\Rightarrow 0 \leq E\left((a-s)^2\right) - E(q).(\mu-\lambda)$$

$$\Rightarrow \boxed{E(q) \leq \frac{E\left((a-s)^2\right)}{\mu-\lambda}}$$

By Little's law,

$$\boxed{W \leq \frac{E\left((a-s)^2\right)}{(\mu-\lambda)\lambda}}$$

How tight is this bound?

$$q(k+1) = q(k) + a(k) - s(k) + \underset{\substack{\uparrow \\ \text{unused} \\ \text{service}}}{u(k)}$$

$$q^2(k+1) = (q+a-s)^2 + u^2 + 2(q+a-s)u$$

$$u = \begin{cases} -(q+a-s) & \text{if } q+a-s < 0 \\ 0 & \text{if } q+a-s \geq 0 \end{cases}$$

$$\Rightarrow (q+a-s)u = -u^2$$

$$\Rightarrow \quad q^2(k+1) = (q + a - s)^2 - u^2$$

So if u is "small", then

$$q^2(k+1) \le (q + a - s)^2$$

becomes nearly an equality. When $\lambda \to \mu$, we can expect $u \approx 0$ and the upper bound on $E(q_\infty)$ will be tight in some appropriate sense. Thus, the upper bound is said to be tight in heavy traffic. We will explore this concept in homework problems.