SSIE 660: Stochastic Systems
Dr. Sung H. Chung
Note 17
Chapter 8. Queueing Theory

Introduction: Queuing System

- Customers arrive in some random manner.
- An arrival will be served by the server (if the server is free).
- They are made to wait in queue (if the server is busy).
- Once served, they are generally assumed to leave the system.

Notation:

L : average number of customers in the system.

L_Q : average number of customers waiting in queue.

W : average amount of time a customer spends in the system.

W_Q : average amount of time a customer spends waiting in queue.

elementary renewal theorem

λ_a : average arrival rate of entering customers.

$$\lambda_a = \lim_{t \to \infty} \frac{N(t)}{t}$$

$$\frac{m(t) = E[N(t)]}{t} = \frac{1}{\mu}$$

Imagine that entering customers are forced to pay money (according to some rule) to the system.

Basic cost identity :

average rate at which the system earns = $\lambda_a \times$ average amount an entering customer $\frac{\text{pays}}{\text{time}}$

Little's formula :

# of customers

$$L = \lambda_a W$$

Also,

$$L_Q = \lambda_a W_Q$$

Let us define:

$E[S]$: average amount of time a customer spends in service.

Then,

$$\lambda_a E[S] = \text{average number of customers in service}$$

Steady-State Probabilities

- $X(t)$: the number of customers in the system at time t

- $P_n = \lim_{t\to\infty} P\{X(t) = n\},\ n \geq 0.$

  Steady-probability that there will be exactly n customers in the system

  = long-run proportion of time that the system contains exactly n customers

- $a_n, n \geq 0$: proportion of customers that find n in the system when they arrive.

- $d_n, n \geq 0$: proportion of customers leaving behind n in the system when they depart.

Example 1. Consider a queueing model in which all customers have service times equal to 1, and where the times between successive customers are always greater than 1.

1. Find a_0, d_0.
2. Is $a_0 = P_0$?

Since every arrival finds the system empty and every departure leaves it empty, we have:

$$a_0 = 1, d_0 = 1$$

However,

$$P_0 \neq 1$$

as the system is not always empty of customers.

Proposition 2. *In any system in which customers arrive and depart one at a time,*

$$\text{rate at which arrivals find } n = \text{rate at which departures leave } n$$

That is,

$$a_n = d_n$$

Proof.

- arrival sees n: systems goes from n to $n+1$.
- departure sees n: systems goes from $n+1$ to n.
- In any interval of time T, # of transitions from n to $n+1$ = # of transitions from $n+1$ to n.
- Now,

$$a_n = \frac{\text{rate at which arrival finds } n}{\text{overall arrival rate}}$$
$$d_n = \frac{\text{rate at which departure leaves } n}{\text{overall departure rate}}$$

- If the overall arrival rate is equal to the overall departure rate, then $a_n = d_n$.
- If the overall arrival rate is greater than the overall departure rate, then $a_n = d_n = 0$.

□

Proposition 3. *Poisson Arrivals always See Time Averages (PASTA principle). In particular, for Poisson arrivals,*

$$P_n = a_n$$

- The total time the system is in state n by time T = $P_n T$.
- The number of arrivals in $[0, T]$ that find the system is in state n is $\lambda P_n T$.
- The long-run rate at which the arrival finds the system is in state n = λP_n.
- The above divided by overall arrival rate = $\lambda P_n / \lambda = P_n \to$ proportion of arrivals that find the system is in state n.

Exponential Models

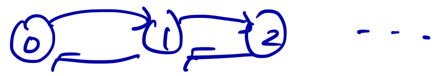
Single Server Exponential Queueing System

- Arrival follows Poisson process with rate λ.
- If the server is free, the arrival will be served by the server.
- If the server is busy, then the arrival will be waiting in the queue.
- Service time is assumed to be independent, exponentially distributed with mean $1/\mu$.
- M/M/1 Markovian Markovian : interarrival & service distributions are exponential → memoryless.

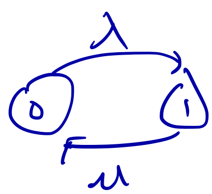
Limiting probabilities: $P_n, n = 0, 1, 2, ...$

rate at which the process enters state n = rate at which the process leaves state n

Figures:



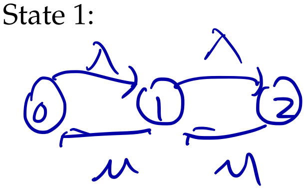
State 0:



the rate at which the process leaves state 0 is λP_0

on the other hand, the rate at which the process enters state 0 is μP_1 $\therefore \lambda P_0 = \mu P_1$

State 1:

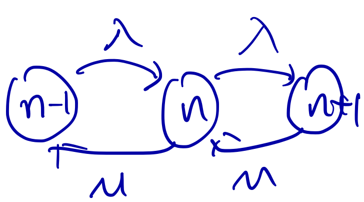


rate out of $(\lambda+\mu)P_1$

rate in $\lambda P_0 + \mu P_2$

$\therefore (\lambda+\mu)P_1 = \lambda P_0 + \mu P_2$

State n:



$$(\lambda+\mu)P_n = \lambda P_{n-1} + \mu P_{n+1}$$

$$\lambda P_0 = \mu P_1$$

$$P_{n+1} = \frac{\lambda}{\mu} P_n + \left(P_n - \frac{\lambda}{\mu} P_{n-1} \right), n \geq 1$$

Solving in terms of P_0 :

$$P_0 = P_0$$

$$P_1 = \frac{\lambda}{\mu} P_0$$

$$P_2 = \frac{\lambda}{\mu} P_1 + \left(P_1 - \frac{\lambda}{\mu} P_0 \right) = \frac{\lambda}{\mu} P_1 = \left(\frac{\lambda}{\mu} \right)^2 P_0$$

$$\vdots$$

$$P_n = \left(\frac{\lambda}{\mu} \right)^n P_0$$

Also,

$$1 = \sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n P_0 = \frac{P_0}{1 - \lambda/\mu} \rightarrow P_0 = 1 - \frac{\lambda}{\mu}$$

Therefore,

$$P_n = \left(\frac{\lambda}{\mu} \right)^n \left(1 - \frac{\lambda}{\mu} \right) \quad n \geq 1$$

Note: we assumed that $\frac{\lambda}{\mu} < 1 \longrightarrow$ the mean service time is less than the mean time between successive arrivals. $\lambda < \mu$

otherwise, if $\lambda > \mu$ then $\underline{P_n = 0}$ $\forall n$

Now,

avg. # of customers in the system

$$L = \sum_{n=0}^{\infty} n P_n$$

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

$$\left(Note: \sum_{n=1}^{\infty} n x^n = \frac{x}{(1-x)^2} \right)$$

Little's law

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

$$W_Q = W - E[S] = W - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}$$

$$L_Q = \lambda W_Q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

Example 4. Suppose that customers arrive at a Poisson rate of one per every 12 minutes, and that the service time is exponential at a rate of one service per 8 minutes. What are L and W?

$$\lambda = \frac{1}{12}$$

$$\mu = \frac{1}{8}$$

$$L = \frac{\lambda}{\mu - \lambda} = 2$$

$$W = \frac{1}{\mu - \lambda} = 24$$

Now suppose that the arrival rate increases 20 percent to $\lambda = 1/10$. What is the corresponding L and W?

$L = 4, \quad W = 40$

(8.4) page 493

Example 5. For an $M/M/1$ queue in steady state, what is the probability that the next arrival finds n in the system?

- If t is the current time, then the time from t until the next arrival is exponentially distributed with rate λ.
- Also, the time from the last arrival before t until the time t is exponentially distributed with rate λ.
- Although, the times between successive arrivals of a Poisson process are exponential with rate λ, the time between the previous arrival before t and the first arrival after t is distributed as the sum of two independent exponentials. (the inspection paradox)
- The length of an inter arrival interval that contains a specified time tends to be longer than an ordinary inter arrival interval.
- N_a: the number found by the next arrival.
- X: the number currently in the system.

Conditioning on X yields:

$$P[N_a = n] = \sum_{k=0}^{\infty} P\{N_a = n \mid X = k\} P\{X = k\}$$

$$= \sum_{k=0}^{\infty} P\{N_a = n \mid X = k\} \left(\frac{\lambda}{\mu}\right)^k (1 - \lambda/\mu)$$

$$= \sum_{k=n}^{\infty} P\{N_a = n \mid X = k\} \left(\frac{\lambda}{\mu}\right)^k (1 - \lambda/\mu)$$

$$= \sum_{i=0}^{\infty} P\{N_a = n \mid X = n+i\} (\lambda/\mu)^{n+i} (1 - \lambda/\mu)$$

By the way

$$P\{N_a = n \mid X = n+i\} = \left(\frac{\mu}{\lambda+\mu}\right)^i \frac{\lambda}{\lambda+\mu}, \quad n > 0$$

Consequently $P\{N_a = n\} = (\lambda/\mu)^{n+1} (1 - \lambda/\mu)$

on the other hand, $P\{N_a = 0\} = \sum_{i=0}^{\infty} \left(\frac{\mu}{\lambda+\mu}\right)^i \left(\frac{\lambda}{\mu}\right)^i (1 - \lambda/\mu)$

$$= (1 + \lambda/\mu)(1 - \lambda/\mu)$$

Single Server Exponential Queueing System Having Finite Capacity

Balance equations:

State	Rate at which the process leaves = Rate at which the process enters
0	$\lambda P_0 = \mu P_1$
$1 \le n \le N-1$	$(\lambda+\mu) P_n = \lambda P_{n-1} + \mu P_{n+1}$
N	$\mu P_N = \lambda P_{N-1}$

Also, $\mu P_n = \lambda P_{n-1}, n = 1, 2, ..., N$ the rate at which departures leave behind n-1 = " arrivals find n-1

Thus,

$$P_n = \frac{\lambda}{\mu} P_{n-1} = \cdots = \left(\frac{\lambda}{\mu}\right)^n P_0$$

$$1 = \sum_{n=0}^{N} P_n = P_0 \sum_{n=0}^{N} \left(\frac{\lambda}{\mu}\right)^n$$

$$P_0 = \frac{(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}}$$

$$P_n = \frac{(\lambda/\mu)^n (1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}} \qquad n = 0, 1, \ldots, N$$

Note: no need to assume $\lambda/\mu < 1$.

$$L = \sum_{n=0}^{N} nP_n = \frac{(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}} \sum_{n\geq 0}^{N} n\left(\frac{\lambda}{\mu}\right)^n$$

$$= \frac{\lambda[1+N(\lambda/\mu)^{N+1} - (N+1)(\lambda/\mu)^N]}{(\mu-\lambda)(1-(\lambda/\mu)^{N+1})}$$

In terms of W, there are two cases.

1. including those arrivals to find the system is full (thereby, not entering the system, W=0)

$$\lambda_a = \lambda$$

2. considering only actual entering arrivals ($W > 0$)

$$\lambda_a = \lambda(1-P_N)$$

$$W = \frac{L}{\lambda_a}$$

Example 6. Suppose that it costs $c\mu$ dollars per hour to provide service at a rate μ. Suppose also that we incur a gross profit of A dollars for each customer served. If the system has a capacity N, what service rate μ maximizes our total profit?

$$\text{profit per hour} = \lambda(1-P_N)A - c\mu$$

$$= \lambda A\left[1 - \frac{(\lambda/\mu)^N(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}}\right] - c\mu$$

$$= \frac{\lambda A[1-(\lambda/\mu)^N]}{1-(\lambda/\mu)^{N+1}} - c\mu$$

$$\frac{d[\quad]}{d\mu} = 10\frac{(2\mu^3-3\mu^2+1)}{(\mu^3-1)^2} - 1$$