Note for lecture 07

λ

Little's Theorem provides us a very useful way to estimate a system's performance regarding N, λ, and T:

$$N = \lambda \cdot T \quad \text{at steady state}$$

But N and T cannot be specified further without knowledge of some statistics of the system. A way to proceed is to try to compute

$$N = \sum_{n=0}^{\infty} n \cdot p_n$$

where p_n stands for the stedy-state probabilities of n customers in the system. Then we can get T using Little's Theorem.

The theory of Markov chains provides us a powerful tool to obtain p_n. In the following, we first introduce the theory of Markov chains, followed by some statistics of the system where the theory is applicable:

√P

☆ Markov chain theory (materials borrowed from textbook 2 by Prof. Mor Harchol Balter)

P Definition: A discrete-time Markov chain (DTMC) is a stochastic process $\{X_n, n = 0, 1, 2, \cdots\}$ where X_n denotes the state at time step n and such that, $\forall n \geq 0$, $\forall i, j$, and $\forall i_0, \ldots, i_{n-1}$,

N

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$$
$$= P\{X_{n+1} = j \mid X_n = i\} = P_{ij}$$

P_{ij} is independent of time step and of past history.

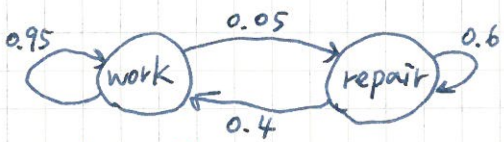
Note: a "stochastic process" is a term for a sequence of random variables.

Definition: The Markovian Property

The conditional distribution of any future state X_{n+1} depends only on the present state X_n.

Definition: The transition probability matrix associated with any DTMC is a matrix $\mathbb{P}$ whose (i,j)-th entry P_{ij} represents the probability of moving to state j on the next transition, given that the current state is i.

Example: Consider a machine is either working or in repair, with the following Markov chain diagram:



○ : state

↷ : transition and its probability

$$\mathbb{P} = \begin{array}{c} \\ W \\ R \end{array} \begin{array}{cc} W & R \\ \left[\begin{array}{c} 0.95 \\ 0.4 \end{array}\right. & \left.\begin{array}{c} 0.05 \\ 0.6 \end{array}\right] \end{array}$$

$$\mathbb{P}^2 = \begin{bmatrix} 0.9225 & 0.0775 \\ 0.62 & 0.38 \end{bmatrix}$$

$$\mathbb{P}^3 = \begin{bmatrix} 0.9073 & 0.0926 \\ 0.741 & 0.259 \end{bmatrix}$$

$\vdots$

$$\mathbb{P}^{10} = \begin{bmatrix} \sim 0.89 & \sim 0.11 \\ \sim 0.89 & \sim 0.11 \end{bmatrix}$$

$$\mathbb{P}^{11} = \begin{bmatrix} \sim 0.89 & \sim 0.11 \\ \sim 0.89 & \sim 0.11 \end{bmatrix}$$

$\vdots$

$$P_{ij}^{2} = \sum_{k=0}^{M-1} P_{ik} \cdot P_{kj}$$

$$=_i [\quad] [\quad]^{j}$$

$$\Rightarrow P_{ij}^{n} = \sum_{k=0}^{M-1} P_{ik} \cdot P_{kj}^{n-1}$$

from i, go to j in n steps

in general, let $\mathbb{P} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$ $\quad 0 < a < 1$, $0 < b < 1$

$$\lim_{n\to\infty} \mathbb{P}^n = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$$

← all rows are the same, implying that the starting state does not matter!

let $\pi_j = \lim_{n\to\infty} P_{ij}^n = (\lim_{n\to\infty} P^n)_{ij}$

for an M-state DTMC,

$\vec{\pi} = (\pi_0, \pi_1, \pi_2, \dots, \pi_M)$ with $\sum_{i=0}^{M-1} \pi_i = 1$

represents the limiting distribution in each state.

(for proofs of existence, etc., see textbook2 Chapters 8 and 9)

It turns out that the above limiting distribution is equal to an unique stationary distribution where $\vec{\pi} \cdot \mathbb{P} = \vec{\pi}$

$\Rightarrow$ in general, we can compute π_j using $\vec{\pi} \cdot \mathbb{P} = \vec{\pi}$ and $\sum_{i=0}^{M-1} \pi_i = 1$

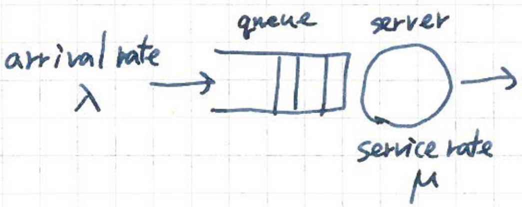
in the previous example, for example,

$$\begin{cases} \pi_W = \pi_W \times 0.95 + \pi_R \times 0.4 \\ \pi_R = \pi_W \times 0.05 + \pi_R \times 0.6 \\ \pi_W + \pi_R = 1 \end{cases}$$

$$\Rightarrow \begin{cases} \pi_W = \frac{8}{9} \approx 0.89 \\ \pi_R = \frac{1}{9} \approx 0.11 \end{cases}$$

Back to our analysis of queueing systems, we may obtain p_n by computing π_j if the system has the Markovian Property! See next page →

★ Arrival statistics and service statistics of queueing system



The Kendall notation to classify the queueing system

X / Y / Z

- X → the distribution of interarrival times for the arrival process
  - M : memoryless
- Y → the distribution of the service time
  - M : memoryless
  - G : general
  - D : deterministic
- Z → # of servers

We start with the M/M/1 queue.

Exponential distribution, Definition :

a random variable X with the probability density function

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Its cumulative distribution function is

$$F_X(s) = P\{X \leq s\} = \int_0^s f(y)\,dy = 1 - e^{-\lambda s}$$

The mean $E[X] = \int_0^\infty x \cdot f(x)\,dx = \frac{1}{\lambda}$, and we call λ "rate"

Exponential distribution is memoryless in the sense that

$$P\{X > s+t \mid X > s\} = \frac{P\{X > s+t\}}{P\{X > s\}} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

$$= P\{X > t\}$$

We model the arrival process as the Poisson process, the most widely used model for its good approximation in describing the aggregate behavior of a large number of low-rate arrivals (e.g., an IoT network with many low-rate data streams).

Poisson process, Definition :

A stochastic process $\{A(t) \mid t \geq 0\}$ is a Poisson process with rate λ if

① $A(t)$ is a counting process that represents the total # of arrivals in $[0, t]$

② # of arrivals occurring in disjoint time intervals are independent

③ # of arrivals in any intervals of length τ is Poisson distributed with parameter $\lambda\tau$. That is, for all $t, \tau > 0$,

See later page for a review of Poisson distribution.

$$\cancel{P\{A(t}\ P\{A(t+\tau) - A(t) = n\} = e^{-\lambda\tau}\frac{(\lambda\tau)^n}{n!}$$

$n = 0, 1, \ldots$

The Poisson process has an important property that interarrival times are independent and exponentially distributed with rate λ.

thus memoryless

Therefore, we may use the Markov chain theory to analyze the M/M/1 queue!

Poisson process with rate λ

service times have an exponential distribution with rate μ

In M/M/1, let $N(t)$ denote the # of customers in the system at time t. The times at which customers will arrive or complete service in the future are independent of ① the arrival times of customers presently in the system, and

of ② how much service the customer currently in service has already received.

We now first focus on discrete times

$$0, \delta, 2\delta, 3\delta, \dots \quad \text{and then let } \delta \to 0$$

Let N_k denote the # of customers in the system at $k\delta$

The transition probabilities of DTMC in this case is

$$P_{ij} = P\{N_{k+1} = j \mid N_k = i\}$$

We can obtain these probabilities by considering # of arrivals and departures in a δ interval.

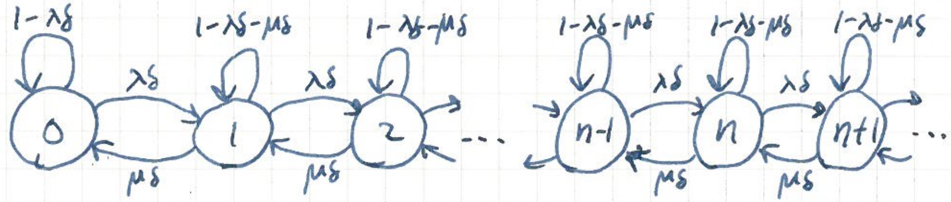
For example, $P_{ii} = (e^{-\lambda\delta})(e^{-\mu\delta})$

$$= e^{-\lambda\delta - \mu\delta}$$

using Taylor expansion $e^{-a} = 1 - a + a^2/2 - \cdots$

$$\Rightarrow P_{ii} = 1 - \lambda\delta - \mu\delta + o(\delta) \quad \text{where } \lim_{\delta \to 0} \frac{o(\delta)}{\delta} = 0$$

The DTMC for the M/M/1 queue is therefore:



where state n corresponds to n customers in the system, and the shown transition probabilities are correct up to an $o(\delta)$ term.

At long term, the frequency of transition from n to $n+1$ is equal to that from $n+1$ to n. This can be derived by considering that the probability of leaving a state is equal to that of entering the state.

$$p_n \lambda\delta + o(\delta) = p_{n+1} \mu\delta + o(\delta)$$

$$\Rightarrow p_n \lambda = p_{n+1} \mu \quad \text{as } \delta \to 0$$

let $\rho = \frac{\lambda}{\mu}$, then $p_{n+1} = \rho p_n$ and $p_{n+1} = \rho^{n+1} p_0$

$$1 = \sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} \rho^n p_0 = \frac{p_0}{1-\rho} \Rightarrow p_n = \rho^n (1-\rho)$$

Finally, $\boxed{N} = \sum_{n=0}^{\infty} n \cdot p_n = \sum_{n=0}^{\infty} n \cdot \rho^n (1-\rho)$

$$= (1-\rho) \sum_{n=0}^{\infty} n \cdot \rho^n = (1-\rho) \cdot \frac{\rho}{(1-\rho)^2} = \boxed{\frac{\rho}{1-\rho}}$$

and from Little's Theorem $\boxed{T = \frac{N}{\lambda} = \frac{1}{\mu - \lambda}}$