THEOREM 4.3.6 (KOLMOGOROV'S FORWARD EQUATION)

Under suitable regularity conditions,

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t)$$

#

PROPOSITION 4.3.7

For a pure birth process,

$$\begin{aligned} P_{ii}(t) &= e^{-\lambda_i t}, \quad i \geq 0 \\ P_{ij}(t) &= \lambda_{j-1} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} P_{i,j-1}(s) ds, \quad j \geq i+1 \end{aligned}$$

#

4.4 Limiting Probabilities

In analogy with DTMC, the probability that a CTMC will be in state j at time t often converges to a limiting value which is independent of the initial state:

$$P_j = \lim_{t \to \infty} P_{ij}(t)$$

To derive a set of equations for the P_j, consider the set of forward equations

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t)$$

Letting t approach ∞, we have

$$\begin{aligned} \lim_{t \to \infty} P'_{ij}(t) &= \lim_{t \to \infty} \left[\sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \right] \\ &= \sum_{k \neq j} q_{kj} P_k - v_j P_j \end{aligned}$$

Since $P'_{ij}(t)$ must converge to 0, we have

$$0 = \sum_{k \neq j} q_{kj} P_k - v_j P_j$$

or

$$v_j P_j = \sum_{k \neq j} q_{kj} P_k, \quad \text{all states } j \tag{4.12}$$

The preceding set of equations, along with this equation,

$$\sum_j P_j = 1 \tag{4.13}$$

can be used to solve the limiting probabilities.

A sufficient condition for existence of limiting probabilities P_j is that

i. All states of the MC communicate.

ii. The MC is positive recurrent.

Eq. (4.12) and (4.13) has a nice interpretation:

$$v_j P_j = \text{rate at which the process leaves state } j$$

$$\sum_{k \neq j} q_{kj} P_k = \text{rate at which the process enters state } j$$

Eq. (4.12) is just a statement of the equality of the rates at which the process enters and leaves the state j. Because it balances these rates, the Eq. (4.12) are sometimes referred to as "balance equations."

Let us now determine the limiting probabilities for a birth and death process. From Eq. (4.12), we obtain

State	Rate at which leave = rate at which enter
0	$\lambda_0 P_0 = \mu_1 P_1$
1	$(\lambda_1 + \mu_1) P_1 = \mu_2 P_2 + \lambda_0 P_0$
2	$(\lambda_2 + \mu_2) P_2 = \mu_3 P_3 + \lambda_1 P_1$
$\cdots$	$\cdots$
$n, n \geq 1$	$(\lambda_n + \mu_n) P_n = \mu_{n+1} P_{n+1} + \lambda_{n-1} P_{n-1}$

By adding to each equation the equation preceding it, we obtain

$$\begin{aligned}
\lambda_0 P_0 &= \mu_1 P_1, \\
\lambda_1 P_1 &= \mu_2 P_2, \\
\lambda_2 P_2 &= \mu_3 P_3, \\
&\cdots \\
\lambda_n P_n &= \mu_{n+1} P_{n+1}, \quad n \geq 0
\end{aligned}$$

Solving in terms of P_0 yields

$$\begin{aligned}
P_1 &= \frac{\lambda_0}{\mu_1} P_0, \\
P_2 &= \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0, \\
P_3 &= \frac{\lambda_2}{\mu_3} P_2 = \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} P_0, \\
&\cdots \\
P_n &= \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_1 \lambda_0}{\mu_n \mu_{n-1} \cdots \mu_2 \mu_1} P_0
\end{aligned}$$

And by using the fact that $\sum_{n=0}^{\infty} P_n = 1$, we obtain

$$1 = P_0 + P_0 \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1}$$

or

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1}}$$

And so

$$P_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1 \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1}\right)}, \quad n \geq 1 \tag{4.14}$$

From the above, the necessary condition for the limiting probabilities to exist is

$$\sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1} < \infty$$

This condition also may be shown to be sufficient.

Ex 4.4.1 (Ex 6.14, pp. 373) M/M/1

#

Ex 4.4.2 (Ex 6.15, pp. 374) Consider the showshine shop and determine the proportion of the time the process is in each of the states 0, 1, 2. Note that it is not a birth and death process.

#

Chapter 5

Queueing Theory

We study a class of models in which customers arrive in some random manner to a service facility. Upon arrival they are made to wait in queue until it is their turn to be served. Once served they are generally assumed to leave the system.

For such models we will be interested in determining such quantities as

i. the average number of customers in the system (or in the queue) and

ii. the average amount of time a customer spends in the system (or spends waiting in the queue).

5.1 Little's Theorem

In a general queueing system when the number of customers is large then so is the waiting time. Here we like to establish a very simple relationship between the mean number in the system, the mean arrival rate, and the mean system time.

Let us look at the input of the queueing system and count how many customers enter as a function of time. We denote this by $\alpha(t)$ where

$$\alpha(t) = \text{number of arrivals in } (0, t)$$

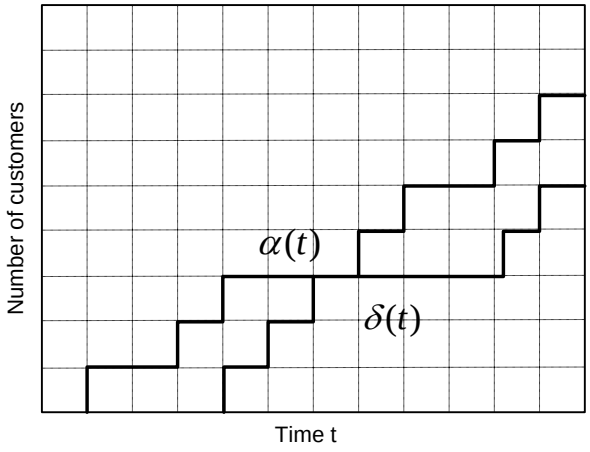
Alternatively, we count the number of departures at the output of the queueing system; we denote this by

$$\delta(t) = \text{number of departures in } (0, t)$$

Clearly $N(t)$, the number in the system at time t, must be given by

$$N(t) = \alpha(t) - \delta(t)$$

The total area between two curves up to some point t represents the total time all customers have spent in the system during the interval $(0, t)$; let us denote this cumulative area by $\gamma(t)$. Moreover let λ_t be



defined as the average arrival rate during the interval $(0, t)$; that is,

$$\lambda_t = \frac{\alpha(t)}{t}$$

We define T_t as the system time per customer averaged over all customers in the interval $(0, t)$; since $\gamma(t)$ represents the accumulated customer-second up to time t, we may divide by the number of arrivals up to that point to obtain

$$T_t = \frac{\gamma(t)}{\alpha(t)}$$

Lastly we define $\overline{N}_t$ as the average number of customers in the queueing system during the interval $(0, t)$; this may be obtained by dividing the accumulated number of customer-seconds by the total interval length t

$$\overline{N}_t = \frac{\gamma(t)}{t}$$

From these last three equations we see

$$\overline{N}_t = \lambda_t T_t$$

Let us now assume that our queueing system is such that the following limits exist as $t \to \infty$.

$$\begin{aligned} \lambda &= \lim_{t\to\infty} \lambda_t \\ T &= \lim_{t\to\infty} T_t \end{aligned}$$

Then,

$$\overline{N} = \lambda T \tag{5.1}$$

This results is known as Little's result. It states that the average number of customers in a queueing system is equal to the average arrival rate of customers to that system, times the average time spent in that system.

Ex 5.1.1 Customers arrive at a fast-food restaurant at a rate of five per minute and wait to receive their order for an average of 5 minutes. Customers eat in the restaurant with prob. 0.5 and carry out their order with prob. 0.5. A meal requires an average of 20 minutes. What is the average number of customers in the restaurant? Ans: 75

#

Ex 5.1.2 Consider a network of transmission lines where packets arrive at n different nodes with corresponding rates $\lambda_1, \cdots, \lambda_n$. If N is the average total number of packets inside the network find the average delay per packet.

#

5.2 The M/M/1 Systems

The M/M/1 queueing system consists of a single queueing system with a single server. Customer arrive according to a Poisson process with rate λ, and the pdf of the service time is exponential with mean $1/\mu$ sec. The name M/M/1 reflects standard queueing theory nomenclature whereby:

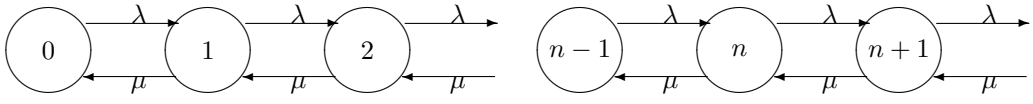
- The first letter indicates the nature of the arrival process
- The second letter indicates the nature of the pdf of the service time
- The last number indicates the number of servers

The letter "M" stands for memoryless, which here means a Poisson process (i.e., exponentially distributed interarrival times), "G" stands for a general distribution of interarrival times, and "D" stands for deterministic interarrival times.

The M/M/1 queue is the simplest queue system and may be described by selecting the birth-death coefficients as follows:

$$\begin{aligned} \lambda_n &= \lambda, \quad n = 0, 1, 2, \cdots \\ \mu_n &= \mu, \quad n = 1, 2, \cdots \end{aligned}$$

The state transition diagram is shown below:



From the previous birth-death results, we have

$$P_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1\left(1+\sum_{n=1}^{\infty}\frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1}\right)}, \quad n \geq 1 \tag{5.2}$$

and

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1}} \tag{5.3}$$

For M/M/1, this becomes

$$P_n = P_0 \left(\frac{\lambda}{\mu}\right)^n \tag{5.4}$$

where

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n}$$

Since $\lambda < \mu$ to have an equilibrium, the P_0 becomes

$$P_0 = 1 - \frac{\lambda}{\mu}$$

Letting $\rho = \lambda/\mu$, we have

$$P_n = (1-\rho)\rho^n, \quad n = 0, 1, 2, \cdots \tag{5.5}$$

This is indeed the solution for steady-state probability of finding n customers in the system.

To find the average number of customers in the system $\overline{N}$,

$$\begin{aligned}
\overline{N} &= \sum_{n=0}^{\infty} nP_n \\
&= (1-\rho)\sum_{n=0}^{\infty} n\rho^n \\
&= (1-\rho)\rho\frac{\partial}{\partial\rho}\sum_{n=0}^{\infty}\rho^n \\
&= (1-\rho)\rho\frac{\partial}{\partial\rho}\frac{1}{1-\rho} \\
&= \frac{\rho}{1-\rho}
\end{aligned} \tag{5.6}$$

We may now apply Little's result to obtain T, the average time in the system as follows:

$$\begin{aligned}
T &= \frac{\overline{N}}{\lambda} \\
&= \frac{\rho}{1-\rho}\cdot\frac{1}{\lambda} \\
&= \frac{1/\mu}{1-\rho}
\end{aligned} \tag{5.7}$$

Question: Find average number of customers in queue, average waiting time in queue.

#