Definitions

A birth refers to the arrival of a new customer and death refers to the departure of a served customer.

State of the system at time $t$, ($t \geq 0$) is given by $N(t) =$ the number of customers in the system at time $t$.

Individual births and deaths occur randomly and the mean occurrence rates depend only upon the current state of the system.

Assumptions

The following assumptions apply:

- Given $N(t) = n$, the current probability distribution of the remaining time until the next birth (arrival) is exponential with parameter $\lambda_n$ ($n = 0, 1, 2, ...$).

- Given $N(t) = n$, the current probability distribution of the remaining time until the next death (service completion) is exponential with parameter $\mu_n$ ($n = 0, 1, 2, ...$).

- The random variable of assumption 1 (the remaining time until the next birth) and the random variable of assumption 2 (the remaining time until the next death) are mutually independent. The next transition in the state of the process is either $n \rightarrow n + 1$ (a single birth) or $n \rightarrow n - 1$ (a single death) depending on whether the former or latter random variable is smaller.

$\lambda_n$ and $\mu_n$ can be different for different values of $n$ if arriving customers become increasingly likely to balk as $n$ increases, and renege as queue size increases respectively.

Balance Equation

$E_n(t) =$ number of times the process enters state $n$ by time $t$.
$L_n(t) =$ number of times the process leaves state $n$ by time $t$.

Mean rate at which process enters state $n$ is given by:

$$\lim_{t \to \infty} \frac{E_n(t)}{t}$$

Mean rate at which process leaves state $n$ is given by:

$$\lim_{t \to \infty} \frac{L_n(t)}{t}$$

Rate in = Rate out. For any state of the system $n$, mean entering rate = mean leaving rate.
Relevant Formulae

\[ P_n = C_n P_0 \] (3)

Given that:

\[ \sum_{n=0}^{\infty} P_n = 1 \] (4)

Where:

\[ C_n = \frac{\lambda_{n-1}\lambda_{n-2}...\lambda_0}{\mu_n\mu_{n-1}...\mu_1} \] (5)

The following formula apply, where the symbols have their usual meaning:

\[ L = \sum_{n=0}^{\infty} n P_n \] (6)

\[ L_q = \sum_{n=0}^{\infty} (n - s) P_n \] (7)

Consider the M/M/1 case:

\[ \lambda_n = \lambda, n = 0, 1... \] (8)

\[ \mu_n = \mu, n = 0, 1... \] (9)

\[ C_n = \left( \frac{\lambda}{\mu} \right)^n = \rho^n \] (10)

\[ P_0 = (1 - \rho) \] (11)

\[ P_n = (1 - \rho)\rho^n \] (12)

\[ L = \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n(1 - \rho)\rho^n = \frac{\lambda}{\mu - \lambda} \] (13)

\[ L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} \] (14)

Consider the M/M/s case:

\[ \lambda_n = \lambda, n = 0, 1... \] (15)

\[ \mu_n = \begin{cases} n\mu, & n \leq s \\ s\mu, & n > s \end{cases} \] (16)

\[ C_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} & \text{for } n = 1, 2,...s \\ \frac{(\lambda/\mu)^n}{s!s^{n-s}} & \text{for } n = s, s+1,... \end{cases} \] (17)

\[ L_q = \frac{P_0(\lambda/\mu)^s\rho}{s!(1 - \rho)^2} \] (18)

\[ P_0 = 1/\left[ \sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \frac{1}{1 - \lambda/(s\mu)} \right] \] (19)

Consider the finite queue variation of the M/M/1 model (M/M/1/K)

\[ \lambda_n = \begin{cases} \lambda & \text{for } n = 0, 1, 2,...K - 1 \\ 0 & \text{for } n \geq K \end{cases} \]
\[ C_n = \begin{cases} \frac{\lambda^n}{\mu} & \text{for } n = 0, 1, 2, \ldots K \\ 0 & \text{for } n = 0, 1, 2, \ldots K - 1 \end{cases} \]

\[ P_0 = \frac{1 - \rho}{1 - \rho^{K+1}} \quad (20) \]

\[ P_n = \frac{1 - \rho}{1 - \rho^{K+1}} \rho^n, \quad \text{for } n = 0, 1, 2, \ldots K \quad (21) \]

\[ L = \frac{\rho}{1 - \rho} - \frac{(K + 1)\rho^{K+1}}{1 - \rho^{K+1}} \quad (22) \]

\[ L_q = L - (1 - P_0) \quad (23) \]

**Applications: Repairman Models**

A system is composed of \( N \) machines of which at most \( M \leq N \) can be operating at one time. The rest are spares. When a machine is operating it operates a random length of time until failure with parameter \( \mu \). When a machine fails it undergoes repair. At most \( R \) machines are in repair at any point of time. The repair time is exponentially distributed with parameter \( \lambda \). Hence a machine can be in any of the four states:

- Operating
- Up but not operating
- In repair
- Waiting for repair

There are a total of \( N \) machines in the system. At most \( M \) can be operating. At most \( R \) can be in repair. Let \( X(t) \) be a random variable denoting the number of up machines at the time \( t \). Hence, we can say:

- Number of machines operating: \( \min\{X(t), M\} \)
- Number of spares: \( \max\{0, X(t) - M\} \)
- Number of down machines: \( Y(t) = N - X(t) \)
- Number in repair: \( \max\{0, Y(t) - R\} \)

\( X(t) = n \) is a finite state birth and death process with the parameters:

\[ \lambda_n = \lambda \times \min\{N - n, R\} = \begin{cases} \lambda R & \text{for } n = 0, 1, \ldots, N - R, \\ \lambda(N - n) & \text{for } n = N - R + 1, \ldots, N \end{cases} \]

and

\[ \mu_n = \mu \times \min\{n, M\} = \begin{cases} \mu n & \text{for } n = 0, 1, \ldots, M, \\ \mu M & \text{for } n = M + 1, \ldots, N \end{cases} \]

All else can be routinely discovered.