대기행렬 이론 개론

VII. M/M/c/m/m Queue

학습에 앞서

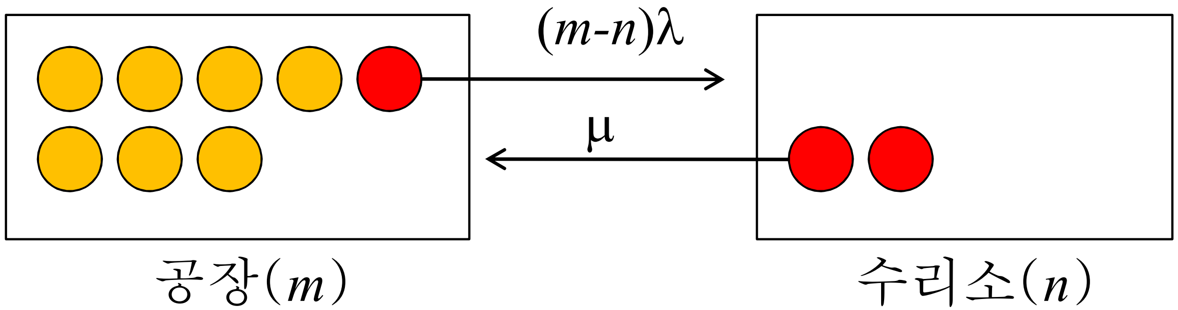
- 학습 목표
  - M/M/c/m/m 대기행렬 모델을 이해하고, 상태전이확률과 정상상태확률을 구하는 능력을 배양한다.
  - 집단도착 대기행렬과 집단서비스 대기행렬 모델을 이해하고, 상태전이확률과 정상상태확률을 구하는 능력을 배양한다.

- 목차
  - 1. The M/M/c/m/m Queue (기계간섭문제)
    - 1.1 The M/M/1/m/m Queue
    - 1.2 The M/M/c/m/m Queue
  - 2. $M^x/M/1$ 집단도착 대기행렬
  - 3. $M/M^B/1$ 집단서비스 대기행렬

1. The *M/M/c/m/m* Queue (기계간섭문제)

m대의 기계가 있는 공장을 생각하자. 각 기계의 공장률은 λ이다. 즉, 고장날 때까지의 시간이 Exp(λ)를 따른다. 고장난 기계는 한 명의 수리공에 의해 수리되는 총 c명의 수리공이 있다. 수리공들의 입장에 보면 고장나는 기계는 고객이다. 또는 고장난 기계는 가상의 수리소로 보내진다고 할 수 있으며, 수리소 입장에서 고장난 기계는 고객이다. 고객이 될 수 있는 자격을 갖는 잠정 고객수는 m($M/M/c/m/m$의 두번째)이고 수리소에 존재할 수 있는 최대 고객수(수리중인 기계를 포함한 고장나 있는 기계 대수) 역시 m이다($M/M/c/m/m$의 첫번째). 이와 같이 유한한 잠정고객집단(모집단)을 갖는 시스템을 유한모집단시스템(system with finite calling population)이라고 부른다.

이 문제는 기계간섭문제(machine interference problem) 또는 기계수리문제(machine repair problem)이라고 불리며 일찍부터 사람들의 관심의 대상이 되어왔다.

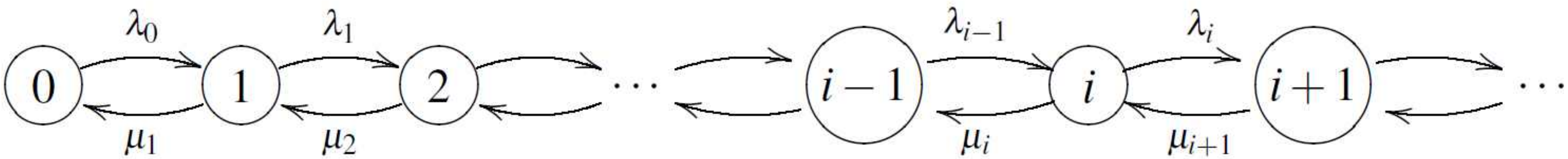


M/M/1/m/m

The *M/M/1/m/m* Queue

이 시스템의 고객수과정은 다음과 같은 전이율을 갖는 출생사멸과정이다.

$$\lambda_n = \begin{cases} (m-n)\lambda, & (0 \le n \le m-1), \\ 0, & (n \ge m), \end{cases} \qquad \mu_n = \begin{cases} \mu, & (1 \le n \le m), \\ 0, & (n \ge m+1). \end{cases}$$



- Birth-death process

$$\pi_0 = \left(1 + \sum_{i=1}^{\infty} \frac{\prod_{k=0}^{i-1} \lambda_k}{\prod_{k=1}^{i} \mu_k}\right)^{-1} \qquad\qquad \pi_i = \frac{\prod_{k=0}^{i-1} \lambda_k}{\prod_{k=1}^{i} \mu_k} \pi_0$$

Steady-state Probability

$N(m)$ m대의 기계로 이루어진 시스템에서 임의시점에서의 고객수

P_n n대의 기계가 고장나 있을 확률이다(수리 중인 것 포함)

P_{m-k} k 대의 기계가 운행 중일 확률

$$P_0 \equiv P_0(m) = Pr[N(m)=0] = \left[\sum_{k=0}^{m} \frac{m!}{(m-k)!} \left(\frac{\lambda}{\mu} \right)^k \right]^{-1}$$

$$P_n \equiv P_n(m) = Pr[N(m)=n] = \frac{\frac{m!}{(m-n)!} \left(\frac{\lambda}{\mu} \right)^n}{\sum_{k=0}^{m} \frac{m!}{(m-k)!} \left(\frac{\lambda}{\mu} \right)^k},$$

$$P_{m-k} = \frac{\frac{e^{-\mu/\lambda}(\mu/\lambda)^k}{k!}}{\sum_{i=0}^{m} \frac{e^{-\mu/\lambda}(\mu/\lambda)^i}{i!}}$$

P_n의 분자, 분모를 $m!$로 나누고
$e^{-\mu/\lambda}(\mu/\lambda)^m$를 곱한 후 n을 $(m-k)$로 바꾸면

Performance Measure (1/2)

(수리 중인 기계의 평균대수(수리 대기 중인 기계 제외)) 수리소에 한 대 이상 있으면 수리 중인 기계가 한 대이므로 다음과 같다.

$$L_{repair} = (0)P_0 + (1)\sum_{n=1}^{m} P_n = 1 - P_0$$

(운행 중인 기계의 평균대수) 운행 중인 평균 기계 대수는 다음과 같다.

$$L_{up} = \sum_{k=0}^{m} kP_{m-k} = \frac{\sum_{k=0}^{m} \frac{ke^{-\mu/\lambda}(\mu/\lambda)^k}{k!}}{\sum_{i=0}^{m} \frac{e^{-\mu/\lambda}(\mu/\lambda)^i}{i!}} = \frac{\mu}{\lambda}(1 - P_0)$$

위 식이 성립되는 것을 보기 위해 평균이 μ/λ인 포아송분포를 $f_k = \frac{e^{-(\mu/\lambda)}(\mu/\lambda)^k}{k!}$ 라고 하고, 그 분포함수를 $F_k = \sum_{i=0}^{k} f_i$로 표현하자. 그러면 $kf_k = (\mu/\lambda)f_{k-1}$의 관계가 성립함을 알 수 있다.

$$\sum_{k=0}^{m} \frac{\frac{ke^{-(\mu/\lambda)}(\mu/\lambda)^k}{k!}}{F_m} = \frac{\mu}{\lambda}\sum_{k=1}^{m} \frac{f_{k-1}}{F_m} = \frac{\mu}{\lambda}\frac{\left[\sum_{k=0}^{m} f_k - f_m\right]}{F_m}$$

$$= \frac{\mu}{\lambda}\left[1 - \frac{f_m}{F_m}\right] = \frac{\mu}{\lambda}[1 - P_0]$$

Performance Measure (2/2)

(수리 대기 중인 기계의 평균대수)

$$L_{q,\,down} = m - L_{up} - L_{repair} = m - \left(1 + \frac{\mu}{\lambda}\right)(1 - P_0)$$

(고장나 있는 기계의 평균대수)

$$L_{down} = L_{q,\,down} + L_{repair} = m - \frac{\mu}{\lambda}(1 - P_0)$$

(도착률(입력률)) 안정상태에서는 도착률과 이탈률이 같아야 한다.

$$\lambda_e = \lambda L_{up} = \mu(1 - P_0)$$

(평균대기시간) $W_q = \dfrac{L_{q,\,down}}{\lambda_e} = \dfrac{m}{\mu(1 - P_0)} - \dfrac{1}{\mu}\left(1 + \dfrac{\mu}{\lambda}\right)$ 〈Little의 법칙〉

(평균체재시간) $W = \dfrac{L_{down}}{\lambda_e} = \dfrac{m}{\mu(1 - P_0)} - \dfrac{1}{\lambda} = W_q + \dfrac{1}{\mu}$ 〈Little의 법칙〉

Time Average (1/2)

- 도착시점확률과 임의시점확률

한 가지 주의할 점은 수리소로의 도착과정은 포아송과정이 아니라는 것이다. 왜냐하면 몇 대가 가동 중이냐에 따라 수리소로의 도착률이 달라지기 때문이다. 이러한 도착과정을 유사랜덤 도착과정(quasi-random arrival process)이라고 부른다. 따라서 PASTA는 적용될 수 없고 도착하는 고객이 보는 고객수확률은 임의시점에서의 고객수확률과 다르다.

도착과정이 포아송과정이 아니므로 임의의 고객이 도착하면서 n명의 고객을 볼 개수평균확률 $\overline{\pi}_n$은 임의시점에서 외부인이 n명을 보는 시간평균확률 P_n과 다르다. 즉,

$$\overline{\pi}_n \neq P_n$$

Time Average (2/2)

그렇지만 다음의 관계가 성립한다

$$\bar{\pi}_n(m) = P_n(m-1)$$

위에서 $\bar{\pi}_n(m)$은 m대의 기계로 구성되는 시스템에서 도착하는 임의의 고객이 n명의 고객을 볼 확률이고 $P_n(m-1)$은 $(m-1)$대로 구성되는 시스템에서 외부인이 임의의 시점에서 n명의 고객을 볼 확률이다.

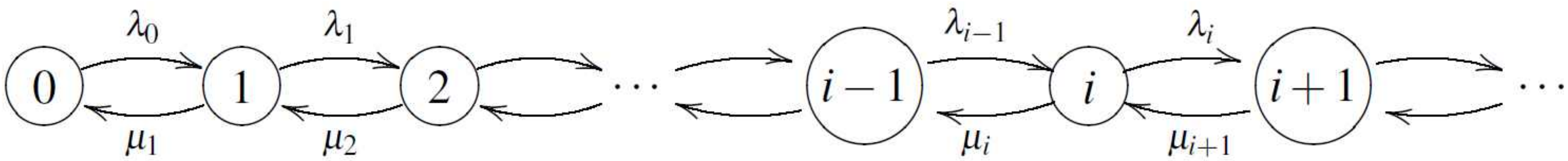
도착하는 고객은 그 자신을 볼 수 없으므로 위의 식이 의미하는 바는 '임의의 도착고객은 마치 외부인의 역할을 한다'는 것이다. 이 해석에 의하면 도착하는 고객은 결국 그 자신을 제외한 시간평균확률을 보는 셈이다. 이 사실로부터 포아송도착이 아니더라도 시간평균을 볼 수가 있음을 알 수 있다(ASTA, Arrivals that See Time Averages, Melamed and Whitt[1990], Green and Melamed[1990]을 참조할 것).

The *M/M/c/m/m* Queue

다음과 같은 파라미터들을 갖는 출생사멸과정이다.

$$\lambda_n = \begin{cases} (m-n)\lambda, & (0 \le n \le m-1), \\ 0, & (n \ge m), \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & (1 \le n \le c), \\ c\mu, & (c+1 \le n \le m). \end{cases}$$



- Birth-death process

$$\pi_0 = \left(1 + \sum_{i=1}^{\infty} \frac{\prod_{k=0}^{i-1} \lambda_k}{\prod_{k=1}^{i} \mu_k}\right)^{-1} \qquad \pi_i = \frac{\prod_{k=0}^{i-1} \lambda_k}{\prod_{k=1}^{i} \mu_k} \pi_0$$

Steady-state Probability

(안정상태 고객수 확률(고장나 있는 기계 대수))

$$P_0 = \left[\sum_{j=0}^{c-1} \frac{m!}{(m-j)!\,j!} (\lambda/\mu)^j + \sum_{j=c}^{m} \frac{m!}{(m-j)!\,c!\,c^{j-c}} (\lambda/\mu)^j \right]^{-1}$$

$$P_n = \begin{cases} \dfrac{m!}{(m-n)!\,n!} (\lambda/\mu)^n P_0, & (1 \le n \le c), \\[2ex] \dfrac{m!}{(m-n)!\,c!\,c^{n-c}} (\lambda/\mu)^n P_0, & (c+1 \le n \le m). \end{cases}$$

Performance Measure (1/2)

(수리 대기 중인 기계의 평균대수)

$$L_{q,\,down} = \sum_{n=c}^{m} (n-c) P_n$$

(고장나 있는 기계의 평균대수)

$$L_{down} = \sum_{n=0}^{m} nP_n = m - \sum_{k=0}^{m} kP_{m-k}$$

(수리 중인 기계의 평균대수(수리 대기 중인 기계 제외))

$$L_{repair} = L_{down} - L_{q,\,down} = \lambda_e E(S) = \lambda_e / \mu$$

Performance Measure (2/2)

(도착률(입력률)) $\lambda_e = \lambda L_{up} = \lambda(m - L_{down}) = \lambda \sum_{k=0}^{m} kP_{m-k}$

(평균대기시간) $W_q = \frac{L_{q,\,down}}{\lambda_e}$ 〈Little의 법칙〉

(평균체재시간) $W = \frac{L_{down}}{\lambda_e} = \frac{m}{\lambda \sum_{k=0}^{m} kP_{m-k}} - \frac{1}{\lambda}$ 〈Little의 법칙〉

(도착시점확률과 임의시점확률의 관계) 여전히 다음의 관계가 성립한다.

$$\overline{\pi}_n(m) = P_n(m-1)$$

2. $M^x/M/1$ 집단도착 대기행렬

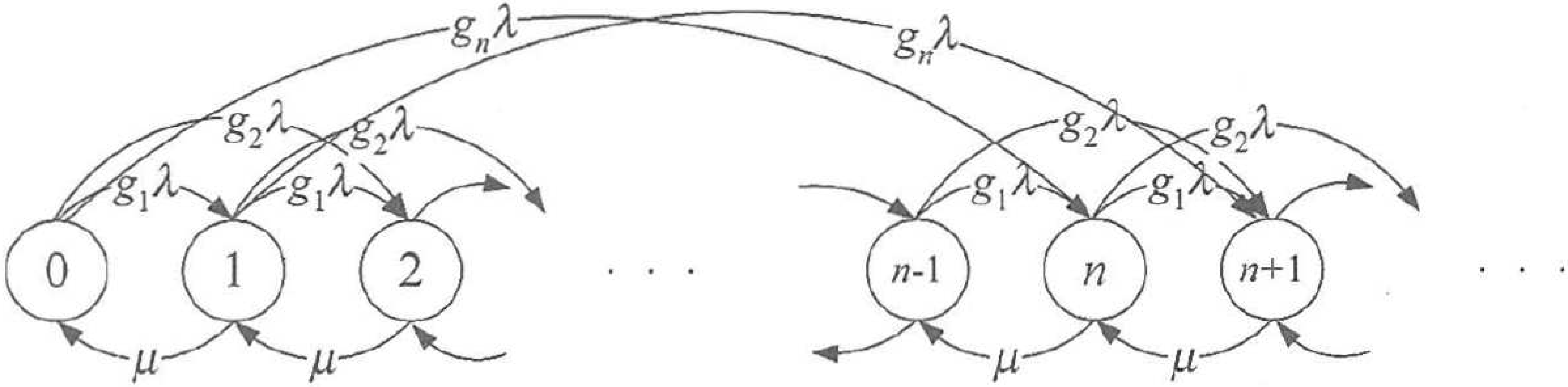
고객들이 택시를 타고 도착하는 상황을 생각하자.
택시들은 포아송과정으로 도착하고 택시에 타고 있는 고객수(집단크기, batch size)는 iid 확률변수일 때 이러한 도착과정을 복합포아송도착과정(compound Poisson arrival process)이라고 부른다

$M^X/M/1$ 대기행렬시스템은 고객들이 복합포아송과정으로 도착하며
서비스시간은 지수분포를 따르는 대기행렬시스템이다. 고객들은 한 명씩 서비스 받는다.

일반적으로 고객들이 집단으로 도착하는 대기행렬을 집단도착 대기행렬(batch arrival queue, bulk arrival queue)이라고 부른다.

State Transition Diagram

임의의 도착집단의 크기를 G, 그 확률을 $Pr(G=k)=g_k$, $(g_0=0)$이라고 하자.
고객수과정은 연속시간 마코프체인(CTMC)이다.



$$\frac{d}{dt}P_0(t)=-\lambda P_0(t)+\mu P_1(t)$$

$$\frac{d}{dt}P_n(t)=-(\lambda+\mu)P_n(t)+\mu P_{n+1}(t)+\lambda\sum_{k=1}^{n}P_{n-k}(t)g_k,\quad (n\geq 1)$$

안정상태방정식은 다음과 같다.

$$0=-\lambda P_0+\mu P_1$$

$$0=-(\lambda+\mu)P_n+\mu P_{n+1}+\lambda\sum_{k=1}^{n}P_{n-k}g_k,\quad (n\geq 1)$$

Steady-state Probability

다음과 같은 PGF들을 정의하자.

$$P(z)=\sum_{n=0}^{\infty}P_n z^n \qquad G(z)=\sum_{n=1}^{\infty}g_n z^n$$

시스템방정식에 z^n을 곱하여 더한 후 다음을 이용하면,

$$\sum_{n=1}^{\infty}\sum_{k=1}^{n}P_{n-k}g_k z^n=\sum_{k=1}^{\infty}g_k z^k\sum_{n=k}^{\infty}P_{n-k}z^{n-k}=G(z)P(z)$$

다음을 얻는다

$$P(z)=\frac{\mu(1-z)P_0}{\mu(1-z)-\lambda z[1-G(z)]}$$

Performance Measure

$P(z)|_{z=1}=1$을 이용하면 P_0는 다음과 같다.

$$P_0=1-\rho,\ (\rho=\lambda E(G)E(S))$$

위에서 ρ는 단위시간당 도착하는 고객들이 갖고 들어오는 평균일량이다.

$$P_{busy}=\rho$$

PASTA에 의하면 이 확률은 도착하는 집단이 바쁜 서어버를 볼 확률과 같다. 따라서 ρ는 또한 임의의 시점에 서비스 받고 있는 평균고객수이다.

평균고객수 $$L=E(N)=\frac{d}{dz}P(z)\Big|_{z=1}=\frac{\lambda[Var(G)+E^2(G)+E(G)]}{2[\mu-\lambda E(G)]}$$

평균대기고객수 $$L_q=E(N_q)=L-\rho=L-\lambda E(G)E(S)$$

3. $M/M^B/1$ 집단서비스 대기행렬

다음과 같은 집단서비스 대기행렬 (batch(bulk) service queue)을 생각하자.

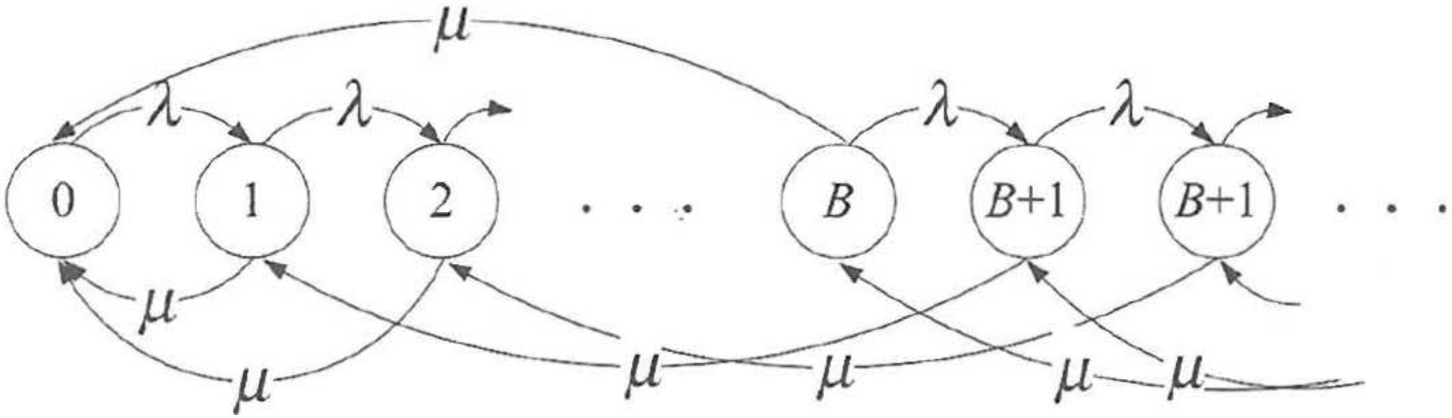
(i) 고객들은 한 명씩 포아송과정으로 도착한다.

(ii) 서어버는 최대 B명의 대기고객을 불러들여 서비스할 수 있다. 만약 B명 미만의 고객이 있으면 고객 전부를 불러들여 서비스한다.

(iii) 서비스 도중 도착하는 고객들은 여분의 자리를 채운다. 예를 들어 $B=5$인데 현재 3명만 서비스를 받고 있으면, 서비스 중에 도착하는 고객들은 나머지 2개의 빈자리를 채운다. 그렇지만 그들의 서비스는 이미 서비스 받고 있던 고객들과 함께 끝난다.

State Transition Probability

고객수과정에 대한 전이율다이어그램.



시스템방정식

$$\frac{d}{dt}P_0(t) = \mu\sum_{k=1}^{B}P_k(t) - \lambda P_0(t)$$

$$\frac{d}{dt}P_n(t) = \lambda P_{n-1}(t) - (\lambda+\mu)P_n(t) + \mu P_{n+B}(t), \quad (n \geq 1)$$

안정상태방정식.

$$\lambda P_0 = \mu\sum_{k=1}^{B}P_k$$

$$(\lambda+\mu)P_n = \lambda P_{n-1} + \mu P_{n+B}, \quad (n \geq 1)$$

Stability Condition

안정상태 조건 $\lambda E(S) < B$, 또는 $\frac{\lambda}{B\mu} < 1$

즉, 단위시간당 시스템에 입력되는 평균일량(실행로드) $\lambda E(S)$는 서어버가 최대한 줄일 수 있는 단위시간당 일량 B보다 작아야 한다. 또는 동일하게 단위시간당 내보낼 수 있는 최대 평균고객수보다 도착하는 평균고객수가 작아야 한다.

Steady-state Probability

다음과 같은 PGF를 정의하자.

$$P(z) = \sum_{n=0}^{\infty} P_n z^n$$

n번째 식에 z^n을 곱하여 더하고 정리하면 다음을 얻는다.

$$P(z) = \frac{\sum_{i=0}^{B-1} P_i \cdot (z^i - z^B)}{\rho z^{B+1} - (1+\rho) z^B + 1}, \quad \left(\rho = \frac{\lambda}{\mu}\right)$$

$$= \frac{1 - z_0}{z - z_0}$$

z_0 : 단위원 밖 해 〈Rouche의 정리〉

$$P_n = \left(1 - \frac{1}{z_0}\right)\left(\frac{1}{z_0}\right)^n, \quad (n \geq 0)$$

이것은 $M/M/1$의 고객수확률 $P_n = (1-\rho)\rho^n$에서 ρ 대신에 $1/z_0$을 취한 것과 같다

Performance Measure

평균고객수 $$L = E(N) = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n(1-\rho) = \frac{\rho}{1-\rho} = \frac{1/z_0}{1-1/z_0} = \frac{1}{z_0 - 1}$$

평균체제시간 $$W = \frac{L}{\lambda} = \frac{1}{\lambda(z_0 - 1)}$$

평균대기시간 $$W_q = W - \frac{1}{\mu} = \frac{1}{\lambda(z_0 - 1)} - \frac{1}{\mu}$$

평균대기고객수 $$L_q = \lambda W_q = L - \rho, \quad (\rho = \lambda/\mu)$$

임의의 시점에 서비스 받고 있는 평균고객수

$$L_{serv} = E(N_{serv}) = \rho = \lambda/\mu = \lambda E(S)$$

도착과정은 포아송과정이므로 PASTA에 의하여 고객이 보는 고객수분포와 임의시점에의 고객수분포가 같다. 즉, $\overline{\pi}_n = P_n$

Summary

- The M/M/c/m/m Queue (기계간섭문제)
  - m대의 기계가 있는 공장; 고장난 기계는 수리공에 의해 수리되는 총 c명의 수리공; 수리소 입장에서의 대기행렬시스템
  - Performance Measure
    - 수리중인 기계의 평균 대수
    - 운행중인 기계의 평균 대수
    - 평균 대기 시간, 평균 체제 시간 등

- $M^x/M/1$ 집단도착 대기행렬
  - 고객들이 택시를 타고 도착하는 대기행렬시스템
  - Performance Measure: 평균 고객수, 평균 대기 고객수 등

- $M/M^B/1$ 집단서비스 대기행렬
  - 집단 서비스 대기행렬시스템(Batch service queue)
  - Performance Measure: 평균 고객수, 평기 대기 고객수, 평균 체제 시간 등