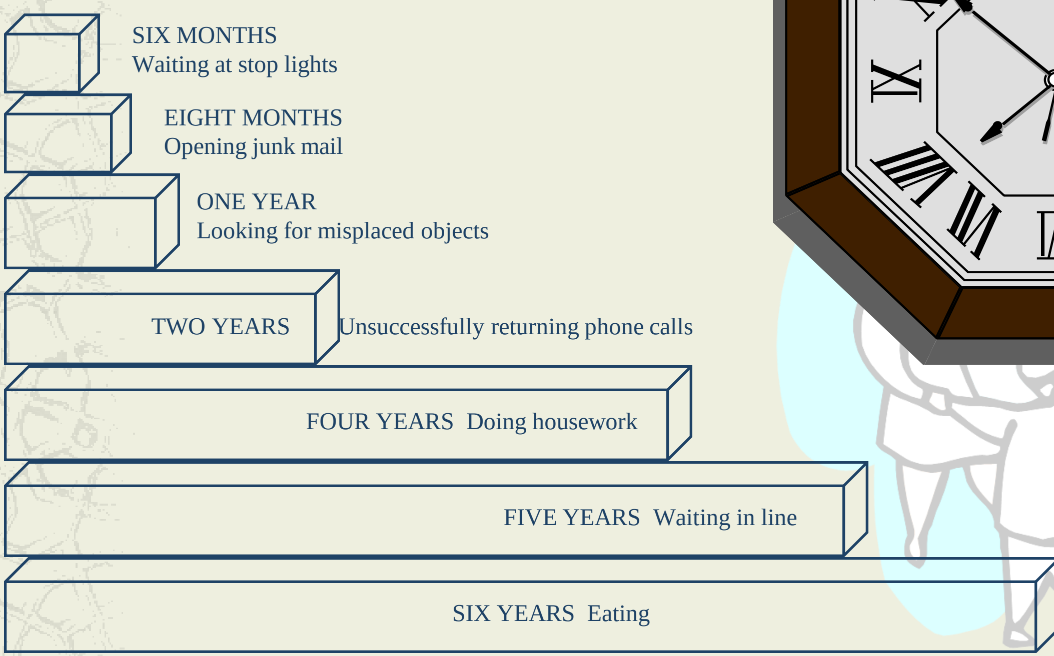
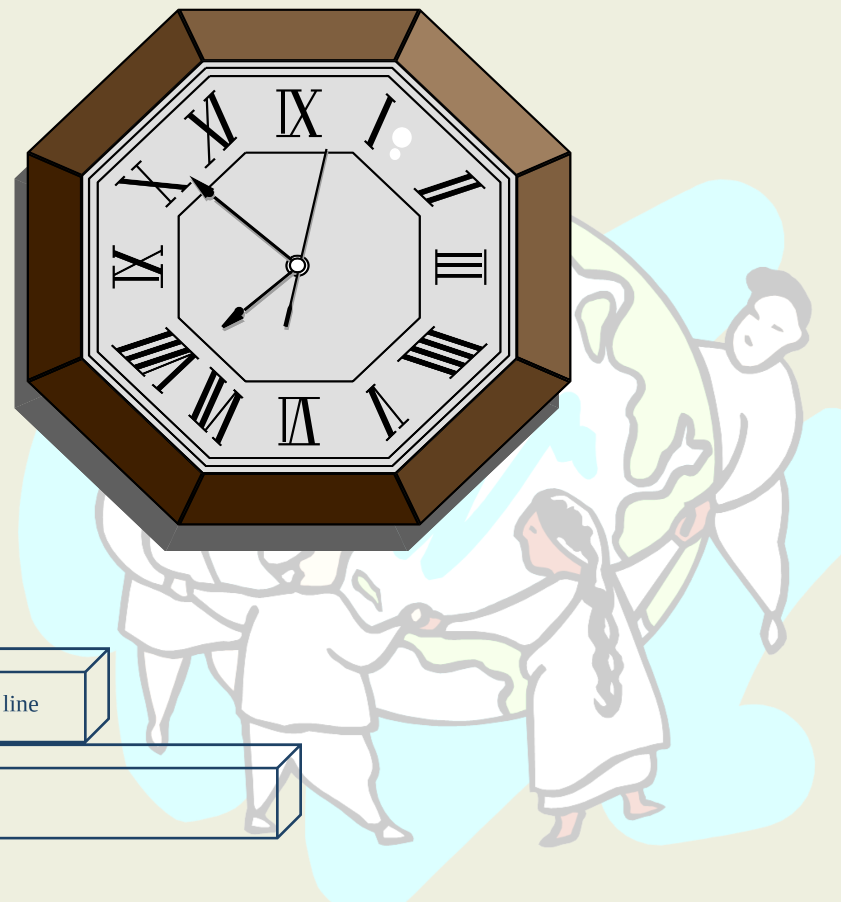
대기행렬 관리

Kwangtae Park, Korea University

Where the Time Goes

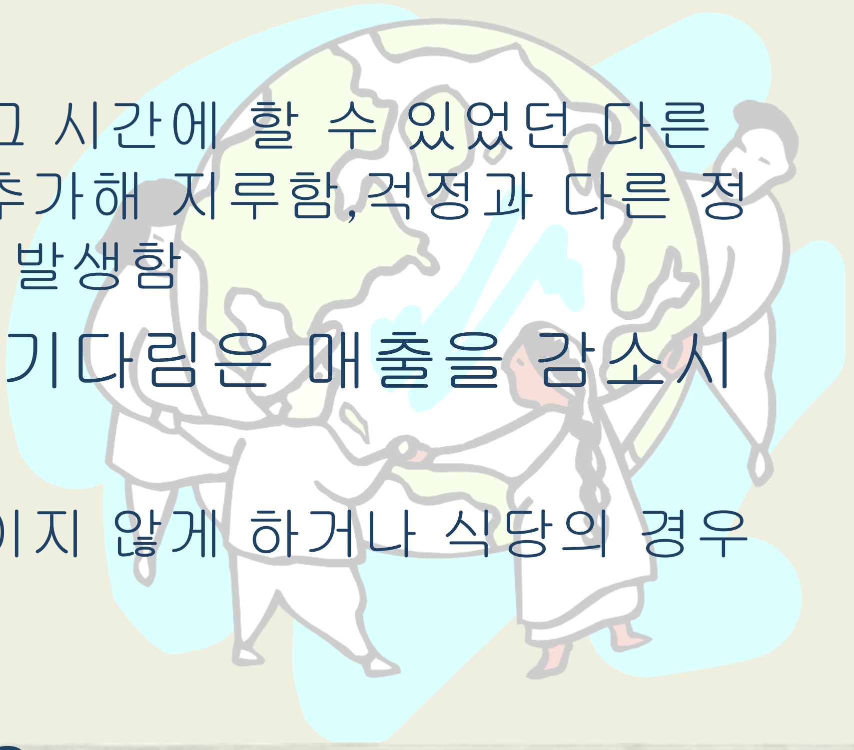
In a life time, the average American will spend--





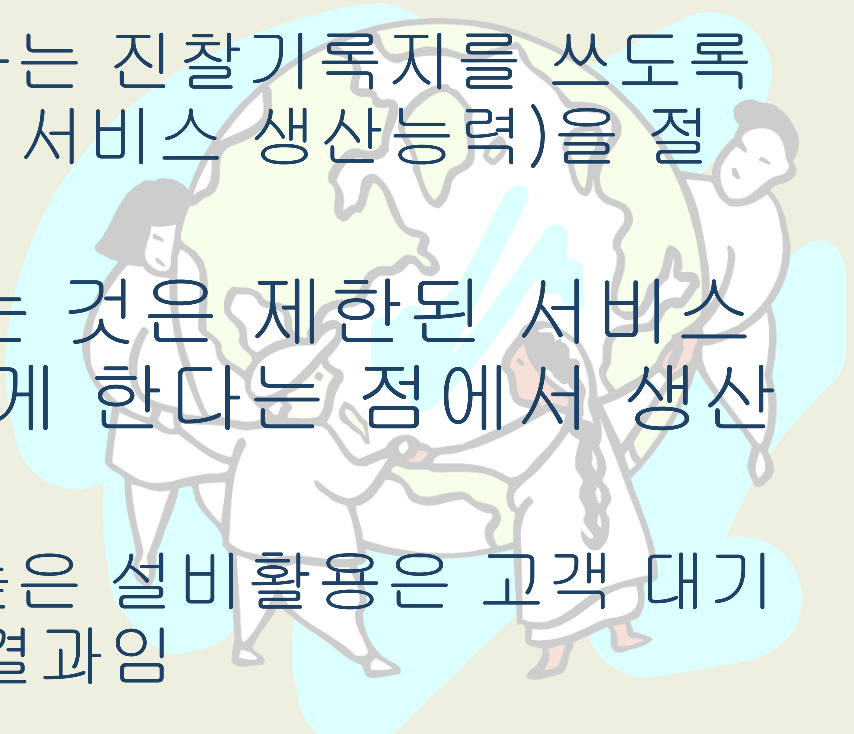
기다림의 경제학

- 기다림의 경제적 비용
  - 종업원을 기다리게 하는 비용은 비생산적 임금으로 측정될 수 있음
  - 외부고객의 대기비용은 그 시간에 할 수 있었던 다른 일의 가치임. 이 비용에 추가해 지루함,걱정과 다른 정신적 근심에 의해 비용이 발생함
- 경쟁시장에서의 지나친 기다림은 매출을 감소시킬 수 있음
  - 고객에게 대기행렬을 보이지 않게 하거나 식당의 경우 고객을 bar로 분산시킴



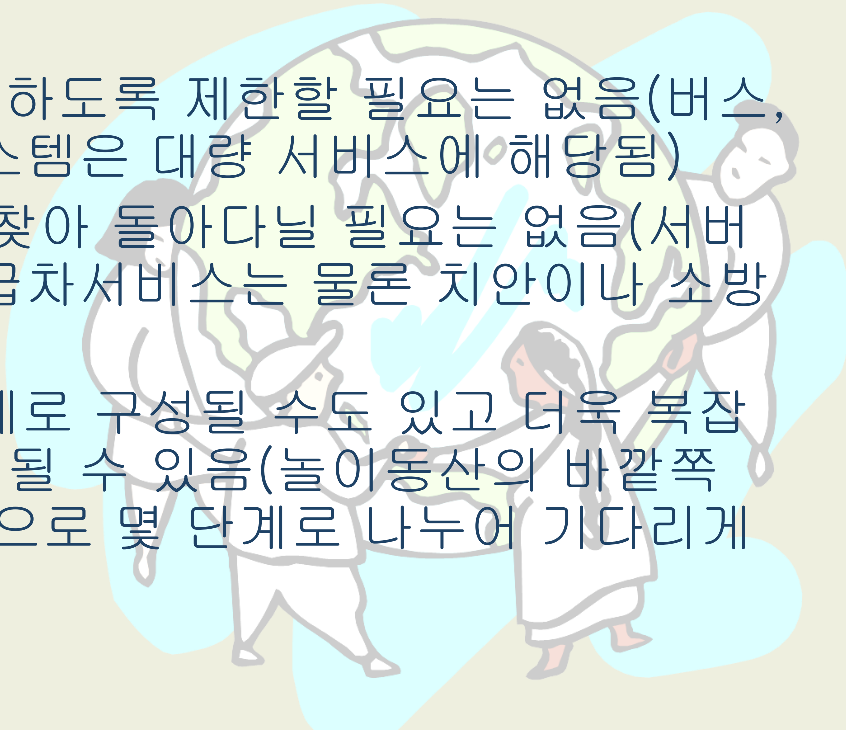
기다림의 경제학

- 고객은 서비스 프로세스에 참여할 수 있는 잠재력을 지닌 자원
  - 의사를 기다리는 환자는 진찰기록지를 쓰도록 하여 의사의 시간(즉, 서비스 생산능력)을 절약할 수 있음
- 고객을 기다리게 하는 것은 제한된 서비스 능력을 더 잘 활용하게 한다는 점에서 생산성 향상에 기여
  - 서비스 시스템에서 높은 설비활용은 고객 대기 비용의 대가를 치룬 결과임



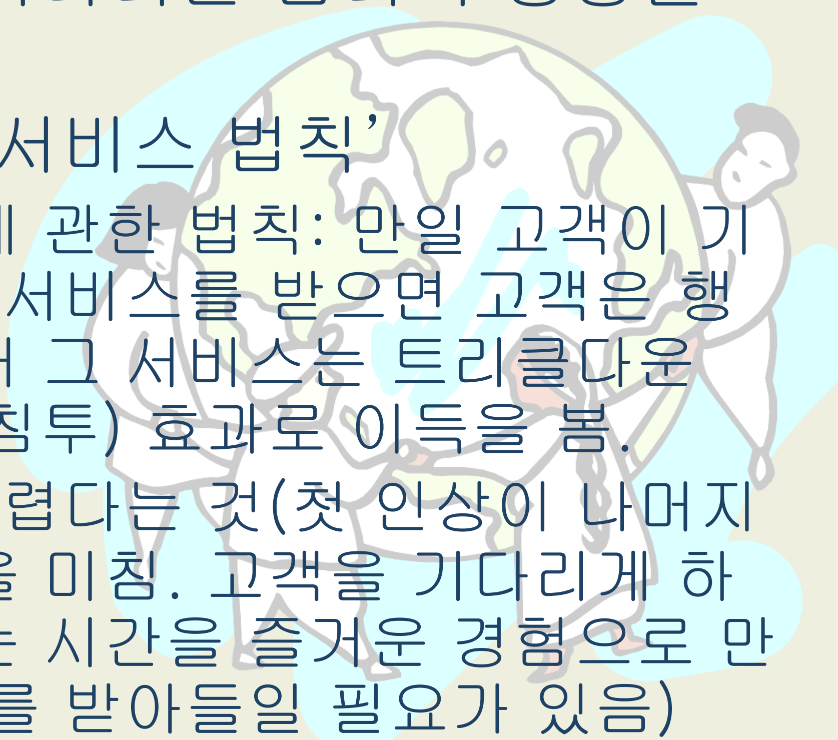
대기행렬 시스템

- 대기행렬 – 하나 혹은 그 이상의 서비스 제공자에게 서비스를 요구하며 기다리는 고객의 줄
- 대기행렬의 변형
  - 서버가 한 번에 한 고객만 상대하도록 제한할 필요는 없음(버스, 비행기, 승강기와 같은 운송시스템은 대량 서비스에 해당됨)
  - 고객이 언제나 서비스 설비를 찾아 돌아다닐 필요는 없음(서버가 실제로 고객을 찾아가는 구급차서비스는 물론 치안이나 소방 활동과 같은 공공서비스)
  - 서비스는 일련의 대기행렬 단계로 구성될 수도 있고 더욱 복잡한 대기행렬의 네트워크로 구성될 수 있음(놀이동산의 바깥쪽 걷기, 대기실, 놀이기구타기 순으로 몇 단계로 나누어 기다리게 함)



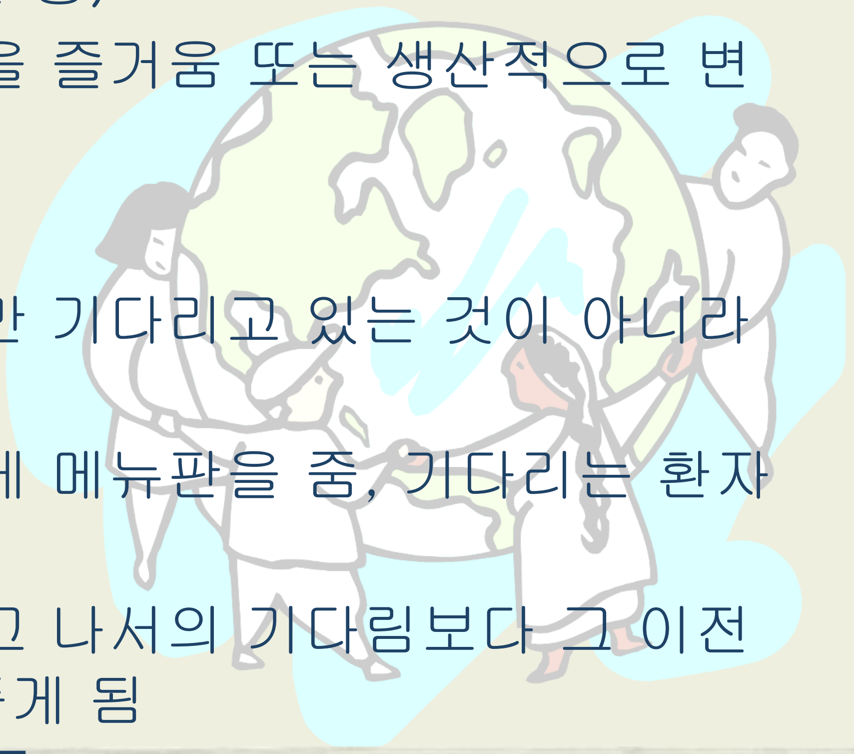
고객 대기관리를 위한 전략-1

- 기다림의 심리학
  - 서비스를 받기 위해 기다리는 심리적 영향을 고려하는 것
  - 마이스터의 두 가지 '서비스 법칙'
    - 고객의 기대와 인지에 관한 법칙: 만일 고객이 기대한 것보다 더 나은 서비스를 받으면 고객은 행복하고 만족하게 되어 그 서비스는 트리클다운(trickle-down, 하향침투) 효과로 이득을 봄.
    - '고객 따라잡기'가 어렵다는 것(첫 인상이 나머지 서비스 경험에 영향을 미침. 고객을 기다리게 하는 서비스는 기다리는 시간을 즐거운 경험으로 만드는 게 좋다는 충고를 받아들일 필요가 있음)



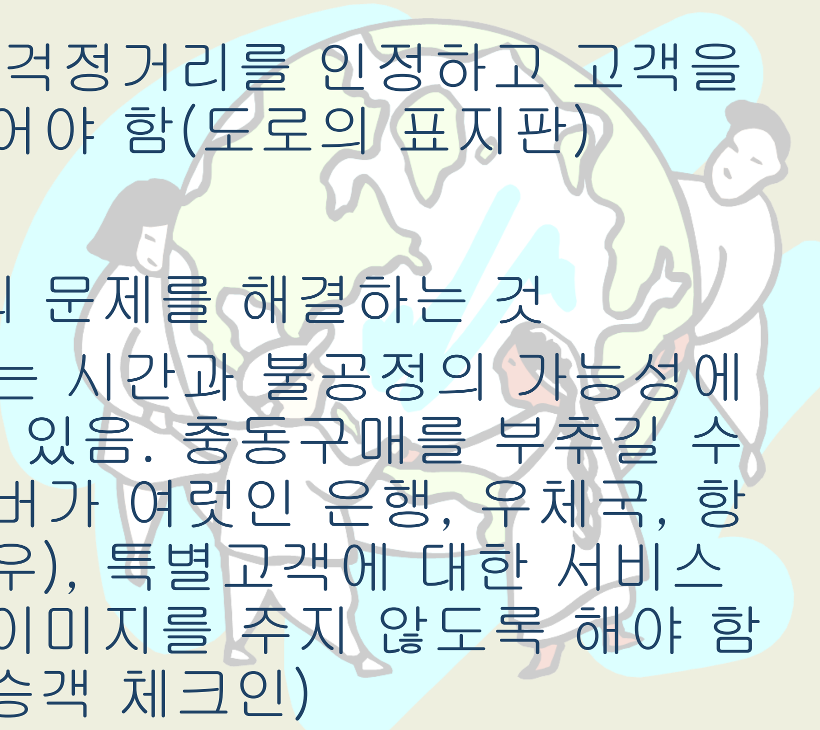
고객 대기관리를 위한 전략-2, 3

- 공허한 감정 누그러뜨리기
  - 대기시간을 긍정적인 방법으로 채워주는 것(편한 의자와 분위기를 좋게 하려고 새로 칠한 페인트, 대기실의 가구, 승강기 근처의 거울 등)
  - 서비스를 통해 대기시간을 즐거움 또는 생산적으로 변화시킬 수 있음
- 서비스의 시작을 알리기
  - 실제로는 기다리고 있지만 기다리고 있는 것이 아니라는 느낌을 주는 것
  - 식사를 기다리는 고객에게 메뉴판을 줌, 기다리는 환자에게 진찰기록지를 줌
  - 고객은 서비스가 시작하고 나서의 기다림보다 그 이전의 기다림에 더 불만을 품게 됨



고객 대기관리를 위한 전략-4, 5

- 터널의 끝이 보임을 알리기
  - 얼마나 기다려야만 하는가?에 대한 고객의 염려를 완화시켜 주는 것
  - 서비스 시작 전의 근심과 걱정거리를 인정하고 고객을 편하게 모실 전략을 만들어야 함(도로의 표지판)
- 자기 순서를 알려주기
  - 전달 과정에서의 공정성의 문제를 해결하는 것
  - 번호표를 나눠 줌(기다리는 시간과 불공정의 가능성에 대한 고객의 우려를 덜 수 있음. 충동구매를 부추길 수 있음), 한 줄로 세우기(서버가 여럿인 은행, 우체국, 항공사의 체크인 카운터 경우), 특별고객에 대한 서비스는 일반고객과 차별화된 이미지를 주지 않도록 해야 함(항공사 카운터의 1등석 승객 체크인)



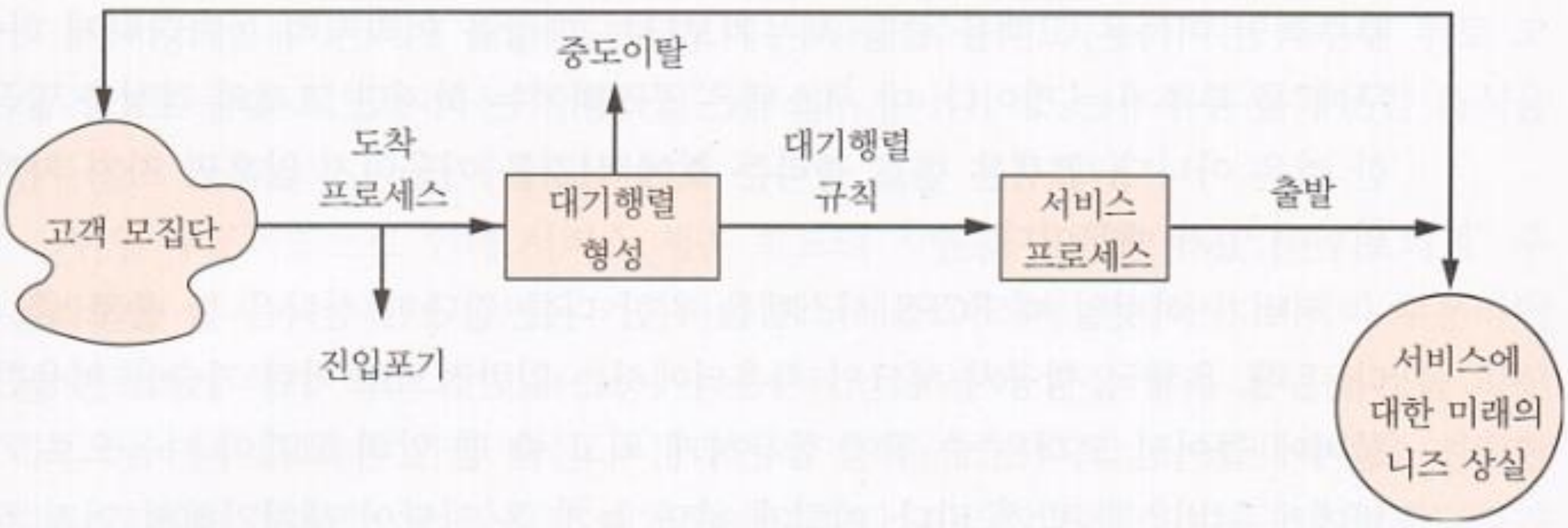
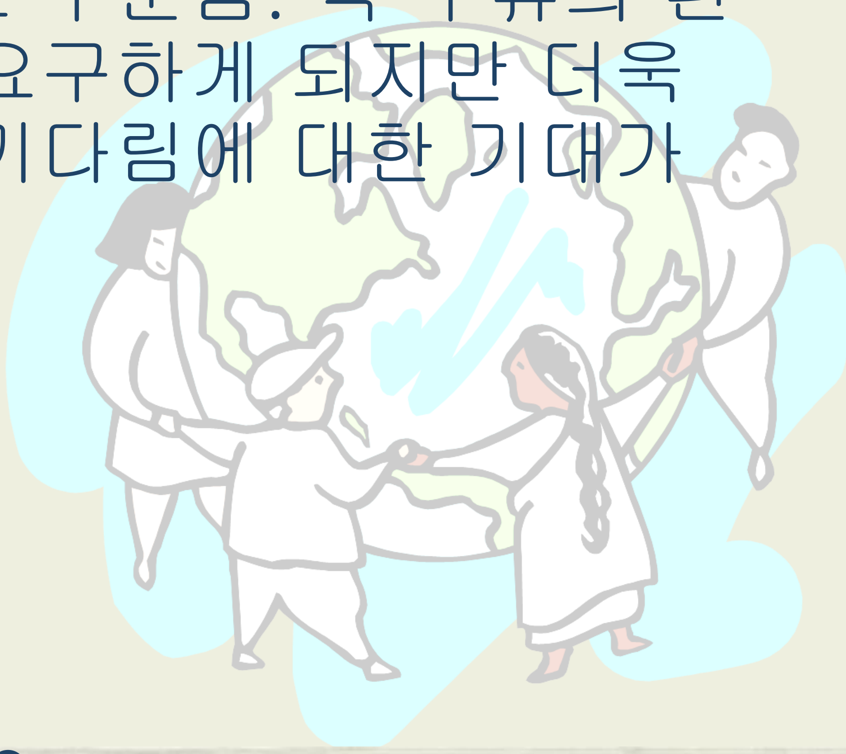


그림 12.1 대기행렬 시스템 모형도

고객 모집단

- 동질적일 필요는 없으며 몇몇 하위 집단으로 구성될 수 있음(외래환자의 도착은 입원환자, 예약환자, 응급환자로 구분됨. 각 부류의 환자는 다른 서비스를 요구하게 되지만 더욱 중요한 것은 각각의 기다림에 대한 기대가 크게 다르다는 점)



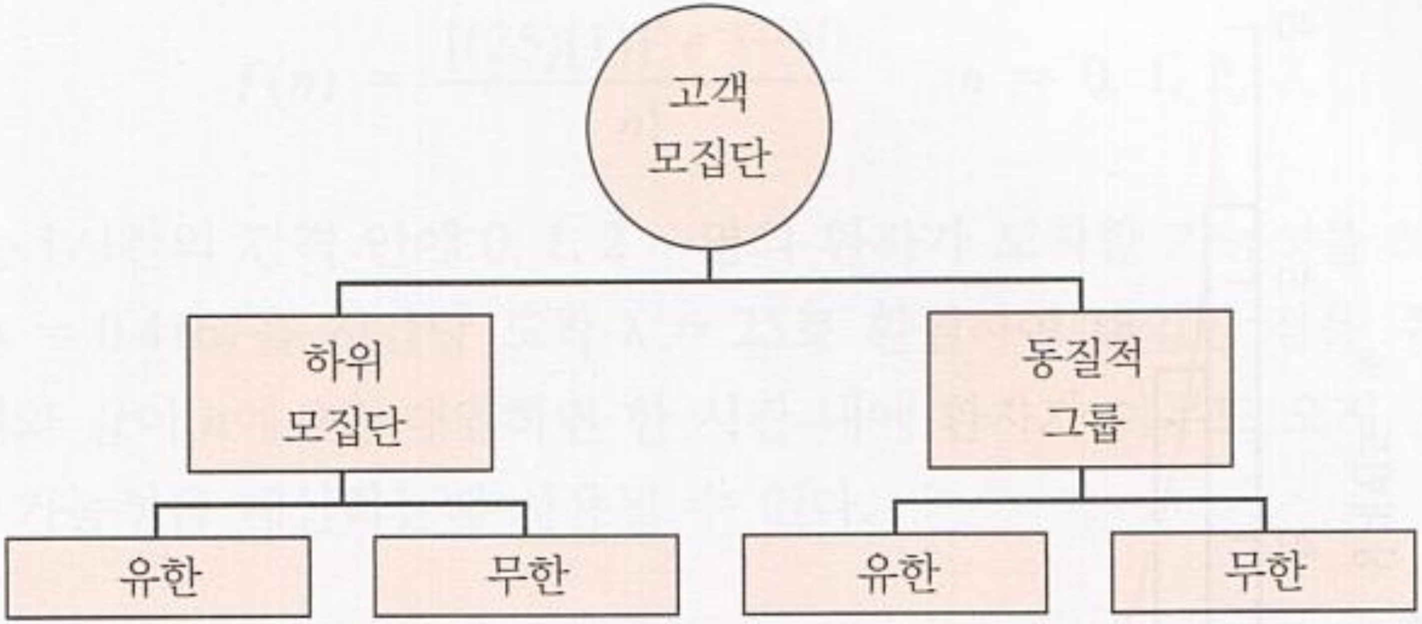


그림 12.2 고객 모집단의 분류

도착 프로세스

도착간격의 분포는 대개 지수분포임

지수분포 : 평균과 표준편차가 같음

$f(t)= \lambda e^{-\lambda t} \quad t \geq 0$

λ =주어진 기간 내의 평균 도착률

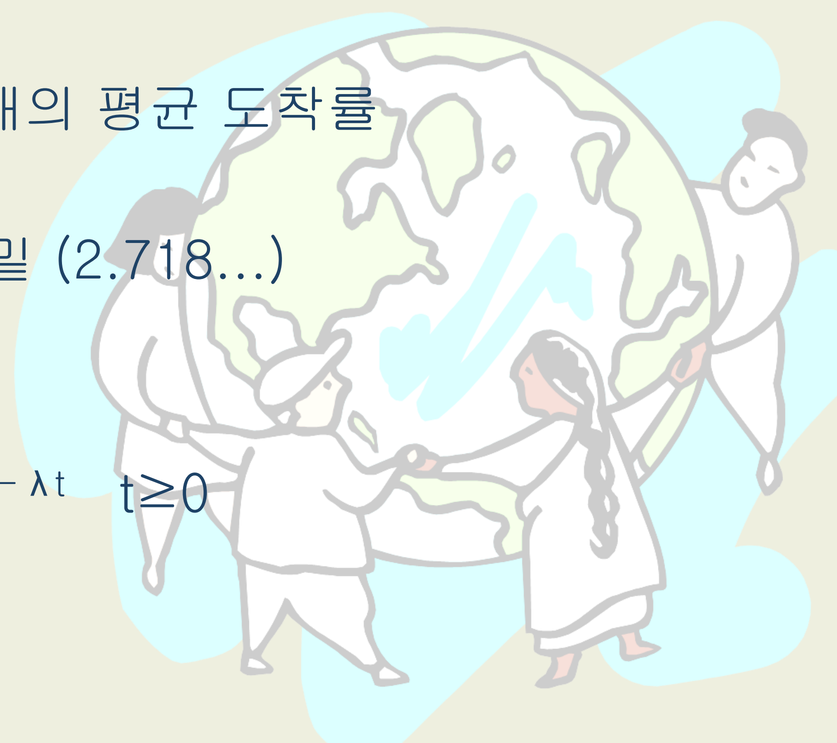
t = 도착간격

e = 자연 대수의 밑 (2.718...)

평균 = $1/\lambda$

분산 = $1/\lambda^2$

누적함수분포 $F(t) = 1- e^{-\lambda t} \quad t \geq 0$



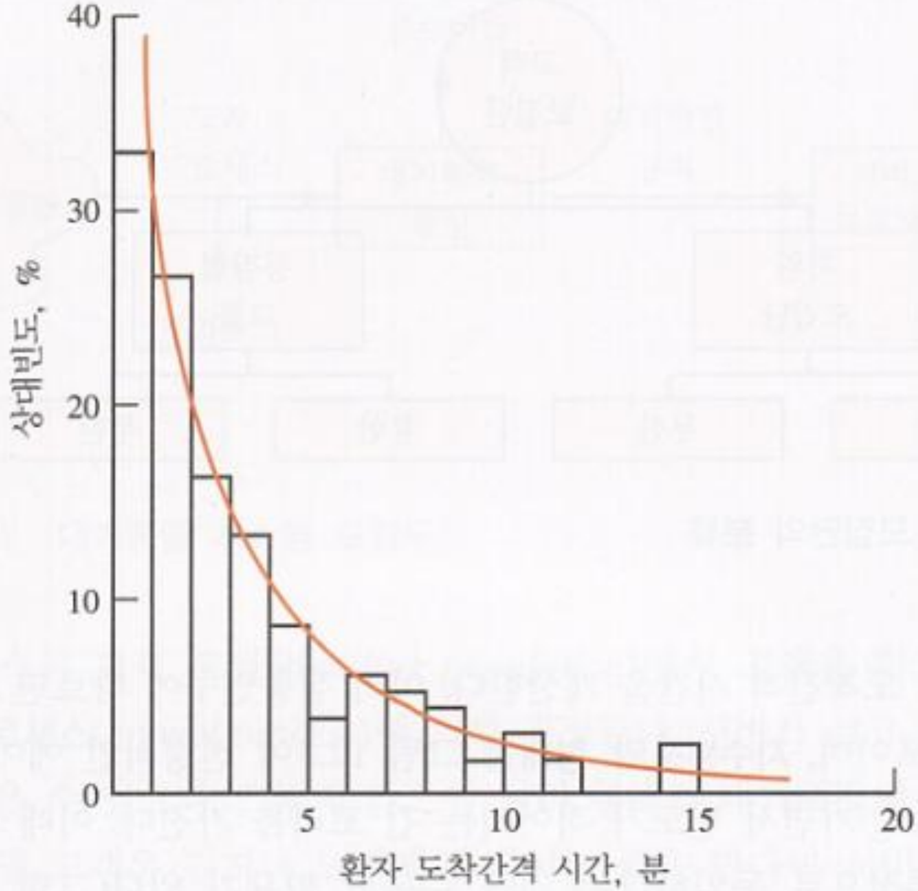


그림 12.3 대학병원에서의 환자 도착간격 시간의 분포

자료원: E. J. Rising, R. Baron, and B. Averill, "A Systems Analysis of a University Health-Service Outpatient Clinic." Reprinted with permission from *Operations Research* 21, no. 5, September? October 1973, p. 1038, Operations Society of America. No further reproduction permitted without the consent of the copyrighted owner.

도착 프로세스

시간간격 t 동안에 n번의 도착이 발생할 확률은 포아송 분포임

포아송 분포는 평균과 분산이 같음

$f(n) = (\lambda t)^n e^{-\lambda t}/n!$

λ =주어진 기간 내의 평균 도착률

t = 관심 기간의 수(보통 t=1)

n = 도착 건수(0,1,2..)

e = 자연 대수의 밑 (2.718...)

평균 = λt

분산 = λt

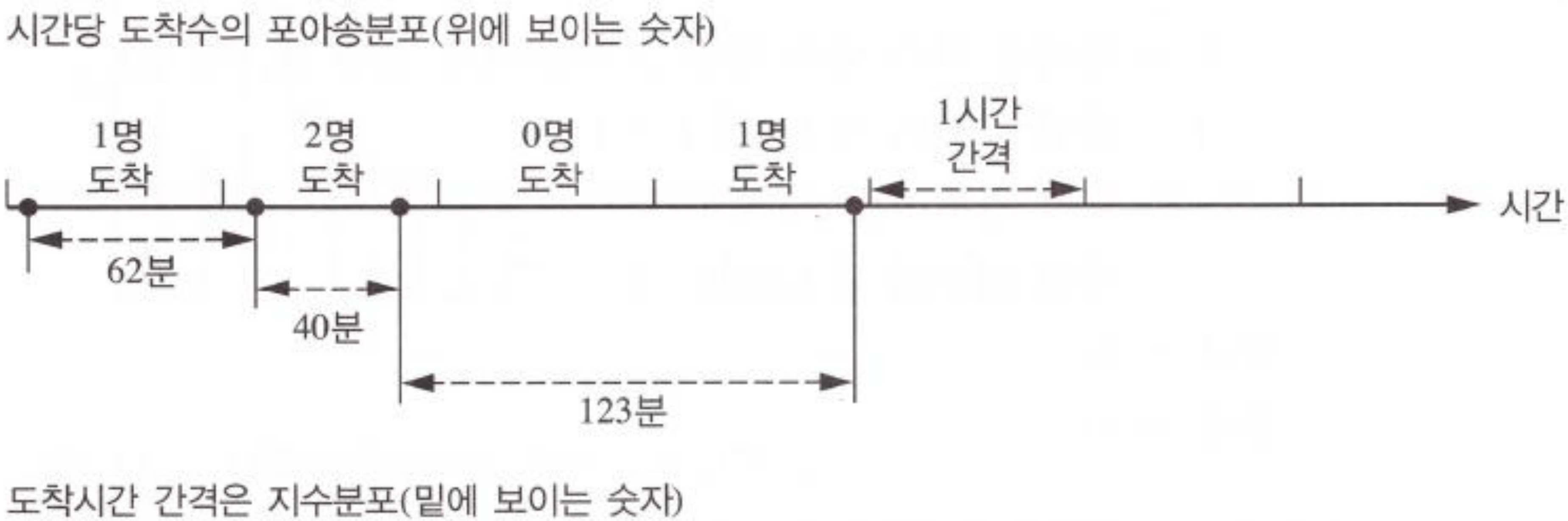


그림 12.4 포아송분포와 지수분포의 동일성

수요율이 시간에 따라 변하는 경우

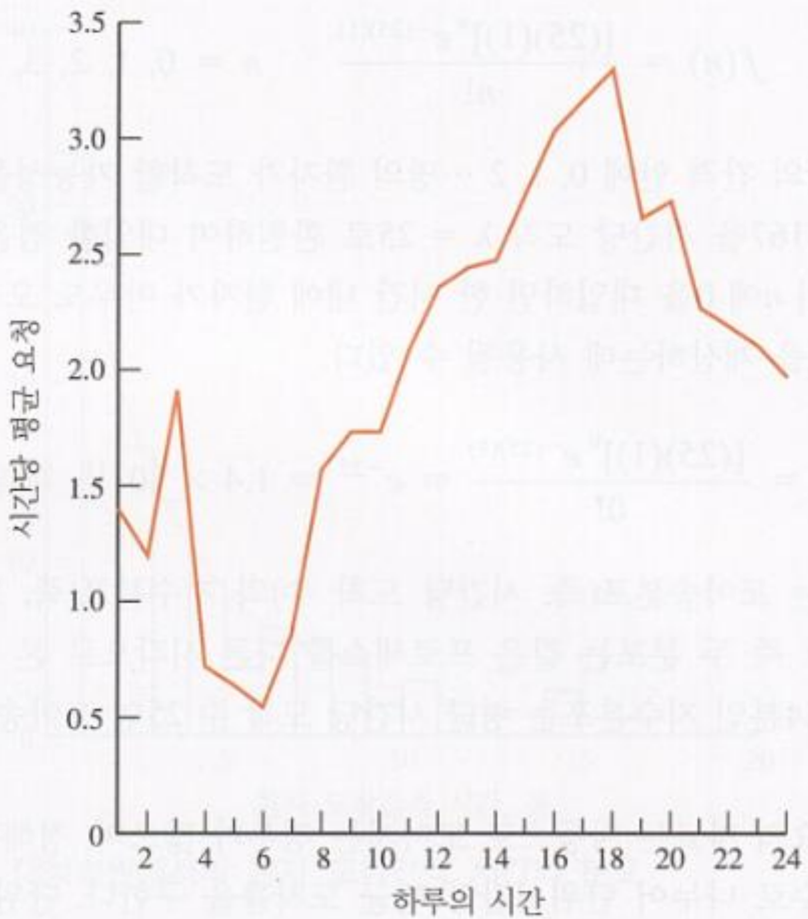


그림 12.5 하루 시간대별 구급차 요청

자료원: Reprinted with permission from James A. Fitzsimmons, "The Use of Spectral Analysis to Validate Planning Models," *Socio-Economic Planning* 8, no. 3, June 1974, p. 127. Copyright © 1974, Pergamon Press Ltd.

수요율이 시간에 따라 변하는 경우

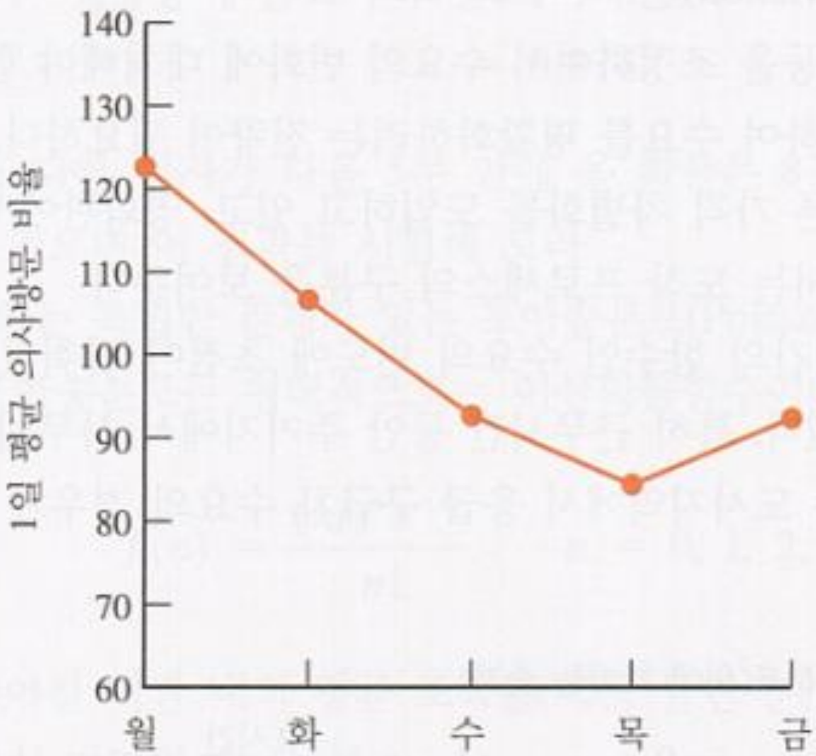


그림 12.6 일주일간 병원의 환자 도착 분포

자료원: E. J. Rising, R. Baron, and B. Averill, "A Systems Analysis of a University Health-Service with permission from *Operations Research* 21, no. 5, September? October 1973, p. 1035, Operations Society of America. No further reproduction permitted without the consent of the copyrighted owner.

수요율이 시간에 따라 변하는 경우

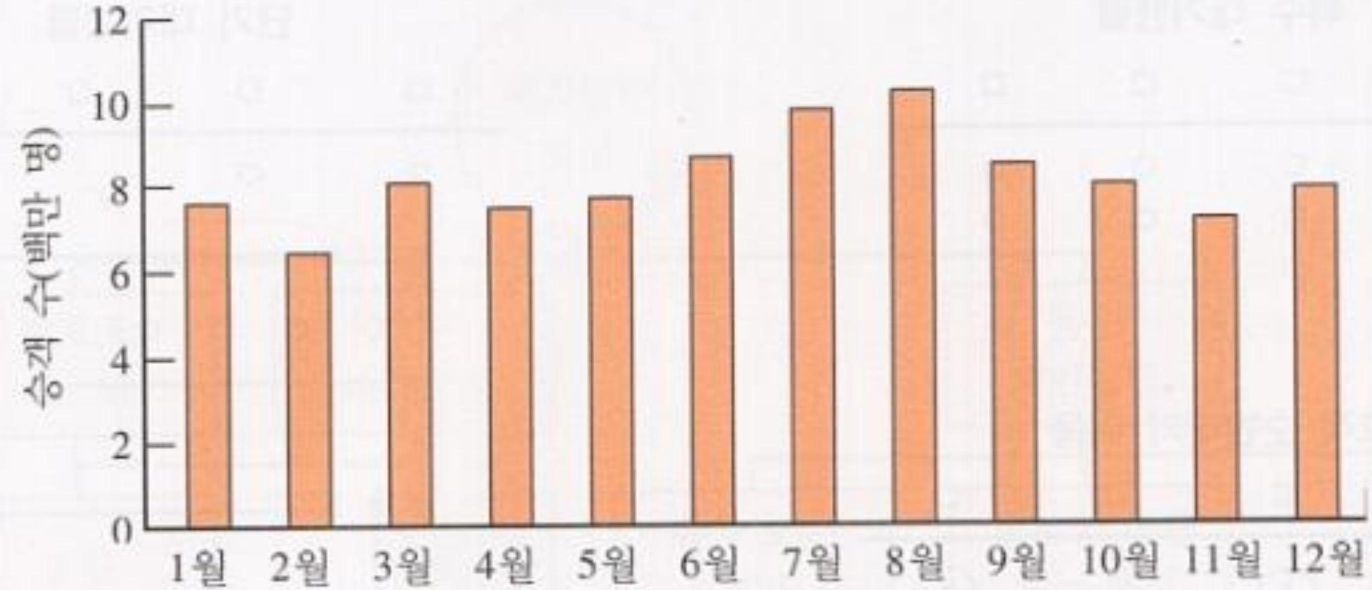


그림 12.7 미국과 전 세계 간 월별 국제 항공여객 분포, 1984

자료원: http://www.bts.gov/oai/international/table1.txt

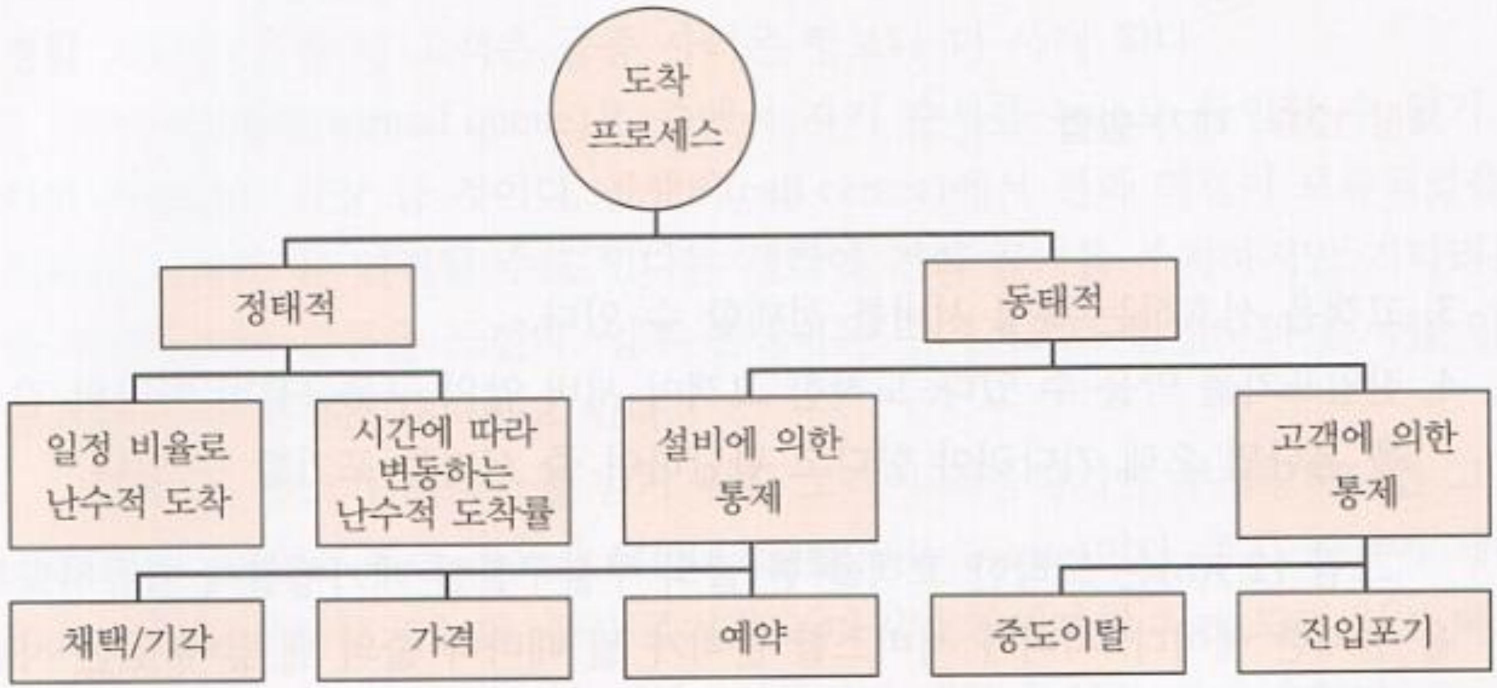
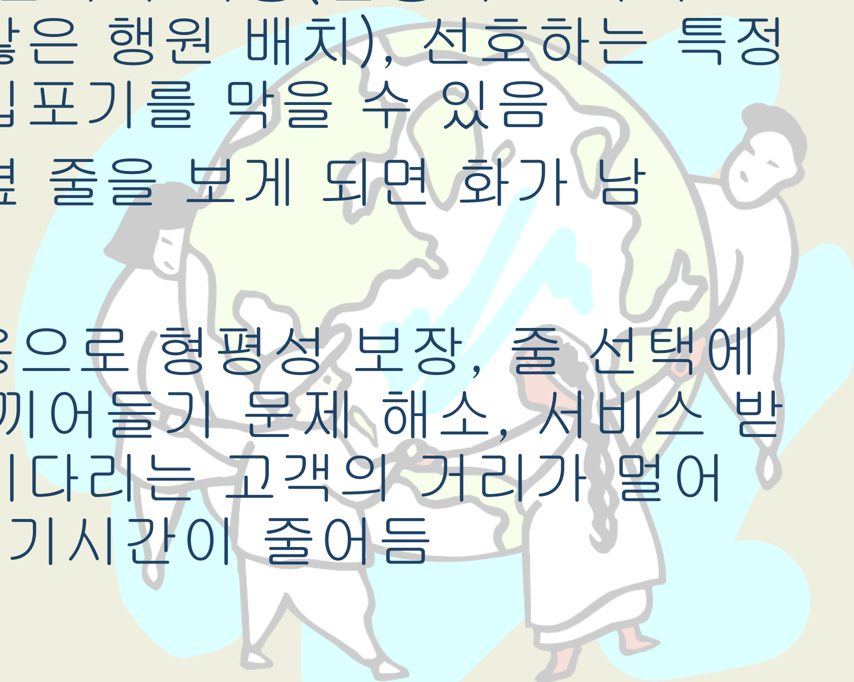


그림 12.8 도착 프로세스의 구분

복수 창구 서비스에 대한 세 가지 대기행렬 형성 방식

- 복수 대기행렬
  - 이점: 제공되는 서비스가 차별화될 수 있음(슈퍼마켓의 소품목 계산대), 노동의 분화가 가능(은행의 드라이브 인 창구에는 보다 경험 많은 행원 배치), 선호하는 특정 서버를 고객이 선택, 진입포기를 막을 수 있음
  - 단점: 빠르게 줄어드는 옆 줄을 보게 되면 화가 남
- 단일 대기행렬
  - 장점 : FCFS의 규칙 적용으로 형평성 보장, 줄 선택에 대한 고민이 필요 없음, 끼어들기 문제 해소, 서비스 받는 고객과 바로 뒤에서 기다리는 고객의 거리가 멀어 프라이버시 향상, 평균대기시간이 줄어듬
- 번호표 시스템
  - 단일 대기행렬의 변형으로 고객은 한결 자유로움



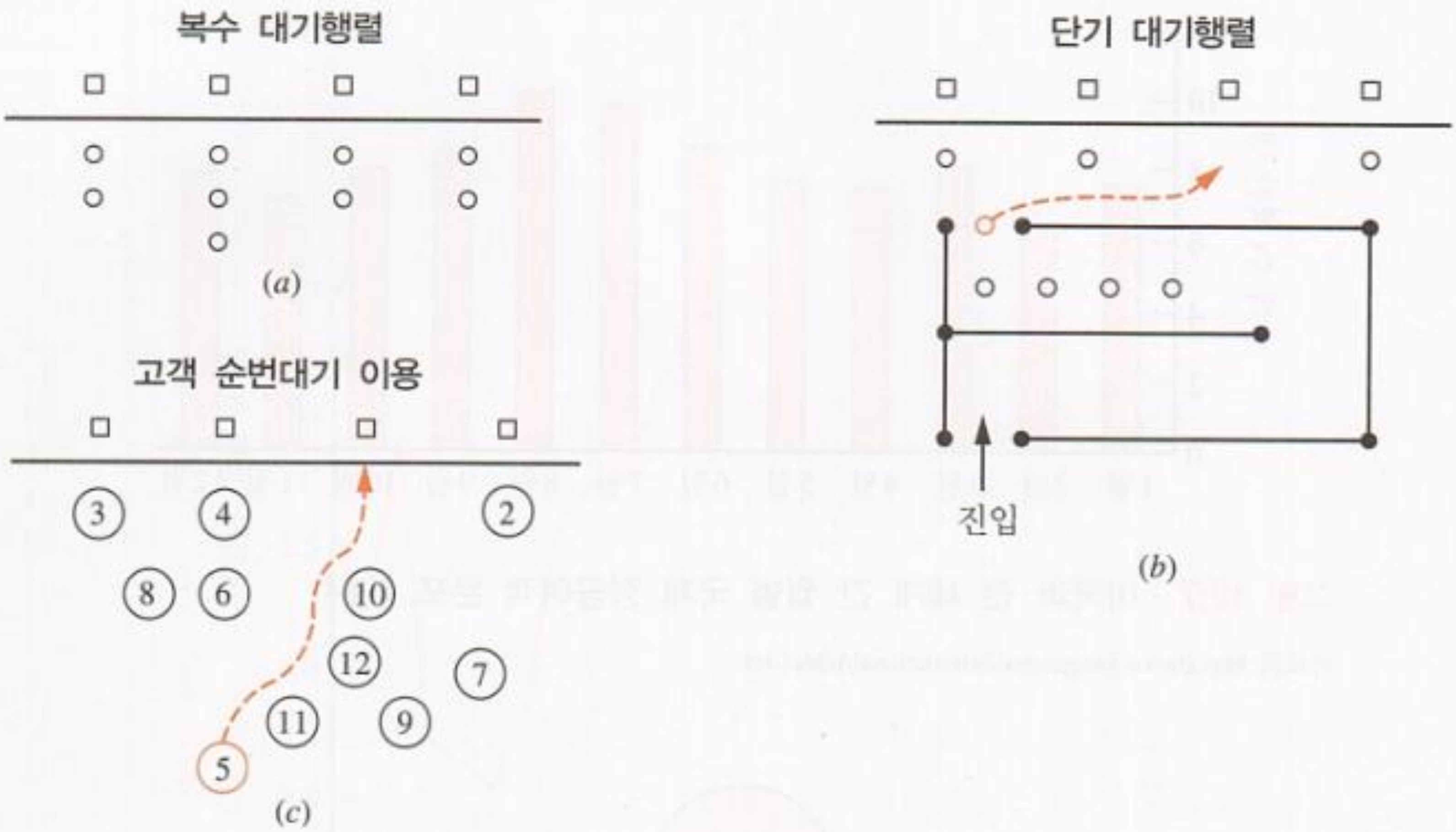


그림 12.9 대기 방법

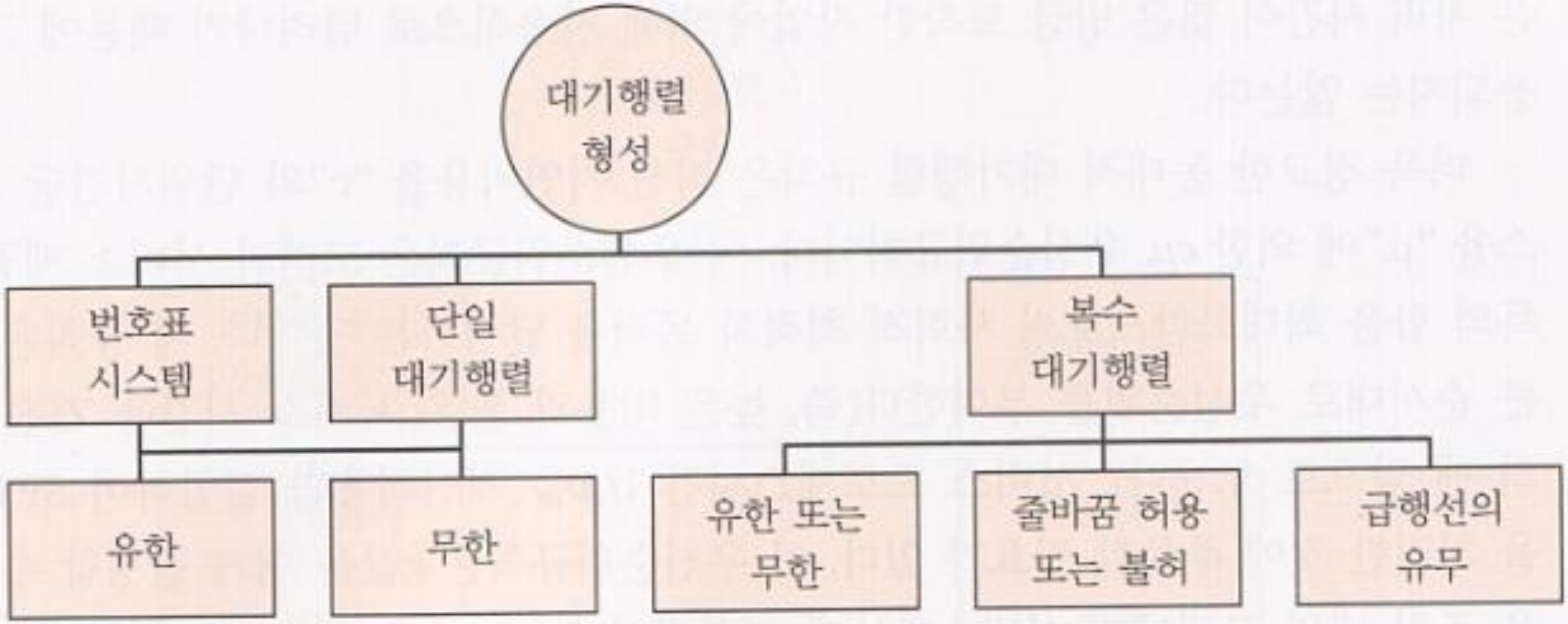
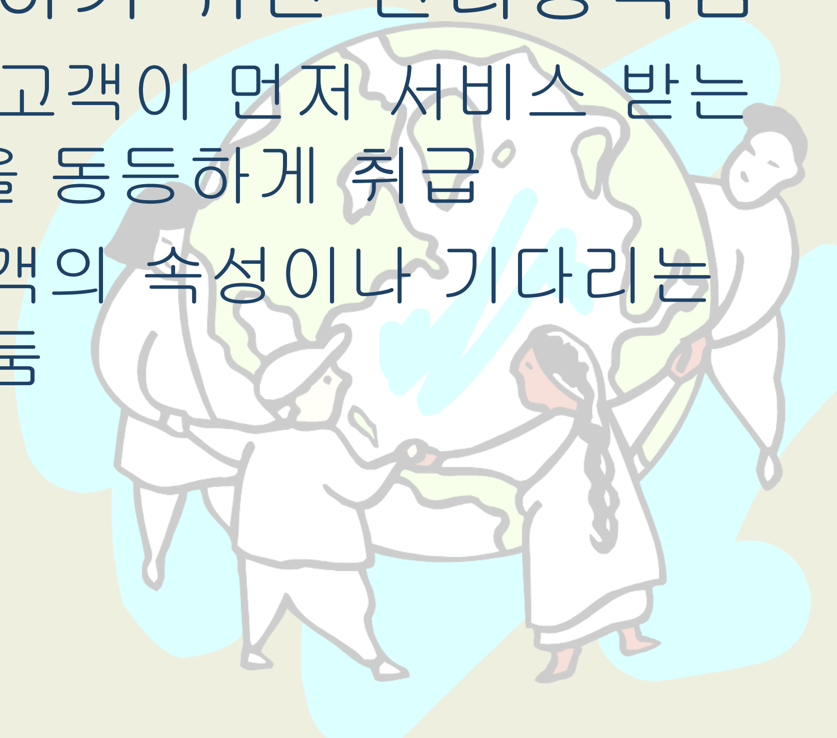


그림 12.10 대기행렬의 형성

대기행렬 규칙

- 대기행렬규칙은 대기행렬에서 다음에 서비스 받을 고객을 선정하기 위한 관리정책임
  - FCFS : 먼저 도착한 고객이 먼저 서비스 받는 규칙으로 모든 고객을 동등하게 취급
  - 동태적 대기규칙: 고객의 속성이나 기다리는 줄의 상태에 근거를 둠



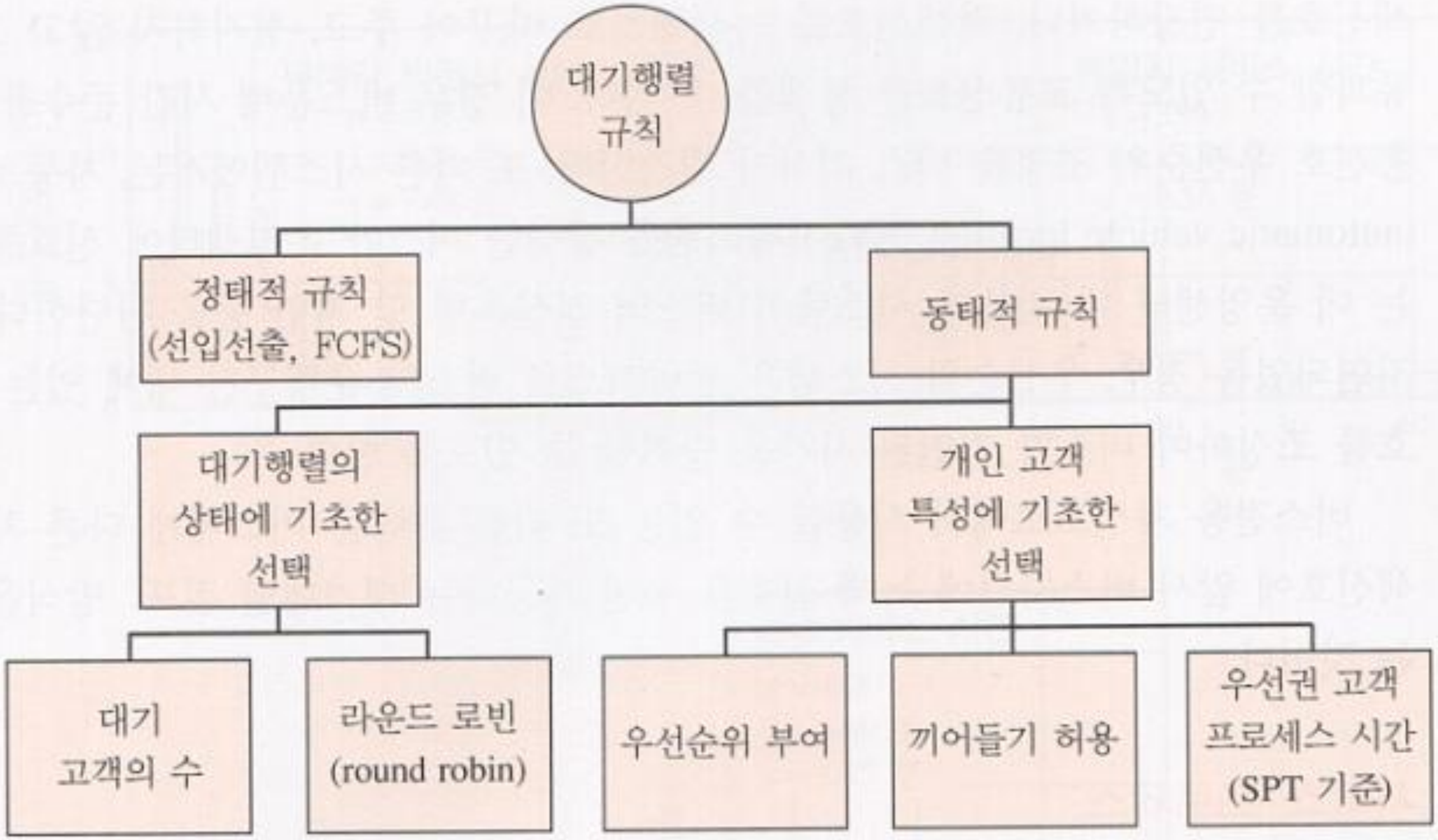


그림 12.11 대기행렬 규칙의 구분

서비스 프로세스

- 서비스 시간의 분포, 서버의 배열, 관리정책과 서버 행동 모두는 서비스 성과에 영향을 줌
- 서비스 시간의 분포는 고객의 요구와 서버 성과의 변화를 반영함

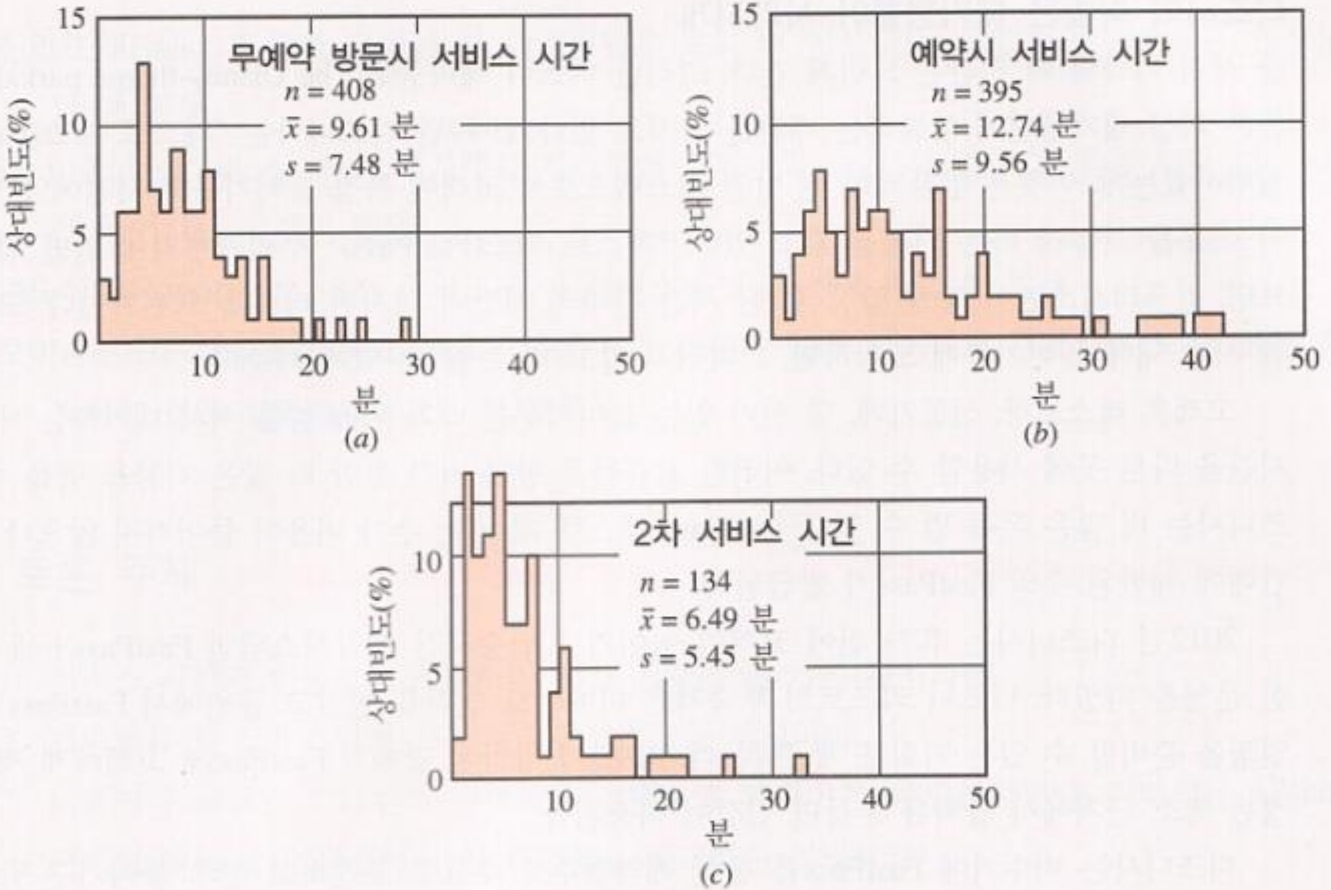


그림 12.12 외래환자 서비스 시간의 히스토그램

자료원: E. J. Rising, R. Baron, and B. Averill, "A Systems Analysis of a University Health-Service Outpatient Clinic." Reprinted with permission from *Operations Research* 21, no. 5, September? October 1973, p. 1039, Operations Society of America. No further reproduction permitted without the consent of the copyrighted owner.

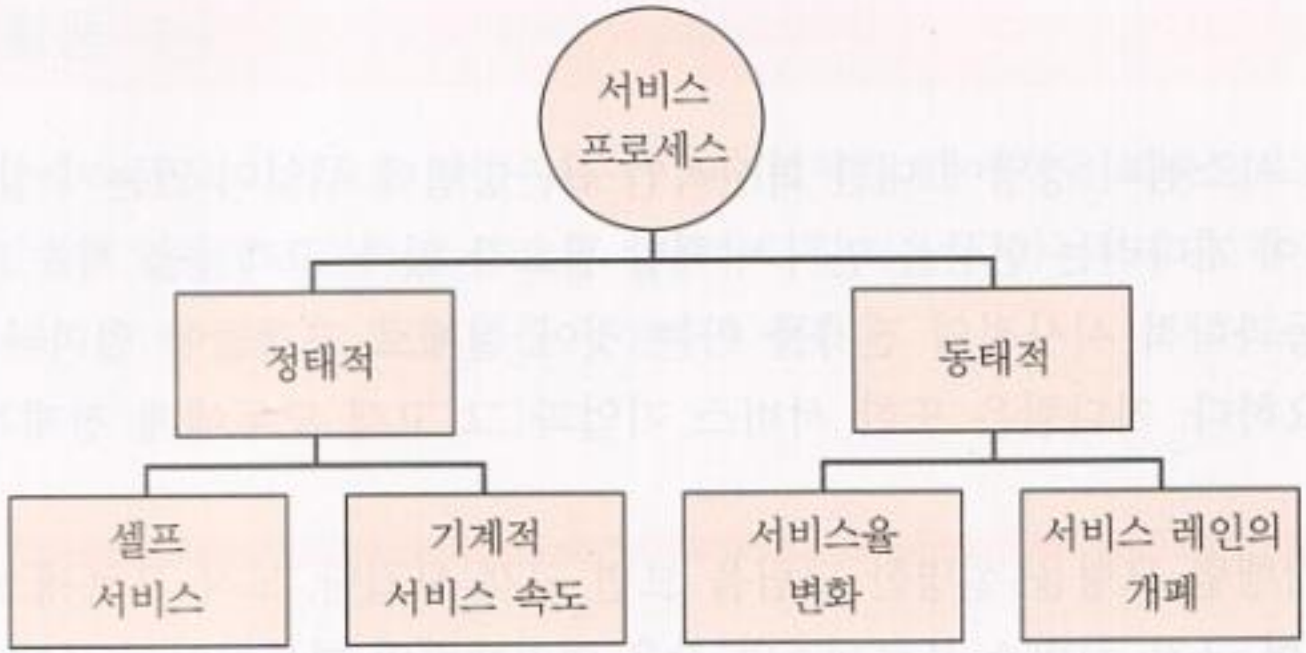


그림 12.13 서비스 프로세스의 분류

표 12.1 서비스 설비 배열

서비스 설비	서비스 배열
주차장	셀프 서비스
식당	직렬 서버
톨 게이트	병렬 서버
슈퍼마켓	첫 장소에 셀프서비스, 둘째 장소에 병렬 서버
병원	많은 서비스 센터가 직렬과 병렬이지만 각각의 환자가 모두 다 사용하는 것은 아님

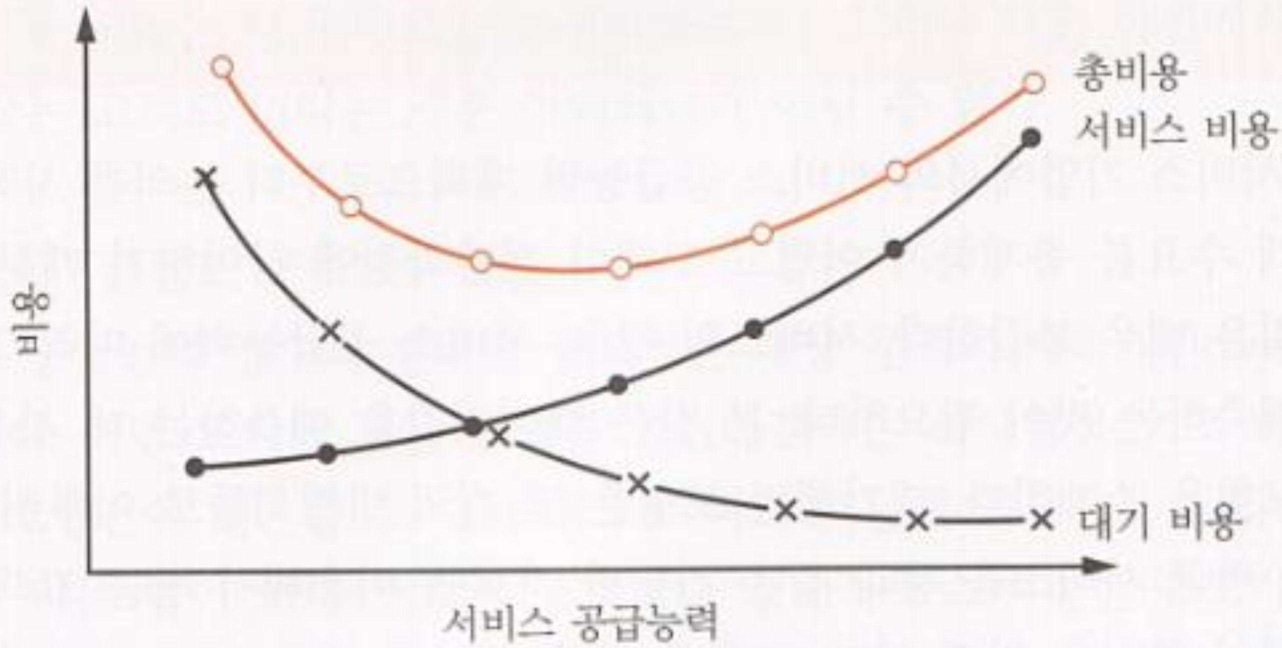


그림 13.1 공급능력 계획에 있어서 경제적 상충관계

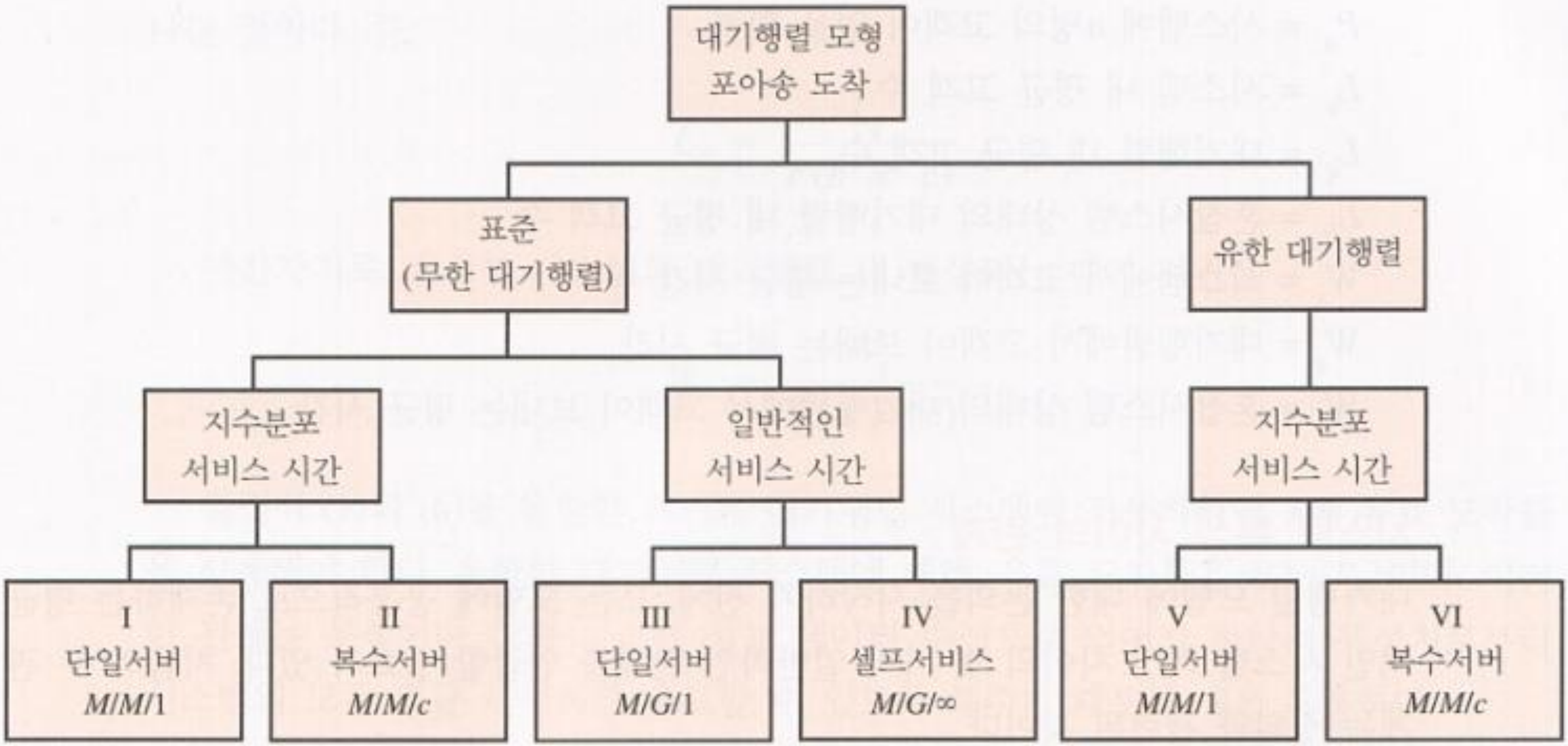


그림 13.2 대기행렬 모형의 분류

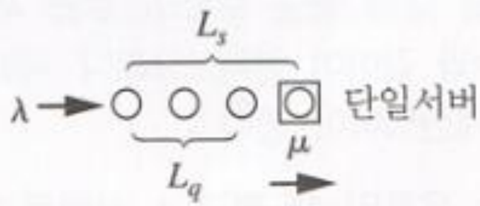


그림 13.3 M/M/1 대기모형의 구성

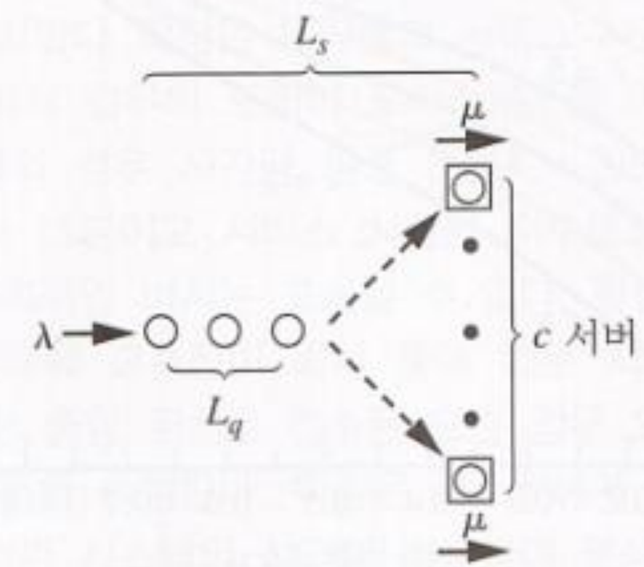


그림 13.4 *M*/*M*/*c* 대기행렬의 구성

Service Management

기호 정의

n = 시스템 내 고객의 수

λ = 평균 도착률(예: 시간당 도착하는 고객 수)

μ = 서버 당 평균 서비스율(예: 시간당 서비스를 제공받는 고객의 수)

ρ = 가동률(λ/μ)

N = 시스템이 허용하는 최대 고객 수

c = 서버의 수

P_n = 시스템에 n명의 고객이 있을 확률

L_s = 시스템 내 평균 고객 수

L_q = 혼잡시스템의 대기행렬 상의 평균 고객 수

L_b = 대기행렬 상의 평균 고객 수

W_s = 시스템에서 고객이 보내는 평균 시간

W_q = 대기행렬에서 고객이 보내는 평균 시간

W_b = 혼잡시스템의 대기행렬에서 고객이 보내는 평균 시간

I. 표준 $M/M/1$ 모형($0 < \rho < 1.0$)

$$P_0 = 1 - \rho \tag{I.1}$$

$$P(n \geq k) = \rho^k \tag{I.2}$$

$$P_n = P_0 \rho^n \tag{I.3}$$

$$L_s = \frac{\lambda}{\mu - \lambda} \tag{I.4}$$

$$L_q = \frac{\rho\lambda}{\mu - \lambda} \tag{I.5}$$

$$L_b = \frac{\lambda}{\mu - \lambda} \tag{I.6}$$

$$W_s = \frac{1}{\mu - \lambda} \tag{I.7}$$

$$W_q = \frac{\rho}{\mu - \lambda} \tag{I.8}$$

$$W_b = \frac{1}{\mu - \lambda} \tag{I.9}$$

II. 표준 $M/M/c$ 모형($0 < \rho < c$)

$$P_0 = \frac{1}{\left(\sum_{i=0}^{c-1} \frac{\rho^i}{i!}\right) + \frac{\rho^c}{c!(1-\rho/c)}} \tag{II.1}$$

$$P_n = \begin{cases} \frac{\rho^n}{n!} P_0 & \text{for } 0 \le n \le c \\ \frac{\rho^n}{c!c^{n-c}} P_0 & \text{for } n \ge c \end{cases} \tag{II.2}$$

$$P(n \ge c) = \frac{\rho^c \mu c}{c!(\mu c - \lambda)} P_0 \tag{II.3}$$

$$L_s = \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} P_0 + \rho \tag{II.4}$$

$$L_q = L_s - \rho \tag{II.5}$$

$$L_b = \frac{L_q}{P(n \ge c)} \tag{II.6}$$

$$W_s = \frac{L_q}{\lambda} + \frac{1}{\mu} \tag{II.7}$$

$$W_q = \frac{L_q}{\lambda} \tag{II.8}$$

$$W_b = \frac{W_q}{P(n \ge c)} \tag{II.9}$$

III. 표준 $M/G/1$ 모형($V(t)$ = 서비스 시간의 분산)

$$L_s = L_q + \rho \tag{III.1}$$

$$L_q = \frac{\rho^2 + \lambda^2 V(t)}{2(1-\rho)} \tag{III.2}$$

$$W_s = \frac{L_s}{\lambda} \tag{III.3}$$

$$W_b = \frac{L_q}{\lambda} \tag{III.4}$$

IV. 표준 $M/G/\infty$ 모형(e = 2.718, 자연로그의 밑수)

$$P_n = \frac{e^{-\rho}}{n!}\rho^n \quad \text{for } n \geq 0 \tag{IV.1}$$

$$L_s = \rho \tag{IV.2}$$

$$W_s = \frac{1}{\mu} \tag{IV.3}$$

V. 유한 대기행렬 $M/M/1$ 모형

$$P_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}} & \text{for } \lambda \neq \mu \\ \dfrac{1}{N+1} & \text{for } \lambda = \mu \end{cases} \tag{V.1}$$

$$P(n > 0) = 1 - P_0 \tag{V.2}$$

$$P_n = P_0\rho^n \quad \text{for } n \leq N \tag{V.3}$$

$$L_s = \begin{cases} \dfrac{\rho}{1-\rho} - \dfrac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} & \text{for } \lambda \neq \mu \\ \dfrac{N}{2} & \text{for } \lambda = \mu \end{cases} \tag{V.4}$$

$$L_q = L_s - (1 - P_0) \tag{V.5}$$

$$L_b = \frac{L_q}{1 - P_0} \tag{V.6}$$

$$W_s = \frac{L_q}{\lambda(1 - P_N)} + \frac{1}{\mu} \tag{V.7}$$

$$W_q = W_s - \frac{1}{\mu} \tag{V.8}$$

$$W_b = \frac{W_q}{1 - P_0} \tag{V.9}$$

VI. 유한 대기행렬 *M/M/c* 모형

$$P_0 = \frac{1}{\left(\sum_{i=0}^{c} \frac{\rho^i}{i!}\right) + \left(\frac{1}{c!}\right)\left(\sum_{i=c+1}^{N} \frac{\rho^i}{c^{i-c}}\right)} \qquad \text{(VI.1)}$$

$$P_n = \begin{cases} \frac{\rho^n}{n!} P_0 & \text{for } 0 \le n \le c \\ \frac{\rho^n}{c!c^{n-c}} P_0 & \text{for } c \le n \le N \end{cases} \qquad \text{(VI.2)}$$

$$P(n \ge c) = 1 - P_0 \sum_{i=0}^{c-1} \frac{\rho^i}{i!} \qquad \text{(VI.3)}$$

$$L_s = \frac{P_0 \rho^{c+1}}{(c-1)!(c-\rho)^2}\left[1 - \left(\frac{\rho}{c}\right)^{N-c} - (N-c)\left(\frac{\rho}{c}\right)^{N-c}\left(1 - \frac{\rho}{c}\right)\right] + \rho(1 - P_N) \qquad \text{(VI.4)}$$

$$L_q = L_s - \rho(1 - P_N) \qquad \text{(VI.5)}$$

$$L_b = \frac{L_q}{P(n \geq c)} \tag{VI.6}$$

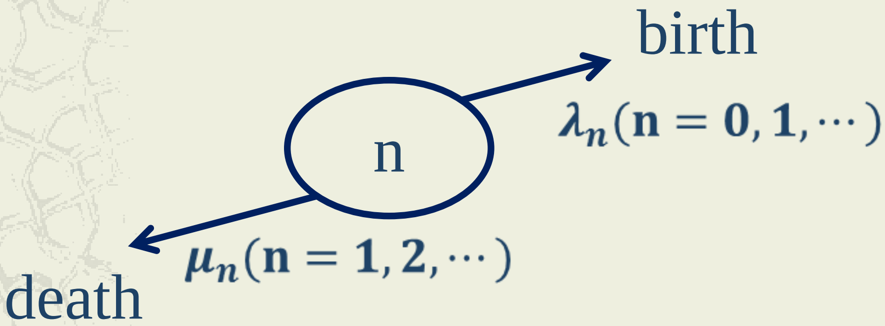
$$W_s = \frac{L_q}{\lambda(1 - P_N)} + \frac{1}{\mu} \tag{VI.7}$$

$$W_q = W_s - \frac{1}{\mu} \tag{VI.8}$$

$$W_b = \frac{W_q}{P(n \geq c)} \tag{VI.9}$$

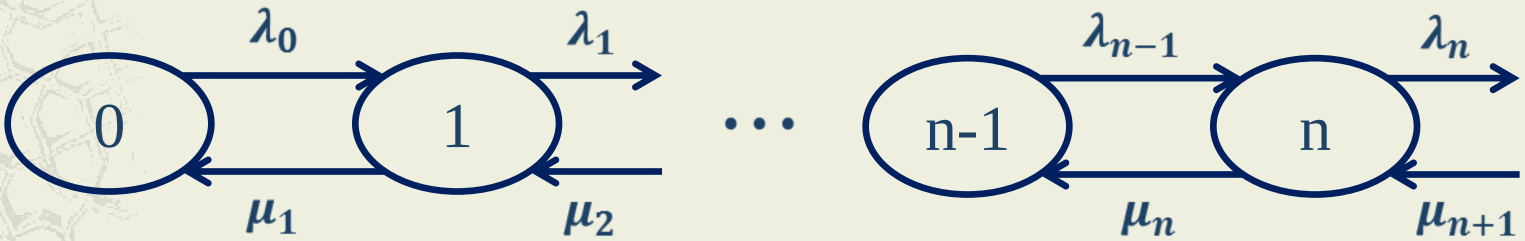
- Birth & Death Process
  - M/M/1
  - M/M/s

Birth & Death Process



Rate in = Rate out

(Balance equation)



$\mathbf{n} = \mathbf{0}$: $\mu_1 P_1 = \lambda_0 P_0$

$\mathbf{n} = \mathbf{1}$: $\lambda_0 P_0 + \mu_2 P_2 = (\mu_1 + \lambda_1) P_1$

… …

$\mathbf{n} = \mathbf{n}$: $\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = (\mu_n + \lambda_n) P_n$

$$\begin{cases} P_1 = \left(\frac{\lambda_0}{\mu_1}\right) P_0 \\ P_2 = \frac{1}{\mu_2}\left[(\mu_1 + \lambda_1)\left(\frac{\lambda_0}{\mu_1}\right) P_0 - \lambda_0 P_0\right] = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0 \\ \cdots \; \cdots \\ P_n = \frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1} P_0 = C_n P_0 \end{cases}$$

$$\sum_{n=0}^{\infty} P_n = 1 \qquad \therefore \; P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1}}$$

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

$$Q = \sum_{n=s}^{\infty} (n - s) P_n$$

# of servers

Queuing Model based on B & D process

M/M/1

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

$$P_n = \rho^n P_0 = \rho^n (1-\rho) \quad \left(\rho = \frac{\lambda}{\mu} < 1\right)$$

$$P_n = C_n P_0$$

traffic intensity

$$L = \sum_{n=1}^{\infty} n P_n = \frac{\rho}{1-\rho}$$

$$Q = \sum_{n=1}^{\infty} (n-1) P_n$$

$$= \sum_{n=1}^{\infty} n P_n - \sum_{n=1}^{\infty} P_n$$

$(1-\rho)$

$$= L - (1 - P_0) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$$

Little's Law

L=Avg. # of customer in system

W=Avg. waiting time in system

Q=Avg. # of customer in queue

d=Avg. delay in queue

$$\boldsymbol{L = \lambda w} \qquad \boldsymbol{w = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu - \lambda}}$$

$$\boldsymbol{Q = \lambda d} \qquad \boldsymbol{d = \frac{L-\rho}{\lambda} = \frac{1}{\mu - \lambda} - \frac{1}{\mu}}$$

$$\boldsymbol{= \frac{\rho^2}{\lambda(1-\rho)} = \frac{\lambda}{\mu^2(1-\rho)}}$$

M/M/s

$$\mu_n = \begin{cases} n\mu & n < s \\ s\mu & n \geq s \end{cases}$$

$$\begin{cases} C_n = \dfrac{\lambda \quad \lambda \quad \cdots \quad \lambda}{n\mu \ (n-1)\mu \ \cdots \ \mu} = \dfrac{1}{n!}\left(\dfrac{\lambda}{\mu}\right)^n & n < s \\ C_n = \dfrac{\lambda \ \cdots \ \lambda \quad \lambda \quad \lambda \quad \cdots \quad \lambda}{s\mu \ \cdots s\mu \ s\mu \ (s-1)\mu \ \cdots \ \mu} = \dfrac{\lambda^n}{(s\mu)^{n-s} s! \, \mu^s} \end{cases}$$

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

M/M/s

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \sum_{n=s}^{\infty}\left(\frac{\lambda}{s\mu}\right)^{n-s} \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s}$$

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

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

$$= \left[\sum_{n=0}^{s-1} \frac{1}{n!}(s\rho)^n + \frac{(s\rho)^s}{s!\,(1-\rho)}\right]^{-1} \qquad \rho = \frac{\lambda}{s\mu} < 1$$

$$\therefore P_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = \frac{1}{n!}(s\rho)^n P_0 & n < s \\ \frac{1}{s^{n-s}s!}\left(\frac{\lambda}{\mu}\right)^n P_0 = \frac{s^s}{s!}\left(\frac{\lambda}{s\mu}\right)^n P_0 & n \geq s \end{cases}$$