

예약기반 서비스 시스템에서의 고객 도착 프로세스 분석

민 대기*†

Analysis of Customer Arrival Process in an Appointment-Based Service System

Daiki Min*

■ Abstract ■

Delivery of a good quality of service in an efficient manner requires matching the supply of capacity with customer demand. Much research has employed queueing models that analyzed the service system on the basis of independent and stationary customer arrivals. However, the appointment system, which is widely used to facilitate customer access to service in many industries including healthcare, has significant influence on the customer arrival process so that the independent and stationary assumption does not hold in an appointment-based service system. In this regard, this paper aims to propose a model for accurate illustration of the appointment-based customer arrival process. The use of the proposed model allows us to evaluate the overall system performance such as mean waiting time and service level under various appointment policies instead of conducting simulation studies.

Keyword : Appointment-Driven Service System, Customer Arrivals, Appointment Policy, Analytic Model

1. 서 론

콜센터, 은행, 병원과 같은 서비스 시스템에서는 방

문 고객에게 적절한 시간 내에 서비스를 제공하는 것이 중요하다. 적정 수준의 고객 대기시간(waiting time)을 만족시키는 서비스 시스템을 설계하기 위해서는 고객 수요와 서비스 공급의 균형을 이루는 것이 중

논문접수일 : 2012년 02월 06일 논문게재확정일 : 2012년 04월 09일

논문수정일(1차 : 2012년 03월 08일, 2차 : 2012년 03월 27일)

* 이화여자대학교 경영학과

† 교신저자

요하다. 대기행렬 모형(queueing models)은 고객 수요를 기반으로 서비스 공급을 결정하기 위한 문제에 광범위하게 사용되고 있다. 일반적으로 대기행렬 이론은 정상상태(stationary state)인 고객 도착 프로세스(customer arrival process)를 가정하여, 고객 대기시간과 자원의 효율성(utilization) 같은 서비스 성능을 계산하고 이를 이용하여 서비스 공급 수준을 결정한다. 따라서 서비스 시스템의 고객 도착 프로세스가 정상상태를 만족하지 않는 경우 일반적인 대기행렬 모형을 활용하는 것은 적절하지 않다.

현실적으로 많은 서비스 시스템에서 정상상태 조건을 만족하지 않으며 시간에 따라 변동하는 고객 도착 프로세스를 확인할 수 있는데, 이와 같이 시간에 따라 변동하는 고객 도착 프로세스를 갖는 서비스 시스템은 기간별 정상상태를 가정하여 분석한다. 예를 들어 Green et al.[8, 9]은 콜센터에서의 인력 수준 결정 문제에 대하여 독립기간별 정상상태(Stationary Independent Period by Period; SIPP) 방법론을 제안하였다. SIPP 방법론은 임의의 짧은 기간에 대하여 평균 고객 도착률(arrival rate)을 계산하고, 각 구간별로 평균 고객 도착률을 갖는 M/M/s 대기행렬 모형을 적용한다. 여기서 M/M/s 대기행렬 모형은 지수분포의 서비스 시간을 갖는 다수의 서버가 포아송 프로세스를 따라 도착하는 고객에게 서비스를 제공하는 시스템을 의미한다. 유사하게 Ingolfsson et al.[17]는 구간별 일정 함수(piecewise constant function)을 이용하여 고객 도착률을 고려하였다. 시간에 따라 고객 도착 프로세스가 변동하는 문제에 대하여 기간별 정상상태를 고려한 대기행렬 모형 연구는 Defraeye and Nieuwenhuysse[6]의 문헌연구를 참고하도록 한다.

개별 연구 모형별로 어느 정도 차이는 존재하지만 대부분의 대기행렬 모형은 공통적으로 고객 도착률을 제어할 수 없는 독립적(independent) 요소로 고려하고 있다. 하지만 예약 기반의 서비스 시스템은 독립적 고객 도착 프로세스와 같이 일반적인 대기행렬 모형이 가정하는 특성과 다른 특성을 갖고 있다. 예약 기반의 서비스 시스템은 예약 정책(appointment

policy)을 이용하여 고객의 도착 프로세스를 조정함으로써 고객 대기시간과 자원의 효율성을 향상시킨다. 따라서 예약기반 서비스 시스템의 전반적인 성과는 예약 정책에 따라 결정되며, 고객 도착 프로세스 또한 독립적인 정상상태가 아닌 고객 예약에 종속적으로 결정된다. 이와 관련하여 효과적인 예약 정책 설계를 위한 많은 연구가 진행되고 있으며 상세한 연구 내용은 다음의 연구를 참고하도록 한다; Ho and Lau[13], Cayirli and Veral[4], Gupta and Denton[11].

예약이 있는 대기행렬 모형은 Winsten[29]에 처음 제시되었으며, Doi et al.[7], 김영한[1], 신양우[2] 등은 확장한 연구를 제시하였다. 하지만 이들의 연구에서는 예정된 시간 구간 내에 도착하는 고객만을 대상으로 하고 있는데 이는 지연 도착이나 조기 도착이 가능한 일반적인 서비스 시스템과는 차이가 존재함을 알 수 있다. 본 연구는 예약시간을 갖는 한정된 고객의 수가 지연 또는 조기 도착이 가능한 시스템을 고려하였다.

본 연구에서 고려하는 예약기반 서비스 시스템에 대한 연구는 최근 의료 영역에서 확인할 수 있다[4, 11, 13, 21]. 의료 영역에서 환자 예약과 관련한 많은 연구는 환자 대기시간을 줄이고 의사의 효율성을 향상시키기 위한 예약 정책의 평가 및 설계를 목적으로 한다. 의료 서비스 시스템의 경우 한정된 수의 환자가 예약 시간에 따라 시스템에 도착하므로 독립적인 정상상태를 만족하지 않는다. 따라서 예약정책의 평가와 분석을 위하여 전통적인 대기행렬 모형보다는 대체적으로 시뮬레이션을 연구 방법론으로 활용하고 있다.

한정된 수의 고객을 대상으로 하는 예약기반의 서비스 시스템에서 시뮬레이션 방법론 대신에 분석 모형을 활용한 연구는 그 수가 많지 않다. 전통적인 대기행렬 모형을 수정하여 외래환자의 예약 스케줄을 결정하거나[12] 병원의 용량계획을 수립하기 위한 연구[5] 등이 제시되었으나 모든 환자는 예약된 순서에 따라 예약 시간에 정시 도착하는 것을 가정하고 있다. Jouini and Benjaafar[18]는 본 논문과

가장 유사한 문제를 대상으로 하고 있는데 예약 기반 서비스 시스템에 대하여 대기행렬 이론을 수정한 모형을 제시하였다. Jouini and Benjaafar[18]의 경우 고객의 비정시도착(non-punctuality)과 노쇼(no-show)를 함께 고려하였지만 모형의 단순화와 계산의 편의성을 고려하여 모든 고객이 예약 순서에 따라 도착함을 가정하였다. 하지만 예약 순서에 따라 모든 고객이 정시 도착하는 것은 비현실적인 가정으로 기존 연구 모형의 한계이라고 할 수 있다.

본 논문은 예약시간을 갖는 고객이 도착하는 서비스 시스템의 분석을 위해서 독립적이고 정상상태를 따르는 고객 도착 프로세스를 기반으로 하는 전통적인 대기행렬 모형과 다른 모형을 제시하는 것을 목적으로 한다. 특히 예약에 따른 순차 도착과 정시 도착 가정을 완화한 보다 일반적인 모형을 제시하도록 한다. 많은 연구가 시뮬레이션 기법을 활용하여 예약 정책을 평가하고 있는데, 본 연구에서 제안한 모형이 다양한 유형의 예약 정책에 적용 가능성을 제시함으로써 효과적으로 예약기반 서비스 시스템의 분석 방안으로 활용될 수 있음을 제시하도록 한다.

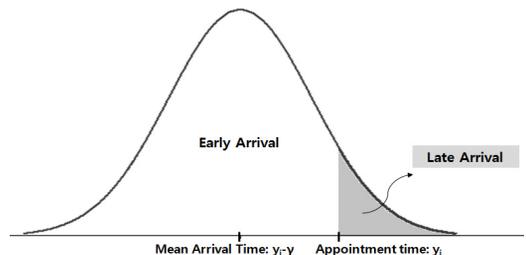
2. 문제의 정의 및 분석

본 논문은 예약 시간을 갖는 다수의 고객이 도착하는 단일 서버로 구성된 서비스 시스템을 고려한다. 전통적인 대기행렬 모형과 비교하여 본 연구에서 고려하는 예약기반 서비스 시스템의 특성을 다음과 같이 요약할 수 있다. 첫째, 예약기반의 서비스 시스템의 경우에도 예약 없이 방문하는 고객을 고려할 수 있으나 본 연구에서는 예약이 안 된 방문 고객(walk-in customer)은 고려하지 않는다. 따라서 하루 동안 도착하는 고객의 수는 한정적이며, 고객의 수가 한정적이므로 정상상태를 가정할 수 없다. 둘째, 실제 고객의 도착 시간은 사전에 정의된 예약 시간에 종속적으로 결정된다. 예약에 종속적인 특성을 활용하여 고객 도착 프로세스에 대한 상세한

이해가 가능하며, 예약 정책을 이용하여 일정 수준에서 고객 도착 프로세스를 통제할 수 있다. 셋째, 일반적으로 고객은 예약 시간에 정시에 도착하는 것이 아니라 예약 시간보다 일찍 또는 늦게 도착한다. 즉, 고객의 정시 도착이 확정적이지 않으며 도착 시간은 확률적으로 결정된다.

2.1 고객 도착 프로세스

이번 장에서는 예약을 갖는 고객의 도착 프로세스를 표현하기 위한 모형을 제시하도록 한다. 모형의 편의를 위하여 단일 서버로 구성된 서비스 시스템을 대상으로 n 명의 예약 고객을 가정하였다. i 번째 예약고객의 예약시간을 y_i 라고 하며, 예약시간은 $y_1 \leq y_2 \leq \dots \leq y_n$ 의 조건을 만족한다. 여기서 y_i 사이의 등호는 여러 명의 고객이 동일 시간대에 예약이 가능한 블록 예약정책(block appointment policy)이 가능함을 나타낸다. 예약시간 y_i 를 갖는 고객의 실제 도착시간은 Y_i 로 표현한다. 앞서 제시한 바와 같이 실제 고객의 도착시간은 예약 시간보다 늦거나 빠를 수가 있는데, 이와 같은 비정시도착 특성은 정규분포를 이용하여 표현할 수 있다[21, 22, 23]. 본 논문에서 도착 시간 Y_i 는 $y_i - \gamma$ 를 평균으로 하고 σ 를 표준편차로 하는 정규분포를 따르는데, γ 는 일반적으로 고객이 예약시간보다 일찍 도착하는 특성을 표현한다. 예를 들어, 환자의 경우 진료예약 시간보다 30분정도 일찍 접수하도록 요청 받는다면, $\gamma = 30$ 분이 된다.



[Figure 1] Distribution of the Customer Arrival Time

[Figure 1]은 고객의 도착시간 Y_i 의 분포를 나타내는데, 정규분포를 가정하는 경우 대부분의 고객은 예약시간보다 일찍 도착하며 일부만이 예약시간보다 늦게 도착함을 알 수 있다. 또한 정규분포를 이용함으로써 예약에 따른 순차도착 가정을 완화하게 되는데, 예를 들어, $y_i \leq y_{i+1}$ 인 경우에 일정 확률로 $Y_i \geq Y_{i+1}$ 이 만족함을 알 수 있다. 하지만 첫 번째 예약고객이 마지막 n 번째 예약고객 보다 늦게 도착할 확률 $P(Y_1 \geq Y_n)$ 이 매우 낮은 것 또한 합리적인 가정이다. 따라서 일반적인 정규분포 대신에 절단정규분포(truncated normal distribution)를 이용하여 도착시간을 표현하였으며, 본 연구에서는 임의적으로 평균에서 2σ 범위를 벗어나는 도착시간은 고려하지 않도록 한다.

시간에 따라 고객 도착 프로세스가 변동하는 서비스 시스템을 분석하기 위한 기존 문헌과 동일하게[10, 17, 28] 본 연구에서도 연속시간을 임의의 짧은 기간 τ 로 구분하고 각 기간에 대하여 고객의 도착 상태를 분석하도록 한다. 우선 기간 $(t, t+\tau)$ 동안 고객 i 가 도착할 확률을 $P_{i,t}$ 라 하면, 절단정규분포 함수 $\Phi^T(\cdot)$ 를 이용하여 다음과 같이 계산된다.

$$\begin{aligned} P_{i,t} &= P(Y_i < t+\tau) - P(Y_i < t) \\ &= \Phi^T\left(\frac{t+\tau-y_i+\gamma}{\sigma}\right) - \Phi^T\left(\frac{t-y_i+\gamma}{\sigma}\right), \\ \text{where } \Phi^T(x) &= \frac{\Phi((x-y_i+\gamma)/\sigma) - \Phi(-2)}{\Phi(2) - \Phi(-2)}. \end{aligned}$$

다음으로 $P_{i,t}$ 를 이용하여 기간 $(t, t+\tau)$ 동안의 고객 도착률을 계산한다. 기간 $(t, t+\tau)$ 동안 새롭게 도착하는 고객의 수 M_t 는 $P_{i,t}$ 에 의하여 다음과 같이 결정된다.

$$\begin{aligned} P(M_t = k) &= \sum_{N(k) \subseteq N_i \in N(k)} \prod_{i \in N_i} P_{i,t} \prod_{j \in N \setminus N(k)} (1 - P_{j,t}), \quad (1) \\ k &= 0, 1, \dots, n. \end{aligned}$$

N 은 n 명으로 구성된 고객의 집합이고, $N(k)$ 는 k 명으로 구성된 고객의 부분집합을 의미한다. 따라서 $\prod_{i \in N(k)} P_{i,t}$ 는 전체 고객 중에서 k 명의 고객이 $(t, t+\tau)$ 기간 동안 도착할 확률을 의미하며, $\prod_{j \in N \setminus N(k)} (1 - P_{j,t})$ 은 나머지 $n-k$ 명의 고객이 도착하지 않을 확률을 의미한다.

식 (1)은 기간 $(t, t+\tau)$ 에서의 고객 도착률을 나타낸다. 여기서 주목할 점은 전통적인 대기행렬 모형에서는 고객 도착률이 단일 값으로 표현되는 결정적 상수임에 비하여, 식 (1)에서 도출한 고객 도착률 M_t 는 예약 시간과 고객의 정시도착 확률에 의하여 결정되는 확률 변수로 표현된다는 것이다. 단일 상수를 이용한 일반적인 대기행렬 모형의 고객 도착률 근사 기법에 비하여, 확률 분포로 제시된 고객의 도착 프로세스를 이용함으로써 서비스 시스템을 보다 정확하게 분석할 수 있다.

2.2 고객 대기시간의 도출

앞서 제시한 고객 도착 프로세스를 활용하여 서비스 시스템에 존재하는 대기 고객의 수와 대기 시간에 대한 확률분포를 도출하도록 한다. $(t, t+\tau)$ 의 기간 초에 서비스를 기다리는 고객의 수를 Q_t 로 정의하고, 동일 기간 동안 서비스를 받고 시스템을 떠난 고객의 수를 S_t 로 정의한다. $t=0$, 즉 서비스 업무의 개시 시점에는 한명의 고객도 존재하지 않으므로 $Q_0=0$ 과 $S_t=0$ 을 각각 만족한다. 앞서 제시한 고객 도착률 M_t 를 활용하면 다음의 관계가 성립한다.

$$\begin{aligned} Q_{t+\tau} &= Q_t + M_t - S_t, \quad (2) \\ t &= 0, \tau, 2\tau, \dots, T-\tau \end{aligned}$$

여기서 T 는 마지막 고객이 서비스를 받고 서비스 시스템을 떠나는 시점을 의미한다.

고객의 서비스 프로세스는 포아송 분포(Poisson

distribution)를 가정하여 S_t 는 서비스율(service rate)이 $1/\mu$, 즉 고객 한명 당 평균 서비스 시간이 μ 인 포아송 분포로 정의하였다. 일반적으로 고객의 예약 시간은 평균 서비스 시간을 고려하여 결정되므로 평균 서비스 시간 μ 는 예약 시간 간격과 유사한 값으로 고려할 수 있다. 또한 서비스가 가능한 고객의 수는 서비스 대기 고객의 수를 넘을 수 없으므로 $S_t \leq Q_t + M_t$ 의 조건을 만족한다.

$$P(S_t = k) = \begin{cases} \frac{e^{1/\mu}(1/\mu)^k}{k!} & \text{if } k = 1, 2, \dots, Q_t + M_t, \\ 1 - \sum_{i=1}^{Q_t + M_t} \frac{e^{1/\mu}(1/\mu)^i}{i!} & \text{if } k = 0. \end{cases}$$

식 (2)를 이용하여 대기 고객 수 Q_t 에 대한 확률 분포를 계산하였다. 기간 초에 x 명의 서비스 대기 고객이 존재할 확률은 다음과 같이 결정된다.

$$\begin{aligned} P(Q_t = x | Q_{t-\tau} = q) & \quad (3) \\ &= \sum_{m|x-q \leq m \leq n-q} P(M_t = m) \\ & \quad \times P(S_t = q + m - x). \end{aligned}$$

$$\begin{aligned} P(Q_t = x) &= \sum_{q=0}^n P(Q_t = x | Q_{t-\tau} = q) \quad (4) \\ & \quad \times P(Q_{t-\tau} = q). \end{aligned}$$

식 (3)은 기간 $(t-\tau, t)$ 에 q 명의 고객이 존재하는 경우 $(t, t+\tau)$ 기간에 x 명의 고객이 존재할 조건부 확률(conditional probability)을 나타내며, 이는 고객 도착 확률과 서비스 확률에 의하여 계산된다. 식 (3)의 조건부 확률을 $Q_{t-\tau}$ 의 주변 확률(marginal probability)과 함께 고려하여 $(t, t+\tau)$ 기간에서의 $P(Q_t = x)$ 를 계산한다. 즉, 기간 0부터 Q_t 의 주변 확률과 조건부 확률을 재귀적으로(recursively) 계산하여 전 기간에서의 확률 값을 계산한다. 예를 들어, $t = \tau$ 시점에서 식 (3)과 식 (4)를 이용하여 Q_t 의 확률 분포를 계산하면 다음과 같다. 서비스 개시 시점에는 고객이 존재하지 않으므로 0보다 큰 모든 q

에 대하여 $P(Q_0 = q) = 0$ 를 만족하므로

$$\begin{aligned} P(Q_\tau = x) &= \sum_{q=0}^n P(Q_\tau = x | Q_0 = q) \cdot P(Q_0 = q) \\ &= P(Q_\tau = x | Q_0 = 0) \text{ 와 같다. 따라서} \end{aligned}$$

$$P(Q_\tau = 0) = \sum_{m=0}^n P(M_\tau = m)P(S_\tau = m),$$

$$P(Q_\tau = 1) = \sum_{m=1}^n P(M_\tau = m)P(S_\tau = m - 1),$$

$$P(Q_\tau = 2) = \sum_{m=2}^n P(M_\tau = m)P(S_\tau = m - 2),$$

...

$$P(Q_\tau = n) = P(M_\tau = n)P(S_\tau = 0).$$

식 (4)를 이용하여 계산한 Q_t 의 확률 분포를 이용하여 기간 $(t, t+\tau)$ 동안 대기 고객 수의 기대값(expected queue length)을 계산한다.

$$\widehat{Q}_t = E[Q_t] = \sum_{x=0}^n x \cdot P(Q_t = x). \quad (5)$$

\widehat{Q}_t 는 시점 t 에서 어느 고객이 도착하는 경우 예상할 수 있는 대기 고객 수의 평균 기대값을 의미한다. 또한 앞서 서비스 프로세스를 $1/\mu$ 의 서비스율을 갖는 포아송 분포로 가정하였으므로, 시점 t 에서 어느 고객의 예상 대기시간은 \widehat{Q}_t 와 $1/\mu$ 를 파라미터로 하는 Erlang 분포를 따르게 된다. 따라서 시점 t 에서 예상 고객 대기시간의 평균과 분산은 각각 $\mu\widehat{Q}_t$ 와 $\mu^2\widehat{Q}_t$ 가 된다.

3. 수치실험을 통한 모형 검증

이번 장에서는 제 2장에서 제시한 모형의 검증을 위한 수치실험 결과를 제시하였다. 모형의 검증을 위하여 다음에 제시되는 문제를 대상으로 본 논문에서 제안한 모형과 시뮬레이션 실험 결과를 비교하였다. 시뮬레이션 실험은 이산사건 시뮬레이션

(discrete event simulation) 소프트웨어인 AutoMod 12.3.1을 이용하여 동일 조건에 대하여 50회 수행하여 자료를 수집하였다. 일반적으로 보다 많은 횟수의 시뮬레이션 실험을 수행하는 경우 대수의 법칙(strong law of large numbers)에 따라 결과의 변동을 줄일 수 있으나 본 논문의 경우 유의미한 결과를 얻었으므로 50회로 실험횟수를 제한하였다.

시뮬레이션 결과와 제안 모형을 각각 실측값과 추정값으로 고려하는 경우 시계열 예측모형의 오차 평가방법을 적용하여 제안 모형의 적합성을 평가할 수 있을 것이다. 일반적으로 시계열 예측모형의 정확도를 평가하는 방법으로 평균절대편차(MAE : Mean Absolute Error), 평균제곱오차(MSE : Mean Squared Error), 평균절대오차백분율(MAPE : Mean Absolute Percentage Error), 평균백분율오차(MPE : Mean Percentage Error)등을 사용한다. 본 연구에서 평가하고자 하는 데이터(도착률, 고객 수 등)의 경우 0의 값이 존재하며, 동일 수준(scale)의 데이터를 분석하므로 Hyndman and Koehler[16]의 연구결과에 의하여 평균절대편차를 활용하도록 한다. \bar{Q}_t 는 시뮬레이션 실험을 통하여 계산한 시간별 대기 고객 수의 평균값을 의미하며, 이를 이용하여 평균절대편차는 다음 식 (6)에 의하여 같이 계산된다.

$$\text{평균절대편차}(MAE) = \frac{1}{T} \sum_{t=1}^T |\hat{Q}_t - \bar{Q}_t|. \quad (6)$$

20명의 예약 고객이 단일 서버에 도착하는 서비스 시스템에서 고객 한명 당 평균 서비스 시간은 15분(즉, $\mu = 15$ 분/고객)으로 결정하였다. $\tau = 5$ 분으로 설정하였으며, 따라서 본 연구에서 제안한 모형의 경우 5분 단위로 고객 도착율과 대기 고객의 수를 분석하게 된다. 시뮬레이션 실험의 경우에도 5분 단위로 동일한 내용을 측정하고 제안 모형의 결과와 비교 분석하여 적합성을 평가하였다. 고객 도착시간 Y_i 에 대한 정규분포의 평균과 표준편차는 각각($y_i - 20$)분과(즉, $\gamma = 20$) 15분($\sigma = 15$)으로 설정하였다.

고객의 예약 시간은 예약 정책에 의하여 결정되는데, 다양한 환경에서 모형을 검증하기 위하여 본 논문에서는 총 12개의 예약 정책에 대하여 실험을 수행하였다. 예약 정책은 블록 크기(block size), 예약 주기(appointment interval) 그리고 Bailey's rule로 알려진 초기중복예약[3] 등의 조합으로 결정된다[13]. 최근 예약주기를 동적으로 결정하는 연구가 있지만[14, 25, 27], 본 논문에서는 실험의 용이성을 위하여 정적으로 고정된 예약주기만을 고려하였다. 블록 크기는 동일한 예약 시간에 할당되는 고객의 수를 의미한다. 예를 들어 블록 크기가 1이고 예약 주기가 10분인 경우 매 10분 마다 1명의 고객을 예약하지만, 블록 크기가 2인 경우 20분 마다 2명을 예약하게 된다. 초기중복예약에서는 예약 주기에 따라 고객 예약 시간을 결정하되 최초 예약 주기에 두 명이상의 고객을 할당한다. 예를 들어 예약 주기가 10분이고 초기중복예약 고객의 수(본 논문에서는 Bailey's Rule로 표현하도록 한다)가 3인 경우 최초 10분에 3명의 고객 예약시간을 결정하되, 나머지 고객은 10분 단위로 예약시간을 결정한다. Bailey's rule에 의하여 초기에 여러 명의 고객을 동시에 예약함으로써 일부 고객이 나타나지 않거나 지연 도착함에 따라 예상되는 서버의 효율 저하를 방지할 수 있다. 하지만 다수의 고객이 동일한 예약 시간을 갖게 되므로 고객의 대기시간이 증가할 수 있다. 본 연구에서 고려한 예약정책을 <Table 1>에 정리하였다.

3.1 고객 도착률

기간($t, t + \tau$) 동안의 평균 고객 도착률 $\hat{\lambda}_t$ 는 수식 (1)을 이용하여 다음과 같이 계산할 수 있다 :

$$\hat{\lambda}_t = \sum_{k=0}^n P(M_t = k) \cdot k. \text{ 고객 도착률이 시간에 따라}$$

변하는 경우 전통적인 대기행렬 모형에서는 기간별 평균 도착률 $\hat{\lambda}_t$ 를 이용한 정상상태 모형을 활용하였다. 하지만 앞서 제시한 바와 같이 본 논문에서 도출한 고객 도착률 M_t 는 확률 변수가 된다.

<Table 1> Appointment Policy

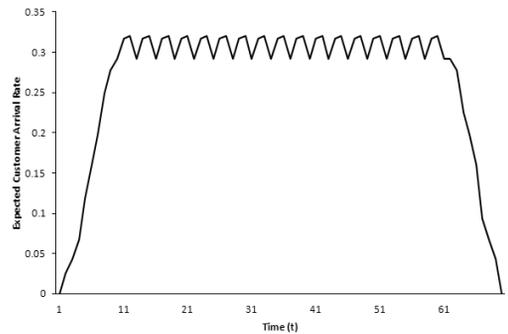
Appt. Policy	Block Size	Appt. Interval(Min.)	Bailey's Rule	Description
A1	1	10	1	$y_{i+1} - y_i = 10$ for all i
A2	1	10	2	$y_1 = y_2, y_{i+1} - y_i = 10$ for $i = 2, 3, \dots$
A3	1	10	3	$y_1 = y_2 = y_3, y_{i+1} - y_i = 10$ for $i = 3, 4, \dots$
A4	1	15	1	$y_{i+1} - y_i = 15$ for all i
A5	1	15	2	$y_1 = y_2, y_{i+1} - y_i = 15$ for $i = 2, 3, \dots$
A6	1	15	3	$y_1 = y_2 = y_3, y_{i+1} - y_i = 15$ for $i = 3, 4, \dots$
A7	1	20	1	$y_{i+1} - y_i = 20$ for all i
A8	1	20	2	$y_1 = y_2, y_{i+1} - y_i = 20$ for $i = 2, 3, \dots$
A9	1	20	3	$y_1 = y_2 = y_3, y_{i+1} - y_i = 20$ for $i = 3, 4, \dots$
A10	2	20	2	$y_{2k-1} = y_{2k}$ and $y_{2k} - y_{2k-2} = 20$ for $k = 1, 2, \dots, \lceil n/2 \rceil$
A11	2	30	2	$y_{2k-1} = y_{2k}$ and $y_{2k} - y_{2k-2} = 30$ for $k = 1, 2, \dots, \lceil n/2 \rceil$
A12	2	40	2	$y_{2k-1} = y_{2k}$ and $y_{2k} - y_{2k-2} = 40$ for $k = 1, 2, \dots, \lceil n/2 \rceil$

[Figure 2]는 예약정책 A4를 활용한 경우에 계산한 고객 도착률의 일부를 제시한다. 결과를 보면 초기에는 고객이 도착하지 않을 확률이 100%이며, 시간에 따라 점차적으로 감소하고 있다. 반대로 1명 이상의 고객이 도착할 확률은 초기 0%에서 점진적으로 증가하여 일정 수준에서 변동하게 된다. 예약

에 기반 한 고객 도착 프로세스의 이해를 돕기 위하여 고객 도착률 기대값의 변화를 [Figure 2]에 함께 제시하였다. 고객 도착률이 시점 11-61 구간에서 동일한 형태로 증가와 감소를 반복하는 것을 확인할 수 있는데, 이는 실험에서 사용한 예약정책(A4)이 일정한 간격으로 고객의 예약시간을 결정하고

t \ M _t	0	1	2	3	4	$\hat{\lambda}_t$
0	1	0	0	0	0	0
1	0.97496	0.02504	0	0	0	0.025
2	0.956579	0.043421	0	0	0	0.044
...	...					
6	0.810803	0.180392	0.008805	0	0	0.198
7	0.768127	0.214616	0.01695	0.000307	0	0.2494
...	...					
10	0.716322	0.251947	0.030281	0.001429	2.07E-05	0.3169
...	...					

(a) Customer Arrival Rate M_t



(b) Expected Customer Arrival Rate $\hat{\lambda}_t$

[Figure 2] Customer Arrival Rate under Appointment Policy A4

있고, 고객의 도착시간 Y_i 가 절단정규분포를 따르는 것에 기인한다.

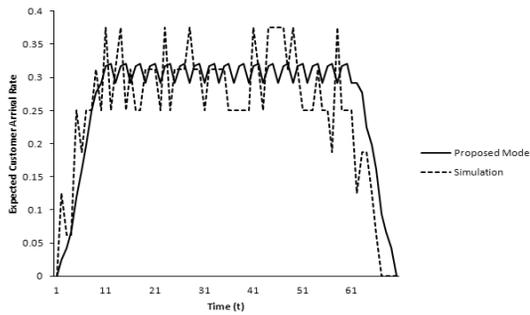
본 논문에서 제안한 고객 도착율의 정합성 검증을 위하여 <Table 1>에서 제시한 12개의 예약 정책에 대하여 시뮬레이션 실험을 수행하고 그 결과를 비교하였다. $\hat{\lambda}_t$ 은 시뮬레이션 결과로 계산한 평균 고객 도착률 $\bar{\lambda}_t$ 과 유사한 결과를 보이고 있는데, 20회의 적은 시뮬레이션 결과에도 예약정책에 따라 평균절대편차가 0.04~0.05로 매우 작은 오차를 갖는다. 따라서 도착율의 정합성 검증을 위한 시뮬레이션 횟수는 20회로 제한하였다. 임의로 선정된 두 개의 예약정책 A4와 A6의 결과를 [Figure 3]에 제시하였으며, 결과를 보면 시뮬레이션 실험에 의하

여 계산한 $\bar{\lambda}_t$ 이 어느 정도 변동하고 있으나 본 연구에서 제시한 모형에 매우 근사한 형태를 갖는 것을 확인할 수 있었다. 시뮬레이션 실험의 횟수를 증가시킨다면 보다 유사한 결과를 예상할 수 있다.

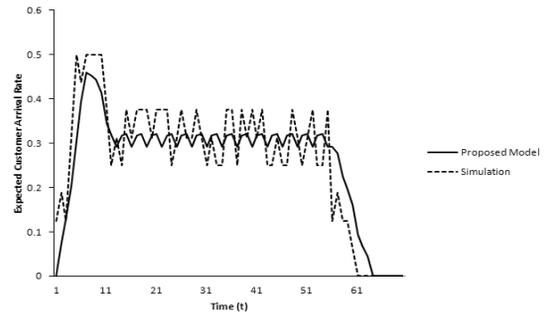
3.2 대기 고객 수

식 (5)에서 제시한 대기 고객 수 \hat{Q}_t 의 검증을 위한 시뮬레이션 실험을 수행하였다. [Figure 4]는 시간대별로 \hat{Q}_t 과 \bar{Q}_t 의 변화를 나타낸 것으로 본 연구에서 제안한 모형이 실제 시뮬레이션 결과와 상당히 일치하는 것을 확인할 수 있다.

보다 상세한 분석을 위하여 <Table 1>에서 제시

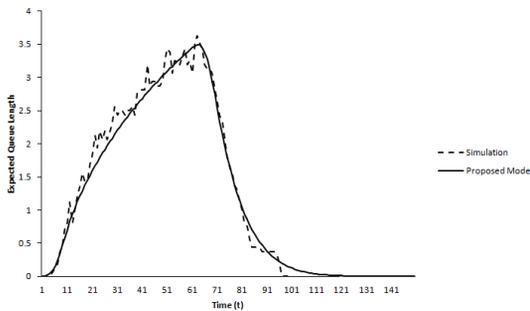


(a) Customer Arrival Rates under Appointment Policy A4

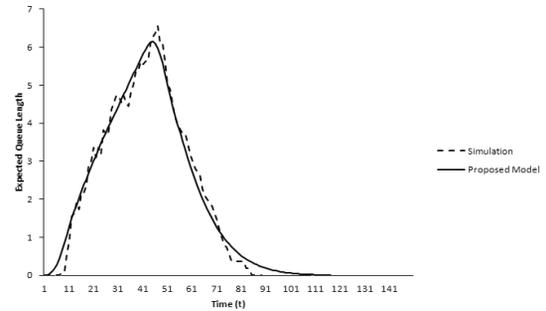


(b) Customer Arrival Rates under Appointment Policy A6

[Figure 3] Model Validation : Customer Arrival Rates



(a) Expected Queue Length under Appointment Policy A4



(b) Expected Queue Length under Appointment Policy A10

[Figure 4] Model Validation : Expected Queue Length

한 12개의 예약정책에 대하여 실행한 실험결과를 <Table 2>에 요약하였다. <Table 2>의 첫 번째 열은 식 (6)에 제시된 평균절대편차(MAE) 값을 나타낸다. 두 번째와 세 번째 열은 제안 모형과 시뮬레이션 실험의 결과로 계산한 \hat{Q}_i 과 \bar{Q}_i 의 평균값을 의미하며, 두 값이 차이를 네 번째 열에 제시하였다. 마지막으로 최대 \hat{Q}_i 을 중심으로 좌우측 시간에 대한 \hat{Q}_i 과 \bar{Q}_i 의 Kolmogorov-Smirnov 검정(K-S 검정) 결과를 제시하였다. K-S 검정은 예측치와 실측치 차이의 누적값을 이용하여 확률분포의 적합성을 판단하는데 사용된다. \hat{Q}_i 은 시계열 자료로 확률분포를 의미하지는 않지만 시간에 따른 오목형태의 특성을 갖는다. 따라서 최대 \hat{Q}_i 을 중심으로 좌우 구간에 대하여 구간별 빈도가 중복되지 않는 누적분포의 특성을 갖는다. 이와 같은 자료의 특성을 기반으로 \hat{Q}_i (예측치)과 \bar{Q}_i (실측치)의 누적 차이값을 이용한 Kolmogorov-Smirnov 검정을 이용하여 \hat{Q}_i 의 적합성을 평가하도록 한다.

실험결과를 보면 평균절대편차(MAE)가 0.09~0.15로 예약정책과 관계없이 일정한 수준의 값을 보

이고 있다. 평균절대편차의 절대 수치를 기준으로 모형의 적합성을 평가할 수는 없지만 일정 수준에서 제안 모형이 시뮬레이션 결과를 근사하고 있음을 알 수 있다. 절대적인 평가 기준을 제시하지는 않지만 매 기간 평균 0.1명 수준의 편차는 예약정책을 평가 분석하는데 있어 비교적 작은 수준의 편차라고 판단된다. 서비스 시스템의 전반적인 성능을 나타내는 평균 대기고객 수(즉, \hat{Q}_i 과 \bar{Q}_i 의 평균값)를 기준으로 모형의 적합성을 살펴보면 대체적으로 모형을 통하여 계산한 결과가 5%이내의 차이로 시뮬레이션 실험결과와 일치하는 것을 확인할 수 있었다.

마지막으로 K-S 검정 결과 또한 모형의 적합성이 높은 것을 제시하고 있다. 5% 유의수준에 대하여 일부 우측 분포를 제외하고 모든 예약정책에 대하여 \hat{Q}_i 과 \bar{Q}_i 의 분포가 동일하다는 귀무가설을 기각할 수 없었다. K-S 검정 결과는 모형의 적합성 향상과 관련하여 중요한 방안을 제시하고 있다. 최대 \hat{Q}_i 를 중심으로 구분하여 실시한 K-S 검정 결과를 보면 대부분의 경우에 있어 우측 시점에 대한 p-value(괄호안의 값)가 작은 것을 확인할 수 있

<Table 2> Model Validation : Expected Queue Length

Policy	MAE	Mean \hat{Q}_i	Mean \bar{Q}_i	$ \bar{Q}_i - \hat{Q}_i / \bar{Q}_i \times 100(\%)$	p-values*
A1	0.13	2.18	2.23	2.24	1.000(0.437)
A2	0.12	2.39	2.45	2.45	0.956(0.030)
A3	0.15	2.41	2.45	1.74	1.000(0.060)
A4	0.09	1.69	1.74	2.87	0.929(0.014)
A5	0.10	1.69	1.74	2.31	0.526(0.032)
A6	0.14	1.71	1.77	3.39	0.324(0.021)
A7	0.09	1.24	1.23	0.81	0.820(0.017)
A8	0.12	1.33	1.32	0.76	0.386(0.130)
A9	0.12	1.33	1.37	2.92	0.475(0.037)
A10	0.16	2.36	2.44	3.28	0.585(0.587)
A11	0.14	1.69	1.77	4.49	0.477(0.120)
A12	0.11	1.43	1.38	3.62	0.685(0.302)

주) * Max \hat{Q}_i 을 기준으로 좌(우)측 시간에 대한 Kolmogorov-Smirnov 검정 결과.

다. 즉, 최대 \hat{Q}_t 이후의 시점에서 본 논문에서 제안한 모형의 정확도가 낮아지는 것을 의미하는데 이와 같은 특징은 [Figure 4]에서도 확인 할 수 있다. 즉, 해당 구간에서는 대체적으로 \hat{Q}_t 가 \bar{Q}_t 보다 큰 값을 갖고 완만한 기울기로 감소하고 있다. 이와 같은 결과에 대한 분석은 추후 보다 상세하게 이루어져야 하겠지만 식 (3)과 식 (4)의 특성에서 일부 원인을 짐작할 수 있다. 식 (3)과 식 (4)는 조건부 확률과 주변 확률을 재귀적으로 적용하여 \hat{Q}_t 를 계산하는데, 매우 작은 확률이라도 많은 수의 고객이 이전 기간에 존재한다면 이후 기간에 대하여 지속적으로 \hat{Q}_t 의 기댓값을 높이는 영향을 주게 된다.

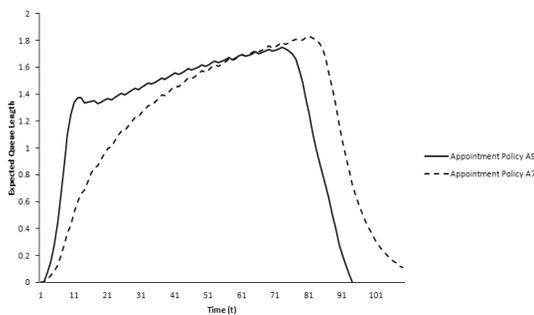
<Table 2>의 결과에 따라 \hat{Q}_t 가 시뮬레이션 결과를 작은 차이로 근사하고 있음을 알 수 있지만, 모형의 적합성을 향상시키기 위한 몇 가지 방안을 추가적으로 고려해 볼 수 있다. 첫째, 본 연구에서는 50회의 시뮬레이션 실험을 통하여 유의미한 결과를 얻었지만 시뮬레이션 횟수를 보다 증가시킨다면 \bar{Q}_t 의 변동성을 줄일 수 있을 것이다. \bar{Q}_t 의 변동성이 모형의 적합성과 절대적인 관계를 갖지는 않지만 \bar{Q}_t 의 변동에 의하여 발생하는 절대오차를 감소시키는 것이 가능할 것이다. 참고로 시뮬레이션 횟수를 20회에서 50회로 증가시킨 경우 \bar{Q}_t 와 \hat{Q}_t 의 차이를 약 1~9%p, 평균 2.22%p 감소시킬 수 있었다. 또한 본 모형에서는 연속 시간을 τ 구간으로

구분하여 서비스 시스템을 분석하였는데, τ 를 작음으로 설정한다면 계산의 복잡도는 증가하지만 정확한 결과를 얻을 수 있을 것이다. 본 연구의 결과 만족할 만한 수준에서 모형의 적합성을 검증하였으므로 τ 와 관련한 추가 실험은 추후 연구로 남기고 다음 장에서는 예약정책에 따른 시스템의 변화를 분석하도록 한다.

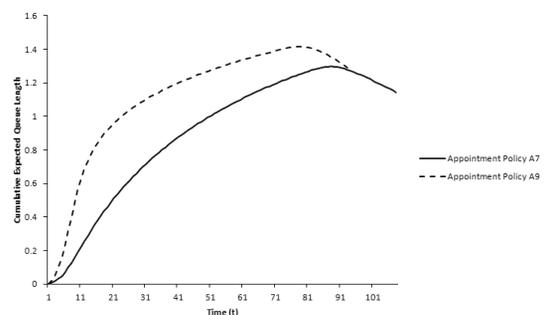
3.3 예약정책에 따른 대기 고객수의 변화

<Table 2>에서 예약정책에 따른 고객 대기시간의 변화를 대기 고객의 수 \hat{Q}_t 를 이용하여 확인할 수 있었다. 일반적으로 Bailey's rule과 블록 예약정책은 자원의 효율성을 향상시키지만 고객의 대기시간을 증가시키는 것으로 알려져 있는데[13], 문헌의 결과와 동일하게 <Table 2>의 \hat{Q}_t 또한 초기 예약시간에 할당된 고객의 수가 증가할수록, 즉 Bailey's rule을 적용한 경우 \hat{Q}_t 가 증가하였다. 예를 들어, (A1, A23, A3), (A4, A5, A6) 그리고 (A7, A8, A9)와 같이 예약주기가 동일한 정책들을 비교하면 \hat{Q}_t 가 증가함을 확인할 수 있다. 또한 A1과 A10, A4와 A11, A7과 A12의 결과를 비교하면 블록예약정책에 의하여 대기 고객이 수가 증가함을 알 수 있다.

Bailey's rule에 의한 영향은 [Figure 5]에 제시한 대기 고객 수의 변화를 통하여 보다 자세하게 확인



(a) Expected Queue Length : \hat{Q}_t



(b) Cumulative Expected Queue Length

[Figure 5] Expected Queue Length under Bailey's Rule

할 수 있다. Bailey's rule을 적용한 예약정책 A9의 경우 초기에 여러 명의 고객이 동시에 도착하게 되므로 A7보다 \hat{Q}_i 와 \hat{Q}_e 의 누적 평균값이 빠르게 증가한다. 하지만 예약정책 A9의 경우 어느 정도 시간이 지나면 도착하는 고객의 수가 감소하므로 예약정책 A7에 비하여 빠른 시간에 모든 고객에 대한 서비스를 완료하게 된다. 본 연구에서는 모든 예약정책에 대하여 20명의 고정된 고객을 단일 서버를 이용하여 처리하는 시스템을 고려하였다. 따라서 서비스의 종료 시간이 줄어든 것은 서버의 효율성(utilization)이 향상된 것으로 해석할 수 있다. [Figure 5](b)에서 보는 바와 같이 Bailey's rule을 활용한 예약정책 A9의 경우 A7과 비교하여 약 70분 정도 종료시간을 단축할 수 있었다.

4. 결론 및 추후 연구 방향

대기행렬 모형은 다양한 종류의 서비스 시스템을 분석 설계함에 있어서 기반 이론으로서 활용되었다. 하지만 대기행렬 모형의 경우 독립적이고 정상 상태를 만족하는 고객 도착 프로세스를 기본 가정으로 하고 있는데, 이와 같은 가정은 예약 기반 서비스 시스템에서는 적합하지 않다. 예약 기반 서비스 시스템의 경우 한정된 수의 고객 도착 프로세스는 예약정책에 종속적으로 결정되므로 독립적인 고객 도착 프로세스를 만족하지 않는다.

본 논문에서는 고객이 예약시간에 따라 도착하는 서비스 시스템을 분석하기 위한 모형을 제시하고, 다양한 예약정책에 대하여 시뮬레이션 실험을 수행하여 제안 모형의 적합성을 검증하였다. 대부분의 예약정책에 대하여 매우 적은 수준의 오차로 시뮬레이션과 동일한 결과를 얻을 수 있었다. 따라서 시뮬레이션과 같이 시간과 노력이 많이 소요되는 방법을 사용하지 않고, 본 논문에서 제시한 모형을 활용함으로써 예약 기반 서비스 시스템을 보다 효율적으로 분석할 수 있을 것이다.

본 연구는 몇 가지 관점에서 보완 및 확장이 요구된다. 우선 예약 기반 서비스 시스템에서 매우 중

요한 요소인 노쇼(no-show)에 대한 고려가 요구된다. 노쇼(No-show)는 사전 공지 없이 예약된 고객이 도착하지 않는 것을 의미하는데 실제 서비스 시스템의 전반적인 성과에 매우 큰 영향을 주는 것으로 알려져 있다[11, 12, 15]. 본 논문의 제안 모형에서 시점별 도착 확률 $P_{i,t}$ 에 노쇼 확률을 곱하여 고객의 노쇼를 표현할 수 있을 것이다.

고객의 특성과 관련하여 본 연구에서는 예약 없이 방문하는 고객을 고려하지 않았다. 예약이 확정된 고객만을 대상으로 하는 것은 대형 외래병원과 같은 환경에서 확인할 수 있으며 예약 기반 서비스 시스템과 관련한 많은 연구가 유사한 가정을 하고 있다[4]. 하지만 최근 예약 없이 방문하는 고객이나[19, 24] 고객 우선순위를 고려한 예약 정책에 대한 연구가 제시되고 있다[26]. 본 연구에서는 예약시간만을 고려한 단일 유형의 고객을 고려하였는데 앞서 제시한 바와 같이 다양한 유형의 고객을 고려한 예약정책을 표현하는 것은 중요한 연구 주제가 될 것이다.

마지막으로 본 논문에서 제안한 모형의 활용 측면에서 연구의 확장 방향을 고려할 수 있다. 본 연구에서는 <Table 1>에 제시한 바와 같이 정적으로 정의된 12개의 예약 정책에 대하여 서비스 시스템을 시점별로 평가하기 위한 모형을 제시하였다. 최근 동적 예약정책, 즉 예약구간과 블록크기 등을 서비스 시스템의 상황에 따라 동적으로 결정하는 연구가 제시되고 있다[20, 25]. 동적 예약정책을 수립하기 위해서는 시점별로 시스템 상태에 대한 평가가 필요한데 관련하여 본 연구의 결과를 활용한 동적 예약정책 수립 절차를 개발하는 것이 가능할 것이다.

예약정책이 수요의 관점에서 활용 방안이라고 하면 공급측면에서 본 연구의 활용방안 또한 고려할 수 있다. 서비스 자원의 공급 수준을 결정함에 있어 고객 도착 프로세스의 정확한 분석은 매우 중요하다. Defraeye and Nieuwenhuys[6]는 고객 수요가 변동하는 상황에서 서비스 용량 결정을 위한 다양한 방법을 제시하고 있으나, 예약 기반 서비스 시스템을 고려한 연구는 존재하지 않는다. 따라서 본 연구의 결과를 이용하여 예약 시스템에서의 서비스

용량 결정 문제에 대한 새로운 해법 제시는 의미 있는 연구가 될 것이다.

참 고 문 헌

- [1] 김영한, 이재용, “인터넷의 실시간 서비스 기술”, 『한국통신학회지』, 제14권, 제4호(1997), pp.367-382.
- [2] 신양우, 예약된 고객이 존재하는 대기체계의 분석과 그 응용, 2001년도 지역대학우수과학자 지원연구 결과보고서, 한국과학재단, 2002.
- [3] Bailey, N., “A study of queues and appointment systems in hospital outpatient departments with special reference to waiting times,” *Journal of the Royal Statistical Society*, Vol. 14(1952), pp.185-199.
- [4] Cayirli, T. and E. Veral, “Outpatient scheduling in health care : A review of literature,” *Production and Operations Management*, Vol.12, No.4(2003), pp.519-549.
- [5] Creemers, S. and M. Lambrecht, “Queueing models for appointment-driven systems,” *Annals of Operations Research*, Vol.178(2010), pp.155-172.
- [6] Defraeye, M. and I. Van Nieuwenhuysse, *Setting staffing levels in systems with time-varying demand : The context of an emergency department*, Working Paper, 2011.
- [7] Doi, M., Y. Chen, and H. Osawa, “A queueing model in which arrival times are scheduled,” *Operations Research Letters*, Vol.21 (1997), pp.249-252.
- [8] Green, L.V., P.J. Kolesar, and J. Soares, “Improving the SIPP approach for stang service systems that have cyclic demands,” *Operations Research*, Vol.49, No.4(2001), pp.549-564.
- [9] Green, L.V., P.J. Kolesar, and J. Soares, “An improved heuristic for staffing telephone call centers with limited operating hours,” *Production and Operations Management*, Vol. 12, No.1(2003), pp.46-61.
- [10] Green, L.V., P.J. Kolesar, and W. Whitt, “Coping with time-varying demand when setting staffing requirements for a service system,” *Production and Operations Management*, Vol.16, No.1(2007), pp.13-39.
- [11] Gupta, D. and B. Denton, “Appointment scheduling in health care : Challenges and opportunities,” *IIE Transactions*, Vol.40, No.9(2008), pp.800-819.
- [12] Hassin, R. and S. Mendel, “Scheduling arrivals to queues : A single-server model with no-shows,” *Management Science*, Vol.54, No.3 (2008), pp.565-572.
- [13] Ho, C. and H. Lau, “Minimizing total cost in scheduling outpatient appointments,” *Management Science*, Vol.38, No.12(1992), pp.1750-1764.
- [14] Ho, C., H. Lau, and J. Li, “Introducing variable-interval appointment scheduling rules in service systems,” *International Journal of Operations and Production Management*, Vol.15, No.6(1994), pp.59-68.
- [15] Hofmann, P. and J. Rockart, “Implications of the no-show rate for scheduling OPD appointments,” *Hospital Progress*, Vol.50, No.8(1969), pp.35-40.
- [16] Hyndman, R.J. and A.B. Koehler, “Another look at measures of forecast accuracy,” Technical Report, 2005.
- [17] Ingolfssona, A., F. Campelloa, X. Wub, and E. Cabral, “Combining integer programming and the randomization method to schedule employees,” *European Journal of Operational Research*, Vol.202, No.1(2010), pp.153-163.

- [18] Jouini, O. and S. Benjaafar, "Queueing systems with appointment-driven arrivals, non-punctual customers, and no-shows," Working Paper, 2010.
- [19] Lee, S. and Y. Yih, "Analysis of open access appointment scheduling system in outpatient clinics—a simulation study," *Simulation-Transactions of the Society for Modeling and Simulation International*, Vol.86, No.8/9(2010), pp.503-518.
- [20] Liu, N., S. Ziya, and V.G. Kulkarni, "Dynamic scheduling of outpatient appointments under patient no-shows and cancellations," *Manufacturing and Service Operations Management*, Vol.12, No.2(2010), pp. 347-364.
- [21] Min, D. and Y. Yih, "A simulation study of registration queue disciplines in an outpatient clinic : a two-stage patient ow model," *European Journal of Industrial Engineering*, Vol.3, No.2(2009), pp.127-145.
- [22] Moorthy, R. and C. Teo, "Berth management in container terminal : the template design problem," *OR Spectrum*, Vol.28, No.4(2006), pp.495-518.
- [23] Muller, R. and G. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," In AIAAs Aircraft Technology, Integration, and Operations (ATIO), 2002.
- [24] Murray, M. and C. Tantau, "Same-day appointments : exploding the access paradigm," *Family Practice Management*, Vol.7, No.8 (2000), pp.45-50.
- [25] Muthuraman, K. and M. Lawley, "A stochastic overbooking model for outpatient clinic scheduling with no-shows," *IIE Transactions*, Vol.40, No.9(2008), pp.820-837.
- [26] Patrick, J., M. Puterman, and M. Queyranne, "Dynamic multi-priority patient scheduling for a diagnostic resource," *Operations Research*, Vol.56, No.6(2008), pp.1507-1525.
- [27] Rohleder, T. and K. Klassen, "Using client-variance information to improve dynamic appointment scheduling performance," *Omega*, Vol.28, No.3(2000), pp.293-302.
- [28] Steckley, S., S. Henderson, and V. Mehrotra, "Service system planning in the presence of a random arrival rate," Technical Report, 2004.
- [29] Winsten, C.B., "Geometric distribution in the theory of queues," *Journal of Royal Statistical Society*, Vol.21(1959), pp.1-35.