2005 한국경영과학회/대한산업공학회 춘계공동학술대회
2005년 5월 13일~14일, 충북대학교

마코비안 서비스 과정을 가지는 대기행렬 모형의 다양한 시점 하에서의 고객수 분포들의 관계에 대한 소고

이상민·채경철
(한국과학기술원 산업공학과)

A Note on Relationship among Queue Lengths at Various Epochs of a Queue with MSP Services

Sang M. Lee·Kyung C. Chae

Abstract

Markovian Service Process(MSP)는 기존의 Markovian Arrival Process(MAP)에서 사용하던 위상 개념을 고객의 서비스 과정에 대응시킨 모형이다. 이는 서버의 상태에 따라 달라질 수 있는 서비스 상태를 위상 변화에 대응시키는 모형이다. 본 논문에서는 대기행렬 모형의 중요한 성능 척도인 고객수 분포에 관하여 임의시점, 고객 도착 직전 시점, 고객 이탈 직후 시점에서의 관계식을 유도한다.

1. 서론

기존의 대기행렬 모형은 대부분 서비스의 속도(rate)가 일정하다는 가정 하에 많은 분석이 이루어져 왔다. MSP는 이러한 가정을 완화하여 서버의 속도(rate)가 서버의 상태 혹은 시간의 흐름에 따라 달라지는 서비스의 과정을 나타내는 모형이다. 이는 기존의 고객 도착과정 중의 하나인 MAP의 개념을 서비스 과정에 도입한 새로운 서비스 과정의 모델링 기법이다.

MAP 도착과정은 통신/컴퓨터 시스템에서 실제 고객들의 도착간격 간의 상관성을 위상의 개념으로 표현한 고객 도착과정 모델링 기법으로서 현재까지 많은 학자들에 의하여 분석되어 왔다. 이에 반하여 MSP 서비스과정은 근래에 대한 논문은 최근에 몇개가 나왔는데 이를 소개하면 다음과 같다.

Baykal-gursoy와 Xiao[1]는 MMSP (Markov modulated service rates)를 가지는 무한서버 대기행렬 시스템을 분석하여 시스템 내의 고객수 분포를 구하였다. Ozawa[2]는 MSP 서비스 과정을 도입하여 서버의 휴가 정책(복수 휴가 정책, N 정책 등)을 표현하였으며 Takine[3]는 MAP과 MSP가 동일한 UMC(underlying Markov chain)을 가지는 모형에서 시간축의 변화를 이용하여 바쁜 시간 분석 및 고객 대기시간의 분석을 수행하였다.

본 논문에서는 고객의 도착과정은 일반 분포를 가지고 고객의 서비스 과정은 MSP를 가지는 GI/MSP/1 모형에서의 근본적인 관계(고객 도착 직전 고객수 분포, 고객 이탈 직후 고객수 분포, 임의시점 고객수 분포들 간의 관계)를 구하고자 한다. 근본적인 관계란 대기 행렬 모형의 분석에 있어 사용되는 중요한 관계로서, 대표적인 대기행렬 모형중의 하나인 M/G/1 대기행렬 모형은 PASTA(Poisson arrivals see time average)와 Burke의 정리라는 관계가 규명되어 있다[4]. 또한 Kim *et al.*[5]은 “rate in=rate out"의 평형 방정식을 이용하여 BMAP/G/1과 D-BMAP/G/1 대기행렬 모형에서의 근본적인 관계를 유도하였다.

본 논문의 구성은 다음과 같다. 먼저 본 논문에서 사용하는 용어를 정의한다. 그리고 임의 시점 고객수와 고객 도착 직전 시점 고객수 분포의 관계와

임의 시점과 고객 이탈 직후 시점 고객수 분포의 관계를 구하며, 마지막으로 결론 및 추후 연구에 관하여 언급한다.

2. 용어 정의

GI/MSP/1 대기행렬 모형의 도착은 평균이 $1/\lambda$인 일반 분포(general distribution)을 가진다. 서비스 과정은 MSP를 따르며, MSP 서비스 종료 과정의 위상 변화는 다음의 두 행렬에 의하여 설명된다.(위상 수는 m)

$$T=\begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ t_{21} & t_{22} & \cdots & t_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mm} \end{pmatrix},$$

$$U=\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mm} \end{pmatrix}$$

행렬 T는 $t_{i,j}(i \neq j)$의 rate에 의하여 스위치 상태가 i인 조건 하에서, 서비스 종료에 의하지 않고, 스위치 자체적으로 상태가 j로 바뀌는 과정을 설명하는 행렬이다. 행렬 U는 이탈률 행렬으로 u_{ij}는 위상 상태가 i인 조건 하에서 고객이 이탈하면서 스위치 상태를 k로 바꾸는 이탈률 행렬이다. $Q=T+U$는 MSP의 UMC의 전이율 핼렬로 정의한다.

GI/MSP/1 대기행렬 모형의 상태를 $\{L(T),S(t),t \geqq 0\}$로 정의하자. $L(t)$와 $S(t)$는 각각 t시점에서의 시스템내의 고객수와 위상의 상태를 나타낸다. 본 논문에서 사용하는 확률들을 $i \geqq 0$, $1 \leqq j \leqq 0$의 범위에서 정의하면 다음과 같다.

$y_{i,j}$: 임의 시점에서 상태 (i,j)에 있을 확률
$y_i=(y_{i,1}, y_{i,2}, \cdots, y_{i,m})$
$y_{i,j}^+$: 고객 이탈(서비스 종료)후에 시스템 내에 i명을 남기고, UMC의 위상은 j일 확률
$y_i^+=(y_{i,1}^+, y_{i,2}^+, \cdots, y_{i,m}^+)$
$x_{i,j}$: 도착하는 고객이 i명을 보고, UMC의 위상은 j일 확률
$x_i=(x_{i,1}, x_{i,2}, \cdots, x_{i,n})$

다음으로 Kim *et al.*[5]이 BMAP/G/1 과 D-BMAP/G/1의 근본적인 관계 유도를 위하여 사용한 “rate in=rate out"의 평형 방정식을 이용하여 GI/MSP/1에서의 임의 시점 고객수와 고객 도착 직전 시점 고객수 분포의 관계식과 임의 시점과 고객 이탈 직후 시점 고객수 분포의 관계를 유도하겠다.

3. 임의 시점 고객수와 고객 도착 직전 시점 고객수 분포의 관계

이 장에서는 “rate in=rate out"의 관계를 이용하여 GI/MPS/1의 임의 시점 고객수 와 도착 직전 시점 고객수의 관계를 유도한다.

Theorem 1.
GI/MSP/1에서 $Y(z,B)$와 $X(z)$는 다음과 같은 관계를 가진다.

$$Y(z,B)(\frac{U+Tz}{z})=\lambda(1-z)X(z) \qquad (1)$$

where

$Y(z,B)=\sum_{i=1}^{\infty} y_i z^i$ (바쁜 기간에서의 임의시점 고객수 PGF)

$X(z)=\sum_{i=0}^{\infty} x_i z^i$ (도착 직전 시점 고객수 PGF)

증명. $\{L(T),S(t),t \geqq 0\}$의 상태 (i,j)에서, (i,j)로 들어오는 속도(rate)와 (i,j)를 빠져나가는 속도는 안정상태에서 항상 동일하다.(“rate in=rate out", [5]) 먼저 상태 (i,j)에서 빠져나가는 속도를 구해보자. (i,j)를 빠져나가는 경우는 고객 도착 혹은 UMC의 변화에 따르는 두 가지 경우가 있다.

먼저 고객 도착에 따른 빠져나가는 속도를 다음과 같이 구한다. 고객들은 단위 시간당 λ명 도착하

고, $x_{i,j}$는 전체 도착하는 고객수 중 상태 (i,j)를 보는 비율이다. 따라서 $\lambda x_{i,j}$는 단위 시간당 상태 (i,j)를 떠나는 속도이다.

두 번째로 UMC의 변화에 의해 빠져나가는 속도는 다음과 같다. 상태 (i,j)에 있을 경우 UMC의 위상 변화는 고객의 이탈과 함께 변하는 경우와 고객의 이탈 없이 위상이 변화는 경우로 나뉘는데 각각 T와 U행렬의 원소의 속도로 변화한다. 두 경우를 고려한 속도를 계산하면 다음과 같다.

$$y_{i,j}\left(\sum_{l=j}^{m}(T)_{j,l}+\sum_{l\neq j}^{m}(U)_{j,l}\right)$$

따라서 상태 (i,j)를 떠나는 속도는

$$y_{i,j}\left(\sum_{l=j}^{m}(T)_{j,l}+\sum_{l\neq j}^{m}(U)_{j,l}\right)+\lambda x_{i,j} \qquad (2)$$

로 나타낼 수 있다.

다음애는 상태 (i,j)로 들어오는 속도를 구한다. 이는 위와 동일하게 고객 도착 혹은 UMC의 변화에 따르는 두 가지 경우가 있다.

먼저 고객 도착에 따라 들어오는 속도를 다음과 같이 구한다. 고객들은 단위 시간당 λ명 도착하고, $x_{i-1,j}$는 전체 도착하는 고객수 중 상태 $(i-1,j)$를 보는 비율이다. 따라서 $\lambda x_{i-1,j}$는 단위 시간당 상태 (i,j)로 들어오는 속도이다.

두 번째로 UMC의 변화에 의해 들어오는 속도는 다음과 같다. 고객이 시스템 내에 $i+1$명 있었을 때, 이탈하면서 UMC의 위상이 j로 바뀌는 경우와 고객이 시스템 내에 i명 있었을 때, 고객 이탈 없이 UMC의 위상이 j로 바뀌는 두 가지 경우가 있다. 두 경우를 고려한 속도를 계산하면 다음과 같다.

$$\sum_{l=1}^{m}y_{i+1,l}(U)_{l,j}+\sum_{l\neq j}^{m}y_{i,l}(T)_{l,j}$$

따라서 상태 (i,j)로 들어오는 속도는

$$\sum_{l=1}^{m}y_{i+1,l}(U)_{l,j}+\sum_{l\neq j}^{m}y_{i,l}(T)_{l,j}+\lambda x_{i-1,j} \qquad (3)$$

로 나타낼 수 있다

(2)과 (3)를 사용하여 “rate in=rate out"으로 놓고 정리하면,

$$\sum_{l=1}^{m}(y_{i+1,l}(U)_{l,j}+y_{i,l}(T)_{l,j})=\lambda(x_{i,j}-x_{i-1,j})$$

이 되고 양변을 PGF화 하면 *Theorem 1*의 결과를 얻는다.

Remark 1.

고객의 서비스 시간이 평균이 $1/\mu$인 지수 분포를 따른 다고 가정하면(GI/M/1의 경우), $U=\mu$, $T=-\mu$를 대입하여 $Y(z,B)=z\frac{\lambda}{\mu}X(z)$를 얻을 수 있다.

4. 임의 시점과 고객 이탈 직후 시점 고객수 분포의 관계

이 장에서는 GI/MPS/1의 임의 시점과 고객 이탈 직후 시점 고객수 분포의 관계를 유도한다.

Theorem 2.

GI/MSP/1에서 $Y(z,B)$와 $Y^{+}(z)$는 다음과 같은 관계를 가진다.

$$Y(z)U=\lambda zY^{+}(z)$$

where

$$Y^{+}(z)=\sum_{i=0}^{\infty}y_i^{+}z^i \text{ (이탈 직후 시점 고객수 PGF)}$$

증명. 고객들은 단위 시간당 λ명이 이탈하고, $y_{i,j}^{+}$는 전체 이탈하는 고객수 중 상태 (i,j)를 남기는 비율이다. 따라서 $\lambda y_{i,j}^{+}$는 단위 시간당 상태 (i,j)를

남기는 고객수이다. 이는 시스템 내에 $i+1$명이 있는 상태에서 U의 원소의 속도로 상태 (i,j)가 되는 횟수와 동일하므로 다음과 같은 식을 얻는다.

$$\lambda y_{i,j}^{+} = \sum_{l=1}^{m} y_{i+1,l}(U)_{l,j}.$$

양변을 PGF화 하면 *Theorem 2*의 결과를 얻는다.

Remark 2.

고객의 서비스 시간이 평균이 $1/\mu$인 지수 분포를 따른 다고 가정하면(GI/M/1의 경우), $U=\mu$, $T=-\mu$를 대입하여 $Y(z,B)=z\frac{\lambda}{\mu}Y^{+}(z)$를 얻을 수 있다. (GI/M/1의 경우 $Y^{+}(z)=X(z)$)

4. 결론 및 추후연구

본 논문에서는 GI/MSP/1 대기행렬 모형의 임의 시점 고객수와 고객 도착 직전 시점 고객수 분포의 관계 및 임의 시점과 고객 이탈 직후 시점 고객수 분포의 관계를 구하였다. 위의 관계를 유도하기 위하여 사용한 "rate in=rate out“의 관계는 매우 유용하여 GI/MSP/1 만이 아닌 MAP/MSP/1, G/MSP/1으로도 확장된다고 사료된다. 또한 무한 용량(infinite buffer) 시스템이 아닌 유한 용량(finite buffer)에서도 동일한 형태의 관계를 얻어 낼 수 있을 것으로 사료되어 더욱 더 큰 틀의 대기행렬모형의 분석을 행하는데 큰 도움이 될 것이다.

참고 문헌

[1] Baykal-Gursoy, M. and W. Xiao, "Stochastic Decomposition in M/M/∞ Queues with Markov Modulated Service Rates", *Queueing Systems*, Vol. 48(2004), 75-88.

[2] Kim, N.K., S.H. Chang and K.C. Chae, "On the Relationships among Queue Lengths at Arrival, Departure, and Random Epochs in the Discrete-time Queue with D-BMAP Arrivals", *Operation Research Letters*, Vol. 30, No.1(2002), 25-32.

[3] Ozawa, T., "Analysis of Queues with Markovian Service Processes", *Stochastic Models*, Vol. 20, No.4(2004), 391-413.

[4] Ross, S.M., *Stochastic Processes*, 2nd edition, John Wiley & Sons, New York, 1996.

[5] Takine, T., "Single-Server Queues with Markov-Modulated Arrivals and Service Speed", *Queueing Systems*, Vol. 49(2005), 7-22.