

계층 전이가 가능한 다계층 대기행렬의 빠른 시뮬레이션

송미정^a, 배경순^a, 이지연^{1,a}

^a영남대학교 통계학과

요약

계층 전이가 가능한 다계층 대기행렬에 대한 빠른 시뮬레이션 방법을 제시하고 이를 이용하여 총 고객의 수가 적정 수준을 초과하는 과부하 발생 확률과 과부하 발생시 각 계층별 평균 고객의 수를 추정한다. 또한 이들 값들이 시스템의 총 로드나 계층별 로드에서 어떤 차이를 보이는지도 빠른 시뮬레이션을 통해 서로 비교한다.

주요용어: 다계층 대기행렬, 계층 전이, 빠른 시뮬레이션, 과부하 확률, 중요 샘플링 추정량.

1. 서론

여러 계층의 고객들이 섞여 도착하는 다계층(multi-class) 대기행렬은 계층에 따라 제공되는 서비스 시간의 분포가 다를 경우에 준역행가능(quasi-reversible, McDonald, 2004)하지 않기 때문에 계층별 고객수에 대한 분포를 일반적인 급의 형태로 구할 수 없다. 다만, 각 계층별 고객수의 분포에 대한 점근적(asymptotic) 결과를 얻거나 혹은 시뮬레이션(simulation)을 통해 분포를 추정할 수 있다. 시뮬레이션을 이용하는 경우에도 안정된(stable) 시스템의 과부하(overflow) 발생처럼 무척 드물게 발생하는 사건의 경우는 일반 시뮬레이션보다는 적절히 확률측도를 변화시켜 과부하 발생을 촉진시킨 후 추정하는 빠른 시뮬레이션(fast simulation, Heidelberger, 1995)을 적용하는 것이 필요하다.

Dabrowski 등 (2008)과 Lee와 Bae (2007)는 고객들이 서비스를 받은 후 바로 시스템을 이탈하는 다계층 대기행렬의 각 계층별 고객수에 대한 확률분포의 점근적 결과를 유도하였고, Song 등 (2007)은 서비스를 마친 후 같은 계층의 고객으로 시스템에 재입장(feedback)하는 고객이 있는 다계층 대기행렬에서 과부하가 발생하는 확률을 빠른 시뮬레이션으로 추정하였다. 하지만 서비스를 마친 고객이 바로 시스템을 이탈하거나 또는 동일한 계층의 고객으로만 재입장할 수 있는 모형은 여러 제약들이 포함되어 있기 때문에 실생활에서 일어날 수 있는 상황을 적용시키기에는 한계가 있다. 예를 들어 은행의 창구에는 비교적 서비스 시간이 짧은 입출금 고객, 시간이 조금 더 필요한 예금의 신규 가입이나 카드 발급 등 상담이 필요한 고객들이 섞여서 도착한다. 이 때 입금하기 위해 창구에 입장한 고객이 서비스를 받고 나서 바로 은행을 나갈 수도 있고, 다른 출금 업무를 해결하기 위해 다시 창구로 입장할 수도 있다. 또한 입출금 업무 외에 새로운 예금에 가입하거나 카드 발급을 위한 상담을 위해 창구로 재입장하는 경우가 발생하기도 한다. 이와 같이 여러 계층의 고객이 시스템으로 입장하여 서비스를 받고 난 후에 시스템을 바로 떠나거나 혹은 동일한 서비스나 다른 종류의 서비스를 받기 위해 재입장하는 경우에도 적용이 가능한 계층 전이가 있는 다계층 대기행렬의 연구가 필요하며, 이 모형은 기존에 분석했던 다계층 대기행렬을 모두 포함하는 확장된 모형이 된다.

이 논문은 2005년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2005-204-C00016).

¹ 교신저자: (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 교수. E-mail: lee@yu.ac.kr

일반적으로 대기행렬 시스템의 성능은 전체 고객의 수가 시스템의 수용가능 수를 초과하는 과부하의 발생 확률로 측정할 수 있다. 안정된 시스템의 과부하 확률을 추정하기 위해서는 먼저 원래의 시스템을 불안정(unstable)한 시스템으로 바꿀 수 있는 h -변환(h -transform, McDonald, 1999)을 찾는다. 그리고 이 변환을 통해 얻어진 불안정한 시스템에서의 과부하 발생 확률의 중요 샘플링 추정량(importance sampling estimator, Heidelberger, 1995)으로 빠른 시뮬레이션을 적용한다. 이 방법을 통해 Lee와 Bae (2007)는 서비스 종료 후 고객들이 바로 이탈하는 다계층 대기행렬에서 과부하 확률을 구하였고, Song 등 (2007)은 동일 계층의 재입장이 있는 다계층 대기행렬에서의 과부하 확률을 추정하였다.

본 논문에서는 기존 연구와 동일한 형태의 h -변환을 이용하여, 계층 전이가 가능한 두 개의 계층의 고객들로 구성된 대기행렬의 과부하 확률 추정에도 적용될 수 있음을 보이고, 이를 이용하여 과부하 발생시 각 계층의 평균 고객수를 추정하고자 한다. 2장에서 계층 전이가 가능한 다계층 대기행렬을 소개하고, McDonald (1999)의 h -변환을 이용하여 빠른 시뮬레이션을 적용하기 위한 변화된 불안정한 시스템을 찾는다. 그리고 불안정한 시스템을 이용하여 과부하 확률과 과부하 발생시 계층별 평균 고객수를 추정하기 위한 중요 샘플링 추정량을 구한다. 3장에서는 예제를 통해 일반 시뮬레이션과 빠른 시뮬레이션 결과를 비교하고, 시스템의 총 로드와 계층별 로드와 따른 과부하 확률과 과부하 발생시 계층별 평균 고객의 수를 비교한다. 마지막으로 4장에서는 그 결론을 정리한다.

2. 계층 전이가 가능한 다계층 대기행렬

서비스 제공자가 한 명 있는 대기행렬 시스템에 서로 다른 두 개의 계층의 고객들이 서로 독립적으로 도착한다. 계층 i , $i = 1, 2$ 의 고객들은 도착률이 λ_i 인 포아송 과정(Poisson process)을 따르며 도착하고 자신의 서비스 차례가 될 때까지 대기열에서 기다린다. 고객에 대한 서비스는 FIFO(First In First Out)로 진행되며, 각 고객의 서비스 시간의 분포는 고객의 계층에 따라 결정되는데 계층 i 의 고객은 평균이 $1/\mu_i$ 인 지수분포(exponential distribution)의 시간만큼 서비스를 받는다. 서비스가 끝나면 r_i (> 0)의 확률로 시스템을 완전히 떠나거나, r_{ii} 의 확률로 동일한 계층의 고객으로 재입장하거나 아니면 r_{ij} 의 확률로 계층 j ($\neq i$)의 고객으로 바뀌어서 대기열의 마지막에 재입장한다. 그러므로 $i = 1, 2$ 에 대해 $r_i + r_{ii} + r_{ij} = 1$ 이 성립한다.

이 때, 계층 i 고객의 총 도착률(total arrival rate) Λ_i 는

$$\Lambda_1 = \lambda_1 + \Lambda_1 r_{11} + \Lambda_2 r_{21}$$

$$\Lambda_2 = \lambda_2 + \Lambda_1 r_{12} + \Lambda_2 r_{22}$$

를 만족하여

$$\Lambda_1 = \frac{(1 - r_{22})\lambda_1 + r_{21}\lambda_2}{(1 - r_{11})(1 - r_{22}) - r_{12}r_{21}}$$

$$\Lambda_2 = \frac{r_{12}\lambda_1 + (1 - r_{11})\lambda_2}{(1 - r_{11})(1 - r_{22}) - r_{12}r_{21}}$$

로 얻어지고, 안정된(stable) 시스템을 고려하기 위해

$$\rho := \frac{\Lambda_1}{\mu_1} + \frac{\Lambda_2}{\mu_2}$$

로 정의되는 총 로드(load, McDonald, 2004) ρ 에 대해 $\rho < 1$ 을 만족한다고 가정한다. 그리고

$$\lambda_1 + \lambda_2 + \mu_1 + \mu_2 = 1$$

을 가정하여 균일화 기법(uniformization method, Walrand, 1988)에 의해 원래의 시스템과 동일한 정상 확률을 갖는 이산 시간형 마코프 체인(Markov chain) S 를 고려한다.

마코프 체인 $S = \{X(t), t = 0, 1, 2, \dots\}$ 의 상태는 대기열 또는 서비스를 받고 있는 고객의 계층을 이용하여 고객이 한 명도 없는 경우는 $\mathbf{0}$ 으로 나타내고, 적어도 1명의 고객이 있는 상태는 벡터 $X(t) = (X_0(t), X_1(t), \dots, X_{n-1}(t))$ 와 $|X(t)| = n$ 으로 나타낸다. 이 때, $X_0(t)$ 는 시점 t 에서 대기열의 맨 앞에서 서비스를 받고 있는 고객의 계층, $X_1(t)$ 는 그 뒤에서 기다리고 있는 고객의 계층, \dots , $X_{n-1}(t)$ 는 마지막에 도착하여 대기열에서 기다리고 있는 고객의 계층을 표시한 것이고, $|X(t)| = n$ 은 시점 t 에서 총 고객의 수가 n 명임을 나타낸다. 마코프 체인 S 의 전이확률(transition probability) $K(x, y)$ 는 $x = \mathbf{0}$ 일 때,

$$K(\mathbf{0}, y) = \begin{cases} \lambda_j, & y = (j), \quad j = 1, 2, \\ \mu_1 + \mu_2, & y = \mathbf{0}, \\ 0, & \text{그 외 } y \end{cases}$$

이고, 각 $i = 1, 2$ 에 대해 $x = (i, x_1, \dots, x_{n-1})$ 일 때,

$$K(x, y) = \begin{cases} \lambda_j, & y = (i, x_1, \dots, x_{n-1}, j), \quad j = 1, 2, \\ \mu_i r_i, & y = (x_1, \dots, x_{n-1}), \\ \mu_i r_{ij}, & y = (x_1, \dots, x_{n-1}, j), \quad j = 1, 2, \\ 1 - \lambda_1 - \lambda_2 - \mu_i, & y = (i, x_1, \dots, x_{n-1}), \\ 0, & \text{그 외 } y \end{cases}$$

가 된다.

안정된 마코프 체인 S 를 불안정한 마코프 체인으로 바꾸기 위해서 마코프 체인의 전이확률 $K(x, y)$ 에 대해,

$$\text{모든 } x \text{에 대해 } h(x) = \sum_y K(x, y)h(y) \tag{2.1}$$

를 만족하는 조화함수(harmonic function) h 를 찾는다 (McDonald, 1999; Dabrowski 등, 2008). 이 조화 함수 h 를 이용하면

$$\tilde{K}(x, y) = K(x, y) \frac{h(y)}{h(x)} \tag{2.2}$$

의 새로운 전이확률 $\tilde{K}(x, y)$ 를 정의할 수 있다.

마코프 체인 S 의 상태벡터 $X(t)$ 의 값 $x = (x_0, x_1, \dots, x_{n-1})$ 에 대하여 $N_i(x)$ 를 $|x| = n$ 의 고객 중 계층 i 인 고객의 수라고 하자. 이를 이용하여 다음과 같은 형태의 함수를 정의할 수 있다 (Dabrowski 등, 2008; Lee와 Bae, 2007; Song 등, 2007);

$$h(x) := \exp(\gamma_1 N_1(x) + \gamma_2 N_2(x)).$$

이 함수가 $\{x : |x| > 0\}$ 에서 조화함수가 되기 위해서는, 즉, 식 (2.1)을 만족하기 위해서는

$$\lambda_1 e^{\gamma_1} + \lambda_2 e^{\gamma_2} + \mu_1 r_1 e^{-\gamma_1} + \mu_1 r_{11} + \mu_1 r_{12} e^{-\gamma_1 + \gamma_2} + \mu_2 = 1 \tag{2.3}$$

$$\lambda_1 e^{\gamma_1} + \lambda_2 e^{\gamma_2} + \mu_2 r_2 e^{-\gamma_2} + \mu_2 r_{22} + \mu_2 r_{21} e^{\gamma_1 - \gamma_2} + \mu_1 = 1 \tag{2.4}$$

이 성립해야 한다.

정리 1. 조건식 (2.3)과 (2.4)를 만족하는

$$(a) \exp(\gamma_1) > 1$$

$$(b) \exp(\gamma_2) > 1$$

인 해 γ_1 과 γ_2 가 항상 존재한다.

증명: $x := \exp(\gamma_1)$, $y := \exp(\gamma_2)$ 라고 두면, 식 (2.3)과 (2.4)는 x 와 y 에 대해

$$f(x) := A_1x^2 + A_2x + A_3 = 0$$

$$g(y) := -\mu_1r_{12}y^2 + \{\mu_1(1-r_{11})x - \mu_2(1-r_{22})x - \mu_1r_{1\cdot}\}y + \mu_2(r_{2\cdot} + r_{21}x)x = 0$$

을 만족한다. 여기서,

$$A_1 := \lambda_1^2\mu_1r_{12} + \lambda_1\lambda_2\{\mu_1(1-r_{11}) - \mu_2(1-r_{22})\} - \lambda_2^2\mu_2r_{21},$$

$$A_2 := -[\lambda_1\mu_1r_{12}\{\lambda_1 + \lambda_2 + \mu_1(1-r_{11}) + \mu_2(1-r_{22})\} + 2\lambda_2\mu_1\mu_2r_{12}r_{21} \\ + \lambda_2\mu_1(1-r_{11})\{\lambda_1 + \lambda_2 + \mu_1(1-r_{11}) - \mu_2(1-r_{22})\}],$$

$$A_3 := \mu_1^2[\mu_2r_{12}\{(1-r_{11})(1-r_{22}) - r_{12}r_{21}\} + r_{1\cdot}\{\lambda_1r_{12} + \lambda_2(1-r_{11})\}]$$

이다.

(a) $f(0) = A_3 > 0$ 이고,

$$f(1) = A_1 + A_2 + A_3 \\ = (\lambda_2 + \mu_1r_{12})[\mu_1\mu_2\{(1-r_{11})(1-r_{22}) - r_{12}r_{21}\}] (1-\rho) > 0$$

이므로

(i) 만약 $A_1 < 0$ 이면, $x > 1$ 인 해 x 가 항상 존재한다.

(ii) 만약 $A_1 > 0$ 이면,

$$f'(1) = \frac{1}{\lambda_1} (A_1 + \lambda_2^2\mu_2r_{21}) \{\lambda_1 - \mu_1(1-r_{11})\} - \{\lambda_1\lambda_2\mu_2(1-r_{22}) \\ + 2\lambda_2^2\mu_2r_{21} + \lambda_1\lambda_2\mu_1r_{12}\lambda_1\mu_1\mu_2r_{12}(1-r_{22}) + 2\lambda_2\mu_1\mu_2r_{12}r_{21} + \lambda_2^2\mu_1(1-r_{11})\} \\ < 0$$

이므로 $x > 1$ 인 해가 항상 존재한다.

(b) (a)에 의해 함수 $g(y)$ 는

$$g(0) = \mu_2(r_{2\cdot} + r_{21}x)x > 0$$

을 만족한다. 또한

$$g(1) = (x-1)\{\mu_2r_{21}x + \mu_1(r_{12} + r_{1\cdot})\} > 0$$

이므로 $g(y) = 0$ 을 만족하는 해 $y > 1$ 가 항상 존재한다. □

조건식 (2.3)과 (2.4)를 만족하는 γ_1 과 γ_2 로 구성된 함수 $h(\mathbf{x}) = \exp(\gamma_1 N_1(\mathbf{x}) + \gamma_2 N_2(\mathbf{x}))$ 를 식 (2.2)에 대입하면 다음과 같은 새로운 전이확률을 얻을 수 있다. 즉, $|\mathbf{x}| = n > 0$ 이고 $\mathbf{x} = (i, x_1, \dots, x_{n-1})$ 인 경우는

$$\bar{K}(\mathbf{x}, \mathbf{y}) = \begin{cases} \lambda_j \exp(\gamma_j), & \mathbf{y} = (i, x_1, \dots, x_{n-1}, j), \quad j = 1, 2, \\ \mu_i r_i \exp(-\gamma_i), & \mathbf{y} = (x_1, \dots, x_{n-1}), \\ \mu_i r_{ii}, & \mathbf{y} = (x_1, \dots, x_{n-1}, i), \\ \mu_i r_{ij} \exp(-\gamma_i + \gamma_j), & \mathbf{y} = (x_1, \dots, x_{n-1}, j), \\ 1 - (\lambda_1 + \lambda_2 + \mu_i), & \mathbf{y} = (i, x_1, \dots, x_{n-1}), \\ 0, & \text{그 외 } \mathbf{y} \end{cases} \quad (2.5)$$

가 된다.

만약 $\mathbf{x} = \mathbf{0}$ 에서의 전이확률 $\bar{K}(\mathbf{0}, \mathbf{y})$ 를

$$\bar{K}(\mathbf{0}, \mathbf{y}) = \begin{cases} \lambda_j \exp(\gamma_j), & \mathbf{y} = (j), \quad j = 1, 2, \\ 1 - \sum_j \lambda_j \exp(\gamma_j), & \mathbf{y} = \mathbf{0}, \\ 0, & \text{그 외 } \mathbf{y} \end{cases} \quad (2.6)$$

로 정의하면 식 (2.5)와 (2.6)의 전이확률을 갖는 새로운 마코프 체인 \tilde{S} 는 기존의 마코프 체인 S 로 설명되는 대기행렬과 동일한 구조를 갖는 대기행렬로서, 계층 i 인 고객이 도착률 $\tilde{\lambda}_i := \lambda_i \exp(\gamma_i)$ 인 포아송 과정을 따르며 독립적으로 도착하고, 계층 i 의 고객의 서비스 차례가 되면 평균이 $1/\tilde{\mu}_i := 1/[\mu_i\{r_i \exp(-\gamma_i) + r_{ii} + r_{ij} \exp(-\gamma_i + \gamma_j)\}]$ 인 지수분포의 시간만큼 서비스를 받는다. 그리고 서비스를 마친 계층 i 의 고객이 시스템을 완전히 떠날 확률이 $\tilde{r}_i := r_i \exp(-\gamma_i)/\{r_i \exp(-\gamma_i) + r_{ii} + r_{ij} \exp(-\gamma_i + \gamma_j)\}$ 이고, 동일한 계층의 고객으로 재입장할 확률은 $\tilde{r}_{ii} := r_{ii}/\{r_i \exp(-\gamma_i) + r_{ii} + r_{ij} \exp(-\gamma_i + \gamma_j)\}$ 이며, 계층 j 인 고객으로 바뀌어서 재입장할 확률이

$$\tilde{r}_{ij} := \frac{r_{ij} \exp(-\gamma_i + \gamma_j)}{r_i \exp(-\gamma_i) + r_{ii} + r_{ij} \exp(-\gamma_i + \gamma_j)}$$

인 새로운 다계층 대기행렬된다.

정리 2. 전이확률 (2.5)와 (2.6)의 새 마코프 체인 \tilde{S} 는 총 로드 $\tilde{\rho}$ 가

$$\tilde{\rho} > 1$$

인 불안정한 시스템이다.

증명: 마코프 체인 \tilde{S} 의 계층 i ($i = 1, 2$) 고객의 총 도착률 $\tilde{\Lambda}_i$ 는

$$\begin{aligned} \tilde{\Lambda}_1 &= \frac{(1 - \tilde{r}_{22})\tilde{\Lambda}_1 + \tilde{r}_{21}\tilde{\Lambda}_2}{(1 - \tilde{r}_{11})(1 - \tilde{r}_{22}) - \tilde{r}_{12}\tilde{r}_{21}} \\ \tilde{\Lambda}_2 &= \frac{\tilde{r}_{12}\tilde{\Lambda}_1 + (1 - \tilde{r}_{11})\tilde{\Lambda}_2}{(1 - \tilde{r}_{11})(1 - \tilde{r}_{22}) - \tilde{r}_{12}\tilde{r}_{21}} \end{aligned}$$

이고 따라서 총 로드 $\tilde{\rho}$ 는

$$\begin{aligned}\tilde{\rho} &:= \frac{\tilde{\Lambda}_1}{\tilde{\mu}_1} + \frac{\tilde{\Lambda}_2}{\tilde{\mu}_2} \\ &= \frac{e^{\gamma_1}(\lambda_1 r_2 e^{\gamma_1} + \lambda_1 r_{21} e^{2\gamma_1} + \lambda_2 r_{21} e^{\gamma_1 + \gamma_2})}{\mu_1(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2})} + \frac{e^{\gamma_2}(\lambda_2 r_1 e^{\gamma_2} + \lambda_2 r_{12} e^{2\gamma_2} + \lambda_1 r_{12} e^{\gamma_1 + \gamma_2})}{\mu_2(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2})}\end{aligned}$$

이다. 정리 1에 의해 $e^{\gamma_1} > 1$ 이고 $e^{\gamma_2} > 1$ 이므로

$$\begin{aligned}\tilde{\rho} &> \frac{\lambda_1 r_2 e^{\gamma_1} + \lambda_1 r_{21} e^{2\gamma_1} + \lambda_2 r_{21} e^{\gamma_1 + \gamma_2}}{\mu_1(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2})} + \frac{\lambda_2 r_1 e^{\gamma_2} + \lambda_2 r_{12} e^{2\gamma_2} + \lambda_1 r_{12} e^{\gamma_1 + \gamma_2}}{\mu_2(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2})} \quad (2.7) \\ &= \left(\frac{\lambda_1}{\mu_1}\right) \left\{ \frac{\mu_2 r_2 e^{\gamma_1} + \mu_2 r_{21} e^{2\gamma_1} + \mu_1 r_{12} e^{\gamma_1 + \gamma_2}}{\mu_2(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2})} \right\} + \left(\frac{\lambda_2}{\mu_2}\right) \left\{ \frac{\mu_1 r_1 e^{\gamma_2} + \mu_1 r_{12} e^{2\gamma_2} + \mu_2 r_{21} e^{\gamma_1 + \gamma_2}}{\mu_1(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2})} \right\} \\ &= \left(\frac{\lambda_1}{\mu_1}\right) \left\{ \frac{e^{\gamma_1} - 1}{r_1(1 - e^{-\gamma_1}) + r_{12}(1 - e^{\gamma_2 - \gamma_1})} \right\} + \left(\frac{\lambda_2}{\mu_2}\right) \left\{ \frac{e^{\gamma_2} - 1}{r_2(1 - e^{-\gamma_2}) + r_{21}(1 - e^{\gamma_1 - \gamma_2})} \right\} \\ &= 1\end{aligned}$$

이다. 여기서 식 (2.7)의 마지막 등식은 식 (2.3)과 (2.4)로부터

$$\mu_1 r_1 (1 - e^{-\gamma_1}) + \mu_1 r_{12} (1 - e^{\gamma_2 - \gamma_1}) = \mu_2 r_2 (1 - e^{-\gamma_2}) + \mu_2 r_{21} (1 - e^{\gamma_1 - \gamma_2}) \quad (2.8)$$

을 얻고, 이것을 식 (2.3)의 양변에 나누어서 얻어지는 다음의 관계식을 적용한 것이다:

$$\frac{\lambda_1 (e^{\gamma_1} - 1)}{\mu_1 r_1 (1 - e^{-\gamma_1}) + \mu_1 r_{12} (1 - e^{\gamma_2 - \gamma_1})} + \frac{\lambda_2 (e^{\gamma_2} - 1)}{\mu_2 r_2 (1 - e^{-\gamma_2}) + \mu_2 r_{21} (1 - e^{\gamma_1 - \gamma_2})} = 1.$$

한편, 식 (2.7)의 두번째 등식의 양쪽 첫 항이 동일한 것은 두 항의 차이가

$$\begin{aligned}&\frac{\mu_2 r_2 e^{\gamma_1} + \mu_2 r_{21} e^{2\gamma_1} + \mu_1 r_{12} e^{\gamma_1 + \gamma_2}}{\mu_2(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2})} - \frac{e^{\gamma_1} - 1}{r_1(1 - e^{-\gamma_1}) + r_{12}(1 - e^{\gamma_2 - \gamma_1})} \quad (2.9) \\ &= r_{12} e^{\gamma_1 + \gamma_2} [\mu_2(r_1 r_2 + r_1 r_{21} e^{\gamma_1} + r_{12} r_2 e^{\gamma_2}) \{r_1(1 - e^{-\gamma_1}) + r_{12}(1 - e^{\gamma_2 - \gamma_1})\}]^{-1} \\ &\quad \{\mu_1 r_1 (1 - e^{-\gamma_1}) + \mu_1 r_{12} (1 - e^{\gamma_2 - \gamma_1}) - \mu_2 r_2 (1 - e^{-\gamma_2}) - \mu_2 r_{21} (1 - e^{\gamma_1 - \gamma_2})\} \\ &= 0\end{aligned}$$

이기 때문이다. 여기서 식 (2.9)의 마지막 등식은 식 (2.8)에 의한 결과이다. 또한 식 (2.7)의 두번째 등식의 두번째 항들도 그 차이의 분자가

$$r_{21} e^{\gamma_1 + \gamma_2} \{\mu_2 r_2 (1 - e^{-\gamma_2}) + \mu_2 r_{21} (1 - e^{\gamma_1 - \gamma_2}) - \mu_1 r_1 (1 - e^{-\gamma_1}) - \mu_1 r_{12} (1 - e^{\gamma_2 - \gamma_1})\} = 0$$

으로 얻어져서 서로 같음을 알 수 있다. \square

2.1. 과부하 확률의 추정량

사건 H_ℓ 을 총 고객수가 0을 벗어난다고 했을 때 다시 0이 되기 전에 처음으로 충분히 큰 수 ℓ 에 먼저 도착하는 사건이라고 하자. ℓ 을 시스템이 수용할 수 있는 최대 고객수라고 하면 사건 H_ℓ 는 시스템에 과부하가 발생하는 사건으로 과부하 확률 $\theta = P(H_\ell)$ 을 추정하고자 한다. 변화된 시스템을 이용한

과부하 확률의 중요 샘플링 추정량은 Lee와 Bae (2007)와 Song 등 (2007)에 자세히 설명되어 있으며 여기서는 본 모형에 해당되는 부분만 간략히 설명하고자 한다.

T 를 고객수가 0을 벗어난다고 했을 때 다시 0이 되기 전에 처음으로 ℓ 이 될 때까지의 총 전이횟수라고 하자. 그러면 사건 H_ℓ 에 해당되는 가능한 경로(trajectory) $\omega := (\mathbf{x}(1) := (x_1), \mathbf{x}(2), \dots, \mathbf{x}(T))$ 에 대해,

$$\begin{aligned} P(\omega) &:= P\{\mathbf{X}(1) = (x_1), \mathbf{X}(2) = \mathbf{x}(2), \dots, \mathbf{X}(T) = \mathbf{x}(T) \mid \text{상태 } \mathbf{0} \text{을 벗어난다}\} \\ &= \frac{\lambda_{x_1}}{\lambda_1 + \lambda_2} \prod_{s=1}^{T-1} K(\mathbf{x}(s), \mathbf{x}(s+1)) \end{aligned}$$

가 얻어진다. 여기서 처음 상태 $\mathbf{0}$ 을 벗어날 때, 제일 먼저 도착하는 고객의 계층이 x_1 일 확률은 $\lambda_{x_1}/(\lambda_1 + \lambda_2)$ 임을 이용하였다. 위의 확률을 식 (2.5)와 (2.6)의 새로운 전이확률 \tilde{K} 를 이용하여 나타내면,

$$\begin{aligned} P(\omega) &= \frac{\lambda_{x_1}}{\lambda_1 + \lambda_2} \prod_{s=1}^{T-1} \tilde{K}(\mathbf{x}(s), \mathbf{x}(s+1)) \frac{h(\mathbf{x}(s))}{h(\mathbf{x}(s+1))} \\ &= \left(\frac{\tilde{\lambda}_1 + \tilde{\lambda}_2}{\lambda_1 + \lambda_2} \right) \frac{1}{h(\mathbf{x}(T))} \tilde{P}(\omega) \end{aligned}$$

가 된다. 여기서 $\tilde{P}(\omega)$ 는 마코프 체인 \tilde{S} 에서의 경로 ω 의 확률이다. 따라서 확률 $\theta = P(H_\ell)$ 의 중요 샘플링 추정량은 N 개의 바쁜 기간(busy period)을 생성했을 때,

$$\hat{\theta} = \frac{1}{N} \left(\frac{\tilde{\lambda}_1 + \tilde{\lambda}_2}{\lambda_1 + \lambda_2} \right) \sum_{i=1}^N 1_{H_\ell}(\tilde{\omega}_i) \exp(-\gamma_1 N_1(\mathbf{X}(T))(\tilde{\omega}_i) - \gamma_2 N_2(\mathbf{X}(T))(\tilde{\omega}_i))$$

로 얻을 수 있다. 이때, $\tilde{\omega}_i$ 는 마코프 체인 \tilde{S} 에서 얻어지는 시행의 결과이다.

2.2. 과부하 발생시 계층별 평균 고객수의 추정량

과부하 사건 H_ℓ 이 발생할 때, 계층 i 의 평균 고객수 $E[N_i(\mathbf{X}(T)) \mid H_\ell]$ 을 구하고자 한다. 우선 과부하가 발생할 때, 각 계층별 고객수에 대한 분포 $p(\ell_1, \ell_2)$ 는 $\ell = \ell_1 + \ell_2$ 인 경우,

$$\begin{aligned} p(\ell_1, \ell_2) &:= P\{\{N_1(\mathbf{X}(T)) = \ell_1, N_2(\mathbf{X}(T)) = \ell_2\} \mid H_\ell\} \\ &= \frac{P[H_\ell \cap \{N_1(\mathbf{X}(T)) = \ell_1, N_2(\mathbf{X}(T)) = \ell_2\}]}{P(H_\ell)} \end{aligned}$$

이고, 새로운 전이확률 \tilde{K} 를 이용하여 나타내면,

$$\begin{aligned} &P[H_\ell \cap \{N_1(\mathbf{X}(T)) = \ell_1, N_2(\mathbf{X}(T)) = \ell_2\}] \\ &= \left(\frac{\tilde{\lambda}_1 + \tilde{\lambda}_2}{\lambda_1 + \lambda_2} \right) \exp(-\gamma_1 \ell_1 - \gamma_2 \ell_2) \tilde{P}[H_\ell \cap \{N_1(\mathbf{X}(T)) = \ell_1, N_2(\mathbf{X}(T)) = \ell_2\}] \end{aligned}$$

가 된다. 그러므로 과부하가 발생했을 때, 각 계층별 고객수에 대한 분포 $p(\ell_1, \ell_2)$ 를 추정하기 위한 중

요 샘플링 추정량은

$$\hat{p}(\ell_1, \ell_2) = \frac{\exp(-\gamma_1 \ell_1 - \gamma_2 \ell_2) \sum_{i=1}^N 1_{H_\ell \cap (N_1(X(T))=\ell_1, N_2(X(T))=\ell_2)}(\tilde{\omega}_i)}{\sum_{i=1}^N 1_{H_\ell}(\tilde{\omega}_i) \exp[-\gamma_1 N_1(X(T))(\tilde{\omega}_i) - \gamma_2 N_2(X(T))(\tilde{\omega}_i)]}$$

로 얻을 수 있다.

따라서 과부하 사건 H_ℓ 이 발생할 때, 계층 1의 평균 고객수는

$$\hat{E}[N_1(X(T))|H_\ell] := \sum_{\ell_1=0}^{\ell} \sum_{\ell_2=0}^{\ell-\ell_1} \ell_1 \hat{p}(\ell_1, \ell_2)$$

로 추정할 수 있고, 계층 2의 평균 고객수는

$$\hat{E}[N_2(X(T))|H_\ell] := \ell - \hat{E}[N_1(X(T))|H_\ell]$$

로 얻을 수 있다.

3. 시뮬레이션

모수가 $\lambda_1 = 0.015, \lambda_2 = 0.035, \mu_1 = 0.05, \mu_2 = 0.9$ 이고 $r_{11} = 0.2, r_{12} = 0.1, r_{11} = 0.7, r_{21} = 0.1, r_{22} = 0.1, r_2 = 0.8$ 인 다계층 대기행렬을 고려한다. 이 시스템은 $\rho_1 = 0.4789$ 이고 $\rho_2 = 0.0461$ 으로 총 로드가 $\rho = 0.525$ 인 안정된 시스템이다. 이 경우에 식 (2.3)와 (2.4)의 해로서 $e^{\gamma_1} = 1.9787$ 과 $e^{\gamma_2} = 1.136$ 이 얻어져서 변화된 확률측도는 $\tilde{\lambda}_1 = 0.02968, \tilde{\lambda}_2 = 0.03976, \tilde{\mu}_1 = 0.03056, \tilde{\mu}_2 = 0.8806$ 이고 $\tilde{r}_{11} = 0.3272, \tilde{r}_{12} = 0.0939, \tilde{r}_{21} = 0.178, \tilde{r}_{22} = 0.1022$ 가 되며 이 변화된 시스템의 총 로드는 $\tilde{\rho} = 1.9412$ 가 되어 불안정한 시스템이 된다.

그림 1과 2는 과부하 확률의 로그값과 과부하 발생시 계층 1의 평균 고객수를 일반 시뮬레이션과 2장의 빠른 시뮬레이션을 적용하여 구한 것이다. 상당히 근사한 두 값을 얻기 위해 과부하 확률의 경우, 일반 시뮬레이션은 $N = 1.5 \times 10^9$ 개의 바쁜 기간을 생성한 반면에 빠른 시뮬레이션은 $N = 1.5 \times 10^5$ 개의 바쁜 기간이 이용되었다. 계층 1의 평균 고객수의 경우에는 일반 시뮬레이션은 $N = 2 \times 10^9$ 개의 바쁜 기간으로 추정되었지만 빠른 시뮬레이션은 $N = 2 \times 10^5$ 개로만 추정되었다.

과부하 확률은 총 로드 ρ 가 클수록 그 값은 증가할 것이고, 계층 1의 로드 ρ_1 이 클수록 계층 1의 평균 고객수가 많아질 것이다. 이를 확인하기 위해 앞의 시스템과 나머지 모수의 값은 동일하고 다음과 같이 도착률 λ_1 과 λ_2 만 차이가 나는 다음의 두 시스템을 살펴보자; $\lambda_1 = 0.0005, \lambda_2 = 0.0495$ 로서 $\rho_1 = 0.1521, \rho_2 = 0.0479$ 이고 총 로드가 $\rho = 0.2142$ 인 시스템과 $\lambda_1 = 0.035, \lambda_2 = 0.015$ 로 $\rho_1 = 0.9296, \rho_2 = 0.0242$ 로서 총 로드가 $\rho = 0.9538$ 인 시스템을 서로 비교하자. 각 시스템의 확률측도를 바꾸고 빠른 시뮬레이션을 적용하여 과부하 확률을 추정한 결과가 그림 3에 나와있다. 로드가 큰 시스템일수록 과부하 확률이 크게 나타남을 확인할 수 있다. 이런 비교를 일반 시뮬레이션으로 추정한다면 로드가 작은 시스템의 경우에는 엄청난 수의 바쁜 기간이 필요하겠지만 빠른 시뮬레이션을 이용하여 추정하면 $N = 10^5$ 개의 바쁜 기간을 이용하여 로드가 작은 시스템에서도 빠르게 추정하여 비교할 수 있다. 평균 고객수의 경우도 빠른 시뮬레이션을 통해 모두 $N = 10^5$ 개의 바쁜 기간으로 추정 비교하였다.

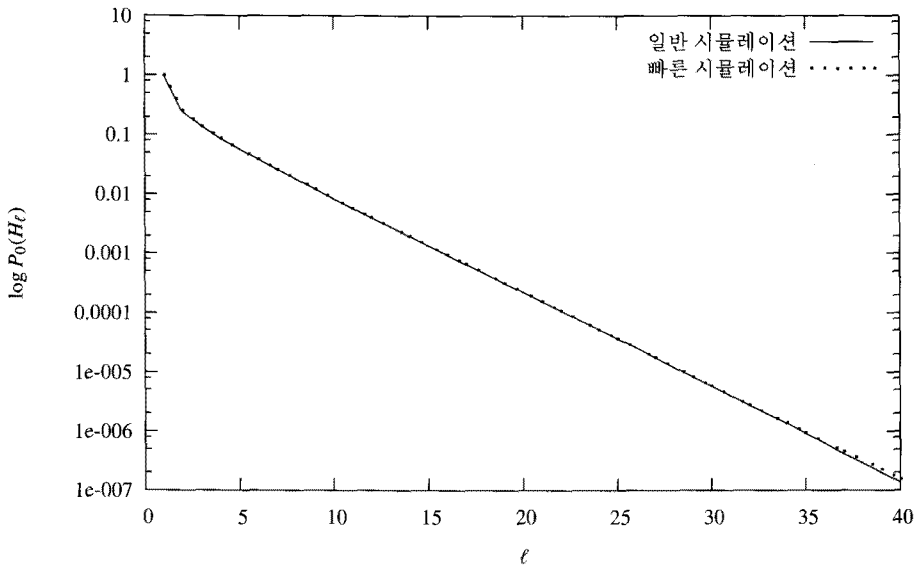


그림 1: 과부하 확률 $P_0(H_\ell)$ 의 로그값

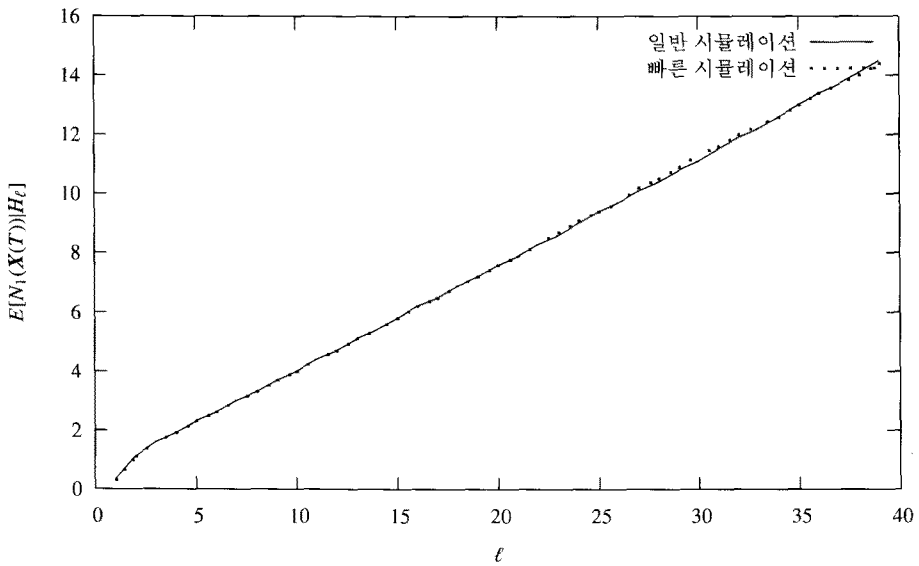


그림 2: 과부하 발생시 계층 1의 평균 고객수

4. 결론

안정된 시스템에서 과부하가 발생하는 것은 거의 일어나지 않는 희귀사건이기 때문에 일반 시뮬레이션으로 확률값을 추정하기 위해서는 엄청난 수의 바쁜 기간을 생성해야 한다. 본 논문은 계층 전이가 가능한 다계층 대기행렬 시스템에서 전체 고객의 수가 일정한 수를 초과하는 과부하가 발생할 확률

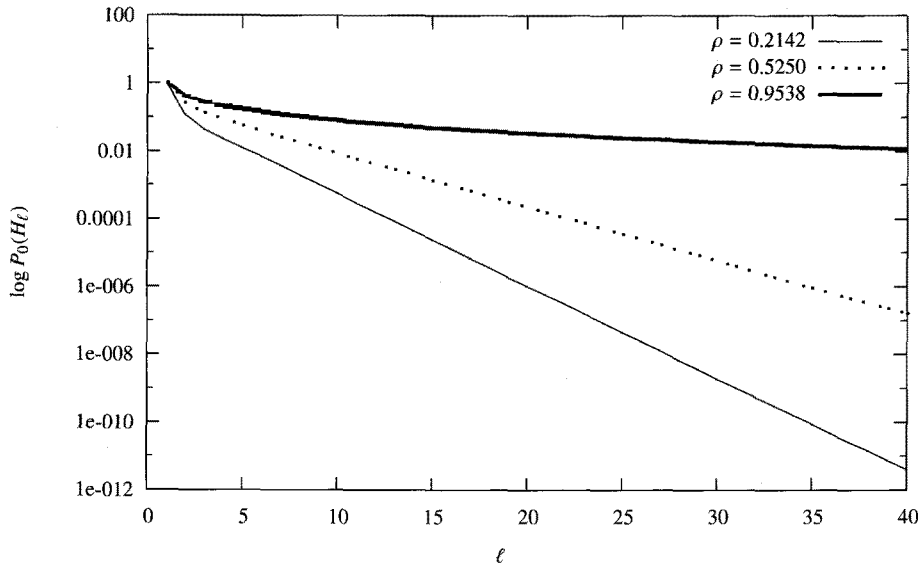


그림 3: 총 로드 ρ 에 따른 과부하 확률 $P_0(H_\ell)$ 의 로그값

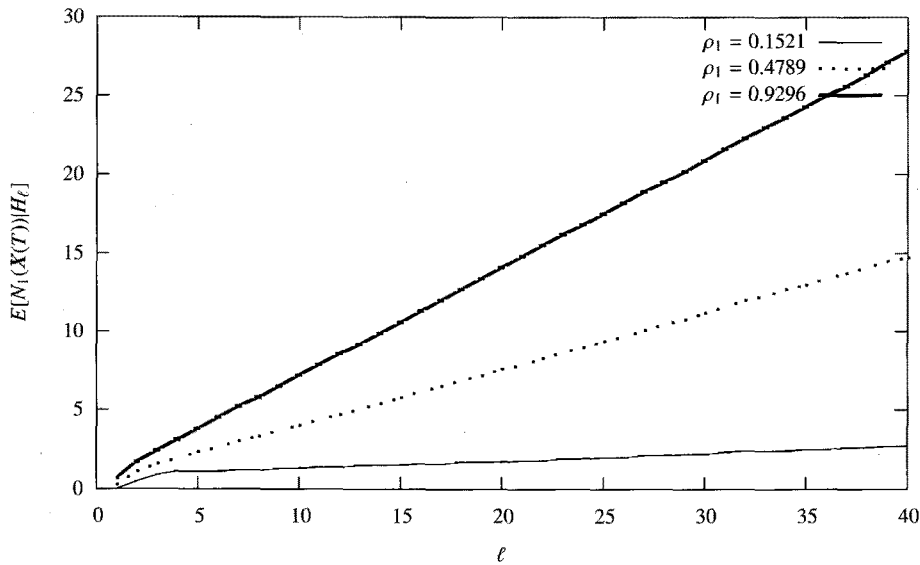


그림 4: 계층 1의 로드 ρ_1 에 따른 과부하 발생시 계층 1의 평균 고객수

과 과부하 발생시 계층별 평균 고객수를 적은 수의 바쁜 기간으로도 빠르게 추정할 수 있는 h -변환을 구하고, 이를 이용한 빠른 시뮬레이션 방법을 제시하였다.

그리고 시스템의 로드 ρ 에 따른 과부하 확률값과 계층별 평균 고객수를 비교하기 위해 빠른 시뮬레이션 방법을 적용하였으며, 로드가 클수록 과부하 확률과 계층별 평균 고객수가 크게 나타남을 확인하였다. 로드가 작은 시스템의 경우 일반 시뮬레이션은 아주 많은 수의 바쁜 기간이 필요하겠지만 빠른 시

물레이션은 로드가 작은 시스템에서도 훨씬 적은 수의 바쁜 기간을 이용하여 과부하 확률을 추정하고 계층별 평균 고객의 수를 비교할 수 있다.

참고 문헌

- Dabrowski, A., Lee, J. and McDonald, D. R. (2008). Large deviations of multi-class M/G/1 queues, *The Canadian Journal of Statistics* (to appear).
- Heidelberger, P. (1995). Fast simulation of rare events in queueing and reliability models, *ACM Transactions on Modeling and Computer Simulation*, **5**, 43–85.
- Lee, J. and Bae, K. (2007). Fast simulation of overflow probabilities in multiclass queues, *The Korean Communications in Statistics*, **14**, 287–299.
- McDonald, D. R. (1999). Asymptotics of first passage times for random walk in an orthant, *The Annals of Applied Probability*, **9**, 110–145.
- McDonald, D. R. (2004). *Elements of Applied Probability for Engineering, Mathematics and Systems Science*, World Scientific, New Jersey.
- Song, M. J., Bae, K. and Lee, J. (2007). Overflow probabilities in multi-class feedback queues, *Journal of the Korean Data & Information Science Society*, **18**, 1045–1056.
- Walrand, J. (1988). *An Introduction to Queueing Networks*, Prentice Hall, New Jersey.

2008년 12월 접수; 2009년 2월 채택

Fast Simulation of Overflow Probabilities in Multi-Class Queues with Class-Transition

Mi Jung Song^a, Kyungsoon Bae^a, Jiyeon Lee^{1,a}

^aDept. of Statistics, Yeungnam Univ.

Abstract

In this paper, we consider a multi-class queueing system in which different classes of customers have different arrival rates, service rates and class-transition probabilities. We use the fast simulation method to estimate the overflow probability and the expected number of customers of each class at the first time the total number of customers hits a high level. We also discuss the overflow probabilities and the expected number of customers at different loads, respectively.

Keywords: Multi-class queues, class-transition, fast simulation, overflow probabilities, importance sampling estimators.

This work was supported by the Korea Research Foundation Grant funded by Korea Government(MOEHRD, Basic Research Promotion Fund)(KRF-2005-204-C00016).

¹ Corresponding author: Professor, Department of Statistics, Yeungnam University, 214-1 Dae-Dong, Kyeongsan, Kyeongbuk 712-749, Korea. E-mail: leejy@yu.ac.kr