2000 Proceedings of the Autumn Conference, Korean Statistical Society

동시입력이 있는 병렬네트워크의 과부하 확률 추정

권민희[1], **이지연**[2]

요 약

동시입력이 있는 병렬 네트워크에서 총 손님의 수가 특정한 값을 초과하여 과부하가 발생하는 확률을 추정하고자 한다. large deviation 이론을 적용하여 추정을 위한 최적의 확률 측도를 찾고 이를 이용하여 과부하 확률의 중요 샘플링 추정량을 구한다.

주요용어 : 병렬 네트워크, 중요 샘플링, 과부하 확률, 시뮬레이션

1. 소개

안정된 시스템에서 과부하가 일어나는 것은 아주 드물게 발생하는 희귀 사건(rate event)이다. 이런 희귀 사건의 확률을 추정하는데, 일반적인 시뮬레이션 방법을 이용하면 시간과 비용이 많이 들게 된다. 이런 경우에 확률 측도를 적당히 변화시켜 시뮬레이션의 속도를 가속화함으로써 희귀 사건의 확률을 빠르게 추정할 수 있다.

가속 시뮬레이션을 사용하기 위해서는 먼저 변화시킬 최적의 확률 측도를 찾아야 한다. Parekh and Walrand(1989)와 Frater et al.(1991)는 잭슨 네트워크(Jackson network)에서 large deviation 이론(Varadran(1984))을 이용한 탐색적인 방법으로, 변화된 확률 측도를 찾는 방법을 제안하였다. Frater and Anderson(1994)은 이 방법을 적용하여 일반적인 서비스분포를 갖는 직렬 네트워크(tandem network)에서 과부하 확률을 추정하였다. 이처럼 large deviation 이론을 이용하여 최적의 확률 측도로 변화시키는 것은 직렬 네트워크나 잭슨 네트워크에서와 같이 제한된 모형에서만 가능하였다. 하지만 본 논문에서 다루는 동시 입력이 있는 병렬 네트워크는 비잭슨 네트워크(non-Jackson network)로서 그 적용범위를 확장하였다.

2장에서는 동시 입력이 있는 병렬 네트워크 모형에 대해 소개하고, 최적의 변화된 확률 측도를 구한다. 그것을 이용하여 과부하 확률의 중요 샘플링 추정량(importance sampling estimator)을 구하고, 마지막으로 시뮬레이션을 통해 일반 시뮬레이션과 가속 시뮬레이션을 비교할 것이다.

2. 동시 입력이 있는 병렬 네트워크

동시 입력이 있는 병렬 네트워크 모형은 노드에 도착하는 손님의 도착율이 각각 λ와 η인 포아송 분포이며, 도착한 손님은 각각 서비스율이 α와 β인 지수 분포를 따르는 서버에 의해 서비스를 받게 된다. 그리고 동시에 두 명의 손님이 도착율 ν인 포아송 분포를 따르면서 도착하여 두 개의 서버에 의해 각각 서비스를 받는다. 본 논문에서는 $\lambda+\nu+\eta+\alpha+\beta=1$로 가정하고 이 모형을 다음 그림과 같이 $(\lambda,\nu,\eta,\alpha,\beta)$로 나타낸다. $(\lambda,\nu,\eta,\alpha,\beta)$가 안정(stable)되기 위

1) 경북 경산시 대동 214-1 영남대학교 통계학과 석사과정
2) 경북 경산시 대동 214-1 영남대학교 통계학과 조교수

해서는 각 노드의 부하량 $\frac{\lambda+\nu}{\alpha}$와 $\frac{\eta+\nu}{\beta}$가 1보다 작아야 한다. (Flatto(1985)와 Wright(1992))

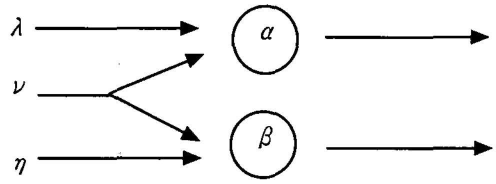

그림. $(\lambda, \nu, \eta, \alpha, \beta)$ - 동시 입력이 있는 병렬 네트워크.

본 논문에서는 네트워크가 비어있는 상태에서 시작하여 다시 빈 상태가 되기 전에 두 노드에 있는 손님수의 합이 큰 수 N을 초과할 확률 α값을 추정하고자 한다. 먼저 최적의 변화된 확률 측도를 구한다.

2.1 최적의 변화된 확률 측도

손님이 도착하거나 서비스를 받고 시스템을 빠져나가는 것을 하나의 사건이라고 하면, 사건이 일어나는 시점에 각 노드에 있는 손님 수 $\{(X_1(n),\ X_2(n))\ |\ n=0,1,2\cdots\}\in R^2$ 는 이산 마코프 체인(Markov chain)이 된다. 여기서 시스템에서 과부하가 일어났을 때의 도착율을 각각 λ', η'이라고 하고, 서비스율을 α',β' 그리고 동시에 입장하는 손님의 도착율은 ν'이라고 하자. 이때 $\lambda'+\nu'+\eta'+\alpha'+\beta'=1$이다. 여기서 시스템의 과부하 확률 α는 다음과 같이 구할 수 있다. (Parekh and Walrand(1989))

$$\alpha\approx\ \exp\Big\{-N\cdot inf\Big[R(\lambda',\nu',\eta',\alpha',\beta')\Big(\lambda'\cdot h_{\lambda}\Big(\frac{1}{\lambda'}\Big)+\nu'\cdot h_{\nu}\Big(\frac{1}{\nu'}\Big)+\eta'\cdot h_{\eta}\Big(\frac{1}{\eta'}\Big)+\alpha'\cdot h_{\alpha}\Big(\frac{1}{\alpha'}\Big)+\beta'\cdot h_{\beta}\Big(\frac{1}{\beta'}\Big)\Big)\Big]\Big\} \tag{2.1}$$

여기에서 $R(\lambda',\nu',\eta',\alpha',\beta')$은

$$R=\begin{cases}\dfrac{1}{\lambda'+\nu'-\alpha'}, & \lambda'+\nu'>\alpha' \text{ 그리고 } \eta'+\nu'<\beta' \text{일 때}\\ \dfrac{1}{\eta'+\nu'-\beta'}, & \lambda'+\nu'<\alpha' \text{ 그리고 } \eta'+\nu'>\beta' \text{일 때}\\ \dfrac{1}{\lambda'+2\nu'+\eta'-\alpha'-\beta'}, & \lambda'+\nu'>\alpha' \text{ 그리고 } \eta'+\nu'>\beta' \text{일 때}\end{cases}$$

이고, h_ξ는 평균 $1/\xi$ 인 지수확률변수의 Cramér 변환으로서

$$h_\xi(u)=\begin{cases}\xi u-1-\log(\xi u), & u>0\\ 0, & \text{그외 } u\end{cases}$$

로 얻어진다. 식 (2.1)에서

$$H:=R\Big(\lambda'\cdot h_{\lambda}(\frac{1}{\lambda'})+\nu'\cdot h_{\nu}(\frac{1}{\nu'})+\eta'\cdot h_{\eta}(\frac{1}{\eta'})+\alpha'\cdot h_{\alpha}(\frac{1}{\alpha'})+\beta'\cdot h_{\beta}(\frac{1}{\beta'})\Big)$$

라 두면, 함수 H가 최소가 되는 모수 λ', ν', η', α', β'가 과부하 현상을 잘 반영하는 최적의 확률 측도가 된다.

(i) $\lambda'+\nu'>\alpha'$ 그리고 $\eta'+\nu'<\beta'$일 때

함수 H는 $\lambda'=\frac{\lambda\alpha}{\lambda+\nu}$, $\nu'=\frac{\nu\alpha}{\lambda+\nu}$, $\eta'=\eta$, $\alpha'=\lambda+\nu$, $\beta'=\beta$에서 최소값

$\log(\frac{\alpha}{(\lambda+\nu)})$를 가진다. 이 모수들이 위의 범위를 만족하기 위해서는 $\eta+\frac{\nu\alpha}{\lambda+\nu}<\beta$이 성립해야 한다. 즉, $\eta+\frac{\nu\alpha}{\lambda+\nu}<\beta$의 경우에는 첫 번째 노드에서만 과부하가 일어나고 두 번째 노드는 안정상태가 된다.

(ii) $\lambda'+\nu'<\alpha'$ 그리고 $\eta'+\nu'>\beta'$일 때

함수 H가 $\lambda'=\lambda$, $\nu'=\frac{\nu\beta}{\eta+\nu}$, $\eta'=\frac{\eta\beta}{\eta+\nu}$, $\alpha'=\alpha$, $\beta'=\eta+\nu$ 에서 최소값 $\log(\frac{\beta}{(\eta+\nu)})$을 갖기 위해서는 $\lambda+\frac{\nu\beta}{\eta+\nu}<\alpha$ 이 성립해야 한다. 이 경우는 두 번째 노드만 과부하가 일어나게 된다.

(iii) $\lambda'+\nu'>\alpha'$ 그리고 $\eta'+\nu'>\beta'$일 때

먼저 λ'을 구하기 위해 함수 H를 λ'에 대하여 미분하면

$$\nu\lambda'^3+\lambda(\lambda+\eta)\lambda'^2-\lambda^2\lambda'+(\alpha+\beta)\lambda^3=0 \tag{2.2}$$

을 얻고, 위 식을 풀면

$\lambda'=\lambda\cdot\frac{-(\lambda+\nu+\eta)+\sqrt{(\lambda+\nu+\eta)^2+4\nu(\alpha+\beta)}}{2\nu}:=\gamma$로서 $\gamma>\lambda$이 만족한다. 나머지 모수는 $\nu'=\left(\frac{\gamma}{\lambda}\right)^2\nu$, $\eta'=\frac{\gamma}{\lambda}\eta$, $\alpha'=\frac{\lambda}{\gamma}\alpha$, $\beta'=\frac{\lambda}{\gamma}\beta$이 된다. 구한 모수가 조건 $\lambda'+\nu'>\alpha'$ 그리고 $\eta'+\nu'>\beta'$을 만족하기 위해서

$$\lambda^2\gamma^2+\nu\gamma^3-\lambda^3\alpha>0 \tag{2.3}$$

$$\lambda\eta\gamma^2+\nu\gamma^3-\lambda^3\beta>0 \tag{2.4}$$

이 성립하여야 한다.

식 (2.3)의 좌변은 식 (2.2)와 조건 $\lambda+\nu+\eta+\alpha+\beta=1$에 의해

$$\lambda^2\gamma^2+\nu\gamma^3-\lambda^3\alpha=\lambda[\alpha\lambda(\gamma-\lambda)+\gamma\{\lambda\beta-(\eta+\nu)\gamma\}]$$

이 된다. 따라서 $\frac{\lambda}{\gamma}>\frac{\eta+\nu}{\beta}$이면 $\lambda'+\nu'>\alpha'$가 성립한다.

식 (2.4)도 마찬가지 방법으로 $\frac{\lambda}{\gamma}>\frac{\lambda+\nu}{\alpha}$이면 $\eta'+\nu'>\beta'$가 성립한다. 이 경우는 한쪽의 노드에서 과부하가 일어나면 다른 한쪽도 과부하가 일어나게 되고, 최소값은 $\log\left(\frac{\gamma}{\lambda}\right)$이다.

H의 최소값을 찾기 위해서 각 범위에서 구한 결과를 이용하면 다음과 같은 4개의 조건으로 정리할 수 있다.

a) $\frac{\alpha}{\lambda+\nu}<\frac{\beta}{\eta+\frac{\alpha}{\lambda+\nu}\nu}<\frac{\beta}{\eta+\nu}$

b) $\frac{\beta}{\eta+\frac{\alpha}{\lambda+\nu}\nu}<\frac{\alpha}{\lambda+\nu}<\frac{\beta}{\eta+\nu}$

c) $\frac{\beta}{\eta+\nu}<\frac{\alpha}{\lambda+\frac{\beta}{\eta+\nu}\nu}<\frac{\alpha}{\lambda+\nu}$

d) $\dfrac{\alpha}{\lambda+\dfrac{\beta}{\eta+\nu}\nu} < \dfrac{\beta}{\eta+\nu} < \dfrac{\alpha}{\lambda+\nu}$

여기서 함수 H는 (a)의 경우에, 최소값 $\log\left(\dfrac{\alpha}{\lambda+\nu}\right)$를 갖고 이 때의 변화된 확률 측도의 모수는 (i)에서 구한 것과 같다. (c)의 경우에는 H의 최소값이 $\log\left(\dfrac{\beta}{\eta+\nu}\right)$가 되어 (ii)에서 구한 모수가 최적의 변화 확률 측도이다. 그리고 (b)와 (d)의 경우에는 $\log\left(\dfrac{\gamma}{\lambda}\right)$가 최소값이 되어 (iii)에서 구한 모수가 최적의 변화 확률 측도이다.

2.2 과부하 확률의 중요샘플링 추정량

우도비(likelihood ratio)를 구하기 위해서 먼저 다음과 같은 변수를 정의한다.

N_1 : 과부하가 발생했을 때 첫 번째 노드에 있는 손님의 수

L_i : i 번째 노드에 손님이 있고 나머지 노드는 비어 있게 되는 회수 $(i=1,2)$

기존의 확률 측도를 P라 하고 변화된 확률 측도를 P^*라고 하면 이 때의 우도비 L (Likelihood ratio)은

$$L = \frac{dP}{dP^*}$$

$$= \begin{cases} \dfrac{\alpha+\eta}{\lambda+\nu+\eta}\cdot\left(\dfrac{\lambda+\nu}{\alpha}\right)^{N_1}\left(\dfrac{\alpha+\eta+\beta}{\lambda+\nu+\eta+\beta}\right)^{L_2} & (a)\text{일때} \\ \dfrac{\beta+\lambda}{\lambda+\nu+\eta}\cdot\left(\dfrac{\eta+\nu}{\beta}\right)^{N-N_1}\left(\dfrac{\alpha+\lambda+\beta}{\lambda+\nu+\eta+\alpha}\right)^{L_1} & (c)\text{일때} \\ \left(\dfrac{\lambda}{\gamma}\right)^{N}\cdot\left(\dfrac{\gamma+\left(\frac{\gamma}{\lambda}\right)^2\nu+\frac{\gamma}{\lambda}\eta}{\lambda+\nu+\eta}\right)\cdot\left(\dfrac{\gamma+\left(\frac{\gamma}{\lambda}\right)^2\nu+\frac{\gamma}{\lambda}\eta+\frac{\lambda}{\gamma}\alpha}{\lambda+\nu+\eta+\alpha}\right)^{L_1} & \\ \quad\cdot\left(\dfrac{\gamma+\left(\frac{\gamma}{\lambda}\right)^2\nu+\frac{\gamma}{\lambda}\eta+\frac{\lambda}{\gamma}\beta}{\lambda+\nu+\eta+\beta}\right)^{L_2} & (b),(d)\text{일때} \end{cases}$$

이다. 따라서 과부하 확률은

$$\alpha^* := \frac{L_1\cdot V_1+L_2\cdot V_2+\cdots+L_n\cdot V_n}{n}$$

으로 구할 수 있다. 이 때 확률변수 V_i는 변화된 확률 측도하에서

$$V_i = \begin{cases} 1, & \text{비어있는 상태에서 시작하여, 다시 비기 전에 손님수의 합이 } N\text{이 먼저 되면} \\ 0, & \text{그 외} \end{cases}$$

로 정의된다.

다음 표는 몇 가지 모수에 대한 일반 시뮬레이션과 가속 시뮬레이션을 비교한 것이다. $\lambda=0.24$, $\nu=0.01$, $\eta=0.01$, $\alpha=0.5$, $\beta=0.24$인 모형에서 $N=12$일 때 정확한 과부하 확률은 $\alpha=3.269\times10^{-4}$이다. 일반 시뮬레이션의 결과를 보면, $n=5000$일 때 $\alpha=3.100\times10^{-4}$이고, 가속화한 시뮬레이션을 결과를 보면 $n=500$일 때 $\alpha=3.216\times10^{-4}$가 되어 훨씬 더 적은 시뮬레이션회수로 더 정확한 과부하 확률을 추정할 수 있음을 알 수 있다.

표 I. 동시입력이 있는 병렬 네트워크의 시뮬레이션

방법	일반 시뮬레이션		가속 시뮬레이션	
Example I (0.24, 0.01, 0.01, 0.5, 0.24) $N=12$ $\alpha=3.269\times10^{-4}$				
n	2000	5000	100	500
$\alpha_n(\alpha_n^*)$	3.000×10^{-4}	3.100×10^{-4}	3.192×10^{-4}	3.216×10^{-4}
Example II (0.01, 0.1, 0.05, 0.59, 0.25) $N=20$ $\alpha=6.608\times10^{-5}$				
n	10000	20000	300	500
$\alpha_n(\alpha_n^*)$	0.0	0.0	6.6336×10^{-5}	6.711×10^{-5}
Example III (0.18, 0.02, 0.05, 0.35, 0.4) $N=23$ $\alpha=2.947\times10^{-6}$				
n	50000	70000	500	700
$\alpha_n(\alpha_n^*)$	0.0	0.0	2.894×10^{-6}	2.9206×10^{-6}

3. 결과

본 논문에서는 동시 입력이 있는 병렬 네트워크의 과부하 확률을 추정하기 위해 가속 시뮬레이션을 사용하였다. 일반적인 병렬 네트워크에서의 과부하 확률 추정은 부하량이 가장 큰 노드의 도착율과 서비스율이 서로 바뀐 것이 최적의 변화된 확률 측도가 되는데(Parekh and Walrand(1989)), 동시 입력이 있는 병렬 네트워크에서는 동시에 입장하는 손님이 두 노드에 동시에 영향을 미치므로 일반적인 병렬 네트워크처럼 두 노드의 부하량만으로 어느 쪽 노드가 먼저 과부하가 일어나는 지를 결정할 수 없다. 또한 두 노드의 부하량의 차이가 아주 작은 경우 즉, 모수가 $\dfrac{\alpha}{\lambda+\dfrac{\beta}{\eta+\nu}\nu} < \dfrac{\beta}{\eta+\nu} < \dfrac{\alpha}{\lambda+\nu}$ 와 $\dfrac{\beta}{\eta+\dfrac{\alpha}{\lambda+\nu}\nu} < \dfrac{\alpha}{\lambda+\nu} < \dfrac{\beta}{\eta+\nu}$ 의 구간에 있을 때는 large deviation이론을 이용한 중요 샘플링 방법이 적합하지 않음을 보였다. 본 논문에서는 중요 샘플링 방법을 유용하게 적용할 수 있는 구간을 설명하였고, 그 구간 안에서는 시스템의 과부하 확률을 일반적인 시뮬레이션보다 가속 시뮬레이션을 사용하여 더 빠르고, 더 효율적으로 추정할 수 있음을 보였다. 그러나, 두 노드의 부하량이 비슷하여 중요샘플링 방법을 적용할 수 없을 때의 과부하 확률을 추정하기 위한 방법은 아직 남아있는 과제로 앞으로의 연구 방향이 될 것이다.

참고문헌

Flatto, L., and Hahn, S. (1984), Two parallel queues created by arrivals with two demands I, *SIAM J. Appl. Math,* Vol. 44., 1041-1053.

Flatto, L. (1985), Two parallel queues created by arrivals with two demands II *SIAM J, Appl. Math.* Vol 45., 861-878.

Frater, M. R., Lennon, T. M. and Anderson, B. D. O. (1991), Optimally Efficient Estimation of the Statistics of Rare Events in Queueing Networks, *IEEE*

Transactions on Automatic control, Vol. 36., 1395-1405.

Frater, M. R. and Anderson, B. D. O. (1994), Fast simulaton of buffer overflows in tandem networks of GI/GI/1 queues, *Annals of Operations Research,* Vol. 49., 207-220.

Parekh, S. and Walrand, J. (1989), A quick simulation method for excessive backlog in networks of queues, *IEEE Transactions on Automatic Control,* Vol. 34., 54-66.

Wright, P. E. (1992), Two parallel processors with coupled inputs, *Advances in Applied Probability,* Vol. 24., 986-1007.

Varadhan, S. R. S. (1984), Large deviations and applications, *SIAM*, Philadelphia, PA.