

Application of Control Variable with Routing Probability to Queueing Network Simulation

Chi-Myung Kwon^{1†} · Sanggyu Lim²

대기행렬 네트워크 시뮬레이션에서 분지확률 통제변수의 응용

권치명 · 임상규

ABSTRACT

This research discusses the application of the control variables to achieve a more precise estimation for the target response in queueing network simulation. The efficiency of control variable method in estimating the response depends upon how we choose a set of control variables strongly correlated with the response and how we construct a function of selected control variables. For a class of queueing network simulations, the random variables that drive the simulation are basically the service-time and routing probability random variables. Most of applications of control variable method focus on utilization of the service time random variables for constructing a controlled estimator. This research attempts to suggest a controlled estimator which uses these two kinds of random variables and explore the efficiency of these estimators in estimating the responses for computer network system. Simulation experiments on this model show the promising results for application of routing probability control variables. We consider the applications of the routing probability control variables to various simulation models and combined control variables using information of service time and routing probability together in constructing a control variable as future researches.

Key words : Queueing Network Simulation, Control Variable Method, Routing Probability Control Variable, Combined Control Variable

요약

본 연구는 대기행렬 네트워크 시뮬레이션에서 통제변수를 활용하여 목표 반응변수를 보다 더 정확히 추정하는 기법을 탐색한다. 반응변수 추정에서 통제변수기법의 효율성은 반응변수와 높은 상관관계를 가지는 통제변수의 선택과 선택된 통제변수를 이용하여 통제추정량을 어떻게 정의하는가에 따라 달라진다. 대기행렬 네트워크 시뮬레이션 모형에서 확률적 모형의 발전과정은 확률적 서비스시간과 분지 확률에 의하여 재현된다. 대부분의 통제변수기법은 통제추정치 구성에서 서비스 시간 확률변수를 사용한다. 본 연구는 서비스 시간 확률변수와 분지 확률변수를 동시에 사용하는 통제 추정량을 제안하고 이를 컴퓨터 네트워크 시스템의 관심 반응변수 추정에 응용하여 그 효율성을 탐색하고자 한다. 시뮬레이션 결과는 반응변수 추정에 있어서 분지확률 통제변수의 활용 가능성을 제시하고 있으며, 서비스-시간과 분지확률을 동시에 이용하는 결합 통제변수의 활용은 향후 연구가 필요한 분야로 판단된다.

주요어 : 대기행렬 네트워크 시뮬레이션, 통제변수기법, 분지확률 통제변수, 결합 통제변수

*This research was supported by research funds from Dong-A University.

접수일(2012년 7월 3일), 심사일(1차 : 2012년 9월 25일), 게재 확정일(2012년 9월 27일)

¹⁾ Dept. of MIS, Dong-A University

²⁾ Dept. of ISE, Gyeongsang National University

주 저 자 : 권치명

교신저자 : 임상규

E-mail; sglim@gnu.ac.kr

1. 서론

Although simulation is frequently the only feasible method for estimating the steady state parameters of a complex queueing system, the resultant computing cost can be serious disadvantage. This research discusses the application of control variables to achieve a more

precise estimation for the target response in queueing networks. Such networks are commonly used to model the contention for resources in job shop type systems. This research is specifically interested in the use of control variables for estimating the response of interest in queueing network system. A control variable is a random variable whose expectation is known and correlated with a statistical estimator for response. In simulation experiments, the realizations of stochastic components of simulation model are obtained from known distributions and some of their outputs may have strong correlations with response. Such variables could be candidates for good control variables (Law, 2007). For instance, suppose we try to estimate the mean sojourn time of customer in the queueing system and service times of server have a strong correlation with sojourn time of customer in this system. Then the service time random variable could be a good control variable for estimating the sojourn time of customer.

In the application of control variable method, a controlled (or new) estimator for the response is usually formed as a direct (or old) estimator plus linear function of control variables. The method of control variable exploits intrinsic correlations between target response and selected auxiliary outputs (control variables), and by using this relationship, reduces the variance of estimator. The efficiency of control variable in estimating the response depends upon how we choose a set of control variables strongly correlated with the response and how we construct a function of selected control variables. For a class of queueing network simulations, the random variables that drive the simulation are basically the service time random variables and routing random variables. The service-time control variable is quite successful in estimating the mean response of system, and under a certain condition, the control variable combined with other variance reduction methods such as antithetic variable and common random variable is very efficient in estimating the response of interest (Kwon, 2005 and Kwon and Tew, 1994a). On the other hand, the routing-probability control variable is rarely applied since the stationary probabilities of stations in general queueing network are not easily obtained. Due

to this reason, most applications of control variable method focus on utilization of the service time random variables for constructing a controlled estimator. This research attempts to suggest a controlled estimator which uses these two kinds of random variables simultaneously and explore the efficiency of this estimator in estimating the responses of interest for computer network systems.

2. Method of Control Variable

Consider a simulation run where we try to estimate a mean response μ of system. We let y be an unbiased estimator for μ , generated from simulation run. A random variable $c(i)$, also generated from simulation run, is the i th control variable if $E[c(i)] = \mu_{c(i)}$ is known and it is strongly correlated with y . We let C be a column vector of q such control variables with elements $c(i)(i = 1, \dots, q)$. Also we let μ_c be the expectation of C . Then for the q -dimensional column vector of constant coefficients, α , the controlled estimator of μ , $y(\alpha)$ given in following equation is an unbiased estimator of μ :

$$y(\alpha) = y - \alpha'(C - \mu_c) \quad (1)$$

The vector α which minimizes the variance of $y(\alpha)$ is given by

$$\alpha' = \sigma_{yc} \Sigma_c^{-1}, \quad (2)$$

where $\sigma_{yc} = Cov(y, C)$ is a row vector of covariance between y and C , and Σ_c is the covariance matrix of C . The resulting minimum variance of controlled estimator of $y(\alpha)$ is given by

$$Var[y(\alpha)] = (1 - R_{yc}^2) \sigma_y^2, \quad (3)$$

where $\sigma_y^2 = Var(y)$ and $R_{yc}^2 = \sigma_{yc} \Sigma_c^{-1} \sigma_{yc}' / \sigma_y^2$ is the square of the multiple correlation coefficient between y and C (Borogobac and Vakili, 2008). The quantity $(1 - R_{yc}^2)$ is called as the minimum variance ratio. It is the factor by which the variance of y could be reduced if the optimal coefficient vector of α is known. Hence it is the

theoretical potential to be achieved by the control variables. In general since the optimal value of α is unknown, we have to estimate it in practice.

At given input parameters of system (or design point), if a simulation run is sufficiently long, then the observed performance of y and control variables C from such a run could be considered as observations from normal distribution, respectively. We suppose that we have n times independent such runs and let y_j be the value of estimator y and $C_i = (c(1)_i, c(2)_i, \dots, c(q)_i)'$ be the output of control variables from the i th independent run. Then y_i and C_i are observations from the $(q+1)$ -variate normal distribution given as follows:

$$\begin{pmatrix} y \\ C \end{pmatrix} \sim N_{(q+1)} \left[\begin{pmatrix} \mu \\ \mu_c \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{yc} \\ \sigma_{yc} & \Sigma_c \end{pmatrix} \right] \quad (4)$$

(Kwon and Tew, 1994b). Thus given C , the conditional distribution of y is normal with expectation

$$E[y|C] = \mu + \alpha'(C - \mu_c) \quad (5)$$

and variance

$$Var(y|C) = (1 - R_{yc}^2) \sigma_y^2 \quad (6)$$

which is the minimum variance given in (3). (Anderson, 1984). Hence conditional on C_i ($i=1, \dots, n$), we have classical regression problem with

$$y_i = \mu + \alpha'(C_i - \mu_c) + \epsilon_i, \quad i=1, \dots, n, \quad (7)$$

where ϵ_i is an independent and normally distributed random variable with mean zero and variance given by (6). The least squares estimators $\hat{\mu}$ and $\hat{\alpha}$ in (7) are given by respectively,

$$\hat{\mu} = \bar{y} - \hat{\alpha}'(\bar{C} - \mu_c) \quad \text{and} \quad \hat{\alpha} = S_c^{-1} S_{yc} \quad (8)$$

where $\bar{y} = \sum_{j=1}^n y_j/n$ is the sample mean of response;

$\bar{C} = \sum_{j=1}^n C_j/n$ is the q -dimensional column vector of control variables whose k th element is given by

$\bar{c}(k) = \sum_{j=1}^n c(k)_j/n$; and S_c and S_{yc} are the sample estimators of Σ_c and σ_{yc} in (4), respectively. Then the variance of $\hat{\mu}$ is given by

$$Var(\hat{\mu}) = \left(\frac{n-2}{n-q-2} \right) (1 - R_{yc}^2) \frac{\sigma_y^2}{n} \quad (9)$$

(Szechtman, 2003). The term $(n-2)/(n-q-2)$ in (9) is the loss factor due to the estimation of α . Thus the effect of control variables is measured by the product of the loss factor and the minimum variance ratio $(1 - R_{yc}^2)$. Due to the trade-off relationship between the loss factor and the multiple correlation coefficient, it is important to select the effective set of control variables not too large.

3. Control Variable with Routing Probability

Queueing networks are often used to model interactive computer systems. For certain networks with exponential service times, steady state behavior can be obtained analytically. However, networks containing such features of real system as scheduling based on priorities, blocking due to capacity limitation, and non-exponential service times are not analytically tractable, in general. Hence, their behaviors must be estimated through simulation. Simulation needs lots of efforts to obtain estimators with high precision. Thus, it is an issue how we can conduct simulation efficiently. For this end, we focus on developing control variables for queueing networks which allow priorities, blocking and arbitrary service time distributions.

For a q -station queueing network, upon arrival at station k , an entity waits for service according to queueing discipline prescribed for that station. At the instant the entity engages a server at station k , the required service time is sampled from service time process $\{s(k)_j : j \geq 1\}$ consisting of independently and identically distributed random variables with mean $\mu_{c(k)}$ and variance σ_k^2 . On the completion of this service time, entity's next going station is determined by independently sampling from the transition probability distribution specified at station k . The considered network

consists of q service stations which are a single server or multiple servers. It is assumed that the K entities circulate service stations infinitely in a following fashion:

- (a) The sequence of service stations visited by an entity is an irreducible Markov chain with state space k which is a subset of $\{1, 2, \dots, q\}$.
- (b) The sequence of service times at the k th service station consists of independently and identically distributed random variables which have finite mean and variance.
- (c) The sequence of service stations visited and the service time sequences are mutually independent random sequences.

The class of queueing network we defined is quite general. We have no assumptions on queueing disciplines or about capacity of service stations.

We next define a set of control variables for this class of queueing networks. The basic random variables which drive the simulation of such network are the service time random variables and the multinomial routing random variables. The latter random variables control the flow of entities in the network and their probability distributions can be given by the Markov transition matrix R . We let a row vector of $\pi = (\pi(1), \dots, \pi(q))$ denote the stationary probability vector for entities. Then the vector π is the unique solution to the equation $\pi R = \pi$ and $\sum_{i=1}^q \pi(i) = 1$. To construct control variables, we need to find functions of these random variables whose expectation are known and which are correlated with estimators for system performance of interest. Intuitively performances of either system sojourn time or average waiting for service are correlated with service times of servers and routing probabilities for entity's flow. So we try to define sets of control variables which characterize these quantities. Over the simulation time period $[0, t]$, we define the following counting variables for the system:

$$f(k, t) = \text{number of service times finished at station}$$

$$k \text{ during } [0, t] \tag{10}$$

$$f(t) = \sum_{k=1}^q f(k, t) \text{ and } s(k, t) = \sum_{j=1}^{f(k,t)} s(k)_j \tag{11}$$

As the simulation time t (or number of service times at service station) increases to infinity, the following asymptotic expectations exist and are known as follow:

$$\lim_{t \rightarrow \infty} E[s(k, t) / f(k, t)] = \mu_{c(k)} \tag{12}$$

$$\lim_{t \rightarrow \infty} E[f(k, t) / f(t)] = \pi(k) \tag{13}$$

$$\lim_{t \rightarrow \infty} E[s(k, t) / f(t)] = \pi(k) \mu_{c(k)} \tag{14}$$

(Lavenberg, Moeller and Welch, 1982).

We first consider the control variables of service-time (CS) at q stations in queueing networks: $CS = (cs(1), \dots, cs(q))'$ with $cs(k) = s(k, t) / f(k, t)$ which is the sample mean of service time at station k obtained from simulation run over time period $[0, t]$. Then we can construct the controlled estimator of μ as follows:

$$y_s = y - \alpha_s' (cs(1) - \mu_{c(1)}, \dots, cs(q) - \mu_{c(q)}) \tag{15}$$

If these control variables are highly correlated with mean response y , it would be a better unbiased estimator for μ . Also we can expect that the flow estimator of routing probability for service station k , $cf(k) = f(k, t) / f(t)$ may have not a small correlation with mean response of y . Since its mean is known as $\pi(k)$, similarly to equation (15), we can construct a routing probability controlled estimator of μ as follows:

$$y_f(\alpha) = y - \alpha_f' (cf(1) - \pi(1), \dots, cf(q) - \pi(q)) \tag{16}$$

Next we develop another new control variable which is the multiplied variable of $cs(k)$ and $cf(k)$. That is,

$$cm(k) = [s(k, t) / f(k, t)] [f(k, t) / f(t)] = s(k, t) / f(t) \tag{17}$$

We expect that $cm(k)$ is also to be correlated with estimator for response and its expectation is known to be the $\pi(k) \mu_{c(k)}$ as t goes to infinity (see equation (14)). Since this type of combined control variables $cm(k) =$

$s(k,t)/f(t)$ ($k=1, \dots, q$) reflect both service time and flow effects together, they appear to be attractive control variables. For a general case of q control variables, we define the unbiased controlled estimator for μ as follows:

$$\begin{aligned}
 ym(\alpha) &= y - \alpha_m' (cm(1) - \pi(1)\mu_{c(1)}, \dots, \\
 &cm(q) - \pi(q)\mu_{c(q)})
 \end{aligned}
 \tag{18}$$

In applying control variables, we have to estimate the unknown parameter vector α_s , α_f and α_m in (15), (16) and (18), respectively. With respect to the obtained information in the course of simulation, we consider the 3 control variables of service-time, routing probability of station, and combined these two information simultaneously. In next section, we explore the efficiencies of these control variables through simulation experiment on computer network model.

4. Simulation Experiment and its Result

We apply the control variable method which utilizes the information of service times and routing probabilities in construction of control variables to computer network model. We conduct a set of simulation experiments on this model to evaluate the performance of control variable method and offer a summary and results.

4.1 Computer network model

Figure 1 shows a model of the multi-programmed computer system. With goal similar to that of time-sharing, a multi-programmed computer system attempts to utilize its resources more efficiently than does a batch processing operation. In particular a number of jobs are permitted simultaneous access to the resources of the system in such a way that the CPU is busy processing one job while various input-output peripheral units are processing some of the others concurrently. In this model, it is of interest to study the utilizations of stations, sojourn time and throughput of system. We consider the central server model that permits the inclusion a number of peripheral devices.

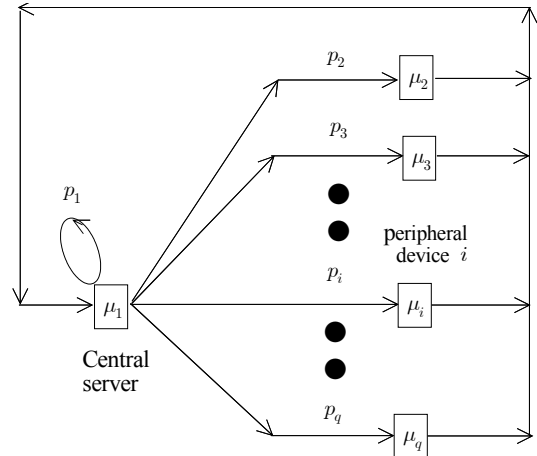


Fig. 1. Central server model

Exactly K jobs (programs) are permitted into computer system and circulate endlessly sharing the q resources of system. The central server, node 1, is meant to represent the CPU and the $q-1$ nodes (or stations) represent peripheral devices, for instance, a rotating disk memory, a swapping drum, a magnetic tape unit, a data cell, and so on. In such a multiprogramming environment, jobs do indeed circulate among these devices in such a way that they require the attention of the CPU followed by the need for some peripheral device, after which they again requiring the attention of some other peripheral device.

The transition probability, r_{ij} represents the probability of going next to node j upon leaving node i , and in this model, we see the constant circulation back to the central server, the CPU. Thus we have the transition probability matrix $R=(r_{ij})$ with

$$r_{ij} = \begin{cases} p_j & i = 1, 1 \leq j \leq q \\ 1 & 2 \leq i \leq q, j = 1 \\ 0 & otherwise \end{cases}
 \tag{19}$$

and $\sum p_j = 1$. In a true multiprogramming environment, most jobs eventually leave, at which time a new job will be inserted into system. This is represented in the model by permitting the job to return directly back to the CPU with probability p_1 , which represents the departure of an old job and insertion of a new job to replace it, with the recognition that this new request for service in the CPU

is the demand by the supervisory system for this job interchange. Thus the number of jobs in the system remains constant at a value of K .

For the simulation of this model, we assume that 1 central server and 5 peripheral units are operating. At the station i , jobs joins a FIFO queue to wait for service of a single server. The service-time distributions of stations 1, 2, 3, 4, 5, and 6 are Expon(10), Expon(50), N(40, 400), Expon(40), Uniform(60,100) and Expon(30), respectively. The probability p_1 representing the departure of an old job and insertion of a new job to replace it is 10% and the number of jobs in the system remains constant at a value of 20. After central service of station 1, jobs go to station i for another peripheral device service with probabilities: $p_2 = 0.35$, $p_3 = 0.25$, $p_4 = 0.15$, $p_5 = 0.10$ and $p_6 = 0.05$. Completing the service of the i th peripheral device, jobs go back to the central server for another service of device or leave from system. This kind of circulations continue repeatedly. For a more detailed description of this system, see the Section 4.13 of Kleinrock (1976).

4.2 Experimental Results

To perform an experimental evaluation of three types of control variables, using AweSim simulation language (Pritsker and O'reilly, 2008), we conduct 40 independent replications for the 10,0000 units time period. During the course of each simulation run, we collect the throughput of system, utilizations of 6 stations and system sojourn time as well as service times at 6 stations and routing probabilities at station 1. Based on correlations

between performances of interest and control variables, we select three controls showing highest correlations with performances. To construct the control variables in equations (16) and (18), we solve the stationary equation $\pi R = \pi$, and compute the stationary probability vector, $\pi = (.526, .184, .131, .079, .0526, .0263)$, and then we establish the controled estimators given in equations (15), (16) and (18).

Table 1 summarizes the simulation results showing the variances of estimators and their percentage reductions in variances (PRV). In estimating the throughput, sojourn time of system, and utilization of stations 1-3, and 5, control variable of routing probability yields superior results to service-time controls and combined controls. The PRV of routing probability control variable is in the range from 16% to 78%. Especially in estimating the throughput and sojourn time of system, routing probability control is very efficient and its PRV is more than 78%. Comparison of performances of three types of control variables gives that a) in estimating the throughput and sojourn time, the efficiencies of service-time and combined controls are similar and their PRVs are around 25% and 22% respectively, and b) in estimating the utilizations of stations 4 and 6, service-time control is more efficient than the routing probability and combined control variables. Contrary to our expectation on combined control, its effectiveness is not good as that of routing probability control. We conjecture that this simulation results of computer network model show the promise for application of routing probability control variable in practice.

Table. 1. Variance of Estimator and Percentage Reduction in Variance (PRV)

Performance Estimator	Without Control		Service Time Control (SC)		Routing Probability Control (RC)		Combined Control (CC)	
	Mean	Var	Var	PRV	Var	PRV	Var	PRV
Throughput	576	1064	792	25.6	229	78.4	791	25.7
Utilization 1	0.57	77E-4	67E-4	12.1	64E-4	16.0	70E-4	8.75
Utilization 2	0.99	18E-5	16E-7	10.4	13E-724E-5	26.8	16E-7	9.8
Utilization 3	0.58	70E-5	60E-5	14.2	21E-5	65.9	54E-5	22.6
Utilization 4	0.35	37E-5	17E-5	53.1	37E-5	41.6	18E-5	50.7
Utilization 5	0.48	53E-5	44E-6	17.2	64E-6	29.7	44E-5	17.5
Utilization 6	0.09	68E-6	21E-6	67.8	7330	54.6	50E-6	25.9
Sojourn Time	3351	33833	26264	22.4		78.3	26263	22.4

5. Conclusion

This research attempts to suggest controlled estimators which use the service-time and routing-probability random variables, and explores their efficiencies in estimating the throughput, utilizations of stations and sojourn time of central server model. Using the stationary probability and service time at each node, we construct the three types of control variables. Simulation results on computer network model show that the routing probability control variable is superior to service time and combined control variables. Most researches concerning about application of control variable method focus on service-time type control variable. This research presents that routing probability random variables may also be good candidates for a set of selected control variables. The routing-probability control variable is promising for applications to various simulation models when the stationary probability of each node is available in simulation model. We consider the applications of the routing probability control variables to various simulation models and combined control variables using information of service time and routing probability together in constructing one control variable as future researches.

References

1. Anderson, T. W., An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, p. 37, 1984.
2. Borogobac, T and Vakili, P., "Control Variates Technique: A Constructive Approach", Proceedings of the 2008 winter Simulation Conference, pp. 320-327, 2008.
3. Kleinrock, L., Queueing Systems Vol. II Computer Applications, John Wiley & Sons, New York, pp.230-232, 1976.
4. Kwon, C., "Efficiency of Estimation for Parameters by Use of Variance Reduction Techniques", Journal of Korean Society for Simulation, Vol. 14, No. 3, pp. 129-136, 2005.
5. Kwon, C. and Tew, J. D., "Strategies for Combining Antithetic and Control Variates in Designed Simulation Experiments", Management Science, Vol. 40, No. 8, pp. 1021-1034, 1994a.
6. Kwon, C. and Tew, J. D., "Combined Correlation Methods for Meta-model Estimation in Multi-population Simulation Experiments", J. of Statistical Computation and Simulation, Vol. 49, pp. 49-75, 1994b.
7. Lavenberg, S., Moeller, T. and Welch, P., "Statistical Results on Control Variables with Application to Queueing Network Simulation", Operations Research Vol. 30, No. 1, pp. 182-202, 1982.
8. Law, A. M., Simulation Modeling and Analysis, Fourth Edition, McGraw-Hill, Boston, pp. 357-361, 2007.
9. Pritsker, A. A. B. and O'reilly, J., Simulation with Visual SLAM and AweSim, John Wiley & sons, New York, 2008.
10. Szechtman, R., "Control Variates Technique for Monte Carlo Simulation", Proceedings of the 2008 winter Simulation Conference, pp. 144-149, 2003.



권 치 명 (cmkwon@dau.ac.kr)

1978 서울대학교 산업공학과 학사
1981 서울대학교 산업공학과 석사
1991 VPI & SU Dept. of ISE 공학박사
1983~현재 동아대학교 교수

관심분야 : 시스템 모델링, Output Analysis, Simulation Optimization



임 상 규 (sglim@gnu.ac.kr)

1978 서울대학교 산업공학과 학사
1986 미국 조지아공대 산업시스템공학과 석사
1990 미국 조지아공대 산업시스템공학과 공학박사
1991~현재 경상대학교 산업시스템공학부 교수

관심분야 : Material Handling, 생산관리