VARIANCE REDUCTION IN ESTIMATING THE MEAN FLOW TIME IN OPEN QUEUEING NETWORK SIMULATIONS

Gerald W. Adkins
The University of Georgia
Athens, Georgia 30602

Abstract

The number of replications required to produce a satisfactory confidence interval for the mean flow time through a queueing network can be prohibitively large unless an effective variance reduction technique is used. A variance reduction technique is developed specifically for the estimation of mean flow time and the corresponding sample variance is analyzed. Several examples are shown with variances reduced to as much as one percent of the variance of the original process.

I. INTRODUCTION

The number of replications required to produce a satisfactory confidence interval for the mean flow time through a queueing network can be prohibitively large unless an effective variance reduction technique is used. A variance reduction technique is developed specifically for the estimation of mean flow time and the corresponding sample variance is analyzed. Several examples are shown with variances reduced to as much as one percent of the variance of the original process.

In Section II the variance reduction technique is formulated and analyzed. Several examples are developed in Section III and conclusions are given in Section IV.

II. FORMULATION AND ANALYSIS

Consider an open network of N nodes with transition probability matrix P. With the nodes labeled 1 to N we have the ijth entry of P, p_{ij}, corresponding to the probability of a transition to node j after departing from node i. In order to avoid trivalities we assume that the average number of visits to each of the N nodes is finite. With this assumption it follows that the number of visits to node i, i = 1, 2, ..., N, is geometrically distributed with mean a_i where a_i is the ith component of the vector

$$A = P_0^T (I - P)^{-1},$$

where I is the N × N identity matrix and

$$P_0 = \begin{pmatrix} P_{01} \\ P_{02} \\ \cdot \\ \cdot \\ \cdot \\ P_{0N} \end{pmatrix}$$

where P_{0i} is the probability of a new arrival to the network entering at node i[2]. Let X_i be the random variable of passage time through node i. Then the mean flow time through the network, $E[X_F]$, is given by

$$E[X_F] = \sum_{i=1}^{N} a_i E[X_i] . \tag{1}$$

From equation 1 it is apparent that tne random variable

$$Y = \sum_{i=1}^{N} a_i X_i \tag{2}$$

can be used to estimate the mean of X_F since $E[X_F] = E[Y]$. The most significant characteristic of Y however, is that the source of variation due to the transition probabilities has been eliminated.

Proceedings of the 1982
Winter Simulation Conference
Highland * Chao * Madrigal, Editors

82CH1844-0/82/0000-0451 $00.75 © 1982 IEEE

From the examples shown in the next section it will be seen that for some networks the variation due to transitions can account for 100% of the variation in X_F.

In using equation (2) to produce an estimate of the mean of X_F, observations are made at each node i and estimates $\hat{X}_i$ are generated for $E[X_i]$, An estimate $\hat{Y}$ of E[Y] can then be obtained by

$\hat{Y} = \sum_{i=1} a_i \hat{X}_i$. The variance of Y is easily

obtained from (2) and is given by

$$\sigma_Y^2 = \sum_{i=1}^{N} a_i^2 \sigma_{x_i}^2 + 2 \sum_{i<j} a_i a_j \sigma_{x_i x_j} . \qquad (3)$$

It is tempting to conclude that the sample variance of $\hat{Y}$, $\hat{\sigma}_Y^2$, can be obtained from (3) by substituting $\hat{\sigma}_{x_i}^2$ and $\hat{\sigma}_{x_i x_j}$, the sample variance and covariances, appropriately. Equation (3) under these conditions overestimates the true sample variance because estimators for the mean values of the X_i's are used in (2) rather than individual random samples. We proceed with the discussion of the sample variance of $\hat{Y}$ after proving the following theorem.

Theorem. If X_i, i = 1, 2, ..., N are independent random variables then

$$\hat{\sigma}_Y^2 = \sum_{i=1}^{N} \beta_i \sigma_{x_i}^2 , \qquad (4)$$

$$\beta_i = \begin{cases} 1, \text{ if } \Pr\{M_i = 1\} = 1 \\ \dfrac{a_i^2 \ln(a_i + 1)}{a_i + 1}, \text{ otherwise} \end{cases}$$

and M_i is the random variable of the number of visits to node i.

Proof. Consider the random variable

$$\hat{Y} = \sum_{i=1}^{N} a_i \hat{X}_i$$

where

$$X_i = \frac{1}{M_i} \sum_{K=1}^{M_i} X_{iK} .$$

It is assumed that the X_{iK} are i.i.d. for each K = 1, 2, ..., M_i. Since each X_i and X_j are independent $\hat{X}_i$ and $\hat{X}_j$ are also. Thus

$$\sigma_{\hat{Y}}^2 = \sum_{i=1}^{N} a_i^2 \sigma_{\hat{x}_i}^2 .$$

To compute $\sigma_{\hat{x}_i}^2$ we use the general conditional relation

$$\mathrm{Var}[U] = \mathrm{Var}[E[U|W]] + E[\mathrm{Var}[U|W]].$$

It follows that

$$\sigma_{\hat{x}_i}^2 = \mathrm{Var}\left[E[\frac{1}{M_i} \sum_{K=1}^{M_i} X_{iK} \, M_i]\right]$$

$$+ E\left[\mathrm{Var}[\frac{1}{M_i} \sum_{K=1}^{M_i} X_{iK} | M_i]\right]$$

$$= \mathrm{Var}\left[E[X_i]\right] + E\left[\frac{1}{M_i} \sigma_{x_i}^2\right] = \sigma_{x_i}^2 E\left[\frac{1}{M_i}\right]$$

For the case when M_i is a random variable it is geometrically distributed with mean a_i. It follows that $E\left|\frac{1}{M_i}\right| = \frac{\ln(1+a_i)}{1+a_i}$. For the case when $\Pr[M_i = 1] = 1$, $E[1/M_i] = 1$ and the result follows.

In comparing equation (3) and (4) it is clear that the significant difference is the multiplicative constants a_i^2 as opposed to $a_i \ln a_i$ for large a_i. The random variables X_i are independent provided that each node is a delay server, i.e., an infinite server. However if queues exist in the network, X_i and X_j will not be independent if a sample path exists between them. An extreme case is where $P_{ij} = 1$. In this case a long delay at i due to queueing increases the likelihood of a similar delay at node j.

For the case of queueing in the network a closed form equation for the variance of Y comparable to (4) could not be obtained. It can be shown, however, that an upper bound for the sample variance of Y is given by

$$\hat{\sigma}_Y^2 = \sum_{i=1}^{N} a_i^2 \hat{\sigma}_{x_i}^2 + 2\sum a_i a_j \hat{\sigma}_{x_i x_j} \qquad (5)$$

where $\sigma_{x_i x_j}$ is the covariance between x_i and x_j. In the simulation cases used to test (5) this estimator was within approximately 30% of the calculated variance and the variance of X_F was consistently 40 to 60% greater than the variance of Y.

III. EXAMPLES AND DISCUSSION

In practice neither equations (4) and (5) need to be used. Equation (2) is recommended with the X_i's replaced by the estimated average value. The usual procedure for computing confidence intervals from a sample set is then used. The network shown at Figure 1 was analyzed numerically by the author in [1]. The transition matrix and analysis for this network is reproduced at Tables 1 and 2 respectively. In Figure 1 node indexes are shown in parenthesis, the average node transit times are shown immediately above the index, and node to node transition probabilities are shown at each transition arrow.

In Table 2 the random variable U_i denotes the time to exit the network starting from the time of arrival at node i. The equation used in the numerical computation for the entries are shown in the Appendix. Table 3 provides a summary of the variance computations for both Y and X_F. The average number of visits to each node is shown, the mean flow time, and a tabulation of the variances of Y and X_F for three different distributions at each node. The first distribution analyzed is the constant with means as shown in Figure 1. From equation (4) it is obvious that $\sigma_Y^2 = 0$. Therefore the variance of X_F is due entirely to variation in transitions through the network. The second distribution shown in Table 3 is uniform where the distribution for node i is uniform on $[0,2E[X_i]]$ where $E[X_i]$ is the average transit time for node i. The third distribution shown is exponential with mean $E[X_i]$. Note that for the uniform and exponential cases the variance of Y is 0.8% and 6.8% of X_F respectively. This correlates closely with the relative differences between the constant and uniform variances and the constant and exponential variances of X_F respectively, i.e., 0.5% and 5.8%. For this example it is true that this variance reduction approach does factor out the source of variation due to transitions.

IV. CONCLUSION

A variance reduction technique for estimating flow time in a general network was developed and analyzed for the case where each server is a delay server. The technique essentially eliminates variation due to transitions. From the examples shown it is clear that a significant reduction in the number of 'samples' required to generate confidence intervals of a specified width is possible with this approach. For simulations of multiple layes of networks, i.e., the nodes of the global network are networks and so forth, repeated application of this technique at each layer would provide a very desirable reduction in variance for estimating the overall flow time in the network.

REFERENCES

1. Adkins, G. The Mean and Variance of Flow Time Through Open Queueing Network Models, Proceedings COMPCON Fall '82, Washington, D.C. (1982).

2. Cinlar, E. Introduction to Stochastic Processes, Prentice Hall, 1975 (Chapter 6).

APPENDIX

The formulas used in producing Tables 2 and 3 are presented in this appendix. The following definitions are required:

I - N×N identify matrix

P - N×N transition matrix

$$P_0 = \begin{pmatrix} P_{01} \\ P_{02} \\ \vdots \\ P_{0N} \end{pmatrix}$$ - probability vector for new arrivals

$$E[U] = \begin{pmatrix} E[U_1] \\ E[U_2] \\ \vdots \\ E[U_N] \end{pmatrix}, E[X] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_N] \end{pmatrix}, Var[U] = \begin{pmatrix} Var[U_1] \\ Var[U_2] \\ \vdots \\ Var[U_N] \end{pmatrix}$$

The mean value of X_F is given as follows

$$E[X_F] = P_0^T E[U], \text{ where}$$

$$E[U] = (I - P)^{-1} E[X].$$

The variance of X_F is given by

Appendix (continued)

$$\mathrm{Var}[X_F] = \sum P_{0i}\mathrm{Var}[U_i] + \sum P_{0i}(1-P_{0i})E[U_i]^2$$

$$- 2 \sum_{i<j} P_{0i}P_{0j}E[U_i]E[U_j]$$

where $\mathrm{Var}[U] = (I - P)^{-1}C$ and C is a vector whose i^{th} component is given by

$$C_i = \mathrm{Var}[X_i] + \sum_j P_{ij}(1-P_{ij})E[U_j]^2$$

$$- 2 \sum_{0<j<K} P_{ij}P_{iK}E[U_j]E[U_K].$$

TABLE 1. Transition Matrix P for Figure 1

Node \ Node	1	2	3	4	5	6	7	8	9
1	0	.7	0	0	.3	0	0	0	0
2	0	0	.01	0	0	.99	0	0	0
3	0	0	0	1.	0	0	0	0	0
4	1.	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	.9	0	0
6	0	0	0	0	1.	0	0	0	0
7	0	0	0	0	0	0	0	.35	.65
8	0	0	0	1.	0	0	0	0	0
9	0	0	0	0	0	1.	0	0	0

TABLE 2. Mean and Variance of U_i for Constant, Uniform, and Exponential Cases

		X_i Constant			X_i Uniform		X_i Exponential	
Node	$E[X_i]$	$\mathrm{Var}[X_i]$	$E[U_i]$	$\mathrm{Var}[U_i]$	$\mathrm{Var}[X_i]$	$\mathrm{Var}[U_i]$	$\mathrm{Var}[X_i]$	$\mathrm{Var}[U_i]$
1	7	0	715.53	517494.47	4.08	520139.56	.49	549234.69
2	44	0	724.31	516926.03	161.33	519617.50	1936	549222.56
3	17	0	740.53	517494.50	24.08	520168.97	289	549587.69
4	8	0	723.53	517494.50	53.67	520144.87	64	549298.69
5	4	0	671.70	516883.28	1.33	519406.59	16	547162.31
6	8	0	679.70	516883.28	53.67	519411.94	64	547226.31
7	.95	0	741.89	519274.41	.08	522076.59	.90	552899.98
8	81	0	804.53	517494.50	546.75	520691.62	6561	555859.69
9	27	0	706.70	516883.31	60.75	519472.66	729	547955.25

TABLE 3. Variance Analysis

node	average number of visits
1	4.18
2	2.93
3	0.0293
4	3.18
5	10.0
6	8.75
7	9.00
8	3.15
9	5.85

$E[Y] = E[X_F] = 716$

	σ^2_Y	$\sigma^2_{X_F}$
Constant	0	517,494
Uniform	4,125	520,140
Exponential	37,124	549,235

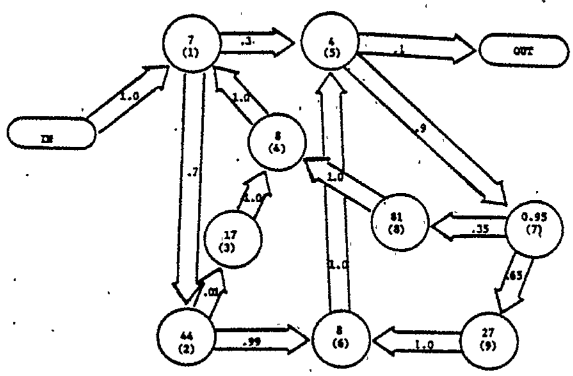

FIGURE 1. NETWORK FLOW EXAMPLE