Proceedings of the 2004 Winter Simulation Conference
R.G. Ingalls, M. D. Rossetti, J.S. Smith, and B.A. Peters, eds.

VARIANCE-BASED SAMPLING FOR CYCLE TIME - THROUGHPUT CONFIDENCE INTERVALS

Rachel T. Johnson
Sonia E. Leach
John W. Fowler
Gerald T. Mackulak

Industrial Engineering Dept.
Arizona State University
PO Box 875906
Tempe, AZ 85287-5906, U.S.A.

ABSTRACT

In the analysis of a manufacturing system, the analyst is often interested in the change in mean cycle time as a function of different throughput (start rate) levels. Since the variance of the mean cycle time generally increases as the start rate increases, an equal allocation of simulation effort at each simulated throughput level will result in confidence intervals of different widths. This paper discusses an approach for generating nearly equal sized mean cycle time confidence intervals at selected throughput levels when the computing budget is fixed (limited). The simulation effort allocation procedure described in this paper determines the proportion of the fixed budget to allocate at each throughput level based on the asymptotic variance.

1 INTRODUCTION

Cycle time – throughput curves (CT-TH) are often used as decision-support tools in manufacturing settings. They allow companies to evaluate the impact of proposed changes in start rate on mean cycle time so as to better plan production. A CT-TH curve plots average cycle time versus throughput (start rate) for a given product or fixed product mix (see Figure 1).

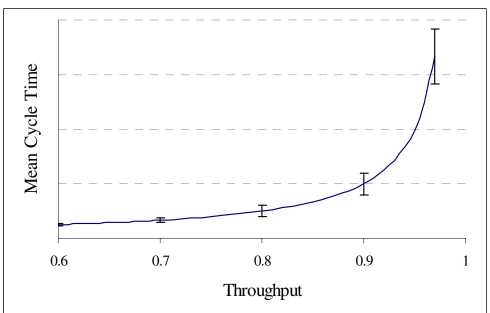

Figure 1: Confidence Intervals along a Cycle Time - Throughput Curve

Since most manufacturing systems are far too complex to model analytically, simulation is often used to generate various points along a CT-TH curve. However, because of many of the issues associated with simulation in complex manufacturing models (starting conditions, correlated data, steady state stopping conditions, etc.), simulation often requires long run lengths and extensive output analysis. In most manufacturing environments, the time and/or budget available for the generation of the CT-TH curve is limited. The typical simulation analyst is ill equipped to effectively allocate the limited budget in a statistically rational way.

We suggest an approach for allocating a fixed budget to previously selected throughput levels so that the resulting confidence interval half widths at these throughput levels are as equal as possible. But before we present our approach we need to convey some fundamental concepts regarding CT-TH curves that influence our allocation procedure.

A CT-TH curve is known to follow a non-linearly increasing trend (Figure 1) resulting in dramatic increases in cycle time variance as the throughput level nears the capacity of the system. A mean CT-TH curve can be estimated using simulation by modeling the system at several different throughput levels (design points) and obtaining their respective average cycle time values. The goal of the simulation is to achieve accurate estimates of mean cycle time, centered on the true value. Ideally, a good CT-TH curve estimation will present equal confidence interval half width precision. The closer to zero the confidence interval half-widths are, the more precise the estimate of the design point. Variance in cycle times and the estimates of mean cycle time prevent reaching a value of zero for confidence interval half-widths, even if the simulation is run for a very long time. Therefore, the analyst must choose a feasible goal to achieve, with respect to the estimates of the cycle time values obtained from simulation, especially when faced with a fixed budget.

When simulating a real world system under a constrained simulation budget, it is important to use a method of allocation that achieves the most accurate and precise mean cycle time estimates. While striving to achieve a confidence interval half width of zero at each design point may be impossible or take an arbitrarily long time, there are several ways of obtaining useful results. One option is to minimize the range of the confidence interval half-widths across the chosen throughput levels as given in equation (1) ,

$$\text{CIHWrange} = \max_{h \in H}\{CIHW_h\} - \min_{h \in H}\{CIHW_h\} \qquad (1)$$

where H is the selected set of design points the analyst chooses to simulate. The ideal minimum for this would be to get as close to zero as possible, where a range of zero indicates equal confidence interval half-widths across each simulated design point.

When faced with the task of modeling points along a CT-TH curve, two important decisions must be made; first, the throughput levels to model (i.e. the location of the design points) and second, the amount of effort to allocate at each throughput level in order to achieve the goal of a CIHW range of zero. There are several options for choosing what throughput levels to use. In order to capture the overall trend of the CT-TH curve it would be desirable to simulate over multiple throughput levels between zero and 100 percent. This method, while good for capturing the entire shape of the curve could lead to a large amount of simulation effort. If there is limited simulation effort available, it would be desirable to choose a small number of design points limiting the effort needed. Park *et al.* (2002) suggests a d-optimal sequential experiment for choosing throughput levels that best capture the shape of the curve and thus the asymptotic capacity of the system. However, in a real world setting, a pragmatic analyst may deem it unnecessary to choose design points that span the entire throughput range, but rather only a small portion. While CT-TH curves are used to characterize system performance over a range of throughput levels, a manager may be only interested in a small range of throughput levels that are feasible for the problem at hand.

In addition to having several options with regards to what design points to choose, it is also important to decide how to allocate simulation effort at each design point. This decision is also metric dependent and a function of the limited simulation budget. Techniques to allocate the effort across different throughput levels are found in the literature (Cheng and Kleijnen, 1999 and Whitt, 1989), however this paper will focus on the method known as naïve sampling as compared to the method proposed by Leach *et al.* (2004). Naive sampling equally allocates the fixed budget to all design points while Leach uses the asymptotic variance constant to determine the allocation. Although Leach's method involves more initial work to allocate the budget, it can be easily applied to any set of design points chosen to represent the system, allocates more effort to higher throughput levels, and makes intuitive sense.

In summary, this paper uses the method of allocation presented in Leach *et al.* (2004) to minimize the range of confidence interval half-widths over a narrow set of throughput levels. The allocation is based on a fixed budget of available simulation effort. The focus is to observe the performance of this method of allocation at achieving nearly equal confidence interval half-widths along a narrow range of throughput levels.

2 FIXED BUDGET

In practice there are often limits placed on the amount of effort allocated to certain tasks in a manufacturing setting. This effort can be measured in terms of time, money spent, or some other form of limited resource availability. This paper focuses on the allocation of simulation effort across several design points (throughput levels) to obtain precise estimates of cycle time, when a fixed computing budget is placed on the analyst carrying out the simulation.

In order to define the concept of a fixed budget, some assumptions must be made. It will be assumed that it takes one unit of computer time to generate and process one simulated observation (Cheng and Kleijnen 1999). Let H be the set of all design points investigated in the simulation experiment. Let n_h be the number of replications and m_h be the number of observations per replication for design point h. The budget allocated for the simulation will then be represented as total effort of the simulation, T, and it will be measured in number of observations, given in (2).

$$T = \sum_{h \in H} n_h m_h \qquad (2)$$

Using this definition of a fixed budget, T, a portion of this total effort can be allocated to each design point based upon a criterion of reaching nearly equal confidence interval half-widths at each design point in absolute or relative terms. The scope of this paper only includes the absolute CIHW case.

3 METHODOLOGY

Several methods for allocating effort across multiple design points to obtain specific confidence interval half-widths are known to exist. The simplest form of these methods is known as naïve sampling. As mentioned earlier, this method allocates equal amounts of effort at each design point without regard to the fact that as the throughput level increases, so does the variance. Another method of allocation, suggested by Law and Kelton (2000), is to expend a portion of the simulation effort available to a set of pilot runs and from these runs obtain a variance estimate of the mean cycle time estimates, S_o^2. The total number of runs, n, can then be calculated as in (3).

$$n \geq \left(\frac{t_{1-\alpha/2, n-1} S_\circ}{\varepsilon} \right)^2 \qquad (3)$$

where ε is the desired half-width, S_o is the standard deviation of the mean response estimates and $t_{1-\alpha/2,n-1}$ is the Student's t distribution quantile. While this method has proven to outperform the naïve method, it requires a pilot run. If a fixed budget is a factor, this pilot run would draw directly from the budget. When the budget is small, this pilot run could be costly, taking away effort needed to obtain desired confidence interval half-widths. A solution to this problem is to use the asymptotic variance of the design points. Whitt (1989) applies heavy-traffic queuing theory limits to estimate asymptotic variance of the desired response, which is then used to estimate the run length of a single simulation run analyzed as 20 batch means. This method completely removes the pilot run step, replacing mathematical calculations to obtain variance estimates of the mean cycle times. As suggested in Whitt (1989), the asymptotic variance approximation for delay time, Wq, for the M/M/1 queuing model is given in equation (4).

$$\tau^2\left(E\left[\overline{W}_q\right]\right) = \frac{\left(2+5\rho-4\rho^2+\rho^3\right)}{\left(1-\rho\right)^4} \tag{4}$$

It has been suggested that many complex systems perform like single channel queuing systems (Whitt 1989). Leach *et al.* (2004) uses the estimate of asymptotic variance for the M/M/1 queue from Whitt (1989) as a variance estimate of the mean cycle time to allocate effort across the design points. A portion of the fixed budget of total effort can be allocated to each design point based upon the confidence interval half-widths desired, ε_h. For the CT-TH curve, the objective is for the half-widths to be equal at each design point investigated, either in absolute or relative terms. Assuming the situation in which the length of the replications at each design point, m_h, are pre-determined, achieving equal absolute confidence interval half-widths at each design point such that all $\varepsilon_h=\varepsilon$ corresponds to the total effort, T, and percent effort for each design point h, π_h, given in (5) and (6).

$$T = \frac{1}{\varepsilon^2}\sum_{k\in H}\left(\left(t_{1-\alpha/2,n_k-1}\right)^2\tau^2(X_k)\right) \tag{5}$$

$$\pi_h = \frac{\left(t_{1-\alpha/2,n_h-1}\right)^2\tau^2(X_h)}{\sum_{k\in H}\left(\left(t_{1-\alpha/2,n_k-1}\right)^2\tau^2(X_k)\right)} \tag{6}$$

This method is used throughout this paper.

While this method has advantages, one problem encountered is due to the sharply increasing nature of the cycle time variance combined with the constrained budget. After calculations are made, the highest throughput level tends to consume nearly all of the simulation effort. Often, the lower design points receive virtually no simulation effort, requiring the analyst carrying out the simulation to give the design points a minimum default value. This occurs when the design points chosen to model the CT-TH curve span a large portion of possible throughput level values. For example, when the design points chosen are at the following throughput levels: 60, 70, 80, and 90 percent, the 90 percent design point claims approximately 99 percent of the budget available for the simulation effort. In a real world manufacturing setting, a manager may be very familiar with the flow of the system and already have rough estimates of points along the entire CT-TH curve. It may then be that the throughput levels of interest cover only a very small range of values, which will have a large impact on pending decisions. Results discussed in this paper focus on how this method of using asymptotic variance as a means for allocating the budget, performs when the design points chosen are from a fairly narrow range.

4 RESULTS

The experimentation to test the proposed method of allocation was done with the C++ discrete event simulator developed in McNeill, Mackulak, and Fowler (2003). All experiments were conducted on an M/M/1 queuing system using a first-in, first-out policy with a service rate of $\mu = 1$. The assumption that the system throughput is equal to the arrival rate of $\lambda < 1$ and a yield of one allows traffic intensity, $\rho = \lambda/\mu$, to be used to calculate system throughput. The design points for our experimentation fall within a range of six percent and a shift of design points was tested. For example, the first set of design points included the throughput levels of 78, 80, 82, and 84 percent. The shift then drops the first design point and adds one to the end. So, the next set of design points tested included 80, 82, 84, and 86 percent. This pattern continues for a total of eight different sets of design points in which the proposed method was tested and compared. Table 1 shows the eight design point scenarios and their corresponding throughput levels.

A total effort of 40 million observations was allotted for the budget used for the simulation of each M/M/1 CT-TH curve and a truncation length of 20% of each run was used to (hopefully) remove initialization bias. The results of the proposed asymptotic variance sampling method are compared to naïve sampling for the four different observations per replication scenarios suggested in Fowler *et al.* (2001). They are denoted Uniform (OU), Increasing (OI), omitting the second design point (O2), and omitting the third design point (O3). The percent of observations per replication assigned to the design points being investigated are shown in Table 2. As suggested by Fowler *et al.* (2001), one million observations were made available for assignment within the observation per replication assignment scenarios. So, the number of observations per replication is given by multiplying the percent given in Table 2 by one million. A minimum number of replications length was set to 10 in order to ensure that reasonable confidence in-

Table 1: Experimental Sets of Design Points with Numbers of Replications for the OI Case Indicated

A8								10	10	12	128
A7							10	10	26	114	
A6						10	16	34	100		
A5					12	20	38	90			
A4				15	23	41	81				
A3			17	25	42	76					
A2		19	27	43	71						
A1	20	29	43	68							
	0.78	0.8	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98

tervals are constructed. The replication assignments for the OI case scenario of the proposed method are shown in Table 1. The replication assignments for the other three replications per observations scenarios are given in Table 3.

Table 2: Observations Per Replication Assignment Scenarios

	Design Points			
	ρ=1	ρ=2	ρ= 3	ρ= 4
OU	25%	25%	25%	25%
OI	12.5%	12.5%	25%	50%
O8	10%	0%	30%	60%
O9	10%	30%	0%	60%

The results from the experiment are given in Table 3. The method for allocating effort to the design points proposed by Leach *et al.* (2004) exhibits superior performance to the naïve method in terms of reducing the absolute confidence interval half-width ranges. Due to the non-linear nature of the values for asymptotic variance, choosing a narrow design point range worked to the analyst advantage. It was previously known that typically all of the simulation effort was put towards the highest design point value. However, when the design points were picked from a narrow range, such as in this experiment, the effort given to each design point showed more of a steady increase across the throughput levels, allocating a reasonable amount of effort to the lower points, as opposed to having the lower design points default to the lowest value.

As seen from Table 3 the proposed method generally outperforms the naïve method. The absolute confidence interval half-widths are lower (indicated in bold) for 27 of the 32 scenarios using the proposed method. Four out of the five cases where the naïve method has a smaller CIHW are scenarios in which one of the design points was omitted. In these cases, the proposed method and the naïve method have relatively similar allocations across the design points varying only slightly in replication length. The one other case where the naïve method outperformed the proposed method included the design points ranging from .92 to .98 (A8). Due to the high variability at these throughput levels, it is possible that the proposed method demonstrated a higher CIHW range simply due to variability in the mean cycle time estimates.

5 CONCLUSIONS

The method presented in this paper demonstrates the ability of the method of Leach *et al.* (2004) when applied to a narrow range of design points. The proposed method generally outperforms the naïve sampling method. In addition to illustrating superior performance through reduced confidence interval half-width ranges, this method, when applied to a narrow range places significant percentage of efforts across all design points being evaluated. This contrasts with choosing a wider range of interest, where the highest point has shown to claim 99% of the simulation effort, forcing the analyst to default to the minimum specified allocations.

REFERENCES

Cheng, R. C. H., and J. P. C. Kleijnen. 1999. Improved design of queuing simulation experiments with highly heteroscedastic responses. *Operations Research* 47 (5):762-777.

Fowler, J. W., S. Park, G. T. Mackulak, and D. L. Shunk. 2001. Efficient cycle time – throughput curve generation using fixed sample size procedure. *International Journal of Production Research* 39 (12):2595-2613.

Law, A. M., and W. D. Kelton. 2000. Confidence intervals for steady-state simulation, II: a survey of sequential procedures. *Simulation Modeling Analysis. Third Edition.* Boston: McGraw-Hill, Inc.

Leach, S., J. W. Fowler, G. T. Mackulak, B. L. Nelson, and R. T. Johnson. 2004. Fixed Budget Variance-Based Sampling Allocation Method for Cycle Time – Throughput Curve Generation. *Work In Progress. Arizona State University.*

McNeill, J. E., G. T. Mackulak, and J. W. Fowler. 2003. Indirect estimation of cycle time quantiles from discrete event simulation models using the Cornish-fisher expansion. *Proceedings of the 2003 Winter Simulation Conference,* ed. S. Chick, P.J. Sanchez, D. Ferrin, D.J. Morrice, 1377-1382.

Park, S., J.W. Fowler, G. T. Mackulak, J.B. Keats, and W. M. Carlyle. 2002. D-Optimal Sequential Experiments for Generating Simulation-Based Cycle Time – Throughput Curve. *Operations Research* 50 (6):981-990.

Table 3: Allocation and Absolute Confidence Interval Range Results (α = 0.05) of the M/M/1 Model Simulated

Replications		Throughput											CIHW Range	
		0.78	**0.80**	**0.82**	**0.84**	**0.86**	**0.88**	**0.90**	**0.92**	**0.94**	**0.96**	**0.98**	**Naïve**	**Proposed**
A8:	**OU**								10	10	12	128	2.5053	**1.2615**
	OI								10	10	12	69	**1.1748**	1.5636
	O2								10	0	10	60	0.9398	**0.1543**
	O3								10	10	0	60	1.0941	**0.4689**
A7:	**OU**							10	10	26	114		0.7839	**0.2344**
	OI							10	18	26	60		0.4475	**0.1580**
	O2							10	0	22	54		0.3545	**0.0969**
	O3							10	10	0	60		0.3685	**0.1746**
A6:	**OU**						10	16	34	100			0.2346	**0.1234**
	OI						15	29	34	52			0.1177	**0.0989**
	O2						19	0	31	48			**0.0271**	0.1104
	O3						22	16	0	55			**0.0443**	0.1808
A5:	**OU**					12	20	38	90				0.1169	**0.0453**
	OI					21	37	39	46				0.0768	**0.0564**
	O2					28	0	36	44				0.0120	**0.0103**
	O3					34	22	0	50				0.0679	**0.0303**
A4:	**OU**				15	23	41	81					0.1189	**0.0539**
	OI				27	43	41	42					0.0377	**0.0108**
	O2				37	0	39	41					0.0128	**0.0017**
	O3				43	25	0	47					**0.0085**	0.0320
A3:	**OU**			17	25	42	76						0.0488	**0.0078**
	OI			31	49	42	39						0.0272	**0.0124**
	O2			46	0	42	38						0.0154	**0.0047**
	O3			52	28	0	44						0.0221	**0.0106**
A2:	**OU**		19	27	43	71							0.0454	**0.0199**
	OI		35	53	42	37							0.0190	**0.0149**
	O2		52	0	42	37							0.0175	**0.0119**
	O3		58	30	0	42							**0.0076**	0.0159
A1:	**OU**	20	29	43	68								0.0225	**0.0111**
	OI	39	55	43	35								0.0206	**0.0105**
	O2	58	0	44	35								0.0265	**0.0170**
	O3	64	32	0	40								0.0233	**0.0109**

Whitt, W. 1989. Asymptotic formulas for Markov process with applications to simulation. *Operations Research* 40 (2):279-291.

AUTHOR BIOGRAPHIES

RACHEL T. JOHNSON is a Masters student in the Industrial Engineering department at Arizona State University. Her research interest is in discrete event simulation methodologies. She currently serves as secretary of the ASU student chapter of INFORMS. She received her B.S. in Industrial Engineering from Northwestern University. She recently was awarded the SRC/Intel Fellowship for the duration of her Masters program.

SONIA E. LEACH is a Doctoral Candidate in Industrial Engineering at Arizona State University. She is an Instructor of Management in the Department of Systems and Engineering Management in the Graduate School of Engineering and Management at the Air Force Institute of Technology. Prior to her current position, she served as an operations research analyst in the Campaign Analysis and Information Superiority Branches of the Air Force Studies and Analysis Agency, Headquarters, United States Air Force, where her responsibilities included the technical and managerial oversight of stochastic warfare simulation models and deterministic resource allocation models.

JOHN W. FOWLER is a Professor of Industrial Engineering at Arizona State University (ASU). Prior to his current position, he was a Senior Member of Technical Staff in the Modeling, CAD, and Statistical Methods Division of SEMATECH. He spent the last year and a half of his doctoral studies as an Intern at Advanced Micro Devices. His research interests include modeling, analysis, and control of manufacturing (especially semiconductor) systems. He is the Co-Director of the Modeling and Analysis of Semiconductor Manufacturing Laboratory at ASU. The lab has had research contracts with NSF, SRC, International SEMATECH, Intel, Motorola, Infineon Technologies, ST Microelectronics, and Tefen, Ltd. Dr. Fowler is a member of ASEE, IIE, INFORMS, POMS, and SCS. He is an Area Editor *for SIMULATION: Transactions of the Society for Modeling and Simulation International* and an Associate Editor of *IEEE Transactions on Electronics Packaging Manufacturing*.

GERALD T. MACKULAK is an Associate Professor in the Department of Industrial Engineering at Arizona State University. He is a graduate of Purdue University receiving his B.Sc., M.Sc., and Ph.D. degrees in the area of Industrial Engineering. His primary area of research is in extending the methodology of simulation to a broader user base. For the past several years he has been concentrating on simulation applied to semiconductor manufacturing.