Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

BATCH SIZE SELECTION FOR THE BATCH MEANS METHOD

Chiahon Chien
Mitsubishi Electric Research Laboratories, Inc.
Sunnyvale, California 94086

ABSTRACT

The batch means method is among the most popular confidence interval techniques for the output analysis of a steady-state simulation. The selection of the batch size for the batch means method affects the quality of the confidence interval estimator. Many existing algorithms, however, are *ad hoc* in nature and lack a rigorous foundation. In this paper, we explore the relationship between the sample size and the optimal batch size ("optimality" is a sense to be defined in this paper). We focus on steady-state analysis and assume that the underlying process is stationary and strongly mixing. Three drastically different choices of batch sizes for the batch means method are discussed. Several empirical results illustrate our findings.

1 INTRODUCTION

An important problem in simulation output analysis is that of forming confidence interval estimators (c.i.e.s) for the mean of a stationary simulation process. The most popular c.i.e. technique is the method of batch means (Schmeiser (1982) and Bratley, Fox and Schrage (1987)). An important parameter in the batch means method is the choice of the batch size. The choice of the batch size has a direct impact on the quality of the variance estimator and the confidence interval (c.i.) (Schmeiser (1982) and Politis and Domano (1992)). Although the selection of an optimum batch size is still an unresolved research problem (Damerdji (1991)), numerous *ad hoc* approaches have been proposed (Bratley, Fox, and Schrage (1987) and Law and Kelton (1991)). While some of these approaches have been empirically tested, these algorithms lack a rigorous support. Extrapolation of these procedures beyond the cases that have been studied should be carefully examined.

On the other hand, Schmeiser (1982) suggested a practical and easy to implement algorithm. A fixed, relatively small number of batches (e.g., 30) is selected. For a given sample size, a test for statistical independence of the batches is checked. If the test result suggests that the batches are "almost" independent, a Student-t based c.i. is generated; otherwise, more samples are needed and the sample size is increased. However, practicality and simplicity is achieved at the expense of consistency for the variance estimator. Glynn and Whitt (1991) had shown that "there does not exist a procedure to consistently estimate the asymptotic variance constant when the number of batches is held constant as run length increases."

Another approach, as proposed by Goldsman and Meketon (1986), Schmeiser and Song (1994), and Chien, Goldsman, and Melamed (1994), is to focus on the the asymptotic properties of the variance estimator. They derive the asymptotic order of a batch size so that the mean square error (MSE) of the variance estimator is minimized. Their results are interesting and remarkable; however, there is no evidence whether their results are pertinent to the problem of generating c.i.e.s. We are unaware of any theoretical or empirical evidence that a better variance estimator would lead to a better confidence interval.

In this paper, we treat the problem of batch size selection from a new point-of-view and suggest a relationship between the optimal batch size and the sample size. Asymptotic properties of some key statistics used in the batch means method are studied. We focus on steady-state analysis. The underlying mathematical model in this study is a stationary stochastic sequence that satisfies certain regularity conditions. We do not assume that the observed data are sampled from independent and identically distributed (i.i.d.), regenerative, or ARMA processes (Schmeiser (1982) and Kang and Goldsman (1990)). This general model is useful for the output analysis of many real-world simulation experiments.

The rest of this paper is organized as follows. Section 2 is a review of the batch means method. Asymptotic properties of some statistics related in the batch means method are presented in Section 3. We discuss three different *optimal* batch sizes in Section 4, where optimal is a sense to be defined. Empirical results from a Monte Carlo study are presented there. Section 5 concludes the paper.

2 THE BATCH MEANS METHOD

Before discussing the batch means method, we digress to review the logic of confidence intervals. Assume that we want to estimate an unknown quantity μ. Suppose that an estimator $\hat{\mu}$ is normally distributed with *unknown* expectation μ and *known* standard error (s.e.) σ, viz., $\hat{\mu} \stackrel{\mathcal{D}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\stackrel{\mathcal{D}}{\sim}$ denote that the random variables (r.v.s) have the same distribution functions (d.f.s) and $\mathcal{N}(\mu, \sigma^2)$ is the normal d.f. with mean μ and variance σ^2. Let $\xi \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0,1)$. The αth percentile of $\mathcal{N}(0,1)$, z_α, is defined by $\mathsf{P}\{\xi \leq z_\alpha\} = \alpha$. By symmetry of $\mathcal{N}(0,1)$, $\mathsf{P}\{\xi \leq -z_\alpha\} = 1-\alpha$. It follows that $\mathsf{P}\{z_\alpha \leq \xi \leq -z_\alpha\} = 1 - 2\alpha$ (if $\alpha < 1/2$). Since $(\hat{\mu} - \mu)\sigma^{-1} \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0,1)$, with simple algebraic substitution, we derive $\mathsf{P}\{\mu \in [\hat{\mu} + z_\alpha \sigma, \hat{\mu} - z_\alpha \sigma]\} = 1 - 2\alpha$.

In practice, however, σ is usually unknown and needs to be estimated as well. Let $\hat{\sigma}$ be some reasonable estimate of s.e. for $\hat{\mu}$. "Under many circumstances it turn out that as the sample size grows large, the distribution of $\hat{\mu}$ becomes more and more normal and variance near σ^2" (Efron and Tibshirani (1993)). Then the random quantity $(\hat{\mu} - \mu)\hat{\sigma}^{-1}$ converges in distribution to a standard normal as the sample size increases, and $\mathsf{P}\{\mu \in [\hat{\mu} + z_\alpha \hat{\sigma}, \hat{\mu} - z_\alpha \hat{\sigma}]\} \approx 1 - 2\alpha$. We say that $[\hat{\mu} + z_\alpha \hat{\sigma}, \hat{\mu} - z_\alpha \hat{\sigma}]$ is a $(1 - 2\alpha) \times 100\%$ c.i.e. for μ. "Taken together, the point estimate $\hat{\mu}$ and the c.i. say what is the best guess for μ, and how far in error that guess might reasonable be" (Efron and Tibshirani (1993)).

It is of interest to obtain a c.i.e. for the mean μ of a discrete-time stationary process $\mathcal{X} = \{X_i : i \geq 1\}$. We assume throughout this paper that $\mathcal{X}$ is *strongly mixing* with mixing constants $\{\alpha_i : i \geq 1\}$. (Roughly speaking, a strongly mixing process $\mathcal{X}$ with mixing constant $\alpha_i \to 0$ as $i \to \infty$ implies that, among others, r.v.s X_1 and X_k are *almost independent* for large k; see Billingsley (1986) for a rigorous treatment.) Without loss of generality, it is assumed that $\mathcal{X}$ has zero-mean. (Otherwise, replace X_i with $X_i - \mu$.)

Assume that m observations, $X_1, \ldots, X_m$, from a time series $\mathcal{X}$ are given. Define $\overline{X}_m = m^{-1} \sum_{i=1}^m X_i$ as the sample mean associated with the sequence. Let $\sigma^2 = \mathsf{Var}(X_1) + 2\sum_{k=1}^\infty \mathsf{Cov}(X_1, X_{k+1})$ be the variance parameter (For many real-world processes, $\sigma^2 = \lim_{m\to\infty} m\mathsf{Var}(\overline{X}_m)$; σ^2 provides a measure for the sample mean's precision.) Suppose these m samples are split into n sub-intervals of length $b = \lfloor m/n \rfloor$ each (where $\lfloor p \rfloor$ denotes the greatest integer that is smaller than or equal to p). Let Y_i denote the mean over the ith interval, i.e., $Y_i = b^{-1} \sum_{j=1}^b X_{(i-1)b+j}$, for $i = 1, \ldots, n$. For the batch means method, n is the *number of batches*, b is the *batch size*, and the Y_i's are the *batch means*. Define $V_{n,b} = n^{-1} \sum_{i=1}^n (Y_i - \overline{X}_m)^2$. as the sample variance for $\{Y_i\}$ and let

$$t_{n,b} = \frac{\overline{X}_m - \mu}{(V_{n,b}/n)^{1/2}}. \tag{1}$$

Under mild regularity conditions (see, e.g., Brillinger (1973)) for fixed n, the r.v. $t_{n,b}$ converges in distribution to a Student-t d.f. with $n-1$ degrees of freedom as $m \to \infty$. The Student-t d.f., in turn, converges to the standard normal d.f. as the degree of freedom increases, so we can use the quantile points of either the Student-t d.f. or the standard normal d.f. to construct a c.i.e. for μ. For example, if z_α is the α-quantile point of the standard normal d.f., a $100(1-2\alpha)\%$ batch means (normal) c.i.e. for μ is

$$[\overline{X}_m + z_\alpha (V_{n,b}/n)^{1/2}, \overline{X}_m - z_\alpha (V_{n,b}/n)^{1/2}]. \tag{2}$$

3 PRELIMINARIES

We say that the sequence $\xi_1, \xi_2, \ldots,$ of r.v.s converges in $\mathcal{L}^p$ or in mean of order p $(0 < p < \infty)$ to the r.v. ξ if $\mathsf{E}[|\xi_s - \xi|^p] \to 0$ as $s \to \infty$. The notation $f(k) = O(g(k))$ means that $|f(k)/g(k)| \leq C$ for some constant C and all $k \geq 1$, and $f(k) = o(g(k))$ means that $f(k)/g(k) \to 0$ as $k \to \infty$.

The batch size b is critical to the quality of the variance estimate and the c.i.e. To see this, note that $\mathsf{Var}(\overline{X}_m) = n^{-1}\mathsf{Var}(Y_1) + 2n^{-2}\sum_{k=1}^{n-1}(n - k)\mathsf{Cov}(Y_1, Y_{k+1})$. If the covariance terms are much smaller than $\mathsf{Var}(Y_1)$, then $\mathsf{Var}(\overline{X}_m) \approx n^{-1}\mathsf{Var}(Y_1)$. Since $V^*_{n,b}$ is a good estimator for $\mathsf{Var}(Y_1)$, the batch means method uses $n^{-1}V^*_{n,b}$ as the estimator for $\mathsf{Var}(\overline{X}_m)$ and a c.i.e. can be generated accordingly. Since $\mathsf{Var}(Y_1) = b^{-1}\mathsf{Var}(X_1) + 2b^{-2}\sum_{k=1}^{b-1}(b - k)\mathsf{Cov}(X_1, X_{k+1})$; $\mathsf{Var}(Y_1)$ and $V^*_{n,b}$ (the estimator for $\mathsf{Var}(Y_1)$) are dependent on b. The choice of the batch size b will have a direct impact on the variance estimator and the quality of the associated c.i.e.

We define

$$V_{n,b} = n^{-1}\sum_{i=1}^n (Y_i - \overline{X}_m)^2, \tag{3}$$

$$\sigma_m^2 = m \cdot \mathsf{Var}(\overline{X}_m), \tag{4}$$

and

$$\Delta_{n,b} = \frac{V_{n,b}}{(\sigma_m^2/b)} - 1. \quad (5)$$

The following Propositions from literture are used for the rest of the paper.

PROPOSITION **3.1** (Carlstein (1986), Chien (1989), Chien, Goldsman, and Melamed (1994))

(1) If $\mathsf{E}|X_1|^{12} < \infty$ *and* $\alpha_i = O(i^{-9})$, *then* $\mathsf{E}(\Delta_{n,b}^2) = O(n^{-1}) + O(b^{-2})$; *therefore,*

(2) $\Delta_{n,b} \to 0$ *in* $\mathcal{L}^2$ *as* $n, b \to \infty$.

(3) For fixed m, $\mathsf{E}(\Delta_{n,b}^2)$ *achieves its lowest possible order, which is* $O(m^{-2/3})$, *when* $n = O(m^{2/3})$ *and* $b = O(m^{1/3})$.

It should be noted that Chien, Goldsman, and Melamed (1994) discuss Proposition 3.1 in detail. Interested readers are referred there for further properties, implications, and a number of analytical and numerical results. On the other hand, Damerdji (1991) uses a more general assumption to obtain $\mathcal{L}^2$ convergence of the variance estimator for the batch means method. Convergence in $\mathcal{L}^2$ of the processes-variance estimators for the overlapping batch means, standardized time series area, and spaced batch means methods are discussed in Damerdji (1991).

PROPOSITION **3.2** (Chien (1989)) *If* $\mathsf{E}|X_1|^{12} < \infty$ *and* $\alpha_i = O(i^{-9})$, *then as* $n, b \to \infty$,

$$\mathsf{Corr}(\Delta_{n,b}, \overline{X}_m) = \frac{\mathsf{Cov}^2(\Delta_{n,b}, \overline{X}_m)}{\mathsf{Var}(\Delta_{n,b})\mathsf{Var}(\overline{X}_m)} = O(b^{-1}).$$

Proposition 3.2 states that the coefficient of correlation between $\Delta_{n,b}$ and $\overline{X}_m$ decreases as the batch size increases, and is asymptotically uncorrelated to the number of batches. It is interesting to compare Proposition 3.2 with a proposition Kang and Goldsman (1990). Notice that Kang and Goldsman (1990) assume that the batch means are independent; a condition perhaps is true only when $\mathcal{X}$ is independent. In contrast, the assumption here allows for a much more general result. The order and leading terms for $\mathsf{Corr}(\Delta_{n,b}, \overline{X}_m)$, as expected, are identical.

Next, consider $t_{n,b} = \overline{X}_m - \mu/(V_{n,b}/n)^{1/2}$. The following proposition demonstrates how to calculate the first four cumulants of $t_{n,b}$.

PROPOSITION **3.3** (Chien (1989))

(1) If $\mathsf{E}|X_1|^8 < \infty$ *and* $\alpha_i = O(i^{-7})$, *then*

$$\kappa_1(t_{n,b}) = -\frac{\sqrt{m}b}{2\sigma_m^3}\kappa_{1,1}(\overline{X}_m, \frac{1}{n}\sum_{i=1}^n Y_i^2) + o(n^{-1/2})o(b^{-1/2}).$$

(2) If $\mathsf{E}|X_1|^{12} < \infty$ *and* $\alpha_i = O(i^{-9})$, *then*

$$\begin{aligned}\kappa_2(t_{n,b}) = {} & 1 + (b/\sigma_m^2)^2\kappa_2(\frac{1}{n}\sum_{i=1}^n Y_i^2) \\ & -(mb/\sigma_m^2)\kappa_{2,1}(\overline{X}_m, \frac{1}{n}\sum_{i=1}^n Y_i^2) \\ & -\mathsf{E}[\frac{b}{n\sigma_m^2}\sum_{i=1}^n Y_i^2 - 1] + 3/n \\ & +o(n^{-1}) + o(b^{-1}).\end{aligned}$$

(3) If $\mathsf{E}|X_1|^{16} < \infty$ *and* $\alpha_i = O(i^{-11})$, *then*

$$\begin{aligned}\kappa_3(t_{n,b}) = {} & (\sqrt{m}/\sigma_m)^3\{\kappa_3(\overline{X}_m) \\ & -\frac{3\sigma_m^2}{m}\kappa_{1,1}(\overline{X}_m, \frac{b}{n\sigma_m^2}\sum_{i=1}^n Y_i^2)\} \\ & +o(n^{-1/2})o(b^{-1/2}).\end{aligned}$$

(4) If $\mathsf{E}|X_1|^{20} < \infty$ *and* $\alpha_i = O(i^{-13})$, *then*

$$\begin{aligned}\kappa_4(t_{n,b}) = {} & -6(mb/\sigma_m^4)\kappa_{2,1}(\overline{X}_m, \frac{1}{n}\sum_{i=1}^n Y_i^2) \\ & +3(b/\sigma_m^2)^2\kappa_2(\frac{1}{n}\sum_{i=1}^n Y_i^2) \\ & +12/n + o(n^{-1}).\end{aligned}$$

Recall that the first four cumulants of a standard normal r.v. are 0, 1, 0, and 0 (Kendall, Stuart, and Ord (1991)). Summing up all the individual terms in Proposition 3.3, Chien (1989) shows that $\kappa_1(t_{n,b}) = O(n^{-1/2})O(b^{-1/2})$, $\kappa_2(t_{n,b}) = 1 + O(n^{-1}) + O(b^{-1})$, $\kappa_3(t_{n,b}) = O(n^{-1/2})O(b^{-1/2})$, and $\kappa_4(t_{n,b}) = O(n^{-1})$. Therefore, the first four cumulants of $t_{n,b}$ do converge to those of the standard normal variate as n and b increase. This is more informal evidence that the d.f. of $t_{n,b}$ converges to a standard normal d.f.

4 BATCH SIZE SELECTION

In this section, we review the results presented in Section 3 and discuss the implications to the selection of batch size. As discussed in Efron and Tibshirani (1993), "a good c.i. should give a dependably accurate coverage probabilities in all situations." In the ensuing discussion, we say that a batch size selection is better than another if the c.i. generated by the batch size has a better coverage characteristic.

Recall that Proposition 3.1 suggests that the MSE of the variance estimator achieves the lowest possible order when the batch size $b = O(m^{1/3})$. This batch size was proposed by Goldsman and Meketon (1986), Schmeiser and Song (1994), and Chien, Goldsman, and Melamed (1994).

On the other hand, Proposition 3.2 states that the order of the correlation between the mean and variance estimator decreases as batch size increases. However, Kang and Goldsman (1990) had shown empirically that "the correlation between the mean and the variance estimators does not share a strong relationship with the coverage of a c.i."

Proposition 3.3 suggests yet another way to choose the batch size. The first four cumulants of a standard normal r.v. are 0, 1, 0, and 0. To use a normal approximation to generate a c.i. for μ, it seems desirable to have the cumulants of $t_{n,b}$ as close as possible to those of the standard normal r.v. This argument leads to two naive, but natural, approaches for the choice of the batch size: (1) minimizing $\max_{1\leq i\leq 4}|\kappa_i(t_{n,b}) - \kappa_i(\xi)|$; and (2) minimizing $\sum_{i=1}^{4}[\kappa_i(t_{n,b}) - \kappa_i(\xi)]^2$. From Proposition 3.3, $\kappa_1(t_{n,b}) - \kappa_1(\xi) = O(n^{-1/2})O(b^{-1/2})$, $\kappa_2(t_{n,b}) - \kappa_2(\xi) = O(n^{-1}) + O(b^{-1})$, $\kappa_3(t_{n,b}) - \kappa_3(\xi) = O(b^{-1})$, and $\kappa_4(t_{n,b}) - \kappa_4(\xi) = O(n^{-1})$. In both cases the optimal b and n should be chosen such that b and n are of the same order, viz., both $b = O(m^{1/2})$ and $n = O(m^{1/2})$ as $m \to \infty$. Let b_m^* denotes the batch size that achieves the best coverage characteristic for a sample size of m. Based on Proposition 3.3, we suggest that

$$b_m^* = cm^{1/2} + o(m^{1/2}), \qquad (6)$$

where c is a constant depends on the underlying stochastic sequence. If eq. (6) is valid, a plot of the optimal batch size vs. the sample size in a log-log chart would be asymptotically linear with a slope of 1/2. This relationship of the batch size and the number of batches gives us a corresponding $t_{n,b}$ that is closer to a standard normal r.v. in the sense that the differences between their respective first four cumulants are smaller.

For the rest of the section, we empirically examine the relationship between the optimal batch size and the sample size. The M/M/1 waiting time process (with different load levels) and an AR(1) process (with different coefficients) were studied. Sample size for the experiments ranged from 32 to 32768. In the Monte Carlo study, 8192 independent replications of the M/M/1 waiting time process or the AR(1) process were drawn. For each replication, a 90% c.i.e. for the average waiting time (for M/M/1 queue) or the average sample value (for AR(1)) was generated (based on the batch means method). Since the theoretical value for the average waiting time for an M/M/1 queue and the mean of an AR(1) process are known, we checked whether the true mean lays within the c.i. The *coverage fraction* is defined as the ratio for the number of replications that the true mean lays within the c.i. and the total number of replications made in the experiment. Sample mean and s.e. of the length of the confidence interval were recorded as well.

Since 90% c.i.e.s were used, we denote that a batch size has a *better coverage characteristic* than another if its (empirical) coverage is closer to 0.9. Since the s.e. of the sample mean of 8192 independent binary r.v.s with parameter 0.9 is 0.0033, difference of coverage that was 0.0033 or smaller was insignificant. We regard that the coverage characteristics for batch sizes that generated such a smaller difference in coverage as the same. The batch size(s) that generated the best coverage characteristic were then denoted as the optimal batch size(s) (among all the batch sizes tried in the experiment).

EXPERIMENT **4.1** For the M/M/1 waiting time process, we used different load levels: 0.25, 0.5, and 0.75. The coverage characteristic, sample mean and s.e. of the length of the c.i. for each batch size were recorded. In Tables 1.a to 1.f, simulation results for sample size ranges from 32 to 32768 for the case that $\rho = 0.25$ are reported. The optimal batch size(s) (among all the batch sizes that has the form 2^i) with the corresponding coverage are marked with an *. For example, experiments with sample size of 2048 were reported in Table 1.d. It is clear that batch size = 32 generated the best empirical coverage (i.e., 0.8806). However, the coverage of batch size = 64 is 0.8773. Since the difference is smaller than 0.0033, both batch sizes are marked as optimal for the particular sample size.

Figure 1 summarizes the relationship between the optimal batch size(s) and the sample size from Tables 1.a to 1.f. X-axis denotes the sample size and Y-axis denotes the batch size. Based on the results in Tables 1 to 11, optimal batch size(s) for each sample size is marked with a circle. For example, the optimal batch size for sample size = 32 is 4, the optimal batch size for sample size = 64 is 8, and so on. Finally, the optimal batch sizes for sample size = 32768 are 64, 128, 256, and 512. It is observed that the optimal batch size increases as the sample size increases. A line with slope 1/2 (in log-log scale) is plotted in the chart. Notice that all the optimal batch sizes (for different sample sizes) are clustered around the line.

Figures 2 and 3 summarize the cases for $\rho = 0.5$ and $\rho = 0.75$. In Figure 2, we observe that the optimal batch sizes for different sample sizes are clustered around a line with slope 1/2 in log-log scale. The intercept of the line on the Y-axis, however, is greater than the one in Figure 1. This implies that, given a fixed sample size, the optimal batch size to simulate the average waiting time of a M/M/1 queue with $\rho = 0.5$ is larger than that of the queue with $\rho = 0.25$. It is in line with the fact that the waiting times from a queue with a higher traffic level are more positively correlated; in general, a larger batch size is required if the data are more correlated (see Bratley, Fox, and Schrage (1987)). In the sense of eq. (6), a more correlated data requires a larger constant c. Figure 3, however, portrays a slightly different picture. Although the optimal batch sizes for sample sizes that are 256 or larger are clustered around a line with slope 1/2 in log-log scale; for sample size = 16, 32, and 64, the optimal batch sizes are smaller than as indicated by the line. This phenomenon is due to the fact that since that data is so positively correlated, it would take a large constant c in eq. (6).

Tables 1.a-f Coverage, sample mean and sample standard error (s.e.) of the length of a 90% c.i. for various batch sizes. The data is based on 8192 independent replications of the M/M/1 waiting time process ($\rho = 0.25$).

Table 1.a Sample size for each replication is 32.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
32	1	32	0.6000	0.4181	0.2571
32	2	16	0.6449	0.4839	0.3499
32	4*	8	0.6674*	0.5379	0.4574
32	8	4	0.6566	0.5624	0.5443
32	16	2	0.5702	0.5368	0.6653

Table 1.b Sample size for each replication is 128.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
128	1	128	0.6365	0.2399	0.08474
128	2	64	0.7097	0.2858	0.1204
128	4	32	0.7551	0.3257	0.1625
128	8*	16	0.7719*	0.3499	0.1994
128	16	8	0.7662	0.3603	0.2283
128	32	4	0.7357	0.3557	0.2551
128	64	2	0.6230	0.3198	0.3027

Table 1.c Sample size for each replication is 512.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
512	1	512	0.6559	0.1259	0.02405
512	2	256	0.7467	0.1523	0.03494
512	4	128	0.8085	0.1768	0.04888
512	8	64	0.8381	0.1929	0.06240
512	16*	32	0.8468*	0.2011	0.07223
512	32*	16	0.8447*	0.2040	0.07918
512	64	8	0.8306	0.2029	0.08804
512	128	4	0.7860	0.1963	0.1053
512	256	2	0.6449	0.1721	0.1427

Table 1.d Sample size for each replication is 2048.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
2048	1	2048	0.6595	0.06379	0.006404
2048	2	1024	0.7533	0.07761	0.009435
2048	4	512	0.8232	0.09073	0.01355
2048	8	256	0.8595	0.09993	0.01797
2048	16	128	0.8756	0.1048	0.02121
2048	32*	64	0.8806*	0.1070	0.02333
2048	64*	32	0.8773*	0.1077	0.02543
2048	128	16	0.8717	0.1073	0.02913
2048	256	8	0.8539	0.1060	0.03507
2048	512	4	0.8035	0.1015	0.04629
2048	1024	2	0.6572	0.08916	0.06962

Table 1.e Sample size for each replication is 8192.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
8192	1	8192	0.6608	0.03207	0.001673
8192	2	4096	0.7524	0.03909	0.002485
8192	4	2048	0.8220	0.04582	0.003606
8192	8	1024	0.8639	0.05068	0.004940
8192	16	512	0.8807	0.05334	0.005993
8192	32	256	0.8876	0.05461	0.006592
8192	64*	128	0.8927*	0.05519	0.007245
8192	128*	64	0.8910*	0.05540	0.008130
8192	256	32	0.8870	0.05536	0.009502
8192	512	16	0.8788	0.05496	0.01173
8192	1024	8	0.8574	0.05400	0.01569
8192	2048	4	0.8013	0.05137	0.02240
8192	4096	2	0.6458	0.04418	0.03433

Table 1.f Sample size for each replication is 32768.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
32768	1	32768	0.6609	0.01605	0.0004212
32768	2	16384	0.7518	0.01957	0.0006269
32768	4	8192	0.8252	0.02296	0.0009153
32768	8	4096	0.8671	0.02542	0.001262
32768	16	2048	0.8867	0.02679	0.001558
32768	32	1024	0.8949	0.02744	0.001737
32768	64*	512	0.8989*	0.02777	0.001893
32768	128*	256	0.8982*	0.02792	0.002114
32768	256*	128	0.8988*	0.02799	0.002472
32768	512*	64	0.8972*	0.02796	0.003033
32768	1024	32	0.8918	0.02786	0.003933
32768	2048	16	0.8795	0.02747	0.005321
32768	4096	8	0.8544	0.02689	0.007477
32768	8192	4	0.8005	0.02577	0.01090
32768	16384	2	0.6624	0.02266	0.01682

EXPERIMENT **4.2** In this experiment, data was drawn from an AR(1) process. The residuals of the process are negative-exponential r.v.s with parameter 1. Four different coefficients were used: 0.25, 0.5, 0.75 and 0.9. Tables 2.a to 2.f list the simulation results for sample size ranges from 32 to 32768 for the case that the coefficient is 0.25. The optimal batch size(s) (among all the batch sizes that has the form 2^i) with the corresponding coverage are marked with an *. Figures 4 to 7 summarize the experimental results. In all cases, we observe that the optimal batch sizes for different sample sizes are all clustered around a line with slope 1/2 in the log-log scale. It is noted that the intercept of the line on the Y-axis becomes bigger as the coefficient increases. This is in line with our intuition since the data becomes more positively correlated as the coefficient increases.

Tables 2.a-f Coverage, sample mean and sample standard error (s.e.) of the length of a 90% c.i. for various batch sizes. The data is based on 8192 independent replications of an AR(1) process (coefficient = 0.25; the residual is exponentially distributed with parameter 1).

Table 2.a Sample size for each replication is 32.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
32	1	32	0.7742	0.5747	0.1387
32	2	16	0.8082	0.6371	0.1742
32	4*	8	0.8157*	0.6746	0.2255
32	8	4	0.7817	0.6748	0.3104
32	16	2	0.6399	0.5940	0.4612

Table 2.b Sample size for each replication is 128.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
128	1	128	0.7877	0.2974	0.03734
128	2	64	0.8304	0.3316	0.04645
128	4*	32	0.8563*	0.3555	0.05930
128	8*	16	0.8589*	0.3652	0.07749
128	16	8	0.8429	0.3640	0.1056
128	32	4	0.7913	0.3506	0.1513
128	64	2	0.6390	0.3015	0.2318

Table 2.c Sample size for each replication is 512.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
512	1	512	0.8043	0.1499	0.009420
512	2	256	0.8507	0.1674	0.01179
512	4	128	0.8804	0.1800	0.01488
512	8*	64	0.8918*	0.1863	0.01935
512	16*	32	0.8921*	0.1887	0.02616
512	32	16	0.8833	0.1890	0.03584
512	64	8	0.8616	0.1864	0.05137
512	128	4	0.8126	0.1785	0.07615
512	256	2	0.6583	0.1533	0.1179

Table 2.d Sample size for each replication is 2048.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
2048	1	2048	0.8104	0.07503	0.002365
2048	2	1024	0.8567	0.08386	0.002979
2048	4	512	0.8816	0.09015	0.003782
2048	8	256	0.8932	0.09345	0.004888
2048	16*	128	0.9006*	0.09501	0.006517
2048	32*	64	0.9015*	0.09568	0.008941
2048	64*	32	0.8970*	0.09558	0.01245
2048	128	16	0.8865	0.09472	0.01769
2048	256	8	0.8634	0.09295	0.02514
2048	512	4	0.8075	0.08917	0.03748
2048	1024	2	0.6566	0.07743	0.06008

Table 2.e Sample size for each replication is 8192.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
8192	1	8192	0.8021	0.03754	0.0005975
8192	2	4096	0.8535	0.04197	0.0007512
8192	4	2048	0.8789	0.04513	0.0009473
8192	8	1024	0.8923	0.04679	0.001225
8192	16	512	0.8966	0.04759	0.001638
8192	32*	256	0.9009*	0.04797	0.002204
8192	64*	128	0.9025*	0.04815	0.003094
8192	128*	64	0.9006*	0.04813	0.004396
8192	256	32	0.8965	0.04800	0.006163
8192	512	16	0.8833	0.04760	0.008895
8192	1024	8	0.8571	0.04658	0.01282
8192	2048	4	0.8036	0.04422	0.01892
8192	4096	2	0.6476	0.03779	0.02870

Table 2.f Sample size for each replication is 32768.

Samp. size	Batch size	# of batches	Coverage	Samp. mean of C.I.	Samp. s.e. of C.I.
32768	1	32768	0.8076	0.01877	0.0001481
32768	2	16384	0.8511	0.02099	0.0001864
32768	4	8192	0.8767	0.02257	0.0002354
32768	8	4096	0.8879	0.02341	0.0003029
32768	16	2048	0.8945	0.02382	0.0004063
32768	32*	1024	0.8973*	0.02402	0.0005566
32768	64*	512	0.8993*	0.02413	0.0007708
32768	128*	256	0.8993*	0.02417	0.001082
32768	256*	128	0.9014*	0.02417	0.001517
32768	512*	64	0.8987*	0.02415	0.002153
32768	1024	32	0.8918	0.02404	0.003078
32768	2048	16	0.8822	0.02364	0.004381
32768	4096	8	0.8564	0.02314	0.006240
32768	8192	4	0.8066	0.02216	0.009233
32768	16384	2	0.6644	0.01957	0.01448

Figure 1. Optimal batch sizes are based on 8192 indep. replications of M/M/1 waiting time process with load 0.25.

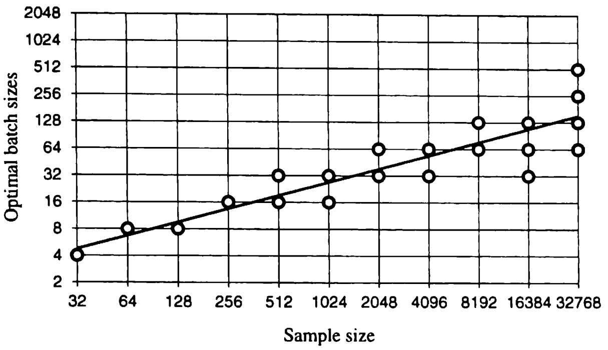


Figure 2. Optimal batch sizes are based on 8192 indep. replications of M/M/1 waiting time process with load 0.5.

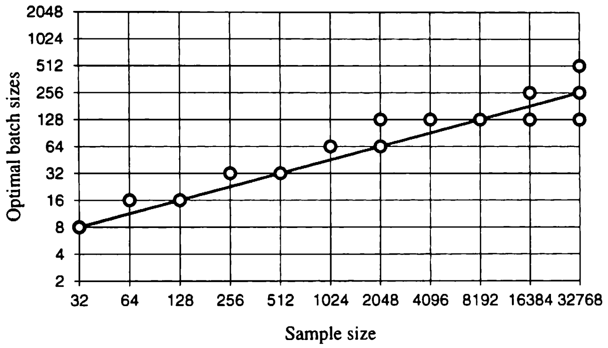


Figure 3. Optimal batch sizes are based on 8192 indep. replications of M/M/1 waiting time process with load 0.75.

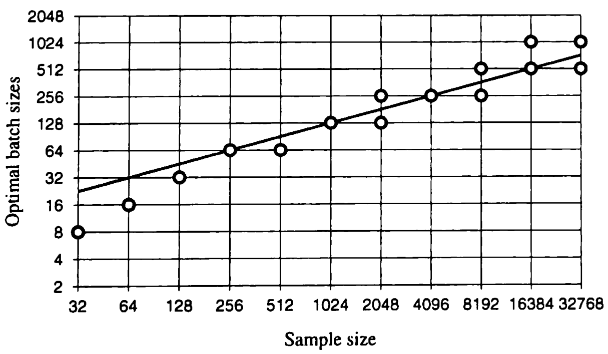


Figure 4. Optimal batch sizes are based on 8192 indep. replications of AR(1) process with parameter 0.25.

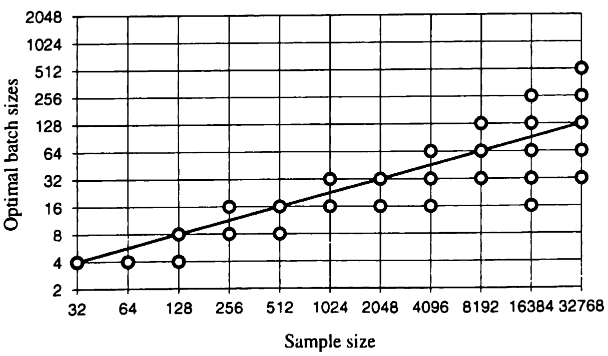


Figure 5. Optimal batch sizes are based on 8192 indep. replications of AR(1) process with parameter 0.5.

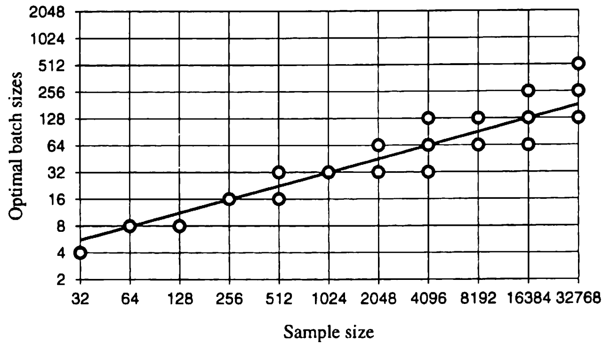


Figure 6. Optimal batch sizes are based on 8192 indep. replications of AR(1) process with parameter 0.75.

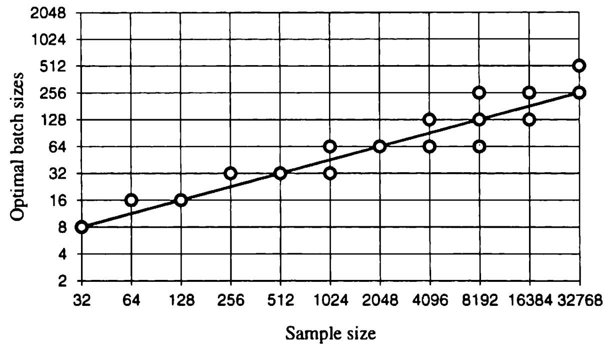
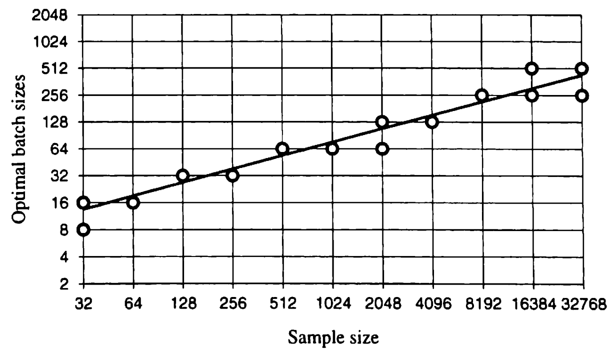


Figure 7. Optimal batch sizes are based on 8192 indep. replications of AR(1) process with parameter 0.9.



5 SUMMARY

Proposition 3.1, Proposition 3.2 and Proposition 3.3 suggest three different methods of reasoning in choosing an *optimal* batch size. However, the resulting batch sizes turn out to be drastically different. While Proposition 3.1 is intended to minimize the MSE of the variance estimator and Proposition 3.2 is intended to minimize the correlation between the mean and variance estimator, Proposition 3.3 can be used to minimize the difference between the cumulants of the statistic of the batch means method and those of the standard normal r.v. If the validity of the c.i. is the main concern, we believe that Proposition 3.3 yields a preferred choice of batch size for the traditional the batch means method. Numerical experiments in Section 4 agree with our findings.

ACKNOWLEDGEMENTS

The author is greatly indebted to Professors Donald Iglehart and Peter Glynn for their guidance, encouragement, and many helpful suggestions on this research. He wishes to thank Professors Halim Damerdji, David Goldsman, and Pierre L'Ecuyer for the helpful comments they made on earlier versions of this work.

REFERENCES

Billingsley, P. 1986. *Probability and Measure*, 2nd edition. John Wiley & Sons, New York.

Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A Guide to Simulation*, 2nd edition. Springer-Verlag, New York.

Brillinger, D. R. 1973. Estimation of the mean of a stationary time series by sampling. *J. Appl. Prob.*, **10**:419–431.

Carlstein, E. 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics*, **14**:1171–1179.

Chien, C. 1989. *Small sample theory for steady-state confidence intervals.* PhD thesis, Dept. of Oper. Res., Stanford University.

Chien, C., D. Goldsman, and B. Melamed. 1994. Large-sample results for batch means. *Submitted for publication.*

Damerdji, H. 1991. Mean square consistency of the variance estimator in steady-state simulation output analysis. *Submitted for publication.*

Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Glynn, P. W. and W. Whitt. 1991. Estimating the asymptotic variance with batch means. *Operations Research Letters*, **10**:431–435.

Goldsman, D. and M. S. Meketon. 1986. A comparison of several variance estimators for stationary increment stochastic processes. Technical report, School of Ind. and Sys. Engr., Georgia Inst. of Tech.

Kang, K. and D. Goldsman. 1990. The correlation between mean and variance estimators in computer simulation. *IIE Trans.*, **22**:15–23.

Kendall, M. G., S. Stuart, and J. K. Ord. 1991. *Kendall's Advanced Theory of Statistics*, 5th edition. Oxford Univ. Press, New York.

Law, A. M. and W. D. Kelton. 1991. *Simulation Modeling and Analysis*, 2nd edition. McGraw-Hill, New York.

Politis, D. N. and J. P. Romano. 1992. A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics*, **20**:1985–2007.

Schmeiser, B. W. 1982. Batch size effects in the analysis of simulation output. *Oper. Res.*, **30**:556–568.

Schmeiser, B. W. and W. T. Song. 1994. Optimal mean-squared-error batch sizes. *To appear in Management Science.*

AUTHOR BIOGRAPHY

Chiahon Chien is a Senior Engineer at Mitsubishi Electric Research Laboratories, Inc. He received B.S. degree in Electrical Engineering from National Taiwan University, M.S. degree in Statistics, and M.S. and Ph.D. degrees in Operations Research from Stanford University. Before joining MERL, he worked for Bell Northern Research and IBM in performance evaluation, and HaL Computer Systems in computer system architecture. His research interests include performance evaluation, simulation methodology, and computer architecture. Dr. Chien is a member of ORSA, ACM and IEEE.