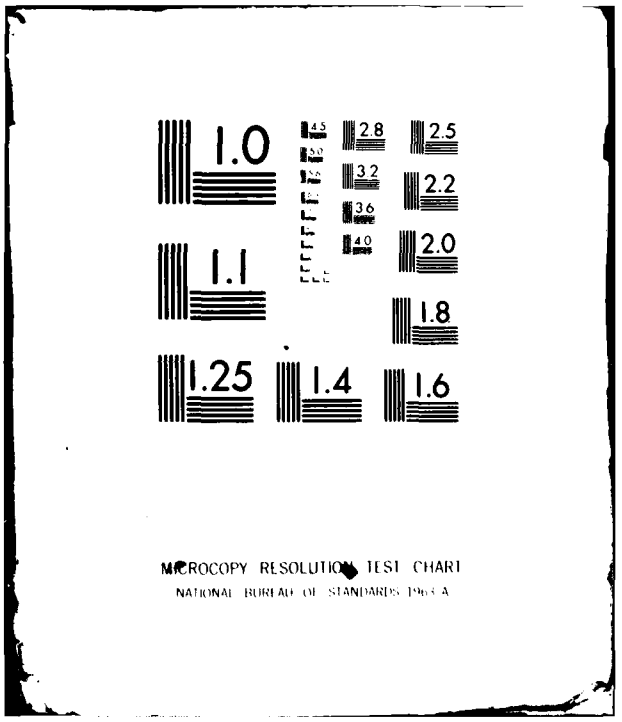
AD-A087 045 PURDUE UNIV LAFAYETTE IND SCHOOL OF INDUSTRIAL ENGIN--ETC F/G 12/1
BATCH SIZE EFFECTS IN THE ANALYSIS OF SIMULATION OUTPUT.(U)
JUN 80 B SCHMEISER N00014-79-C-0832
UNCLASSIFIED NL

END
DATE FILMED
9 80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

(6)

1

LEVEL II

BATCH SIZE EFFECTS IN THE ANALYSIS OF SIMULATION OUTPUT

Bruce Schmeiser

School of Industrial Engineering

Purdue University

N00014-79-C-0832

June 1980

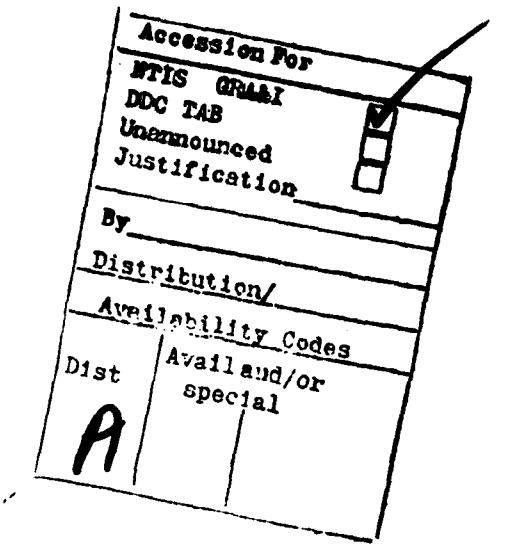
ABSTRACT

Batching is a commonly used method for calculating confidence intervals on the mean of a sequence of correlated observations arising from a simulation experiment. Several recent papers have considered the effect of using too many batches. The use of too many batches fails to satisfy assumptions of normality and/or independence, resulting in incorrect probabilities of the confidence interval covering the mean.

This paper considers the effects of using fewer batches than are necessary to satisfy normality and independence assumptions. Using too few batches results in 1) correct probability of covering the mean, 2) an increase in expected half length, 3) an increase in the standard deviation of the half length, and 4) an increase in the probability of covering incorrect values of the mean (analogous to Type II error in hypothesis testing). These effects, quantified here, are shown to be small when at least eight to ten batches are used, with least effect on confidence intervals having low confidence values. With the effects of using too few batches quantified, a simulation practitioner can make the trade-off between the ease of using very few batches with known independence and normality versus using a batching algorithm to squeeze some remaining information from the data. For researchers developing batching algorithms, the results are useful in selecting initial batch sizes. The results may also be useful in the context of using independent replications to establish confidence intervals on the mean.

next page

Finally, some criteria and a procedure are suggested for Monte Carlo comparison of confidence interval procedures. These suggestions are not restricted to batch mean algorithms.



1. INTRODUCTION

The determination of confidence intervals on the mean of a process arising from simulation experiments has been a problem of long standing interest for computer simulation practitioners and researchers. Five approaches have evolved: independent replications, batching, regeneration, autoregressive representation, and spectral analysis; as discussed, for example, in Fishman [3], Kleijnen [4] and Law and Kelton [8, 9]. We discuss only the first two here, with emphasis on batching.

Consider observations $X_1, X_2, \ldots, X_n$ from a simulation experiment. We assume that the output is a covariance stationary process; i.e., all initial transient effects have been removed. Let μ denote the process mean, and let R_h denote the h lag covariance $E\{(X_i-\mu)(X_{i+h}-\mu)\}$. We assume that $R_h < R_{h+j}$ for $j = 1,2,\ldots$. The variance of the process is R_0, also denoted in this paper as σ^2. The h lag autocorrelations are $\rho_h = R_h/R_0$. The point estimator of μ considered here is the sample mean $\overline{X} = \sum_{i=1}^{n} X_i/n$, which has expected value $E(\overline{X})=\mu$ and variance $V(\overline{X})=c\sigma^2/n$, where $c = 1 + 2 \sum_{h=1}^{n} (1-(h/n))\rho_h$ is the number of correlated observations containing the same information as one independent observation.

Batching, discussed as early as 1963 by Conway [1], is a conceptually straightforward method for computing confidence intervals on μ by transforming correlated observations into fewer (almost) independent and (almost) normally distributed observations. Define the k batch means $\overline{X}_i = \sum_{j=(i-1)m+1}^{im} X_j/m$ for $i=1,2,\ldots,k$; where m is the batch size n/k.

We assume either that the problem of n/k not being integer is insignificant or that n/k is integer. The mean of each batch mean is $E\{\overline{X}_i\}=\mu$ and the sample variance is $S_k^2=(\sum_{i=1}^{k} \overline{X}_i^2 - k\overline{X}^2)/(k-1)$. If k is chosen small enough that the dependence between the batch means and the nonnormality of the batch means are negligible, then S_k^2/k is an unbiased estimator of $V\{\overline{X}\}$ and a valid $(1-\alpha)100\%$ confidence interval on μ is $\overline{X} \pm H_k$. Here $H_k = t_{\alpha/2,k-1}S_k/\sqrt{k}$ is the half length of the confidence interval, with $t_{\alpha/2,k-1}$ denoting the $1-(\alpha/2)$ quantile of the t distribution with $k-1$ degrees of freedom.

The primary question facing both practitioners and researchers is the selection of the appropriate number of batches k. If the only goal were to have a confidence interval which has a probability of $1-\alpha$ of covering the mean, then $k=2$ would always be optimal, since the two batches would each contain the longest possible number of observations (allowing the central limit theorem to create normality) and would tend to be less correlated (since the observations in the batches are farther apart), thereby best satisfying normality and independence assumptions. However, other measures of goodness are important. Probably the most used criterion, other than probability of coverage, is the expected half length of the confidence interval, $E\{H_k\}$. The loss of information which occurs with the extreme batching associated with $k=2$ causes $E\{H_k\}$ to be much larger than if more batches are used, as seen in Section 2. In general, the more batches used, the less information lost and the shorter the expected half length. Thus there is a tradeoff between expected length and coverage which makes the selection of number of batches difficult.

We examine here the penalty of using fewer batches than are necessary to satisfy normality and independence assumptions. When k is smaller than necessary, normality and independence still hold, but the loss of information causes deteriorating performance of the confidence intervals. In addition to measuring this deterioration of the performance by the probability of covering μ and the expected half length $E\{H_k\}$, two other measures are suggested here: standard deviation of the half length, $\sqrt{V\{H_k\}}$, and the probability $\beta(\mu_1)$ of covering points $\mu_1 \neq \mu$. The variance of the half length is important since a confidence interval procedure with high variance gives false signals as to the accuracy of the estimate on a large fraction of the simulation runs. The probability of covering points which are not the true mean is analogous to type II error in hypothesis testing -- the lower the probability the better the procedure. Curves analogous to operating characteristic curves are the subject of Section 3. Section 2 gives properties of the half length H_k. Section 4 discusses implications of the results of Sections 2 and 3 for both practitioners and researchers. Section 5 suggests that the probabilities of coverage be used to compare confidence interval procedures, analogous to comparing alternative tests of hypothesis with operating characteristic curves.

2. PROPERTIES OF THE HALF LENGTH

To discuss the effects of too few batches, we need to first establish a base point for comparison. Let k^* and m^* denote the number of batches and batch size, respectively, that are necessary for the nonnormality of the batch means and the dependence of the batch means to be negligible. Establishing values for these quantities is difficult, but

for our purposes they need not be actually determined. For simplicity in the following analysis, we assume that k^* batches are sufficient to provide exact normality and independence and therefore exact $1-\alpha$ level confidence intervals. This assumption of exact normality and independence is relaxed at the end of Section 4.

The assumption that R_h decreases as h increases, made in Section 1, implies that using fewer than k^* batches will also result in normality and independence. This is usually a valid assumption, but, for example, Schriber and Andrews [12] consider sequences of independent trivariate normal observations for which the results of this paper do not hold, since there $m^* = 3$ satisfies normality and independence exactly, but $m = 4, 5, 7, 8, 10, \ldots$ do not.

Consider the expected half length resulting from k batches. $E\{H_k\}$ is inversely proportional to the square root of the sample size n when observations are independent, and even for correlated observations a quadrupling of the sample size will cut the expected half length to about one-half its original value. The same is not true for the number of batches, k, when n remains constant. This is because S_k, the standard deviation of the k batch means, is a function of k. Similar results are true of the variance of the half length, $V\{H_k\}$. Nevertheless, changing k does affect these properties, as shown in Table 1.

Table 1 shows $E\{H_k\}$ and $\sqrt{V\{H_k\}}$ for $k = 2, 3, 4, 5, 6, 10, 30, 61, 121$, and ∞ and for $\alpha = .10, .05$, and $.01$. The units are $V\{\overline{X}\} = c\sigma/\sqrt{n}$, making the tables valid for all values of μ, σ^2 and n. The correlation structure between the batches is not a factor so long as $k \leq k^*$, since then the batch means are independent and normally distributed. The as-

sociated t distribution quantiles are shown, as well as the dimensionless bias ratio $r = E\{S_k/\sqrt{k}\}/\sqrt{V\{\bar{X}\}}$, upon which the other quantities depend. The values in Table 1 are derived in Appendix A. Note that the values are deterministically calculated, rather than being the result of Monte Carlo experiments.

Table 1 about here

As expected, $E\{H_k\}$ decreases monotonically as k is increased for all values of α. The rate of decrease is much larger for small values of k than for large values. In fact, the decreases in $E\{H_k\}$ associated with increasing k from ten to infinity is only about twelve percent for α=.05. The correct comparison is not between ten and infinity, however, but between ten and k^*, since more batches than k^* do not result in valid confidence intervals. The decrease in length is about ten percent when k^*=61 and about eight percent when k^*=30.

Although the expected half length is robust for $k \geq 10$, the standard deviation exhibits a different pattern. While also decreasing as k increases, and decreasing more rapidly for small values of k then for large values of k, $\sqrt{V\{H_k\}}$ is affected more by k than is $E\{H_k\}$, indicating that the stability of the confidence interval associated with less variance may be a reason to exert more effort to use many batches. However, again k should not be compared with the limiting results at infinity, but rather with k^*. For α=.05 and k=10, the variance is decreased about 48 percent if k^*=30 and about 65 percent if k^*=61. Thus the major benefit of using more than ten batches may be more in reducing $V\{H_k\}$ than in reducing $E\{H_k\}$.

3. PROBABILITIES OF COVERAGE

Although consideration of the moments of the half length are intuitively appealling, a more comprehensive criterion is to compute the probability, $\beta(\mu_1)$, that the confidence interval covers μ_1, as a function of μ_1, i.e., $\beta(\mu_1) = P(\overline{X} - H_k \leq \mu_1 \leq \overline{X} + H_k)$. When $\mu_1 = \mu$ this is the commonly considered probability of coverage of the mean, analogous to one minus the probability of type I error when testing hypotheses. When $\mu_1 \neq \mu$ this probability is analogous to the type II error. The coverage function is more comprehensive than the expected value and variance of the half length because the coverage function considers the performance of $\overline{X}$ and S_k together while the half length is a function of S_k only. In addition the coverage function is directly related to the final product -- covering the mean of the process.

We are interested in calculating the probability, $\beta(\mu_1)$, of covering μ_1 as a function of μ, σ^2, n, α and k when $k \leq k^*$. Results for $\alpha = .10, .05$ and $.01$ are shown in Figures 1, 2, and 3, respectively. Each figure is valid for all values of μ, μ_1, σ^2 and n by plotting the probabilities of coverage as a function of $\delta_k = |\mu_1 - \mu|(n/(c\sigma^2))^{1/2}$. The derivation of the values, based on the noncentral Student's t distribution, is given in Appendix B.

Figures 1, 2 and 3 about here

For $k \leq k^*$, the following patterns emerge from Figures 1, 2, and 3:

1. The probability of covering μ (corresponding to $\delta=0$) is $(1-\alpha)$ for $k = 2, 3, \ldots, k^*$.
2. The decrease in $\beta(\mu_1)$ due to incrementing k by one decreases as k increases.

3. The decrease in $\beta(\mu_1)$ due to incrementing k by one decreases as α increases.

4. The decrease in $\beta(\mu_1)$ due to incrementing k by one is small when $\delta < 1$.

5. The decrease in $\beta(\mu_1)$ due to incrementing k by one is independent of n, other than that k^* increases with n.

As with moments of the half length of the intervals, it is important here to distinguish between the effect of increasing the sample size n and increasing the number of batches k. For an increase in the number of observations n, there is a corresponding increase in δ by a factor of $n^{1/2}$, and a corresponding decrease in the probability of covering any point μ_1 other than $\mu_1=\mu$. The effect of increasing the number of batches k is much less. For example, when $\alpha=.05$ and k=10 tripling δ from $\delta=1$ to $\delta=3$ (i.e., increasing n by a factor of eight) reduces the coverage from .87 to .23. However tripling the number of batches from k=10 to k=30 reduces the coverage from .86 to .83 when $\delta=1$ and from .23 to .18 when $\delta=3$.

4. IMPLICATIONS

The results of Sections 2 and 3 quantify the effects of using too few batches. Here we discuss the implications these results have for both practitioners and researchers interested in developing algorithms to determine the number of batches to be used.

When running a simulation experiment, the practitioner faces two constraints: (1) The run must be long enough to provide the desired accuracy and (2) the run must be long enough to calculate a valid confidence interval. (For a related discussion, see Lavenberg and Sauer [5,

p. 555].) Now if the latter constraint comes into play, then using a small number of batches will allow the simulation run to be terminated earlier than if more batches are demanded, while still resulting in a confidence interval with correct coverage. The results of the last two sections can be used to determine the penalty for using a specific (small) number of batches.

Probably a much more common situation is when the accuracy required forces the run to be longer than the number of observations necessary to establish a valid confidence interval. The results of Sections 2 and 3 show that there is seldom a need to use more than k=30 batches and that k=10 batches contain almost all the information in the data regardless of the number of observations n. Thus the practitioner should seldom exert much effort to increase the number of batches when k>30, and hardly ever when k>60.

The implication for researchers interested in constructing algorithms for determining batch sizes is to place less emphasis on obtaining very large numbers of batches. Four batching algorithms have appeared in the literature: Law and Carson [7], Mechanic and McKay [10], Fishman [2], and Schriber and Andrews [12]. We discuss the implications of the results of Sections 2 and 3 on each.

Law and Carson require a minimum of k=40 batches. This research shows that a minimum of k=10 batches will be almost as effective and result in shorter runs. The algorithm could be modified to try initially for 40 batches, but before doubling n, k=20 batches could be checked. If k=20, fails, then try k=10. If k=10 batches fail, then double the sample size n, keeping k=10, since it appears that the first constraint

above applies. Similar comments apply to Mechanic and McKay who use a minimum of k=25 batches.

Fishman's algorithm requires $k \geq 8$, not because of considerations concerning the interval, but because the test used to detect correlation between the batch means fails for small values of k. This algorithm begins with k=n and iteratively halves k. Samples of size n=2048, 4096, 8192 and 16384 are used for experimenting with the algorithm, and n=111, 716 is discussed as being necessary in not unreasonable situations. The results of Sections 2 and 3 indicate the initial value of k could be substantially smaller with almost no deterioration in the confidence intervals.

Schriber and Andrews modify Fishman's algorithm by considering every possible batch size yielding $k \geq 8$ and selecting the value of k corresponding to the test statistic least indicating correlation between the batch means. Their algorithm could be modified to consider all possible batch numbers between eight and some value between thirty and sixty. Rather than selecting the number of batches with the test statistic value closest to zero, the algorithm could be modified to consider the advantages of larger values of k. For example, if k=8 is indicated, but k=30 also easily passes the correlation test, then k=30 should be considered because of its better properties.

Another implication concerns calculating confidence intervals on the mean based on the use of independent replications, for which notations similar to batch means may be defined. Let $\bar{X}_i$ denote the sample average of the i^{th} simulation run having m observations, i=1, 2, ..., k. Then the point estimate of μ resulting from the k runs is $\bar{\bar{X}} = \sum_{i=1}^{k} \bar{X}_i/k$

$= \sum_{i=1}^{n} X_i/n$. A confidence interval may be calculated based on $\bar{X}$ and S_k exactly as with batch means. For a detailed comparison of replications and batch means, see Law [6].

Since each replication has an associated overhead of initializing the run and removing the initial transient effects, k=2 replications have an advantage compared to using more replications. The m=n/2 observations per run give $\bar{X}_1$ and $\bar{X}_2$ the best chance of being normally distributed. Independence is guaranteed for all values of k by using different random number seeds. Therefore k=2 minimizes computation and has the best chance of satisfying the assumptions necessary for obtaining the desired level of coverage. However, the resulting confidence intervals have a larger expected half length than when larger values of k are used. Thus a trade off between initialization cost and information loss must be made, just as with batch means.

The results of Sections 2 and 3 imply that every effort should be made to use at least eight to ten replications, but that little gain results from using many more than ten or twenty replications. Since initial transients are often a major factor when using independent replications, our recommendation is to use more than ten batches only when the cost of dealing with the initial transient effect is very small. This is a very general recommendation, but hopefully the quantification of effects in Sections 2 and 3 will be useful for practitioners when making the tradeoff between few and many replications.

A final implication concerns the lack of knowledge about k*, the number of batches for which nonnormality and dependence of the batch means are negligible. In the analysis of Sections 2 and 3 we assumed

that k* batches were sufficient to provide normality and independence exactly. Since in most simulations, some violation of the assumptions occurs for all batch sizes, there is some advantage to using smaller numbers of batches which is not reflected in the analysis. That advantage is that the smaller number of batches will more closely satisfy normality and independence, thereby yielding more exact coverage probabilities for the mean.

5. COMPARING CONFIDENCE INTERVAL PROCEDURES

The analysis of coverage functions in Section 3 leads to a general method for comparing confidence interval procedures. Just as two tests of hypothesis can be compared by calculating operating characteristic curves, procedures for calculating confidence intervals can be compared by empirically estimating the coverage function for various values of α and μ_1. Before stating the empirical procedure explicitly, first consider Figure 4, which summarizes the information in Figures 1, 2, and 3 for k=10 and k=∞. Contours of the coverage probability β are plotted as functions of α and δ. (Recall that δ and μ_1 differ only in location and scaling, so alternatively the δ axis may be thought of a rescaled μ_1 axis.) The solid curves corresponding to k = ∞ are all lower than the dashed curves corresponding to k=10, indicating that the larger numbers of batches are preferable, as long as the normality and independence assumptions hold. Also as long as these two assumptions hold, the contour curves intersect the α axis at $1-\beta$. This property corresponds to the coverage function described by Schruben [13], who suggests that confidence interval procedures be compared by empirically determining the probability of coverage of the mean as a function of α. Thus the

suggestion made here is to generalize Schruben's coverage function to be a function of μ_1 as well as α.

The Monte Carlo estimation of the coverage function is straightforward. Let α_j, j=1, 2, ..., J, denote the α values of interest, such as .2, .1, .05, and .01. Let β_ℓ, ℓ=1, 2, ..., L, denote the β values of interest, say .7, .8, .9, and .95. Perform R replications. In replication i, calculate the J confidence intervals (V_{ij}, W_{ij}) using the procedure of interest, and store these values either explicitly or in 2J histograms. Then for j=1, 2, ..., J and ℓ=1, 2, ..., L; estimate $\mu_{j\ell}$ such that $P(V_j \leq \mu_{j\ell} \leq W_j) = \beta_\ell$, where V_j and W_j are random variables denoting the lower and upper bounds of the confidence interval. The confidence contour corresponding to each β_ℓ can be plotted using the points $(\mu_{j\ell}, \alpha_j)$, j=1, 2, ..., J. Note that there are two values of $\mu_{j\ell}$, one less that μ and one greater than μ, as shown in Figure 5, but that symmetry allows Figure 4 to show only the greater value.

The estimation of $\mu_{j\ell}$ involves the estimation of quantiles, which is a bit more involved than the estimation of fractiles. If the results are not to be plotted, the above procedure can be modified to simply increment a counter c_{jm} for each replication that confidence interval j covers μ_m, where μ_m, m=1, 2, ..., M, denote the values of μ_1 of interest. Then the estimator for β_{jm}, the probability of a $(1-\alpha_j)100\%$ confidence interval covering μ_m, is c_{jm}/R.

APPENDIX A

We derive here the bias ratios r, the expected half lengths $E\{H_k\}$, and the standard deviations of the half lengths $\sqrt{V\{H_k\}}$ needed for Table 1. When $k \leq k^*$, the assumptions of independence and normality of the batch means are satisfied. Then S_k has a chi distribution with mean $E\{S_k\} = \sigma_k (2/(k-1))^{1/2} \Gamma(k/2) / \Gamma((k-1)/2)$, where σ_k is the standard deviation of the k batch means and $\Gamma(\cdot)$ denotes the gamma function. Also when $k \leq k^*$, $\sigma_k/\sqrt{k} = \sqrt{V\{X\}}$. Since by definition $r = E\{S_k\sqrt{k}\}/\sqrt{V\{X\}}$, we have $r = (2k/(k-1))^{1/2} \Gamma(k/2) / \Gamma((k-1)/2)$, which is dimensionless.

The expected half length is $E\{H_k\} = t_{\alpha/2,k-1} E\{S_k\}/\sqrt{k}$. From the definition of r, we have directly that $E\{H_k\} = t_{\alpha/2,k-1} r$ in units of $\sqrt{V\{X\}}$.

The variance of the half length is $V\{H_k\} = E\{H_k^2\} - E^2\{H_k\} = t^2_{\alpha/2,k-1} E\{S_k^2\} / k - (t_{\alpha/2,k-1} r \sqrt{V\{X\}})^2$. Recalling that $E\{S_k^2\} = \sigma_k^2$ for all $k \leq k^*$ and taking the square root to obtain the standard deviation yields $\sqrt{V\{H_k\}} = t_{\alpha/2,k-1} (1-r^2)^{1/2}$ in units of $\sqrt{V\{X\}}$.

APPENDIX B

Figures 1, 2, and 3 show the probability of covering μ_1 as a function of μ, n, k, σ^2. These probabilities may be derived by noting that the probability of coverage $P(|\bar{X}-\mu_1| \leq H_k)$ is equal to

$$P\left(-t_{\alpha/2,k-1} \leq \frac{\frac{\bar{X}-\mu}{\sigma_k/\sqrt{k}} - \frac{\mu-\mu_1}{\sigma_k/\sqrt{k}}}{s_k/\sigma_k} \leq t_{\alpha/2,k-1}\right) \qquad (1)$$

where σ_k is the standard deviation of the k batch means. For all $k \leq k^*$, $(\bar{X}-\mu)/(\sigma_k/\sqrt{k})$ is a standard normal random variable and the batch means are independent. Therefore the entire center expression in (1) follows a noncentral Student's t distribution with k-1 degrees of freedom and noncentrality parameter $\delta_k=(\mu-\mu_1)/(\sigma_k/\sqrt{k})$. Again since $k \leq k^*$, $\sigma_k/\sqrt{k} = \sqrt{V(\bar{X})}$, yielding $\delta_k=(\mu-\mu_1)/\sqrt{V(\bar{X})}$. Using the expression for $V(\bar{X})$ in the introduction, we have $\delta_k=(\mu-\mu_1)\ (n\ /c\sigma^2))^{1/2}$. Since the probability of coverage is the same for both $\pm\delta_k$, Figures 1, 2, and 3 are drawn using $|\delta_k|$ to save space. The required noncentral t values may be found in Owen [11][1].

[1] This research was supported by the Naval Analysis Program of the Office of Naval Research under Contract N00014-C-79-0832. This paper benefited from discussions with Andy Andrews, Averill Law, Alan Pritsker, Steve Roberts, Lee Schruben, Kathy Stecke, and Jim Swain.

REFERENCES

1. Conway, R.W., "Some Tactical Problems in Digital Simulation," Management Science, Vol. 10, No. 1 (1963), pp. 47-61.

2. Fishman, G.S., "Grouping Observations in Digital Simulation," Management Science, Vol. 24, No. 5 (1978), pp. 510-521.

3. ----------, Principles of Discrete Event Simulation, Wiley, New York, 1978.

4. Kleijnen, J.P.C., Statistical Techniques in Simulation, Part II, Marcel Dekker, New York, 1975.

5. Lavenberg, S.S. and Sauer, C.H., "Sequential Stopping Rules for the Regenerative Method of Simulation," IBM J. Res. Development, (November 1977), pp. 545-558.

6. Law, A.M., "Confidence Intervals in Discrete Event Simulation: A Comparison of Replication and Batch Means," Naval Res. Logist. Quart., Vol. 24, No. 4 (1977), pp. 667-678.

7. ---------- and Carson, J.S., "A Sequential Procedure for Determining the Length of a Steady State Simulation," Operations Research, Vol. 27, No. 5 (1979), pp. 1011-1025.

8. --------- and Kelton, W.D., "Confidence Intervals for Steady-state Simulations, I: A Survey of Fixed Sample Size Procedures," Technical Report 78-5, Department of Industrial Engineering, University of Wisconsin, Madison, Wisconsin (1978).

9. -------- and --------, "Confidence Intervals for Steady-state Simulations, II: A Survey of Sequential Procedures," Technical Report 78-6, Department of Industrial Engineering, University of Wisconsin, Madison, Wisconsin (1978).

10. Mechanic, H. and McKay, W., "Confidence Intervals for Averages of Dependent Data in Simulations II," Technical Report ASDD 17-202, IBM Corp., Yorktown Heights, New York (1966).

11. Owen, D.B., "The Power of Student's t-Test," J. Amer. Statist. Assoc., Vol. 60, No. 309 (1965), pp. 320-333.

12. Schriber, T.J. and Andrews, R.W., "Interactive Analysis of Simulation Output by the Method of Batch Means," Proc. Winter Sim. Conf., pp. 513-525 (1979).

13. Schruben, L.W., "A Coverage Function for Interval Estimators of Simulation Response," Management Science, Vol. 26, No. 1 (1980), pp. 18-27.

		$\alpha = .10$			$\alpha = .05$			$\alpha = .01$		
k	Bias Ratio r	$t_{\alpha/2,k-1}$	$E\{H_k\}$	$\sqrt{V\{H_k\}}$	$t_{\alpha/2,k-1}$	$E\{H_k\}$	$\sqrt{V\{H_k\}}$	$t_{\alpha/2,k-1}$	$E\{H_k\}$	$\sqrt{V\{H_k\}}$
2	.7979	6.314	5.04	3.81	12.706	10.1	7.66	63.657	50.8	38.4
3	.8862	2.920	2.59	1.35	4.303	3.81	1.99	9.925	8.80	4.60
4	.9213	2.353	2.17	.91	3.182	2.93	1.24	5.841	5.38	2.27
5	.9400	2.132	2.00	.73	2.776	2.61	.95	4.604	4.33	1.57
6	.9515	2.015	1.92	.62	2.571	2.45	.79	4.032	3.84	1.24
10	.9727	1.833	1.78	.43	2.262	2.20	.52	3.250	3.16	.75
30	.9914	1.699	1.68	.22	2.045	2.03	.27	2.756	2.73	.36
61	.9959	1.671	1.66	.15	2.000	1.99	.18	2.660	2.65	.24
121	.9979	1.658	1.65	.11	1.980	1.98	.13	2.617	2.61	.17
∞	1.0	1.645	1.645	0.0	1.960	1.960	0.0	2.576	2.576	0.0

Table 1. The effect of number of batches, k, on the expected value $E\{H_k\}$ and standard deviation $\sqrt{V\{H_k\}}$ of the half width of confidence intervals on the mean.

Figure 1. Comparison, by number of batches k, of the probability of covering points $\delta\sqrt{V\{\bar{X}\}}$ from μ when $\alpha=.10$.

Figure 2. Comparison, by number of batches k, of the probability of covering points $\delta\sqrt{V\{\bar{X}\}}$ from μ when $\alpha=.05$.

Figure 3. Comparison, by number of batches k, of the probabilities of covering points $\delta\sqrt{V\{\bar{X}\}}$ from μ when $\alpha=.01$.

Figure 4. Comparison of k=10 batches with the limiting case $k=\infty$.

Figure 5. Cumulative distribution functions for the lower and upper confidence interval bounds for k=10 independent and normally distributed batch means, $\alpha=0.10$.

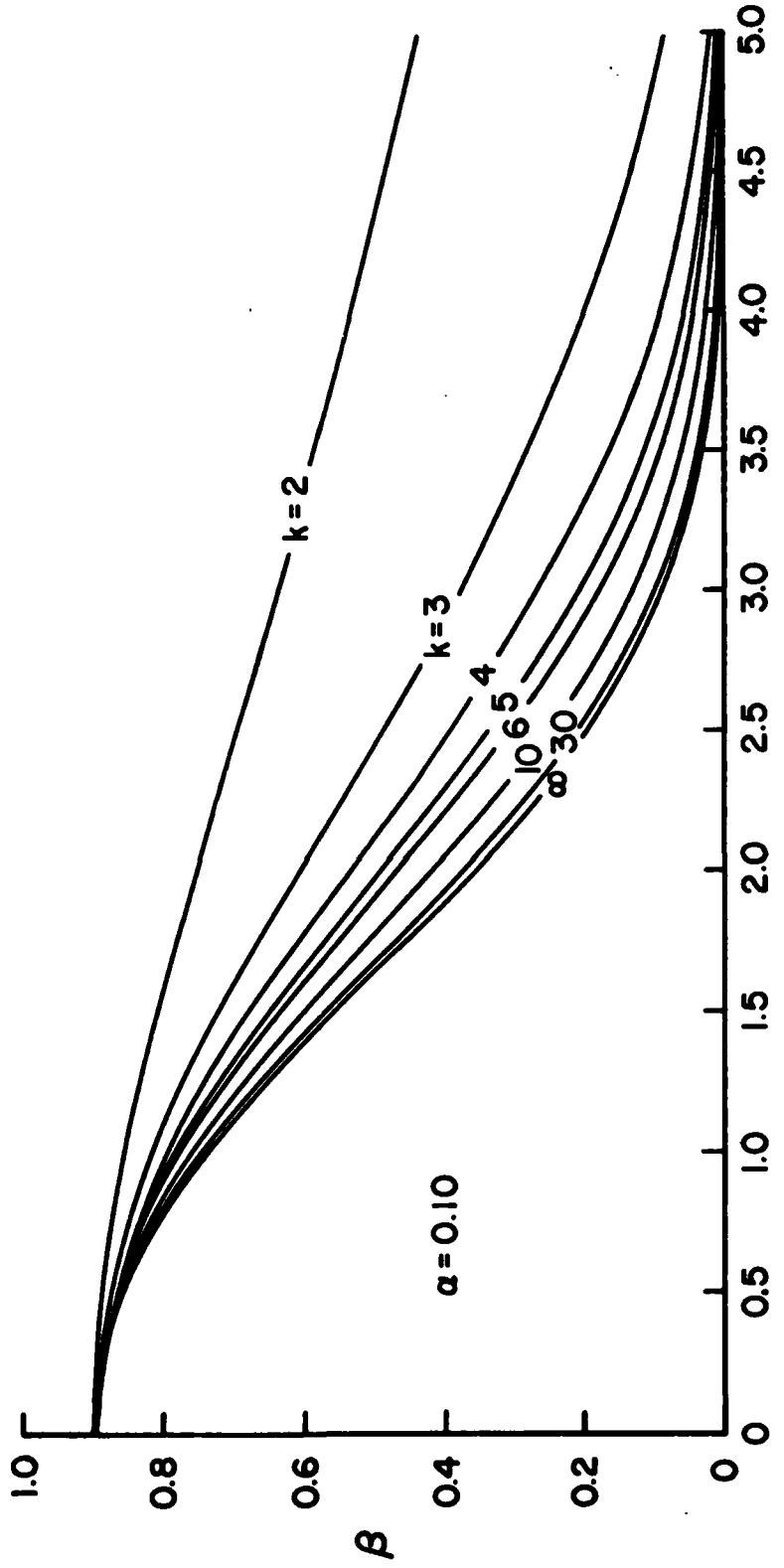


Figure 1. Comparison, by number of batches k, of the probability of covering points $\delta\sqrt{V\{\bar{X}\}}$ from μ when $\alpha=.10$.

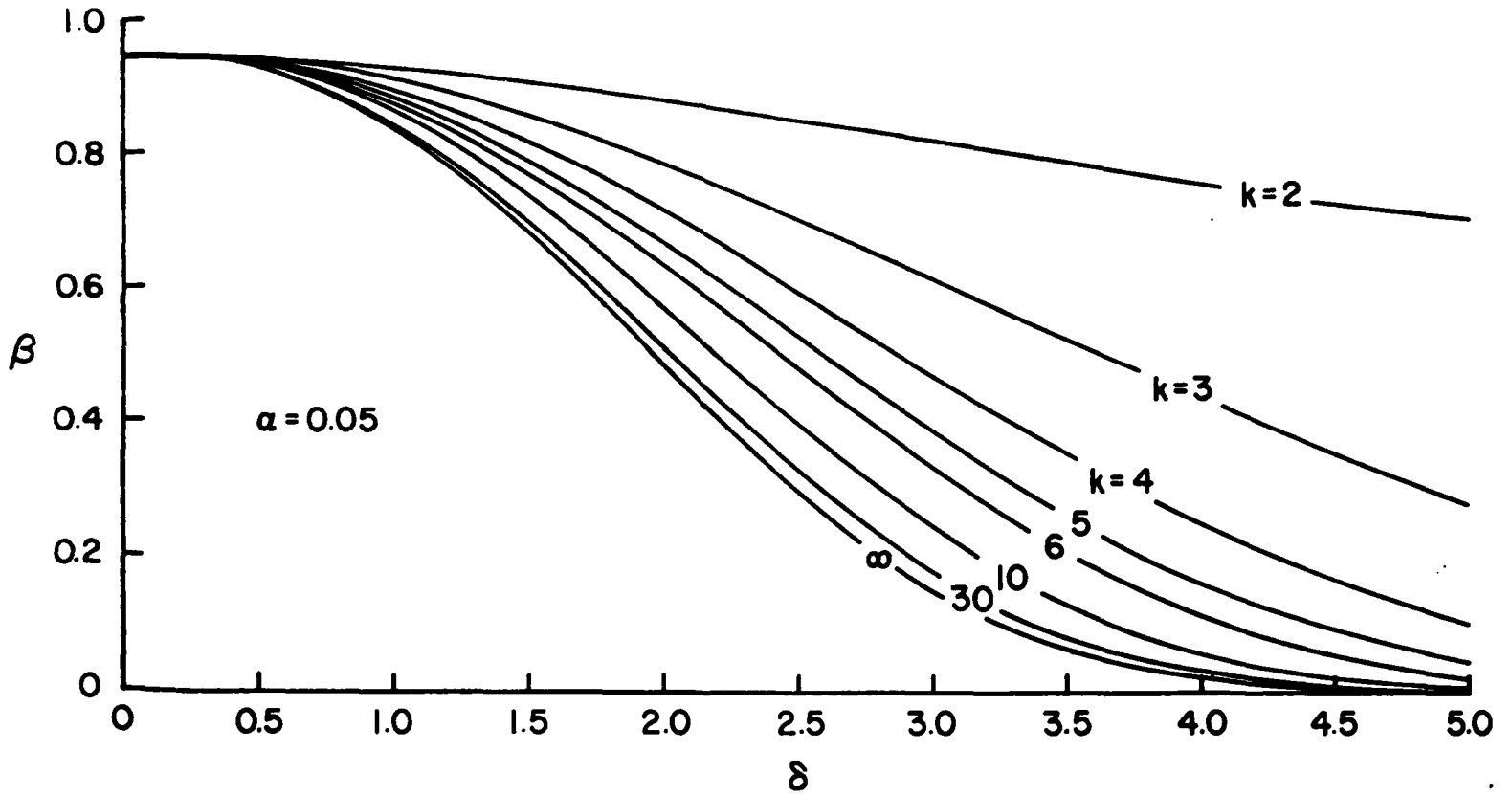


Figure 2. Comparison, by number of batches k, of the probability of covering points $\delta\sqrt{V(\bar{X})}$ from μ when α=.05.

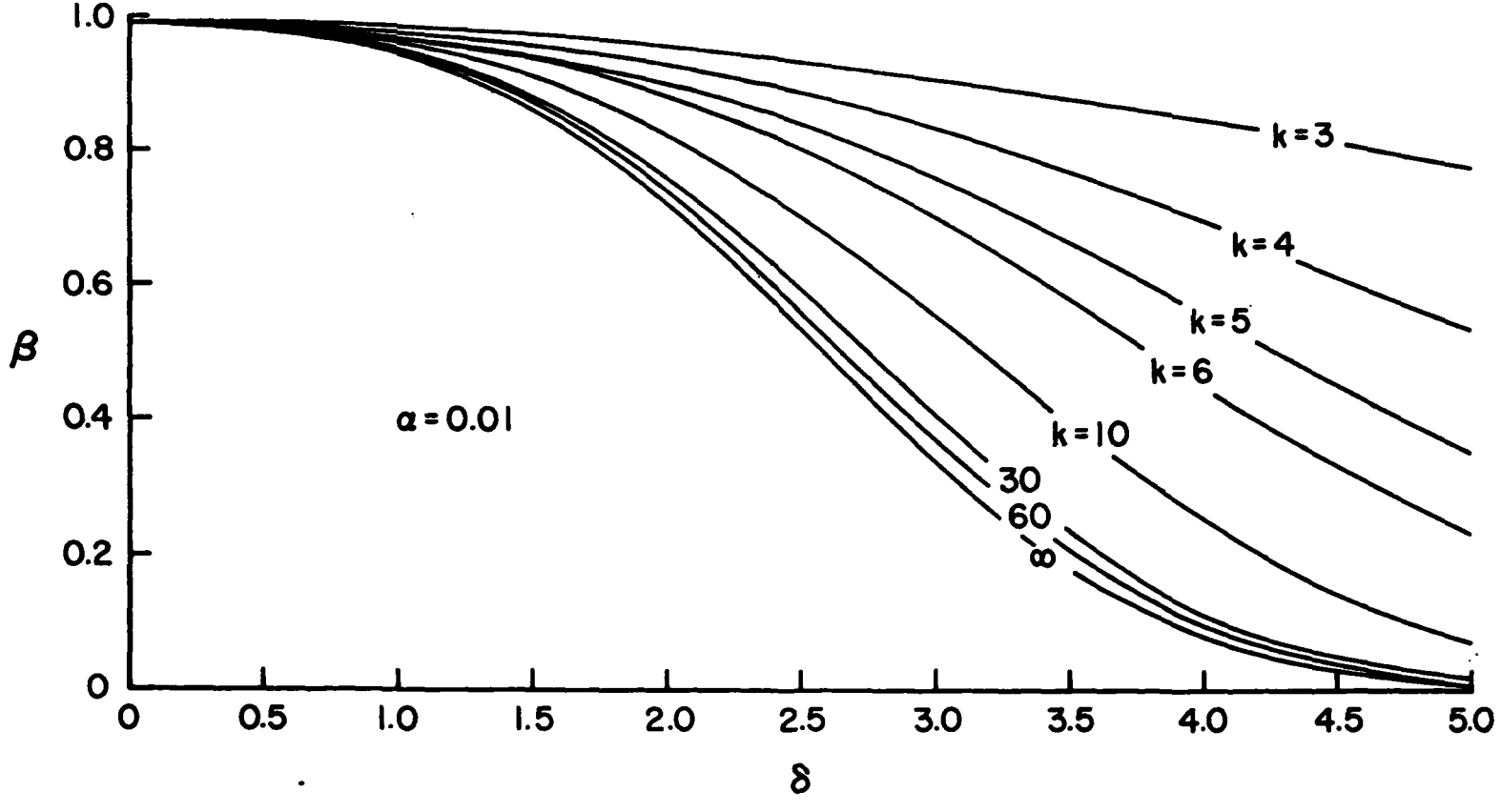


Figure 3. Comparison, by number of batches k, of the probabilities of covering points $\delta\sqrt{V\{\bar{X}\}}$ from μ when α=.01.

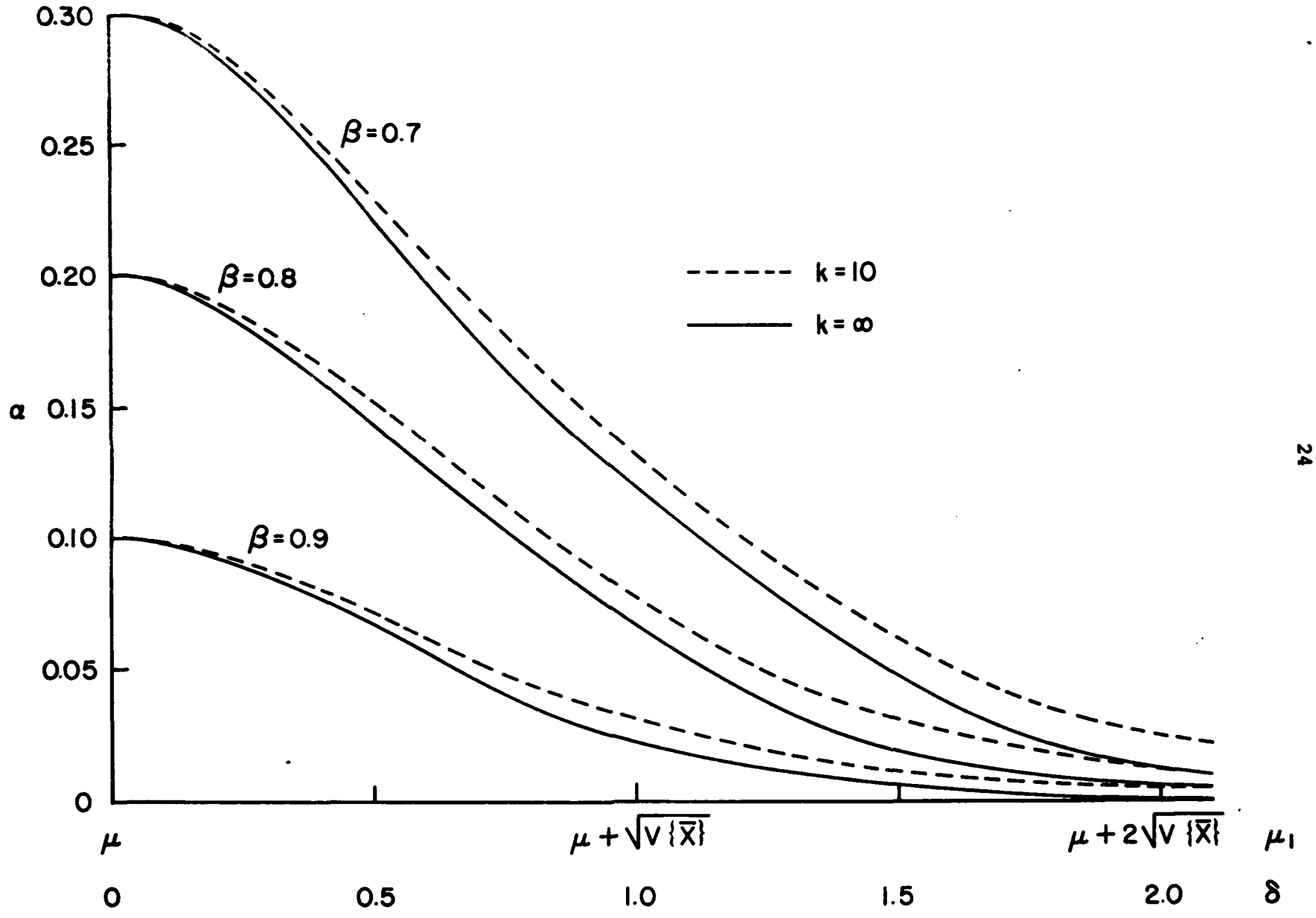


Figure 4. Comparison of k=10 batches with the limiting case k=∞.

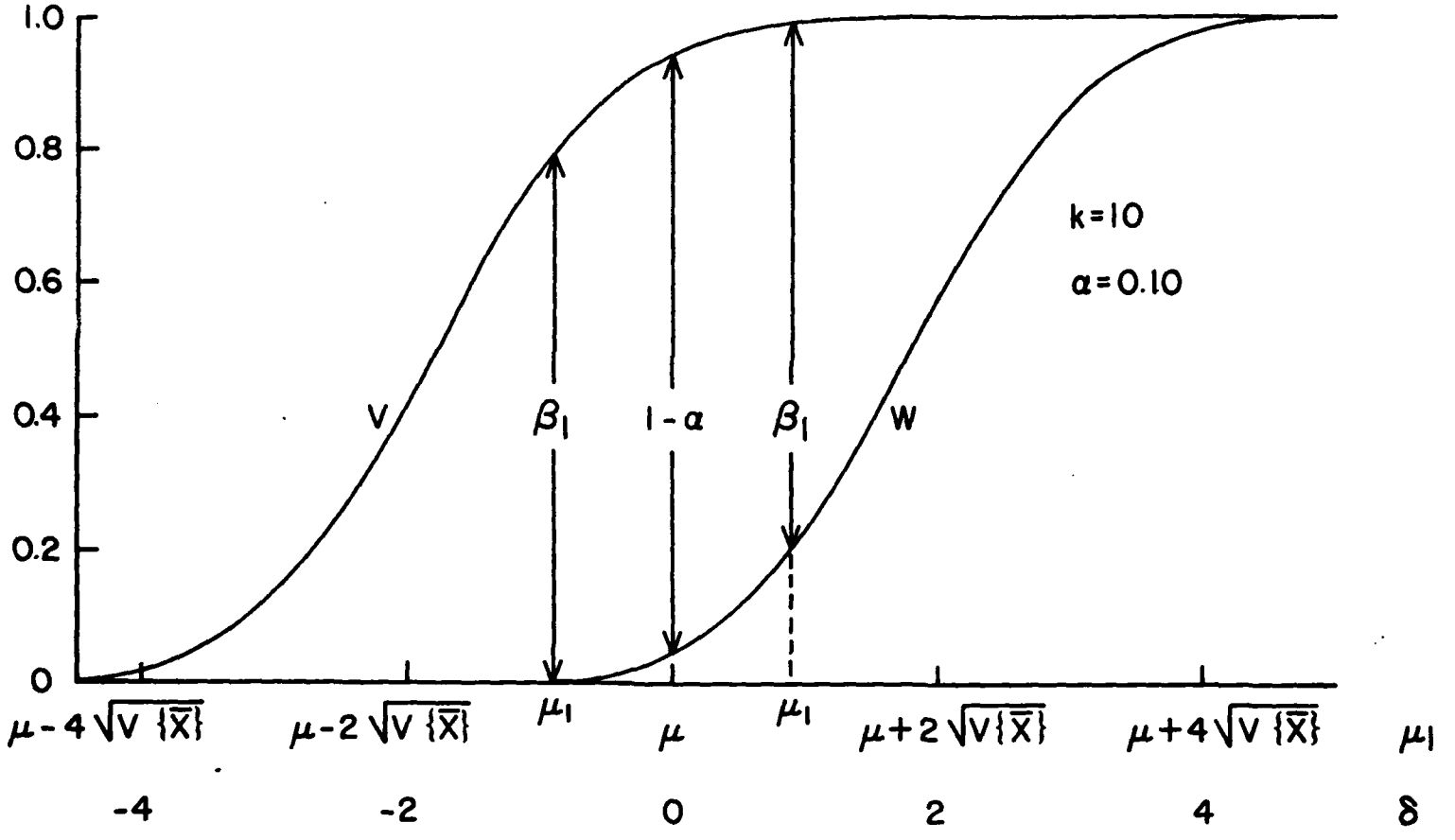


Figure 5. Cumulative distribution functions for the lower and upper confidence interval bounds for k=10 independent and normally distributed batch means, α=0.10.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: None	2. GOVT ACCESSION NO.: AD-A087 045
	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): Batch size effects in the analysis of simulation output.	5. TYPE OF REPORT & PERIOD COVERED: Technical report
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Bruce Schmeiser	8. CONTRACT OR GRANT NUMBER(s): N00014-79-C-0832
9. PERFORMING ORGANIZATION NAME AND ADDRESS: School of Industrial Engineering, Purdue University, West Lafayette, IN 47906	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 277-236 / 6-6-79 (431)
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Analysis Program, Office of Naval Research, Arlington, VA 22217	12. REPORT DATE: June 1980
	13. NUMBER OF PAGES: 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited.

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Monte Carlo
Simulation
Confidence interval
Batch means

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Batching is a commonly used method for calculating confidence intervals on the mean of a sequence of correlated observations arising from a simulation experiment. Several recent papers have considered the effects of using too many batches. The use of too many batches fails to satisfy assumptions of normality and/or independence, resulting in incorrect probabilities of the confidence interval covering the mean.

This paper considers the effects of using fewer batches than are necessary to satisfy normality and independence assumptions. Using too few batches

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract (continued)

results in 1) correct probability of covering the mean, 2) an increase in expected half length, 3) an increase in the standard deviation of the half length, and 4) an increase in the probability of covering incorrect values of the mean (analagous to Type II error in hypothesis testing). These effects, quantified here, are shown to be small when at least eight to ten batches are used, with least effect on confidence intervals having low confidence values. With the effects of using too few batches quantified, a simulation practitioner can make the trade-off between the ease of using very few batches with known independence and normality versus using a batching algorithm to squeeze some remaining information from the data. For researchers developing batching algorithms, the results are useful in selecting initial batch sizes. The results may also be useful in the context of using independent replications to establish confidence intervals on the mean.

Finally, some criteria and a procedure are suggested for Monte Carlo comparison of confidence interval procedures. These suggestions are not restricted to batch mean algorithms.


DATE
FILME
-8