

N-SKART: A NONSEQUENTIAL SKEWNESS- AND AUTOREGRESSION-ADJUSTED BATCH-MEANS PROCEDURE FOR SIMULATION ANALYSIS

Ali Tafazzoli

Metron Aviation, Inc.
45300 Catalina Ct. Suite 101
Dulles, VA 20166, U.S.A.

James R. Wilson

Edward P. Fitts Department of
Industrial & Systems Engineering
North Carolina State University
Raleigh, NC 27695, U.S.A.

ABSTRACT

We discuss N-Skart, a nonsequential procedure designed to deliver a confidence interval (CI) for the steady-state mean of a simulation output process when the user supplies a single simulation-generated time series of arbitrary size and specifies the required coverage probability for a CI based on that data set. N-Skart is a variant of the method of batch means that exploits separate adjustments to the half-length of the CI so as to account for the effects on the distribution of the underlying Student's t -statistic that arise from skewness (nonnormality) and autocorrelation of the batch means. If the sample size is sufficiently large, then N-Skart delivers not only a CI but also a point estimator for the steady-state mean that is approximately free of initialization bias. In an experimental performance evaluation involving a wide range of test processes and sample sizes, N-Skart exhibited close conformance to the user-specified CI coverage probabilities.

1. INTRODUCTION

A long-standing problem in the analysis of an output process generated by a steady-state simulation is the formulation of a robust and efficient procedure to construct a valid confidence interval (CI) for the steady-state process mean. Three primary problems impede successful analysis (Law 2007). The first problem is the influence of the simulation's initial condition on the output process, which often results in a transient that induces substantial bias in the sample mean of a simulation-generated time series. The second problem is the effect of correlation between successive simulation responses on the conventional estimator of the standard error of the sample mean, which often results in substantial underestimation of the variability of the sample mean. The third problem is the effect of highly nonnormal simulation responses on the distribution of the usual Student's t -ratio underlying conventional CIs for the steady-state mean. In many types of applications, these problems can give the user a misleading picture of the true accuracy and reliability of simulation-based results—that is, the bias and variance of the point estimator of the steady-state mean as well as the probability that the associated CI will cover the steady-state mean. A good CI procedure requires the solution of these three problems to provide the following:

- (a) an accurate point estimator of the steady-state mean that is approximately free of initialization bias;
- (b) a sufficiently stable estimator of the standard error of the point estimator (a) that adequately accounts for any correlation among the simulation responses used in computing the point estimator; and
- (c) a suitable adjustment to the usual critical value of Student's t -distribution that adequately accounts for any departures from normality in the simulation responses used in computing the point estimator (a) and the standard error estimator (b).

Exploiting (a)–(c), we are then able to construct not only an accurate point estimator for the steady-state mean but also a CI estimator whose actual coverage probability is close to the user-specified nominal confidence level.

In this paper we discuss N-Skart, a new nonsequential procedure for steady-state simulation output analysis which can be considered as an extension of the classical method of nonoverlapping batch means. This procedure is designed for simulation experiments in which the size of the output data set is fixed because of a limited computing budget, a constraint on the time available for the user to complete the simulation study, or other restrictions that prevent the user from resuming the current run of the simulation model. Therefore, N-Skart is specifically designed for the situation in which the user merely supplies

a single simulation-generated time series of an arbitrary fixed length and requests a CI with a specific coverage probability based on all the available data.

The rest of this paper is organized as follows. Section 2 provides a brief overview of N-Skart. Section 3 contains a formal algorithmic statement of N-Skart. In Section 4 we present selected results from our experimental performance evaluation. In Section 5 we present our main conclusions and recommendations for future work. The slides for the oral presentation of this article are available online via www.ise.ncsu.edu/jwilson/files/wsc09nskart.pdf. N-Skart is a simplified version of Skart, a fully sequential procedure designed to deliver a CI for the steady-state mean that satisfies user-specified requirements concerning not only the CI's coverage probability but also the absolute or relative precision provided by its half-length. A complete discussion of N-Skart and Skart is given in Tafazzoli (2009).

2. OVERVIEW OF N-SKART

We begin by introducing some notation required to state the problem and to describe the operation of N-Skart. Let $\{X_i : i = 1, 2, \dots, N\}$ denote the output time series of length N generated by a single run of a nonterminating (infinite-horizon) probabilistic simulation. If the simulation is in steady-state operation, then the random variables $\{X_i\}$ will have the same steady-state marginal cumulative distribution function (c.d.f.) $F_X(x) = \Pr\{X_i \leq x\}$ for $i = 1, 2, \dots, N$ and for all real x .

Usually in a nonterminating simulation, we are interested in constructing point and CI estimators for some parameter of the steady-state c.d.f. $F_X(\cdot)$. In this article, we are primarily interested in estimating the steady-state mean, $\mu_X = E[X] = \int_{-\infty}^{\infty} x dF_X(x)$; and we limit the discussion to output processes for which $E[|X_i|^3] < \infty$ so that the marginal mean μ_X , marginal variance $\sigma_X^2 = \text{Var}[X_i] = E[(X_i - \mu_X)^2]$, and marginal skewness $\text{Sk}[X_i] = E\{[(X_i - \mu_X)/\sigma_X]^3\}$ are well defined. We also assume that the variance parameter

$$\gamma_X = \sum_{\ell=-\infty}^{\infty} \text{Cov}[X_i, X_{i+\ell}] = \sum_{\ell=-\infty}^{\infty} E[(X_i - \mu_X)(X_{i+\ell} - \mu_X)] \tag{1}$$

is positive and well defined in the sense that (1) is absolutely convergent.

To construct point and CI estimators for μ_X based on the time series $\{X_i : i = 1, \dots, N\}$, N-Skart addresses the start-up problem by successively applying the randomness test of von Neumann (1941) to spaced batch means with progressively increasing batch sizes and interbatch spacer sizes. When the randomness test is finally passed with a batch size m and spacer size dm for sufficiently large integers m and d (where $m \geq 1$ and $d \geq 0$), the data-truncation point (warm-up period) is defined by the initial spacer so that the initial dm observations are truncated (ignored) in calculating the point and CI estimators of μ_X . N-Skart addresses the normality problem by a modified Cornish-Fisher expansion for the classical batch-means Student's t -ratio that incorporates a term due to Willink (2005) accounting for any skewness in the set of truncated, nonspaced (adjacent) batch means that are finally delivered. N-Skart addresses the correlation problem by using an autoregressive approximation to the autocorrelation function of the delivered set of truncated, nonspaced batch means.

Beyond the data-truncation point dm , N-Skart computes the k' truncated, nonspaced batch means with batch size m ,

$$Y_j(m) = \frac{1}{m} \sum_{i=(d+j-1)m+1}^{(d+j)m} X_i \text{ for } j = 1, \dots, k'; \tag{2}$$

and then N-Skart computes the truncated grand mean of the batch means,

$$\bar{Y}(m, k') = \frac{1}{k'} \sum_{j=1}^{k'} Y_j(m). \tag{3}$$

Next N-Skart computes an asymptotically valid $100(1 - \alpha)\%$ skewness- and autoregression-adjusted CI for μ_X having the form

$$\left[\bar{Y}(m, k') - G(t_{1-\alpha/2, k''-1}) \sqrt{\frac{AS_{m, k''}^2}{k'}}, \bar{Y}(m, k') - G(t_{\alpha/2, k''-1}) \sqrt{\frac{AS_{m, k''}^2}{k'}} \right], \tag{4}$$

where

$$G(\zeta) \equiv \begin{cases} [\sqrt[3]{1 + 6\beta(\zeta - \beta)} - 1]/(2\beta), & \text{if } \beta = \hat{B}_{m,k''}/(6\sqrt{k'}) \neq 0, \\ \zeta, & \text{otherwise,} \end{cases} \quad (5)$$

and

$$\left(\begin{matrix} S_{m,k''}^2 \\ \hat{B}_{m,k''} \end{matrix} \right) = \left\{ \begin{matrix} \text{approx. unbiased est. of marginal} \left(\begin{matrix} \text{variance} \\ \text{skewness} \end{matrix} \right) \text{ of the truncated, nonspaced batch means (2)} \\ \text{computed from } k'' \text{ spaced batch means with batch size } m \text{ that are separated by spacers of size } dm, \end{matrix} \right\} \quad (6)$$

where for $q \in (0, 1)$, the quantity $t_{q,\nu}$ denotes the q quantile of Student's t -distribution with ν degrees of freedom. (Note that in Equation (5), the indicated cube root $\sqrt[3]{1 + 6\beta(\zeta - \beta)}$ is understood to have the same sign as the quantity $1 + 6\beta(\zeta - \beta)$.) Thus we see that $G(t_{1-\alpha/2,k''-1})$ and $G(t_{\alpha/2,k''-1})$ are skewness-adjusted quantiles of Student's t -distribution for the left and right half-lengths of the proposed CI; and the autoregression (correlation) adjustment A is applied to the sample variance $S_{m,k''}^2$ described in Equation (6) so as to compensate for any residual correlation between the truncated, nonspaced batch means (2) that are used to compute the truncated grand mean (3). The correlation adjustment A is computed as

$$A = [1 + \hat{\varphi}_{Y(m)}] / [1 - \hat{\varphi}_{Y(m)}], \quad (7)$$

where the standard estimator of the lag-one correlation of the truncated, nonspaced batch means is

$$\hat{\varphi}_{Y(m)} = \widehat{\text{Corr}}[Y_j(m), Y_{j+1}(m)] = \frac{1}{k'-1} \sum_{j=1}^{k'-1} [Y_j(m) - \bar{Y}(m, k')][Y_{j+1}(m) - \bar{Y}(m, k')] / S_{m,k'}^2, \quad (8)$$

and

$$S_{m,k'}^2 = \frac{1}{k'-1} \sum_{j=1}^{k'} [Y_j(m) - \bar{Y}(m, k')]^2 \quad (9)$$

denotes the usual sample variance of the truncated, nonspaced batch means defined by Equation (2).

3. DETAILED ALGORITHMIC STATEMENT OF N-SKART

Figure 1 depicts a high-level flowchart of N-Skart. To invoke this procedure, the following user-supplied inputs are required:

- A simulation-generated time series $\{X_i : i = 1, \dots, N\}$ of length at least 1,280 from which the corresponding steady-state mean μ_X is to be estimated;
- The desired CI coverage probability $1 - \alpha$, where $0 < \alpha < 1$.

A formal algorithmic statement of N-Skart for a data set of fixed size N is given in Figure 2. In the rest of this section, we explain the logic of the steps of N-Skart that are detailed in Figure 2.

First we discuss the initialization of N-Skart (Step [1] of Figure 2). Using the given time series $\{X_i : i = 1, \dots, N\}$, N-Skart first computes the sample skewness \hat{B} of the last 80% of the raw (unbatched) observations as follows:

$$\left. \begin{aligned} N_0 = [0.8N], \quad \bar{X} &\leftarrow \frac{1}{N_0} \sum_{i=N-N_0+1}^N X_i, \quad S^2 \leftarrow \frac{1}{N_0-1} \sum_{i=N-N_0+1}^N (X_i - \bar{X})^2, \\ \hat{B} &\leftarrow \frac{N_0}{(N_0-1)(N_0-2)} \sum_{i=N-N_0+1}^N (X_i - \bar{X})^3 / S^3. \end{aligned} \right\} \quad (10)$$

If $|\hat{B}| > 4.0$, then N-Skart sets the initial batch size m according to $m \leftarrow \min\{16, \lfloor N/1,280 \rfloor\}$. This extreme case only happens when the observations are highly nonnormal. Usually, we have $|\hat{B}| \leq 4.0$; and in this situation N-Skart assigns

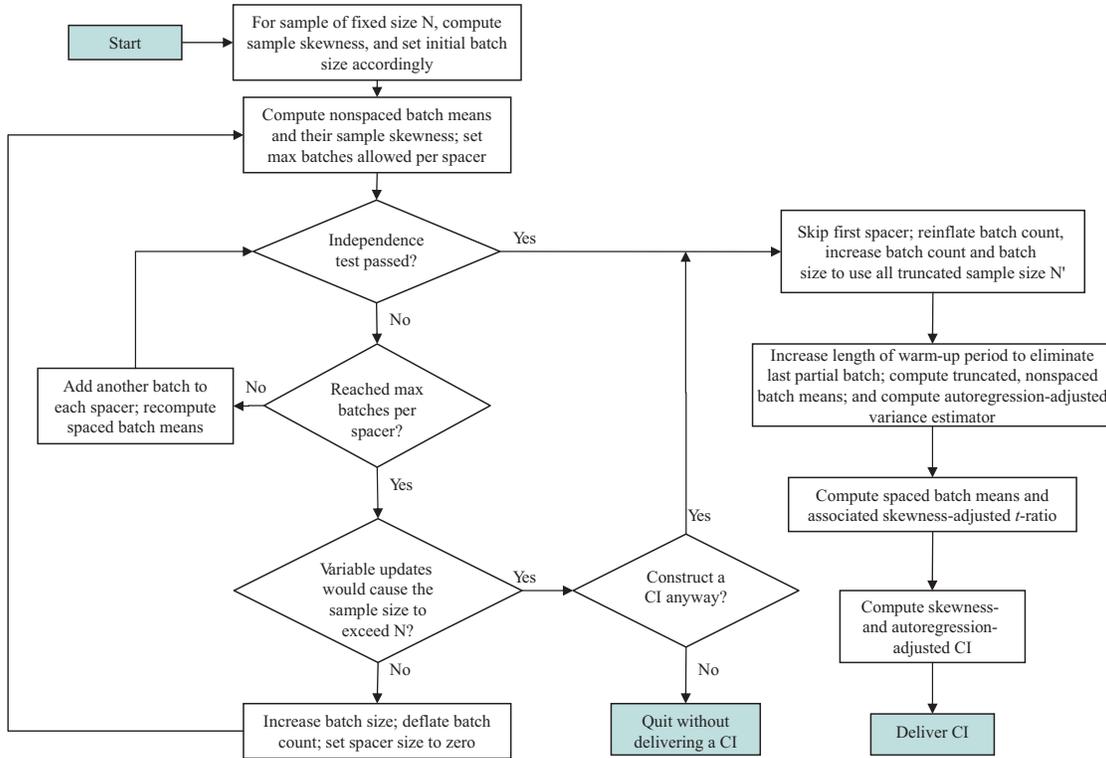


Figure 1: High-level flowchart of N-Skart

the initial batch size $m \leftarrow 1$. To complete its initialization step, N-Skart makes the following assignments: $d \leftarrow 0$ is the current number of batches per spacer; $d^* \leftarrow 10$ is the maximum number of batches allowed in a spacer in subsequent steps; $k \leftarrow 1,280$ is the current number nonspaced (adjacent) batch means; $\alpha_{\text{ran}} \leftarrow 0.20$ is the level of significance used on each iteration of the randomness test; and $b \leftarrow 0$ is the number of times the batch count has been deflated and the batch size has been inflated in successive iterations of the randomness test.

At key points in testing the spaced batch means for randomness (that is, in Step [2] of Figure 2), N-Skart must reassign d^* , the maximum number of batches per spacer, as a function of the skewness of the nonspaced batch means with the current batch size. First N-Skart computes the current set of nonspaced batch means with the latest batch size m according to

$$Y_j(m) = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} X_i \quad \text{for } j = 1, \dots, k; \quad (11)$$

then N-Skart computes the sample skewness of the last 80% of the batch means defined by Equation (11) to reduce the effect of any initialization bias that may be present. Specifically, N-Skart performs the following calculations:

$$\left. \begin{aligned} \ell &\leftarrow \lfloor 0.8k \rfloor, \quad \bar{Y}(m, \ell) \leftarrow \frac{1}{\ell} \sum_{j=k-\ell+1}^k Y_j(m), \quad S_{m,\ell}^2 \leftarrow \frac{1}{\ell-1} \sum_{j=k-\ell+1}^k [Y_j(m) - \bar{Y}(m, \ell)]^2, \\ \hat{B}_m &\leftarrow \frac{\ell}{(\ell-1)(\ell-2)} \sum_{j=k-\ell+1}^k [Y_j(m) - \bar{Y}(m, \ell)]^3 / S_{m,\ell}^3. \end{aligned} \right\} \quad (12)$$

If the estimated skewness \hat{B}_m of the batch means satisfies $|\hat{B}_m| > 0.5$, then N-Skart performs the reassignment $d^* \leftarrow 3$; otherwise N-Skart retains the assignment $d^* \leftarrow 10$. By doing this, N-Skart forces the randomness test to increase the batch size more frequently for highly skewed data sets.

Next N-Skart applies the randomness test of von Neumann (1941) to the current set of nonspaced batch means to determine the required batch count, batch size, spacer size, and data-truncation point beyond which all computed spaced

batch means are approximately independent not only of each other but also of the simulation's initial condition (see Step [3] of Figure 2). After each iteration of the randomness test, we let k' denote the current number of spaced batch means, where each spacer consists of d ignored batches. Each time the randomness test is failed, N-Skart adds an additional batch to each spacer and increments the number of batches per spacer, $d \leftarrow d + 1$ (up to the computed limit of d^* batches per spacer), and updates the count of spaced batch means, $k' \leftarrow \lfloor n/\{(d + 1)m\} \rfloor$; then N-Skart reapplies the randomness test to the new reduced set of k' spaced batch means.

If the randomness test is failed with a spacer consisting of d^* batches (see Step [4] of Figure 2), then some key status variables of the procedure must be suitably updated before reappling the randomness test. Because only a limited number of simulation-generated observations are available in N-Skart, a feasibility check is performed in this step of the procedure to determine if the updates to the batch size m and batch count k would cause the sample size n for the next iteration of the randomness test to exceed the available sample size N as follows—

- If $\lceil \sqrt{2}m \rceil \times \lceil 0.9k \rceil \leq N$, then we perform the following updates: the batch size is inflated according to $m \leftarrow \lceil \sqrt{2}m \rceil$; the batch count is deflated according to $k \leftarrow \lceil 0.9k \rceil$, where the assignment $b \leftarrow b + 1$ updates the total number of times the batch count is deflated in the randomness test (b is of course initialized to 0); the overall sample size is updated according to $n \leftarrow km$; and finally we take $d \leftarrow 0$ and $d^* \leftarrow 10$. Next N-Skart reperforms the following operations:
 - (i) the computation of the nonspaced batch means according to Equation (11) and the assignment of d^* based on the sample skewness of the batch means as computed from Equation (12) (that is, Step [2] in Figure 2); and
 - (ii) the randomness testing procedure outlined in the paragraph immediately preceding this paragraph (that is, Step [3] of Figure 2).
- On the other hand if $\lceil \sqrt{2}m \rceil \times \lceil 0.9k \rceil > N$, then N-Skart issues a warning to the user, stating that the randomness test could not be passed because of insufficient data. The warning also notes that if the user decides to continue the procedure under the given circumstances, then the delivered CI might not provide the target confidence level. Here the user has two choices:
 - (i) quit the procedure without delivering a CI; or
 - (ii) continue with construction of the requested CI by ignoring the warning.

Once the randomness test is passed (or bypassed owing to an inadequate data-set size and at the user's explicit request), N-Skart recomputes the truncated, spaced batch means with the final values of the warm-up period, the batch count, and the batch size (see Step [5] of Figure 2). N-Skart skips the first $w = dm$ observations in the warm-up period, so that

$$N' = N - w$$

approximately steady-state observations are available to build a CI for μ_X . Next the batch count k' is reinflated according to the formula

$$k' \leftarrow \min \{ \lceil k'(1/0.9)^b \rceil, k \}$$

to compensate for the total number of times the batch count was deflated in successive iterations of the randomness test. Then N-Skart computes a multiplier

$$f = \sqrt{N'/(k'm)}$$

to increase both the batch count k' and the batch size m so as to use all the available N' observations, subject to the constraint that $k' \leq 1,024$. Thus, N-Skart updates the count of truncated, nonspaced batch means according to

$$k' \leftarrow \min \{ \lfloor fk' \rfloor, 1,024 \};$$

and the associated batch size is updated as follows:

$$m \leftarrow \begin{cases} \lfloor fm \rfloor, & \text{if } k' < 1,024, \\ \lfloor N'/1,024 \rfloor, & \text{if } k' = 1,024. \end{cases}$$

Then, N-Skart computes k' nonspaced batch means for batches of size m according to

$$Y_j(m) \leftarrow \frac{1}{m} \sum_{i=N-(k'-j+1)m+1}^{N-(k'-j)m} X_i \quad \text{for } j = 1, \dots, k'; \quad (13)$$

so that there is no partial batch left at the end of the data set; and N-Skart adds the extra $N' - mk'$ observations (where $N' - mk' < m$) to the end of warm-up period,

$$w \leftarrow w + (N' - k'm),$$

so that the initial observations $\{X_i : i = 1, \dots, w\}$ are the only unused items in the entire data set of size N . This step enhances the removal of any transient effect, especially in the problem instances in which the provided sample size N is fairly small and the randomness test of N-Skart is bypassed at the user's request.

Next N-Skart computes the correlation adjustment to be used in computing the CI for μ_X (see Step [6] of Figure 2). N-Skart computes the following sample statistics from the final set of truncated, nonspaced batch means defined by Equation (13): the grand mean $\bar{Y}(m, k')$; the sample variance $S_{m,k'}^2$; and the lag-one correlation $\hat{\phi}_{Y(m)}$. Then the correlation adjustment A is computed according to Equation (7).

The final step of N-Skart is to construct the skewness- and correlation-adjusted CI for μ_X (see Step [7] of Figure 2). N-Skart makes separate adjustments to the classical batch-means CI based on the corresponding effects of nonnormality and correlation of the batch means on the distribution of the usual Student's t -ratio,

$$t_1 = \sqrt{k'} \left[\bar{Y}(m, k') - \mu_X \right] / S_{m,k'}, \quad (14)$$

that underlies the batch-means method. To do this, N-Skart must first compute approximately unbiased estimators of the marginal variance and skewness of the truncated, nonspaced batch means $\{Y_j(m) : j = 1, \dots, k'\}$ with the current batch size m . From all the individual observations in the current simulation-generated data set, N-Skart temporarily forms a set of approximately i.i.d. spaced batch means with batch size m , where the spacer size is the smallest multiple of m exceeding the size of the warm-up period. Let k'' denote the resulting number of spaced batch means. From this approximate random sample of size k'' , N-Skart computes $S_{m,k''}^2$ and $\hat{B}_{m,k''}$, the usual unbiased estimators of the associated marginal variance and skewness of batch means with batch size m as specified in Equations (17)–(19) in Figure 2.

For the skewness adjustment, N-Skart exploits the results of Willink (2005). To construct a CI for the mean μ of a nonnormal population based on a random sample of size n from that population, Willink (2005) derived the following modified t -statistic based on the sample mean \bar{X} , the sample standard deviation S , and the sample third central moment $\hat{\mu}_3$ computed from the given data set:

$$t_2 = \frac{(\bar{X} - \mu) + \hat{\mu}_3 / (6S^2n) + [\hat{\mu}_3 / (3S^4)](\bar{X} - \mu)^2 + [\hat{\mu}_3^2 / (27S^8)](\bar{X} - \mu)^3}{\sqrt{S^2/n}}, \quad (15)$$

which has approximately Student's t -distribution with $n - 1$ degrees of freedom under widely applicable conditions—provided that (15) is computed from observations that are independent and identically distributed (and hence uncorrelated). To obtain the skewness adjustment appropriate for N-Skart, we make the following substitutions in the numerator of Willink's modified t -statistic (15): (i) \bar{X} is replaced by $\bar{Y}(m, k')$; (ii) S^2 is replaced by $S_{m,k''}^2$; and (iii) $\hat{\mu}_3$ is replaced by $\hat{B}_{m,k''} S_{m,k''}^3$.

For the correlation adjustment, N-Skart makes the following substitution in the denominator of Willink's modified t -statistic (15): S^2 is replaced by $AS_{m,k''}^2$, where the correlation-adjustment factor A is computed according to Equations (7)–(9). Usually the batch-means process can be adequately modeled by an autoregressive-moving average (ARMA) process, at least for the purpose of estimating the autocorrelation structure of the batch means—see Box, Jenkins, and Reinsel (2008); Steiger et al. (2004, 2005a, 2005b); and Lada and Wilson (2007, 2008). In N-Skart, the autocorrelation adjustment A is applied to the variance estimator $S_{m,k''}^2$ to compensate for any residual correlation between the truncated batch means. For a detailed explanation of the correlation-adjustment used in N-Skart, see Appendix A of Tafazzoli (2009). If k' and k'' are sufficiently large, then we can treat $AS_{m,k''}^2/k'$ as an approximately unbiased estimator of $\text{Var}[\bar{Y}(m, k')]$ with $k'' - 1$ degrees of freedom. The substitutions of the last two paragraphs yield N-Skart's $100(1 - \alpha)\%$ skewness- and autoregression-adjusted CI for μ_X that is given by Equations (16)–(19) in Figure 2.

- [1] From the given sample data set of size N , compute the sample skewness \widehat{B} of the last 80% of the observations according to Equation (10). If $|\widehat{B}| > 4.0$, then set the initial batch size $m \leftarrow \min\{16, \lfloor N/1,280 \rfloor\}$; otherwise set $m \leftarrow 1$. Set the current number of batches in a spacer, $d \leftarrow 0$, and the maximum number of batches allowed in a spacer, $d^* \leftarrow 10$. Then divide the initial sample into $k \leftarrow 1,280$ nonspaced (adjacent) batches of size m . Set the randomness test size, $\alpha_{\text{ran}} \leftarrow 0.20$, and the number of times the batch count has been deflated in the randomness test, $b \leftarrow 0$.
- [2] Use Equation (11) to compute the current set of nonspaced batch means $\{Y_j(m) : j = 1, \dots, k\}$; and from the last 80% of $\{Y_j(m) : j = 1, \dots, k\}$, compute the sample skewness \widehat{B}_m as specified by Equation (12). If $|\widehat{B}_m| > 0.5$, then reassign the maximum number of batches per spacer according to $d^* \leftarrow 3$.
- [3] Apply the von Neumann test for randomness to the current set of k batch means with significance level α_{ran} .
- [3a] If the randomness test is passed, then set $k' \leftarrow k$ and go to [5]; otherwise go to [3b].
- [3b] Insert spacers each with m observations (one ignored batch) between the $k' \leftarrow \lfloor k/2 \rfloor$ remaining batches; assign the values of the k' spaced batch means; and set the total number of batches in a spacer, $d \leftarrow 1$.
- [3c] Apply the randomness test to the current set of k' spaced batch means with significance level α_{ran} . If the randomness test is passed, then go to [5]; otherwise go to [3d].
- [3d] If $d = d^*$ so that the current number of batches per spacer equals the maximum number of batches per spacer, then go to [4]; else add another ignored batch to each spacer so that the total number of batches per spacer and the number of spaced batches are respectively updated according to

$$d \leftarrow d + 1 \text{ and } k' \leftarrow \lfloor n / \{(d + 1)m\} \rfloor,$$

respectively. Reassign the values of the k' spaced batch means, and go to [3c].

- [4] If $\lceil \sqrt{2}m \rceil \times \lceil 0.9k \rceil \leq N$, then update the batch size m , the total batch count k , the overall sample size n , and N-Skart's other status variables according to

$$m \leftarrow \lceil \sqrt{2}m \rceil, \quad k \leftarrow \lceil 0.9k \rceil, \quad n \leftarrow km, \quad d \leftarrow 0, \quad b \leftarrow b + 1, \text{ and } d^* \leftarrow 10;$$

and go to [2]. Otherwise, issue a warning that the randomness test could not be passed due to insufficient data, and ask if user wishes to continue. If the user chooses to continue with constructing a CI, then go to [5]; otherwise quit the procedure without delivering a CI.

- [5] Skip the first $w = d \times m$ observations in the overall sample of size N , and take $N' = N - w$. First reinflate the batch count $k' \leftarrow \min\{\lceil k'(1/0.9)^b \rceil, k\}$; then compute the additional inflation factor $f \leftarrow \sqrt{N'/(k'm)}$ for both the batch size and batch count, and reset the truncated batch count $k' \leftarrow \min\{\lfloor fk' \rfloor, 1,024\}$. If $k' < 1,024$, then reset the batch size according to $m \leftarrow \lfloor fm \rfloor$; otherwise take $m \leftarrow \lfloor N'/1,024 \rfloor$ because the maximum of $k' = 1,024$ truncated, nonspaced batch means has been reached. Compute the corresponding truncated, nonspaced batch means $\{Y_j(m) : j = 1, \dots, k'\}$ according to Equation (13) so that there is no partial batch left at the end of the data set.
- [6] From the current set of truncated, nonspaced batch means $\{Y_j(m) : j = 1, \dots, k'\}$ defined by Equation (13), compute the following: the grand mean $\bar{Y}(m, k')$ defined by Equation (3); the sample variance $S_{m, k'}^2$ defined by Equation (6); the sample estimator $\widehat{\varphi}_{Y(m)}$ of the lag-one correlation of the truncated, nonspaced batch means defined by Equation (8); and finally the correlation adjustment A defined by Equation (7).

Figure 2: Algorithmic statement of N-Skart

- [7] Compute the correlation-adjusted $100(1-\alpha)\%$ CI for μ_X using the skewness-adjusted critical values $G(t_{1-\alpha/2, k''-1})$ and $G(t_{\alpha/2, k''-1})$ of Student's t -ratio for $k'' - 1$ degrees of freedom,

$$\left[\bar{Y}(m, k') - G(t_{1-\alpha/2, k''-1}) \sqrt{\frac{AS_{m, k''}^2}{k'}}, \bar{Y}(m, k') - G(t_{\alpha/2, k''-1}) \sqrt{\frac{AS_{m, k''}^2}{k'}} \right], \quad (16)$$

where to evaluate (16) we must first compute spaced batch means with $d' = \lceil w/m \rceil$ batches per spacer so we have $k'' = 1 + \lfloor (k' - 1)/(d' + 1) \rfloor$ spaced batches of size m with corresponding spaced batch means $Y_j(m, d') \equiv Y_{(j-1)(d'+1)+1}(m)$ as defined in (13) for $j = 1, \dots, k''$ with grand mean

$$\bar{Y}(m, k'', d') \leftarrow \frac{1}{k''} \sum_{j=1}^{k''} Y_j(m, d')$$

and sample variance and sample skewness respectively given by

$$S_{m, k''}^2 \leftarrow \frac{1}{k'' - 1} \sum_{j=1}^{k''} [Y_j(m, d') - \bar{Y}(m, k'', d')]^2 \quad (17)$$

and

$$\mathcal{J}_{m, k'', d'} \leftarrow \frac{k''}{(k'' - 1)(k'' - 2)} \sum_{j=1}^{k''} [Y_j(m, d') - \bar{Y}(m, k'', d')]^3. \quad (18)$$

From the latter statistics (17) and (18), compute

$$\hat{B}_{m, k''} = \frac{\mathcal{J}_{m, k'', d'}}{S_{m, k''}^3}, \quad \beta \equiv \frac{\hat{B}_{m, k''}}{6\sqrt{k'}}, \quad \text{and} \quad G(\zeta) \equiv \begin{cases} [\sqrt[3]{1 + 6\beta(\zeta - \beta)} - 1]/(2\beta), & \text{if } \beta \neq 0, \\ \zeta, & \text{if } \beta = 0. \end{cases} \quad (19)$$

Deliver the CI (16) and stop.

Figure 2 (continued): Algorithmic statement of N-Skart

4. PERFORMANCE EVALUATION OF N-SKART

To examine the performance of N-Skart with respect to coverage probability and the mean and variance of the half-length of its CIs, we applied N-Skart to a set of test problems including processes resembling practical applications with realistic levels of complexity and processes exhibiting extremes of stochastic behavior. For each of the test processes, the steady-state mean is known; therefore for a given test process, we can compute the empirical coverage probabilities for the CIs delivered by each output procedure in order to evaluate the performance of the procedure and compare its performance with that of N-Skart. We used the following sample sizes in our experiments: 10,000; 20,000; 50,000; and 200,000. These particular values were singled out to evaluate the performance of N-Skart for what might be considered “very small,” “small,” “medium,” and “large” sample sizes.

Each experiment includes 1,000 independent replications of N-Skart applied to each test process; and on each replication, N-Skart delivered 90% and 95% CIs for the selected steady-state mean response. Beyond CI coverage probability, the performance of each output procedure is reported with respect to the following criteria:

- (a) average CI relative precision—that is, sample average taken over all 1,000 replications of N-Skart of the larger of the left and right CI half-lengths expressed as fraction of the magnitude of the CI midpoint;
- (b) average CI half-length—that is, the sample average taken over all 1,000 replications of N-Skart of the larger of the left and right CI half-lengths; and
- (c) variance of the CI half-length—that is, the sample average taken over all 1,000 replications of N-Skart of the larger of the left and right CI half-lengths.

The standard error of each CI coverage estimator for CIs with nominal coverage probability 90% is approximately 0.95%; and the standard error of each CI coverage estimator for CIs with nominal coverage probability 95% is approximately 0.69%.

We used the following test processes in part of our performance evaluation of N-Skart. For a description of all the test processes used in the full performance evaluation of N-Skart, see Tafazzoli (2009).

- ***M/M/1 Queue Waiting Times, 90% Server Utilization*** The test process $\{X_i\}$ is the sequence of waiting times in the *M/M/1* queue with an empty-and-idle initial condition, an interarrival rate of $\lambda = 0.9$ customers per time unit, and a service rate of $\mu = 1.0$ customers per time unit. In this system the steady-state server utilization is $\tau = 0.9$, and the steady-state expected waiting time is $\mu_X = 9.0$ time units.
- ***M/M/1 Queue Waiting Times, 80% Server Utilization*** The test process $\{X_i\}$ is defined in the same way as for the previous test process, except the interarrival rate is $\lambda = 0.8$ customers per time unit so that the steady-state server utilization is $\tau = 0.8$, and the steady-state expected waiting time is $\mu_X = 3.2$ time units.
- ***AR(1)-to-Pareto (ARTOP) Process*** The test process $\{X_i\}$ is generated from an underlying (or base) AR(1) process $\{Z_i : i = 1, 2, \dots\}$ with autoregressive parameter $\rho = 0.995$ and white-noise variance $1 - \rho^2 = 0.9975 \times 10^{-2}$ so that in steady-state the $\{Z_i\}$ are standard normal random variables with lag-one correlation 0.995. The corresponding observations $\{X_i : i = 0, 1, \dots\}$ of the target ARTOP process are generated from the Pareto c.d.f.

$$F_X(x) \equiv \Pr\{X \leq x\} = \begin{cases} 1 - (\xi/x)^\psi, & x \geq \xi, \\ 0, & x < \xi, \end{cases}$$

with location parameter $\xi = 1.0$ and shape parameter $\psi = 2.1$ as follows: $X_i = F_X^{-1}[\Phi(Z_i)] = \xi/[1 - \Phi(Z_i)]^{1/\psi}$ for $i = 0, 1, \dots$, where $\Phi(z) = \int_{-\infty}^z (2\pi)^{-1/2} e^{-\xi^2/2} d\xi$ for all real z denotes the c.d.f. of the standard normal distribution. This scheme provides a test process $\{X_i : i = 1, 2, \dots\}$ whose steady-state marginal distribution has mean, standard deviation, and skewness given by $\mu_X = 1.9091$, $\sigma_X = 4.1660$, and $\text{Sk}(X_i) = \infty$, respectively. By taking $Z_0 = 3.4$, we obtain the initial condition $X_0 = F_X^{-1}[\Phi(Z_0)] = 43.5689$, which induces a pronounced positive bias in this test process. Clearly this process also exhibits pronounced correlation among successive observations as well as severe nonnormality.

- ***M/M/1/LIFO Queue Waiting Times*** The test process $\{X_i\}$ is the sequence of queue waiting times for the *M/M/1/LIFO* queue, with customers in the queue being served in last-in-first-out (LIFO) order, a mean interarrival time of 1.0, a mean service time of 0.8, and an empty-and-idle initial condition. Thus in steady-state operation this system has a server utilization of $\tau = 0.8$ and a mean queue waiting time $\mu_X = 3.20$. This test process was selected mainly because in steady-state operation, batch means computed from the waiting times are highly skewed, even for batch sizes that are sufficiently large to ensure the batch means are nearly uncorrelated.

Table 1 shows the result of applying N-Skart to the selected test processes to construct nominal 90% and 95% CIs. The experimentation for each test problem included 1,000 independent replications of N-Skart.

The results in Table 1 and in Section 4.2 of Tafazzoli (2009) indicate that the coverage probabilities provided by N-Skart for the given sample sizes were close to their nominal levels in almost all test problems, except for the queue-waiting-time process in the *M/M/1/LIFO* queue and the ARTOP process, where N-Skart experienced some minor undercoverage for the sample sizes 10,000 and 20,000. The pronounced level of nonnormality and stochastic dependence exhibited by the *M/M/1/LIFO* and ARTOP processes prevented N-Skart from working effectively with such unrealistically small sample sizes as 10,000 and 20,000. In general, we concluded that N-Skart performed better when it was applied to processes with limited marginal skewness. In the cases of the *M/M/1/LIFO* queueing system and the ARTOP process, when the sample size was small, the batch size could not increase sufficiently to reduce the batch-means skewness to a reasonable level. It should be mentioned here that in all the experimentation reported in Table 1, we simply ignored the warning message issued by N-Skart for test problems in which the randomness test could not be passed due to insufficient data; and we requested that N-Skart deliver a CI on all 1,000 independent replications of each test problem.

Table 1: Performance of N-Skart for selected test problems based on 1,000 replications

Confidence Level	Performance Measure	M/M/1, 90% Utilization				M/M/1, 80% Utilization			
		10,000	20,000	50,000	200,000	10,000	20,000	50,000	200,000
90%	CI coverage	87.60%	88.40%	90.30%	90.00%	90.40%	90.70%	90.40%	89.60%
	Avg. rel. precision	33.32%	25.77%	17.66%	8.64%	20.57%	14.77%	9.23%	4.39%
	Avg. CI half-length	3.1309	2.4018	1.6195	0.7791	0.8395	0.5972	0.3702	0.1757
	Var. CI half-length	3.7402	2.1298	0.4448	0.0542	0.1038	0.0522	0.0115	0.0008
95%	CI coverage	92.20%	93.10%	94.00%	94.90%	94.90%	95.80%	94.20%	94.90%
	Avg. rel. precision	39.79%	30.93%	20.83%	10.46%	24.06%	17.91%	10.85%	5.28%
	Avg. CI half-length	3.7681	2.8757	1.8963	0.9461	0.9764	0.7231	0.4363	0.2114
	Var. CI half-length	4.847	2.2475	0.5176	0.0582	0.1076	0.0503	0.015	0.0012
Confidence Level	Performance Measure	ARTOP				M/M/1/LIFO			
		10,000	20,000	50,000	200,000	10,000	20,000	50,000	200,000
90%	CI coverage	83.10%	85.30%	87.40%	88.80%	84.50%	87.20%	88.50%	89.00%
	Avg. rel. precision	26.97%	22.03%	14.69%	9.17%	17.76%	13.54%	8.89%	4.40%
	Avg. CI half-length	0.5632	0.4505	0.2878	0.179	0.5757	0.4368	0.2855	0.1411
	Var. CI half-length	0.405	0.2802	0.0587	0.049	0.0534	0.0291	0.006	0.0008
95%	CI coverage	88.40%	90.60%	91.80%	93.60%	90.70%	92.70%	93.10%	95.10%
	Avg. rel. precision	33.15%	28.63%	23.08%	11.31%	21.87%	16.64%	10.84%	5.32%
	Avg. CI half-length	0.6715	0.5897	0.4816	0.2213	0.7094	0.5373	0.3481	0.1706
	Var. CI half-length	0.3698	0.9588	1.1747	0.1093	0.092	0.0526	0.0104	0.0013

To put some of the results in Table 1 into perspective, we note from Figures 2 and 4 of Fishman and Yarberr (1997) that when ABATCH was applied to waiting times in the $M/M/1$ queue with 90% server utilization to compute a nominal 90% CI for the steady-state mean waiting time, ABATCH delivered the following coverage probabilities over 1,000 independent replications of data sets with the following fixed sizes: (i) $N = 16,384$, 80%; (ii) $N = 65,536$, 83%; and (iii) $N = 262,144$, 86%. Clearly N-Skart outperformed ABATCH in this test process. Steiger (1999) also contains results on the performance of ABATCH in some of the test processes used in this article; but the results in Steiger (1999) are based on only 100 independent replications of ABATCH, and the sample sizes are based on a relative-precision stopping rule and thus are not fixed over all replications. It is nevertheless true that the performance of ABATCH reported in Steiger (1999) is closely similar to that reported in Fishman and Yarberr (1997); and on the basis of the results in Table 1 for N-Skart and the results reported in Fishman and Yarberr (1997) and Steiger (1999) for ABATCH, we concluded that N-Skart performed at least as well as ABATCH in all the given test problems. A definitive comparison of the performance of N-Skart with that of ABATCH and MSER-5 (Franklin and White 2008) is the subject of ongoing work.

In general, when we are working with N-Skart, a CI with abnormally large half-length or high relative precision should alert us regarding potential problems with the delivered CIs and a possible need for bigger sample size.

5. CONCLUSIONS AND RECOMMENDATIONS

In this paper we developed a new, completely automated nonoverlapping batch-means method, called N-Skart, for constructing a correlation- and skewness-adjusted CI for the steady-state mean of a simulation output process for handling the test problems in which the user supplies a single simulation-generated series of arbitrary length, and the user specifies the desired coverage probability for a CI based on that series. From the experimental results presented in Section 4, it is evident that N-Skart provides close conformance to the user specified CI coverage probabilities.

REFERENCES

- Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 2008. *Time series analysis: Forecasting and control*. 4th ed. Englewood Cliffs, NJ: Prentice Hall.
- Fishman, G. S., and L. S. Yarberr. 1997. An implementation of the batch means method. *INFORMS Journal on Computing* 9 (3): 296–310.

- Franklin, W. W., and K. P. White, Jr. 2008. Stationarity tests and MSER-5: Exploring the intuition behind mean-squared-error reduction in detecting and correcting initialization bias. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 541–546. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via www.informs-sim.org/wsc08papers/064.pdf [accessed July 17, 2009].
- Lada, E. K., and J. R. Wilson. 2007. SBatch: A spaced batch means procedure for simulation analysis. In *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 463–471. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via www.informs-sim.org/wsc07papers/055.pdf [accessed July 17, 2009].
- Lada, E. K., and J. R. Wilson. 2008. SBatch: A spaced batch means procedure for steady-state simulation analysis. *Journal of Simulation* 2:170–185.
- Law, A. M. 2007. *Simulation modeling and analysis*. 4th ed. New York: McGraw-Hill, Inc.
- Steiger, N. M. 1999. Improved batching for confidence interval construction in steady-state simulation. Doctoral dissertation, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available online via www.lib.ncsu.edu/etd/public/etd-19231992992670/etd.pdf [accessed May 3, 2009].
- Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman. 2004. Steady-state simulation analysis using ASAP3. In *Proceedings of the 2004 Winter Simulation Conference*, ed. R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 672–680. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via www.informs-sim.org/wsc04papers/081.pdf [accessed July 17, 2009].
- Steiger, N. M., E. K. Lada, and J. R. Wilson. 2005a. Performance evaluation of ASAP3 for steady-state output analysis. In *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 781–789. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via www.informs-sim.org/wsc05papers/090.pdf [accessed July 17, 2009].
- Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman. 2005b. ASAP3: A batch means procedure for steady-state simulation output analysis. *ACM Transactions on Modeling and Computer Simulation* 15 (1): 39–73.
- Tafazzoli, A. 2009. *Skart: A skewness- and autoregression-adjusted batch-means procedure for simulation analysis*. Ph.D. thesis, Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, North Carolina. Available via www.lib.ncsu.edu/theses/available/etd-01122009-153054/unrestricted/etd.pdf [accessed May 3, 2009].
- von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics* 12:367–395.
- Willink, R. 2005. A confidence interval and test for the mean of an asymmetric distribution. *Communications in Statistics—Theory and Methods* 34:753–766.

AUTHOR BIOGRAPHIES

ALI TAFAZZOLI is an Analyst at Metron Aviation, Inc. He is a member of IIE and INFORMS. His e-mail address is tafazzoli@metronaviation.com.

JAMES R. WILSON is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. He is a member of AAUW, ACM, and SCS; and he is a Fellow of IIE and INFORMS. His e-mail address is jjwilson@ncsu.edu, and his Web page is www.ise.ncsu.edu/jwilson.