2012

Finite-sample performance of absolute precision stopping rules

Singham, Dashi I.

INFORMS

INFORMS Journal on Computing, v. 24, no. 4, Fall 2012, pp. 624-635
http://hdl.handle.net/10945/47790

INFORMS Journal on Computing
Vol. 24, No. 4, Fall 2012, pp. 624–635
ISSN 1091-9856 (print) | ISSN 1526-5528 (online)

http://dx.doi.org/10.1287/ijoc.1110.0471

Finite-Sample Performance of Absolute Precision Stopping Rules

Dashi I. Singham
Operations Research Department, Naval Postgraduate School, Monterey, California 93943, dsingham@nps.edu

Lee W. Schruben
Department of Industrial Engineering and Operations Research, University of California at Berkeley, Berkeley, California 94720, lees@berkeley.edu

Absolute precision stopping rules are often used to determine the length of sequential experiments to estimate confidence intervals for simulated performance measures. Much is known about the asymptotic behavior of such procedures. In this paper, we introduce coverage contours to quantify the trade-offs in interval coverage, stopping times, and precision for finite-sample experiments using absolute precision rules. We use these contours to evaluate the coverage of a basic absolute precision stopping rule, and we show that this rule will lead to a bias in coverage even if all of the assumptions supporting the procedure are true. We define optimal stopping rules that deliver nominal coverage with the smallest expected number of observations. Contrary to previous asymptotic results that suggest decreasing the precision of the rule to approach nominal coverage in the limit, we find that it is optimal to increase the confidence coefficient used in the stopping rule, thus obtaining nominal coverage in a finite-sample experiment. If the simulation data are independent and identically normally distributed, we can calculate coverage contours analytically and find a stopping rule that is insensitive to the variance of the data while delivering at least nominal coverage for any precision value.

Key words: simulation; statistical analysis; design of experiments
History: Accepted by Marvin Nakayama, Area Editor for Simulation; received December 2009; revised June 2010, November 2010, April 2011; accepted April 2011. Published online in *Articles in Advance* August 25, 2011.

1. Introduction

Stopping rules are used to determine how many observations (replications, batch means, regenerative cycles, etc.) to collect in a sequential sampling experiment. These are typically used in sequential confidence interval procedures (CIPs) designed to help assess the risks in making a decision based on simulation results. A conventional method for deciding when to stop collecting observations is to estimate confidence intervals with nominal coverage probability, η, after each observation. If the half-width of a confidence interval drops below a threshold, δ, the experiment is stopped, and a confidence interval is reported. This is called an absolute precision stopping rule, which is used, for example, when δ is the minimal performance improvement required to make a proposed change worth its cost or to ensure that a system's performance is within a contracted tolerance. The confidence level η is a component of the risk taken in discontinuing the simulation experiment and making a decision. We will consider two-sided confidence intervals for the mean of some stochastic simulation output.

Although such sequential CIPs produce confidence interval estimates with appropriate half-widths, the intervals generated in practice typically do not cover the true parameter value as often as intended, underestimating risk. Stopping may occur when the sample variance is small, but the observations may not be centered near the true mean. Of course, coverage can be affected by many factors other than the stopping rule: these include the lack of independence or normality in the data, bias in the variance estimator, and output that is biased because of run initialization. In this paper, we focus on the loss of coverage as a result of the choice of stopping rule.

Chow and Robbins (1965) and Glynn and Whitt (1992) give conditions where the coverage of a CIP designed to estimate the sample mean approaches η as δ is decreased toward 0. Encouraged by this asymptotic result, methods have been proposed to find values of δ that are small enough to provide the desired coverage (see Heidelberger and Welch 1983, Law and Carson 1979). Sproule (1985) extends the work of Chow and Robbins (1965) to procedures designed to estimate the mean of U-statistics. We will see that such stopping rules cannot return the desired coverage. Some rules allow for the possibility of early stopping, which can lead to a loss in coverage, whereas other rules produce intervals wider than what is necessary, resulting in coverage higher than expected. Generally, a loss in coverage occurs even when very small values of δ are used.

Stopping rules are typically incorporated in CIPs that seek to return valid confidence intervals for data that may be dependent or nonnormal (Chen and Kelton 2007, Hoad et al. 2009, Steiger and Wilson 2002, Tafazzoli et al. 2011). Sequential ad hoc stopping rules are sometimes tailored to a specific set of simulation test models to provide better results. To study how the stopping rules affect coverage, we will need a more general framework to analyze the coverage of CIPs employing sequential stopping rules that is independent of any specific simulation model. This is necessary in order to isolate the effect of the stopping rule on the loss in coverage and to determine the actual coverage for practical (finite-sample) experiments.

Schruben (1980) introduces coverage functions as a way of evaluating the coverage performance of a CIP over the space of possible confidence coefficients. Schmeiser and Yeh (2002) introduce a metric for evaluating CIPs based on the deviations of coverage functions from nominal coverage. Here, we introduce coverage contours as a way of viewing coverage over the space of possible confidence coefficients η and precision values δ. We calculate the contours for the most basic type of absolute precision rule studied in the above references (where stopping occurs when the half-width of the confidence interval falls below the threshold precision value). Other variations on CIPs use relative precision rules, which involve stopping when the half-width is less than some fraction of the sample mean. These types of stopping rules are described in Law (2007). In this paper, we focus on absolute precision stopping rules.

We analytically derive the coverage contours for the case when the underlying data are independent and normally distributed, providing a basis for understanding coverage in the finite-sample domain, whereas much of the previous literature focused on asymptotic limits. We find that coverage can be determined by integrating over the range of possible values of the sample mean and sample variance as each observation is collected in order to find the distribution of the stopping time. Our results show that coverage is usually less than nominal for popular parameter choices (high values of η and low values of δ) even though the data are independent and identically distributed (i.i.d.) normal, providing evidence of the systematic bias inherent in these types of sequential stopping rules. For nonnormal data, these coverage contours can be estimated by applying stopping rules repeatedly to streams of simulated data. Simulation of stopping rules applied to i.i.d. normal data produces the same results as those calculated using our method (within some level of numerical error), encouraging us to use simulation to calculate coverage contours for nonnormal data if analytical methods are not tractable.

We also calculate the contours of the expected stopping time of absolute precision rules over the same space of η and δ. (For normally distributed data, we derive the distribution of the stopping time analytically.) These results, together with the coverage contours, can be used to define an optimal stopping rule that delivers nominal coverage with a minimal number of expected replications. Contrary to what is suggested by the asymptotic theory on stopping rules, we find that better results can be achieved by varying the confidence coefficient used in a CIP instead of the half-width threshold. We then find a stopping rule that is insensitive to the variance and delivers at least nominal coverage for any precision value when the data are normally distributed.

Currently, CIPs are tested by repeatedly applying the rule (usually with confidence coefficients of 0.90 or 0.95 and small values of δ) against various types of data with known distributions. Schruben (1980) and Schmeiser and Yeh (2002) show that it is much more informative to test coverage at all levels of η in order to evaluate CIPs. In this paper, we show how testing a rule over the entire parameter space of η and δ can reveal potential problems in CIPs. We focus on the loss in coverage associated with the stopping component of CIPs, and we provide evidence of this problem by analytically deriving the coverage when a CIP is applied to i.i.d. normal data. The optimization method proposed provides a systematic way of choosing a stopping rule that will achieve nominal coverage while minimizing computational effort.

The rest of this paper is organized as follows: the first half of the paper gives the derivation of the analytical coverage contours for the normal distribution, and the second half shows how to use coverage contours to find optimal stopping rule parameters in general. Section 2 provides background on the sequential stopping rule literature and notation needed to set up a basic stopping rule. Section 3 reviews the procedure for calculating coverage contours of CIPs analytically for normally distributed data as in Singham and Schruben (2009); §4 discusses the coverage contours using our procedure from §3. Section 5 defines an optimal policy for choosing the stopping rule based on coverage results. Section 6 shows how our analysis can be applied to data from nonnormal and dependent output distributions, and it provides some recommendations for designing stopping rules. Future research topics and concluding remarks are given in §7.

2. Background and Notation

In this section, we introduce some notation for the stopping rule used for the analysis in the rest of the paper, and we review the literature on sequential stopping rules. We focus on a simple absolute precision stopping rule in order to calculate coverage analytically. Simulations often employ CIPs to estimate the unknown mean performance of some system, μ, by generating independent replications of estimates of the performance. The goal of a CIP is to obtain an η-confidence interval for μ with an absolute precision of δ. The data can generically represent such outputs as a replication average, a batch mean, the mean difference between runs of two competing systems, etc. For now, let these outputs X_i be estimates for μ that are assumed to be independent and normally distributed with mean μ and known variance σ^2.

We make these assumptions for two main reasons. The first and most important reason for this paper is to isolate the effects of the stopping rule on the actual coverage probability by avoiding the problems associated with dependence and nonnormality. The second reason is that the analytical contours provide insight for making stopping rule decisions. Coverage contours can be estimated for any distribution using simulation, but in the next two sections, we focus on the analytical derivation of coverage contours for normally distributed data. The fact that these contours closely match those calculated by simulation for normal data encourages the use of simulated contours from other distributions. Section 6 gives examples of some coverage contours for stopping rules applied to simulated nonnormal and dependent data.

Let $\overline{X}_k$ be the cumulative sample mean of the first k i.i.d. $\mathcal{N}(\mu, \sigma^2)$ performance estimators $X_1, \ldots, X_k$. We can construct symmetric confidence intervals of the form $[\overline{X}_k - H_{\eta,k}, \overline{X}_k + H_{\eta,k}]$, where $H_{\eta,k}$ is the half-width of the confidence interval and is calculated in the usual manner by

$$H_{\eta,k} = t_{\eta,k-1}\sqrt{\frac{S_k^2}{k}}, \tag{1}$$

where the sample variance is denoted as S_k^2. The term $t_{\eta,k-1}$ is the $(1+\eta)/2$ quantile of the t-distribution with $k-1$ degrees of freedom.

One could fix a sample size k, generate k i.i.d. simulation observations, and calculate a confidence interval with the correct (nominal) coverage probability using (1). However, the half-width of this confidence interval might be too large to provide a meaningful estimate of μ. Let δ be chosen according to some desired absolute precision criteria. A sequential stopping rule would involve generating observations until a confidence interval with a half-width no greater than δ is obtained. Let k^* be a random variable that is the stopping time of a sequential procedure that stops and generates a confidence interval when the half-width is less than δ. Then k^* is determined by

$$k^* = \arg\min_k H_{\eta,k} \leq \delta. \tag{2}$$

The CIP would return the confidence interval:

$$[\overline{X}_{k^*} - \delta, \overline{X}_{k^*} + \delta]. \tag{3}$$

It is possible that the half-width at stopping is less than δ, but here we assume that the user takes intervals of the form (3) instead of $[\overline{X}_{k^*} - H_{\eta,k^*}, \overline{X}_{k^*} + H_{\eta,k^*}]$. Confidence intervals using δ have coverage greater than or equal to those using the half-width at stopping and meet the precision requirement exactly.

Asymptotically, sequential procedures using (3) have adequate coverage probability at the limits of both small and large values of δ. As δ approaches 0, the observed coverage of sequential CIPs applied to simulated i.i.d. $\mathcal{N}(\mu, \sigma^2)$ data approaches the nominal coverage η as the required sample size approaches infinity, as expected from the results in Chow and Robbins (1965) and Glynn and Whitt (1992). On the other hand, as δ becomes very large, stopping occurs quickly, and the coverage probability will necessarily approach 1 as the intervals become wide. If the output confidence intervals used the half-width H_{η,k^*} instead, the coverage would approach η as δ increases and the expected stopping time decreases. Even though (3) results in wider intervals than those that take the half-width at stopping, the coverage is still usually different from η.

In practice, sequential stopping rules using (3) with smaller values of δ have been consistently observed to yield coverage that is less than nominal, even when the data are i.i.d. $\mathcal{N}(\mu, \sigma^2)$. Problems with coverage have been observed in stopping rules in Anscombe (1954), Ray (1957), and Starr (1966), with the last two references focusing on the case where the data are i.i.d. normal. The research by Anscombe (1954) and Ray (1957) studied variations of our standard sequential procedure in order to evaluate coverage and stopping time distributions. For some of the stopping rules, the authors are able to approximate the expected stopping time, and Ray develops a rule that stops after every two observations (starting with the first three observations) that allows for exact computation of the distribution of the stopping time. Ray is only able to compute coverage for stopping rules that have relatively small expected stopping times (under 30 observations). With improved computing power, we can analyze rules that stop after thousands of observations, allowing us to explore finite-sample behavior closer to the limit. Starr's (1966) work expands on Ray's computations and also proves the asymptotic consistency and efficiency of absolute precision stopping rules applied to normal data with finite variance. Law (1983) surveyed many sequential procedures applied to different types of data and reported on coverage results, with many procedures returning less than nominal coverage. Given that previous research has shown that there is often a loss in coverage using sequential stopping rules, our objective is to provide a systematic way of evaluating this loss and optimizing parameter choice to improve coverage.

In the rest of this paper, we will consider CIPs using stopping rules of the form (2) that produce confidence intervals given by (3). Although sequential stopping rules are asymptotically valid under our assumptions, real simulation experiments must stop for any positive value of δ. This paper focuses on determining the loss in coverage associated with the resulting finite random-sized sample. Most of the previous references focus on coverage performance when η is high. The quality of a CIP can be evaluated using coverage functions, which plot the coverage of a procedure for all possible values of η. In addition to coverage, CIPs can also be evaluated by their effectiveness according to criteria such as the expected length of the half-width (de Peretti and Siani 2010). We model the stopping process to assess the quality of CIPs according to these and other criteria. By evaluating rules according to their coverage, precision, and expected stopping time, we can analyze the trade-offs between these performance measures and see whether an optimal solution can be reached.

3. Method for Calculating Coverage

In this section, we summarize the method of analytically calculating coverage for sequential CIPs producing intervals (3) that is introduced in Singham and Schruben (2009) for i.i.d. normal data. As stated in the previous section, we focus on the normal distribution to isolate the stopping rule bias and to obtain analytical coverage results. We calculate the actual coverage, $\eta^*(\eta, \delta)$, and expected stopping time, $Ek^*(\eta, \delta)$, associated with parameters η and δ. Coverage is calculated for CIPs employing stopping rules of the form (2).

We condition the coverage probability on the random stopping time k^* and calculate the distribution of k^* to determine coverage. Knowledge of the distribution of k^* also provides information on the quality of the procedure, because premature stopping contributes to lower coverage. The value of $Ek^*(\eta, \delta)$ is calculated from the distribution of k^* to measure the expected work required by the CIP.

The actual coverage of the procedure $\eta^* \equiv \eta^*(\eta, \delta)$ is computed by conditioning on the CIP stopping time:

$$\eta^* = P(Cover) = \sum_k P(Cover \mid Stop\ at\ k) P(Stop\ at\ k). \quad (4)$$

To calculate the probability of coverage given stopping at k, we need the following theorem.

Theorem 1 (Robbins 1959). *Suppose that X_i are i.i.d. normal random variables. The distribution of $\overline{X}_{k^*}$ given that $k^* = k$, where k is fixed, is the same as the unconditional distribution of the sample mean $\overline{X}_k$.*

Proof. See Robbins (1959, p. 237). A more detailed proof is provided in Appendix A. □

The proof of this theorem shows that the sample mean at stopping is independent of the history of the sample variances leading up to stopping (for i.i.d. normal data). The unconditional distribution of $\overline{X}_k$ is $\mathcal{N}(\mu, \sigma^2/k)$. Therefore, (4) can be written as

$$P(Cover) = \sum_{k=2}^{\infty} P(\mu - \delta \le \overline{X}_k \le \mu + \delta) P(k^* = k). \quad (5)$$

At least two replications are needed to calculate a confidence interval, although most practitioners start with a higher value of k to avoid stopping early. If stopping occurs at k, then $H_{\eta,k}$ must be less than or equal to δ, so we rewrite the stopping condition as

$$S_k^2 \le \frac{\delta^2 k}{t_{\eta,k-1}^2}.$$

The distribution of the sample variance of normally distributed data is related to the chi-squared distribution, and because the starting sample size is 2, $P(k^* = 2)$ can be defined as $P(S_2^2 \le 2\delta^2/t_{\eta,1}^2)$. Rearranging the terms and using the fact that for i.i.d. normal data, S_2^2 has the distribution of the random variable $\sigma^2\chi_1^2$, we write

$$P(k^* = 2) = P\left(\chi_1^2 \le \frac{2\delta^2}{\sigma^2 t_{\eta,1}^2}\right).$$

For values of k^* greater than 2, the probability of stopping must be calculated conditional on not stopping earlier. The probability of stopping at three samples, for example, is

$$P(k^* = 3) = P(k^* = 3 \mid k^* > 2) P(k^* > 2).$$

Continuing the recursion for larger sample sizes, we multiply the probability of not stopping at $k-1$ by the probability of stopping at k given that stopping has not occurred by $k-1$:

$$P(k^* = k) = P(k^* = k \mid k^* > k-1) \cdot \left[\prod_{i=3}^{k-1} P(k^* > i \mid k^* > i-1)\right] P(k^* > 2). \quad (6)$$

The next step is to calculate the conditional probabilities $P(k^* = k \mid k^* > k-1)$ and $P(k^* > k \mid k^* > k-1)$ for $k \ge 3$. The conditional probabilities $P(k^* = k \mid k^* > k-1)$ are calculated by integrating over the possible ranges of $\overline{X}_{k-1}$ and S_{k-1}^2 given that stopping has not occurred before k, and determining the values of the next observation (X_k) that result in stopping. The random variable $\bar{X}_{k-1}$ has an infinite range, but the range of possible values of S^2_{k-1}, given that $k^* > k-1$, is bounded.

To see this, consider the fact that the sample variance for the first $k-1$ observations must be large enough to avoid meeting the stopping rule at $k-1$ but small enough that the next observation X_k could lower the variance enough to meet the stopping rule at time k. Focusing on this range of S^2_{k-1} helps us understand the conditions that lead to stopping, although we could perform the calculations over the infinite range. Let $V_{\min}(k)$ denote the greatest lower bound on the sample variance S^2_{k-1} that ensures the stopping rule is not met before k. We solve for $V_{\min}(k)$ by setting $H_{\eta,k-1}$ greater than or equal to δ and solving for S^2_{k-1}:

$$S^2_{k-1} = \frac{\delta^2(k-1)}{t^2_{\eta,k-2}} \triangleq V_{\min}(k).$$

Next, we find the smallest upper bound on S^2_{k-1}. To stop at time k, the sample variance S^2_{k-1} must be low enough so that it is possible for S^2_k to result in $H_{\eta,k} \le \delta$. To relate the variances S^2_{k-1} and S^2_k, we use the following recursion (Welford 1962):

$$S^2_k = \frac{k-2}{k-1}S^2_{k-1} + \frac{(X_k - \bar{X}_{k-1})^2}{k}. \qquad (7)$$

The sample variance S^2_k must be less than $\delta^2 k/t^2_{\eta,k-1}$ for stopping to occur. Setting S^2_k in (7) less than $\delta^2 k/t^2_{\eta,k-1}$ and solving for $|X_k - \bar{X}_{k-1}|$ yields the maximum absolute difference between X_k and $\bar{X}_{k-1}$ that would allow stopping for a given S^2_{k-1}. Call this value $X^b_k(S^2_{k-1})$:

$$X^b_k(S^2_{k-1}) = \sqrt{\frac{\delta^2 k^2}{t^2_{\eta,k-1}} - \frac{(k-2)kS^2_{k-1}}{k-1}}. \qquad (8)$$

To find the maximum sample variance at $k-1$ that will allow stopping, consider the value of $X^b_k(S^2_{k-1})$ in (8). If S^2_{k-1} is too large, the term under the radical becomes negative. The maximum value of S^2_{k-1} that allows $X^b_k(S^2_{k-1})$ to be real-valued is the maximum value of S^2_{k-1} that allows for the possibility of stopping. Setting the terms under the radical to be greater than or equal to 0 and solving for S^2_{k-1} yields the upper bound on the sample variance at observation k, $V_{\max}(k)$:

$$S^2_{k-1} = \frac{\delta^2 k(k-1)}{t^2_{\eta,k-1}(k-2)} \triangleq V_{\max}(k).$$

For values of S^2_{k-1} that are larger than $V_{\max}(k)$, the sample variance is so large that even if $X_k = \bar{X}_{k-1}$, S^2_k will not be small enough to meet the stopping rule.

We use the ranges $S^2_{k-1} \in [V_{\min}(k), V_{\max}(k)]$ and $X_k \in [\bar{X}_{k-1} - X^b_k(S^2_{k-1})\ \bar{X}_{k-1} + X^b_k(S^2_{k-1})]$ to construct an integral representation for the conditional probability of stopping. Let $f_{\bar{X}_{k-1}}$ be the probability density function for the random variable $\bar{X}_{k-1}$, and let $f_{S^2_{k-1}\mid k^*>k-1}$ be the density function for the sample variance conditional on the stopping rule not being met before time k. We calculate the conditional probability of stopping at k, $P(k^* = k \mid k^* > k-1)$, as

$$\int_x \int_{y=V_{\min}(k)}^{V_{\max}(k)} P(x - X^b_k(y) \le X_k \le x + X^b_k(y)) f_{S^2_{k-1}\mid k^*>k-1}(y) f_{\bar{X}_{k-1}}(x)\,dy\,dx. \qquad (9)$$

The outer integral is over the range of the random variable $\bar{X}_{k-1}$, and the inner integral is over the random variable S^2_{k-1}, given that $k^* > k-1$. These variables are independent by Theorem 1. Once the distribution of the stopping time has been calculated as in (6), the probabilities can be used to calculate the coverage probabilities in (5). One important point to note is that the distribution of S^2_{k-1}, given that stopping has not yet occurred, is different from the unconditional distribution of S^2_{k-1}. The history of the sample variances determines whether stopping has occurred and provides information on the value of the sample variance at $k-1$. The conditional density function for S^2_{k-1} is derived in Appendix B.

We evaluate the integral (9) numerically using a Newton-Cotes method with a rectangular rule approximation. We divide the space of $\bar{X}_{k-1}$ and S^2_{k-1} into equally sized rectangles and approximate the term in the integral using the function values at the centers of the rectangles. As smaller rectangles are used, the distribution of k^* calculated from (6) and (9) appears to converge to the distribution calculated via simulation. From these results, we learn that coverage of a stopping rule can be broken down according to the distribution of the sample mean and sample variance as each observation is collected, and that it is possible to characterize the distribution of the stopping time as the product of conditional probabilities as in (6).

Theorem 1 allows us to treat the distribution of the sample mean at stopping as independent of the history leading up to stopping for normally distributed data. We are not aware of a similar result existing for other distributions. It may still be possible to calculate coverage analytically for nonnormal data using integration, although it might be much harder if the sample mean and sample variance are not independent. However, our ability to derive coverage for the normal case provides insight into the factors affecting coverage for any distribution.

4. Analytical Results

Using the method in the previous section, we calculate the expected stopping time and actual coverage probability of our stopping rule for a range of values of η and δ. The same quantities can be estimated for nonnormal distributions by simulation. As suggested in Schmeiser and Yeh (2002), the quality of a procedure should be evaluated by considering all possible values of η. We plotted the contours of $\eta^*(\eta, \delta)$ and $Ek^*(\eta, \delta)$ to observe which stopping rules resulted in better coverage (higher values of η^*) and which ones required lower numbers of replications (lower values of Ek^*). Figure 1 shows the contours of η^* and Ek^* calculated for i.i.d. $\mathcal{N}(0, 1)$ data. The x axes contain values of η, and the y axes show the range of values of δ considered.

In the left plot of Figure 1, the coverage contours are plotted. Coverage was calculated numerically using the distribution of k^* (found by the method in §3) in (5). In Region 3, coverage behaves as expected from the asymptotic theory. As δ approaches 0, coverage approaches η from below, and as η increases, coverage improves. However, coverage is not always less than η; for large δ, coverage approaches 1. This is because with large values of δ, stopping occurs early, and the output confidence interval $[\bar{X}_{k^*} - \delta, \bar{X}_{k^*} + \delta]$ could be much wider than the half-width that led to stopping.

Region 1 contains the parameters for which the probability of stopping immediately (after two replications) is almost 1. That is, the expected stopping time is approximately 2 according to the calculations to compute the distribution of k^*. Region 2 contains the parameters where increasing δ leads to improved coverage and the expected stopping time is greater than 2. Coverage for a given value of η is lowest on the boundary between Regions 2 and 3. The dashed line shows the parameters for which nominal coverage is achieved. Above the dashed line, $\eta^*(\eta, \delta) > \eta$, whereas below the line, $\eta^*(\eta, \delta) < \eta$.

The right plot shows the contours of Ek^* for the same parameter space. These values are calculated numerically using the distribution function for k^* discussed in §3. The function Ek^* appears to be convex and increasing as η increases and δ decreases. The contours (spaced by 100 for $Ek^* \geq 100$) reveal that the expected stopping time increases rapidly as δ is decreased and η is increased. It is important to note that if batching or some other aggregation method is used to generate each observation k, then the actual computation cost might be greater than what these contours suggest.

From the literature it is known that stopping rules in CIPs often lead to less-than-nominal coverage. These figures reveal how the relationships between η and δ affect coverage performance. Of course, the figures in this section only pertain to the $\mathcal{N}(0, 1)$ distribution. However, similar contours can be generated for other distributions using simulation (see §6), and generally, the same relationships hold. If the user is willing to use large values of δ, coverage may be better than nominal. Usually, however, small values of δ and high values of η are used, resulting in subnominal coverage and high numbers of replications. This provides the motivation to use optimization to determine whether better parameter choices can be found.

5. Optimal Policies for Coverage

Using the results of the previous section, we determine an optimal policy for obtaining a confidence interval with coverage level η^o within a specified half-width δ^o. Again, we focus on the contours calculated for normally distributed data using the method in §3 because those are calculated analytically. Our

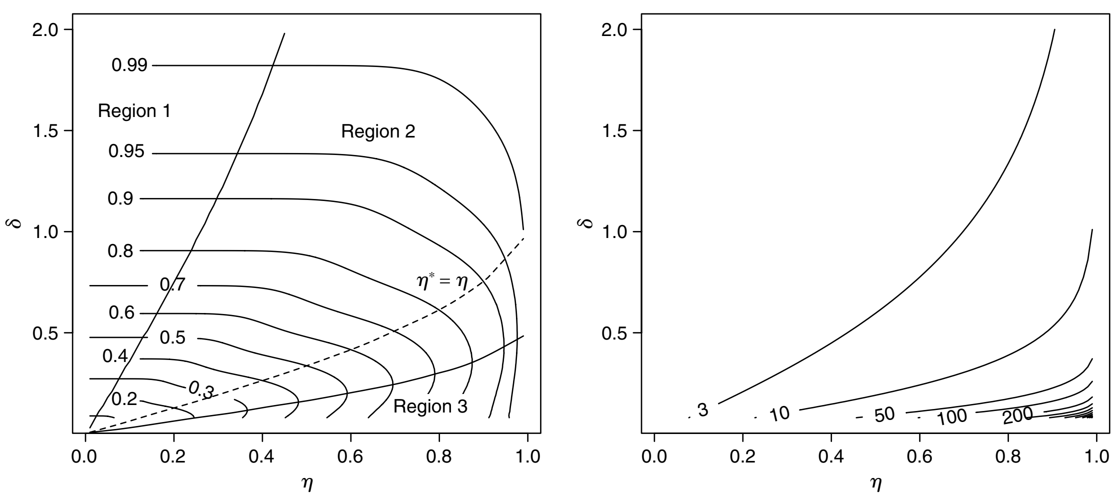

Figure 1 Coverage Contours for Various η, δ (Left); Contours of $Ek^*(\eta, \delta)$ for Various η, δ (Right)

optimization method can also apply to contours generated by simulation for other distributions if they have the same general shape as those for the normal distribution. Even if the assumptions of our optimization model do not hold for the contours in question, a search over the feasible space can yield the minimum-cost rule that meets the confidence and precision requirements of the user.

Suppose the user wishes to design a CIP that returns intervals with coverage of at least η^o while having a half-width of at most δ^o. Setting a stopping rule with parameters (η^o, δ^o) will result in coverage $\eta^*(\eta^o, \delta^o)$ that is most likely different from η^o. If the parameters are below the dashed line in the left plot of Figure 1, the coverage obtained is less than η^o. For the moment, assume that we are below the dashed line in the area we refer to as the subnominal coverage region. We can either increase the value of the requested confidence coefficient η or decrease the value of the requested precision δ to improve actual coverage. However, decreasing the value of δ alone, as suggested from the asymptotic theory, will not bring the coverage to at least η—only closer to it. In fact, our results suggest quite the opposite—that nominal coverage can be achieved (at least cost) by increasing the confidence level requested in the procedure.

There are various possibilities involving combinations of decreasing δ and increasing η to bring the coverage up to what is desired. We suggest that the least-cost solution (with cost measured in terms of the expected stopping time Ek^*) is to change η while leaving δ the same. If (η^o, δ^o) is in the subnominal region, the optimal choice is to increase η, whereas η should be decreased if it is above the nominal boundary. We first find an optimal policy for parameters in Region 3, which is a subset of the subnominal region.

Assume that Ek^* is a convex function of η and δ and that the function η^* has contours of the form in Figure 1. To achieve a confidence interval with coverage η^o and half-width less than or equal to δ^o in Region 3, the optimal parameter choices (that minimize Ek^*) are $\eta' > \eta^o$ such that $\eta^*(\eta', \delta^o) = \eta^o$, and $\delta = \delta^o$. To show this, we formulate the minimization as follows:

$$\min_{\eta,\,\delta} \; Ek^*(\eta, \delta)$$
$$\text{s.t.} \quad \eta^*(\eta, \delta) \geq \eta^o,$$
$$\delta \leq \delta^o.$$

We seek a set of procedure parameters (η, δ) that will minimize the expected stopping time of the procedure while meeting the sequential CIP precision-and coverage-level requirements. Because Ek^* is a convex function with respect to η and δ, the solution will be a Karush-Kuhn-Tucker (KKT) point of the feasible space, i.e., a point that meets the KKT conditions for optimality in a nonlinear program (Luenberger 2003). Consider the point (η', δ^o), where $\eta^*(\eta', \delta^o) = \eta^o$. There is a one-to-one mapping between η and $\eta^*(\eta, \delta^o)$, and $\eta^*(\eta, \delta^o)$ increases with η. In the subnominal region, $\eta^*(\eta^o, \delta^o) < \eta^o$ for $\delta^o > 0$. Therefore, η' must be greater than η^o in order to have $\eta^*(\eta', \delta^o) = \eta^o$.

Our solution, (η', δ^o), meets both constraints at equality. The dual feasibility condition is

$$\begin{pmatrix} \partial Ek^*/\partial\eta \\ \partial Ek^*/\partial\delta \end{pmatrix} + u_1 \begin{pmatrix} -\partial\eta^*/\partial\eta \\ -\partial\eta^*/\partial\delta \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0, \qquad (10)$$

where u_1 and u_2 are required to be nonnegative for (η', δ^o) to be a solution. Solving for u_1 and u_2 yields

$$u_1 = \frac{\partial Ek^*/\partial\eta}{\partial\eta^*/\partial\eta},$$
$$u_2 = \left(\frac{\partial Ek^*/\partial\eta}{\partial\eta^*/\partial\eta}\right)(\partial\eta^*/\partial\delta) - (\partial Ek^*/\partial\delta).$$

For u_1 to be nonnegative, the derivatives of Ek^* and η^* with respect to η must be either both positive or both negative. Both the coverage and the expected stopping time are increasing in η, so $u_1 \geq 0$.

Next, we show that $u_2 \geq 0$ in Region 3. Keeping in mind that Ek^* and η^* decrease with increasing values of δ, to have $u_2 \geq 0$, we need the following to hold:

$$\frac{\partial Ek^*/\partial\eta}{\partial\eta^*/\partial\eta} \leq \frac{\partial Ek^*/\partial\delta}{\partial\eta^*/\partial\delta}. \qquad (11)$$

The ratio of the increase in expected cost to the increase in coverage gained should be higher for decreasing δ than for increasing η. Figure 2 shows the

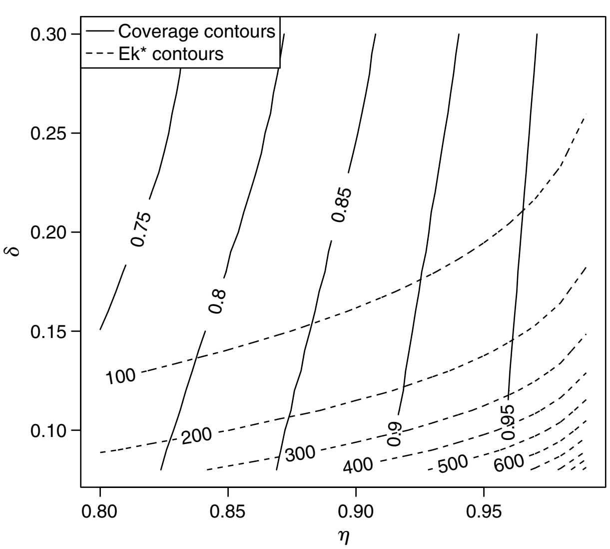

Figure 2 Contours of $\eta^*(\eta, \delta)$ and $Ek^*(\eta, \delta)$ in Region 3

contours of $\eta^*(\eta, \delta)$ and $Ek^*(\eta, \delta)$ in part of Region 3. We see that decreasing δ requires relatively more computations for the amount of coverage gained than the amount of computations required for increasing η. Additionally, calculating the partial derivatives numerically confirms that $u_2 \geq 0$. Therefore, within Region 3, the optimal parameter choice to obtain coverage of η^o and half-widths of δ^o is to choose η' and δ^o such that $\eta^*(\eta', \delta^o) = \eta^o$. We summarize the optimization stopping rule in the following result.

Result 1. Assume that the contours of Ek^* are convex in η and δ, that Ek^* and η^* are increasing in η, and that (11) holds. Then the optimal parameter choice is to use δ^o, and the value of η' higher than η^o such that the actual coverage using (η', δ^o) is η^o.

A similar analysis shows that the optimal policy is the same for (η^o, δ^o) outside of Region 3. If (η^o, δ^o) lies above the nominal boundary, then the optimal choice of η will be less than η^o. The main result of this analysis is that it is more effective to change η to improve coverage than to change δ.

Result 2. Nominal coverage can be achieved at least cost by modifying the confidence coefficient of the procedure while decreasing the precision value results in approaching nominal coverage from below. For coverage contours and expected stopping time contours having the properties listed in Result 1, the optimal choice of η can be defined.

For data with different distributions, the same optimal solution applies if the contours of Ek^* are convex and the relative cost of decreasing δ to achieve an incremental increase in coverage is greater than the cost of increasing η. If these conditions are not met, a different optimization approach may be available to solve the nonlinear program. Most of the current literature suggests values of δ that lead to adequate coverage; here, we suggest changing η instead. The next section examines the coverage results for data that are not standard normal and provides recommendations on choosing stopping rules based on the optimal policies calculated.

6. Results for Different Distributions and Recommendations

Our work in the previous sections addresses the coverage for stopping rules applied to i.i.d. normal data. We computed the coverage and expected stopping time contours for the case where the variance of the data was 1 and showed how an optimal solution existed. We now show the results for data with distributions different from $\mathcal{N}(0, 1)$. We modify the variance of the data, using i.i.d. normal data with $\sigma^2 = 1/4$, 4, and 100. Our numerical integration routine was slow for large values of σ^2, so we used Monte Carlo methods to compute the coverage contours when the data were normally distributed with variance 100. Coverage contours for these different distributions are provided in Figures 3 and 4. The contours have the same shape for different variances, but they scale differently along the δ axis. Intervals using a particular precision value will have greater coverage over data with a smaller variance than data with a larger variance.

We also evaluate the coverage of the stopping rule applied to exponentially distributed random variables with mean 1, and dependent data from an autoregressive model with lag 1 and autocorrelation coefficient $\rho = 0.5$. Here, the stopping rule assumes that the data are i.i.d. normal in calculating confidence intervals, but the data actually have a different distribution. Figure 5 shows the coverage contours for exponential

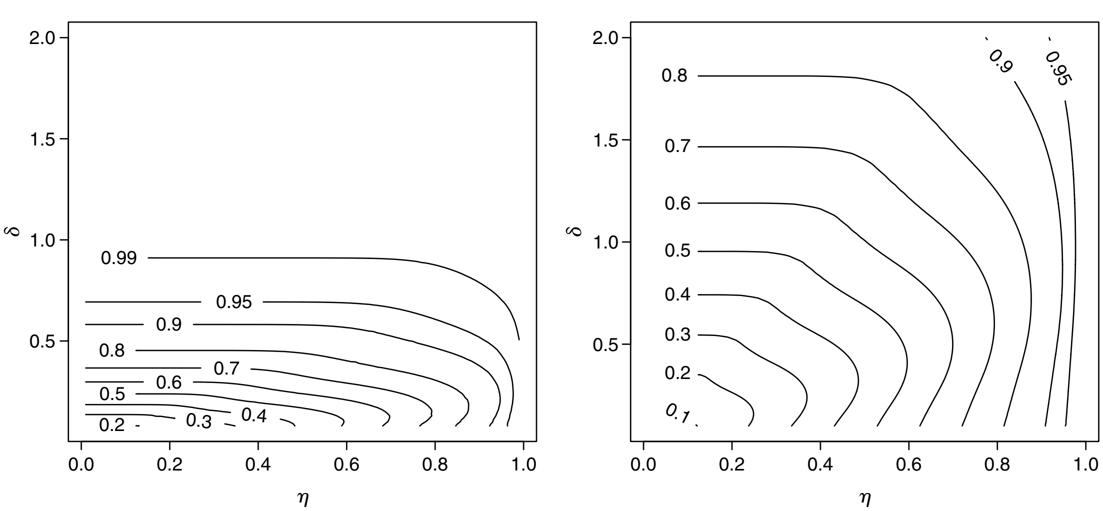

Figure 3 **Coverage of Stopping Rules for i.i.d. Normal Data with $\sigma^2 = 1/4$ and 4**

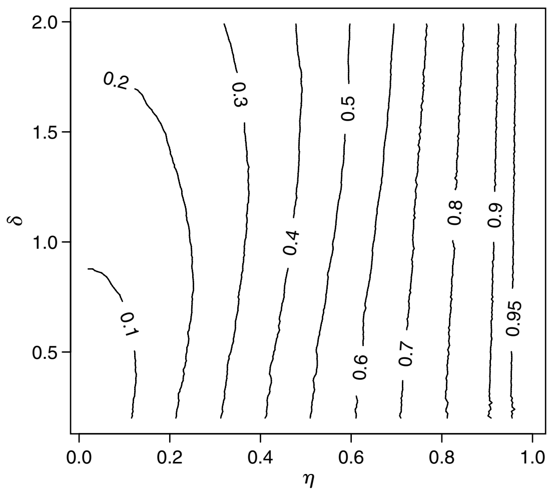

Figure 4 **Coverage of Stopping Rules for i.i.d. Normal Data with** $\sigma^2 = 100$

data (which violates the normality assumption) and for autoregressive data (which violates the independence assumption). We see that the coverage is worse than what it would be for standard normal data, because the assumptions of the procedure are not met. Monte Carlo methods were used to generate the plots for these distributions where coverage could not be calculated analytically.

Based on the optimal policy for choosing stopping rules derived in §5, we calculate optimal values of η over the parameter space for different distributions of data. Let $\eta'(\eta, \delta)$ be the optimal value of η for given parameters (η, δ). Suppose that for a set of data from a given distribution, we fix η and vary δ. As δ approaches 0, η' approaches η from above; as δ approaches infinity, η' approaches 0. The value of η' is bounded by 1, so for a fixed η, we can find the maximum value of η' over all values of δ. This maximum value of η' as a function of η is

$$\hat{\eta}(\eta) = \max_{\delta} \eta'(\eta, \delta).$$

This maximum optimal value η' would provide at least nominal coverage for any choice of δ. We compute values of $\hat{\eta}(\eta)$ for normally distributed data with various values of σ^2 and display them in Table 1.

We find that $\hat{\eta}(\eta)$ remains relatively independent of the variance of normally distributed data. For example, if 90% confidence intervals are desired, the maximum value of η needed to obtain coverage of 90% for any value of δ is around 94.5% for the possible values of σ^2 tested. This implies that if the data are normally distributed, using a stopping rule with $\eta = 95\%$ should allow us to obtain coverage of at least 90% for *any desired precision* value of δ. These results can be useful in choosing stopping rules for experiments where the underlying variance of the data is not known. The value of $\hat{\eta}(\eta)$ appears to be insensitive to the variance of normally distributed data, presumably because we are searching along the same contours for the worst possible δ, which will vary based on the rescaling of the graph. There is variation in the values of $\hat{\eta}(\eta)$ across the columns because of numerical error in our calculations (or finite-sample Monte Carlo error in the last column). These values should be taken as guides for choosing η to achieve the required coverage level when the variance is not known. If the variance is known, then the exact optimal parameter choice can be found using the method in §5.

However, if the data are nonnormal or dependent, then the maximum value of η required to achieve a desired coverage level may be substantially different. Table 2 shows the values of $\hat{\eta}(\eta)$ for the exponential distribution and the AR(1) data, which are much higher than those for the normally distributed data.

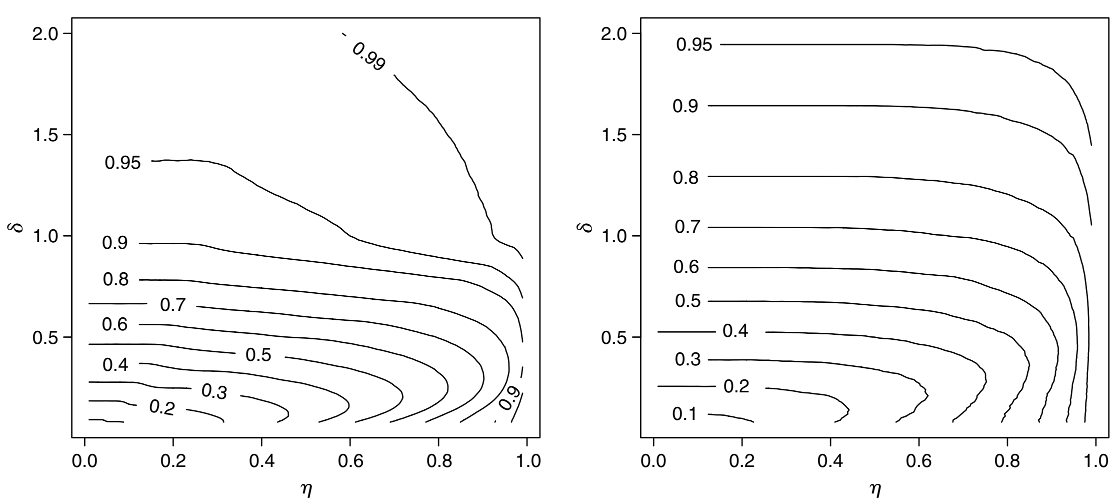

Figure 5 **Coverage of Stopping Rules for i.i.d. Exponentially Distributed Data with Mean 1 and AR(1) Data with Autocorrelation of** $\rho = 0.5$

Table 1 Values of $\hat{\eta}(\eta)$ for Normal Distributions with Different σ^2

η (%)	$\sigma^2 = 1$	$\sigma^2 = 1/4$	$\sigma^2 = 4$	$\sigma^2 = 100$
80	0.874	0.874	0.867	0.865
85	0.913	0.913	0.912	0.904
90	0.947	0.946	0.944	0.942
95	0.976	0.976	0.977	0.976
99	0.996	0.996	0.996	0.996

Table 2 Values of $\hat{\eta}(\eta)$ for Exponential and Autoregressive Data Compared to the Normal Distribution

η (%)	$\mathcal{N}(0, 1)$	Exp(1)	AR(1), $\rho = 0.5$
80	0.874	0.961	0.986
85	0.913	0.979	0.993
90	0.947	0.992	0.997
95	0.976	0.998	>0.999
99	0.996	>0.999	>0.999

The confidence coefficient of the procedure must be much higher than the coverage desired to compensate for the nonnormality or dependence of the underlying data. This suggests another payoff in efforts to achieve approximately i.i.d. normal observations (say, replication or batch averages).

We also consider modifications to the basic rule that requires stopping as soon as the half-width is less than δ. One rule checks the stopping condition every five observations instead of after each one. The second rule requires that the precision requirement be met for two observations, so stopping occurs after $\mathrm{HW}_{\eta,k} \le \delta$ for two separate k (not necessarily in sequential order). Table 3 shows the values of $\hat{\eta}(\eta)$ for these rules, and we see that the maximum confidence coefficients that are required to obtain at least nominal coverage for all possible δ are much lower than the values in Table 1. However, the expected stopping times for these rules are generally higher because they do not stop as soon as the precision requirement is met.

These results suggest that nominal coverage can be achieved by increasing the value of η used to calculate the confidence intervals. However, most simulation experiments do not have known output distributions. More research is needed to determine what adjustments must be made in the cases where the output distribution is completely unknown. What we can suggest is that the stopping rule component be tested against i.i.d. normally distributed data to see the potential effect on coverage. Testing the rule against other distributions can also provide some idea of what might happen if the distribution is different from that assumed. Additionally, the variance of the underlying data plays a large role in the scale of the contours along the δ axis. Therefore, if contours can be estimated for an approximate data distribution, the user might have a better idea of the ranges of η and δ values that avoid the regions with subnominal coverage.

Table 3 Values of $\hat{\eta}(\eta)$ for Modified Sequential Procedures Applied to i.i.d. $\mathcal{N}(0, 1)$ Data

η (%)	Check every 5 obs.	Meets rule twice
80	0.813	0.830
85	0.862	0.875
90	0.911	0.920
95	0.958	0.963
99	0.992	0.994

7. Conclusion

This research analyzes the effects of sequential stopping rules on confidence interval procedures. By applying the stopping rules to i.i.d. normal data, we isolate the effect of the rules on coverage. Generally speaking, the coverage returned by a CIP using a sequential stopping rule is different from the confidence sought by the user. To achieve the target coverage, the user can change the confidence level and precision requested of the procedure. We generate contour plots of the coverage and expected stopping time over a space of parameters to evaluate the rules. To balance improved performance against cost, we formulate an optimization model to choose stopping rules and find an optimal policy. We find that it is cheaper to increase the confidence coefficient to improve coverage rather than to decrease the half-width threshold for the particular distributions analyzed. By analyzing the optimal policies for the normal distribution with different variance parameters, we are able to find confidence coefficients for stopping rules that appear to deliver at least nominal coverage for any desired precision level.

There is generally a trade-off between coverage and the computational cost of stopping rules. Coverage and expected stopping time contours provide information on approximately how many observations are needed to obtain better coverage or a smaller precision level. Optimization methods can be used to manage these trade-offs in order to provide results that meet the user's objectives. If the data are independent and normally distributed with unknown variance, the values in Table 1 can be used to achieve at least nominal coverage with minimal replications. If the variance of the data is known, optimal solutions can be found using the results of §5. If the data have a known nonnormal distribution, similar contours can be estimated to see whether an optimal solution exists. We see from the exponential and autoregressive contours that even higher confidence coefficients must be used if the data do not meet the assumptions of the CIP.

In practice, batching techniques are often used to obtain observations that are approximately i.i.d. normal, so the real computational cost associated with each observation can be high. Future research will determine how batching techniques used with stopping rules can be evaluated using coverage contours. If the data are dependent or nonnormal, batching and applying the stopping rule to approximately i.i.d. normal data might be more efficient than using the optimal stopping rule parameters derived for the data directly. Additionally, many stopping rules used in practice are more complex than the ones considered here and may have different requirements for stopping. As seen in Table 3, it is possible to improve coverage by designing procedures that have more strict stopping requirements. However, these procedures will likely have a higher expected stopping time, so it would be interesting to see whether there exists an optimal class of procedures. We hope that we have provided the motivation to explore an alternative method of improving coverage for many stopping rules under different distributional settings.

Acknowledgments

The authors thank J. George Shanthikumar and David R. Brillinger for their assistance with this research. They are also grateful to two anonymous reviewers for their valuable comments and suggestions that greatly improved this paper.

Appendix A

Proof of Theorem 1. We need to show that the distribution of $\bar{X}_{k^*}$, given that the stopping time is k, is $\mathcal{N}(\mu, \sigma^2/k)$, which is the unconditional distribution of the sample mean for a fixed sample size experiment with k replications. The probability of stopping at k depends on the history of sample variances $S_2^2, S_3^2, \ldots, S_k^2$. Thus, we seek to show that the distribution of $\bar{X}_k$, given the sample variance history, is the same as its unconditional distribution.

First, we show that $\bar{X}_k$ is independent of S_i^2 for $i \le k$. To do this, we rely on a version of the proof of independence between $\bar{X}_k$ and S_k^2 presented in Davison (2003). Take $X_1, X_2, \ldots, X_k$ to be i.i.d. $\mathcal{N}(\mu, \sigma^2)$. Consider the $(k \times k)$ matrix $\mathbf{B}^T$:

$$\mathbf{B}^T = \begin{pmatrix} \frac{1}{k^{1/2}} & \frac{1}{k^{1/2}} & \frac{1}{k^{1/2}} & \cdots & \frac{1}{k^{1/2}} \\ \frac{1}{2^{1/2}} & \frac{-1}{2^{1/2}} & 0 & \cdots & 0 \\ \frac{1}{6^{1/2}} & \frac{1}{6^{1/2}} & \frac{-2}{6^{1/2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{[k(k-1)]^{1/2}} & \frac{1}{[k(k-1)]^{1/2}} & \frac{1}{[k(k-1)]^{1/2}} & \cdots & \frac{-(k-1)}{[k(k-1)]^{1/2}} \end{pmatrix}.$$

Note that $\mathbf{B}^T\mathbf{B}$ is the identity matrix, $\mathbf{I}_k$ and that $\mathbf{B}^T 1_k = (\sqrt{k}, 0, \ldots, 0)^T$. Let $\mathbf{X}$ be the vector of k observations $(X_1, X_2, \ldots, X_k)^T$. Let $\mathbf{U} = \mathbf{B}^T\mathbf{X}$. The distribution of $\mathbf{U}$ is multivariate normal with mean vector $(\sqrt{k}\mu, 0, \ldots, 0)^T$ and covariance matrix $\sigma^2\mathbf{I}_k$.

Then, $U_1 = 1/\sqrt{k}\sum_{i=1}^k X_i = \sqrt{k}\bar{X}_k$. Using the fact that $U_1 \sim \mathcal{N}(\sqrt{k}\mu, \sigma^2)$, we see that $\bar{X}_k \sim \mathcal{N}(\mu, \sigma^2/k)$. This establishes the distribution of $\bar{X}_k$, and so we now show that it is independent of historical values of the sample variance S_i^2 for $i < k$.

For $j = 2, \ldots, k$, $U_j = \sqrt{(j-1)/j}(\bar{X}_{j-1} - X_j)$. Next, rewrite the variance recursion Equation (7) as the following:

$$\begin{aligned}(k-1)S_k^2 &= (k-2)S_{k-1}^2 + \frac{k-1}{k}\left(X_k - \bar{X}_{k-1}\right)^2 \\ &= (k-3)S_{k-2}^2 + \frac{k-2}{k-1}\left(X_{k-1} - \bar{X}_{k-2}\right)^2 \\ &\quad + \frac{k-1}{k}\left(X_k - \bar{X}_{k-1}\right)^2 \\ &= S_2^2 + \sum_{j=3}^{k}\frac{j-1}{j}\left(X_j - \bar{X}_{j-1}\right)^2.\end{aligned}$$

For a particular historical variance S_{k-i}^2, $i = 1, \ldots, k-2$, we have

$$(k-1)S_k^2 = (k-i-1)S_{k-i}^2 + \sum_{j=k-i+1}^{k}\frac{j-1}{j}\left(X_j - \bar{X}_{j-1}\right)^2.$$

Next, note that $\sum_{j=1}^k X_j^2 = \mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{B}^T\mathbf{B}\mathbf{X} = \mathbf{U}^T\mathbf{U}$ because $\mathbf{B}^T\mathbf{B}$ is the identity matrix. Then write

$$\begin{aligned}(k-1)S_k^2 &= \sum_{j=1}^{k} X_j^2 - k\bar{X}_k^2 \\ &= \sum_{j=1}^{k} U_j^2 - k\bar{X}_k^2 = (k-i-1)S_{k-i}^2 \\ &\quad + \sum_{j=k-i+1}^{k}\frac{j-1}{j}\left(X_j - \bar{X}_{j-1}\right)^2.\end{aligned}$$

However, we should note that $U_1^2 = k\bar{X}_k^2$ and $U_j^2 = (j-1/j)(X_j - \bar{X}_{j-1})^2$. Therefore, $(k-i-1)S_{k-i}^2 = U_2^2 + \cdots + U_{k-i}^2$, and these values of U_j are independent of U_1, so S_{k-i}^2 is independent of $\bar{X}_k$.

Finally, we need to show that the distribution of the mean at stopping given that we stop at k is the same as the distribution of the mean for a fixed sample size k. Write the distribution $P(\bar{X}_{k^*} \le z \mid k^* = k)$ as

$$P\big(\bar{X}_k \le z \mid S_2^2 > V_{\min}(2), S_3^2 > V_{\min}(3), \ldots, S_{k-1}^2 > V_{\min}(k-1), S_k^2 \le V_{\min}(k)\big).$$

Because $\bar{X}_k$ is independent of the history of the sample variances, this distribution simplifies to $P(\bar{X}_k \le z)$, which is the unconditional distribution of the mean for a fixed sample size k. □

Alternatively, the following explanation shows how $\bar{X}_k$ and S_i^2 (for $i < k$) are independent. The sample variance S_i^2 consists of the squared terms $\bar{X}_i - X_j$ for $j = 1, \ldots, i-1, i$. Because both $\bar{X}_k$ and $\bar{X}_i - X_j$ are normally distributed, if they are uncorrelated, they are independent. We rewrite the covariance of $\bar{X}_k$ and $\bar{X}_i - X_j$ as

$$\begin{aligned}&\mathrm{Cov}\left(\frac{X_1 + \cdots + X_k}{k}, \frac{X_1 + \cdots + X_i}{i} - X_j\right) \\ &\quad = \frac{1}{ik}\mathrm{Cov}(X_1 + \cdots + X_k, X_1 + \cdots + (1-i)X_j + \cdots + X_i).\end{aligned}$$

Because the variables X_m, $m = 1, \ldots, k$, are independent, most of the cross terms cancel, and we are left with

$$\frac{1}{ik}\left((1-i)\operatorname{Var}(X_j) + \sum_{m=1,\ldots,i,\, m\neq j} \operatorname{Var}(X_m)\right)$$
$$= \frac{1}{ik}((1-i)\sigma^2 + (i-1)\sigma^2) = 0.$$

Appendix B

The following analysis also appears in Singham and Schruben (2009), although the notation has been corrected here. We wish to calculate the density function of S_{k-1}^2 given $k^* > k-1$, i.e., given that the stopping rule has not yet been met by time $k-1$. Consider the distribution of the sample variance of a set of $k-1$ i.i.d. normal random variables with variance σ^2.

Relate the distribution of S_k^2 to S_{k-1}^2 by (7), and rewrite it as

$$S_k^2 \stackrel{d}{=} \left[\frac{S_{k-1}^2(k-2)}{\sigma^2} + \frac{(k-1)(X_k - \overline{X}_{k-1})^2}{k\sigma^2}\right]\frac{\sigma^2}{(k-1)}.$$

Because X_k and $\overline{X}_{k-1}$ are independent of each other and of S_k^2 and S_{k-1}^2 by Theorem 1, we can simplify the above by noting that $(X_k - \overline{X}_{k-1})$ has a $\mathcal{N}(0, \sigma^2 k/(k-1))$ distribution. Let Z_{k-1}^2 be a squared standard normal random variable, and write the distribution of S_k^2 as

$$S_k^2 \stackrel{d}{=} \left[\frac{S_{k-1}^2(k-2)}{\sigma^2} + Z_{k-1}^2\right]\frac{\sigma^2}{(k-1)}. \tag{B1}$$

Using (B1), we can write the distribution of $S_k^2 \mid S_{k-1}^2$ in terms of Z_{k-1}^2:

$$P(S_k^2 \le x \mid S_{k-1}^2) = P\left(\left[\frac{S_{k-1}^2(k-2)}{\sigma^2} + Z_{k-1}^2\right]\frac{\sigma^2}{(k-1)} \le x \;\Big|\; S_{k-1}^2\right)$$
$$= P\left(Z_{k-1}^2 \le \frac{1}{\sigma^2}[x(k-1) - S_{k-1}^2(k-2)] \;\Big|\; S_{k-1}^2\right).$$

We index Z_{k-1}^2 according to the number of observations in the history prior to X_k because there will be a different squared normal random variable associated with the transition from S_{k-1}^2 to S_k^2 for each k. Recall that to have $k^* > k-1$, S_{k-1}^2 must be bounded from below by $V_{\min}(k)$. By integrating over the possible values of S_{k-1}^2 according to its distribution conditional on not having stopped yet, we recursively calculate the conditional distribution of the variance as

$$f_{S_k^2 \mid k^* > k-1}(x) = \int_{y=V_{\min}(k)}^{\infty} f_{Z_{k-1}^2}\left(\frac{1}{\sigma^2}[x(k-1) - y(k-2)]\right) f_{S_{k-1}^2 \mid k^* > k-1}(y)\, dy, \tag{B2}$$

where f is again used to represent the density function of the random variable in its subscript. Because the variance is calculated sequentially, at time k we have the value of S_{k-1}^2. We calculated (B2) numerically and included the results in (9) to estimate the distribution of the stopping time.

References

Anscombe, F. J. 1954. Fixed-sample-size analysis of sequential observations. *Biometrics* **10**(1) 89–100.

Chen, E. J., W. D. Kelton. 2007. A procedure for generating batch-means confidence intervals for simulation: Checking independence and normality. *Simulation* **83**(10) 683–694.

Chow, Y. S., H. Robbins. 1965. On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36**(2) 457–462.

Davison, A. C. 2003. *Statistical Models*. Cambridge University Press, Cambridge, UK.

de Peretti, C., C. Siani. 2010. Graphical methods for investigating the finite-sample properties of confidence regions. *Comput. Statist. Data Anal.* **54**(2) 262–271.

Glynn, P. W., W. Whitt. 1992. The asymptotic validity of sequential stopping rules for stochastic simulations. *Ann. Appl. Probab.* **2**(1) 180–198.

Heidelberger, P., P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Oper. Res.* **31**(6) 1109–1144.

Hoad, K., S. Robinson, R. Davies. 2009. Automating discrete event simulation output analysis—Automatic estimation of number of replications, warm-up period and run length. L. H. Lee, M. E. Kuhl, J. W. Fowler, S. Robinson, eds. *Proc. 2009 INFORMS Simulation Soc. Res. Workshop, Coventry, UK*, INFORMS, Hanover, MD, 75–79.

Law, A. M. 1983. Statistical analysis of simulation output data. *Oper. Res.* **31**(6) 983–1029.

Law, A. M. 2007. *Simulation Modeling and Analysis*, 4th ed. McGraw-Hill, New York.

Law, A. M., J. S. Carson. 1979. A sequential procedure for determining the length of a steady-state simulation. *Oper. Res.* **27**(5) 1011–1025.

Luenberger, D. G. 2003. *Linear and Nonlinear Programming*, 2nd ed. Springer, New York.

Ray, W. D. 1957. Sequential confidence intervals for the mean of a normal population with unknown variance. *J. Roy. Statist. Soc. Ser. B* **19**(1) 133–143.

Robbins, H. 1959. Sequential estimation of the mean of a normal population. U. Grenander, ed. *Probability and Statistics: The Harald Cramér Volume*. John Wiley & Sons, New York, 235–245.

Schmeiser, B., Y. Yeh. 2002. On choosing a single criterion for confidence-interval procedures. E. Yücesan, C. H. Chen, J. L. Snowdon, J. M. Charnes, eds. *Proc. 2002 Winter Simulation Conf*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 345–352.

Schruben, L. W. 1980. A coverage function for interval estimators of simulation response. *Management Sci.* **26**(1) 18–27.

Singham, D. I., L. W. Schruben. 2009. Analysis of sequential stopping rules. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, R. G. Ingalls, eds. *Proc. 2009 Winter Simulation Conf.*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 724–730.

Sproule, R. N. 1985. Sequential nonparametric fixed-width confidence intervals for *U*-statistics. *Ann. Statist.* **13**(1) 228–235.

Starr, N. 1966. The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37**(1) 36–50.

Steiger, N. M., J. R. Wilson. 2002. An improved batch means procedure for simulation output analysis. *Management Sci.* **48**(12) 1569–1586.

Tafazzoli, A., N. M. Steiger, J. R. Wilson. 2011. N-Skart: A nonsequential skewness and autoregression-adjusted batch-means procedure for simulation analysis. *IEEE Trans. Automatic Control* **56**(1) 1–11.

Welford, B. P. 1962. Note on a method for calculating corrected sums of squares and products. *Technometrics* **4**(3) 419–420.