A Sequential Procedure for Simultaneous Estimation of Several Means

Kimmo E. E. Raatikainen
University of Helsinki, Finland

Sequential procedures for controlling the length of a simulation run are widely used when a single mean is estimated. In many practical situations, however, the analyst is simultaneously interested in several means. We propose a sequential procedure for controlling the length of a simulation run when several means are simultaneously estimated. A lower bound on the overall level of confidence for a set of interval estimates on each mean is obtained with the Bonferroni inequality. The primary advantage of applying the Bonferroni inequality is that the precisions are easy to interpret. In addition, we can use the existing methods for estimating the variances of the means, since the covariances of estimates are not needed.

Categories and Subject Descriptors: I.6.6 [**Simulation and Modeling**]: Simulation Output Analysis—*sequential estimation*

General Terms: Algorithms, Experimentation, Performance

Additional Key Words and Phrases: Run length control, multiple response variables

---

*Author's address: University of Helsinki, Department of Computer Science, P. O. Box 26 (Teollisuuskatu 23), FIN-00014 University of Helsinki, Finland. E-mail:Kimmo.Raatikainen@Helsinki.FI
Manuscript of the paper published in *ACM Transactions on Modeling and Computer Simulation 3, 2* (April 1993): 108–133

1. INTRODUCTION

The methods for controlling the precision of simulation results have concentrated on a single mean, for example on the mean residence or waiting time in a queueing system. These methods are well-known and described in many textbooks and survey articles, for example in [20] and in [23]. As explained in [32] the methods can also be used when other indices based on cumulative statistics are estimated.

In many practical applications, however, the analyst is simultaneously interested in several means, for example in the means of the waiting times at each server in a queueing network model and in the mean residence time in the whole model. Another example of several means of interest is a queueing system with several classes of customers when the mean residence time of each class is estimated.

The output sequences in the two examples above are different. In the first example, the output sequence $\{\boldsymbol{x}_i\}_{i=1}^{n}$ is a multivariate one: $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$. In the second example we have one sequence for each class. These output sequences, $\{x_{ij}\}_{i=1}^{n_j}$, $j = 1, \ldots, p$, are simultaneous in the sense that they are generated by a single simulation run. However, the sequences are asynchronous, since observations with the same observation number from different sequences are not necessarily related to each other in any way. In some situations we may have a mixture of the two cases, asynchronous multivariate sequences. An example of such a case is a queueing system with several classes of customers when the residence time and mean throughput of each class is estimated.

In recent years some papers concerning multivariate output sequences have been published. The research reported includes methods for estimating the confidence region and the covariance matrix [4, 7, 15, 29, 13, 30], comparisons between different methods [5, 6], methods for detecting the initialization bias [28], and variance reduction techniques [3, 31, 33, 34]. Validation of multivariate simulation models is discussed in [2].

The importance of sequential procedures for estimating a single mean is widely recognized. Clearly, the same is true for simultaneously estimating several means. However, the methods proposed for estimating the confidence region are based on predefined (fixed) length of the output sequence. In this paper we propose a sequential procedure for simultaneous estimation of several means. Our procedure is based on the Bonferroni inequality. The primary advantage of using the Bonferroni inequality is that the precisions of estimated means are easy to interpret. Secondly, the variances of estimates alone are needed. Therefore, we can use the methods developed for estimating a single mean, such as the spectral method introduced in [12] or the method of batch means introduced in [18]. Thirdly, the method can be used to control the run length even if the output sequences are asynchronous.

In Section 2 we introduce the sequential procedure, the requirement for the precisions, and the termination rule which automatically terminates the simulation run as soon as the estimated precisions meet the prespecified requirement. In Section 3 we examine the properties of the proposed termination rule. We show that our sequential procedure has a sound theoretical basis. In Section 4 we report experimental results from four queueing systems. Finally, in Section 5 we discuss some alternative methods that are based on the estimated covariance matrix.

2. A SEQUENTIAL PROCEDURE FOR MULTIPLE MEANS

There are two fundamental questions in developing a sequential procedure for multiple means. The first one is how we specify the desired precisions for the estimated means. The second one is the termination rule, that is how we conclude whether the precisions are high enough.

The requirement for the precision must be meaningful and easy to interpret. This suggests that the requirement should be a natural generalization of the relative half-width criterion, which is widely used when a single mean is estimated. The termination rule must be as simple and general as possible. Ideally, the rule should be computationally as easy as the relative half-width criterion. In addition, the assumptions about the output sequences should be the same as in the case of estimating a single mean.

We start this section by introducing a precision requirement that generalizes the relative half-width criterion. Next we introduce the termination rule. Since the rule is based on the Bonferroni inequality, our method needs the same assumptions as the methods developed for estimating a single mean. The assumption is that each (marginal) output sequence $\{x_{ij}\}_{i=1}^{\infty}$, $j = 1, \ldots, p$, is covariance stationary. Finally, we show that the sequential procedure proposed is a straightforward extension of procedures developed for a single mean.

2.1 Precision Requirement

When a single mean θ is estimated, the precision requirement is usually specified using the relative half-width criterion. The analyst specifies the desired relative precision, ε, of the $100(1-\alpha)\%$ confidence interval. Hence, the simulation is then continued until $\widehat{\Pr}\{|\hat{\theta}-\theta| \leq \varepsilon|\hat{\theta}|\} \geq 1-\alpha$,which implies that with an estimated probability of $1-\alpha$ (at least) the relative error of the estimate $\hat{\theta}$, $|\hat{\theta}-\theta|/|\theta|$, is at most $\varepsilon/(1-\varepsilon)$; see e.g. [23].

When several means, $\theta_1, \ldots, \theta_p$, are estimated, there are several possible ways to generalize the relative half-width criterion. When the means are of interest both as a vector and separately, it is natural to require that with a predefined probability the relative error of each estimate should be less than a predefined tolerance. Therefore, we use the following straightforward generalization of the relative half-width criterion: The simulation is continued until

$$\widehat{\Pr}\{|\theta_1 - \hat{\theta}_1| \leq \varepsilon_1 |\hat{\theta}_1| \ \wedge \ \ldots \ \wedge \ |\theta_p - \hat{\theta}_p| \leq \varepsilon_p |\hat{\theta}_p|\} \geq 1 - \alpha \ , \tag{2.1}$$

which implies that with an estimated probability of $1 - \alpha$ (at least), the relative errors of the estimates are at most $\varepsilon_j/(1-\varepsilon_j)\,,\ j = 1, \ldots, p$.

2.2 Termination Rule

The methods for controlling the precision of a single mean assume that the estimate $\hat{\theta}$ has a normal limiting distribution. Therefore, they use the Student t-distribution to approximate the distribution of $(\hat{\theta} - \theta)/s_{\hat{\theta}}$ and they terminate the simulation when

$$s_{\hat{\theta}}/|\hat{\theta}| \leq \varepsilon/t_\nu(1 - \alpha/2)\,. \tag{2.2}$$

In (2.2) $s^2_{\hat{\theta}}$ is the estimated variance of $\hat{\theta}$ with ν degrees of freedom and $t_\nu(x)$ is the $100x^{\text{th}}$ percentile of the Student t-distribution with ν degrees of freedom.

When several means are estimated, we can use the Bonferroni inequality (see e.g. [20, Ch. 9.7]) in the following manner. If each $\hat{\theta}_j$ satisfies $\widehat{\Pr}\{|\theta_j - \hat{\theta}_j| \leq \varepsilon_j|\hat{\theta}_j|\} \geq 1 - \alpha_j$ and $\sum_{j=1}^{p} \alpha_j \leq \alpha$, we can conclude that the requirement (2.1) is satisfied.

The α_j's in the Bonferroni inequality may be different. However, it is usually difficult in practice to prespecify optimal α_j's. Therefore, the most common way of applying the Bonferroni inequality is to set $\alpha_j = \alpha/p\,,\ j = 1, \ldots, p$. The simulation is then continued until the $100(1-\alpha/p)\%$ confidence interval for each θ_j is narrow enough. When the Student t-distribution is used to approximate the distributions of $(\hat{\theta}_j - \theta_j)/s_{\hat{\theta}_j}$, the simulation should be continued until

$$s_{\hat{\theta}_j}/|\hat{\theta}_j| \leq \varepsilon_j/t_{\nu_j}(1 - \alpha/2p)\,,\ \forall\, j = 1, \ldots, p\,, \tag{2.3}$$

where $s^2_{\hat{\theta}_j}$ is the estimated variance of $\hat{\theta}_j$ with ν_j degrees of freedom.

The termination rule (2.3) usually leads to simulation runs which are unnecessarily long to fulfil the requirement (2.1). The reason is that the standard deviations of the estimates are usually not proportional to $\varepsilon_j|\hat{\theta}_j|/t_{\nu_j}(1-\alpha/2p)$. Instead, a typical situation is that only few of the estimates dominate the run length. Therefore, we apply the Bonferroni inequality in the same way as in [10].

Instead of constructing the $100(1 - \alpha/p)\%$ confidence intervals, we estimate the α_j's, that is the probabilities of $\{|\hat{\theta}_j - \theta_j| > \varepsilon_j|\hat{\theta}_j|\}$. When the Student t-distribution with ν_j degrees of freedom is used to approximate the distribution of $(\hat{\theta}_j - \theta_j)/s_{\hat{\theta}_j}$, we obtain $\hat{\alpha}_j = 2F_{\nu_j}\left(-\varepsilon_j|\hat{\theta}_j|/s_{\hat{\theta}_j}\right)$, where

F_{ν_j} is the cumulative distribution function of the Student t-distribution with ν_j degrees of freedom. The termination rule we propose is that the simulation is continued until

$$\sum_{j=1}^{p} \hat{\alpha}_j = 2 \sum_{j=1}^{p} F_{\nu_j} \left(-\varepsilon_j |\hat{\theta}_j| / s_{\hat{\theta}_j} \right) \leq \alpha \ . \tag{2.4}$$

The termination rule above implies that the requested joint confidence region is the rectangular bounded through lines $\hat{\theta}_j \pm \varepsilon_j \hat{\theta}_j$, $j = 1, \ldots, p$.

2.3 Implementation Notes

The sequential procedures developed for estimating a single mean estimate the mean and its variance ($\hat{\theta}$ and $s^2_{\hat{\theta}}$) at prespecified checkpoints $m_1, m_2, \ldots, m_{max}$. Typically, the sequence of checkpoints is geometrically increasing, i.e. $m_{k+1} = I\, m_k$, where I is a prespecified constant, usually $1 < I \leq 2$. The simulation is terminated at the first checkpoint in which $s_{\hat{\theta}}/|\hat{\theta}| \leq \varepsilon / t_\nu(1 - \alpha/2)$; see e.g. [12]. Our sequential procedure is a straightforward generalization of this procedure.

As discussed in the Introduction there are two different kinds of output sequences when several means are estimated. Firstly, the output sequence may be a multivariate one: $\{x_{ij}\}_{i=1}^{n}$, $j = 1, \ldots, p$. Secondly, the output sequences may be asynchronous: $\{x_{ij}\}_{i=1}^{n_j}$, $j = 1, \ldots, p$. The two types of output sequences require slightly different implementations.

The multivariate case is a straightforward generalization of the procedure outlined above. The means and their variances ($\hat{\theta}_j$ and $s^2_{\hat{\theta}_j}$, $j = 1, \ldots, p$) are estimated at prespecified checkpoints $m_1, m_2, \ldots, m_{max}$; $m_{k+1} = I\, m_k$. The simulation is terminated at the first checkpoint in which $2 \sum_{j=1}^{p} F_{\nu_j}(-\varepsilon_j |\hat{\theta}_j| / s_{\hat{\theta}_j}) \leq \alpha$.

When the output sequences are asynchronous, the situation is slightly more complex. There are several possible ways of implementing the procedure. We have decided to estimate the mean ($\hat{\theta}_j$), its variance ($s^2_{\hat{\theta}_j}$), and the probability $\hat{\alpha}_j = \widehat{\Pr}\{|\hat{\theta}_j - \theta_j| > \varepsilon_j |\hat{\theta}_j|\} = 2F_{\nu_j}(-\varepsilon_j |\hat{\theta}_j| / s_{\hat{\theta}_j})$, when the sequence reaches the checkpoint m_k ($n_j = m_k$). These estimates are regarded as the current estimates until the sequence reaches the next checkpoint. This implies that we ignore some observations available in all but one sequences. The procedure is initialized by setting $\hat{\alpha}_j = 1$ and $n'_j = m_1$, $j = 1, \ldots, p$. When n_j reaches the checkpoint ($n_j = n'_j$), we estimate $\hat{\theta}_j$, ($s^2_{\hat{\theta}_j}$), and $\hat{\alpha}_j$. If $\hat{\alpha}_1 + \ldots + \hat{\alpha}_p \leq \alpha$, we terminate the simulation. Otherwise, we set $n'_j = I n'_j$.

The decision of ignoring some observations was based on the fact that the adopted method for estimating the variances is considerably more efficient (computationally) when the checkpoints are powers of 2. Other approaches are also plausible, particularly if the efficiency of the method for estimating the variances does not depend on the sequence length.

3. PROPERTIES OF THE PROPOSED TERMINATION RULE

In this section we examine the properties of the termination rule proposed in Eq. (2.4). We show that the confidence intervals constructed are asymptotically valid provided that the method of estimating the variances of the means is asymptotically reliable. We also show that the proposed termination rule is asymptotically the most efficient way of applying the Bonferroni inequality. In addition to the asymptotic properties we examine the expected run length in small samples. We compare the proposed rule to the traditional way of applying the Bonferroni inequality and to the case of estimating a single mean.

3.1 Asymptotic Coverage of Constructed Confidence Intervals

In this section we examine the asymptotic coverage of the confidence intervals constructed. Our objective is to show that the termination rule (2.4) provides confidence intervals that satisfy

$$1-\alpha \le \lim_{\varepsilon_1,\ldots,\varepsilon_p \to 0} \Pr\{|\hat\theta_1-\theta_1| \le \varepsilon_1|\hat\theta_1| \wedge \ldots \wedge |\hat\theta_p-\theta_p| \le \varepsilon_p|\hat\theta_p|\} < 1\,. \tag{3.1}$$

When we can show that (3.1) holds, we can trust that for small ε_j's the lower bound for the true probability that all relative errors of estimates are at most $\varepsilon_j/(1-\varepsilon_j)$ is close to $1-\alpha$. Unfortunately, the asymptotic coverage does not tell us how small the ε_j's should be. The primary objective of the asymptotic analysis is to show that the proposed method has a sound theoretical basis. Moreover, the assumptions must be precisely formulated.

The termination rule in Eq. (2.2) for estimating a single mean can be written as

$$N_j = \text{smallest integer } n_j \text{ such that } F_{\nu_j}\left(-\varepsilon_j|\hat\theta_j(n_j)|/s_{\hat\theta_j(n_j)}\right) \le \alpha_j/2\,, \tag{3.2}$$

where $\hat\theta_j(n_j)$ is the estimate of θ_j and $s^2_{\hat\theta_j(n_j)}$ is the estimated variance of $\hat\theta_j(n_j)$ with ν_j degrees of freedom. The termination rule (3.2) provides an asymptotically consistent confidence interval, that is

$$\lim_{\varepsilon_j\to 0} \Pr\{|\hat\theta_j(N_j)-\theta_j| \le \varepsilon_j|\hat\theta_j(N_j)|\} = 1-\alpha_j\,,$$

if the output sequence is covariance stationary and if $s^2_{\hat\theta_j(n_j)}$ is a strongly consistent estimator of $\mathrm{Var}(\hat\theta_j(n_j))$. The strong consistency means that $\nu_j s^2_{\hat\theta_j(n_j)}/\mathrm{Var}(\hat\theta_j(n_j))$ converges in distribution to $\chi^2_{\nu_j}$-distribution (with ν_j degrees of freedom) and $\nu_j \to \infty$ as $n_j \to \infty$. The proof of the asymptotic consistency is a straightforward application of results given in [8, 17, 22]; see also [9, 21].

Most of the methods for estimating $\mathrm{Var}(\hat\theta_j(n_j))$ that are currently used do not satisfy the condition that $\nu_j \to \infty$ as $n_j \to \infty$. However, if $s^2_{\hat\theta_j(n_j)}$ converges (almost surely) to $\mathrm{Var}(\hat\theta_j(n_j))$ as $n_j \to \infty$ and the distribution of $\nu_j s^2_{\hat\theta_j(n_j)}/\mathrm{Var}(\hat\theta_j(n_j))$ is approximated by the $\chi^2_{\nu_j}$ distribution with constant (ν_j) degrees of freedom, then the termination rule in Eq. (3.2) provides a confidence interval that satisfies

$$1-\alpha_j \le \lim_{\varepsilon_j \to 0} \Pr\{|\hat\theta_j(N_j)-\theta_j| \le \varepsilon_j|\hat\theta_j(N_j)|\} < 1\,. \tag{3.3}$$

If the method for estimating the variance of the mean provides confidence intervals that satisfy (3.3), we say that the method is asymptotically reliable and that the confidence intervals are asymptotically valid. For details, see [26].

We assume that the method for estimating the variance of the means leads to confidence intervals that are asymptotically valid. Let α_j, $j=1,\ldots,p$, be arbitrary positive constants such that $\sum_{j=1}^p \alpha_j = \alpha$. Using the Bonferroni inequality we obtain

$$1-\alpha = 1-\sum_{j=1}^p \alpha_j \le \lim_{\varepsilon_1,\ldots,\varepsilon_p \to 0} \Pr\{|\hat\theta_1(N_1)-\theta_1| \le \varepsilon_1|\hat\theta_1(N_1)| \wedge \ldots \wedge |\hat\theta_p(N_p)-\theta_p| \le \varepsilon_p|\hat\theta_p(N_p)|\}\,.$$

Since $\lim_{\varepsilon_j \to 0} \Pr\{|\hat\theta_j(N_j)-\theta_j| \le \varepsilon_j|\hat\theta_j(N_j)|\} < 1$, $j=1,\ldots,p$, we obtain

$$\lim_{\varepsilon_1,\ldots,\varepsilon_p \to 0} \Pr\{|\hat\theta_1(N_1)-\theta_1| \le \varepsilon_1|\hat\theta_1(N_1)| \wedge \ldots \wedge |\hat\theta_1(N_p)-\theta_p| \le \varepsilon_p|\hat\theta_1(N_p)|\} < 1\,.$$

Therefore any choice of α_j's such that $\sum_{j=1}^p \alpha_j = \alpha$ provides confidence intervals that are asymptotically valid, if N_j is determined as in (3.2), $\hat\theta_j = \hat\theta_j(N_j)$, and the method for estimating the variance of the means is asymptotically reliable. However, the length of simulation run depends on the choice of α_j's.

3.2 Best Choice of Individual Confidence Levels in Bonferroni Inequality

In this section we show that the proposed termination rule implicitly chooses the α_j's which minimize the length of simulation run when the ε_j's are small. When the output sequence is a multivariate one, the length of the simulation run is $N=\max\{N_1,\ldots,N_p\}$. When the output sequences are asynchronous and generated by a single simulation run, the length of the simulation run is $N=N_1+\ldots+N_p$.

In the multivariate case the best choice of the α_j's is the solution of

$$\begin{array}{rl} \underset{\alpha_1,\ldots,\alpha_p}{\text{minimize}} & \max\{N_1,\ldots,N_p\} \\ \text{subject to} & 2F_{\nu_j}\big(-\varepsilon_j|\hat\theta_j(N_j)|/s_{\hat\theta_j(N_j)}\big) \le \alpha_j\,,\ j=1,\ldots,p, \\ \text{and} & \displaystyle\sum_{j=1}^p \alpha_j = \alpha\,. \end{array}$$

Because of the definition of N_j in Eq. (3.2) each N_j is an implicit function of α_j. Therefore, the solution of the following integer optimization problem,

$$\begin{aligned} &\underset{N_0,\ldots,N_p}{\text{minimize}} && N_0 \\ &\text{subject to} && \alpha - 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j|\hat{\theta}_j(N_j)|/s_{\hat{\theta}_j(N_j)}\big) \geq 0 \\ &\text{and} && N_0 - N_j \geq 0\,,\ j=1,\ldots,p\,, \end{aligned} \tag{3.4}$$

provides the best choice of α_j's.

In a general case the analytic solution of an integer optimization problem with non-linear constraints is unknown. However, under the assumptions we have made the solution of (3.4) is

$$N_0 = \text{smallest integer } n \text{ such that } 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j|\hat{\theta}_j(n)|/s_{\hat{\theta}_j(n)}\big) \leq \alpha\,.$$

Individual N_j's are not necessarily unique, but at least one N_j is equal to N_0. The assumptions we have made are: 1) the (marginal) output sequences are covariance stationary, 2) each $s^2_{\hat{\theta}_j(n)}$ converges (almost surely) to $\mathrm{Var}(\hat{\theta}_j(n))$ as $n \to \infty$, and 3) the ε_j's are small enough. For details, see the Appendix.

When the output sequences are asynchronous, then for any given run length $N = N_1 + \ldots + N_p$ the individual N_j's can be written as $N_j = N\pi_{j|N}$, $j = 1,\ldots,p$, and $\pi_{1|N} + \ldots + \pi_{p|N} = 1$. Usually the $\pi_{j|N}$'s are random variables. Therefore, we must assume, in addition to the three assumptions above, that the sequences $\{\pi_{j|N}\}_{N=1}^{\infty}$ converge (almost surely) to positive constants π_j as $N \to \infty$ and $\pi_1 + \ldots + \pi_p = 1$. The practical implication of this assumption is that further observations from each sequence can be obtained.

The best choice of α_j's is implied through the solution of the following integer optimization problem

$$\begin{aligned} &\underset{N_0,\ldots,N_p}{\text{minimize}} && N_0 \\ &\text{subject to} && \alpha - 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j|\hat{\theta}_j(N_j)|/s_{\hat{\theta}_j(N_j)}\big) \geq 0 \\ &\text{and} && N_0\pi_{j|N_0} - N_j \geq 0\,,\ j=1,\ldots,p\,. \end{aligned} \tag{3.5}$$

We show in the Appendix that under the assumptions we have made the solution of (3.5) is $N_j = N_0\pi_{j|N_0}$, $j = 1,\ldots,p$, and

$$N_0 = \text{smallest integer } n = n_1 + \ldots + n_p \text{ such that } 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j|\hat{\theta}_j(n_j)|/s_{\hat{\theta}_j(n_j)}\big) \leq \alpha\,.$$

Therefore the termination rule proposed in Eq. (2.4) is asymptotically the most efficient way of applying the Bonferroni inequality. In addition, an immediate consequence of the non-linear constraints in (3.4) and (3.5) is that no choice of α_j's is more efficient than the choice implied through (2.4) when the ε_j's are small enough. The ε_j's are small enough when the coefficients of variation, $s_{\hat{\theta}_j(n)}/|\hat{\theta}_j(n)|$, are non-increasing for each $n \geq n_j$ and for each $(n_1, \ldots, n_p)$ such that $2\sum_{j=1}^{p} F_{\nu_j}(-\varepsilon_j|\hat{\theta}_j(n_j)|/s_{\hat{\theta}_j(n_j)}) \leq \alpha$.

3.3 Expected Run Length in Small Samples

In this section we examine the expected run length of the proposed termination rule (Eq. 2.4) when the ε_j's are positive constants. The analysis is based on the assumption that $\hat{\theta}_j \sim N(\theta_j, \sigma_j^2/n)$ and s_j^2 is an estimate of σ_j^2 with ν degrees of freedom. The expected run lengths are analytically evaluated. We compare the proposed rule to the traditional way of applying the Bonferroni inequality (the termination rule in Eq. 2.3) and to the case in which only a single mean (θ_1) is estimated (the termination rule in Eq. 2.2).

Figure 1 shows the relative decreases (per cent) in the expected run length, when the proposed rule is used instead of the termination rule in Eq. (2.3). The relative decrease,

$$100\big(1 - \mathrm{E}\{N|\sum_{j=1}^{p}\alpha_j = \alpha\}/\,\mathrm{E}\{N|\alpha_1 = \ldots = \alpha_p = \alpha/p\}\big)\ \%\,,$$

represents how much we save through using the proposed way of applying the Bonferroni inequality instead of the traditional way.

Figure 2 shows the relative increases (per cent) in the expected run length, when additional means $(\theta_2, \ldots, \theta_p)$ are estimated using the proposed rule. The relative increase,

$$100\big(\,\mathrm{E}\{N|\sum_{j=1}^{p}\alpha_j = \alpha\}/\,\mathrm{E}\{N|\alpha_1 = \alpha\} - 1\big)\ \%\,,$$

indicates how much we must pay for the additional means.

The expected run lengths are affected by factors: α, ν, and $\varepsilon_j|\theta_j|/\sigma_j$. The last factor is the allowed relative half-width (ε_j) divided by the coefficient of variation ($\sigma_j/|\theta_j|$). The squared inverse of the factor, $C_j = \sigma_j^2/(\varepsilon_j\theta_j)^2$, may be called *the run length factor* of the estimate. Loosely speaking we can say that the component j is dominant with respect to the component k, if the ratio $C_j/C_k \geq 1$.

The cases examined are a two-dimensional case, a three-dimensional case, and a general multidimensional case. In the two-dimensional case, we show the relative decreases and increases as a function of C_1/C_2 for $\alpha = 0.10$, 0.05, and 0.01. In the three-dimensional case, we show the relative

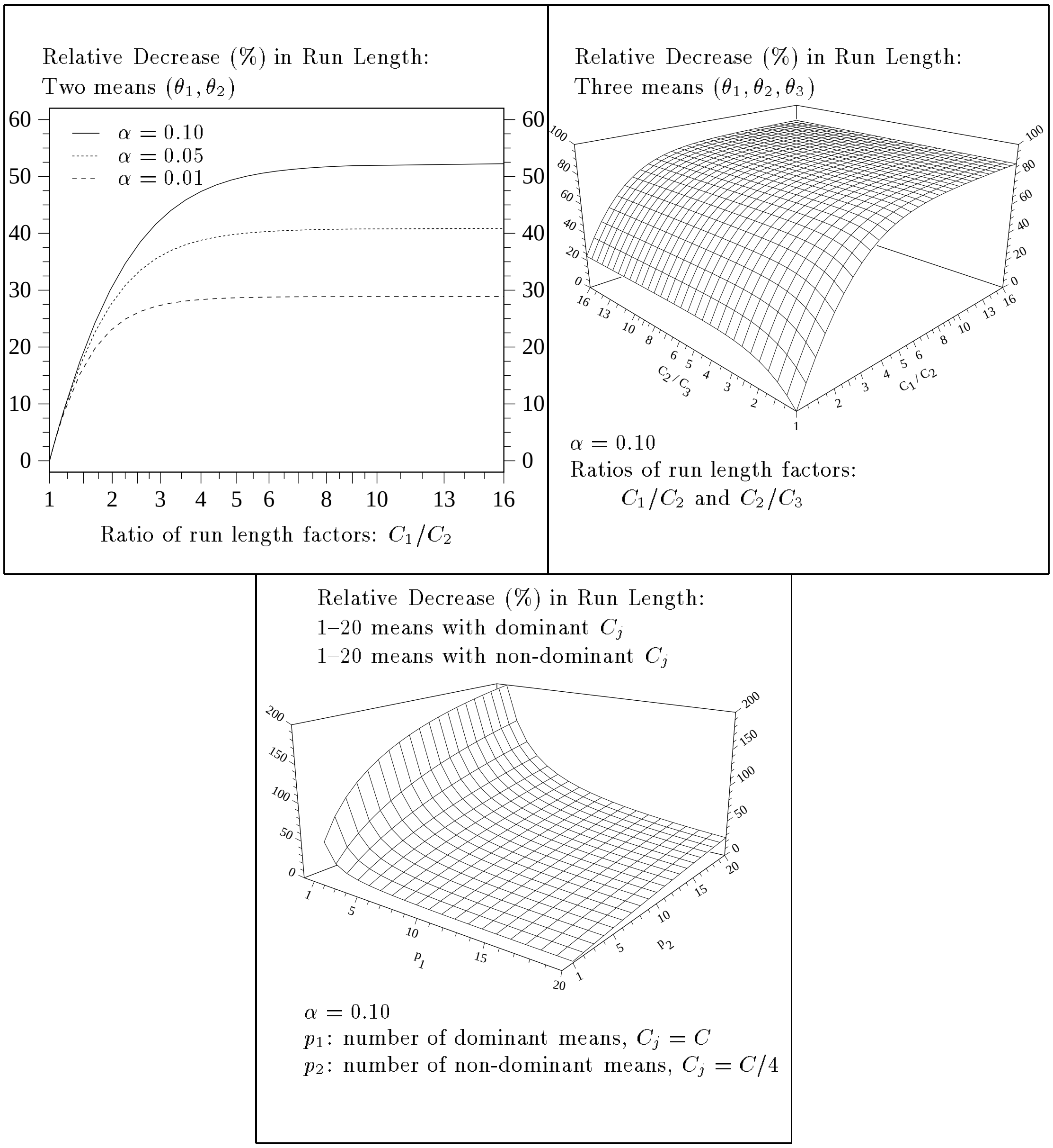


Fig. 1: Expected Relative Decrease in Run Length,
when the proposed way of applying the Bonferroni inequality is used instead of the traditional way

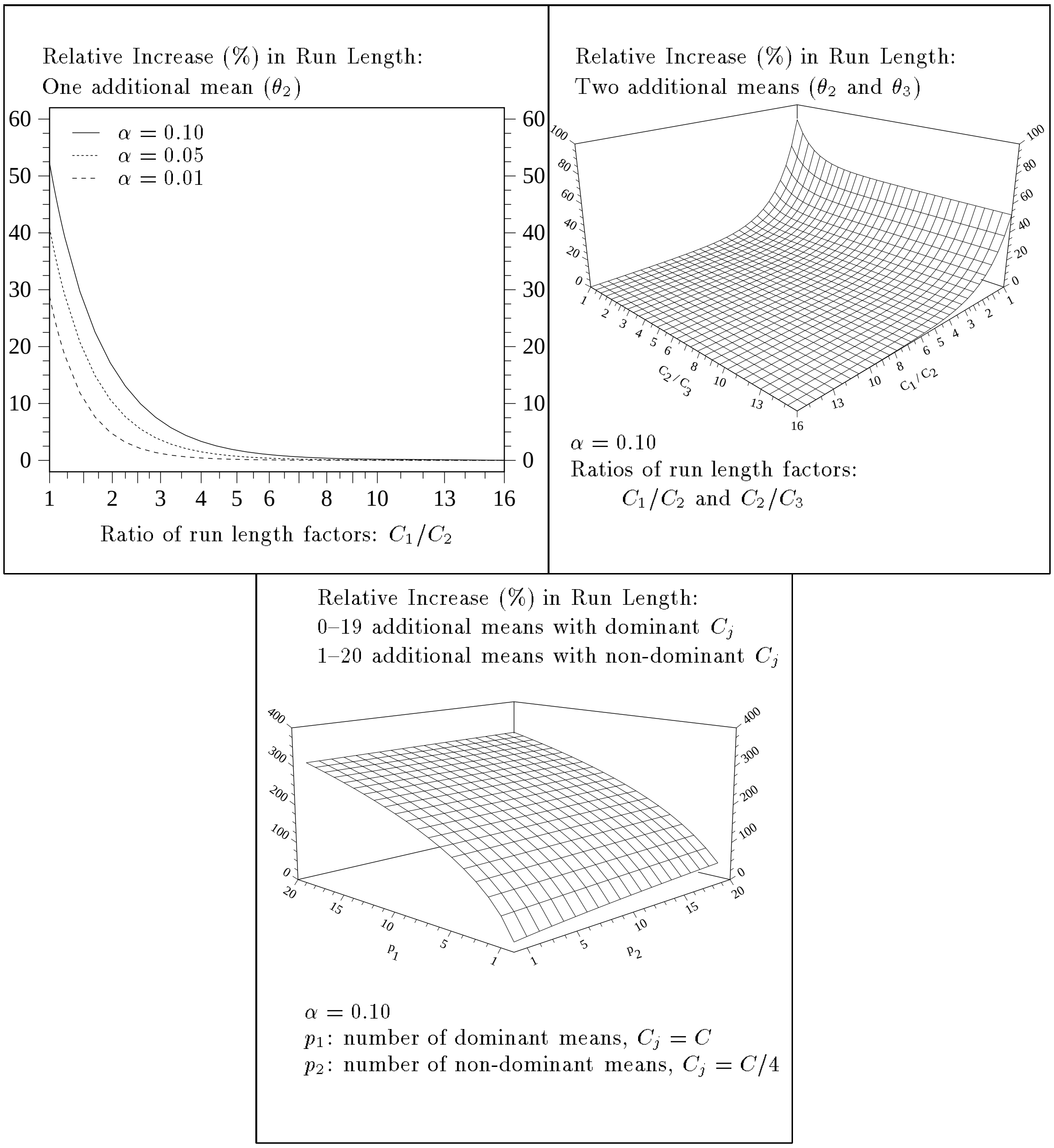


Fig. 2: Expected Relative Increase in Run Length,
when additional means are estimated using the proposed method

decreases and increases as a function of C_1/C_2 and C_2/C_3. In the general multidimensional case, the relative decreases and increases are shown as a function of p_1 and p_2, where p_1 is the number of dominant components and p_2 is the number of non-dominant components: $C_1 = \ldots = C_{p_1} = C$ and $C_{p_1+1} = \ldots = C_{p_1+p_2} = C/4$.

In the three-dimensional and the multidimensional case we examined only the confidence level 0.90, i.e. $\alpha = 0.10$. In all three cases the degrees of freedom examined are 9. The reason for $\nu = 9$ is that the spectral method which we use in the experiments provides variance estimates with 9 degrees of freedom. When the degrees of freedom are higher, then the relative decreases and increases are slightly lower.

The figures clearly demonstrate the efficiency of the proposed termination rule. Particularly, the relative increase in the expected run length (when compared to the estimation of one dominant component) is significant only if two or more components are dominant.

4. EXPERIMENTATION

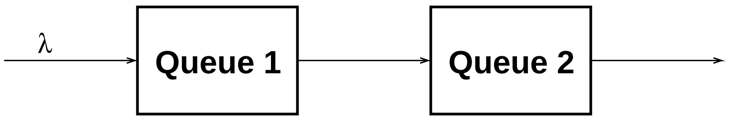
We examined empirically the proposed sequential procedure. We give results from the proposed and the traditional way of applying the Bonferroni inequality. In addition, we give results from experiments when a single mean was estimated. Before the results are presented, we describe the experiments and summarize the spectral method introduced in [12] that we used to estimate the variances of the means.

4.1 Description of Experiments

In the experiments we examined the four different queueing systems shown in Figure 3. The first three queueing systems are tandem queues of two exponential FIFO-servers. The fourth system is an open queueing network model with two classes of customers and two FIFO-servers. The means estimated in Models 1–3 (the tandem queues with $\rho_1 = 0.5$, 0.8, and 0.9, respectively) are the mean waiting times in **Queue 1** (W_1) and in **Queue 2** (W_2), and the mean residence time in the system (R). In these models we have a multivariate output sequence. The means estimated in Model 4 are the mean response time for Class A (R_A) and Class B (R_B). In this model we have two asynchronous output sequences, one for each class.

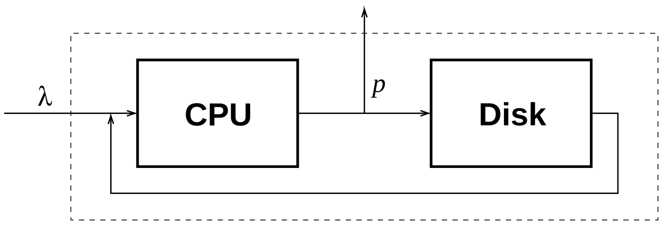
The parameters controlling the simulation run length were $\alpha = 0.10$, 0.05, and $\varepsilon = 0.20$, 0.10, 0.05. For each mean estimated the same tolerance was used ($\varepsilon_j = \varepsilon\,,\ j = 1, \ldots, p$). In an experiment the same output of simulation was examined using the proposed way of applying the Bonferroni inequality (termination rule in Eq. 2.4) and the traditional way (termination rule in Eq. 2.3), and the spectral method for a single mean (termination rule in Eq. 2.2). The experiment was independently repeated 101 times.

Models 1–3: Tandem Queue



Parameters	
λ	1
ρ_1	0.5, 0.8, and 0.9
ρ_2	0.5

Model 4: Queueing Network Model with Two Classes of Customers



Class	λ	p	μ_{cpu}	μ_{disk}
A	3/19	0.1	10	3
B	2/19	0.2	2.5	1

Fig. 3: Queueing Systems Examined

The simulations were initialized to the steady-state phase. The initial state of the system was generated according to the steady-state distribution. In the tandem queues the waiting time of the first arriving customer at **Queue 1** and at **Queue 2** was generated from the known steady-state distribution; see e.g. [16, p. 401]. Model 4 does not have a product-from solution. Therefore, we carried out a pilot study. The objectives of the pilot study were to obtain independent estimates for the unknown means of response times and to obtain the distribution of queue lengths at arrival instants. The mean response times obtained are regarded as the exact means when the results of actual experiments are evaluated. The observed frequencies of queue lengths at arrival instants were used when the waiting time of the first arriving customer was generated.

In the pilot study the initial state was the idle and empty system. The initialization bias was removed in the following manner. We modified our sequential procedure so that the number of batches was $3M$ for both output sequences. The mean response time and its variance was estimated from batches $M+1,\ldots,2M$ and from batches $2M+1,\ldots,3M$. The simulation was continued until the precisions of the four means (two means for Class A and two means for Class B) were high enough ($\alpha = 0.01$, $\varepsilon = 0.01$) and the relative difference between the two means for Class A and between the two means for Class B were both less than 1%. The queue lengths at arrival instants were recorded during the second half of the simulation.

In the actual experiments of Model 4, the waiting time of the first arriving customer at **CPU** and at **Disk** was generated in the following way. First we generated (using the observed frequencies of queue lengths) the number of customers in Class A and Class B at the server. The waiting time was then generated as the sum of services times for these customers. (Please, note that customers of Class A see a different queue length distribution than customers of Class B.)

4.2 Summary of Spectral Method

The variances of the means were estimated using the spectral method described in [12]. The method is regarded as one of the most reliable and robust methods for estimating the variance of means [23]. The primary advantage of the spectral method is that the batch means need not be uncorrelated. This is in contrast to the method of batch means introduced in [18].

The spectral method uses batch means of the output sequences. The number of batches varies from M to $2M-1$. A polynomial of order d is fitted to the first K ordinates of the bias corrected logarithms of smoothed periodogram. Periodogram ordinates are evaluated using Fast Fourier transform and smoothing is done by averaging two adjacent ordinates. The variance is estimated at prespecified checkpoints $m_1, m_2, \ldots, m_{max}$; $m_{k+1} = I m_k$. We have used the following values: $M = 512$, $d = 2$, $K = 31$, $m_1 = 512$, $m_{max} = 2^{30}$, and $I = 2$. The degrees of freedom in the variance estimate are 9. For details, see [12] or [23].

4.3 Empirical Results

The results of the experiments are given in Tables I–IV for Models 1–4, respectively. The two first columns of the tables give the parameters controlling the run length, ε and α. Column 3 gives the equation number of the termination rule: 2.4 denotes the proposed way and 2.3 denotes the traditional way of applying the Bonferroni inequality while $2.2/X$ denotes that a single mean, X, is estimated. Columns 4 and 5 give the mean and the standard deviation of the simulation run length.

The observed precisions are summarizes in Column 6 (Joint Coverage). We report the fraction of runs in which the estimated means are close enough to the true means, i.e. $|\hat{\theta}_j - \theta_j| \leq \varepsilon_j |\hat{\theta}_j|$ for each estimated mean. The coverage reported measures the observed precision using the the precision requirement in Eq. (2.1). We give the observed joint probability of $\{|\hat{\theta}_j - \theta_j| \leq \varepsilon_j |\hat{\theta}_j|\ ,\ j = 1, \ldots, p\}$ for termination rules (2.3) and (2.4). When a single mean is estimated, the observed probability of $\{|\hat{\theta}_j - \theta_j| \leq \varepsilon_j |\hat{\theta}_j|\}$ for that mean is given.

In addition to the coverage and mean run length, the volume of the estimated confidence region or the (relative) half-width of the estimated confidence interval has been summarized in empirical studies. The comparison of the estimated half-width (or volume) to the request value indicates the efficiency of the method for estimating the variance and how efficiently the termination rule applies the estimated variances. For the proposed method the volume of the rectangular confidence region is $\varepsilon_1|\hat{\theta}_1| \times \ldots \times \varepsilon_p|\hat{\theta}_p|$, which does not characterize the merits of the method. Because the simulation is continued until the estimated confidence level is high enough, the estimated α, see Eq. (2.4), is a natural substitute for the estimated confidence interval. Since $\hat{\alpha}$ is a function of the estimated means and their estimated standard deviations, it can also be evaluated for termination rules ((2.3) and (2.2). These rules can also be interpreted in the way that the simulation is continued until the estimated confidence level is high enough. Therefore the estimated α's characterize the same properties as the estimated half-widths or volumes.

Columns 7 and 8 in Tables I-IV summarize the estimated α's. We provide the mean and the standard deviation (reported as % of the nominal α). Because the objective of the termination rules is to terminate the simulation as soon as the estimated confidence level is high enough, that is as soon as $\hat{\alpha} \leq \alpha$, the ideal value of $\hat{\alpha}$ is 100% of α.

The observed coverage, that is the observed probability of $\{|\hat{\theta}_j - \theta_j| \leq \varepsilon_j|\hat{\theta}_j|\,,\ j = 1, \ldots, p\}$, exceeds the requested nominal level of $1 - \alpha$ in most of the cases. When the Bonferroni inequality is used this is not surprising since the inequality provides a lower bound.

When the experiment is independently repeated, we can use the binomial distribution and the observed coverages in testing the hypothesis that the true (unknown) coverage is high enough. We conclude that the termination rule works satisfactorily for the triple $(model, \varepsilon, \alpha)$ if the null-hypothesis H_0 can not be rejected at significance level α in the statistical test of

$$\begin{array}{lll} H_0: & \text{true coverage } \geq 1-\alpha & \quad vs. \\ H_A: & \text{true coverage } < 1-\alpha & \end{array}$$

This is a modification of the approach used in [17] to evaluate the observed coverage.

The null-hypothesis of acceptable coverage can only be rejected in six cases. These cases are indicated by '†' in Tables I–IV. When the null-hypothesis was rejected, the tolerance ε was 0.20 and a single mean was estimated. An obvious explanation is that the run length is not long enough for the t-approximation of $(\hat{\theta}_j - \theta_j)/s_{\hat{\theta}_j}$ to be appropriate; see also [19]. The overall conclusion is that the proposed termination rule controls the run length so that the estimates have the requested precision.

The means of run lengths are summarized in Figure 4. The figure displays the relative run lengths with respect to the proposed rule for the other rules examined. The figure clearly demonstrates the savings (13–40 %) when the proposed way of applying the Bonferroni inequality is used instead of the traditional way. The figure also indicates that one or two additional means do not always considerably increase the run length. The increase in run length is 6–83 % (mean 31.4 %, median 29 %), when compared to the estimation of W_1 in Models 1–3 and R_A in Model 4.

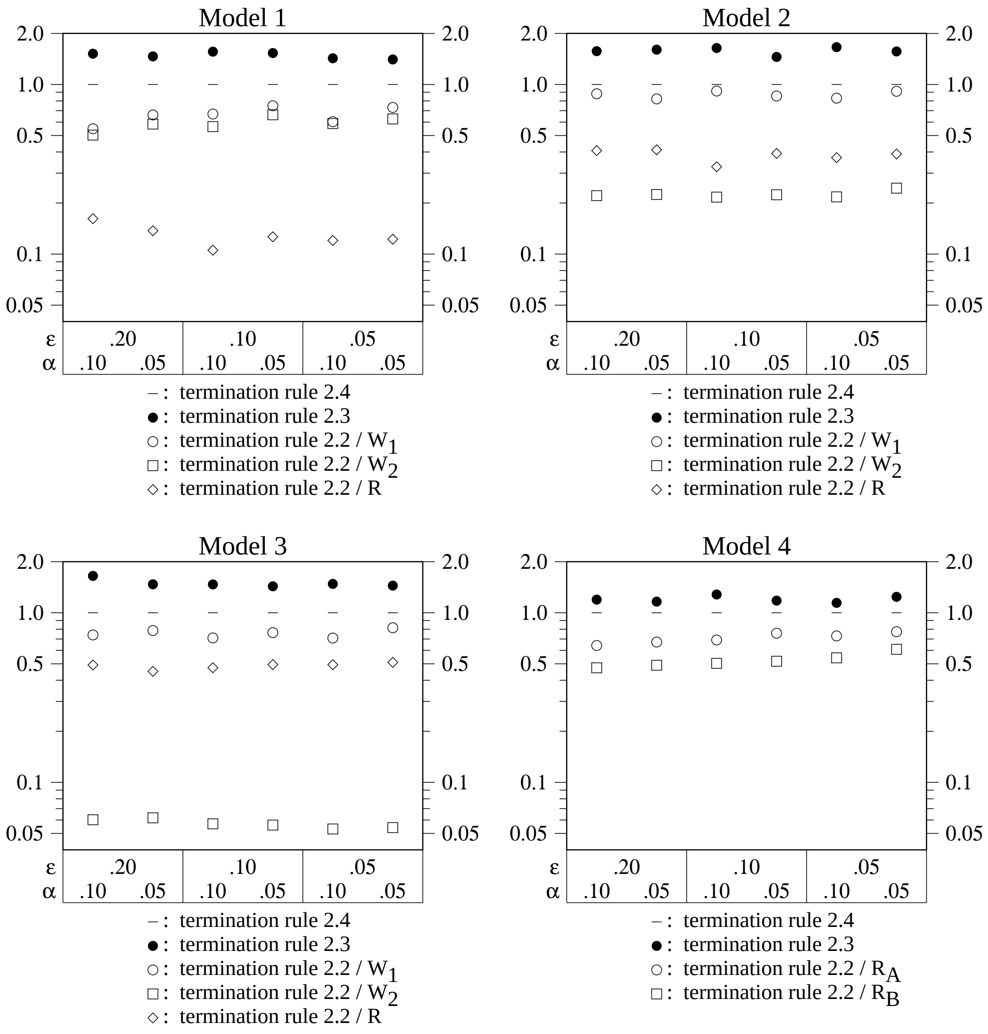


Fig. 4: Relative Means of Simulation Run Lengths

Table I: Summary of Results from Model 1

ε	α	Term. Rule	Run Length Mean	Run Length St.dev.	Joint Coverage	Estimated α (% of nominal α) Mean	Estimated α (% of nominal α) St.dev.
0.20	0.10	2.4	6773	3206	1.00	49.38	27.05
		2.3	10281	4805	1.00	19.56	13.26
		$2.2/W_1$	3711	2423	0.90	55.23	26.88
		$2.2/W_2$	3407	2074	0.91	53.93	28.16
		$2.2/R$	1095	260	1.00	30.10	25.48
0.20	0.05	2.4	9753	4026	1.00	42.96	26.75
		2.3	14275	6144	1.00	16.30	12.80
		$2.2/W_1$	6448	4015	0.95	45.32	28.06
		$2.2/W_2$	5698	2986	0.95	43.82	28.16
		$2.2/R$	1338	590	1.00	32.84	25.80
0.10	0.10	2.4	25874	10920	1.00	48.59	26.44
		2.3	40392	20949	1.00	19.61	13.08
		$2.2/W_1$	17317	8290	0.94	52.13	27.47
		$2.2/W_2$	14600	6897	0.95	51.41	24.17
		$2.2/R$	2727	1464	0.91	52.53	24.88
0.10	0.05	2.4	35688	14654	1.00	46.87	27.31
		2.3	54667	26617	1.00	19.01	13.28
		$2.2/W_1$	26725	13252	0.99	45.05	27.97
		$2.2/W_2$	23643	13243	0.99	48.85	28.76
		$2.2/R$	4512	2408	0.95	42.01	29.45
0.05	0.10	2.4	108686	42744	0.99	47.58	27.27
		2.3	155080	63037	0.99	21.76	14.46
		$2.2/W_1$	65536	31947	0.95	52.96	25.85
		$2.2/W_2$	63914	32483	0.97	51.38	27.40
		$2.2/R$	13079	7135	0.98	47.55	27.65
0.05	0.05	2.4	152485	59129	0.99	43.83	28.18
		2.3	214128	85469	0.99	18.67	13.29
		$2.2/W_1$	111606	55940	0.99	39.57	28.67
		$2.2/W_2$	95546	46288	0.99	48.02	30.25
		$2.2/R$	18655	9636	0.98	45.58	25.05

Table II: Summary of Results from Model 2

ε	α	Term. Rule	Run Length Mean	Run Length St.dev.	Joint Coverage	Estimated α (% of nominal α) Mean	Estimated α (% of nominal α) St.dev.
0.20	0.10	2.4	16749	9490	0.89	44.86	26.22
		2.3	26320	15330	0.92	21.34	12.61
		$2.2/W_1$	14762	8926	0.87	44.70	26.51
		$2.2/W_2$	3701	1817	0.92	47.79	23.97
		$2.2/R$	6823	5178	$0.83^{\dagger}$	42.55	27.07
0.20	0.05	2.4	24373	13790	0.93	47.51	26.38
		2.3	39094	21960	0.99	17.55	12.61
		$2.2/W_1$	20034	11541	$0.90^{\dagger}$	50.90	28.72
		$2.2/W_2$	5475	3089	0.96	50.64	28.87
		$2.2/R$	10037	6734	$0.89^{\dagger}$	39.79	29.50
0.10	0.10	2.4	72998	32485	0.95	45.81	27.56
		2.3	119717	63966	0.97	14.94	11.65
		$2.2/W_1$	66753	31892	0.94	43.78	25.12
		$2.2/W_2$	15816	8193	0.95	50.28	26.15
		$2.2/R$	23887	14201	0.90	54.52	27.76
0.10	0.05	2.4	107388	51784	0.97	39.17	27.86
		2.3	156054	81858	0.99	15.89	12.97
		$2.2/W_1$	91815	44413	0.96	47.06	30.29
		$2.2/W_2$	24008	11600	0.98	48.20	28.81
		$2.2/R$	42096	23485	0.97	47.39	30.32
0.05	0.10	2.4	293939	131917	0.96	49.68	27.25
		2.3	487951	228233	0.99	18.04	13.55
		$2.2/W_1$	243976	121691	0.93	52.39	27.03
		$2.2/W_2$	63914	29661	0.93	53.81	25.98
		$2.2/R$	108848	60014	0.93	49.54	28.30
0.05	0.05	2.4	424362	187325	0.99	46.09	31.39
		2.3	664444	288852	0.99	15.70	11.31
		$2.2/W_1$	386727	189112	0.97	44.42	29.29
		$2.2/W_2$	103819	46127	0.98	44.03	27.92
		$2.2/R$	165138	88687	0.99	43.59	29.64

†: Hypothesis of acceptable coverage rejected at significance level α

Table III: Summary of Results from Model 3

ε	α	Term. Rule	Run Length Mean	Run Length St.dev.	Joint Coverage	Estimated α (% of nominal α) Mean	Estimated α (% of nominal α) St.dev.
0.20	0.10	2.4	61481	38727	0.90	48.01	27.89
		2.3	101386	64790	0.96	20.51	13.06
		$2.2/W_1$	45482	31475	0.86	49.37	28.55
		$2.2/W_2$	3701	2204	0.92	51.39	26.66
		$2.2/R$	30233	27070	0.84†	51.19	28.99
0.20	0.05	2.4	100088	65329	0.96	43.32	27.26
		2.3	147294	92524	0.99	17.91	13.24
		$2.2/W_1$	78605	56415	0.93	41.54	26.96
		$2.2/W_2$	6185	4073	0.98	46.96	27.81
		$2.2/R$	45188	31813	0.92	44.23	29.21
0.10	0.10	2.4	308214	156925	0.98	45.56	27.08
		2.3	452912	220967	0.99	19.12	11.93
		$2.2/W_1$	218994	118992	0.93	54.62	26.16
		$2.2/W_2$	17560	9494	0.96	47.32	29.99
		$2.2/R$	146158	92816	0.92	50.90	26.06
0.10	0.05	2.4	447721	222305	0.99	38.90	24.94
		2.3	641085	334501	1.00	18.11	13.50
		$2.2/W_1$	341955	204268	0.96	48.72	29.45
		$2.2/W_2$	25063	12671	0.99	44.32	27.38
		$2.2/R$	221590	122706	0.94	48.81	27.61
0.05	0.10	2.4	1276979	549213	0.99	48.87	29.70
		2.3	1889514	953641	1.00	19.27	11.96
		$2.2/W_1$	905824	441934	0.96	57.15	24.18
		$2.2/W_2$	67807	35942	0.95	45.64	26.42
		$2.2/R$	630703	304032	0.95	46.14	26.39
0.05	0.05	2.4	1816840	868459	1.00	41.87	25.69
		2.3	2626630	1381230	1.00	18.47	12.86
		$2.2/W_1$	1482022	741855	0.98	44.69	28.27
		$2.2/W_2$	98304	50089	0.97	43.31	27.24
		$2.2/R$	926588	471779	0.97	50.88	25.97

†: Hypothesis of acceptable coverage rejected at significance level α

Table IV: Summary of Results from Model 4

ε	α	Term. Rule	Run Length Mean	Run Length St.dev.	Joint Cover- age	Estimated α (% of nominal α) Mean	Estimated α (% of nominal α) St.dev.
0.20	0.10	2.4	64484	53517	0.90	58.56	24.99
		2.3	77063	58980	0.92	37.79	20.09
		$2.2/R_A$	41341	31019	0.89	47.52	29.83
		$2.2/R_B$	30539	26037	$0.84^{\dagger}$	46.55	29.05
0.20	0.05	2.4	97498	74004	0.94	55.73	27.45
		2.3	113408	80630	0.94	34.89	21.34
		$2.2/R_A$	65496	46312	0.91	42.59	26.89
		$2.2/R_B$	47817	41937	$0.87^{\dagger}$	42.10	28.88
0.10	0.10	2.4	268726	131471	0.94	63.91	22.62
		2.3	344321	175516	0.97	38.51	21.40
		$2.2/R_A$	185672	108138	0.90	51.65	27.41
		$2.2/R_B$	135213	78891	0.87	54.79	26.89
0.10	0.05	2.4	412283	212838	0.98	55.63	25.77
		2.3	486150	235248	0.99	36.08	22.22
		$2.2/R_A$	312199	174075	0.96	41.59	28.37
		$2.2/R_B$	213529	124191	0.93	46.50	28.83
0.05	0.10	2.4	1217815	534716	0.97	59.32	23.08
		2.3	1393086	594533	0.98	41.80	20.23
		$2.2/R_A$	889174	486154	0.94	47.57	27.63
		$2.2/R_B$	660277	339676	0.96	50.65	26.94
0.05	0.05	2.4	1630909	751435	0.99	62.10	23.81
		2.3	2024348	992181	0.99	36.49	22.77
		$2.2/R_A$	1261165	632538	0.96	46.16	27.78
		$2.2/R_B$	992924	466138	0.99	45.81	28.12

†: Hypothesis of acceptable coverage rejected at significance level α

The advantage of simultaneous estimation can easily be seen when we compare the run length of the proposed method to the sum of run lengths when a single mean alone is estimated. The run length in simultaneous estimation is 62–88 % of the sum of run lengths for single means. In addition, the joint confidence level is $1-\alpha$, when the means are simultaneously estimated, and $1-p\alpha$, when the means are separately estimated.

When means of estimated α's are examined, we can see that all termination rules fall below the ideal value of 100%. This is not surprising since the sequence of checkpoints is geometrically increasing. Our primary reason for providing the estimated α's is to compare the termination rules based on the Bonferroni inequality to the termination rule for estimating a single mean. Since the means of estimated α's for the proposed rule are close to the means for the rule of estimating a single mean, we can conclude that the proposed termination rule chooses the individual α_j's in an efficient way. The means also indicate that the traditional way of applying the Bonferroni inequality is not usually efficient.

The standard deviation of estimated α's is smallest, when the Bonferroni inequality is applied in the traditional way. However, the price is the large departure from the ideal mean level. When the proposed method is used, the standard deviations are about the same as in estimating a single mean. This is another indication that the method chooses the individual α_j's in an efficient way.

5. ALTERNATIVE APPROACHES

As discussed in the Introduction the multivariate simulation output analysis has obtained increasing attention during the recent years. The research reported primarily concerns methods for estimating the confidence region. Because of the fundamental difference in the problem settings, the comparison between the method proposed in this paper and the methods based on the estimated covariance matrix is difficult.

In Section 2 we stated that the simulation should be continued until

$$\widehat{\Pr}\{|\hat{\theta}_1-\theta_1|\le\varepsilon_1|\hat{\theta}_1| \;\wedge\; \ldots \;\wedge\; |\hat{\theta}_p-\theta_p|\le\varepsilon_p|\hat{\theta}_p|\}\ge 1-\alpha\,. \tag{5.1}$$

In papers examining the estimation of the covariance matrix the statement has been that the analyst is interested in the mean vector and in the joint confidence region.

If the analyst is interested in the mean vector and in the joint confidence region, when the length of output sequence is predefined, the volume of the constructed confidence region characterizes the merits of the method. The comparison in [6] clearly indicates that the volume of the rectangular confidence region based on the Bonferroni inequality is large when compared to the volumes of the ellipsoidic confidence regions based on the T^2-approximation with the estimated covariance matrix. However, if we are interested in criterion (5.1), the volume of the confidence region is immaterial.

The importance of sequential procedures for a single mean is widely recognized. We do not see any reason why the same would not be true when several means are simultaneously estimated. Therefore we discuss some possible ways of developing a sequential procedure that is based on the estimated covariance matrix.

The methods based on the estimated covariance matrix assume that the mean vector $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)'$ of the p-dimensional multivariate output sequence has a multinormal limiting distribution: $\hat{\boldsymbol{\theta}} \rightarrow N_p(\boldsymbol{\theta}, \Sigma_{\hat{\boldsymbol{\theta}}})$. Therefore, an approximative $100(1-\alpha)$ % confidence region for the unknown true mean $\boldsymbol{\theta}$ is the ellipsoid

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' S_{\hat{\boldsymbol{\theta}}}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \frac{p\nu}{\nu - p + 1} F_{\nu, \nu-p+1}(1-\alpha) \,, \tag{5.2}$$

where $S_{\hat{\boldsymbol{\theta}}}$ is the estimated covariance matrix of $\hat{\boldsymbol{\theta}}$ with ν degrees of freedom and $F_{\nu_n, \nu_d}(x)$ is the $100x^{\text{th}}$ percentile point of the F distribution with ν_n numerator and ν_d denominator degrees of freedom.

Of the methods for estimating $S_{\hat{\boldsymbol{\theta}}}$ the multivariate batch-means method introduced in [7] has received most attention; see also [4, 5, 6]. The multivariate batch-means method increases the batch size until auto- and cross-correlations at lag 1 are small enough. Various statistical tests for detecting significant correlations at lag 1 are compared in [4]. Other methods proposed for estimating $S_{\hat{\boldsymbol{\theta}}}$ include the method based on spectral estimation introduced in [15] and the method based on regenerative processes examined in [29, 13, 30].

As discussed in Section 2 the first fundamental question in developing a sequential procedure for a mean vector is how we specify the desired precision. If we use the requirement in Eq. (5.1) we should integrate the joint probability density function. The joint distribution of the $(\hat{\theta}_j - \theta_j)/s_{\hat{\theta}_j}$'s can be assumed to be the general multidimensional t-distribution; see [14, Ch. 37]. However, the integration of the joint pdf of the multivariate t-distribution is not a simple task. It requires numerical integration that can be both time-consuming and inaccurate.

If we can assume that $\Sigma_{\hat{\boldsymbol{\theta}}} = \sigma^2 \Omega$, where σ^2 is the unknown common variance and Ω is a known $p \times p$ matrix, then we can use the results given in [27]:

$$\Pr\left\{ |\hat{\theta}_j - \theta_j| \leq t^*_{p,\nu}(1-\alpha) s \sqrt{\omega_{jj}} \,,\ \forall\, j \in \{1, \ldots, p\} \right\} \geq 1 - \alpha \,,$$

where $t^*_{p,\nu}(x)$ is the $100x^{\text{th}}$ percentile of the Studentized maximum modulos distribution, s^2 is an estimate of σ^2 with ν degrees of freedom, and ω_{jj} is the j^{th} diagonal element of Ω. When $t^*_{p,\nu}(1-\alpha) s \sqrt{\omega_{jj}} \leq \varepsilon_j |\hat{\theta}_j|$ for all $j \in \{1, \ldots, p\}$, we can conclude that the requirement in Eq. (5.1) is satisfied. The fundamental problem with this approach is that the matrix Ω must be known.

A third alternative to infer that the requirement (5.1) is satisfied is when the ellipsoid in Eq. (5.2) is inside the rectangular defined through the requirements that $|\hat{\theta}_j - \theta_j| \leq \varepsilon_j |\hat{\theta}_j|$, $j = 1, \ldots, p$.

The confidence region is inside the rectangular, when

$$\max_{j,k=1,\ldots,p} \{|q_{kj}|\sqrt{\lambda_k}\} \leq \sqrt{\frac{\nu - p + 1}{p\nu F_{\nu,\nu-p+1}(1-\alpha)}}\ , \tag{5.3}$$

where λ_k is the k^{th} eigenvalue of $\Theta^{-1} S_{\hat{\boldsymbol{\theta}}} \Theta^{-1}$, $\Theta = \text{diag}(\varepsilon_1|\hat{\theta}_1|, \ldots, \varepsilon_p|\hat{\theta}_p|)$, and q_{kj} is the j^{th} element in the k^{th} normalized eigenvector of $\Theta^{-1} S_{\hat{\boldsymbol{\theta}}} \Theta^{-1}$. However, our experience is that the run length is usually longer when (5.3) is used as the termination rule than when the proposed termination rule is used.

If we do not want to solve the eigenvectors, we can use the fact that $|q_{kj}| \leq 1$. Therefore, we can conclude that the ellipsoidic confidence region is inside the rectangular defined through $|\hat{\theta}_j - \theta_j| \leq \varepsilon|\hat{\theta}_j|$, $j = 1, \ldots, p$, when

$$\lambda_{max} \leq \frac{\nu - p + 1}{p\nu F_{\nu,\nu-p+1}(1-\alpha)}\ , \tag{5.4}$$

where λ_{max} is the largest eigenvalue of $\Theta^{-1} S_{\hat{\boldsymbol{\theta}}} \Theta^{-1}$. If we use (5.4) as the termination rule, then we know that

$$\widehat{\Pr}\left\{ \sum_{j=1}^{p} \left(\frac{\hat{\theta}_j - \theta_j}{\varepsilon_j \theta_j} \right)^2 \leq 1 \right\} = 1 - \alpha\ .$$

This is an essentially stronger requirement than (5.1). In [24] we used a criterion similar to (5.4) to control the simulation run length, when several percentiles were simultaneously estimated. However, we found that the simulation runs were quite expensive. Therefore, we replaced the criterion (5.4) by criterion (5.1) and used the Bonferroni inequality; see [25].

An alternative way of generalizing the half-width criterion is to specify an upper bound for the volume of the $100(1-\alpha)$ % confidence region. This generalization is more suitable for the methods based on the estimated covariance matrix than the generalization in Eq. (5.1). For example, we can require that the volume of the confidence region must be less than $(2\varepsilon)^p \prod_{j=1}^{p} |\hat{\theta}_j|$. The volume of the ellipsoidic confidence region in Eq. (5.2) is proportional to $|S_{\hat{\boldsymbol{\theta}}}|^{1/2}$; see e.g. [1, p. 170, 176].

If the analyst specifies the required precision as an upper bound for the $100(1-\alpha)$ % confidence region, the inference about the individual means must be done very carefully. The reason is that when the volume is controlled, the precisions of individual means can not be controlled. We believe that in many practical applications the analyst is primarily interested in the individual means. Of course, there are also situations in which the volume is an appropriate requirement for the precision. In these situations the analyst should estimate the covariance matrix instead of using the Bonferroni inequality.

6. SUMMARY

We have proposed a sequential procedure for controlling the length of a simulation run when several means are simultaneously estimated. The precision requirement is specified as a rectangular confidence region having desired widths of each side. The Bonferroni inequality is used to estimate a lower bound for the confidence level of the region. The procedure is a straightforward generalization of the procedures developed for estimating a single mean.

The empirical results reported indicate that the method controls the run length so that the estimates have the precision requested. In addition, the results show that the simultaneous estimation of several means does not necessarily require long simulation runs when compared to the estimation of a single mean. If we compare the computing times needed in estimating several means and a single mean, multiple estimation is automatically more expensive. This is due to the variance estimation. For each mean we need an estimated variance. However, the spectral method can be implemented so that multiple variance estimates can be calculated efficiently.

APPENDIX: SOLUTIONS OF THE INTEGER OPTIMIZATION PROBLEMS FOR THE BEST CHOICE OF INDIVIDUAL CONFIDENCE LEVELS IN BONFERRONI INEQUALITY

The best choice of the α_j's in the Bonferroni inequality is implied through the solution of the following integer optimization problem

$$\begin{aligned} &\underset{N_0,\ldots,N_p}{\text{minimize}} && N_0 \\ &\text{subject to} && \alpha - 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j|\hat{\theta}_j(N_j)|/s_{\hat{\theta}_j(N_j)}\big) \geq 0 \\ &\text{and} && N_0 - N_j \geq 0\,,\ j = 1,\ldots,p\,, \end{aligned} \tag{A.1}$$

when the output sequence is a multivariate one.

In a general case the analytic solution of (A.1) is unknown. However, when the (marginal) output sequences are covariance stationary, the $\hat{\theta}_j(N_j)$'s converge (almost surely) to θ_j and the $\mathrm{Var}(\hat{\theta}_j(N_j))$'s converge to $h_j(0)/N_j$, where the $h_j(0)$'s are the spectral densities at zero frequency. We have also assumed that each $s^2_{\hat{\theta}_j(N_j)}$ converges (almost surely) to $\mathrm{Var}(\hat{\theta}_j(N_j))$. Under these assumptions there exists N^* such that $\alpha - 2\sum_{j=1}^{p} F_{\nu_j}(-\varepsilon_j|\hat{\theta}_j(N_j)|/s_{\hat{\theta}_j(N_j)})$ is (almost surely) monotonically increasing function of each N_j, when $N_j > N^*$, $j = 1,\ldots,p$. Now the problem is so well-behaving that we can analytically solve it through the *branch and bound method*.

The solution of (A.1) is obtained through the solution of the following optimization problem

$$\begin{aligned} &\underset{x_0,\dots,x_p}{\text{minimize}} \quad x_0 \\ &\text{subject to} \quad \alpha - 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j g_j(x_j)\big) \geq 0 \\ &\text{and} \quad x_0 - x_j \geq 0\,,\ j = 1,\dots,p\,, \end{aligned} \tag{A.2}$$

where $g_j(x_j) = (N_j+1-x_j)|\hat{\theta}_j(N_j)|/s_{\hat{\theta}_j(N_j)} + (x_j-N_j)|\hat{\theta}_j(N_j+1)|/s_{\hat{\theta}_j(N_j+1)}$, when $N_j \leq x_j \leq N_j+1$. If $x_0,\dots,x_p$ is the solution of (A.2), then the solutions of (A.1) are $N_0 = \lceil x_0 \rceil$ and N_j is either $\lceil x_j \rceil$ or $\lfloor x_j \rfloor$ but at least one N_j is $\lceil x_j \rceil$. For details, see e.g. [11, Ch. 13].

The Lagrangian function of problem (A.2) is

$$\mathcal{L}(x_0,\dots,x_p,\lambda_0,\dots,\lambda_p) = x_0 - \lambda_0\left(\alpha - 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j g_j(x_j)\big)\right) - \sum_{j=1}^{p} \lambda_j (x_0 - x_j)\,.$$

The Kuhn-Tucker conditions (see e.g. [11, Ch. 9.1]) for a local minimum are:

$$\begin{aligned} &\text{KT-1:} \quad 1 - \sum_{j=1}^{p} \lambda_j = 0\,, \\ &\text{KT-2:} \quad \lambda_j - 2\lambda_0 \varepsilon_j g_j'(x_j) f_{\nu_j}(-\varepsilon_j g_j(x_j)) = 0\,,\ j = 1,\dots,p\,, \\ &\text{KT-3:} \quad \text{either } \lambda_j > 0 \text{ and } x_j = x_0 \text{ or } \lambda_j = 0 \text{ and } x_j < x_0\,,\ j = 1,\dots,p\,, \\ &\text{KT-4:} \quad \text{either } \lambda_0 > 0 \text{ and } 2\sum_{j=1}^{p} F_{\nu_j}(-\varepsilon_j g_j(x_j)) = \alpha \\ &\qquad\qquad \text{or } \lambda_0 = 0 \text{ and } 2\sum_{j=1}^{p} F_{\nu_j}(-\varepsilon_j g_j(x_j)) < \alpha\,. \end{aligned}$$

The assumptions that we have made imply that there exists x^* such that each $g_j(x_j)$ is a monotonically increasing function of x_j, when $x_j > x^*$, $j = 1,\dots,p$. When $g_j(x_j)$ is a monotonically increasing function of x_j, then its derivative $g_j'(x_j)$ is positive. Since $F_\nu(t)$ is the cumulative distribution function of Student t distribution, the corresponding probability density function $f_\nu(t)$ is positive. Therefore, the Kuhn-Tucker conditions are (almost surely) satisfied only if $\lambda_j > 0$, $j = 0,\dots,p$. If $\lambda_0 = 0$, then KT-2 implies that $\lambda_j = 0$, $j = 1,\dots,p$, but this violates KT-1. On the other hand, if $\lambda_j = 0$ for some $j \in \{1,\dots,p\}$, then KT-2 implies that $\lambda_0 = 0$, but then $\lambda_j = 0$ for all $j \in \{1,\dots,p\}$ and KT-1 is not satisfied. Therefore, $x_j = x_0$, $j = 1,\dots,p$, and $2\sum_{j=1}^{p} F_{\nu_j}(-\varepsilon_j g_j(x(j))) = \alpha$ is the only local minimizer of (A.2), when the ε_j's are so small that the non-linear constraint is not satisfied for any $(x_1,\dots,x_p)$ such that $x_j < x^*$ for some $j \in \{1,\dots,p\}$.

When the output sequences are asynchronous and generated by a single simulation run, the run length is $N = N_1+\dots+N_p$. For any given N the N_j 's can be written as $N_j = N\pi_{j|N}$, $j = 1,\dots,p$.

We assume that the $\pi_{j|N}$'s are random variables such that $\pi_{1|N} + \ldots + \pi_{p|N} = 1$ for all N and each sequence $\{\pi_{j|N}\}_{N=1}^{\infty}$ converges (almost surely) to a positive constant π_j as $N \to \infty$. This implies that further observations from each sequence can be obtained.

The best choice of α_j's is implied through the solution of the following integer optimization problem

$$\begin{aligned} &\underset{N_0,\ldots,N_p}{\text{minimize}} && N_0 \\ &\text{subject to} && \alpha - 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j|\hat{\theta}_j(N_j)|/s_{\hat{\theta}_j(N_j)}\big) \geq 0 \\ &\text{and} && N_0\pi_{j|N_0} - N_j \geq 0\,,\ j = 1,\ldots,p\,. \end{aligned} \tag{A.3}$$

Under the assumptions we have made there exists N^* such that $\alpha - 2\sum_{j=1}^{p} F_{\nu_j}(-\varepsilon|\hat{\theta}_j(N_j)|/s_{\hat{\theta}_j(N_j)})$ is (almost surely) monotonically increasing function of each N_j when $N_j > N^*\pi_{j|N^*}$, $j = 1,\ldots,p$. Therefore, we can solve the problem (A.3) in a similar fashion as the problem (A.1). The solution of (A.3) is obtained through solving the following optimization problem

$$\begin{aligned} &\underset{x}{\text{minimize}} && x \\ &\text{subject to} && \alpha - 2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j g_j(h_j(x))\big) \geq 0\,, \end{aligned} \tag{A.4}$$

where $g_j(x_j)$ is as in (A.2) and the $h_j(x)$'s are continuous and monotonically increasing functions of x, $\sum_{j=1}^{p} h_j(x) = x$, and for integer values of x, $h_j(N) = N_j = N\pi_{j|N}$, $j = 1,\ldots,p$. The exact form of the $h_j(x)$'s depends on the sequence $\{\pi_{j|N}\}_{N=1}^{\infty}$. The sequences satisfy that $(N+1)\pi_{j|N+1} = N\pi_{j|N} + 1$ for exactly one $j \in \{1,\ldots,p\}$ and $(N+1)\pi_{j|N+1} = N\pi_{j|N}$ for the rest of j's. Therefore the $h_j(x)$'s can be uniquely determined for any possible set of the $\{\pi_{j|N}\}$ in the following manner. Suppose that $N\pi_{j|N} = 1+(N-1)\pi_{j|N-1}$, $N\pi_{j|N} = \ldots = (N+k-1)\pi_{j|N+k-1}$, and $(N+k)\pi_{j|N+k} = 1+(N+k-1)\pi_{j|N+k-1}$, then $h_j(x) = \big((N+k-x)N\pi_{j|N} + (x-N)(N+k)\pi_{j|N+k}\big)/k$, when $N \leq x \leq N+k$.

If x is the solution of (A.4), then the solution of (A.3) is $N_0 = \lceil x \rceil$ and $N_j = h_j(N_0) = N_0\pi_{j|N_0}$, $j = 1,\ldots,p$. Our assumptions imply that there exists x^* such that each $g_j(h_j(x))$ is (almost surely) a monotonically increasing function of x, when $x > x^*$. Hence, the implicit solution of (A.4) is $2\sum_{j=1}^{p} F_{\nu_j}\big(-\varepsilon_j g_j(h_j(x))\big) = \alpha$, if the ε_j's are so small that the constraint in (A.4) is not satisfied for any $x < x^*$.

References

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Wiley, New York, 1958.

[2] O. Balci and R. G. Sargent. Validation of simulation models via simultaneous confidence intervals. *Am. J. Math. Manag. Sci.*, 4(3&4):375–406, 1984.

[3] K. W. Bauer, S. Venkatraman, and J. R. Wilson. Estimation procedures based on control variates with known covariance matrix. In *Proceedings of the 1987 Winter Simulation Conference*, pages 334–341. IEEE, 1987.

[4] J. M. Charnes. Power comparisons for the multivariate batch-means method. In *Proceedings of the 1990 Winter Simulation Conference*, pages 281–287. IEEE, 1990.

[5] J. M. Charnes. Multivariate simulation output analysis. In *Proceedings of the 1991 Winter Simulation Conference*, pages 187–193. IEEE, 1991.

[6] J. M. Charnes and W. D. Kelton. A comparison of confidence region estimators for multivariate simulation output. In *Proceedings of the 1988 Winter Simulation Conference*, pages 458–465. IEEE, 1988.

[7] D. R. Chen and A. F. Seila. Multivariate inference in stationary simulation using batch means. In *Proceedings of the 1987 Winter Simulation Conference*, pages 302–304. IEEE, 1987.

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

[9] H. Damerdji. On strong consistency of the variance estimator. In *Proceedings of the 1987 Winter Simulation Conference*, pages 305–308. IEEE, 1987.

[10] O. J. Dunn. Multiple comparisons among means. *J. Amer. Statist. Assoc.*, 56(293):52–64, March 1961.

[11] R. Fletcher. *Practical Methods of Optimization, Volume 2: Constrained Optimization.* Wiley, Chichester, 1981.

[12] P. Heidelberger and P. D. Welch. A spectral method for confidence interval generation and run length control in simulation. *Commun. ACM*, 24(4):233–245, April 1981.

[13] IEEE. *Multivariate Estimation in Simulation*, 1983.

[14] N.L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York, 1972.

[15] P. Kabaila and G. Nelson. On confidence regions for the mean of a multivariate time series. *Commun. Statist. -Simula.*, B14(3):735–753, 1985.

[16] L. Kleinrock. *Queueing Systems, Volume I: Theory.* Wiley, New York, 1975.

[17] S. S. Lavenberg and C. H. Sauer. Sequential stopping rules for the regenerative method of simulation. *IBM J. Res. Develop.*, 21(6):545–558., November 1977.

[18] A. M. Law and J. S. Carson. A sequential procedure for determining the length of a steady-state simulation. *Operations Res.*, 27(6):1011–1025, November 1979.

[19] A. M. Law and W. D. Kelton. Confidence intervals for steady-state simulations, ii: A survey of sequential procedures. *Management Sci.*, 28(5):550–562, May 1982.

[20] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis.* McGraw-Hill, New York, 2^{nd} ed. edition, 1991.

[21] M. Law, A., W. D. Kelton, and L. W. Koenig. Relative width sequential confidence intervals for the mean. *Commun. Statist. -Simula.*, B10(1):29–39, 1981.

[22] A. Nadas. An extension of a theorem of chow and robbins on sequential confidence intervals for the mean. *Ann. Math. Statist.*, 40(2):667–671, September 1969.

[23] K. Pawlikowski. Steady-state simulation of queueing processes: A survey of problems and solutions. *ACM Comput. Surv.*, 22(2):123–170, June 1990.

[24] K. E. E. Raatikainen. Run length control for simultaneous estimation of several percentiles in dependent sequences. In O. Balci, editor, *Methodology and Validation, 1987*, volume 19, 1 of *Simulation Series*, pages 54–59, San Diego, 1987. Society for Computer Simulation.

[25] K. E. E. Raatikainen. Sequential procedure for simultaneous estimation of several percentiles. *Trans. Soc. Comput. Simul.*, 7(1):21–44, March 1990.

[26] K. E. E. Raatikainen. Simultaneous sequential confidence intervals of fixed widths for several means using bonferroni inequality. Technical report, University of Helsinki, Department of Computer Science, February 1992.

[27] J. Richmond. A general method for constructing simultaneous confidence interval. *J. Amer. Statist. Assoc.*, 77(378):455–460, June 1982.

[28] L. W. Schruben. Control of initialization bias in multivariate simulation response. *Commun. ACM*, 24(4):246–252, April 1981.

[29] A. F. Seila. Multivariate estimation in regenerative simulation. *Oper. Res. Lett.*, 1(4):153–156, September 1982.

[30] A. F. Seila. Multivariate simulation output analysis. *Am. J. Math. Manag. Sci.*, 4(3&4):313–334, 1984.

[31] S. Venkatraman and J. R. Wilson. Estimation procedures based on control variates with known covariance matrix. *Oper. Res. Lett.*, 5(1):37–42, June 1986.

[32] P. D. Welch. The statistical analysis of simulation results. In S. S. Lavenberg, editor, *Computer Performance Modeling Handbook*, chapter 6. Academic Press, New York, 1983.

[33] W.-N. Yang and B. L. Nelson. Multivariate estimation and variance reduction in terminating and steady-state simulation. In *Proceedings of the 1988 Winter Simulation Conference*, pages 458–465, December 1988.

[34] W.-N. Yang and B. L. Nelson. Multivariate batch means and control variates. *Management Sci.*, 38(10):1415–1431, October 1992.