# SIMULTANEOUS SEQUENTIAL CONFIDENCE INTERVALS OF FIXED WIDTHS FOR SEVERAL MEANS USING BONFERRONI INEQUALITY

KIMMO E. E. RAATIKAINEN
University of Helsinki, Department of Computer Science
P. O. Box 26 (Teollisuuskatu 23), FIN-00014 University of Helsinki, Finland
*e-mail:* Kimmo.Raatikainen@cs.Helsinki.FI

*Sequentially constructed confidence intervals for a single mean are widely used to control the length of a simulation run. When the analyst is simultaneously interested in several response variables, the control of a simulation run needs sequential confidence regions. In this paper we examine asymptotic properties of sequentially constructed confidence regions based on the Bonferroni inequality. We also report empirical results when simultaneous confidence intervals for normal and transformed normal means are constructed.*
(Simultaneous Estimation, Sequential Confidence Intervals)

## 1. Introduction

Sequentially constructed confidence intervals are widely used to control the precision of a single estimated mean in simulation studies. The methods proposed (see e.g. Pawlikowski, 1990) to control the precision assume that the output sequence $\{X_i\}_{i=1}^n$ is covariance stationary. In addition, they assume that the output sequence satisfies certain regularity conditions which are fairly general (see e.g. Anderson, 1971, p. 478) so that the estimator $\hat{\theta}(n) = (X_1 + \ldots + X_n)/n$ has a normal limiting distribution:

$$\sqrt{n}\left(\hat{\theta}(n) - \theta\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathrm{N}(0, \gamma)\,,$$

where $\theta = \mathrm{E}(X_i)$ and $\gamma = \lim_{n\to\infty} n\,\mathrm{Var}[\hat{\theta}(n)]\,,\ 0 < \gamma < \infty$. When $s^2_{\hat{\theta}(n)}$ denotes the estimated variance of $\hat{\theta}(n)$ with $\nu(n)$ degrees of freedom, the methods approximate the distribution of $(\hat{\theta}(n) - \theta)/s_{\hat{\theta}(n)}$ by the Student $t$-distribution with $\nu(n)$ degrees of freedom. Therefore, the simulation is terminated as soon as

$$s_{\hat{\theta}(n)} \le \delta/t_{\nu(n)}(1 - \alpha/2) \tag{1.1}$$

where $\delta$ is the prespecified upper bound for the half-width of the confidence interval, $I(n) = [\hat{\theta}(n) - \delta, \hat{\theta}(n) + \delta]$, and $1 - \alpha$ is the prespecified confidence level. The termination rule in (1.1) implies that $\widehat{\Pr}\{|\hat{\theta}(n) - \theta| \leq \delta\} \geq 1 - \alpha$, i.e. the error in the estimator $\hat{\theta}(n)$ is at most $\delta$ with an estimated probability of at least $1 - \alpha$.

In Raatikainen (1994) we proposed a sequential procedure for controlling the simultaneous precision of several estimated proportions. The procedure is based on the Bonferroni inequality (see e.g. Kleijnen, 1987, p. 42, or Law and Kelton, 1991, Ch. 9.7) and on the following precision requirement. The simulation is continued until

$$\widehat{\Pr}\{|\hat{\theta}_1(n) - \theta_1| \leq \delta_1 \ \wedge \ \ldots \ \wedge \ |\hat{\theta}_p(n) - \theta_p| \leq \delta_p\} \geq 1 - \alpha \tag{1.2}$$

implying that the error in each estimate $\hat{\theta}_j(n)$ is less than $\delta_j$ with an estimated probability of at least $1 - \alpha$. In (1.2) the $\hat{\theta}_j(n)$'s are the estimates of the unknown $\theta_j$'s based on a $p$-dimensional sequence $\{\boldsymbol{X}_i\}_{i=1}^n$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})'$, the $\delta_j$'s are the prespecified upper bounds for the half-widths of the (marginal) confidence intervals, $I_j(n) = [\hat{\theta}_j(n) - \delta_j, \hat{\theta}_j(n) + \delta_j]$, and $1 - \alpha$ is the prespecified lower bound for the joint coverage of the confidence intervals.

The objective of this paper is to provide the theoretical foundation for the termination rule

$$N = \text{smallest integer } n \text{ such that } 2\sum_{j=1}^{p} F_{\nu_j(n)}\left(-\delta_j / s_{\hat{\theta}_j(n)}\right) \leq \alpha\,. \tag{1.3}$$

In (1.3) $F_{\nu_j(n)}(\cdot)$ is the cumulative distribution function of the Student $t$-distribution with $\nu_j(n)$ degrees of freedom and the $s^2_{\hat{\theta}_j(n)}$'s are the estimated variances of the $\hat{\theta}_j(n)$'s, each with $\nu_j(n)$ degrees of freedom. The primary advantage of using the Bonferroni inequality is that the covariances between the $\hat{\theta}_j(n)$'s are not needed. Therefore, our method only needs the same assumptions as the methods developed for estimating a single mean.

In Section 2 we examine the properties of the proposed termination rule. We examine the asymptotic simultaneous coverages of the (marginal) confidence intervals. We also show that the termination rule (1.3) is an efficient way of applying the Bonferroni inequality. In Section 3 we report our experiments of constructing the intervals when independent observations are generated from two- and ten-dimensional normal distributions and from transformed five-dimensional normal distributions. The reason for using the normal and transformed normal distributions (and not stochastic processes typical to simulation studies) is that we want to examine the effects that the proposed way of applying the Bonferroni inequality has. Therefore, we selected output sequences in which the autocorrelations do not affect the estimation of variances.

## 2. Properties of the Proposed Termination Rule

In this section we examine the properties of the termination rule proposed in (1.3). We examine the asymptotic joint coverage of the confidence intervals constructed. We also show that under our assumptions no other way of applying the Bonferroni inequality is more efficient than the proposed one.

### *2.1. Asymptotic Coverage of Constructed Confidence Intervals*

Our objective is to show that for a given $\alpha$, $0 < \alpha < 1$, the termination rule provides confidence intervals that satisfy

$$\lim_{\delta_1,\ldots,\delta_p \to 0} \Pr\{|\hat{\theta}_1 - \theta_1| \leq \delta_1 \wedge \ldots \wedge |\hat{\theta}_p - \theta_p| \leq \delta_p\} \geq 1 - \alpha\,. \tag{2.1}$$

When we can show that (2.1) holds, we can trust that for small $\delta_j$'s the lower bound for the true probability that errors in the estimates are at most $\delta_j$ is close to $1 - \alpha$. Unfortunately, the asymptotic coverage does not tell us how small the $\delta_j$'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 sequence length determined through the termination rule proposed in (1.3) is a random variable that depends on $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$:

$$N(\boldsymbol{\delta}) \equiv \text{smallest integer } n \text{ such that } 2\sum_{j=1}^{p} F_{\nu_j(n)}\left(-\delta_j / s_{\hat{\theta}_j(n)}\right) \leq \alpha\,. \tag{2.2}$$

Now the estimates $\hat{\theta}_j(n)$, their estimated variances $s^2_{\hat{\theta}_j(n)}$, and the degrees of freedom $\nu_j(n)$ depend on $\boldsymbol{\delta}$ through $N(\boldsymbol{\delta})$. Let the notations $\hat{\theta}_j[N(\boldsymbol{\delta})]$, $s^2_{\hat{\theta}_j}[N(\boldsymbol{\delta})]$, and $\nu_j[N(\boldsymbol{\delta})]$ indicate this dependency.

Suppose that the marginal output sequences are covariance stationary and satisfy the regularity conditions so that each $\sqrt{n}(\hat{\theta}_j(n) - \theta_j)$ converges in distribution to $\mathrm{N}(0, \gamma_j)$, where $\gamma_j \equiv \lim_{n\to\infty} n \operatorname{Var}[\hat{\theta}_j(n)]$, $0 < \gamma_j < \infty$. These convergences are for a nonrandom $n$. It is shown in Billingsley (1968, Ch. 17) that under certain conditions the convergences for a nonrandom $n$ imply that each $\sqrt{N(\boldsymbol{\delta})}\left(\hat{\theta}_j[N(\boldsymbol{\delta})] - \theta_j\right)$, $j = 1, \ldots, p$, converges in distribution to $\mathrm{N}(0, \gamma_j)$ as $\delta_1, \ldots, \delta_p \to 0$. The conditions are satisfied if we can find for any sequence of $p$-dimensional positive constants $\{\boldsymbol{\delta}(m) = (\delta_1(m), \ldots, \delta_p(m))' :\ m = 1, 2, \ldots\}$, each $\delta_j(m)$ going to zero, a sequence of $p$-dimensional constants $\{\boldsymbol{a}(m) = (a_1(m), \ldots, a_p(m))' :\ m = 1, 2, \ldots\}$, each $a_j(m)$ going to infinity, such that for each $j = 1, \ldots, p$ the ratio $N[\boldsymbol{\delta}(m)]/a_j(m)$ converges in probability to a positive limit $\psi_j^2$ as $m$ goes to infinity:

$$\frac{N[\boldsymbol{\delta}(m)]}{a_j(m)} \xrightarrow[m\to\infty]{\mathcal{P}} \psi_j^2 \quad \text{for some } 0 < \psi_j^2 < \infty\,,\ j = 1, \ldots, p\,. \tag{2.3}$$

In order to establish the asymptotic coverages of the confidence intervals $[\hat{\theta}_j[N(\boldsymbol{\delta})]-\delta_j, \hat{\theta}_j[N(\boldsymbol{\delta})]+\delta_j]$, we need to assume that the method of estimating $\mathrm{Var}[\hat{\theta}_j(n)]$'s fulfills the following three assumptions. Firstly, we must assume that each $ns^2_{\hat{\theta}_j(n)} > 0$ (almost surely). Secondly, we must assume that each $ns^2_{\hat{\theta}_j(n)}$ is a consistent estimator of the corresponding variance parameter $\gamma_j$:

$$\lim_{n\to\infty} \Pr\{|ns^2_{\hat{\theta}_j(n)} - \gamma_j| > \epsilon\} = 0 \text{ , for every } \epsilon > 0 \text{ .} \tag{2.4}$$

Thirdly, we must assume that there are (possibly infinite) constants $\nu_j(\infty)$ , $j = 1, \ldots, p$, to which the $\nu_j(n)$'s, that is the degrees of freedoms for the estimators $ns^2_{\hat{\theta}_j(n)}$, converge in probability as $n \to \infty$:

$$\nu_j(n) \underset{n\to\infty}{\xrightarrow{\mathcal{P}}} \nu_j(\infty) \le \infty \text{ , } j = 1, \ldots, p \text{ .} \tag{2.5}$$

The following lemma is a generalization of Lemma 1 given in Chow and Robbins (1965).

**Lemma 1.** Let $\{z_j(m): \ m = 1, 2, \ldots\}$ , $j = 1, \ldots, p$, be sequences of positive random variables that converge in probability to positive limits $\psi_j^2$,

$$z_j(m) \underset{m\to\infty}{\xrightarrow{\mathcal{P}}} \psi_j^2 \text{ , } j = 1, \ldots, p, \text{ where } \ 0 < \psi_j < \infty \text{ .} \tag{2.6}$$

Let $\{f_j(m): \ m = 1, 2, \ldots\}$ , $j = 1, \ldots, p$, be sequences of constants such that for each $j = 1, \ldots, p$

$$f_j(m) > 0 \text{ , } \quad \lim_{m\to\infty} f_j(m) = \infty \text{ , } \quad \lim_{m\to\infty} f_j(m)/f_j(m-1) = 1 \text{ .} \tag{2.7}$$

For each $\tau_j > 0$ , $j = 1, \ldots, p$, define

$$M_j(\tau_j) \equiv \text{smallest } m \ge 1 \text{ such that } z_j(m) \le f_j(m)/\tau_j \text{ .} \tag{2.8}$$

Each $M_j(\tau_j)$ is well defined and nondecreasing as a function of $\tau_j$ such that

$$M_j(\tau_j) \underset{\tau_j\to\infty}{\xrightarrow{\mathcal{P}}} \infty \quad \text{and} \quad f_j[M_j(\tau_j)]/\tau_j \underset{\tau_j\to\infty}{\xrightarrow{\mathcal{P}}} \psi_j^2 \text{ .} \tag{2.9}$$

**Proof of Lemma 1.** Proving that the $M_j(\tau_j)$'s converge in probability to infinity is a straightforward application of the definition of $M_j(\tau_j)$'s and the notion of convergence in probability. To prove the second part of (2.9), we observe that for $M_j(\tau_j) > 1$,

$$z_j[M_j(\tau_j)] \le f_j[M_j(\tau_j)]/\tau_j < \{f_j[M_j(\tau_j)]/f_j[M_j(\tau_j)-1]\}z_j[M_j(\tau_j)-1] \text{ .} \tag{2.10}$$

Now each ratio $f_j[M_j(\tau_j)]/\tau_j$ has a stochastic lower and upper bound. The first part of (2.9), (2.6), and (2.7) together with Application D in Serfling (1980, p. 26) imply that the bounds converge in probability to $\psi_j^2$. Hence, $f_j[M_j(\tau_j)]/\tau_j$ converges in probability to $\psi_j^2$. $\square$

Let $\{\alpha_j(m) : \ m = 1,2,\ldots\}$, $j = 1,\ldots,p$, be sequences of positive constants such that $\alpha_1(m) + \ldots + \alpha_p(m) = \alpha$ for every $m = 1,2,\ldots$ and $\lim_{m\to\infty} \alpha_j(m) = \alpha_j$, $j = 1,\ldots,p$. Set

$$\left.\begin{array}{rcl} z_j(m) & = & m\dfrac{s^2_{\hat{\theta}_j(m)}}{\gamma_j}\left(\dfrac{t_{\nu_j(m)}(1-\alpha_j(m)/2)}{t_m(1-\alpha_j(m)/2)}\right)^2 \\ f_j(m) & = & m\left(\dfrac{z(1-\alpha_j(m)/2)}{t_m(1-\alpha_j(m)/2)}\right)^2 \\ \tau_j(m) & = & \dfrac{[z(1-\alpha_j(m)/2)]^2\gamma_j}{[\delta_j(m)]^2} \end{array}\right\} . \tag{2.11}$$

Now (2.8) implies that

$$M_j[\tau_j(m)] \equiv \text{smallest } m \geq 1 \text{ such that } \ 2F_{\nu_j(m)}\left(-\delta_j(m)/s_{\hat{\theta}_j(m)}\right) \leq \alpha_j(m) . \tag{2.12}$$

When we choose the $\alpha_j(m)$'s so that each $\alpha_j(m)$ is proportional to $F_{\nu_j(m)}\left(-\delta_j(m)\Big/s_{\hat{\theta}_j(m)}\right)$, that is

$$\alpha_j(m) = \alpha\frac{F_{\nu_j(m)}\left(-\delta_j(m)\Big/s_{\hat{\theta}_j(m)}\right)}{\sum_{k=1}^p F_{\nu_k(m)}\left(-\delta_k(m)\Big/s_{\hat{\theta}_k(m)}\right)}$$

for $j = 1,\ldots,p$ and $m = 1,2,\ldots$ , then it is easy to verify that

$$M_1[\tau_1(m)] = \ldots = M_p[\tau_p(m)] = N[\boldsymbol{\delta}(m)] \quad \text{for} \quad m = 1,2,\ldots \tag{2.13}$$

and that the assumptions of Lemma 1 are satisfied with

$$\psi_j = t_{\nu_j(\infty)}(1-\alpha_j/2)/z(1-\alpha_j/2) \geq 1 . \tag{2.14}$$

Now we observe that $\psi_j = 1$ only if $\nu_j(\infty) = \infty$. However, most methods that are currently used to estimate the variance, such as the method of batch means or the spectral method, provide a finite $\nu_j(\infty)$. In such a case we have $\psi_j > 1$.

For a given $\{\boldsymbol{\delta}(m) = (\delta_1(m),\ldots,\delta_p(m))' \ : \ m = 1,2,\ldots\}$ such that $\lim_{m\to\infty}\delta_j(m) = 0$, $j = 1,\ldots,p$, we define

$$a_j(m) \equiv \frac{[t_m(1-\alpha_j(m)/2)]^2\gamma_j}{[\delta_j(m)]^2} \quad \text{for} \quad j = 1,\ldots,p \text{ and } m = 1,2,\ldots \tag{2.15}$$

so that $\lim_{m\to\infty} a_j(m) = \infty$ , $j = 1,\ldots,p$.

Now we can show that each ratio in (2.3) converges in probability to a positive limit for each $j = 1,\ldots,p$:

$$\begin{array}{rcll} \dfrac{N[\boldsymbol{\delta}(m)]}{a_j(m)} & = & \dfrac{N[\boldsymbol{\delta}(m)]}{[t_m(1-\alpha_j(m)/2)]^2\gamma_j/[\delta_j(m)]^2} & \text{by (2.15)} \quad (2.16) \\ & = & \dfrac{f_j\{N[\boldsymbol{\delta}(m)]\}}{\tau_j(m)} & \text{by (2.11)} \quad (2.17) \\ & = & \dfrac{f_j\{M_j[\tau_j(m)]\}}{\tau_j(m)} & \text{by (2.13)} \quad (2.18) \\ & \underset{m\to\infty}{\xrightarrow{\mathcal{P}}} & \psi_j^2 & \text{by (2.9)} . \quad (2.19) \end{array}$$

Thus the conditions of Theorem 17.1 in Billingsley (1968) are satisfied so that we have

$$\sqrt{N(\boldsymbol{\delta})}\left(\hat{\theta}_j[N(\boldsymbol{\delta})] - \theta_j\right) \underset{\delta_1,\ldots,\delta_p\to 0}{\xrightarrow{\mathcal{D}}} \mathrm{N}(0,\gamma_j)\,,\; j = 1,\ldots,p\,. \tag{2.20}$$

Now we can establish the asymptotic coverages of the confidence intervals $[\hat{\theta}_j[N(\boldsymbol{\delta})]-\delta_j, \hat{\theta}_j[N(\boldsymbol{\delta})]+\delta_j]$. We observe that (2.19) implies that

$$\delta_j\sqrt{N(\boldsymbol{\delta})} \underset{\delta_1,\ldots,\delta_p\to 0}{\xrightarrow{\mathcal{P}}} t_{\nu_j(\infty)}(1-\alpha_j/2)\sqrt{\gamma_j}\,,\; j = 1,\ldots,p\,, \tag{2.21}$$

where the $\nu_j(\infty)$'s are defined in (2.5). Inserting (2.21) into

$$\begin{aligned} &\lim_{\delta_1,\ldots,\delta_p\to 0} \Pr\left\{\theta_j < \hat{\theta}_j[N(\boldsymbol{\delta})] - \delta_j\right\} \\ =\; &\lim_{\delta_1,\ldots,\delta_p\to 0} \Pr\left\{\delta_j\sqrt{N(\boldsymbol{\delta})} < \sqrt{N(\boldsymbol{\delta})}\left(\hat{\theta}_j[N(\boldsymbol{\delta})] - \theta_j\right)\right\} \end{aligned} \tag{2.22}$$

and applying (2.20) together with (2.14) and Pólya's Theorem (Bickel and Doksum 1977, p. 462), we obtain

$$\begin{aligned} &\lim_{\delta_1,\ldots,\delta_p\to 0} \Pr\{\theta_j < \hat{\theta}_j[N(\boldsymbol{\delta})] - \delta_j\} \\ =\; &\lim_{\delta_1,\ldots,\delta_p\to 0} \Pr\left\{t_{\nu_j(\infty)}(1-\alpha_j/2)\sqrt{\gamma_j} < N(0,1)\sqrt{\gamma_j}\right\} && (2.23) \\ \le\; &\Pr\left\{z(1-\alpha_j/2) < \mathrm{N}(0,1)\right\} && (2.24) \\ =\; &\alpha_j/2\,. && (2.25) \end{aligned}$$

A similar argument shows that

$$\lim_{\delta_1,\ldots,\delta_p\to 0} \Pr\{\hat{\theta}_j[N(\boldsymbol{\delta})] + \delta_j < \theta_j\} \le \alpha_j/2\,. \tag{2.26}$$

Combining (2.25) and (2.26) we obtain

$$\lim_{\delta_1,\ldots,\delta_p\to 0} \Pr\{|\hat{\theta}_j[N(\boldsymbol{\delta})] - \theta_j| < \delta_j\} \ge 1-\alpha_j\,. \tag{2.27}$$

Applying the Bonferroni inequality we obtain the desired result

$$\lim_{\delta_1,\ldots,\delta_p\to 0} \Pr\{|\hat{\theta}_1 - \theta_1| \le \delta_1 \wedge \ldots \wedge |\hat{\theta}_p - \theta_p| \le \delta_p\} \ge 1 - \sum_{j=1}^{p}\alpha_j = 1-\alpha\,. \tag{2.28}$$

### *2.2. Efficiency of the Termination Rule*

The termination rule proposed in (1.3) is a new way of applying the Bonferroni inequality. Our objective is to show that no other way of applying the Bonferroni inequality is more efficient than the proposed one, when the precision requirement (1.2) is used and when the distributions of $\left(\hat{\theta}_j(n) - \theta_j\right) \Big/ s_{\hat{\theta}_j(n)}$ are approximated by the Student $t$-distributions, each with $\nu_j(n)$ degrees of freedom.

Let us examine a simulation run that produces the $p$-dimensional output sequence $\{\boldsymbol{X}_i : i = 1, 2, \ldots\}$. Let $\{s^2_{\hat{\theta}_j(n)} : n = 2, 3, \ldots\}$, $j = 1, \ldots, p$, be the sequences of the estimated $\mathrm{Var}[\hat{\theta}_j(n)]$'s. Let $\nu_j(n)$ be the degrees of freedom for the estimator $ns^2_{\hat{\theta}_j(n)}$. Observe that the $s^2_{\hat{\theta}_j(n)}$'s and $\nu_j(n)$'s depend on the output sequence and on the method used to estimate the $\mathrm{Var}[\hat{\theta}_j(n)]$'s. When we examine the efficiency of different ways to apply the Bonferroni inequality in the termination rule, we can consider the sequences of the $s^2_{\hat{\theta}_j(n)}$'s and $\nu_j(n)$'s as sequences of constants. This implies that our comparison is restricted to situations in which the same method for estimating the $\mathrm{Var}[\hat{\theta}_j(n)]$'s is used.

Now, for a given $\alpha$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$, $0 < \alpha < 1$ and $\delta_j > 0$, $j = 1, \ldots, p$, the proposed termination rule terminates the simulation at $n_*$:

$$n_* = \text{smallest } n \text{ such that } 2\sum_{j=1}^{p} F_{\nu_j(n)}\left(-\delta_j \Big/ s_{\hat{\theta}_j(n)}\right) \le \alpha \, .$$

Let $n'$ present the length of the simulation run when we use any termination rule based on the Bonferroni inequality such that $\widehat{\Pr}\{|\hat{\theta}_1(n') - \theta_1| \le \delta_1 \wedge \ldots \wedge |\hat{\theta}_p(n') - \theta_p| \le \delta_p\} \ge 1 - \alpha$ when the distributions of $\left(\hat{\theta}_j(n) - \theta_j\right) \Big/ s_{\hat{\theta}_j(n)}$ are approximated by the Student $t$-distributions, each with $\nu_j(n)$ degrees of freedom. Under these assumptions the termination point $n'$ is determined by

$$n' = \text{smallest } n \text{ such that } 2F_{\nu_j(n)}\left(-\delta_j \Big/ s_{\hat{\theta}_j(n)}\right) \le \alpha_j(n) \, , \text{ for every } j = 1, \ldots, p \, ,$$

where the sequences $\{\alpha_j(m) : m = 1, 2, \ldots\}$, $j = 1, \ldots, p$, are sequences of positive constants such that $\alpha_1(m) + \ldots + \alpha_p(m) = \alpha$ for every $m = 1, 2, \ldots$ . The difference between the ways of applying the Bonferroni inequality is in the sequences of the $\alpha_j(m)$'s.

Let $\mathcal{A}(\alpha)$ be the set of all $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ such that $\alpha_j > 0$, $j = 1, \ldots, p$ and $\sum_{j=1}^{p} \alpha_j = \alpha < 1$. Let $\mathcal{N}(\alpha)$ be the set of all integers $m \ge 1$ such that $s_{\hat{\theta}_j(m)} \le \delta_j / t_{\nu_j(m)}(1 - \alpha_j(m)/2)$, for each $j = 1, \ldots, p$ and for all $\boldsymbol{\alpha}(m) \in \mathcal{A}(\alpha)$. It is straightforward to verify that the termination point $n'$ belongs to the set $\mathcal{N}(\alpha)$. The definition of $\mathcal{N}(\alpha)$ implies that, for every $m \in \mathcal{N}(\alpha)$, $\alpha_j(m) \ge 2F_{\nu_j(m)}\left(-\delta_j / s_{\hat{\theta}_j(m)}\right)$, $j = 1, \ldots, p$, and $\boldsymbol{\alpha}(m) \in \mathcal{A}(\alpha)$. Hence, for every $m \in \mathcal{N}(\alpha)$, we have $2\sum_{j=1}^{p} F_{\nu_j(m)}\left(-\delta_j / s_{\hat{\theta}_j(m)}\right) \le \sum_{j=1}^{p} \alpha_j(m) = \alpha$. Since $n' \in \mathcal{N}(\alpha)$, we obtain $n_* \le n'$. In other words, no way of applying the Bonferroni inequality is more efficient than the proposed one. However, some other ways may be as efficient as the proposed way.

## 3. Empirical Results

The asymptotic properties given in the previous section guarantee that the simultaneous coverage of the constructed confidence intervals will not be much lower than $1-\alpha$ when the $\delta_j$'s are small. In this section we report experimental results of constructing simultaneous confidence intervals for several means. The objective of these experiments is to find out how small the $\delta_j$'s should be.

We conducted three different sets of experiments. In the first set the observations were from 35 different binormal distributions. In the second set the observations were from 100 different ten-dimensional normal distributions. In the third set we used observations from the so-called $g$-and-$h$ distributions. A random variable $X$ is said to belong to the family of $g$-and-$h$ distributions (with parameters $a$, $b$, $g$, and $h$) when $X$ can be expressed as

$$X = a + \frac{b}{g}\left(\exp(gZ) - 1\right)\,\exp(hZ^2/2)\,,$$

where $Z$ is the standardized normal variable. The parameter $g$ controls the amount of skewness and the parameter $h$ controls the amount of elongation. The parameters $a$ and $b$ adjust the location and scale of the distribution, respectively. For details, see Hoaglin (1985). We examined 100 different five-dimensional $g$-and-$h$ distributions.

As stated in the Introduction, the normal and transformed normal distributions are used because we want to concentrate on the effects that the proposed way of applying the Bonferroni inequality has. Therefore, we selected output sequences for which the estimation of $\mathrm{Var}[\hat{\theta}_j(n)]$'s can be based on sample variances. The first and second set represent a situation in which the $t$-approximation is known to be exact. Therefore, in these cases observed low coverages must be due to the way the Bonferroni inequality is applied. The third set represents a situation in which the skewness and elongation affects the accuracy of the $t$-approximation. In this case the observed low coverages can also be due to the inaccuracies of the $t$-approximation.

In all three sets of experiments we used three different values of $\delta$ ($\delta_j = \delta\,,\ j = 1,\ldots,p$) and $\alpha$: $\delta = 0.20\,,\ 0.10\,,\ 0.05$ and $\alpha = 0.10\,,\ 0.05\,,\ 0.01$. The sequence lengths were determined as

$$N = \text{smallest } n > 1 \text{ such that } 2\sum_{j=1}^{p} F_{n-1}\left(-\delta/\sqrt{v_j(n)/n}\right) \le \alpha\,,$$

where $v_j(n) = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j(n))^2/(n-1)$. In order to reduce the amount of computing same sequences were used with different pairs of $(\delta,\alpha)$.

In the binormal case we conducted a systematic study for various $\Sigma$'s:

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 1 & \rho_{12}\sigma_2 \\ \rho_{12}\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

TABLE 1

*Observed Coverages: Output Sequences from 35 different binormal distributions*
1 000 *independent replications with each distribution*

| | | $\alpha$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | | | 0.05 | | | 0.01 | | |
| | | $\delta$ | | | $\delta$ | | | $\delta$ | | |
| $\sigma_2^2$ | $\rho_{12}$ | 0.20 | 0.10 | 0.05 | 0.20 | 0.10 | 0.05 | 0.20 | 0.10 | 0.05 |
| 1 | .00 | .903 | .899 | .908 | .961 | .947 | .953 | .993 | .996 | .991 |
| | .50 | .910 | .912 | .898 | .955 | .963 | .957 | .994 | .988 | .992 |
| | -.50 | .901 | .911 | .906 | .952 | .959 | .966 | .993 | *.985* | .991 |
| | .90 | .918 | .917 | .932 | .972 | .968 | .960 | .991 | .992 | .990 |
| | -.90 | .923 | .928 | .929 | .958 | .956 | .958 | .995 | .990 | .994 |
| | .99 | .942 | .943 | .953 | .968 | .975 | .971 | .995 | .990 | .997 |
| | -.99 | .927 | .939 | .932 | .971 | .977 | .969 | .993 | .992 | .987 |
| 4 | .00 | .910 | .901 | .892 | .960 | .942 | .955 | .991 | .987 | .993 |
| | .50 | .898 | *.879* | .901 | *.936* | *.940* | .950 | *.985* | .986 | .986 |
| | -.50 | .900 | .890 | .908 | .953 | .953 | .951 | .990 | .994 | .991 |
| | .90 | .900 | .900 | .909 | .948 | .944 | .954 | *.982* | .993 | .988 |
| | -.90 | **.876** | .897 | .900 | .942 | .943 | .947 | .989 | .994 | .988 |
| | .99 | *.886* | .903 | *.879* | .953 | .948 | .942 | .991 | *.985* | .986 |
| | -.99 | **.871** | .897 | .899 | .941 | .950 | .958 | .989 | .991 | .993 |
| 9 | .00 | .895 | *.883* | .888 | *.937* | .948 | .945 | .990 | .991 | *.984* |
| | .50 | .895 | .906 | .897 | .945 | .950 | .952 | .987 | .993 | .992 |
| | -.50 | .896 | .900 | .916 | .957 | .948 | .969 | .995 | .996 | .993 |
| | .90 | .907 | .911 | *.883* | .956 | .962 | *.933* | .995 | .988 | .989 |
| | -.90 | .903 | .896 | .908 | .943 | .955 | .957 | .991 | .993 | .986 |
| | .99 | .895 | .894 | .898 | .947 | .953 | .946 | .991 | .990 | .989 |
| | -.99 | .898 | .908 | *.881* | .949 | .946 | .945 | .991 | .989 | .990 |
| 16 | .00 | .915 | .909 | .891 | .955 | .941 | *.935* | .994 | .994 | .992 |
| | .50 | .893 | .902 | .894 | .941 | .946 | .944 | .989 | .991 | .986 |
| | -.50 | .896 | .889 | *.878* | .949 | .941 | .947 | .993 | .991 | .990 |
| | .90 | .888 | .900 | .905 | .953 | *.936* | .953 | .989 | .994 | .991 |
| | -.90 | .888 | .906 | .900 | .953 | .948 | .949 | .993 | .991 | .990 |
| | .99 | *.886* | *.886* | .902 | *.939* | .953 | .945 | .989 | .991 | **.980** |
| | -.99 | .893 | .897 | .897 | .945 | .943 | *.936* | .994 | .989 | *.982* |
| 25 | .00 | .898 | .902 | .903 | .957 | .942 | .955 | .989 | .996 | .990 |
| | .50 | .899 | .908 | .906 | .951 | .963 | .969 | .992 | .994 | .994 |
| | -.50 | *.886* | .900 | .907 | .947 | .947 | .943 | .990 | .994 | .993 |
| | .90 | .892 | .903 | .920 | *.940* | .954 | .965 | .993 | .994 | .988 |
| | -.90 | .892 | .902 | .896 | .956 | .957 | .947 | .992 | .991 | .990 |
| | .99 | .898 | .914 | .914 | .941 | .954 | .941 | .987 | .995 | .992 |
| | -.99 | *.884* | .888 | .894 | *.938* | .955 | .944 | .988 | .992 | .990 |

**bold:** hypothesis of *true coverage* $\geq 1-\alpha$ rejected at significance level 1%
*italics:* hypothesis of *true coverage* $\geq 1-\alpha$ rejected at significance level 5%
*italics:* hypothesis of *true coverage* $\geq 1-\alpha$ rejected at significance level 10%

The values of $\sigma_2$ examined were 1, 2, 3, 4, and 5. For each $\sigma_2$ the values of $\rho_{12}$ examined were 0, $\pm 0.50$, $\pm 0.90$, and $\pm 0.99$. The number of replications was 1 000 for each pair of $(\sigma_2, \rho_{12})$.

The observed coverages, that is the fraction of runs in which $|\bar{x}_1| \leq \delta$ and $|\bar{x}_2| \leq \delta$, are given in Table 1. Based on the observed coverages in independent replications and on the binomial distribution, we can construct a statistical test for the unknown true coverage. The hypothesis of acceptable true coverage

$$\begin{aligned} H_0 &: \Pr\{|\bar{x}_1| \leq \delta \wedge |\bar{x}_2| \leq \delta\} \geq 1 - \alpha \\ H_A &: \Pr\{|\bar{x}_1| \leq \delta \wedge |\bar{x}_2| \leq \delta\} < 1 - \alpha \end{aligned}$$

can be rejected only in a few test points. The observed coverages that reject null-hypothesis at significance levels 0.01, 0.05, and 0.10 are highlighted in Table 1. When we count the rejections in each column, we get only one frequency that is noticeably greater than the expected number of rejections when the null-hypothesis is true. This is the two rejections at level 0.01 for $\delta = 0.20$ and $\alpha = 0.10$. The probability of detecting two or more rejections in 35 independent tests at level 0.01 is 0.048 when the null-hypothesis is true. This is a slight indication that $\delta = 0.20$ (with $\alpha = 0.10$) may be too large.

In the case of ten-dimensional normal distribution we carried out a randomized experiment. The number of distributions, $\mathrm{N}_{10}(0, \Sigma)$, examined was 100. The covariance matrices were randomly generated: $\Sigma = U'U$, where $U$ is a $10 \times 10$ matrix of independent and uniformly in $[-1, 1]$ distributed random variables. The number of replications was 100 for each $(0, \Sigma)$.

The frequencies of observed coverages are given in Table 2. The hypothesis of acceptable true coverage

$$\begin{aligned} H_0 &: \Pr\{|\bar{x}_1| \leq \delta \wedge \ldots \wedge |\bar{x}_{10}| \leq \delta\} \geq 1 - \alpha \\ H_A &: \Pr\{|\bar{x}_1| \leq \delta \wedge \ldots \wedge |\bar{x}_{10}| \leq \delta\} < 1 - \alpha \end{aligned}$$

can be rejected only in a few test points. The observed coverages that reject null-hypothesis at significance levels 0.01, 0.05, and 0.10 are highlighted in Table 2. When we count the rejections in each column, we do not get any frequency that is noticeably greater than the expected number of rejections when the null-hypothesis is true.

In the case of five-dimensional $g$-and-$h$ distribution we also carried out a randomized experiment. The number of distributions examined was 100. For each distribution, the number of replications was 100. The distributions examined were constructed as follows. Firstly, we generated a $5 \times 5$ covariance matrix $\Sigma = (\sigma_{ij}) = U'U$, where $U$ is a $5 \times 5$ matrix of independent and uniformly in $[-1, 1]$ distributed random variables. The correlations between the standardized

TABLE 2
*Frequencies of Observed Coverages: Output Sequences from* $\mathrm{N}_{10}(0, \Sigma)$
*100 different* $\Sigma$*'s; 100 replications with each* $\Sigma$

| Obs. coverage | $\alpha$ = 0.10, $\delta$ = 0.20 | 0.10 | 0.05 | $\alpha$ = 0.05, $\delta$ = 0.20 | 0.10 | 0.05 | $\alpha$ = 0.01, $\delta$ = 0.20 | 0.10 | 0.05 |
|---|---|---|---|---|---|---|---|---|---|
| .81 | | | **1** | | | | | | |
| .82 | | | *1* | | | | | | |
| .83 | | *1* | | | | | | | |
| .84 | *3* | | | | | | | | |
| .85 | *1* | | *6* | | | | | | |
| .86 | 1 | 3 | 4 | | | | | | |
| .87 | | 5 | 4 | | | **1** | | | |
| .88 | 8 | 8 | 7 | | | **1** | | | |
| .89 | 10 | 8 | 8 | | *2* | | | | |
| .90 | 12 | 11 | 8 | *1* | *2* | | | | |
| .91 | 15 | 12 | 19 | *2* | *4* | | | | |
| .92 | 12 | 14 | 11 | 6 | 4 | 7 | | | |
| .93 | 18 | 20 | 10 | 8 | 8 | 13 | | | |
| .94 | 8 | 10 | 5 | 13 | 9 | 11 | | | |
| .95 | 7 | 4 | 9 | 14 | 13 | 17 | | | **1** |
| .96 | 4 | 2 | 5 | 20 | 23 | 14 | | *1* | *2* |
| .97 | 1 | 2 | 2 | 18 | 19 | 22 | *4* | *5* | *9* |
| .98 | | | | 12 | 12 | 10 | 15 | 14 | 15 |
| .99 | | | | 6 | 4 | 3 | 42 | 44 | 36 |
| 1.0 | | | | | | 1 | 39 | 36 | 37 |

**bold:** hypothesis of *true coverage* $\geq 1 - \alpha$ rejected at significance level 1%
*italics:* hypothesis of *true coverage* $\geq 1 - \alpha$ rejected at significance level 5%
*italics:* hypothesis of *true coverage* $\geq 1 - \alpha$ rejected at significance level 10%

normal variables $(Z_1, \ldots, Z_5)$ were chosen as $\mathrm{Corr}(Z_i, Z_j) = \sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}$. Secondly, the parameters controlling the amount of skewness $(g_1, \ldots, g_5)$ were generated from the uniform distribution $(0, 1)$. Thirdly, the parameters controlling the amount of elongation $(h_1, \ldots, h_5)$ were generated from the uniform distribution $(0, 0.5)$. Finally, the parameters adjusting the scales $(b_1, \ldots, b_5)$ and the locations $(a_1, \ldots, a_5)$ were determined so that

$$\begin{aligned} \mathrm{Var}(X_i) &= \frac{b_i^2}{g_i^2\sqrt{1-2h_i}} \left[ \exp\left(\frac{2g_i^2}{1-2h_i}\right) - 2\exp\left(\frac{g_i^2}{2(1-2h_i)}\right) + 1 \right] \\ &\quad - \frac{b_i^2}{g_i^2(1-h_i)} \left[ \exp\left(\frac{g_i^2}{2(1-h_i)}\right) - 1 \right]^2 \\ &= \sigma_{ii}\,, \end{aligned}$$

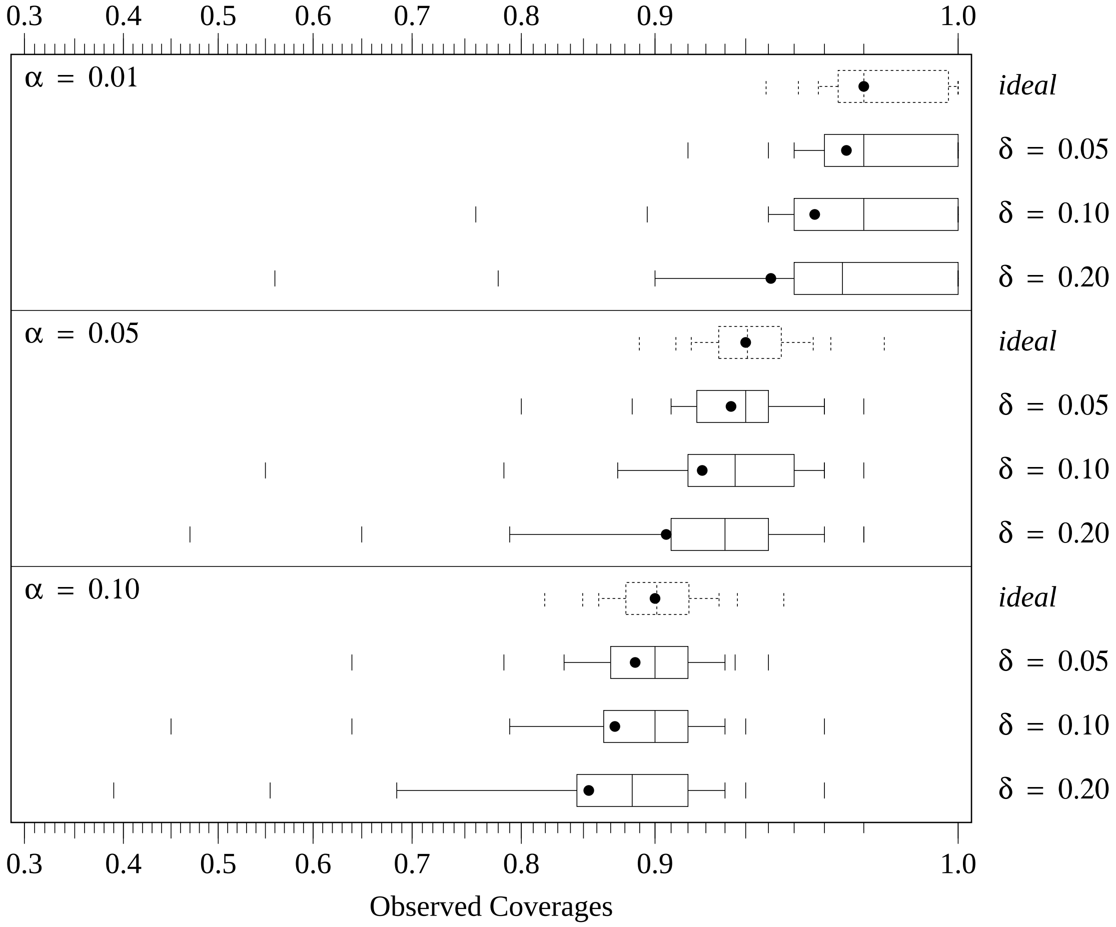


FIGURE 1: Boxplots of Observed Coverages: Output Sequences from 100 different five-dimensional $g$-and-$h$ Distributions; 100 replications with each distribution Dashed boxes represent the expected boxplots based on 100 observations from the binomial distribution $Bin(100, 1-\alpha)$.

$$\begin{aligned} \mathrm{E}(X_i) &= a + \frac{b_i}{g_i\sqrt{1-h_i}}\left[\exp\left(\frac{g_i^2}{2(1-h_i)}\right) - 1\right] \\ &= 0\,. \end{aligned}$$

Above $\mathrm{E}(X_i)$ and $\mathrm{Var}(X_i)$ are based on expressions given in Hoaglin (1985, p. 503). The $\sigma_{ii}$'s are the diagonal elements of $\Sigma = U'U$ generated in the first step.

The observed coverages are summarized in Figure 1 which shows the boxplots. In the boxplot a bullet '•' denotes the mean of observed coverages. The three quartiles are visualized by a box '  '. The whiskers '  ' reach from the 25$^{\mathrm{th}}$ to the 10$^{\mathrm{th}}$ percentile and from the 75$^{\mathrm{th}}$ to the 90$^{\mathrm{th}}$ percentile. The 5$^{\mathrm{th}}$ and 95$^{\mathrm{th}}$ percentile together with the smallest and largest observed coverage are marked by ticks '|'. In addition to the boxplots for the nine pairs of $(\alpha, \delta)$, the figure shows expected or ideal boxplots for each $\alpha$. These boxplots are dashed and labeled as "ideal". They are based on the expected values of the order statistics when 100 independent observations are sampled from the binomial distribution $Bin(100, 1-\alpha)$.

Based on Figure 1 we can draw the following three conclusions. Firstly, the number of observed low coverages highly exceeds the expectations; especially when $\delta = 0.20$. Secondly, the upper halves in the observed distributions of coverages are quite close to the expectations. Thirdly, the boxplots converge towards the ideals as the error tolerance $\delta$ decreases.

As stated earlier in this section, we have two primary explanations for the low observed coverages. The alternatives are: 1) The proposed way of applying the Bonferroni inequality is inadequate. 2) The convergence to the normal limiting distribution is so slow that the $t$-approximation is inaccurate. In order to verify that the second explanation is more plausible than the first one, we examined the marginal output sequences using the termination rule

$$N_j = \text{smallest } n > 1 \text{ such that } 2F_{n-1}\left(-\delta \Big/ \sqrt{v_j(n)/n}\right) \leq \alpha \ .$$

This termination rule is the same as the traditional termination rule given in Eq. (1.1). For each five-dimensional distribution examined we obtain five observed marginal coverages, that is the fraction of runs for which $|\bar{x}_j(N_j)| \leq \delta$.

Since each component of the five-dimensional distribution affects the proposed termination rule, we compare the lowest marginal coverage to the simultaneous coverage. The comparison is visualized in Figure 2 that provides the boxplots of paired differences. The conclusion is self-evidence. The observed simultaneous coverage when the proposed termination rule is applied highly exceeds the observed lowest marginal coverage when the traditional termination rule is separately applied to the marginal sequences. This is a clear indication that the inaccuracy of the $t$-approximation is the primary reason to the low observed coverages.

The overall conclusion of these experiments is that the proposed way of applying the Bonferroni inequality is appropriate even when the $\delta$'s are small provided that the $t$-approximation of $(\hat{\theta}-\theta)/s_{\hat{\theta}}$ is appropriate. It should be noted that we have applied the method only in cases of independent observations. However, the asymptotic coverage and the results of these experiments do not give any indication that our way of applying the Bonferroni inequality would introduce serious errors in realistic simulation studies. The observed low coverages in the case of some $g$-and-$h$ distributions suggest that the termination rule should also verify that the $t$-approximation is appropriate.

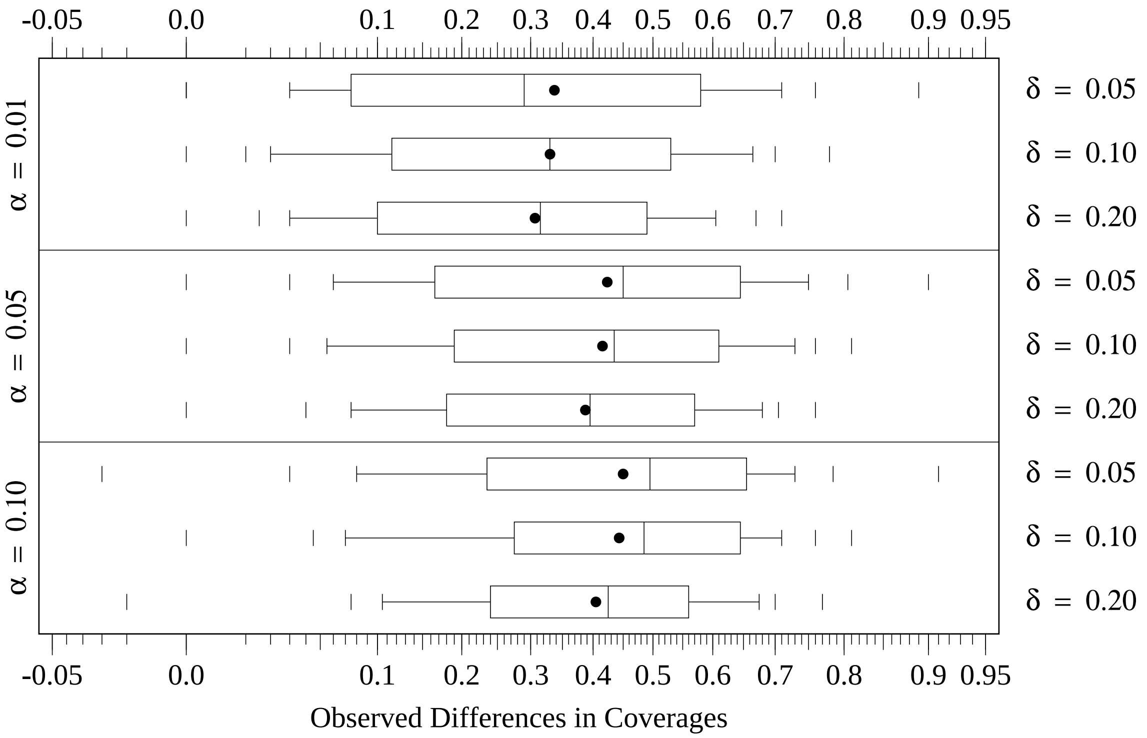


FIGURE 2: Boxplots of Observed Paired Difference Between Simultaneous Coverage and Lowest Marginal Coverage
Output Sequences from 100 different five-dimensional $g$-and-$h$ Distributions;
100 replications with each distribution

## Summary

We have proposed a new way of applying the Bonferroni inequality when simultaneous confidence intervals of fixed-widths for several means are constructed. We have shown that the simultaneous coverage of intervals $I_j(N) = [\hat{\theta}_j(N) - \delta_j, \hat{\theta}_j(N) + \delta_j]$, $j = 1, \ldots, p$, is at least $1 - \alpha$ when $\delta_j \to 0$, $j = 1, \ldots, p$, and

$$N = \text{smallest } n \text{ such that } 2 \sum_{j=1}^{p} F_{\nu_j(n)} \left( -\delta_j / s_{\hat{\theta}_j(n)} \right) \leq \alpha \,,$$

where $s^2_{\hat{\theta}_j(n)}$ is the estimated variance of $\hat{\theta}_j(n)$ with $\nu_j(n)$ degrees of freedom. The result holds under the following three assumptions. Firstly, the (marginal) output sequences must be covariance stationary and satisfy certain regularity conditions (see e.g. Anderson, 1971, p. 478). Secondly, the $ns^2_{\hat{\theta}_j(n)}$'s must be (almost surely) positive, and they must also be consistent estimators of the $n\,\mathrm{Var}[\hat{\theta}_j(n)]$'s. Thirdly, the degrees of freedom for the estimator $ns^2_{\hat{\theta}_j(n)}$ must converge in probability to (possibly infinite) constants. The first assumption solely depends on the simulation model. The last two assumptions also depend on the method used to estimate the Var$[\hat{\theta}_j(n)]$'s.

In addition to the asymptotic coverage, we have shown that our way of applying the Bonferroni inequality is the most efficient one. We have also reported satisfactory observed coverages for small positive $\delta_j$'s when confidence intervals for normal means are constructed. When we estimated transformed normal means, we also observed coverages that are highly too low. However, our verification based on analyzing the marginal sequences indicates that the observed low coverages were primarily due to the $t$-approximation of $\hat{\theta}_j(n) - \theta_j)/s_{\hat{\theta}_j(n)}$; not due to the proposed way of applying the Bonferroni inequality. Additional experimental results for more complex stochastic systems can be found in Raatikainen (1993, 1994).

## References

ANDERSON, T. W. 1971. *The Statistical Analysis of Time Series*. John Wiley & Sons, New York.

BICKEL, P. J., AND DOKSUM, K. A. 1977. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, San Francisco.

BILLINGSLEY, P. 1968. *Convergence of Probability Measures*. John Wiley & Sons, New York.

CHOW, Y., AND ROBBINS, H. 1965. "On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean". *Ann. Math. Statist. 36,* 2, 457–462.

HOAGLIN, D. C. 1985. "Summarizing Shape Numerically: The $g$-and-$h$ Distributions". In *Exploring Data Tables, Trends, and Shapes*, D. C. Hoaglin, F. Mosteller, and J. W. Tukey, Eds. John Wiley & Sons, New York.

KLEIJNEN, J. P. C. 1987. *Statistical Tools for Simulation Practitioners*. Marcel Dekker, Inc., New York.

LAW, A. M., AND KELTON, W. D. 1991. *Simulation Modeling and Analysis*, 2nd Ed. McGraw-Hill, New York.

PAWLIKOWSKI, K. 1990. "Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions". *ACM Comput. Surv. 22,* 2, 123–170.

RAATIKAINEN, K. E. E. 1993. "A Sequential Procedure for Simultaneous Estimation of Several Means". *ACM Trans. on Mod. Comp. Simul. 3,* 2, 108–133.

RAATIKAINEN, K. E. E. 1994. "Simulation-Based Estimation of Proportions". *Manag. Sci.*, to appear.

SERFLING, R. J.. 1980. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.