# SIMULATION-BASED ESTIMATION OF PROPORTIONS

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

*Simulated estimates for several proportions are needed in many simulation studies. We propose a method for controlling the length of a simulation run so that the proportions estimated satisfy a prespecified precision requirement.*
*The method applies the arcsin transform in order to generate confidence intervals with the desired width. The Bonferroni inequality is applied in a new way to construct a rectangular confidence area for the proportions being estimated. Since we are using the Bonferroni inequality, we can use the existing methods developed for estimating the variance of a single mean.*
*The properties of the method proposed depend on three factors: 1) the arcsin transform, 2) the proposed way of applying the Bonferroni inequality, and 3) the method for estimating the variance. We examine the effects that the transform and the way of applying the Bonferroni inequality have.*
*Empirical results indicate that the method provides estimates that satisfy the prespecified precision requirement, when the spectral method is used to estimate the variances of the proportions. In addition, simulation runs, when estimating either several proportions or the corresponding mean, are about the same lengths.*
(DISCRETE-EVENT SIMULATION, SIMULTANEOUS ESTIMATION, RUN LENGTH CONTROL)

## 1. Introduction

Most methods for controlling the precision of simulation results have concentrated on the mean of one response variable. In many practical applications, however, the analyst is interested not only in the mean but also in other characteristics of the simulation output. These characteristics include the variance, quantiles or percentiles, and proportions or percentages. In this paper we propose a method for controlling the precision when several proportions are simultaneously estimated. In queueing network analysis, which is our primary application area of simulation, the need for several

Manuscript of paper to appear in *Management Science*

simultaneous proportions arises, for example, when the distribution of queue length is estimated or when the distribution of the system response times is approximated through a histogram.

When simulation is used to estimate several proportions, the following two problems arise: 1) How do we specify the precision requirement for individual proportions? 2) How do we conclude that the individual requirements are simultaneously satisfied? The second problem can be solved by applying the Bonferroni inequality; see e.g. Kleijnen (1987, pp. 41–43). The first problem is discussed below.

When the mean of a response variable is estimated, the precision requirement is usually specified using the so-called relative half-width criterion. For simplicity in the following discussion, we assume that the unknown mean $\theta$ and its estimate $\hat{\theta}$ are both positive. When the relative half-width criterion is used, the analyst specifies the desired relative half-width $\varepsilon$ of the $100(1-\alpha)\%$ confidence interval for the mean. Hence, the parameters controlling the precision are $\varepsilon$ and $\alpha$. The desired confidence interval for the unknown mean $\theta$ is $[(1-\varepsilon)\hat{\theta},(1+\varepsilon)\hat{\theta}]$, where $\hat{\theta}$ is the estimated mean. The simulation is continued until the estimated confidence interval $[\hat{\theta}_{\mathrm{L}},\hat{\theta}_{\mathrm{U}}]$ is inside the desired confidence interval, i.e. $(1-\varepsilon)\hat{\theta} \leq \hat{\theta}_{\mathrm{L}} < \hat{\theta}_{\mathrm{U}} \leq (1+\varepsilon)\hat{\theta}$.

The methods for controlling the precision assume that the estimator $\hat{\theta}$ has a normal limiting distribution. Therefore, they use the Student $t$-distribution and terminate the simulation when $s_{\hat{\theta}}/\hat{\theta} \leq \varepsilon/t_{\nu}(1-\alpha/2)$ , where $s^2_{\hat{\theta}}$ is the estimated variance of $\hat{\theta}$ with $\nu$ degrees of freedom and $t_{\nu}(x)$ is the $100x^{\mathrm{th}}$ percentile of the Student $t$-distribution with $\nu$ degrees of freedom.

The primary reason for using relative half-width (instead of the absolute one) is that the analyst is interested in the magnitude and in one or two leading digits of the mean. However, the situation is different when proportions are estimated. Typically, the closer to the end point of the interval $(0,1)$ the proportion is, the narrower the confidence interval should be. In addition, the distribution of an estimated proportion is, in general, skew. We use the arcsin transform both to take the skewness into account in the confidence interval estimation and to select the width of the desired confidence interval.

In Section 2 we discuss the estimation of a single proportion. In Section 3 the method described in Section 2 is generalized to control the simultaneous precision of several proportions. In Section 4 we report empirical results for the proposed method. We provide results of estimating several proportions for the arrival queue length distribution in an M/M/1 queue and for two system response time distributions in a closed queueing network model. In order to learn how expensive the estimation of several proportions is, the results reported also include the run lengths when the corresponding means are estimated.

## 2. Estimating a Single Proportion

In this section we discuss the estimation of a single proportion, say $p$. We use the arcsin transform, $\hat{y} = 2\sin^{-1}\left(\sqrt{\hat{p}}\right)$, to construct an approximate confidence interval for the proportion being estimated. The transform is also used to select the desired width of the confidence interval. This method is not able to control the accuracy if the estimated proportion is very close to 0 or 1. Therefore, we complete the section by discussing briefly the problem of small proportions (rare events).

When a proportion is estimated, we are interested in the probability $p$ that the response variable $X$ belongs to a prespecified category $\mathcal{C}$: $p = \Pr\{X \in \mathcal{C}\}$. An estimate of $p$ is based on a transformation of the output sequence $\{X_j\}_{j=1}^{n}$:

$$\hat{p} = \frac{1}{n}\sum_{j=1}^{n} I_j \ , \tag{2.1}$$

where $I_j = 1$ if $X_j \in \mathcal{C}$, and $I_j = 0$ if $X_j \notin \mathcal{C}$.

### *2.1. Confidence Interval for the Proportion*

If the $I_j$'s are independent and identically distributed, $\mathrm{E}(I_j) = p$ and $\mathrm{Var}(I_j) = p(1-p)$, then an exact confidence interval for the estimated proportion $\hat{p}$ is obtained using the binomial distribution. Although $\hat{p}$ has a normal limiting distribution, the confidence interval based on the normal approximation is not appropriate if $p$ is close to 0 or 1. One of the methods proposed to improve the approximate confidence interval $[\hat{p}_{\mathrm{L}}, \hat{p}_{\mathrm{U}}]$ is based on the arcsin transform; see e.g. Hald (1952, p. 699):

$$\left.\begin{array}{rcl} 2\sin^{-1}\left(\sqrt{\hat{p}_{\mathrm{L}}}\right) & = & 2\sin^{-1}\left(\sqrt{\hat{p} - 1/(2n)}\right) - z(1-\alpha/2)/\sqrt{n} \\ 2\sin^{-1}\left(\sqrt{\hat{p}_{\mathrm{U}}}\right) & = & 2\sin^{-1}\left(\sqrt{\hat{p} + 1/(2n)}\right) + z(1-\alpha/2)/\sqrt{n} \end{array}\right\} \ , \tag{2.2}$$

where $z(x)$ is the $100x^{\mathrm{th}}$ percentile of the standardized normal distribution. The confidence interval (2.2) is based on the normal approximation, $\hat{y} \equiv 2\sin^{-1}\left(\sqrt{\hat{p}}\right) \sim \mathrm{N}(y, 1/n)$, where $y \equiv 2\sin^{-1}\left(\sqrt{p}\right)$ , and on a continuity correction.

We use the same approach to generate a confidence interval for the estimated proportion in steady-state simulation. However, now we cannot assume that the $I_j$'s are independent. Instead, we assume that the sequence $\{I_j\}$ is covariance stationary. In this case

$$\mathrm{Var}(\hat{p}) = \frac{p - p^2}{n}\left(1 + 2\sum_{t=1}^{n-1}(1 - t/n)\rho_t\right) \ ,$$

where $\rho_t = \mathrm{Corr}(I_j, I_{j+t})$, see e.g. Welch (1983, pp. 304–305). Since $\hat{p}$ is based on a cumulative statistic, we can use any method developed for estimating the variance of the mean to estimate

$\mathrm{Var}(\hat{p})$. Available methods are described in several textbooks and survey articles including Law and Kelton (1991), Pawlikowski (1990), and Welch (1983).

Let $s_{\hat{p}}^2$ denote the estimated variance of $\hat{p}$ with $\nu_p$ degrees of freedom. The degrees of freedom depend on the method of estimating the variance. For example, if we group the indicator variables into $b$ nonoverlapping batches each consisting $g = \lfloor n/b \rfloor$ observations, we have $\nu_p = b-1$ and

$$s_{\hat{p}}^2 = \frac{1}{b(b-1)} \sum_{i=1}^{b} (\bar{I}_i - \hat{p})^2$$

where $\bar{I}_i = \sum_{j=1}^{g} I_{(i-1)g+j} / g$, $i = 1, \ldots, b$; see e.g. Welch (1983, pp. 307–309).

When the sequence $\{I_j\}$ is covariance stationary, then the limiting distribution of $\sqrt{n}\,(\hat{p}-p)$ is normal if the sequence $\{I_j\}$ satisfies certain regularity conditions, which are fairly general; see e.g. Anderson (1971, p. 478). Since the transform $g(x) \equiv 2\sin^{-1}(\sqrt{x})$ and its derivates are continuous when $0 < x < 1$, we obtain through the Taylor expansion; see e.g. Cox and Hinkley (1974, pp. 261 and 275): $\mathrm{E}(\hat{y}) \approx 2\sin^{-1}(\sqrt{p})$ and $\mathrm{Var}(\hat{y}) \approx \mathrm{Var}(\hat{p})/(p-p^2)$ when $n$ is large and $0 < p < 1$. Therefore, $\sqrt{n}\,(\hat{y}-y)$ converges, as $n \to \infty$, in distribution to $N(0, C)$, where $C = 1 + 2\sum_{t=1}^{\infty} \rho_t$.

Based on the normal limiting distribution we can use, when $n$ is large, the Student $t$-distribution to approximate the distribution of $(\hat{y}-y)/s_{\hat{y}}$, where $s_{\hat{y}}^2$ is the estimated variance of $\hat{y}$. A natural estimate of $\mathrm{Var}(\hat{y})$ is obtained using the approximation $\mathrm{Var}(\hat{y}) \approx \mathrm{Var}(\hat{p})/(p-p^2)$. Therefore, the natural estimate of $\mathrm{Var}(\hat{y})$ is $s_{\hat{p}}^2/(\hat{p}-\hat{p}^2)$, i.e. $s_{\hat{y}}^2 = s_{\hat{p}}^2/(\hat{p}-\hat{p}^2)$.

In order to use the Student approximation we need the degrees of freedom in $s_{\hat{y}}^2$. A common practice, which is used for example in Fishman (1978, p. 252), is to approximate the distribution of $s_{\hat{y}}^2/\mathrm{Var}(\hat{y})$ by a $\chi^2$-variate with $\nu_y$ degrees of freedom: $\nu_y = 2\left\{\mathrm{E}\left[s_{\hat{y}}^2/\mathrm{Var}(\hat{y})\right]\right\}^2 \Big/ \mathrm{Var}[s_{\hat{y}}^2/\mathrm{Var}(\hat{y})]$. When $s_{\hat{p}}^2$ and $\hat{p}$ are independent and $n$ is large, we have

$$\nu_y \approx \frac{2\left\{\mathrm{E}\left[\dfrac{s_{\hat{p}}^2}{\mathrm{Var}(\hat{p})} \Big/ \dfrac{\hat{p}-\hat{p}^2}{p-p^2}\right]\right\}^2}{\mathrm{Var}\left[\dfrac{s_{\hat{p}}^2}{\mathrm{Var}(\hat{p})} \Big/ \dfrac{\hat{p}-\hat{p}^2}{p-p^2}\right]} \approx \frac{2[\mathrm{E}(\,s_{\hat{p}}^2\,)]^2}{\mathrm{Var}(\,s_{\hat{p}}^2\,)} \equiv \nu_p\,.$$

Therefore, we can approximate the degrees of freedom in $s_{\hat{y}}^2$ by $\nu_p$, that is $\nu_y = \nu_p = \nu$. Now we obtain the following approximate confidence interval for the proportion being estimated

$$\left.\begin{array}{rcl} \hat{p}_{\mathrm{L}} &=& \{\sin(\max(0, \hat{y}_{\mathrm{L}}/2))\}^2 \\ \hat{p}_{\mathrm{U}} &=& \{\sin(\min(\pi/2, \hat{y}_{\mathrm{U}}/2))\}^2 \end{array}\right\}, \tag{2.3}$$

where

$$\left.\begin{array}{rcl} \hat{y}_{\mathrm{L}} &=& 2\sin^{-1}\left(\sqrt{\hat{p}}\right) - t_\nu(1-\alpha/2)s_{\hat{p}}/\sqrt{\hat{p}(1-\hat{p})} \\ \hat{y}_{\mathrm{U}} &=& 2\sin^{-1}\left(\sqrt{\hat{p}}\right) + t_\nu(1-\alpha/2)s_{\hat{p}}/\sqrt{\hat{p}(1-\hat{p})} \end{array}\right\}, \tag{2.4}$$

and $s_{\hat{p}}^2$ is the estimated variance of $\hat{p}$ with $\nu$ degrees of freedom.

### *2.2. Precision Requirement for the Estimated Proportion*

As discussed in the Introduction, the relative half-width criterion, which is commonly used to specify the precision requirement for a mean, is not appropriate for a proportion. Typically, the closer to the end points of the interval $(0,1)$ the proportion is, the narrower the confidence interval requested is. There are several ways of selecting the desired width of the confidence interval. We use the arcsin transform to scale the desired width.

We specify the precision requirement for the proportion as the desired half-width $\delta_y$ of the $100(1-\alpha)\%$ confidence interval for $y = 2\sin^{-1}\left(\sqrt{p}\right)$. When the desired confidence interval for $y$ is $[\hat{y}-\delta_y, \hat{y}+\delta_y]$, then the desired confidence interval for $p$ is $[\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$:

$$\left.\begin{array}{rcl} \tilde{p}_{\mathrm{L}} & = & \left\{\sin\left(\max(0, \sin^{-1}\left(\sqrt{\hat{p}}\right) - \delta_y/2)\right)\right\}^2 \\ \tilde{p}_{\mathrm{U}} & = & \left\{\sin\left(\min(\pi/2, \sin^{-1}\left(\sqrt{\hat{p}}\right) + \delta_y/2)\right)\right\}^2 \end{array}\right\} . \tag{2.5}$$

Figure 1 demonstrates how the desired confidence interval for $p$ depends on $\hat{p}$ when $\delta_y = 2^{-i}$, $i = 1, \ldots, 5$. The analyst may find it difficult to select a suitable $\delta_y$. One solution is to specify $\tilde{p}_{\mathrm{L}}$ or $\tilde{p}_{\mathrm{U}}$ for a specific $\hat{p}$ and to solve $\delta_y$ from (2.5). For example, if we want to have $\tilde{p}_{\mathrm{L}} = 1/2 - \delta_p$ when $\hat{p} = 1/2$, then we take $\delta_y = \sin^{-1}(2\delta_p)$.

The simulation will be continued until the desired confidence interval (2.5) includes the estimated confidence interval (2.3), that is $\tilde{p}_{\mathrm{L}} \le \hat{p}_{\mathrm{L}} < \hat{p}_{\mathrm{U}} \le \tilde{p}_{\mathrm{U}}$. This requirement implies that the simulation will be continued until

$$s_{\hat{p}} \le \delta_y \sqrt{\hat{p}(1-\hat{p})} \Big/ t_\nu(1-\alpha/2) . \tag{2.6}$$

The sequential termination rules that are derived from a confidence interval of fixed sample size, such as (2.6), are common in the practice of simulation. Their motivation is based on the asymptotic consistency and efficiency of sequential confidence intervals of fixed-width; see Chow and Robbins (1965) and Damerdji (1987).

### *2.3. Properties of the Termination Rule Proposed*

The properties of the termination rule (2.6) depend on the method for estimating $\mathrm{Var}(\hat{p})$ – see the left side of (2.6) – and on the approximation $\hat{y} \sim N(y, C/n)$. Since we base our method on existing methods for estimating $\mathrm{Var}(\hat{p})$, we concentrate on the effects that the approximation $\hat{y} \sim N(y, C/n)$ has. We examine the case of independent indicator sequence and the case of dependent indicator sequence with known correlation structure. Finally, we briefly analyze the case of unknown correlation structure.

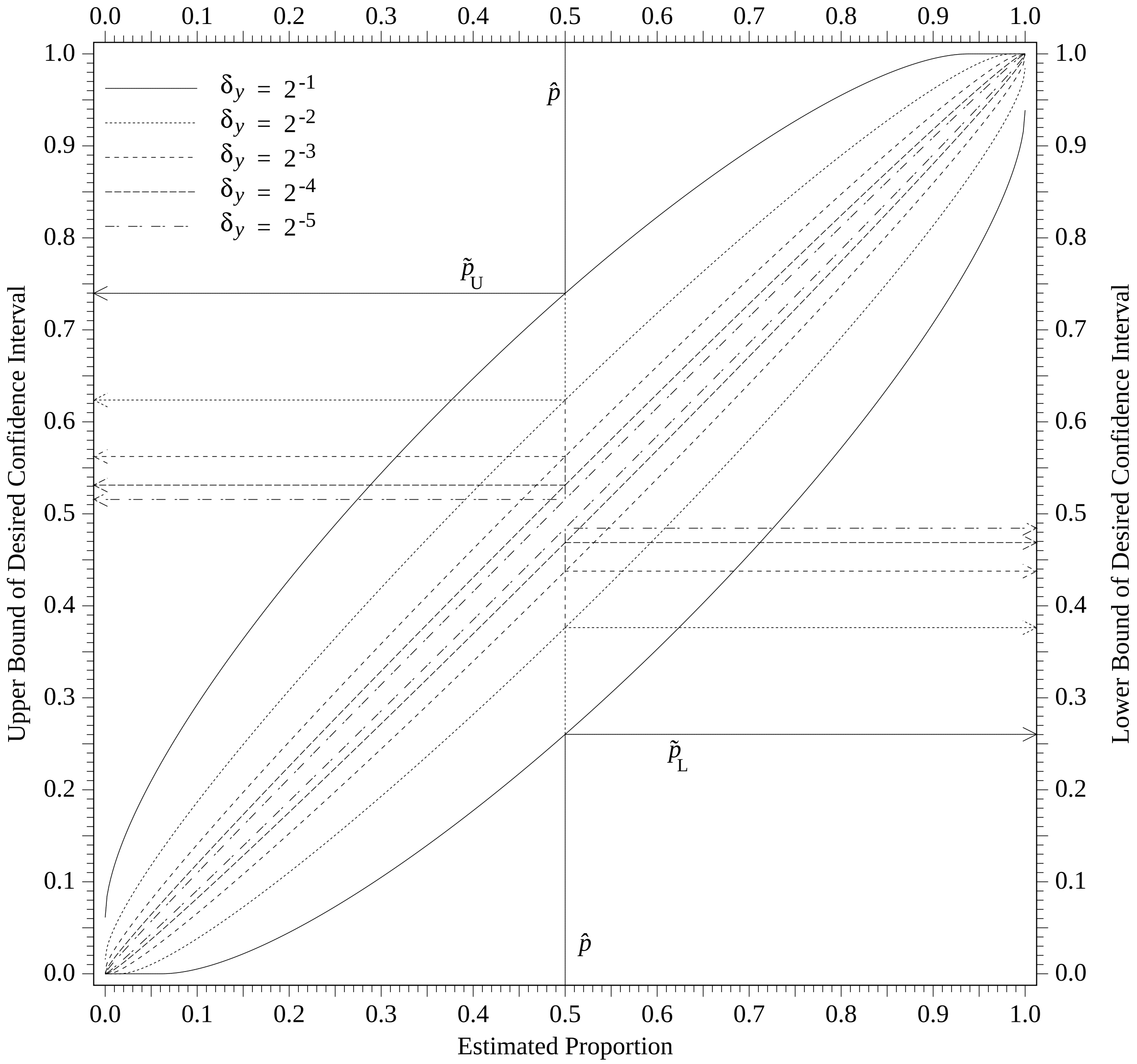



FIGURE 1: Desired confidence intervals $[\tilde{p}_L, \tilde{p}_U]$ for proportion $p$ when the desired confidence interval for $y = 2\sin^{-1}\left(\sqrt{p}\right)$ is $[\hat{y} - \delta_y, \hat{y} + \delta_y]$.

### *2.3.1. Sequence of Independent Indicators*

Let the indicator sequence be independent. This implies that $\hat{p} = k/n$ and $k \sim \mathrm{Bin}(n,p)$. The variance is now known, and the termination rule provides the fixed sample size

$$n = \left\lceil (z(1-\alpha/2)/\delta_y)^2 \right\rceil . \tag{2.7}$$

From the Taylor expansion of $\hat{y} = 2\sin^{-1}\left(\sqrt{\hat{p}}\right)$ about the point $p$ we obtain

$$\begin{aligned}\mathrm{Var}(\hat{y}) &= \mathrm{E}\left(\hat{y} - \mathrm{E}(\hat{y})\right)^2 \\ &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} g^{(j)}(p)g^{(k)}(p)(\mu_{j+k} - \mu_j\mu_k)/(j!k!) ,\end{aligned}$$

where $g^{(k)}(x)$ is the $k^{\mathrm{th}}$ derivate of $g(x) \equiv 2\sin^{-1}\left(\sqrt{x}\right)$ and $\mu_k = \mathrm{E}(\hat{p}-p)^k$. By taking the terms that contain $n^{-1}$ or $n^{-2}$ we obtain the following approximation

$$\begin{aligned}\mathrm{Var}(\hat{y}) \;\approx\; & \left[g^{(1)}(p)\right]^2 \mu_2 + g^{(1)}(p)g^{(2)}(p)\mu_3 \\ & +\frac{1}{4}\left[g^{(2)}(p)\right]^2(\mu_4 - \mu_2^2) \\ & +\frac{1}{3}g^{(1)}(p)g^{(3)}(p)\mu_4 .\end{aligned}$$

By substituting

$$\begin{aligned} g^{(1)}(p) &= (p-p^2)^{-1/2} \\ g^{(2)}(p) &= -\frac{1}{2}(1-2p)(p-p^2)^{-3/2} \\ g^{(3)}(p) &= (p-p^2)^{-3/2} + \frac{3}{4}(1-2p)^2(p-p^2)^{-5/2} \\ \mu_2 &= p(1-p)/n \\ \mu_3 &= p(1-p)(1-2p)/n^2 \\ \mu_4 &= 3p^2(1-p)^2/n^2 + \mathcal{O}(p(1-p)n^{-3}) \end{aligned}$$

we obtain

$$n\,\mathrm{Var}(\hat{y}) = 1 + \frac{3-4p(1-p)}{8np(1-p)} + \mathcal{O}\left(\{np(1-p)\}^{-2}\right) .$$

Hence $n\,\mathrm{Var}(\hat{y}) \longrightarrow 1$ as $n \to \infty$, provided $0 < p < 1$.

In practice, it is important to learn how large $n$ must be in order to have $n\,\mathrm{Var}(\hat{y}) \approx 1$. Figure 2 shows $n\,\mathrm{Var}(\hat{y})$ as a function of $n$ for $p = 2^{-i}$, $i = 1,\ldots,15$. The curves are based on exact values of $n\,\mathrm{Var}(\hat{y})$ for $n = 2^i$, $i = 1,\ldots,20$. The figure indicates that $n\,\mathrm{Var}(\hat{y}) \approx 1$ when $np \geq 16$ and $p \leq 1/2$. Because the situation is symmetric for $p$ and $1-p$, we conclude that $n\,\mathrm{Var}(\hat{y}) \approx 1$, when $n(1-p) \geq 16$ and $p \geq 1/2$. Based on the figure and on the Taylor expansion, we conclude that the approximation $\mathrm{Var}(\hat{y}) = 1/n$ does not introduce serious error when $n \geq 16/\min(p, 1-p)$.

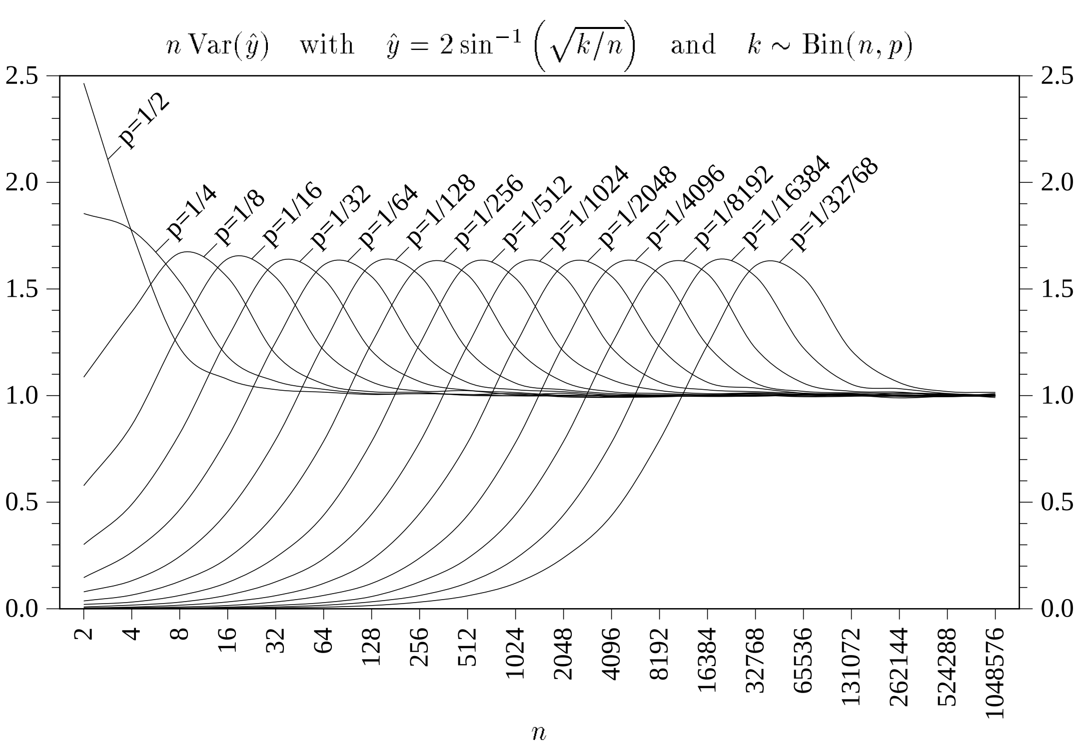



FIGURE 2: Accuracy of Approximation $\mathrm{Var}(\hat{y}) = 1/n$ when $\hat{y} = 2\sin^{-1}\left(\sqrt{k/n}\right)$ and $k \sim \mathrm{Bin}(n, p)$

The desired confidence interval $[\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$ is asymptotically consistent in the sense that

$$\lim_{\delta_y \to 0} \Pr\{p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]\} = 1 - \alpha \,.$$

In order to show the asymptotic consistency, we use (2.5). When $\{\sin(\delta_y/2)\}^2 \le \hat{p} \le 1 - \{\sin(\delta_y/2)\}^2$, we have

$$\tilde{p}_{\mathrm{L}} = \left\{\sin\left[\sin^{-1}\left(\sqrt{\hat{p}}\right) - \delta_y/2\right]\right\}^2$$

and

$$\tilde{p}_{\mathrm{U}} = \left\{\sin\left[\sin^{-1}\left(\sqrt{\hat{p}}\right) + \delta_y/2\right]\right\}^2 \,.$$

Applying the basic trigonometric formulas we obtain

$$\begin{aligned}
&\left\{\sin\left[\sin^{-1}\left(\sqrt{\hat{p}}\right) \pm \delta_y/2\right]\right\}^2 = \\
&= \left\{\sin\left[\sin^{-1}\left(\sqrt{\hat{p}}\right)\right]\cos(\delta_y/2) \pm \cos\left[\sin^{-1}\left(\sqrt{\hat{p}}\right)\right]\sin(\delta_y/2)\right\}^2 \\
&= \hat{p}\,\{\cos(\delta_y/2)\}^2 \pm 2\sin(\delta_y/2)\cos(\delta_y/2)\sqrt{\hat{p}}\cos\left[\cos^{-1}\left(\sqrt{1-(\sqrt{\hat{p}})^2}\right)\right] \\
&\quad + \left(1 - \left\{\sin\left[\sin^{-1}\left(\sqrt{\hat{p}}\right)\right]\right\}^2\right)\{\sin(\delta_y/2)\}^2
\end{aligned}$$

$$\begin{aligned} &= \hat{p}\frac{1+\cos(\delta_y)}{2} \pm \sin(\delta_y)\sqrt{\hat{p}-\hat{p}^2} + (1-\hat{p})\frac{1-\cos(\delta_y)}{2} \\ &= \hat{p}\cos(\delta_y) \pm \sqrt{\hat{p}-\hat{p}^2}\,\sin(\delta_y) + \frac{1-\cos(\delta_y)}{2} \,. \end{aligned}$$

Using the expansions $\sin(x) = x + \mathcal{O}(x^3)$ and $\cos(x) = 1 + \mathcal{O}(x^2)$ we obtain

$$\tilde{p}_{\mathrm{L}} = \hat{p} - \sqrt{\hat{p}-\hat{p}^2}\,\delta_y + \mathcal{O}\left(\delta_y^2\right)$$

and

$$\tilde{p}_{\mathrm{U}} = \hat{p} + \sqrt{\hat{p}-\hat{p}^2}\,\delta_y + \mathcal{O}\left(\delta_y^2\right) \ .$$

Since (2.7) implies that $z(1-\alpha/2)/\sqrt{n} = \delta_y + o(\delta_y^2)$, we have

$$\tilde{p}_{\mathrm{L}} = \hat{p} - z(1-\alpha/2)\sqrt{\hat{p}-\hat{p}^2}/\sqrt{n} + \mathcal{O}\left(\delta_y^2\right)$$

and

$$\tilde{p}_{\mathrm{U}} = \hat{p} + z(1-\alpha/2)\sqrt{\hat{p}-\hat{p}^2}/\sqrt{n} + \mathcal{O}\left(\delta_y^2\right) \ .$$

Therefore,

$$\lim_{\delta_y\to 0} \Pr\{p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]\} = \Pr\{\sqrt{n}\,|\hat{p}-p|/\sqrt{\hat{p}(1-\hat{p})} \le z(1-\alpha/2)\} \ .$$

The limiting distribution of $\sqrt{n}(\hat{p}-p)/\sqrt{\hat{p}(1-\hat{p})}$ is the standard normal distribution, and $\hat{p} \to p$ a.s. as $n \to \infty$ if $0 < p < 1$. Since $\delta_y \to 0$ implies that $n \to \infty$, we have

$$\lim_{\delta_y\to 0} \Pr\{p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]\} = 1 - \alpha \,.$$

The asymptotic consistency implies that for small $\delta_y$'s, the probability of $p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$ will be close to $1-\alpha$. In practice, we are interested in small but fixed $\delta_y$'s. Table 1 gives the probabilities of $p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$ for $\delta_y = 2^{-i}$, $i = 1, \ldots, 10$, $p = 2^{-i}$, $i = 1, \ldots, 10$, and $\alpha = 0.10$, $0.05$, $0.01$. The overall conclusion of the approximation $\hat{y} \sim N(y, 1/n)$ is that the approximation does not introduce any serious error when $n \ge 16/\min(p, 1-p)$ and $k \sim \mathrm{Bin}(n, p)$.

### *2.3.2. Sequence of Dependent Indicators with Known Correlation Structure*

When the sequence $\{I_j\}$ is covariance stationary (not i.i.d.) and the autocorrelations are known, the termination rule (2.6) provides a fixed sample size (which replaces Eq. 2.7):

$$n = \left\lceil C(n)(z(1-\alpha/2)/\delta_y)^2 \right\rceil \ , \qquad (2.8)$$

where $C(n) = 1 + 2\sum_{t=1}^{n-1}(1-t/n)\rho_t$. In this case $\sqrt{C(n)}\,z(1-\alpha/2)/\sqrt{n} = \delta_y + o(\delta_y^2)$. If $0 < p < 1$ the limiting distribution of $\sqrt{n}(\hat{p}-p)/\sqrt{p(1-p)}$ is $N(0, C)$, where $C = C(\infty)$, and $\hat{p} \to p$ a.s. as $n \to \infty$, if $0 < p < 1$. Therefore the desired confidence interval is asymptotically consistent.

TABLE 1

*Probabilities of* $p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$ *when* $\hat{p} = k/n$ , $k \sim \mathrm{Bin}(n, p)$, *and* $n = \lceil (z(1-\alpha/2)/\delta_y)^2 \rceil$

| $\alpha$ | $\delta_y$ | $p$ = 1/2 | 1/4 | 1/8 | 1/16 | 1/32 | 1/64 | 1/128 | 1/256 | 1/512 | 1/1024 | $np$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| .10 | 1/2 | .9346 | .9234 | .7308 | .4808 | .9555 | .9878 | .9968 | .9992 | .9787 | .9893 | |
| | 1/4 | .9039 | .9159 | .8833 | .8868 | .7045 | .4685 | .9535 | .9871 | .9966 | .9991 | |
| | 1/8 | .8889 | .9029 | .8868 | .8769 | .8633 | .9151 | .6960 | .4625 | .9540 | .9872 | |
| | 1/16 | .9054 | .9045 | .9042 | .8992 | .8990 | .8672 | .8583 | .9130 | .6936 | .4606 | |
| | 1/32 | .8977 | .9001 | .8982 | .9007 | .8971 | .8907 | .8949 | .8649 | .8571 | .9126 | |
| | 1/64 | .8997 | .9001 | .9013 | .9005 | .9040 | .8998 | .9056 | .8886 | .8939 | .8643 | $np < 16$ |
| | 1/128 | .9007 | .9001 | .9014 | .8987 | .9013 | .9002 | .9000 | .8978 | .9046 | .8881 | $np > 16$ |
| | 1/256 | .8997 | .9001 | .8999 | .9007 | .8999 | .9003 | .9001 | .8982 | .8990 | .8973 | |
| | 1/512 | .8999 | .8998 | .8999 | .9002 | .9000 | .9004 | .9001 | .9004 | .9018 | .8977 | |
| | 1/1024 | .9001 | .9000 | .8999 | .8999 | .9000 | .8999 | .9001 | .9004 | .8991 | .8999 | |
| .05 | 1/2 | .9232 | .9094 | .8720 | .6288 | .9874 | .9982 | .9932 | .9982 | .9996 | .9999 | |
| | 1/4 | .9441 | .9429 | .9405 | .8916 | .8475 | .6071 | .9872 | .9981 | .9933 | .9983 | |
| | 1/8 | .9521 | .9531 | .9451 | .9518 | .9265 | .9630 | .8412 | .6017 | .9872 | .9981 | |
| | 1/16 | .9482 | .9532 | .9512 | .9508 | .9559 | .9462 | .9362 | .9619 | .8401 | .6010 | |
| | 1/32 | .9501 | .9511 | .9519 | .9518 | .9506 | .9450 | .9533 | .9447 | .9353 | .9616 | |
| | 1/64 | .9501 | .9501 | .9491 | .9500 | .9511 | .9497 | .9478 | .9436 | .9526 | .9546 | $np < 16$ |
| | 1/128 | .9501 | .9501 | .9498 | .9499 | .9498 | .9500 | .9484 | .9483 | .9471 | .9432 | $np > 16$ |
| | 1/256 | .9499 | .9498 | .9502 | .9500 | .9498 | .9501 | .9497 | .9486 | .9502 | .9480 | |
| | 1/512 | .9500 | .9501 | .9500 | .9500 | .9502 | .9501 | .9498 | .9506 | .9504 | .9503 | |
| | 1/1024 | .9500 | .9500 | .9500 | .9500 | .9500 | .9501 | .9501 | .9496 | .9498 | .9502 | |
| .01 | 1/2 | .9808 | .9933 | .9686 | .8238 | .9986 | .9992 | .9988 | .9998 | .9999 | .9999 | |
| | 1/4 | .9884 | .9885 | .9913 | .9891 | .9646 | .8130 | .9984 | .9991 | .9999 | .9998 | |
| | 1/8 | .9884 | .9897 | .9898 | .9867 | .9896 | .9871 | .9621 | .8089 | .9984 | .9991 | |
| | 1/16 | .9899 | .9900 | .9897 | .9907 | .9890 | .9914 | .9887 | .9866 | .9616 | .8082 | |
| | 1/32 | .9899 | .9900 | .9898 | .9901 | .9901 | .9901 | .9881 | .9910 | .9885 | .9865 | $np < 16$ |
| | 1/64 | .9899 | .9898 | .9900 | .9898 | .9901 | .9897 | .9901 | .9896 | .9908 | .9909 | $np > 16$ |
| | 1/128 | .9901 | .9901 | .9900 | .9900 | .9901 | .9901 | .9897 | .9899 | .9899 | .9895 | |
| | 1/256 | .9900 | .9900 | .9900 | .9901 | .9900 | .9899 | .9900 | .9900 | .9900 | .9898 | |
| | 1/512 | .9900 | .9900 | .9900 | .9900 | .9900 | .9900 | .9900 | .9900 | .9900 | .9899 | |
| | 1/1024 | .9900 | .9900 | .9900 | .9900 | .9900 | .9900 | .9900 | .9899 | .9900 | .9901 | |

The properties of the termination rule are now difficult to analyze when $\delta_y$ is a fixed positive constant. The reason is that the accuracy of approximation $\mathrm{Var}(\hat{y}) = C/n$ depends also on $\mathrm{E}(\hat{p}-p)^k$, $k = 3, 4, \ldots$. In addition, the higher moments affect the rate of convergence to the normal limiting distribution.

As an example of autocorrelated sequences we examined a two-stage Markov chain with $\Pr\{I_{j+1} = 1 | I_j = 0\} = a$ and $\Pr\{I_{j+1} = 0 | I_j = 1\} = d$. In this case $p = \Pr\{I_j = 1\} = a/(a+d)$ and $C = 2/(a+d) - 1$; see Glynn (1984). Tables 2 and 3 give the probabilities of $p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$ for $p = 1/2, 1/5, 1/10, 1/20, 1/50, 1/100, 1/200, 1/500, 1/1000$, $C = 2, 4, 8, 16, 32, 64$, $\alpha = 0.10, 0.05, 0.01$, $\delta_y = 0.20, 0.10, 0.05$, and $n = \lceil C(z(1-\alpha/2)/\delta_y)^2 \rceil$.

### *2.3.3. Sequence of Dependent Indicators with Unknown Correlation Structure*

When the correlation structure of $\{I_j\}$ is unknown, the properties of the termination rule (2.6) depend also on the method of estimating $\mathrm{Var}(\hat{p})$. The sequence length determined through the termination rule is a random variable that depends on $\delta_y$:

$$N(\delta_y) \equiv \text{smallest integer } n \text{ such that } t_\nu(1-\alpha/2) s_{\hat{p}} / \sqrt{\hat{p} - \hat{p}^2} \leq \delta_y \,. \tag{2.9}$$

Now the estimate $\hat{p}$, its estimated variance $s^2_{\hat{p}}$, the desired confidence interval $[\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$, and the degrees of freedom $\nu$ depend on $\delta_y$ through $N(\delta_y)$. Let the notations $\hat{p}[N(\delta_y)]$, $s^2_{\hat{p}}[N(\delta_y)]$, $\tilde{p}_{\mathrm{L}}[N(\delta_y)]$, $\tilde{p}_{\mathrm{U}}[N(\delta_y)]$, and $\nu[N(\delta_y)]$ indicate this dependency.

Suppose that the indicator sequence $\{I_j\}$ is covariance stationary and satisfies the regularity conditions so that $\sqrt{n}(\hat{p} - p)$ converges in distribution to $N(0, \gamma)$, where $\gamma \equiv \lim_{n\to\infty} n \,\mathrm{Var}(\hat{p})$, $0 < \gamma < \infty$. This convergence is for a nonrandom $n$. It is shown in Billingsley (1968, Ch. 17) that under certain conditions the convergence for a nonrandom $n$ implies that $\sqrt{N(\delta_y)}\,\big(\hat{p}[N(\delta_y)] - p\big)$ converges in distribution to $N(0, \gamma)$ as $\delta_y \to 0$. The conditions are satisfied, if we can find for any sequence of positive constants $\{\delta_y(m) : m = 1, 2, \ldots\}$ going to zero a sequence of constants $\{a_m : m = 1, 2, \ldots\}$ going to infinity such that the ratio $N[\delta_y(m)]/a_m$ converges in probability to a positive limit $\theta^2$ as $m$ goes to infinity:

$$\frac{N[\delta_y(m)]}{a_m} \underset{m\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} \theta^2 \quad \text{for some} \quad 0 < \theta^2 < \infty \,. \tag{2.10}$$

In order to establish the asymptotic coverage of the confidence interval $[\tilde{p}_{\mathrm{L}}[N(\delta_y)], \tilde{p}_{\mathrm{U}}[N(\delta_y)]]$, we need to assume that the method of estimating $\mathrm{Var}(\hat{p})$ fulfills the following assumptions. Firstly, we must assume that $ns^2_{\hat{p}} > 0$ (almost surely). Secondly, we must assume that $ns^2_{\hat{p}}$ is a consistent estimator of the variance parameter $\gamma$:

$$\lim_{n\to\infty} \Pr\{|ns^2_{\hat{p}} - \gamma| > \epsilon\} = 0 \text{ , for every } \epsilon > 0 \,. \tag{2.11}$$

TABLE 2

*Probabilities of $p \in [\tilde{p}_L, \tilde{p}_U]$ when $\hat{p} = k/n$, $p = \Pr\{state\ 1\}$,*
*$k$ = number of visits to state 1, and $n = \lceil C(z(1-\alpha/2)/\delta_y)^2 \rceil$*

| | | $\delta_y$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.20 | | | 0.10 | | | 0.05 | | |
| | | $\alpha$ | | | $\alpha$ | | | $\alpha$ | | |
| $p$ | $C$ | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 |
| 1/2 | 2 | .8994 | .9477 | .9887 | .9054 | .9476 | .9897 | .8993 | .9490 | .9899 |
| | 4 | .9001 | .9480 | .9893 | .8994 | .9507 | .9900 | .8994 | .9505 | .9900 |
| | 8 | .8969 | .9482 | .9899 | .8995 | .9499 | .9898 | .9002 | .9498 | .9900 |
| | 16 | .9003 | .9498 | .9898 | .8996 | .9500 | .9899 | .8998 | .9500 | .9900 |
| | 32 | .8987 | .9495 | .9899 | .9000 | .9500 | .9899 | .8998 | .9500 | .9900 |
| | 64 | .8988 | .9495 | .9898 | .8998 | .9499 | .9900 | .8999 | .9500 | .9900 |
| 1/5 | 2 | .8879 | .9506 | .9883 | .8977 | .9481 | .9899 | .9015 | .9501 | .9899 |
| | 4 | .8966 | .9487 | .9892 | .9017 | .9500 | .9900 | .8998 | .9502 | .9901 |
| | 8 | .8976 | .9483 | .9898 | .8998 | .9502 | .9899 | .8997 | .9497 | .9900 |
| | 16 | .8971 | .9486 | .9898 | .8996 | .9497 | .9899 | .8997 | .9499 | .9900 |
| | 32 | .8978 | .9491 | .9897 | .8997 | .9495 | .9899 | .8999 | .9499 | .9900 |
| | 64 | .8977 | .9488 | .9897 | .8995 | .9496 | .9899 | .8999 | .9499 | .9900 |
| 1/10 | 2 | .8935 | .9458 | .9875 | .8928 | .9489 | .9902 | .8996 | .9518 | .9903 |
| | 4 | .8890 | .9486 | .9890 | .8940 | .9495 | .9898 | .8976 | .9495 | .9901 |
| | 8 | .8917 | .9457 | .9893 | .9001 | .9493 | .9899 | .8989 | .9495 | .9900 |
| | 16 | .8947 | .9472 | .9894 | .8998 | .9499 | .9899 | .8996 | .9498 | .9900 |
| | 32 | .8953 | .9479 | .9895 | .8991 | .9496 | .9899 | .8995 | .9498 | .9900 |
| | 64 | .8950 | .9476 | .9895 | .8988 | .9495 | .9899 | .8997 | .9499 | .9900 |
| 1/20 | 2 | .8565 | .9255 | .9855 | .9021 | .9459 | .9898 | .9053 | .9502 | .9901 |
| | 4 | .8788 | .9375 | .9882 | .8945 | .9498 | .9901 | .9018 | .9500 | .9898 |
| | 8 | .8893 | .9408 | .9892 | .8958 | .9494 | .9898 | .9000 | .9499 | .9899 |
| | 16 | .8921 | .9434 | .9890 | .8990 | .9485 | .9898 | .9001 | .9495 | .9899 |
| | 32 | .8907 | .9441 | .9888 | .8980 | .9489 | .9898 | .8997 | .9497 | .9899 |
| | 64 | .8894 | .9445 | .9889 | .8979 | .9489 | .9898 | .8994 | .9497 | .9899 |
| 1/50 | 2 | .7998 | .9122 | .9858 | .8922 | .9421 | .9878 | .8954 | .9498 | .9894 |
| | 4 | .8523 | .9391 | .9818 | .8860 | .9494 | .9890 | .9014 | .9493 | .9896 |
| | 8 | .8784 | .9264 | .9853 | .8898 | .9482 | .9897 | .8985 | .9490 | .9899 |
| | 16 | .8657 | .9325 | .9867 | .8933 | .9475 | .9894 | .8982 | .9492 | .9898 |
| | 32 | .8592 | .9291 | .9860 | .8941 | .9471 | .9895 | .8980 | .9493 | .9898 |
| | 64 | .8559 | .9305 | .9864 | .8937 | .9466 | .9894 | .8984 | .9493 | .9899 |

TABLE 3

*Probabilities of $p \in [\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$ when $\hat{p} = k/n$, $p = \Pr\{state\ 1\}$, $k$ = number of visits to state 1, and $n = \lceil C(z(1-\alpha/2)/\delta_y)^2 \rceil$*

| | | $\delta_y$ 0.20 | | | $\delta_y$ 0.10 | | | $\delta_y$ 0.05 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\alpha$ | | | $\alpha$ | | | $\alpha$ | | |
| $p$ | $C$ | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 |
| 1/100 | 2 | .5717 | .7098 | .8895 | .8779 | .9579 | .9859 | .8865 | .9439 | .9902 |
| | 4 | .6311 | .7730 | .9275 | .8955 | .9496 | .9891 | .8959 | .9485 | .9901 |
| | 8 | .6718 | .8047 | .9448 | .8839 | .9444 | .9882 | .8976 | .9479 | .9897 |
| | 16 | .6906 | .8210 | .9529 | .8880 | .9426 | .9888 | .8968 | .9490 | .9897 |
| | 32 | .7017 | .8300 | .9569 | .8861 | .9441 | .9886 | .8964 | .9484 | .9898 |
| | 64 | .7073 | .8344 | .9588 | .8851 | .9434 | .9887 | .8970 | .9485 | .9898 |
| 1/200 | 2 | .9647 | .9866 | .9973 | .7955 | .9095 | .9852 | .8895 | .9470 | .9872 |
| | 4 | .9699 | .9880 | .9974 | .8490 | .9372 | .9811 | .8832 | .9511 | .9890 |
| | 8 | .9724 | .9866 | .9971 | .8755 | .9243 | .9851 | .8906 | .9481 | .9894 |
| | 16 | .9737 | .9859 | .9970 | .8627 | .9318 | .9866 | .8924 | .9465 | .9896 |
| | 32 | .9730 | .9856 | .9970 | .8574 | .9283 | .9859 | .8933 | .9474 | .9895 |
| | 64 | .9733 | .9857 | .9970 | .8547 | .9298 | .9862 | .8941 | .9472 | .9894 |
| 1/500 | 2 | .9749 | .9938 | .9972 | .9640 | .9817 | .9965 | .9083 | .9165 | .9806 |
| | 4 | .9793 | .9913 | .9973 | .9685 | .9835 | .9966 | .8924 | .9354 | .9857 |
| | 8 | .9812 | .9878 | .9975 | .9709 | .9845 | .9968 | .8779 | .9436 | .9877 |
| | 16 | .9801 | .9883 | .9976 | .9695 | .9851 | .9966 | .8836 | .9410 | .9883 |
| | 32 | .9807 | .9886 | .9975 | .9701 | .9841 | .9967 | .8798 | .9392 | .9880 |
| | 64 | .9804 | .9888 | .9975 | .9698 | .9843 | .9967 | .8812 | .9401 | .9881 |
| 1/1000 | 2 | .9887 | .9938 | .9981 | .9746 | .9913 | .9970 | .7223 | .8518 | .9686 |
| | 4 | .9848 | .9913 | .9981 | .9781 | .9883 | .9971 | .7859 | .8966 | .9827 |
| | 8 | .9864 | .9921 | .9978 | .9755 | .9868 | .9973 | .8191 | .9179 | .9877 |
| | 16 | .9856 | .9916 | .9979 | .9741 | .9861 | .9971 | .8359 | .9081 | .9858 |
| | 32 | .9860 | .9913 | .9980 | .9748 | .9864 | .9972 | .8275 | .9133 | .9847 |
| | 64 | .9858 | .9914 | .9980 | .9751 | .9866 | .9972 | .8239 | .9158 | .9841 |

Thirdly, we must assume that there is a (possibly infinite) constant $\nu(\infty)$ to which the degrees of freedom $\nu(n)$ for the estimator $ns_{\hat{p}}^2$ converge in probability as $n \to \infty$:

$$\nu(n) \underset{n\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} \nu(\infty) \le \infty \ . \tag{2.12}$$

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

**Lemma 1.** Let $\{z_m : m = 1, 2, \ldots\}$ be a sequence of positive random variables that converge in probability to a positive limit $\theta^2$,

$$z_m \underset{m\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} \theta^2 \ , \text{ where } \ 0 < \theta < \infty \ . \tag{2.13}$$

Let $\{f(m) : m = 1, 2, \ldots\}$ be any sequence of constants such that

$$f(m) > 0 \ , \quad \lim_{m\to\infty} f(m) = \infty \ , \quad \lim_{m\to\infty} f(m)/f(m-1) = 1 \ . \tag{2.14}$$

For each $\tau > 0$, define

$$M(\tau) \equiv \text{smallest } m \ge 1 \text{ such that } z_m \le f(m)/\tau \ . \tag{2.15}$$

Then $M(\tau)$ is well defined and nondecreasing as a function of $\tau$ such that

$$M(\tau) \underset{\tau\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} \infty \quad \text{and} \qquad f[M(\tau)]/\tau \underset{\tau\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} \theta^2 \ . \tag{2.16}$$

**Proof of Lemma 1.** Proving that $M(\tau)$ converges in probability to infinity is a straightforward application of the definition of $M(\tau)$ and the notion of convergence in probability. To prove the second part of (2.16), we observe that for $M(\tau) > 1$,

$$z_{M(\tau)} \le f[M(\tau)]/\tau < \{f[M(\tau)]/f[M(\tau)-1]\} z_{M(\tau)-1} \ . \tag{2.17}$$

Now the ratio $f[M(\tau)]/\tau$ has a stochastic lower and upper bound. The first part of (2.16), (2.13), and (2.14) together with Application D in Serfling (1980, p. 26) imply that the bounds converge in probability to $\theta^2$. Hence, $f[M(\tau)]/\tau$ converges in probability to $\theta^2$. $\Box$

Set

$$\left.\begin{aligned} z_m &= \left(\frac{t_{\nu(m)}(1-\alpha/2)}{t_m(1-\alpha/2)}\right)^2 \frac{ms_{\hat{p}}^2}{\gamma} \Bigg/ \frac{\hat{p}-\hat{p}^2}{p-p^2} \\ f(m) &= m\left(\frac{z(1-\alpha/2)}{t_m(1-\alpha/2)}\right)^2 \\ \tau_m &= \frac{[z(1-\alpha/2)]^2\gamma}{[\delta_y(m)]^2(p-p^2)} \end{aligned}\right\} . \tag{2.18}$$

It is easy to verify that

$$M(\tau_m) = N[\delta_y(m)] \quad \text{for} \quad m = 1, 2, \ldots \tag{2.19}$$

and that the assumptions of Lemma 1 are satisfied with

$$\theta = t_{\nu(\infty)}(1-\alpha/2)/z(1-\alpha/2) \geq 1 \; . \tag{2.20}$$

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

For a given $\{\delta_y(m) : m = 1, 2, \ldots\}$ such that $\lim_{m\to\infty} \delta_y(m) = 0$, we define

$$a_m \equiv \frac{[t_m(1-\alpha/2)]^2\gamma}{[\delta_y(m)]^2(p-p^2)} \quad \text{for} \quad m = 1, 2, \ldots \tag{2.21}$$

so that $\lim_{m\to\infty} a_m = \infty$.

Now we can show that the ratio in (2.10) converges in probability to a positive limit:

$$\begin{aligned}
\frac{N[\delta_y(m)]}{a_m} &= \frac{N[\delta_y(m)]}{[t_m(1-\alpha/2)]^2\gamma/\{[\delta_y(m)]^2(p-p^2)\}} \qquad \textit{by } (2.21) && (2.22)\\
&= \frac{f[N(\delta_y(m))]}{\tau_m} \qquad \textit{by } (2.18) && (2.23)\\
&= \frac{f[M(\tau_m)]}{\tau_m} \qquad \textit{by } (2.19) && (2.24)\\
&\xrightarrow[m\to\infty]{\mathcal{P}} \theta^2 \qquad \textit{by } (2.16)\,. && (2.25)
\end{aligned}$$

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

$$\sqrt{N(\delta_y)}\left(\hat{p}[N(\delta_y)] - p\right) \xrightarrow[\delta_y\to 0]{\mathcal{D}} N(0,\gamma)\,. \tag{2.26}$$

Now we can establish the asymptotic coverage of the desired confidence interval. From Section 2.3.1 we recall that

$$\tilde{p}_{\mathrm{L}} = \hat{p}\cos(\delta_y) - \sqrt{\hat{p}-\hat{p}^2}\,\sin(\delta_y) + [1-\cos(\delta_y)]/2\,,$$

when $[\sin(\delta_y/2)]^2 \leq \hat{p} \leq 1 - [\sin(\delta_y/2)]^2$. Since $\lim_{\delta_y\to 0}\cos(\delta_y) = 1$ and $\lim_{\delta_y\to 0}\sin(\delta_y)/\delta_y = 1$, we have

$$\left|\tilde{p}_{\mathrm{L}} - \left(\hat{p} - \delta_y\sqrt{\hat{p}-\hat{p}^2}\right)\right| \xrightarrow[\delta_y\to 0]{\mathcal{P}} 0\,.$$

Therefore, we can use Theorem 4.1 of Billingsley (1968) to obtain

$$\begin{aligned}
&\lim_{\delta_y\to 0} \Pr\{p < \tilde{p}_{\mathrm{L}}[N(\delta_y)]\} && (2.27)\\
&= \lim_{\delta_y\to 0} \Pr\left\{p < \hat{p} - \delta_y\sqrt{\hat{p}-\hat{p}^2}\right\} && (2.28)\\
&= \lim_{\delta_y\to 0} \Pr\left\{\delta_y\sqrt{(\hat{p}-\hat{p}^2)N(\delta_y)} < \sqrt{N(\delta_y)}\left[\hat{p}-p\right]\right\}\,. && (2.29)
\end{aligned}$$

Notice that in (2.28) and (2.29) we have suppressed the dependence of $\hat{p}$ on $N(\delta_y)$ to simplify the expressions on the right-hand sides. Next we observe that (2.25) implies that

$$\delta_y \sqrt{(\hat{p} - \hat{p}^2) N(\delta_y)} \underset{\delta_y \to 0}{\xrightarrow{\mathcal{P}}} t_{\nu(\infty)}(1-\alpha/2)\sqrt{\gamma}\ , \tag{2.30}$$

where $\nu(\infty)$ is defined in (2.12). Inserting (2.30) into (2.29) and applying (2.26) together with (2.20) and Pólya's Theorem (Bickel and Doksum 1977, p. 462), we obtain

$$\begin{aligned} \lim_{\delta_y \to 0} \Pr\{p < \tilde{p}_{\mathrm{L}}[N(\delta_y)]\} &= \lim_{\delta_y \to 0} \Pr\left\{t_{\nu(\infty)}(1-\alpha/2)\sqrt{\gamma} < N(0,1)\sqrt{\gamma}\right\} & (2.31)\\ &\leq \Pr\left\{z(1-\alpha/2) < N(0,1)\right\} & (2.32)\\ &= \alpha/2\ . & (2.33) \end{aligned}$$

A similar argument shows that

$$\lim_{\delta_y \to 0} \Pr\{\tilde{p}_{\mathrm{U}}[N(\delta_y)] < p\} \leq \alpha/2\ . \tag{2.34}$$

Combining (2.33) and (2.34), we obtain the desired result:

$$\lim_{\delta_y \to 0} \Pr\left\{p \in [\tilde{p}_{\mathrm{L}}[N(\delta_y)], \tilde{p}_{\mathrm{U}}[N(\delta_y)]]\right\} \geq 1 - \alpha\ . \tag{2.35}$$

In other words the asymptotic coverage of the desired confidence interval is at least the requested nominal level. The interval is asymptotically consistent only if $\theta = 1$, i.e. $\nu(\infty) = \infty$.

The overall conclusion from the situations examined in Section 2.3 is that the termination rule (2.6) provides estimates of $p$ that satisfy the required precision if $\delta_y / t_\nu(1-\alpha/2)$ is small enough. Primarily $p$ but also to some extent the autocorrelations determine how small $\delta_y / t_\nu(1-\alpha/2)$ should be. The smaller $p$ is, the smaller $\delta_y / t_\nu(1-\alpha/2)$ must be.

## *2.4. Small Proportions*

The termination rule (2.6) cannot be used if $\Pr\{X \in \mathcal{C}\}$ is very close to 0 or 1. In such a case it will be highly probable that $\hat{p}(1-\hat{p}) = 0$, which implies, as (2.6) shows, that the simulation should be continued until $s_{\hat{p}} = 0$. Moreover, the sequence $\{0, \ldots, 0\}$ or $\{1, \ldots, 1\}$ does not provide any information about $\mathrm{Var}(\hat{p})$.

When we have observed the sequence $I_j = 0\,,\ j = 1, \ldots, n$ and we assume that the $I_j$'s are independent, then the confidence interval for $p$ can be based on the fact that

$$\Pr\{I_1 = 0 \ \&\ \cdots\ \&\ I_n = 0\} = (1-p)^n\ . \tag{2.36}$$

The $100(1-\alpha)$ percent confidence interval for $p$ is $[0, \hat{p}_{\mathrm{U}}]$, where $\hat{p}_{\mathrm{U}} = 1 - \sqrt[n]{\alpha}$: see e.g. Bickel and Doksum (1977, p. 180); also 1) Brenner and Quan (1990) and 2) Louis (1981).

When the $I_j$'s are autocorrelated, Eq. (2.36) does not hold. Instead, we have

$$\Pr\{I_1 = 0 \;\&\; \cdots \;\&\; I_n = 0\} = \prod_{j=1}^{n}(1 - A_j p)$$

where $A_1 = 1$ and $A_j p = \mathrm{E}(I_j | I_1 = \ldots = I_{j-1} = 0)$, $j = 2, 3, \ldots$. Hence, the $\hat{p}_{\mathrm{U}}$ in the $100(1-\alpha)$ percent confidence interval, $[0, \hat{p}_{\mathrm{U}}]$, is the solution of $\prod_{j=1}^{n}(1 - A_j \hat{p}_{\mathrm{U}}) - \alpha = 0$.

Since in steady-state simulation it is dangerous to assume that the observations are independent, we should learn the $A_j$'s that characterize the dependencies between the $I_j$'s. However, the sequence $\{0, \ldots, 0\}$ does not contain any information about the $A_j$'s. Therefore, we propose a practical treatment of small proportions.

We propose that the analyst specify a tolerance of small proportion $p_{\mathrm{small}}$. If $\hat{p} < p_{\mathrm{small}}$ or $\hat{p} > 1 - p_{\mathrm{small}}$, then the simulation is terminated without estimating the confidence interval if $np_{\mathrm{small}} \geq 16$, whereas the simulation is continued if $np_{\mathrm{small}} < 16$. The constant 16 is based on Section 2.3. The objective of this treatment is to obtain reliable estimates for all $p \in [p_{\mathrm{small}}, 1 - p_{\mathrm{small}}]$. It should be noted that this treatment does not take the dependencies into account. As discussed above, a sequence that contains a few 1's (or a few 0's) does not provide sufficient information to take the possible dependencies into account.

If the analyst is interested in rare events, say $\Pr\{X \in \mathcal{C}\} \ll 0.01$, importance sampling should be used. For example, *Management Science 35, 11* (Nov. 1989) is a special issue devoted to importance sampling; see also Heidelberger (1992).

## 3. Estimating Several Proportions Simultaneously

In many applications a single proportion is not sufficient. For example, when approximating the distribution of a response variable by a histogram, the analyst typically wants to estimate from 5 to 25 proportions simultaneously.

Suppose we are interested in estimating the probabilities $p_1, \ldots, p_m$ with $p_k = \Pr\{X \in \mathcal{C}_k\}$ and prespecified (not necessarily disjoint nor exhaustive) categories $\mathcal{C}_1, \ldots, \mathcal{C}_m$ , $k = 1, \ldots, m$ , $m \geq 2$. Each estimate is based on a transformation of the output sequence $\{X_j\}_{j=1}^{n}$, strictly analogous to (2.1):

$$\hat{p}_k = \frac{1}{n}\sum_{j=1}^{n} I_{k,j} \;, \tag{3.1}$$

where $I_{k,j} = 1$ if $X_j \in \mathcal{C}_k$, and $I_{k,j} = 0$ if $X_j \notin \mathcal{C}_k$.

### *3.1. Precision Requirement and Termination Rule*

We specify the precision requirement for the estimated proportions as in the case of a single proportion, by giving the parameters $\delta_y$ and $\alpha$. Instead of a common $\delta_y$, we might specify a separate $\delta_{y_k}$ for each $p_k$. However, we believe that in most cases a common $\delta_y$ is appropriate since the arcsin transform adjusts the desired half-widths of the confidence intervals; see Figure 1.

The difference with the one-dimensional case is that instead of a single confidence interval, $[\tilde{p}_{\mathrm{L}}, \tilde{p}_{\mathrm{U}}]$, we want several confidence intervals, $[\tilde{p}_{\mathrm{L}_k}, \tilde{p}_{\mathrm{U}_k}]$, $k = 1, \ldots, m$, such that

$$\left.\begin{array}{l}
\tilde{p}_{\mathrm{L}_k} = \left\{\sin(\max(0, \sin^{-1}\left(\sqrt{\hat{p}_k}\right) - \delta_y/2))\right\}^2 \\
\tilde{p}_{\mathrm{U}_k} = \left\{\sin(\min(\pi/2, \sin^{-1}\left(\sqrt{\hat{p}_k}\right) + \delta_y/2))\right\}^2 \\
\Pr\{p_1 \in [\tilde{p}_{\mathrm{L}_1}, \tilde{p}_{\mathrm{U}_1}] \;\&\; \ldots \;\&\; p_m \in [\tilde{p}_{\mathrm{L}_m}, \tilde{p}_{\mathrm{U}_m}]\} \geq 1 - \alpha
\end{array}\right\} . \tag{3.2}$$

In other words, we are interested in a rectangular $100(1-\alpha)\%$ confidence area.

The termination rule is very similar to the one-dimensional case. The simulation will be continued until each desired confidence interval $[\tilde{p}_{\mathrm{L}_k}, \tilde{p}_{\mathrm{U}_k}]$ includes the corresponding estimated confidence interval $[\hat{p}_{\mathrm{L}_k}, \hat{p}_{\mathrm{U}_k}]$, that is $\tilde{p}_{\mathrm{L}_k} \leq \hat{p}_{\mathrm{L}_k} < \hat{p}_{\mathrm{U}_k} \leq \tilde{p}_{\mathrm{U}_k}$, $k = 1, \ldots, m$, where

$$\left.\begin{array}{lcl}
\hat{p}_{\mathrm{L}_k} &=& \{\sin(\max(0, \hat{y}_{\mathrm{L}_k}/2))\}^2 \\
\hat{p}_{\mathrm{U}_k} &=& \{\sin(\min(\pi/2, \hat{y}_{\mathrm{U}_k}/2))\}^2 \\
\hat{y}_{\mathrm{L}_k} &=& 2\sin^{-1}\left(\sqrt{\hat{p}_k}\right) - t_{\nu_k}(1-\alpha_k/2)s_{\hat{p}_k}\Big/\sqrt{\hat{p}_k - \hat{p}_k^2} \\
\hat{y}_{\mathrm{U}_k} &=& 2\sin^{-1}\left(\sqrt{\hat{p}_k}\right) + t_{\nu_k}(1-\alpha_k/2)s_{\hat{p}_k}\Big/\sqrt{\hat{p}_k - \hat{p}_k^2}
\end{array}\right\} , \tag{3.3}$$

and $s^2_{\hat{p}_k}$ is the estimated variance of $\hat{p}_k$ with $\nu_k$ degrees of freedom. The $\alpha_k$'s are arbitrary positive constants such that $\sum_{k=1}^{m} \alpha_k = \alpha$ (Bonferroni inequality).

The most common way is to choose $\alpha_k = \alpha/m$, $k = 1, \ldots, m$. Hence, the simulation is continued until

$$s_{\hat{p}_k} \leq \delta_y \sqrt{\hat{p}_k(1-\hat{p}_k)} \Big/ t_{\nu_k}(1-\alpha/(2m)) \,, \text{ for every } k = 1, \ldots, m \,. \tag{3.4}$$

The termination rule (3.4) usually leads to simulation runs which are unnecessarily long to fulfil the requirement (3.2). The reason is that only in a special case do we have $\mathrm{Var}(\hat{p}_1)/(p_1 - p_1^2) = \ldots = \mathrm{Var}(\hat{p}_m)/(p_m - p_m^2)$. The special case arises only when the binary sequences $\{I_{k,j}\}$ have the same sum of autocorrelations:

$$\sum_{t=-\infty}^{\infty} \mathrm{Corr}(I_{k,j}, I_{k,j+t}) = C \,, \text{ for every } k = 1, \ldots, m \,.$$

Usually only a few proportions dominate the run length. Therefore, we propose an alternative way of applying the Bonferroni inequality. Instead of estimating the $100(1-\alpha/m)\%$ confidence intervals, we estimate the confidence levels of the desired confidence intervals $[\tilde{p}_{L_k}, \tilde{p}_{U_k}]$ or $\hat{\alpha}_k = 1 - \widehat{\Pr}\{p_k \in [\tilde{p}_{L_k}, \tilde{p}_{U_k}]\}$; see (3.2). When we use the Student $t$-distribution to approximate the distribution of $(\hat{y}_k - y_k)/s_{\hat{y}_k}$, we obtain

$$\hat{\alpha}_k = 2F_{\nu_k}\left(-\delta_y \sqrt{\hat{p}_k(1-\hat{p}_k)} \Big/ s_{\hat{p}_k}\right) \tag{3.5}$$

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

$$\sum_{k=1}^{m} \hat{\alpha}_k \le \alpha \tag{3.6}$$

where the $\hat{\alpha}_k$'s are computed through (3.5).

### *3.2. Properties of the Termination Rule*

The properties of the termination rule (3.6) depend on three different factors. Two of them have already been discussed in Section 2.3: the method for estimating $\mathrm{Var}(\hat{p}_k)$ and the approximation $\hat{y}_k \sim N(y_k, C_k/n)$. In addition, the proposed way of applying the Bonferroni inequality affects the properties of the termination rule.

In Raatikainen (1992) we have examined the sequentially constructed simultaneous confidence intervals of fixed widths for several *means* $\theta_k$, $k = 1, \ldots, m$. The desired confidence interval for the mean $\theta_k$ is $[\hat{\theta}_k - \delta_k, \hat{\theta}_k + \delta_k]$, where $\hat{\theta}_k$ is the estimated mean and $\delta_k$ is the desired half-width specified by the analyst. We have shown that the termination rule

$$n = \text{smallest integer such that } \sum_{k=1}^{m} F_{\nu_k}(-\delta_k / s_{\hat{\theta}_k}) \le \alpha/2 \tag{3.7}$$

is asymptotically the most efficient way of applying the Bonferroni inequality so that the asymptotic probability of simultaneous coverages is at least $1 - \alpha$:

$$\lim_{\delta_1,\ldots,\delta_m \to 0} \Pr\{\theta_k \in [\hat{\theta}_k - \delta_k, \hat{\theta}_k + \delta_k]\,, \text{ for every } k = 1, \ldots, m\} \ge 1 - \alpha\ .$$

In (3.7) $s^2_{\hat{\theta}_k}$ is the estimated variance of $\hat{\theta}_k$ with $\nu_k$ degrees of freedom. The results given in Raatikainen (1992) indicate that the termination rule (3.7) is appropriate even for small positive $\delta_k$'s.

Combining the results given in Raatikainen (1992) and in Section 2.3, we expect that the termination rule (3.6) is appropriate if the joint effects of the approximation and the way of applying the Bonferroni inequality do not cause serious errors. Therefore, we can concentrate on the case of $\hat{p}_k = n_k/n$, where the $n_k$'s have a multinomial distribution. In this case the $\mathrm{Var}(\hat{p}_k)$'s are known and we obtain a fixed sample size $n$:

$$n = \text{ smallest integer such that } \sum_{k=1}^{m} \Phi(-\delta_y \sqrt{n}\,) \le \alpha/2\ , \tag{3.8}$$

TABLE 4

*Probabilities of* $\{p_k \in [\tilde{p}_{\mathrm{L}_k}, \tilde{p}_{\mathrm{U}_k}]$ *for every* $k = 1, \ldots, m\}$*, where* $p_k = 1/m$, $k = 1, \ldots, m$ *and* $n\hat{p}_k$*'s have a multinomial distribution with* $n$

| | $\delta_y$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | .20 | | | .10 | | | .05 | | | |
| | $\alpha$ | | | $\alpha$ | | | $\alpha$ | | | |
| $m$ | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 | $np_m$ |
| 2 | .9583 | .9745 | .9957 | .9474 | .9743 | .9951 | .9475 | .9758 | .9948 | |
| 3 | .9263 | .9559 | .9894 | .9086 | .9559 | .9904 | .9186 | .9559 | .9904 | |
| 4 | .8927 | .9518 | .9910 | .9156 | .9538 | .9900 | .9109 | .9536 | .9901 | |
| 5 | .8880 | .9401 | .9885 | .9050 | .9514 | .9895 | .9110 | .9524 | .9896 | |
| 6* | *.8915* | *.9478* | *.9897* | *.8969* | *.9520* | *.9903* | *.9048* | *.9530* | *.9905* | |
| 7* | *.8886* | *.9459* | *.9878* | *.9075* | *.9527* | *.9898* | *.9056* | *.9508* | *.9903* | |
| 8* | *.8818* | *.9482* | *.9870* | *.9022* | *.9463* | *.9902* | *.9098* | *.9519* | *.9902* | |
| 9* | *.8745* | *.9497* | *.9838* | *.9114* | *.9556* | *.9904* | *.9043* | *.9530* | *.9902* | |
| 10* | *.8492* | *.9471* | *.9836* | *.9080* | *.9490* | *.9875* | *.9083* | *.9525* | *.9899* | $np_m > 16$ |

* results based on Monte-Carlo simulations; 1 000 000 replications

or $n = \left\lceil \{z(1-\alpha/(2m))/\delta_y\}^2 \right\rceil$. In (3.8) $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

We examine the probability of $\{p_k \in [\tilde{p}_{\mathrm{L}_k}, \tilde{p}_{\mathrm{U}_k}]$, for every $k = 1, \ldots, m\}$ with two sets of the $p_k$'s:

1) $p_k = 1/m$, $k = 1, \ldots, m$
2) $p_k = 2^{-k}$, $k = 1, \ldots, m-1$ and $p_m = p_{m-1}$.

These probabilities are given in Tables 4 and 5 when $m = 2, \ldots, 10$, $\delta_y = 0.20$, 0.10, 0.05, and $\alpha = 0.10$, 0.05, 0.01. The results are obtained analytically for $m = 2, \ldots, 5$ and from Monte Carlo simulations for $m = 6, \ldots, 10$. In the Monte Carlo simulations 1 000 000 observations were generated from the multinomial distributions. Therefore, the probability that the first three digits are correct is more than 95%. Based on these results and on those discussed in Section 2.3, we conclude that the termination rule (3.6) is appropriate provided that $\delta_y$ is small enough.

### *3.3. Small Proportion*

To overcome the problems due to small estimated proportions, $\widehat{\Pr}\{X \in \mathcal{C}_k\}$, we make a practical suggestion similar to that in Section 2.4. If $\hat{p}_k < p_{\mathrm{small}}$ or $\hat{p}_k > 1 - p_{\mathrm{small}}$, we do not use (3.5) but we set

$$\hat{\alpha}_k = \begin{cases} 0 & \text{if } np_{\mathrm{small}} \geq 16 , \\ 1 & \text{if } np_{\mathrm{small}} < 16 . \end{cases}$$

TABLE 5

*Probabilities of* $\{p_k \in [\tilde{p}_{L_k}, \tilde{p}_{U_k}]$ *for every* $k = 1, \ldots, m\}$ *where*
$p_k = 2^{-k}$, $k = 1, \ldots, m-1$, $p_m = p_{m-1}$
*and* $n\hat{p}_k$*'s have a multinomial distribution with* $n$

| | $\delta_y$ .20 | | | $\delta_y$ .10 | | | $\delta_y$ .05 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | | | $\alpha$ | | | $\alpha$ | | | |
| $m$ | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 | $np_m$ |
| 2 | .9583 | .9745 | .9957 | .9474 | .9743 | .9951 | .9475 | .9758 | .9948 | |
| 3 | .9078 | .9570 | .9902 | .9164 | .9568 | .9905 | .9154 | .9554 | .9906 | |
| 4 | .8957 | .9473 | .9894 | .9093 | .9539 | .9908 | .9145 | .9553 | .9902 | |
| 5 | .8825 | .9457 | .9887 | .9176 | .9528 | .9901 | .9073 | .9520 | .9903 | |
| 6* | *.9046* | *.9079* | *.9857* | *.9025* | *.9472* | *.9877* | *.9030* | *.9522* | *.9901* | |
| 7* | *.7594* | *.8342* | *.9544* | *.8966* | *.9401* | *.9886* | *.8992* | *.9482* | *.9897* | |
| 8* | *.7782* | *.8534* | *.9579* | *.8984* | *.9176* | *.9868* | *.8963* | *.9451* | *.9899* | |
| 9* | *.6286* | *.7115* | *.8552* | *.7798* | *.8515* | *.9584* | *.9008* | *.9437* | *.9895* | $np_m > 16$ |
| 10* | *.4059* | *.4924* | *.6397* | *.7965* | *.8666* | *.9615* | *.9076* | *.9290* | *.9861* | $np_m < 16$ |

* results based on Monte-Carlo simulations; 1 000 000 replications

By setting $\hat{\alpha}_k = 1$ in (3.6) we force the simulation to be continued so that reliable estimates for all $p_k \in [p_{\mathbf{small}}, 1 - p_{\mathbf{small}}]$ may be obtained.

## 4. Example Applications

We empirically examined the method proposed. Three simulation experiments were executed. In the first one we estimated the probabilities of the queue lengths at arrival instants in the M/M/1 queue. In the two other experiments we estimated a histogram of the system response time in a closed queueing network model. We also estimated the corresponding means: the mean queue length at arrival instants in the M/M/1 queue and the mean system response time in the closed queueing network model. The objective of also estimating the means was to learn how hard a problem the simultaneous estimation of several proportions is when compared to the estimation of a single mean.

Each experiment consisted of 1,000 independent simulation runs. The parameters controlling the precision were $\alpha = 0.1$, $\delta_y = \sin^{-1}(0.1)$ implying that at $\hat{p} = 0.5$ the desired confidence interval is $[0.45, 0.55]$. The tolerance of small probability was $p_{\mathbf{small}} = 0.001$. When the means were estimated, the parameters controlling the precision were $\alpha = 0.1$ and $\varepsilon = 0.1$.

We use the observed coverage and the run length to characterize the performance of the proposed method. The observed coverage is the fraction of simulation runs, in which the true probability $p_k$ lies within the desired confidence interval $[\tilde{p}_{\mathrm{L}_k}, \tilde{p}_{\mathrm{U}_k}]$ for every $k = 1, \ldots, m$. The run length is the sample size determined by the termination rule (3.6).

### *4.1. Method for Estimating Variances*

The variances were estimated using the spectral method introduced in Heidelberger and Welch (1981). The method maintains batch means of the output sequences. When proportions are estimated, the batch means of the indicator sequences are maintained. The method has six parameters: $M$, $d$, $K$, $j_1$, $j_{\max}$, and $J$ explained below.

Parameter $M$ controls the number of batches, which varies from $M$ to $2M$. When the number of batches grows to $2M$, the batch size is doubled and the number of batches is reduced to $M$ by averaging two adjacent batch means. A polynomial of order $d$ is fitted to the first $K$ ($K < M/4$) ordinates of the bias corrected logarithms of the smoothed periodogram. Periodogram ordinates are evaluated using Fast Fourier transforms and smoothing is done by averaging two adjacent ordinates. The variance is estimated at prespecified checkpoints $j_1, j_2, \ldots, j_{\max}$; $j_{i+1} = J\, j_i$. We have used the following values: $M = 512$, $d = 2$, $K = 31$, $j_1 = 512$, $j_{\max} = 8\,388\,608$, and $J = 2$. The degrees of freedom in the variance estimate are 9. For details, see Heidelberger and Welch (1981), or Pawlikowski (1990).

### *4.2. M/M/1 queue*

The traffic intensity of the M/M/1 queue examined was $\rho = 0.8$. We estimated 21 probabilities of the queue lengths that an arriving customer sees ("arrival queue lengths"): $\Pr\{Q = k\}$, $k = 0, \ldots, 19$ and $\Pr\{Q > 19\}$. The simulation runs were initialized using the known distribution of the queue lengths at arrival instants, $\Pr\{Q = k\} = (1 - \rho)\rho^k$.

The results are summarized in Figure 3, which provides a summary of run lengths and boxplots of estimated probabilities. The summary of run lengths (upper right corner) includes mean, standard deviation, and fractions of different run lengths. In the boxplot a bullet ‘ • ’ denotes the estimated mean of estimates. The three estimated quartiles are visualized by a box ‘ ’. The whiskers ‘ ’ reach from the 25th to the 10th estimated percentile and from the 75th to the 90th estimated percentile. The 1st, 5th, 95th, and 99th estimated percentile together with the smallest and largest estimate are marked by ticks ‘ | ’. The dashed boxes denotes the values of $\hat{p}_k$ for which $p_k \in [\tilde{p}_{\mathrm{L}_k}, \tilde{p}_{\mathrm{U}_k}]$. The exact value of $p_k = \Pr\{Q = k\}$ is $(1 - \rho)\rho^k$, $k = 0, \ldots, 19$, and $p_{20} = \Pr\{Q > 19\} = \rho^{20}$.

The observed coverage is 85.2% (852/1000) if we count all experiments, or 86.7 % (852/983) if we exclude the estimates that are less than $p_{\mathrm{small}}$. These coverages are less than the requested

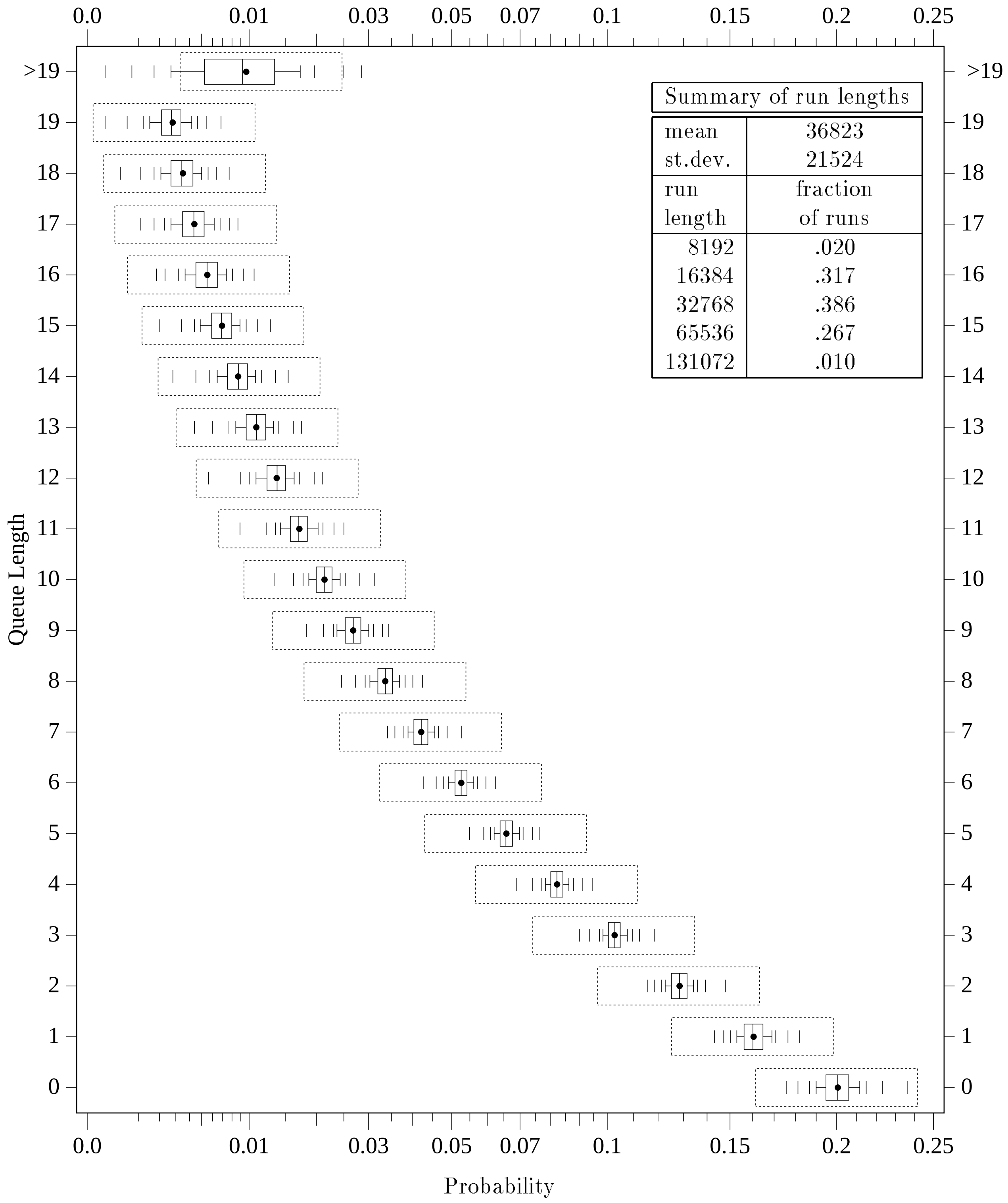
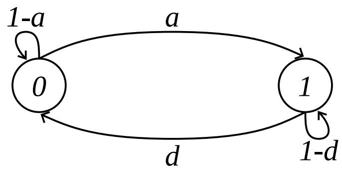



FIGURE 3: Boxplots of estimated probabilities of queue lengths at arrival instants in M/M/1 queue with traffic intensity $\rho = 0.8$.

Requested precision: 90% of probabilities simultaneously inside the dashed boxes ($\alpha = 0.1,\ \delta_y = \sin^{-1}(0.1),\ p_{\text{small}} = 0.001$)




| State 1 | $a$ | $d$ |
|---|---|---|
| $Q = 0$ | $(1-\rho)/(1+\rho)$ | $\rho/(1+\rho)$ |
| $Q = k > 0$ | $d \times \rho^k(1-\rho))/(1-\rho^k(1-\rho))$ | $1-\rho/(1+\rho)^2$ |
| $Q \geq k > 0$ | $d \times \rho^k/(1-\rho^k)$ | $(1-\rho)/(1+\rho)$ |

FIGURE 4: Two-Stage Markov-Chain for Arrival Queue Lengths in M/M/1 queue

TABLE 6
*Factors $C = \lim_{n\to\infty} n\,\mathrm{Var}(\hat{p})/(p-p^2)$ for the estimated proportions of arrival queue lengths in M/M/1 queue with $\rho = 0.8$*

| State 1 | $C$ | State 1 | $C$ | State 1 | $C$ |
|---|---|---|---|---|---|
| $Q = 0$ | 2.6000 | $Q = 7$ | 1.5443 | $Q = 14$ | 1.6324 |
| $Q = 1$ | 1.2308 | $Q = 8$ | 1.5666 | $Q = 15$ | 1.6370 |
| $Q = 2$ | 1.3158 | $Q = 9$ | 1.5844 | $Q = 16$ | 1.6408 |
| $Q = 3$ | 1.3838 | $Q = 10$ | 1.5987 | $Q = 17$ | 1.6438 |
| $Q = 4$ | 1.4382 | $Q = 11$ | 1.6101 | $Q = 18$ | 1.6462 |
| $Q = 5$ | 1.4817 | $Q = 12$ | 1.6192 | $Q = 19$ | 1.6481 |
| $Q = 6$ | 1.5165 | $Q = 13$ | 1.6265 | $Q > 19$ | 16.7925 |

coverage of 90%. In addition, the departure from the nominal level is statistically significant (the $p$-value less than $10^{-5}$). All of the 148 estimated probabilities for which the true probability did not fall into the desired intervals are estimates of $p_{20}$.

In order to learn reasons for the low observed coverage, we examined the arrival queue lengths in the M/M/1 queue. Since the distributions of interarrival times and service times are independent and memoryless, we can construct a two-stage Markov-chain corresponding with an arrival queue length. Let 'State 1' be the arrival queue length of interest. When we are estimating $p = \Pr\{\text{State 1}\}$, then $\lim_{n\to\infty} n\,\mathrm{Var}(\hat{p})/(p-p^2) = C = 2/(a+d) - 1$. The Markov chain and its parameters are shown in Figure 4.

Table 6 shows the factors $C = \lim_{n\to\infty} n\,\mathrm{Var}(\hat{p})/(p-p^2)$ for the proportions estimated. The effects of the factor are discussed in Section 2.3. Since the factor $C$ when 'State 1' is $\{Q \geq 20\}$ is

TABLE 7
*Fractions of runs in which* $p \in [\tilde{p}_L, \tilde{p}_U]$
*M/M/1 queue with* $\rho = 0.8$; $p = \Pr\{\textit{arrival queue length} > 19\}$

| $\delta_y$ | $\alpha$ 0.10 | 0.05 | 0.01 |
|---|---|---|---|
| 0.10 | .760 | .858 | .933 |
| 0.05 | .897 | .962 | .996 |

much larger than all the others, the estimation of $p_{20}$ dominates the run length. Therefore, it is not surprising that all the estimates found outside the dashed boxes in Figure 3 were estimates of $p_{20}$.

In order to find the roots of the problems in estimating $p_{20}$, we conducted a further simulation study. In this study we only estimated the probability that the queue length at an arrival instance is greater than 19. The simulation was repeated 1 000 times with $\delta_y = 0.10\,,\ 0.05$ and $\alpha = 0.10\,,\ 0.05\,,\ 0.01$. The observed coverages are shown in Table 7. The observed coverages suggest that $\delta_y = \sin^{-1}(0.1) \approx 0.10$ was too large in the initial experiment.

In order to verify this conclusion, we examined the accuracy of estimated variances. Figure 5 shows the boxplots of $ns^2_{\hat{p}}/(\hat{p}-\hat{p}^2)$ for $n = 512\,,\ 1\,024\,,\ \ldots\,,\ 262\,144$. This boxplot shows the mean, minimum, maximum, and the $1^{\text{th}}$, $5^{\text{th}}$, $10^{\text{th}}$, $20^{\text{th}}$, $50^{\text{th}}$, $75^{\text{th}}$, $90^{\text{th}}$, $95^{\text{th}}$, and $99^{\text{th}}$ percentiles. The boxplots are based on 1 000 replications. The fractions of replications in which the sequence $\{0,\ldots,0\}$ was observed, were 0.497, 0.271, 0.060, and 0.003 for $n = 512$, 1024, 2048, and 4096, respectively. The conclusion is that the estimation of $\mathrm{Var}(\hat{p})$ is accurate enough.

We also examined the normality of $\hat{p}$ and $\hat{y} = 2\sin^{-1}\left(\sqrt{\hat{p}}\right)$ using the same 1,000 replications as in the analysis of $ns^2_{\hat{p}}/(\hat{p}-\hat{p}^2)$ for $n = 512\,,\ 1\,024\,,\ \ldots\,,\ 262\,144$. The objective of this analysis was to verify that the arcsin transform does not disturb the convergence to the normal limiting distribution. Table 8 contains the $p$-values of Shapiro and Wilk's test for normality. The test statistic $W$ and its $p$-value were evaluated using the NAG Library routine G01DDF; for details see NAG (1991) and Royston (1982a, 1982b). The table clearly indicates that the distribution of $\hat{y}$ converges more quickly towards the normal distribution than the distribution of $\hat{p}$ does. These two verifications support our conclusion that $\delta_y = \sin^{-1}(0.1)$ was too large in the initial experiment.

### *4.3. Closed Queueing Network Model*

In the two other experiments we estimated a histogram of system response time in a closed queueing network model. The model shown in Figure 6 originates in Heidelberger and Welch (1981). The **TTY** node is an infinite server and all the other nodes (**CPU**, **IO-1**, **IO-2**, **IO-3**, and **IO-4**) are FCFS servers. The number of customers is 25. All the service times are exponentially distributed;

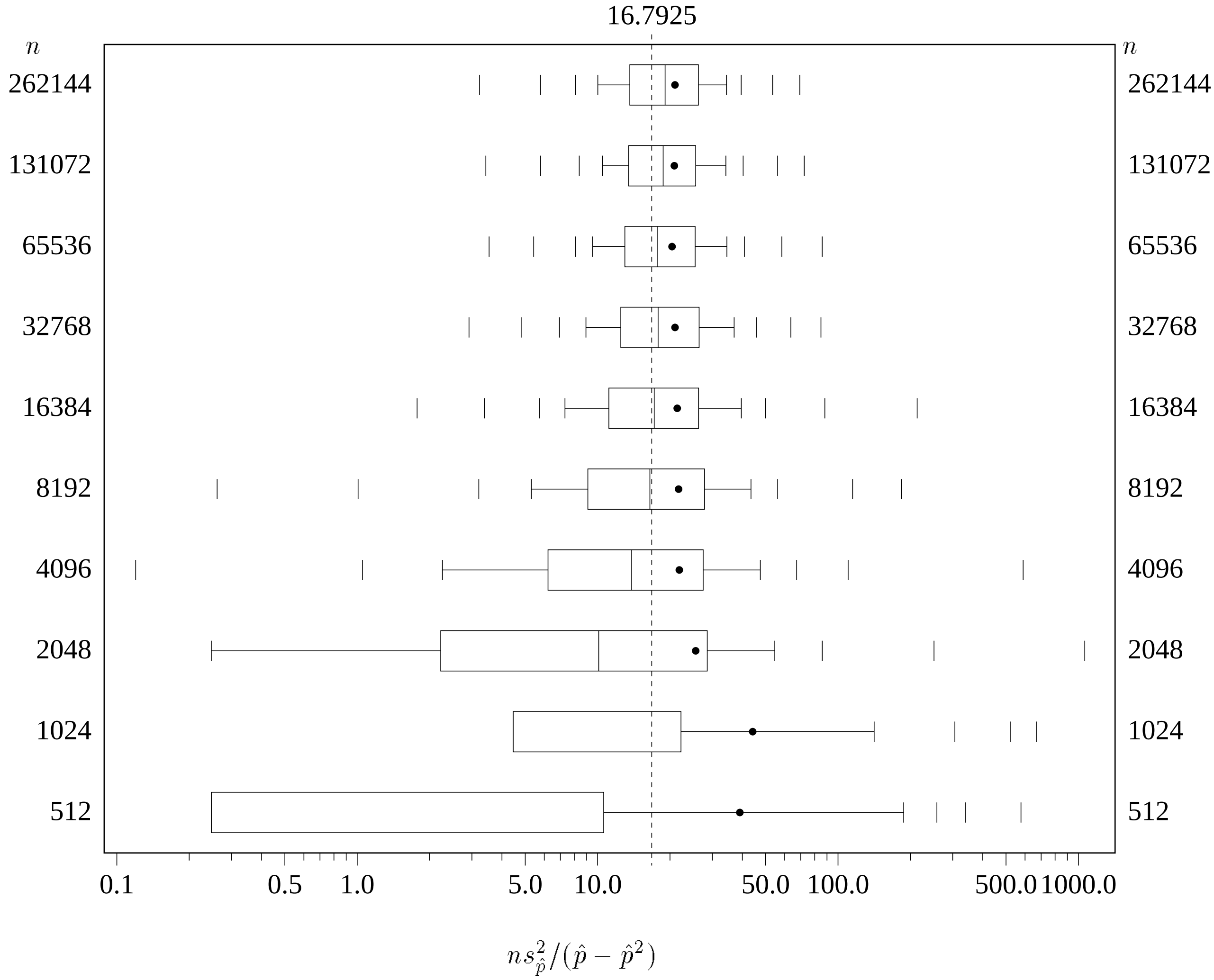
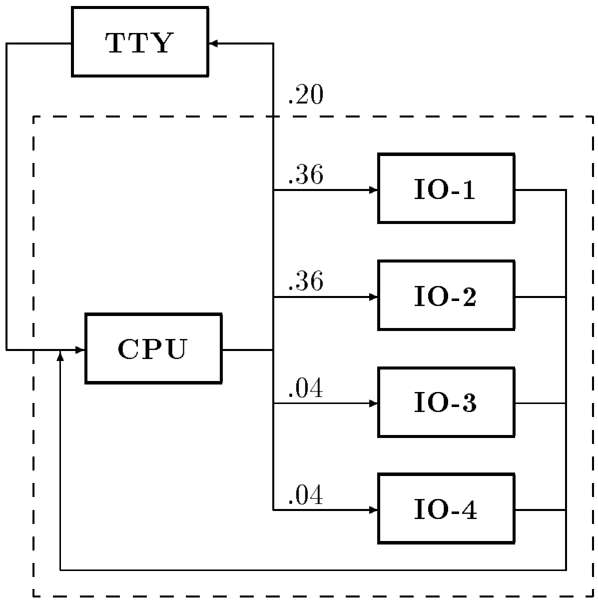



FIGURE 5: Boxplots of $ns_{\hat{p}}^2/(\hat{p}-\hat{p}^2)$; 1 000 replications
$p = \Pr\{\text{arrival queue length} > 19\}$ in M/M/1 queue with $\rho = 0.8$

TABLE 8
*p-values of Shapiro and Wilk's test for normality*
$p = \Pr\{$*arrival queue length* $> 19\}$ *in M/M/1 queue with* $\rho = 0.8$;
$\hat{y} = 2\sin^{-1}\left(\sqrt{\hat{p}}\right)$ ; *1,000 replications*

| | $n$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 512 | 1024 | 2048 | 4096 | 8192 | 16384 | 32768 | 65536 | 131072 | 262144 |
| $\hat{p}$ | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0009 | .0072 | .0718 |
| $\hat{y}$ | .0000 | .0000 | .0000 | .4318 | .6907 | .2571 | .2199 | .0726 | .0340 | .1194 |




| Service Rates | | |
|---|---|---|
| **Case** | **A** | **B** |
| TTY | 0.01 | 0.01 |
| CPU | 1.0 | 1.0 |
| IO-1, IO-2 | 0.72 | 0.18 |
| IO-3, IO-4 | 0.08 | 0.04 |

FIGURE 6: Queueing network model examined

two different sets of service rates are used, denoted as case **A** and case **B**. The simulation runs were initialized using the known load distribution.

One practical problem in estimating a histogram is the specification of class boundaries. We use a $\log_2$ related classification, in which the boundaries are powers of 2. This classification is usually a good ad hoc guess if the analyst believes that the distribution may be long-tailed. In case **A** we used 10 bins and in case **B** 11 bins.

The results are summarized in Figures 7 and 8. The figures contain a summary of run lengths and boxplots of estimated proportions in the system response-time classes. Since the true distributions of the system response-times are unknown, the dashed boxes represent the values of $\hat{p}_k$ such that $\bar{p}_k \in [\tilde{p}_{\mathrm{L}_k}, \tilde{p}_{\mathrm{U}_k}]$. The means $\bar{p}_k$ are estimated from 1,000 independent simulation runs. In these runs we generated response time sequences of length 8,192. Each sequence provides estimates for $p_0, \ldots, p_m$. Each $\bar{p}_k$ is the mean of 1,000 estimates for $p_k$.

In both experiments the observed coverage, 99.6 % in case **A** and 99.4 % in case **B**, is much higher than the requested nominal level of 90 %. However, the run lengths are quite short and stable. Since the Bonferroni inequality provides an lower bound, it is not surprising that the observed coverage exceeds the requested nominal level. However, the observed coverages are almost 100%. Therefore, we conducted additional studies in order to learn reasons for the high coverages.

Firstly, we verified the accuracy of estimated $\mathrm{Var}(\hat{p}_k)$'s through 1,000 independent sequences of length 8,192. These experiments indicated that the estimation of the $\mathrm{Var}(\hat{p}_k)$'s using the spectral

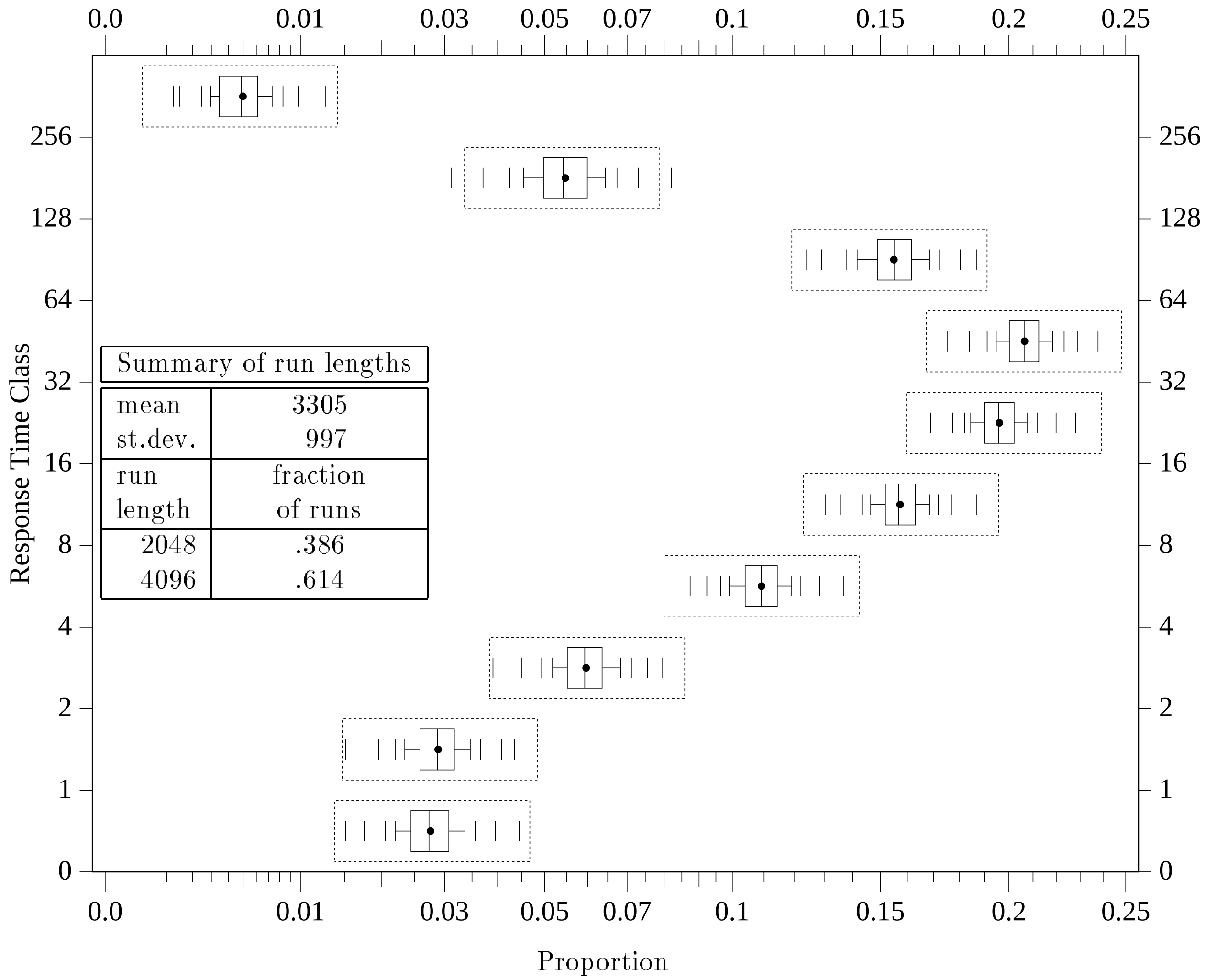



FIGURE 7: Boxplots of estimated proportions in response time classes; case **A**
Requested precision: 90% of proportions simultaneously inside the dashed boxes
$(\alpha = 0.1,\ \delta_y = \sin^{-1}(0.1),\ p_{\text{small}} = 0.001)$

TABLE 9

*Estimated factors $n\,\mathrm{Var}(\hat{y}_k)$ where $\hat{y}_k = 2\sin^{-1}\left(\sqrt{\hat{p}_k}\right)$ and $\hat{p}_k$ is the estimated probability that system response time belongs to response time class $\mathcal{C}_k$. Estimates are based on 1,000 independent sequences of length 8,192.*

| Case | Response Time Class (sec.) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | |
| A | 1.617 | 1.427 | 1.423 | 1.207 | 1.166 | 1.123 | 1.278 | 1.696 | 2.370 | 1.652 | |
| B | 1.040 | 1.006 | 0.993 | 1.102 | 1.123 | 1.315 | 1.610 | 1.097 | 1.009 | 1.272 | 1.968 |

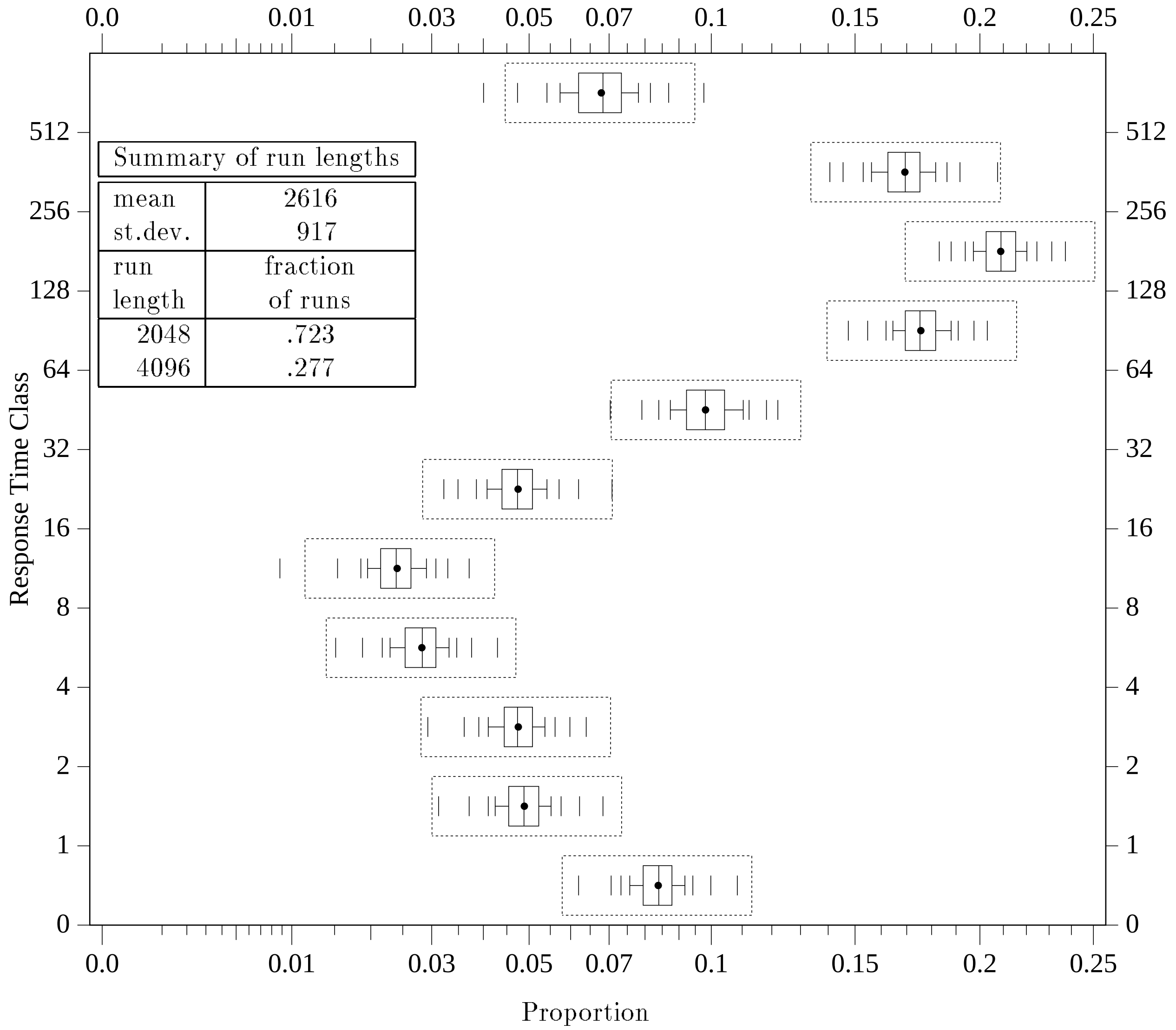



FIGURE 8: Boxplots of estimated proportions in response time classes; case **B**
Requested precision: 90% of proportions simultaneously inside the dashed boxes
($\alpha = 0.1,\ \delta_y = \sin^{-1}(0.1),\ p_{\mathrm{small}} = 0.001$)

method is accurate enough.

Secondly, we examined the factors $n\,\mathrm{Var}(\hat{y}_k)$'s. Since the exact dependency structure of the proportions is unknown, we generated additional 1,000 independent sequences of length 8,192. The estimates of the factors, $ns^2_{\hat{y}_k}$, are shown in Table 9. All of the factors are of the same magnitude. Therefore, we do not have a few components that dominate the run length. In such a case the proposed way of applying the Bonferroni inequality is roughly the same as the traditional way (i.e. choosing $\alpha_k = \alpha/m\,,\ k = 1, \ldots, m$); for details, see Raatikainen (1993). Therefore, we assume that the high coverage is primarily due to the Bonferroni inequality.

### *4.4. Comparison of Run Lengths for Proportions and Means*

In order to assess the amount of computing required in estimating several proportions, we also simulated the corresponding means. In the M/M/1 queue we estimated the mean queue length at arrival instants. In the closed queueing network model we estimated the mean system response-time. These experiments consisted of 1000 independent simulation runs. Each simulation was continued until the half-width of the 90% estimated confidence interval was less than 10% of the estimated mean. The variance was estimated using the spectral method with parameters given in Section 4.1.

The run lengths are compared in Table 10. The table indicates that the estimation of several proportions requires about the same length as the corresponding mean does when $\alpha = 0.1$, $\varepsilon = 0.1$, and $\delta_y = \sin^{-1}(0.1)$. It should be noted that the $\varepsilon = 0.1$ and $\delta_y = \sin^{-1}(0.1)$ are not necessarily equivalent. However, the comparison shows that the estimation of a histogram does not require simulation runs of great lengths.

If we compare the computer times needed to estimate several proportions and a single mean, proportions are more expensive. This is due to the variance estimation. For each proportion we need an estimated variance. However, the spectral method can be implemented so that multiple variance estimates can be calculated efficiently.

## Summary

We have proposed a method for controlling the length of a simulation run when several proportions are simultaneously estimated. The precision requirement is specified as a rectangular confidence area having desired widths of each side. The method uses the arcsin transform in estimating a confidence interval and in selecting the desired widths of the intervals. The Bonferroni inequality is applied in a new way to construct the rectangular confidence area.

TABLE 10

*Comparison of run lengths when proportions and means are estimated*

*proportions:* $\alpha = 0.1,\ \delta_y = \sin^{-1}(0.1)$

*mean:* $\alpha = 0.1,\ \varepsilon = 0.1$

| | M/M/1 | | QNM-A | | QNM-B | |
|---|---|---|---|---|---|---|
| Run length | propor-tions | mean | propor-tions | mean | propor-tions | mean |
| Mean | 24866 | 52609 | 2284 | 3558 | 2040 | 2208 |
| St.dev. | 15554 | 31163 | 672 | 1530 | 373 | 1301 |
| | Fraction of runs | | | | | |
| 512 | | | | .012 | | .099 |
| 1024 | | | .008 | .033 | .049 | .237 |
| 2048 | | | .873 | .300 | .930 | .401 |
| 4096 | .001 | | .119 | .601 | .021 | .259 |
| 8192 | .131 | .018 | | .054 | | .004 |
| 16384 | .452 | .112 | | | | |
| 32768 | .336 | .365 | | | | |
| 65536 | .078 | .424 | | | | |
| 131072 | .002 | .079 | | | | |
| 262144 | | .002 | | | | |

The analysis and example applications indicate that the method controls the run length so that the estimated proportions have the precision requested, when the desired half-width ($\delta_y$) of the confidence intervals for the transformed proportions is narrow enough. In addition, the results show that the simultaneous estimation of several proportions does not require long simulation runs when compared with the estimation of a single mean.[1]

## References

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

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

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

BRENNER, D. J., AND QUAN, H, "Exact Confidence Limits for Binomial Proportions – Pearson and Hartley Revisited," *The Statistician*, 39 (1990), 391–397.

[1]The author would like to thank two anonymous referees, the Associate Editor, and Simulation Departmental Editor James R. Wilson for valuable comments and suggestions that led to the improvement of this paper.

CHOW, Y. AND ROBBINS, H. "On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean," *Annals of Mathematical Statistics,* 36 (1965), 457–462.

COX, D. R., AND HINKLEY, D. V., *Theoretical Statistics*. Chapman and Hall, London, 1974.

DAMERDJI, H., "On Strong Consistency of the Variance Estimator," *Proceedings of the 1987 Winter Simulation Conference*, pp. 305–308.

FISHMAN, G. S., *Principles of Discrete Event Simulation.* John Wiley & Sons, New York, 1978.

GLYNN, P., "Some Asymptotic Formulas for Markov Chains with Applications to Simulation," *Journal of Statistical Computation and Simulation,* 19 (1984), 97–112.

HALD, A., *Statistical Theory with Engineering Applications*. John Wiley & Sons, New York, 1952.

HEIDELBERGER, P., "Simultaneous and Efficient Simulation of Highly Dependable Systems with Different Underlying Distributions". *Proceedings of the 1992 Winter Simulation Conference*, pp. 458–465.

HEIDELBERGER, P., AND WELCH, P. D., "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulation". *Communications of the ACM,* 24 (1981), 233-245.

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

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

LOUIS, T. A., "Confidence Intervals for a Binomial Parameter After Observing No Successes," *The American Statistician,* 35 (1981), 154.

*The NAG Fortran Library Manual, Mark 15, Volume 7.* The Numerical Algorithms Group Ltd, Oxford, UK, 1991.

PAWLIKOWSKI, K., "Steady-state Simulation of Queueing Process: A Survey of Problems and Solutions," *ACM Computing Surveys,* 22 (1990), 123–170.

RAATIKAINEN, K. E. E., "Simultaneous Sequential Confidence Intervals of Fixed Widths for Several Means Using Bonferroni Inequality," University of Helsinki, Department of Computer Science, 1992.

RAATIKAINEN, K. E. E., "A Sequential Procedure for Simultaneous Estimation of Several Means," *ACM Transactions on Modeling and Computer Simulation,* 3 (1993), 108–133.

ROYSTON, J. P., "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples," *Applied Statistics,* 31 (1982a), 115–124.

ROYSTON, J. P., "Algorithm AS181: The W Test for Normality," *Applied Statistics,* 31 (1982b), 176–180.

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

WELCH, P. D., "The Statistical Analysis of Simulation Results," *Computer Performance Modeling Handbook*, S. S. Lavenberg, ed. Academic Press, New York, 1983.