Journal of Quality Technology
A Quarterly Journal of Methods, Applications and Related Topics

ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: http://www.tandfonline.com/loi/ujqt20

The Performance of Bootstrap Control Charts

L. Allison Jones & William H. Woodall

The Performance of Bootstrap Control Charts

L. ALLISON JONES

University of Miami, Coral Gables, FL 33124-8237

WILLIAM H. WOODALL

University of Alabama, Tuscaloosa, AL 35487-0226

The bootstrap is a statistical technique that substitutes computing-power for traditional parametric assumptions. Recently, several authors have considered the application of the bootstrap to statistical quality control charts. Simulation studies have been used to evaluate the performance of the bootstrap control charts. In most cases, the performance measures do not consider the average run lengths of the charts. In this paper, we discuss the proposed bootstrap control chart procedures. We provide extensive computer simulation results and evaluate the performance of each control chart in terms of the average run length.

Introduction

The bootstrap, originally proposed and named by Efron (1979), is a computational technique that can be used to effectively estimate the sampling distribution of a statistic. In particular, one can use the nonparametric bootstrap to estimate the sampling distribution of a statistic, while assuming only that the sample is representative of the population from which it is drawn and that the observations are independent and identically distributed. In its simplest form, the nonparametric bootstrap does not rely on any distributional assumptions about the underlying population.

To see how the nonparametric bootstrap works, suppose we use a random variable, X, to evaluate the performance of a process. Although we do not have any information regarding the distribution of X, we wish to estimate some parameter, θ, that characterizes the performance of the process. For example, θ may be the mean, median, or standard deviation of the population. A sample of n observations is drawn from the population and denoted by $x_1, x_2, \ldots, x_n$. An estimate of the parameter of interest can be computed from this sample and referred to as $\widehat{\theta}$. Since there is no information about the underlying population, there is no information about the sampling distribution of $\widehat{\theta}$. The sampling distribution of $\widehat{\theta}$ is of particular interest because one may wish to construct a confidence interval for θ, or, as in our case, a control chart for monitoring process performance using future values of $\widehat{\theta}$.

According to the nonparametric bootstrap, the empirical distribution function, EDF, can be used to estimate the underlying population cumulative distribution function. The EDF simply assigns a probability of $1/n$ to each value observed in the sample and is written

$$F_n(x) = \frac{1}{n} \cdot (\text{number of } X_i \leq x) \ .$$

A simple random sample of size n can be drawn from the EDF and denoted by $x_1^*, x_2^*, \ldots, x_n^*$. This sample is called the bootstrap sample. Note that the bootstrap sample is equivalent to resampling n observations with replacement from the original n observations. An estimate of $\widehat{\theta}$ can be computed from the bootstrap sample. This estimate is denoted by $\widehat{\theta}^*$ and is called the bootstrap estimate. This resampling procedure can be repeated multiple times, for example, B times. The B bootstrap estimates,

Dr. Jones is an Assistant Professor in the Department of Management Science. She is a Member of ASQ.

Dr. Woodall is the Russell Professor of Statistics in the Department of Management Science and Statistics. He is a Senior Member of ASQ.

$\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$, can be computed from the resamples. A histogram of $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ provides an estimated sampling distribution of $\hat{\theta}$. For a thorough look at bootstrap methods, the interested reader is referred to Efron (1979); Efron and Gong (1983); Gunter (1991, 1992); Mooney and Duval (1993); or Young (1994).

Recently, three papers have proposed the application of the nonparametric bootstrap to statistical process control charts. Only one of these papers evaluates the performance of the bootstrap control chart in terms of the resulting average run length (ARL). A control chart's in-control ARL is the average number of samples before a signal is given, assuming the process is in control. A standard Shewhart control chart's in-control ARL is equal to the inverse of the Type I error probability or false alarm rate.

Bajgier (1992) proposed a bootstrap control chart for subgrouped data. He evaluated the performance of his chart by graphically comparing the simulated distribution of the ARL's of the respective standard Shewhart control chart and his proposed bootstrap control chart. Bajgier's (1992) technique gives a meaningful evaluation of his bootstrap chart's performance. Seppala et al. (1995) proposed a technique called the subgroup bootstrap that can be used to monitor the mean or standard deviation of subgroups of data. Seppala et al. (1995) used a measure called the simulated coverage to evaluate the performance of the subgroup bootstrap method. Liu and Tang (1996) proposed a similar technique for constructing control charts for subgrouped data. They evaluated the performance of their control chart by the difference in the bootstrap control limits and the true control limits. Neither the simulated coverage nor the difference in the bootstrap limits and the true limits gives an adequate picture of the performance of the bootstrap control charts.

In this paper, the three techniques for constructing bootstrap control charts are discussed in the order in which they were proposed. Each chart is discussed in the context of monitoring the mean of a process, assuming independence of the process observations. Historical data are used to determine control limits which then are used prospectively with future data. The discussion of the charts is accompanied by extensive simulation results. These simulations are used to evaluate the performance of each bootstrap control chart in terms of the resulting in-control ARL.

Bajgier's (1992) Bootstrap Control Chart

Bajgier (1992) proposed a bootstrap control chart to monitor the mean of a process. This bootstrap chart is intended as an alternative to the Shewhart $\overline{X}$ control chart. The bootstrap control chart is constructed as follows.

1. Observe k subgroups of size n for a total of $n \cdot k$ observations.
2. Draw a random sample of size n, with replacement, from the pooled sample of $n \cdot k$ observations. This sample, $x_1^*, x_2^*, \ldots, x_n^*$, is a bootstrap sample.
3. Compute the sample mean $(\bar{x}^*)$ from the bootstrap sample drawn in Step 2.
4. Repeat Steps 2–3 a large number of times, say, B times.
5. Sort the B bootstrap estimates, $\bar{x}_1^*, \bar{x}_2^*, \ldots, \bar{x}_B^*$.
6. Find the smallest ordered $\bar{x}^*$ such that $(\alpha/2) \cdot B$ values are below it. This is the bootstrap control chart's lower control limit, LCL.
7. Find the smallest ordered $\bar{x}^*$ such that $(1 - \alpha/2) \cdot B$ values are below it. This is the bootstrap control chart's upper control limit, UCL.

Here, α is the desired false alarm rate, also referred to as the Type I error probability. For a 3-σ Shewhart control chart with the mean and variance assumed known, $\alpha = 0.0027$.

Bajgier (1992) made the observation that it is impossible to compare the performance of the bootstrap control chart to the standard Shewhart $\overline{X}$ control chart based on a single sample of $n \cdot k$ observations. Because each set of $n \cdot k$ observations will produce different bootstrap and standard control limits, the ARL of each chart is a random variable. Bajgier (1992) simulated 1,000 sets of control limits for $k = 20$ samples of size $n = 5$ from a normal distribution and a χ^2 distribution with five degrees of freedom. He computed bootstrap control limits according to the above algorithm. He compared the bootstrap control limits to standard Shewhart control limits computed based on

$$(\text{LCL,UCL}) = \overline{\overline{X}} \pm A_2\, \overline{R}$$

and

$$(\text{LCL,UCL}) = \overline{\overline{X}} \pm A_3\, \overline{S}\ . \qquad (1)$$

Here $\overline{\overline{X}}$ is the grand average of k subgroups of size n, $\overline{R}$ is the average of the k subgroup ranges, $\overline{S}$ is the average of the k subgroup sample standard deviations, and A_2 and A_3 are the standard control chart constants. Bajgier (1992) concluded that his bootstrap control chart performs comparably to the standard methods in terms of the distribution of the ARL's, although no summary statistics are reported from his simulation study.

The Subgroup Bootstrap

Seppala et al. (1995) pointed out an obvious limitation of Bajgier's (1992) bootstrap control chart. The approach implicitly assumes that the process is stable and in control when the control limits are computed. If this assumption is violated, the control limits computed according to Bajgier's (1992) method will be too wide. Seppala et al. (1995) attempted to prevent the necessity of this assumption with their subgroup bootstrap. The subgroup bootstrap assumes that the observations are described by the model

$$X_{ij} = \mu_i + \varepsilon_{ij} ,$$

where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n$. Here, μ_i is the true mean of the i^{th} subgroup, and ε_{ij} is a random error term. An algorithm for the subgroup bootstrap follows:

1. Observe k subgroups of size n for a total of $n \cdot k$ observations.
2. Compute $e_{ij} = x_{ij} - \bar{x}_i$ for $i = 1,\ 2, \ldots, k$ and $j = 1, 2, \ldots, n$, where $\bar{x}_i$ is the sample mean of the i^{th} subgroup.
3. Draw a random sample of size n, with replacement, from the pooled sample of $n \cdot k$ residuals computed in Step 2. This sample, $e_1^*, e_2^*, \ldots, e_n^*$, is a bootstrap sample.
4. Compute $x_j^* = \bar{\bar{x}} + a \cdot e_j^*$ for $j = 1, 2, \ldots, n$. Here, $a = \sqrt{n/(n-1)}$ is a correction factor used to adjust the variance of the resampled subgroups.
5. Compute the sample mean, $\bar{x}^*$, from x_1^*, x_2^*, ..., x_n^*.
6. Repeat Steps 3–5 a large number of times, say, B times.
7. Sort the B bootstrap estimates, $\bar{x}_1^*, \bar{x}_2^*, \ldots, \bar{x}_B^*$.
8. Find the smallest ordered $\bar{x}^*$ such that $(\alpha/2) \cdot B$ values are below it. This is the bootstrap control chart's LCL.
9. Find the smallest ordered $\bar{x}^*$ such that $(1 - \alpha/2) \cdot B$ values are below it. This is the bootstrap control chart's UCL.

Here again, α is the desired false alarm rate for the control chart. Seppala et al. (1995) suggested replacing the control limits found in Steps 8 and 9 with interpolated percentiles. This has little effect on the control limits if B is large and has no effect on the control limits if $(\alpha/2) \cdot B$ and $(1-\alpha/2) \cdot B$ are integer values. Seppala et al. (1995) also suggested a modification to the subgroup bootstrap algorithm that is referred to as the balanced bootstrap technique. The balanced bootstrap is discussed by Davison, Hinkley, and Schectman (1986).

Seppala et al. (1995) evaluated the performance of the subgroup bootstrap control chart and the standard Shewhart $\overline{X}$ chart by a measure they referred to as the simulated coverage of 1,000 sets of each type of control limits. They simulated samples of various sizes from either the standard normal or the exponential distribution with mean equal to one. The exact coverage probability of each set of control limits can be computed according to

$$\text{CVG} = P\left(\text{LCL} < \overline{X}_n < \text{UCL}\right) , \qquad (2)$$

where UCL and LCL are the given upper and lower control limits of the subgroup bootstrap chart and where $\overline{X}_n$ represents a future sample mean based on a sample of size n from the in-control process. A control chart is considered to perform well if CVG is near $1 - \alpha$. Seppala et al. (1995), however, computed the simulated coverage for the upper control and lower control limits separately and did not specify how they computed their measure of simulated coverage. Thus, results consistent with theirs could not be reproduced when conducting similar simulations, even for the standard $\overline{X}$ control chart. For the simulations detailed in the next section, the average coverage was computed by finding the coverage for each set of control limits according to Equation (2). The number of simulated sets of control limits is denoted by NSIM. The average of these coverage values then gives us the average coverage, that is,

$$\text{CVG}_{avg} = \frac{1}{\text{NSIM}} \sum_{i=1}^{\text{NSIM}} \text{CVG}_i \qquad (3)$$

when CVG_i is the observed coverage value for the i^{th} simulated set of control limits.

Results consistent with those reported by Seppala et al. (1995) were found when the NSIM upper and lower control limits were averaged (UCL_{avg}, LCL_{avg}, respectively). The coverage was then computed using

$$\mathrm{CVG}_{avg} = P\left(\overline{X}_n < \mathrm{UCL}_{avg}\right)$$

and

$$\mathrm{CVG}_{avg} = P\left(\overline{X}_n < \mathrm{LCL}_{avg}\right) .$$

However, this method of computing the coverage probability gives a distorted view of the performance of a control chart. Seppala et al. (1995) reported, for example, the simulated coverage of the upper side of the Shewhart $\overline{X}$ control chart, where the mean and standard deviation are estimated from $k = 5$ samples of size $n = 5$ from a normal distribution and where the desired false alarm rate for the upper side of the chart when μ and σ are known is set at $\alpha = 0.0013$. Their results suggest that the simulated coverage for the estimated control limits is 0.9986. This would correspond to an approximate false alarm rate of 0.0014. This implies that the control limits are estimated nearly exactly even though μ and σ are replaced with sample estimates. One might assume that this approximate false alarm rate corresponds to an in-control ARL of $1/0.0014 \approx 714$. This assumption is incorrect for two reasons: 1) the results reported in Seppala et al. (1995) regarding the standard control chart methods are not consistent with other published works concerning the effects of estimating μ and σ on the false alarm rate and the in-control ARL of the Shewhart $\overline{X}$ chart. For example, Quesenberry (1993) and Ghosh, Reynolds, and Hui (1981) showed that it is difficult to estimate control chart limits accurately with small data sets; 2) even if the simulated coverages were correctly computed according to Equations (2) and (3), one cannot compute the in-control ARL for the chart in this manner, that is,

$$\mathrm{ARL}_{avg} \neq \frac{1}{1 - \mathrm{CVG}_{avg}} .$$

To obtain a correct estimate of the average in-control ARL for the control chart, one would need to compute it as follows:

$$\mathrm{ARL}_{avg} = \frac{1}{\mathrm{NSIM}} \sum_{i=1}^{\mathrm{NSIM}} \frac{1}{1 - \mathrm{CVG}_i} . \tag{4}$$

In an attempt to compare the simulation results reported in Seppala et al. (1995) to other published works regarding the performance of a standard $\overline{X}$ control chart, we conducted a simulation study based on 40,700 sets of control limits. The control limits were computed according to

$$\overline{\overline{x}} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) , \tag{5}$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution and where

$$s^2 = \frac{1}{k(n-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 .$$

Derman and Ross (1995) recommended this approach, which uses the pooled subgroup variance estimate for computing the control limits of an $\overline{X}$ chart. The control limits in our simulation were based on $k = 5$ samples of size $n = 5$ from a standard normal distribution, with a desired false alarm rate of $\alpha = 0.0026$. The coverage probabilities for the upper and lower side of each simulated control chart were computed separately according to

$$\mathrm{CVG}_U = P\left(\overline{X}_n < \mathrm{UCL}\right)$$

and

$$\mathrm{CVG}_L = P\left(\overline{X}_n < \mathrm{LCL}\right) .$$

The average coverages for the upper and lower sides of the control charts were computed separately according to

$$\mathrm{CVG}_{U,avg} = \frac{1}{\mathrm{NSIM}} \sum_{i=1}^{\mathrm{NSIM}} \mathrm{CVG}_{U,i}$$

and

$$\mathrm{CVG}_{L,avg} = \frac{1}{\mathrm{NSIM}} \sum_{i=1}^{\mathrm{NSIM}} \mathrm{CVG}_{L,i} ,$$

where $\mathrm{CVG}_{U,i}$ and $\mathrm{CVG}_{L,i}$ are the upper and lower coverage values, respectively, for the i^{th} simulated set of control limits. The average ARL's for the upper and lower sides of the control charts were also computed separately using

$$\mathrm{ARL}_{U,avg} = \frac{1}{\mathrm{NSIM}} \sum_{i=1}^{\mathrm{NSIM}} \frac{1}{1 - \mathrm{CVG}_{U,i}}$$

and

$$\mathrm{ARL}_{L,avg} = \frac{1}{\mathrm{NSIM}} \sum_{i=1}^{\mathrm{NSIM}} \frac{1}{\mathrm{CVG}_{L,i}} ,$$

respectively. Since two-sided control charts are more common in practice, the coverage probabilities in Equation (2), the CVG_{avg} in Equation (3), and the ARL_{avg} in Equation (4) were also computed for the two-sided $\overline{X}$ chart. This allows us to compare these simulation results to those in Quesenberry (1993), who reported the average ARL performance of 40,700 two-sided $\overline{X}$ control charts with limits computed from Equation (1). Although the variance estimate used by Quesenberry (1993) is slightly different than

TABLE 1. Comparison of Simulated Performance of $\overline{X}$ Control Chart

		Simulation Results	Quesenberry (1993)	Seppala et al. (1995)
One-sided Chart Upper	Avg. Control Limit	1.33 (0.0015)	not reported	1.33
	Avg. Coverage	0.9937 (0.0001)	not reported	0.9986
	Avg. ARL	25786.8 (1957)	not reported	not reported
One-sided Chart Lower	Avg. Control Limit	−1.33 (0.0015)	not reported	−1.32
	Avg. Coverage	0.0062 (0.0001)	not reported	0.0016
	Avg. ARL	27883.7 (2906)	not reported	not reported
Two-sided Chart	Avg. Coverage	0.9874 (0.0001)	not reported	not reported
	Avg. ARL	1138.8 (54)	1347 (41)	not reported

Note: Standard errors are reported in parentheses where available. Simulation results and Quesenberry (1993) values are based on 40,700 sets of control limits. Seppala et al. (1995) values are based on 1,000 sets of control limits.

the pooled estimate used in our simulation, the results are comparable. Seppala et al. (1995) also computed the limits for the $\overline{X}$ chart according to Equation (5). Our simulation results, as well as those reported by Quesenberry (1993) and Seppala et al. (1995), are displayed in Table 1 which shows that parameter estimation greatly increases the resulting ARL of the $\overline{X}$ chart.

Simulated Performance of the Subgroup Bootstrap

Seppala et al. (1995) used an experimental design approach to evaluate the effect of the number of subgroups, k; the subgroup size, n; the false alarm rate, α; and the population distribution on the performance of the subgroup bootstrap and the standard Shewhart control chart. In this section, the experimental design is repeated and the subgroup bootstrap algorithm presented in the previous section is used for the simulation results. Our simulations also used the balanced bootstrap approach suggested by Seppala et al. (1995); similar results were obtained using either algorithm. The control limits for the standard Shewhart control chart were computed using Equation (5). The same factor levels considered by Seppala et al. (1995) were considered:

- Number of subgroups: $k = 5$, $k = 20$
- Subgroup size: $n = 5$, $n = 10$
- False alarm rate: $\alpha = 0.10$, $\alpha = 0.01$, and $\alpha = 0.0026$
- Population Distribution: Normal (0,1) and Exponential (1).

In this simulation, $B = 2{,}000$ bootstrap resamples are used to compute each set of bootstrap control limits. Seppala et al. (1995) replicate their experiment 1,000 times and do not report standard errors for their average estimates. We found that 1,000 replications were not enough to estimate the ARL of the subgroup bootstrap accurately. For some factor level combinations, 1,000 simulations produced ARL_{avg} values with standard error estimates as large as 16,000. It is clearly impossible to draw valid conclusions regarding the performance of the subgroup bootstrap with this amount of variability in the estimate of ARL_{avg}. Therefore, for each factor level combination, the number of replications was varied from 1,000 to 100,000. The number of replications (NSIM) was chosen to achieve a reasonably small

TABLE 2. Comparison of Subgroup Bootstrap Control Chart to Standard Shewhart Control Chart With 90% Limits ($\alpha = 0.10$)

Distribution		EXP(1)				N(0,1)			
Subgroup Size	n	5	5	10	10	5	5	10	10
# Subgroups	k	5	20	5	20	5	20	5	20
UCL_{avg}	standard	1.7106	1.7271	1.5117	1.5156	0.7231	0.7288	0.5229	0.5192
		(0.0116)	(0.0059)	(0.0071)	(0.0011)	(0.0069)	(0.0037)	(0.0047)	(0.0024)
	bootstrap	1.7668	1.7956	1.5512	1.5581	0.7246	0.7279	0.5221	0.5182
		(0.0123)	(0.0066)	(0.0075)	(0.0019)	(0.0070)	(0.0038)	(0.0047)	(0.0025)
	desired	1.8307	1.8307	1.5705	1.5705	0.7356	0.7356	0.5201	0.5201
LCL_{avg}	standard	0.0290	0.2754	0.4850	0.4819	−0.7363	−0.7296	−0.5156	−0.5181
		(0.0044)	(0.0025)	(0.0030)	(0.0016)	(0.0073)	(0.0036)	(0.0047)	(0.0025)
	bootstrap	0.3574	0.3527	0.5336	0.5326	−0.7363	−0.7306	−0.5169	−0.5187
		(0.0037)	(0.0020)	(0.0028)	(0.0014)	(0.0074)	(0.0037)	(0.0047)	(0.0025)
	desired	0.3940	0.3940	0.5425	0.5425	−0.7356	−0.7356	−0.5201	−0.5201
CVG_{avg}	standard	0.8586	0.9032	0.8654	0.8982	0.8546	0.8856	0.8629	0.8888
		(0.0032)	(0.0014)	(0.0026)	(0.0011)	(0.0025)	(0.0010)	(0.0019)	(0.0007)
	bootstrap	0.8531	0.8964	0.8582	0.8907	0.8544	0.8850	0.8628	0.8882
		(0.0030)	(0.0013)	(0.0023)	(0.0009)	(0.0025)	(0.0010)	(0.0019)	(0.0008)
	desired	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000
ARL_{avg}	standard	19.98	13.38	12.54	11.10	9.31	9.43	8.65	9.38
		(2.64)	(0.28)	(1.27)	(0.15)	(0.19)	(0.07)	(0.12)	(0.06)
	bootstrap	13.14	11.66	9.19	9.83	9.29	9.42	8.69	9.38
		(0.86)	(0.20)	(0.19)	(0.09)	(0.19)	(0.09)	(0.12)	(0.07)
	desired	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
SRL_{avg}	standard	83.63	8.86	40.11	4.76	6.05	2.72	3.80	1.95
	bootstrap	27.26	6.22	6.22	2.79	6.05	2.81	3.89	2.11

Note: The values in parentheses are standard errors.

standard error of ARL_{avg}. For example, when $k = 5$, $n = 5$, and $\alpha = 0.10$ and when the population is Normal (0,1), the simulation was conducted as follows:

1. Generate $n \cdot k = 25$ observations from a Normal (0,1) distribution.
2. Construct the subgroup bootstrap control limits based on $B = 2{,}000$ resamples from these 25 observations.
3. Construct the standard Shewhart control limits based on these 25 observations.
4. Compute CVG_i for the subgroup bootstrap limits and the Shewhart limits from Equation (2).
5. Repeat the entire simulation NSIM = 1,000 times.

The simulations were conducted in FORTRAN using the IMSL library. For each treatment combination and each control chart type, the following performance measures are reported:

- The average upper and lower control limits, defined by

$$\text{UCL}_{avg} = \frac{1}{\text{NSIM}} \sum_{i=1}^{\text{NSIM}} \text{UCL}_i$$

and

$$\text{LCL}_{avg} = \frac{1}{\text{NSIM}} \sum_{i=1}^{\text{NSIM}} \text{LCL}_i \,.$$

TABLE 3. Comparison of Subgroup Bootstrap Control Chart to Standard Shewhart Control Chart With 99% Limits ($\alpha = 0.01$)

Distribution		EXP(1)				N(0,1)			
Subgroup Size	n	5	5	10	10	5	5	10	10
# Subgroups	k	5	20	5	20	5	20	5	20
UCL_{avg}	standard	2.1120	2.1416	1.79983	1.8104	1.1401	1.1464	0.8178	0.8124
		(0.0015)	(0.0011)	(0.0009)	(0.0006)	(0.0038)	(0.0019)	(0.0024)	(0.0011)
	bootstrap	2.2676	2.3568	1.9152	1.9494	1.1132	1.1329	0.8083	0.8058
		(0.0018)	(0.0014)	(0.0010)	(0.0008)	(0.0040)	(0.0021)	(0.0025)	(0.0013)
	desired	2.5188	2.5188	1.9998	1.9998	1.1519	1.1519	0.8146	0.8146
LCL_{avg}	standard	−0.1112	−0.1410	0.2011	0.1891	−1.1431	−1.1473	−0.8091	−0.8145
		(0.0007)	(0.0005)	(0.0004)	(0.0006)	(0.0038)	(0.0019)	(0.0024)	(0.0012)
	bootstrap	0.0850	0.0477	0.3397	0.3319	−1.1307	−1.1487	−0.8107	−0.8183
		(0.0004)	(0.0004)	(0.0002)	(0.0002)	(0.0040)	(0.0021)	(0.0025)	(0.0013)
	desired	0.2156	0.2156	0.3717	0.3717	−1.1519	−1.1519	−0.8146	−0.8146
CVG_{avg}	standard	0.9538	0.9758	0.9655	0.9813	0.9718	0.9859	0.9769	0.9872
		(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0005)	(0.0001)	(0.0003)	(0.0001)
	bootstrap	0.9607	0.9853	0.9707	0.9867	0.9671	0.9845	0.9752	0.9865
		(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0005)	(0.0001)	(0.0004)	(0.0001)
	desired	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900
ARL_{avg}	standard	877.81	81.35	397.38	95.19	149.44	102.42	98.68	93.44
		(202.24)	(0.50)	(34.37)	(0.46)	(5.71)	(1.04)	(1.87)	(0.59)
	bootstrap	8,397.31	280.84	663.63	130.80	124.18	97.07	94.20	92.78
		(868.08)	(8.52)	(262.20)	(0.82)	(4.38)	(1.07)	(1.75)	(0.68)
	desired	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SRL_{avg}	standard	63,954.54	111.98	10,868.91	102.95	403.95	82.65	132.20	41.62
	bootstrap	274,510.68	1905.13	82,914.51	184.41	309.98	90.31	123.40	48.10

Note: The values in parentheses are standard errors.

- The average coverage, computed from Equation (3).
- The average in-control ARL, computed using Equation (4).
- The standard deviation of the ARL, defined as

$$\mathrm{SRL}_{avg} = \sqrt{\frac{1}{\mathrm{NSIM}-1}} \times \sqrt{\sum_{i=1}^{\mathrm{NSIM}} \left(\frac{1}{1-\mathrm{CVG}_i} - \mathrm{ARL}_{avg}\right)^2}.$$

Although Seppala et al. (1995) considered only one-sided control charts, all performance measures are reported here for two-sided control charts since two-sided charts are more frequently used in practice. Standard errors are also reported for all average measurements. The results of these simulations appear in Tables 2–4.

Figures 1–3 display a graphical comparison of the ARL_{avg} for the subgroup bootstrap and the standard control charts. Desired in-control ARL values ($1/\alpha$) are also displayed as a point of reference. In most cases, the subgroup bootstrap and the standard control charts have ARL_{avg} values that are comparable to the desired values when the control limits are estimated from normally distributed samples. When sampling from an exponential distribution, however, the ARL_{avg} of the subgroup bootstrap is sensitive to both the sample size and the false alarm rate of the control chart. For example, the ARL_{avg} for the subgroup bootstrap is very large for most sample sizes when the desired false alarm rate is 0.0026 and 0.01. The ARL_{avg} is closer to the desired ARL in these

TABLE 4. Comparison of Subgroup Bootstrap Control Chart to Standard Shewhart Control Chart With 99.74% Limits ($\alpha = 0.0026$)

Distribution		EXP(1)				N(0,1)			
Subgroup Size	n	5	5	10	10	5	5	10	10
# Subgroups	k	5	20	5	20	5	20	5	20
UCL_{avg}	standard	2.2984	2.3345	1.9340	1.9480	1.3295	1.3396	0.9513	0.9516
		(0.0017)	(0.0012)	(0.0010)	(0.0007)	(0.0029)	(0.0015)	(0.0017)	(0.0009)
	bootstrap	2.4599	2.5900	2.0660	2.1149	1.2538	1.2869	0.9139	0.9190
		(0.0020)	(0.0017)	(0.0012)	(0.0009)	(0.0030)	(0.0017)	(0.0018)	(0.0010)
	desired	2.8886	2.8886	2.2237	2.2237	1.3468	1.3468	0.9523	0.9523
LCL_{avg}	standard	−0.2975	−0.3340	0.0659	0.0522	−1.3307	−1.3421	−0.9453	−0.9514
		(0.0008)	(0.0006)	(0.0004)	(0.0003)	(0.0029)	(0.0015)	(0.0017)	(0.0009)
	bootstrap	−0.0246	−0.0813	0.2618	0.2501	−1.2929	−1.3312	−0.9392	−0.9508
		(0.0005)	(0.0005)	(0.0003)	(0.0002)	(0.0032)	(0.0018)	(0.0019)	(0.0011)
	desired	0.1570	0.1570	0.3068	0.3068	−1.3468	−1.3468	−0.9523	−0.9523
CVG_{avg}	standard	0.9694	0.9862	0.9802	0.9907	0.9876	0.9956	0.9916	0.9963
		(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0002)	(0.0000)	(0.0001)	(0.0000)
	bootstrap	0.9740	0.9922	0.9845	0.9947	0.9817	0.9937	0.9889	0.9950
		(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0003)	(0.0001)	(0.0001)	(0.0000)
	desired	0.9974	0.9974	0.9974	0.9974	0.9974	0.9974	0.9974	0.9974
ARL_{avg}	standard	2,013.83	180.52	2,382.26	233.91	1,072.99	436.78	430.96	370.50
		(149.58)	(1.60)	(247.95)	(1.58)	(78.44)	(4.84)	(9.09)	(2.35)
	bootstrap	26,129.26	1,726.82	7,220.42	745.03	614.83	343.61	370.03	319.80
		(1,551.47)	(120.32)	(805.64)	(49.11)	(70.27)	(5.08)	(19.12)	(2.88)
	desired	384.62	384.62	384.62	384.62	384.62	384.62	384.62	384.62
SRL_{avg}	standard	47,302.65	357.01	78,408.67	353.75	7,444.01	483.98	909.28	235.43
	bootstrap	490,619.16	26,905.38	254,766.34	10,980.88	7,027.30	508.31	1,912.46	288.41

Note: The values in parentheses are standard errors.

cases when $k = 20$ subgroups of size $n = 10$ are used to construct the control limits. When the desired false alarm rate is 0.10, the ARL_{avg} of the subgroup bootstrap is close to the desired ARL of 10. It is unlikely, however, that control limits with such a large false alarm rate will be useful in practice.

The performance of the bootstrap and standard methods are not nearly as good as reported by Seppala et al. (1995). The average in-control ARL can actually be far from the desired in-control ARL, and the average ARL can be quite variable when using either method. The values in Tables 2–4 show a clearer picture of the performance of the subgroup bootstrap and the standard method. For example, when estimating 99.74% control limits ($\alpha = 0.0026$) from $k = 20$ samples of size $n = 5$ from an exponential distribution, the ARL_{avg} of the subgroup bootstrap is 1,727. The corresponding ARL_{avg} values for the standard method and the desired ARL are 181 and 385, respectively. The SRL_{avg} values are 26,906 for the subgroup bootstrap and 357 for the standard limits in this scenario. That is, in this case, the ARL for the subgroup bootstrap is nearly ten times larger and more than seventy-five times more variable than the ARL of the standard method. Variability in the performance of the subgroup bootstrap when estimating control limits with a small false alarm probability is a severe deterrent to the practicality of this method. As already mentioned, control charts are usually desired to have small false alarm rates, such as 0.0026. Furthermore, a nonparametric method is most useful when the population follows a skewed distribution. The simulation results show that, in this important and practical case, the performance of the subgroup bootstrap is not as stable as the performance of the standard method.

An obvious problem with the subgroup bootstrap method is that it is sensitive to extreme values.

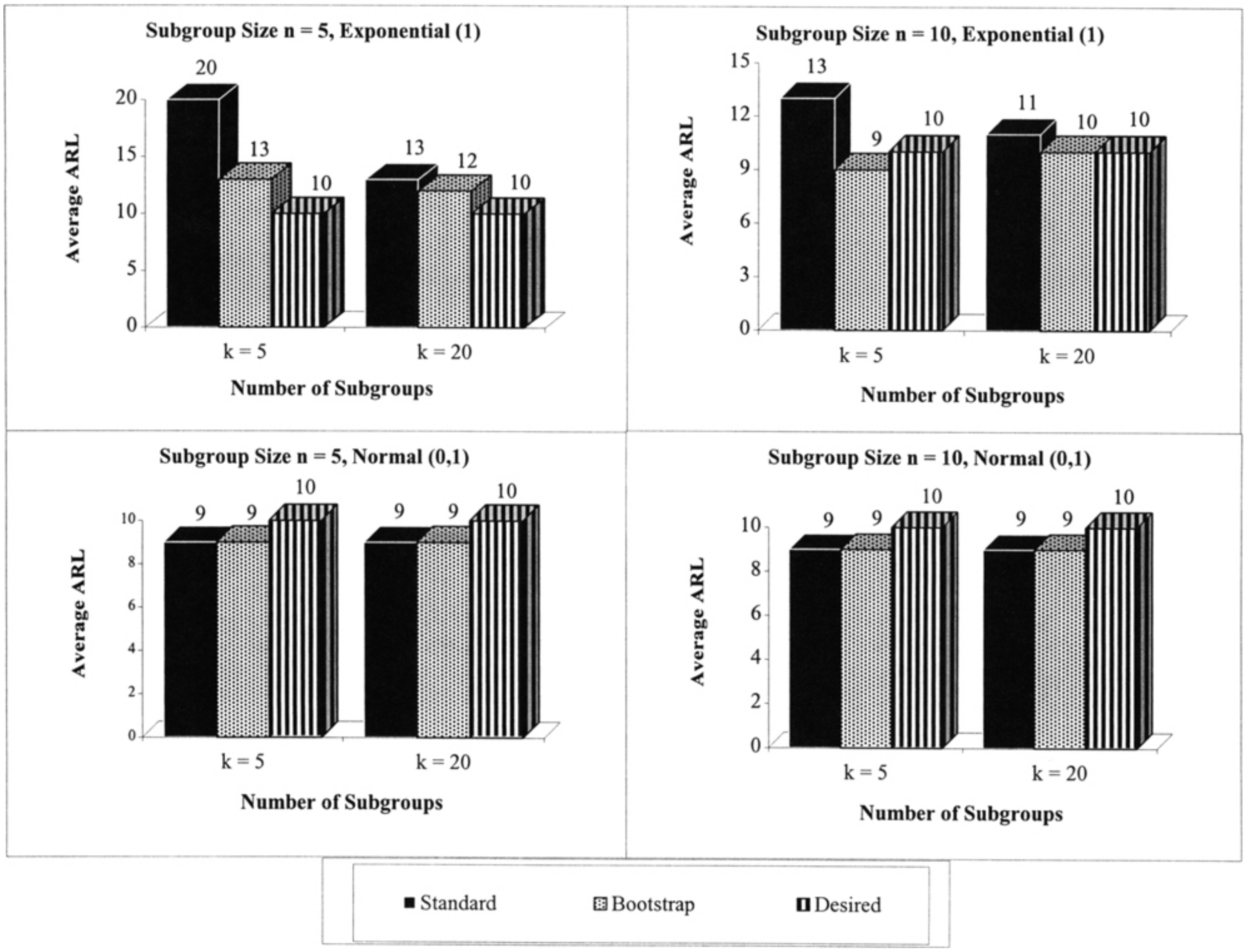

FIGURE 1. Comparison of Simulated ARL Between Subgroup Bootstrap and Standard Control Charts With 90% Control Limits ($\alpha = 0.10$).

Seppala et al. (1995) attempted to adjust for out-of-control observations by subtracting the subgroup mean from each observation and subsequently adding back the grand mean. It is well known that the sample mean is sensitive to extreme values; simply adding the grand mean to the subgroup residuals without checking for out-of-control observations does not lessen the impact of the extreme points. Clearly, the standard method is also sensitive to out-of-control points. When using either the subgroup bootstrap or the standard control chart, one must verify that the process is in control when the historical data are collected before constructing the control limits for prospective use.

Liu and Tang's (1996) Bootstrap Control Chart for Independent Observations

Liu and Tang (1996) proposed bootstrap control charts to monitor the mean of both independent and dependent processes; however, this paper discusses only their technique for monitoring independent processes. Their bootstrap control chart for independent processes is constructed as follows.

1. Observe k subgroups of size n for a total of $N = n \cdot k$ observations.
2. Compute $\overline{\overline{X}}_N = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}$.
3. Draw a random sample of size n, with replacement, from the pooled sample of N observations. This sample, $x_1^*, x_2^*, \ldots, x_n^*$, is a bootstrap sample.
4. Compute the sample mean $\bar{x}^* = \frac{1}{n} \sum_{i=1}^{n} x_i^*$ and $t^* = \sqrt{n}(\bar{x}^* - \overline{\overline{X}}_N)$ from the bootstrap sample drawn in Step 3.
5. Repeat Steps 3 and 4 a large number of times, say, B times.

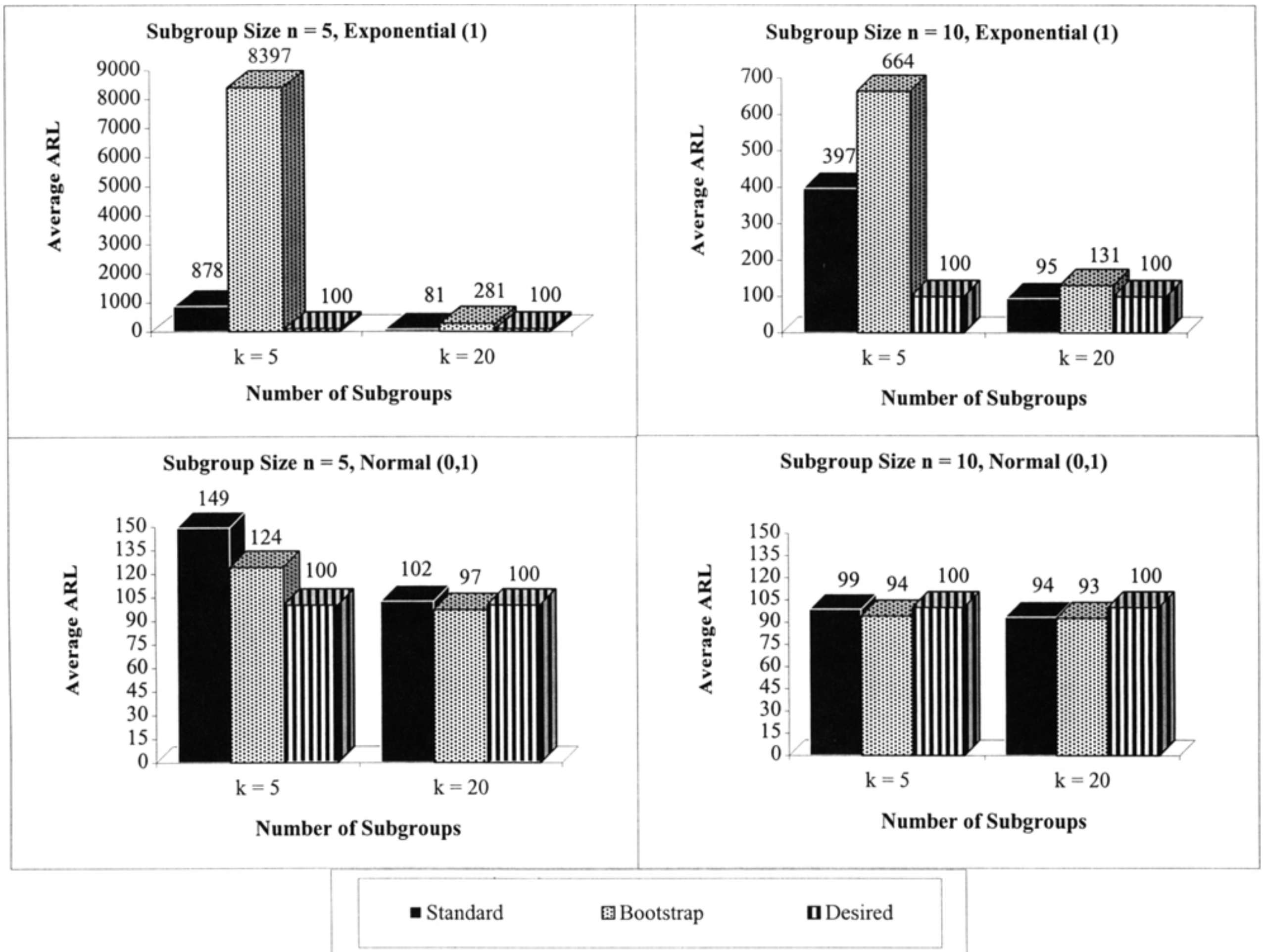

FIGURE 2. Comparison of Simulated ARL Between Subgroup Bootstrap and Standard Control Charts With 99% Control Limits ($\alpha = 0.01$).

6. Sort the B bootstrap values, $t_1^*, t_2^*, \ldots, t_B^*$.
7. Find the smallest ordered t^* such that $(\alpha/2) \cdot B$ values are below it. Denote this value by $t_{\alpha/2}^*$.
8. Find the smallest ordered t^* such that $(1-\alpha/2) \cdot B$ values are below it. Denote this value by $t_{1-\alpha/2}^*$.
9. Compute the lower and upper control limits using

$$\mathrm{LCL} = \overline{\overline{X}}_N + \frac{t_{\alpha/2}^*}{\sqrt{n}} \tag{6}$$

and

$$\mathrm{UCL} = \overline{\overline{X}}_N + \frac{t_{1-\alpha/2}^*}{\sqrt{n}} . \tag{7}$$

It is worth noting that

$$t_{\alpha/2}^* = \sqrt{n}\left(\bar{x}_{\alpha/2}^* - \overline{\overline{X}}_N\right) , \tag{8}$$

where $\bar{x}_{\alpha/2}^*$ is the smallest ordered value of $\bar{x}^*$ such that $(\alpha/2) \cdot B$ values are below it. Similarly,

$$t_{1-\alpha/2}^* = \sqrt{n}\left(\bar{x}_{1-\alpha/2}^* - \overline{\overline{X}}_N\right) , \tag{9}$$

where $\bar{x}_{1-\alpha/2}^*$ is the smallest ordered value of $\bar{x}^*$ such that $(1-\alpha/2) \cdot B$ values are below it. Substituting Equation (8) into Equation (6), it is clear that

$$\mathrm{LCL} = \overline{\overline{X}}_N + \frac{\sqrt{n}\left(\bar{x}_{\alpha/2}^* - \overline{\overline{X}}_N\right)}{\sqrt{n}},$$

and thus,

$$\mathrm{LCL} = \bar{x}_{\alpha/2}^* .$$

By substituting Equation (9) into Equation (7) it can also be shown that

$$\mathrm{UCL} = \bar{x}_{1-\alpha/2}^* .$$

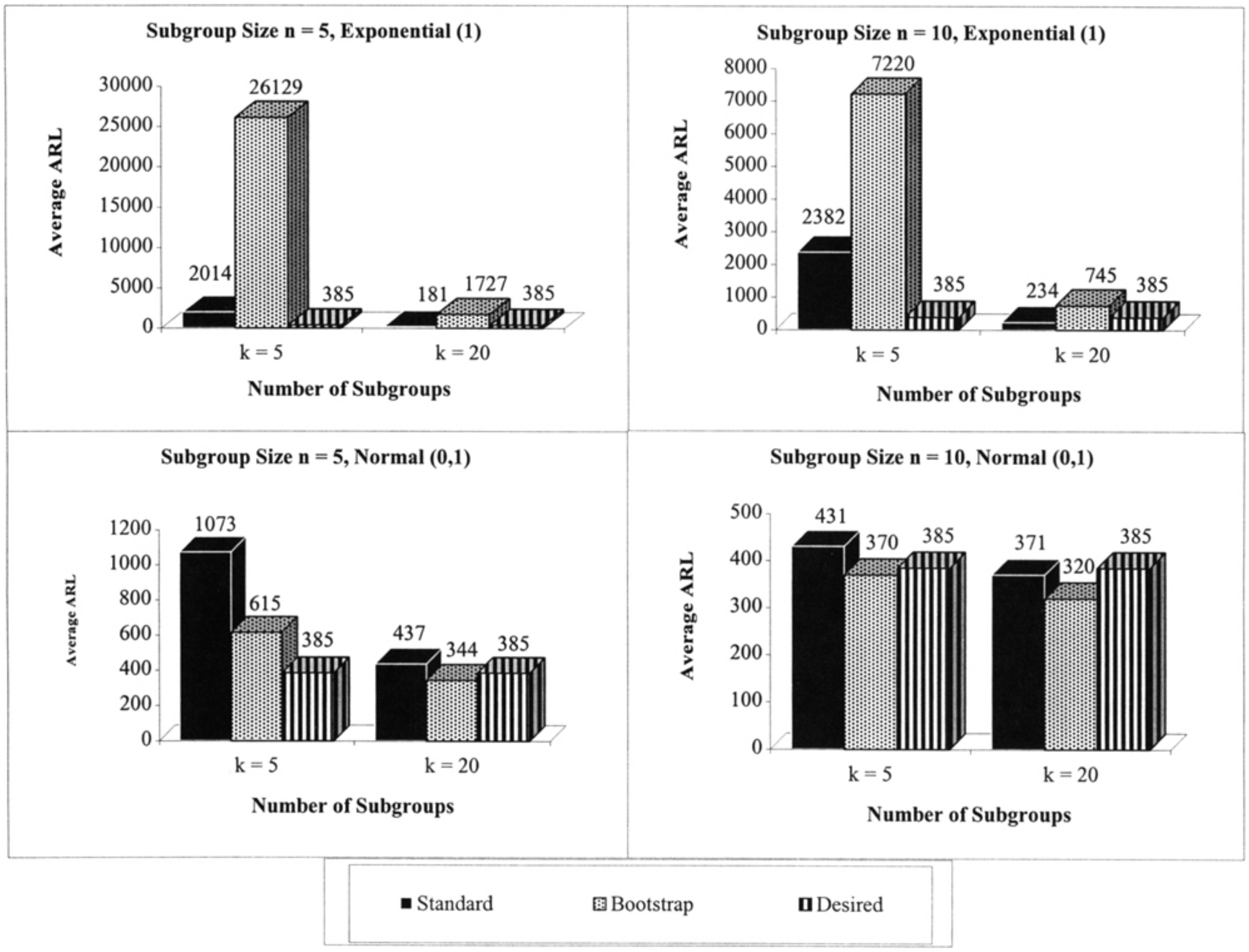

FIGURE 3. Comparison of Simulated ARL Between Subgroup Bootstrap and Standard Control Charts With 99.74% Control Limits ($\alpha = 0.0026$).

Thus, Liu and Tang's (1996) bootstrap control chart for monitoring independent data is equivalent to the method proposed by Bajgier (1992).

Simulated Performance of Bajgier's (1992) and Liu and Tang's (1996) Bootstrap Control Chart

Liu and Tang (1996) discuss the asymptotic properties of their bootstrap control chart and use a simulation study to evaluate the performance of their bootstrap control chart in particular situations. They generate $k = 25$ samples of size $n = 4$ observations from both the standard normal distribution and the exponential distribution with mean equal to one. They compute the bootstrap control limits, based on $B = 200$ bootstrap resamples, according to Equations (6) and (7) and the standard Shewhart control limits according to Equation (5) at the 10%, 5%, and 2% Type I error rates. To evaluate the performance of the charts, they compare the bootstrap and the standard control limits to the exact quantiles of the appropriate distributions.

Using only one simulated set of control limits of each type for each distribution, Liu and Tang (1996) conclude that the bootstrap limits are superior to the standard limits. As discussed by Bajgier (1992), it is impossible to draw such conclusions about the performance of the bootstrap control chart based on a single sample of size $n \cdot k$ since both the bootstrap and the standard control limits are random variables.

A simulation similar to the one performed in Liu and Tang (1996) is conducted to study the performance of this bootstrap control chart. At each of the 10%, 5%, 2%, and 0.27% Type I error rates, this simulation is conducted as follows:

TABLE 5. Comparison of Bajgier's (1992) and Liu and Tang's (1996) Bootstrap Control Chart to Standard Shewhart Method

Distribution		EXP(1)				N(0,1)			
False Alarm Rate	α	10%	5%	2%	0.27%	10%	5%	2%	0.27%
UCL_{avg}	standard	1.8056	1.9621	2.1482	2.4821	0.8209	0.9838	1.1619	1.5045
		(0.0060)	(0.0066)	(0.0077)	(0.0091)	(0.0037)	(0.0039)	(0.0044)	(0.0051)
	bootstrap	1.9194	2.1560	2.4321	2.9114	0.8177	0.9731	1.1483	1.4383
		(0.0068)	(0.0087)	(0.0103)	(0.0010)	(0.0037)	(0.0039)	(0.0045)	(0.0057)
	desired	1.9384	2.1918	2.5113	3.1701	0.8224	0.9800	1.1632	1.5000
LCL_{avg}	standard	0.1889	0.0318	−0.1421	−0.4804	−0.8276	−0.9782	−1.1565	−1.5019
		(0.0026)	(0.0031)	(0.0036)	(0.0051)	(0.0038)	(0.0040)	(0.0044)	(0.0050)
	bootstrap	0.3418	0.2750	0.2099	0.1191	−0.8252	−0.9749	−1.1515	−1.4860
		(0.0015)	(0.0014)	(0.0012)	(0.0011)	(0.0038)	(0.0040)	(0.0046)	(0.0059)
	desired	0.3416	0.2725	0.2058	0.1163	−0.8224	−0.9800	−1.1632	−1.5000
CVG_{avg}	standard	0.9094	0.9452	0.9658	0.9853	0.8912	0.9429	0.9746	0.9958
		(0.0012)	(0.0009)	(0.0007)	(0.0004)	(0.0009)	(0.0007)	(0.0004)	(0.0001)
	bootstrap	0.8872	0.9378	0.9703	0.9921	0.8902	0.9731	0.9731	0.9941
		(0.0010)	(0.0008)	(0.0005)	(0.0002)	(0.0009)	(0.0039)	(0.0004)	(0.0002)
	desired	0.9000	0.9500	0.9800	0.9973	0.9000	0.9500	0.9800	0.9973
ARL_{avg}	standard	13.95	25.73	46.89	138.34	9.90	20.26	51.42	480.40
		(0.32)	(0.97)	(1.49)	(5.27)	(0.09)	(0.26)	(0.97)	(16.12)
	bootstrap	9.63	18.81	43.72	263.64	9.69	19.18	46.47	339.57
		(0.09)	(0.27)	(0.80)	(9.18)	(0.08)	(0.22)	(0.79)	(15.54)
	desired	10.00	20.00	50.00	370.37	10.00	20.00	50.00	370.37
SRL_{avg}	standard	10.14	30.60	47.08	166.63	2.80	8.32	30.71	509.81
	bootstrap	2.99	8.56	25.24	290.18	2.50	7.09	25.07	491.28

Note: The values in parentheses are standard errors.

1. Generate $k = 25$ samples of size $n = 4$ from a Normal (0,1) distribution.
2. Construct the bootstrap control limits according to Equations (6) and (7) based on $B = 2{,}000$ resamples from the 100 observations. (Note we chose to use 2,000 resamples rather than 200 as used by Liu and Tang (1996) in order to obtain more accurate percentile estimates.)
3. Construct the standard Shewhart control limits based on the 100 observations using Equation (5).
4. Compute CVG_i for the bootstrap limits and the Shewhart limits from Equation (2).
5. Repeat the entire simulation 1,000 times.

The entire simulation is repeated for data generated from an exponential distribution with mean equal to one. The performance measures are the same measures for the simulated performance of the subgroup bootstrap, with 1,000 replications for each case. The results of these simulations appear in Table 5.

Figure 4 displays a graphical comparison of the ARL_{avg} for the bootstrap and the standard control charts. The simulated performance of Bajgier's (1992) and Liu and Tang's (1996) bootstrap control chart is comparable to the simulated performance of the standard method in most cases. When the false alarm rate is small, this bootstrap method seems to perform relatively well in terms of both the ARL_{avg} and the SRL_{avg}. This bootstrap method does not display the unusual variability that the subgroup bootstrap displayed when estimating extreme tails of skewed distributions. Perhaps this is due to the somewhat unreasonable assumption of Bajgier (1992) and Liu and Tang (1996) that the samples are taken from an in-control population. All simulations were conducted under this assumption. In using Bajgier's (1992) and Liu and Tang's (1996) bootstrap

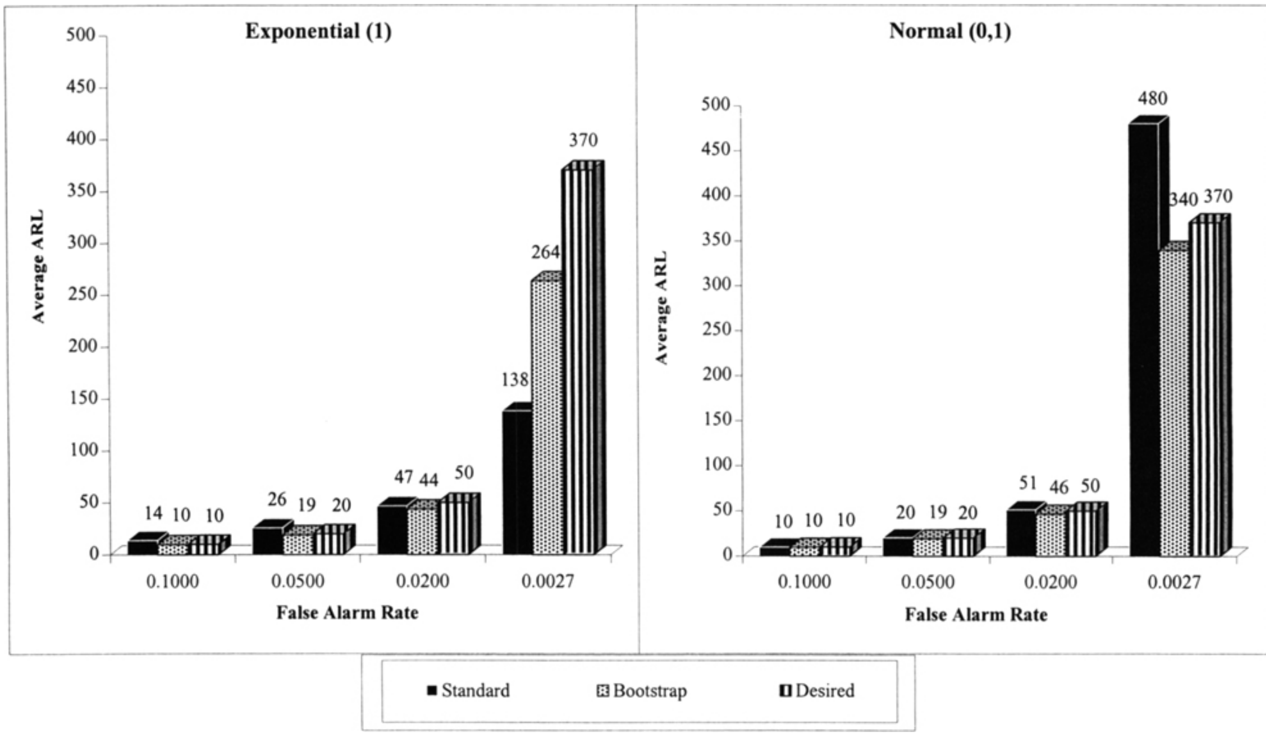

FIGURE 4. Comparison of Simulated ARL Between Bajgier's (1992) and Liu and Tang's (1996) Bootstrap Chart and Standard Control Chart.

control chart, however, the reader is cautioned to verify that the data come from populations with the same mean, that is, that the process is in statistical control.

Conclusion

The results of these simulation studies show that the bootstrap control charts do not perform substantially better than the standard method when the performance of the charts is evaluated in terms of the resulting in-control ARL. When estimating the tails of an extremely skewed distribution, the bootstrap techniques discussed here seem to produce estimates that are closer on average to the true quantile values than the standard Shewhart method. This performance benefit, however, does not translate into superior performance in terms of a predictable in-control ARL. The run length distributions of these bootstrap control charts are quite variable, especially when estimating the extreme tails of a skewed distribution. The quality practitioner who seeks a nonparametric control chart with a predictable in-control ARL should use these bootstrap control charts with caution.

References

BAJGIER, S. M. (1992). "The Use of Bootstrapping to Construct Limits on Control Charts". *Proceedings of the Decision Science Institute,* San Diego, CA, pp. 1611–1613.

DAVISON, A.; HINKLEY, D.; and SCHECTMAN, E. (1986). "Efficient Bootstrap Simulation". *Biometrika* 73, pp. 555–556.

DERMAN, C. and ROSS, S. (1995). "An Improved Estimator of σ in Quality Control". *Probability in the Engineering and Informational Sciences* 9, pp. 411–415.

EFRON, B. (1979). "Bootstrap Methods: Another Look at the Jackknife". *Annals of Statistics* 7, pp. 1–26.

EFRON, B. and GONG, G. (1983). "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation". *The American Statistician* 37, pp. 36–48.

GHOSH, B. K.; REYNOLDS, JR., M. R.; and HUI, Y. V. (1981). "Shewhart $\overline{X}$-Charts with Estimated Process Variance". *Communications in Statistics—Theory and Methods* 18, pp. 1797–1822.

GUNTER, B. (1991). "Bootstrapping: How to Make Something From Almost Nothing and Get Statistically Valid Answers, Part I". *Quality Progress* pp. 97–103.

GUNTER, B. (1992). "Bootstrapping: How to Make Something From Almost Nothing and Get Statistically Valid Answers, Part III". *Quality Progress* pp 119–122.

LIU, R. Y. and TANG, J. (1996). "Control Charts for Dependent and Independent Measurements Based on the Bootstrap". *Journal of the American Statistical Association* 91, pp. 1694–1700.

MOONEY, C. Z. and DUVAL, R. D. (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference.* SAGE Publications, Newbury Park, CA.

QUESENBERRY, C. P. (1993). "The Effect of Sample Size on Estimated Limits for $\overline{X}$ and X Control Charts". *Journal of Quality Technology* 25, pp. 237–247.

SEPPALA, T.; MOSKOWITZ, H.; PLANTE, R.; and TANG, J. (1995). "Statistical Process Control via the Subgroup Bootstrap". *Journal of Quality Technology* 27, pp. 139–153.

YOUNG, G. A. (1994). "Bootstrap: More than a Stab in the Dark?". *Statistical Science* 9, pp. 382-415.

Key Words: *Bootstrap Methods, Control Limits, Quality Control, Shewhart Control Charts, Statistical Process Control.*