Communications for Statistical Applications and Methods
2020, Vol. 27, No. 1, 65–77

https://doi.org/10.29220/CSAM.2020.27.1.065
Print ISSN 2287-7843 / Online ISSN 2383-4757

Geometric charts with bootstrap-based control limits using the Bayes estimator

Minji Kim[a], Jaeheon Lee[1,a]

[a]Department of Applied Statistics, Chung-Ang University, Korea

Abstract

Geometric charts are effective in monitoring the fraction nonconforming in high-quality processes. The in-control fraction nonconforming is unknown in most actual processes; therefore, it should be estimated using the Phase I sample. However, if the Phase I sample size is small the practitioner may not achieve the desired in-control performance because estimation errors can occur when the parameters are estimated. Therefore, in this paper, we adjust the control limits of geometric charts with the bootstrap algorithm to improve the in-control performance of charts with smaller sample sizes. The simulation results show that the adjustment with the bootstrap algorithm improves the in-control performance of geometric charts by controlling the probability that the in-control average run length has a value greater than the desired one. The out-of-control performance of geometric charts with adjusted limits is also discussed.

Keywords: Bayes estimator, bootstrap algorithm, control limits, geometric chart, statistical process control

1. Introduction

Control charts are highly effective for monitoring the quality of manufacturing processes. The basic assumption of the control chart is that in-control process parameters are known or can be accurately estimated; however, they should be estimated using the Phase I sample if the parameters are unknown. Then they will be used in Phase II to detect a process change. Jensen *et al*. (2006) and Psarakis *et al*. (2014) reviewed the performance of charts with estimated parameters.

Geometric charts are particularly useful for monitoring high-quality processes. There have been several studies on the performance and estimation effects of geometric charts for which parameters have been estimated. Yang *et al*. (2002) investigated the performance of geometric charts with estimated control limits and showed that the effect on the alarm probability can be significant even with sample sizes as large as 10,000. Tang and Cheong (2004) studied the effects of estimation error on geometric chart performance when a sequential sampling scheme was employed.

A widely used method to measure control chart performance is to use the average run length (ARL), where the run length is defined as the number of chart statistics plotted until the chart signals. ARL is constant under the known parameters assumption; however, this metric becomes a random variable when estimating in-control parameters and determining control limits. The control chart performance will vary among practitioners because it depends on the estimated parameters when the in-control parameters are estimated. This is because practitioners use different Phase I data sets, which

[1] Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: jaeheon@cau.ac.kr

Published 31 January 2020 / journal homepage: http://csam.or.kr

result in different parameter estimates, control limits, and ARL values. Therefore, charts are evaluated and the amount of Phase I data necessary for the desired chart performance is determined based on the expected value of the ARL (AARL) and the standard deviation of the ARL (SDARL). The SDARL metric accounts for the practitioner-to-practitioner chart variability, with lower values indicating less variation in the ARL values between practitioners. More details about the SDARL metric can be referred to Saleh *et al*. (2015).

Zhang *et al*. (2013) evaluated the in-control performance of the geometric chart using the SDARL, and showed that larger Phase I sample sizes were needed to obtain small SDARL values. They also recommended using Bayesian methods to address the lack of observable nonconforming items in the Phase I sample and to use practitioner knowledge about in-control parameter values.

Saleh *et al*. (2015) evaluated the in-control performance of an exponentially weighted moving average (EWMA) chart in terms of SDARL and percentiles of the ARL distribution when process parameters were estimated. They showed that the EWMA chart required a larger volume of Phase I data than recommended in previous studies to sufficiently reduce the variation in the chart performance. Because of practical limitations in the amount of Phase I data, they recommended a procedure based on the bootstrap approach proposed by Jones and Steiner (2012) and Gandy and Kvaløy (2013). The approach involves applying a bootstrap algorithm to the control chart to ensure that the probability that the in-control ARL value exceeds the target value is a constant. Since Saleh *et al*. (2015), many studies have used the bootstrap algorithm to adjust the control limits of other control charts. Faraz *et al*. (2015) proposed designing a method for the S^2-chart and investigated the effect from adjusting the control limits. Zhao and Driscoll (2016) adjusted the control limits of the c-chart using the bootstrap approach, resulting in improved in-control ARL performance. Faraz *et al*. (2017) recently applied the bootstrap algorithm to adjust the np-chart and show that in-control ARL values were greater than the desired value with a certain probability.

For high-quality processes, more accurate parameters can be estimated with a larger the sample size. However, this is extremely difficult in real processes. Therefore, in this paper, we apply the bootstrap algorithm to the geometric chart to accurately estimate parameters with small sample sizes. We use the Bayes estimator instead of the maximum likelihood estimator (MLE) when applying a bootstrap approach. This enables us to construct control limits even when nonconforming items are not observed in the Phase I sample. We also evaluate the in-control and the out-of-control performance of geometric charts with adjusted limits.

In Sections 2 and 3, we give an overview of the geometric chart with known parameter and unknown parameter. Section 4 describes the MLE, which is generally used to estimate the unknown parameter of a geometric chart, and the Bayes estimator, which complements the MLE, considering its limitations. In Sections 5 and 6, we present the in-control performance of the geometric chart with the estimated parameter and introduce the bootstrap method for obtaining adjusted control limits for geometric charts. In Section 7, a simulation study is performed to compare the performance of the geometric chart with and without bootstrap adjusted control limits. Finally, we present the conclusions in Section 8.

2. The geometric chart with known p_0

With the continuous advancements in manufacturing technology, many processes are now characterized by a very small proportion p of nonconforming items. For these high-quality processes, many control charting methods have been recommended for monitoring the proportion of nonconforming (Szarka and Woodall, 2011). The geometric chart is more effective than the traditional p and np charts

when monitoring high-quality processes. This chart is also frequently called the cumulative count of conforming chart in an statistical process control (SPC) literature, because it is based on the number of items sampled between pairs of nonconforming items.

Let Y_i be the number of conforming items between the $(i-1)^{th}$ and i^{th} nonconforming item with the in-control probability of a nonconforming item p_0. Then, Y_i is a geometric random variable with parameter p_0. Thus, the probability mass function of Y_i is given by the following:

$$g(y_i) = (1-p_0)^{y_i} p_0, \quad y_i = 0, 1, \ldots,$$

with $P(Y_i \geq y_i) = (1-p_0)^{y_i}$.

To establish a control chart, it is necessary to set lower control limits (LCL) and upper control limits (UCL). The geometric chart will give a signal at i if $Y_i \leq$ LCL or $Y_i \geq$ UCL, where LCL and UCL are integer values. Let $\alpha_{p_0,L}$ denote the false alarm rate for being below the lower control limit and $\alpha_{p_0,U}$ denote the false alarm rate for exceeding the upper control when p_0 is known. We cannot obtain probability control limits that satisfy the exact desired value for the false alarm rate, α because of the discrete nature of the geometric distribution. Instead, the probability control limits, which have a false alarm rate less than the desired false alarm rate α, can be obtained from the following equations:

$$\alpha_{p_0,L} = P(Y_i \leq \text{LCL} \mid p_0) = \sum_{y_i=0}^{\text{LCL}} (1-p_0)^{y_i} p_0 \leq \frac{\alpha}{2},$$

$$\alpha_{p_0,U} = P(Y_i \geq \text{UCL} \mid p_0) = \sum_{y_i=\text{UCL}}^{\infty} (1-p_0)^{y_i} p_0 \leq \frac{\alpha}{2},$$

which are equivalent to

$$1-(1-p_0)^{\text{LCL}+1} \leq \frac{\alpha}{2} \quad \text{and} \quad (1-p_0)^{\text{UCL}} \leq \frac{\alpha}{2}. \tag{2.1}$$

From (2.1), we can obtain the control limits for the geometric chart as:

$$\text{LCL} = \left\lfloor \frac{\ln(1-\alpha/2)}{\ln(1-p_0)} - 1 \right\rfloor \quad \text{and} \quad \text{UCL} = \left\lceil \frac{\ln(\alpha/2)}{\ln(1-p_0)} \right\rceil, \tag{2.2}$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x (called a floor function), and $\lceil x \rceil$ denotes the least integer greater than or equal to x (called a ceiling function).

3. The geometric chart with unknown p_0

The control limits in (2.2) are valid only when parameter p_0 is known. When p_0 is unknown, it must be estimated through the Phase I sample to set the control limits. Note that the estimated control limits can differ from the actual values when setting the control limit with a small sample. As parameter p_0 in (2.2) is replaced by $\hat{p}_0$, the control limits for the geometric chart when parameter p_0 is unknown are:

$$\widehat{\text{LCL}}(N) = \left\lfloor \frac{\ln(1-\alpha/2)}{\ln(1-\hat{p}_0)} - 1 \right\rfloor \quad \text{and} \quad \widehat{\text{UCL}}(N) = \left\lceil \frac{\ln(\alpha/2)}{\ln(1-\hat{p}_0)} \right\rceil, \tag{3.1}$$

where N is the number of nonconforming items among a total of m items sampled in Phase I.

In Phase II, if the fraction nonconforming is p, the alarm probability $\gamma(N)$ using the control limits in (3.1) can be expressed as:

$$\gamma(N) = P\left\{Y_i \leq \widehat{\mathrm{LCL}}(N) \mid N = n\right\} + P\left\{Y_i \geq \widehat{\mathrm{UCL}}(N) \mid N = n\right\},$$

where Y_i is a geometric random variable with parameter p, and N is a binomial random variable with parameters m and p_0.

To evaluate the performance of control charts with estimated limits, we use the average of ARL (AARL) and the SDARL. The run length (denoted by R) is the number of points plotted on the chart until an out-of-control signal is given. The conditional distribution of R given N is geometric with parameter $\gamma(N)$. Hence, the ARL is

$$\mathrm{ARL}(N) = \frac{1}{\gamma(N)}.$$

In addition, the AARL and the SDARL are observed to have the form:

$$\mathrm{AARL} = E_N\left[\frac{1}{\gamma(N)}\right] \quad \text{and} \quad \mathrm{SDARL} = \sqrt{E_N\left[\frac{1}{\gamma^2(N)}\right] - {E_N}^2\left[\frac{1}{\gamma(N)}\right]}, \tag{3.2}$$

respectively, where

$$E_N\left[\frac{1}{\gamma(N)}\right] = \sum_{n=1}^{m} \frac{1}{\gamma(n)}\binom{m}{n} p_0^n (1-p_0)^{m-n} + (1-p_0)^m$$

and

$$E_N\left[\frac{1}{\gamma^2(N)}\right] = \sum_{n=1}^{m} \frac{1}{\gamma^2(n)}\binom{m}{n} p_0^n (1-p_0)^{m-n} + (1-p_0)^m.$$

When $N = 0$ in above equations, we use the strategy used in Zhang *et al.* (2013) to signal at each nonconforming item, that is $\gamma(N = 0) = 1$.

4. Estimators of p_0

When the fraction of nonconforming p_0 is unknown, p_0 must be estimated through the Phase I sample to establish the control limits. Many studies have used the MLE as an estimator for p_0, which is:

$$\hat{p}_0 = \frac{N}{m}. \tag{4.1}$$

As in Zhang *et al.* (2013) and Lee *et al.* (2013), the problem with using the MLE is that control limits are not defined when nonconforming items are not observed in the Phase I sample of when $N = 0$. This is especially common in the high-quality processes that are the focus of this paper. To solve this problem, Zhang *et al.* (2013) used the Bayes estimator instead of the MLE. Other studies that also used the Bayes estimator for the fraction nonconforming are: Hong and Lee (2015), Zhang *et al.* (2017), and Han *et al.* (2018).

To perform the bootstrap algorithm in this paper, we use the Bayes estimator instead of the MLE, as the bootstrap algorithm cannot be carried out when nonconforming items are absent in the Phase I sample, which results in $\hat{p}_0 = 0$. We discuss this further in Section 6. In addition, the Bayesian approach is advantageous, since it incorporates prior information about the fraction nonconforming into the Bayes estimator through a prior distribution based on practitioner beliefs or experiences. However, it is updated to a posterior distribution because the prior distribution is set to the practitioner's prior knowledge and the data is observed.

The prior distribution for p_0 is usually assumed by a beta distribution with parameters a and b, Beta(a, b). When we use a beta prior distribution after observing the Phase I sample, our posterior distribution becomes a beta distribution Beta$(a+N, b+m-N)$. Thus, the mean of the prior distribution denoted by $\hat{p}_{0,p}$ is as:

$$\hat{p}_{0,p} = \frac{a}{a+b}, \tag{4.2}$$

and the Bayes estimator of p_0 is the mean of the posterior distribution, denoted by $\hat{p}_{0,B}$, which is

$$\hat{p}_{0,B} = \frac{N+a}{m+a+b}. \tag{4.3}$$

The beta distribution has various forms depending on parameters a and b: it is strictly decreasing when $a \leq 1$ and $b > 1$; U-shaped when $a < 1$ and $b < 1$; and unimodal when $a > 1$ and $b > 1$. Because p_0 is very small in this paper, we consider only 1 and 2 for the values of a to facilitate arriving at a low value for the prior mean.

Besides using the Bayes estimator of the in-control parameter, recent studies for the Bayesian approach in SPC are as follows. Pan and Rigdon (2012) used a Bayesian approach to select possible change point models for multivariate process. Tan and Shi (2012) also proposed Bayesian approach to identify the means that shifted and the direction of the shifts for multivariate charts. Apley (2012) proposed a Bayesian method for graphically monitoring process means. Kumar and Chakraborti (2017) proposed a Bayesian approach to establish control limits for control charts to monitor the times between events following an exponential distribution.

5. In-control performance of the geometric chart with the estimated parameter

When setting the control limits with the estimate parameter, we consider the in-control performance of the geometric charts using the AARL and SDARL to investigate problems. Table 1 gives the in-control AARL (AARL$_0$) and SDARL (SDARL$_0$) values of geometric charts for the m and p_0 values. The values of AARL$_0$ and SDARL$_0$ were computed using (3.2). For these results, we use the MLE for p_0 and assume the desired in-control ARL (ARL$_0$) is 200. The last row in the table, $m = \infty$, refers to the case in which the in-control process parameter p_0 is known. We note that, the ARL$_0$ values for each p_0 can deviate from 200 (more precisely, values are greater than 200) since we use the control limits in (3.1) in this paper.

For fixed p_0 values, as the sample size m increases, the AARL$_0$ values do not approximate the desired ARL$_0$ values even when m is sufficiently large. For example, when $p_0 = 0.0005$ and $m =$ 2,000,000, the AARL value is 209.8, which does not come close to the value for the known-parameter case. As m increases, the SDARL$_0$ values tend to decrease; however, the convergence rate is very slow, especially when p_0 is very small ($p_0 = 0.0001$). Zhang *et al.* (2014) also suggested that SDARL$_0$ values within 10% of ARL$_0$ values may be adequate for practitioners to have confidence about the

Table 1: Values of $AARL_0$ (upper entry) and $SDARL_0$ (lower entry)

m	p_0		
	0.0001	0.0005	0.001
10,000	77.7	163.6	195.8
	93.6	88.3	91.5
20,000	119.6	183.7	214.6
	88.7	81.3	88.9
50,000	160.9	203.3	223.2
	85.9	74.1	74.2
100,000	179.8	207.5	225.5
	79.0	61.0	62.1
200,000	191.2	209.4	226.0
	70.0	47.8	49.6
2,000,000	201.6	209.8	222.8
	33.3	13.6	16.5
∞	200.1	200.1	222.3
	0.0	0.0	0.0

$AARL_0$ = the in-control AARL (the expected value of the average run length); $SDARL_0$ = the in-control SDARL (the standard deviation of the average run length).

Table 2: The percentage of geometric charts with ARL_0 values below the targeted ARL_0

m	p_0		
	0.0001	0.0005	0.001
10,000	64.01	51.10	48.23
20,000	46.58	44.33	44.81
30,000	55.43	43.53	45.31
40,000	46.71	44.21	45.60
50,000	51.11	44.50	45.98
60,000	45.56	45.25	46.19
70,000	48.37	38.97	46.69
80,000	44.71	39.35	46.94
90,000	46.66	40.11	47.37
100,000	44.33	40.33	47.45

ARL_0 = the in-control ARL (average run length).

predictability of the Phase I performance, although it may still reflect significant variation. Table 1 shows that a smaller p_0 value requires a larger sample size to satisfy this condition. For example, the value of m that is required to satisfy the condition is 2,000,000 when values of p_0 are 0.0005 and 0.001; and an m of more than 2,000,000 is required when p_0 is 0.0001. In practice, it is desirable to select a sample size that is large enough to obtain acceptable $SDARL_0$ values. However, a very large sample size is often not realistic because of time and cost.

We also perform a simulation study to investigate the variability of the in-control ARL values when using the estimated control limits. In this study, for each Phase I sample of size m and p_0, an ARL_0 value is obtained that repeats 10,000 times. The desired ARL_0 is assumed to be 200, and the MLE is used to estimate p_0. Table 2 shows the percentage of geometric charts with ARL_0 values below the targeted ARL_0, which is the ARL value for the case in which the in-control process parameter p_0 is known. As in Table 1, the targeted ARL_0 values for $p_0 = 0.0001, 0.0005$, and 0.001 are 200.1, 200.1, and 222.3, respectively. Table 2 shows that in all cases except when $p_0 = 0.0005$ and $m = 70{,}000$ and 80,000, more than 40% of geometric charts have ARL_0 values below the targeted ARL_0. This means that more than 40% of practitioners would have a false alarm rate greater than the desired value. This is a serious problem that requires other ways to construct control charts with estimated parameters. In

the next section, we introduce the bootstrap method to adjust the control limits of geometric charts.

6. The bootstrap approach

To overcome the problem of low ARL_0 values when using estimated parameters, we apply the bootstrap algorithm proposed by Jones and Steiner (2012) and Gandy and Kvaløy (2013). This algorithm can adjust the geometric control limits so that the in-control ARL values are equal to or greater than the targeted ARL_0 value, A_0, with at least a certain probability, say $1-\rho$. Therefore, it is expressed as:

$$P\Big\{\mathrm{ARL}_0 > \mathrm{A}_0\Big\} \geq 1-\rho. \tag{6.1}$$

In this paper, we set $\rho = 0.1$ so that more than 90% of the values are larger than the targeted ARL_0 value.

Saleh *et al*. (2015) and Faraz *et al*. (2015) recently adjusted the control limits of the EWMA chart and S^2- chart, respectively, using the bootstrap algorithm. Zhao and Driscoll (2016) and Faraz *et al*. (2017) also adjusted the control limits of the *c*-chart and the *np*-chart. The bootstrap algorithm for adjusting the control limits of the geometric chart is summarized as:

1. Set the prior distribution for p_0 as Beta(a, b), and from the Phase I sample, estimate the fraction defective p_0 as $\hat{p}_0 = (N+a)/(m+a+b)$ in (4.3).

2. Generate B bootstrap samples $N_1^*, N_2^*, \ldots, N_B^*$ from $B(m, \hat{p}_0)$, and calculate $\hat{p}_{0,i}^* = (N_i^*+a)/(m+a+b)$ for $i = 1, 2, \ldots, B$, where B is a large number, for example $B = 1{,}000$.

3. Find the ρ^{th} percentiles of the $\hat{p}_{0,i}^*$ (say, $\hat{p}_L^*$) and find $(1-\rho)^{th}$ percentiles of the $\hat{p}_{0,i}^*$ (say, $\hat{p}_U^*$).

4. Calculate the adjusted control limits as

$$\widehat{\mathrm{LCL}}^* = \left\lfloor \frac{\ln(1-\alpha/2)}{\ln(1-\hat{p}_U^*)} - 1 \right\rfloor \quad \text{and} \quad \widehat{\mathrm{UCL}}^* = \left\lceil \frac{\ln(\alpha/2)}{\ln(1-\hat{p}_L^*)} \right\rceil.$$

   Therefore, the probability of observation out-of-control limits $\alpha_{\hat{p}}^*$ and the average run length ARL^* are computed as

$$\alpha_{\hat{p}}^* = (1-p)^{\widehat{\mathrm{UCL}}^*} - (1-p)^{\widehat{\mathrm{LCL}}^*+1} + 1 \quad \text{and} \quad \mathrm{ARL}^* = \frac{1}{\alpha_{\hat{p}}^*},$$

   respectively.

Note that if $N = 0$ when using the MLE for p_0 in step 1, we cannot generate the valid bootstrap samples in step 2. However, this problem does not occur when using the Bayes estimator.

7. The performance of the geometric chart with unadjusted and adjusted control limits

In this section, a simulation study is performed to compare geometric chart performance with and without bootstrap adjusted control limits. First, we compare the in-control performance for each p_0 and m. As a result of the simulation repeating 10,000 times without adjustment (denoted by "Unadj.") and with adjustment (denoted by "Adj."), we have 10,000 ARL_0 values for "Unadj." and "Adj."

Table 3: Comparison of the percentage of geometric charts with in-control ARL values below the targeted ARL_{p_0} for charts with unadjusted and adjusted control limits

	Mean			m			
				10,000	20,000	50,000	100,000
$p_0 = 0.0001$ ($ARL_{p_0} = 200.12$)	Unadj.			64.01	46.58	51.11	44.33
	Adj.	0.00005	Beta(1, 19999)	0.00	0.00	0.65	0.65
			Beta(2, 39998)	0.00	0.00	0.27	0.02
		0.0001	Beta(1, 9999)	0.00	0.35	1.95	4.17
			Beta(2, 19998)	0.00	0.02	0.97	2.87
		0.0002	Beta(1, 4999)	8.16	12.74	12.95	5.26
			Beta(2, 9998)	7.70	12.99	13.29	13.52
$p_0 = 0.0005$ ($ARL_{p_0} = 200.10$)	Unadj.			51.10	44.33	44.50	40.33
	Adj.	0.00025	Beta(1, 3999)	0.77	0.29	1.53	1.40
			Beta(2, 7998)	0.22	0.08	0.25	0.63
		0.0005	Beta(1, 1999)	1.99	4.12	3.56	2.98
			Beta(2, 3998)	0.97	1.17	3.30	2.85
		0.001	Beta(1, 999)	9.65	6.20	4.92	4.70
			Beta(2, 1998)	12.91	13.41	8.13	5.15
$p_0 = 0.001$ ($ARL_{p_0} = 222.34$)	Unadj.			48.23	44.81	45.95	47.45
	Adj.	0.0005	Beta(1, 1999)	0.95	1.15	1.49	1.40
			Beta(2, 3998)	0.23	0.21	0.56	0.82
		0.001	Beta(1, 999)	4.17	3.56	3.12	2.22
			Beta(2, 1998)	2.89	2.44	2.40	2.35
		0.002	Beta(1, 499)	5.89	7.28	4.69	2.98
			Beta(2, 998)	13.31	7.70	5.29	3.84

ARL = average run length; Unadj. = unadjusted limits; Adj. = adjusted limits.

respectively. We use $B = 1{,}000$ for the bootstrap algorithm. We set the targeted ARL_0 value as $A_0 = 200$, but when p_0 is 0.0001, 0.0005, and 0.001, the true ARL_0 values (denoted by ARL_{p_0}) are 200.12, 200.10, and 222.34, respectively.

Table 3 shows the proportion of the 10,000 ARL_0 values that are less than the ARL_{p_0} values. The purpose of the bootstrap adjustment is to achieve good performance even with a small amount of Phase I data, but since we consider a very small fraction nonconforming in the high quality process, the simulation should be performed with a somewhat larger m to attain a degree of accuracy. Therefore, we set m as 10,000, 20,000, 50,000, and 100,000. We note that the MLE in (4.1) is used in the case of "Unadj.", whereas the Bayes estimator in (4.3) is used in the case of "Adj.". It is clear that the choice of a prior distribution should incorporate the practitioner's knowledge of p_0 as accurately as possible. In this simulation, combinations of $(1, b)$ and $(2, b)$ for parameters of the beta prior distribution, are used to ensure that the prior mean is equal to p_0. For example, when $p_0 = 0.0001$, if the parameter a is set to 1, b is determined at 9999 because the mean of the prior distribution in (4.2) is the same as p_0. We also simulated a combination of cases that the prior mean is $(1/2)p_0$ (that is, the prior mean is less than p_0) and $2p_0$ (that is, the prior mean is greater than p_0). We will investigate the effect of the parameters of the prior distribution in a future study.

Table 3 shows that, in the case of "Unadj.", the percentages of geometric charts that result in ARL_0 values below ARL_{p_0} are all over 40%. However, for "Adj.", in the prior distribution except for the prior mean greater than p_0, the percentage of values is less than 10%. The results show that the desired in-control performance in (6.1) is well satisfied; however, there are some cases that the percentage value is greater than 10% when the prior mean is greater than p_0. This result indicates that the effect of parameter estimation increases when we set the parameters of the prior distribution to have an overestimated prior mean.

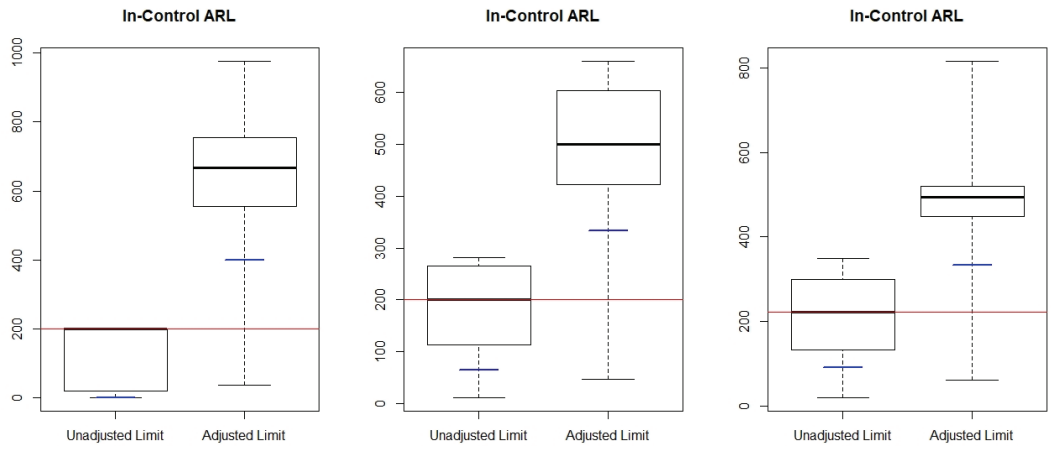

Figure 1: *Boxplots of in-control ARL with unadjusted and adjusted control limits when* $m = 20{,}000$, *(a)* $p_0 = 0.0001, ARL_{p_0} = 200.12$, *Beta(1, 9999) for the prior distribution; (b)* $p_0 = 0.0005, ARL_{p_0} = 200.10$, *Beta(1, 1999) for the prior distribution; (c)* $p_0 = 0.001, ARL_{p_0} = 222.34$, *Beta(1, 999). The boxplots show the 0th, 10th, 25th, 50th, 75th, and 100th percentiles. ARL = average run length.*

To compare the in-control performance of geometric charts with unadjusted and adjusted control limits, boxplots of the in-control ARL for each $p_0 = 0.0001, 0.0005$, and 0.001 at $m = 20{,}000$ are illustrated in Figure 1. The reference line (the red line) in Figure 1 corresponds to the targeted ARL_{p_0} values for each p_0. Parameter a of the prior distribution is set to 1. When comparing "Unadj." and "Adj." for each p_0, the plots show that all the 10th percentiles (the blue line) of "Unadj." are below the reference line, while all of "Adj." are above the reference line. Table 1 shows that the Phase I sample of $m = 20{,}000$ without adjustment does not provide an acceptable in-control performance; however, an adjustment using the bootstrap approach improves in-control ARL performance by reducing the percentage of charts with ARL_0 values below the targeted ARL_{p_0} value.

An interesting finding in previous studies about adjustment using the bootstrap approach is that the control charts with adjusted control limits have more variable in-control ARL distribution than control charts based on unadjusted control limits. Figure 1 shows that the in-control ARL distribution with adjusted control limits also has a larger variability than the distribution with unadjusted control limits. Saleh *et al*. (2015) therefore maintained that control charts based on adjusted control limits result in an acceptable increased variability in the in-control ARL distribution as long as the out-of-control performance is not considerably worse than that based on unadjusted control limits. We note that the purpose of adjustment using the bootstrap approach is to improve in-control performance.

Table 4 shows a comparison of the out-of-control performance of geometric charts with unadjusted and adjusted control limits. Similar to the in-control performance, the simulation is repeated 10,000 times, and the number of bootstrap samples B is set to 1,000. Moreover, the MLE is used for unadjusted control limits, and the Bayes estimator is used for adjusted control limits. The out-of-control ARL is used as a metric for confirming the out-of-control performance for the case of shift from $p_0 = 0.0001, 0.0005$, and 0.001 to each p_1 when $m = 10{,}000, 20{,}000$, and $50{,}000$. The prior distribution for each p_0 is used as a beta distribution with parameter a of 1. In Table 4, as the shift size increases in all p_0 and m cases, the out-of-control ARL decreases for control charts with both unadjusted and adjusted control limits. A larger shift size allows for a faster detection and prediction of the

Table 4: The out-of-control ARL performance when $p_0 = 0.0001, 0.0005$, and 0.0001

	p_1	m					
		10,000		20,000		50,000	
		Unadj.	Adj.	Unadj.	Adj.	Unadj.	Adj.
$p_0 = 0.0001$ with Beta (1, 9999)	0.0002	114.06	314.99	153.91	343.85	187.19	308.44
	0.0003	113.25	210.18	129.38	229.60	136.97	206.16
	0.0004	97.35	157.75	101.91	172.32	103.49	154.74
	0.0005	81.62	126.30	82.32	137.95	82.92	123.88
$p_0 = 0.0005$ with Beta (1, 1999)	0.001	194.65	344.68	212.30	323.37	220.43	287.72
	0.0015	144.96	230.63	148.16	216.14	148.76	192.17
	0.002	109.73	173.06	111.36	162.19	111.68	144.22
	0.0025	87.92	138.51	89.17	129.82	89.42	115.45
	0.003	73.33	115.48	74.37	108.24	74.58	96.26
$p_0 = 0.001$ with Beta (1, 999)	0.002	253.01	411.91	252.71	379.29	249.84	330.89
	0.0025	212.51	329.96	206.54	303.85	201.29	265.12
	0.003	178.48	275.01	172.56	253.25	167.88	220.98
	0.0035	153.19	235.74	147.98	217.09	143.94	189.44
	0.004	134.09	206.28	129.52	189.97	125.98	165.78
	0.005	107.32	165.04	103.66	152.00	100.83	132.66

ARL = average run length; Unadj. = unadjusted limits; Adj. = adjusted limits.

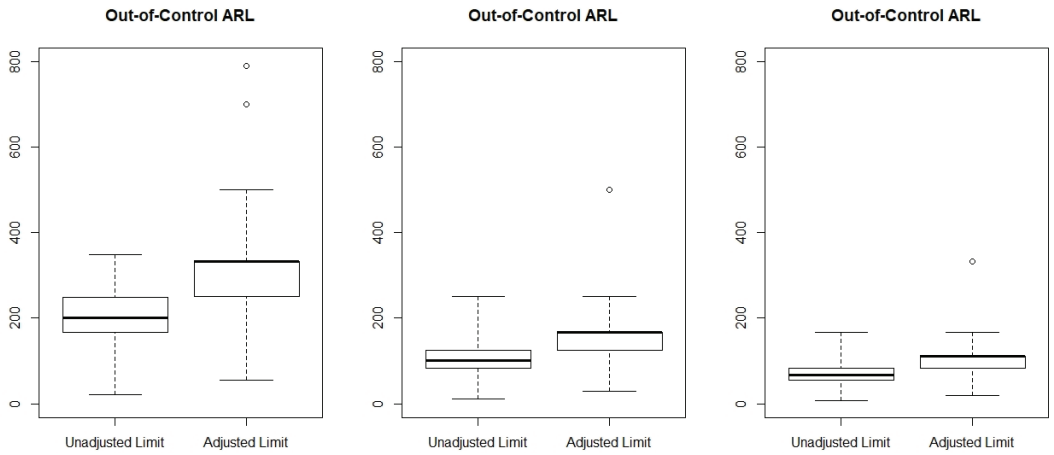

Figure 2: *Boxplots of the out-of-control ARL with unadjusted and adjusted control limits when m = 20,000, $p_0 = 0.0005$, using Beta(1, 1999), and (a) $p_1 = 0.001$, (b) $p_1 = 0.002$, (c) $p_1 = 0.003$. ARL = average run length.*

out-of-control state. Further, in this case, it is confirmed that the difference between the out-of-control ARL of "Unadj." and "Adj." is reduced.

Boxplots of the out-of-control ARL in Figures 2 and 3 confirm the out-of-control ARL performance. To visualize the effect of the shift size, Figure 2 illustrates shifts from $p_0 = 0.0005$ to $p_1 = 0.001, 0.002$, and 0.003 when $m = 20{,}000$. Also, to visualize the effect of the Phase I sample size, in Figure 3, we set $m = 10{,}000, 20{,}000$, and $50{,}000$ when $p_0 = 0.0005$ shifts to $p_1 = 0.001$. Similar to the pattern in Table 4, as the shift size increases from (a) to (c) in Figure 2, the overall out-of-control ARL values decrease, and the differences in out-of-control ARL values based on "Unadj." and "Adj." also decrease. We can see that the variability (including the outliers) for the control limits with adjustment is greater than that without adjustment; in addition, variability decreases as the shift

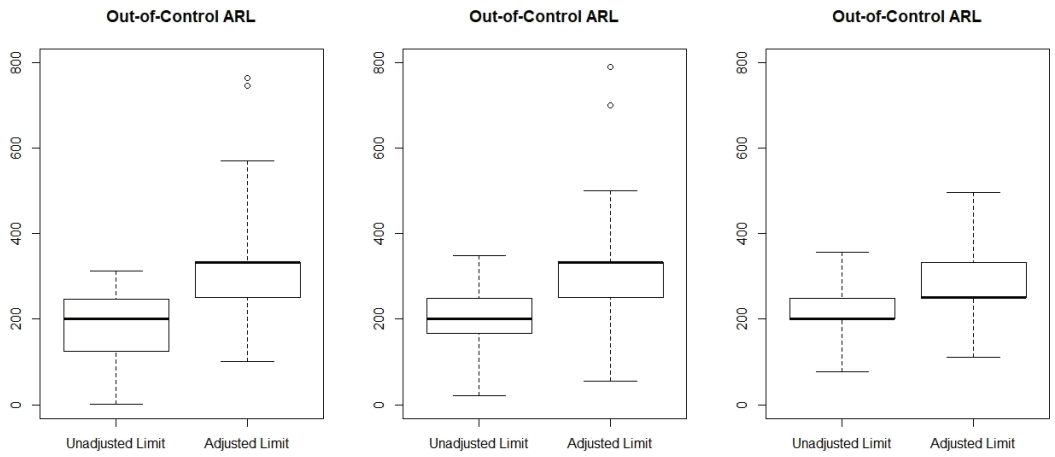

Figure 3: *Boxplots of the out-of-control ARL with unadjusted and adjusted control limits when* $p_0 = 0.0005$, $p_1 = 0.001$, *using Beta(1, 1999), and (a)* $m = 10{,}000$, *(b)* $m = 20{,}000$, *(c)* $m = 50{,}000$. *ARL = average run length.*

Table 5: The mode of adjusted control limits to guarantee that $P(\text{ARL}_0 > 200) \geq 0.9$

p_0	LCL	UCL	m	mode of $\widehat{\text{LCL}}^*$	mode of $\widehat{\text{UCL}}^*$
0.0001 with Beta (1, 9999)	24	59912	10,000	15(40.6)	119827(98.0)
			20,000	17(27.4)	179741(66.8)
			50,000	17(17.2)	119827(34.1)
0.0005 with Beta (1, 1999)	4	11980	10,000	2(49.5)	23963(34.1)
			20,000	2(50.8)	21966(20.4)
			50,000	2(46.5)	15575(9.4)
0.001 with Beta (1, 999)	1	5989	10,000	0(66.4)	10982(20.4)
			20,000	0(52.7)	8386(11.0)
			50,000	1(66.6)	7637(6.1)

ARL = average run length; LCL = lower control limits; UCL = upper control limits.

size increases. In Figure 3, as m increases, the difference between the out-of-control ARL values with unadjusted and adjusted control limits decreases; in addition, the variability of the out-of-control ARL values also decreases.

Table 5 shows the mode of the control limits obtained by repeating the bootstrap algorithm 10,000 times with the number of bootstraps $B = 1{,}000$ so that practitioners can refer to the following table without performing the bootstrap algorithm according to each p_0 and m. In the table, values of LCL and UCL denote the unadjusted control limits, and values in parentheses denote the percentage of occurrence at the mode. We hope that practitioners can more efficiently use geometric charts (with fewer samples) by referring to Table 5.

8. Conclusion

In a high-quality process in which defects are rarely observed, the geometric chart can quickly detect process deterioration by observing the number of conforming items between the two nonconforming items. Since the in-control fraction defective p_0 is not known in most actual processes, the fraction defective should be estimated with the Phase I sample. If the sample size is insufficient, estimation

errors occur when the parameter is estimated; therefore, a practitioner may not achieve the desired in-control performance. However, due to time and cost limitations, it may be difficult to use sufficient data in an actual process. In this paper, we suggest an adjusting method for control limits of a geometric chart using the bootstrap algorithm proposed by Jones and Steiner (2012) and Gandy and Kvaløy (2013) to improve the performance of a geometric chart with fewer Phase I data.

Many studies have used the MLE to estimate fraction defective p_0. However, from Zhang *et al*. (2013) and Lee *et al*. (2013), it is understood that when using the MLE, control limits cannot be defined when nonconforming items are not observed in the Phase I sample. This problem occurs frequently when estimating the fraction defective p_0 as the MLE since the geometric chart is often used in high-quality processes. Therefore, we have suggested an adjustment method based on the Bayes estimator.

From the simulation results comparing the performance of the geometric chart with unadjusted and adjusted control limits, we confirmed that using adjusted control limits improved the in-control ARL performance, as the percentage of geometric charts with in-control ARL values below the targeted ARL value was achieved above a certain probability. It was also observed that when using adjusted control limits, the ARL distribution had a larger variability than with unadjusted control limits and that the adjustment worsened the out-of-control performance. However, there was no large difference between instances with and without adjustment.

In conclusion, we recommend using adjusted control limits with the bootstrap algorithm when using a geometric chart in a high-quality process with a relatively small Phase I sample size. We hope that the adjustment method for control limits using the bootstrap algorithm is applicable to other types of control charts in addition to the geometric chart of this paper.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03029035).

References

Apley DW (2012). Posterior distribution charts: A Bayesian approach for graphically exploring a Process Mean, *Technometrics*, **54**, 279–293.

Faraz A, Heuchenne C, and Saniga E (2017). The np chart with guaranteed in-control average run lengths, *Quality and Reliability Engineering International*, **33**, 1057–1066.

Faraz A, Woodall WH, and Heuchenne C (2015). Guaranteed conditional performance of the S^2 control chart with estimated parameters, *International Journal of Production Research*, **53**, 4405–4413.

Gandy A and Kvaløy JT (2013). Guaranteed conditional performance of control charts via bootstrap methods, *Scandinavian Journal of Statistics*, **40**, 647–668.

Han SW, Lee J, and Park J (2018). A Bernoulli GLR chart based on Bayes estimator, *Journal of the Korean Data & Information Science Society*, **29**, 37–47.

Hong H and Lee J (2015). Comparisons of the performance with Bayes estimator and MLE for control charts based on geometric distribution, *The Korean Journal of Applied Statistics*, **28**, 907–920.

Jensen WA, Jones-Farmer LA, Champ CW, and Woodall WH (2006). Effects of parameter estimation on control chart properties: a literature review, *Journal of Quality Technology*, **38**, 349–364.

Jones MA and Steiner SH (2012). Assessing the effect of estimation error on the risk-adjusted CUSUM chart performance, *International Journal for Quality in Health Care*, **24**, 176–181.

Kumar N and Chakraborti S (2017). Bayesian monitoring of times between events: The Shewhart t_r−chart, *Journal of Quality Technology*, **49**, 136–154.

Lee J, Wang N, Xu L, Schuh A, and Woodall WH (2013). The effect of parameter estimation on upper-sided Bernoulli cumulative sum charts, *Quality and Reliability Engineering International*, **29**, 639–651.

Pan R and Rigdon SE (2012). A Bayesian approach to change point estimation in multivariate SPC, *Journal of Quality Technology*, **44**, 231–248.

Psarakis S, Vyniou AK, and Castagliola P (2014). Some recent developments on the effects of parameter estimation on control charts, *Quality and Reliability Engineering International*, **30**, 1113–1129.

Saleh NA, Mahmoud MA, Jones-Farmer LA, Zwetsloot IM, and Woodall WH (2015). Another look at the EWMA control chart with estimated parameters, *Journal of Quality Technology*, **47**, 363–382.

Szarka JL III and Woodall WH (2011). A review and perspective on surveillance of high quality Bernoulli processes, *Quality and Reliability Engineering International*, **27**, 735–752.

Tan MHY and Shi J (2012). A Bayesian approach for interpreting mean shifts in multivariate quality control, *Technometrics*, **54**, 294–307.

Tang LC and Cheong WT (2004). Cumulative conformance count chart with sequentially updated parameters, *IIE Transactions*, **36**, 841–853.

Yang Z, Xie M, Kuralmani V, and Tsui K-L (2002). On the performance of geometric charts with estimated control limits, *Journal of Quality Technology*, **34**, 448–458.

Zhang M, Hou X, He Z, and Xu Y (2017). Performance comparison for the CRL control charts with estimated parameters for high-quality processes, *Quality Technology & Quantitative Management*, **14**, 31–43.

Zhang M, Megahed FM, and Woodall WH (2014). Exponential CUSUM charts with estimated control limits, *Quality and Reliability Engineering International*, **30**, 275–286.

Zhang M, Peng Y, Schuh A, Megahed FM, and Woodall WH (2013). Geometric charts with estimated control limits, *Quality and Reliability Engineering International*, **29**, 209–223.

Zhao MJ and Driscoll AR (2016). The c-chart with bootstrap adjusted control limits to improve conditional performance, *Quality and Reliability Engineering International*, **32**, 2871–2881.

Received July 27, 2019; Revised October 17, 2019; Accepted November 9, 2019