

Adjustment of Control Limits for Geometric Charts

Byung Jun Kim^a, Jaeheon Lee^{1,b}

^aDepartment of Statistics, Virginia Tech, USA;

^bDepartment of Applied Statistics, Chung-Ang University, Korea

Abstract

The geometric chart has proven more effective than Shewhart p or np charts to monitor the proportion nonconforming in high-quality processes. Implementing a geometric chart commonly requires the assumption that the in-control proportion nonconforming is known or accurately estimated. However, accurate parameter estimation is very difficult and may require a larger sample size than that available in practice in high-quality process where the proportion of nonconforming items is very small. Thus, the error in the parameter estimation increases and may lead to deterioration in the performance of the control chart if a sample size is inadequate. We suggest adjusting the control limits in order to improve the performance when a sample size is insufficient to estimate the parameter. We propose a linear function for the adjustment constant, which is a function of the sample size, the number of nonconforming items in a sample, and the false alarm rate. We also compare the performance of the geometric charts without and with adjustment using the expected value of the average run length (ARL) and the standard deviation of the ARL (SDARL).

Keywords: average run length (ARL), control limits, geometric chart, high-quality process, Phase I, statistical process control

1. Introduction

Control charts have been primary and effective tools in statistical process control (SPC) and improvement. Generally, there are two phases in control charts; the first phase (Phase I) parameters of the process in the in-control state are estimated to construct the control limits. In the second phase (Phase II) the control limits established in Phase I are utilized to monitor the process continuously to detect parameter shifts from the in-control state.

In most SPC, implementing a control chart is based on the assumption that in-control parameters are known or accurately estimated. However, the error of parameter estimation may exist in practice and result in wrong conclusions by practitioners because the performance in Phase II relies on control limits established in Phase I. In high-quality processes, there are some difficulties to accurately estimate the process parameter (such as the proportion nonconforming) because nonconforming items are rare. Under such circumstances, an accurate estimation may need a very large sample size which is rarely attainable. Thus, it is important for practitioners to understand the effect of parameter estimation when using a control chart. Jensen *et al.* (2006) and Jones-Farmer *et al.* (2014) presented review papers on the effect of estimation error on control chart performance.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2054200).

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 30 September 2015 / journal homepage: <http://csam.or.kr>

© 2015 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

For high-quality processes, the geometric chart has been proved to be more useful than Shewhart p or np charts (Woodall, 1997; Xie and Goh, 1997; Szarka and Woodall, 2011). In the geometric chart, the count of conforming items between two consecutive nonconforming items is used to monitor the process. Therefore, the geometric chart does not require aggregating the data overtime unlike p or np charts. The problem of parameter estimation still remains despite the advantage of the geometric chart. Other studies on control charts for monitoring high-quality processes can be found in Goh (1987), Goh and Xie (1994), Nelson (1994), McCool and Joyner-Motley (1998), Chang and Gan (2001), and Steiner and Mackay (2004).

Yang *et al.* (2002) and Tang and Cheong (2004) discussed the effects of estimation error on the performance of geometric charts depending on each combination of the sample size in Phase I and the in-control proportion nonconforming. They used some metrics to evaluate the performance of the charts. Zhang *et al.* (2013) recommended a Bayes estimator for the in-control proportion nonconforming to avoid estimation problems when no nonconforming items are observed in the Phase I sample. They also considered the standard deviation of the average run length (SDARL) in order to check the robustness of average run length (ARL) in the geometric charts. Lee *et al.* (2013) studied the effects of estimation error on the Bernoulli cumulative sum (CUSUM) chart by using the same metric proposed by Zhang *et al.* (2013).

From previous studies on the geometric chart, we see that the actual ARL can deviate significantly from the desired value when the in-control proportion nonconforming p_0 is estimated from the Phase I sample, especially when p_0 is small. In this paper, we focus on relieving the estimation error effects on the geometric charts. In Section 2, we provide some background of the geometric chart and some metrics discussed in the works of Yang *et al.* (2002) and Zhang *et al.* (2013). We propose the method to adjust the control limits in the geometric chart in Section 3. We then compare the chart performance without and with adjustment of the control limits. Lastly, our conclusions are provided in Section 4.

2. Geometric Chart

2.1. Geometric chart with known p_0

Let Y_i be the count of conforming items between $(i - 1)^{th}$ and i^{th} nonconforming items from a process with the in-control probability of a nonconforming item p_0 . Assume that the process produces a sequence of independent Bernoulli items inspected without error and assume that the proportion nonconforming p_0 is known. Then, Y_i is a geometric random variable with parameter p_0 and the probability mass function of Y_i is

$$g(y_i) = (1 - p_0)^{y_i} p_0,$$

where $y_i = 0, 1, 2, \dots$

We let the lower-sided and the upper-sided control limits of the geometric chart be LCL and UCL, respectively. The geometric chart will signal if $Y_i < \text{LCL}$ or $Y_i > \text{UCL}$. Suppose the false alarm rate is α , then LCL and UCL must satisfy

$$\sum_{y_i=0}^{\text{LCL}-1} (1 - p_0)^{y_i} p_0 = \frac{\alpha}{2} \quad \text{and} \quad \sum_{y_i=\text{UCL}+1}^{\infty} (1 - p_0)^{y_i} p_0 = \frac{\alpha}{2}.$$

Thus, the control limits of the geometric charts can be expressed as

$$\text{LCL} = \frac{\ln(1 - \alpha/2)}{\ln(1 - p_0)}, \quad (2.1)$$

$$\text{UCL} = \frac{\ln(\alpha/2)}{\ln(1 - p_0)} - 1. \quad (2.2)$$

Yang *et al.* (2002) explained that there exists the difference between the desired false alarm rate and the actual one from the chart because of the discrete characteristic of the geometric distribution. They also stated that the difference is negligible unless p_0 is relatively large.

2.2. Geometric chart with unknown p_0

When p_0 is unknown, we need to establish the chart after estimating p_0 by inspecting a predetermined number of items in Phase I. In the practice, the estimated value of the parameter may be different with the true value due to the estimation error, especially when using the Phase I sample of a small size.

To estimate p_0 from an initial sample, we usually use the maximum likelihood estimator (MLE) which is

$$\hat{p}_0 = \frac{N}{m},$$

where N is the number of nonconforming items among a total of m items sampled from the Bernoulli process. Thus, N is a random variable of binomial distribution with parameters m and p_0 .

When we replace p_0 in (2.1) and (2.2) by \hat{p}_0 , the estimated control limits for the geometric chart are obtained as

$$\widehat{\text{LCL}}(N) = \frac{\ln(1 - \alpha/2)}{\ln(1 - N/m)}, \quad (2.3)$$

$$\widehat{\text{UCL}}(N) = \frac{\ln(\alpha/2)}{\ln(1 - N/m)} - 1. \quad (2.4)$$

2.3. Alarm rate when using \hat{p}_0

In this subsection, we discuss the alarm rate considered in Yang *et al.* (2002) and Zhang *et al.* (2013). In the work of Yang *et al.* (2002), they let X_i be the number of conforming items between two consecutive nonconforming items in Phase II, with the probability of a nonconforming item likely shifted from p_0 to p . They defined the event B_i as

$$B_i = \{X_i < \widehat{\text{LCL}}(N) \text{ or } X_i > \widehat{\text{UCL}}(N)\}.$$

Then, $P(B_i)$ is the alarm rate (AR) and it can be expressed as

$$P(B_i) = \sum_{n=0}^m P(B_i | N = n) P(N = n) = \sum_{n=0}^m P(B_i | N = n) \binom{m}{n} p_0^n (1 - p_0)^{m-n}, \quad (2.5)$$

where

$$\begin{aligned} P(B_i | N = n) &= P\{X_i < \widehat{\text{LCL}}(N) | N = n\} + P\{X_i > \widehat{\text{UCL}}(N) | N = n\} \\ &= 1 - (1 - p)^{\ln(1 - \frac{\alpha}{2}) / \ln(1 - \frac{n}{m})} + (1 - p)^{\ln(\frac{\alpha}{2}) / \ln(1 - \frac{n}{m})}. \end{aligned}$$

Note that $P(B_i)$ becomes the false alarm rate (FAR) when $p = p_0$. Yang *et al.* (2002) investigated how the values of FAR differ for various combinations of m and p_0 . From the investigation, they stated that the actual FAR can deviate significantly from the desired value of FAR when p_0 is estimated from m sampled items, and the amount of the deviation is considerably reduced as the sample size m increases. They concluded that the effect on the FAR can be significant even with the large sample size that we think is adequate, especially when p_0 is very small.

Equation (2.5), however, does not reflect the case when nonconforming items are unobserved in the Phase I sample. When $N = 0$, the control limits of (2.3) and (2.4) cannot be defined. To complement this problem, Zhang *et al.* (2013) modified (2.5) as

$$P(B_i) = \sum_{n=1}^m P(B_i | N = n) \binom{m}{n} p_0^n (1 - p_0)^{m-n} + (1 - p_0)^m. \quad (2.6)$$

In (2.6), they put $P(B_i | N = 0) = 1$. That is, we give a signal when a nonconforming item is observed because $\hat{p}_0 = 0$ when $N = 0$. We use (2.6) in this paper when we calculate the FAR, since it seems more reasonable.

2.4. Metrics for evaluating the performance

Now we introduce the metrics used in this paper and review the works of Yang *et al.* (2002) and Zhang *et al.* (2013).

First, we introduce two metrics used in Yang *et al.* (2002). The first metric is the average run length (ARL) and the second metric is the standard deviation of the run length (SDRL) where the run length is the number of points plotted on the chart until a signal is given. When $P(B_i | N = n)$ is denoted by $\alpha(N)$, ARL and SDRL are respectively defined as

$$\text{ARL} = E_N \left(\frac{1}{\alpha(N)} \right) \quad (2.7)$$

and

$$\text{SDRL} = \sqrt{\text{Var}_N \left[\frac{1}{\alpha(N)} \right] + E_N \left[\frac{1 - \alpha(N)}{\alpha^2(N)} \right]},$$

where

$$E_N \left(\frac{1}{\alpha(N)} \right) = \sum_{n=0}^m \frac{1}{\alpha(n)} \binom{m}{n} p_0^n (1 - p_0)^{m-n}$$

and

$$E_N \left(\frac{1}{\alpha^2(N)} \right) = \sum_{n=0}^m \frac{1}{\alpha^2(n)} \binom{m}{n} p_0^n (1 - p_0)^{m-n}.$$

Zhang *et al.* (2013) pointed out that SDRL does not reflect the robustness of the performance of the charts because practitioners would usually take one Phase I sample to estimate p_0 , the performance of the charts constructed by each practitioner would be different. To compensate this problem, they

used alternate metrics such as the expected value of the ARL (AARL) and the standard deviation of the ARL (SDARL). $E_N [E_R(R|N)] = E_R(R)$ for the run length R ; therefore, the AARL has the same expression as the ARL in (2.7). The SDARL is defined as

$$\text{SDARL} = \sqrt{E_N(\text{ARL}^2) - E_N^2(\text{ARL})} = \sqrt{E_N\left(\frac{1}{\alpha^2(N)}\right) - E_N^2\left(\frac{1}{\alpha(N)}\right)}.$$

The SDARL metric was proposed by Jones and Steiner (2012), who used it to evaluate the effect of estimation error on the risk-adjusted CUSUM chart.

From evaluation with the SDARL, Zhang *et al.* (2013) stated that larger sample sizes are needed than those stated by Yang *et al.* (2002), because the SDARL is relatively large compared to the AARL even when m is quite large so that the robustness of the ARL values seem not satisfied. Therefore, they pointed out that the SDARL is also needed to evaluate the performance of the geometric charts.

3. Adjustment of the Control Limits

In this section, we propose a method for adjustment of the control limits to yield the desired FAR (α) or the desired in-control ARL (ARL_0). Note that the relationship such as $\text{ARL}_0 = 1/\alpha$ exists in a geometric chart. It is expected that the amount for adjustment depends on the sample size m , the number of nonconforming items N , and the desired α (or the desired ARL_0).

3.1. Adjustment constants of the control limits

First, we established the amount to adjust the control limits. From the work of Yang *et al.* (2002), we can see that the AARL_0 for given values of m is always smaller than the desired value. To make the AARL_0 approximate to the desired ARL_0 , the UCL must move upward while the LCL move downward. Thus, when we denote the lower-sided and the upper-sided constants for the adjustment by Δ_L and Δ_U respectively, the adjusted control limits of the charts can be expressed as

$$\begin{aligned}\widehat{\text{LCL}}^*(N) &= \frac{\ln(1 - \alpha/2)}{\ln(1 - N/m)} - \Delta_L, \\ \widehat{\text{UCL}}^*(N) &= \frac{\ln(\alpha/2)}{\ln(1 - N/m)} - 1 + \Delta_U\end{aligned}$$

by using (2.3) and (2.4). If a value of the desired ARL_0 is given as Z_0 , the AARL_0 using the adjusted control limits must satisfy

$$E_N\left(\frac{1}{\alpha_0(N)}\right) \approx Z_0, \quad (3.1)$$

where

$$E_N\left(\frac{1}{\alpha_0(N)}\right) = \sum_{n=1}^m \frac{1}{\alpha_0(n)} \binom{m}{n} p_0^n (1 - p_0)^{m-n} + (1 - p_0)^m \quad (3.2)$$

and

$$\begin{aligned}\alpha_0(n) &= P_0(B_i | N = n) \\ &= P_0\{X_i < \widehat{\text{LCL}}^*(N) | N = n\} + P\{X_i > \widehat{\text{UCL}}^*(N) | N = n\} \\ &= 1 - (1 - p_0)^{\{\ln(1 - \frac{\alpha}{2}) / \ln(1 - \frac{n}{m})\} - \Delta_L} + (1 - p_0)^{\{\ln(\frac{\alpha}{2}) / \ln(1 - \frac{n}{m})\} + \Delta_U}.\end{aligned} \quad (3.3)$$

In (3.2), we use a similar technique used in (2.6) for the case of $N = 0$. We aim to find the optimal combination of Δ_L and Δ_U which satisfies (3.1). However, it is complicated to determine which combination is the best. Therefore, we consider the relation that the ratio of Δ_U to Δ_L becomes approximately equal in order to calculate the constants easily. From (2.3) and (2.4), we have

$$\ln\left(1 - \frac{N}{m}\right) = \frac{\ln(1 - \alpha/2)}{\widehat{\text{LCL}}(N)} = \frac{\ln(\alpha/2)}{\widehat{\text{UCL}}(N) + 1}.$$

We then obtain a relation between $\widehat{\text{LCL}}(N)$ and $\widehat{\text{UCL}}(N)$ as

$$\widehat{\text{LCL}}(N) = \frac{\ln(1 - \alpha/2)}{\ln(\alpha/2)} \{\widehat{\text{UCL}}(N) + 1\}.$$

From the relation above, we get the approximate equation as $\Delta_L \approx c \Delta_U$, where

$$c = \frac{\ln(1 - \alpha/2)}{\ln(\alpha/2)}.$$

If we substitute $\Delta_U = \Delta$ and $\Delta_L = c \Delta$ into (3.3), $\alpha_0(n)$ can be expressed as

$$\alpha_0(n) = 1 - (1 - p_0)^{\{\ln(1 - \frac{\alpha}{2}) / \ln(1 - \frac{n}{m})\} - c\Delta} + (1 - p_0)^{\{\ln(\alpha/2) / \ln(1 - \frac{n}{m})\} + \Delta}. \quad (3.4)$$

If we can estimate the constant Δ , the AARL_0 which is closer to the desired value can be obtained by using (3.1), (3.2), and (3.4).

3.2. Estimation of the adjustment constants

In this subsection, the adjustment constants are calculated in order to approximate the AARL_0 to the specified values through the simulation. We consider the situation where the constants depend on the following factors: sample size m , number of nonconforming items N , and false alarm rate α .

Simulation in this paper designed to consider the following: there are the sample sizes m per 10,000 units from 7,000 to 2,000,000, the in-control fraction nonconforming p_0 per 0.0001 from 0.0001 to 0.01, and the false alarm rate $\alpha = 0.001, 0.00125, 0.002, 0.0027, 0.005, \text{ and } 0.01$ (that is $Z_0 = 1000, 800, 500, 370.4, 200, \text{ and } 100$) to examine how the change of these factors affects the constants. The constants Δ satisfying (3.1), (3.2), and (3.4) for given values of $m, N, \text{ and } \alpha$ are founded using a bisection search algorithm. Then, various regression models are fitted where a response variable is the constant (Δ), there are 3 factors to adjust the control limits and explanatory variables that include sample size, number of nonconforming items in a sample, and false alarm rate. After evaluating the adequacy of each regression model, we conclude that the best model is

$$\ln \Delta = 0.337 + 1.026 \ln m - 2.288 \ln N - 0.1732 \ln \alpha,$$

where the coefficient of determination R^2 is 0.991. This equation can be expressed as

$$\Delta = \exp\{0.337 + 1.026 \ln m - 2.288 \ln N - 0.1732 \ln \alpha\}. \quad (3.5)$$

Using the above model, the lower-sided and the upper-sided constants can be estimated and used to adjust the control limits of the chart to yield the desired performance.

It can be seen that both the lower-sided and the upper-sided constants tend to be smaller and approximate to 0 when sample size m increases. However, when p_0 is very small (for example $p_0 = 0.0001$) the adjustment constants seem to be very large and be unstable unless m is sufficiently large. This is due to the MLE restriction that demands a sufficiently large sample size when the in-control fraction nonconforming p_0 is very small. We do not seriously consider cases where m is also very small when p_0 is very small.

Table 1: Values of $AARL_0$ without (the 1st row) and with (the 2nd row) adjustment of the control limits when $Z_0 = 370.4$

m	p_0							
	0.0001	0.0002	0.0003	0.0004	0.0005	0.0007	0.001	0.005
20,000	217.44	277.54	302.27	316.69	326.28	338.32	348.22	367.51
	704.86	432.64	381.28	369.45	366.22	365.22	366.06	370.18
30,000	255.78	302.27	321.92	333.15	340.45	349.37	356.40	368.77
	520.17	382.41	367.86	365.67	365.56	366.41	367.69	370.40
40,000	277.54	316.69	333.15	342.35	348.22	355.22	360.55	369.31
	438.63	370.52	365.92	365.83	366.39	367.54	368.68	370.46
50,000	291.84	326.28	340.46	348.22	353.09	358.78	363.01	369.60
	401.73	367.22	365.90	366.50	367.22	368.34	369.29	370.47
70,000	310.29	338.32	349.37	355.22	358.78	362.81	365.69	369.89
	375.95	366.12	366.80	367.72	368.42	369.30	369.92	370.47
90,000	321.92	345.58	354.56	359.18	361.93	364.96	367.07	370.03
	369.29	366.53	367.69	368.57	369.15	369.80	370.22	370.45
100,000	326.28	348.23	356.41	360.56	363.01	365.69	367.53	370.07
	368.00	366.83	368.05	368.88	369.40	369.97	370.30	370.44
120,000	333.15	352.27	359.18	362.61	364.60	366.74	368.18	370.14
	366.90	367.43	368.64	369.35	369.77	370.18	370.41	370.43
150,000	340.46	356.41	361.93	364.60	366.12	367.72	368.78	370.20
	366.65	368.18	369.25	369.80	370.09	370.36	370.48	370.42
170,000	344.06	358.36	363.20	365.50	366.80	368.16	369.05	370.22
	366.80	368.57	369.53	369.99	370.22	370.43	370.50	370.41
200,000	348.23	360.56	364.60	366.48	367.53	368.62	369.32	370.25
	367.17	369.03	369.84	370.19	370.35	370.48	370.52	370.40
300,000	356.41	364.60	367.08	368.18	368.78	369.39	369.78	370.29
	368.39	369.89	370.32	370.46	370.51	370.53	370.51	370.39
500,000	363.01	367.53	368.78	369.32	369.61	369.90	370.08	370.33
	369.68	370.42	370.54	370.55	370.53	370.50	370.46	370.38
700,000	365.69	368.62	369.39	369.72	369.90	370.07	370.18	370.34
	370.19	370.55	370.56	370.53	370.51	370.47	370.43	370.37
900,000	367.08	369.15	369.69	369.91	370.03	370.15	370.23	370.35
	370.41	370.57	370.54	370.51	370.48	370.45	370.42	370.37
1,000,000	367.53	369.32	369.78	369.98	370.08	370.18	370.25	370.35
	370.48	370.57	370.53	370.50	370.47	370.44	370.41	370.37
1,200,000	368.18	369.56	369.91	370.06	370.14	370.22	370.27	370.35
	370.55	370.56	370.51	370.48	370.45	370.42	370.40	370.37
1,500,000	368.79	369.78	370.03	370.14	370.20	370.26	370.30	370.36
	370.59	370.54	370.49	370.45	370.43	370.41	370.39	370.37
2,000,000	369.32	369.98	370.14	370.21	370.25	370.29	370.32	370.36
	370.59	370.51	370.46	370.43	370.41	370.40	370.38	370.37

3.3. Performance of the geometric chart with adjustment

To check whether the $AARL_0$ of the chart using the adjusted control limits yields the desired value of $Z_0 (= 1/\alpha)$, values of $AARL_0$ are computed by substituting Δ in (3.4) with the value estimated from (3.5). In calculation, we consider the values of the desired ARL_0 , Z_0 , are 200, 370.4, and 800. The variation of ARL_0 is also an important factor for the performance of the charts; therefore, the values of $SDARL_0$ used by Zhang *et al.* (2013) are also computed.

We first investigate the improvement of the performance on the charts by comparing the values of $AARL_0$ without and with adjustment of the control limits for each combination of m , p_0 , and the desired ARL_0 (Z_0). We provide Tables 1 and 2 which represent the values of $AARL_0$ without and with adjusting the control limits when $Z_0 = 370.4$ and 800, respectively. For each combination m and

Table 2: Values of $AARL_0$ without (the 1st row) and with (the 2nd row) adjustment of the control limits when $Z_0 = 800$

m	p_0							
	0.0001	0.0002	0.0003	0.0004	0.0005	0.0007	0.001	0.005
20,000	460.99	590.62	644.67	676.31	697.53	724.47	746.97	792.79
	1614.78	948.35	824.92	796.72	788.92	786.36	788.30	799.14
30,000	542.94	644.67	687.87	712.86	729.29	749.61	765.95	795.96
	1162.50	827.57	792.84	787.49	787.13	789.14	792.31	799.88
40,000	590.62	676.32	712.86	733.59	746.98	763.18	775.79	797.34
	963.17	799.16	788.07	787.75	789.06	791.91	794.88	800.11
50,000	621.89	697.54	729.29	746.98	758.23	771.57	781.69	798.08
	873.91	791.23	787.91	789.31	791.10	793.95	796.50	800.19
70,000	662.25	724.48	749.62	763.18	771.57	781.21	788.24	798.82
	812.07	788.45	790.03	792.33	794.15	796.51	798.29	800.21
90,000	687.87	740.94	761.65	772.51	779.08	786.46	791.68	799.17
	796.11	789.36	792.23	794.52	796.09	797.92	799.17	800.19
100,000	697.54	746.99	765.96	775.80	781.69	788.24	792.82	799.28
	793.01	790.09	793.16	795.34	796.77	798.39	799.44	800.18
120,000	712.86	756.33	772.51	780.72	785.56	790.84	794.46	799.44
	790.30	791.55	794.68	796.61	797.79	799.04	799.79	800.15
150,000	729.30	765.97	779.08	785.56	789.30	793.31	795.99	799.58
	789.64	793.45	796.31	797.86	798.74	799.61	800.07	800.12
170,000	737.46	770.58	782.15	787.78	790.99	794.40	796.66	799.65
	789.99	794.47	797.09	798.42	799.15	799.83	800.16	800.11
200,000	746.99	775.80	785.56	790.21	792.83	795.57	797.36	799.71
	790.88	795.70	797.95	799.01	799.56	800.04	800.24	800.09
300,000	765.97	785.56	791.68	794.47	795.99	797.55	798.54	799.83
	793.95	798.09	799.41	799.92	800.14	800.28	800.29	800.05
500,000	781.70	792.83	795.99	797.37	798.10	798.83	799.29	799.90
	797.42	799.73	800.20	800.31	800.33	800.29	800.22	800.02
700,000	788.25	795.57	797.55	798.39	798.83	799.27	799.55	799.93
	798.92	800.19	800.35	800.34	800.31	800.24	800.16	800.01
900,000	791.69	796.93	798.30	798.88	799.18	799.48	799.68	799.95
	799.64	800.34	800.36	800.31	800.26	800.19	800.12	800.01
1,000,000	792.83	797.37	798.54	799.03	799.29	799.55	799.72	799.96
	799.85	800.37	800.35	800.29	800.24	800.17	800.11	800.01
1,200,000	794.47	797.98	798.88	799.25	799.45	799.65	799.78	799.96
	800.13	800.39	800.32	800.25	800.20	800.14	800.09	800.00
1,500,000	796.00	798.54	799.18	799.45	799.59	799.74	799.83	799.97
	800.32	800.37	800.28	800.21	800.16	800.11	800.07	800.00
2,000,000	797.37	799.03	799.45	799.62	799.72	799.81	799.88	799.98
	800.42	800.31	800.22	800.16	800.12	800.07	800.04	800.00

p_0 in the tables, the first and the second rows contain the $AARL_0$ without and with adjustment of the control limits, respectively.

Tables 1 and 2 show that the values of $AARL_0$ (with adjustment) are very closer to the desired values than those without the adjustment, especially for small m and p_0 . In the results, some values of $AARL_0$ with adjustment are larger than the desired value when both m and p_0 are very small (such as $m \leq 70,000$ when $p_0 = 0.0001$, $m \leq 30,000$ when $p_0 = 0.0002$, and $m = 20,000$ when $p_0 = 0.0003$). This is because the sample size is inadequate to accurately estimate p_0 and this makes the estimation of the adjustment constants difficult, as mentioned on Subsection 3.2. However, we can see that except these cases for very small p_0 , the adjustment of the control limits improves the performance of the control charts even if the sample size is inadequate. For example, when $Z_0 = 370.4$ and $p_0 = 0.0001$,

Table 3: Values of $SDARL_0$ without (the 1st row) and with (the 2nd row) adjustment of the control limits when $Z_0 = 370.4$

m	p_0							
	0.0001	0.0002	0.0003	0.0004	0.0005	0.0007	0.001	0.005
20,000	167.63	169.46	162.88	157.86	153.65	146.52	137.81	88.19
	630.82	249.46	193.88	177.93	168.38	155.59	143.14	88.64
30,000	171.52	162.88	155.68	149.91	144.93	136.53	126.52	75.56
	361.57	194.43	172.92	161.39	153.20	141.52	129.43	75.80
40,000	169.46	157.86	149.91	143.40	137.81	128.53	117.77	67.18
	255.69	178.31	161.47	150.97	143.23	131.79	119.65	67.33
50,000	165.96	153.64	144.92	137.81	131.76	121.89	110.68	61.08
	213.84	168.73	153.31	143.26	135.65	124.22	112.03	61.19
70,000	160.23	146.51	136.52	128.53	121.89	111.33	99.75	52.65
	185.50	155.88	141.63	131.83	124.24	112.73	100.56	52.71
90,000	155.68	140.50	129.58	121.03	114.05	103.19	91.58	46.96
	173.43	146.99	133.09	123.29	115.66	104.15	92.13	47.00
100,000	153.64	137.81	126.51	117.76	110.68	99.75	88.19	44.73
	169.00	143.36	129.52	119.70	112.05	100.57	88.65	44.76
120,000	149.90	132.90	121.03	111.99	104.78	93.81	82.41	41.08
	161.80	137.13	123.31	113.46	105.82	94.43	82.76	41.10
150,000	144.92	126.51	114.05	104.78	97.50	86.63	75.56	36.97
	153.55	129.55	115.68	105.82	98.24	87.07	75.81	36.99
170,000	141.92	122.77	110.04	100.70	93.43	82.67	71.84	34.83
	149.10	125.28	111.39	101.56	94.04	83.03	72.04	34.84
200,000	137.81	117.76	104.78	95.40	88.19	77.65	67.17	32.22
	143.46	119.73	105.83	96.07	88.66	77.93	67.33	32.23
300,000	126.51	104.78	91.57	82.41	75.56	65.81	56.40	26.47
	129.60	105.84	92.14	82.77	75.81	65.95	56.48	26.48
500,000	110.68	88.19	75.56	67.17	61.08	52.65	44.73	20.61
	112.11	88.67	75.81	67.33	61.19	52.71	44.76	20.61
700,000	99.75	77.65	65.81	58.14	52.65	45.15	38.20	17.46
	100.61	77.94	65.96	58.23	52.71	45.18	38.22	17.46
900,000	91.57	70.18	59.07	51.98	46.96	40.15	33.89	15.42
	92.16	70.37	59.17	52.04	47.00	40.17	33.90	15.42
1,000,000	88.19	67.17	56.40	49.56	44.73	38.20	32.22	14.63
	88.68	67.34	56.48	49.61	44.76	38.22	32.23	14.63
1,200,000	82.41	62.17	51.98	45.58	41.08	35.03	29.50	13.36
	82.78	62.29	52.04	45.62	41.10	35.04	29.51	13.36
1,500,000	75.56	56.40	46.96	41.08	36.97	31.47	26.47	11.96
	75.82	56.48	47.00	41.10	36.99	31.48	26.48	11.96
2,000,000	67.17	49.56	41.08	35.85	32.22	27.38	23.00	10.37
	67.34	49.61	41.10	35.87	32.23	27.38	23.00	10.37

the values of $AARL_0$ with adjustment are 369.29, 368.00, and 366.90, respectively, for $m = 90,000$, 100,000, and 120,000, whereas those without adjustment are 321.92, 326.28, and 333.15.

We then compute the values of $SDARL_0$ for each combination of m , p_0 , and Z_0 , and make a comparison with the values without and with adjustment in order to show how the adjustment affects the variation of ARL_0 . We provide the values of $SDARL_0$ without and with adjusting the control limits in Tables 3 and 4 when $Z_0 = 370.4$ and 800, respectively. The first and the second rows show the $SDARL_0$ without and with the adjustment for each combination of m and p_0 in the tables, respectively.

Tables 3 and 4 indicate that the values of $SDARL_0$ with adjustment are larger (or much bigger) for some cases than for those without the adjustment when both m and p_0 are small. We posit that this

Table 4: Values of $SDARL_0$ without (the 1st row) and with (the 2nd row) adjustment of the control limits when $Z_0 = 800$

m	p_0							
	0.0001	0.0002	0.0003	0.0004	0.0005	0.0007	0.001	0.005
20,000	370.30	383.70	373.01	363.65	355.47	341.08	322.82	211.97
	1484.27	576.77	444.96	409.68	389.18	361.82	335.00	213.02
30,000	383.66	373.01	359.45	348.00	337.79	320.08	298.40	182.57
	849.29	446.39	398.93	374.25	356.71	331.52	305.07	183.13
40,000	383.70	363.65	348.00	334.61	322.81	302.81	279.09	162.83
	593.83	410.59	374.44	351.93	335.22	310.27	283.43	163.19
50,000	378.44	355.47	337.79	322.81	309.83	288.23	263.26	148.38
	492.53	389.98	356.96	335.29	318.74	293.58	266.36	148.63
70,000	368.12	341.07	320.07	302.80	288.23	264.71	238.51	128.24
	426.50	362.49	331.76	310.37	293.62	267.94	240.38	128.39
90,000	359.45	328.53	305.09	286.32	270.81	246.33	219.78	114.57
	400.10	343.36	313.13	291.52	274.51	248.54	221.06	114.67
100,000	355.47	322.81	298.40	279.09	263.25	238.51	211.96	109.20
	390.61	335.51	305.28	283.53	266.41	240.40	213.05	109.28
120,000	348.00	312.28	286.32	266.19	249.93	224.92	198.57	100.38
	375.20	321.98	291.56	269.57	252.34	226.35	199.39	100.44
150,000	337.78	298.40	270.80	249.93	233.37	208.36	182.57	90.43
	357.50	305.35	274.55	252.35	235.08	209.38	183.15	90.47
170,000	331.52	290.17	261.82	240.66	224.04	199.18	173.84	85.23
	347.94	295.94	264.93	242.66	225.46	200.02	174.32	85.26
200,000	322.81	279.08	249.93	228.56	211.96	187.46	162.83	78.88
	335.73	283.61	252.37	230.12	213.07	188.11	163.20	78.91
300,000	298.40	249.93	219.78	198.57	182.57	159.60	137.21	64.88
	305.47	252.39	221.10	199.41	183.16	159.94	137.40	64.90
500,000	263.25	211.96	182.57	162.83	148.38	128.24	109.19	50.56
	266.54	213.09	183.17	163.21	148.64	128.39	109.28	50.57
700,000	238.51	187.46	159.60	141.37	128.24	110.21	93.41	42.84
	240.49	188.13	159.95	141.59	128.39	110.30	93.46	42.85
900,000	219.78	169.92	143.59	126.65	114.57	98.13	82.95	37.84
	221.14	170.38	143.83	126.80	114.67	98.19	82.98	37.84
1,000,000	211.96	162.83	137.21	120.83	109.19	93.41	78.88	35.91
	213.11	163.21	137.41	120.95	109.28	93.46	78.91	35.92
1,200,000	198.57	150.96	126.65	111.25	100.38	85.71	72.28	32.81
	199.44	151.25	126.80	111.34	100.45	85.75	72.29	32.81
1,500,000	182.57	137.21	114.57	100.38	90.43	77.06	64.88	29.37
	183.18	137.41	114.67	100.45	90.47	77.09	64.90	29.37
2,000,000	162.83	120.83	100.38	87.71	78.88	67.09	56.40	25.45
	163.22	120.95	100.45	87.75	78.91	67.11	56.41	25.45

is because the values of the adjustment constants are relatively large and unstable when both m and p_0 are small, and this results in an increase in the variation of ARL_0 . However, as the sample size becomes large, the difference of the $SDARL_0$ without and with adjustment becomes negligible even when p_0 is very small.

We also obtained the values of $AARL_0$ and $SDARL_0$ when $Z_0 = 100, 200, 500,$ and 1000 . But we do not include these results in this paper because they have a similar trend with Tables 1 to 4. From comparisons of the $AARL_0$ and the $SDARL_0$, we conclude that the use of the adjusted control limits can improve the performance of the geometric charts, especially when monitoring the high-quality processes.

We now provide an example of how the proposed method can be applied in an industrial process.

Table 5: Values of $P(N)$, \hat{p}_0 , $(\widehat{LCL}, \widehat{UCL})$, and ARL_0 without and with adjustment when $Z_0 = 370.4$, $m = 90000$, and $p_0 = 0.0001$

N	$P(N)$	\hat{p}_0	Without Adjustment		With Adjustment	
			$(\widehat{LCL}, \widehat{UCL})$	ARL_0	$(\widehat{LCL}^*, \widehat{UCL}^*)$	ARL_0
1	0.00111	0.00001	(121.58, 594684.25)	82.75	(25.00, 1067071.13)	400.42
2	0.00500	0.00002	(60.79, 297339.97)	164.99	(41.02, 394065.47)	244.30
3	0.01499	0.00003	(40.53, 198225.22)	247.24	(32.71, 236476.34)	306.24
4	0.03373	0.00004	(30.39, 148667.84)	329.45	(26.35, 168473.28)	380.04
5	0.06072	0.00006	(24.32, 118933.41)	410.58	(21.89, 130819.92)	456.96
6	0.09109	0.00007	(20.26, 99110.45)	482.17	(18.66, 106942.73)	529.89
7	0.11712	0.00008	(17.37, 84951.20)	515.53	(16.24, 90455.65)	574.42
8	0.13176	0.00009	(15.20, 74331.77)	474.01	(14.37, 78387.11)	546.50
9	0.13176	0.00010	(13.51, 66072.20)	370.37	(12.88, 69169.55)	439.14
10	0.11859	0.00011	(12.16, 59464.55)	261.16	(11.66, 61898.43)	311.07
11	0.09702	0.00012	(11.05, 54058.29)	178.78	(10.65, 56015.29)	210.28
12	0.07277	0.00013	(10.13, 49553.07)	124.13	(9.80, 51156.80)	143.27
13	0.05037	0.00014	(9.35, 45740.97)	88.91	(9.08, 47076.32)	100.70
14	0.03238	0.00016	(8.68, 42473.45)	65.94	(8.45, 43600.54)	73.43
15	0.01943	0.00017	(8.10, 39641.60)	50.53	(7.91, 40604.10)	55.47
16	0.01093	0.00018	(7.60, 37163.73)	39.88	(7.43, 37994.10)	43.25
17	0.00578	0.00019	(7.15, 34977.38)	32.29	(7.00, 35700.20)	34.66
18	0.00289	0.00020	(6.75, 33033.95)	26.72	(6.62, 33668.16)	28.45
19	0.00137	0.00021	(6.40, 31295.09)	22.54	(6.28, 31855.51)	23.82

We assume that the in-control proportion nonconforming of the process is $p_0 = 0.0001$. Suppose that we want the desired ARL_0 to be 370.4 and we use 90,000 items in Phase I. Table 5 gives values of $P(N)$, \hat{p}_0 , the control limits without and with the adjustment, and ARL_0 when using the obtained control limits, respectively. For several values of the number of nonconforming items N . $P(N)$ denotes the probability that the number of nonconforming items equals N , and note that we display values of N whose probability is greater than 0.001. Table 5 provides values of the control limits $(\widehat{LCL}, \widehat{UCL})$ when the value of N in the Phase I sample is determined.

Table 5 shows that the performance with adjustment is worse than that without adjustment for some values of N , but the performance with adjustment is better in the most values of N . Note that the better performance means the actual ARL_0 is close to the desired ARL_0 , and the overall performance is not good because we use a relatively small sample size even though p_0 is very small.

4. Conclusion

A geometric chart is very useful to monitor high-quality processes. However, the low p_0 values inherent to high-quality processes imply that a very large sample size may be needed to accurately estimate p_0 in Phase I.

This paper proposes a method to abate the effect of the estimation error and improve the performance of the geometric chart when a sufficiently large sample size is unavailable. We adjust the control limits using the estimated constants computed from a regression model which considers the sample size in Phase I (m), the number of nonconforming items (N), and the false alarm rate (α) as explanatory variables. From the calculation results, we conclude that the chart with the adjusted control limits provide superior performance compared to the chart with the traditional control limits. We also believe that the use of the adjusted control limits can help practitioners choose a sample size when they monitor high-quality processes.

We expect that the proposed method would facilitate the adjustment of the control limits to save

time and cost in monitoring high-quality processes by decreasing the required sample size to estimate the parameter. The parameter estimation can be still unstable (especially when both the sample size and the in-control fraction nonconforming are very small) despite the adjustment. We believe that additional research on this problem is needed in order to improve the performance of the geometric charts. One may consider other methods to estimate the adjustment constants instead of the proposed method. According to the works of Zhang *et al.* (2013) and Lee *et al.* (2013), using the Bayes estimator may yield better performance than using the MLE, when monitoring high-quality processes.

References

- Chang, T. C. and Gan, F. F. (2001). Cumulative sum charts for high yield processes, *Statistica Sinica*, **11**, 791–805.
- Goh, T. N. (1987). A control chart for very high yield processes, *Quality of Assurance*, **13**, 18–22.
- Goh, T. N. and Xie, M. (1994). New approach to quality in a near-zero defect environment, *Total Quality Management*, **5**, 3–10.
- Jensen, W. A., Jones-Farmer, L. A., Champ, C. W. and Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review, *Journal of Quality Technology*, **38**, 349–364.
- Jones, M. A. and Steiner, S. H. (2012). Assessing the effect of estimation error on the risk-adjusted CUSUM chart performance, *International Journal for Quality in Health Care*, **24**, 176–181.
- Jones-Farmer, L. A., Woodall, W. H., Steiner, S. H. and Champ, C. W. (2014). An overview of phase I analysis for process improvement and monitoring, *Journal of Quality Technology*, **46**, 265–280.
- Lee, J., Wang, N., Xu, L., Schuh, A. and Woodall, W. H. (2013). The effect of parameter estimation on upper-sided Bernoulli CUSUM charts, *Quality and Reliability Engineering International*, **29**, 209–223.
- McCool, J. and Joyner-Motley, T. (1998). Control charts applicable when the fraction nonconforming is small, *Journal of Quality Technology*, **30**, 240–247.
- Nelson, L. S. (1994). A control chart for parts-per-million nonconforming items, *Journal of Quality Technology*, **26**, 239–240.
- Steiner, S. H. and MacKay, R. J. (2004). Effective monitoring of processes with parts per million defective: A hard problem!, In H. J. Lenz and P. Wilrich (Eds.), *Frontiers in Statistical Quality Control 7*, Physica-Verlag, Heidelberg, 140–149.
- Szarka, J. L. and Woodall, W. H. (2011). A review and perspective on surveillance of Bernoulli processes, *Quality and Reliability Engineering International*, **27**, 735–752.
- Tang, L. C. and Cheong, W. T. (2004). Cumulative conformance count chart with sequentially updated parameters, *IIE Transactions*, **36**, 841–853.
- Woodall, W. H. (1997). Control charts based on attribute data: bibliography and review, *Journal of Quality Technology*, **29**, 172–183.
- Xie, M. and Goh, T. N. (1997). The use of probability limits for process control based on geometric distribution, *International Journal of Quality & Reliability Management*, **14**, 64–73.
- 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., Peng, Y., Schuh, A., Megahed F. M. and Woodall, W. H. (2013). Geometric charts with estimated control limits, *Quality and Reliability Engineering International*, **29**, 209–223.