The Korean Communications
in Statistics Vol. 5, No. 3, 1998
pp. 645-657

The Effect of Estimated Control Limits[1]

JaiWook Baik[2] and TaiYon Won[3]

Abstract

During the start-up of a process or in a job-shop environment conventional use of control charts may lead to erroneous results due to the limited number of subgroups used for the construction of control limits. This article considers the effect of using estimated control limits based on a limited number of subgroups. Especially we investigate the performance of $\overline{X}$ and R control charts when the data are independent, and X control chart when the data are serially correlated in terms of average run length(ARL) and standard deviation run length(SDRL) using simulation. It is found that the ARL and SDRL get larger as the number of subgroups used for the construction of the chart becomes smaller.

1. Introduction

Control charts have been widely used to check whether the production process has fallen out of control. Especially, $\overline{X}$ and R control charts are used to control the process mean and variation. Usually it is assumed that there are a large number of subgroups available before we construct control limits. In this case, we can obtain relatively accurate probability limits. However, a production run may be too short to have a large number of subgroups. Or we may want to use control charts for monitoring the process as soon as possible. In fact, there has recently been considerable interest in using SPC charting techniques in the job-shop environment. It is shown in Hillier (1964, 1969) that the effect of using control limits based on a small number of samples is to erroneously indicate trouble more frequently than assumed for future samples in the process.

During the start-up of a process just brought into statistical control, parameter may not be known. Ghosh, Reynolds and Hui (1981) and Quesenberry (1993) showed that using estimated control limits with limited number of subgroups results in charts for which the events that future individual points exceed the control limits are dependent. They also showed that the result of this dependence is to increase the ARL which is the expected time to signal. In

1) This paper was financially supported by Korea National Open University Research Fund, 1997.
2) Associate Professor, Department of Applied Statistics, Korea National Open University, Seoul, 110-791, Korea
3) Instructor, Department of Applied Statistics, Chung Ang University, Seoul, 156-756, Korea

Quesenberry's paper, $\overline{X}$ chart with σ estimated by $\overline{S}/c_4$ was investigated. However, $\overline{R}/d_2$ is also widely used to estimate σ in the construction of $\overline{X}$ and R charts. Hence, we investigate the performance of $\overline{X}$ and R control charts with σ estimated by $\overline{R}/d_2$.

The traditional assumption in quality control charts is that the sequential observations are identically and independently distributed. But in practice, it is common to have correlation in the data; autocorrelations and other systematic time series effects are often substantial. In particular, serial correlation is a common feature in environmental, biological and chemical data. Therefore, we consider the situation where the consecutive observations are correlated according to a first order autoregressive process and investigate the performance of X control charts when the control limits are based on a small number of observations.

2. $\overline{X}$ Control Chart

Suppose that a quality characteristic X_{ij} is normally distributed with both known mean μ and known standard deviation σ. Then the sample mean $\overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$ is normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$. Hence, $\overline{X}_i$ for $i=1,\cdots,m$ is plotted on a chart with control limits

$$UCL = \mu + 3\sigma/\sqrt{n} \tag{2-1}$$

$$LCL = \mu - 3\sigma/\sqrt{n} \tag{2-2}$$

for monitoring a process mean. If there is no point outside the control limits then the process mean is considered to be in statistical process control. If any future sample mean falls out of the control limits then it is an indication that the process mean is out of control and corrective action is required. If the process parameters μ and σ are not known then the estimates of them such as

$$\overline{\overline{X}} = \frac{1}{m}\sum_{i=1}^{m}\overline{X}_i \tag{2-3}$$

$$\widehat{\sigma} = \frac{\overline{R}}{d_2} = \frac{1}{d_2}\frac{1}{m}\sum_{i=1}^{m}R_i \tag{2-4}$$

are widely used to give estimated control limits of

$$\widehat{UCL} = \overline{\overline{X}} + 3\frac{1}{\sqrt{n}}\frac{\overline{R}}{d_2} \tag{2-5}$$

$$\widehat{LCL} = \overline{\overline{X}} - 3\frac{1}{\sqrt{n}}\frac{\overline{R}}{d_2} \tag{2-6}$$

where d_2 is a function of n as can be found in Montgomery (1991).

Let B_i be the event such that a future sample mean $\overline{X}_i$ goes out of control limits which are based on previous m subgroups as defined in Qusenberry (1993). Then the probability that event B_i occurs is

$$P(B_i) = P(\overline{X}_i > \widehat{UCL} \text{ or } \overline{X}_i < \widehat{LCL}).$$

Since both $\overline{X}_i$ and $\widehat{UCL}$ are approximately normally distributed the difference $\overline{X}_i - \widehat{UCL}$ is also approximately normally distributed with mean and variance of

$$E(\overline{X}_i - \widehat{UCL}) = -3\sigma/\sqrt{n}$$

$$Var(\overline{X}_i - \widehat{UCL}) = \frac{\sigma^2}{n}[1 + \frac{1}{m}(1 + 9\frac{d_3^2}{d_2^2})]$$

where d_3 is also a function of n as found in Montgomery (1991). Hence,

$$P(B_i) = 2\left[1 - \Phi\left(\frac{3}{\left\{1 + \frac{1}{m}\left(1 + \frac{9d_3^2}{d_2^2}\right)\right\}^{1/2}}\right)\right].$$

Table I shows the result for specific values of m and n. It shows that for small values of m and n the false alarm probability is larger than expected when we have known parameters or when we have a very large number of subgroups available for the construction of control limits. Especially for a smaller number of subgroups, say m=10, the false alarm probability is a lot larger than 0.0027 which is $P(\overline{X}_i > UCL$ or $\overline{X}_i < LCL)$. For example, if m=10 and n=5 the false alarm probability is 0.0067. However, as the number of subgroups or(and) the subgroup size grows larger, the false alarm probability approaches 0.0027.

Table I. False signal probabilities for various values of m and n
($\overline{X}$ chart with σ estimated by $\overline{R}/d_2$)

m \ n	2	3	4	5	6	7	8	9	10
5	.0445	.0212	.0153	.0127	.0113	.0104	.0098	.0094	.0090
10	.0182	.0098	.0076	.0067	.0062	.0059	.0056	.0055	.0054
30	.0063	.0045	.0040	.0038	.0037	.0036	.0035	.0035	.0035
100	.0036	.0032	.0031	.0030	.0030	.0030	.0029	.0029	.0029
500	.0029	.0028	.0028	.0028	.0028	.0027	.0027	.0027	.0027
∞	.0027	.0027	.0027	.0027	.0027	.0027	.0027	.0027	.0027

As explained in Quesenberry (1993), the events B_i and B_j for $i \neq j$ are not independent.

Instead, the correlation between the random variables $\overline{X}_i - \widehat{UCL}$ and $\overline{X}_j - \widehat{UCL}$ is

$$Corr(\overline{X}_i - \widehat{UCL}, \overline{X}_j - \widehat{UCL})$$

$$= \frac{Var(\widehat{UCL})}{Var(\overline{X}_i - \widehat{UCL})}$$

$$= \frac{\frac{\sigma^2}{mn}(1+9\frac{d_3^2}{d_2^2})}{\frac{\sigma^2}{n}+\frac{\sigma^2}{mn}(1+9\frac{d_3^2}{d_2^2})} = [1+m(\frac{1+9d_3^2}{d_2^2})^{-1}]^{-1} .$$

Table II gives some values of the correlation for specific values of m and n. For example, if m=10 and n=5 the correlation is 0.1831. Table II shows that the correlation between events B_i and B_j gets smaller as the number of subgroups or(and) subgroup size grows larger.

Table II. Correlation for various values of m and n
($\overline{X}$ chart with σ estimated by $\overline{R}/d_2$)

m \ n	2	3	4	5	6	7	8	9	10
5	.5514	.4101	.3459	.3096	.2865	.2705	.2589	.2499	.2428
10	.3807	.2579	.2091	.1831	.1672	.1564	.1487	.1428	.1382
30	.1700	.1038	.0810	.0695	.0627	.0582	.0550	.0526	.0507
100	.0579	.0336	.0258	.0219	.0197	.0182	.0172	.0164	.0158
500	.0121	.0069	.0053	.0045	.0040	.0037	.0035	.0033	.0032
∞	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Since events B_i and B_j are not independent the distribution of run length is not a geometric distribution. Therefore, simulation is used to find the mean and standard deviation of the run length (ARL and SDRL). The following is the procedure used to obtain simulated ARL(estimated standard error of the simulated ARL) and SDRL of Table III.

1) For each entry in Table III, m limited subgroups of size n(=5) are generated from a N (μ, σ^2) distribution where we assume that μ=0 and σ=1 without loss of generality.
2) Both $\widehat{UCL}$ and $\widehat{LCL}$ for the $\overline{X}$ chart are computed.
3) Samples are generated from a N($\mu + \delta\sigma/\sqrt{n}, \sigma^2$) distribution and an appropriate statististic($\overline{x}$ in Table III) is calculated until it falls outside of control limits. Then the number of times to the out of control signal is one observation for the run length distribution.

4) Repeat 1), 2) and 3) until we obtain a sufficiently small estimated standard error of the estimate of the ARL. In Table III the number of simulations is 10,000.

Then we can get average and standard deviation from those 10,000 simulated run lengths which are simulated ARL and SDRL, respectively. Finally estimated standard error of the simulated ARL can be obtained from 'simulated SDRL / sqrt(number of simulations)'.

For comparison purpose we include ARL and SDRL for infinite m with

$$ARL = 1/(1-\beta)$$

$$SDRL = \sqrt{\beta}/(1-\beta)$$

where $\beta = P(LCL < \overline{X}_i < UCL)$. The first value in Table III is the estimated ARL and the value with parenthesis is its estimated standard error. The value below the estimated ARL is the estimated SDRL. For instance, if the control limits are based on 10 subgroups of size 5 using formula (2-5) and (2-6) then it will take about 646.5 unit hours when there is no shift in the process mean before the $\overline{X}$ chart signals as opposed to 370.4 unit hours which would be obtained using formula (2-1) and (2-2) when the parameters are known or when there is a very large number of subgroups available before the construction of control limits. In this case the standard error of the estimated ARL is 29.5 and the estimated SDRL is 2953.3.

As can be seen in Table III, the ARL based on a small number of subgroups is larger than the one based on a large number of subgroups. Therefore, one may want to shorten the

Table III. Simulated ARL (standard error of the simulated ARL) and SDRL of the $\overline{X}$ chart based on m subgroups of size n=5 ($\overline{X}$ chart with σ estimated by $\overline{R}/d_2$)

m \ δ	0.0	0.25	0.5	1.0	2.0	3.0
5	1512.1(136.9)	437.0(416.0)	1110.5(138.4)	439.9(157.9)	19.3(1.1)	3.0(0.1)
	13694.1	41598.0	13838.6	15787.8	105.1	5.6
10	646.5(29.5)	528.4(16.9)	344.0(11.0)	98.4(3.2)	9.9(0.3)	2.3(0.0)
	2953.3	1686.0	1101.3	320.1	25.8	2.5
30	418.1(7.0)	339.4(5.6)	201.2(3.3)	56.5(0.8)	7.2(0.1)	2.1(0.0)
	695.5	561.4	334.0	84.9	9.0	1.7
100	375.0(4.3)	298.4(3.5)	167.5(2.0)	6.6(0.5)	6.5(0.1)	2.0(0.0)
	434.0	347.6	202.4	51.8	6.4	1.5
500	367.6(3.8)	280.4(2.9)	156.7(1.6)	44.1(0.5)	6.4(0.1)	2.0(0.0)
	375.9	287.2	161.1	45.3	6.0	1.4
∞	370.4	281.2	155.2	43.9	6.3	2.0
	369.9	280.6	154.7	43.4	5.8	1.4

control limits when the number of subgroups is not large in order to get a desired ARL of 370.4 for instance. Trial and error may be needed in adjusting the control limits to get a desired level of in control ARL.

3. R Control Chart

Process variability may be controlled by plotting ranges from successive samples on an R control chart where control limits are usually determined by

$$UCL = (d_2 + 3d_3)\sigma \tag{3-1}$$

$$LCL = (d_2 - 3d_3)\sigma . \tag{3-2}$$

If the process standard deviation σ is not known the estimate $\widehat{\sigma} = \overline{R}/d_2$ is usually used to give estimated control limits of

$$\widehat{UCL} = D_4\overline{R} \tag{3-3}$$

$$\widehat{LCL} = D_3\overline{R} \tag{3-4}$$

where D_4 and D_3 are $1 + 3d_3/d_2$ and $1 - 3d_3/d_2$, respectively. False alarm probabilities based on $\widehat{UCL}$ and $\widehat{LCL}$ with some values of m are shown in Hillier (1969).

Let B_i be the event that the sample range R_i goes out of control limits which are based on m subgroups. Then the events B_i and B_j for $i \neq j$ are not independent. Instead, the correlation between the random variables $R_i - \widehat{UCL}$ and $R_j - \widehat{UCL}$ is

$$Corr(R_i - \widehat{UCL},\ R_j - \widehat{UCL})$$

$$= \frac{Var(\widehat{UCL})}{Var(R_i - \widehat{UCL})}$$

$$= \frac{D_4^2 \frac{d_3^2}{m} \sigma^2}{(1 + \frac{D_4^2}{m}) d_3^2 \sigma^2} = [1 + (\frac{D_4^2}{m})^{-1}]^{-1} \quad .$$

Table IV gives some values of the correlation for specific values of m and n. Note that the correlation gets smaller as m or(and) n increases. However, even for m equal to 30 and n equal to 5, which is usually required for setting up control limits, the correlation is as high as 0.1298.

Since events B_i and B_j are not independent simulation is used to find the ARL and SDRL of the run length distribution. The following is the procedure used to obtain simulated ARL(estimated standard error of the simulated ARL) and SDRL of Table V.

Table IV. Corr $(R_i - \widehat{UCL},\ R_j - \widehat{LCL})$ for various values of m and n

(R chart with σ estimated by $\overline{R}/d_2$)

m \ n	2	3	4	5	6	7	8	9	10
5	.6810	.5701	.5102	.4722	.4454	.4254	.4100	.3974	.3871
10	.5163	.3987	.3424	.3091	.2865	.2702	.2579	.2480	.2400
30	.2624	.1810	.1479	.1298	.1181	.1098	.1038	.0990	.0952
100	.0964	.0622	.0495	.0428	.0386	.0357	.0336	.0319	.0306
500	.0209	.0131	.0103	.0089	.0080	.0073	.0069	.0066	.0063
∞	0	0	0	0	0	0	0	0	0

1) For each entry in Table V, m limited subgroups of size n(=5) are generated from a N(μ, σ^2) distribution where we assume that μ=0 and σ=1 without loss of generality.
2) Both $\widehat{UCL}$ and $\widehat{LCL}$ for the R chart are computed.
3) Samples are generated from a N(μ, σ_1^2) distribution, where σ_1/σ is taken to be 1, 1.1, 1.25, 2, 5 as in Table V, and the statistic R is calculated until it falls outside of control limits. Then the number of times to the out of control signal is one observation for the run length distribution.

Table V. Simulated ARL (standard error of the simulated ARL) and SDRL of R chart based on m subgroups of size n=5

m \ σ_1/σ	1.0	1.10	1.25	1.5	2.0	5.0
5	3817.5(1092.4)	1271.8(348.7)	65.8(5.9)	13.7(.9)	2.8(.1)	1.0(.0)
	34545.3	11025.3	186.2	29.7	2.7	.2
10	820.0(151.3)	216.8(27.0)	37.2(2.2)	8.7(.4)	2.7(.1)	1.0(.0)
	4785.1	854.7	68.4	11.1	2.7	.2
30	306.4(19.2)	97.7(4.2)	29.7(1.3)	7.6(.2)	2.4(.1)	1.1(.0)
	608.0	132.5	41.3	7.8	1.9	.2
100	245.5(8.9)	80.3(3.0)	23.5(.8)	7.5(.2)	2.6(.1)	1.0(.0)
	279.9	93.6	23.8	7.1	2.1	.2
500	230.4(7.4)	76.0(2.5)	24.5(.7)	7.4(.2)	2.3(.1)	1.0(.0)
	234.8	79.4	23.7	6.7	1.7	.2
∞	204.8	72.5	23.2	7.2	2.4	1.0
	204.3	72.0	22.7	6.7	1.9	0.2

4) Repeat 1), 2) and 3) until we obtain a relatively small estimated standard error of the estimate of the ARL. In Table V, the number of simulations is 1,000.

Then we can obtain simulated ARL(estimated standard error of the simulated ARL) and SDRL as described in Section 2.

For comparison purpose the ARL for infinite m was calculated using Pearson (1942). Note that the ARL based on a small number of subgroups using control limits of formula (3-3) and (3-4) is a lot larger than the ARL that would be obtained using formula (3-1) and (3-2) if the parameters are known or if there is a very large number of samples available for the construction of the chart. For example the estimated ARL when m is only 30 and when there is 25% increase in σ is 29.7 with its estimated standard error of 1.3. The value 41.3 is the simulated SDRL. On the other hand the ARL and SDRL based on known parameters or on an infinite number of subgroups are 23.2 and 22.7, respectively.

As in Table III, the ARL based on a small number of subgroups is larger than the one based on a large number of subgroups. Hence, one may want to shorten the control limits when the number of subgroups available is not large in order to get a desired level of ARL. However, it will take some time before we can get adjusted control limits using simulation.

4. Sequentially Dependent Data

Shewhart control charts have been used extensively for process control with the assumption that the sequential observations are independent. However, in practice, serial correlation is frequently not negligible. Therefore, it is important to know how the serial correlation affects the performance of control chart. The most popular method to incorporate the dependence in the data is to use a time series model. In this paper, we will consider the simple case of the first order autoregressive process.

Assume that the observation at time t is $X'(t)=X(t)+\delta\sigma_x$, where $X(t)= \phi X(t-1)+a(t)$. Note that the shift $\delta\sigma_x$ is expressed in units of the process standard deviation σ_x. The $a(t)$ may be regarded as a series of random shocks with $E(a(t))=0$ and $Var(a(t)) = \sigma^2_a$. The parameter ϕ must satisfy the condition that $-1 < \phi < 1$ for the process to be stationary. Note that we are dealing with a stationary process. The autocorrelation function of the AR(1) process is $\rho(t)=\phi\rho(t-1)$. Thus, $\rho(k)= \phi^k$, $k \geq 0$. The correlation between consecutive observations X(t-1) and X(t) is ϕ. However, if the time between the observations is k time units apart, then the correlation between them comes down to ϕ^k. The variance of the process is $\sigma_x^2 = \sigma^2_a/(1-\phi^2)$ which is a function of the parameter ϕ. Therefore, the process variance increases as the correlation ϕ increases in absolute value.

We will only investigate the performance, especially the average run length, of the X control chart when the correlation between the consecutive means can be modeled as the first order autoregressive process. In this case, the X(t) in the above model can be interpreted as the mean of a subgroup of size n. Note that the consecutive observations are not independent. Also, the events B_i and B_j defined in sections 2 and 3 are not independent either. Therefore, the distribution of the run length is not a geometric distribution. Hence, simulation is used to find the mean of the run length along with its estimated standard error. The following is the procedure used to obtain simulated ARL(estimated standard error of the simulated ARL) and SDRL of Table VI.

1) For each entry in Table VI, a value for $X(t)$ is generated from a $N(0, Var(X(t))$ distribution, where we assume that $Var(X(t)) = \sigma^2{}_a/(1-\phi^2)$ in order to get rid of the dependence on t. From then on observations for $a(t+1)$, $a(t+2), \cdots, a(t+m)$ are generated from a $N(0, \sigma_a^2)$ distribution where we assume that $\sigma_a^2=1$ and the values for $X(t+1), X(t+2), \cdots, X(t+m)$ are calculated.
2) Both $\widehat{UCL}$ and $\widehat{LCL}$ are computed using moving ranges.
3) Observations for $a(t+m+1), a(t+m+2), \cdots$ are generated and the values for $X'(t+m+1)$, $X'(t+m+2), \cdots$ are calculated and compared with $\widehat{UCL}$ and $\widehat{LCL}$ until the chart signals. This provides one observation for the run length distribution.
4) Repeat 1), 2) and 3) until we obtain a relatively small estimated standard error of the estimate of the ARL. In Table VI, the number of simulations is 10,000.

Then we can obtain simulated ARL(estimated standard error of the simulated ARL) and SDRL as in Table VI. Tables VI gives the results for the estimated ARL and SDRL for each of ϕ=0.0, 0.2, 0.5, 0.8. Each of the values for ϕ=0.0, 0.2, 0.5, 0.8 shows that the ARL tends to increase as the number of subgroups available for the construction of the charts decreases. This reflects the impact of the number of subgroups available for the construction of the chart limits. However, as seen in values for ϕ=0.8, for very highly correlated data the number of subgroups available for the construction of the chart limits does not seem to affect the performance of the chart.

Table VI essentially gives the same values of ARL and SDRL as in the Table 4 of Quesenberry (1993). The ARL based on the small number of subgroups tends to become large. For instance, the simulated ARL based on m=50 subgroups when there is no shift in the process mean is 1162.1 while the simulated ARL based on m=150 is 499.8.

The results in Table VI show that the ARL gets smaller as the correlation between consecutive observations becomes larger. For instance, the simulated ARLs based on m=2,000 observations when there is no shift in the process mean are 377.0, 139.6, 34.0, 10.1 for ϕ=0, 0.2, 0.5, 0.8, respectively. This phenomena reflects the fact that the estimate $\overline{MR}/d_2$ of the

Table VI. Simulated ARL (standard error of the simulated ARL) and SDRL of the X chart (X chart with σ estimated by $\overline{MR}/d_2$)

m \ δ	0.0	0.25	0.5	1.0	2.0	3.0
			(Ø=0)			
30	6272.3(2243.0)	4265.1(1715.5)	1740.4(567.1)	223.7(30.7)	11.1(0.3)	2.6(0.0)
	224297.3	171554.7	56710.1	3069.8	25.9	4.1
50	1162.1(91.0)	782.3((31.6)	460.3(49.7)	90.5(5.2)	8.5(0.1)	2.3(0.0)
	9096.1	3162.4	4970.7	518.1	13.0	2.2
150	499.8(11.0)	370.6(6.5)	200.3(3.2)	53.8(0.8)	6.9(0.1)	2.1(0.0)
	1097.2	645.7	322.9	80.4	7.6	1.6
500	404.9(4.7)	305.9(3.6)	168.0(1.9)	46.3(0.5)	6.5(0.1)	2.0(0.0)
	472.2	360.5	189.8	50.0	6.2	1.5
2000	377.0(3.9)	285.9(2.9)	157.6(1.6)	44.0(0.4)	6.3(0.1)	2.0(0.0)
	391.6	293.5	163.6	44.1	5.8	1.4
			(Ø=0.2)			
30	473.3(76.6)	327.8(26.3)	193.0(13.2)	49.7(2.9)	6.2(0.1)	1.9(0.0)
	7662.6	2630.7	1318.1	294.7	10.6	2.0
50	233.6(7.8)	180.6(4.2)	118.6(5.0)	33.9(.7)	5.4(.1)	1.8(.0)
	781.8	424.2	504.0	67.5	6.9	1.5
150	162.4(2.4)	132.8(1.8)	78.7(1.1)	25.6(.3)	4.7(.0)	1.7(.0)
	236.7	183.6	109.2	30.0	4.7	1.2
500	145.3(1.6)	115.5(1.3)	68.7(.7)	23.6(.2)	4.5(.0)	1.7(.0)
	161.0	128.4	73.5	24.8	4.2	1.2
2000	139.6(1.4)	113.1(1.2)	67.3(.7)	23.1(.2)	4.5(.0)	1.7(.0)
	143.3	115.6	68.6	22.9	4.1	1.2
			(Ø=0.5)			
30	40.6(.8)	37.3(.7)	29.7(.5)	14.5(.3)	3.5(.0)	1.4(.0)
	78.9	66.7	49.6	26.2	4.5	1.2
50	37.5(.5)	33.9(.5)	26.5(.4)	12.5(.2)	3.2(.0)	1.4(.0)
	53.8	47.2	37.9	16.7	3.7	1.0
150	35.7(.4)	31.2(.3)	23.7(.3)	11.2(.1)	3.1(.0)	1.4(.0)
	38.7	34.2	25.8	12.4	3.2	.9
500	34.2(.4)	30.6(.3)	22.7(.2)	10.7(.1)	2.9	1.3(.0)
	35.2	31.9	23.1	10.9	2.9	.9
2000	34.0(.3)	30.8(.3)	22.3(.2)	10.4(.1)	2.9(.0)	1.3(.0)
	33.9	31.2	22.0	10.3	2.8	.9
			(Ø=0.8)			
30	9.9(.1)	9.5(.1)	8.6(.1)	6.2(.1)	2.2(.0)	1.1(.0)
	11.6	11.4	11.0	8.6	3.2	.8
50	9.6(.1)	9.5(.1)	8.5(.1)	5.8(.1)	2.1(.0)	1.1(.0)
	10.9	10.6	9.8	7.6	2.9	.6
150	10.1(.1)	9.5(.1)	8.4(.1)	5.5(.1)	2.0(.0)	1.1(.0)
	10.8	10.2	9.4	6.8	2.5	.6
500	10.0(.1)	9.7(.1)	8.6(.1)	5.6(.1)	1.9(.0)	1.1(.0)
	10.6	10.2	9.3	6.9	2.3	.6
2000	10.1(.1)	9.7(.1)	8.5(.1)	5.5(.1)	1.9(.0)	1.1(.0)
	10.5	10.5	9.6	6.6	2.4	.6

process standard deviation σ_x when there is correlation in the data underestimates the process standard deviation σ_x. In fact, it is known that $E(\overline{MR}/d_2)=\sqrt{1-\phi^2}\sigma_x$.

For practical use of the results in Table VI we have to know (or estimate) ϕ for the sequential observations, and based on the number of subgroups available we can have an idea of the size of the ARL for in control condition. Next, if the size of the in control ARL is too small then we can entend the control limits a little bit so that we can obtain a desirable level of the in control ARL. However, it will take some time before we can find the right level of the control limts since we are now using simulation methods.

Table VII. ARL of X chart with control limits of $\mu \pm 3\sigma_x$

ϕ \ δ	0.0	0.25	0.5	1.0	2.0	3.0
0	370.4	281.2	155.2	43.9	6.3	2.0
0.2	372.7	284.1	158.3	45.8	6.9	2.2
0.5	396.3	307.0	176.3	54.4	8.9	2.6
0.8	555.2	444.6	271.1	92.4	16.3	4.0

So far it is assumed that we have a limited number of observations and use $\overline{MR}/d_2$ which is a biased estimate of the process standard deviation σ_x. However, we may be able to use an unbiased estimate of σ_x such as the sample standard deviation. In this case as the number of observations increases the sample standard deviation tends to σ_x. Therefore, with infinite number of observations and with an unbiased estimate of σ_x, we can construct the control limits of UCL= $\mu+3\sigma_x$ and LCL= $\mu-3\sigma_x$. In this case a Markov chain representation can be used to obtain the ARL of the chart. Table VII, which can be found in Baik (1991), gives the ARLs for each value of ϕ=0, 0.2, 0.5, 0.8. Note that the ARLs for ϕ=0, 0.2, 0.5, 0.8 when there is no shift in the process mean are 370.4, 372.7, 383.5, 419.4 which is increasing as ϕ increases. This is what we would expect if we have positively correlated data and if an unbiased estimate of σ_x is used for the construction of the chart limits. However, this does not happen when we estimate the parameter σ_x with a biased estimate $\overline{MR}/d_2$ as seen in the last rows of the values for ϕ=0, 0.2, 0.5, 0.8.

5. Conclusions

We have considered some aspects of $\overline{X}$, R and X control charts. In particular, we have looked at the ARLs of each chart assuming that the process parameters are not known and that a limited number of subgroups are used to construct the control limits. We have found the following.

The performance of the $\overline{X}$ chart with process standard deviation σ estimated by $\overline{R}/d_2$ is very similar to that with σ estimated by $\overline{S}/c_4$ in Quesenberry (1993). However, we present the ARLs in Table III in terms of the shift in units of the standard deviation of $\overline{X}$. In Table III it is found that the ARL and SDRL get larger as the number of subgroups used for the construction of the chart becomes smaller.

The performance of R chart with estimated control limits based on a small number of subgroups is very much different from what is expected with a large number of subgroups or with known process parameters. Specifically, the ARL and SDRL get larger as the number of subgroups used for the construction of the chart becomes smaller.

The performance of X chart with estimated parameters is different from that with known parameters. Especially if the number of observations for the construction of the chart is small the ARL tends to increase. However, when there is correlation in the data the ARL gets smaller as the correlation between consecutive observations becomes larger due to the underestimation of the process standard deviation with moving ranges. On the other hand, if an unbiased estimate such as sample standard deviation is used then the ARL gets larger as the correlation between consecutive observations increases.

So far we have found that during the start-up of a process or in a job-shop environment the use of control charts may lead to erroneous results. Hence, there has to be some adjustment in the determination of control limits. In addition, there may have to be some other alternatives to the application of control charts (see Q charts (1995), Seppala, et al. (1995), Castillo and Montgomery (1995) to name a few).

References

[1] Baik, JaiWook. (1991), "$\overline{X}$ control charts in the presence of correlation," Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

[2] Castillo, E. D. and Montgomery, D. C. (1995), "A Kalman filtering process control scheme with an application in semiconductor short run manufacturing," *Quality and Reliability Engineering International*, Vol 11, 101-105.

[3] Ghosh, B. K., Reynolds, M. R. and Hui, Y. V. (1981), "Shewhart $\overline{X}$ charts with estimated process variance," *Communications in Statistics-Theory and Methods*, A10(18), 1797-1822.

[4] Hillier, F. S. (1964), "$\overline{X}$ control limits based on a small number of subgroups," *Journal of Quality Technology*, January, 17-26.

[5] Hillier, F. S. (1969), "$\overline{X}$ and R control limits based on a small number of subgroups," *Journal of Quality Technology*, January, 24-29.

[6] Montgomery, D. C. (1991), Introduction to Statistical Quality Control, 2nd Edition, Wiley & Sons, New York, NY.

[7] Pearson, E. S. (1942), "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32, 301-310.

[8] Quesenberry, C. P. (1993), "The effect of sample size on estimated limits for $\overline{X}$ and X control charts," *Journal of Quality Technology*, Vol. 25, 237-247.

[9] Quesenberry, C. P. (1995), "Geometric Q charts for high quality processes," *Journal of Quality Technology*, Vol. 27, 304-315.

[10] Seppala, T, Moskowitz, H., Plante, R and Tang, J. (1995), "Statistical process control via the subgroup bootstrap," *Journal of Quality Technology*, Vol. 27, 139-153.