The Korean Communications
in Statistics Vol. 5, No. 2, 1998
pp. 373-381

A Statistical Control Chart for Process with Correlated Subgroups

Kwang Ho Lee[1)]

Abstract

In this paper a new control chart which accounts for correlation between process subgroups will be proposed. We consider the case where the process fluctuations are autocorrelated by a stationary AR(1) time series and where $n(\geq 1)$ items are sampled from the process at each sampling time. A simulation study is presented and shows that for correlated subgroups, the proposed control chart makes a significant improvement over the traditionally employed X-bar chart which ignores subgroup correlations. Finally, we illustrate the proposed chart by comparing the standardized residuals and X-bar chart on a data set of motor shaft diameters.

1. Introduction

Statistical process control techniques are widely used in industry for process monitoring and quality improvement. Various statistical control charts have been developed to monitor the process mean and variance. Traditional statistical process control methodology is based on the independence assumption between the process subgroups. However the process subgroups are not always statistically independent from each other. In the continuous industries most process data are autocorrelated. For example, current automated measurement and recording technology, subgroup samples may be taken with high frequency, and with consecutive samples being similar in nature. The motor shaft diameter data of Devor *et al.*(1992) shows that the autocorrelation between process subgroups occur when items made by a worker exhibit similar characteristic due to the way that the machine is handled, and the process data shows seasonal patterns due to materials or weather, or the alertness of a work changes over time.

It has been shown that subgroup correlations markedly affect the performance of a Shewhart control chart. Padgett *et al.*(1992) demonstrated that the true probabilities of at least

1) Professor, Department of Statistics, Yeungnam University, Kyungsan, Kyungsangbukdo, 712-749, Korea

one sample statistic exceeding the control limits differ from those calculated under the assumption of independent subgroups. Many authors including Spurrier and Thombs (1990), Harris and Ross(1991), Alwan(1992), Yaschin(1993), Superville and Adams (1994), Wardell *et al.*(1994), Tatum(1996) and Zhang(1998) investigated the control charts for correlated process. But, most of the above authors investigate the case where one measurement is sampled from the process at each sampling time. In these cases, a typical approach to monitoring correlated processes is to fit an ARIMA time series model to the process data and then monitor one-step-ahead forecasts with a control chart such as X-bar chart, CUSUM chart.

In this paper, a subgroup of $n(\geq 1)$ measurements is sampled from the process at equally spaced points in time. We will propose and investigate a new control chart for this situation. The performance of the proposed chart will be compared to the standard Shewhart X-bar chart which ignores subgroup correlations. The proposed control chart will be developed in section 2. Through the simulation study, its performance will be investigated in section 3. An example will be presented in section 4. Finally, we summmarize this paper with conclusions.

2. A Proposed Control Chart

Suppose that at each time t , a measurement sampled from a process is given by

$$X_{tj} = \mu_t + \varepsilon_{tj}, \quad t=1,2,\cdots,k \ , \quad j=1,2,\cdots,n, \tag{1}$$

where $\{\mu_t\}$ and $\{\varepsilon_{tj}\}$ are independent zero mean Gaussian processes. Here, μ_t is mean of process at time t that fluctuates randomly with t , and $\varepsilon_{tj}(j=1,2,\cdots,n)$ are measurement errors at time t. We also assume that variance of ε_{tj} is σ_X^2. Thus conditional upon the process mean is μ_t at time t, the subgroup sample $\{X_{tj,}\ 1\leq j\leq n\}$ is an independent identically distributed Gaussian random sample with mean μ_t and variance σ_X^2.

The process fluctuation $\{\mu_t\}$ is assumed a stationary causal first order autoregressive (AR(1)) time series which is defined by

$$\mu_t = \emptyset \mu_{t-1} + a_t \tag{2}$$

where $|\emptyset|<1$ is assumed for stationarity and causality of $\{\mu_t\}$, and $\{a_t\}$ is an independent identically distributed mean zero Gaussian random sequence with variance σ_a^2 that is independent of $\{\varepsilon_{tj}\}$.

Since $E[X_{tj}]=0$ for all t and j, the correlation between the process measurements at time t and s $(s\neq t)$ is given by

$$Corr(X_{ti}, X_{sj}) = \frac{\emptyset^{|t-s|}\sigma_a^2}{\sigma_a^2 + (1-\emptyset^2)\sigma_X^2}$$

when $1 \le i, j \le n$. This correlation is nonzero whenever $\varnothing \neq 0$. From this, the model (1) allows for correlated subgroups. Moreover, the correlation between process subgroups decays at a geometric rate to zero as the lag between subgroups approaches infinity. From $Corr(X_{ti}, X_{tj}) \neq 0$ for $i \neq j$, we note that the subgroup $\{X_{tj},\ 1 \le j \le n\}$ is not independent. In fact, it is only conditionally independent.

Now, we define the subgroup measurement mean and error mean at time t ,respectively, as follows

$$\overline{X}_t = \sum_{j=1}^{n} \frac{X_{tj}}{n}, \qquad \overline{\varepsilon}_t = \sum_{j=1}^{n} \frac{\varepsilon_{tj}}{n}.$$

Then from the equations (1) and (2), we obtain

$$\overline{X}_t = \mu_t + \overline{\varepsilon}_t, \text{ and } \mu_t = \varnothing \mu_{t-1} + a_t , \tag{3}$$

which is the Kalman filter equations. Manipulation with (3) provide

$$\overline{X}_t - \varnothing\ \overline{X}_{t-1} = \overline{\varepsilon}_t - \overline{\varepsilon}_{t-1} + a_t .$$

That is, $\{\overline{X}_t\}$ is ARMA process. Although $\{\overline{X}_t\}$ is stationary, it is not easy to fit the process measurements. Thus we will use the Kalman filtering forecasting techniques to develope a control chart for this process. The first two moments of $\{\overline{X}_t\}$ are easily obtained as follows.

$$E[\overline{X}_t] = 0 \quad , \text{ and } \quad Var[\overline{X}_t] = \frac{\sigma_a^2}{1-\varnothing^2} + \frac{\sigma_X^2}{n} .$$

Let $\widehat{\mu_t} = E[\mu_t \mid \overline{X}_1, \overline{X}_2, \cdots, \overline{X}_t]$ be the best mean squared error predictor of μ_t based upon $\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_t$ and define its prediction error as

$$v_t^2 = E[(\widehat{\mu_t} - \mu_t)^2 \mid \overline{X}_1, \overline{X}_2, \cdots, \overline{X}_t] . \tag{4}$$

From the Kalman recursion(Brockwell and Davis, 1996), the best mean square error predictor of μ_t is given by

$$\widehat{\mu_t} = \varnothing\ \widehat{\mu}_{t-1} + k_t(\overline{X}_t - \varnothing\ \widehat{\mu}_{t-1}) , \tag{5}$$

where

$$k_t = \frac{\varnothing^2 v_{t-1}^2 + \sigma_a^2}{n^{-1}\sigma_X^2 + \varnothing^2 v_{t-1}^2 + \sigma_a^2} . \tag{6}$$

To compute the quantities in (4) ~ (6) iteratively in t, we update v_t^2 with

$$v_t^2 = \frac{n^{-1}\sigma_X^2(\varnothing^2 v_{t-1}^2 + \sigma_a^2)}{n^{-1}\sigma_X^2 + \varnothing^2 v_{t-1}^2 + \sigma_a^2} \tag{7}$$

and begin the iterations with $\widehat{\mu}_0 = 0$ and $v_0^2 = \sigma_a^2 / (1 - \varnothing^2)$.

Now, as the one step ahead prediction residual for process sample means at time t, we can define the control chart as follows

$$R_t = \overline{X}_t - E[\overline{X}_t \mid \overline{X}_1, \overline{X}_2, \cdots, \overline{X}_{t-1}] \\ = \overline{X}_t - \varnothing\ \hat{\mu}_{t-1} \tag{8}$$

A large R_t in absolute value is associated with a process that is out of control. From (3) and (8),

$$R_t = \mu_t - \varnothing\ \hat{\mu}_{t-1} + \overline{\varepsilon}_t \tag{9}$$

This $\{R_t\}$ is also mean zero Gaussian process and the conditional variance of R_t given $\overline{X}_1, \cdots, \overline{X}_{t-1}$ obtain as follows.

$$\sigma_{R_t}^{\ 2} = Var[R_t \mid \overline{X}_1, \overline{X}_2, \cdots, \overline{X}_{t-1}] = \sigma_a^2 + \varnothing^2 v_{t-1}^2 + \sigma_X^2/n\ . \tag{10}$$

In fact, $\sigma_{R_t}^{\ 2}$ does not depend on $\overline{X}_1, \cdots, \overline{X}_{t-1}$. Thus the standardized residuals,

$$Z_t = R_t / \sigma_{R_t} \tag{11}$$

are independent identically standard normal random variables. If all values of the parameters are known, $\{Z_t\}$ can be used to monitor the process.

This monitoring method is the proposed control chart for process. Although the values of the parameters are seldom known in practice, it is not difficult to estimate the parameters by using the maximum likelihood method. The usual 3σ control limits can be applied to this control chart. That is, the process is out of control at time t if $|Z_t| > 3$. It is obvious that the probability of a "false" out of control signal at any time t is $P[|Z_t| > \varepsilon] \approx 0.0027$.
To investigate the performance of the proposed control chart we compare our results to those in Padgett *et al.*(1992) by using the α_k-risks. The α_k-risk is the probability of obtaining at least one false out of control signal in a process where k total subgroups, each of size n, are observed when the process is actually in control. Hence the α_k-risk can be viewed as the overall false alarm rate for a control chart of k subgroups and is analogous to a type I error probability.

Computations with the standard normal distribution show that $\alpha_1 \approx 0.0027$. If the parameters are known and the subgroups are independent, then α_k-risk can be computed from the binomial distribution. That is,

$$\alpha_k = 1 - (1 - \alpha_1)^k \tag{12}$$

Padgett *et al.*(1992) demonstrated the gross inadequacies of using (12) when process subgroups are correlated.

3. Simulation Study

In this section we investigate the performance of the proposed control chart through the Monte Carlo simulation. To study a variety of correlation properties between the subgroups, we examine four different parameter sets of $\emptyset$, σ_a^2 and σ_X^2. Sample α_k-risks will be simulated for a variety of choices of the subgroup numbers k and the subgroup size n, and compared to the results reported in Padgett *et al.*(1992) where a Shewhart chart was applied to correlated subgroups without accounting for subgroup correlations.

Table 1 shows sample α_k-risks for the parameters $\emptyset = 0.5$, $\sigma_a^2 = 0.375$, $\sigma_X^2 = 0.50$ and for each and $k = 20, 25, 30$ and 50 and $n = 3, 4, 5$ and 6. The values of σ_a^2 and σ_X^2 were selected to make $Var(\overline{X}_t) = 1$ when $n = 1$. And the values of n and k reflect subgroup sizes that are typically encounted in practice. All sample α_k-risks are computed from 2000 simulations. The sample α_k-risks are the proportion of simulations where Z_t exceeds 3 in absolute value for at least one t satisfying $1 \le t \le k$.

We first note that sample α_k-risks computed with the exact value of the parameters are very closed to the theoretical α_k-risks for normal data in Table 5. Hence, we find that the proposed control chart is functioning well. We contrast this results to Table 6 - 10 of Padgett *et al.*(1992). Their tables show that the sample α_k-risks are closed to unity when a traditional Shewhart X-bar chart, unmodified for subgroup correlation, is used. From these facts, we obtain that the proposed control chart shows a clear improvement over a traditional Shewhart X-bar chart when subgroups are correlated.

Table 2 - 4 show sample α_k-risks for the same values of k and n and different choices of $\emptyset$, σ_X^2 and σ_a^2. These three tables show a similar structure to Table 1. The value $\emptyset = -0.5, -0.8$ in Table 2 and Table 4 provides negative correlation between subgroups.

4. An Example

In this section, we apply the proposed control chart to the motor shaft diameter data of Devor *et al.*(1992, pp 186-188). This data has process subgroups of size $n = 5$ observed at $k = 60$ sampling times separated by 30 minutes each. Figure 1 plots the sample mean motor shaft diameter at each sampling time and the traditional Shewhart X-bar chart upper and lower control limits, computed assuming independent subgroups. From Figure 1, we know that process sample mean fall out side of the control limits for subgroups 28 and 30. Hence a

traditional Shewhart X-bar chart would signal "out of control" for this process.
To fit the proposed model (3) to this data, it need to know the values of the parameters. Since {Zt} are independent identically standard normal random variable, we obtain the log likelihood function as follows.

$$\ln L(\mu, \phi, \sigma_a^2, \sigma_X^2 \mid \overline{X}_1, \cdots, \overline{X}_n) = -\ln(2\pi)^{k/2} - \sum_{t=1}^{k} \sigma_{R_t}^2 - \frac{1}{2}\sum_{t=1}^{k} z_t^2$$

From above equation, we can obtain the maximum likelihood estimates $\widehat{\mu}$, $\widehat{\phi}$, $\widehat{\sigma_a}$ and $\widehat{\sigma_X^2}$ by numerical method. For computationally convenience, we use the EM algorithm. For the motor shaft diameter data, we obtain the parameter estimates $\widehat{\mu}=48.31$, $\widehat{\phi}=0.74$, $\widehat{\sigma_a^2}=2.59$ and $\widehat{\sigma_X^2}=38.62$. The value of $\widehat{\phi}$ strongly suggest that the process subgroups exhibit positive correlation.
To assess the performance of the proposed process model for this data, we examine the estimated standardized residuals $\{\widehat{Z}_t\}$ plotted in Figure 2. The estimated standardized residuals in Figure 2 all lie inside the control limit ± 3 . Hence, contrary to a traditional Shewhart X-bar chart, the proposed chart for correlated subgroups does not signal that the process is out of control. We also note that the estimated standardized residuals in Figure 2 follow a similar pattern as the Shewhart X-bar chart in Figure 1, but have a smaller variability.

5. Remarks and Conclusions

This paper presents a new control chart that account for subgroup correlations in the case where multiple observations are sampled from the process at each sampling time. Specifically, the case where process mean fluctuations are governed by an AR(1) time series is explored. Extensions to more general time series may be considered.

This paper also show that by modeling rather then ignoring subgroup correlations, a more efficient control chart can be achieved. This improved chart acts to greatly minimized the number of unwarranted process shut-down. From the simulation study, the proposed control chart greatly improved upon a Shewhart X-bar chart that ignores subgroup correlations.

Table 1 α_k-risks for $\varnothing = 0.5$, $\sigma_X^2 = 0.50$, $\sigma_a^2 = 0.375$

n \ k	20	25	30	50
3	0.0435	0.0710	0.0830	0.1255
4	0.0520	0.0645	0.0730	0.1410
5	0.0575	0.0615	0.0855	0.1210
6	0.0480	0.0715	0.0840	0.1185

Table 2 α_k-risks for $\varnothing = -0.5$, $\sigma_X^2 = 0.50$, $\sigma_a^2 = 0.375$

n \ k	20	25	30	50
3	0.0465	0.0665	0.0770	0.1120
4	0.0505	0.0650	0.0750	0.1265
5	0.0605	0.0750	0.0755	0.1335
6	0.0540	0.0645	0.0910	0.1225

Table 3 α_k-risks for $\varnothing = 0.8$, $\sigma_X^2 = 0.50$, $\sigma_a^2 = 0.315$

n \ k	20	25	30	50
3	0.0495	0.0580	0.0830	0.1190
4	0.0645	0.0665	0.0745	0.1205
5	0.0470	0.0800	0.0735	0.1140
6	0.0470	0.0665	0.0795	0.1195

Table 4 α_k-risks for $\varnothing = -0.8$, $\sigma_X^2 = 0.50$, $\sigma_a^2 = 0.315$

n \ k	20	25	30	50
3	0.0505	0.0645	0.0750	0.1310
4	0.0435	0.0770	0.0840	0.1195
5	0.0470	0.0600	0.0890	0.1375
6	0.0640	0.0645	0.0775	0.1115

Table 5 Theoretical α_k-risks.

k	20	25	30	50	100
α_k	0.05263	0.06535	0.07790	0.12643	0.23688

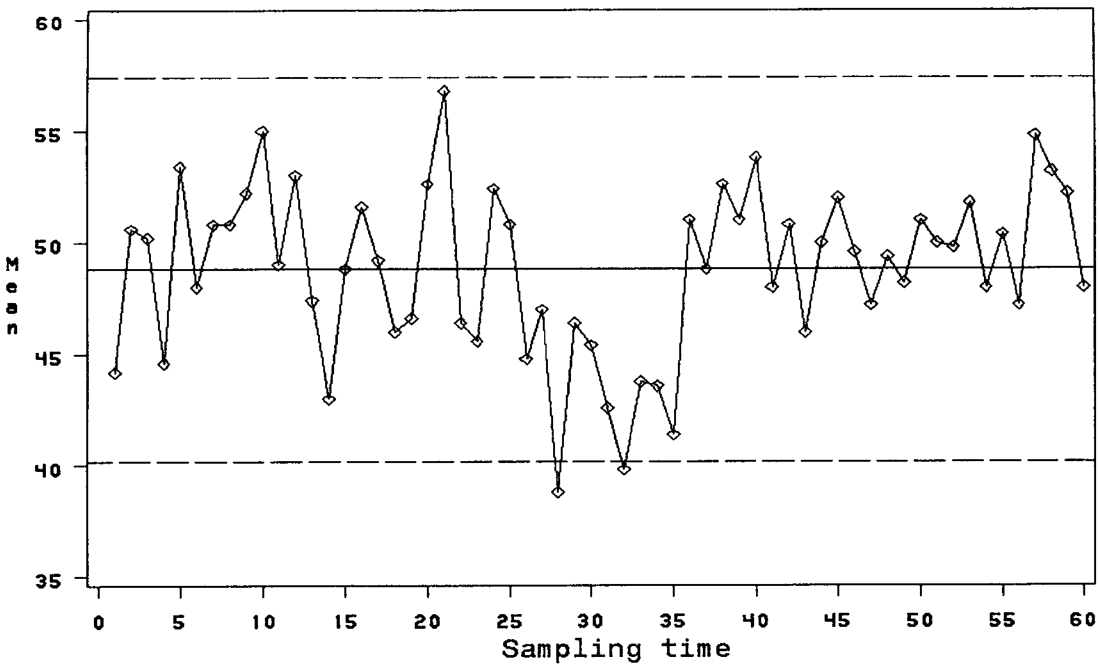

Figure 1. Shewhart X-bar chart for motor shaft diameter data.

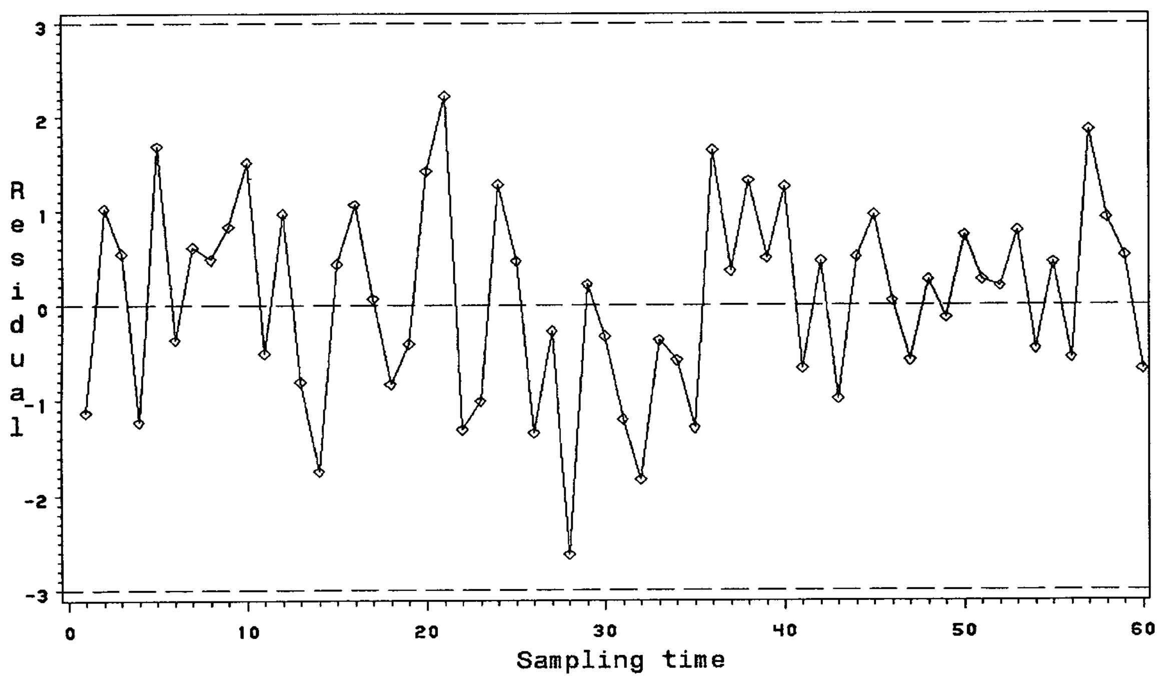

Figure 2. Estimated standardized residuals chart

References

[1] Alwan, L.C.(1992). Effects of autocorrelation on control chart performance, *Communications in Statistics, Theory and Methods,* 21, 1025-1049.

[2] Brockwell, P.J. and Davis, R.A.(1996). *Introduction to Time Series and Forecasting,* Spring-Verlag, New York.

[3] Dever, R.E., Chang, T.H. and Sutherland, J.W.(1992). *Statistical Quality Design and Control,* Macmillan, New York.

[4] Harris, T.J. and Ross, W.H.(1991). Statistical process control procedure for correlated observation, *Canadian Journal of Chemical Engineering,* 69, 48-57.

[5] Padgett, C.S., Thombs, L.A. and Padgett, W.J.(1992). On the α-risks for Shewhart control charts, *Communications in Statistics, Simulation and Computation,* 21, 1125-1147

[6] Spurrier, J.D. and Thombs, L.A.(1990). Control charts for detecting cyclical behavior, *Technometrics,* 32, 163-171.

[7] Superville, C.R. and Adams, B.M.(1994). An evaluation of forecast-based quality control schemes, *Communications in Statistics, Simulation and Computation,* 23, 645-661.

[8] Tatum, L.G.(1996). Control Charts for Detection of a Periodic Component, *Technometrics,* 38, 152-160.

[9] Wardell, D.G., Moskowitz,H. and Plante, R.D.(1994). Modification of control chart limits in the presence of data correlation, *Journal of Quality Technology,* 10, 20-30.

[10] Yaschin, E.(1993). Performance of CUSUM control schemes for serially correlated observations, *Technometrics,* 35, 37-52.

[11] Zhang, N.F.(1996). A statistical Control Chart for Stationary Process Data, *Technometrics,* 40. 24-38.