⊠연구논문

Control Charts Based on Self-critical Estimation Process

Won, Hyung Gyoo
Dept. of Industrial Eng., Hansung University

Abstract

Shewhart control chart is a basic technique to monitor the state of a process. We observe samples of size four or five and plot some statistic(e.g., mean or range) of each sample on the chart. When setting up the chart, we need to obtain upper and lower control limits. It is common practice that those limits are calculated from the preliminary 20-40 samples presumed to be homogeneous. However, it may happen in practice that the samples are contaminated by outlying observations caused by various reasons. The presence of outlying observations make the control limits wider and hence decrease the sensitivity of the charts. In this paper, we introduce robust control charts with tighter control limits when outlying observations are present in the preliminary samples. Examples will be given via simulation study.

1. Introduction

Shewhart control chart is one of the most powerful technique to analyze the behavior of any process. We collect a series of a subgroup of size four or five in a rational way from the process and plot some statistic(e.g., mean or range), obtained from each of the subgroup, on the chart. Serially plotted points in such a way in time order reveal various information about the process through many types of patterns. If points scatters naturally in some statistical law, we say the process is in statistical control, and if there are unnatural patterns, we say the process is out of control. In the state of statistical control, we may expect the process produce more uniform products.

Among the various unnatural patterns, one basic unnatural one is the point falling beyond some statistical limits. To support our decision on this pattern, three statistical lines are drawn on the control chart. They are the upper control limit, the lower control limit, and the center line for process average. Since the three lines play important roles in interpreting and deciding other natural or unnatural patterns also, the usefulness of the chart much depends on their correct setting.

It is a common practice to set the positions of the lines from a series of preliminary 20-40 groups of samples, which are obtained from presumably in-control state of process. In other words, the assumption of in-control implies statistical naturalness or homogeneity of the set of data. However, when constructing the chart either for a new process or for any existing process for the first time because of quality problems, it is practically difficult to assume the process is in statistical control. In addition, errors of measurement or sampling due to lack of experience may happen when collecting and recording the data. Therefore, it is very likely that the preliminary set of data may contain some nonhomogeneous observations from group to group or within each group. They are different because they generally pose themselves far from the other most of the observations. Therefore, we sometimes call them outliers.

When outliers are present, they tend to set the control limits wider through the inflated statistics and estimates of parameters required for obtaining the control limits. Then the control chart loses its sensitivity against unnatural patterns. The usual methods of handling outliers are either to get rid of them and plot again, or to winsorize observations in each group before estimating the required parameters, or to use some robust estimators such as median[1, 2, 3, 4, 7, 9, 10]. In this paper, we use a robust estimating procedure, called self-critical estimates, for parameter estimation and suggest modified $\overline{x}$-S type control charts with the control limits set by the self-critical estimates.

2. Control chart and morphosis

Since W.A. Shewhart proposed the general theory of the control chart, the chart has the general format with the center line(CL), the upper control limit(UCL), and the lower control limit(LCL) as follows:

$$\text{UCL} = \mu_W + k\sigma_W$$
$$\text{CL} = \mu_W$$
$$\text{LCL} = \mu_W - k\sigma_W$$

where W is a sample statistic measuring some characteristic of our interest about the process. μ_W is the mean of W and σ_W is the standard deviation of W. k is the distance of control limits from the center line in the units of standard deviation of W.

W is the statistic reacting very sensitively against any disturbances in the process average and spread. Generally, for the problem of process shift in average, we use the sample mean, $\bar{x}$(or x-bar for notational convenience), and for the spread we use either the sample range, R, or the sample standard deviation, S.

When outliers are present in any preliminary subgroup of observations, these statistics lose their statistical efficiency very quickly[Huber, 1981]. It is our belief that the fact that they react very sensitively against any change in the process is very useful in the purpose of identifying any unnatural pattern in the on-going process, but is inappropriate for the purpose of setting the control limits. Possible outliers in the preliminary subgroups may inflate the width of control limits and hence decrease the sensitivity of the control charts when the process is unduly disturbed by some causes. The best compromise in this contradictory situation is this: first, use robust estimators when estimating the parameters from the preliminary samples and thereby keep the control limits tight and second, plot the sensitive statistics on this charts. In the following sections, we will show this strategy is useful for detecting some unnatural patterns, especially arising from sampling error. To obtain robust estimates we use the self-critical estimates.

3. Self-critical estimation process

Self-critical estimate is obtained by maximizing the objective function, ℓ_c,

$$\ell_c(\theta) = (1/c)\sum_{i=1}^{n} f^c(x_i, \theta)/Q^{c/(1+c)} - 1$$

where

$$Q = Q(\theta, c) = \int_{-\infty}^{\infty} f^{1+c}(x_i, \theta)dx_i$$

Equivalently, by taking derivative of $\ell_c(\theta)$ with respect to θ and setting them equal to zero, the estimates are obtained by solving the estimating the equation

$$\sum_{i=1}^{n} f^c(x_i, \theta)\left\{(1+c)\frac{\partial \log f(x_i, \theta)}{\partial \theta} - \frac{\partial \log Q(\theta, c)}{\partial \theta}\right\} = 0.$$

When c=0, the objective function $\ell_c(\theta)$ is equivalent to the likelihood function and hence is so for the estimating equation. The estimator, $\hat{\theta}$(c), resulting from the estimating equation, is a robust estimator and has an unbiased score. The reason that the specific reality of this estimator is called self-critical estimate is because the estimate is obtained in such a way that best optimize the information in the data in face to the assumed model. We are not forgiving any single observation, but instead, we use it as much as it fits the model[8].

For the normal distribution with its density form

$$f(x:\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = 0,$$

we have estimating equations

$$\sum_{i=1}^{n}(x_i - \mu) e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = 0$$

$$\sum_{i=1}^{n}\{(1+c)(x_i-\mu)^2 - \sigma^2\} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = 0.$$

The self-critical estimators, $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$, are obtained from the implicit equations

$$\mu = \frac{\sum x_i v_{ic}}{\sum v_{ic}}$$

$$\sigma^2 = (1+c)\frac{\sum (x_i - \mu)^2 v_{ic}}{\sum v_{ic}}$$

where

$$v_{ic} = e^{-\frac{c}{2}(\frac{x-\mu}{\sigma})^2}.$$

For c in a neighborhood of zero, the estimators $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$ are M-estimators and consistent for μ and σ^2, when the $x_1, x_2, \ldots, x_n$ are a random sample from a normal distribution with mean μ and variance σ^2, provided the consistent zeros of estimating equations are chosen. It also has been shown that $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$ are asymptotically independently distributed. The asymptotic efficiencies remain high over the range of values of c. The influence function for $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$ are bounded and redescended to zero for all c>0.

4. self-critical $\bar{x}$-S control charts

$\bar{x}$-S control charts can be used without much difficulty in plants equipped with computer facility. They are preferable to the popular $\bar{x}$-R charts in statistical points of view, especially when the sample size is moderately large, say greater than 10. We obtain the 3-sigma control limits with the center line for the $\bar{x}$ chart as follows:

$$\text{UCL} = \bar{\bar{x}} + A_3 \bar{S}$$
$$\text{CL} = \bar{\bar{x}}$$
$$\text{LCL} = \bar{\bar{x}} - A_3 \bar{S}$$

where $\bar{\bar{x}}$ represents mean of the subgroup(sample) means, $\bar{x}_i$, i=1, m and $\bar{S}$ mean of the subgroup standard deviations, S_i, i=1, m, where m is the number of subgroups. Also A_3 is a constant read from the table as a function of the size of subgroup. In the same context, the lines in the S chart is obtained from the following equations:

$$\text{UCL} = B_4 \bar{S}$$
$$\text{CL} = \bar{S}$$
$$\text{LCL} = B_3 \bar{S}$$

Also, B_3 and B_4 are constants which can be read in a table as a function of the size of each subgroup.

While the parameters obtained in this way for the control limits have such a good statistical property of unbiasness and efficiency for the process mean and standard deviation, their influence functions are unbounded in the presence of outliers. When outliers are present, they lose all the statistical good properties quite easily[Huber, 1981]. If there is any outlier-generating mechanism within subgroups and its influence prolongs over the period of collecting preliminary set of data, the sensitivity of $\bar{x}$ and S by the mechanism is so great that the control limits will be enlarged. The same scenario will take place more severely for the $\bar{x}$-R charts.

Therefore, we propose replacing $\bar{x}$ and S with corresponding self-critical estimators, $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$ defined in section 3. Since they have robustness against outliers, we expect tighter control limits. The self-critical $\bar{x}$ and S control charts, symbolized with $\bar{x}(c)-S(c)$, are defined as follows. In $\bar{x}(c)$ chart, control limits are obtained in the following equations:

$$\mathrm{UCL} = \bar{\mu}_{(c)} + A_3 \bar{\sigma}_{(c)}$$
$$\mathrm{CL} = \bar{\mu}_{(c)}$$
$$\mathrm{LCL} = \bar{\mu}_{(c)} - A_3 \bar{\sigma}_{(c)}$$

where A_3 is the same constant as in equation for $\bar{x}$-S charts. In S(c) chart, the limits are calculated similarly as follows:

$$\mathrm{UCL} = B_4 \bar{\sigma}_{(c)}$$
$$\mathrm{CL} = \bar{\sigma}_{(c)}$$
$$\mathrm{LCL} = B_3 \bar{\sigma}_{(c)}$$

Once we have the limits, we plot sample mean and sample standard deviation of each subgroup as in the $\bar{x}$-S control charts. Since $\bar{x}$-S control limits are easily obtained in the process of obtaining self-critical estimates by setting self-critical coefficient to zero, we can easily identify any different behavior of patterns between the two charts. We can summarize the steps for constructing the suggested $\bar{x}$(c)-S(c) control charts as follows:

Initial step : Obtain 20-40 subgroups of appropriate size.
Decide the value of the self-critical coefficient, c.

Step 1: Obtain self-critical estimates, $\mu_i(c)$ and $\sigma_i(C)$, for each subgroup i, i = 1, m, for c=0 and c given in the initial step.

Step 2: Estimate parameters by taking averages of self-critical estimates obtained at step 1.

Step 3: Calculate control limits of $\bar{x}$(c)-S(c) control charts.

Step 4: Plot $\mu_i(c=0)$ and $\sigma_i(c=0)$, for i = 1, m

These steps require more computational effort of programming than the traditional control charts. But with easy access to the computational power in recent organizational environment, it is a tradeoff between computational effort and increased sensitivity of the suggested control charts.

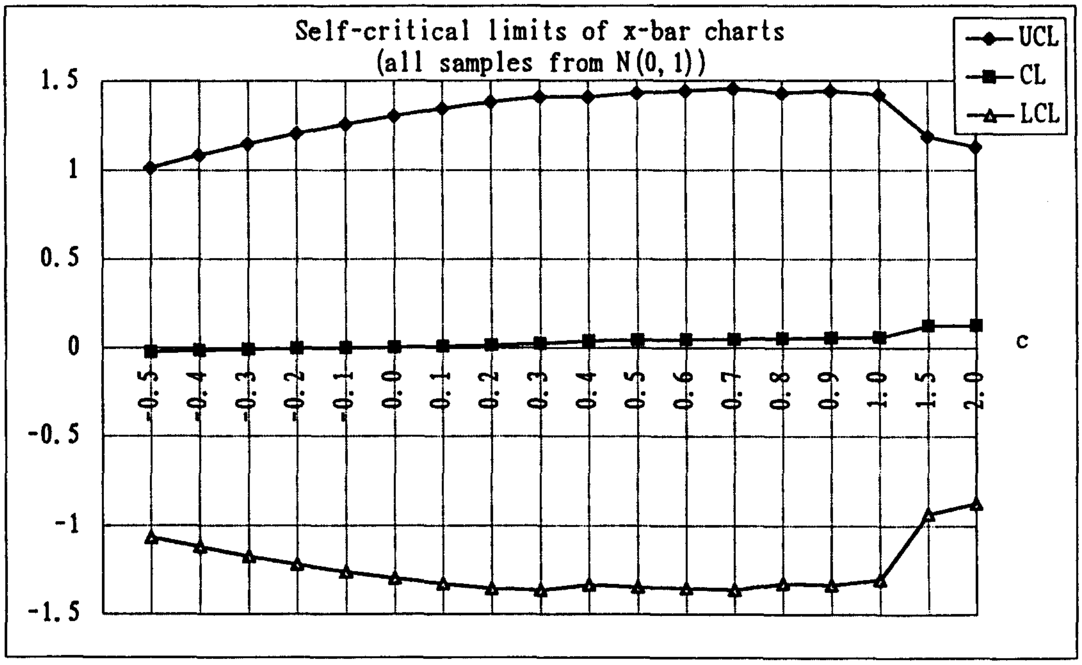
5. Examples by simulation

We compare the $\bar{x}$(c)-S(c) control charts with the popular $\bar{x}$-R and $\bar{x}$-S control charts by using Monte Carlo simulations. We simulate preliminary set of data in the format of 25 subgroups with each size of 5 from normal random numbers using the function random_normal() of IMSL(1994) with the initial seed of 123457. We consider two examples, one for all preliminary data generated from standard normal distribution and the other for the case of stratification. Stratification is a type of pattern in which points are clustered around the center line and few points appear near the control limits. This phenomenon happens due to systematic sampling from two or more different causes of systems[11].

Example 1. all preliminary data generated by N(0, 1)

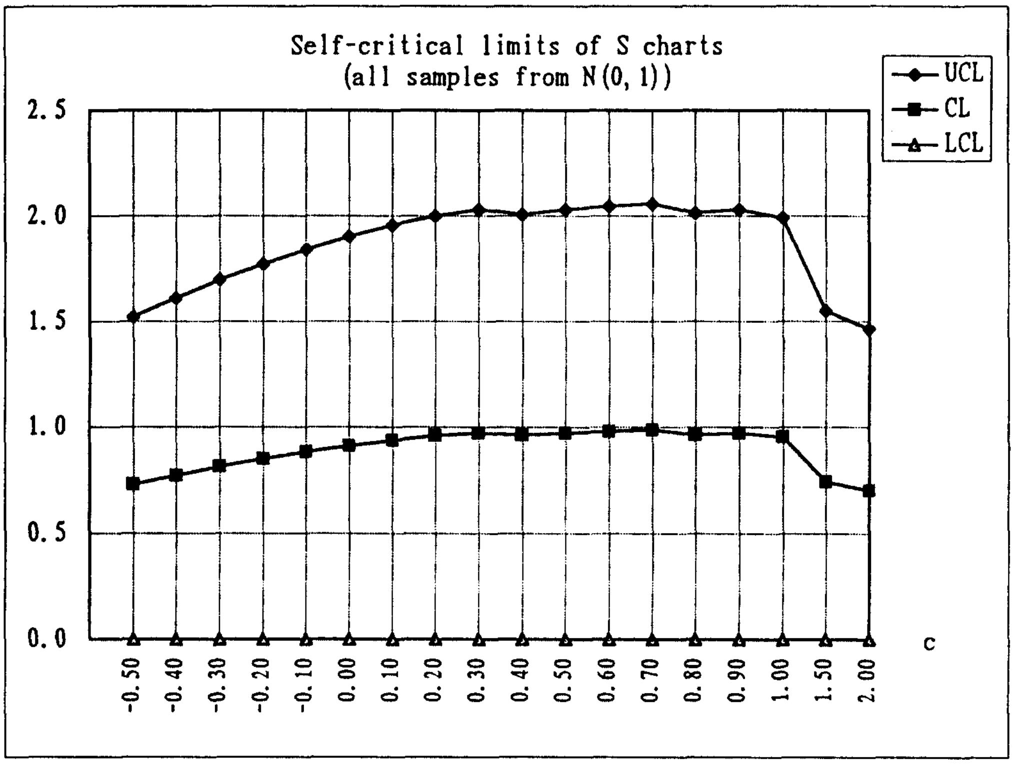
We generate standard normal random deviates of 25 subgroups, where each subgroup possesses 5 random observations. <Figure 1 (a), (b)> show control limits of $\bar{x}$(c)-S(c) charts as a function of c in the range of c=-0.5 to c=2.0. The control limits show similar shape for 10 observations in each subgroup. Also in the same figures, we see the number of subgroups whose summary statistics have been plotted beyond the control limits. We observe 1 subgroup for the case of $\bar{x}$ (c = -0.5) and some subgroups for the cases of S(c) over c $\leq$ -0.3 and c $\geq$ 1.5. This implies the fact that if c increases either negatively or positively from 0, the control limits will have larger type I errors. Note that the control limits of c=0 are equal to the limits of ordinary $\bar{x}$-S control charts.

c	UCL	CL	LCL	freq
-0.5	1.015	-0.025	-1.065	1
-0.4	1.084	-0.016	-1.117	0
-0.3	1.148	-0.011	-1.171	0
-0.2	1.206	-0.006	-1.219	0
-0.1	1.257	-0.002	-1.260	0
0.0	1.303	0.003	-1.297	0
0.1	1.345	0.008	-1.328	0
0.2	1.381	0.014	-1.353	0
0.3	1.410	0.023	-1.364	0
0.4	1.410	0.038	-1.333	0
0.5	1.430	0.043	-1.344	0
0.6	1.443	0.047	-1.350	0
0.7	1.454	0.047	-1.359	0
0.8	1.429	0.053	-1.323	0
0.9	1.442	0.055	-1.332	0
1.0	1.421	0.059	-1.302	0
1.5	1.188	0.127	-0.933	0
2.0	1.129	0.129	-0.871	0



< Figure 1(a) > Control limits of x-bar (c) with n=5

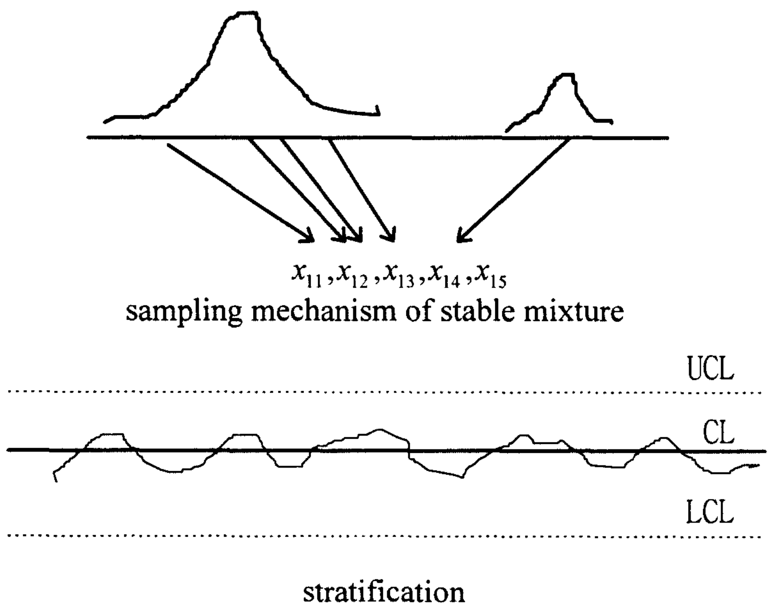
c	UCL	CL	LCL	freq
-0.5	1.523	0.729	0	2
-0.4	1.611	0.771	0	1
-0.3	1.698	0.813	0	1
-0.2	1.774	0.849	0	0
-0.1	1.843	0.882	0	0
0.0	1.903	0.911	0	0
0.1	1.956	0.937	0	0
0.2	2.001	0.958	0	0
0.3	2.030	0.972	0	0
0.4	2.008	0.961	0	0
0.5	2.030	0.972	0	0
0.6	2.045	0.979	0	0
0.7	2.059	0.986	0	0
0.8	2.015	0.964	0	0
0.9	2.031	0.972	0	0
1.0	1.993	0.954	0	0
1.5	1.552	0.743	0	2
2.0	1.464	0.701	0	4



< Figure 1(b) > Control limits of S(c) for n=5, m=25

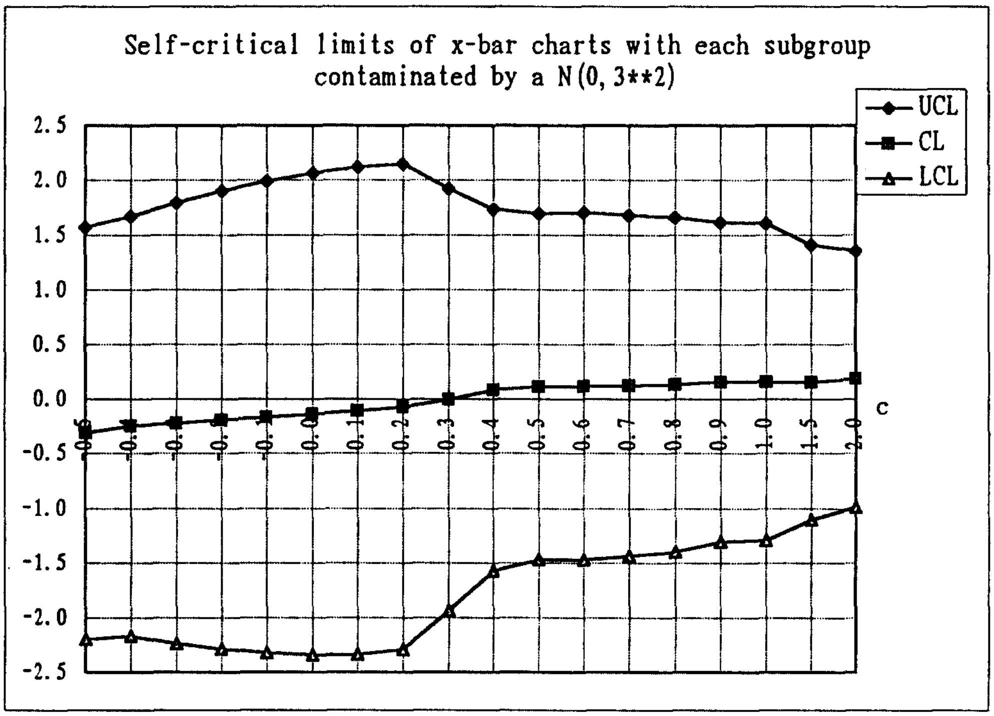
Example 2. case of stable mixture

5th observation in each subgroup is generated by non-standard normal distribution, which is known as a case of stable mixture. This case is not the same as mixture distributions, where contaminated distribution can place itself in such a random way that any specific subgroup may or may not have one or more observations. <Figure 2> shows the systematic sampling and its resulting unnatural pattern in $\bar{x}$ chart. <Figure 3 (a), (b)> shows the control limits and the number of subgroups out of control as a function of c, when the fifth observation is contaminated by a nonstandard normal distribution with mean 0 and standard deviation 3. We observe tighter control limits and some number of out-of-control subgroups for c $\geq$ 0.4 approximately, compared with the limits of c=0. <Figure 4> shows two control charts on the same graph for c=0 and c=0.4. In $\bar{x}$-S charts, it is hard to see any unnatural pattern quickly, but in $\bar{x}$(0.4)-S(0.4) charts, the unnaturalness is magnificent owing to the 6th and 23rd points, which fall beyond the control limits in S(0.4) chart. It is also notewothy that $\bar{x}$-R charts in <Figure 5> fail to show any stratification phenomina without difficulty of analysis in detail.



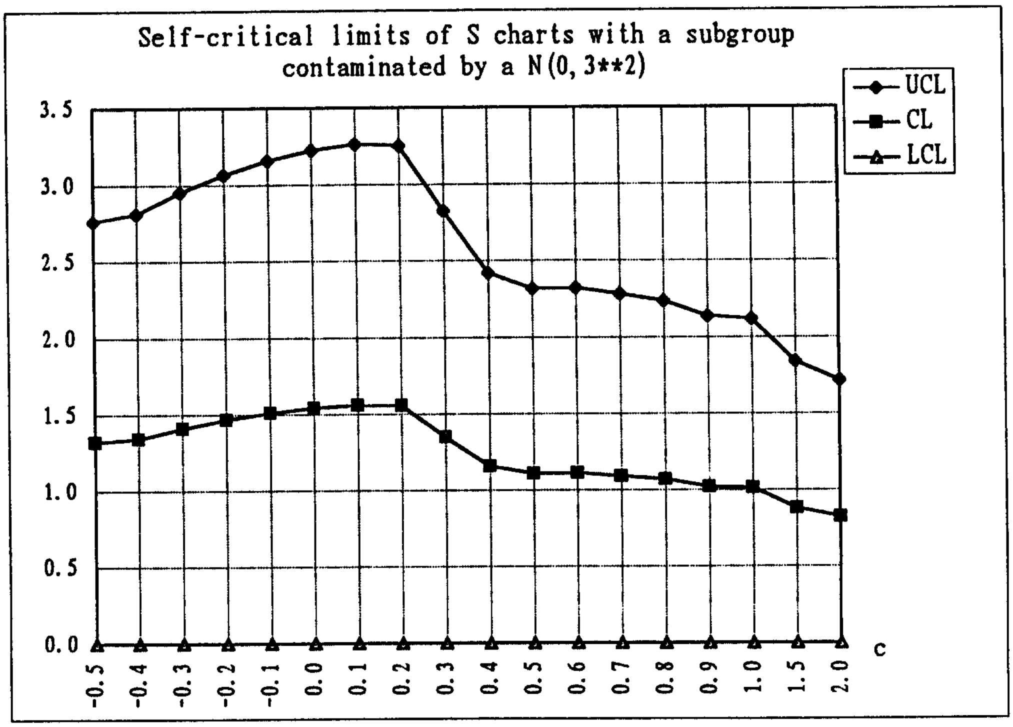
< Figure 2 > Sampling mechanism of stable mixture and pattern of stratification

c	UCL	CL	LCL	freq
-0.5	1.574	-0.313	-2.200	0
-0.4	1.672	-0.249	-2.170	0
-0.3	1.795	-0.221	-2.238	0
-0.2	1.901	-0.194	-2.289	0
-0.1	1.991	-0.166	-2.324	0
0.0	2.066	-0.138	-2.341	0
0.1	2.122	-0.107	-2.336	0
0.2	2.151	-0.072	-2.295	0
0.3	1.925	-0.006	-1.937	0
0.4	1.734	0.082	-1.571	0
0.5	1.697	0.114	-1.470	1
0.6	1.705	0.118	-1.468	1
0.7	1.682	0.123	-1.436	1
0.8	1.661	0.134	-1.394	1
0.9	1.615	0.157	-1.301	1
1.0	1.606	0.159	-1.288	1
1.5	1.412	0.155	-1.101	4
2.0	1.358	0.186	-0.986	5

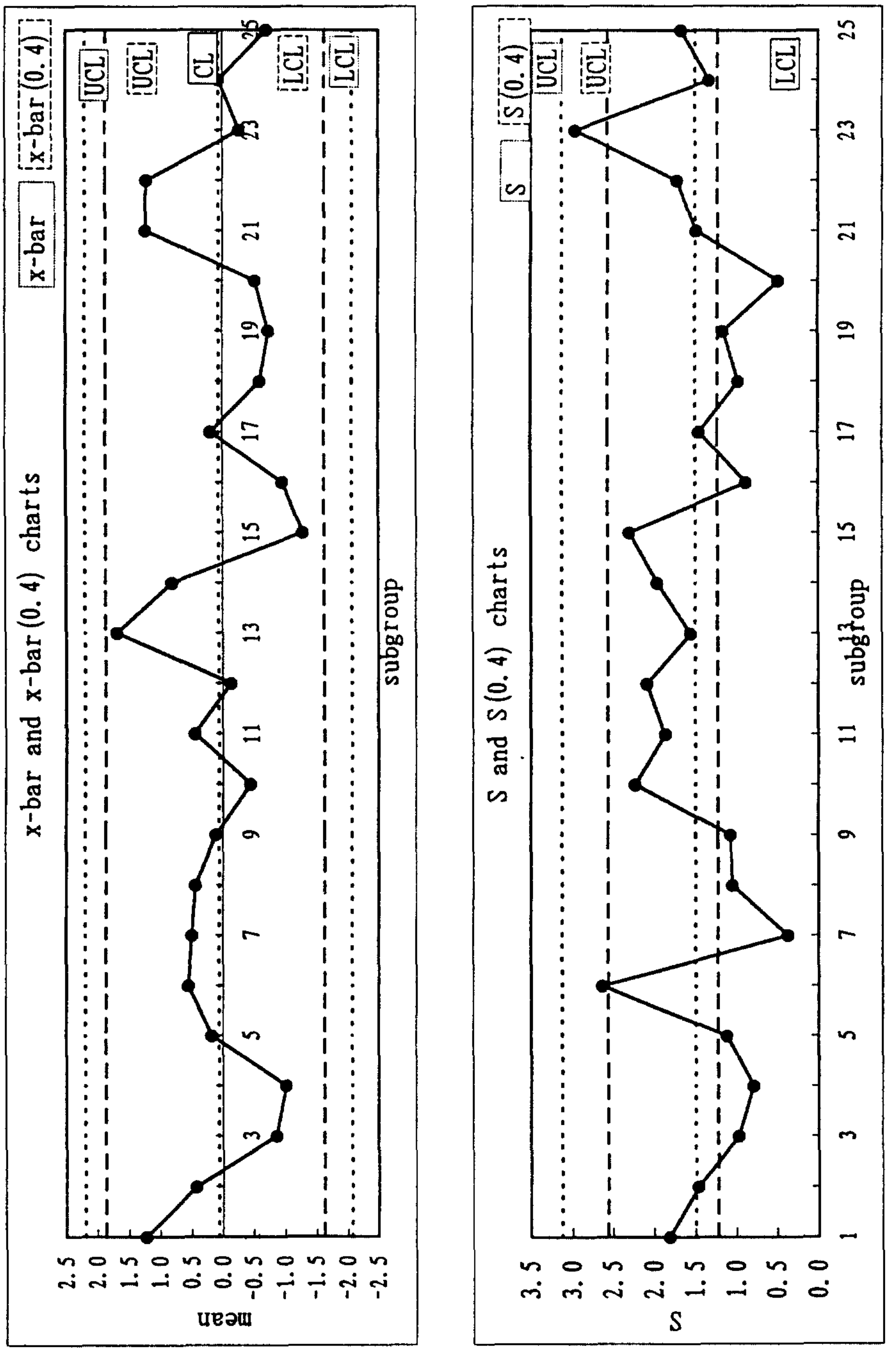


< Figure 3(a) > Control limits of x-bar (0.4) with a contaminant by N(0, 3*3)

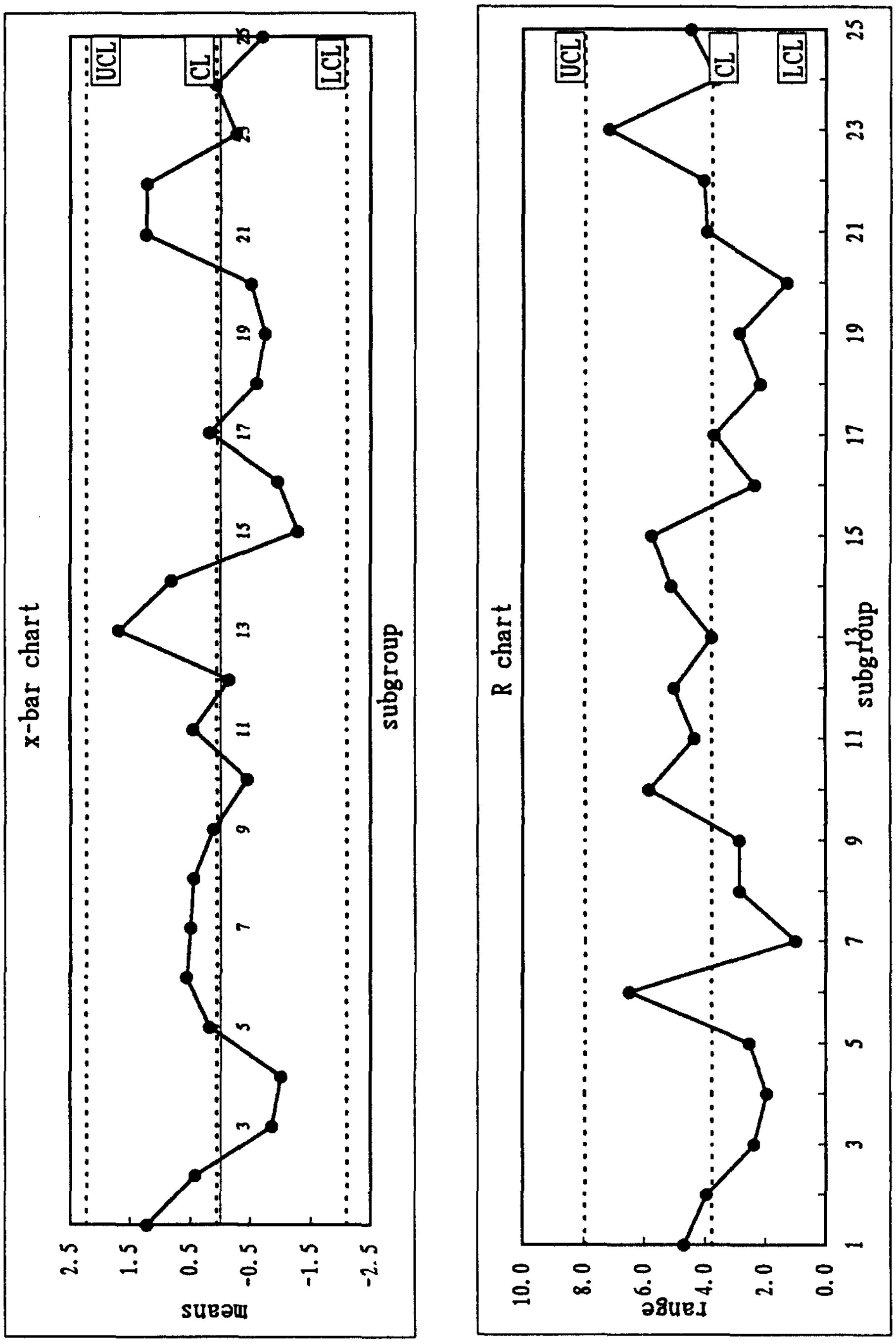
c	UCL	CL	LCL	freq
-0.5	2.763	1.323	0.0	1
-0.4	2.812	1.346	0.0	1
-0.3	2.952	1.413	0.0	1
-0.2	3.067	1.468	0.0	1
-0.1	3.159	1.512	0.0	1
0.0	3.226	1.544	0.0	1
0.1	3.263	1.562	0.0	1
0.2	3.254	1.558	0.0	1
0.3	2.827	1.353	0.0	1
0.4	2.419	1.158	0.0	4
0.5	2.318	1.110	0.0	4
0.6	2.322	1.112	0.0	4
0.7	2.282	1.092	0.0	4
0.8	2.236	1.070	0.0	5
0.9	2.135	1.022	0.0	5
1.0	2.119	1.014	0.0	5
1.5	1.839	0.880	0.0	6
2.0	1.716	0.822	0.0	8



< Figure 3(b) > Control limits of S (0.4) charts with a contaminant by N(0, 3*3)



< Figure 4 > Control charts of x-bar(c) and S(c) for c=0, 0.4 for the simulated example



< Figure 5 > Control charts of x-bar and R for the simulated example

We extended the Monte Carlo simulation by considering lager ranges of parameters of the non-standard normal distributions. <Table 1> and <Table 2> show the number of subgroups whose summary statistics are positioned out of the control limits for the subgroup size of n=5 and n=10. The ranges we considered in this table as a contaminating distribution are μ=0, 3, 5 and σ=1, 3, 5. This table shows that for the case of stable mixture the value of critical coefficient from 0.3 to 0.5 are good enough to detect unnaturalness for the subgroup sizes of 5 and 10. For smaller subgroup sizes, for example, such as 3 or 4, we need to consider the coefficient larger than 0.6 and thereby increase the robustness of the self-critical estimators.

< Table 1 > Number of subgroups out of the control limits for the case of stratification (n = 5, m = 25)

		N(0, 1)	N(0, 3)	N(0, 5)	N(3, 1)	N(3, 3)	N(3, 5)	N(5, 1)	N(5, 3)	N(5, 5)
$\bar{x}$-R charts	$\bar{x}$	0	0	0	0	0	0	0	0	0
	R	0	1	1	0	0	0	0	0	1
$\bar{x}$-S charts	$\bar{x}$	0	0	0	0	0	0	0	0	0
	S	0	1	1	0	1	0	0	0	1
$\bar{x}$(c)-S(c) charts	$\bar{x}$(0.3)	0	0	1	0	0	1	0	1	6
	S(0.3)	0	1	6	0	1	9	1	8	11
	$\bar{x}$(0.4)	0	0	5	0	4	6	1	4	7
	S(0.4)	0	4	10	0	7	11	7	10	12
	$\bar{x}$(0.5)	0	1	8	0	4	7	2	5	8
	S(0.5)	0	4	15	1	8	13	15	12	14

< Table 2 > Number of subgroups out of the control limits for the case of stratification (n = 10, m = 25)

		N(0, 1)	N(0, 3)	N(0, 5)	N(3, 1)	N(3, 3)	N(3, 5)	N(5, 1)	N(5, 3)	N(5, 5)
$\bar{x}$-R charts	$\bar{x}$	0	0	0	0	0	0	0	0	0
	R	0	2	2	0	1	1	0	1	2
$\bar{x}$-S charts	$\bar{x}$	0	0	0	0	0	0	0	0	0
	S	0	1	2	0	1	1	0	1	2
$\bar{x}$(c)-S(c) charts	$\bar{x}$(0.3)	0	0	3	0	2	4	0	3	4
	S(0.3)	0	3	8	0	8	11	9	13	14
	$\bar{x}$(0.4)	0	1	3	0	3	4	0	4	4
	S(0.4)	0	3	11	0	8	11	12	13	14
	$\bar{x}$(0.5)	0	2	5	0	3	5	2	4	5
	S(0.5)	0	4	11	0	9	11	15	13	14

6. Conclusions

When we set up control charts in order to monitor the state of any process, it is quite often that preliminary set of data may posses errors of observations. These observational errors will make the control limits inflated and hence decrease the sensitivity of the charts. To keep the limits tight, we suggested use of a robust estimators, called self-critical estimators. We can control the degree of influence by the observational errors to the estimators by changing the user-given self-critical coefficient. For the case of stable mixture the values of c ranged from 0.3 to 0.5 are recommended in most of the practical situations of subgroup sizes more than 5.

References

[1] Alloway, J.A. JR, and Raghavachari, M.(1991), "Control Charts Based on the Hodges-Lehmann Estimator," *Journal of Quality Technology*, vol. 23, no. 4, pp. 336-347.

[2] Amin, R.W., and Miller, R.W.(1993), "A Robustness Study of bar X Charts with Variable Sampling Intervals," *Journal of Quality Technology*, vol. 25, no. 1, pp. 36-44.

[3] Ferrel, E.B.(1953), "Control Charts Using Midrange and Medians," *Industrial Quality Control*, vol. 9, pp. 30-34.

[4] Hawkins, D.M.(1993), "Robustification of Cumulative Sum Charts by Winsorization," *Journal of Quality Technology*, vol. 25, no. 4, pp. 248-261.

[5] Huber, P.J.(1981), *Robust Statistics*, John Wiley & Sons Inc., New York.

[6] IMSL(1994), *User's manual*(C functions for scientific programming), Visual Numerics, Inc.

[7] Langenberg P. and Iglewicz, B(1986), "Trimmed Mean bar X and R Charts," *Journal of Quality Technology*, vol. 18, no. 3, pp. 152-161.

[8] Paulson, A.S., Presser, M.A., and Nicklin, E.H., "Self-Critical And Robust Procedures For the Analysis of Univariate Complete Data," unpublished research paper, Rensselaer Polytechnic Institute.

[9] Quesenberry C.P.(1986), "Screening Outliers in Normal Process Control Data with Uniform Residuals," *Journal of Quality Technology*, vol. 18, no. 4, pp. 226-233.

[10] Rocke, D.M.(1989), "Robust Control Charts," *Technometrics*, vol. 31, no. 2, pp. 173-184.

[11] *Statistical Quality Control Handbook*, Western Electric Co., Inc., 1984.