〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 14, No. 2, 1985

Some Control Procedures Useful for One-sided Asymmetrical Distributions+

Changsoon Park*

ABSTRACT

Shewhart $\bar{X}$-chart, which is most widely used in practice, is shown to be inappropriate for the cases where the process distribution is one-sided asymmetrical, and thus some nonparametric Shewhart type charts are developed instead. These schemes may be applied usefully when there is not enough information in determining the process distribution. The average run lengths are obtained to compare the efficiency of control charts for various shifts of the location parameter and for some typical one-sided asymmetrical distributions.

1. Introduction

In a continuous production process, one of the main purposes of using statistical control charts is to minimize the loss by taking some rectifying action as soon as the process is out of control. Examples of such control charts are Shewhart chart originated by Shewhart (1931), modified Shewhart chart with warning lines by Page (1955), and cumulative sum control chart by Page (1954).

Suppose that independently identically distributed (i.i.d.) random samples of fixed size are observed at regular time intervals during the process. Let $\underline{X}_i=(X_{i1}, ..., X_{in})$, $i=1, 2...$, be the i-th observed sample and let θ be the parameter which specifies the quality of the product. Also suppose that a statistic $S_i=S(X_i, \theta_0)$, where S is determined only by $\underline{X}_i$ and the given control value θ_0 of θ, is obtained each time a sample is observed. We assume that large values of S tend to indicate positive shifts of θ and small values

* Department of Applied Statistics, Chung-ang University, Seoul 151, Korea

+ This research was supported in part by the Research Fund of the Ministry of Education, Korea, 1985.

negative shifts. Then the stopping rule of the Shewhart chart for detecting positive shifts is to;

$$\text{Stop at the first } i \text{ for which } S_i \geq c \tag{1.1}$$

for some constant c which is called the control limit.

The run length of a control chart is the number of samples required to stop the control chart. The run length N of the Shewhart chart follows a geometric distribution with a parameter $P(S \geq c)$, and thus the average run length (ARL) is

$$E_\theta N = 1/P(S \geq c). \tag{1.2}$$

Let Θ be the parameter space of θ, and let Θ_0 and Θ_1 be two mutually exclusive subsets of Θ. If $\theta \in \Theta_0$, we say the process is in control, and if $\theta \in \Theta_1$, out of control. Also we define ARL in control (ARL_0) as $\min_\theta E_\theta N$ for $\theta \in \Theta_0$ and the control value as the parameter value $\theta \in \Theta_0$ which minimizes $E_\theta N$.

Most standard control procedures are designed on the assumption that the distribution of the observations is of a specified form, usually the normal distribution. But for the cases where there is insufficient information to completely determine the distribution, it is appropriate to use nonparametric procedures which require fewer assumptions than parametrice ones. One example of such nonparametric procedures is the procedure suggested by Bakir and Reynolds (1979) which uses nonparametric statistics in cumulative sum control charts. One advantage of nonparametric procedures is that the variance of the process does not need to be known or estimated, moreover the variance need not to be stationary. Another advantage is that we may obtain an exact control limit for the desired ARL_0.

In this paper, some nonparametric Shewhart type charts for controlling location shifts are studied for some one-sided asymmetrical distributions as parent distributions. For convenience, only positive shifts are to be considered throughout this paper.

2. Comparison of control charts

When the process is out of control the ARL gives the expected time elapsed until a signal, and thus measures the amount of scrap produced before a rectifying action is taken. On the other hand, when the process is in control any signal by the control chart is a false alarm, and thus the ARL is a measure of the frequency of false alarms. As long as the process is in control, the ARL should be large so that production may continue

uninteruppeted as long as possible, but if the process is out of control the ARL should be small so that the change is detected quickly.

When control charts are to be compared, the usual procedure is to compare their ARL's as in the following definition.

Definition: For two control charts T_1 and T_2 whose ARL's are denoted by $E_\theta N_1$ and $E_\theta N_2$, respectively, T_1 is defined to be more efficient than T_2 iff

$$E_\theta N_1 \leq E_\theta N_2 \text{ for all } \theta \in \Theta_1$$

for $\min_\theta E_\theta N_1 \leq \min_\theta E_\theta N_2$ for $\theta \in \Theta_0$. ■

This method of comparison is similar to that of comparing the powers of two tests having the same type I error in a hypothesis testing problem. The following theorem gives a method of obtaining the most efficient Shewhart chart.

Theorem 1 : Let T be a Shewhart chart which stops at the first i for which $S_i \geq c$. If there exists the uniformly most powerful size α test for $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ which rejects H_0 iff $S \geq c$, then T is more efficient than any other Shewhart chart whose $\text{ARL}_0 \geq 1/\alpha$

Proof: Let T^* be any other Shewhart chart which stops at the first i for which $S^* \geq c^*$ where c^* is determined so that $\text{ARL}_0 \geq 1/\alpha$ Also let N and N^* be the run lengths of T and T^*, respectively. Then by the uniformly most powerfulness,

$$P_\theta(S \geq c) \geq P_\theta(S^* \geq c^*) \text{ for all } \theta \in \Theta_1$$

Thus by (1.2),

$$E_\theta N \leq E_\theta N^* \text{ for all } \theta \in \Theta_1. \blacksquare$$

3. Inappropriateness of $\overline{X}$-chart for asymmetrical distributions

When the distribution of the process is normal with known variance σ^2, we see that the Shewhart $\overline{X}$-chart which stops at the first i for which $(\overline{X}_i - \theta_0)/(\sigma/\sqrt{n}) \geq c$ is most efficient by Theorem 1. However for the cases where the distribution is quite different from normal such as one-sided asymmetrical distributions and the sample size is relatively small, we have to be cautious about applying $\overline{X}$-chart because the central limit theorem does not give a good approximation for such cases.

Three well-known one-sided asymmetrical distributions, Weibull, Gamma and Lognormal, are considered as the parent distributions of the process. The probability density functions (p.d.f.'s) of the three distributions are

$$f(x)=(2x/\lambda^2)\exp(-(x/\lambda)^2)I_{(0,\infty)}(x), \tag{3.1}$$

$$f(x)=(x/\lambda^2)\exp(-x/\lambda)I_{(0,\infty)}(x), \tag{3.2}$$

and

$$f(x)=\{1/(x\sqrt{2\pi})\}\exp\{-(\log x-\lambda)^2/2\}I_{(0,\infty)}(x), \tag{3.3}$$

respectively. The p.d.f.'s of the three distributions when the process is in control are the corresponding expressions (3.1), (3.2), and (3.3) where λ is replaced by 1 for Weibull and Gamma, and 0 for Lognormal. The mean, median, and variance of the three distributions when the process is in control are $(\sqrt{\pi}/2,\ \sqrt{\log 2},\ 1-\pi/4)$, $(2, 1.67853, 2)$, and $(\sqrt{e}, 1, e^2-e)$, respectively. The graphs of the three p.d.f.'s are shown in Figure 1.

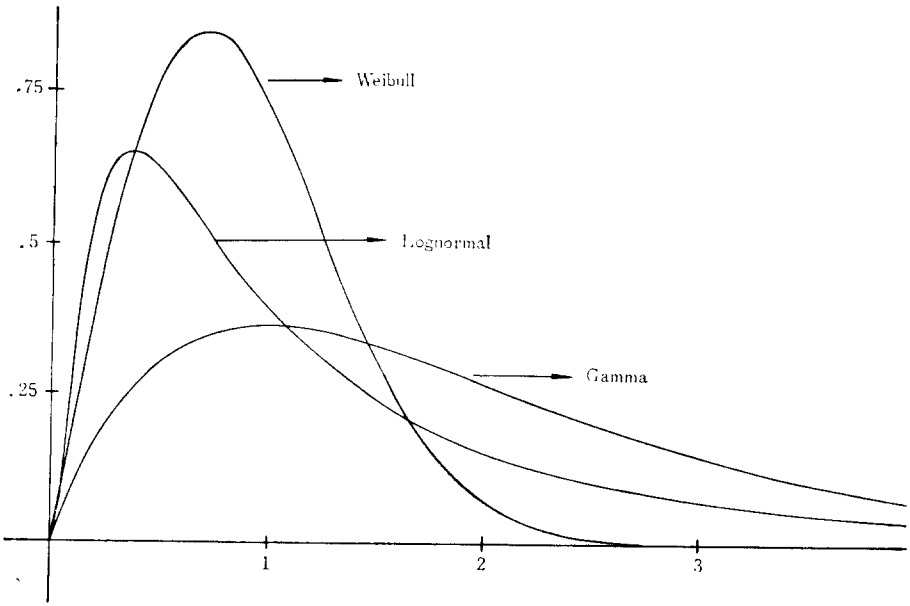

Figure 1. The probability density functions of Weibull, Lognormal, and Gamma distributions

In Table 1, the ARL's of $\overline{X}$-chart are obtained for various shifts of the mean assuming the process variance is known. The parameter λ is changed appropriately for each amount of the mean shift. The ARL's for the three distributions, Weibull, Gamma, and Lognormal, are obtained by simulation. The 500 simulated run lengths are averaged to produce the ARL. The values of shift can be thought of as being expressed in units of standard deviation when the process is in control and the number in ( ) denotes the standard deviation of estimator of the ARL. From the table we see that ARL's for the three

distributions are quite different from that for normal. Especially when the process is in control, ARL's for all the three distributions are quite less than expected (i.e. 500) and this discrepancy seems to be more serious as the distribution becomes heavier-tailed. Another difficulty is the problem of deciding the control limit for a desired ARL_0 for such distributions. Therefore we may conclude that the result indicates that $\overline{X}$-chart is not appropriate for one-sided asymmetrical distributions when the size of the observed sample is relativelly small.

Table 1. The ARL of the Shewhart $\overline{X}$-chart for $n=10$ and $c=2.88$

Shift	Normal	Weibull	Lognormal	Gamma
0.0	500.00	282.33(11.66)	62.70(2.93)	171.73(7.74)
0.25	98.04	25.06(1.09)	14.71(0.65)	19.66(0.86)
0.5	25.58	6.46(0.27)	5.88(0.24)	5.74(0.22)
1.0	3.85	1.76(0.06)	2.24(0.08)	1.87(0.06)
2.0	1.06	1.05(0.01)	1.13(0.02)	1.08(0.01)

4. Nonparametric Shewhart charts

Since a Shewhart chart is completely determined by the statistic to use, a nonparametric Shewhart chart is developed by simply using a nonparametric statistic in the chart. For the distributions which are not symmetric and the distributions whose means do not exist, the median of the distribution may be used as the parameter to be controlled in place of the mean to achieve nonparametric properties of the control chart. Thus we now reformulate the problem of controlling the mean into the median.

Two nonparametric statistics, sign and signed rank statistics, are considered for the statistic used in the Shewhart chart for controlling the median of the process.

Let the i-th observed sign statistic be

$$S_i=\sum_{j=0}^{n} \Psi(X_{ij}-\theta_0) \tag{4.1}$$

where $\Psi(t)=1, 0$ as $t>, \leq 0$ and θ_0 is the median of the process in control. The ARL of a Shewhart chart using the statistic (4.1) is

$$E_\theta N=1/P_\theta(S\geq c)$$
$$=1/\left[\sum_{k=c}^{n}\binom{n}{c}\{G(\theta_0)\}^{n-k}\{1-G(\theta_0)\}^{k}\right] \tag{4.2}$$

where G denotes the distribution function of the observations. The ARL's for various

shifts of the median are obtained analytically in Table 2 by use of the expression (4.2). The parameter λ in (3.1), (3.2), and (3.3) is changed appropriately to produce each amount of the median shift. From the table, it seems that Shewhart chart using sign statistic becomes more efficient as the distribution becomes heavier-tailed in the order of Weibull, Gamma, and Lognormal.

Table 2. The ARL of the Shewhart chart using sign statistic for $n=8$ and $c=8$

Shift	Weibull	Lognormal	Gamma
0.0	256.00	256.00	256.00
0.25	71.78	26.24	62.39
0.5	29.79	8.24	24.77
1.0	9.87	2.91	8.22
2.0	3.46	1.47	3.05

Usually in hypothesis testing problems for location parameter, signed rank statistic is more powerful than sign statistic except for very heavy-tailed distributions [Randles and Wolfe (1979)]. Thus signed rank statistic is also considered as the statistic to use to improve the sensitivity of the chart. Log transform is applied to the distributions to make the distributions more symmetric because application of signed rank statistic assumes symmetry of the distribution. Log transform of Lognormal becomes normal which is symmetric. For Gamma distribution, log transform is used often to make the distribution more nearly normal [Johnson and Kotz (1970)]. Log transform of Weibull becomes extreme value distribution and it does not seem to make the distribution more symmetric, but log transform is also applied to Weibull for consistency, Notice that median is invariant to transform, i.e. the median of the transformed distribution is the same transform of the median of the original distribution.

Let $Z_{ij}=\log X_{ij}$, for $i=1,2,\ldots$ and $j=1,2,\ldots,n$, then the i-th observed signed rank statistic is

$$SR_i=\sum_{j=1}^{n} \Psi(Z_{ij}-\log\theta_0)R_{ij} \tag{4.3}$$

where R_{ij} is the rank of $|Z_{ij}-\log\theta_0|$ among $|Z_{i1}-\log\theta_0|,\ldots,|Z_{in}-\log\theta_0|$. The ARL's of a Shewhart chart using the statistic (4.3) are obtained for various shifts in Table 3 by averaging 500 simulated run lengths. It seems that the use of signed rank statistic improves sensitivity of the Shewhart chart considerably when compared to the use of sign statistic. Thus it is recommended to use signed rank statistic in the Shewhart chart

unless the distribution is known to be extremely heavy-tailed. The ARL_0's of Weibull and Gamma are almost the same as that of Lognormal and this result shows that log transform of Weibull and Gamma makes the distributions nearly symmetric as long as signed rank statistic is concerned.

Table 3. The ARL of the Shewhart chart using signed rank statistic for $n=9$ and $c=44$

Shift	Weibull	Lognormal	Gamma
0.0	270.39(11.56)	256.00	242.14(11.32)
0.25	62.41(2.73)	21.45(0.87)	60.96(2.69)
0.5	23.95(1.15)	6.32(0.27)	20.20(0.87)
1.0	7.51(0.31)	2.14(0.07)	6.30(0.26)
2.0	2.73(0.10)	1.16(0.02)	2.41(0.08)

5. Nonparametric Shewhart charts based on a standard sample

Before there is enough information about the distribution of the process, the procedures described in the previous section may be used usefully. Although the process is considered to be in control the control value of the median may not be determined easily if there are not sufficient observations available for the population value. In such a case we obtain a standard sample $\underline{X}_0=(X_{01}, \ldots, X_{0m})$, and use the sample median $\tilde{X}_0$ as a control value of the true median. Control values are often estimated from a standard sample in practice for the cases where the control value is not known. A standard sample is assumed to be obtained when the process is in control.

A modified sign statistic is defined as

$$S_i^*=\sum_{j=1}^{n}\Psi(X_{ij}-\tilde{X}_0) \tag{5.1}$$

for $i=1, 2, \ldots$ The statistic (5.1) is equivalent to the percentile placement statistic using the Bernoulli scoring function which is defined by Orban and Wolfe (1982). A modified signed rank statistic is defined as

$$SR_i^*=\sum_{j=1}^{n}\Psi(Z_{ij}-\log\tilde{X}_0)R_{ij}^* \tag{5.2}$$

for $i=1, 2, \ldots$ where R_{ij}^* is the rank of $|Z_{ij}-\log\tilde{X}_0|$ among $|Z_{i1}-\log\tilde{X}_0|, \ldots, |Z_{in}-\log\tilde{X}_0|$.

Either the sequence of statistics $\{S_i^*;\ i=1, 2, \ldots\}$ or $\{SR_i^*;\ i=1, 2, \ldots\}$ is not independent because they depend on the same sample median $\tilde{X}_0$. But for given $\tilde{X}_0$, both the sequence

of statistics $\{S_i^*;\ i=1,2,\ldots\}$ and $\{SR_i^*;\ i=1,2,\ldots\}$ are independent. Also if the standard sample size m goes to infinity, the sample median converges to a constant, i.e. the population median, and thus both the sequence of statistics $\{S_i^*;\ i=1,2,\ldots\}$ and $\{SR_i^*;\ i=1,2,\ldots\}$ are asymptotically $(m\to\infty)$ independent. That is, for any positive integer k and any real numbers $s_1,\ldots,s_k$,

$$\lim_{m\to\infty} P(S_1\leq s_1,\ldots,S_k\leq s_k)=\lim_{m\to\infty} P(S_1\leq s_1)\ldots P(S_k\leq s_k) \tag{5.3}$$

where S_i denotes either S_i^* or SR_i^*.

Although a nonparametric statistic is used in a Shewhart chart, it is not obvious whether the properties of the chart are distribution-free or not because of dependency among the sequence of statistics. The following theorem shows that the properties are still distribution-free when the process is in control even though the statistics are not independent.

Theorem 2: Let $\underline{X}_0=(X_{01},\ldots,X_{0m})$ and $\underline{X}_i=(X_{i1},\ldots,X_{in})$, $i=1,2,\ldots$ be i.i.d. random samples with a continuous distribution function F. Then for any positive integer k, the joint distributin function of $S_1,\ldots,S_k$ is the same for all F where S denotes the statistic (5.1) or (5.2). ■

Proof: Let $V_{0j}=F(X_{0j})$ for $j=1,\ldots,m$, $V_{ij}=F(X_{ij})$ for $i=1,\ldots,k$, $j=1,\ldots,n$, and $\tilde{V}_0$ be the median of $V_{01},\ldots,V_{0m}$. Since the distribution of the statistics (5.1) and (5.2) are invariant to the transform F of $\underline{X}_i$, for $i=0,1,\ldots,k$,

$$P(S_1\leq s_1,\ldots,S_k\leq s_k)=P(Q_1\leq s_1,\ldots,Q_k\leq s_k)$$

where Q_i's are the corresponding statistics S_i's for which X_{ij}'s and $\tilde{X}_0$ are replaced by V_{ij}'s and $\tilde{V}_0$. Because each of the random variables V_{ij}'s follows a uniform (0, 1) distribution for every continuous distribution F, the Q_i's are always functions only of uniform (0, 1) random variables and the proof is done. ■

This theorem ensures that the run length distribution of any control chart which uses only the statistics (5.1) and (5.2) for stopping rule is always the same regardless of the underlying distribution when the process is in control.

In the case of the Shewhart chart with a standard sample there may not be an upper bound for ARL which was shown by Park (1984), thus $T=1000$ is used as a truncation point which specifies the upper bound for the run length of the Shewhart chart. To lessen the difficulties of the dependent structure of the sequence of the statistics, we use the fact that the statistics are i.i.d. random variables conditioned on the standard sample. For the Shewhart chart using the modified sign statistic, the ARL can be expressed as

$$E_0N=\sum_{t=0}^{T-1}P(N>t)$$

$$=1+\sum_{t=1}^{T-1}P(S_1<c,\ldots,S_t<c)$$

$$=1+\sum_{t=1}^{T-1}\int_0^{\infty}[P(S<c\,|\,\tilde{X}_0=x)]^t h(x)\,dx \tag{5.4}$$

$$=1+\int_0^{\infty}\sum_{t=1}^{T-1}[1-\sum_{k=c}^{n}\binom{n}{k}\{1-G(x)\}^k\{G(x)\}^{n-k}]^t h(x)\,dx \tag{5.5}$$

where h is the density function of $\tilde{X}_0$ and G is the distribution function of X_{ij}'s for $i=1,2,\ldots,\ j=1,\ldots,n$.

Applying simple trapezoid rule to the expression (5.5), the ARL's of the Shewhart chart using the modified sign statistic are calculated in Table 4. For the Shewhart chart using the modified signed rank statistic, the ARL's are obtained in Table 5 by averaging 500 simulated run lengths. In both Table 4 and 5, the ARL is greater than the corresponding one in Table 2 and 3 for each distribution and shift. This can be verified as follows.

Table 4. The ARL of the Shewhart chart using modified sign statistic for $m=49$, $n=8$, $c=8$, and $T=1000$

Shift	Weibull	Lognormal	Gamma
0.0	318.68	318.68	318.68
0.25	109.55	38.67	96.48
0.5	40.99	10.56	34.25
1.0	11.51	3.24	9.66
2.0	3.65	1.53	3.24

Table 5. The ARL of the Shewhart chart using modified signed rank statistic for $m=49$, $n=9$, $c=44$, and $T=1000$

Shift	Weibull	Lognormal	Gamma
0.0	662.15(70.31)	331.54(16.05)	323.85(15.95)
0.25	116.16(11.21)	33.81(2.74)	91.70(6.93)
0.5	37.07(2.69)	8.45(0.48)	28.67(2.57)
1.0	9.07(0.54)	2.55(0.11)	7.46(0.38)
2.0	2.97(0.13)	1.28(0.03)	2.42(0.09)

By Jensen's inequality,

$$\int_0^{\infty}[P(S<c\,|\,\tilde{X}_0=x)]^t h(x)\,dx\geq[P(S<c)]^t \tag{5.6}$$

Applying (5.6) to (5.4),

$$E_\theta N \geq \sum_{t=0}^{T-1} [P(S<c)]^t \tag{5.7}$$

where the right hand side (r.h.s.) of (5.7) is the expression of the ARL with the truncation point T when the control value of the median is given. It can be easily seen from r.h.s. of (5.7) that, when the control value is given, the ARL is almost the same whether there is a truncation point or not if it is a large number such as $T=1000$. The discrepancy of the ARL's between the two cases where the control value is given and not given is getting smaller as the size m of the standard sample increases so that the ARL's are equal when m goes to infinity by the asymptotic independency (5.3).

It is a rather surprising result that the ARL_0 of Weibull in Table 5 is much greater than that of Lognormal while the values are almost the same in Table 3. This may be because log transform of Weibull does not make the distribution symmetric enough to be applied to a dependent sequence of modified signed rank statistics. It also seems as the results in Section 4 that modified signed rank statistic is more sensitive than modified sign statistic in the Shewhart chart.

6. Conclusions

It was shown that the Shewhart $\overline{X}$-chart is not an efficient scheme for controlling the process mean of asymmetrical distributions. Instead of the sample mean it is useful to use the sign and signed rank statistic in the Shewhart chart for controlling the process median in case of insufficient information. If the control value of the parameter is estimated from a standard sample, we have to be careful in estimating the ARL of the control chart because the ARL will be greater than that for the case where the control value is given. The basic idea of the proposed procedure is to replace the parametric statistic by a nonparametric one in the existing control scheme to achieve nonparametric properties.

References

(1) Bakir, S.T. and Reynolds, M.R., Jr. (1979). A Nonparametric Procedure for Process Control Based on Within-Group Ranking, *Technometrics*, Vol. 21, No. 2, 175~183.

(2) Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate*

Distributions-1, 181, John Wiley and Sons, New York.

(3) Orban, J. and Wolfe, D.A. (1982). A Class of Distribution-free Two-Sample Tests based on Placements, *Journal of American Statistical Association*, Vol. 77, 666~670.

(4) Page, E.S. (1954). Continuous Inspection Schemes, *Biometrika*, Vol. 41, 100~114.

(5) Page, E.S. (1955). Control Charts with Warning Lines, *Biometrika*, Vol. 42, 243~257.

(6) Park, C. (1984). Nonparametric Procedures for Process Control When the Control Value is not Specified, 49, Ph. D. Dissertation, Department of Statistics, Virginia Polytechnic Institute and State University.

(7) Randles, R.H. and Wolfe, D.A. (1979). *Introduction to the Theory of Nonparametric Statistics*, 116, John Wiley and Sons, New York.

(8) Shewhart, W.A. (1931). *The Economic Control of Quality of Manufactured Product*, Van Nostrand, New York.