Journal of the Korean Data & Information Science Society
2012, **23**(5), 1037–1044

http://dx.doi.org/10.7465/jkdi.2012.23.5.1037
한국데이터정보과학회지

Multivariate control charts for monitoring correlation coefficients in dispersion matrix

Duk-Joon Chang[1] · Sunyeong Heo[2]

[12]Department of Statistics, Changwon National University

Received 23 August 2012, revised 13 September 2012, accepted 25 September 2012

Abstract

Multivariate control charts for effectively monitoring every component in the dispersion matrix of multivariate normal process are considered. Through the numerical results, we noticed that the multivariate control charts based on sample statistic V_i by Hotelling or W_i by Alt do not work effectively when the correlation coefficient components in dispersion matrix are increased. We propose a combined procedure monitoring every component of dispersion matrix, which operates simultaneously both control charts, a chart controlling variance components and a chart controlling correlation coefficients. Our numerical results show that the proposed combined procedure is efficient for detecting changes in both variances and correlation coefficients of dispersion matrix.

Keywords: Average run length, false alarm rate, likelihood ratio test statistic, Markov chain approach.

1. Introduction

Statistical control chart is used for continuously monitoring the production process to quickly detect any shifts that may produce deterioration in the quality of the output. Hence, the purpose of a control chart is to detect assignable causes of variation so that these causes of variation can be found and eliminated. During the control process, one wishes to detect any departure from a target state as quickly as possible and identify which attributes are responsible for the deviation.

The ability of a control chart to detect process changes is determined by the length of time required for the chart to signal when the process is out of control, and frequency of false alarm. Therefore a good control chart should quickly detect changes while producing few false alarms. Traditional practice in using a control chart is to take samples from the process at fixed sampling interval (FSI), and the length of the sampling interval between sampling times t_i and t_{i-1}, is constant for all i $(i = 1, 2, \cdots)$ in FSI chart.

The run length (RL) is defined as the random number of samples required for the chart to signal and the ARL is the expected value of the RL. Therefore, the expected time to signal is simply the product of the ARL and the length of the fixed sampling interval d in FSI chart, so the ARL can be thought of as the expected time to signal. Researchers have

[1] Professor, Department of Statistics, Changwon National University, Changwon 641-773, Korea.
[2] Corresponding author: Associate professor, Department of Statistics, Changwon National University, Changwon 641-773, Korea. E-mail : syheo@changwon..ac.kr

studied the modified control procedures to reduce average time to signal when a change of the production process has occurred such as EWMA chart, CUSUM chart, variable sampling interval (VSI) procedure, and so on.

Many situations in industrial quality control involve a vector of measurements of two or more related quality variables rather than single quality variable or single process parameter. When the quality variables are correlated, one could obtain better sensitivity by using multivariate control procedures than separate control variables for each of the quality variables.

The original work on multivariate control chart was introduced by Hotelling (1947) and became popular in recent years. Alt (1984) reviewed much of the study on multivariate control charts. A multivariate extension of the EWMA control chart was presented by Lowry *et al.* (1992). Multivariate control charts were studied by Cho (2010), Chang and Heo (2012). Most of the studies on multivariate control charts have been concentrated on monitoring mean vector of multivariate normal process.

The changes in the correlation coefficients of quality variables may often largely affect the quality of outputs like chemical industry. Through the numerical performances, we found that the multivariate control charts based on sample statistic V_i by Hotelling (1947) or W_i by Alt (1984) well respond to the changes of variance components, but they do not detect well on the changes of correlation coefficients in dispersion matrix.

In addition, our numerical performances show that the control charts based on their sample statistics have a limit that in some region where a component of correlation coefficients increases from the target value, the out of state ARL value is greater than the in control state ARL value even though the process has changed.

In this paper, we propose a combined procedure to overcome the limit which operates two multivariate EWMA charts at the same time for monitoring both the variances and the correlation coefficients in dispersion matrix. Through numerical results, we found that the proposed control procedure effectively responds to the changes from both the variances and correlation coefficients in dispersion matrix.

2. Notation and types of shifts

Suppose that the production process of interest has p quality variables whose distribution is multivariate normal with mean vector $\underline{\mu}$ and dispersion matrix Σ, and $(\underline{\mu}_0, \Sigma_0)$ is the known target process values for $(\underline{\mu}, \Sigma)$. The target dispersion matrix of p quality variables is represented as

$$\Sigma_0 = \begin{pmatrix} \sigma_{10}^2 & \rho_{120}\sigma_{10}\sigma_{20} & \cdots & \rho_{1p0}\sigma_{10}\sigma_{p0} \\ \rho_{210}\sigma_{10}\sigma_{20} & \sigma_{20}^2 & \cdots & \rho_{2p0}\sigma_{20}\sigma_{p0} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p10}\sigma_{10}\sigma_{p0} & \rho_{p20}\sigma_{20}\sigma_{p0} & \cdots & \sigma_{p0}^2 \end{pmatrix}.$$

At each sampling time i $(i = 1, 2, \cdots)$, we take a sequence of random vector $\underline{X}_i' = (\underline{X}_{i1}', \underline{X}_{i2}', \cdots, \underline{X}_{in}')$ where $\underline{X}_{ij}' = (X_{ij1}, X_{ij2}, \cdots, X_{ijp})$. Thus $\underline{X}_i$ is an $np \times 1$ column vector. Then the jk th element X_{ijk} of $\underline{X}_i$ is the j th observation for k th quality variable at each sampling time i $(j = 1, 2, \cdots, n; k = 1, 2, \cdots, p)$.

When the process is in control then the ARL should be large so that the frequency of false alarms is low. When the process is out of control then the ARL should be small so that the changes will be detected quickly. To compare the performances of the considered charts, the charts should be set up so that both have the same ARL when the process is in control. This means that the false alarm rates when the process is in control are the same for the considered charts.

For simplicity in our numerical computation, we assume that target mean vector $\underline{\mu}_0 = \underline{0}'$, all diagonal elements of Σ_0 are 1 and off-diagonal elements of Σ_0 are 0.4. The numerical results were obtained when the ARL of in control state was approximately equal to 400, and the sample size for each quality variable is 3 for $p = 2$ or $p = 4$.

To investigate and compare the performances of proposed multivariate control chart, there are various forms in which the shifts could take place on dispersion matrix. But, for simplicity in this study, we consider the following typical types of shifts and evaluate the performances of such types in the process. In fact, a shift can occur only on the process parameters of dispersion matrix:

(1) V_i : σ_{10} in Σ_0 is increased to $[1 + 0.1i]$, $i = 1, 2, \cdots, 12$.
(2) C_i : ρ_{120} and ρ_{210} of Σ_0 are changed to $[0.4 + 0.i]$, $i = \pm 4, \pm 3, \cdots, 0$.
(3) (V_4, C_{-4}), (V_3, C_{-3}), $\cdots$, (V_3, C_3), (V_4, C_4) are considered.

3. Single control chart procedure

To monitor process dispersion of single quality variable, one can use S^2 chart or R chart, where S^2 denote an unbiased sample variance for a random sample of size n from a process. The S^2 chart signals for large values of S_i^2 or equivalently for large values of $T_i = (n-1)S_i^2/\sigma_0^2$ where σ_0^2 is target value of the process dispersion σ^2 and S_i^2 is obtained at sampling occasion i.

For multivariate case, one possible multivariate version of T_i is

$$V_i = \sum_{j=1}^{n} (\underline{X}_{ij} - \underline{\bar{X}_i})' \Sigma_0^{-1} (\underline{X}_{ij} - \underline{\bar{X}_i}) \tag{3.1}$$

and the $p \times p$ sample dispersion matrix S_i is $A_i/(n-1)$ where $A_i = \sum_{j=1}^{n} (\underline{X}_{ij} - \underline{\bar{X}}_i)(\underline{X}_{ij} - \underline{\bar{X}}_i)'$. Hotelling (1947) proposed that Lawley-Hotelling statistic V_i can be used to monitor the process dispersion matrix Σ of p quality variables. Lawley (1938) and Hotelling (1951) studied the distribution of V_i. When the process is in control then the control statistic V_i has a chi-squared distribution with $(n-1)p$ degrees of freedom.

Multivariate Shewhart chart based on V_i would be set as $\left\{0, \chi^2_{1-\alpha}[(n-1)p]\right\}$ and a Shewhart chart based on V_i signals whenever

$$V_i \geq \chi^2_{1-\alpha}[(n-1)p]. \tag{3.2}$$

And multivariate EWMA chart based on Lawley-Hotelling statistic V_i is given by

$$Y_{V,i} = (1-\lambda)Y_{V,i-1} + \lambda V_i, \tag{3.3}$$

where $Y_{V,0} = \omega_V \cdot I_{(\omega_V \geq 0)}$ and $0 < \lambda \leq 1$. This chart signals whenever $Y_{V,i} \geq h_V$. The upper control limit (UCL) h_V can be obtained by Markov chain approach where h_V can be obtained to satisfy a specified in control ARL. When the process is out of control state, then it is difficult to obtain the exact distribution of V_i. Thus, in order to evaluate the performances of the considered charts it is necessary to use computer simulations. Markov chain approach for multivariate chart can be referred from Chang *et al.* (2003) or Im and Cho (2009).

Another sample statistic for monitoring dispersion matrix can be constructed by using likelihood ratio test (LRT) statistic for testing $H_0 : \Sigma = \Sigma_0$ vs $H_1 : \Sigma \neq \Sigma_0$. In this case, the region above the UCL corresponds to the LRT rejection region. The LRT statistic W_i for monitoring dispersion martix is given by

$$W_i = tr(A_i\Sigma_0^{-1}) - n\ \ln|A_i| + n\ \ln|\Sigma_0| + np\ \ln n - np, \tag{3.4}$$

Alt (1984) proposed that LRT statistic W_i can be used to monitor Σ of p quality variables. Multivariate Shewhart chart based on LRT statistic W_i signals whenever

$$W_i \geq h_W. \tag{3.5}$$

And multivariate EWMA chart based on W_i is given by

$$Y_{W,i} = (1-\lambda)Y_{W,i-1} + \lambda W_i, \tag{3.6}$$

where $Y_{W,0} = \omega_W \cdot I_{(\omega_W \geq 0)}$ and $0 < \lambda \leq 1$. If the chart statistic $Y_{W,i}$ exceeds UCL h_W, the process is deemed out-of-control state and assignable causes are sought. Since it is difficult to obtain the exact distribution of W_i when the process is in control or out of control state, UCL h_W and performances of the charts are obtained by simulation.

Numerical results in Table 3.1 show that multivariate control chart based on sample statistic V_i is superior to the corresponding charts based on LRT statistic W_i, and performances of CUSUM chart are more efficient than those of Shewhart chart when shifts of variance components in Σ have occurred.

And Table 3.2 shows that ARL values of multivariate chart based on V_i are larger in some changes of process than in control state, and so the multivariate chart based on V_i is not proper to use for detecting the changes of correlation coefficients in Σ. Table 3.2 also shows that LRT statistic W_i does not properly detect the changes of correlation coefficients. ARL performances of our proposed combined procedure are also given in the last column of Table 3.1 through Table 3.3.

Table 3.1 ARL performances when σ_1 has changed ($p = 2$, $\sigma_{10} = 1.0$)

types of shift	Shewhart chart		EWMA chart ($\lambda = 0.2$)		
	V_i	W_i	V_i	W_i	Combined
in control	400.00	399.99	400.00	399.99	400.03
V_1	188.64	372.98	137.13	348.80	147.28
V_2	93.10	293.65	60.43	237.25	64.39
V_3	50.28	187.65	33.16	133.46	34.73
V_4	30.56	108.27	21.63	71.13	21.93
V_5	19.88	63.04	15.74	40.58	15.52
V_6	13.87	38.34	12.21	25.89	11.82
V_7	10.38	24.54	9.92	18.07	9.48
V_8	8.13	16.69	8.35	13.45	7.85
V_9	6.48	12.01	7.22	10.62	6.71
V_{10}	5.38	9.07	6.32	8.65	5.84
V_{11}	4.54	7.15	5.62	7.23	5.18
V_{12}	3.96	5.76	5.07	6.21	4.64

Table 3.2 ARL performances when ρ_{12} and ρ_{21} have changed ($p = 2$, $\rho_{120} = 0.4$)

types of shift	Shewhart chart		EWMA chart ($\lambda = 0.2$)		
	V_i	W_i	V_i	W_i	Combined
C_{-4}	77.76	194.07	62.59	124.70	49.75
C_{-3}	112.04	260.87	88.85	195.55	92.29
C_{-2}	170.51	328.28	137.01	284.31	184.26
C_{-1}	265.13	383.27	229.17	364.16	334.36
in control	400.00	399.99	400.00	399.99	400.03
C_0	508.19	383.44	677.45	361.29	238.51
C_2	543.42	334.86	996.03	263.12	110.36
C_3	505.68	268.12	1241.01	149.19	56.62
C_4	434.53	184.54	1380.77	63.52	32.67

Table 3.3 ARL performances when both σ_1 and ρ_{21} have changed ($p = 2$, $\sigma_{10} = 1.0$, $\rho_{120} = 0.4$)

types of dhift	Shewhart chart		EWMA chart ($\lambda = 0.2$)		
	V_i	W_i	V_i	W_i	Combined
$(V_4.C_{-4})$	14.86	39.21	13.05	26.55	13.11
$(V_3.C_{-3})$	25.93	84.62	19.50	54.40	21.05
$(V_2.C_{-2})$	52.63	191.49	35.02	135.78	42.36
$(V_1.C_{-1})$	134.35	339.63	90.94	300.17	121.91
in control	400.00	399.99	400.00	399.99	400.03
$(V_1.C_1)$	239.17	364.51	208.09	331.23	108.02
$(V_2.C_2)$	128.82	266.52	110.14	192.56	37.20
$(V_3.C_3)$	72.34	151.97	64.66	85.01	17.38
$(V_4.C_4)$	44.30	75.22	42.64	33.85	9.88

4. Combined control charts procedure

In this section, we consider a combined procedure which uses two separate multivariate EWMA charts for p variance components and for $s = p(p-1)/2$ correlation coefficient components, respectively. The proposed combined procedure signals if one of the two charts signals.

From the numerical performances in Table 3.1 through Table 3.3, we found that multivariate EWMA chart based on Lawley-Hotelling V_i statistic works well when variance components in dispersion matrix has changed. On the other hand, we noticed that the chart based on V_i does not work well under the changes of correlation coefficients in dispersion matrix. Therefore, we propose a multivariate chart to monitor all correlation coefficients simultaneously, and also propose a combined procedure which is operating EWMA control charts for both variance components based on V_i and correlation coefficients based on T_i^2 simultaneously.

Multivariate control chart for simultaneously monitoring s correlation coefficients $\underline{\rho} = (\rho_{12}, \rho_{13}, \cdots, \rho_{1p}, \rho_{23}, \cdots, \rho_{2p}, \cdots, \rho_{p-1,p})'$ of p quality variables can be constructed by forming multivariate extension of the univariate EWMA chart statistic based on the estimator r_{mu} of ρ_{mu} for the bivariate normal population $N_2(\mu_m, \mu_u, \sigma_m^2, \sigma_u^2, \rho_{mu})$ as

$$r_{mu} = \frac{\sum_{j=1}^{n}(X_{jm} - \mu_{m0})(X_{ju} - \mu_{u0})}{n\sigma_{m0}\sigma_{u0}}.$$

Then, an EWMA chart based on the control statistic r_{mu} for ρ_{mu} can be constructed by using the statistic

$$Y_i = (1-\lambda)Y_{i-1} + \lambda r_{mu},\ i = 1, 2, \cdots \tag{4.1}$$

where $0 < \lambda \leq 1$. For multivariate case, if we let the control statistic for ρ_{lm} be r_{lm}, then the vectors of EWMA's can be defined as

$$\begin{aligned} \underline{Y_i}' &= (r_{12}, r_{13}, \cdots, r_{1p}, r_{23}, \cdots, r_{2p}, \cdots, r_{p-1,p-2}, r_{p-1,p})' \\ &= (Y_{i1}, Y_{i2}, \cdots, Y_{i,p-1}, Y_{ip}, \cdots, Y_{i,2p-3}, \cdots, Y_{i,s-1}, Y_{i,s})', \end{aligned} \tag{4.2}$$

where the corresponding component for ρ_{mu} can be represented as

$$(1-\lambda)^i Y_{m,u,0} + \sum_{k=1}^{i} \lambda(1-\lambda)^{i-k} Z_{imu},$$

and $Z_{imu} = \left(\sum_{j=1}^{n}(X_{ijm} - \mu_{m0})(X_{iju} - \mu_{u0})\right)/(n\sigma_{m0}\sigma_{u0}) - \rho_{\mu 0}$. Then the multivariate EWMA vector can be expressed as

$$\underline{Y}_i = (1-\lambda)\underline{Y}_{i-1} + \lambda \underline{Z}_i, \tag{4.3}$$

where $\underline{Z}_i = (Z_{i1}, Z_{i2}, \cdots, Z_{is})' = (Z_{i12}, Z_{i13}, \cdots, Z_{i1p}, Z_{i23}, \cdots, Z_{i2p}, \cdots, Z_{i,p-1,p})'$. Then, the dispersion matrix of $\underline{Y}_i$ with dimension $s \times s$ when the process is in control and $\underline{Y}_0 = \underline{0}$ is given by

$$\Sigma_{\underline{Y_i}} = \left\{ \frac{\lambda[1-(1-\lambda)^{2i}]}{2-\lambda} \right\} \cdot \Sigma_{\underline{Z}} \tag{4.4}$$

and

$$\Sigma_{\underline{Z}} = \begin{pmatrix} Var(Z_{i12}) & Cov(Z_{i12}, Z_{i13}) & \cdots & Cov(Z_{i12}, Z_{i,p-1,p}) \\ & Var(Z_{i13}) & \cdots & Cov(Z_{i13}, Z_{i,p-1,p}) \\ & & \ddots & \vdots \\ & Sym & & Var(Z_{i,p-1,p}) \end{pmatrix}, \tag{4.5}$$

where

$$\begin{aligned} Var(Z_{ipq}) &= \frac{1+\rho_{pq0}^2}{n}, \\ Cov(Z_{ipq}, Z_{ipr}) &= \frac{\rho_{qr0} + \rho_{pq0}\rho_{pr0}}{n}, \\ Cov(Z_{ipq}, Z_{irw}) &= \frac{\rho_{pr0}\rho_{qw0} + \rho_{pw0}\rho_{qr0}}{n} \end{aligned}$$

and the subscripts p, q, r and w are different each other.

Therefore, we propose a combined procedure to monitor every component in dispersion matrix, which applies the following two EWMA charts simultaneously, and the combined procedures signals if or if a multivariate EWMA chart for $\underline{\rho}$

$$T_i^2 = \underline{Y}_i' \Sigma_{\underline{Y}_i}^{-1} \underline{Y}_i > h,$$

where h is chosen to achieve a specified in control ARL by simulation. And, the overall false alarm probability based on (V_i, T_i^2) is $(1-\alpha_v)(1-\alpha_{T^2})$ where the signal probability for V_i and T_i^2 are α_v and α_{T^2}, respectively.

5. Concluding remarks

Up to the present, most of the studies on multivariate control charts have been concentrated on monitoring mean vector of multivaiate normal process and little research to monitor correlation coefficients in dispersion matrix is found.

In this paper, we noticed that multivariate control procedure based on sample statistic V_i by Hotelling (1947) or W_i by Alt (1984) can not be recommended for monitoring both variance components and correlation coefficient components in Σ. These control charts do not perform properly especially when the correlation coefficient of quality variables is increasing.

To overcome this limit, we propose a combined procedure to simultaneously operate both EWMA chart for variance components and EWMA chart for the vector of correlation coefficients. Through our numerical results, we found that the proposed combined procedure is an efficient scheme for detecting both variances and correlation coefficients in dispersion matrix.

Table 5.1 ARL values of the considered multivariate Shewhart and EWMA charts

	$p=2$					$p=4$				
	Shewhart		EWMA ($\lambda=0.2$)			Shewhart		EWMA ($\lambda=0.2$)		
	V_i	W_i	V_i	W_i	combined	V_i	W_i	V_i	W_i	combined
in control	400.0	400.0	400.0	400.0	400.0	400.0	400.0	400.0	400.0	400.0
V_1	188.6	373.0	137.1	348.8	147.3	239.0	307.3	182.9	240.2	187.4
V_3	50.3	187.6	33.2	133.5	34.7	77.7	119.8	49.8	80.7	47.5
V_5	19.9	63.0	15.7	40.6	15.5	29.6	43.3	22.7	36.4	19.5
V_7	10.4	24.5	9.9	18.1	9.5	14.6	18.8	13.8	22.7	11.2
V_9	6.5	12.0	7.2	10.6	6.7	8.6	10.1	9.9	16.8	7.5
V_{11}	4.5	7.1	5.6	7.2	5.2	5.8	6.2	7.6	13.5	5.5
C_{-3}	112.0	260.9	88.9	195.5	92.3	186.2	244.3	167.3	195.2	90.6
C_{-1}	265.1	383.3	229.2	364.2	334.4	332.1	358.8	306.4	324.3	296.8
C_1	508.2	383.4	677.5	361.3	238.5	431.6	427.3	489.7	454.9	357.1
C_3	505.7	268.1	1241.0	149.2	56.6	373.4	376.3	585.5	484.3	101.3
$(V_3.C_{-3})$	25.9	84.6	19.5	54.4	21.1	45.3	68.4	33.5	50.9	24.9
$(V_1.C_{-1})$	134.3	339.6	90.9	300.2	121.9	197.8	267.7	145.7	197.1	148.2
$(V_1.C_1)$	239.2	364.5	208.1	331.2	108.0	258.6	324.7	219.6	283.4	166.6
$(V_3.C_3)$	72.3	152.0	64.7	85.0	17.4	83.1	130.1	68.8	105.9	23.1

References

Alt, F. B. (1984). Multivariate control charts. In *Encyclopedia of Statistical Sciences*, edited by S. Kotz and M. L. Johnson, Wiley, New York.

Chang, D. J. and Heo, S. (2012). Switching properties of CUSUM charts for controlling mean vector. *Journal of the Korean Data & Information Science Society*, **23**, 859-866.

Chang, D. J., Kwon, Y. M. and Hong, Y. W. (2003). Markovian EWMA control chart for several correlated quality characteristics. *Journal of the Korean Data & Information Science Society*, **14**, 1045-1053.

Cho, G. Y. (2010). Multivariate Shewhart control charts with variable sampling intervals. *Journal of the Korean Data & Information Science Society*, **21**, 999-1008.

Hotelling, H. (1947). *Multivariate quality control, techniques of statistical analysis*, McGraw-Hill, New York, 111-114.

Hotelling, H. (1951). A generalized t test and measure of multivariate dispersion, *Proceedings of Second Berkely Symposium on Mathematical Statistics and Probability*, University of California Press, 23-42.

Im, C. D. and Cho, G. Y. (2009). Multiparameter CUSUM charts with variable sampling intervals. *Journal of the Korean Data & Information Science Society*, **20**, 593-599.

Lawley, D. N. (1938). A generalization of Fisher's z test. *Biometrika*, **30**, 180-187.

Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control charts. *Technometrics*, **34**, 46-53.