Journal of the Korean Data & Information Science Society
2015, **26**(4), 1027–1034
http://dx.doi.org/10.7465/jkdi.2015.26.4.1027
한국데이터정보과학회지

A statistical quality control for the dispersion matrix[†]

Jinnam Jo[1]

[1]Department of Statistics and Information Science, Dongduk Women's University

Received 30 June 2015, revised 17 July 2015, accepted 20 July 2015

Abstract

A control chart is very useful in monitoring various production process. There are many situations in which the simultaneous control of two or more related quality variables is necessary. When the joint distribution of the process variables is multivariate normal, multivariate Shewhart control charts using the function of the maximum likelihood estimator for monitoring the dispersion matrix are considered for the simultaneous monitoring of the dispersion matrix. The performances of the multivariate Shewhart control charts based on the proposed control statistic are evaluated in term of average run length (ARL). The performance is investigated in three cases, where the variances, covariances, and variances and covariances are changed respectively. The numerical results show that the performances of the proposed multivariate Shewhart control charts are not better than the control charts using the trace of the covariance matrix in the Jeong and Cho (2012) in terms of the ARLs.

Keywords: Average run length, dispersion matrix, maximum likelihood estimator, multivariate Shewhart control chart.

1. Introduction

Recently there has been a growing interest in multivariate statistical process control. The multivariate control charts for monitoring the mean vector has already been studied in depth. However, very little attention has been paid to monitor the dispersion matrix. There are various approaches to constructing control charts for multivarite data. The original work in multivariate control charts was introduced by Hotelling (1947) which is the multivariate Shewhart chart based on Hotelling's T^2 statistic. Jackson (1959), and Ghare and Torgerson (1968) presented multivariate Shewhart control charts based on Hotelling's T^2 statistic. Other multivariate Shewhart control charts are discussed at Alt (1984), Wierda (1994), and Lowry and Montgomery (1995). Simultaneously monitoring the means and variances in the production processes in the univariate case was studied by Im and Cho (2009). The multivariate control charts for monitoring dispersion matrix were studied by Chang and Shin (2009), Aparisi *et al.* (2009), Na *et al.* (2010), Jeong and Cho (2012). In this paper, we study multivariate Shewhart control charts using the function of the maximum likelihood estimator for monitoring the dispersion matrix.

[†] This research was supported by the Dongduk Women's University Grant 2013.
[1] Professor, Department of Statistics and Information Science, Dongduk Women's University, Seoul 136-714, Korea. E-mail: jinnam@dongduk.ac.kr

2. Notation and assumptions

It will be convenient to let σ represent the vector of standard deviations of the p variables in the multivariate normal distribution of the observation. According to the notation in Jeong and cho (2012), let μ_0, Σ_0, and σ_0 be the in-control values of μ, Σ, and σ. And we assume that the in-control parameter values are known. Suppose that we take a sample of $n \geq 1$ independent observation vectors at each sampling point, where the sampling points are d time units apart. Let X_{kij} be the jth observation for variable i at sampling point k for $k = 1, 2, \cdots,\ i = 1, 2, \cdots, p$ and $j = 1, 2, \cdots, n$, and let the corresponding standardized observation be

$$Z_{kij} = (X_{kij} - \mu_{0i})/\sigma_{0i},$$

where μ_{0i} is the ith component of μ_0 and σ_{0i} is the ith component of σ_0. Also let

$$\boldsymbol{Z}_{kj} = (Z_{k1j}, Z_{k2j}, \cdots, Z_{kpj})',\ j = 1, 2, \cdots, n$$

be the vector of standardized observations for the jth observation at sampling point k. Let Σ_Z be the covariance matrix of $\boldsymbol{Z}_{kj}$, and let Σ_{Z0} be the in-control value of Σ_Z. The in-control distribution of Z_{kij} is standard normal, so Σ_{Z0} is also the in-control correlation matrix of the unstandardized observations.

Let $\bar{X}_{ki} = \sum_{j=1}^{n} X_{kij}/n$ be the sample mean for variable i at sampling point k, and define the standardized sample mean to be

$$Z_{ki} = \sqrt{n}(\bar{X}_{ki} - \mu_{0i})/\sigma_{0i},\ i = 1, 2, \cdots, p.$$

When $n \geq p$, some control statistics used for monitoring Σ are functions of the sample estimates $\widehat{\Sigma}_Z$. At sampling point k, let $\widehat{\Sigma}_{Z_k}$ be the maximum likelihood estimator of Σ_Z, where the (i, i') element of $\widehat{\Sigma}_{Z_k}$ is $\sum_{j=1}^{n} Z_{kij} Z_{ki'j}/n$.

We investigate a number of multivariate Shewhart control charts for Simultaneously monitoring the dispersion matrix.

3. Multivariate Shewhart control charts

We will discussmultivariate Shewhart control charts based on the statistic L_k proposed by Hui (1980) which use the sample generalized variances for monitoring the process dispersion matrix

$$L_k = \frac{|\widehat{\Sigma}_k|}{|\Sigma_0|}, \tag{3.1}$$

where Σ_0 is known. It is known that the distribution of $\sqrt{n}(L_k - 1)$ is asymptotically normal with mean 0 and variance $2p$ (Anderson, 1958). But, we can not get the exact distribution. Because of using small sample size in the control chart study, we need to calculate control limits and ARLs by using simulations. When sample size n is large, the approximate expected value of L_k is 1 and the variance of L_k is $2np$.

For large sample size n, the approximate three-sigma control limits for L_k for monitoring Σ are then

$$\mathrm{UCL} = 1 + 3\sqrt{2np},$$
$$\mathrm{LCL} = 1 - 3\sqrt{2np}.$$

For a Shewhart control chart, control limits based on the statistic L_k would be set by using percentage point of L_k, and signals whenever

$$L_k > h, \tag{3.2}$$

where h can be obtained to satisfy a specified in-control ARL.

If the process shifts from $\Sigma_0$then it is difficult to obtain the exact distribution of L_k. Thus it is necessary to use simulations in order to obtain the percentage points of L_k when the process is out-of-control state.

In this paper, we evaluated the performance of the control charts based on the average run length performance. The ARLs for the multivariate control chart byusing (3.1) when the process is in control can be obtained by using 10,000 replications.

In our computation, each control chart was calibrated so that the on-target ARL was approximately equal to 800.0 and the sample size for each control chart was $n \geq p$ for $p = 2$ and $p = 4$. The performance of the control charts for monitoring the dispersion matrix depends on the value of Σ. The following types of shifts were considered;

(1) variances are changed and covariances are not changed,
(2) covariances are changed and variances are not changed,
(3) variances and covariances are simultaneously changed.

The control limits h and ARLs for multivariate Shewhart control charts based on the generalized variances are obtained by using 10,000 replications. Table 3.1 gives the values of h for $n \geq p$ for $p = 2, 4$ and $\rho = 0.9, 0.5, 0.3$ when the ARL is approximately 800.

Table 3.1 Values of h for multivariate Shewhart control charts for various values of p and ρ when the in-control ARL is approximately 800

p	n	ρ_0		
		$\rho_0 = 0.9$	$\rho_0 = 0.5$	$\rho_0 = 0.3$
2	2	11.2002	11.1482	11.1682
	4	7.5265	7.4995	7.5165
4	4	3.9115	3.9155	3.9106

4. Numerical performances and concluding remarks

The multivariate Shewhart control charts with control statistic given by (3.1) are compared on the basis of their ARLs. The ARLs of multivariate control chart when the process is in-control are fixed to be 800.

Table 4.1 gives p ARLs in each cell when one, two, $\cdots$, p variances are changed and covariances are not changed respectively. Here standard deviations are changed from σ_0 to $\sigma = \sqrt{c}\sigma_0$, for $c = 1.21, 1.44, 1.69, 4.00$. Table 4.1 shows that the performances of the control charts for the three different correlations are very similar.

Table 4.2 gives ARLs for $n = 2, 4$, $p = 2, 4$ and three different in-control correlation coefficients $\rho_0 = 0.9, 0.5, 0.3$ when covariances are changed and variances are not changed. Table 4.2 shows that the performances of the control charts for the larger $\rho_0 = 0.9$ are better than the other cases.

As shown in Tables 4.1-4.2, the multivariate Shewhart control charts for monitoring the dispersion matrix are effective in detecting changes in Σ. In case of larger $\rho_0 = 0.9$, multivariate Shewhart control charts are very effective in detecting changes in covariances in Σ.

For $n = 2, 4$ and $p = 2, 4$, Tables 4.3-4.5 give p ARLs in each cell when one, two, $\cdots$, p variances and p covariances are simultaneously changed, respectively. Here standard deviations are changed from σ_0 to $\sigma = \sqrt{c}\sigma_0$ for $c = 1.21, 1.44, 1.69, 4.00$, and covariances are changed from $\rho_0 = 0.9$ to $\rho = 0.72, 0.54$, $\rho_0 = 0.5$ to $\rho = 0.4, 0, 3$, and $\rho_0 = 0.3$ to $\rho = 0.21, 0.18$.

As shown in Tables 4.3-4.5, the multivariate Shewhart control charts for monitoring the dispersion matrix are also effective in detecting simultaneously changes in variances and covariances in Σ. In case of larger $\rho_0 = 0.9$, this chart is very effective in detecting changes in variances and covariances in Σ.

From Tables 4.1-4.5, the performances of the multivariate Shewhart control charts using the function of the maximum likelihood estimator for monitoring the dispersion matrix are not better than the control charts using the trace of the covariance matrix in the Jeong and Cho (2012) in terms of the ARLs.

Table 4.1 ARL when variances are changed and covariances are not changed

c	$\rho_0 = 0.9$			$\rho_0 = 0.5$			$\rho_0 = 0.3$		
	$n = 2$ $p = 2$	$n = 4$ $p = 2$	$n = 4$ $p = 4$	$n = 2$ $p = 2$	$n = 4$ $p = 2$	$n = 4$ $p = 4$	$n = 2$ $p = 2$	$n = 4$ $p = 2$	$n = 4$ $p = 4$
$c = 1.00$	800.00	800.20	800.14	800.00	800.20	800.14	800.00	800.20	800.14
$c = 1.21$	434.94 252.34	355.01 171.31	518.30 349.22 242.41 169.41	428.75 247.68	347.95 165.97	526.43 355.21 240.11 168.07	436.86 251.06	357.00 168.92	527.59 348.44 244.13 169.59
$c = 1.44$	259.17 106.63	184.19 53.78	364.34 178.58 96.53 55.29	260.96 103.73	179.58 53.91	364.07 180.73 97.06 55.00	266.48 105.08	176.68 53.70	362.20 183.05 96.09 54.80
$c = 1.69$	171.13 52.95	102.49 23.02	263.10 102.72 46.15 23.49	169.37 52.00	101.74 22.72	263.18 102.39 46.27 23.70	169.78 52.62	104.43 23.13	264.46 101.70 46.55 23.79
$c = 4.00$	28.34 5.44	11.46 2.06	61.09 11.56 4.19 2.23	28.05 5.33	11.34 2.08	60.89 11.56 4.11 2.18	28.22 5.31	11.28 2.04	60.77 11.73 4.11 2.25

Table 4.2 ARL when covariances are changed and variances are not changed

ρ	$\rho_0 = 0.9$			ρ	$\rho_0 = 0.5$			ρ	$\rho_0 = 0.3$		
	$n = 2$ $p = 2$	$n = 4$ $p = 2$	$n = 4$ $p = 4$		$n = 2$ $p = 2$	$n = 4$ $p = 2$	$n = 4$ $p = 4$		$n = 2$ $p = 2$	$n = 4$ $p = 2$	$n = 4$ $p = 4$
$\rho = 0.90$	799.97	800.03	800.01	$\rho = 0.50$	800.03	799.97	799.96	$\rho = 0.30$	800.02	799.99	799.91
$\rho = 0.81$	146.47	80.10	32.10	$\rho = 0.45$	650.22	98.11	492.97	$\rho = 0.27$	754.77	726.30	675.02
$\rho = 0.72$	66.71	30.88	10.20	$\rho = 0.40$	546.44	85.64	324.78	$\rho = 0.24$	712.12	684.26	580.62
$\rho = 0.63$	42.60	17.91	5.55	$\rho = 0.35$	482.88	13.08	236.18	$\rho = 0.21$	673.64	648.89	505.60
$\rho = 0.54$	32.09	12.75	4.00	$\rho = 0.30$	427.37	340.48	176.18	$\rho = 0.18$	647.71	601.97	454.94
$\rho = 0.45$	26.13	10.33	3.14	$\rho = 0.25$	393.80	301.41	142.38	$\rho = 0.15$	629.76	572.04	417.92
$\rho = 0.36$	22.64	8.77	2.72	$\rho = 0.20$	368.00	271.00	118.91	$\rho = 0.12$	623.18	560.71	379.24
$\rho = 0.27$	20.71	7.81	2.49	$\rho = 0.15$	347.88	259.82	103.05	$\rho = 0.09$	698.08	551.83	347.75
$\rho = 0.18$	19.51	7.38	2.30	$\rho = 0.10$	339.78	246.03	94.04	$\rho = 0.06$	586.08	540.54	334.35
$\rho = 0.09$	18.77	6.93	2.22	$\rho = 0.05$	331.25	239.92	89.36	$\rho = 0.03$	570.05	520.34	324.06

Table 4.3 ARL when variances and covariances are changed ($\rho_0 = 0.9$)

c	ρ	n, p		
		$n=2,\ p=2$	$n=4,\ p=2$	$n=4,\ p=4$
$c=1.00$	$\rho=0.90$	799.97	800.03	800.01
$c=1.21$	$\rho=0.72$			8.50
		46.51	19.47	7.47
		31.71	11.84	6.27
				5.55
	$\rho=0.54$			3.57
		23.45	8.97	3.28
		17.63	6.54	2.95
				2.70
$c=1.44$	$\rho=0.72$			7.30
		33.44	13.76	5.68
		18.66	6.99	4.51
				3.63
	$\rho=0.54$			3.26
		17.76	6.74	2.77
		11.11	4.04	2.36
				2.08
$c=1.69$	$\rho=0.72$			6.52
		25.38	9.90	4.55
		12.05	4.41	3.41
				2.68
	$\rho=0.54$			2.97
		14.62	5.25	2.45
		7.63	2.86	2.05
				1.78
$c=1.96$	$\rho=0.72$			3.78
		8.09	3.04	2.10
		2.87	1.33	1.49
				1.24
	$\rho=0.54$			2.18
		5.65	2.18	1.51
		2.38	1.21	1.25
				1.12
$c=4.00$	$\rho=0.72$			1.02
		1.30	1.06	1.00
		1.14	1.01	1.00
				1.00
	$\rho=0.54$			1.01
		1.25	1.04	1.00
		1.10	1.01	1.00
				1.00

Table 4.4 ARLs when variances and covariances are changed ($\rho_0 = 0.5$)

c	ρ	$n,\ p$		
		$n=2,\ p=2$	$n=4,\ p=2$	$n=4,\ p=4$
$c=1.00$	$\rho=0.50$	800.03	799.97	799.96
$c=1.21$				226.13
	$\rho=0.40$	312.19	222.56	155.98
		182.95	111.54	113.96
				82.40
				126.84
	$\rho=0.30$	247.69	164.79	92.24
		148.47	85.36	67.78
				45.23
$c=1.44$				163.23
	$\rho=0.40$	190.75	116.74	87.40
		79.29	38.84	51.32
				31.06
				94.25
	$\rho=0.30$	157.42	91.14	55.16
		68.20	31.29	32.97
				21.12
$c=1.69$				123.15
	$\rho=0.40$	129.74	70.26	54.40
		41.44	17.72	27.01
				14.93
				74.59
	$\rho=0.30$	104.98	56.41	35.15
		36.33	14.75	18.64
				11.18
$c=1.96$				34.19
	$\rho=0.40$	23.87	8.87	8.16
		4.85	1.91	3.36
				1.19
				22.94
	$\rho=0.30$	21.17	7.86	6.28
		4.58	1.84	2.85
				1.78
$c=4.00$				1.39
	$\rho=0.40$	2.11	1.36	1.05
		1.32	1.05	1.01
				1.00
				1.34
		2.05	1.35	1.04
	$\rho=0.30$	1.30	1.04	1.00
				1.00

Table 4.5 ARLs when variances and covariances are changed ($\rho_0 = 0.3$)

c	ρ	$n,\ p$		
		$n=2,\ p=2$	$n=4,\ p=2$	$n=4,\ p=4$
$c = 1.00$	$\rho = 0.3$	800.00	800.20	800.08
$c = 1.21$	$\rho = 0.21$			338.95
		374.29	283.66	232.38
		218.74	142.58	163.21
				117.67
	$\rho = 0.18$			302.31
		365.10	271.82	211.05
		213.15	136.20	147.89
				106.65
$c = 1.44$	$\rho = 0.21$			241.08
		234.06	150.63	123.16
		93.15	46.46	69.29
				40.86
	$\rho = 0.18$			215.85
		221.35	140.64	112.87
		90.30	45.25	64.64
				37.69
$c = 1.69$	$\rho = 0.21$			178.43
		152.01	87.13	73.70
		48.06	20.61	34.85
				18.58
	$\rho = 0.18$			163.97
		146.99	83.30	68.91
		46.98	19.87	32.76
				17.58
$c = 1.96$	$\rho = 0.21$			46.08
		26.45	10.04	9.71
		5.13	1.98	3.69
				2.06
	$\rho = 0.18$			41.71
		25.52	9.97	9.08
		5.04	1.99	3.55
				2.02
$c = 4.00$	$\rho = 0.21$			1.59
		2.32	1.45	1.07
		1.34	1.06	1.01
				1.00
	$\rho = 0.18$			1.52
		2.30	1.45	1.07
		1.33	1.06	1.01
				1.00

References

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

Anderson, T. W. (1958). An introduction to multivariate statistical analysis, Wiley, New York.

Aparisi, F., Jabaioyes, J. and Carrion, A. (2009). Statistical properties of the $|S|$ multivariate control chart. *Communications in Statistics-theory and methods*, **28**, 2671-2686.

Chang, D. J. and Shin, J. K. (2009). Variable sampling interval control charts for variance-covariance matrix. *Journal of the Korean Data & Information Science Society*, **21**, 999-1008.

Ghare, P. H. and Torgerson, P. E. (1968). The multicharacteristic control chart. *Journal of Industrial Engineering*, **19**, 269-272.

Hotelling, H. (1947). Multivariate quality control, techniques of statistical analysis, McGraw-Hill, New York, 111-184.

Hui, Y. V. (1980). Topics in statistical process control, Ph. D. Thesis, Virginia Politechnic Institute and State University, Blacksberg, Virginia.

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.

Jackson, J. S. (1959). Quality control methods for several related variables. *Technometrics*, **1**, 359-377.

Jeong, J. I. and Cho, G. Y. (2012). Multivariate Shewhart control charts for monitoring the variance-covariance matrix. *Journal of the Korean Data & Information Science Society*, **23**, 617-628.

Lowry, C. A. and Montgomery, D. C. (1995). A review of multivariate control charts. *IIE Transactions*, **27**, 800-810.

Na, O., Ko, B. and Lee, S. (2010). Cusum of squares test for discretely observed sample from multidimensional diffusion processes. *Journal of the Korean Data & Information Science Society*, **21**, 555-560.

Wierda, S. J. (1994). Multivariate statistical process control - recent results and directions for future research. *Statistica Neerlandica*, **48**, 147-168.