A Multivariate Process Variability Monitoring Based on Individual Observations

Maman A. Djauhari (Corresponding author)
Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia
81310 UTM Skudai, Johor Bahru, Malaysia
Tel: 60-7-553-4273 E-mail: maman@utm.my

The research is financed by Government of Malaysia under FRGS vote no. 78484 and Foreign Academic Research Grant UTM vote no. 77536

Abstract

In order to have a better understanding whether or not an additional observation has changed the covariance structure, a new statistic will be introduced. This statistic will be defined as the scatter matrix issued from augmented data set subtracted by that from historical data set. Under normality assumption, the distribution of its Frobenius norm will be derived and, for practical purpose, a chi-square approximation will be presented. This statistic and Wilks' will be used to construct a new procedure for monitoring process variability based on individual observations. The performance of this procedure in providing information about the effect of an additional observation on covariance structure is promising. An industrial application will be presented to illustrate its advantage. .

Keywords: Frobenius norm, Generalized variance, Multivariate dispersion, Scatter matrix

1. Introduction

Since last decades, the notion of manufacturing process quality becomes more and more complex. This is the main reason why quality experts have been considering process quality in multivariate setting. In this setting, one of the most important parameters is process variability. The stability of this shape parameter, which is numerically represented by covariance matrix, must be monitored. In general, there are three scenarios in multivariate process variability monitoring. First, is based on sub-grouping where the sub-group size m is greater than the number of quality characteristics p. Articles in this scenario include Yeh, Huwang and Wu (2004), Djauhari (2005) and Yeh, Lin and McGrath (2006). Second, is based on individual observations, i.e., $m = 1$ such as presented, for example, in Tracy, Young and Mason (1992), Sullivan and Woodall (1996), Khoo and Quah (2003), Huwang, Yeh and Wu (2007), and very recently Mason, Chou and Young (2009, 2010). In this scenario, the main problem is to test the effect of an additional observation on covariance structure. Third, the most recent scenario introduced in Mason, Chou and Young (2009), is based on sub-grouping where $1 < m < p$.

The idea behind the present paper was inspired by the use of Wilks' statistic (1963) for the second scenario. This monitoring procedure was originally introduced by Mason, Chou and Young (2009) and developed in Mason, Chou and Young (2010) in order to identify the quality characteristics that contribute to the out-of-control signal. What makes Wilks's statistic important in this area of industrial application is that it has direct and simple geometrical interpretation and it is easy to implement in practice especially when p is not too large. Based on Wilks' statistic, the effect of an additional observation on covariance structure is measured as the ratio of the scatter matrix determinant issued from a historical data set (HDS) and that issued from the augmented data set (ADS). The latter data set consists of HDS and an additional observation. It is thus proportional to the ratio of the generalized variance (GV) of HDS and that of ADS. Geometrically, see Anderson (2003), it is the ratio of the volume of the p-dimensional parallelotope related to HDS and that related to ADS.

Since the covariance structure is absolutely determined by the eigenvalues and eigenvectors of covariance matrix, then the use of Wilks' statistic to detect the effect of an additional observation on covariance structure might be misleading. This is caused by the fact that GV is only the product of all eigenvalues. It might happen then that Wilks' statistic fails to detect that effect whereas actually the covariance structure has changed. To illustrate the situation, it is sufficient to consider two different covariance matrices having the same GV. Let us consider the following two hypothetical covariance matrices Σ_1 and Σ_2,

$$\Sigma_1 = \begin{pmatrix} 15 & 3 \\ 3 & 3 \end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix} 20 & 8 \\ 8 & 5 \end{pmatrix}.$$

These covariance matrices represent two different covariance structures. The variance of the first and the second variables and also the correlation coefficient between them represented by Σ_1 are totally different from those represented by Σ_2. Both matrices have different set of eigenvalues. They have different Frobenius norm, i.e., $\sqrt{252}$ for Σ_1 and $\sqrt{553}$ for Σ_2. However, they have the same GV which is equal to 36. In Djauhari, Mashuri and Herwindiati (2008) we can see the use of Frobenius norm of covariance matrix as another multivariate dispersion measure besides GV. In that paper this measure is used in process variability monitoring under the first scenario.

The above illustration indicates that the use of Wilks' statistic alone might not be sufficient to describe the effect of an additional observation on covariance structure. This is a logical consequence of the use of GV as a multivariate dispersion measure. This measure has serious limitations as mentioned in Montgomery (2005) and discussed in details in Alt and Smith (1988). Therefore, another statistic is needed to have a better understanding about that effect. This is what we intend to discuss in this paper.

In what follows we introduce a new statistic that can be used, besides Wilks' statistic, for monitoring process variability based on individual observations. That statistic will be constructed based on the matrix D defined as the scatter matrix issued from ADS subtracted by that from HDS. The distribution of its Frobenius norm will be derived and, for practical purpose, a chi-square approximation will be presented. Based on these results we propose a new monitoring procedure which will give a better understanding about the effect of an additional observation on covariance structure. These are the topic in the next section. In the third section, an industrial example will be reported to illustrate the advantage of this procedure. In the last section, additional remarks will close the presentation.

2. Proposed control charting procedure

Let $X_1, X_2, \ldots, X_n, X_{n+1}$ be a random sample drawn from a p-variate normal distribution with covariance matrix Σ positive definite. The realization of $X_1, X_2, \ldots, X_n$ will be used as HDS and the union of a realization of X_{n+1} and HDS is called ADS. See Mason, Chou and Young (2009) for further details. Let

$$SS_k = \sum_{i=1}^{k} \left(X_i - \bar{X}_k\right)\left(X_i - \bar{X}_k\right)^t \text{ with } \bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i \quad \ldots\ldots (1)$$

where $k = n, n+1$. SS_k is the scatter matrix issued from HDS if $k = n$ and from ADS if $k = n+1$. Wilks (1963) proposes to use the following statistic to measure the effect of X_{n+1} on covariance structure,

$$W = \frac{|SS_n|}{|SS_{n+1}|} \text{ or, equivalently, } W = \left(\frac{n-1}{n}\right)^p \frac{|S_n|}{|S_{n+1}|} \quad \ldots\ldots (2)$$

where $|SS_k|$ is the determinant of SS_k and $|S_k|$ is the GV issued from HDS if $k = n$ and from ADS if $k = n+1$. Wilks also shows that W follows *Beta* distribution with parameters $(n-p)/2$ and $p/2$.

Due to the limitations of GV as a multivariate dispersion measure mentioned above, a careful attention must be paid when we use Wilks' statistic; two different scatter matrices might have the same value of W. To escape from this situation, in the next paragraph we define a matrix D as the scatter matrix issued from ADS subtracted by that from HDS. We will see that the use of Wilks' statistic together with the Frobenius norm of D will give a better understanding about the effect of an additional observation on covariance structure.

From (2) we know how to quantify the effect of an additional observation on covariance structure using Wilks' statistic. In the following proposition we present another quantification method based on the Frobenius norm of D. Those who are interested in the mathematical derivation are pleased to contact the author.

Proposition 1 Let $X_1, X_2, \ldots, X_n, X_{n+1}$ be a random sample of a p-variate normal distribution with covariance matrix Σ positive definite. If $D = SS_{n+1} - SS_n$ where SS_n and SS_{n+1} are defined in (1), then $\sqrt{Tr\left(D^2\right)}$ has the same distribution as $\sum_{k=1}^{p} \omega_k z_k^2$ where $z_1, z_2, \ldots, z_k$ are i.i.d. standard normal $N(0,1)$ and ω_k is the k-th eigenvalue of Σ.

2.1 A new statistic

Proposition 1 shows that the statistic

$$F = \sqrt{Tr\left(D^2\right)} \quad \ldots\ldots (3)$$

represents the effect of X_{n+1} on scatter matrix measured using the Frobenius norm of *D*. Like Wilk's statistic, it can be used to test whether or not X_{n+1} has significantly changed the covariance structure. However, *W* in (2) and *F* in (3) are two different statistics. Therefore, they might give different statistical decision. This indicates that the use of both statistics will provide a better understanding about the effect of X_{n+1} on covariance structure.

The statistic *F* is still difficult to implement in practice because its distribution in Proposition 1 is still impractical except ω_1, ω_2, …, ω_p are equal to each other. In order to handle this problem, a chi-square approximation will be discussed in the next sub-section.

2.2 A chi-square approximation

The result in Proposition 1 is still difficult to use in practice. To make it more practical, in what follows a chi-square approximation will be presented. Since many decades ago, see Solomon and Stephens (1977), it is common in practice to approximate the distribution of a linear combination of independent chi-square distributions $\sum_{k=1}^{p} \omega_k z_k^2$ by $c\chi_r^2$, where *c* is a positive constant and *r* is the corresponding degree of freedom, satisfying

$$E\left(c\chi_r^2\right) = E\left(\sum_{k=1}^{p} \omega_k z_k^2\right) \text{ and } Var\left(c\chi_r^2\right) = Var\left(\sum_{k=1}^{p} \omega_k z_k^2\right).$$

These equalities give

$$c = \frac{Tr\left(\Sigma^2\right)}{Tr\left(\Sigma\right)} \text{ and } r = \frac{\left\{Tr\left(\Sigma\right)\right\}^2}{Tr\left(\Sigma^2\right)}.$$

In the case where Σ is unknown, it is customary to replace Σ by the sample covariance matrix issued from HDS, i.e., S_n. See, for example, Montgomery (2005). Therefore, the distribution of *F* can be further approximated by

$$c\chi_r^2 \text{ with } c = \frac{Tr\left(S_n^2\right)}{Tr\left(S_n\right)} \text{ and } r = \frac{\left\{Tr\left(S_n\right)\right\}^2}{Tr\left(S_n^2\right)}. \quad \ldots\ldots (4)$$

2.3 Proposed procedure

The procedure to monitor multivariate process variability based on Wilks' statistic consists of plotting (i) the observed value of *W* and (ii) the lower control limit (LCL) which is equal to the α-th quantile of *Beta* distribution with parameters $(n - p)/2$ and $p/2$. An out-of-control signal occurs if the observed value of *W* is less than LCL. Here, α is the probability of false alarm. Instead of using Wilks' statistic, we can also use the statistic *F* in Proposition 1 where its distribution is approximated by (4). In this case, the control procedure consists of plotting (i) the observed values of *F* and (ii) the upper control limit (UCL) which is equal to the $(1-\alpha)$-th quantile of $c\chi_r^2$. An out-of-control signal occurs if the observed value of *F* exceeds UCL.

Since both statistics *W* and *F* are different, in order to handle the limitation of *W*, we propose to use both control charting procedures one after another. In the next section, an industrial example will illustrate the advantage of this procedure.

3. Industrial example

We use Wilks' statistic (2) in monitoring the process variability of B-complex vitamin production at a pharmaceutical industry based on individual observations. There are two quality characteristics under consideration, i.e., x_1 (Thickness of the tablet in mm) and x_2 (Hardness of the tablet in kg/cm^2). A HDS of size $n = 40$ gave

$$\bar{X}_n = \begin{pmatrix} 4.310 \\ 7.751 \end{pmatrix} \text{ and } S_n = \begin{pmatrix} 0.0371 & -0.0197 \\ -0.0197 & 0.0254 \end{pmatrix}.$$

During process variability monitoring, 20 individual observations were observed. These data and their

corresponding values of W and F are presented in Table 1.

To construct W and F charts, we use α = 0.0027. Therefore, the lower control limit of W chart is LCL = 0.7325. If this value is plotted together with the observed values of W in the third column of Table 1, we get the W chart such as presented in Figure 1. In this figure, an out-of-control signal occurs at the forth sample. On the other hand, based on (3) and (4), the upper control limit of F chart is UCL = 0.4032 and the corresponding F chart is presented in Figure 2. From this figure we see that F chart gives different message than W chart. According to F chart, an out-of-control signal occurs at the fifth sample and not at the forth.

This example shows that the use of W chart alone will be misleading as we have mentioned earlier. If it is used together with F chart, one after another, we get more information about the history of process variability.

4. Additional remarks

If we look more in-depth at the data in Table 1 and observe the effect of each additional observation on covariance structure in relation with the shift in correlation and also with the shift in variance, we arrive at the following interesting phenomenon. The W chart is more sensitive than F chart to the shift in correlation while F chart is more sensitive than W chart to the shift in variance. This strengthens our conclusion that the use of both charts will be more useful than a single chart.

References

Alt F. B., & Smith, N.D. (1988). Multivariate process control. P.R. Krishnaiah and C.R. Rao Eds., *Handbook of Statistics* (7, pp. 333-351). Elsevier Science Publishers.

Anderson, T. W. (2003). *Introduction to multivariate statistical analysis* (3rd ed.). New York: Wiley-Interscience.

Djauhari, M. A., Mashuri, M., & Herwindiati, D. E. (2008). A multivariate process variability monitoring. *Communication in Statistics – Theory and Methods*, *37*, 1742-1754.

Djauhari, M. A. (2005). Improved monitoring of multivariate process variability. *Journal of Quality Technology, 37*, 32-39.

Huwang, L., Yeh, A. B., & Wu, C. W. (2007). Monitoring multivariate process variability for individual observations. *Journal of Quality Technology, 39*, 258-278.

Khoo, M. B. C., & Quah, S. H. (2003). Multivariate control chart for process dispersion based on individual observations. *Quality Engineering*, *15(4)*, 639-642.

Mason, R. L., Chou, Y. M., & Young, C. J. (2009). Monitoring variation in a multivariate process when the dimension is large relative to the sample size. *Communication in Statistics – Theory and Methods, 38*, 939-951.

Mason, R. L., Chou, Y. M., & Young, C. J. (2010). Decomposition of scatter ratios used in monitoring multivariate process variability. *Communication in Statistics – Theory and Methods, 39*, 2128-2145.

Montgomery, D. C. (2005). *Introduction to statistical quality control* (5th ed.). New York: John Wiley & Sons, Inc.

Solomon, H., & Stephens, M. A. (1977). Distribution of a sum of weighted chi-square variables. *Journal of American Statistical Association, 72*, 881-885

Sullivan, J. H., & Woodall, W. H. (1996). A comparison of multivariate control charts for individual observation. *Journal of Quality Technology, 28*, 398-408.

Tracy, N. D., Young, J. C., & Mason, R. L. (1992). Multivariate control chart for individual observations. *Journal of Quality Technology, 24*, 88-95.

Wilks, S. S. (1963). Multivariate statistical outliers. *Sankhya A, 25*, 407-426.

Yeh, A. B., Huwang, L., & Wu, Y. F. (2004). A likelihood-ratio-based EWMA control chart for monitoring variability of multivariate normal processes. *IIE Transaction, 36*, 865-879.

Yeh, A. B., Lin, D. K., & McGrath, R. N. (2006). Multivariate control chart for monitoring covariance matrix. *Quality Technology & Quantitative Management, 3*, 415-436.

Table 1. Individual observations and the value of W and F statistics

No.	x_1	x_2	W	F
1	4.305	8.150	0.7921	0.1553
2	4.320	7.640	0.9816	0.0121
3	4.330	7.750	0.9996	0.0004
4	4.310	7.130	0.6081	0.3762
5	3.890	8.310	0.7649	0.4770
6	4.300	8.130	0.8108	0.1402
7	4.370	8.030	0.8592	0.0795
8	4.360	7.540	0.9447	0.0459
9	4.130	7.865	0.9781	0.0443
10	4.310	7.440	0.8609	0.0944
11	4.270	7.740	0.9972	0.0017
12	4.274	7.640	0.9717	0.0133
13	4.380	7.440	0.8861	0.0991
14	4.278	8.150	0.8035	0.1563
15	4.258	8.050	0.8890	0.0899
16	4.312	7.640	0.9802	0.0120
17	4.328	8.150	0.7818	0.1556
18	4.300	7.740	0.9995	0.0002
19	4.320	7.540	0.9339	0.0435
20	4.342	7.876	0.9668	0.0162

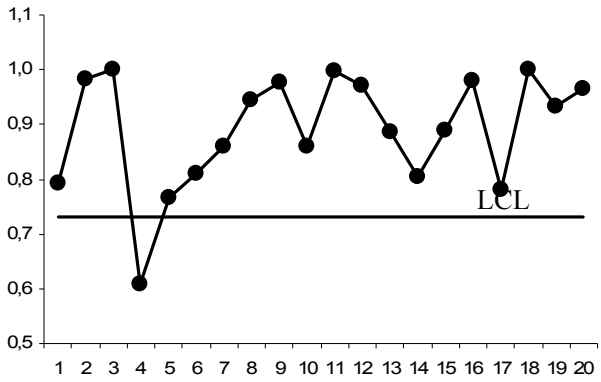

Figure 1. W chart

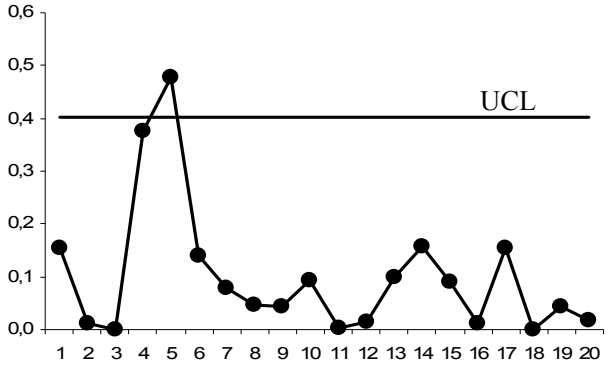

Figure 2. *F* chart