MATEMATIKA, 2011, Volume 27, Number 1, 51–57

Geometric Interpretation of Vector Variance

Maman A. Djauhari
Department of Mathematics, Universiti Teknologi Malaysia, 81310 Skudai, Johor.
e-mail: maman@utm.my

Abstract Multivariate dispersion is difficult to measure, and thus to manage, because of the complexity of covariance structure. There is no single measure that can properly represent the whole structure. The most popular and widely used measure is the generalized variance. Unfortunately, it has some serious limitations. An alternative measure that features good properties is the vector variance. However, its geometric interpretation in terms of random sample is still vague. This paper is to clarify the geometric meaning of vector variance which will ensure the proper application of this measure in practice. For that purpose we use Escoufier's operator, an operator representation of random vector, to show that sample vector variance is equal to the squared Frobenius norm of that operator in random sample setting.

Keywords Escoufier's operator, Frobenius norm, generalized variance, multivariate dispersion, vector variance

2010 Mathematics Subject Classification 62H10, 62H25, 62H86.

1 Introduction

Understanding multivariate dispersion is as important as understanding location. However, the former has received far less attention in the literature. This dearth probably results from the fact that there is no strongly suitable measure that can explain the whole covariance structure. The total variance [1,2] concerns only on the sum of all variances, the generalized variance, [3–5], possesses serious limitations, and the minimum volume of ellipsoid [6, 7] depends on the generalized variance [8]. Other examples, such as the effective variance [8,9], the square root of generalized variance [3, 5], and the relative generalized variance [10], are variants of the generalized variance (GV).

Sample GV is geometrically easy to interpret. This is probably the most important reason why GV is very popular and widely used in many applications even though it is an imprecise measure and not apt to show the difference between two covariance structures. It is easy to find examples where, according to GV, two different covariance structures might be declared equal to each other. This imprecision is then exploited in Djauhari [11] in order to introduce the so-called vector variance (VV) as a new measure. Later on, Djauhari et al. [12] show that this new measure, if it is simultaneously used together with GV, provides better understanding about the shift in covariance structure. Despite this advantage, unlike GV, its geometric interpretation is still vague. From the classical books of multivariate analysis such as Anderson [13], we learn that sample GV is proportional to the squared volume of parallelotope spanned by a multivariate data set. See also Mason et al. [14] for a recent application in monitoring multivariate process variability. On the other hand, the geometric interpretation of sample VV needs to be clarified. This is the main topic of this paper.

The rest of the paper is organized as follows. In the next section we will discuss the important properties of VV. Later on, in Section 3, we present the geometric interpretation of its sample version. We start by using Escoufier's operator to characterize random vector and then consider VV as a metric in the space of all such operators. Later on, we show that sample VV is equal to the squared Frobenius norm of an operator associated to a random sample. This will lead us to the geometric meaning of sample VV. Additional remarks, in the last section, will close the presentation of this paper.

2 Properties of Vector Variance

Consider $\mathbf{X}$ and $\mathbf{Y}$, two random vectors of $p$and$q$ dimensions, $p$and$q$ are not necessarily equal. The covariance matrix of the random vector $\left(\mathbf{X}^T\ \mathbf{Y}^T\right)^T$ of $(p+q)$ dimension can be written in the form of a partition matrix,

$$\begin{pmatrix} \Sigma_{\mathbf{XX}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{YY}} \end{pmatrix}$$

where $\Sigma_{\mathbf{XX}}$ and $\Sigma_{\mathbf{YY}}$ are covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively, and

$$\Sigma_{\mathbf{XY}} \quad = \quad \Sigma_{\mathbf{YX}}^{\mathbf{T}}$$

is the covariance matrix between $\mathbf{X}$ and $\mathbf{Y}$.

To measure the linear relationship between the random vectors $\mathbf{X}$ and $\mathbf{Y}$, Escoufier ([15], [16]) defined the correlation coefficient between them as

$$\rho_V\left(\mathbf{X},\ \mathbf{Y}\right) \quad = \quad \frac{Tr\ \left(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}}\right)}{\sqrt{Tr\ \left(\Sigma_{\mathbf{XX}}^2\right)\ Tr\ \left(\Sigma_{\mathbf{YY}}^2\right)}}$$

where Tr is the operator "trace". See also Cleroux and Ducharme ([17]) for its inferential discussion. This measure is called "vector correlation." Accordingly, $Tr\left(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}}\right)$ is the vector covariance between $\mathbf{X}$ and $\mathbf{Y}$ and $Tr\left(\Sigma_{\mathbf{XX}}^2\right)$ and $Tr\left(\Sigma_{\mathbf{YY}}^2\right)$ the VV of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

Let us focus on the random vector $\mathbf{X}$. The VV of $\mathbf{X}$ is the sum of all diagonal elements of $\Sigma_{\mathbf{XX}}^2$. Like the GV of $\mathbf{X}$, $|\Sigma_{\mathbf{XX}}|$, it is a monotonic increasing function of all eigenvalues of$\Sigma_{\mathbf{XX}}$. If GV is the product of all eigenvalues, VV is their sum of squares. Therefore, like GV, VV can also be used as a measure of multivariate dispersion. Here are some important properties of sample VV.

Property 1: Computational Complexity

VV is equal to the sum of squares of all elements of $\Sigma_{\mathbf{XX}}$ while GV is the determinant of that matrix. Therefore, the former is a quadratic form and the latter is a multilinear form. Consequently, (i) VV can be efficiently used for high-dimension data sets and (ii) the computational complexity of VV is $O\left(p^2\right)$; far less than that of GV which is of order $O\left(p^3\right)$.

Property 2: Non Singularity of Covariance Matrix

Unlike GV, VV does not need the condition that $\Sigma_{\mathbf{XX}}$ must be non singular. Therefore, VV can be used although the distribution is degenerate in R^p.

Property 3: Average Run Length (ARL)

If VV is used to monitor process variability, VV performs much better than GV in detecting the small shift in covariance structure (Djauhari et al., 2008). Its out-of-control ARL is far less than that of GV for small shift and they have the same performance for large shift.

Property 4: Robust Estimation

Instead of minimum covariance determinant, minimum vector variance can also be used as the stopping rule in the C-step of fast minimum covariance determinant (FMCD) algorithm ([18]). As long as the algorithm is convergent, this new stopping rule is as effective as the original one but its computational complexity is far lower (see Djauhari et al., [12]).

Property 5: Asymptotic Distribution

Let$\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$ be a random sample of size n of $\mathbf{X}$ and $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^T$ be the sample mean vector and sample covariance matrix, respectively. The asymptotic distribution of sample VV is ([12]),

$$\sqrt{n-1}\left\{Tr\left(S^{\mathbf{2}}\right) - Tr\left(\Sigma^{\mathbf{2}}_{XX}\right)\right\} \xrightarrow{d} \quad N\left(0, 8Tr\left(\Sigma^{\mathbf{4}}_{\mathbf{XX}}\right)\right).$$

For practical purposes, we recommend to use

$$\sqrt{\frac{(n-1)^2}{n}}\left\{Tr\left(S^2\right) - \frac{n+1}{n-1}Tr\left(\Sigma^2_{XX}\right)\right\} \xrightarrow{d} \quad N\left(0, 8Tr\left(\Sigma^4_{\mathbf{XX}}\right)\right)$$

because the sample size n takes part in the determination of the mean and variance estimates. This distribution is simpler than the asymptotic distribution of sample GV (see Djauhari [19]),

$$\frac{1}{\sqrt{b_2}}\left\{|S| - b_1|\Sigma|\right\} \xrightarrow{d} \quad N\left(0, |\Sigma|^2\right)$$

where

$$b_1 = \frac{1}{(n-1)^p}\prod_{k=1}^{p}(n-k)$$

and

$$b_2 = b_1\left\{\frac{1}{(n-1)^p}\prod_{k=1}^{p}(n-k+2) - b_1\right\}.$$

Those properties show the significant advantages of VV compared to GV. However, in general, we cannot claim that one measure is superior to the other. It is not difficult to find examples where two different covariance matrices might have the same VV but different GV. Therefore, to have a better understanding about multivariate dispersion, the best way is to use both simultaneously.

3 Geometric Interpretation

Algebraically, GV is a "weak" measure in the sense that $|\Sigma_{\mathbf{XX}}| = 0$ if and only if at least one variable has zero variance or there is at least one variable that is a linear combination of the others. In other words, the rank of $\Sigma_{\mathbf{XX}}$ is less than p. On the other hand, VV is "strong" in the sense that $Tr\left(\Sigma^2_{\mathbf{XX}}\right) = 0$ if and only if the rank of $\Sigma_{\mathbf{XX}}$ is 0 which means that all variables have zero variance or, equivalently, the distribution is degenerate at the mean vector. This outcome is similar to that of the univariate case.

It is not difficult to give the geometric interpretation of VV in univariate case because in this case VV is the square of the classical variance. However, to the knowledge of the author, no geometric interpretation in multivariate setting is available in the literature. This is what we intent to discuss in the rest of the paper. The discussion begins by exploiting the classical result in Escoufier [16] on the characterization of a random vector. Later on, we focus on sample random setting.

3.1 Characterization of Random Vector

Let us consider $L_2(\Omega, A, P)$ the set of all centered random variables of finite variance defined on a probability space (Ω, A, P) and $\mathbf{X} = (X_1, X_2, ..., X_p)^T$ a random vector of p dimension where X_k is in $L_2(\Omega, A, P)$; $k = 1, 2, \ldots, p$. Associated to that random vector, Escoufier ([16]) defined the following operator from $L_2(\Omega, A, P)$ into itself

$$\varphi_{\mathbf{X}}(\mathbf{Y}) = \sum_{k=1}^{p} E(X_k Y) X_k \text{for all } Y \text{ in } L_2(\Omega, \mathcal{A}, P)$$

or, in matrix multiplication form,

$$\varphi_{\mathbf{X}}(\mathbf{Y}) = E(\mathbf{X}Y)\mathbf{X}.$$

It is interesting to note that if the covariance matrix of $\mathbf{X}$ is positive definite, then $\varphi_{\mathbf{X}}(\mathbf{Y})$ puts Y in the coordinate system defined byX_1, X_2, ..., X_p. The coordinate of Y on X_k is $E(X_k Y)$ the covariance of Y andX_k. The operator $\varphi_{\mathbf{X}}(\mathbf{Y})$ plays a fundamental role in multivariate analysis because it characterizes the random vector $\mathbf{X}$. In other words, there exists a one-to-one correspondence between the set of all random vectors

$$\mathcal{X} = \left\{\mathbf{X} : \mathbf{X} = (X_1, X_2, ..., X_p)^T, X_k \text{ in } L_2(\Omega, \mathcal{A}, P) \text{ for all } k\right\}$$

and the set of all operators

$$\mathcal{P} = \left\{\varphi_{\mathbf{X}} : \varphi_{\mathbf{X}}(Y) = E(\mathbf{X}Y)\mathbf{X}, \mathbf{X} \text{ in } \mathcal{X} \text{ and } Y \text{ in } L_2(\Omega, \mathcal{A}, P)\right\}.$$

Specifically, if $\mathbf{u}$ is in R^p such that $\Sigma_{\mathbf{XX}}\mathbf{u} = \lambda\mathbf{u}$, then $\mathrm{Y} = \mathbf{u}^T\mathbf{X}$ satisfies $\varphi_{\mathbf{X}}(\mathbf{Y}) = \lambda\mathrm{Y}$. Conversely, if Y is in $L_2(\Omega, \mathcal{A}, P)$ such that $\varphi_{\mathbf{X}}(\mathbf{Y}) = \lambda\mathrm{Y}$, then $Y = \mathbf{u}^T\mathbf{X}$ where $\mathbf{u}$ satisfies $\Sigma_{\mathbf{XX}}\mathbf{u} = \lambda\mathbf{u}$. The operator $\varphi_{\mathbf{X}}$ is the so-called Escoufier's operator associated with the random vector $\mathbf{X}$.

We see that: (i) $\varphi_{\mathbf{X}}$ and $\Sigma_{\mathbf{XX}}$ have the same positive eigenvalues and (ii) if $\lambda \neq 0$, Y is the principal component of $\mathbf{X}$ with varianceλ. Therefore, the characterization of $\mathbf{X}$ by $\varphi_{\mathbf{X}}$ is not only based on the magnitudes (variances) of the principal components but also their directions. This result will be exploited to explore the geometric interpretation of sample VV. First, we consider the VV as a metric in $\mathcal{P}$.

3.2 VV as a Metric

Let C_k be the principal component of $\mathbf{X}$ associated with eigenvalue λ_k; $k = 1, 2, \ldots, p.$. Then, $\mathbf{C} = (C_1, C_2, ..., C_p)^T$ and $\mathbf{X}$ have the same operators; $\varphi_{\mathbf{C}} = \varphi_{\mathbf{X}}$. Consequently, two random vectors having the same sets of principal components have the same operator. This motivates us to focus our investigation on the set of such operators. Now, let us consider the set P of all operators. We know that$L_2(\Omega, A, P)$ completed with the product scalar $< X, Y > = E(XY)$, the covariance of X and Y, is a Hilbert space. That scalar product induces in P the following scalar product

$$< \varphi_{\mathbf{X}}, \varphi_{\mathbf{Y}} > = Tr(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}})$$

for all $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{X}$. With this scalar product, $\mathcal{P}$ is also a Hilbert space. Therefore, we now know how to study the linear relationship among two elements in X via the linear relationship among their respective operators in $\mathcal{P}$. Based on the definition of $< \varphi_{\mathbf{X}}, \varphi_{\mathbf{Y}} >$, we obtain that

$$\|\varphi_{\mathbf{X}}\|^2 = < \varphi_{\mathbf{X}}, \varphi_{\mathbf{Y}} > = Tr\left(\Sigma^2_{\mathbf{XX}}\right).$$

Thus, the squared norm of $\varphi_{\mathbf{X}}$ is the squared Frobenius norm of $\Sigma_{\mathbf{XX}}$ which is the VV of $\mathbf{X}$. Now, we focus our study on the sampling form of VV.

3.3 VV as a Measure of Dispersion

Consider again $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ the random sample of size n of $\mathbf{X}$, and $\bar{\mathbf{X}}$ and $\mathbf{S}$ the sample mean vector and sample covariance matrix, respectively. We write

$$\mathbf{X}_k = (X_{1k}, X_{2k}, ..., X_{pk})^T \text{ and } \bar{\mathbf{X}} = \left(\bar{X}_1, \bar{X}_2, ..., \bar{X}_p\right)^T.$$

If we construct $\mathbf{X} = \left(X_{ij} - \bar{X}_i\right)$ a matrix of size $(p \times n)$, then the sample covariance matrix $\mathbf{S}$ can be written in matrix multiplication form as follows

$$\mathbf{S} = \frac{1}{n-1} X X^{\mathbf{T}}.$$

Accordingly, $\mathbf{W} = \frac{1}{n-1} X^{\mathbf{T}} X$ is a matrix of size $(n \times n)$ satisfying:

(i) If $\mathbf{u}$ in R^p is the solution of $\mathbf{Su} = \lambda\mathbf{u}$, then $\mathrm{C} = \mathbf{u}^T X$ satisfies $\mathbf{W}\mathrm{C} = \lambda\mathrm{C}$. Conversely, if C is such that $\mathbf{W}\mathrm{C} = \lambda\mathrm{C}$, then $\mathrm{C} = \mathbf{u}^T X$ where $\mathbf{u}$ satisfies $\mathbf{Su} = \lambda\mathbf{u}$. Thus, $\mathbf{W}$ is the sample version of Escoufier's operator associated with data matrix $\mathbf{X}$.

(ii) If $\lambda \neq 0$, then C is the principal component with varianceλ.

(iii) $$Tr\left(\mathbf{W}^2\right) = Tr\left(\frac{1}{(n-1)^2}\mathbf{X}^T\mathbf{X}\mathbf{X}^T\mathbf{X}\right) = Tr\left(\frac{1}{(n-1)^2}\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{X}^T\right) = Tr\left(\mathbf{S}^2\right).$$

The last property (iii) shows that the sample VV,$Tr\left(\mathbf{S}^2\right)$, can be interpreted by means of $Tr\left(\mathbf{W}^2\right)$. Let us consider a realization of the random sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. It forms a cloud of n observation points in R^p where the sample mean vector $\bar{\mathbf{X}}$ is the center of gravity of the cloud. From the definition of $\mathbf{W}$ we see that the geometric configuration of all points in the cloud with respect to the center of gravity is numerically summarized in that matrix. More specifically, if we write $\mathbf{W} = (W_{ij})$, then

(i) W_{ij} is the scalar product of observation points $\mathbf{X}_i$ and $\mathbf{X}_j$ with respect to $\bar{\mathbf{X}}$ divided by $(n-1)$.

(ii) W_{ii} is the squared distance from $\mathbf{X}_i$ to $\bar{\mathbf{X}}$ divided by $(n-1)$.

From this investigation we conclude that sample VV is equal to the squared Frobenius norm of $\mathbf{W}$. It is the sum of squares of all elements of $\mathbf{W}$. Therefore, the value of sample VV is a numerical indicator of how the cloud of n observation points in R^p is dispersed around the sample mean vector. The smaller the value of sample VV the more concentrated the cloud around$\bar{\mathbf{X}}$ and the larger the value of sample VV the more dispersed the cloud around$\bar{\mathbf{X}}$ in a subspace of dimension k; $k \leq p$.

4 Additional Remarks

GV and VV have different behaviours. We can easily find examples of two different matrices having the same GV but different VV. Conversely, we can also find examples of two different matrices having the same VV but different GV. Thus, algebraically, we cannot claim that GV is superior than VV or vice versa. Therefore, a simultaneous use of both will be very helpful. However, if GV needs the condition that the covariance matrix is non singular, VV does not need that condition. VV is more appropriate than GV to deal with high dimensional data sets.

Geometrically, sample GV is proportional to the squared volume of parallelotope spanned by all rows of $\mathbf{X}$. However, no geometric interpretation of sample VV is available in the literature. In this paper we have showed that sample VV is equal to $Tr\left(\mathbf{W}^2\right)$; the squared Frobenius norm of the scalar product matrix $\mathbf{W}$. By means of this statistic we see that sample VV is a numerical indicator of how the cloud of n observation points in R^p are dispersed around the sample mean vector. Sample VV equals 0 if and only if the n observation points are equal to each other. The larger the value of sample VV, the more dispersed the cloud around the sample mean vector in a subspace of R^p.

Acknowledgment

This research is supported by the Government of Malaysia under Fundamental Research Grant Scheme vote number 78484 and Universiti Teknologi Malaysia under Foreign Academic Research Grant vote number 77356. The author is grateful to the Government of Malaysia and Universiti Teknologi Malaysia for the sponsorship. Special thanks go to Professor Yves Escoufier, Laboratoire de Probabilités et Statistiques, Université de Montpellier II, France, for fruitful discussion. The author also thanks the Editor and reviewers for their efforts that led to improvements of the presentation of this paper.

References

[1] Chatterjee, S., and Hadi, A. S. *Sensitivity Analysis in Linear Regression*. New York: John Wiley & Sons. 1988.

[2] Mardia, K. V., Kent, J. T., and Bibby, J. M. *Multivariate Analysis*. New York: Academic Press. 2000.

[3] Alt, F. B., and Smith, N. D. *Multivariate Process Control. Handbook of Statistics (Krishnaiah P.R. and Rao C.R. Eds)*. New York: Elsevier Science Publishers. 7: 333–351. 1988.

[4] Montgomery, D. C. *Introduction to Statistical Quality Control*. 4th Edition. New York: John Wiley and Sons, Inc. 2001.

[5] Djauhari, M. A. Improved monitoring of multivariate process variability. *Journal of Quality Technology*. 2005. 37(1): 32 – 39.

[6] Rousseeuw, P.J. Multivariate estimation with high breakdown point. In Grossman W., Pflug G., Vincze I. and Wertz W. (Ed). *Mathematical Statistics and Applications, Volume B*. New York: D. Reidel Publishing Company. 1985 .283 – 297.

[7] Grambow, S., and Stromberg, A. J. Combining the EID and FSA for Computing the Minimum Volume Ellipsoid. Technical Report. Department of Statistics, University of Kentucky. 1998.

[8] Serfling, R. J. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons. 1980.

[9] Pena, D., and Rodriguez, J. Descriptive measures of multivariate scatter and linear dependence. 2005. Retrieved on July 22, 2009. http://halweb.uc3m.es/esp/ personal/dpena/article/JMVA03.PDF

[10] Tang, P. F., and Barnett, N. S. Dispersion control for multivariate processes. *Australian Journal of Statistics*. 1996. 38(3): 235 – 251.

[11] Djauhari, M. A. A measure of multivariate data concentration. *Journal of Applied Probability*. 2007. 2(2): 139 – 155.

[12] Djauhari, M. A., Mashuri, M., and Herwindiati, D.E. A multivariate process variability monitoring. *Communications in Statistics–Theory and Methods*. 2008. 37: 1742 – 1754.

[13] Anderson, T.W. *Introduction to Multivariate Statistical Analysis*. 1st Edition. New York: John Wiley & Sons. 1966.

[14] Mason, R. L., Chou, Y-M., and Young, J. C. Monitoring variation in a multivariate process when the dimension is large relative to the sample size. *Communications in Statistics–Theory and Methods*. 2009. 38(6): 939 – 951.

[15] Escoufier, Y. Le traitement des variables vectorielles. *Biometrics*. 1973. 29: 751 – 760.

[16] Escoufier, Y. Opérateur associé à un tableau de données. *Annales de l'INSEE*. 1976. 22-23: 342 – 346.

[17] Cleroux, R., and Ducharme, G. R. Vector correlation for elliptical distributions. *Communications in Statistics–Theory and Methods*. 1989. 18(4): 1441 – 1454.

[18] Rousseeuw, P. J., and van Driessen, K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*. 1999. 41:212 – 223.

[19] Djauhari, M. A. Asymptotic distribution of sample covariance determinant. *Matematika*. 2009. 25(1): 79 – 85.