Angelina Rajda-Tasior [*], *Grzegorz Kończak* [**]

ON THE MONITORING COMPLEX MULTIVARIATE PROCESSES

Abstract: This article presents a proposal of the method of monitoring complex multi-dimensional processes. The problem relates to monitoring the quality of production with some attribute variables when the production is performed by some operators. To describe the quality status we used the matrix in which elements are the numbers of defective units.

The proposed method uses permutation tests. The "out-of-order" signal is obtained by comparing the matrix in period t to the matrix from stable process. The test statistic used in permutation test is based on a function of distance between matrices. The properties of the proposed method have been described using computer simulation.

Keywords: multivariate processes, process monitoring, permutation tests, Monte Carlo study
JEL: C12, C14

1. INTRODUCTION AND BASIC NOTATIONS

The control charts were introduced in 1924 (Montgomery, 2009). These charts are a graphical representation of the sequence of statistical tests. Univariate control charts are used to monitor a single process characteristic. There are some multivariate control charts which could be used to monitor multiple process characteristics. Santos-Fernández (2012, p. 17) describes the following control charts for monitoring multivariate processes:

– the control ellipsoid or chi-square control chart,
– the T^2 or Hotelling chart,
– the Multivariate Exponentially Weighted Moving Average (MEWMA) chart,
– the Multivariate Cumulative Sum (MCUSUM) chart,
– the chart based on Principal Components Analysis (PCA).

All the mentioned above control charts are used to monitor product characteristics that are variables. The proposal presented in this paper deals with monitoring multidimensional attribute data.The presented solution can be used in a variety of real cases for monitoring complex multivariate processes such as:

---

[*] Ph. D Student, University of Economics in Katowice, Faculty of Management, angelina.rajda@gmail.com
[**] Associate Professor, University of Economics in Katowice, Faculty of Management, grzegorz.konczak@ue.katowice.pl

– conformity assessment of products with requirements due to the k characteristics for m positions,

– conformity assessment of products with requirements due to the k characteristics for m employees,

– conformity assessment of the product due to k characteristics for the different products of the same type (eg. different product types)

Let us assume that the product is performed by m employees and is described by k attribute characteristics. Let us assume that the observations of the monitored process could be written as a sequence of matrices $\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_n$ where $\mathbf{X}_t$ ($t = 1, 2, …, n$) is the matrix containing the elements n_{tij} where n_{tij} is the number of discrepancies in time t for the j-th characteristic ($j = 1, 2, …, k$) and for i-th ($i = 1, 2, …, m$) place, position or operator. For example n_{tij} – is a number of j-th ($j = 1, 2, …,k$) type of defects for i-th ($i = 1, 2, …, m$) person in t-th sample, where k is the number of types of defects and m is the number of operators.

Let us assume that for the time $t = 1, 2, …, n$ the process was stable and "in-control". There are given matrices $\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_n$ with data from this process. Let $N_t = [N_{t1}, N_{t2}, …, N_{tm}]$ be a vector of numbers of controlled parts by places (positions or operators).

Table 1. The number of defects ($\mathbf{X}_t$ matrix)and the number of controlled elements

Position i	The type of defects j				The number of elements
	1	2		K	
1	n_{t11}	n_{t12}		n_{t1k}	N_{t1}
2	n_{t21}	n_{t22}		n_{t2k}	N_{t2}
m	n_{tm1}	n_{tm2}		n_{tmk}	N_{tm}

Source: own elaboration.

Then, based on the data from matrix $\mathbf{X}_t$ (see Table 1) let

$$\mathbf{Y}_t = \begin{bmatrix} \hat{p}_{t11} & \hat{p}_{t12} & & \hat{p}_{t1k} \\ \hat{p}_{t21} & \hat{p}_{t22} & & \hat{p}_{t2k} \\ & & & \\ \hat{p}_{tm1} & \hat{p}_{tm2} & & \hat{p}_{tmk} \end{bmatrix} \qquad (1)$$

where $\hat{p}_{tij} = \dfrac{n_{tij}}{N_{ti}}$. Therefore, the process at time t can be described by the matrix $\mathbf{Y}_t$. Based on the data obtained from the stable process which is described by matrices $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ it should be assessed whether the analyzed process is "out-of-order".

When considering information of the total number of nonconformities in a given unit of time t (sum of all observations in the matrix $\mathbf{X}_t$) the classical control chart c could be used.This method allows to observe the total number of nonconformities in the unit or the set of the products (Kończak 2007). The total number of nonconformities can be determined by the following formula:

$$c_t = \sum_{i=1}^{m} \sum_{j=1}^{k} n_{tij}$$

The use of the control chart c will not track changes in the structure of discrepancies but only their total number.The proposed method refers to the permutation test and avoids the inconvenience. A permutation test will be applied to evaluate the stability of the process By a permutation test the form of the test statistics should be determined. The test statistic will be defined as a function of the distance between the matrix which represents the stable process, and the matrix from the monitored process.

2. PERMUTATION TESTS – THE IDEA AND DISTANCE APPROACH

Permutation tests were introduced by R.A. Fisher and E.J.G. Pitman in 1930's (see Berry at al. 2014). In permutation tests the observed value of the test statistic is compared with the empirical distribution of this statistics under the null hypothesis. Lehmann (2009) shows that permutation tests are generally asymptotically as good as the best parametric ones. The concept of permutation tests is simpler than the concept of tests based on normal distribution. The main application of these tests is a two-sample problem (Efron, Tibshirani1993). The following steps should be taken when dealing with permutation tests Good 2005):

1. Identify the null hypothesis and the alternative hypothesis.

2. Choose a test statistic T.

3. Compute the value T_0 of the test statistic for the original data.

4. Determine by the series of permutationsthe frequency distribution of the test statistic under the null hypothesis ($T_1, T_2, \ldots, T_N$, where $N \geq 1000$).

5. Make a decision using this empirical distribution as a guide.

The ASL (*Achieved Significance Level*) has the following form:

$$ASL = P\left(T \geq T_0\right). \tag{2}$$

The ASL is unknown and could be estimated by the following formula

$$ASL \approx \frac{card\left(i : T_i \geq T_0\right)}{N}. \tag{3}$$

This notation applies, where the H_0 rejection area is right-sided. In the case of the left-sided rejection area in the above notation inequality should be changed. If the value of *ASL* is lower than the assumed level of significance α, then H_0 will be rejected.

A typical problem considered with permutation test is comparing two populations (Kończak 2012). Let S_1 and S_2 be two samples of the sizes n_1 and n_2. These hypotheses have the following forms:

$$H_0 : m_1 = m_2$$

$$H_1 : m_1 \neq m_2$$

A typical test statistic for comparing means in two populations has the following form:

$$T = \overline{X}_1 - \overline{X}_2 \tag{4}$$

The high values or the small (negative) values are against the hypothesis H_0. The T statistic deals with univariate variable. It is not possible to use the test statistic (4) for the multivariate comparison. Therefore, the distance approach will be applied.

Mielke and Berry (2007) considered a completely different approach than the described above in determining the form of the test statistics. In this approach the distance between the elements in different groups is considered. For all the possible divisions of the data set should be calculated the sum of distances for all pairs of elements in each group, and then the weighted sum of the distance with weights n_1 and n_2 is determined. In this approach, the aim is to obtain the most homogeneous groups.

The distance between the element r and s will be denoted by Δ_{rs}, which for the case of one-dimensional data may be written as follows:

$$\Delta_{rs} = |x_r - x_s| \tag{5}$$

Then the average distance between all possible pairs of objects should be calculated separately in each group: δ_1 and δ_2. The last step is calculating a weighted average of both distances:

$$\delta = \frac{n_1}{n_1 + n_2}\delta_1 + \frac{n_2}{n_1 + n_2}\delta_2. \tag{6}$$

The function (6) will be taken as the test statistic for the permutation test. As the purpose is to obtain the two most homogeneous clusters, small value of this statistics will be against the hypothesis H_0. In this case the rejection area is left sided.

The distance Δ_{rs} between observations x_r and x_s may be defined in a completely different way from the above. This approach is particularly interesting, if verified hypothesis refers to the comparison of multidimensional population distributions. This measure the distance between the points $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in R^n space we can use a variety of distance functions such as (see Kuratowski, 1980, p. 103, Walesiak, 2011, p. 24):

– Euclidean (E) $\delta(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$

– Manhattan (M) $\delta(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n}|x_i - y_i|$

– Chebyshev (Ch) $\delta(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i|$

The list of other potential distance functions can be find in Deza, Deza (2013).

3. THE TEST PROCEDURE – MONTE CARLO STUDY

Let us assume that the matrices $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ are determined on the basis of observations from the "in-control" process. In the performed computer simulations it was assumed $n = 5$. For the next two periods a number of nonconformities was generated and on that basis matrices $\mathbf{Y}_{n+1}$ and $\mathbf{Y}_{n+2}$ were obtained. The hypothesis H_0 says that the process for monitored periods was "in-control" (the probabilities of discrepancies in the structure of the matrix $\mathbf{Y}_t$ for $t > n$ was the same as in matrix for periods from 1 to n).The hypothesis H_1 says that the matrices $\mathbf{Y}_t (t > n)$ represent the "out-of-order" process.

The aim of the simulations was to describe the characteristics of the proposed method of monitoring multivariate processes with attribute data. In the Monte Carlo study, the quality of elements was assessed at m positions (in the simulations it was assumed that the number of stations $m = 6$). Each component was assessed of the k characteristics (in the simulations it was assumed $k = 5$) and assessed conformity with the assumptions for all products. The vector of probabilities of occurrence of nonconformities on the i-th position will be denoted by $\mathbf{p}_i$ ($i = 1, 2, ..., m$). For each position samples of the size N_i were taken. For the stable process probability of occurrence of the j-th defect on the i-th position was denoted as p_{ij}. According to the description, we generated matrices with the number of nonconformities for the j-th characteristics.

In computer simulations the probabilities of occurrence of the "out-of-order" signal for non-stable processes were estimated. The parameters which are used in the simulations are presented in Table 2.

Table 2. The probabilities of occurrence the j-th defect on the i-th position

Position i	Type of defects j					Number of elements
	A	B	C	D	E	
1	0.01	0.02	0.02	0.03	0.03	100
2	0.02	0.02	0.03	0.03	0.03	200
3	0.02	0.02	0.04	0.04	0.06	300
4	0.02	0.02	0.04	0.04	0.06	400
5	0.03	0.03	0.05	0.04	0.06	400
6	0.03	0.04	0.05	0.05	0.07	500

Source: own elaboration.

In the simulation analysis the characteristics of the proposed method and the results which were obtained by applying the control chart c were compared. An example of the matrix $\mathbf{X}$ which represents the number of j-th nonconformities on the i-th position is as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 2 & 2 & 4 \\ 2 & 5 & 8 & 3 & 3 \\ 3 & 8 & 15 & 13 & 20 \\ 3 & 8 & 12 & 8 & 19 \\ 13 & 12 & 20 & 11 & 22 \\ 17 & 21 & 31 & 26 & 41 \end{bmatrix}$$

In the simulation analysis the probability of occurrence of an "out-of-order" signal by using a permutation test and control chart c was assessed. For this purpose $n = 5$ matrices from "in-order" process were generated, and then the two tables with the same distribution (for the assessment of the probability of occurrence of a signal for the regular process) or from the distributions of the higher probability of nonconformities (for assessment of the probability of occurrence a signal for the deregulated process) and a permutation test was carried out on the basis of the test statistic δ by the formula (6).

The following variants were considered in the Monte Carlo study:

0) without changing parameters (the "in-order" process)

a) a double increase in the probabilities of occurrence of defect C type at position no. 2.

b) 50% increase in probability of occurrence of all type of nonconformities at position no. 3.

c) 20% increase in probability of occurrence of nonconformities type A at every position.

d) 20% increase in probabilities of the occurrence of all nonconformities at position no 3 with reducing the probability of the occurrence nonconformities by 20% at position no 4.

e) 50% increase in probabilities of the occurrence of all nonconformities at position no 3 with reducing the probability of occurrence of nonconformities by 50% at position no 4.

The first case enables us to estimate the size of the permutation test (see Domański et al., p. 17, 2014). It was assumed that the level of significance $\alpha = 0,05$. The permutation test procedure that was used for testing the hypothesis has the following steps:

1. The matrices $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ from stable process (see Table 2) were generated.

2. The matrices $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ were obtained (see 1)

3. The probabilities p_{ij} were estimated on the basis of $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$.

4. The matrices $\mathbf{Y}_n$ and $\mathbf{Y}_{n+1}$ were generated.

5. The values δ_1 (mean distance of matrices from stable process) and δ_2 (the distance between of matrices from monitored process) were calculated.

6. The value δ_0 of the test statistics δ for the sample data was calculated.

7. The matrices $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n, \mathbf{Y}_{n+1}, \mathbf{Y}_{n+2}$ were permuted $N = 1000$ times and the values of the statistics δ were obtained.

8. On the basis of empirical distribution of the test statistic δ, the ASL value is obtained.

If $ASL < \alpha$, then H_0 is rejected, otherwise H_0 hypothesis cannot be rejected.

To estimate the probabilities in each case there were performed 1000 repetitions of the simulation. The results of the assessment of the probability of the "out-of-order" signal occurrence for three considered distance functions and control chart c are shown in Table 3.

Table 3. The probabilities of occurrence of the "out-of-order" signal for the permutation test
and with using control chart c

Model	Control chart c	Permutation test with distance statistic		
		E	M	Ch
0)	0.055	0.043	0.081	0.082
a)	0.077	0.086	0.115	0.162
b)	0.319	0.284	0.360	0.448
c)	0.099	0.077	0.101	0.120
d)	0.053	0.121	0.144	0.221
e)	0.060	0.477	0.528	0.899

Source: own elaboration.

The results of the simulation study are presented in Table 3 and also in Figure 1. When the process is stable the results from simulations for all the considered methods were similar. A little too often we have got a signal about deregulation of process with using the statistics based on Manhattan and Chebyshev distance formula.

In the analyzed cases of the deregulated processes much more effective than the control chart c was the use of the proposed method based on permutation test with statistic based on Euclidean or Chebyshev distance formula. The best results were obtained using a Chebyshev distance formula. The use of this form of statistics allows for a fast detection of deterioration even at a single position.
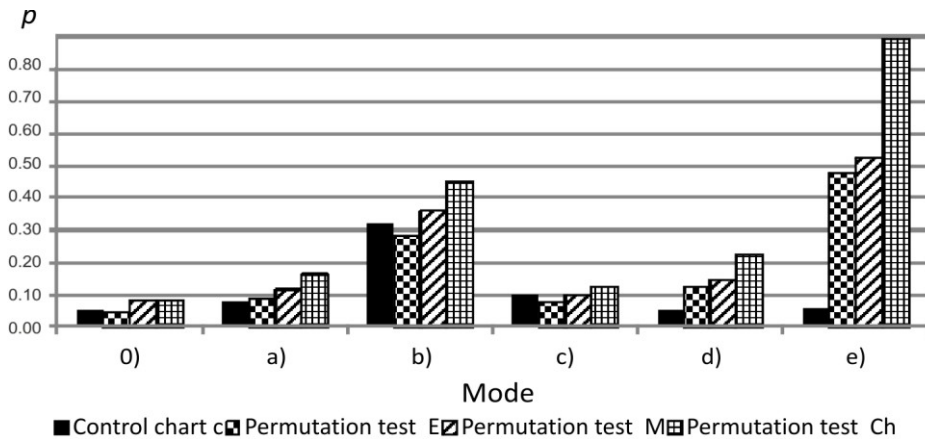

Figure 1. The probabilities of occurrence of the "out-of-order" signal for the permutation test
and with using control chart c

Source: own elaboration based on Table 3.

4. CONCLUDING REMARKS

This article presents a proposal of the method of monitoring complex multi-dimensional processes with the attribute characteristics.In order to compare multivariate processes the permutation tests with statistics based on distances between matrices was used. The efficiency of detection deregulation of the process for the proposed method and the control chart c were compared in Monte Carlo study. For control chart c the total number of nonconformities was determined.The proposed method can be used by managers who are interested in multipack results of process monitoring.When the deregulation of the process is detected it is necessary to carry out the decomposition of the "out-of-order" signal to identify the variable which was the cause of deregulation of the process.

REFERENCES

Berry K.J., Johnston J.E., Mielke Jr. P.W. (2014), *A Chronicle of Permutation Statistical Methods*, Springer International Publishing, New York.

Deza M.M., Deza E. (2013), *Encyclopedia of Distances*, Springer Verlag, Berlin.

Domański Cz., Pekasiewicz D., Baszczyńska A., Witaszczyk A. (2014), *Testy statystyczne w procesie podejmowania decyzji*, Wydawnictwo Uniwersytetu Łódzkiego.

Efron B., Tibshirani R. (1993), *An Introduction to the Bootstrap*, Chapman & Hall. New York.

Good P. (2005), *Permutation, Parametric and Bootstrap Tests of Hypotheses*, Science Business Media, Inc.

Kończak G. (2007), *Metody statystyczne w sterowaniu jakością produkcji*, Akademia Ekonomiczna w Katowicach.

Kończak G. (2012), *Wprowadzenie do symulacji komputerowych*, Uniwersytet Ekonomiczny w Katowicach.

Kuratowski K. (1980), *Wstęp do teorii mnogości i topologii*, PWN Warszawa.

Lehmann E.L. (2009), *Parametric vs. nonparametric: Two alternative methodologies*, "Journal of Nonparametric Statistics", vol. 21, pp. 397–405.

Mielke P.W., Berry K.J. Jr. (2007), *Permutation Methods. A Distance Function Approach*. Springer Science+Business Media, LLC. New York.

Montgomery D.C. (2009), *Introduction to Statistical Quality Control*, John Wiley & Sons, Inc., Arizona State University.

Santos-Fernández E. (2012), *Multivariate Statistical Quality Control Using R*, Springer Science+Business Media, LLC., New York.

Walesiak M. (2011), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.

Angelina Rajda-Tasior, *Grzegorz Kończak*

O MONITOROWANIU ZŁOŻONYCH WIELOWYMIAROWYCH PROCESÓW

Streszczenie. W artykule przedstawiono propozycję metody monitorowania złożonych wielowymiarowych procesów produkcyjnych. Rozważany problem dotyczy monitorowania jakości produkcji przy stosowaniu oceny alternatywnej jednocześnie względem wielu charakterystyk, gdy produkcja wykonywana jest na wielu różnych stanowiskach. Do opisu stanu jakości w czasie t wykorzystywana jest macierz, w której elementami są liczby wadliwych sztuk wykonanych na danym stanowiskuwedług ocenianych wielu charakterystyk.

Proponowana metoda odwołuje się do testu permutacyjnego. Sygnał o nieprawidłowym przebiegu produkcji jest uzyskiwany na podstawie porównania macierzy z bieżącego okresu dla monitorowanego procesu oraz macierzy danych uzyskanej z procesu ustabilizowanego. Ze względu na dużą liczbę charakterystyk rejestrowanych na skali porządkowej konstrukcja statystyki testowej została oparta o funkcję odległości macierzy. Własności proponowanej metody zostały poddane analizie z wykorzystaniem symulacji komputerowych. Przeprowadzono również porównania wyników uzyskanych z zastosowaniem proponowanej metody i karty kontrolnej c.

Słowa kluczowe: wielowymiarowe procesy, monitorowanie procesów, testy permutacyjne, symulacje komputerowe.