ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 206, 2007

*Grzegorz Kończak**

JOINT MONITORING OF THE MEAN, DISPERSION AND ASYMMETRY IN PROCESS CONTROL

Abstract. In this paper the quadratic form of sample mean, sample variance and sample asymmetry is considered. This quadratic form can be used for testing hypothesis on the expected value, variance and asymmetry of observed variable. This statistic can be used for constructing a control chart for monitoring these three parameters in process control or in acceptance sampling by variables.

It is very difficult to find the exact distribution of the proposed statistic. The asymptotic distribution of the proposed statistic is presented. The quantiles of exact distributions have been derived using Monte Carlo study for the case of normal distribution of the monitored variable.

Key words: test, mean, variance, asymmetry, distribution.

1. INTRODUCTION

The control charts are developed by W. A. Shewhart and were first used to monitor a process mean. The charts as $\overline{X}-S$ and $\overline{X}-R$ were used to monitor the process mean and its variance in two separated charts. There were some propositions for joint monitoring of these two parameters. F. F. Gan (2000) considered Hotelling T type control chart and F. F. Gan (1995) considered joint monitoring of sample mean and sample variance using exponentially weighted moving average control charts. G. Kończak and J. Wywiał (2001) proposed a modification of control chart $\overline{X}-S$. They added to the standard chart a third chart to monitor asymmetry.

* Ph.D., Department of Statistics, University of Economica Katowice, e-mail: koncz-@ae.katowice.pl.

The quadratic form of the sample mean, sample variance and sample asymmetry is considered in this paper. This quadratic form can be applied as the test statistic for the hypothesis on expected value, variance and asymmetry of normal distribution. The distribution function of this statistic leads to χ^2 distribution with 3 degrees of freedom, if sample size leads to infinity. The critical values for this statistic in the case of small size sample were found using Monte Carlo method. The table with approximated critical values of the test statistic has been prepared for three standard significance levels. Using these tables the hypothesis on expected value, variance and asymmetry of a diagnostic variable can be tested when sample size is greater or equal to 4.

2. BASIC DEFINITIONS

Let X be the random variable which has moments of at least 6^{th} order. Let us consider the random variable vector

$$\mathbf{X} = \begin{bmatrix} X \\ (X - \mathrm{E}X)^2 \\ (X - \mathrm{E}X)^3 \end{bmatrix} \tag{1}$$

The first component of the vector **X** measures the mean of the random variable X, the second one dispersion and the third one asymmetry. Let us denote the expected value of the random variable X by

$$\mathrm{E}(X) = \mu \tag{2}$$

and the r-th central moment of the variable X

$$\mathrm{E}(X - \mathrm{E}X)^r = \eta_r \tag{3}$$

The second central moment is denoted by

$$\mathrm{E}(X - \mathrm{E}X)^2 = \eta_2 = \sigma^2 \tag{4}$$

On the basis of expressions (2)–(4) we have

$$\mathbf{E(X)} = \begin{bmatrix} \mathrm{E}X \\ \mathrm{E}(X - \mathrm{E}X)^2 \\ \mathrm{E}(X - \mathrm{E}X)^3 \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma^2 \\ \eta_3 \end{bmatrix} \tag{5}$$

The covariance matrix of this vector is as follows

$$\Sigma(\mathrm{X}) = \begin{bmatrix} \eta_2 & \eta_3 & \eta_4 \\ \eta_3 & \eta_4 - \eta_2^4 & \eta_5 - \eta_2\eta_3 \\ \eta_4 & \eta_5 - \eta_2\eta_3 & \eta_6 - \eta_3^2 \end{bmatrix} \tag{6}$$

Let $X_1, X_2, ..., X_n$ be the simple sample from the population. Basing on this sample we would like to test a hypothesis

$$H_0: \begin{bmatrix} \mathrm{E}X \\ \mathrm{E}(X - \mathrm{E}X)^2 \\ \mathrm{E}(X - \mathrm{E}X)^3 \end{bmatrix} = \begin{bmatrix} \mu_0 \\ \sigma_0^2 \\ \eta_{3,0} \end{bmatrix} \tag{7}$$

where the alternative hypothesis is as follows

$$H_1: \sim H_0.$$

Let us consider the following vector $\mathbf{V}_n$

$$\mathbf{V}_n = \begin{bmatrix} \overline{X} \\ S^2 \\ C_3 \end{bmatrix} \tag{8}$$

where $\overline{X}$ is a sample mean given by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{9}$$

S^2 is a sample variance

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{10}$$

C_3 is the sample third central moment

$$C_3 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^3 \tag{11}$$

In the case of small size of the sample we can use following variance and third moment estimators

$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{12}$$

$$\hat{C}_3 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(X_i - \overline{X})^3 \tag{13}$$

To find the covariance matrix of vector $\mathbf{V}_n$ we use the formulas from H. Cramer (1958)

$$\operatorname{cov}(\overline{X}, C_k(x)) = \frac{1}{n}(\eta_{k+1} - k\eta_2\eta_{k-1}) + O(n^{-2}) \tag{14}$$

$$\operatorname{cov}(C_k(x), C_k(x)) = \frac{1}{n}(\eta_{k+s} - \eta_k\eta_s + ks\eta_2\eta_{k-1}\eta_{s-1} - k\eta_{k-1}\eta_{s+1} - + s\eta_{k+1}\eta_{s-1}) + O(n^{-2}) \tag{15}$$

where C_r is the r-th sample central moment given by

$$C_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^r.$$

The covariance matrix of the vector $\mathbf{V}_n$ takes form

$$\mathbf{\Sigma}_n = \frac{1}{n}\begin{bmatrix} \eta_2 & \eta_3 & \eta_4 - 3\eta_2^2 \\ \eta_3 & \eta_4 - \eta_2^2 & \eta_5 - 4\eta_2\eta_3 \\ \eta_4 - 3\eta_2^2 & \eta_5 - 4\eta_2\eta_3 & \eta_6 - 6\eta_2\eta_4 + 9\eta_2^3 - \eta_3^2 \end{bmatrix} \tag{16}$$

If the parameters μ and σ of the random variable X are unknown the covariance matrix (16) can be estimated using the following estimator of $\mathbf{\Sigma}_n$:

$$\hat{\mathbf{\Sigma}}_n = \frac{1}{n}\begin{bmatrix} C_2 & C_3 & C_4 - 3C_2^2 \\ C_3 & C_4 - C_2^2 & C_5 - 4C_2C_3 \\ C_4 - 3S^4 & C_5 - 4C_2C_3 & C_6 - 6C_2C_4 + 9C_2^3 - C_3^2 \end{bmatrix} \tag{17}$$

To test the hypothesis (7) we can use the Wald type statistic Z_n

$$Z_n = [\overline{X} - \mu, S_n^2 - \sigma^2, C_3 - \mu_3]\, \Sigma_n^{-1} \begin{bmatrix} \overline{X} - \mu \\ S_n^2 - \sigma^2 \\ C_3 - \mu_3 \end{bmatrix} \tag{18}$$

This statistic has asymptotic (if n leads to infinity) chi square distribution with 3 degrees of freedom. If the parameters of the random variables X are unknown we can construct the test statistic Z_n^* based on the estimator of the covariance matrix

$$Z_n^* = [\overline{X} - \mu, S_n^2 - \sigma^2, C_3 - \mu_3]\, \hat{\Sigma}_n^{-1} \begin{bmatrix} \overline{X} - \mu \\ S_n^2 - \sigma^2 \\ C_3 - \mu_3 \end{bmatrix} \tag{19}$$

If n leads to infinity the distribution leads to T^2 Hotelling distribution (e.g. Fuchs, Kenett, 1998)

$$Z_n^* \underset{n \to \infty}{\rightarrow} T^2.$$

Using the following dependecy

$$T^2 \sim \frac{fp}{f - p + 1} F_{p, f-p+1} \tag{20}$$

where $f = n - 1$ and $n - 1 > p$, the critical values for the statistics (19) can be found.

3. THE CASE OF THE NORMAL DISTRIBUTION

Let us assume that the simple sample is taken from the normal distribution with the mean μ and the standard deviation σ. We can write this as follows $X\colon \mathrm{N}(\mu, \sigma^2)$. In this case the expected value and the covariance matrix of the vector $\mathbf{V}_n$ statistic are following:

$$\mathrm{E}(\mathbf{V}_n) = [\mu, \sigma^2, 0]^{\mathrm{T}} \tag{21}$$

and

$$\boldsymbol{\Sigma}_n = \frac{1}{n}\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & 2\sigma^4 & 0 \\ 0 & 0 & 15\sigma^6 \end{bmatrix} \tag{22}$$

The inverse matrix of the covariance matrix is following:

$$\boldsymbol{\Sigma}_n^{-1} = \frac{n}{30\sigma^6}\begin{bmatrix} 30\sigma^4 & 0 & 0 \\ 0 & 15\sigma^2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{23}$$

The statistic (18) takes form

$$Z_n = \frac{n}{30\sigma^6}\,[30\sigma^4(\overline{X} - \mu)^2 + 15\sigma^2(S_n^2 - \sigma^2)^2 + 2C_3^2] \tag{24}$$

or equally it can be written

$$Z_n = n\left[\frac{30(\overline{X} - \mu)^2}{30\sigma^2} + \frac{15(S_n^2 - \sigma^2)^2}{30\sigma^4} + \frac{2C_3^2}{30\sigma^6}\right] \tag{25}$$

This statistic has assymptotic *chi* square distribution with 3 degrees of freedom (eg. Fuchs, Kenett, 1998).

4. MONTE CARLO STUDY

Because it is very difficult to find the exact distribution of the statistic Z_n there were made Monte Carlo study. For sample sizes of $n = 4, 5, \ldots, 20$ there were found quantiles of this statistic. The quantiles were found for three significance levels 0.10, 0.05 and 0.01. The run of computer simulations was following:

1) there were generated values of random variables $X_1, X_2, \ldots, X_n$ from normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 5$,

2) for each n from 4 to 20 step 1 was repeated 10 000 times. The values of the statistic Z_n were calculated,

3) the quantiles 0.90, 0.95 and 0.99 from empirical distribution were accepted as estimates of quantiles of statistic Z_n.

The results of Monte Carlo study are presented in Tab. 1. The Fig. 1 presents the quantiles of the statistics Z_n for the significance levels 0.1, 0.05 and 0.01. These quantiles were estimated for sample sizes from 4 to 20.

Table 1

Quantiles of the statistic Z_n – the results of Monte Carlo study

Sample size	Quantil $q(1-\alpha)$		
n	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
4	4.25	5.85	10.97
5	4.29	5.78	11.25
6	4.39	6.07	11.92
7	4.43	6.19	11.75
8	4.53	6.28	11.49
9	4.59	6.36	11.54
10	4.67	6.41	11.42
11	4.84	6.52	11.66
12	4.77	6.39	11.26
13	4.78	6.51	11.52
14	4.8	6.36	11.05
15	4.69	6.41	10.76
16	4.93	6.51	11.20
17	4.83	6.32	11.69
18	4.94	6.74	11.17
19	4.93	6.43	11.01
20	4.87	6.37	10.85

Source: own calculations.

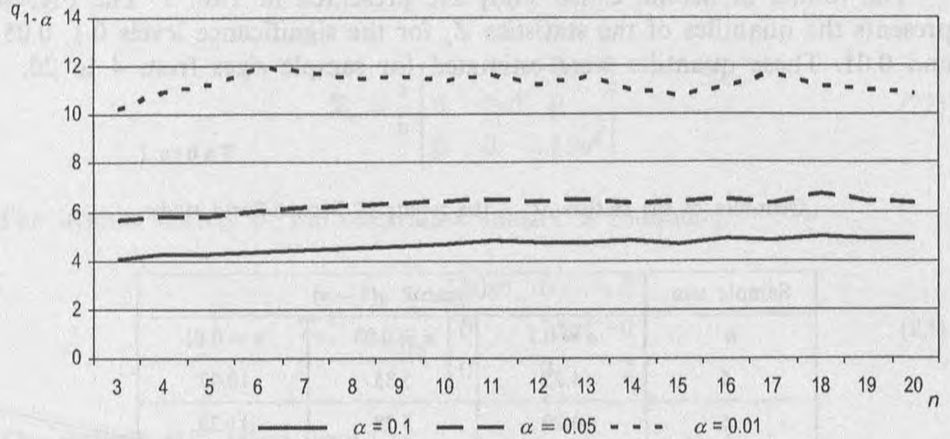

Fig. 1. Quantiles of the statistic Z_n

Source: own calculations based on Tab. 1.

5. APPLICATIONS

Acceptance sampling is commonly used by manufacturers to screen incoming lots of material for an excessive number of defective units. There are two types of acceptance sampling plans: for attributes and for variables. The proposed statistic can be used in acceptance sampling plans by variables. Let us assume that there is a mixture of two subsamples in the incoming lot. Let the characteristic of the first subsample has normal distribution with mean μ_1 and standard deviation σ_1 and the second subsample with mean μ_2 and standard deviation σ_2. We can then write that $X_1 : \mathrm{N}(\mu_1, \sigma_1)$ and $X_2 : \mathrm{N}(\mu_2, \sigma_2)$. Let the fraction of units from the first subsample be denoted by α and the fraction of units from the second subsample by $1 - \alpha$. Let $f_1(x)$ and $f_2(x)$ are the density functions of random variables X_1 and X_2. The density functions of the mixture of random variables will be then as follows:

$$f(x) = \alpha f_1(x) + (1 - \alpha)f_2(x) \tag{26}$$

The expected value of this mixture will be given by

$$\mathrm{E}(X) = \alpha\mu_1 + (1 - \alpha)\mu_2 \tag{27}$$

and the variance

$$D^2(X) = \alpha\sigma_1^2 + (1-\alpha)\sigma_2^2 + \alpha(1-\alpha)(\mu_1 - \mu_2)^2 \quad (28)$$

The third central moment of the random variable X is following:

$$C_3(X) = \alpha(1-\alpha)(\mu_2 - \mu_1)[(2\alpha - 1)(\mu_2 - \mu_1)^2 + 3(\sigma_2^2 - \sigma_1^2)] \quad (29)$$

The Fig. 2 presents the density functions of random variables X_1, X_2 and their mixtures in the case where $\mu_1 < \mu_2$ and $\sigma_1 > \sigma_2$. Under these assumptions and if $\alpha = 0.5$ we have $C_3(x) < 0$. The third central moment will inform us that the incoming lot isn't homogeneous and the hypothesis (7) may be rejected.

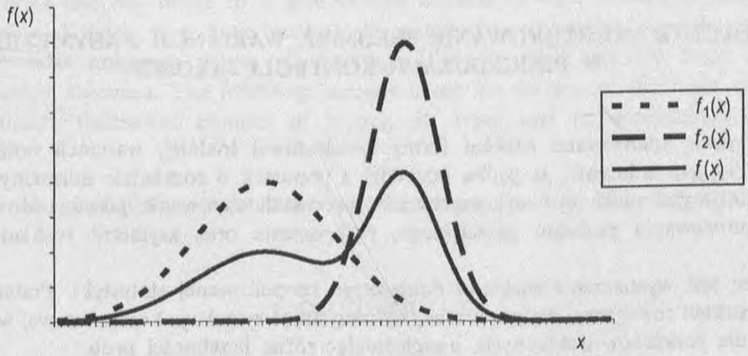

Fig. 2. The density function of the mixture of two random variables

Source: own calculations.

6. SUMMARY

The proposed statistic Z_n we can use to test the hypothesis on expected value, variance and asymmetry of random variable. We can use this test in quality control procedures especially in acceptance sampling. When we have a sample which is a mixture of some subsamples with different means and standard deviations it is possible that mean is equal to μ and variance is equal to σ^2. In such situations the third central moment can give information to reject the sample.

REFERENCES

Cramer H. (1958), *Metody matematyczne w statystyce*, PWN, Warszawa.

Fuchs C., Kenett R. S. (1998), *Multivariate Quality Control*, Marcel Dekker, New York–Basel–Hong Kong.

Gan F. F. (1995), *Joint Monitoring of Process Mean and Variance Using Exponentially Weighted Moving Average Control Charts*, "Technometrics", 37, 446–453.

Gan F. F. (2000), *Joint Monitoring of Process Mean and Variance Based on the Exponentially Weighted Moving Averages*, [in:] *Statistical Process Monitoring and Optimization*, Marcel Dekker, New York–Basel, 189–208.

Kończak G., Wywiał J. (2001), *O pewnej modyfikacji karty kontrolnej X-S*, [in:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych – teoria i zastosowania*, "Taksonomia", 8, (AE, Wrocław), 140–148.

Grzegorz Kończak

ŁĄCZNE MONITOROWANIE ŚREDNIEJ, WARIANCJI I ASYMETRII W PROCEDURACH KONTROLI JAKOŚCI

W artykule analizowano rozkład formy kwadratowej średniej, wariancji oraz asymetrii z próby. Przyjęto założenie, że próba pochodzi z populacji o rozkładzie normalnym. Proponowana statystyka może być wykorzystana w procesach sterowania jakością do jednoczesnego monitorowania poziomu przeciętnego, rozproszenia oraz asymetrii rozkładu badanej zmiennej.

Trudne jest wyznaczenie rozkładu dokładnego proponowanej statystyki. Podano asymptotyczny rozkład rozważanej zmiennej oraz, wykorzystując symulacje komputerowe, wyznaczono kwantyle dla rozkładów dokładnych, uwzględniając różne liczebności prób.