ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 196, 2006

*Janusz Wywiał**, *Grzegorz Kończak***

ON THE DISTRIBUTION OF QUADRATIC FORM OF SAMPLE MEAN AND VARIANCE

Abstract. The quadratic form of the sample mean and sample variance was considered. The sample is from normal distribution. The density function of the quadratic form has been derived. The quadratic form can be applied as the test statistic for the hypothesis on expected value and variance of normal distribution. The table with approximated critical values of the test statistic has been derived.

Key words: quadratic form, mean, variance, distribution.

1. INTRODUCTION

The quadratic form of the sample mean and sample variance was considered. The density function of the quadratic form has been derived for the case the sample is taken from normal distribution. The quadratic form can be applied as the test statistic for the hypothesis on expected value and variance of normal distribution. It is very difficult to find exactly distribution function for this statistic. The critical values for this statistic were found using numerical integration. The table with approximated critical values of the test statistic has been prepared for three standard significance levels. The distribution function of this statistic leads to χ^2 distribution with 2 degrees of freedom if sample size leads to infinity. Using these tables we can tests hypothesis on expected value and variance of a diagnostic variable when sample size is greater or equal to 3.

* Professor, Department of Statistics, The Karol Adamiecki University of Economics, Katowice.

** Ph.D., Department of Statistics, The Karol Adamiecki University of Economics, Katowice.

2. BASIC DEFINITIONS

Let X_1, X_2, ..., X_n be the simple sample from a distribution which has moments of at least four-order. Let us consider the following statistics Z_n:

$$Z_n = [\overline{X}_n, \hat{S}_n^2] \tag{1}$$

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

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2}$$

and $\hat{S}_n^2$ is a sample variance

$$\hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2. \tag{3}$$

The parameters of these statistics are as follows:

$$E(\overline{X}_n) = \mu, \quad D^2(\overline{X}_n) = \frac{\sigma^2}{n}, \tag{4}$$

$$E(\hat{S}_n^2) = \sigma^2, \quad D^2(\hat{S}_n^2) = \frac{\eta_4 - \eta_2^2}{n} + O(n^{-2}), \tag{5}$$

$$\mathrm{Cov}(\overline{X}_n, \hat{S}_n^2) = \frac{\eta_3}{n} + O(n^{-2}). \tag{6}$$

The well-know theorem of Cramèr (1958) leads to the following. If $n \to \infty$ then $Z_n \to Z \sim N(E(Z), \Sigma(Z))$ where:

$$E(Z) = [\mu, \sigma^2] \tag{7}$$

and the covariance matrix:

$$\Sigma(Z) = \frac{1}{n}\begin{bmatrix} \sigma^2 & \eta_3 \\ \eta_3 & \eta_4 - \sigma^4 \end{bmatrix}. \tag{8}$$

The consistent estimator of the covariance matrix $\Sigma(Z)$ is as follows:

$$\Sigma_n = \frac{1}{n}\begin{bmatrix} \hat{S}_n^2 & C_{3n} \\ C_{3n} & C_{4n} - \hat{S}_n^2 \end{bmatrix}. \tag{9}$$

Let us consider the following Wald-type statistic Q_n

$$Q_n = [\overline{X} - \mu, S_n^2 - \sigma^2]\Sigma_n^{-1}\begin{bmatrix} \overline{X} - \mu \\ S_n^2 - \sigma^2 \end{bmatrix}. \tag{10}$$

If $n \to \infty$ then Q_n leads to chi square distribution with 2 degrees of freedom.

3. THE CASE OF THE NORMAL DISTRIBUTION

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

$$E(Z_n) = [\mu, \sigma^2] \tag{11}$$

and

$$\Sigma(Z_n) = \frac{1}{n}\begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \tag{12}$$

The statistic Q_n (10) can be rewritten in the following way:

$$Q_n = Q_{1n} + Q_{2n}, \tag{13}$$

where

$$Q_{1n} = n\frac{(\overline{X}_n - \mu)^2}{\sigma^2} \tag{14}$$

$$Q_{2n} = n\frac{(\hat{S}_n^2 - \sigma^2)^2}{2\sigma^4}. \tag{15}$$

B.U.Ł.

It is well known that the random variable Q_{1n} has χ_1^2 distribution. The well known Geary (1936) theorem leads to conclusion that the statistics Q_{1n} and Q_{2n} are independent.

The statistic Q_{2n} can be rewritten in the following way:

$$Q_{2n} = \frac{n}{2}\left(\frac{1}{n-1}U - 1\right)^2, \tag{16}$$

where

$$U = \frac{(n-1)\hat{S}_n^2}{\sigma^2} \sim \chi_{n-1}^2. \tag{17}$$

The density function of the random variable U is chi square distributions with $n-1$ degrees of freedom. This density function we can write (e.g. Wilks, 1962):

$$f(u)=\frac{1}{2^{\frac{n-1}{2}}\cdot\Gamma\left(\frac{n-1}{2}\right)}u^{\frac{n-3}{2}}e^{-\frac{u}{2}}\cdot I_{(0,\infty)}(u),\quad \text{dla}\quad n\geqslant 2. \tag{18}$$

The statistic (16) we can be expressed as follows:

$$Q_{2n}=\left(\sqrt{\frac{n}{2}}\frac{u}{n-1}-\sqrt{\frac{n}{2}}\right)^2, \tag{19}$$

where statistic u has chi square distribution with $n-1$ degrees of freedom.
If

$$z=\sqrt{\frac{n}{2}}\frac{u}{n-1}-\sqrt{\frac{n}{2}}, \tag{20}$$

then

$$u=(n-1)\cdot\left(\sqrt{\frac{2}{n}}z+1\right)\quad \text{and}\quad du=(n-1)\sqrt{\frac{2}{n}}\,dz \tag{21}$$

and

$$u\in <0,\infty)\quad \text{and}\quad z\in\left\langle -\sqrt{\frac{n}{2}},\infty\right)$$

The general theorems on functions of random variables (e.g. Krzyśko, 1996) lead to derivation the density function $f_1(z)$ of random variable Z in the following way:

$$f_1(z)=(n-1)\sqrt{\frac{2}{n}}f\left((n-1)\left(\sqrt{\frac{2}{n}}z+1\right)\right)\cdot I_{\left(-\sqrt{\frac{n}{2}};\infty\right)}=$$

$$=\frac{(n-1)\sqrt{\frac{2}{n}}}{\Gamma\left(\frac{n-1}{2}\right)\cdot 2^{\frac{n-1}{2}}}(n-1)^{\frac{n-3}{2}}\left(\sqrt{\frac{2}{n}}z+1\right)^{\frac{n-3}{2}}e^{-\frac{1}{2}(n-1)\left(\sqrt{\frac{2}{n}}z+1\right)}\cdot I_{\left(-\sqrt{\frac{n}{2}};\infty\right)}=$$

$$=\frac{(n-1)^{\frac{n-1}{2}}\sqrt{\frac{2}{n}}}{\Gamma\left(\frac{n-1}{2}\right)\cdot 2^{\frac{n-1}{2}}}e^{-\frac{1}{2}(n-1)}\left(\sqrt{\frac{2}{n}}z+1\right)^{\frac{n-3}{2}}e^{-\frac{1}{2}(n-1)\sqrt{\frac{2}{n}}z}\cdot I_{\left(-\sqrt{\frac{n}{2}};\infty\right)}. \tag{22}$$

Now we consider the statistic ν given by

$$V = Z^2. \tag{23}$$

Let

$$h(z) = z^2$$

$$D = \left\langle -\sqrt{\frac{n}{2}}, \infty \right), \quad H = \langle 0, \infty),$$

$$D_1 = \left\langle -\sqrt{\frac{n}{2}}, 0 \right), \quad H_1 = \left(0, \frac{n}{2} \right\rangle, \quad h_1^{-1}(\nu) = -\sqrt{\nu},$$

$$D_2 = \langle 0, \infty), \quad H_2 = \langle 0, \infty), \quad h_2^{-1}(\nu) = \sqrt{\nu}.$$

Then, the density function $f_2(\nu)$ of random variable V is as follows:

$$f_2(\nu) = \sum_{i=1}^{2} f_1(h_i^{-1}(\nu)) \cdot \left| \frac{d}{d\nu} h_i^{-1}(\nu) \right| I_{H_i}(\nu) \tag{24}$$

$$f_2(\nu) = f_1(h_1^{-1}(\nu)) \cdot \frac{1}{2\sqrt{\nu}} I_{(0, \frac{n}{2})}(\nu) + f_2(h_2^{-1}(\nu)) \cdot \frac{1}{2\sqrt{\nu}} I_{(0, \infty)}(\nu)$$

$$f_2(\nu) = \frac{1}{2\sqrt{\nu}} \left(a \left(\sqrt{\frac{2}{n}} \sqrt{\nu} + 1 \right)^{\frac{n-3}{2}} e^{-\frac{1}{2}(n-1)\sqrt{\frac{2}{n}}(-\sqrt{\nu})} I_{(0, \frac{n}{2}\rangle}(\nu) + \right.$$

$$\left. + a \left(\sqrt{\frac{2}{n}} (-\sqrt{\nu}) + 1 \right)^{\frac{n-3}{2}} e^{-\frac{1}{2}(n-1)\sqrt{\frac{2}{n}}\sqrt{\nu}} I_{(0, \infty)}(\nu) \right) =$$

$$= \frac{a}{2\sqrt{\nu}} \left(\left(1 - \sqrt{\frac{2\nu}{n}} \right)^{\frac{n-3}{2}} e^{\frac{1}{2}(n-1)\sqrt{\frac{2\nu}{n}}} I_{(0, \frac{n}{2}\rangle}(\nu) + \right.$$

$$\left. + \left(1 + \sqrt{\frac{2\nu}{n}} \right)^{\frac{n-3}{2}} e^{-\frac{1}{2}(n-1)\sqrt{\frac{2\nu}{n}}} I_{(0, \infty)}(\nu) \right),$$

where the constant a is following:

$$a = \frac{(n-1)^{\frac{n-1}{2}} \sqrt{\frac{2}{n}}}{\Gamma\left(\frac{n-1}{2} \right) \cdot 2^{\frac{n-1}{2}}} e^{-\frac{1}{2}(n-1)}. \tag{25}$$

The density function $f_2(v)$ can be rewritten in the following way, too:

$$f_2(v)=\begin{cases}0 & \text{for} \quad v\in(-\infty;0)\\ f_{2a}(v) & \text{for} \quad v\in\left[0;\frac{n}{2}\right\rangle\\ f_{2b}(v) & \text{for} \quad v\in\left[\frac{n}{2};\infty\right\rangle\end{cases}, \tag{26}$$

where:

$$f_{2a}(v)=\frac{a}{2\sqrt{v}}\left(\left(1-\sqrt{\frac{2v}{n}}\right)^{\frac{n-3}{2}}e^{\frac{1}{2}(n-1)\sqrt{\frac{2v}{n}}}\left(1+\sqrt{\frac{2v}{n}}\right)^{\frac{n-3}{2}}e^{-\frac{1}{2}(n-1)\sqrt{\frac{2v}{n}}}\right)$$

and

$$f_{2b}(v)=\frac{a}{2\sqrt{v}}\left(\left(1+\sqrt{\frac{2v}{n}}\right)^{\frac{n-3}{2}}e^{-\frac{1}{2}(n-1)\sqrt{\frac{2v}{n}}}\right.$$

Finally, we consider the distribution of the random variable given by the expression (10). Let q, t, and w are the values of the random variables Q_n, Q_{1n} and Q_{2n} respectively. When

$$q=t+w$$

then

$$t=q-w.$$

The statistic Q_{1n} has χ_1^2 distribution and its density function is as follows:

$$f_0(t)=\frac{1}{2^{\frac{1}{2}}\cdot\Gamma\left(\frac{1}{2}\right)}t^{\frac{-1}{2}}e^{-\frac{1}{2}t} \quad t>0$$

Hence:

$$h(q)=\int_0^{\infty}f_0(q-w)f_2(w)dw, \tag{27}$$

$$h(q)=\int_0^{\sqrt{\frac{n}{2}}}f_0(q-w)f_{2a}(w)dw+\int_{\sqrt{\frac{n}{2}}}^{\infty}f_0(q-w)f_{2b}(w)dw. \tag{28}$$

The density function of the statistic Q_n is determined by the expression (28). The distribution function of the random variable Q_n in a point q is defined by the expression:

$$F(q) = \int_{-\infty}^{q} \int_{0}^{\infty} f_0(q-w) f_2(w) dw dq. \tag{29}$$

This integral can be evaluated by means of the appriopriate numeric method. Using numerical methods we can find the cumulative distribution function for given sample number (n).

4. APPROXIMATION OF THE DISTRIBUTION OF THE STATISTIC Q_{2n}

The quantils $q(\alpha)$ of the random variable Q_n are determined on the basis of equation $F(q(\alpha)) = \alpha$. The left side of this equation is evaluated by means of numerical integration (cf. Dahlquist, Bjorck 1983). The equation $F(q(\alpha)) = \alpha$. can be rewritten in the following way:

$$\int_{q(\alpha)}^{\infty} f^*(q) dq = \alpha, \tag{30}$$

where:

$$f^*(q) = \int_{0}^{\infty} f_0(q-w) f_2(w) dw. \tag{31}$$

For sample size greater than 3 and $\alpha = 0.01$, 0.05, 0.1 the quantils $q(\alpha)$ were found. In this case the location of $q(\alpha)$ is presented by Figure 1.

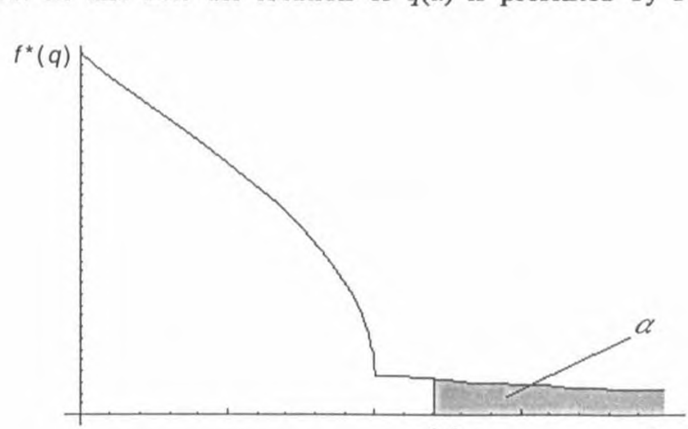

Fig. 1. The distribution function of statistic Q_n

The results of the numerical integration were presented in Tables 1–3. The tables present critical values of the proposed statistic for given significance level α (where $\alpha = 0.01$, 0.5 and 0.1) and for various sample size. These values can be treated as critical values for testing the hypothesis that the population mean is equal to μ and the population variance is equal to σ_2. The significantly large value of the statistic Q_n leads to rejection the null hypothesis. Such hypothesis is considered e. g. in problems of statistical quality control.

Table 1. Rejection values for significance level $\alpha = 0.1$

Sample size n	3	4, 5	6, 7	8–13	14 and more
Quantil $q(\alpha)$	5.2	4.9	4.8	4.7	4.6

Table 2. Rejection values for significance level $\alpha = 0.05$

Sample size n	3	4	5	6	7, 8	9	10, 11	12–14	15–18	19–62	62 and more
Quantil $q(\alpha)$	8.0	7.3	7.0	6.7	6.6	6.5	6.4	6.3	6.2	6.1	6.0

Table 3. Rejection values for significance level $\alpha = 0.01$

Sample size n	3	4	5	6	7	8	9	10	11	12	13	14, 15	16	17
Quantil $q(\alpha)$	20.1	16.6	14.7	13.6	12.9	12.4	12.0	11.6	11.4	11.3	11.1	10.9	10.7	10.6

Sample size n	18	19–21	22	23, 24	25, 27	28–32	33–41	42–45	46–52	53–69	70–104	105–213	214 and more
Quantil $q(\alpha)$	10.5	10.4	10.3	10.2	10.1	10.0	9.9	9.8	9.7	9.6	9.5	9.4	9.3

REFERENCES

Geary R. C. (1936), "Distribution of Student's Ratio for Non-normal Samples", *Journal of the Royal Statistical Society B*, **3**.

Cramer H. (1958), *Metody matematyczne w statystyce*, Państwowe Wydawnictwo Naukowe, Warszawa.

Dahlquist G., Bjorck A. (1983), *Metody numeryczne*, Państwowe Wydawnictwo Naukowe, Warszawa.

Krzyśko M. (1996), *Statystyka matematyczna*, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań.

Wilks C. S. (1962), *Mathematical Statistics*, John Wiley and Sons, New York–London.

Janusz Wywiał, Grzegorz Kończak

O ROZKŁADZIE FORMY KWADRATOWEJ ŚREDNIEJ I WARIANCJI Z PRÓBY

(Streszczenie)

W artykule rozważano statystykę Q_n, która jest formą kwadratową średniej i wariancji z n elementowej próby z rozkładu normalnego. Statystykę tę proponuje się wykorzystać do weryfikacji hipotezy o wartości oczekiwanej i wariancji zmiennej o rozkładzie normalnym. W oparciu o twierdzenia pozwalające wyznaczyć funkcje gęstości dla funkcji zmiennych losowych wyznaczono gęstość wspomnianej zmiennej losowej. Ze względu na skomplikowaną postać funkcyjną nie jest praktycznie możliwe wyznaczenie kwantyli dla tej zmiennej losowej, które byłyby podstawą do konstrukcji obszarów krytycznych proponowanego testu. Przybliżone wartości kwantyli rzędu $1-\alpha$ ($\alpha = 0{,}1$; $0{,}05$; $0{,}01$) wyznaczono wykorzystując metody całkowania numerycznego. Odpowiednie tabele wartości krytycznych przy różnych wielkościach próbki zostały przedstawione w artykule.

Otrzymane rezultaty mogą być wykorzystywane w zagadnieniach kontroli jakości lub teorii niezawodności.