CMST 14(1) 39-46 (2008) DOI:10.12921/cmst.2008.14.01.39-46

COMPUTATIONAL METHODS IN SCIENCE AND TECHNOLOGY **14(1)**, 39-46 (2008)

On Differently Defined Skewness

P. Sulewski

Pomeranian Academy, Arciszewskiego 22, 76-200 Slupsk, Poland
e-mail: sulewski@zis.pap.edu.pl

(Received: 8 November 2007; accepted: 20 January 2008, published online: 1 April 2008)

Abstract: Four definitions of skewness are discussed: classic skewness, two Pearson's skewnesses and Bowley's skewness. The ability of these skewnesses to express asymmetry is compared as well as the accuracy of their estimation from normal distribution is assessed.
Key words: skewness, Johnson's distribution, method of Parzen, estimator density function.

I. INTRODUCTION

In statistical literature four different definitions of the skewness exist. Beside the classic definition presented in section III, also two Pearson's skewnesses defined in sections IV and V, as well as Bowley's skewness described in section VI occur. To compare these skewnesses certain probability distribution will be useful, in which through the change of parameter values in a wide range their asymmetry change is possible. Distributions derived from the Gaussian distribution fit superbly for numerical experiments of this type. Johnson's distribution of type S_R and S_U, in which changing the parameter values makes it possible to get the transition from the negative skewness to the positive one, deserves special attention in this aspect. In the present work Johnson's distribution of type S_U was used for numerical experiments, whose domain – in contrast to Johnson's distribution of type S_R – is a set of real numbers. In section VII ability of these skewnesses to express asymmetry was compared as well as the accuracy of their estimation from normal distribution was assessed.

II. DISTRIBUTIONS DERIVED FROM THE GAUSSIAN DISTRIBUTION

A cumulative distribution function of distributions derived from the normal distribution is given by [4]

$$F(x) = \Phi\left[\vartheta\left(x, \bar{\theta}\right)\right], \tag{1}$$

where $\Phi(.)$ is a normal cumulative distribution function, $\vartheta(x)$ is an increasing function of argument x, whereas $\bar{\theta}$ is a vector of parameters of the discussed distribution.

Examples of distributions derived from the normal distribution are:

- Lognormal distribution,
- Chhikary's distribution,
- Birnbaum–Saunders's distribution,
- Johnson' distribution of type S_L, S_B, S_U.

The family of Johnson's distribution describes the formula [5]

$$z = \gamma + \eta \cdot \psi\left(\frac{x-\varepsilon}{\lambda}\right) \tag{2}$$

transforming random variable x into random variable z dependent on normal distribution $N(0,1)$. A detailed discussion of Johnson's distributions of type S_L, S_B, S_U together with numerous numerical examples is possible to find in a book by Drapella [4].

As it is difficult to investigate properties of distributions with four parameters, in further considerations we will accept $\varepsilon = 0, \lambda = 1$ and $\gamma = -a/b$, $\eta = 1/b$. With these assumptions (2) takes the form

$$z = \frac{\psi(x) - a}{b}. \tag{3}$$

The cumulative distribution function for Johnson's distribution of type S_U random variable x is given by

$$F(x)=\Phi\left[\frac{\ln\left(x+\sqrt{x^2+1}\right)-a}{b}\right] \tag{4}$$

whereas the density function is given by

$$f(x)=\frac{1}{\sqrt{2\pi}\cdot b\cdot\sqrt{x^2+1}}\times$$
$$\times\ \exp\left[-\frac{1}{2}\times\left(\frac{\ln\left(x+\sqrt{x^2+1}\right)-a}{b}\right)^2\right]. \tag{5}$$

The density function of Johnson's distribution of type S_U for combinations of parameter values presented in Table 1 was exemplified in Figs. 1 and 2.

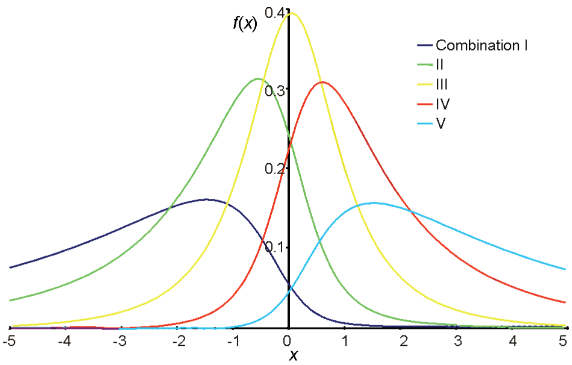

Fig. 1. Density function of Johnson's distribution of type S_U for combinations I-V of parameter values presented in Table 1

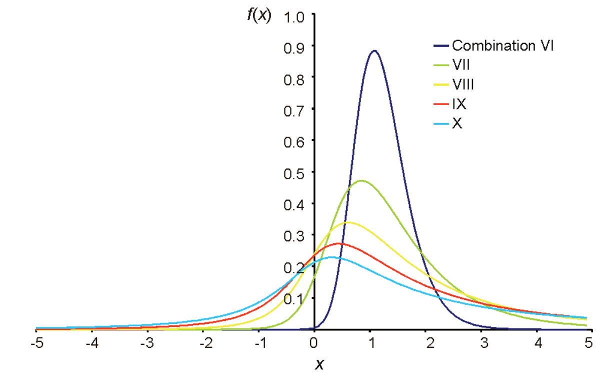

Fig. 2. Density function of Johnson's distribution of type S_U for combinations VI-X of parameter values presented in Table 1

Table 1. Combinations of parameter values

Combination	a	b	Combination	a	b
I	-2	1	VI	1	0.3
II	-1	1	VII	1	0.6
III	0	1	VIII	1	0.9
IV	1	1	IX	1	1.2
V	2	1	X	1	1.5

It follows from Figure 1 that in Johnson's distribution of type S_U it is possible to change the value of parameters to get the transition from negative skewness to a positive one.

III. THE CLASSIC SKEWNESS

Let us note that the classic skewness is calculated as [2, 6]

$$\gamma_1^K=\frac{\mu_3}{\mu_2^{3/2}}, \tag{6}$$

where μ_k – for continuous distribution – are central moments of the k-th order in form of [1, 2, 6]

$$\mu_k=\int_{-\infty}^{\infty}(x-\alpha_1)^k f(x)\,dx. \tag{7}$$

Central moments of Johnson's distribution of type S_U are impossible to define analytically, therefore mathematical environment Mathcad counting the value of indefinite integrals was used. The model computer implementation of classic skewness, written in Mathcad, was introduced below.

$a:1 \quad b=1$

$$f(x):=\frac{\exp\left[\frac{\left(\frac{\ln\left(x+\sqrt{x^2+1}-a\right)}{b}\right)^2}{-2}\right]}{\sqrt{2\pi}\cdot b\cdot\sqrt{x^2+1}}$$

$$\alpha_1:=\int_{-\infty}^{\infty}x\cdot f(x)dx$$

$$\alpha_2:=\int_{-\infty}^{\infty}x^2\cdot f(x)dx$$

$$\alpha_3:=\int_{-\infty}^{\infty}x^3\cdot f(x)dx$$

$$\mu_2:=\alpha_2-(\alpha_1)^2$$

$$\mu_3:=\alpha_3-3\cdot\alpha_1\cdot\alpha_2+2\cdot(\alpha_1)^3$$

$$\gamma_1:=\frac{\mu_3}{\sqrt{(\mu_2)^3}} \qquad \gamma_1=5.363.$$

The relation between classic skewness for Johnson' distribution of type S_U and values of parameters a and b was presented in Fgs. 3 and 4. It is notable that

$$\lim_{b\to 0}\gamma_1^K=0,\ \lim_{b\to\infty}\gamma_1^K=\infty. \tag{8}$$

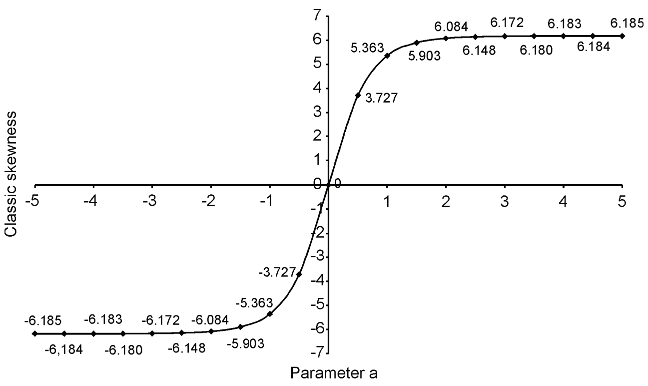

Fig. 3. Relation between classic skewness for Johnson' distribution of type S_U and parameter a

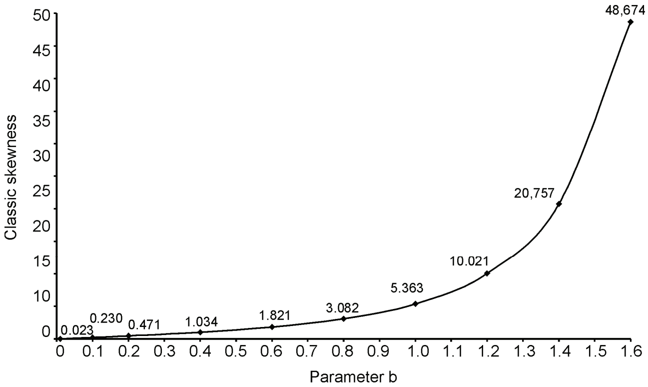

Fig. 4. Relation between classic skewness for Johnson' distribution of type S_U and parameter b

IV. PEARSON'S SKEWNESS

Pearson's skewness (the mode skewness) is calculated as [7]

$$\gamma_1^P = \frac{\alpha_1 - x_{\text{mod}}}{\sqrt{\mu_2}}, \tag{9}$$

where α_1 is a mean value, x_{mod} – mode of distribution.

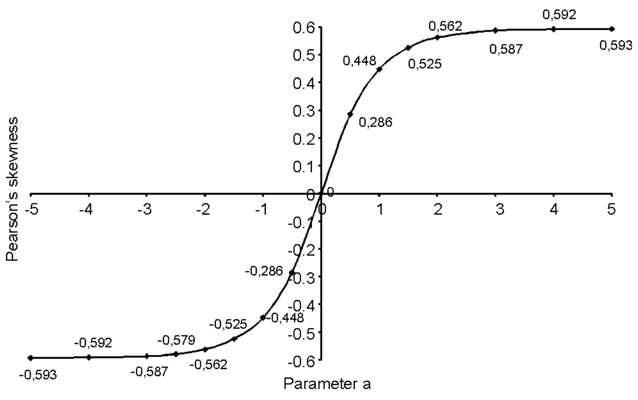

Fig. 5. Relation between Pearson's skewness for Johnson' distribution of type S_U and parameter a

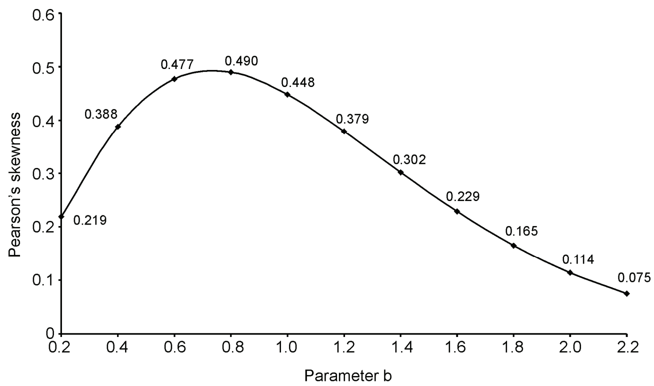

Fig. 6. Relation between Pearson's skewness for Johnson' distribution of type S_U and parameter b

To calculate the values of this coefficient computational environment Mathcad was used. The relation between Pearson's skewness for Johnson' distribution of type S_U and values of parameters a and b was presented in Figs. 5 and 6. One should notice that

$$\lim_{b \to \infty} \gamma_1^P = 0 . \tag{10}$$

V. MEDIAN SKEWNESS

The median skewness is given by

$$\gamma_1^M = \frac{\alpha_1 - x_{0.5}}{\sqrt{\mu_2}}, \tag{11}$$

where α_1 is a mean value, $x_{0.5}$ – a median of distribution. This coefficient is well-known in literature as Pearson's second skewness coefficient [7].

Values of this coefficient in Mathcad were counted. The relation between median skewness for Johnson' distribution of type S_U and values of parameters a and b are presented in Fgs. 7 and 8. Let us notice that

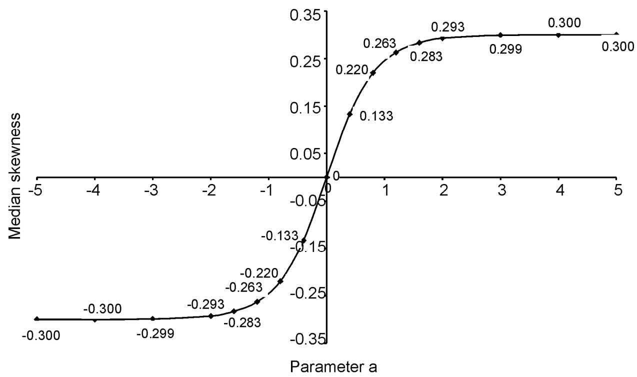

Fig. 7. Relation between median skewness for Johnson' distribution of type S_U and parameter a

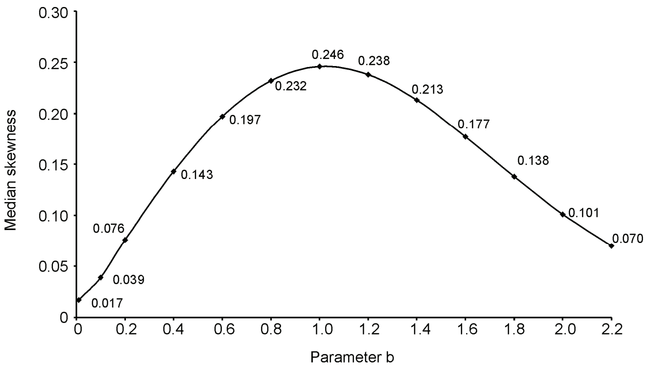

Fig. 8. Relation between median skewness for Johnson' distribution of type S_U and parameter b

$$\lim_{a\to-\infty}\gamma_1^M = 0.3\,,\quad \lim_{a\to\infty}\gamma_1^M = 0.3\,, \tag{12}$$

$$\lim_{b\to 0}\gamma_1^M = 0\,,\quad \lim_{b\to\infty}\gamma_1^M = 0\,. \tag{13}$$

VI. BOWLEY'S SKEWNESS

Bowley's skewness is defined as [7]

$$\gamma_1^B = \frac{\left(x_{0,75} - x_{0,5}\right) - \left(x_{0,5} - x_{0,25}\right)}{\left(x_{0,75} - x_{0,25}\right)}\,, \tag{14}$$

where x_k are quantiles of the k-th order of distribution $\left(0 < k < 1\right)$.

Values of quantiles of Johnson' distribution were calculated in Solver, which is located in Microsoft Excel. The relation between Bowley's skewness for Johnson' distribution of type S_U and values of parameters a and b are presented in Fgs. 9 and 10. It is worthwhile marking that

$$\lim_{a\to-\infty}\gamma_1^B = -1\,,\ \lim_{a\to\infty}\gamma_1^B = 1\,, \tag{15}$$

$$\lim_{b\to 0}\gamma_1^B = 0\,,\ \lim_{b\to\infty}\gamma_1^B = 1\,. \tag{16}$$

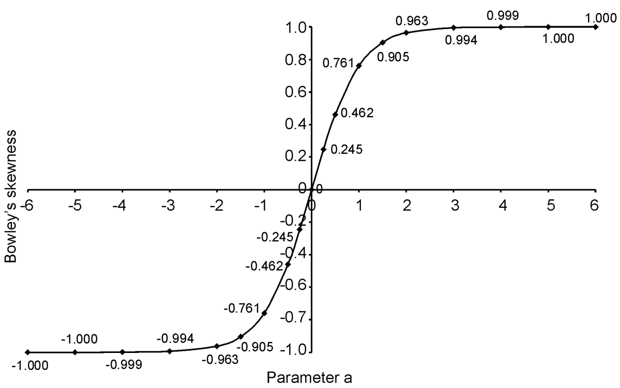

Fig. 9. Relation between Bowley's skewness for Johnson' distribution of type S_U and parameter a

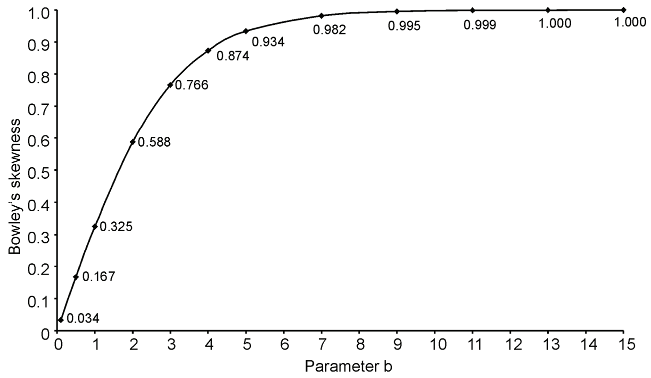

Fig. 10. Relation between Bowley's skewness for Johnson' distribution of type S_U and parameter b

VII. THE COMPARISON OF SKEWNESS

The ability of discussed skewnesses to express asymmetry is shown in Figs. 11 and 12. Skewness for Johnson' distribution of type S_U and value parameters contained in Table 1 were compared on them, thanks to which distribution about negative, zero and positive asymmetry was received.

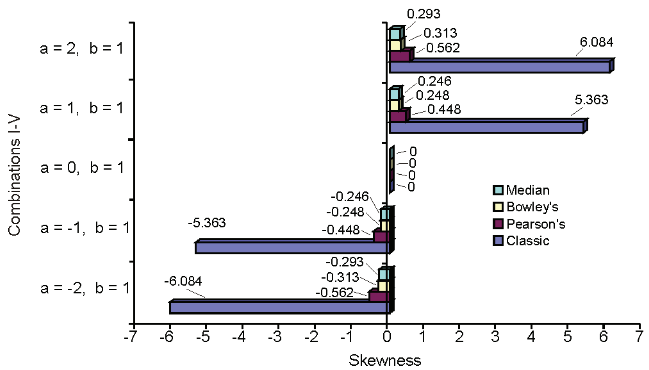

Fig. 11. The ability of skewness to express asymmetry for combinations I-V

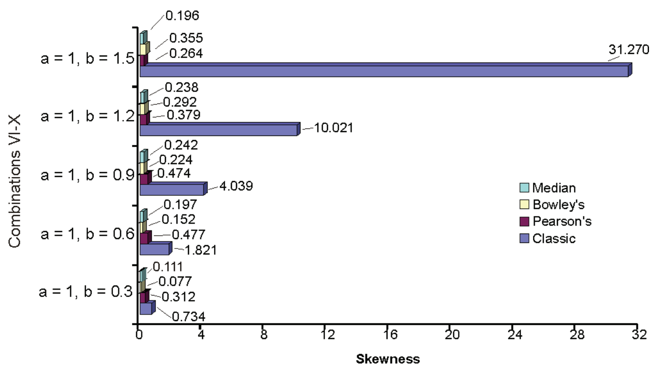

Fig. 12. The ability of skewness to express asymmetry for combinations VI-X

Assessments of estimation accuracy of these skewnesses were executed too, when a sample $x_1^*, x_2^*, \ldots, x_n^*$ was drawn from Gaussian distribution. Values x_i^* $(i = 1, \ldots, n)$ were generated by means of a function NormLos, which was created in Visual Basic for Applications (VBA).

```
Function NormLos(m As Single, s As Single)
      Dim i As Integer
      Dim sum As Single
      sum = 0
      For i = 1 To 12
          Let sum = sum + Rnd
      Next i
      Let NormLos = s * (sum - 6) + m
End Function
```

The estimate of the sample moment of *k*-th order is given by

$$\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^*\right)^k ; \tag{17}$$

however, the unbiased estimator of the central moment of 2-nd and 3-th order are calculated as [2]

$$\hat{\mu}_2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^* - \hat{\alpha}_1\right)^2 , \tag{18}$$

$$\hat{\mu}_3 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(x_i^* - \hat{\alpha}_1\right)^3 . \tag{19}$$

Unknown values of quantiles were replaced by appropriate order statistics [3]

$$\text{number} = \text{int}\left[n * \alpha\right] + 1 . \tag{20}$$

The sample mode is, according to the definition, a position of maximum of the empirical density function.

Figures 13 and 14 present the relation between variance calculated on the basis of 10 240 estimations each from skewness and a sample size *n*.

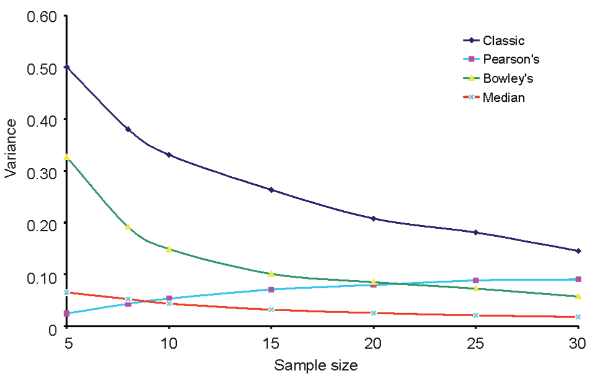

Fig. 13. The relation between variance and sample size

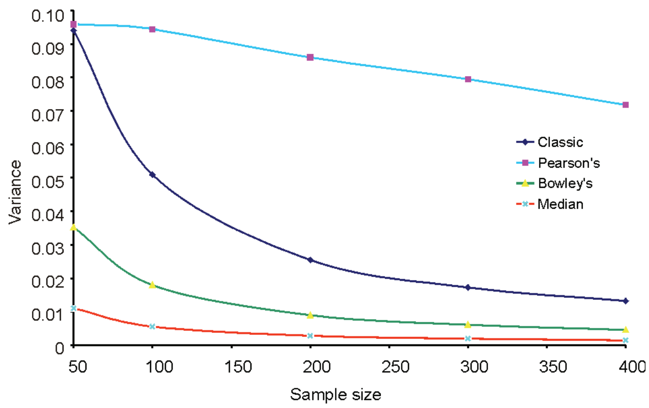

Fig. 14. The relation between variance and sample size

For $n \le 7$ the smallest variance has the mode skewness. The classic skewness for $n \le 48$ has the biggest variance, because – as it is widely known – the accuracy of estimation worsens significantly alongside with the increase of the order of central moments. Among the analysed skewnesses, median skewness should be taken into account, which for $n \ge 8$ has the smallest variance. To confirm the above-quoted facts as well as in order to smooth-out empirical density functions, the author employed the Parzen Method also known as the kernel method [8, 9]. The empirical density function is composed of "kernels". In this paper each kernel is of Gaussian form

$$K(z) = \frac{1}{\sqrt{2\pi}\,h}\exp\left(-\frac{1}{2}z^2\right) \qquad z = \frac{x - x_{(i)}^*}{h} , \tag{21}$$

therefore the empirical density function is given by

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{nh}\sum_{i=1}^{n}\exp\left[-\frac{1}{2}\left(\frac{x - x_{(i)}^*}{h}\right)^2\right] . \tag{22}$$

The parameter *h* is a function of sample size

$$h(n) = 3\frac{\sigma}{\sqrt{n}} = 3\sqrt{\frac{S_2^* - S_1^{*2}}{n}} , \tag{23}$$

where

$$S_1^* = \frac{1}{n}\sum_{i=1}^{n} x_{(i)}^* , \qquad S_2^* = \frac{1}{n}\sum_{i=1}^{n} x_{(i)}^{*2} \tag{24}$$

The computer implementation of estimation of four skewnesses, written in VBA, was introduced below. Comments were placed after apostrophes.

```
Sub Estimate()
      'declaration of tables
      Dim edf(50, 2) As Double
      Dim x() As Double
      Dim skewness() As Double
      'declaration of variables
      Dim m As Single, s As Single, mode As Double
      Dim xd As Double, xg As Double, krok As Double
      Dim q1 As Double, q2 As Double, q3 As Double
      Dim xc As Double, b1 As Double, index As Long
      Dim i As Long, n As Long, k As Long
      Dim sr As Double, m2 As Double, m3 As Double
      Dim c1 As Double, hor As Byte, cc As Double
      Dim s1 As Double, s2 As Double, ds As Double
      Dim h As Double, j As Long, xx As Double
      Dim t As Double, max As Double

      Randomize Timer
      Worksheets("estimate").Select   'selecting worksheet "estimate"
      'introduction of cells to variables
      Let m = Cells(3, 1).Value   'mean value
      Let s = Cells(3, 2).Value   'standard deviation
      Let n = Cells(3, 3).Value   'sample size
      ReDim x(n)
      ReDim skewness(10240, 4)
      For k = 1 To 10240
              For i = 1 To n
                  Let x(i) = NormLos(m, s) 'recall to the function
              Next i
          'sorting
      powrot:
          Let hor = 0
          For i = 1 To n - 1
              If (x(i) <= x(i + 1)) Then GoTo dalej
              Let b1 = x(i)
              Let x(i) = x(i + 1)
              Let x(i + 1) = b1
              Let hor = 1
      dalej:
          Next i
          If hor = 1 Then GoTo powrot
          'the empirical density function - method of Parzen
          Let s1 = 0
          Let s2 = 0
          For i = 1 To n
              Let ds = x(i)
              Let s1 = s1 + ds
              Let s2 = s2 + ds * ds
          Next i
          Let s1 = s1 / n
          Let s2 = s2 / n
          Let s2 = s2 - s1 * s1
          Let h = 3 * Sqr(s2 / n)
          Let c2 = 1 / (Sqr(2 * Application.Pi()))
          Let xd = x(1) - 3 * h
          Let xg = x(n) + 3 * h
          Let krok = (xg - xd) / 50
          For j = 0 To 50
              Let xx = xd + j * krok
              Let edf(j, 1) = xx
              Let edf(j, 2) = 0
              For i = 1 To n
                  Let t = (xx - x(i)) / h
                  Let edf(j, 2) = edf(j, 2) + Exp(-0.5 * t * t)
              Next i
              Let edf(j, 2) = c2 * edf(j, 2) / n / h
          Next j
          'the sample mode
          max = edf(1, 2)
          For j = 2 To 50
              If edf(j, 2) > max Then max = edf(j, 2): index = j
          Next j
          mode = edf(index, 1)
          'the sample quantiles
```

```
    q1 = x(Int(n * 0.25) + 1)
    q2 = x(Int(n * 0.5) + 1)
    q3 = x(Int(n * 0.75) + 1)
    'the sample moments
    sr = 0
    For i = 1 To n
        sr = sr + x(i)
    Next i
    sr = sr / n
    For i = 1 To n
        xc = x(i) - sr
        m2 = m2 + xc ^ 2
        m3 = m3 + xc ^ 3
    Next i
    m2 = m2 / n
    m3 = m3 / n
    c1 = Sqr(n * (n - 1)) / (n - 2)
    skewness(k, 1) = (c1 * m3) / m2 ^ (1.5)                    'classic skewness
    skewness(k, 2) = (sr - mode) / Sqr(m2 * n / (n - 1))       'Pearson's skewness
    skewness(k, 3) = ((q3 - q2) - (q2 - q1)) / (q3 - q1)       'Bowley's skewness
    skewness(k, 4) = (sr - q2) / Sqr(m2 * n / (n - 1))         'median skewness
Next k
'introduction of results to cells
For i = 1 To 10240
        Let Cells(i, 1) = skewness(i, 1)
        Let Cells(i, 2) = skewness(i, 2)
        Let Cells(i, 3) = skewness(i, 3)
        Let Cells(i, 4) = skewness(i, 4)
Next i
End Sub
```

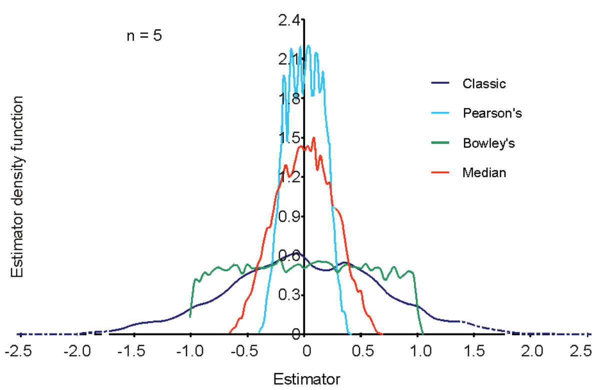

Fig. 15. The estimator density function of skewness obtained with the Parzen Method for $n = 5$

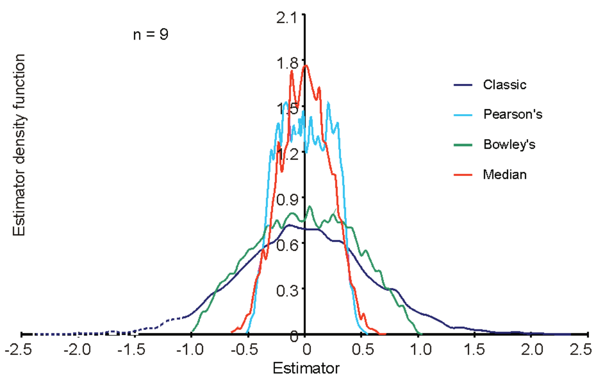

Fig. 16. The estimator density function of skewness obtained with the Parzen Method for $n = 9$

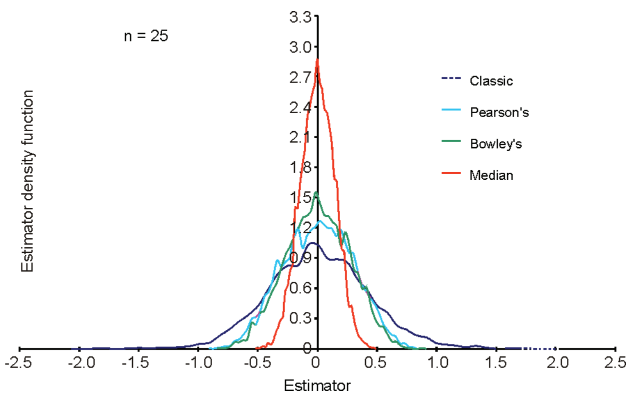

Fig. 17. The estimator density function of skewness obtained with the Parzen Method for $n = 25$

Figures 15-17 present the estimator density function of skewness obtained with the Parzen Method for $n \in \{5; 9; 25\}$. The sampling mode influences – particularly for small n – wavy 'shape of density function of estimate of skewness.

References

[1] S. Brandt, *Data analysis*, Warsaw 1998 (in Polish).

[2] H. Cramer, *Mathematical method in statistics*, Warsaw 1958 (in Polish).

[3] H. A. David, *Order statistics*, Wiley, New York, 1970.

[4] A. Drapella, *Statistical inference on the base skewness and kurtosis*, Pomeranian Pedagogical Academy, Slupsk 2004 (in Polish).

[5] N. L. Johnson, *System of frequency curves generated by methods of translation*, Biometrika 36 (1949).
[6] M. G. Kendal, A. Stuart, *The advanced theory of statistics*, Distribution theory, 1, London 1958.
[7] J. F. Kenney, E. S. Keeping, *Mathematics of statistics*, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, 101-102 (1962).
[8] E. Parzen, *On estimation of a probability density and mode*, Ann. Math. Statist, **33**, 1065-1076 (1962).
[9] B. W. Silverman, *Density estimation*, Chapman and Hall, London – New York, 1986.

PIOTR SULEWSKI graduated in Mathematics in 1996. Since then he has been working at the Institute of Mathematics at Pomeranian Academy in Słupsk. He received the PhD in reliability theory in 2001 from the Systems Research Institute of Polish Academy of Sciences in Warsaw. His research interests concern reliability mathematics and computational methods in statistics.