

REMARKS ON MODAL VALUE

Andrzej Wilkowski

Abstract. In this paper we talk about modal value, ideal modal and the relationship between stable distributions and the statistical characteristics like modal and ideal modal. The second part of this article is about the properties of normal and skew-normal density. In the third part of the article we present multiaverage. Multiaverage is an approximation of the random variable with more than just one point at the same time (which is important when we talk about random variables, which distributions are mixtures, or about multi-modal densities). While defining multiaverage, we use standard moments method and some facts from orthogonal polynomial theory.

Keywords: modal value, ideal modal, skew-normal density, multiaverage.

JEL Classification: C19.

DOI: 10.15611/me.2013.9.08.

1. Modal value and ideal modal value

Modal value is one of the central measures. According to the classical definition, it indicates the most probable value (if we talk about discrete distribution), or the most probable value in the sample. For differentiable distribution, it is the maximum density function.

Stable distributions are one of the most crucial. Their characteristic functions' form is given by:

$$\varphi(t) = \exp\left\{imt - |ct|^\alpha (1 - i\beta l(t))\right\},$$

where

$$l(t) = \begin{cases} \operatorname{sgn}(t)tg\left(\frac{\pi\alpha}{2}\right) & \text{for } \alpha \neq 1, \\ -\operatorname{sgn}(t)\frac{2}{\pi}\ln|t| & \text{for } \alpha = 1, \end{cases}$$

Andrzej Wilkowski

Department of Mathematics and Cybernetics, Wrocław University of Economics, Komandorska Street 118/120, 53-345 Wrocław, Poland.

E-mail: andrzej.wilkowski@ue.wroc.pl

and

$$0 < \alpha \leq 2, |\beta| \leq 1, m \in \mathbb{R}, c > 0,$$

where

$$\operatorname{sgn}(t) = \begin{cases} 1 & \text{for } t > 0, \\ 0 & \text{for } t = 0, \\ -1 & \text{for } t < 0. \end{cases}$$

All stable distributions are continuous functions. Only densities of normal, Cauchy's and Levy's distributions can be described by elementary functions.

The fact which is given below shows the relation between modal value and stable distributions.

Theorem (Yamazato 1978). All stable distributions are unimodal.

A short proof of this fact can be found in (Simon 2011).

The proposal of the new concept, connected with modal value and called ideal modal value, appeared in the paper (Smoluk 1997). Let function f be:

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

$$f_{\mu,x}(h) = \begin{cases} 0 & \text{for } h < 0, \\ \int_{x-h}^{x+h} d\mu(t) & \text{for } h \geq 0, \end{cases}$$

where μ is a real probabilistic measure in \mathbb{R} , x is a fixed real number and $h \in \mathbb{R}$. Let F be:

$$F(\mu) = \{f_{\mu,x} : x \in \mathbb{R}\}.$$

It is an ordered set, which means that

$$f_{\mu,x} \leq f_{\mu,y} \Leftrightarrow \forall h \in \mathbb{R} \quad f_{\mu,x}(h) \leq f_{\mu,y}(h).$$

Definition 1. The m is an ideal modal value of probability measure μ , if there only exists the element $f_{\mu,m}$ in ordered set F , such that

$$\forall x \in \mathbb{R} \quad f_{\mu,x} \leq f_{\mu,m}.$$

The existence of ideal modal value guarantees unimodality of distribution. Ideal modal value is then equal to modal value. It is possible that distribution has one modal value and does not have ideal modal value. The necessary condition on the existence of ideal modal value is the symmetry of distribution. For random variables which have symmetric distributions and finite mean value, ideal modal value, mean value, median and symmetry point distribution are equal. The theorem given below describes distributions which have ideal modal value.

Theorem (Smoluk 2000). Closure of a set of convex and linear combinations of uniform distributions, which have equal ideal modal value m , is equal to the class of all distributions which have ideal modal value m .

2. Normal and skew-normal distribution

Normal distribution is one of the most important probability distribution formulas used in the theory and practice of probability science and statistics. Normal distribution was originally introduced by de Moivre (Cramer 1958) in 1733, in his examination of limes forms of binomial distribution. This initial postulate went largely unnoticed, leading to the re-discovery of normal distribution in the works of Gauss in 1809 and Laplace in 1812 (Cramer 1958). The authors arrived at normal distribution principles in the course of their analyses of experiment error theory.

Definition 2. Random variable X is considered as falling into normal distribution with parameters m and s (in principle, $X \sim N(m, s)$), where $s > 0$, $m \in R$, if its density function takes the form of

$$f_{m,s}(x) = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2s^2}\right), x \in R.$$

As seen in the above formula, the resulting curve is symmetrical and unimodal, reaching maximum at point $x = m$, which at the same time is the mean ($E(X) = m$), median and modal value of the distribution. Variance of random variable X is expressed by a second parameter: $Var(X) = s^2$.

This section addresses selected properties that characterize normal distribution. Variants of the central border theorem as well as the infinite divisibility property are omitted, with discussion centered on some of the lesser-known characteristics (Wilkowski 2008).

• If U and V are independent random variables defined on the same probability space, monotonously distributed on $(0,1)$, then

$$X = \sqrt{-2\log U} \cos(2\pi V)$$

and

$$Y = \sqrt{-2\log U} \sin(2\pi V)$$

are independent and distributed along $N(0, 1)$ (Jakubowski, Sztencel 2000). This property is frequently used in normal distribution random number generators (random numbers in monotonous distribution can be generated fairly easily).

• Cramér's theorem. If normally distributed random variable X is a sum of two independent random variables Y and Z , then both those variables are normally distributed as well (Cramer 1958).

• Let random variables U, V be independent, and

$$X = aU + bV,$$

$$Y = cU + dV.$$

If X and Y are independent, then all four variables are normal, unless $b = c = 0$ or $a = d = 0$ (Feller 1978).

It must be noted that the above property allows to define Gaussian random variables in infinite-dimension Banach spaces or groups (in the latter case, it is enough to define the sum).

• Let $R(X, Y)$ be defined as: $R(X, Y) = \sup r\{f(X), g(Y)\}$, where r is a correlation coefficient of respective random variables, while supremum applies to all functions f and g , for which $0 < \text{Var}\{f(X)\} < \infty$, $0 < \text{Var}\{g(Y)\} < \infty$. If random vector (X, Y) is normal, then $R(X, Y) = r_{X, Y}$. Proof of this theorem can be found in (Lancaster 1957, Yu 2008).

• Assume that random variables X and Y are independent and identically distributed.

Then

$$\frac{2XY}{\sqrt{X^2 + Y^2}} \sim N(0, s) \quad \text{only if } X \sim N(0, s).$$

Analogical characteristics applies to symmetrical Bernoulli distribution.

While keeping the above assumptions,

$$\frac{2XY}{\sqrt{X^2 + Y^2}}$$

is a random variable symmetrically, Bernoulli distributed only if variable X is distributed in the same way. It is worth noting that Poisson distribution and standard Bernoulli distribution do not share the above property. For proof on that, see (Novak 2007).

The skew-normal distribution (Azzalini 1985) can be got from normal distribution.

Definiton 3. Random variable X_λ has skew-normal distribution with $\lambda \in \mathbb{R}$ (which we write as $X_\lambda \sim SN(\lambda)$), if its density function is given by

$$f_\lambda(x) = 2\varphi(x)\Phi(\lambda x), \quad x \in \mathbb{R}, \quad (1)$$

where φ and Φ are density and distribution of normal distribution.

Skew-normal distribution is not a stable distribution, has one modal value, and does not have ideal modal value (it is not a symmetric distribution). As a conclusion from (1), random variable from this distribution has all moments. Its mean value and variance are given by (Jamalizadeh et al. 2008):

$$E(X_\lambda) = \frac{2\lambda}{\sqrt{2\pi(1+\lambda^2)}}, \quad (2)$$

$$Var(X_\lambda) = 1 + \frac{2\lambda^2(2+\lambda^2)}{\pi(1+\lambda^2)^{\frac{3}{2}}} - E^2(X_\lambda).$$

Generalisation of skew-normal distribution can be found in (Jamalizadeh et al. 2008; Sharafi, Behboodiani 2008; Satheesh Kumar, Anusree 2011; Bernardi 2013).

Example 1. Density function of distribution $SN(2)$ (Figure 1).

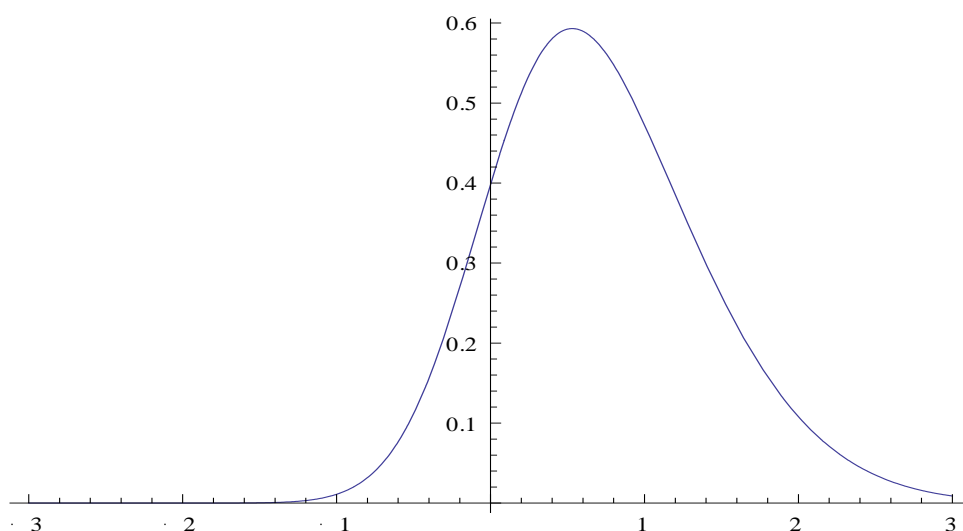


Fig. 1. Density function of distribution $SN(2)$

Source: own study.

In this case, according to (2),

$$E(X_2) = \frac{4}{\sqrt{10\pi}}$$

and modal value $x \approx 0.7$. As we can see, mean value approximately equal to modal value.

3. Multiaverage

The paper (Wilkowski 2011) is used in this section. Moments m_k of random variable X (on probability space (Ω, \mathcal{F}, P)) are important characteristics, used in statistic and theory of probability:

$$m_k = E(X^k) = \int_{\Omega} X^k(\omega)P(d\omega),$$

where $k = 1, 2, \dots$, while integrals are unconditionally convergent.

First moment $m_1 = E(X)$ is called mean value, average. Linear combination of first second moments are defined by

$$V_1(X) = \text{Var}(X) = E(X - E(X))^2 = m_2 - m_1^2,$$

and is called variance. Polynomial $X - E(X)$ minimizes root-mean-square norm, which is given below:

$$\min_{a \in \mathbb{R}} E(X - a)^2 = E(X - E(X))^2 = V_1(X).$$

According to approximation by least squares, mean value is the best one point approximation of random variable.

We now assume that random variable X has density function f_X . Consecutive moments can then be calculated from

$$m_k = \int_{-\infty}^{+\infty} x^k f_X(x) dx,$$

$k = 1, 2, \dots$, where integrals are unconditionally convergent.

Maximums of function f_X , so called modal values, are also important in statistic research. They mark concentration points of probability. In unimodal density case, average $E(X)$ is a good modal value approximation (in symmetric case, both average and modal value are equal).

Let random variable X has finite moments rank $2n - 1$:

$$E(X^k) = m_k < \infty, \quad k = 1, 2, \dots, 2n - 1.$$

Normed polynomial p_n , which minimizes norm:

$$\min_{a, b, \dots, c \in \mathbb{R}} E(X^n + aX^{n-1} + bX^{n-2} + \dots + c)^2 = E(p_n(X))^2,$$

is given below (Cramer 1958, Laurent 1975):

$$p_n(x) = K \begin{vmatrix} 1 & m_1 & \dots & m_n \\ \dots & \dots & \dots & \dots \\ m_{n-1} & m_n & \dots & m_{2n-1} \\ 1 & x & \dots & x^n \end{vmatrix}, \text{ where } K \neq 0 \quad (3)$$

While p_n is orthogonal polynomial rank n (Szegő 1975), we have:

$$p_n(x) = (x - s_1) \dots (x - s_n), \text{ where } s_1 < \dots < s_n \quad (4)$$

Definition 4. Ordered n values $(s_1; \dots; s_n) = E_n(X)$ is called n -mean (multiaverage) of random variable X (Antoniewicz 2005). It is obvious that $E_1(X) = E(X)$.

This vector is a square root approximation of random variable by n points. Equivalents of variance and standard deviation are:

$$V_n(X) = E((X - s_1) \dots (X - s_n))^2,$$

$$\sqrt[2n]{V_n(X)} = \sqrt[2n]{E((X - s_1) \dots (X - s_n))^2}. \quad (5)$$

These characteristics measure mean-square deviation of random variable X , from n probability concentration points. Other characteristics, related with multiaverage can be found in (Antoniewicz, Wilkowski 2004, Antoniewicz 2005).

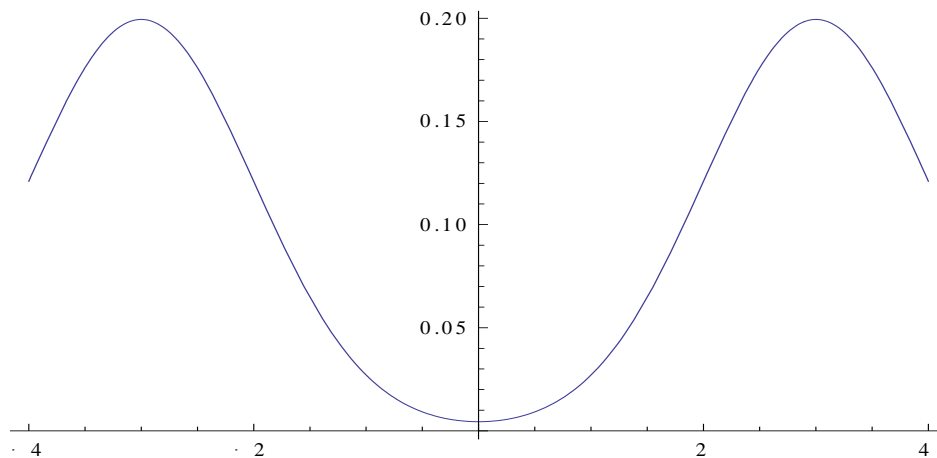


Fig. 2. Density function of distribution $\frac{1}{2}N(-3,1) + \frac{1}{2}N(3,1)$

Source: own study.

Example 2. Let random variable X have distribution, which is a mixture of two normal distributions: $X \sim \left(\frac{1}{2}N(-3,1) + \frac{1}{2}N(3,1) \right)$.

Density function of this distribution is given by:

$$f_X(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x+3)^2}{2}} + \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}}.$$

A diagram of this function can be found in Figure 2.

Mean value ($E(X) = 0$) is, in this case, "unexpected value", because probability is concentrated in points $x_1 = -3$, $x_2 = 3$. This distribution does not have ideal modal value. From (3) and (4) we have p_2 :

$$p_2(x) = (x + \sqrt{10})(x - \sqrt{10}).$$

Biaverage $E_2 = (s_1; s_2) = (-\sqrt{10}; \sqrt{10})$ approximates modal values better than mean value. Standard deviation and its generalization for 2-mean (according to (5)) are: $\sqrt{V_1(X)} = \sqrt{10}$, $\sqrt[4]{V_2(X)} = \sqrt[4]{38}$, which also suggest that biaverage is a more precise characteristic than mean value. Polynomial p_3 , 3-mean and its generalization of standard deviation are given by:

$$p_3(x) = (x + 3.7148)x(x - 3.7148),$$

$$E_3(X) = (s_1; s_2; s_3) = (-3.7148; 0; 3.7148),$$

$$\sqrt[6]{V_3(X)} = 2.7799 > \sqrt[4]{V_2(X)}.$$

Calculation of next n -means $E_n(X)$ increases approximation error of modal values. It turns out, that the most precise characteristic in this case is biaverage.

Example 3. Let random variable X have distribution which is the mixture of two skew-normal distributions: $X \sim \left(\frac{1}{2} SN(-1) + \frac{1}{2} SN(12) \right)$.

Its density function is given by:

$$f_X(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}} \left(\int_{-\infty}^{-x} e^{-\frac{t^2}{2}} dt + \int_{-\infty}^{12x} e^{-\frac{t^2}{2}} dt \right).$$

A diagram of this function can be found in Figure 3. As we can see, this distribution is not stable, does not have ideal modal value and has two

modal values in: $x_1 \approx -0.6$, $x_2 \approx 0.3$. Mean value and 2-average are, respectively, equal:

$$E(X) = 0.199471, E_2(X) = (s_1; s_2) = (-1.121760; 0.091124).$$

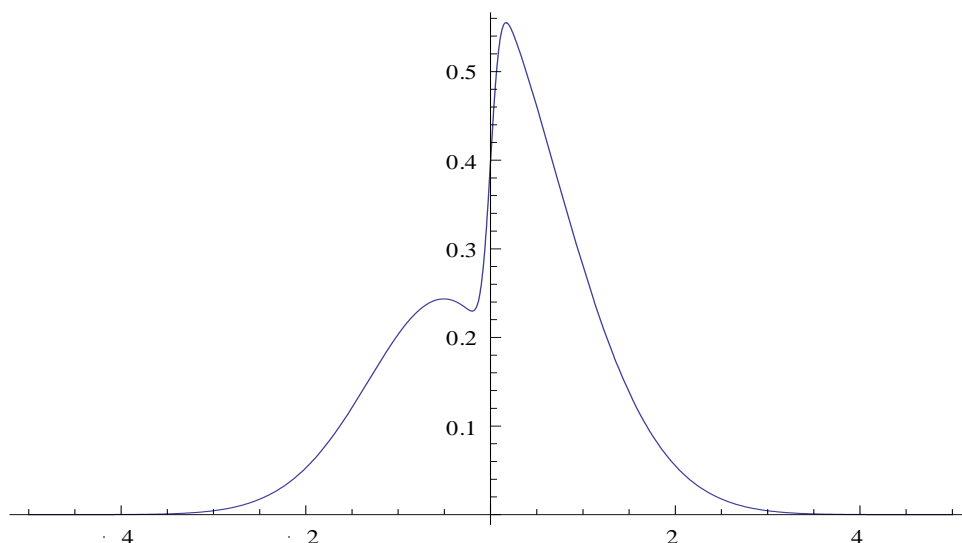


Fig. 3. Density function of distribution $\frac{1}{2} SN(-1) + \frac{1}{2} SN(12)$

Source: own study.

Standard deviation and its generalization in 2-average case are equal: $\sqrt{V_1(X)} = 0.979904$, $\sqrt[4]{V_2(X)} = 1.450051$. It is hard to indicate which of these characteristics (mean value or 2-average) approximate better modal value, due to the fact that both modal values are almost equal. For next n -averages the error of approximation (given by $\sqrt[2n]{V_n(X)}$) increases.

4. Summary

The oldest, general method of making an estimation of distribution parameters with the use of the range in the sample, is the Pearson's moment method. It is widely used by him and his successors (Cramer 1958). It consists in comparing some of the moments in the sample to moments of distribution, which are functions of unknown parameters. After the calculation of this system of equations we get their estimators. This method is

usually easy to solve. The calculation of n-averages uses this method. That is why it seems that n-average may be a useful tool for more precise data analysis (for example to the localization of concentration points of probability (modal values), in the case of multimodal and mixture distributions).

References

- Antoniewicz R. (2005). *O średnich i przeciętnych*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
- Antoniewicz R., Wilkowski A. (2004). *O pewnym rozkładzie dwumodalnym*. Przegląd Statystyczny. Tom 51. I. Warszawa.
- Azzalini A. (1985). *A class of distributions which includes the normal ones*. Scandinavian Journal of Statistics 12.
- Bernardi M. (2013). *Risk measures for skew normal mixtures*. Statistics and Probability Letters 83.
- Billingsley P. (1987). *Prawdopodobieństwo i miara*. PWN. Warszawa.
- Cramer H. (1958). *Metody matematyczne w statystyce*. PWN. Warszawa.
- Feller W. (1978). *Wstęp do rachunku prawdopodobieństwa*. Tom II. PWN. Warszawa.
- Jakubowski J., Sztencel R. (2000). *Wstęp do teorii prawdopodobieństwa*. Script. Warszawa.
- Jamalzadeh A., Behboodian J., Balakrishnan N. (2008). *A two-parameter generalized skew-normal distribution*. Statistics and Probability Letters 78.
- Lancaster H.O. (1957). *Some properties of the bivariate normal distribution considered in the form of a contingency table*. Biometrika 44.
- Laurent P.J. (1975). *Aproksymacja i optymalizacja* (tłumaczenie z francuskiego) Wydawnictwo „Mir”. Moskwa.
- Novak S.Y. (2007). *A new characterization of the normal law*. Statistics and Probability Letters 77. Elsevier.
- Satheesh Kumar C., Anusree M.R. (2011). *On a generalized mixture of standard normal and skew normal distributions*. Statistics and Probability Letters 81.
- Sharafi M., Behboodian J. (2008). *The Balakrishnan skew-normal density*. Statistical Papers 49.
- Simon T. (2011). *A multiplicative short proof for the unimodality of stable densities*. Electronic Communications in Probability 16.
- Smoluk A. (1997). *O definicji wartości modalnej*. Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 750. Wrocław.
- Smoluk A. (2000). *Moda idealna a prognozy*. Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 838. Wrocław.
- Szego G. (1975). *Orthogonal Polynomials*. Colloquium Publications XXIII. American Mathematical Society. Providence.
- Wilkowski A. (2008). *Notes on normal distribution*. Didactics of Mathematics 5(9). The Publishing House of Wrocław University of Economics. Wrocław.

- Wilkowski A. (2011). *Notes on line dependent coefficient and multiaverage*. Mathematical Economics 7(14). The Publishing House of Wrocław University of Economics. Wrocław.
- Yamazato M. (1978). *Unimodality of infinitely divisible distribution functions of class L*. Annals Probability. Vol. 6. No (4).
- Y. Yu (2008). *On the maximal correlation coefficient*. Statistics and Probability Letters 78. Elsevier.