EDUCATIONAL STANDARDS IN STATISTICS

Antoni Smoluk
Wroclaw School of Economics
Poland

In the paper we propose some educational standards for modal value and random sample. Those standards should facilitate deeper understanding of these statistical notions.

INTRODUCTION

Modal value, a dominant, the most frequent value or simply a mode, is a positional parameter informing about distribution of probability. In place of distribution of probability we shall be using a synonymous expression, a probabilistic measure, shortened to one word, measure – since only probabilistic measures will be mentioned.

And so what is a mode of measure μ? To start with, before giving a precise definition of this notion we should mention that not all the positional parameters are defined for each measure. There are measures for which an average value is defined, and there are measures for which it is not so – because it does not exist. Similarly this is the case with standard deviation and moments of higher order. So for certain measures a modal value exists, and for others it does not. Before we propose a new definition of a mode, and an attempt to generalise it into multidimensional distribution, let us consult the literature. This notion does not figure in many textbooks on probability. Why? Because it is not properly defined and there is not a precise definition of what a mode is. "The mode is the value of the variable corresponding to the maximum of the ideal curve which gives the closest possible fit to the actual distribution" (Yule & Kendall, 1948). Further on Yule explained this definition of modal value: "It represents the value which is most frequent or typical, the value which is, in fact, the fashion (*la mode*)". In theory, for continuous distribution with a single maximum of the density function, such a definition is satisfactory. In practice however, we have only histograms. Yule warned his readers loyally. "It is, in fact, difficult to determine the mode for such distributions as arise in practice, particularly by elementary methods". The entry *mode*, written by A.V. Prochorov, can be found in a well prepared, much valued by mathematicians and practitioners, Soviet six-volume *Encyclopaedia of Mathematics*. "If function $g(x)$ is a density of a random variable, then every point x_0 in which $g(x)$ reaches the maximum value is called a mode. A mode can be also defined for discrete distributions: if a random variable takes on value x_k with a probability p_k and when $x_k < x_{k+1}$ then point x_m is called a mode, when $p_{m-1} \le p_m$ and $p_{m+1} \le p_m$. Distributions with a single, two or larger number of modal values are called respectively unimodal, bimodal and multimodal. The most important for the theory of probability and for statistics are unimodal distributions" (Vinogradov, 1982). The entry *mode* can also be found in the popular *Mathematics Dictionary* of Glenn James. "The member of a series of measurements or observations that occurs most often, if there is such a member; there is no useful definition if more than one member occurs most often. If more students in a given class make 75 than any other one grade, then 75 is the mode. For a continuous random variable with probability density function *f*, the mode is the point at which *f* has its greatest value, if there is exactly one such point. Sometimes any point at which *f* has a local maximum is called a mode, but multimodality is unusual in practise" (James, 1976). Similar information can be found in a pedantic German *Lexicon der Mathematik* under *Modawert* (*Modus*, *Mode*), (Gellert, Kästner, & Neuber, 1977). There is no mention of a mode in a selection of articles valuable for historical studies, *Studies in the History of Statistics and Probability* (Pearson & Kendall, 1970), which can indicate additionally a low opinion of the theoretical and practical importance of a mode. The above mentioned statements suggest that the notion of a mode has been depreciated. Is it not useful? Certainly it is useful and important. It has been pushed aside due to its imprecise definition, and not because of its small usefulness. Karl Pearson introduced the idea of a mode in 1895 in his work *Skew variation in homogenous material*. This notion is still important and occurs frequently in research and applications (e.g. Wywial, 2000). I think it is high time to review and make more precise this undoubtedly useful – especially in prediction theory – notion. Statistics is not just a language and a method but above all a science of the real world, formulating laws of nature, and every science is a mathematical, exact discipline.

DEFINITION OF IDEAL MODAL VALUE

Let μ signify a probabilistic measure in a set of real numbers R. In a symbolic language it means that $\mu \in \mathrm{Prob}(R)$. Let x is any but fixed real number. To the number $x \in R$ and the measure $\mu \in \mathrm{Prob}(R)$ is assigned a function $f(\mu, x)$ transforming the set of real numbers R in R by a formula

$$f(\mu, x)(h) = \begin{cases} 0, & \text{if} \quad h < 0, \\ \int\limits_{x-h}^{x+h} d\mu(t), & \text{if} \quad 0 \le h, \end{cases}$$

where $h \in R$, and an integral appearing in a definition of the function $f(\mu, x)$ is a probability that any random variable with a distribution μ takes on values from the closed interval $[x - h , x + h]$. Naturally the function $f(\mu, x)$ is a certain distribution function. The symbol $F(\mu)$ designates the family of all such distribution functions

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

The family $F(\mu)$ is an order set. It is

$$f(\mu, x) \le f(\mu, y)$$

if and only if, for every $h \in R$ it is

$$f(\mu, x)(h) \le f(\mu, y)(h).$$

If in the family $F(\mu)$ exists a function $f(\mu, m)$ such, that for every $x \in R$ is

$$f(\mu, x) \le f(\mu, m),$$

then the value m is called *an ideal mode* of the measure μ. The ideal mode of the measure μ exists only when in the family $F(\mu)$ exists the largest element – supremum. *Theorem.* An ideal mode is unique. *Argument*. If an ideal mode exists, it is only one because the supremum of the set is unique. *Comment*. Every normal distribution has a mode and an ideal mode; both of these values are equal to the average.

REMARKS AND MAIN RESULT

A mode for a uniform distribution can be any point of an interval of positive density. In such case the notion has no value. However an ideal mode of uniform distribution is unique, and obviously just like for normal distribution, it equals to the mean. The necessary condition to exist an ideal mode is the symmetry of distribution. The distribution has to be symmetrical in order to have an ideal mode. Naturally many a symmetrical distribution exist without an ideal mode. At the same time we should stress the great importance of multimodal distribution in the cluster analysis and in the theory of subdivision of the population into classes – just as it is done in calibration. Multimodality cannot be fought against because the distributions of such kind take place in natural and social sciences. The notion of a mode is not very useful in cases when the distribution is multimodal. Simply a mode carries little information about distribution. I have even come across a task to measure an informational value of a mode. An ideal mode gives far more information than a mode. Full information about an ideal mode m is contained in a distribution function $f(\mu, m)$ defining the mode. Such function carry also information about an ordinary mode, because to each modal value m – in a traditional sense – in a set of distribution functions $F(\mu)$ corresponds a maximal element $f(\mu, m)$. This modal value m which corresponds to distribution function $f(\mu, m)$, and the function is maximum element in the set $F(\mu)$, deserves special attention as their informative value is greater. An ideal mode is a stable value due to a weak convergence of random variables.

Theorem. If a sequence (f_n) of random variables with distributions (μ_n) is weakly convergent to a random variable f with distribution μ, and the measure μ_n for each natural n has an ideal mode m_n, then the measure μ has an ideal mode m and $m = \lim (m_n)$.

Examples: A. Measure μ with a density

$$g(x) = \begin{cases} 0\,, & \text{if} \quad \dfrac{\pi}{2} \le |x|, \\ c\,\dfrac{\cos^2 x}{1+x^2}\,, & \text{if} \quad |x| \le \dfrac{\pi}{2}, \end{cases}$$

and number $0 < c$ is selected in such a way that function g is the density, has a mode equal 0. It is a symmetrical unimodal distribution, which has also an ideal mode.

B. A symmetrical distribution

$$\mu_o = (\,1/10,\ 3/20,\ 1/2,\ 3/20,\ 1/10),$$

associating with the natural numbers 0, 1, 2, 3 and 4 respectively these probabilities, has modal value equal 2. It is naturally symmetrical distribution with a centre of symmetry 2. Such distribution has no ideal modal value since it is a purely atomic distribution and has no density. If we extend this distribution to the density function g given by the following formula, we shall obtain a continuous symmetrical distribution with an ideal mode equal 2 (Fig. 1).

$$g(t) = \begin{cases} 0, & \text{if} \quad t < -\dfrac{1}{2} \quad \text{or} \quad \dfrac{9}{2} < t, \\ \dfrac{1}{10}, & \text{if} \quad -\dfrac{1}{2} \le t < \dfrac{1}{2}, \\ \dfrac{3}{20}, & \text{if} \quad \dfrac{1}{2} \le t < \dfrac{3}{2}, \\ \dfrac{1}{2}, & \text{if} \quad \dfrac{3}{2} \le t < \dfrac{5}{2}, \\ \dfrac{3}{20}, & \text{if} \quad \dfrac{5}{2} \le t < \dfrac{7}{2}, \\ \dfrac{1}{10}, & \text{if} \quad \dfrac{7}{2} \le t \le \dfrac{9}{2}, \end{cases}$$

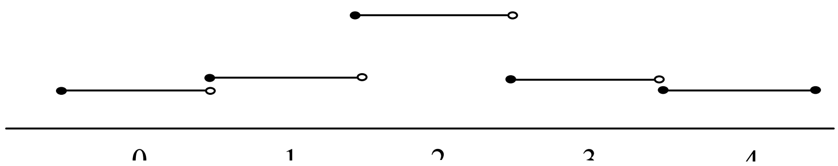

Figure1

C. Distribution of a student with n degrees of freedom, $n \in N$, $1 \le n$, with density given by a formula

$$g_n(t) = \frac{1}{\sqrt{n}\,B\left(\dfrac{1}{2}, \dfrac{n}{2}\right)} \; \frac{1}{\left(1 + \dfrac{t^2}{n}\right)^{\frac{n+1}{2}}}\,, \quad t \in R,$$

where $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$, $x, y \in R$, $0 < x$, $0 < y$, has an ideal mode equal to an ordinary modal value.

Generalisation. The notion of an ideal mode can be easily extended onto multi-dimensional distributions. In such cases a term *centre* is more appropriate because an ideal mode is a certain point m in the space R^n. Therefore if $\mu \in \text{Prob}(R^n)$, so μ is a n-dimensional distribution, then analogously to one-dimensional case, we can define function $f(\mu, x)$ using a formula

$$f(\mu, x)(h) = \begin{cases} 0, \ if \ h < 0, \\ \displaystyle\int_{K(x,h)} d\mu(t), \ \text{if} \ 0 \le h, \end{cases}$$

where $K(x, h)$ is a closed ball $\{z \in R^n : |x - z| \leq h\}$, h is a real number, x a vector in R^n, and the number $| y |$ is a norm of vector y. The ideal mode – a centre – of distribution $\mu \in \mathrm{Prob}(R^n)$ we can call a point $m \in R^n$ such, that function $f(\mu, m)$ is the biggest element – the supremum – of family $F(\mu)$. The mode of distribution in R^n is a point $m = (m_1, ..., m_n)$ therefore a certain vector. It is not a single number but a series of numbers. Multidimensional normal distribution has an ideal mode and it is a point in which density reaches the highest value. This mode does not depend on the choice of a norm in R^n and it is a vector consisting of modal values of the marginal distributions.

RANDOM SAMPLE

A sample is a multidimentional stochastic variable

$$f : X \to R^n ,$$

where X is a set of elementary events and $f(x) = (f_1(x), ..., f_n(x))$; the distributions of stochastic variables f_i are the same and are equal the distribution of the general population. The sample is a specific measurement of general population. General theory of probability can use language of stochastic variables and, equivalently, the language of the measure theory. Statistics is founded on the notion of the stochastic variable. We assume that general population is a probabilistic measure p in the set of elementary events X. If an event B is possible, that means that probability of B is greater than zero, then event B can be thought as sample. The result of a measurement of the probability p is the measure q which is defined by the formula of the conditional probability

$$p(AB) / p(B) .$$

A stochastic sample is then some application of the probabilistic space $(M(X), p)$ into the probabilistic space $(X(B), q)$, where $M(X)$ is a Boolean algebra of a subset of the set X, and $X(B)$ is the image of this algebra relatively to B. The measurement of event $A \in M(X)$ is the event $S(A) = AB$ in probabilistic space $X(B)$, moreover we assume that

$$q(S(A)) = p(AB) / p(B) .$$

That both definitions are some educational and statistical standards. First of them is founded on the notion of quality and number, the second uses properties which are equivalent to the sets. So, in the first case the sample is a multidimentional stochastic variable, and in the second the function S. A sample structure is harmonized with the structure of general population $(A_0, ..., A_k)$, here events A_i are disjoint and cover all the space X, if vectors

$$(p(A_0), ..., p(A_k)), \; (p(A_0B), ..., p(A_kB))$$

are linearly depended. So the structure of general population is given by some discrimination of the set X. Sample structure is rarely harmonized with the structure of general population.

REFERENCES

Gellert, W., Kästner, H., & Neuber, S. (1977). *Lexikon der mathematik*. Leipzig: Bibliographisches Institut.

James, R.C. (1976). *Mathematics – Dictionary*. New York: Van Nostrand Reinhold.

Pearson, E. S. & Kendall, M.G. (1970). *Studies in the history of statistics and probability*. London: Griffin.

Vinogradov, I.M. (1982). *Mathematical Encyclopaedia.* Moscow: Sovietskaya Encyclopaedia.

.Wywial, J. (2000). Estimation of mode on the basis of a truncated sample. *Acta Universitatis Lodziensis. Folia Oeconomica 152,* 73-81.

Yule, G.U., & Kendall, M.G. (1948). *An introduction to the theory of statistics*. London: Charles Griffin & Company.