13th International Congress on Mathematical Education
Hamburg, 24-31 July 2016

RANDOM VARIABLE AND ITS RELATIONSHIP WITH STATISTICAL VARIABLE: AN EDUCATIONAL PERSPECTIVE FROM AN ANALYSIS OF THE CONCEPT

Blanca Ruiz

Tecnologico de Monterrey

The random variable is a probabilistic concept which allows the passage of probability of events to the study of the distributions. One of its main aspects is that the basis for the study of statistical inference, but also its links with the statistical variable. The random variable mathematical object and its relations with the statistical variable is analyzed using content analysis of several books both college textbook on probability and statistics as other introductory probability theory to analyze.

INTRODUCTION

Unlike Algebra or Calculus, Probability exists on a formal definition of the 'variable' that is linked with chance. The crucial question that arises immediately after noticing this seemingly superficial fact is why, what is the meaning of the definition of the random variable?, what was it that caused the need for such a definition? Adjective (random) not only responds to the need to distinguish between two scientific fields, since it is one of the key concepts of the axiomatization of probability theory proposed by Kolmogorov (1956). Nor is a concept that has been used only recently, since from the first attempts to quantify the chances of winning in gambling (Bellhouse, 2000), began to define implicitly 'variables' that made more accessible such quantification.

This concept is also remarkable because, despite its long history and its impact on the probability theory is not very important in the teaching of probability and statistics. In the classroom it is made explicit its existence to the university level and its definition, when not omitted, simplified or distorts (Miller, 1998). One might think that should not be as necessary or it is not possible to teach it in all its complexity. However, Heitele (1975) proposed it as one of the ten fundamental ideas of teaching stochastic. A fundamental idea, according to Heitele is one that is likely to develop from basic education and can be taken up throughout the school education with different cognitive levels and deepening. However, it is such an abstract concept that we might ask ourselves the need to introduce it in school or learning problems that can lead to teach according to its definition. Moreover, since the conceptual baggage of his definition it is inaccessible even for college students what conception can be more accessible to college students without distorting the nature of its formal definition?

Systemic and cognitive complexity of the random variable is trivialized in the teaching of probability and statistics and has been recently treated for educational research. In this article, we summarize some of the epistemological elements that show the complex nature of this concept to give meaning to the question of what is the random variable from the discipline. The answer intended to assist in better treatment of the probability and the link between probability and statistics.

AN ANALYSIS OF ITS CONCEPT

The random variable is a function that associates a numerical value to each event in the sample space associated with a random experiment (Walpole, Myers and Myers, 1999). In numeric context it is feasible to work with the probabilities of the values that the random variable takes instead of the probabilities of the events because the probability in the sample space must be previously defined. This means that this seemingly simple concept transforms the events (in the form of sets) in numbers and can model the relationship of the sample space by the probability distribution in functional form within the real numbers. But its formal definition is not easy to interpret, since the role of the random variable is quite complex because it establishes a link between a probability space (defined in an algebra of events and a probability measure) and real numbers. However, the mathematical definition of the random variable is not explicit on how the rule of linking the probability space and real numbers is selected since it is not necessarily implied in the sample points. The definition of this standard is determined by the context of the problem to be solved.

Parzen (1971) proposes to introduce the random variable through random phenomena with numerical results (p. 172). Subsequently, random variable is defined and then it is clarified that random phenomena with numerical results, the random variable is the identity function. He emphasizes that we must learn to recognize and formulate mathematical objects that are described verbally random variables as functions. Parzen thus emphasizes that we must clarify how the linkage rule that defines the random variable is a function and its role. So, in a finite sample space, random variable, like any mathematical function, has three components: The sample space associated with the experiment (domain); the matching rule, which define how the sample space and real numbers are linked; and the set of numerical values (image), which takes the random variable. It is also noted that the image of the random variable is a set of real numbers that relate to the sample space and also meet all field properties of all real. Under these conditions, it is possible to establish a functional relationship (the distribution function) between R (image of the random variable) and the interval [0, 1]. In the distribution function, the image of the random variable plays the role of independent variable and the probability assigned to each of these values, the dependent variable with which it is possible to make use of the tool of analysis. So in the process of construction of the distribution function, the random variable plays different roles. Once you define the random variable, the distribution function is also defined. Indeed the ultimate goal of the process linked to the random variable is the definition of this function because it is useful model for analysis of the problem situation.

Importantly, the way how some authors conceptualize the random variable. Krickeberg (1973), Wackerly, Mendenhall and Schaeffer (2002) and Devore (2011) are more or less consistent with the axioms of probability: the probability distribution of a discrete random variable is introduced immediately after (or both) of the definition of the random variable, then the definition and calculation of the expectation, variance and moments of a discrete random variable and finally the continuous random variable, distribution and moments are treated. Parzen (1971) proposes a different order. He defines the distribution function of a random phenomenon numerical results, the expectation, variance and moments and introduces some more known distribution functions and then define the concept of random variable. In Parzen, distribution functions after this issue involving several variables or composite functions. Ríos (1967) rescues the frequentist probability and statistics variable.

In either case, their positions are based on the analysis of students' knowledge, but rather an analysis of knowledge itself, so we can assume that each book is based on different epistemic postures to develop students apparently the same subject. Only the different reflections of these authors, demonstrate the epistemic complexity of the random variable.

While the random variable refers to all possible values (theoretical) of a variable in a random experiment, the statistical variable describing the set of data values obtained by performing a given number n of times the experiment. This variable is linked in any way to chance, but is more related to the data. Its nature is empirical and therefore cannot be considered rather than theoretical infinity. Thus, the random variable by a process of convergence of the statistical variable can also define (Ríos, 1967). This shows that the definition of random variable given by Kolmogorov (1957) does not care about assigning probabilities, but is consistent with the definition of classical probability, which also implies that the definition of the random variable is displayed differently depending of assigning probabilities established.

The statistical variable is also used in the analysis of a sample and plays an important role in the frequency inference, where it is used to infer characteristics of the variable corresponding random. This is consistent with the definition established by Rios.

However, assuming convergence of a statistical variable to a random also implies the existence of an infinite population where there is a random phenomenon linked. But if the population is finite a census can be done. In that case a random phenomenon not intervene because all data will be collected. The variable involved is statistical, but can become random variable when a random phenomenon is involved, that is, when asked what will happen if an individual of the population is randomly selected. The probability distribution would seemingly be the same as the relative frequency distribution: your graphics have the same shape and values would be the same, but conceptually would be very different, the first being random variable and the second statistical variable.

Then, the statistical variable is related to the collection of a sample of data and random variable with a knowledge of the probability space (which can be obtained empirically through a census or theoretically). The random variable is always associated with a random experiment, but not necessarily the statistical variable. The random variable is linked with various probability assignments (classical, frequency, subjective or otherwise), however the statistical variable is linked to the frequencies. Both variables are linked in various ways: in the convergence in infinite populations, the insertion of a random experiment in finite population censuses and sampling inference, which influence the modeling of random phenomena.

The analysis of this relationship also illuminates the importance of the definition of the random variable in the different probability assignments, since the definition of the random variable passes through different times and will coordinate different notions depending on the probability assignment to take into account. Thus, assigning probabilities combines and links epistemic disputes and set out on the random variable.

Thus, assigning probabilities combines and links epistemic disputes and set out on the random variable since the definition of a random variable passes through different times and different notions together depending on the probability assignment is taken into account.

CONCLUSIONS

This study allowed us to show that the development of statistical thinking involves two seemingly contradictory processes. On the one hand the domain of abstract mathematical tools are required and the other a break with the deterministic thinking to accept another kind of knowledge which is characterized by providing results with a degree of uncertainty is needed. It should be noted, therefore, the importance of the random variable in modeling random situations and the opportunity offered by the field of probability to teach modeling in mathematics, because of the many everyday situations that can be modeled through this tool.

As a rule of correspondence and from a purely mathematical perspective, the random variable could be arbitrary when the sample space satisfies the conditions defined in the definition (Mood and Graybill, 1963). But from the perspective of modeling, mapping rule must have a sense from your question it seeks to solve. Therefore, in a context, it cannot be so arbitrary: the random variable to be explained in the context of the problem. This makes the random phenomena modeling not only the random variable takes a leading role but also the relationship he has with the statistical variable.

The probability theory does not describe how the probability space is formed. The terms of the axiomatic probability (or formal) and thus provides certain conditions upon which rests the theory. However, the relationship between this theory and reality is given by the way in which the probability value is assigned. This assignment will also indicate how it operates and the random variable arises. That is, the meanings brings theory to reality and vice versa comes from that assignment. Thus, modeling processes are conditioned by the probability assignment: classical, frequency, subjective and axiomatic, since they condition the appearance of mathematical and probabilistic concepts associated with the random variable.

References

Bellhouse, D. R. (2000). De Vetula: a medieval manuscript containing probability calculations. *International Statistical Review*, 68(2), 123-136.

Devore, J. (2011). *Probabilidad y estadística para ingeniería y ciencias*. México: Cengage Learning.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, 6(3), 187-205.

Kolmogorov A. (1956). *Foundations of the theory of probability*, 2° edition. New York: Chelsea Publishing Company (original work in 1933).

Krickeberg, K. (1973). *Teoría de la probabilidad*. Barcelona: Teide.

Miller, T. K. (1998). The random variable concept in introductory statistics. In L. Pereira- Mendoza (Ed.). *Proceedings of the Fifth International Conference on Teaching Statistics*. (pp. 1221-1222). Singapur: IASE.

Mood, A. y Graybill, F. (1963). *Introducción a la teoría de la estadística*. Madrid: Aguilar.

Parzen, E. (1971). *Teoría moderna de probabilidades y sus aplicaciones*. México: Limusa.

Ríos, S. (1967). Métodos estadísticos. Madrid: Del Castillo.

Wackerly, D., Mendenhall, W. y Scheaffer L. (2002). *Estadística matemática con aplicaciones*. México: Thompson.

Walpole, R., Myers, R. y Myers, Sh. (1999). Probabilidad y estadística para ingenieros (6ª Ed.) (trad. R. Cruz). México: Prentice-Hall Hispanoamericana