KNOWLEDGE OF BINOMIAL DISTRIBUTION IN PRE-SERVICE MATHEMATICS TEACHERS

Nicolás Fernández, Jaime García-García, Elizabeth H. Arredondo, and Isaac Imilpán
Departamento de Ciencias Exactas, Universidad de Los Lagos, Osorno, Chile
nicolasalonso.fernandez@alumnos.ulagos.cl

The binomial distribution is one of the most important discrete distributions in probability and statistics; however, research identifies weaknesses in teachers' and students' application of binomial distributions for solving tasks beyond the direct use of the formula. Based on historical epistemological study and notions from the onto-semiotic approach to mathematical knowledge and instruction, we designed and administered a questionnaire to secondary school mathematics teachers in training. In our results, we identify and describe a lack of articulation among historical epistemological elements of the binomial distribution. Teachers can correctly use concepts such as combinatorics, probability, and the binomial distribution formula to model and identify binomial phenomena but cannot answer questions about a random variable or expected value. They show a lack of consideration of alternative representations.

INTRODUCTION

Probability allows human beings to better understand random events of everyday life through tasks such as the identification of possible cases, the calculation of probabilities for event occurrences, and the modeling of random phenomena, which implies a need to educate probabilistically literate students (Gal, 2005). For this, understanding the binomial distribution is essential because it allows for modeling daily life situations and facilitates the understanding of other important probabilistic and mathematical notions such as the normal distribution and the law of large numbers. However, for both students and teachers, it is possible to identify conflicts in aspects of probability such as modeling, case counting, and the application of constructs such as combinatorics (Alvarado & Batanero, 2007; Sánchez & Carrasco, 2018) that require the articulation of the different components of the binomial distribution and its nature as a mathematical and probabilistic object (Batanero, 2005). These are of particular concern for prospective teachers, given their role in the educational process. Teachers in training should be prepared to teach the binomial distribution from this perspective, but because they have been trained in a more theoretical way, such a perspective is unfamiliar to them (Batanero, 2020). Promoting it would also facilitate teachers developing understanding of students' learning conflicts, the different elements and ideas that give shape and meaning to the binomial distribution, and the background behind the proposals in school curriculum and textbooks.

The Onto-semiotic Approach (OSA) to mathematical knowledge and instruction provides a theoretical-methodological framework that addresses epistemological, anthropological, sociocultural, cognitive, and semiotic aspects of mathematics education (Gordillo & Pino-Fan, 2016). From a historical-epistemological study (Anacona, 2003) based on the notions of this approach, it is possible to identify key historical elements for developing the meaning of the binomial distribution, which can be associated with the reasoning and mathematical practices of students when facing problem situations of the same nature. The identified components and relationships can provide indicators for generating hypothetical learning trajectories, resolving learning conflicts, and analyzing reasoning.

In order to contribute to the early state of research on the didactics of the binomial distribution (García-García & Sánchez, 2015), we explored the historical-epistemological knowledge evidenced by teachers in training of Pedagogy in Secondary Education in Mathematics and Computer Science, from a questionnaire designed based on the meaning of the binomial distribution and reconstructed based on the notions of OSA and historical-epistemological study. We considered the following research questions:

- What knowledge do prospective high school mathematics teachers have about the binomial distribution?
- What elements of the binomial distribution do pre-service mathematics teachers mobilize when faced with binomial problems?

The identified historical-epistemological strengths and weaknesses can be used for educational purposes in the proposal of didactic interventions or general proposals for teacher education.

THEORETICAL FRAMEWORK

OSA is "a modular and inclusive theoretical system for mathematics education that includes principles and methodological tools to address the central issues involved in the processes of teaching and learning mathematics" (Godino et al., 2020). In OSA, mathematics is defined as a human activity centered on the resolution of problem-situations that can be analyzed by considering elements required for resolution and for which global understanding requires understanding relationships among those elements. From this approach, we focus on notions of practice, object, and configuration.

A mathematical practice corresponds to "any action or expression (verbal, graphic, etc.) carried out by someone to solve mathematical problems, communicate the solution obtained to others, validate it or generalize it to other contexts and problems" (Godino & Batanero, 1994, p. 334). When a person solves a problem, he or she puts into play various sequences of practices that correspond to the meaning that the problem addresses. From these practices emerges another series of elements, the mathematical objects, "any material or immaterial entity that intervenes in the mathematical practice, supporting and regulating its realization" (Godino et al., 2020, p.6). The primary elements include situations and their accompanying problems, procedures, concepts, and their definitions, properties and propositions, arguments, and linguistic elements. The global or holistic meaning of an object then is the potential system of practices that a subject can manifest, attainable from a historical-epistemological study, and manifested in different degrees of completeness depending on the person or human creation (partial meanings).

In conclusion, we can consider the binomial distribution as a probabilistic-mathematical object that originated from the solving practices of problem-situations of a binomial nature in relation to other objects such as probabilistic principles and combinatorics.

METHOD

Our study is qualitative (Pérez, 1994) because it addresses the analysis and description of essential elements of the binomial distribution present in the answers of teachers in training—22 students of Pedagogy in Secondary Education in Mathematics and Computer Science from the Universidad de Los Lagos, Chile. These teachers in training have taken at least one course in statistics and probability that addressed the concept of binomial distribution and its associated concepts. The study also offers the possibility of obtaining numerical results by quantifying aspects of the study such as the number of teachers who identify specific elements of the binomial distribution.

For the analysis and description of the elements evidenced by the teachers, we rely on an exploratory-descriptive design (Abreu, 2012; Cazau, 2006). Exploratory research aims to examine little-studied or unknown problems, whereas descriptive research consists of data collection for analysis and synthesis to describe a phenomenon. The design allows for identification of patterns and relationships among elements. We employ the descriptive exploratory design in four phases: (a) previous work, in which the elements to be studied for understanding binomial distribution are diagnosed; (b) the design and validation of an instrument and a modality for analyzing prospective teachers' understanding of binomial distribution; (c) the description of the data collected, based on the principles pointed out in (a); and (d) the identification, synthesis, and representation of new ideas regarding the understanding of binomial distribution by teachers in training.

ABOUT THE MEANING OF THE BINOMIAL DISTRIBUTION

For our work, we consider the meaning of binomial distribution, which originated in the 18th century by Jacob Bernoulli. Bernoulli used inductive reasoning, the multiplicative principle of probability, and combinatorics to develop the formula for calculating probabilities from binomial distributions. From this historical development, and based on the notions of OSA, it was possible to reconstruct this partial meaning of the binomial distribution, that is, the resolutive practices of problem situations that can only be effectively addressed from the formula (García-García et al., 2022).

The problem-situations characteristic of this meaning are those for which there are large or infinite numbers of trials or variable values of probabilities. The definitions and concepts that characterize this meaning are related to the formal theory of probability applied to binomial distributions: the binomial distribution and its formula, probability function, parameter, probabilistic inference, degree of knowledge, mean, variance, and expected value. Relevant procedures deal with the use of the binomial distribution for analyzing different probabilistic phenomena and to approximate values such as mean and variance or compare theoretical and empirical values. In the same way, the propositions and arguments deal with the binomial distribution as an object for analyzing situations, involving propositions such as: in a binomial trial, the probability of obtaining *m* outcomes in *n* trials with a probability *p* of success is given by the binomial formula. Finally, the use of symbolic language predominates, approaching the binomial distribution from its formula and other mathematical concepts such as functions; however, tabular or graphical representations are still used, for example, to represent phenomena and compare them with the binomial distribution or to deliver specific values of interest such as cumulative binomial distributions (Fernández et al., 2022).

DESIGN OF INSTRUMENT OF STUDY

The questionnaire that was developed and validated by expert judgment and pilot application consists of a problem–situation with two questions requesting justification and considering essential elements of those identified in the historical study, as shown in Table 1.

Table 1. Problem–situation and expected answers

Problem–situation: Defective light bulbs	Expected answer
1. If it is known that the probability that a light bulb works correctly is 9 times the probability that it is defective: (a). What is the mathematical expression that represents the probability that, out of 100 light bulbs, 50 do not work? (b). Of these 100 light bulbs, how many would you expect to be defective? Give ample justification. 2. What is the maximum number of bulbs that can be manufactured if you want a high probability (greater than 0.5) that they will all work properly?	1(a). $\binom{100}{50}\left(\frac{1}{10}\right)^{50}\left(\frac{9}{10}\right)^{50}$ 1(b). Ten of the bulbs would be expected to be defective. 2. 6, because with 7 there is only a 47.8% chance that all of them will work correctly.

This problem-situation seeks to evaluate the handling of practical binomial situations with some undefined variables (or *p* other than ½) or a high number of cases. In the answers, direct use of the general formula, parameters, and the mean is sought, in addition to exploring the concepts of discrete random variable and probability that belong to an informal state of the binomial. As procedures, the calculation of parameter values, the application of the binomial distribution formula, the calculation of the mean using the formula, and the identification of the situation as binomial were considered. The rules (properties and arguments) that participants were expected to use are deductive reasoning and the use of representations to relate the binomial phenomenon to the formula and to relate expected value to the mean of the probability distribution. Regarding the language used, a tree diagram, symbolic language (arithmetic-algebraic), and tabular language are possible. Therefore, considering simple and probabilistic language as written language, we consider the possibility of identifying four types of language in the answers.

RESULTS

Addressing the more formal meaning of binomial distribution, answering the questions involves its use for analyzing and modeling binomial phenomena, once identified as such. These are presented in order of complexity, with question 1(a) solved by constructing the mathematical expression that represents the requested probability. In question 1(b), the teacher in training is asked to obtain the expected value, using probabilistic concepts associated with the binomial situation, the mean, and its formula. Because the last question involves using the formula to model a phenomenon and calculate an unknown parameter, we consider it one of the most complex addressed.

In Table 2 we present the frequencies of desired and undesired responses. Based on this, we conclude that although most of the participants were able to solve questions involving direct use of the formula of the binomial distribution or its mean (68% and 73%, respectively), almost all of them were not able to use the binomial distribution and its principles to model the phenomenon of the last question. In other words, there are difficulties in using the binomial distribution to answer questions beyond the calculation of probabilities or the formulas associated with it.

Table 2. Frequencies of answers given by the participants to the problem-situation

Question	Desired answer	Unwanted answer/no answer
1(a)	15 (68%)	7 (32%)
1(b)	16 (73%)	6 (27%)
2	2 (9%)	20 (91%)

Regarding the essential elements proposed (Table 3), most of them are present (P) in most of the responses with no or few errors (PwE). However, it is possible to identify that the elements that present greater errors (PwE) or that are absent (A) are those associated with the formal binomial distribution: parameter and mean.

Table 3. Concept–definitions identified in the answers given by the participants

Concept–definition	P	PwE	A
Discrete random variable	17 (77%)	0	2 (9%)
Probability	20 (91%)	0	2 (9%)
Binomial distribution formula	18 (82%)	2 (9%)	2 (9%)
Parameter	19 (86%)	1 (5%)	2 (9%)
Mean (expected value)	15 (68%)	2 (9%)	5 (23%)

Regarding the procedures involved (Table 4), the incorrect use of parameters will mean an erroneous application of the formula, the only procedure in which conflicts were identified (14%). Regarding the properties (Table 5), we identified that most of the participants explicitly or implicitly handle characteristics of the binomial phenomenon, probabilistic principles, and the relationship between the formula of the distribution and its probabilities. On the other hand, the entire group does not address the relationship between the mean and expectation, even though in some cases the calculation is performed. This means that the participants consider the mean of a binomial distribution and its expected value as equivalent terms, even though the former is a measure that can be calculated from the formula and the latter depends on the context being studied.

Table 4. Frequencies of procedures identified in the answers given by participants

Procedure	P	PwE	A
Identifying binomial situation	20 (91%)	0	2 (9%)
Calculation of parameter values	18 (82%)	0	4 (18%)
Application of binomial distribution formula	16 (73%)	3 (14%)	3 (14%)
Calculation of the mean ($\mu = np$)	14 (64%)	0	8 (36%)

Table 5. Frequencies of properties–propositions in the answers given by participants

Properties–propositions	P	PwE	A
Additive Principle	12 (55%)	1 (5%)	9 (41%)
Multiplicative Principle	19 (86%)	1 (5%)	2 (9%)
There are n observations	21 (95%)	0	1 (5%)
The observations are independent	20 (91%)	0	2 (9%)
There are only 2 outcomes: success and failure	21 (95%)	0	1 (5%)
The probabilities p and q are constant	20 (91%)	0	2 (9%)
Relationship between binomial phenomenon and formula	16 (73%)	3 (14%)	3 (14%)
Relationship between expectation and mean	0	0	22 (100%)

Despite reaching the correct answer in some questions and evidencing primary objects of different meanings, it was not possible to identify the presence of justification statements (arguments).

Likewise, despite applying the formula and calculating the mean, there were no instances in which reference was made to theory or reasons why the properties could be applied, that is, deductive reasoning without statements and, therefore, considered as absent (see Table 6).

Table 6. Frequency of arguments in the answers given by the participants

Arguments	P	PwE	A
Deductive reasoning	0	0	22 (100%)
Use of representations	0	0	22 (100%)

Finally, there is no evidence of the use of language other than the written and symbolic, which are associated with the fact that the problem situations include very large values and, therefore, are difficult to represent or simply required the application of the formula (see Table 7).

Table 7. Frequencies of language in the answers given by the participants

Language	P	A
Written	21 (95%)	1 (5%)
Symbolic	20 (91%)	22 (9%)
Tabular	0	22 (100%)
Graphic	0	22 (100%)

DISCUSION AND CONCLUSION

Our work consisted of analyzing high school mathematics teachers in training's understanding of the meaning of the binomial distribution, following the current research that evidences it as one of the components of probabilistic literacy.

As preliminary work, we approached one of the partial meanings of the binomial distribution, taking as a basis the theoretical-methodological notions of the OSA, by identifying the primary objects originated from the problem situations associated with the binomial distribution throughout history. After these elements were identified and categorized, they were used for the design, validation, and application of a questionnaire that allowed us to explore and analyze their presence in the answers of teachers in training and the level of understanding of the binomial distribution from a historical-epistemological perspective.

In the results, it is evident that the teachers in training present correct handling of concepts such as binomial distribution formula, parameter, and mean. In addition, most of them explicitly or implicitly manifest the resolution procedures. However, even though most of the answers show properties associated with identifying the binomial phenomenon and applying the binomial formula, there are no propositions that address the relationship between expectation of a modelable phenomenon and the mean of its probability distribution, and, even worse, no argumentation. Furthermore, in the problem situation addressed, most of the teachers in training were able to answer the questions involving the correct use of parameters in the formula of the binomial distribution and the direct calculation of its mean knowing n and p. However, only 9% of the teachers managed to answer the last question correctly, which involves use of the binomial distribution or principles on which it is built, to identify the value of n necessary to arrive at a certain probability. The latter question, which involves the binomial distribution with other mathematical and probabilistic objects such as the equation and the functional meaning of probability, requires a complete understanding of the principles of the former, which is already evident as not achieved if we consider the absence of arguments or alternative representations.

The promotion of these components in the design of learning experiences focused on understanding and critical thinking would facilitate reaching a higher level of understanding of the meaning of the binomial distribution. In particular, we are greatly interested in designing and implementing learning experiences in which the binomial distribution mean is related to expectation of a probabilistic phenomenon from the traditional formula and its reconstruction from the additive and multiplicative principles of probability. The analysis of real-life situations (such as those with monetary value) would also allow generating representations not identified in this study (tabular and graphical) and arguments, which are essential components in teaching. For this, we consider it necessary to improve our questionnaire, to apply it to other groups of people, and to address the handling of other essential components of the binomial distribution such as combinatorics in order to explore whether the reported problems are caused by conflicts with previous concepts. This would allow us to solve the limitation of the number of participants and to identify resolving patterns that associate understanding of the binomial distribution with other notions of mathematics and probability theory, necessary to design a longitudinal proposal for learning from a historical-epistemological perspective and extending it to the study of other concepts such as the law of large numbers and the normal distribution with the support of technology.

REFERENCES

Abreu, J. (2012). Hipótesis, método & diseño de investigación [Hypothesis, method & research design]. *Daena: International Journal of Good Conscience*, *7*(2), 187–197.

Alvarado, H., & Batanero, C. (2007). Dificultades de comprensión de la aproximación normal a la distribución binomial [Difficulties in understanding the normal approximation to the binomial distribution]. *Números: Revista de Didáctica de las Matemáticas*, *67*, 1–7.

Anacona, M. (2003). La historia de las matemáticas en la educación matemática [The history of mathematics in mathematics education]. *Revista Ema*, *8*(1), 30–46.

Batanero, C. (2005). Significados de la probabilidad en la educación secundaria [Meanings of probability in secondary education]. *Revista Latinoamericana de Investigación en Matemática Educativa, RELIME*, *8*(3), 247–263.

Batanero, C. (2020). Probability teaching and learning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 682 –686). Springer. https://doi.org/10.1007/978-3-030-15789-0_128

Cazau, P. (2006). *Introducción a la investigación en ciencias sociale*s. Buenos Aires.

Fernández, N., García-García, J., Arredondo, E. H., & Araya, I. (2022). Epistemic configurations and holistic meaning of the binomial distribution. *Mathematics*, *10*(10), 1748. https://doi.org/10.3390/math10101748

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school* (pp. 39–63). Springer. https://doi.org/10.1007/0-387-24530-8_3

García-García, J., Fernández, N., Pino-Fan, L., Arredondo, E. H., & Imilpán, I. (2022). The binomial distribution: Historical origin and evolution of its problem situations. *Mathematics*, *10*(15), 2680. https://doi.org/10.3390/math10152680

García-García, J., & Sánchez, E. (2015). El desarrollo del razonamiento probabilístico de estudiantes de bachillerato sobre la noción de la distribución binomial [The development of the probabilistic reasoning of high school students on the notion of the binomial distribution]. In *Memorias del III Coloquio de Doctorado del Departamento de Matemática Educativa* (pp. 1–13). Cinvestav.

Godino, J. D., & Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos [Institutional and personal meaning of mathematical objects]. *Recherches en didactique des Mathématiques*, *14*(3), 325-355.

Godino, J. D., Font, V., & Batanero, C. (2020). El enfoque ontosemiótico: Implicaciones sobre el carácter prescriptivo de la didáctica [The ontosemiotic approach: Implications on the prescriptive nature of didactics]. *Revista Chilena de Educación Matemática*, *12*(2), 47–59. https://doi.org/10.46219/rechiem.v12i2.25

Gordillo, W., & Pino-Fan, L. R. (2016). Una propuesta de reconstrucción del significado holístico de la antiderivada [A proposal to reconstruct the holistic meaning of the antiderivative]. *Bolema*, *30*(55), 535–558. https://doi.org/10.1590/1980-4415v30n55a12

Pérez, G. (1994). *Investigación cualitativa. Retos e* interrogantes [Qualitative research. Challenges and questions]. Madrid.

Sánchez, E., & Carrasco, G. (2018). El razonamiento probabilistico de estudiantes de bachillerato en actividades de distribución binomial [The probabilistic reasoning of high school students in binomial distribution activities]. In L. J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, A. Aguilar-González, P. Alonso, F. J. García & A. Bruno (Eds.), *Investigación en Educación Matemática XXII* (pp. 535–543). SEIEM.