Binomial distribution at high school: An analysis based on learning trajectory

I Taufiq1, F Sulistyowati1 and A Usman2

1Pendidikan Matematika, Universitas Sarjanawiyata Tamansiswa Yogyakarta, Jl.
Kusumanegara No. 157, Yogyakarta 55165, Indonesia
2SMAIT Abu Bakar Boarding School Kulon Progo, Jl. Nyi Ageng Serang, Triharjo,
Wates 55651, Indonesia

*Corresponding author’s e-mail : irham.taufiq@ustjogja.ac.id

Abstract. The binomial distribution was introduced in mathematics learning in high school. The problem is that literacy does not optimize the concept of binomial distribution itself. This research is a qualitative descriptive that aims to describe the problem of binomial distribution in a certain high school mathematics text book. Data were collected by the documentation method and analyzed descriptively. It was found several problems, such as (1) unstructured discussion schemes and some important concepts are not given; (2) the statistical description of the binomial distribution is not detailed (mean and variance). The results of this study are expected to become a reference for book writers in developing a book that can optimize conceptual understanding, especially in a binomial distribution.

1. Introduction

The binomial distribution is one of the discrete probability distributions that the application is easy to found in everyday life. For example, in a headman selection, a candidate will have two possibilities, that is failing as the headman or being selected as the headman. The basic competencies (BC) that must be held by students related to binomial distribution include: (BC.3.5.) Explaining and determining binomial probability distribution related to binomial probability functions; (BC.4.5.) Problem-solving was related to the binomial probability distribution from an experiment (random) and its conclusions. These basic competencies BC will be reached with learning that according to the applicable curriculum and supported by good literacy.

One of the learning principles in Indonesia is learning based on various learning literacy. Books are learning literacy provided by schools and teachers often use them in learning. On the other hand, learning is expected to be based on disclosure/research (discovery/inquiry learning). Therefore, books as learning literacy must be able to support inquiry learning so that it can achieve learning goals. Inquiry-based learning emphasizes the activeness and responsibility of students in finding new knowledge[1]. It can be said that inquiry-based learning as the process of finding a new causal relation through the formulation of hypotheses and testing in experiments case or in-depth observations[2]. In this process, students carry out self-regulated learning as an inductive process and conduct experiments as a deductive process by investigating the relations between the dependent and independent variables[3]. Inquiry learning as learning that gives experience to students to discover and organize new knowledge by the process of experimentation or observation.
New knowledge may be concepts, characteristics or ways to solve problems. If examined further, (BC.3.5) emphasizes the concept, while (BC.4.5) emphasizes more on how to solve the problem by applying some characteristics of the binomial distribution. These two basic competences certainly affected each other, that is (BC.4.5) will not be achieved if students not held (BC.3.5). Therefore, it is necessary to optimize the concept to achieve (BC.3.5), one of which is a learning literacy (in this case is books) that can support teachers in optimizing the concept.

Several math books have been tested at senior high school. One of the books is a book that can be bought by government funds. So, the book will be used by almost all high schools in Indonesia, because most senior high schools in Indonesia do not give burden students to buy books as a learning literacy. The effort certainly has a good goal, but good goals should be complemented by good quality books, such as books that can support inquiry learning by optimizing concepts.

Related to the binomial distribution, of course, problems will occur when students only use one book while the book still less to optimizing the concept of the binomial distribution. As a result (BC.4.5) it is not reached optimally because students did not hold (BC.3.5). For example, when students are asked about "what is a binomial distribution?" They will answer:

\[
\text{"binomial distribution is } b(n; p, x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, ..., n \text{".}
\]

There are not false, but the definition of the answer is not right. The answer shows that students do not understand what the probability distribution is and what is Binomial. This is certainly a new problem which will provide another more complex problem in binomial distribution learning. Therefore, it is necessary to avoid these problems, one of which is by analyzing discussion errors on the binomial distribution at senior high school mathematics books based on optimizing concepts in the learning trajectory. The results of the analysis are expected to be used as a suggestion in the preparation of creating the discussion of the binomial distribution.

2. Methods
This research is a qualitative descriptive that aims to describe the problem of binomial distribution in one of the high school mathematics books based on the learning trajectory in optimizing conceptual understanding. The subjects in this study were the binomial distribution in one of the high school mathematics books. In qualitative research, data takes the form of words or pictures rather than numbers. Usually, the descriptive data contains quotations said by informants to describe and support the presenting findings. Data can include field notes, photographs, transcripts, video recordings, memos, audio recordings, and personal documents[4]. Data collected by the documentation method and analyzed descriptively. The data used in this research is a document that includes theories related to conceptual understanding, binomial distribution and learning trajectory. Also, some research can be optimizing and support the results of the analysis.

3. Result and Discussion
Based on the analysis that has been done, there are some errors in the binomial distribution in the book. The errors are based on the analysis of the learning trajectory in conceptual understanding of the binomial distribution and their statistics (mean and variance).

3.1. The learning trajectory of the binomial distribution
The learning trajectory can help us answer some questions, such as: (1) what goals should we set? (2) where will we start learning? (3) how do we know where to go next? (4) how do we get there? On the other hand, the learning trajectory has three parts: a) objectives; b) the path through which students develop to achieve that goal; and c) several teaching activities, or assignments, which are matched to their level of thinking.[5]

The expected learning outcomes (mathematic goals) are students can hold (BC.3.5) and (BC.4.5). In the process of achieving these competencies, activities are needed that give a chance for students to
develop the conceptual understanding, characteristics and solving binomial distribution problem abilities. Books as a learning literacy should be a supporter of these activities to reach competence (BC.3.5) and (BC.4.5). The problem is that the books that have been analyzed show that the scheme of the discussion of the binomial distribution is less structured. Why does it happen? Let’s discuss it.

The scheme of the binomial distribution discussion in the book analyzed is inferential statistics which are divided into two plots, such as (A1) sample and distribution function; (A2) the hypothesis test of the binomial distribution. Plot (A1) in the scheme, beginning with the discussion of random variables, the function of probability, the binomial experiment to its function of probability. If examined more closely, some basic concepts are not given in plot (A1), such as (1) discrete and continuous random variables; and (2) discrete and continuous probability distribution. The concept needs to be explained in advance so that students are easier to understand the binomial distribution which is a special discrete probability distribution. The binomial distribution is explained after optimizing the concept of probability, random variables and their distribution both discrete and continuous [6,7]. In senior high school, the discussion of probability has been taught in the previous chapter, while the random variable and its distribution both discrete and continuous have not been discussed in detail. This makes students only understand about binomial distribution without knowing that actually, the binomial distribution is one of the special discrete probability distributions. As a result, when students are asked the question "what is a binomial distribution?" They are likely to say that "binomial distribution is an experiment that has two possibilities, were failure and success" or according to the previous explanation that:

\[
b(n; p, x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, ..., n
\]

Certainly, it is not false, but the meaning of the distribution itself does not reach the students. Therefore, the plot (A1) at P1 needs to be preceded by a discrete and continuous probability distribution concept. In addition to the problems in the plot (A1), there is a problem with the plot (A2) is P2. Plot (A2) in the scheme, divided into several sections, namely (1) population and sample; (2) level of significance; (3) binomial distribution; (4) decision. In the plot (A2) there are not too many errors when viewed from the scheme of the discussion, which is related to the structure of sub-chapters that need to be rearranged.

Based on several problems that have been explained, the learning trajectory is designed for binomial distribution learning that is adapted with (BC.3.5) and can be seen in Figure 1. Figure 1 shows that the learning trajectory can be used by the teacher to set a learning activity in binomial distribution learning. While for students it can be used as a reference in studying the binomial distribution. At least, the learning trajectory clarifies the discussion scheme to be taught and does not miss some important concepts that should be given to students.

3.2. Conceptual understanding of the binomial distribution learning
The concept is said to be an abstract idea that is generalized based on a special event [8]. On another hand said that the concept of knowledge is conceptual knowledge itself [9–12]. This knowledge is usually not related to a specific problem and can be either implicit or explicit. Therefore, this knowledge is not required to be delivered verbally [13]. Meanwhile, the National Research Council defines conceptual knowledge of mathematics as an understanding of mathematical concepts, operations, and relationships based on a review of the mathematics education research literature [14]. Sometimes this knowledge is called conceptual understanding of principled knowledge. Conceptual understanding can be said as an ability to understand abstract ideas as a result of generalizing a particular activity. Therefore, activities that are necessary to generalize events that arise to obtain conceptual understanding capabilities are needed. In binomial distribution learning, it can be a real experiment or a problem based on real life.
Problem 1. lack of discussion related to the probability distribution. The distribution function of the random variable $X$ written $F(x)$, is defined as the probability function of the random variable $X$ less than $x$ and written with $P(X < x)$. The function of $F(x)$ is given by the formula $F(x) = \sum_{x_i < x} p_i$ in which the summation is extended over all values of $i$ such that $x_i < x$[16]. $F(x)$ of a discrete random variable $X$ with probability distribution $f(x)$ is $F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i[8,9]$. On another hand, in the book, it is explained that probability distribution is a list of probability values of each random variable. If we look further, the definition in the book is not adapted to the definitions previously. Maybe the meaning can be said the same, but not all students can hold that meaning. Submission using different sentences can give different meanings if interpreted by different people. A sentence can have several meanings which are interpreted based on the word components that make up the sentence. Furthermore, sentences may have more than one literal meaning or cannot be interpreted. The literal meaning of a sentence needs to be clarified its meaning by the speaker because it can deviate from its true meaning if it is written in sentences that have many or empty meanings [15]. Errors in hold the meaning of a sentence make a mistake in understanding the concept that has been given. In the case of the definition of distribution, maybe the students understand the probability distribution as a list of probability values of a random variable. When students are asked, "what is a binomial distribution?" They will answer "binomial distribution is a list of probability values of a random variable". If I asked again "what is a binomial random variable?" Students will have difficulty answering because, before the discussion, students understand random variables as variables whose values are determined based
on the results of the experiment. While other books say that a random variable is a function that pairs real numbers with each member in the sample space[6].

Generally, a mistake in the explanation of a definition can affect the concepts that will be accepted by students. Therefore, books as learning literacy should be able to choose sentences that are easy to understand but also not deviate from the concept. At least the explanation based on literacy that can be responsible. In this case, it is recommended to use a literacy that is already used in other countries or are often used at another college in Indonesia. The concepts in the book can be explained in clear and simple language in high school so that students will more easily understand without deviating far from the actual concept.

**Problem 2.** The lack of discussion on statistics of the binomial distribution. In the books analyzed, the statistics discussed are mean and variance. In the book, it is said that if \( x \) is a random variable and \( f(x) \) is a probability for each of these variables, then the mean \( x \) (denoted by \( \mu \)) can be calculated as follows:

\[
\mu = \sum_{i=1}^{n} x_i f(x_i) = x_1 f(x_1) + x_2 f(x_2) + x_3 f(x_3) + \ldots + x_n f(x_n) \quad \text{(i)}
\]

There were no errors to found the mean value of \( x \). But so far students understand that to find the mean is calculated by the formula:

\[
\text{Mean} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} \quad \text{(ii)}
\]

So, while the students are given an equation (i) students will confusion, because it is different from equation (ii). Why is this a problem? Let's discuss it. Reasonably if students are asked to calculate for the mean of numbers 20, 10 and 20, they will an answer \( \text{mean} = (20+10+20)/3 = 16.66 \). In this case, students find for the mean using equation (ii), and clearly cannot be solved using equation (i) because the form of function \( f(x) \) is unclear. Therefore, to understand equation (i) requires equation (ii) as a concept link that students have with the concept that will be accepted by students. As reasoning, the following will be given one of the cases that can connect the concept of mean in equation (ii) with the concept of mean in equation (i).

Figure 2 shows that 4/16, 7/16, and 5/16 are fractions of the number of heads that appear from the total tosses, as many as 0, 1, and 2 heads. In this experiment, the fractions are relative frequencies for the different values of \( X \). Therefore, if 4/16, or 1/4, of the tosses, result in no heads, 7/16 one head, and 5/16 two heads, the mean number of heads per toss would be 1.06 no matter whether the total number of tosses were a thousand tosses. This result means that we toss 2 coins over and over again until a thousand tosses, we will get the average 1 head per toss.

The method described above suggests that the mean, or expected value, of any discrete random variable, may be obtained by multiplying each of the values \( x_1, x_2, x_3, \ldots, x_n \) of the random variable \( X \) by its corresponding probability \( f(x_1), f(x_2), f(x_3), \ldots, f(x_n) \). And then, we can be summing the products. So, the mean, or expected value, of \( X \) can be written as:

\[
\mu = \sum_{i=1}^{n} x_i f(x_i) \quad \text{(iii)}
\]

Conceptual understanding to find the mean by using the case in Figure 2 will be more acceptable because the reasoning is easy to hold. Related to the binomial distribution, in the book it is explained that to find the mean in binomial distribution is formulated as follows:
If two coins are tossed 16 times and \( X \) is the number of heads that occur per toss, then the values of \( X \) are 0, 1, and 2. Suppose that the experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively. The average number of heads per toss of the two coins is then
\[
\frac{0(4)+1(7)+2(5)}{16}
\]
Let us now restructure our computation for the average number of heads so as to have the following equivalent form:
\[
0\left(\frac{4}{16}\right) + 1\left(\frac{7}{16}\right) + 2\left(\frac{5}{16}\right)
\]

**Figure 2.** The case of the mean concept link

\[
\mu = np \quad \text{(iv)}
\]

with \( n \) stated the sum of data while \( p \) states the probability of success of the random variable. The formula is not false, but for students, the formula is certainly still not acceptable. How can the mean distribution that can be searched by equation (iii) change to equation (iv)? Therefore, the reasoning is needed in finding equation (iv) by using equation (iii) which exists in Figure 3.

\[
\begin{align*}
\mu &= \sum_{x=0}^{n} x f(x) = \sum_{x=0}^{n} x \left(\frac{n!}{x!(n-x)!}\right) p^x q^{n-x} \\
&= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=0}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} = \sum_{x=0}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \\
&= np \sum_{x=0}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} = np \sum_{x=0}^{n} \frac{(n-1)!}{(n-1)!(x-1)!} p^{x-1} q^{n-(x-1)} \\
&= np \sum_{x=0}^{n} (n-1) p^{x-1} q^{n-(x-1)} = np(p + q)^{n-1} = np(p + (1-p))^{n-1} = np
\end{align*}
\]

**Figure 3.** How to find the mean of the binomial distribution

Figure 3 shows how to find the mean of binomial distribution through the mean concept of the probability distribution. Obtaining mean with these steps will be more easily accepted by students. Also, the concept can last longer in the cognitive structure of students because of the experience of finding equations (iv). Variances and standard deviations in the book are also given like the mean of the binomial distribution. This of course also needs to be corrected according to the previous case.

4. **Conclusion**

Based on the analysis that has been done, there are several problems such as (1) the scheme of the binomial distribution is less structured and there are several important concepts that are not given; (2) The discussion related statistics on binomial distribution is less detailed (in this case mean and variance). Therefore, there needs to be further action related to the problem, for example, revisions to some discussions or teachers as facilitators can add some things that can reduce the lack of the book.

5. **References**

[1] De Jong T and Van Joolingen W R 1998 *Review of Educational Research* 68
Acknowledgments

Thank you to all parties who supported the completion of this research, such as parents, family, friends, and colleagues. Also, thank you to ICMScE 2019 for allowing the author to present the results of this research. Without the support of these parties, of course, this research cannot be completed optimally. This research is still not perfect so that it needs to be improved. This is left to the reader to do further research.