An analysis of difficulties in learning Probability in high school

SHAO Xiayan, East China Normal University, China

Introduction

Probability is a study of the rules of random phenomenon. It provides methods of problem solving and a thinking model for people to know the objective world. Probability also offers the foundational theory for the development of statistics. Thus, probability is an important part of high school mathematics, and it reflects the applicability of mathematics. Although, currently, the content of probability taught in high school is very simple, yet it provides high school students with a good platform to know the applicability of mathematics, and lays the foundation for their future study.

Analysis framework

This paper analyzes the concept of probability from three aspects: the logical structure, 、 the historical development and the cognition process. Logical process is based on the knowledge structure of mathematics: it analyses the definition of probability, and the status and role of probability in the conceptual system. Historical process uses historical analysis methods: this paper expounds the ancient games and the scientific background of probability. Psychological process is based on literature analysis: the research on the concept map of probability and the difficulties and obstacles on understanding the concept of probability.

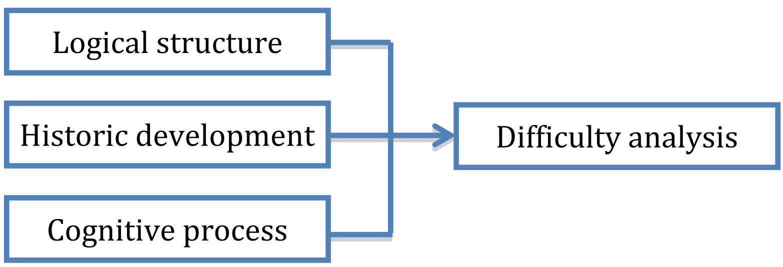

Figure 1. Analysis framework

Analysis of the difficulty of concepts

Teaching the concept of probability is difficult because of the high degree of abstractness and generality, as well as teachers' insufficient theoretical studies and inexperience.

History of the formation of probability

There are interesting examples of ancient games involving probability, or example, the Egyptian game "Hounds and Jackals." The famous Greek historian, Herodotus, wrote in his great work *History*: early in 1500 BC, that in order to forget their hunger, the Egyptians often played dice and a board game called "Hounds and Jackals," which moves pieces on the board according to the throw and based on certain rules. From 1200

BC, people made the cubic dice from natural bone. They ground the bone into a rough cube, forming the six surfaces of the dice, and engraved different numbers on each surface of the dice. Throwing the dice is the random generator used frequently in the game. The dice used nowadays, with two partitions of seven, imprinted on opposite faces, was made in Egypt in 1400 BC.

The Italian Mathematician, Fra Luca Pacioli, published a book about arithmetic skills in 1494. It described the following problem in probability: A team plays ball so that a total of 60 points is required to win the game and the stakes are 22 ducats. By some incident, they cannot finish the game and one side has 50 points, and the other 30.What share of the prize money belongs to each side? Fra Luca's response to the "problem of points" is that the stakes should be divided in the proportion 5:3. At that time, many people didn't think Fra Luca's distribution fair, because the player who had 50 points just needed 10 points to win, and the player who had 30 points needed to win 30 points more, which is much more difficult. Nevertheless, people still can't find a better solution. After 100 years, there were several mathematicians who studied the problem successively, but no one got the right answer.

Scientific background of the emergence of probability

The most important concepts of probability theory are independence and randomness. Probability theory started from the study of classical probability, the research objects are masses of independent random processes. Through problem solving using the processes of independence and randomness, probability theory was established.

The emergence of independent random processes was not the decisive factor in the development of probability theory. Probability theory is a branch of mathematics, which utilizes the principles of addition and multiplication (combinatorial mathematics principle) to do pure digit arithmetic during the calculation process. To people today, the calculation of probability is not a difficult thing. However, for people before the 16th century, calculation was very difficult due to the lack of a simple calculation system. Arab numbers can easily express the fraction and the decimal. This allowed the clear expression of probability in the form of symbols and allowed the gradual development of the theoretical system.

The formation of probability theory also needs inductive thought. Probability uses induction as the basic research method, drawing conclusions from many observed patterns, but also establishing a methodology and theory based on these observations. The formation of probability theory became possible only due to the premise that induction was been widely used.

Probability theory is the inevitable result of a large amount of accidental factors. Firstly, gambling, an opportunity game, provides a good independent stochastic process. Secondly, advanced counting systems removed hindrances to the development of probability. New methods and new ways of thinking were necessary for the emergence of probability.

The logical process of the formation of probability

In the Shanghai Education Press, 2008 edition of teaching materials for the first semester of grade 12, classical probability is defined as follows:

On a randomized trial the number of all basic events is finite, and the likelihood of each basic event appears equal, the probabilistic model with these two characteristics is called classical probability (page 85). This definition is substantially similar to the 2007 People Education Press Edition. It offers the following definition:

The probability of event A is defined as:

$$P(A)=\frac{\text{The number of basic events included in event A}}{\text{The number of all the basic events included in the trial}}$$

A third definition is offered in the Dictionary of Mathematics:

Classical probability randomized trials have the following characteristics: (a) Sample space contains only a finite number of sample points; (b) The probability of each sample point appears equally.

In this case for each event $A\subset\Omega$, preferably $P(\mathrm{A})=\frac{n(A)}{n(\Omega)}$, we call this $P(\mathrm{A})$ the classical probability of event A. $n(A)$ here represents the number of elements in finite set A.

This definition is difficult to translate into Chinese. The meanings of "sample space," "sample points" and other words seem to be more difficult to understand. Table 1 shows a comparison of the three definitions

Table 1. Brief comparison of the three definitions

<table>
<tr><th rowspan="2"></th><th colspan="2">Mathematical concepts involved</th><th colspan="2">Mathematical symbols involved</th><th rowspan="2">Chinese words</th></tr>
<tr><th>Common</th><th>Difference</th><th>Common</th><th>Difference</th></tr>
<tr><td>Shanghai Education Press Edition</td><td rowspan="3">Limited, equal possibility</td><td rowspan="2">Basic events</td><td rowspan="3">The nature of the formula</td><td rowspan="2">$P(A)=\frac{\text{The number of basic events included in event A}}{\text{The number of all the basic events included in the trial}}$</td><td rowspan="2">89</td></tr>
<tr><td>People Education Press Edition</td></tr>
<tr><td>Dictionary of Mathematics</td><td>Sample points</td><td>$A\subset\Omega$
$P(\mathrm{A})=\frac{n(A)}{n(\Omega)}$</td><td>104</td></tr>
</table>

Comparing the three different definitions we find that:

(a) Three definitions are involved in the expression of a large number of mathematical concepts and mathematical symbols, and they are composed of multiple clauses and complex sentences. Definitions include at least four concepts: randomized trials, the basic events, finite numbers, and equal possibilities.

(b) The definition of Shanghai Education Press Edition and People Education Press Edition adopt the statements “basic event of event A” and “all the basic events of the trial.” These are in accordance with the students' ability to understand. The Dictionary of Mathematics uses mathematical language and abstract symbols to express the definition. The statement is rigorous and specific, but students have great difficulties in understanding.

In Shanghai Education Press, 2008 edition of teaching materials in the first semester of grade 12, page 91-92, statistical probability is defined as follows:

Statistical probability: If random event E occurs m times in n trials ($0 \leq m \leq n$), then number m is called the frequency of event E, while $\frac{m}{n}$ is called the probability of event E occurring: the number of times the event occurs in the experimental trials.

The 2007 People Education Press Edition defined probability as:

Statistical probability: In n times repeated trials, the frequency of event A is $\frac{m}{n}$, n is the total number of trials. Thus, when n is large, $\frac{m}{n}$ approaches stability.

Comparing the two different definitions, we can find that the definition is basically the same, but the People Education Press Edition definition emphasizes, "n is large."

The definition of probability depends on the definition of specific words. Some of their meanings are the same as used in everyday speech. However, significantly some words are mathematical terminologies and have special mathematical meanings. The following examples demonstrate this:

(1) Random events

A random phenomenon is a phenomenon with statistical regularity, which is to be or not to be under certain conditions. The outcome of the experimental trial is called random events, abbreviated to events.

(2) Fundamental events

The possible outcomes of a trial are Fundamental events.

(3) Equal possibility

The possibility of every event is the same.

The analysis of knowledge structure

For senior high school, there are some subordinate concepts taught in the theory of possibility (Figure 2).

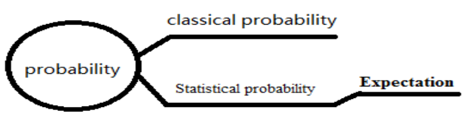

Figure 2. Sub-concept of probability

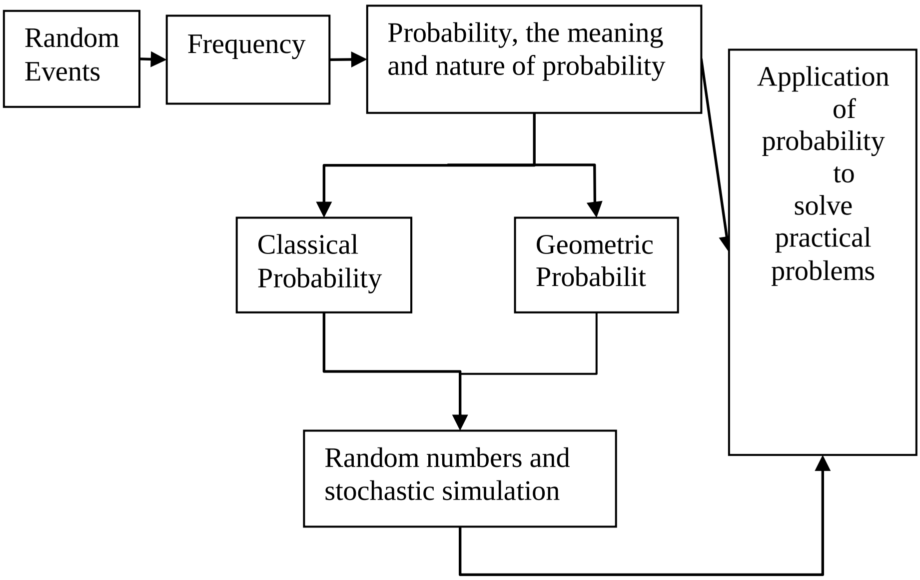

Figure 3. Knowledge structure of probability

In high school, the knowledge structure of probability is shown in Figure 3. The high school curriculum requires students to: understand the meaning of probability, and the difference between frequency and probability, and understand the uncertainty of random events and the stability of frequency. During the study, understanding the significance of the difference between frequency and probability is foundational for further learning of probability.

The psychology process of the formation of concepts

The use of concept maps is a teaching skill devised in the 1960s by the American doctor J. D. Novak, based on David P. Ausubel's learning theory. They establish concepts and relations between the concepts through net structures. The concept map of probability is as follows:

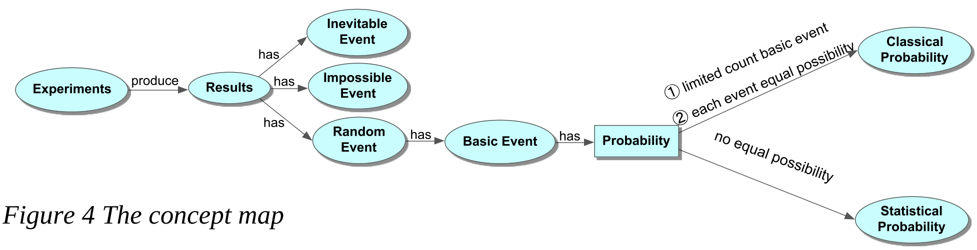

Figure 4 The concept map

Understanding of the statistical definition of probability

According to the statistical definition of probability, the frequency of the random event fluctuates around a certain constant. With the increase of the experiment times, the amplitude of the fluctuation gets smaller and smaller, and then we call the constant the probability of the random event.

Li (2012) divided the errors of students' understanding of the statistical definition of probability into four categories:

1. The probability is the average value obtained from the frequency of the experiment;
2. The probability is the approximate value of frequency;
3. The probability is the limiting value of frequency;
4. The probability is the frequency value of certain large number experiments, that is when the experiment time is large enough, the frequency slowly nears constant, until it reaches the final value, then this constant is the probability.

The studies of Fu (2014), Zhou (2012), Liu (2011), and Wang (2009) also arrived at similar conclusions. With regard to the reason for these common errors, Li (2012) indicated that it is difficult for the students to master the "law of great numbers," some of them don't trust the rules, some don't believe in the experiments, and with the limitation of teaching, the students find it hard to comprehend abstracts or directly observe the stability of the probability, they have no choice but accept the concept of probability passively. Fu (2014) argues that some teachers do not take account of the knowledge formation of the students, difficult to observe, and find the constant rules changing frequently if the students do not do experiments by themselves. Many teachers do not pay attention to the precision of classroom instruction; they mislead the students in class.

Understanding of classical probability

(a) The sample space is a basic and important concept in probability theory. It is a set of all possible outcomes of random events, and is the base of calculation using classical probability. The study of Yuan (2011) showed Shanghai nine grade students' understanding of sample space, but there is no domestic study of teaching and learning sample space in high school.

When students learn sample space, three major errors occur:

The equi-probability bias when students list all the possible outcomes. They think some different outcomes are the same;

Students cannot use the exact words to describe the sample space;

Students make mistakes when using a sequence array. When students use a sequence array to list the sample space, there are two types of errors: students cannot list the correct sequence and there are omissions.

(b) Misconceptions in Chance Interpretation

Equi-probability means that each of the possible outcomes of an experiment has an equal probability of occurring. Students think that all the outcomes are equi-probable, and have nothing to do with the objective conditions.

The Outcome Approach: some students do not think that frequentist information obtained from a series of repetitions can be used for estimating the probability for a single trial. In other words, they do not have the understanding of interpreting probability in a frequentist way.

(c) Misconceptions in Chance Comparison
Comparison in One-Stage Experiments: Besides the equi-probability bias discussed previously, students who have not received instruction in probability often use self-designed rules when they are required to compare likelihood.

(d) Chance Comparison in Two-Stage Experiments: one main misconception in two-stage experiments is to split multi-stage tasks into several independent one-stage tasks. This is defined as the compound approach by Li (2000). The compound approach is a misconception used in measuring and comparing probability in multi-stage experiments. It splits multi-stage experiments into several one-stage experiments and interprets the likelihood of a multi-stage event as just a combination of the likelihood of its elements.

Understanding the definition of the geometric probability
After the new curriculum reform, high school mathematics added the content of geometric probability. People Education Press Edition's definition is that if the probability of every event is only proportional to the length (surface or volume) of the region containing the event, then we can call this kind of probability model as geometric models of probability, shortened as geometric probability.

Zhou (2012) investigated 1176 students of grade 11, from a key high school of Daqing city, and found that the typical errors existing in the study of the geometric probability included:

1. The students thought that if the probability satisfied the two conditions "basic event is infinite, [and] every case has the equal possibility," then it must be geometric probability, and they don't take into consideration that the whole basic event must be composed out of a region which has length, surface or volume;
2. The students did not clearly analyze the basic event from the question and the region formed by the basic events.

Conclusion
According to the analysis above, it is difficult for some students to master "the law of large numbers" and some teachers are not precise when they introduce probability in the classroom. These reasons cause students to have difficulties in accepting the statistical definition of probability. On the understanding of the classical definition of probability, the equi-probability bias is a common misconception.

Based on our analysis, the implications for teaching probability are:

1. Ignoring the process of the formation of students' knowledge in teaching causes a lack of a deep understanding of the relationship between probability and frequency. Teachers can use multimedia technology or provide rich real life examples to help students understand the conception of probability.
2. Because of the lack of understanding of the historical background and the practical application of probability, students are not interested in learning probability. Introducing some history and culture of probability in the classroom can arouse students' learning interest, and thus deepen their understanding of probability.
3. Instructional designs based on our analysis will be developed in our later study.

References

DING, H. F. S. (2009). Review on the study of psychology of learning in probability and statistic. *Journal of Chongqing University of Arts and Sciences (Natural Science Edition), Vol.28* No.2, 92-96.

FU, M. (2014). Research on High School Students' Understanding of the Concept about Probability Based on the New Curriculum (In Chinese). Unpublished thesis (M. ed). Capital Normal University, Beijing.

GUO, B. (2006). Analyses on common mistakes of middle school students' probability studying. Unpublished thesis (M. ed). Northeast Normal University, Jilin.

LI, J. (2000). Chinese Students' Understanding of Probability. Ph.D. thesis. School of Science Nanyang Technological University.

LI, Y. (2012). The Research and Analysis of High-school Students' Probability Learning (In Chinese). Unpublished thesis (M. ed). Capital Normal University, Beijing.

LIU, G. (2011). The Error Causes Analysis and Countermeasures Research of High-school Students' Probability and Statistics Learning (In Chinese). Unpublished thesis (M. ed). Shandong Normal University, Shandong.

SU, X. (2008). Teaching Research of Probability in the New Mathematics Curriculum of Senior High School Based on Historical- genetic-Principle (In Chinese). Unpublished thesis (M. ed). Zhejiang Normal University, Zhejiang.

WANG, L. (2011). The Analysis of High School Students' Cognitive Obstacles in Probability Learning and The Solving Strategies (In Chinese). Unpublished thesis (M. ed). Shandong Normal University, Shandong.

WANG, Y. (2009). The Research of High-school Student Probability Study Condition. (In Chinese).Unpublished thesis (M. ed). Northeast Normal University, Jilin.

YUAN, W. (2011). Research on Shanghai Grade Nine Students' Understanding of Probability (In Chinese). Unpublished thesis (M. ed). East China Normal University, Shanghai.

ZHANG, J. (2009). Analyses on Common Understanding of High School Students in The Process of Learning Probability (In Chinese). Unpublished thesis (M. ed). Northeast Normal University, Jilin.

ZHOU, F. (2012). Research on Senior High School Students' Understanding of the Concept about Probability Based on the New Curriculum (In Chinese). Unpublished thesis (M. ed). Northeast Normal University, Jilin.

SHAO Xiayan
East China Normal University
shaoyan2009@126.com