# Re-exploring teaching and learning of probability and statistics using Excel 

Seung-Bum Lee*, Jungeun Park**, Sang-Ho Choi***, Dong-Joong Kim****


#### Abstract

The law of large numbers, central limit theorem, and connection among binomial distribution, normal distribution, and statistical estimation require dynamics of continuous visualization for students' better understanding of the concepts. During this visualization process, the differences and similarities between statistical probability and mathematical probability that students should observe need to be provided with the intermediate steps in the converging process. We propose a visualization method that can integrate intermediate processes and results through Excel. In this process, students' experiences with dynamic visualization help them to perceive that the results are continuously changed and extracted from multiple situations. Considering modeling as a key process, we developed a classroom exercise using Excel to estimate the population mean and standard deviation by using a sample mean computed from a collection of data out of the population through sampling.


- Keyword : binomial distribution, Excel, normal distribution, the central limit theorem, the law of large numbers


## I. Introduction

Mathematical thinking often involves various visuals that aid in developing an understanding of mathematical objects help students make sense of abstract mathematics concepts and thus motivate their learning and participation in the classroom ([1], [2]). Multiple representations as visual mediators facilitate students' deeper understanding of mathematical concepts ([3]). An efficient way of representing visual mediators is to employ multiple representations through the use of technology. However, there has been limited awareness of the need for using technology in the classroom, where many mathematical concepts are taught in a lecture style that may not reflect students' various ways of learning.
Among various mathematical subjects, the ambiguity and uncertainty of the concepts of probability has been particularly difficult for students to learn. Moreover, teachers tend to teach concepts in probability and statistics
by providing definitions and theorems in an abstract way and focusing on computation which may not help students understand the concept of probability deeply or may even hinder students to see the concept's connection to real-life situations ([4], [5], [6]).
The appropriate use of technology for continuous visualization in teaching and learning concepts in probability and statistics may help overcome these difficulties. In particular, in teaching and learning the differences and similarities between statistical probability and mathematical probability, which is a difficult but crucial concept, technology can be an instrument in generating dynamic processes where statistical probability converges to mathematical probability. To this end, it would be helpful to design experiments involving simulations of such processes with the use of random numbers, in which students can predict the mathematical probability. This would make the relation between the two types of probabilities more accessible ([7], [8]).

Various class materials have been developed that

[^0]visualize the convergence of statistical probability to mathematical probability using technology ([1], [6]). However, it is rare to find class materials that emphasize dynamic conversions of statistical probability visually represented in diagrams to a normal distribution curve that mediates the mathematical probability as trials repeat. Moreover, once the normal distribution is standardized, most class materials mainly address computation based on the standard normal distribution table, but rarely mention the relation between statistical probability and mathematical probability. In addition, the number of trials are finite ([9]. [10], [11], [12]), basic commands in Excel are used and fixed data are visualized ([13], [14]) in the majority of developed programs. Even after learning the similarities and differences between statistical probability and mathematical probability, mere mechanical repetition of mathematical probability may lead students to conceive of probability as being the result of computation, which may hamper students' understanding about the relation between probability and real-life situations (see [15], [16]). For this reason, emphasizing the dynamics of the relation between statistical probability and mathematical probability through tables and diagrams which depict the converging nature from statistical probability to mathematical probability and revisiting the similarities and differences between these two types of probabilities constantly throughout the probability and statistics unit would help enhance students' understanding of the concept and its real-life applications ([17], [18], [19]). To this end, this paper provides a class activity using Excel.

## II. DESIGN OF THE ACTIVITY

## 1. Activity for the Law of Large Number

The law of large numbers explains the relation between probability and real-life situations, but it can be difficult for students to understand because it involves the concept of the limit. The following activity applies the law of large numbers to die toss. Through this activity, students will experience the law of large numbers by observing how the statistical probability of rolling the number 1 on a die approaches mathematical probability ( $1 / 6$ ) when the number of trials is large enough.
As shown in Fig 1, students can set the number of rolls of the dice, P . Then, by clicking the button "rolling a die P times," they toss a die P times. They will observe the
differences between statistical probability and mathematical probability decrease as they increase the number of trials, P. Most textbooks address this relationship only with words, which may not help students see the relationship. However, through this activity, students may experience the relation by changing the numbers of trials and observing the subsequent probabilities and their differences.


Fig. 1. Statistical probability and mathematical probability

## 2. Activity for Central Limit Theorem

The central limit theorem in general tells us that $\bar{X}$, the mean of a sample generated from a sufficiently large number of iterations of independent random variables, will follow a normal distribution when the sample size is large enough ([20]). Most textbooks introduce and explain this theorem, but rarely provide data that students can visually make sense of. Students can observe how this theorem can be visually understood using Excel. For example, let $\bar{X}$ be the mean of numbers of a die when it is tossed 100 times (which will create a sample consisting of 100 numbers), and then create more samples and thus generate several $\bar{X}$ by iterating 500 tosses. Such results can be cumulated and represented as a discrete graph (Fig 2) or as a continuous graph (Fig 3). Fig 2 represents 100 values of $\bar{X}$ as a simple mean of a set consisting of 100 numbers. In Fig 3, P represents the number of tosses, and N represents the number of samples consisting of P numbers.


Fig. 2. Distribution of sample mean, when $P=100, N=100$
Students can easily observe that a discrete variable shown in Fig 2 follows a normal distribution as the number of times generating $\bar{X}$ increases from 30 to 300, to 1000,
and to a large enough number. All these processes happen simultaneously when the results in Fig 1 are found. In other words, students will observe all these results on one Excel sheet in real time. It shows the dynamics of the graph that reflects each $\bar{X}$ found by pushing a button (i. e., Shift + N in Fig 3) rather than showing the results while omitting the process. Such continuous visualizations without omitting the intermediate steps in the converging process can provide students a better understanding of the dynamic aspect of the central limit theorem. Specifically, emphasizing the dynamic aspect of the theorem that shows the relation between statistical probability and mathematical probability via its visualization (as P is fixed and N increases, the graph approaches a normal curve as shown in Fig 3) can help various types of students understand the theorem, and also the concept of probability in general.


Fig. 3. As N is increasing, graph-shapes approach to a normal distribution

## 3. Connection between the binomial distribution and normal distribution

In probability and statistics, students tend to spend a considerable amount of time solving problems involving a normal distribution curve. They also learn about binomial distribution, which is discrete probability distribution, and then normal distribution, which is a continuous probability distribution. Then, when a random variable X is binomially distributed, its probability is similar to normal distribution when " $n$ " (i. e., number of trials) is large enough. In other words, the key is that X that follows the binomial distribution becomes similar to the normal distribution. However, there is not much class material that visualizes this process. Thus, students seem to conceive of it only numerically and then just focus on computation by following guided procedures and algorithms. Even if they learn the similarities and differences between statistical probability and mathematical probability in the beginning of the probability and statistics unit, their understanding might be limited in terms of appreciating the use/meaning of statistical probability. To
this end, we developed a problem(shown Table 1)and class material(shown in Fig4).

Table 1. Problem statements used in the activity

|  | Statement |
| :--- | :--- |
| Problem | At a company that produces batteries, the <br> probability of making failed product is 10\%. <br> If they produce 1000 batteries today, what is <br> the probability of producing greater than <br> equal to 90 failed products and less than <br> equal to 105 failed products?). |
| Extended <br> problem 1 | When the same company produces 1000 <br> batteries, find the probability of making <br> greater than or equal to a failed product, <br> and less than equal to "b"" products. Now <br> let's change the values of "a" and "b", then <br> find a mathematical probability, and then <br> check how close it is to the statistical <br> probability.). |
| Extended | At the same company, let's change the <br> number of batteries that they produce and <br> the probability of making a failed product. Is <br> statistical probability still similar to the <br> mathematical probability?). |
| problem 2 |  |

In these data, after students compute the probability based on the assumption that the binomial distribution follows the normal distribution as the number of sample means is large enough, they see that the statistical probability and mathematical probability computed are similar. Thus, they can see that they can assume that the binomial distribution follows a normal distribution and the probability that they found from the data as shown in Fig 4 is similar to the mathematical probability.


Fig. 4. A program for understanding the similarities and differences between statistical probability and mathematical probability

The main idea of the program shown in Fig 4 is as follows. First, the failure rate for new batteries that the company produces is the simplest fraction representing probability. Then the program generates random numbers from 1 to the denominator of the fraction and these generating processes are as many as the number of the batteries that the company produces. For example, if the failure rate for new products
is $4 / 13$, the program can generate random numbers from 1 to 13 as many as the number of batteries that the company produces. Then, among all these generated random numbers, the way of counting numbers from 1 through the numerator 4 (except 5 through 13) reflects the probability of producing failed products given the problem. In the case presented in Fig 4, the failure rate for new products was $1 / 10$, so the generated random numbers 1 through 10 were recorded on the first column, and among those generated numbers, 1 would be the number of failed products produced.

To apply this idea, first enter the number of batteries produced, the failure rate for new products (denominator " p " and numerator " $q$ "), and the interval for failed products, and generate random numbers as many as the number of batteries produced by clicking the "Number of batteries produced" button(Fig 5). Then, when one clicks the "Number of failed batteries" button, the program counts numbers that represent failed products among random numbers, and records the number(frequency) of such numbers in the second column. The process so far is considered as one trial. If one clicks the "Number of failed batteries" but to "n"multiple times, the program repeats the trial, and records the number of failed products from each trial in the third column. Then, the frequencies of such numbers are recorded in the fourth column. For example, Fig5 shows the case when one clicked "Number of failed batteries" five times. For these trials, the numbers of failed batteries were $119,98,80,125$, and 120 in turn. Thus, the frequency of each of those numbers, shown in the fourth column, increased to 1 in the fifth column. Because among these five trials, there was one case(i. e., 98) between 90 and 105, the statistical probability asked in problem1 became $1 / 5$.


Fig. 5. Mid-process of the program
Because 5 was the small number, one can also increase the number of trials. To increase the number of trials,
enter the number right next to "the trial number" N , and click "find a trial N at once." For example, Fig 6 shows a case of $\mathrm{N}=1000$. It shows the process of generating 1000 random numbers and counting the number of failed batteries from N equals $1,2, \cdots, 1000$. The fifth column shows the frequency of the failed battery whose number was between 90 and 105, and the sixth column shows the statistical probability of such cases.


Fig. 6. A result, when $N=1000$
Fig 6 also includes mathematical probability computed from a normal distribution table. Here, because the total number of the batteries produced was 1000, the failure rate for new batteries was $1 / 10$, if one lets X be the number of failed batteries, $X \sim B(1000,1 / 10)$, and thus $E(X)=100$ and $\operatorname{Var}(X)=90$. Thus, $X \sim N\left(100,(3 \sqrt{10})^{2}\right)$. Based on this information, one can find the mathematical probability, $\mathrm{P}(90$ $\leqq X \leqq 105$ ) using the normal distribution function in Excel, which is recorded in the 13th row. The value of the mathematical probability will change depending on the number of trials, the failure rate for products, $\mathrm{q} / \mathrm{p}$, and the minimum and maximum for the interval of failed products. Then, the statistical probability can be found by repeating the process of generating 1000 random numbers and counting the number of failed products, once, twice $\cdots$ to one thousand times. Thus, students can find the statistical probability for making failed products. Fig7 shows the details of this process.


Fig. 7. Statistical probability and mathematical probability using a normal distribution

In Fig 7, which shows that | statistical probability mathematical probabilityl is almost 0 , one can find that the mathematical probability computed from the normal distribution table is valid, and the value computed by approximating the binomial distribution to normal distribution is also valid. Thus, the purpose of this designed program can be effectively attained. Students can see that |statistical probability-mathematical probability is very close to 0 , but not zero. Through this "almost zero" value, students may conceive of the importance and meaning of mathematical fact, and by emphasizing that "almost zero" is not zero, students can think of the statistical theory as an approximation of reality, not as reality itself. Furthermore, because one can enter different values in "the number of batteries produced," "the failure rate for products (denominator " p " and numerator " q ")," and the minimum and maximum of failed products, one can apply this activity to various situations including extended problems 1 and 2. For example, in the case of extended problem 1, one can let "a" be the minimum number of failed products and "b" be the maximum number of failed products of the given interval.

## 4. Statistical estimation

It is well known that statistical estimation is such a difficult concept that students mechanically apply predetermined formulae in solving problems ([21]). In addition, there is a lack of opportunities for students to use statistical estimation in a meaningful way. However, statistical estimation is applied to a great deal of real-life data such as public opinion polls and election predictions. Knowing how to determine whether these data are reliable and how they are modeled mathematically can help students to interpret and comprehend diverse social phenomena through the mathematical modeling process. Considering modeling as a key process, we developed a classroom exercise using Excel to estimate the population mean and standard deviation by using a sample mean computed from a collection of data out of the population through sampling.

In the developed material, which will be applied to diverse situations as a meaning-making tool, the population mean and standard deviation acquired from the trials " n " and "p" are presented in column G and row 17(i. e., G[17]) and column $H$ and row 17(i. e., H[17]) respectively(see Fig 8). In addition, a related normal
distribution curve is drawn below them(see Fig 9). We can input a value referring to reliability a into cell L[2]. Then the boundary values belonging to a confidence interval computed from sample mean $\bar{X}$ are placed in cells $M$ and $N$. For instance, the values in M[2] and N[2] are computed from sample mean $\overline{X_{2}}$. In other words, two values $n \times\left(\widehat{P}_{2}-\alpha \sqrt{\frac{\widehat{P}_{2}\left(1-\widehat{P}_{2}\right)}{n}}\right)$ and $n \times\left(\widehat{P}_{2}-\alpha \sqrt{\frac{\widehat{P}_{2}\left(1-\widehat{P}_{2}\right)}{n}}\right)$ are computed with the sample probability $\widehat{P}_{2}=\overline{X_{2}} / n$ and then entered into M[2] and $\mathrm{N}[2]$ respectively.


Fig. 8. Statistical estimation

The reason to use multiplication by n instead of simply using population mean is because it is more intuitive to use sample mean for students than to use population mean. If this method is repeated and then computed values are entered into cells $\mathrm{M}[\mathrm{k}]$ and $\mathrm{N}[\mathrm{k}]$ where $1 \leq \mathrm{k}$ $\leq \mathrm{n}$, the n numbers of M and N cells are represented with their boundary values. In these processes, if an expected value $\mathrm{E}(\mathrm{x})$ belongs to the confidence interval, the symbol O is presented in column O , and if not, the symbol X is presented. Finally, the ratio of the symbol O to the total, namely, the ratio to measure amount of confidence intervals to which the population mean belongs out of a string of $n$ sample means, is presented in cell P[2] in Fig 8. This value is a meaningful one to show reliability. All these processes are represented on a normal distribution curve for students to visually see and intuitively understand, as shown in Fig 9. If the processes are continuously represented on a normal distribution curve, students can experience the continuity of thinking through visualization. In addition, they can experience the meaningfulness of statistical estimation because all data inputs as variables can be arbitrarily created and
manipulated.


Fig. 9. Statistical estimation by 95\% confidence interval

## III. CONCLUSION

With the activity introduced in this paper that emphasizes the similarities and differences between statistical probability, mathematical probability through their converging relations, and statistical estimation, we conclude the followings. First, students can have a deeper understanding about probability by experiencing the similarities and differences between statistical probability and mathematical probability with the visual instrument that shows the dynamic processes between the two ([1]). The data including tables and graphs visualized through technology can accommodate student preference in learning tools and thus help their mathematical reflection. Specifically, observing dynamics in which the graphs approach a normal curve would help them understand the concept that probability results from continuous change in variables in various cases, not just as a fixed mathematical probability. Thus, this program can give students a more holistic understanding of probability and statistics in general. Eventually, it positively impacts on students' interest and attitudes in their mathematical learning ([13], [14]).

Second, students can revisit the mathematical meaning of probability through the relationship between binomial distribution, normal distribution, and statistical estimation. Through the program introduced in this paper, students can observe that the difference between statistical probability and mathematical probability is almost zero, but not zero, and can then speculate and explain the reason by examining the data. This experience can help students understand that (a) the meaning of probability and statistical theory as well as its limitation that it cannot be exactly the same as what happens in reality,
and (b) mathematical models that explain reality can have their limitations in general. With this in mind, emphasizing the importance of understanding and reducing the difference between reality and theory becomes an overarching goal for teaching and learning statistics.

Based on this conclusion, we can also discuss the use of technology in general. Use of technology can promote a dynamic learning environment where communication among learners and between teachers and students is considered part of the learning process. In other words, using technology in teaching and learning mathematics can motivate the interest of students, promote student participation in class, and eventually help the class become a dynamic environment where the communication between teacher and students can naturally and frequently happen ([22], [23]). Second, the activity that involves various representations and cases helps bring the mathematical variability principle and perceptual variability principle ([24]) into the students' learning process. Specifically, activities with various representations through real-time technology help promote the perceptual variability principle, and including the manipulation of various factors that are included in one mathematical concept while explaining the concept would address the mathematical variability principle.

Our conclusions and discussions above provide the following suggestions. First, teachers should understand various aspects of technology, so that they can use it in various teaching and learning situations appropriately. The appropriate use of technology to promote students' understanding about mathematical concepts and motivate their learning depends significantly on when and how teachers incorporate the technology in their classes. For example, using a graphing calculator in a class where the goal is to learn computation (e. g., for finding the derivative using the definition) may not help teachers achieve that goal. Similarly, teachers' decisions about how they address mathematical concepts with or without technology would be crucial to maximize the effect of their teaching method in student learning ([8], [25]). Second, curriculum developers and policy makers may need to provide concrete and realistic principles for the use of technology in teaching and learning. Many textbooks provide a fixed picture as an example of using technology. With such pictures students may not visualize the dynamics that technology can provide but conceive of it as one of the representations that textbooks provide
such as a diagram. Activities in which students can manipulate numbers and variables may help the class become a place where students can experience mathematical ideas and thus promote deeper understanding of the concepts. Third, we need to develop more activities and materials by which students can gain a deeper understanding of mathematical concepts through the use of technology. Providing various representations through technology is important, but developing activities with which students can experience how advanced and difficult mathematical concepts are defined and work would help teachers and students even more in the long run. In this study, we developed a classroom exercise using Excel to explore how to improve probability and statistics education, but we did not verify its effectiveness. It is needed to the effectiveness of the developed material and its applicability to educational contexts.

## REFERENCES

[1] B. Shin, and K. Lee, "A Study on the Statistical Probability Instruction through Computer Simulation," Journal of Educational Research in Mathematics, vol. 16, no. 2, pp. 139-156. May 2006.
[2] L. S. Vygotsky, "Thought and language," Cambridge, MA: MIT Press, 1986.
[3] J. Kilpatrick, J. Swafford, and B. Findell, "Adding It Up: Helping Children Learn Mathematics," Washington, DC: National Academies Press, 2001.
[4] K. Chang, "Visual Representations and Visual Thinking in Mathematics Education," Journal of the Korea Society of Educational Studies in Mathematics, vol. 2, no. 2, pp. 53-63, December 1992.
[5] H. Lew, and J. Lee, "Visualization in Mathematics Education and LoGo," Journal of the Korea Society of Educational Studies in Mathematics, vol. 3, no. 1, pp. 75-85, July 1993.
[6] H. Seo, S. Kang, and H. Lim, "The Application of Excel to Probability and Statistics Teaching," Communications of Mathematical Education, vol. 9, pp. 299-316, July 1999.
[7] R. C. Caulfield, P. E. Smith, and K. McCormick, "The spreadsheet: A vehicle for connecting proportional reasoning to the real world in a middle school classroom," In W. Masalski, \& P. C. Elliott(Eds.),

Technology-supported mathematics learning environments, Reston, VA: NCTM, pp. 165-173, 2005.
[8] National Council of Teachers of Mathematics. "Principles and standards for school mathematics," Reston, VA: NCTM, 2000.
[9] D. Lee, "Application of excel on statistics education," Hong-ik University of the graduate school of education master's thesis, 2005.
[10] M. Lee, "A study on teaching high school statistics with excel," Dankook University of the graduate school of education master's thesis, 2010.
[11] J. Lee, "Studying on education of probability and statistics by using excel," Yonsei University of the graduate school of education master's thesis, 2007.
[12] A. Graham, and M. Thomas, "Representational versatility in learning statistics," International Journal for Technology in Mathematics Education, vol. 12, pp. 3-14, July 2005.
[13] Y. Kim, "A study on teaching and learning probability and statistics education using excel," Ajou University of the graduate school of education master's thesis, 2005.
[14] M. Lee, "Studying on education of probability and statistics by using microsoft excel," Kangwon National University of the graduate school of education master's thesis, 2008.
[15] D. Nolan, and D. Temple Lang, "Dynamic interactive documents for teaching statistical practice," International Statistical Review, vol. 75, pp. 295-321, December 2007.
[16] C. Schwartz, "Computer-aided statistical instruction: multimediocre techno-trash?," International Statistical Review, vol. 75, pp. 348-364, December 2007.
[17] S. Lewis, "Laptops in the middle school mathematics classroom," In W. Masalski, \& P. C. Elliott(Eds.), Technology-supported mathematics learning environment, Reston, VA: NCTM, pp. 263-275, 2005.
[18] C. S. Parke, "Using spreadsheet software to explore measures of central tendency and variability," In W. Masalski, \& P. C. Elliott(Eds.), Technology-supported mathematics learning environments, Reston, VA: NCTM, pp. 175-188, 2005.
[19] J. B. Plymouth, "Developing mathematical modelling skills: The role of CAS," The International Journal on Mathematics Education, vol. 34, no. 5, pp. 212-220, 2002.
[20] J. Woo, K. Park, K. Park, K. Lee, N. Kim, J. Yim, J.

Lee, and M. Kim, "Integral Calculus and Statistics," Seoul: Dongapublishing, 2010.
[21] W. Kim, S. Moon, and J. Byun, "Mathematics teachers' knowledge and belief on the high school probability and statistics," The Mathematical Education, vol. 45, no. 4, pp. 381-406, November 2006.
[22] P. H. Dunham, and T. P. Dick, "Research on graphing calculators," The Mathematics Teacher, vol. 87, no. 6, pp. 440-445, 1994.
[23] M. K. Heid, and M. T. Edwards, "Computer algebra systems: Revolution or retrofit for today's mathematics classrooms?" Theory Into Practice, vol. 40, no. 2, pp. 128-136, April 2010.
[24] Z. Dienes, "Building up mathematics," London, UK: Hutchinson Educational Ltd, 1960.
[25] L. Ball, and K. Stacey, "Teaching Strategies for Developing Judicious Technology Use," In W. Masalski, \& P. C. Elliott(Eds.), Technology-supported mathematics learning environments, Reston, VA: NCTM, pp. 3-15, 2005.


He worked as a teacher in Bromsgrove International School from U.K. in Thailand. He was also a teacher at Dawon International Middle School in Korea.


Jungeun Park received the B.S. in Mathematics in 2001 and M.A. in Education in 2005 from Korea University, Korea. Jungeun Park also received M.S. in Mathematics in 2007 and PhD in Mathematics Education in 2011 from Michigan State University, U.S.A.

Dr. Park joined the faculty of the Department of Mathematics at the University of Central Arkansas in 2011. She is currently an assistant professor in the Department of Mathematical Sciences at the University of Delaware. She is interested in undergraduate mathematics education for STEM majors, discourse analysis for mathematical teaching and learning, and promoting students' mathematical thinking through the use of technology.


Sang-Ho Choi received the B. S. in Education in 2006 from Chonbuk National University and the M. S. in mathematics education from Korea National University of Education, Korea, in 2013.
Mr. Choi is currently a graduate student in the Department of Curriculum \& Instruction, Korea University. He is interested in mathematics discourse and technology.


Dong-Joong Kim received his bachelor's degree in mathematics education from Korea University, his master's degree in mathematics from the University of Georgia, and his doctoral degree in mathematics education from Michigan State University.
Dr. Kim is an associate professor in the Department of Mathematics Education at Korea University. His research areas include discourse analysis and the use of technologies to promote student engagement in learning mathematics. He seeks to engage his students in a search for relating mathematics learning and teaching to discourse.


[^0]:    - First Author: Seung-Bum Lee, Corresponding Author: Dong-Joong Kim
    *Seung-Bum Lee(ttazzan@naver.com), Department of Mathematics Education, Korea University
    **Jungeun Park(jungeunpark124@gmail.com), Department of Mathematical Sciences, University of Delaware
    ***Sang-Ho Choi(shchoi83@hanmail.net), Department of Curriculum \& Instruction, Korea University Graduate School
    ****Dong-Joong Kim(dongjoongkim@korea.ac.kr), Department of Mathematics Education, Korea University
    - Received: 2016. 05. 12, Revised: 2016. 06. 23, Accepted: 2016. 07. 12.

