The Central Limit Theorem in the Sociology Classroom: A Constructivist Approach

Brigitte H. Bechtold

Department of Sociology, Anthropology, and Social Work
Central Michigan University

Abstract

To undergraduate students of sociology, and who are required to take at least one course in research methods, understanding of the CLT requires the successful negotiation of a number of hurdles, lack of training in mathematics in general and mathematical statistics in particular, and a sociological aversion to the "bell curve" or normal distribution. This paper critiques the "sweetening" approaches to conduct experiments that construct the Central Limit Theorem in the classroom. It proceeds to outline a simple type of experiment based on a discrete rectangular population distribution, and offers a proof, understandable to sociology undergraduates, of how a discrete rectangular population distribution gives rise to a continuous sampling distribution.

1. Introduction

The Central Limit Theorem (CLT) is a fundamental theorem of mathematical statistics. It has the characteristic of fundamental theorems in other branches of mathematics in that it conveys a profound and partly surprising finding. While some statisticians point out that there are two fundamental theorems in mathematical statistics—the Law of Large Numbers (LLN) and the CLT—and refer to the CLT as the "second" fundamental theory of statistics (e.g. Grimstead and Snell 1997), they would not dispute the much more surprising result of the CLT compared to the LLN (Wicklin 2014). Indeed, the LLN says that large random samples reflect the population. This result is not all that unexpected. It is also often confused with the much more extensive result of the CLT, that in repeated large (simple) random samples taken from a population, the distribution of the sample mean is normal, with a mean equal to that of the population (μ) and a standard deviation equal to that of the population divided by the square root of the sample size. The most exhilarating part of the result is that this holds *regardless of the population distribution.* When we consider that the normal distribution is a continuous distribution, with the range of real numbers on the horizontal axis, while the population distribution may be very non-normal (e.g. a rectangular or uniform distribution), with only a few discrete values, this results is definitely not intuitively obvious.

For undergraduate students of sociology, and who are required to take at least one course in research methods, understanding of the CLT requires the successful negotiation of a number of hurdles, lack of training in mathematics in general and mathematical statistics in particular, and a sociological aversion to the "bell curve" or normal distribution. As a representation of actual populations, the normal distribution

has known its heyday. Notably, the famous Belgian statistician and sociologist Adolphe Quetelet (1796-1874) utilized it to compute deviations in human behavior and characteristics (see Landau and Lazarsfeld 1968). Today, the average sociology student is understandably weary of the use of the normal distribution in describing human traits as average or deviating from the average in a negative or positive direction, and the research methods professor's task partly involves re-introduction of the normal distribution as a vital one for quantitative data analysis. The larger task is to convey the "truth" of the CLT to sociology students without having to rely on the formal proof. While some form of experimentation is often used in "constructivist" approaches that constitute a relatively easy alternative to a formal proof, this method requires a leap from the classroom experiment to the situation where many more and larger samples are taken from the population. While this leap can be taken constructively by applying combinatorics, sociology students do not typically possess skills in this area.

During the 15 years I spent as an economics professor, I found it rather easy to introduce students to the Central Limit Theorem, by means of (1) a simple experiment of drawing chips or marbles of four colors from a closed container of a fixed total number of them, usually 100; (2) using combinatorics to demonstrate that the rudimentary normal-looking curve generated via our experiment becomes more "normal" as the number of samples increases indefinitely; and (3) relying on a statistical program—such as TSP—to simulate the sample selection of larger-size samples out of non-normal populations. I often relied on a section in my own textbook (Author 1989, pp. 257-261) to show the outcome of the second step and to formulate the Central Limit Theorem. For other situations than the CLT, such as confidence intervals, it is relatively easy to incorporate computer simulations in the classroom, and these are also covered in the literature, as exemplified by Dambolena (1986), Kennedy, Olinsky, and Schumacher (1990), Ng and Wong (1995), West and Ogden (1998), and Paret and Martz (2008). A review of the relevant literature is provided by Mills (2002). The CLT is not so straightforward to convey by simulation, since the computer simulation has to include a routine that selects random samples of a given size. When students can only see the end result of such random sample selection, they may not necessarily be convinced that the outcome is truly illustrating the theorem or concept at hand.

After joining the sociology group of a multidisciplinary department in 1997, I began to look for ways to explain the meaning of the CLT using essentially only the first of these three steps. At the same time, I noticed that professors teaching research methods to sociology students often "sweetened" the CLT by taking samples from a population of M&Ms, with the accompanying objective that the students can eat the experiment. Nearly two decades later, I remain convinced that M&Ms do not lend themselves well to such experiments, as explained in the next section.

2. Sweetening the CLT?

Let us start with the consideration that the majority of students would only want to eat M&Ms that only they have touched. That does not leave many options for a constructivist experiment. Suppose that we purchase a number of small packages of the popular candy, let each student open their bag and count the

number of reds out of the total. That result can then be compared with the number of reds that are in the population produced by the Mars Company, a number it is willing to share with the public. There are three problems with this experiment. The first problem is that we are making the implicit assumption that each bag constitutes a simple random sample of the total M&Ms produced. Since the bags are filled by a machine that distributes colored candies according to preset percentages for each color, this is not a valid assumption. The second problem is that we would not clearly be able to illustrate how the continuous normal distribution is associated with non-normal and discrete distributions. The third problem relates to the first, and that is that this experiment is more appropriate to construct the law of large numbers than the CLT, i.e. in repeated trials, the population frequency of any color of the M&Ms is approximated, as the mean of the means for the students' relative frequencies of "red" will reveal.

M&Ms can be used for experiments in the classroom, in experiments that require each candy bag to be examined by one student only, and only once (they do get sticky after a while, despite their hard shell). For example, one can test by means of a Chi-squared test whether the counts for each color per bag are the same. This is a fun exercise, and the candy does not get germ-covered or sticky in the process, culminating in an enjoyable snack in addition to increased understanding of the Chi-squared test. Paret and Martz (2008) provide a nice overview of ways in which statistics can be sweetened, using Minitab 16. Not surprisingly, the Central Limit Theorem is not one of the five tests reviewed in their paper. This does not mean that M&Ms cannot be used at all in a constructionist experiment to demonstrate the CLT. However, the M&Ms will (1) be touched by more than one student (yuck) and (2) become sticky before the experiment is concluded. Here are the steps of one such experiment: (1) open a number of small M&M bags, and empty them in a container that does not reveal the colors (There is no need to have a bag for each student); (2) count the number of each color represented by the M&Ms in the bowl; (3) give a numerical code to each color (e.g. red = 1, brown = 2, and so on); (3) construct the frequency distribution for this population and compute the population mean and variance; (4) let each student draw a given number of M&Ms at random, with replacement (here is where the stickiness comes in); (5) shake the container and move on to the next student; (6) let each student compute $\overline{X}$ using the numerical codes associated with each color[1]; (7) let students plot their own value of $\overline{X}$ in a graph on the blackboard, and voila. It is doubtful that students will want to eat any of the pieces of candy after the experiment is concluded. To be fair, such experiment can be conducted with candies wrapped in solid-color cellophane, which would allow step (8): divide the candies among the students and eat away.

3. A Useful Constructionist Approach

It is easier to do the experiment described at the end of the previous section if one uses a well-defined population that is non-normal. After all, while a normal population distribution also results in a normal sampling distribution, it is much more exhilarating for students (and their teachers) to begin with a very non-normal distribution. For CLT experiments, I keep a bag of 100 glass beads in my filing cabinet, with

---

[1] This part of the experiment is also a useful way to show how qualitative information such as a color can be given a numerical code.

equal numbers of each of the colors clear (white), yellow, blue and green.[2] The population is easily visualized and plotted, and its mean and standard deviation easily computed. In fact, this is a good review for students who have difficulty computing variances and standard deviations. We assign codes as follows: clear = 1, yellow = 2, blue = 3, and green = 4. The population mean μ equals 25 x (1 + 2 + 3 + 4) divided by 100, which yields 2.5. The variance and standard deviation are then computed as in the following familiar worksheet:[3]

X	μ	X – μ	(X – μ) squared
1	2.5	-1.5	2.25
2	2.5	-0.5	.25
3	2.5	0.5	.25
4	2.5	1.5	2.25

This is a good place to remind students that the sum of the absolute deviations (X – μ) will always be zero, so we square the deviations prior to computing the variance, which is the sum of the last column divided by 4, or 5/4 = 1.25. The standard deviation is √1.25 or 1.12, rather large, considering that the range of values is from 1 to 4, and the difference between the highest (or lowest) values and the mean is 1.5 in absolute value. The outcome that for the rectangular distribution the variance is identical to this difference is not to be expected for all rectangular distributions. The formula for the variance of a discrete uniform distribution is $(k^2 -1)/12$, where k = highest value – lowest value + 1 (in our case, 4 – 1 + 1 = is 4. 1/12 of k^2 -1 equals 1.25, which is the same as we computed in our little worksheet above.

The beads are placed in a dark-colored envelope or bag, which is mixed before each student draws the assigned number of beads in one grab, counts the beads, records the colors, and computes her $\overline{X}$ with the number codes. The results are recorded in tables, such as Tables 1-4 that are the outcome of four such experiments. We can easily see that the mean of the sampling distribution is very close to that of the population, and we can compute the standard deviation using SPSS or another computer or calculator, and see that it is indeed substantially smaller than that of the population. Plotting our results often yields a frequency distribution that indeed seems to approach a normal distribution. Let us not forget that the experiments presented in Tables 1-4 are based on very small samples. If n increases to the rule of thumb value for a large sample, 30, the results will be very striking indeed. However, such experiment would require the aid of a computer program.[4]

---

[2] I have used variants of this experiment since the 1980s (Bechtold and Johnson 1989), and do not make a claim to its originality.
[3] This provides an opportunity to review the fact that, if there are 25 beads of each color in this rectangular distribution, the variance and standard deviation can be computed as if there were only one of each color.
[4] This can be done with Minitab, for example.

TABLE 1
25 Simple Random Draws from a Rectangular Population of 100 Beads
(25 Each White, Yellow, Blue, and Green; Coded as W=1, Y=2, B=3, G=4)
n = 3

Draw	Colors	Numbers	$\overline{X}$	Draw	Colors	Numbers	$\overline{X}$
1	WBG	1 3 4	2.67	14	WWG	1 1 4	2
2	WBY	1 3 2	2	15	YYW	2 2 1	1.67
3	BBY	3 3 2	2.67	16	GGY	4 4 2	3.33
4	WBB	1 3 3	2.33	17	YGB	2 4 3	3
5	WWY	1 1 2	1.33	18	GGB	4 4 3	3.67
6	WGB	1 4 3	2.67	19	WYY	1 2 2	1.67
7	WGB	1 4 3	2.67	20	YGB	2 4 3	3
8	YGB	2 4 3	3	21	WWB	1 1 3	1.67
9	WYB	1 2 3	2	22	YYB	2 2 3	2.33
10	YYY	2 2 2	2	23	WGB	1 4 3	2.67
11	WGG	1 4 4	3	24	GGB	4 4 3	3.67
12	YGB	2 4 3	3	25	GBY	4 3 2	3
13	WWW	1 1 1	1				

$\mu_{Xbar} = 2.48$ $\mu = 2.5$

TABLE 2
25 Simple Random Draws from a Rectangular Population of 100 Beads
(25 Each White, Yellow, Blue, and Green, Coded as W=1, Y=2, B=3, G=4)
n = 4

Draw	Colors	Numbers	$\overline{X}$	Draw	Colors	Numbers	$\overline{X}$
1	BBYY	3 3 2 2	2.5	14	GBBB	4 3 3 3	3.25
2	GGGY	4 4 4 2	3.5	15	WWWG	1 1 1 4	1.75
3	WYGB	1 2 4 3	2.5	16	WYGG	1 2 4 4	2.75
4	GGGW	4 4 4 1	3.25	17	WYGG	1 2 4 4	2.75
5	WYGB	1 2 4 3	2.5	18	GGYB	4 4 2 3	3.25
6	WGGY	1 4 4 2	2.75	19	GGYB	4 4 2 3	3.25
7	WGGB	1 4 4 3	3	20	GGGW	4 4 4 1	3.25
8	BBBB	3 3 3 3	3	21	BBWW	3 3 1 1	2
9	GYBW	4 2 3 1	2.5	22	GGGB	4 4 4 3	3.75
10	WYBB	1 2 3 3	2.25	23	WYGB	1 2 4 3	2.5
11	WYBG	1 2 3 4	2.5	24	GGGY	4 4 4 2	3.5
12	WYGG	1 2 4 4	2.75	25	WWYY	1 1 2 2	1.5
13	WWYB	1 1 2 3	1.75				

μ_{Xbar} = 2.51   **μ** = 2.5

μ$_{Xbar}$ = 2.51   **μ** = 2.5

TABLE 3

25 Simple Random Draws) From a Rectangular Population of 100 Beads

(25 Each White, Yellow, Blue, and Green, Coded as W=1, Y=2, B=3, G=4)

n = 5

Draw	Colors	Numbers	$\overline{X}$	Draw	Colors	Numbers	$\overline{X}$
1	BBWWY	3 3 1 1 2	2	14	WWWBB	1 1 1 3 3	1.8
2	GWYBB	4 1 2 3 3	2.6	15	GBBYY	4 3 3 2 2	2.8
3	YYYYG	2 2 2 2 4	2.4	16	WWWBB	1 1 1 3 3	1.8
4	WWBBB	1 1 3 3 3	2.2	17	WWYBG	1 1 2 3 4	2.2
5	YBBBG	2 3 3 3 4	3	18	WYBBG	1 2 3 3 4	2.6
6	YYBBG	2 2 3 3 4	2.8	19	YYGGB	2 2 4 4 3	3
7	YYYGB	2 2 2 4 3	2.6	20	BBGGW	3 3 4 4 1	3
8	WBBGG	1 3 3 4 4	2.8	21	GGGBY	4 4 4 3 2	3.4
9	GGBYW	4 4 3 2 1	2.8	22	YYYGB	2 2 2 4 3	2.6
10	YYWGB	2 2 1 4 3	2.4	23	WWWBY	1 1 1 3 2	1.6
11	WWYYG	1 1 2 2 4	2	24	WGGGB	1 4 4 4 3	3.2
12	GYWWB	4 2 1 1 3	2.2	25	YBGGG	2 3 4 4 4	3.4
13	WYYYG	1 2 2 2 4	2.2				

μ_{Xbar} = 2.55   **μ** = 2.5

TABLE 4

25 Simple Random Draws) From a Rectangular Population of 100 Beads

(25 Each White, Yellow, Blue, and Green, Coded as W=1, Y=2, B=3, G=4)

n = 6

Draw	Colors	Numbers	$\overline{X}$	Draw	Colors	Numbers	$\overline{X}$
1	YYBBGG	2 2 3 3 4 4	3	14	WWYBBG	1 1 2 3 3 4	2.33
2	WWWBBB	1 1 1 3 3 3	2	15	WWBBBG	1 1 3 3 3 4	2.5
3	YYYBBG	2 2 2 3 3 4	2.67	16	YBBBGG	2 3 3 3 4 4	3.17
4	WYYGGB	1 2 2 4 4 3	2.67	17	WWWYYB	1 1 1 2 2 3	1.67
5	WWWYYB	1 1 1 2 2 3	1.67	18	WWYGGG	1 1 2 4 4 4	2.67
6	WYYYBG	1 2 2 2 3 4	2.33	19	WWWYBG	1 1 1 2 3 4	2
7	WWYYBG	1 1 2 2 3 4	2.17	20	WWYBBG	1 1 2 3 3 4	2.33
8	WWYBGG	1 1 2 3 4 4	2.5	21	WYBBGG	1 2 3 3 4 4	2.83
9	WYYBBG	1 2 2 3 3 4	2.33	22	YYBBGG	2 2 3 3 4 4	3

10	WYBGGG	1 2 3 4 4 4	3	23	WYYBBB	1 2 2 3 3 3	2.33
11	WWWYBB	1 1 1 2 3 3	1.3	24	WYYYBB	1 2 2 2 3 3	2.17
12	WYYBGG	1 2 2 3 4 4	2.67	25	WYYBBB	1 2 2 3 3 3	2.33
13	WYBBGG	1 2 3 3 4 4	2.83				

μ$_{\text{Xbar}}$ = 2.49   μ = 2.5

4. The Standard Deviation of the Mean, or the Standard Error

While the experiment also clearly shows that the standard error of the mean is much smaller than the population standard deviation, students are not automatically convinced that it equals $\sigma/\sqrt{n}$. This part of the theorem can be demonstrated very straightforwardly, however. Many sociology students possess sufficient mathematical skills to understand that (1) the variance of a constant times a variable, or $\text{Var}(CX)$ equals the square of the constant times the variance, or $C^2 \text{Var}(X)$. Most also remember the quadratic formula from their high school years, so that it becomes clear that the X need to be independent (as in simple random samples) for the covariances to disappear. Accordingly, $\text{Var}(\overline{X}) = 1/n^2 \text{Var}(X_1 + \ldots + X_n) = n/n^2 \text{Var}(X)$, so that the standard deviation of the sample mean equals $\sigma/\sqrt{n}$. Most of the students in a sociology research methods class can derive this result.

5. Another and More Exciting Part of the Theorem

Our experiment outlined in Section 3 and illustrated by four actual classroom experiments presented in Tables 1-4 does not provide a clear way to show how the discrete population distribution gets transformed into a sampling distribution of the mean that is continuous. It is clear from the experiments, however, that increasingly large sample sizes yield an increasing number of possible values on the $\overline{X}$ axis. We can do better than this, however, and introduce an actual proof that is within the realm of understanding of the typical sociology student.

The method of mathematical proof by induction relies directly on the properties of natural numbers, which progress in a clear pattern 1, 2, 3, and so on. Proof by induction can be envisaged when something needs to be proved where a progression of the natural numbers plays a crucial role. Such is the case when we think about taking larger and larger samples out of a population. The sample size thus increases from the smallest possible size of 1 to 2, 3, 4, and going on to infinity. We need to convince students that, even though there are only four distinct (discrete) values on the horizontal axis for the population, an infinite number of values on the horizontal axis results for an infinitely large sample size.

It is certainly possible to construct part of this argument experimentally. Table 1 shows that when n = 3, we have values for $\overline{X}$ that include the original four values in the population, 1, 2, 3, and 4, and the additional

values of 1.33, 1.67, 2.33, 2.67, 3.33, and 3.67. While 6 new values are occupied on the X-axis, they are nicely spread out, with two new values between any given values in the population. Thus, a sample size of n = 3 has added n – 1 additional values between each population value. This finding carries over to the larger sample sizes in Tables 2-4. When n = 4, the values between the population values are now 1.25 (not all possible outcomes for the sample mean appear in the experiment shown in the table, but they are easily explained as possible outcomes for $\overline{X}$), 1.5, 1.75, 2.25, 2.5, 2.75, … , 3.25, 3.5, and 3.75. By now, students have also noticed how evenly spread these "in-between" $\overline{X}$ outcomes are in relation to the population values. It is easy to verify in a class experiment that n = 4 leads to n – 1 = 3 such in-between values, n = 5 leads to 4 in-between values, and n = 6 has 5 in-between values. For example, in this latter case, we have between 2 and 3 the five new values 2.17, 2.33, 2.5, 2.67, and 2.87.

The formal extension and proof that each addition to the sample size also adds one additional outcome for $\overline{X}$ is fairly straightforward. Any proof by induction has only two steps. The first is to demonstrate the result for the smallest value of the natural number involved, and the second (harder) is to show that the result holds for any two consecutive values of the natural number involved. If these two steps are demonstrated, the result holds for any value of the natural number. This proof is outlined below for the case of the 4-value discrete uniform distribution, with the incremental sample size following the progression of natural numbers, that is, a sample size of 1, or 2, or 3, or 30, or 5000, etc.

In the first step of the proof, we demonstrate that the result holds for n = 1. Indeed, when n = 1, only the outcomes 1, 2, 3, or 4 are possible, and zero additional values are possible for $\overline{X}$. This completes the first step.

In the second step, we pick a sample size n = k. The sample mean $\overline{X}$ is the sum of k natural numbers for our uniform discrete distribution. The smallest possible $\overline{X}$ occurs when each item picked is the white one, with a value of 1. So 1 occurs k times, and $\overline{X}$ = k (1)/k = 1. The next highest value for $\overline{X}$ occurs when all but one of the items picked is the white one (value of 1), and one is yellow (value of 2), so that we get $\overline{X}$ = (k – 1 + 2)/k = (k + 1)/k = 1 + 1/k. The next largest value is (k + 2)/k = 2 – 1/k, and so on, until we get the largest possible outcome of 2k/k = 2. Thus, gathering the results together, there are k – 1 new values for $\overline{X}$ between the values 1 and 2. Extending the same reasoning to the possible values between 2 and 3 again yields k – 1 new values, and the same goes for the values between 3 and 4. Since we have demonstrated the result for steps 1 and 2 of the proof, we have established the proof for the rectangular distribution. The proof can easily be extended to other types of discrete distributions.

6. Conclusion

The meaning of the Central Limit Theorem is notoriously difficult to convey to students who are not well versed in mathematical statistics, and undergraduate sociology students certain are among these. A well-designed experiment, however, can help students to "construct" the theory rather than to formally prove it. The portion of the theorem that implies that discrete population distributions nevertheless have sample distributions for the mean that approach a normal and therefore continuous distribution, can be understood by the constructivist approach of the classroom experiment, but students can learn to prove this part of the theorem with tools no more extensive than those obtained in the very basic mathematics competency program that exists in all universities.

References

Bechtold, B. H. and Johnson, R. H. (1989), *Statistics for Business and Economics.* Boston: PWS-Kent.

Dambolena, I. G. (1986), "Using Simulation in Statistics Courses," *Collegiate Microcomputer* 4:　339-344.

Goertzel, T. and Fashing, J. (1981), "The Myth of the Normal Curve: A Theoretical Critique and Examination of its Role in Teaching and Research," *Humanity and Society* 5: 14-31.

Grinstead, C. M., and Snell, J. L. (1997), *Introduction to Probability.* 2nd rev. ed. Providence RI: American Mathematical Society.

Kennedy, K., Olinskey, A. and Schumacher, P. (1990), "Using Simulation as an Integrated Teaching Tool in the Mathematics Classroom," *Education* 111: 275-296.

Landau, D.　and Lazarsfeld, P. F. (1968), "Adolphe Quetelet," In *International Encyclopedia of the Social Sciences.* Vol 13. New York: Macmillan and Free Press.

Mills, J. D. (2002), "Using Computer Simulation Methods to teach Statistics: A Review of the Literature," *Journal of Statistics Education,* 10 (n.p.). Available online at http://www.amstat.org/publications/jse/v10n1/mills.html.

Ng, V. M., and Wong, K. Y. (1999), "Using simulation on the Internet to teach Statistics," *The Mathematics Teacher* 92: 729-733.

Paret, M. and Martz, E. (2008), "Sweetening Statistics with Minitab 16," *Minitab News,* August. Online at http://www.minitab.com/uploadedFiles/Content/Academic/sweetening_statistics.pdf.

West, R. W. and Ogden, R. T. (1998), "Interactive Demonstrations for Statistics Education on the World Wide Web," *Journal of Statistics Education,* 6 (n.p.). Available online at http://www. amstat.org/publications/jse/v6n3/west.html.

Wicklin, R. (2014), *Fundamental Theorems of Mathematics and Statistics.* Online at http:// www.blogs.sas.com/content/iml/2014/02/12/fundamental-theorems-of-mathematics-and-statistics/ (n.p.).

Yu, C. H., Behrens, J. T., and Anthony, S. (1995), "Identification of Misconceptions in the Central Limit Theorem and Related Concepts and Evaluation of Computer Media as a Remedial Tool," *ERIC Document Reproduction Service* No. 395 989. New Orleans LA.