A PRACTICAL APPROACH TO THE CENTRAL LIMIT THEOREM

Michael J. Glencross
University of the Witwatersrand
South Africa

1. Introduction

The Central Limit Theorem has been described as "one of the most remarkable results in all of mathematics" and "a dominating personality in the world of probability and statistics" (Adams, 1974, p. 2). It is "one of the oldest results in probability theory" (Araujo & Gine, 1980, p. v), "occupies a unique position at the heart of probabilistic limit theory" (Hall 1982, p. 1) and "plays a central role in the theory of statistical inference" (Keeping, 1962, p. 90). Much of its importance "stems from its proven adaptability and utility in many areas of mathematics, probability theory and statistics" (Hall, 1982, p. 1), while it "accounts very largely for the importance of the normal distribution in theoretical investigations" (Keeping, 1962, p. 90).

2. The theorem

The Central Limit Theorem may be stated as follows:

> Let $X_1, X_2 \ldots, X_n$ be a sequence of independent random variables each having the same distribution with finite mean μ and finite variance of σ^2. If $\bar{X}_n$ is the mean of $X_1, \ldots, X_n$, then the distribution of the standardized variable $Z_n = (\bar{X}_n - \mu)/(\sigma/\sqrt{n})$ converges to the Normal (0 , 1) distribution as $n \to \infty$.
>
> (Hogg & Tanis, 1977, p. 155; Mood, Graybill & Goes, 1974, p. 195)

The point of the theorem is that no matter what the original distribution, the mean of a large enough sample will have a nearly normal distribution. Note that the population from which we are sampling could have a distribution that is uniform, skewed, or non-normal in some other way. It is the sample mean that behaves as a random variable with an approximately normal distribution. This theoretical sampling distribution has a mean equal to μ, the mean of the population and a standard deviation equal to / n. This powerful result may be used to explain why, for example, the observations in many engineering, physical and psychological processes follow the characteristic bell-shaped curve of the normal distribution.

3. Some learning principles

For students taking an introductory statistics course which includes inferential statistics, the Central Limit Theorem is clearly a *sine qua non*. However, although this theorem may be regarded as the cornerstone of statistical inference, it has been the writer's experience over a number of years that many students find it difficult to understand and that the general concept of a theoretical sampling distribution may be too abstract to be fully appreciated when first met. Even students with a good mathematical background, who can appreciate a proof of the theorem, may have difficulty understanding its significance.

In searching for ways to overcome such difficulties and to be able to present the essential features of the theorem to students in an understandable manner, let us turn to some experts for help. Professor Richard Skemp, a mathematician turned psychologist who has played a major role in the development of a theory of mathematics learning has stated a fundamental principle of the learning of mathematics:

> *Concepts of a higher order than those which a person already has cannot be communicated to him by a definition, but only by arranging for him to encounter a suitable collection of examples* (Skemp, 1971, p.32).

Another mathematician-cum-psychologist, Professor Zoltan Dienes, after investigating the learning process in children, identified a number of principles on which he believes the learning process is based. Two of these principles are particularly relevant here. They are

(i) <u>The principle of perceptual variability</u> (later called "multiple embodiment"): "To abstract a mathematical structure effectively one must meet it in a number of different situations to perceive its purely structural properties";

(ii) <u>The principle of mathematical variability</u>: "As every mathematical concept involves essential variables, all these mathematical variables need to be varied if full generality of the mathematical concept is to be achieved" (Dienes, 1963, pp. 156-158; 164, p. 40).

As Skemp notes, it is the teacher rather than the learner who most needs to know such principles (Skemp, 1971). Thus, these three principles have been summarised for the teacher. A teacher should:

1. Provide a suitable collection of examples which exemplify the concept.

2. Ensure that the concept is met in a number of different situations.

3. Ensure that all the variables involved in the concept are seen to vary.

It was decided to use these three principles as guidelines for the design of suitable learning activities. Also, the experience of many teachers has shown that, irrespective of the mathematical background of the students, the effectiveness of a course in statistics is greatly improved when augmented by some kind of experimental work (Murdoch & Barnes, 1966), while the eminent members of the Joint Education Committee of the Royal Statistical Society and the Institute of Statisticians have noted that a criticism often expressed by practising statisticians is that, at this level, "statistics too frequently ignores the practical situation and concentrates on formal manipulation" (Barnett et al, 1979, p. 3). All of the above point clearly to the appropriateness of a practical approach to the Central Limit Theorem.

4. Practical sampling from different distributions

There are a number of sources that give details of simple experiments involving, for example, dice-rolling to illustrate the Central Limit Theorem (e.g. SMP 1970). The purpose of this section is to outline further experiments suitable for demonstrating this important theorem.

Apparatus: A box of wooden counters is required, each counter numbered according to a given frequency distribution.

Method:

1. Each student chooses, at random and with replacement, a sample of n = 2, 5, 10 counters, records the numbers obtained and then calculates the mean of the sample.

2. This process is repeated around the class until, say, 100 such samples have been selected. We now have 100 sample means for samples of size n = 2, 5, 10. These results may be graphed for easy comparison.

3. The mean and standard deviation of the 100 sample means for the different sample sizes are now calculated and the experimental results compared with the theoretical results obtained by applying the Central Limit Theorem.

Examples

(i) Sampling from a normal distribution

A box of 200 counters is required, each counter marked according to the distribution below:

Number:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Frequency:	1	2	3	4	7	9	12	15	18	19	20	19	18	15	12	9	7	4	3	2	1

This distribution is approximately normal with a mean of 10 and a standard deviation of 3.93. If necessary, the calculation of these values could be done by the students.

(ii) Sampling from a uniform distribution

A box of 210 counters, with 10 marked '0', 10 marked '1', . . . , 10 marked '20', is needed. This represents a uniform distribution with a mean of 10 and a standard deviation of 6.06. The calculation of this mean and standard deviation could again form a class exercise.

(iii) Sampling from a triangular distribution

For this experiment, 210 counters are again required with 1 marked '1', 2 marked '2', 20 marked '20'. This represents a triangular distribution having a mean of 13.7 and a standard deviation of 4.82. Variations on this theme are virtually unlimited, including showing differences between sampling with and without replacement. Other possibilities include the use of tables of random numbers. For example, take the population as the set of integers from 1 to 99. Students draw random samples of a fixed size and calculate the sample means. Further samples are taken and the mean and standard deviation of this sample of means are then found. Additional variations would be to take the population as the squares of integers from 1 to 99. The above experiments, even when shared out on a class basis, tend to be time-consuming and in this age of instant everything, teachers and students alike may be reluctant to devote the necessary time. However, these kinds of experiments which involve "getting one's hands dirty" provide invaluable experience for which there is no substitute.

5. Using computer-generated random numbers

The use of computer facilities, in particular the microcomputer, enables the rapid generation of many sample means for samples of varying sizes. A computer program, designed partly by the writer, allows samples of varying size to be selected from a number of different populations with known distributions. The heart of the program lies in the ability of modern microcomputers to generate random numbers (strictly speaking, pseudo-random numbers). The interactive program permits sampling from, for instance, a normal, a uniform, a binomial, or an exponential distribution. Some typical examples of the different probability distributions and the sampling distribution for varying sample sizes may be found in Thomas (1984) (for the Apple II) and Bloom et al (1986) (for the BBC model B).

6. Conclusion

The most important aspect of the Central Limit Theorem is that no stipulation is made concerning the population from which one is sampling. From a pedagogical point of view, a student needs to draw a random sample from a population with a known distribution and then to compare the sample mean with the population mean to see "how close, he or she comes. Any student who does this will know the difference between the two. Students will also be led to understand the difference between the population mean and the mean of the sample means. It is not enough for a teacher to talk about these ideas – concrete experience with sampling is necessary for success. It is hoped that these experiments go some way towards enabling students to observe the central limit phenomenon operating, as well as providing empirical evidence of the truth of the theorem.

References

Adams, W.J. (1974). The life and times for the Central Limit Theorem. New York: Kaedmon.

Araujo, A., & Gine, E. (1980). The Central Limit Theorem for real and Banach valued random numbers. New York: Wiley.

Barnett, V., Downton, F., Kemp, C.D., Holmes, P., Thomas, D.A.H., & Wetherill G.B. (1979). Submission to the Committee of Enquiry into the teaching of mathematics in schools, . . . , made jointly by the Royal Statistical Society and the Institute of Statisticians. A Teaching Statistics Supplement, Teaching Statistics Trust, University of Sheffield.

Bloom, L.M., Comber, G.A., & Cross, J.M. (1986). Using the microcomputer to simulate the binomial distrubution and to illustrate the central limit theorem. International Journal of Mathematical Education in Science and Technology, 17(2), 229-237.

Dienes, Z.P. (1963). An experimental study of mathematics learning. London: Hutchinson Educational.

Dienes, Z.F. (1964). The power of mathematics. London: Hutchinson Educational.

Hall, F. (1982). Rates of convergence in the Central Limit Theorem.

Hogg, R.V., & Tanis, E.A. (1977). Probability and statistical inference. New York: Macmillan.

Keeping, E.S. (1962). Introduction to statistical inference. London: Van Nostrand.

Mood, A.M., Graybill, F.A., & Boes, D.C. (1974) (3rd ed.). Introduction to the theory of statistics. Tokyo: McGraw-Hill Kokagusha.

Murdoch, J., & Barnes, J.A. (1966). Basic statistics. London: Macmillan.

Ruckdeschel, F.R. (1981). BASIC Scientific subroutines (Vol. 1). New York: McGraw-Hill.

SMF (School Mathematics Project) (1970). Advanced Mathematics, Book 4. Cambridge: Cambridge University Press.

Skemp, R.R. (1971). The psychology of learning mathematics. Harmondsworth: Penguin.

Thomas, D.A. (1984). Understanding the Central Limit Theorem. Mathematics Teacher, 77(7), 542-543.