Central limit theorem: The cornerstone of modern statistics

2 authors:

Sang Gyu Kwak
Catholic University of Daegu
98 PUBLICATIONS   **564** CITATIONS

SEE PROFILE

Jong Hae Kim
Catholic University of Daegu
37 PUBLICATIONS   **388** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Heart rate and blood pressure variability and baroreflex sensitivity in liver transplantation View project

Effect of a mother's recorded voice on emergence from general anesthesia in pediatric patients: Study protocol for a randomized controlled trial View project

Central limit theorem: the cornerstone of modern statistics

Sang Gyu Kwak[1] and Jong Hae Kim[2]

Departments of [1]Medical Statistics, [2]Anesthesiology and Pain Medicine, School of Medicine, Catholic University of Daegu, Daegu, Korea

According to the central limit theorem, the means of a random sample of size, n, from a population with mean, μ, and variance, σ^2, distribute normally with mean, μ, and variance, $\frac{\sigma^2}{n}$. Using the central limit theorem, a variety of parametric tests have been developed under assumptions about the parameters that determine the population probability distribution. Compared to non-parametric tests, which do not require any assumptions about the population probability distribution, parametric tests produce more accurate and precise estimates with higher statistical powers. However, many medical researchers use parametric tests to present their data without knowledge of the contribution of the central limit theorem to the development of such tests. Thus, this review presents the basic concepts of the central limit theorem and its role in binomial distributions and the Student's t-test, and provides an example of the sampling distributions of small populations. A proof of the central limit theorem is also described with the mathematical concepts required for its near-complete understanding.

Key Words: Normal distribution, Probability, Statistical distributions, Statistics.

Introduction

The central limit theorem is the most fundamental theory in modern statistics. Without this theorem, parametric tests based on the assumption that sample data come from a population with fixed parameters determining its probability distribution would not exist. With the central limit theorem, parametric tests have higher statistical power than non-parametric tests, which do not require probability distribution assumptions. Currently, multiple parametric tests are used to assess the statistical validity of clinical studies performed by medical researchers; however, most researchers are unaware of the value of the central limit theorem, despite their routine use of parametric tests. Thus, clinical researchers would benefit from knowing what the central limit theorem is and how it has become the basis for parametric tests. This review aims to address these topics. The proof of the central limit theorem is described in the appendix, with the necessary mathematical concepts (e.g., moment-generating function and Taylor's formula) required for understanding the proof. However, some mathematical techniques (e.g., differential and integral calculus) were omitted due to space limitations.

Basic Concepts of Central Limit Theorem

In statistics, a population is the set of all items, people, or events of interest. In reality, however, collecting all such elements of the population requires considerable effort and is often not possible. For example, it is not possible to investigate the proficiency of every anesthesiologist, worldwide, in performing awake nasotracheal intubations. To make inferences regarding the population, however, a subset of the population (sample)

Corresponding author: Jong Hae Kim, M.D.
Department of Anesthesiology and Pain Medicine, School of Medicine, Catholic University of Daegu, 33, Duryugongwon-ro 17-gil, Nam-gu, Daegu 42472, Korea
Tel: 82-53-650-4979, Fax: 82-53-650-4517
Email: usmed@cu.ac.kr
ORCID: http://orcid.org/0000-0003-1222-0054

can be used. A sample of sufficient size that is randomly selected can be used to estimate the parameters of the population using inferential statistics. A finite number of samples are attainable from the population depending on the size of the sample and the population itself. For example, all samples with a size of 1 obtained at random, with replacement, from the population {3,6,9,30} would be {3},{6},{9}, and {30}. If the sample size is 2, a total of $4 \times 4 = 4^2 = 16$ samples, which are {3,3},{3,6},{3,9},{3,30}, {6,3},{6,6}…{9,30},{30,3},{30,6},{30,9}, and {30,30}, would be possible. In this way, 4^n samples with a size of n would be obtained from the population. Here, we consider the distribution of the sample means.

For example, here is the asymmetric population with a size of 4 as presented above:

$$\{3, 6, 9, 30\}$$

The population mean, μ, and variance, σ^2, are:

$$\mu = \frac{3 + 6 + 9 + 30}{4} = 12$$

$$\sigma^2 = \frac{(3 - 12)^2 + (6 - 12)^2 + (9 - 12)^2 + (30 - 12)^2}{4} = 112.5$$

A simple random sampling with replacement from the population produces $4 \times 4 \times 4 = 4^3 = 64$ samples with a size of 3 (Table 1). The mean and variance of the 64 sample means are 12 (the

population mean) and $37.5 = \frac{\sigma^2}{n} = \frac{112.5}{3}$, respectively; however, the distribution of the means of samples is skewed (Fig. 1A).

When a simple random sampling with replacement is performed for samples with a size of 6, $4 \times 4 \times 4 \times 4 \times 4 \times 4 = 4^6 = 4{,}096$ samples are possible (Table 2). The mean and variance of the 4,096 sample means are 12 (the population mean) and 18.75 $= \frac{\sigma^2}{n} = \frac{112.5}{6}$, respectively. Compared to the distribution of the means of samples with a size of 3, that of the means of samples with a size of 6 is less skewed. Importantly, the sample means also gather around the population mean. (Fig. 1B). Thus, the larger the sample size (n), the more closely the sample means gather symmetrically around the population mean (μ) and have a corresponding reduction in the variance $\left(\frac{\sigma^2}{n}\right)$ (Figs. 1C and 1D). If Figs. 1A to 1D are converted to the probability density function by replacing the variable "frequency" with another variable "probability" on the vertical axis, their shapes remain unchanged.

In general, as the sample size from the population increases, its mean gathers more closely around the population mean with a decrease in variance. Thus, as the sample size approaches infinity, the sample means approximate the normal distribution with a mean, μ, and a variance, $\frac{\sigma^2}{n}$. As shown above, the skewed distribution of the population does not affect the distribution of the sample means as the sample size increases. Therefore, the central limit theorem indicates that if the sample size is sufficiently large, the means of samples obtained using a random sampling with replacement are distributed normally with the

Table 1. Samples with a Size of 3 and Their Means

Number	Sample			Sample mean
1	3	3	3	3
2	3	3	6	4
3	3	3	9	5
4	3	3	30	12
5	3	6	3	4
6	3	6	6	5
7	3	6	9	6
8	3	6	30	13
9	3	9	3	5
10	3	9	6	6
	Truncated			
54	30	6	6	14
55	30	6	9	15
56	30	6	30	22
57	30	9	3	14
58	30	9	6	15
59	30	9	9	16
60	30	9	30	23
61	30	30	3	21
62	30	30	6	22
63	30	30	9	23
64	30	30	30	30

Table 2. Samples with a Size of 6 and Their Means

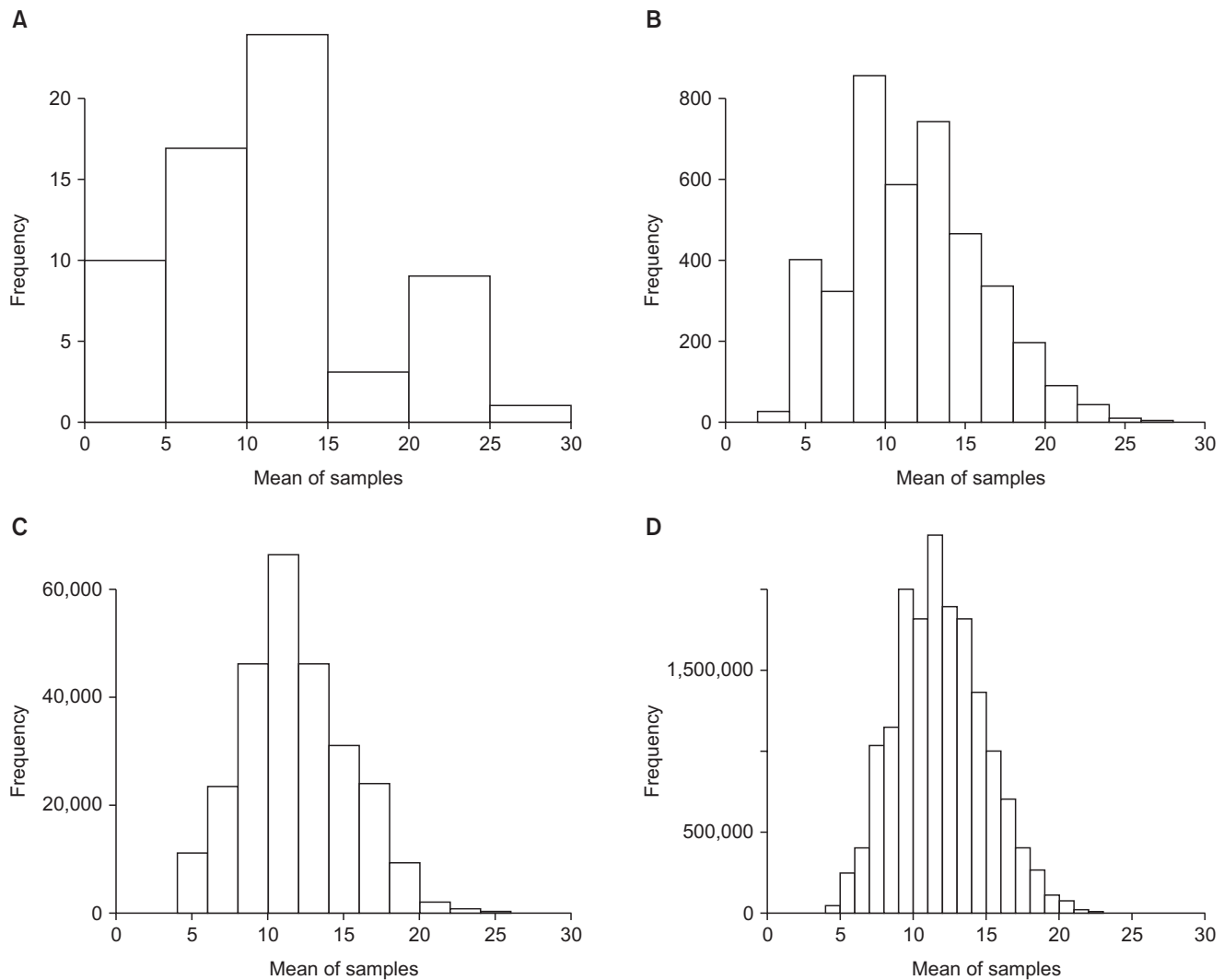| Number	Sample						Sample mean
1	3	3	3	3	3	3	3
2	3	3	3	3	3	6	3.5
3	3	3	3	3	3	9	4
4	3	3	3	3	3	30	7.5
5	3	3	3	3	6	3	3.5
6	3	3	3	3	6	6	4
7	3	3	3	3	6	9	4.5
8	3	3	3	3	6	30	8
9	3	3	3	3	9	3	4
10	3	3	3	3	9	6	4.5
	Truncated						
4087	30	30	30	30	6	9	22.5
4088	30	30	30	30	6	30	26
4089	30	30	30	30	9	3	22
4090	30	30	30	30	9	6	22.5
4091	30	30	30	30	9	9	23
4092	30	30	30	30	9	30	26.5
4093	30	30	30	30	30	3	25.5
4094	30	30	30	30	30	6	26
4095	30	30	30	30	30	9	26.5
4096	30	30	30	30	30	30	30

A

B

C

D

Fig. 1. Histogram representing the means for samples of sizes of 3 (A), 6 (B), 9 (C), and 12 (D).

mean, μ, and the variance, $\frac{\sigma^2}{n}$, regardless of the population distribution. Refer to the appendix for a near-complete proof of the central limit theorem, as well as the basic mathematical concepts required for its proof.

Central Limit Theorem in the Real World

An unbiased, symmetric, 6-sided dice is rolled at random n times. The probability of rolling the number 3 x times in n successive independent trials has the following probability density distribution, which is called the binomial distribution:

$$f(x) = \binom{n}{x}\left(\frac{1}{6}\right)^x \left(1 - \frac{1}{6}\right)^{n-x}$$

n: number of independent trials (rolling a dice), x: number of times the number 3 is rolled in each trial, $\frac{1}{6}$: the probability of

rolling the number 3 in each trial, $1 - \frac{1}{6}$: the probability of rolling a number other than 3 in each trial.

The mathematical expectation of the random variable, X, (i.e., the number of times that the number 3 is rolled in each trial), which is also referred to as the mean of the distribution, is:

$$E(X) = n \times \frac{1}{6}$$

And the variance is:

$$Var(X) = n \times \frac{1}{6} \times \left(1 - \frac{1}{6}\right)$$

When $n = 10$, the probability has a skewed distribution (Fig. 2A); however, as n increases, the distribution becomes symmetric with respect to its mean (Figs. 2B–2D). As n approaches infinity, the binomial distribution approximates the normal dis-

A

Number of cases to obtain number 3 upon 10 successive trials of rolling an unbiased 6-sided dice

B

Number of cases to obtain number 3 upon 20 successive trials of rolling an unbiased 6-sided dice

C

Number of cases to obtain number 3 upon 30 successive trials of rolling an unbiased 6-sided dice

D

Number of cases to obtain number 3 upon 100 successive trials of rolling an unbiased 6-sided dice
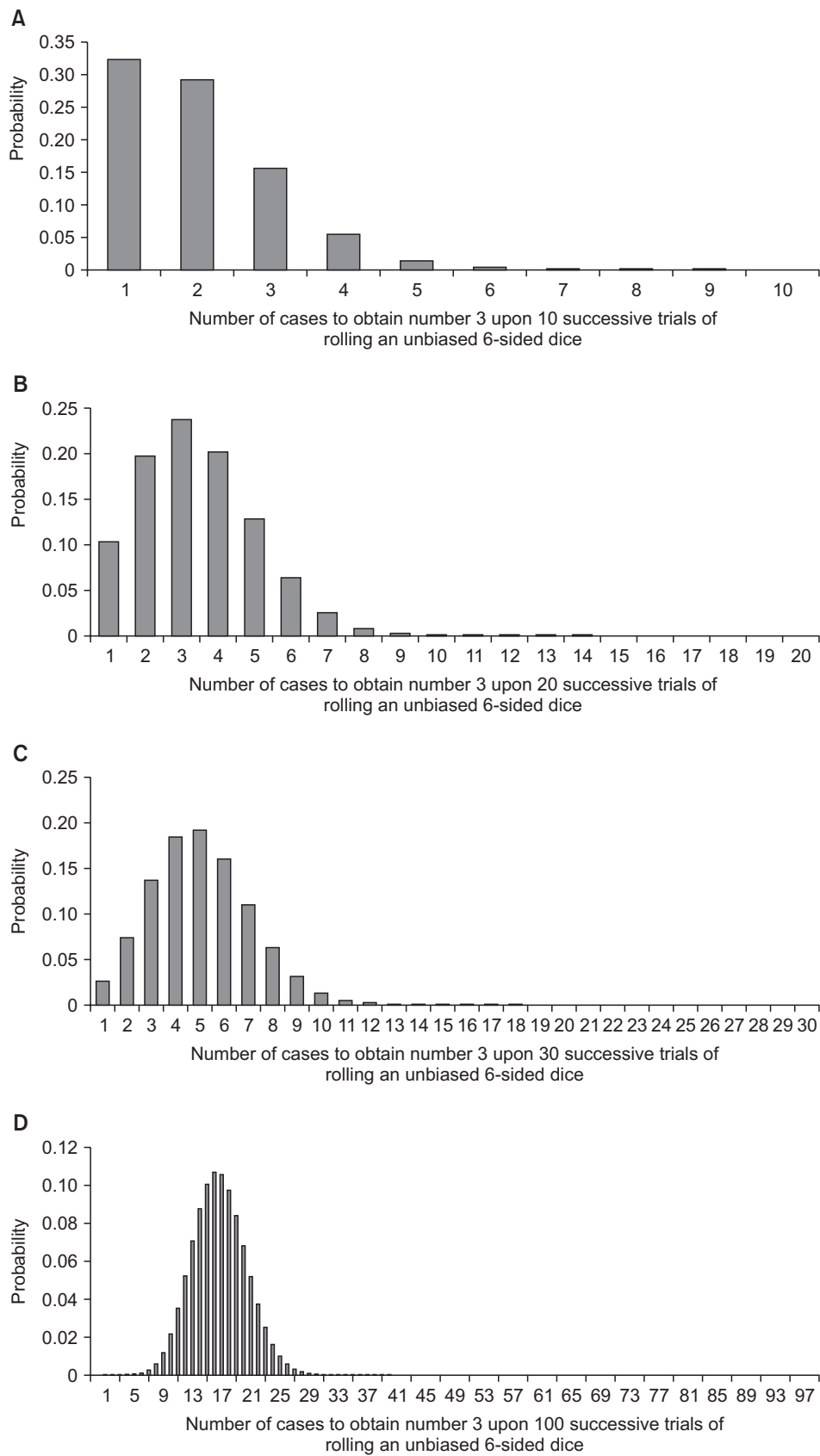
Fig. 2. The probability density function of a binomial distribution with a probability parameter of $\frac{1}{6}$ (i.e., the probability of rolling the number 3 in each trial), based on to the number of trials.

tribution with a mean, np, and a variance, $np(1 - p)$, where p is the probability constant for the occurrence of the specific event during each trial.

Central Limit Theorem in the Student's t-test

Since the central limit theorem determines the sampling distribution of the means with a sufficient size, a specific mean ($\bar{X}$) can be standardized $\left(z = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}} \right)$ and subsequently identified against the normal distribution with mean of 0 and variance of 1^2. In reality, however, the lack of a known population variance (σ^2) prevents a determination of the probability density distribution.

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_N - \mu)^2}{N}$$

X_i ($i = 1, 2, \ldots, n$): a sample from the population, N: the size of the population, μ: the mean of the population.

Notably, the Student's t-distribution was developed to use a sample variance (S) instead of a population variance (σ^2).

$$S^2 = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \cdots + (x_n - \bar{X})^2}{n - 1}$$

x_i ($i = 1, 2, \ldots, n$): a random sample from the population, n: sample size, $\bar{X}$: the mean of the samples.

The specific mean ($\bar{X}$) is studentized $\left(t = \frac{(\bar{X} - \mu)}{\frac{S}{\sqrt{n}}} \right)$ and its location is evaluated on the Student's t-distribution, based on the degree of freedom ($n - 1$). The shape of the Student's t-distribution is dependent on the degree of freedom. A low degree of freedom renders the peak of the Student's t-distribution lower than that of a normal distribution, although at some points, the tails have higher values than those of the normal distribution.

As the degree of freedom increases, the Student's t-distribution approaches the normal distribution. At a degree of freedom of 30, the Student's t-distribution is regarded as equaling the normal distribution [1]. The underlying assumption for the Student's t-test is that samples should be obtained from a normally distributed population. However, since the distribution of population is not known, it should be determined whether the sample is normally distributed. This is particularly true for small sample sizes. If small sample sizes are normally distributed, the studentized distribution of the sample means is equal to the Student's t-distribution with a degree of freedom corresponding to the sample size. If the small samples are not normally distributed, non-parametric tests should be performed instead of the Student's t-test since they do not require assumptions about the distribution of population. If the sample size is 30, the studentized sampling distribution approximates the standard normal distribution and assumptions about the population distribution are meaningless since the sampling distribution is considered normal, according to the central limit theorem. Therefore, even if the mean of a sample of size > 30 is studentized using the variance, a normal distribution can be used for the probability distribution.

Conclusions

A comprehensive understanding of the central limit theorem will assist medical investigators when performing parametric tests to analyze their data with high statistical powers. The use of this theorem will also aid in the design of study protocols that are based on best-fit statistics.

ORCID

Sang Gyu Kwak, http://orcid.org/0000-0003-0398-5514
Jong Hae Kim, http://orcid.org/0000-0003-1222-0054

Reference

1. Kim TK. T test as a parametric statistic. Korean J Anesthesiol 2015; 68: 540-6.

Appendix

Moment-generating function

Since the central limit theorem is proven using a moment-generating function for a random variable, regardless of its distribution, it is necessary to understand the moment-generating function. Prior to that, however, it is necessary to understand the probability density function and mathematical expectation. When a fair coin is tossed at random on two independent and successive trials, the following probability distribution describes the expected number of head(s) (Table 1).

Here, we derive a function, $f(x)$, detailing the probability according to the number of the head(s).

$$f(x) = \binom{2}{x} \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{2-x}$$

2: number of independent trials (coin tosses), x: number of head(s),
$\frac{1}{2}$: the probability of having a head on each trial, $1 - \frac{1}{2}$: the probability of having a tail on each trial.

The $f(x)$ is called the probability density function.
Based on Table 1, the mathematical expectation (or expected value) of the random variable, X, $E(X)$ is:

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$$

Table 1. Probability Distribution

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

x : number of head(s), $P(X = x)$: probability to have x head(s).

Generally, it is expressed as:

$$E(X) = \sum_{x=1}^{n} xf(x) \text{ for a discrete variable,}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \text{ for a continuous variable (1)}$$

Therefore, E(X) is equivalent to the mean, μ, of a random variable. $E(X) = \mu$.

The moment-generating function for a random continuous variable, X, is defined as the mathematical expectation of e^{tX}, $E(e^{tX})$, and denoted as $M_X(t)$. Thus,

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx, -h < t < h \ (h > 0)$$

Since the moment-generating function is a unique determinant for the distribution of a random sample, it is always true that two variables with a common moment-generating function have the same distribution. To get its first order derivative, tx is replaced by u. Then, $\frac{du}{dt}$ is obtained by:

$$u = tx, \quad \frac{du}{dt} = x$$

$$M_X{}'(t) = \frac{dM(t)}{dt} = \frac{\int_{-\infty}^{\infty} e^{tx} f(x)dx}{dt} = \frac{\int_{-\infty}^{\infty} e^u f(x)dx}{dt} = \frac{\int_{-\infty}^{\infty} e^u f(x)dx}{du} \times \frac{du}{dt} = \frac{\int_{-\infty}^{\infty} e^u f(x)dx}{du} \times \frac{du}{dt}$$

Since $\dfrac{e^u}{du} = e^u$, $\dfrac{du}{dt} = x$, and $u = tx$,

$$M_X'(t) = \int_{-\infty}^{\infty} e^u f(x)dx \times x = \int_{-\infty}^{\infty} xe^u f(x)dx = \int_{-\infty}^{\infty} xe^{tx} f(x)dx \quad (2)$$

When $t = 0$,

$$M_X'(0) = \int_{-\infty}^{\infty} xe^0 f(x)dx = \int_{-\infty}^{\infty} xf(x)dx = E(X) = \mu \ (\because \text{please, refer to equation (1).}) \ (3)$$

The second order derivative of $M_X(t)$ is,

$$M_X''(t) = \frac{M'(t)}{dt} = \frac{\int_{-\infty}^{\infty} xe^{tx} f(x)dx}{dt} = \frac{\int_{-\infty}^{\infty} x \times e^{tx} f(x)dx}{dt}$$

According to the equation (2),

$$\frac{\int_{-\infty}^{\infty} e^{tx} f(x)dx}{dt} = \int_{-\infty}^{\infty} xe^{tx} f(x)dx$$

Therefore,

$$M_X''(t) = \frac{\int_{-\infty}^{\infty} x \times e^{tx} f(x)dx}{dt} = \int_{-\infty}^{\infty} x \times xe^{tx} f(x)dx = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)dx$$

$$\therefore M_X''(0) = \int_{-\infty}^{\infty} x^2 e^0 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx = E(X^2) \ (4)$$

Moment-generating function of a normal distribution

The derivation of the probability density function for a normal distribution is beyond the scope of this review. However, the probability density function for a normal distribution with mean, μ, and variance, σ^2, is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty$$

The sum of the probabilities under the normal distribution should be 1.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1$$

As a reference, $\exp(x) = e^x$. According to the definition of the moment-generating function, $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$, $-h < t < h$ $(h > 0)$, the moment-generating function for a normal distribution with mean, μ, and variance, σ^2, is:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \int_{-\infty}^{\infty} \exp(tx) \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp(tx) \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[tx - \frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{2\sigma^2 tx}{2\sigma^2}\right.$$

$$\left. - \frac{(x-\mu)^2 + (\sigma^2 t)^2 - (\sigma^2 t)^2 - 2(x-\mu)\sigma^2 t + 2(x-\mu)\sigma^2 t}{2\sigma^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{2\sigma^2 tx}{2\sigma^2} - \frac{(x-\mu)^2 - 2(x-\mu)\sigma^2 t + (\sigma^2 t)^2}{2\sigma^2}\right.$$

$$\left. - \frac{-(\sigma^2 t)^2 + 2(x-\mu)\sigma^2 t}{2\sigma^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{-2\sigma^2 tx - (\sigma^2 t)^2 + 2(x-\mu)\sigma^2 t}{2\sigma^2}\right.$$

$$\left. - \frac{(x-\mu)^2 - 2(x-\mu)\sigma^2 t + (\sigma^2 t)^2}{2\sigma^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{-2\sigma^2 tx - (\sigma^2 t)^2 + 2\sigma^2 tx - 2\mu\sigma^2 t}{2\sigma^2}\right.$$

$$\left. - \frac{(x-\mu)^2 - 2(x-\mu)\sigma^2 t + (\sigma^2 t)^2}{2\sigma^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{-(\sigma^2 t)^2 - 2\mu\sigma^2 t}{2\sigma^2} - \frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right] dx$$

$$= \exp\left[-\frac{-(\sigma^2 t)^2 - 2\mu\sigma^2 t}{2\sigma^2}\right] \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right] dx$$

$$= \exp\left[\frac{2\mu\sigma^2 t + (\sigma^2 t)^2}{2\sigma^2}\right] \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{[x-(\mu-\sigma^2 t)]^2}{2\sigma^2}\right] dx$$

The last integral, $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{[x-(\mu-\sigma^2 t)]^2}{2\sigma^2}\right] dx$, is the probability density function for the normal distribution with mean, $\mu - \sigma^2 t$, and variance, σ^2. Since the sum of the probabilities under the normal distribution should be 1,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{[x-(\mu-\sigma^2 t)]^2}{2\sigma^2}\right] dx = 1$$

Therefore, the moment-generating function for the normal distribution with mean, μ, and variance, σ^2, is:

$$M_X(t) = \exp\left[\frac{2\mu\sigma^2 t + (\sigma^2 t)^2}{2\sigma^2}\right] = \exp\left[\frac{2\mu\sigma^2 t}{2\sigma^2} + \frac{(\sigma^2 t)^2}{2\sigma^2}\right] = \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right] \quad (5)$$

Mathematical expectation of multiplication between stochastically independent random variables

Let the discrete random variables, X and Y, be stochastically independent.

$$X = \{X_1, X_2, \cdots X_n\}, \quad Y = \{Y_1, Y_2, \cdots Y_m\}$$

$$E(XY) = \frac{X_1 Y_1 + X_1 Y_2 + \cdots + X_1 Y_m + X_2 Y_1 + X_2 Y_2 + \cdots + X_2 Y_m + \cdots + X_n Y_1 + X_n Y_2 + \cdots X_n Y_m}{nm}$$

$$= \frac{X_1(Y_1 + Y_2 + \cdots + Y_m) + X_2(Y_1 + Y_2 + \cdots + Y_m) + \cdots + X_n(Y_1 + Y_2 + \cdots + Y_m)}{nm}$$

$$= \frac{(X_1 + X_2 + \cdots + X_n)(1 + Y_1 + Y_2 + \cdots + Y_m)}{nm} = E(X)E(Y) \quad (6)$$

The above equation also applies to stochastically independent continuous random variables.

Taylor's formula

Assume that the function, f, is continuous on the closed interval, $[a, x]$, and differentiable on the open interval, (a, x).

$$f(x) = f(a) + [f(x) - f(a)] = f(a) + \int_a^x f'(t)dt$$

Let t be replaced by $a + u(x - a)$.

$$t = a + u(x - a) \Longleftrightarrow \frac{dt}{du} = x - a$$

$$\therefore dt = du(x - a)$$

$$a < t < x \Longleftrightarrow a < a + u(x - a) < x \Longleftrightarrow a - a < u(x - a) < x - a \Longleftrightarrow \frac{a - a}{x - a} < u < \frac{x - a}{x - a}$$

$$\therefore 0 < u < 1$$

Then,

$$f(x) = f(a) + \int_a^x f'(t)dt = f(a) + \int_0^1 f'\big(a + u(x - a)\big)du(x - a)$$

As is known,

$$\int_a^b f'(x)g(x)dx = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)dx$$

Using the above formula,

$$\int_0^1 f'\big(a + u(x-a)\big)du = \int_0^1 (-(1-u))'f'\big(a + u(x-a)\big)du$$

$$= \Big[-(1-u)f'\big(a + u(x-a)\big)\Big]_0^1 - \int_0^1 -(1-u)\frac{d}{du}f'\big(a + u(x-a)\big)du$$

$$= -(1-1)f'\big(a + 1 \times (x-a)\big) - \Big[-(1-0)f'\big(a + 0 \times (x-a)\big)\Big]$$

$$+ \int_0^1 (1-u)f''\big(a + u(x-a)\big)\big(a + u(x-a)\big)du$$

$$= f'(a) + \int_0^1 (1-u)f''\big(a + u(x-a)\big)(x-a)du$$

Therefore,

$$f(x) = f(a) + \int_0^1 f'\big(a + u(x-a)\big)du(x-a)$$

$$= f(a) + \left[f'(a) + \int_0^1 (1-u)f''\big(a + u(x-a)\big)(x-a)du\right](x-a)$$

$$= f^{(0)}(a) + f^{(1)}(a)(x-a) + \int_0^1 (1-u)f^{(2)}\big(a + u(x-a)\big)(x-a)^2 du$$

When the above process is generalized,

$$f(x) = f^{(0)}(a) + \frac{f^{(1)}(a)(x-a)}{1!} + \frac{f^{(2)}(a)(x-a)^2}{2!} + \cdots + \frac{f^{(n)}(a)(x-a)^n}{n!}$$

$$+ \int_0^1 \frac{(1-u)^n}{n!} f^{(n+1)}\big(a + u(x-a)\big)(x-a)^{n+1} du$$

The last term, $\int_0^1 \frac{(1-u)^n}{n!} f^{(n+1)}\big(a + u(x-a)\big)(x-a)^{n+1} du$, is called the remainder term and can be expressed as $\frac{f^{(n+1)}(\xi)(x-a)^{n+1}}{n+1!}$, $a < \xi < x$. Finally, the Taylor formula is:

$$f(x) = f^{(0)}(a) + \frac{f^{(1)}(a)(x-a)}{1!} + \frac{f^{(2)}(a)(x-a)^2}{2!} + \cdots + \frac{f^{(n)}(a)(x-a)^n}{n!}$$

$$+ \frac{f^{(n+1)}(\xi)(x-a)^{n+1}}{n+1!} \quad (7)$$

Central limit theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean, μ, and variance, σ^2. Here, $\overline{X}$ is defined as:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Assuming the existence of the moment-generating function, $M_X(t) = E(e^{tX})$, $-h < t < h$, for the distribution, the moment-generating function for $\frac{(\overline{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$ is:

$$M_{Xn}(t) = E\left[\exp\left(t\frac{(\overline{X} - \mu)}{\sigma/\sqrt{n}}\right)\right] = E\left[\exp\left(t\frac{\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu}{\sigma/\sqrt{n}}\right)\right]$$

$$= E\left[\exp\left(t\frac{\left(\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right) \times n}{\sigma/\sqrt{n} \times n}\right)\right]$$

$$= E\left[\exp\left(t\frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}}\right)\right]$$

Xi is independently sampled at random from the population. So,

$$= E\left[\exp\left(t\frac{(X_1 + X_2 + \cdots + X_n) - (\mu + \mu + \cdots + \mu)}{\sigma\sqrt{n}}\right)\right]$$

$$= E\left[\exp\left(t\frac{(X_1 - \mu) + (X_2 - \mu) + \cdots + (X_n - \mu)}{\sigma\sqrt{n}}\right)\right]$$

$$= E\left[\exp\left(t\left\{\frac{(X_1 - \mu)}{\sigma\sqrt{n}} + \frac{(X_2 - \mu)}{\sigma\sqrt{n}} + \cdots + \frac{(X_n - \mu)}{\sigma\sqrt{n}}\right\}\right)\right]$$

$$= E\left[\exp\left(t\frac{(X_1 - \mu)}{\sigma\sqrt{n}} + t\frac{(X_2 - \mu)}{\sigma\sqrt{n}} + \cdots + t\frac{(X_n - \mu)}{\sigma\sqrt{n}}\right)\right]$$

$$= E\left[\exp\left(t\frac{X_1 - \mu}{\sigma\sqrt{n}}\right)\exp\left(t\frac{X_2 - \mu}{\sigma\sqrt{n}}\right)\ldots\exp\left(t\frac{X_n - \mu}{\sigma\sqrt{n}}\right)\right]$$

Since $X_1, X_2, \ldots, X_n$ are from the same population with mean, μ, and variance, σ^2, they have a common moment-generating function.

$$= E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\ldots\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right]$$

$$= E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right]E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right]\ldots E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right] = \left\{E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right]\right\}^n$$

$$\because E(XY) = E(X)E(Y), \quad \text{Please refer to the equation (6).}$$

When we define the moment-generating function for $X - \mu$,

$$m_X(t) = E\big[e^{t(X-\mu)}\big] = E[e^{tX-t\mu}] = E[e^{Xt} \times e^{-\mu t}] = e^{-\mu t}E(e^{Xt}) = e^{-\mu t}M_X(t), \quad -h < t < h$$

Then,

$$\left\{E\left[\exp\left(t\frac{X-\mu}{\sigma\sqrt{n}}\right)\right]\right\}^n = \left\{E\left[\exp\left(\frac{t}{\sigma\sqrt{n}}(X-\mu)\right)\right]\right\}^n = \left\{E\left[\exp\left(-\mu\frac{t}{\sigma\sqrt{n}} + X\frac{t}{\sigma\sqrt{n}}\right)\right]\right\}^n$$

$$= \left\{E\left[\exp\left(-\mu\frac{t}{\sigma\sqrt{n}}\right)\exp\left(X\frac{t}{\sigma\sqrt{n}}\right)\right]\right\}^n = \left\{\exp\left(-\mu\frac{t}{\sigma\sqrt{n}}\right)E\left[\exp\left(X\frac{t}{\sigma\sqrt{n}}\right)\right]\right\}^n$$

$$= \left\{\exp\left(-\mu\frac{t}{\sigma\sqrt{n}}\right)M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n = \left[m_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n, \quad -h < \frac{t}{\sigma\sqrt{n}} < h$$

Using Taylor's formula (equation (7)), the moment-generating function for $X - \mu$, $m_X(t)$ is:

$$m_X(t) = m_X(0) + \frac{m_X{}'(0)(t-0)^1}{1!} + \frac{m_X{}''(\xi)(t-0)^2}{2!}, \quad -h < t < h, \quad 0 < \xi < t$$

Since $m_X(0) = E[e^{0\times(X-\mu)}] = 1$ and $m_X{}'(0) = E(X-\mu) = E(X) - \mu = \mu - \mu = 0$, according to equation (3),

$$m_X(t) = 1 + \frac{m_X{}''(\xi)t^2}{2!}$$

By adding and subtracting $\sigma^2 t^2/2$,

$$m_X(t) = 1 + \frac{m_X{}''(\xi)t^2}{2!} + \frac{\sigma^2 t^2}{2} - \frac{\sigma^2 t^2}{2} = 1 + \frac{\sigma^2 t^2}{2} + \frac{m_X{}''(\xi)t^2}{2} - \frac{\sigma^2 t^2}{2}$$

$$= 1 + \frac{\sigma^2 t^2}{2} + \frac{[m_X{}''(\xi) - \sigma^2]t^2}{2}$$

When t is replaced by $\frac{t}{\sigma\sqrt{n}}$,

$$m_X\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2}{2} + \frac{[m_X{}''(\xi) - \sigma^2]\left(\frac{t}{\sigma\sqrt{n}}\right)^2}{2} = 1 + \frac{\sigma^2 \times \frac{t^2}{\sigma^2 n}}{2} + \frac{[m_X{}''(\xi) - \sigma^2]\frac{t^2}{\sigma^2 n}}{2}$$

$$= 1 + \frac{t^2}{2n} + \frac{[m_X{}''(\xi) - \sigma^2]t^2}{2n\sigma^2}, \quad -h < \frac{t}{\sigma\sqrt{n}} < h, \quad 0 < \xi < \frac{t}{\sigma\sqrt{n}}$$

Therefore,

$$M_{Xn}(t) = \left[m_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 + \frac{t^2}{2n} + \frac{[m_X{}''(\xi) - \sigma^2]t^2}{2n\sigma^2}\right]^n, \quad -h\sigma\sqrt{n} < t < h\sigma\sqrt{n},$$

$$0 < \xi < \frac{t}{\sigma\sqrt{n}}$$

In $0 < \xi < \frac{t}{\sigma\sqrt{n}}$, as $n \to \infty$, $\xi \to 0$. Since $m_X''(0) = E[(X - \mu)^2] = \sigma^2$, based on equation (4) and the definition of the variance of X (the mean of the square of the deviation of a random variable, X, from its mean, μ), $\lim_{n\to\infty} m_X''(\xi) = \sigma^2 \Leftrightarrow \lim_{n\to\infty}[m_X''(\xi) - \sigma^2] = 0$.

As n becomes infinite, the moment-generating function for $\frac{(\bar{x}-\mu)}{\frac{\sigma}{\sqrt{n}}}$ is defined as $M_{Xn}(t)$. Therefore,

$$\lim_{n\to\infty} M_{Xn}(t) = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + \frac{[m_X''(\xi) - \sigma^2]t^2}{2n\sigma^2}\right]^n = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n}\right]^n \quad \because \lim_{n\to\infty}[m_X''(\xi) - \sigma^2] = 0$$

As is generally known, $\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = \lim_{n\to\infty}\left(1 + 1 \times \frac{1}{n}\right)^n = e^1$. Thus,

$$\lim_{n\to\infty} M_{Xn}(t) = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n}\right]^n = \lim_{n\to\infty}\left[1 + \frac{t^2}{2} \times \frac{1}{n}\right]^n = e^{t^2/2} = \exp(0 \times t + \frac{1^2 \times t^2}{2}), \quad -\infty < t < \infty$$

According to equation (5), the moment-generating function for the normal distribution with mean, μ, and variance, σ^2, is:

$$M_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})$$

We conclude that the distribution of $\frac{(\bar{X}-\mu)}{\frac{\sigma}{\sqrt{n}}}$, with a fixed natural number, n, approximates the normal distribution with mean 0 and variance 1. Thus, if the size (n) of $\bar{X}$ (the mean of a random sample from a distribution with mean, μ, and variance, σ^2) is sufficiently large, the distribution of $\bar{X}$ follows the normal distribution with mean, μ, and variance, $\frac{\sigma^2}{n}$, regardless of the population distribution.