

Improving Accuracy of Normal Approximation

Jan Vrbik

Department of Mathematics, Brock University, St. Catharines, Ontario, Canada

Email: jvrbik@brocku.ca

How to cite this paper: Vrbik, J. (2025) Improving Accuracy of Normal Approximation. *Open Journal of Statistics*, 15, 473-491.

<https://doi.org/10.4236/ojs.2025.156025>

Received: October 17, 2025

Accepted: December 13, 2025

Published: December 16, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Central Limit Theorem occasionally results in extremely slow convergence to the Normal distribution, making the corresponding approximation virtually useless (a typical example being the sample correlation coefficient). In most of these cases, there is a relatively easy way of substantially improving the approximation by including a few extra terms of the corresponding Edgeworth expansion to match not only the distribution's mean and variance, but also its skewness and kurtosis. A further improvement is possible by transforming the studied sample statistic to fully remove the skewness; this simplifies the resulting approximation, making it also more accurate. All this is amply demonstrated by many practical examples.

Keywords

Central Limit Theorem, Distribution's Cumulants, Edgeworth Series, Sample-Mean Transformation, Skewness Removal

1. Introduction

This article demonstrates how the distribution of any sample statistic meeting the conditions of the Central Limit Theorem (CLT) [1] and having the first four *finite* moments can be approximated with high accuracy by considering two more terms of the corresponding Edgeworth expansion [2]; this then allows us to use the correspondingly improved approximation even with relatively small sample sizes. The key component essential for the new technique to achieve this goal lies in the proper utilization of *cumulants* of a distribution; the article's first section provides a summary of relevant formulas.

The second section then explains how cumulants help to find higher moments of a *sample mean*, which then enables us to extend the usual Normal approximation of its distribution by two extra terms (as done in Section 3), to match the distribution's *skewness* and *kurtosis*. This yields a substantial improvement in the

resulting accuracy, as illustrated by several examples.

We then proceed (in Section 4) to show how the same approach applies to any *function* of a sample mean, and how to utilize this method to find a specific function to remove the resulting skewness (in Section 5), automatically achieving a further, often dramatic, improvement. Section 6 demonstrates how to extend the new technique to a function of *several* sample means.

2. Distribution's Cumulants

It is well known that expanding the *moment generating function* (MGF), defined as $M_X(t) := \mathbb{E}(e^{tX})$, of a *random variable* (RV) X in powers of t provides an easy way of computing *simple* moments, defined by $\mu_j := \mathbb{E}(X^j)$, of the corresponding distribution, thus

$$M_X(t) = 1 + \sum_{j=1}^{\infty} \frac{\mu_j t^j}{j!} \tag{1}$$

This implies that, when similarly expanding

$$M_{X-\mu}(t) = M_X(t)e^{-\mu t} = 1 + \sum_{j=2}^{\infty} \frac{\mu_j t^j}{j!} \tag{2}$$

where $\mu := \mu_1$ is the *expected value* of X , the coefficients of this expansion are $\mu_j := \mathbb{E}[(X - \mu)^j]$, i.e. *central* moments of the same distribution.

Finally, expanding the following *cumulant generating function* (CGF)

$$C(t) := \ln M_X(t) = \sum_{j=1}^{\infty} \frac{\kappa_j t^j}{j!} \tag{3}$$

defines the distribution's *cumulants* [3], denoted κ_j . Since MGF of a sum of independent RVs is a *product* of the individual MGFs, the corresponding CGF becomes the *sum* of the individual $C(t)$ functions; this implies that, to get the j^{th} cumulant of an *independent sum*, all we have to do is to *add* the individual κ_j 's.

The following formula

$$\ln M_{X-\mu}(t) = \ln M_X(t) - \mu t \tag{4}$$

implies that

$$\ln \left(1 + \sum_{j=2}^{\infty} \frac{\mu_j t^j}{j!} \right) = \sum_{j=2}^{\infty} \frac{\kappa_j t^j}{j!} \tag{5}$$

which shows that all cumulants (with the exception of $\kappa_1 = \mu$) are functions of *central* moments. For the first several such cumulants we get: $\kappa_2 = \mu_2 = \sigma^2$ (the variance), $\kappa_3 = \mu_3$, $\kappa_4 = \mu_4 - 3\mu_2^2$, etc.

When similarly expanding $\exp \left(\sum_{j=2}^{\infty} \frac{\kappa_j t^j}{j!} \right)$ in powers of t , the coefficient of t^j yields $\frac{\mu_j}{j!}$, resulting in

$$\begin{aligned}
 \mu_2 &= \kappa_2 & (6) \\
 \mu_3 &= \kappa_3 \\
 \mu_4 &= \kappa_4 + 3\kappa_2^2 \\
 \mu_5 &= \kappa_5 + 10\kappa_3\kappa_2 \\
 \mu_6 &= \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3 \\
 &\vdots
 \end{aligned}$$

Note that the coefficients of individual terms in each such μ_j expansion equal to the number of different *partitionings* of a group of j distinct objects into subsets whose sizes are given by the corresponding κ indices, assuming that interchanging subsets of the same size is inconsequential (e.g. dividing 6 people into three groups of 2 people each can be done in $\binom{6}{2,2,2}/3! = 15$ ways, giving us the coefficient of κ_2^3 in the expansion of μ_6). Also note that subsets of size *one* are not allowed.

Often, we need cumulants of a *function* of X , say $Y = h(X)$; in many such cases it becomes impossible to find the MGF of Y in an analytic form. Nevertheless, it is usually still possible to compute the first few moments of Y (by *numerical* integration, if necessary) and convert them to the same number of cumulants.

EXAMPLE 1: To get the first four cumulants of $Y := \frac{X^2}{1+X^2}$, where X has Cauchy distribution with the median of 0 and semi-inter-quartile range of 1, we first find the first four simple moments of Y by

$$\dot{\mu}_j = \frac{1}{\pi} \int_{-\infty}^{\infty} \left(\frac{x^2}{1+x^2} \right)^j \frac{dx}{1+x^2}$$

(in this case, it can be done analytically), getting $\dot{\mu}_j = \frac{1}{2}, \frac{3}{8}, \frac{5}{16}$ and $\frac{35}{128}$ when $j = 1, 2, 3$ and 4 respectively. Expanding

$$C_y(t) = \ln \left(1 + \sum_{j=1}^4 \frac{\dot{\mu}_j t^j}{j!} \right) = \frac{1}{2} \cdot t + \frac{3}{4} \cdot \frac{t^2}{2} + \frac{128}{3} \cdot \frac{t^4}{4!}$$

then yields the first four cumulants of Y (note that, coincidentally, $\kappa_3 = 0$).

EXAMPLE 2: Similarly, when sampling Gamma distribution whose PDF is given by

$$\frac{x^{\alpha-1} \exp(-x)}{\Gamma(\alpha)} \text{ when } x > 0$$

and aiming to get the first four cumulants of $Y := \ln X$, we first evaluate the corresponding first four simple *moments* of $\ln X$ and then convert them, in the same manner, to cumulants; this results in $\kappa_j = \psi^{(j-1)}(\alpha)$, $\psi'(\alpha)$, $\psi''(\alpha)$, $\psi'''(\alpha)$ for $j = 1$ to 4, where $\psi(\alpha)$ is the first derivative of $\ln \Gamma(\alpha)$.

3. Moments of a Sample Mean

Suppose that, instead of considering a single X , we take a *random independent sample* (RIS) of size n from the X distribution and seek the moments of the corresponding *sample mean*

$$\bar{X} := \frac{\sum_{i=1}^n X_i}{n} \quad (7)$$

To find the k^{th} central moment of \bar{X} (a non-trivial task, unless cumulants are used) which we denote $\bar{\mu}_k$, we first notice that finding the corresponding cumulants (up to and including the k^{th}) is substantially easier, since the j^{th} *cumulant* of $\sum_{i=1}^n X_i$ is simply $n\kappa_j$. This result must be further divided by n^j to account for the denominator in the definition of \bar{X} , thus leading to the following simple formula

$$\bar{\kappa}_j = \frac{\kappa_j}{n^{j-1}} \quad (8)$$

We can now find $\bar{\mu}_k$ since we know how to express moments in terms of cumulants (what was true for X is still true for \bar{X}). Thus, for example

$$\bar{\mu}_6 = \bar{\kappa}_6 + 15\bar{\kappa}_4\bar{\kappa}_2 + 10\bar{\kappa}_3^2 + 15\bar{\kappa}_2^3 = \frac{\kappa_6}{n^5} + 15\frac{\kappa_4\kappa_2}{n^4} + 10\frac{\kappa_3^2}{n^4} + 15\frac{\kappa_2^3}{n^3}$$

where we note that individual terms of this expansion decrease with n in a different manner (the first term does it a lot faster than the last one). Also note that, when the distribution of X is *symmetric* (its PDF being an *even* function of $x - \mu$), both μ_{2k+1} and κ_{2k+1} (and their \bar{X} counterparts) are equal to zero for every positive k .

In summary, $\bar{\mu}_j := \mathbb{E}\left[(\bar{X} - \mu)^j\right]$ can be found as the coefficient of $\frac{t^j}{j!}$ in the expansion of

$$\exp\left(n \ln M\left(\frac{t}{n}\right) - \mu t\right) \quad (9)$$

where $M(t)$ is the MGF of a single X .

4. Central Limit Theorem (CLT)

Consider RIS of size n from a distribution with a mean of μ , standard deviation of σ (both must be finite), and its MGF denoted $M(t)$. CLT states that the distribution of

$$Z_n := \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \quad (10)$$

tends, as $n \rightarrow \infty$, to that of the *standardized* (0 mean, variance of 1) Normal distribution.

Proof: The MGF of Z_n is found by taking the MGF of $X - \mu$, namely

$$\exp\left(\sum_{j=2}^{\infty} \frac{\kappa_j t^j}{j!}\right) \tag{11}$$

and raising it to the power of n while (at the same time) replacing t by $\frac{t}{\sigma\sqrt{n}}$; this results in

$$\exp\left(\sum_{j=2}^{\infty} \frac{\kappa_j t^j}{n^{j/2-1} \sigma^j j!}\right) = \exp\left(\frac{t^2}{2}\right) \cdot \exp\left(\sum_{j=3}^{\infty} \frac{\kappa_j t^j}{n^{j/2-1} \sigma^j j!}\right) \tag{12}$$

where the RHS clearly tends, as $n \rightarrow \infty$, to $\exp\left(\frac{t^2}{2}\right)$ of the $\mathcal{N}(0,1)$ distribution.

5. Edgeworth Expansion

To achieve a better approximation (called *Edgeworth*), we expand the *second* factor of (12) in powers of $\frac{1}{\sqrt{n}}$ up to and including $\frac{1}{n}$ -proportional terms, thus getting

$$1 + \frac{\kappa_3 t^3}{3! \sigma^3 \sqrt{n}} + \frac{\kappa_3^2 t^6}{2 \cdot 3!^2 \sigma^6 n} + \frac{\kappa_4 t^4}{4! \sigma^4 n} = 1 + \frac{\gamma_3 t^3}{6\sqrt{n}} + \frac{3\gamma_4 t^4 + \gamma_3^2 t^6}{72n} \tag{13}$$

where $\gamma_3 := \frac{\kappa_3}{\sigma^3}$ and $\gamma_4 := \frac{\kappa_4}{\sigma^4}$ are the distribution's *skewness* and *excess kurtosis*, respectively.

To extend CLT by these terms, we need to convert the resulting MGF to the corresponding *probability density function* (PDF); in our case, it is sufficient to find out how to convert $\exp\left(\frac{t^2}{2}\right) \cdot t^j$, needed with $j = 3, 4$ and 6 only. To do that, we start with the well known result (used to compute the MGF of standardized Normal distribution) of

$$\int_{-\infty}^{\infty} \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \exp(zt) dz = \exp\left(\frac{t^2}{2}\right) \tag{14}$$

and realize that, in general

$$\begin{aligned} \int_{-\infty}^{\infty} H(z) \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \exp(zt) dz &= \int_{-\infty}^{\infty} H(z) \frac{\exp\left(-\frac{z^2}{2}\right) d \exp(zt)}{\sqrt{2\pi} dz} dz \\ &= \int_{-\infty}^{\infty} (zH(z) - H'(z)) \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \exp(zt) dz \end{aligned} \tag{15}$$

(using by-part integration) when $H(z)$ is a *polynomial*. Starting with $H_0(z) = 1$, this yields $H_1(z) = zH_0(z) - H'_0(z) = z$, thus proving that

$\exp\left(\frac{t^2}{2}\right) \cdot t$ converts to $\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot z$. Using (15) repeatedly, we get

$H_2(z) = zH_1(z) - H_1'(z) = z^2 - 1$
$H_3(z) = zH_2(z) - H_2'(z) = z(z^2 - 3)$
$H_4(z) = z^4 - 6z^2 + 3$
$H_5(z) = z(z^4 - 10z^2 + 15)$
$H_6(z) = z^6 - 15z^4 + 45z^2 - 15$

(16)

letting us convert $\exp\left(\frac{t^2}{2}\right) \cdot t^j$ to $\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot H_j(z)$ for all the j values needed in our approximation ($H_j(z)$ is usually referred to as *Hermite* polynomial of degree j).

The new, improved (even though still approximate) PDF of Z_n is thus

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot \left(1 + \frac{\kappa_3 H_3(z)}{6\sigma^3 \sqrt{n}} + \frac{\kappa_3^2 H_6(z)}{72\sigma^6 n} + \frac{\kappa_4 H_4(z)}{24\sigma^4 n}\right) \tag{17}$$

The formula's *error* decreases with $\frac{1}{n^{3/2}}$, unlike that of the basic Normal approximation, whose error decreases with only $\frac{1}{\sqrt{n}}$. To be able to construct the new approximation, the sampled distribution's first *four* moments need to be finite.

The corresponding *cumulative distribution function* (CDF) of Z_n is given by

$$\Phi(z) - \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot \left(\frac{\kappa_3 H_2(z)}{6\sigma^3 \sqrt{n}} + \frac{\kappa_3^2 H_5(z)}{72\sigma^6 n} + \frac{\kappa_4 H_3(z)}{24\sigma^4 n}\right) \tag{18}$$

where Φ is the CDF of the standardized Normal distribution. This is clear from

$$\begin{aligned} \int H_{k+1}(z) \exp\left(-\frac{z^2}{2}\right) dz &= \int (zH_k(z) - H_k'(z)) \exp\left(-\frac{z^2}{2}\right) dz \\ &= -H_k(z) \exp\left(-\frac{z^2}{2}\right) \end{aligned}$$

where the last step is a result of a by-part integration of $\int -H_k'(z) \exp\left(-\frac{z^2}{2}\right) dz$.

Using (18) to find a specific *quantile* of the \bar{X} distribution (not a goal of this article) would lead to a modified (due to different standardization of \bar{X}) version of Cornish-Fisher expansion [4].

It is possible to derive more accurate formulas by including further terms of the Edgeworth expansion; this would lead to results of high complexity and dimin-

ishing returns (we do not pursue it any further).

EXAMPLE 3: Suppose X has an exponential distribution with $\mu = 1$. This time, rather uniquely and deliberately (to test the accuracy of the new approximation), we know the exact PDF of

$$Z_n := \frac{\sum_{i=1}^n X_i - n}{\sqrt{n}}$$

to be

$$\frac{(z + \sqrt{n})^{n-1} \exp\left(-\left(z + \sqrt{n}\right)\sqrt{n}\right)}{(n-1)!} n^{n/2} \quad \text{for } z > -\sqrt{n} \tag{19}$$

Based on $\ln M(t) = -\ln(1-t)$, we can easily find $\kappa_2 = 1$, $\kappa_3 = 2$ and $\kappa_4 = 6$; the corresponding Edgeworth approximation to the PDF of Z_n then reads

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \left(1 + \frac{z(z^2 - 3)}{3\sqrt{n}} + \frac{2z^6 - 21z^4 + 36z^2 - 3}{36n} \right) \tag{20}$$

which can be readily transformed into the PDF of \bar{X} , if desired.

In the following table, we compare the maximum errors of Edgeworth and Normal approximations for three different values of n .

n	Edgeworth	Normal
10	0.42%	6.7%
20	0.14%	4.6%
30	0.045%	3.2%

One can see that the improvement achieved by the new technique is truly substantial; the table also gives a good demonstration of the new error decreasing with $n^{-3/2}$, instead of $n^{-1/2}$ of the last column. Note that a graph of the difference between exact and approximate PDFs always alternates between positive and negative areas; the largest of these (in absolute value) defines the maximum error.

Also note that (20) can be confirmed by directly expanding (19) in powers of $\frac{1}{\sqrt{n}}$.

EXAMPLE 4: Sampling Uniform distribution (with values from 0 to 1) whose MGF is

$$M(t) = \frac{\exp(t) - 1}{t}$$

(note the removable singularity at $t = 0$) and whose first four cumulants are $\frac{1}{2}$,

$\frac{1}{12}$, 0 (due to its symmetry with respect to the mean value) and $-\frac{1}{120}$, the ap-

proximate PDF of

$$Z_n := \frac{\bar{X} - \kappa_1}{\sqrt{\frac{\kappa_2}{n}}} = \frac{\bar{X} - \frac{1}{2}}{\sqrt{\frac{1}{12n}}}$$

is thus

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \left(1 + \frac{\kappa_4 H_4(z)}{24\sigma^4 n}\right) = \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \left(1 - \frac{H_4(z)}{20n}\right)$$

Since the exact PDF of $\sum_{i=1}^n X_i$ is also known (it is called *Irwin-Hall* distribution) and easily transformable to Z_n , we can still compare the maximum error of the Normal approximation, which is 0.56% when $n = 10$ (tolerable but large) to that of the last formula, which equals 0.016% for the same value of n . Note that this large improvement has been achieved by only a single extra $\frac{1}{n}$ -proportional term; not surprisingly, *symmetric* distributions result in simpler and more accurate approximations.

6. Transforming \bar{X}

In this section, we discuss how to find a similar approximation for the distribution of a *function* of \bar{X} (still asymptotically Normal under most circumstances, specified shortly).

The first step is to expand the function (denoted g) with respect to its argument \bar{X} (now taken to be a strictly Mathematical variable) at $\bar{X} = \mu$ (the corresponding expected value), getting

$$g(\bar{X}) = g(\mu) + g'(\mu)(\bar{X} - \mu) + \frac{g''(\mu)}{2}(\bar{X} - \mu)^2 + \frac{g'''(\mu)}{3!}(\bar{X} - \mu)^3 + \dots \quad (21)$$

To be able to do this, we require the function to be sufficiently smooth (the three derivatives must exist and be finite); for our purpose, the value of $g'(\mu)$ must also *not* equal to zero. Note that to achieve $\frac{1}{n}$ accuracy of the resulting PDF (i.e. having an error proportional to $n^{-3/2}$), we do not need higher than *cubic* terms in this expansion.

We already know that

$$\begin{aligned} \bar{\mu}_2 &= \frac{\kappa_2}{n} \\ \bar{\mu}_3 &= \frac{\kappa_3}{n^2} \\ \bar{\mu}_4 &= 3 \frac{\kappa_2^2}{n^2} + \frac{\kappa_4}{n^3} \\ \bar{\mu}_5 &= 10 \frac{\kappa_3 \kappa_2}{n^3} + \dots \end{aligned} \quad (22)$$

$$\bar{\mu}_6 = 15 \frac{\kappa_2^3}{n^3} + \dots$$

where $+\dots$ stands for terms whose contribution goes beyond the desired $\frac{1}{n}$ accuracy, and which can be thus discarded. Also note that only the first *four* cumulants of the sampled distribution are needed. The last set of formulas is then used to find the first four cumulants of $g(\bar{X})$, each to sufficient accuracy in terms of the corresponding $1/n$ expansion. Denoting them K_1 to K_4 we get, after some algebra (all g derivatives are evaluated at μ)

$$K_1 =: \mathbb{E}[g(\bar{X})] = g(\mu) + \frac{1}{2} \frac{g'' \kappa_2}{n} + \dots \tag{23}$$

$$K_2 =: \mathbb{E}[(g(\bar{X}) - K_1)^2] = \frac{(g')^2 \kappa_2}{n} + \frac{g' g'' \kappa_3 + g' g''' \kappa_2^2 + \frac{1}{2} (g'')^2 \kappa_2^2}{n^2} + \dots$$

$$K_3 =: \mathbb{E}[(g(\bar{X}) - K_1)^3] = \frac{(g')^2 (g' \kappa_3 + 3g'' \kappa_2^2)}{n^2} + \dots$$

$$K_4 =: \mathbb{E}[(g(\bar{X}) - K_1)^4] - 3K_2^2 = \frac{(g')^4 \kappa_4 + 12(g')^3 g'' \kappa_3 \kappa_2 + 4(g')^3 g''' \kappa_2^3 + 12(g')^2 (g'')^2 \kappa_2^3}{n^3} + \dots$$

For efficient execution of the algorithm, the following is of utmost importance: to achieve the overall $1/n$ accuracy of the resulting approximation, K_1 needs to be only $1/n$ accurate (thus ignoring terms beyond *quadratic* in the $g(\bar{X})$ expansion), while both K_2 and K_3 require first expanding the corresponding powers of $g(\bar{X}) - K_1$ up to and including *quartic* terms in $\bar{X} - \mu$ (to make each result $1/n^2$ accurate). Finally, to reach the $1/n^3$ accuracy of K_4 , the 4th power of $(g(\bar{X}) - K_1)^4$ needs to be expanded up to *hexic* terms, before taking its expected value.

The approximate PDF of

$$Z_n := \frac{g(\bar{X}) - K_1}{\sqrt{K_2}} \tag{24}$$

is then given by

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot \left(1 + \frac{K_3 H_3(z)}{6\hat{K}_2^{3/2}} + \frac{K_3^2 H_6(z)}{72\hat{K}_2^3} + \frac{K_4 H_4(z)}{24\hat{K}_2^2}\right) \tag{25}$$

where

$$\hat{K}_2 := \frac{(g')^2 \kappa_2}{n} \tag{26}$$

is sufficiently accurate for this part of the result (but *not* when computing Z_n).

Note that the *basic Normal* approximation in this case states that

$$Z_n := \frac{g(\bar{X}) - g(\mu)}{\sqrt{\frac{(g')^2 \kappa_2}{n}}} \tag{27}$$

has, approximately, the standardized Normal distribution.

We should point out that, to find Edgeworth approximation to the PDF of $g(\bar{h}(X))$, where $h(X)$ is a function of X , requires the κ_j cumulants in (23) to be the cumulants of $h(X)$, *not* of X ; the rest of the procedure remains the same.

EXAMPLE 5: To get an approximate PDF of

$$Z_n := \frac{\ln(\bar{X}) - K_1}{\sqrt{K_2}}$$

(implying that $g = \ln$) when sampling Exponential distribution with $\mu = 1$, we start with $g' = 1$, $g'' = -1$ and $g''' = 2$. Since we already know the first four cumulants of the sampled distribution, it is now routine to compute $K_1 = -\frac{1}{2n}$, $K_2 = \frac{1}{n} + \frac{1}{2n^2}$, $K_3 = -\frac{1}{n^2}$ and $K_4 = \frac{2}{n^3}$, resulting in the following approximate PDF of Z_n

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot \left(1 - \frac{H_3(z)}{6\sqrt{n}} + \frac{6H_4(z) + H_6(z)}{72n}\right)$$

The maximum error of this approximation (when $n = 10$) is 0.11%; a significant improvement over the error of basic Normal approximation, which equals 4.3%.

EXAMPLE 6: Similarly, when sampling Gamma distribution of Example 2, the *maximum-likelihood* (ML) estimator of α is the (numerical) solution to

$$\psi(\hat{\alpha}) = \overline{\ln X} = \psi(\alpha) + (\overline{\ln X} - \psi(\alpha)) \tag{28}$$

where $\hat{\alpha}$ (a function of $\overline{\ln X}$) denotes this estimator. Since this function is defined only *implicitly* (ψ does not have an analytic inverse), getting the corresponding g_i values (coefficients of the (21) expansion, where $\hat{\alpha}$ stands for the g function, i is the order of its derivative and the evaluation is done at $\mu = \psi(\alpha)$ which, in this context, are all implicit operations) is now more difficult. Expanding the RHS of

$$\psi(\hat{\alpha}) = \psi\left(g_0 + g_1x + g_2\frac{x^2}{2} + g_3\frac{x^3}{6}\right)$$

where $x := \overline{\ln X} - \psi(\alpha)$, results in

$$\psi(g_0) + g_1\psi'x + \left(g_1^2\psi'' + g_2\psi'\right)\frac{x^2}{2} + \left(g_1^3\psi''' + 3g_1g_2\psi'' + g_3\psi'\right)\frac{x^3}{6}$$

with the ψ derivatives deemed evaluated at the *true* value of α . Matching this to the RHS of (28) yields $g_0 = \alpha$, $g_1 = \frac{1}{\psi'}$, $g_2 = -\frac{\psi''}{(\psi')^3}$ and

$$g_3 = -\frac{\psi'''}{(\psi')^4} + \frac{3(\psi'')^2}{(\psi')^5}.$$

Based on (23), we can now build the corresponding Edgeworth approximation for the PDF of $\hat{\alpha}$, given values of α and n (we stop here, as the point of this example was only to show how to expand an implicit function of a sample mean; the rest is routine). We should also mention that there is a more direct way of building an approximate PDF of ML estimators [5], not to be discussed in this article. ■

7. Removing Skewness

The technique of the previous section can be utilized in the following way: instead of being given a specific function of \bar{X} , we may seek a function g which *eliminates* (to the level of our approximation) the *third* central moment (and thus skewness) of $g(\bar{X})$. The reason for doing this is that the resulting PDF is simpler and more accurate than the one built directly for \bar{X} , thus providing the most accurate approximation yet for the distribution of \bar{X} itself (after the corresponding back transformation).

When the X distribution is *parameter-free* and *positive*, this goal can always be achieved by g being a simple *power* of \bar{X} .

Proof: This can be seen from the fact that the K_3 cumulant of $g(\beta\bar{X})$ is proportional to $g'\beta^3\kappa_3 + 3g''(\beta^2\kappa_2)^2$ and can be made equal to zero by solving

$$g'(\beta\mu)\kappa_3 + 3g''(\beta\mu)\beta\kappa_2^2 = 0$$

or, equivalently,

$$g'(x)\mu\kappa_3 + 3xg''(x)\kappa_2^2 = 0 \tag{29}$$

The simplest solution is

$$g(x) = x^{1-\mu\kappa_3/(3\kappa_2^2)} \tag{30}$$

EXAMPLE 7: For Exponential distribution with $\mu = 1$, $\kappa_2 = 1$ and $\kappa_3 = 2$, this would be achieved by

$$g(\bar{X}) = \bar{X}^{1/3}$$

which leads to $g_1 = \frac{1}{3}$, $g_2 = -\frac{2}{9}$ and $g_3 = \frac{10}{27}$, resulting in $K_1 = 1 - \frac{1}{9n}$,

$K_2 = \frac{1}{9n} + \frac{0}{n^2}$, $K_3 = 0$ (by design) and $K_4 = -\frac{2}{(9n)^3}$. Therefore

$$Z_n := \left(\bar{X}^{1/3} - 1 + \frac{1}{9n} \right) \sqrt{9n} \tag{31}$$

has the following (still approximate, but now highly accurate) PDF

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot \left(1 - \frac{H_4(z)}{108n}\right) \tag{32}$$

This result is easily transformed back to a PDF of \bar{X} (substitute the RHS of (31) for z in (32) and multiply the result by $\frac{dz}{d\bar{x}}$). The maximum error of this approximation when $n = 10$ is 0.036%; a very substantial improvement yet over the approximation of Example 3.

To use results of this example when sampling Exponential distribution with a mean of β , we just replace $\bar{X}^{1/3}$ in (31) by $\bar{X}^{1/3}/\beta^{1/3}$, the rest stays the same.

■

Another possibility of removing skewness of a *parameter-free* distribution (no longer necessarily positive) is to realize that the K_3 cumulant of $g(X + a)$ is proportional to

$$g'(\mu + a)\kappa_3 + 3g''(\mu + a)\kappa_2^2$$

and can thus be made equal to zero by

$$g(x) = \exp\left(-\frac{K_3x}{3\kappa_2^2}\right) \tag{33}$$

EXAMPLE 8: For Gumbel distribution whose PDF is

$$\exp(x - e^x)$$

and whose first four cumulants are $-\gamma$ (Euler gamma), $\frac{\pi^2}{6}$, $\psi''(1)$ and $\frac{\pi^4}{15}$ respectively, this implies that

$$\exp(A \cdot \bar{X})$$

where

$$A = -\frac{12\psi''(1)}{\pi^4} \approx -0.29613$$

results in $K_1 = \exp(-A\gamma) + \frac{0.07876}{n}$, $K_2 = \frac{0.17195}{n} - \frac{0.0372}{n^2}$ and

$K_4 = -\frac{0.01436}{n^3}$ while eliminating skewness. This means that the PDF of

$$Z_n := \frac{\exp(A \cdot \bar{X}) - K_1}{\sqrt{K_2}}$$

is, to a good accuracy, given by

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \cdot \left(1 + \frac{K_4 H_4(z)}{24\hat{K}_2^2}\right) \tag{34}$$

easily transformed into an approximate PDF of \bar{X} .

Unfortunately, this time we cannot compare our solution with an exact answer

which is impossible to derive analytically; the only way to establish its maximum error is to generate an *empirical distribution* of \bar{X} , using as many RIS's as possible (in this case, we have used four million samples of size 10). Luckily, the resulting error of this distribution is smaller than the error of our Edgeworth approximation; this enables us only to *estimate* the maximum error of the latter not to exceed 0.25% (when $n = 10$), while that of the basic Normal approximation is almost 3.5% (beyond acceptable).

We should mention that the above simulation took less than one second to complete, while converting the resulting grand sample of four million $\exp(A \cdot \bar{X})$ values into the corresponding empirical PDF required about 35 seconds; both computations were done by Mathematica. ■

Note that the same skewness removal (and resulting accuracy improvement) can be achieved by applying a similarly designed g function to $h(X, Y)$ (easily extended to sample mean of functions of three or more random variables).

EXAMPLE 9: Consider a RIS of size n from a bivariate Normal distribution with both means equal to 0, both variances equal to 1, and a correlation coefficient equal to $\frac{1}{2}$. To find a highly accurate approximation to the PDF of $\frac{X^2}{X^2 + 2Y^2}$,

we first compute the first four cumulants of $\frac{X^2}{X^2 + 2Y^2}$, getting

$\kappa_1 = \mu = 0.40825$, $\kappa_2 = 0.11237$, $\kappa_3 = 0.01547$ and $\kappa_4 = -0.01563$; this implies, based on (30), that $g(x) = x^{0.83333}$ is the corresponding skewness-removing function, leading to $K_1 = 0.47399 - \frac{0.02219}{n}$, $K_2 = \frac{0.10519}{n} + \frac{0.00887}{n^2}$ and $K_4 = -\frac{0.01287}{n^3}$. The final result is that

$$Z_n = \frac{\left(\frac{X^2}{X^2 + 2Y^2}\right)^{0.83333} - K_1}{\sqrt{K_2}}$$

has a PDF of (34); its maximum error (established from the corresponding empirical distribution) when $n = 10$ is about 0.2%.

8. Function of Two Sample Means

The Edgeworth approximation is somehow more difficult to construct for a function of *two* sample means; the main complication arises from the fact that this time, instead of a single μ_k , there are several ($k + 1$, to be specific) k -order central moments, defined by

$$\mu_{ij} := \mathbb{E}\left[(X - \mu_x)^i (Y - \mu_y)^j\right] \tag{35}$$

(where $i + j = k$), and the same applies to cumulants.

The central moments can be found from a joint MGF of $X - \mu_x$ and $Y - \mu_y$, namely from

$$M_c(t, u) := \exp(-\mu_x t - \mu_y u) \cdot \mathbb{E}[\exp(tX + uY)] = 1 + \sum_{i+j \geq 2} \frac{\mu_{ij} t^i u^j}{i! j!} \tag{36}$$

by differentiating it i times with respect to t and j times with respect to u , and evaluating at $t = u = 0$. When analytic form of this MGF is not available, we may first expand $\exp(tX + uY)$ to the required order and then take the expected value (by numerical integration, if necessary) of the resulting expansion.

Cumulants are similarly found by either expanding or differentiating $\ln M_c(t, u)$; for the second and third-order cumulants, we still get $\kappa_{ij} = \mu_{ij}$, the forth-order formulas become more involved, namely

$$\begin{aligned} \kappa_{40} &= \mu_{40} - 3\mu_{20}^2 & (37) \\ \kappa_{31} &= \mu_{31} - 3\mu_{20}\mu_{11} \\ \kappa_{22} &= \mu_{22} - \mu_{20}\mu_{02} - 2\mu_{11}^2 \end{aligned}$$

We can then get $\bar{\mu}_{ij}$, from an obvious generalization of (9), namely

$$\exp\left(n \ln M\left(\frac{t}{n}, \frac{u}{n}\right) - \mu_x t - \mu_y u\right) \tag{38}$$

as the coefficient of $\frac{t^i u^j}{i! j!}$ in the corresponding bivariate expansion of this function.

This results in the following joint central moments of \bar{X} and \bar{Y} (quoting only terms needed for our $1/n$ accurate approximation)

$$\begin{aligned} \bar{\mu}_{20} &= \frac{\kappa_{20}}{n} & (39) \\ \bar{\mu}_{11} &= \frac{\kappa_{11}}{n} \\ \bar{\mu}_{31} &= \frac{3\kappa_{20}\kappa_{11}}{n^2} + \frac{\kappa_{31}}{n^3} \\ \bar{\mu}_{22} &= \frac{\kappa_{20}\kappa_{02} + 2\kappa_{11}^2}{n^2} + \frac{\kappa_{22}}{n^3} \\ \bar{\mu}_{41} &= \frac{10\kappa_{21}\kappa_{20}}{n^3} + \dots \\ \bar{\mu}_{32} &= \frac{\kappa_{30}\kappa_{02} + 6\kappa_{21}\kappa_{11} + 3\kappa_{20}\kappa_{12}}{n^3} + \dots \\ \bar{\mu}_{51} &= \frac{15\kappa_{20}^2\kappa_{02}}{n^3} + \dots \\ \bar{\mu}_{42} &= \frac{3\kappa_{20}^2\kappa_{02} + 12\kappa_{20}\kappa_{11}^2}{n^3} + \dots \\ \bar{\mu}_{33} &= \frac{9\kappa_{20}^2\kappa_{02} + 6\kappa_{20}\kappa_{11}^2}{n^3} + \dots \end{aligned}$$

in analogy with (22). Note that the coefficients in each numerator equal the number of ways to divide i males and j females into subgroups of size specified by

the indices (e.g. the coefficient of $\kappa_{21}\kappa_{11}$ is 3×2 since we need to select one male out of 3 and one female out of 2 to form the smaller subgroup; the coefficient of $\kappa_{20}\kappa_{11}^2$ is 6×2 since we need to select 2 males out of 4 for the first subgroup, followed by creating two mixed subgroups from the remaining 2 males and 2 females).

At this point it should be obvious that the complexity of these formulas is forcing us to abandon any further by-hand computation and start using computers (equipped with Mathematica, in our case).

EXAMPLE 10: To find Edgeworth approximation to PDF of

$$g := \frac{\overline{\ln X}}{\pi^2 + X^2}$$

when sampling Exponential distribution with $\mu = 1$, we start by computing the joint MGF of $\ln X$ and X^2 (these are the X and Y of the general theory), by expanding (up to cubic terms in t and u) the integrand of

$$M(t, u) := \int_0^\infty \exp(-x + t \ln x + ux^2) dx$$

before completing the integration. The answer is then converted into a joint MGF of central moments of $\overline{\ln X}$ and X^2 ($\bar{\mu}_{ij}$ of our notation) by expanding (38), this time in powers of $1/n$ (up to cubic) while using $-\gamma$ for the mean of $\ln X$ and 2 for the mean of X^2 (note that the last operation will create some quartic, quintic and hexic terms in t and u).

Next, we need a bivariate Taylor expansion our g function in $\overline{\ln X}$ at $-\gamma$ and in X^2 at 2 (up to cubic terms), resulting in

$$g = \frac{-\gamma}{2 + \pi^2} + (\overline{\ln X} + \gamma) + \frac{\gamma(\overline{X^2} - 2)}{(2 + \pi^2)^2} - \frac{\gamma(\overline{X^2} - 2)^2}{(2 + \pi^2)^3} + \frac{\gamma(\overline{X^2} - 2)^3}{(2 + \pi^2)^4}$$

Using $\bar{\mu}_{ij}$ values found in the previous step, we are then well positioned to compute the K_1 to K_4 cumulants of g , getting

$$K_1 = -0.048630 - \frac{0.028197}{n}$$

$$K_2 = \frac{0.014082}{n} + \frac{0.001486}{n^2}$$

$$K_3 = -\frac{0.003444}{n^2}$$

$$K_4 = -\frac{0.001050}{n^3}$$

The approximate PDF of

$$Z_n := \frac{\overline{\ln X} - K_1}{\sqrt{K_2} + X^2}$$

is given by (25). When $n = 10$, its maximum error is about 0.25%, again deter-

mined from an empirical distribution of 4 million random values of Z_n ; this is again a major improvement over the basic Normal approximation whose maximum error is 4.4%.

9. Final Extension

The Edgeworth approximation becomes increasingly more difficult to construct (due to the number of terms needed) for a function of *several* sample means, yet the layout of the procedure is a simple generalization of what has been done in the previous sections. Let us summarize it:

- Find a joint MGF of all RVs whose sample means are in the studied function, and expand it up to and including *quartic* terms (as shown already: we may need to expand before integrating).
- Based on this expansion, denoted $M_e(t)$, build

$$\exp\left(n \ln M_e\left(\frac{t}{n}\right) - \boldsymbol{\mu} \cdot t\right)$$

where $\boldsymbol{\mu}$ is a vector of the variables' expected values, and expand it in terms of $1/n$ up to and including *cubic* terms. Note that getting a joint central moment of the sample means, say $\bar{\mu}_{ijk\dots}$, is (to a sufficient accuracy) given by the coefficient of $t_1^i t_2^j t_3^k \dots$ in the last expansion, further divided by $i!j!k!\dots$.

- The given function is then expanded in the sample means at their respective expected values, up to and including *cubic* terms (let us denote this expansion g_e). We then proceed to compute K_1 as the expected value of g_e (ignoring the cubic term of this expansion), K_2 as the expected value of $(g_e - K_1)^2$, after expanding it up to quartic terms, K_3 as the expected value of $(g_e - K_1)^3$, after expanding it up to quartic terms, and finally K_4 as the expected value of $(g_e - K_1)^4$, after expanding it up to hexic terms; this time we must also subtract $3K_2^2$ to convert the result from central moment to cumulant. The resulting approximation is then given by (24), (25) and (26), where $g(\bar{X})$ now changes to the given function of several sample means).

EXAMPLE 11: Assume sampling from a general bivariate Normal distribution; the ML estimator of its correlation coefficient ρ is the following sample statistic

$$r := \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sqrt{(\overline{X^2} - \bar{X}^2)(\overline{Y^2} - \bar{Y}^2)}} \tag{40}$$

To find its approximate PDF, we first realize (skipping a routine proof) that the distribution of r remains the same when sampling bivariate Normal distribution with both means equal to 0 and both variances equal to 1, as assumed from now on [6].

Associating X, X^2, Y, Y^2 and XY with t_1, t_2, u_1, u_2 and v respectively, $\ln M(t_1, t_2, u_1, u_2, v)$ of these five RVs is given by

$$\frac{1}{2} \cdot \frac{t_1^2 + u_1^2 - (1 - \rho^2)(t_1^2 u_2 + u_1^2 t_2) + 2\rho t_1 u_1 - 2(1 - \rho^2)t_1 u_1 v}{1 - 2(t_2 + u_2) + 4(1 - \rho^2)t_2 u_2 - 2\rho v - (1 - \rho^2)v^2} - \frac{1}{2} \cdot \ln(1 - 2(t_2 + u_2) + 4(1 - \rho^2)t_2 u_2 - 2\rho v - (1 - \rho^2)v^2) \tag{41}$$

which follows from the corresponding bivariate integration. This then provides a way of computing any joint central moment of the five sample means found in (40) by expanding the 5-variable version of (38).

We also need to expand r in the same sample means, treating them as Mathematical variables, at their expected values, up to *cubic* terms; in principle, this may already result in a total of $1 + 5 + 15 + 35 = 56$ terms but, in this case, some of them turn out to be zero (we get 27 nonzero terms; still too large to quote). As before, we compute the first simple moment of this expansion, followed by computing the next three central moments; the last of these is then easily converted into the corresponding cumulant. Using the previous notation, these turn out to be

$$K_1 = \rho \left(1 - \frac{1 - \rho^2}{2n} \right)$$

$$K_2 = \frac{(1 - \rho^2)^2}{n} \left(1 + \frac{2 + 11\rho^2}{n} \right)$$

$$K_3 = -\frac{6\rho(1 - \rho^2)^3}{n^2}$$

$$K_4 = \frac{(72\rho^2 - 6)(1 - \rho^2)^4}{n^3}$$

The approximate PDF of

$$Z_n := \frac{r - K_1}{\sqrt{K_2}}$$

is then given by (25). Note that to complete the computation, 452 values of $\mu_{ijk\ell m}$ were needed, while the expansion of $(r - K_1)^4$ had 4134 terms!

This time we can explore and quote *exact* errors of this approximation, as an analytic formula for the PDF of r exists (but it is rarely used, due to its complexity). The maximum error, when $n = 20$ and $\rho = \frac{1}{2}$, is 2.5%; too large to make it acceptable.

Nevertheless, when we run the same program using $g(r)$ instead of r , the resulting K_3 is proportional to

$$2\rho g'(\rho) - (1 - \rho^2)g''(\rho)$$

It is easy to verify that $g(\rho) := \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho}$ makes this expression (and therefore K_3 itself) equal to zero [7] [8], while the remaining three cumulants are found to be

$$K_1 = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2n}$$

$$K_2 = \frac{1}{n} + \frac{6-\rho^2}{2n^2}$$

$$K_4 = \frac{2}{n^3}$$

making the approximate PDF of

$$Z_n := \frac{\frac{1}{2} \ln \frac{1+r}{1-r} - K_1}{\sqrt{K_2}}$$

given by (34) [9]; this result is not only much simpler but also substantially more accurate. Its maximum error, when $n = 20$ and $\rho = \frac{1}{2}$, is now a tolerable 0.5%.

10. Conclusions

We have delineated a procedure for extending the CLT beyond its usual $\propto n^{-1/2}$ accuracy by adding cubic and quartic terms to the MGF of \bar{X} and consequently computing two extra moments of the sampled distribution: this then substantially improves the resulting approximation to the corresponding PDF (making its error proportional to $n^{-3/2}$). The improvement is due to matching the sampling distribution's skewness and kurtosis, in addition to the usual mean and variance. The same procedure can be applied to a function of \bar{X} , and eventually extended to a function of several sample means, while sampling any univariate or multivariate distribution. This eventually leads to having to deal with a large number of terms of various intermediate expansions, but this should not be seen as a major obstacle in our age of powerful and fast computers.

When the sampled distribution is either parameter-free or has only a single parameter, the technique enables us to find a transformation of any sample statistic of the previous paragraph, to eliminate its distribution's skewness; this automatically results in an additional and often quite dramatic further improvement in the accuracy of the resulting approximation (making it also simpler, as an extra benefit).

We have demonstrated the new procedure using the classic example of Fisher z transformation, but its main applicability is to sample statistics having no analytic formula for their distribution, especially when the basic Normal approximation of the CLT proves woefully inadequate.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] (2001) Central Limit Theorem. In: *Encyclopedia of Mathematics*, EMS Press. http://encyclopediaofmath.org/index.php?title=Central_limit_theorem&oldid=18508

-
- [2] Edgeworth, F.Y. (1905) The Law of Error I. *Proceedings of the Cambridge Philosophical Society*, **20**, 36-65.
- [3] Kendall, M.G. and Stuart, A. (1969) *The Advanced Theory of Statistics*. Volume 1, 3rd Edition, Griffin.
- [4] Cornish, E.A. and Fisher, R.A. (1938) Moments and Cumulants in the Specification of Distributions. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, **5**, 307-320. <https://doi.org/10.2307/1400905>
- [5] Vrbik, J. (2014) Asymptotic Expansion of Sampling Distributions of an MLE to $1/n$ Accuracy. *Advances and Applications in Statistics*, **38**, 37-50.
- [6] Fisher, R.A. (1915) Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika*, **10**, 507-521. <https://doi.org/10.1093/biomet/10.4.507>
- [7] Hotelling, H. (1953) New Light on the Correlation Coefficient and Its Transforms. *Journal of the Royal Statistical Society Series B. Statistical Methodology*, **15**, 193-225. <https://doi.org/10.1111/j.2517-6161.1953.tb00135.x>
- [8] Winterbottom, A. (1979) A Note on the Derivation of Fisher's Transformation of the Correlation Coefficient. *The American Statistician*, **33**, 142-143. <https://doi.org/10.1080/00031305.1979.10482682>
- [9] Vrbik, J. (2022) Fisher Transformation via Edgeworth Expansion. <https://arxiv.org/abs/2208.05070v1>