Advances in Methodology and Statistics, 2025, *22*(2), 51–61. https://doi.org/10.51936/ummr9138

Severe problems with kurtosis

Gaj Vidmar[a,b,c,*], Bor Vratanar[b]

[a] *University Rehabilitation Institute, Ljubljana, Slovenia*
[b] *University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia*
[c] *University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia*

Abstract

Kurtosis is routinely taught in introductory statistics courses, where it is sometimes still wrongly interpreted as "peakedness" even though the statistical profession has dismissed that notion. Furthermore, the internet abounds with wrong illustrations of kurtosis. Moreover, our literature search found only two actual published practical applications of the kurtosis statistic. To assess the merits of kurtosis, or lack thereof, we review the definition of kurtosis and present simulations that demonstrate extreme instability of kurtosis estimates, especially when viewed jointly with skewness. The main simulations were conducted with 100 000 draws of samples of size 30, 100 and 1000 from standard normal, standard uniform, lognormal, arcsine and standard triangular distributions. Results are presented graphically for the marginal and joint distributions of the skewness and kurtosis estimates. With the lognormal distribution, which is highly skewed, the vast majority of skewness-kurtosis pairs fell very far from the theoretical population values even in samples of size 1000. At the same time, high asymmetry of the sampling distribution of kurtosis when sampling from a symmetric distribution was observed for the arcsine and the standard triangular distribution. In addition, graphical assessment of the coverage of confidence intervals for the studied distributions based on 1000 samples of size 100 revealed several problems. We also made two sets of didactic simulations with data from standard normal distribution to illustrate the huge sampling variability of kurtosis estimates as compared to estimates of lower moments. Hence, we believe that kurtosis should be avoided in non-specialist introductory statistics courses, it should not be routinely calculated as part of numerical data description (whereby its standard error is particularly misleading), and it should not serve routinely as criterion for assessing appropriateness of using the normal distribution as the model for an empirical dataset.

Keywords: probability distributions, central moments, kurtosis, sampling distribution, statistical education

*Corresponding author
Email addresses: gaj.vidmar@mf.uni-lj.si (GV), bor.vratanar@mf.uni-lj.si (BV)
ORCID iDs: 0000-0002-5682-3124 (GV), 0000-0002-6518-4475 (BV)

1. Introduction

Rather than immediately diving into a standard mathematical introduction with formulae and theorems, we begin with some history of the concept in question. A pioneer of both statistical theory and practice, William Henry Gosset, better known under his pseudonym Student, wrote:

> In case any of my readers may be unfamiliar with the term ‘kurtosis’ we may define mesokurtic as ‘having β_2 equal to 3’, while platykurtic curves have $\beta_2 < 3$ and leptokurtic > 3. The important property which follows from this is that platykurtic curves have shorter ‘tails’ than the normal curve of error and leptokurtic longer ‘tails’ (Student [W. S. Gosset], 1927, p. 160).[1]

So far so good, a contemporary statistician would comment; however, effective and ingenuous as it was, Student’s *memoria technica* that followed might have, alas, caused the subsequent spread of misconceptions about kurtosis. The sympathetic animal illustration, accompanied by the witty words “the first figure represents platypus, and the second kangaroos, noted for ‘lepping’, though, perhaps, with equal reason they should be hares” (Student [W. S. Gosset], 1927, p. 160), might have namely induced the wrong notion that leptokurtic distributions have lighter tails as compared to heavier tails of platykurtic distributions, whilst the truth is (broadly speaking) the opposite (as illustrated below in Figure 3).

It is also unfortunate that mainstream statistical teaching has overlooked the insightful words of Walther Andrew Shewhart, the pioneer of statistical quality control, who wrote even earlier that

> We may divide observed distributions into two classes—those that have and those that have not arisen under controlled conditions. For distributions of the first class, the three simple statistics, average, standard deviation, and skewness contain almost all of the information in the original distribution. For those of the second class the most useful statistics are the average and standard deviation. These contain a large part of the total information in the original distribution (Shewhart, 1923, p. 98).

A contemporary leading author in the same field, Donald J. Wheeler, stated even more emphatically that kurtosis is a highly problematic statistic:

> Based on the slow convergence of shape statistics to theoretical shape parameters, clearly thousands of data are needed to do a decent job. Moreover, these thousands of data will need to be collected while the process displays a reasonable degree of statistical control. [...] In short, skewness and kurtosis statistics are practically worthless (Wheeler, 2004, p. 54).

Assuming that the reader is familiar with skewness, i.e., the standardised third central moment, we now present the essential mathematics of kurtosis. It is defined as the standardised fourth central moment:

$$K = \frac{\mu_4}{\sigma^4}.$$

[1]“Normal curve of error” is nowadays usually called normal distribution.

If Z denotes a standardised random variable with a finite fourth moment, then K is the expected value of Z^4. In this form, kurtosis can be interpreted as reflecting the dispersion of Z^2 around its mean, which is 1 (hence it reflects the dispersion of Z around $+1$ and -1).

In practice, this definition is usually replaced by excess kurtosis, where the expectation of K for a normal distribution is subtracted. This is also what the `KURT` function in Microsoft® Excel and kurtosis functions or commands in the majority of statistical packages return:

$$K_E = K - 3.$$

Based on excess kurtosis, theoretical probability distributions can be called (following Student [W. S. Gosset], 1927) mesokurtic (with $K_E = 0$), leptokurtic (with $K_E > 0$) or platykurtic (with $K_E < 0$). For empirical data distributions, more realistic and meaningful expressions are $K_E \approx 0$, $K_E \gg 0$ and $K_E \ll 0$, respectively.

Because our study addresses the issue of joint distribution of skewness and kurtosis, we must emphasise that kurtosis is bounded by squared skewness plus one:

$$\frac{\mu_4}{\sigma^4} \geq \left(\frac{\mu_3}{\sigma^3}\right)^2 + 1.$$

This relationship is the basis for the diagram proposed by Cullen and Frey (1999) for identifying distributions and distribution families. A general example produced in R software (R Core Team, 2024) using the fitdistrplus package (Delignette-Muller & Dutang, 2015) is presented in Figure 1. In this diagram, a distribution whose skewness and kurtosis are constant regardless of its parameters (e.g., the normal and the uniform distribution) is represented by a point. If there is a deterministic relationship between skewness and kurtosis that depends on one of the parameters, the distribution is represented by a (straight or curved) line (e.g., for the gamma distribution with parameter alpha, excess kurtosis equals skewness times 3 divided by square-root of alpha). For the beta distribution, both skewness and kurtosis depend on its two parameters in a non-trivial way, and the shape of the distribution changes greatly with changing parameters, so the skewness^2 – kurtosis combinations can cover any value within a region between two lines.

The first problem with kurtosis is that for nearly one hundred years, it has been wrongly interpreted as a measure of "peakedness" of the distribution. That notion was already rightfully challenged more than fifty years ago (Chissom, 1970; Darlington, 1970; Hildebrand, 1971), it was already reviewed about forty years ago (Balanda & MacGillivray, 1988), and well-known nearly thirty years ago (DeCarlo, 1997). However, it appears to have been unquestionably settled—at least within the statistical profession—only about ten years ago (Westfall, 2014).

Nevertheless, the internet abounds with wrong illustrations of kurtosis, such as two normal distributions with the same mean and different variance presented as one leptokurtic and one platykurtic (Figure 2). At the same time, an image-search on kurtosis retrieves very few correct illustrations (Figure 3).

Furthermore, we found only one actual published practical application of kurtosis (in seismology; Liang et al., 2008), and one submitted for publication (in statistical process control; Zago, 2026). This is practically negligible given the unfathomable amount of scientific publications (even if only those easily identifiable online are considered). Hence, we conducted a series of simulations in order to explore some fundamental properties of the standard kurtosis estimator.

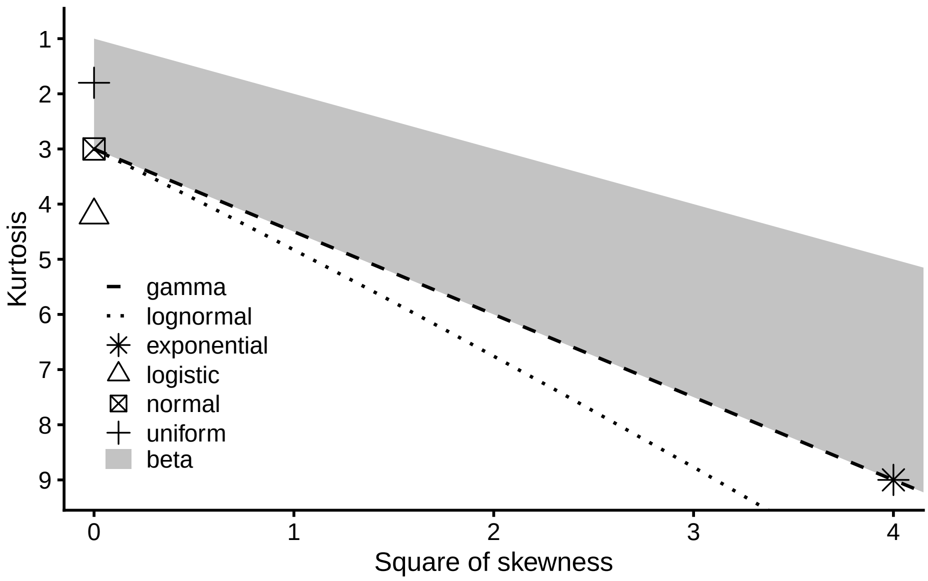

Figure 1. Cullen and Frey (1999) diagram of the relationship between kurtosis and squared skewness for common probability distributions

2. Methods

We conducted two sets of simulations. The first set of simulations was generated by drawing 100 000 samples of size 30, 100 and 1000 from five distributions (Figure 4):

- standard normal ($N_{(0,1)}$),
- standard uniform ($U_{[0,1]}$),
- lognormal (with parameters $\mu = 0$ and $\sigma^2 = 1$),
- arcsine (U-shaped, equivalent to beta distribution with parameters $\alpha = \beta = 0.5$) and
- standard triangular (triangular with parameters $a = -1, c = 1$ and $b = 0$, hence $f(b) = 1$).

Distributions of skewness and kurtosis were depicted using density strip plots, and their joint distributions were depicted using 95 % and 80 % density contours. Strip plot is a good compromise between a detailed presentation of distribution shape as conveyed by histograms, and a stylised representation as applied in boxplots. It is also suitable for conveying uncertainty (Jackson, 2008), including that of a sampling distribution obtained via simulation.

Coverage of confidence intervals (calculated as estimate $\pm 1.96 \times$ asymptotic standard error) for the studied five distributions was graphically assessed by the second set of simulations, based on 1000 samples of size 100. The results were depicted using zip plots (Morris et al., 2019).

Two additional illustrations were made based on samples from the standard normal distribution to compare the first four moments. Line charts were used to show

- how the variability of the estimates depends on sample size (based on 100 repetitions of a series of increasing sample sizes from 10 to 1000) and
- variation between repeated samples (based on 10 repetitions of a series of 10 samples of size 100).

All simulations and graphics were produced using R (R Core Team, 2024). The code is available from the second author upon request.

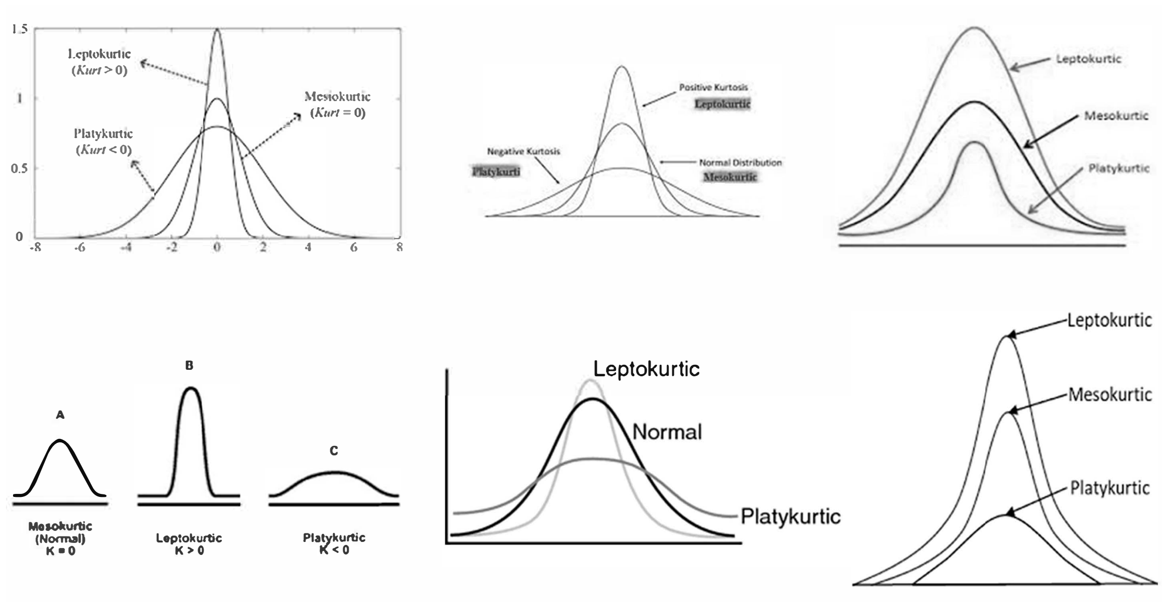

Figure 2. Wrong illustrations retrieved by image-search for kurtosis on the internet (source URLs are not provided because the aim is to show the breadth and variety of misconceptions, not identify their authors)

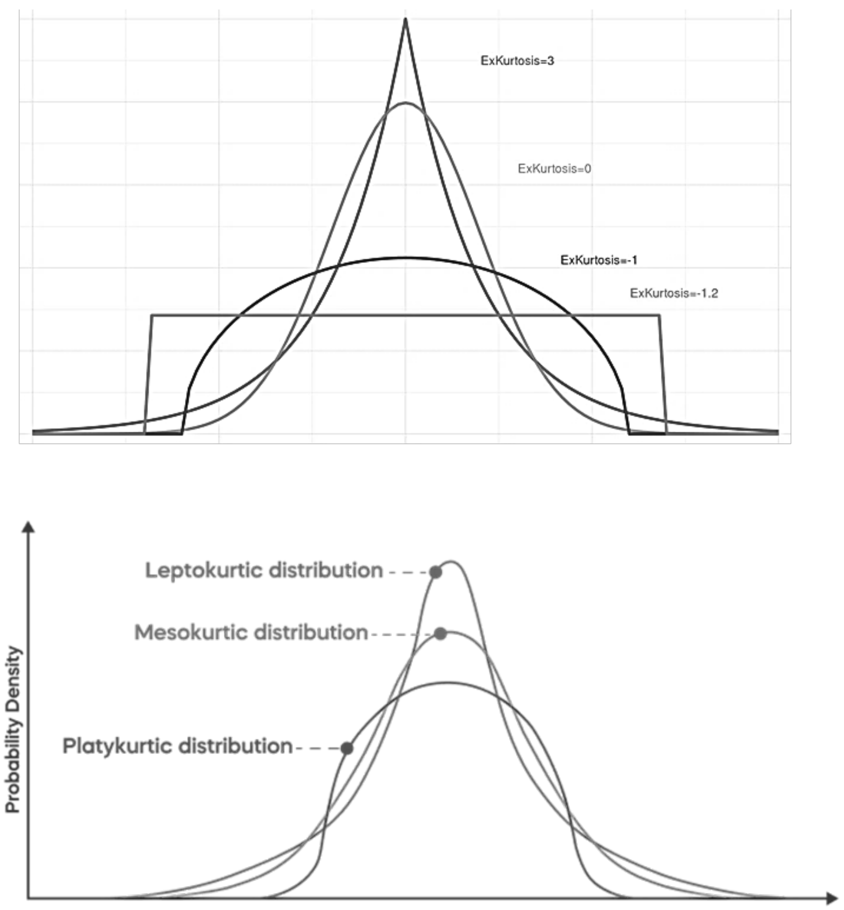

Figure 3. Two correct illustrations of kurtosis retrieved by image-search for kurtosis on the internet (ExKurtosis denotes excess kurtosis)

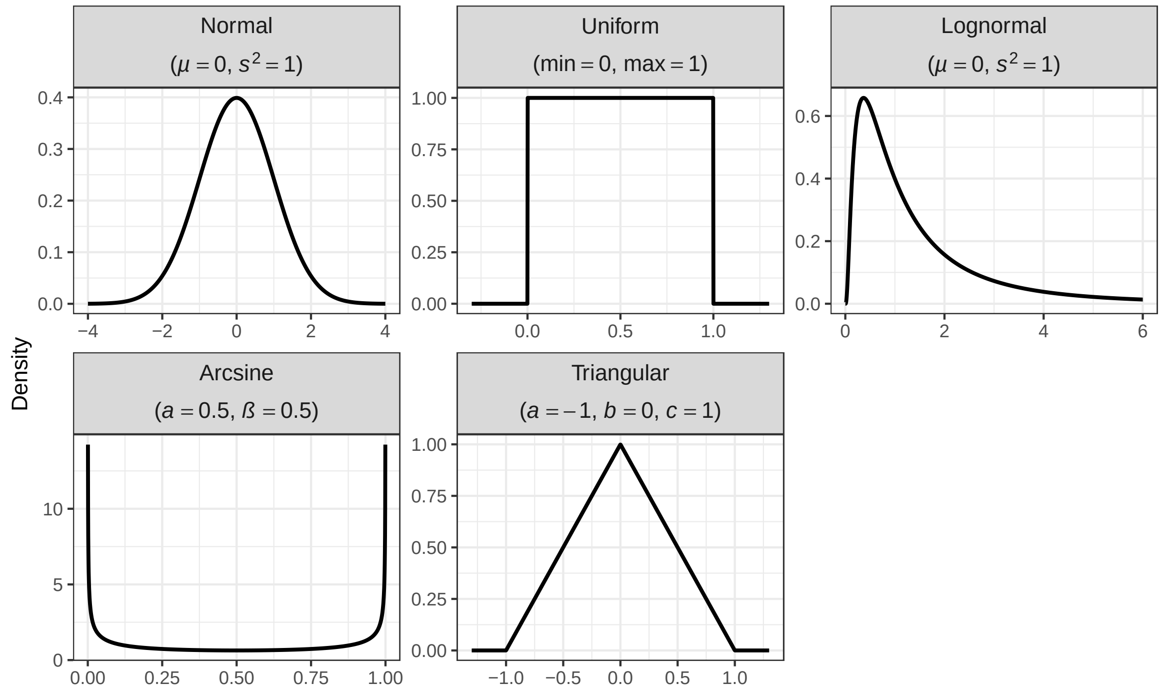

Figure 4. Probability distributions used in the simulations

3. Results

Marginal distributions of skewness and kurtosis are shown in Figure 5.[2] While some degree of asymmetry of the sampling distribution of skewness and kurtosis in the case of lognormal distribution is expected, notable asymmetry of the sampling distribution of kurtosis in the case of arcsine and standard triangular distribution, which are symmetric, is very problematic. What is more, there is some asymmetry of kurtosis estimates (with overrepresentation of too-high estimates) even in the case of the normal and uniform distribution.

Joint distributions of skewness and kurtosis are shown in Figure 6 and Figure 7 in terms of 95 % and 80 % density contours, respectively. For the symmetric distributions (i.e., normal, uniform, arcsine and standard triangular), the contours are characteristically heart-shaped. This was already noticed by Wheeler (2004) for the normal distribution. The key observation is that for the lognormal distribution, even the 80 % contours for samples of size 1000 are nowhere near the population centroid.

The coverage of asymptotic 95 % confidence intervals for excess kurtosis is displayed in Figure 8. Within each panel, the intervals are ordered in the vertical direction from those most congruent with the true value at the bottom to those least congruent at the top. The green dashed vertical line denotes the true value; if a confidence interval includes that value, it is shown as a black line, otherwise it is show by a red line. The horizontal blue line corresponds to the 95-th centile; if the confidence intervals had worked perfectly, all the black lines would have been below that line, and all the red lines above it. Asymmetry is clearly visible for the normal distribution, because the inadequate confidence intervals mainly overestimate the true (i.e., zero) excess kurtosis. For the lognormal distribution, the confidence intervals are completely uninformative because they miss the true value for all

[2]See online article for color versions of the figures.

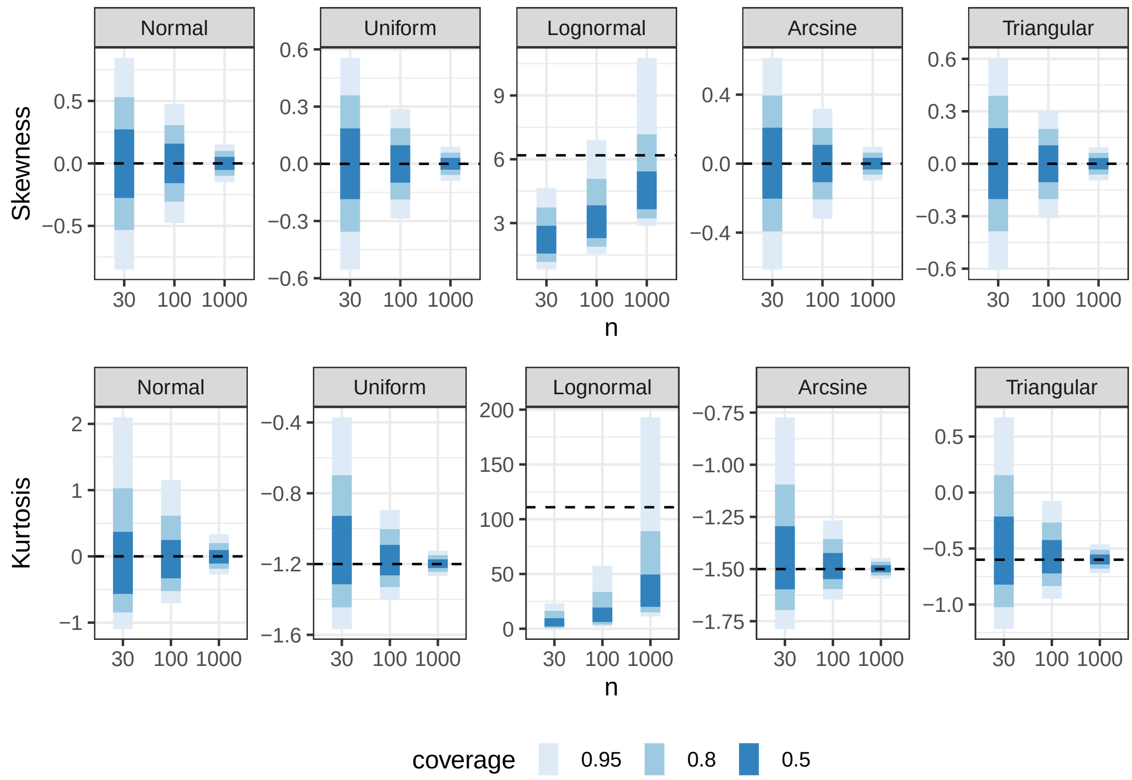

Figure 5. Marginal distributions of skewness and kurtosis estimates depicted using strip-plots (dashed lines indicate population values)

simulated samples. For the other three distributions (i.e., uniform, arcsine and triangular), the confidence intervals are far too wide and therefore extremely conservative.

Both additional illustrations (Figure 9, Figure 10) show that there is a dramatic increase in the variability of the estimates from the third (skewness) to the fourth moment (kurtosis). Figure 9 illustrates much slower convergence of kurtosis to its population value even compared to skewness, let alone mean. Figure 10 mimics hypothetical meta-analyses and illustrates that they could reach very different conclusions despite a relatively high amount of data.

4. Discussion

Our simulations exposed very serious problems with the kurtosis statistic. The first one is bias. With the lognormal distribution, which is highly skewed, the vast majority of skewness–kurtosis pairs were far away from the theoretical population values even in samples of size 1000. The combination of those two statistics, which is routinely examined in statistical applications and presented in countless publications, is therefore of questionable utility because it is supposed to help with judging whether a distribution differs substantially from the normal distribution, yet it can be highly biased precisely if data arise from such a distribution.

A related problem is high asymmetry of the sampling distribution of kurtosis (but not skewness) even though the population distribution is symmetric, which occurs with the arcsine as well as the standard triangular distribution. In our simulations, kurtosis estimates of the arcsine distribution, for which the population value is -1.5, virtually never fell below -2 even in samples of size 30 (where the sampling variability was supposed to be the largest),

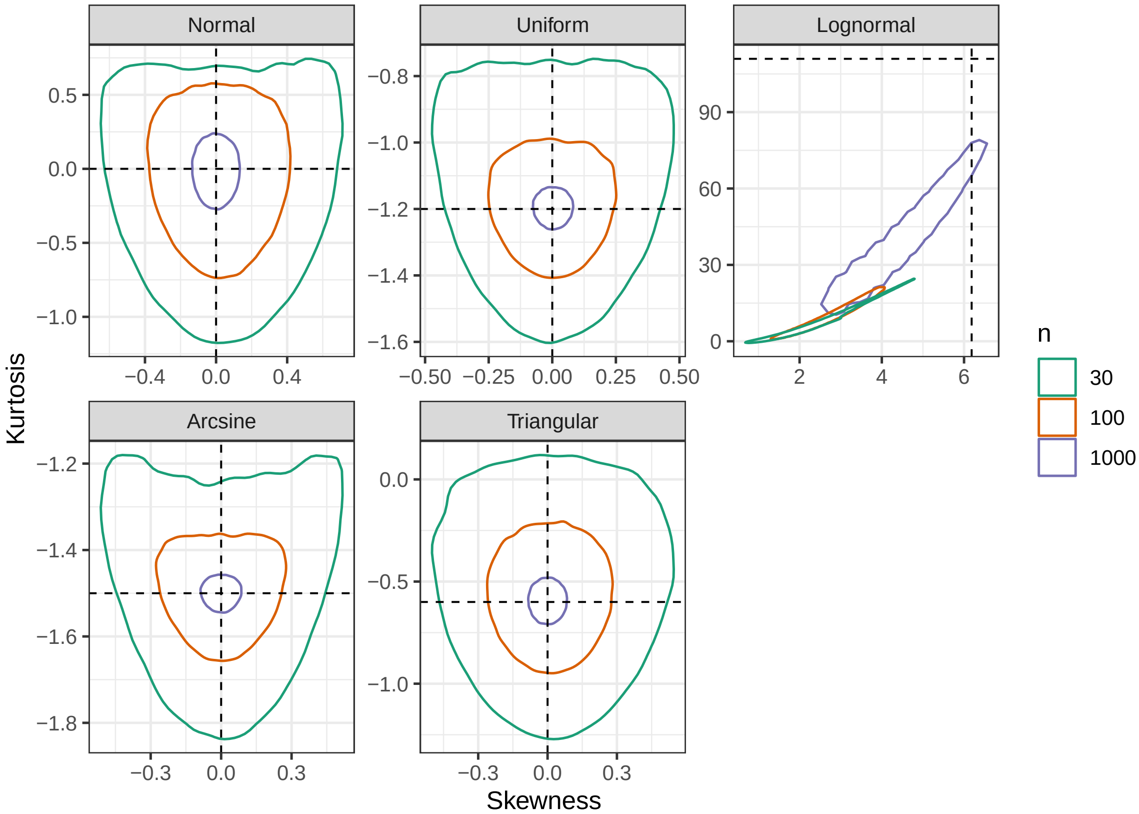

Figure 6. Joint distributions of skewness and kurtosis estimates depicted using 95 % density contours (dashed lines indicate population values)

while they did exceed −1 in about 10 % of such samples. Similarly, kurtosis estimates of the standard triangular distribution, for which the population value is −0.6, practically never fell below −1.2 even in samples of size 30, while they did exceed 0 in about 15 % of such samples.

The asymmetry of the sampling distribution implies that normal-approximation-based confidence intervals for kurtosis are unreliable even when sample sizes are relatively large. This is illustrated by the second set of simulations, where the coverage is asymmetric for the normal distribution, the confidence intervals for the lognormal distribution never include the true parameter value, and the confidence intervals for the uniform, arcsine and triangular distributions are exceedingly wide.

Because of the problems highlighted above, the widespread z-test of the null hypothesis of zero excess kurtosis is also problematic. It is telling and regrettable that a most modern general statistics textbook that corrects the error of interpreting kurtosis as peakedness from its previous editions (Field, 2024) nevertheless presents that test and applies it in several examples.

5. Conclusion

Our simulations demonstrated extreme instability of kurtosis estimates. Wheeler's (2004) judgement is excessively harsh on skewness, but we agree that "thousands of data" are required for reliably estimating kurtosis, especially for heavy-tailed distributions or distributions prone to extreme values. The combination of skewness and kurtosis is particularly unstable and even clearly biased in the case of an asymmetric underlying distribution, such as the lognormal. Together with persisting misconceptions of kurtosis and its extremely

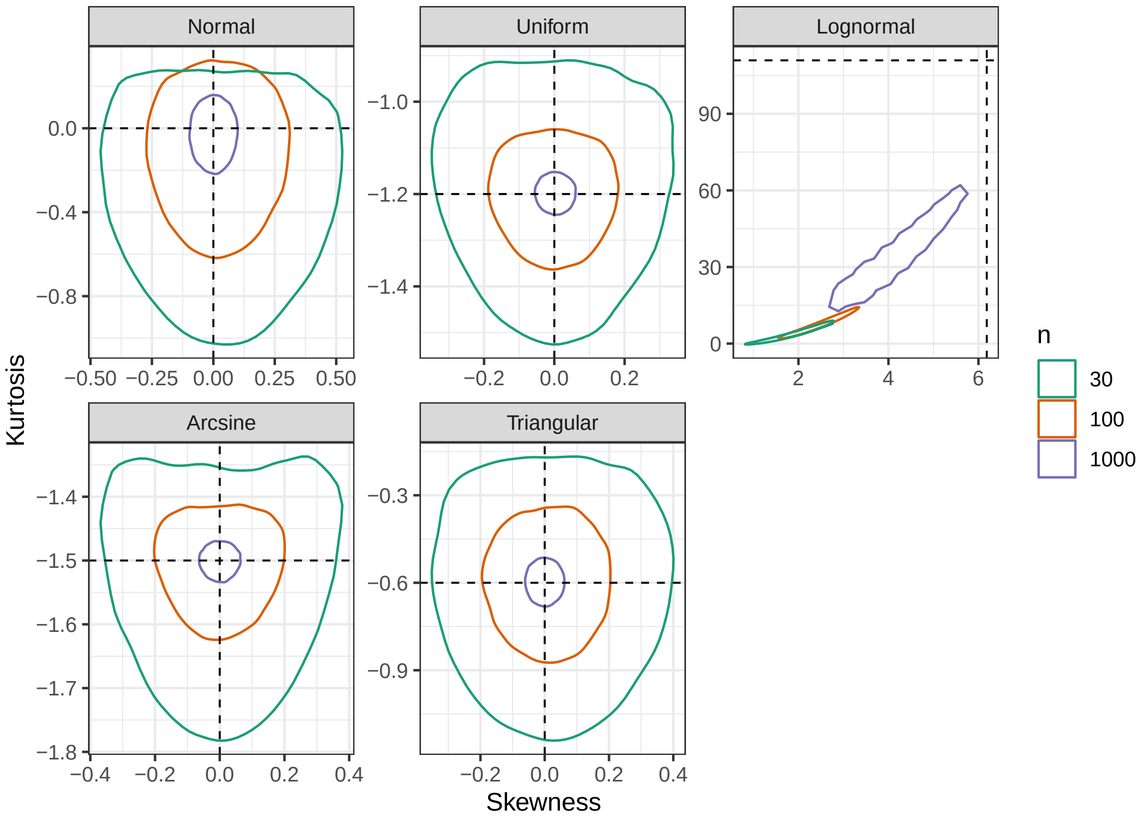

Figure 7. Joint distributions of skewness and kurtosis estimates depicted using 80 % density contours (dashed lines indicate population values)

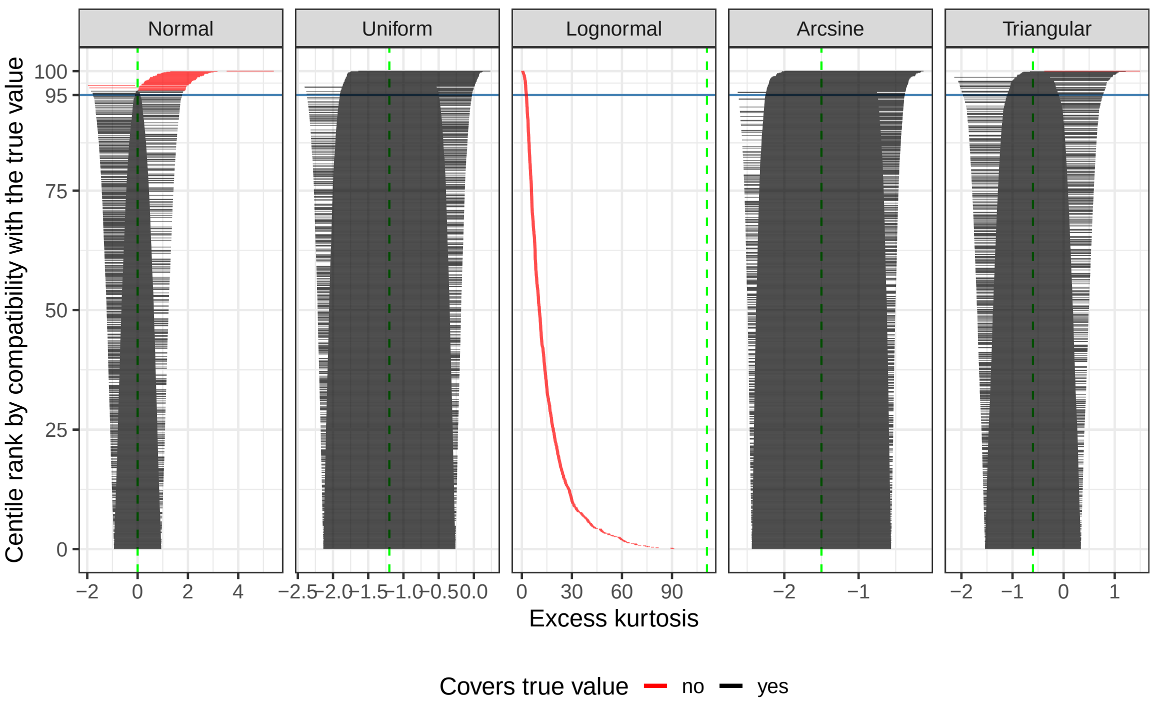

Figure 8. Coverage of asymptotic 95 % confidence intervals for excess kurtosis depicted using zip plots (black lines denote CI's missing the true value, denoted by green vertical dashed line; horizontal blue line denotes the 95-th centile)

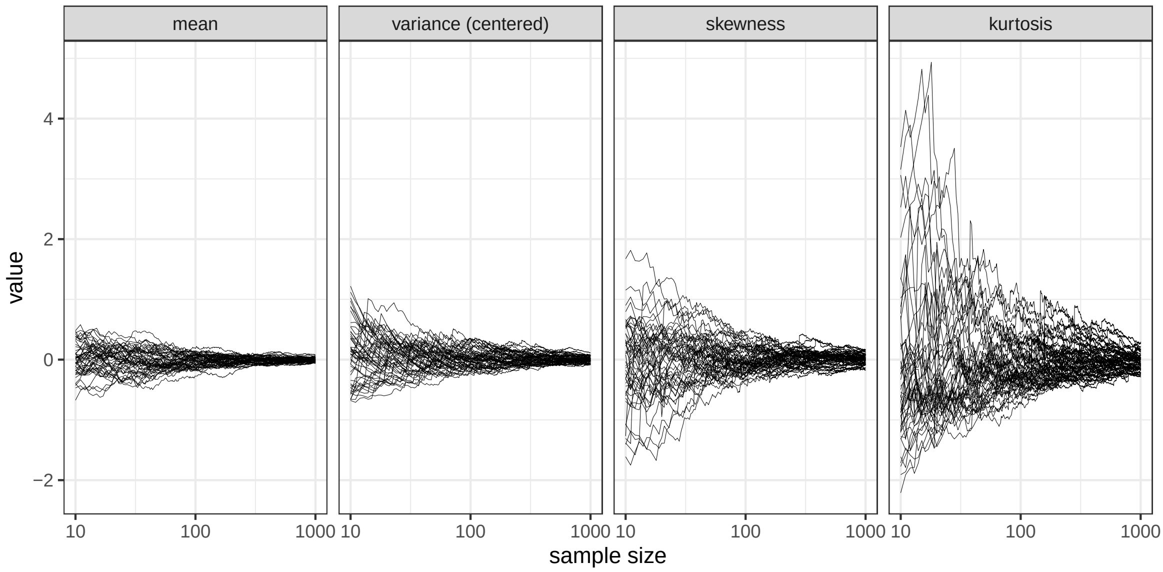

Figure 9. Traces of estimates of the first four moments of the standard normal distribution based on 100 repetitions of a series of increasing sample sizes from 10 to 1000

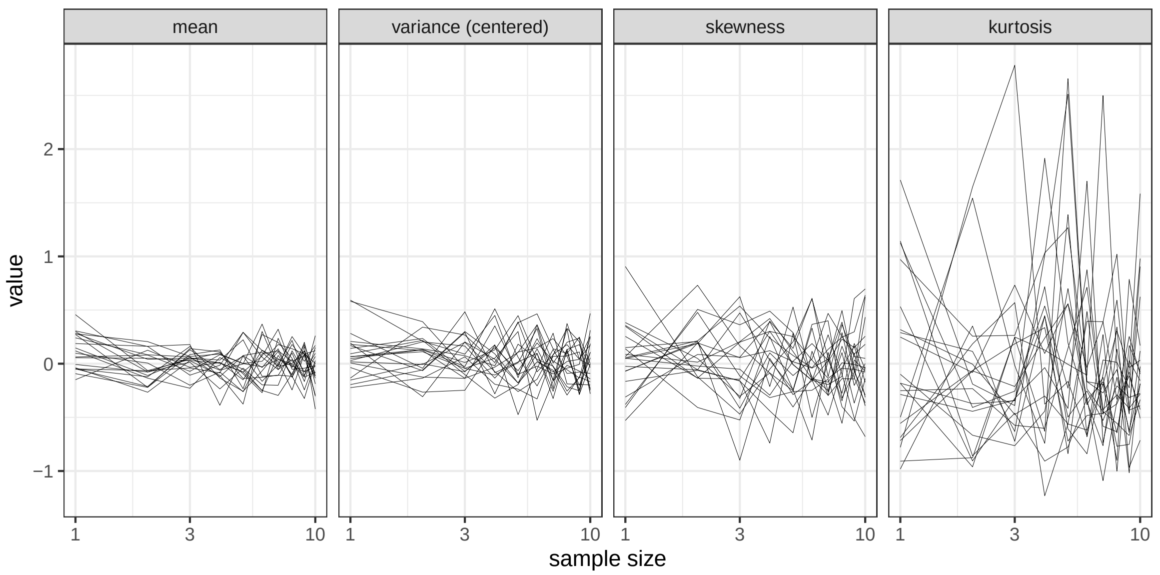

Figure 10. Traces of estimates of the first four moments of the standard normal distribution based on 10 repetitions of a series of 10 samples of size 100

limited practical utility, this leads us to recommend that kurtosis should be avoided in introductory statistics courses, at least for non-specialists. Furthermore, we believe that kurtosis should not be routinely calculated as part of numerical data description. Its standard error is particularly misleading not only in the case of an asymmetric population distribution such as the lognormal, but also because of possibly asymmetric sampling distribution despite a symmetric population distribution (not only bimodal, such as the arcsine, but also unimodal one, such as the standard triangular), which makes the associated confidence intervals very unreliable. Consequently, we believe that kurtosis should not routinely serve as a criterion for assessing the appropriateness of using the normal distribution as the model for an empirical dataset.

References

Balanda, K. P., & MacGillivray, H. L. (1988). Kurtosis: A critical review. *The American Statistician*, *42*(2), 111–119. https://doi.org/10.1080/00031305.1988.10475539

Chissom, B. S. (1970). Interpretation of the kurtosis statistic. *The American Statistician*, *24*(4), 19–22.

Cullen, A. C., & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs*. Springer.

Darlington, R. B. (1970). Is kurtosis really “peakedness”? *The American Statistician*, *24*(2), 19–20. https://doi.org/10.1080/00031305.1970.10478885

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*(3), 292–307. https://doi.org/10.1037/1082-989X.2.3.292

Delignette-Muller, M. L., & Dutang, C. (2015). Fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34. https://doi.org/10.18637/jss.v064.i04

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th ed.). Sage.

Hildebrand, D. K. (1971). Kurtosis measures bimodality? *The American Statistician*, *25*(1), 42–43. https://doi.org/10.1080/00031305.1971.10477241

Jackson, C. H. (2008). Displaying uncertainty with shading. *The American Statistician*, *62*(4), 340–347. https://doi.org/10.1198/000313008X370843

Liang, Z., Wei, J., Zhao, J., Liu, H., Li, B., Shu, J., & Zheng, C. (2008). The statistical meaning of kurtosis and its new application to identification of persons based on seismic signals. *Sensors*, *8*(8), 5106–5119. https://doi.org/10.3390/s8085106

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

R Core Team. (2024). *R: A language and environment for statistical computing* [Computer Program]. R Foundation for Statistical Computing. https://www.R-project.org/

Shewhart, W. A. (1923). *Economic control of quality of manufactured product* (7th ed.). Van Nostrand.

Student [W. S. Gosset]. (1927). Errors of routine analysis. *Biometrika*, *19*(1/2), 151–164. https://doi.org/10.2307/2332181

Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. R.I.P. *The American Statistician*, *68*(3), 191–195. https://doi.org/10.1080/00031305.2014.917055

Wheeler, D. J. (2004). *Advanced topics in statistical process control* (2nd ed.). SPC Press.

Zago, D. (2026, April 14). *Monitoring of complex geometrical shapes* [Manuscript submitted for publication]. https://dedzago.github.io/