SPC

On the asymptotic distribution of an alternative measure of kurtosis

Kagba Suaray

California State University, Long Beach, USA
Email: Kagba.Suaray@csulb.edu

Abstract

Pearson defined the fourth standardized moment of a symmetric distribution as its kurtosis. There has been much discussion in the literature concerning both the meaning of kurtosis, as well as the effectiveness of the classical sample kurtosis as a reliable measure of peakedness and tail weight. In this paper, we consider an alternative measure, developed by Crow and Siddiqui [1], used to describe kurtosis. Its value is calculated for a number of common distributions, and a derivation of its asymptotic distribution is given. Simulations follow, which reveal an interesting connection to the literature on the ratio of normal random variables.

Keywords: *Kurtosis; Quantile Estimator; Ratio of Normal Variables; Spread Function.*

1. Introduction

Consider $X_1, \ldots, X_n \sim F$, where F is an absolutely continuous cdf with density f. In practice, it can be important either to classify data as heavy tailed, or to compare the tail heaviness and/or peakedness of two different samples. The primary descriptor for these purposes is the kurtosis. There has been debate (see[2]) as to whether the population kurtosis (and its variants), essentially defined as the fourth standardized moment

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{E\left[X - E(X)\right]^4}{\left[E\left[X - E(X)\right]^2\right]^2}$$

is a reliable way to describe the movement of probability mass from the "shoulders" to the peak and tails of a density function. If so, how should it be estimated and interpreted?

Much work has been done in the classical theory in the way of discussion on kurtosis. Fisher himself recommended its use as a means to determine whether a sample mean or median is preferable as an estimate for a location parameter see [3]. The desire to use the normal distribution (which has $\beta_2 = 3$) as a baseline recommends the definition of the excess kurtosis, $\beta_2' = \beta_2 - 3$. The law of large numbers motivates the asymptotically unbiased estimator

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^4}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2},$$

with $\hat{\beta}'_2 = \hat{\beta}_2 - 3$ an estimator for the excess kurtosis. However, being essentially a function of the sample mean and sample deviations, $\hat{\beta}_2$, and $\hat{\beta}'_2$ can be sensitive to the outlying observations often characteristic of data from heavy tailed distributions. It has been noted that notions of risk and variation that arise in applications when making data based decisions must take into account not only the variance, but also higher order characteristics of the shape of the underlying distribution (see [4] for one such application in economics). Thus robust measures of tail heaviness and peakedness based on quantiles have been developed. These measures provide a richer understanding of heavy tailed data and distributions in a manner analogous to the way the sample median is preferable to the sample mean for the Cauchy distribution (see [5]). The most popular of these take the form

$$\frac{\sum_{i=1}^{n} \lambda_{1i} F^{-1}(\alpha_i) - \sum_{i=1}^{n} \lambda_{2i} F^{-1}(1 - \alpha_i)}{\sum_{i=1}^{n} \lambda_{3i} F^{-1}(\gamma_i) - \sum_{i=1}^{n} \lambda_{4i} F^{-1}(1 - \gamma_i)}.$$

One such alternative statistic introduced by Hogg [6], is

$$Q = \frac{\bar{U}(.05) - \bar{L}(.05)}{\bar{U}(.5) - \bar{L}(.5)},$$

where $\bar{U}(\alpha)$ and $\bar{L}(\alpha)$ are the average of the upper and lower $n\alpha$ order statistics respectively.
Another statistic in this class was developed by Crow and Siddiqui [1]:

$$\beta_{CS} = \frac{F^{-1}(0.95) - F^{-1}(0.05)}{F^{-1}(0.75) - F^{-1}(0.25)}. \tag{1}$$

Using Ballanda and MacGillivray's [7] notion of a *spread function* of a cdf F,

$$S_F(u) = F^{-1}(0.5 + u) - F^{-1}(0.5 - u) \text{ for } 0 \leq u < \tfrac{1}{2},$$
we can write

$$\beta_{CS} = \frac{S_F(0.45)}{S_F(0.25)}.$$

Ballanda and MacGillivray [2] give examples of more statistics of this type and describe how they are used, including detecting heavy tails, ranking distributions in order of tail thickness, and testing for normality.

A rescaling similar to that imposed on the classical estimator guaranteeing a value of zero for the normal distribution motivates the definition $\beta'_{CS} = \beta_{CS} - 2.44$. It should be noted that both [6] as well as [4] use a slightly modified definition of $\hat{\beta}_{CS}$, with numerator $S_F(0.475)$ replacing $S_F(0.45)$.

Note that the denominator is the interquartile range (IQR). It has been noted in [8] that for distributions with thicker tails, IQR based estimates for the population variance can be more efficient. Related to this is the fact that spread functions are more informative when exploring the relative tail thickness of various distributions, and can be used to relate kurtosis for skewed distributions to that of their symmetrized versions (see [7]). These results recommend β_{CS} as a measure of kurtosis worthy of further study.

2. Calculating and Estimating β_{CS}

2.1. β'_{CS} for Some Common Distributions

We now consider the value of β'_{CS} for some common distributions. Much has been made in the literature concerning the shortcomings of β'_2 as a comparative measure of peakedness and/or tail thickness. Without getting into the details of that discussion, we direct the reader to [7] for a thorough discussion on the topic. Table 1 lists the excess kurtosis values for some distributions using the traditional as well as the Crow and Siddiqui notions of kurtosis. As kurtosis is not impacted by location and scale, the parameters that appear in the lower part of the table are shape parameters. Figures 1 and 2 indicate the relationship between the kurtosis and the shape parameter for both β'_2 and β'_{CS}. It is interesting to note the asymptotic nature of the kurtosis measures for the distributions that tend toward symmetry as the shape parameter increases, as well as the similarity in shape for the two notions of kurtosis for each of the selected distributions.

Table 1: Excess Kurtosis Measures

Distribution	β'_2	β'_{CS}
$Exponential(\sigma)$	6	0.241
$Laplace$	3	0.89
$Uniform$	-1.20	-0.64
$Logistic$	1.20	0.24
$Cauchy$	∞	3.88
$Weibull(\sigma, \xi)$	$\frac{-6\Gamma_1^4 + 12\Gamma_1^2\Gamma_2 - 3\Gamma_2^2 - 4\Gamma_1\Gamma_3 + \Gamma_4}{[\Gamma_2 - \Gamma_1^2]^2}$	$\frac{[-ln(0.05)]^{1/\xi} - [-ln(0.95)]^{1/\xi}}{[-ln(0.25)]^{1/\xi} - [-ln(0.75)]^{1/\xi}} - 2.44$
$GEV(\mu, \sigma, \xi)$	$\frac{g_4 - 4g_1g_3 + 6g_2g_1^2 - 3g_1^4}{(g_2 - g_1^2)^2} - 3$	$\frac{[-ln(0.05)]^{-\xi} - [-ln(0.95)]^{-\xi}}{[-ln(0.25)]^{-\xi} - [-ln(0.75)]^{-\xi}} - 2.44$
$\ln N(\mu, \sigma^2)$	$e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6$	$\frac{\sinh 1.645\sigma}{\sinh 0.674\sigma} - 2.44$
$Pareto(\alpha, \beta)$	$\frac{6(\alpha^3 + \alpha^2 - 6\alpha - 2)}{\alpha(\alpha-3)(\alpha-4)}$	$\frac{(0.05)^{-1/\alpha} - (0.95)^{-1/\alpha}}{(0.25)^{-1/\alpha} - (0.75)^{-1/\alpha}} - 2.44$
$N(\mu, \sigma^2)$	0	0
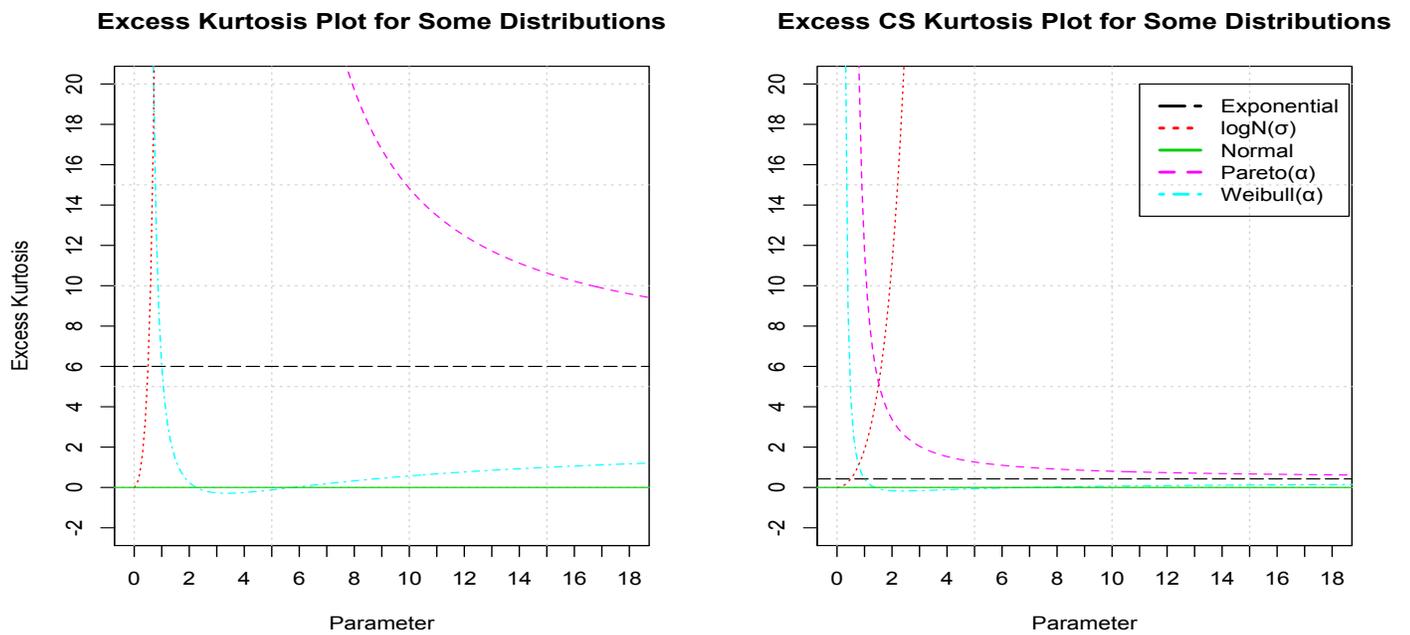
Figure 1: Plots of excess kurtosis vs shape parameter for some distributions

2.2. Estimating β_{CS}

The un-centered parameter β_{CS} given in (1) may be viewed as a function of quantiles, and a reasonable statistic that may be used to estimate it involves replacing the population quantiles with one of several quantile estimators (see [9]). In this paper, we use the classical estimator for $F^{-1}(\alpha) = \xi_\alpha$, i.e.

$$\hat{\xi}_\alpha = \begin{cases} X_{(\lfloor n\alpha \rfloor)}, & \text{if } np \text{ is an inetger} \\ X_{(\lfloor n\alpha \rfloor + 1)}, & \text{otherwise} \end{cases} \tag{2}$$

This yields the statistic

$$\hat{\beta}_{CS} = \frac{\hat{\xi}_{0.95} - \hat{\xi}_{0.05}}{\hat{\xi}_{0.75} - \hat{\xi}_{0.25}}. \tag{3}$$

Yang [10] stated in simulation studies with the Cauchy distribution that some of the more computationally intensive quantile statistics, such as the kernel quantile estimator were outperformed by the sample quantile. In the next section we develop the properties of $\hat{\beta}_{CS}$ that will lead to a discovery of its asymptotic distribution.

3. Asymptotic Distribution of $\hat{\beta}_{CS}$

The following is a well known statement concerning the asymptotic normality of (2). See [8] for a statement and development of the theorem.

Theorem 3.1 Let $X_{k_i:n}$, $1 \leq i \leq r$, be r specified order statistics, and suppose, for some $0 < q_1 < q_2 < \cdots < q_r < 1$, that $\sqrt{n}\left(\frac{k_i}{n} - q_i\right) \to 0$ as $n \to \infty$. Then

$$\sqrt{n}[(X_{k_1:n}, X_{k_2:n}, \ldots, X_{k_r:n}) - (\xi_{q_1}, \xi_{q_2}, \ldots, \xi_{q_r})] \to N_r(\mathbf{0}, \Sigma), \tag{4}$$

where for $i \leq j$, $\sigma_{ij} = \frac{q_i(1-q_j)}{f(\xi_{q_i})f(\xi_{q_j})}$, provided $f(\xi_{q_i})$ exists and is > 0 for each $i = 1, 2, \ldots, r$.

Thus in particular, we can establish the following result concerning the numerator and denominator of $\hat{\beta}_{CS}$, and generally any two values of the spread function, $S_F(\alpha - 0.5)$ and $S_F(\gamma - 0.5)$ obtained from the same random sample.

Corollary 3.2 Let $\mathbf{Y} = \begin{pmatrix} \hat{\xi}_\alpha - \hat{\xi}_{1-\alpha} \\ \hat{\xi}_\gamma - \hat{\xi}_{1-\gamma} \end{pmatrix}$ and $\xi = \begin{pmatrix} \xi_\alpha - \xi_{1-\alpha} \\ \xi_\gamma - \xi_{1-\gamma} \end{pmatrix}$ then

$$\sqrt{n}\left(\mathbf{Y} - \xi\right) \to N\left(\mathbf{0}, \Upsilon\right)$$

where $\Upsilon =$

$$\begin{pmatrix} \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)} - \frac{2(1-\alpha)^2}{f(\xi_\alpha)f(\xi_{1-\alpha})} + \frac{\alpha(1-\alpha)}{f^2(\xi_{1-\alpha})} & \frac{\alpha(1-\gamma)}{f(\xi_\alpha)f(\xi_\gamma)} - \frac{\alpha\gamma}{f(\xi_\alpha)f(\xi_{1-\gamma})} - \frac{(1-\alpha)(1-\gamma)}{f(\xi_{1-\alpha})f(\xi_\gamma)} + \frac{(1-\alpha)(\gamma)}{f(\xi_{1-\alpha})f(\xi_{1-\gamma})} \\ \frac{\alpha(1-\gamma)}{f(\xi_\alpha)f(\xi_\gamma)} - \frac{\alpha\gamma}{f(\xi_\alpha)f(\xi_{1-\gamma})} - \frac{(1-\alpha)(1-\gamma)}{f(\xi_{1-\alpha})f(\xi_\gamma)} + \frac{(1-\alpha)(\gamma)}{f(\xi_{1-\alpha})f(\xi_{1-\gamma})} & \frac{\gamma(1-\gamma)}{f^2(\xi_\gamma)} - \frac{2(1-\gamma)^2}{f(\xi_\gamma)f(\xi_{1-\gamma})} + \frac{\gamma(1-\gamma)}{f^2(\xi_{1-\gamma})} \end{pmatrix}$$

This yields a correlation coefficient between the numerator and denominator of $\hat{\beta}_{CS}$ of

$$\rho_{\alpha\gamma} = \frac{\alpha(1-\gamma)f(\xi_{1-\alpha})f(\xi_{1-\gamma}) - \alpha\gamma f(\xi_{1-\alpha})f(\xi_\gamma) - (1-\alpha)(1-\gamma)f(\xi_\alpha)f(\xi_{1-\gamma}) + (1-\alpha)(\gamma)f(\xi_\alpha)f(\xi_\gamma)}{\left[\alpha\left(f^2(\xi_\alpha) + f^2(\xi_{1-\alpha})\right) - \alpha^2\left(f(\xi_\alpha) + f(\xi_{1-\alpha})\right)^2\right]\left[\gamma\left(f^2(\xi_\gamma) + f^2(\xi_{1-\gamma})\right) - \gamma^2\left(f(\xi_\gamma) + f(\xi_{1-\gamma})\right)^2\right]}.$$

Proof: Define the following

$$\hat{\xi} = \begin{pmatrix} \hat{\xi}_{1-\alpha} \\ \hat{\xi}_{1-\gamma} \\ \hat{\xi}_\gamma \\ \hat{\xi}_\alpha \end{pmatrix} \text{ and } A = \begin{pmatrix} 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

where the elements of $\hat{\xi}$ are chosen so that the probabilities in the subscripts are in increasing order. The corollary follows from elementary methods and the theorem, since $\mathbf{Y} = A\hat{\xi}$ and $\Upsilon = A\Sigma A^T$. □

We will denote the variances of the numerator and denominator of $\hat{\beta}_{CS}$, which are the elements on the main diagonal of Υ, as σ_α^2 and σ_γ^2, respectively.

Thus both the numerator and denominator of $\hat{\beta}_{CS}$ are asymptotically normally distributed with means and variances indicated by the above corollary, taking $\alpha = 0.75$ and $\gamma = 0.95$. It is known that the convergence in law of the quantile function is relatively slow. This being said, the corollary indicates that for a large enough sample size, we expect $\hat{\beta}_{CS}$ to be approximately distributed as the ratio of two correlated normal variates.

A number of authors have explored the nature of the distribution of the ratio of normals, with a variety of motivations. Both Marsaglia ([11],[12]) and Pham-Gia et. al [13], give closed form expressions of the resulting density and distribution functions. The most applicable form results from the fact that the denominator of $\hat{\beta}_{CS}$, i.e. the IQR, is a positive random variable for nontrivial large samples. We apply Marsaglia's notation to yield the following

Proposition 3.3 *Under the assumptions of the theorem, the density function of $\hat{\beta}_{CS}$ is asymptotically equivalent to*

$$g(x) = \phi\left(\frac{\mu_\alpha x - \mu_\gamma}{\sqrt{\sigma_\gamma^2 - 2cx + \sigma_\alpha x^2}}\right)\frac{\mu_\alpha \sigma_\alpha^2 - c\mu_\gamma + (\mu_\gamma \sigma_\alpha^2 - c\mu_\alpha)x}{(\sigma_\gamma^2 - 2cx + \sigma_\alpha^2 x^2)^{3/2}} \tag{5}$$

where ϕ represents the standard normal pdf, and $c = \rho_{\alpha\gamma}\sigma_\alpha\sigma_\gamma$ represents the covariance between the numerator and denominator of $\hat{\beta}_{CS}$.

Depending on the values of the parameters, this distribution can be unimodal or bimodal (although in many of the bimodal cases, the second mode is very far to the left of the primary mode, and often encompasses a relatively trivial amount of area). Many of the unimodal instances actually end up being surprisingly close to familiar distributions. As stated in [12], "for many practical situations, the densities of the ratios of normals... may be considered to have the familiar kinds of shapes often seen in Statistics- normal, t, chi, etc". In the simulations that follow, we explore the distributional properties of $\hat{\beta}_{CS}$, under normal sampling, and will see that its asymptotic distribution is itself approximately normal.

4.   Simulations

In this section, we assume we have a random sample, $X_1 \ldots, X_n \sim N(\mu, \sigma^2)$. The first fact we sought to establish was that the $\hat{\beta}_{CS}$ statistic indeed followed a distribution with density given in (5). The corresponding cdf is

$$G(x) = \Phi\left(\frac{\mu_\alpha x - \mu_\gamma}{\sqrt{\sigma_\gamma^2 - 2cx + \sigma_\alpha x^2}}\right) \tag{6}$$

where Φ represents the standard normal cumulative distribution function. One way to verify this empirically is given in the following procedure:

1. Generate M standard normal simple random samples of size n.
2. For the jth sample, calculate the $\hat{\beta}_{CS}$ value, $j = 1 \ldots M$.
3. Evaluate the cdf $G(x)$ in (6) at the M values obtained in step 2.
4. Employ a goodness-of-fit test to compare the outcome from step 3 to the uniform distribution on $(0, 1)$.

If the goodness-of-fit test finds the distributions sufficiently close, we can conclude that $\hat{\beta}_{CS}$ follows the distribution $G(x)$. We employed a Chi-Square goodness of fit test, as well as the Kolmogorov-Smirnoff test to get an indication as to what number of replications were large enough to yield agreement with the theoretical distribution (as should be expected, M had a much more discernible impact on the results than n, though a larger sample size had a stabilizing effect on the distribution of p-values). The results of Table 2 were obtained for simulations with a sample size of $n = 50,000$.

Table 2: Goodness of Fit p-Values

M	Chi-Sq	K-S
100	0.1320	0.2499
500	0.9999	0.2604
1000	0.9999	0.2422

It is clear that the Monte Carlo sample of $\hat{\beta}_{CS}$ starts behaving close to (6) relatively quickly, with replication sizes of a little less than 200 leading to very high p-values for the Chi-Square goodness of fit test. This is encouraging, given it is known that the sample quantile function converges to normality at a sluggish pace.

The question now remains, does this distribution take an asymptotic form that is approximately equivalent to a familiar density?

We conducted simulations to compare $g(x)$, the theoretical pdf of $\hat{\beta}_{CS}$, i.e. (5), with the normal pdf, $f(x)$ with parameters estimated by their corresponding maximum likelihood estimators. The results for three simulations may be seen in Fig. 2 to 4. There seems to be greater agreement with the normal distribution as the number of trials

increases. The results of the Shapiro-Wilk and Shapiro-Francia tests for normality are given in Table 3. It should be noted that any given realization of the Monte Carlo simulations may lead to an acceptance or rejection of the null hypothesis of normality for the $\hat{\beta}_{CS}$ statistic. The simulations presented were typical for their M values.
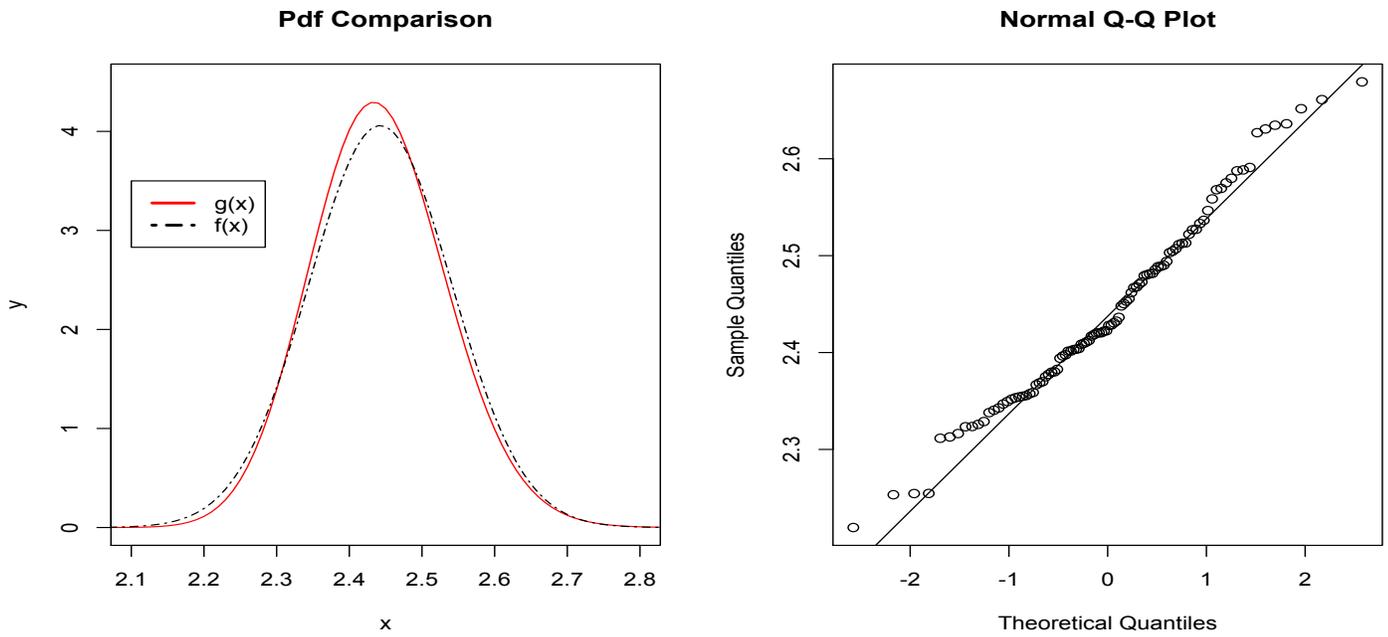
Figure 2: Simulation with $M = 100$ and $n = 1000$
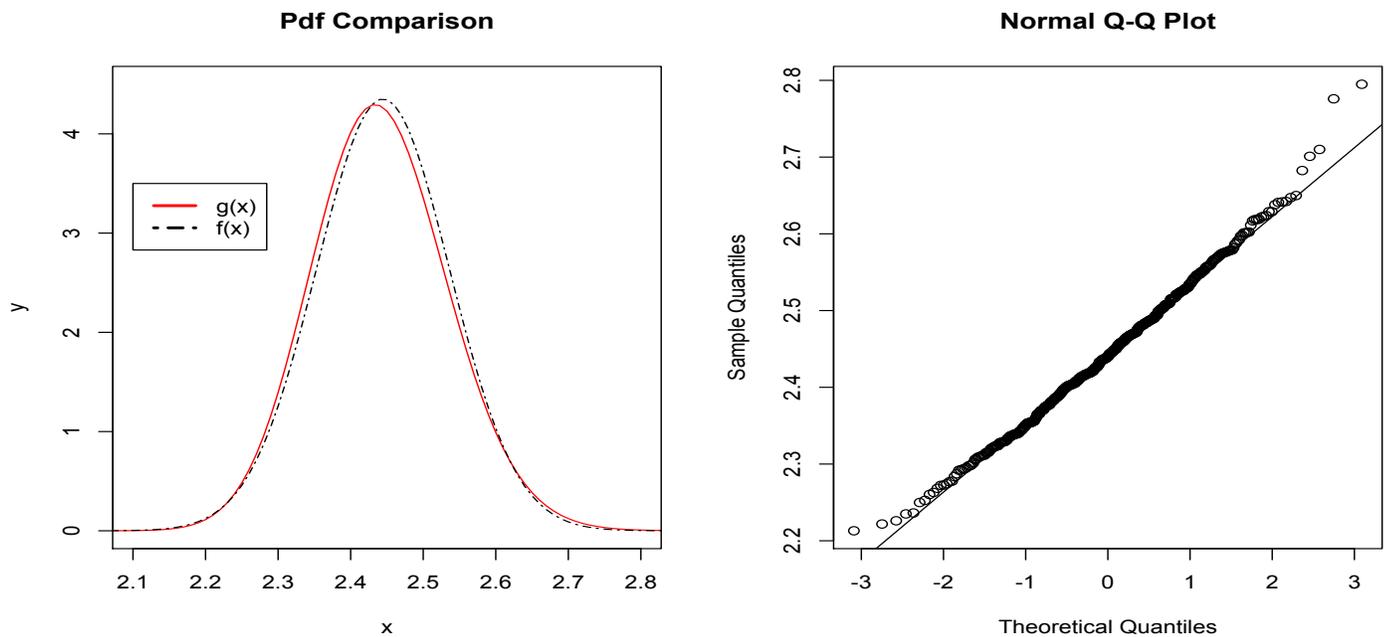
Figure 3: Simulation with $M = 500$ and $n = 1000$

Pdf Comparison   **Normal Q-Q Plot**
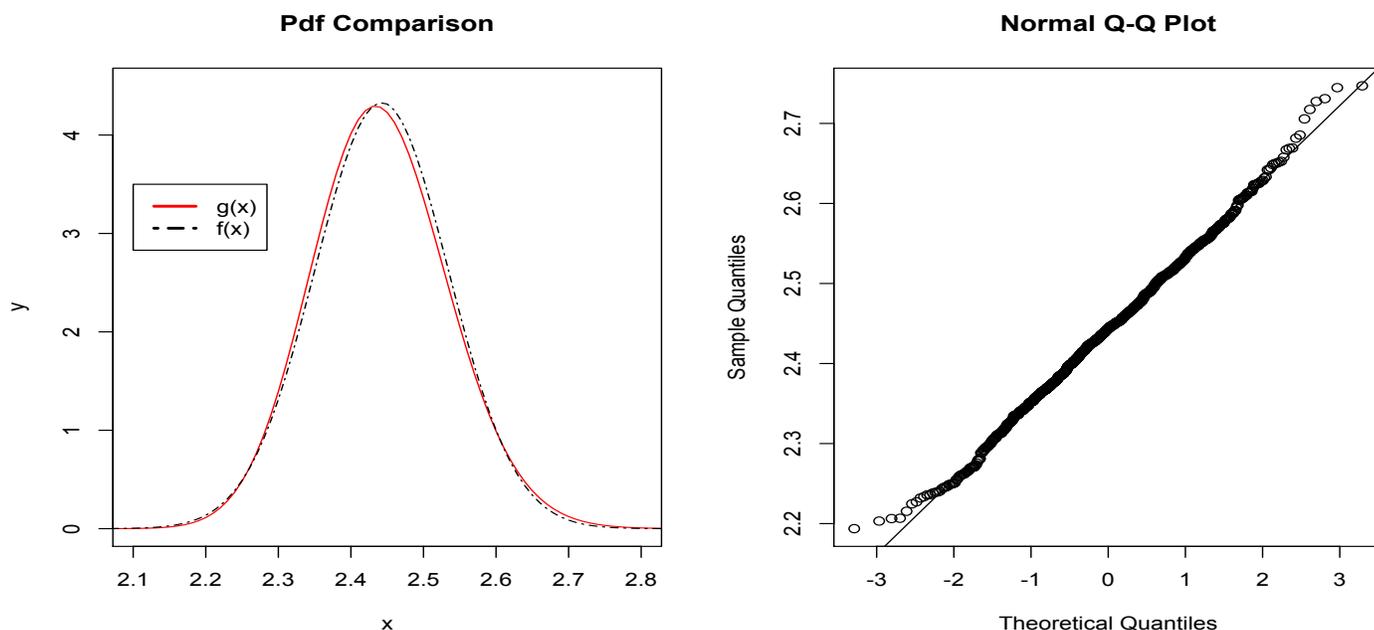
Figure 4: Simulation with $M = 1000$ and $n = 1000$

Table 3: Normality Testing p-Values

M	S-F	S-W
100	0.2353	0.1824
500	0.0459	0.0534
1000	0.1878	0.1310

5.   Conclusion

Kurtosis is a foundational parameter used to describe the shape of statistical distributions from the earliest days of the discipline. In spite of its potential for comparing and categorizing distributions, the classical notion of kurtosis, defined by β_2 has been shown to have some drawbacks as a measure of the location and scale free movement of probability mass from the "shoulders" to the peak and tails of a probability density. Crow and Siddiqui's β_{CS} kurtosis parameter is based on functions of quantities with desirable consistency and efficiency properties for heavy tailed distributions. The fact that for some important cases the corresponding statistic $\hat{\beta}_{CS}$ has an asymptotic distribution that is approximately normal indicates that this parameter can be increasingly used to determine and discover properties of distributions in both a theoretical as well as a data analytic framework.

References

[1] E. Crow and M. Siddiqui, Robust Estimation of Location, Journal of the American Statistical Association, 62, 318 (1967) 353-389.

[2] K. Balanda and H.L. MacGillivray, Kurtosis: A Critical review, The American Statistician, 42, 2, (1988) 111-119.

[3] S. Stigler, Studies in the History of Probability and Statistics. XXXII: Laplace, Fisher and the Discovery of the Concept of Sufficiency, Biometrika, 60, 3 (1973) 439-445.

[4] T. Kim and H. White, On More Robust Estimation of Skewness and Kurtosis, Finance Research Letters, 1, 1 (2004) 56-73.

[5] R. Serfling, Approximation Theorems of Mathematical Statistics, John Wiley, New York, 1980.

[6] R. Hogg, Adaptive Robust Procedures: A partial review and some suggestions for future applications and theory, Journal of the American Statistical Association, 69, 348 (1974) 909-923.

[7] K. Balanda and H.L. MacGillivray, Kurtosis and Spread, The Canadian Journal of Statistics, 18, 1, (1990) 17-30.

[8] A. DasGupta, Asymptotic Theory of Statistics and Probability, Springer, New York, 2008.

[9] R. Hyndman and Y. Fan, Sample Quantiles in Statistical Packages, The American Statistician, 50, 4, (1996) 361-365.

[10] S. Yang, A Smooth Nonparametric Estimator of a Quantile Function, Journal of the American Statistical Association, 80, 392 (1985) 1004-1011.

[11] G. Marsaglia, Ratios of Normal Variables and Ratios of Sums of Uniform Variables, Journal of the American Statistical Association, 60, (1965) 193-204.

[12] G. Marsaglia, Ratios of Normal Variables, Journal of Statistical Software, 16, 4 (2006) 1-10.

[13] T. Pham-Gia, N. Turkkan, E. Marchand, Density of the Ratio of Two Normal Random Variables and Applications, Communications in Statistics, Theory and Methods, 35, (2006) 1569-1591.