



Measuring skewness: We do not assume much

A.A. Khan^{a,*}, S.A. Cheema^b, Z. Hussain^a, and G.A. Abdel-Salam^c

a. *Department of Statistics, Quaid-i-Azam University 45320, Islamabad 44000, Pakistan.*

b. *Department of Mathematical and Physical Sciences, Newcastle University 2308, Australia.*

c. *Department of Mathematics, Statistics and Physics, Qatar University, Doha, Qatar.*

Received 26 November 2018; received in revised form 21 December 2019; accepted 27 January 2020

KEYWORDS

Distribution function;
 Mean;
 Moment;
 Influence function;
 Skewness.

Abstract. Since skewness plays a vital role in different engineering phenomena, its accurate measurement gains significance. Several measures have been taken to quantify the extent of skewness in distributions over the years, but each measure is subject to some serious limitations. In this regard, the present study aims to propose a new skewness measuring functional based on distribution function evaluated at mean with minimal assumptions and limitations. Four well-recognized properties for an appropriate measure of skewness were verified and demonstrated for the new measure. A comparison was made between the new measure and the conventional moment-based measure using both functionals over the range of distributions available in the literature. Furthermore, the robustness of the proposed measure against unusual data points was explored using influence function. The mathematical findings were verified through meticulous simulation studies; further, they were verified by real data sets derived from diverse fields of inquiries. As observed, compared to the classical moment-based measure, the proposed one passed all the checks with distinction. Given the computational simplicity, applicability in a more general environment, and preservation of *c-ordering* of distribution, the proposed measure may be regarded as an attractive addition to the family of skewness measures.

© 2021 Sharif University of Technology. All rights reserved.

1. Introduction

As a pioneering effort, a study was conducted by Pearson [1] that introduced the concept of skewness and proposed a measure for skewness by standardizing the difference between the mean and mode of a distribution. The comprehension of the idea of skewness immediately earned a core position in statistics and allied literature (see [2]), and this concept has been extensively used in different fields (e.g., see [3–12]

and the other cited references). Zwet [13] introduced the concept of ordering two functions in relevance to skewness and brought the idea of *c-preceding* or *c-ordering* into the lime light. A number of researchers have drawn attention to the vibrancy of the concept of skewness and applied it to (i) develop tests of normality, (ii) investigate the robustness of the standard normal theory procedure, and (iii) select a member of family such as from the Karl Pearson family (see [14,15]). A number of recent high-profile academic articles including those authored by Aucremanne et al. [16], Brys et al. [17], Doane and Seward [18], Hosking [19], and Li et al. [20] from various fields of inquiry, witnessed the ongoing glamour of skewness in research community. It has been widely accepted that an appropriate measure of skewness, say $\gamma(X)$, must satisfy the following characteristics (see [21]):

*. *Corresponding author. Tel.: +92-345-9796 413*

E-mail addresses: akhan@stat.qau.edu.pk (A.A. Khan);

saqvn.cheema@gmail.com (S.A. Cheema);

zhlangah@yahoo.com (Z. Hussain); abdo@qu.edu.qa (G.A. Abdel-Salam)

- i. For a symmetric distribution, $\gamma(X) = 0$;
- ii. An appropriate skewness measure is insensitive to linear transformation; in other words, when dealing with transformation of the form $Y = cX + d$, we have $\gamma(cX + d) = \gamma(X)$, where c and d are subject to the conditions that $c > 0$ and $-\infty < d < +\infty$;
- iii. For $Y = -X$, $\gamma(Y) = -\gamma(X)$;
- iv. While comparing two distributions with respect to skewness, if $F_X(\cdot) \prec_c G_Y(\cdot)$, that is, if $F_X(\cdot) <_c G_Y(\cdot)$, then $\gamma(F) \leq \gamma(G)$.

Several skewness functionals and their assessed performances under different conditions are available in multidisciplinary research literature. For instance, Brys et al. [22] remarked the shortcomings of using the moment-based measures of skewness in the presence of outliers. They compared several alternative robust measures of skewness based on quantiles which are less sensitive to outliers. The authors rather suggested using the median, double median, and triple median of actual data points in the range of median and quantiles to measure skewness of the data. A comparison revealed that the skewness measure, “*medcouple*” based on double median, was less sensitive to other moment or quantile-based measures of skewness.

Kim and White [23] compared the single-outlier robust quantile-based and octile-based measures of skewness using the stock market SP & 500 index data. The authors, however, ranked their effort as a starting point for measuring skewness in the financial market data and further modeling asset prices. Holgersson [24] proposed a modified version of the conventional measure of skewness based on the third central moment. The author suggested considering the difference between the mean and median as a base for calculating the third central moment instead of using only the mean (which is extremely sensitive to outliers). Moreover, Tajuddin [25] extended the study carried out by Brys et al. [22] and investigated the *medcouple* as a robust measure of skewness. However, mixed results were found regarding the robustness of the *medcouple* based on the severity of skewness. Yet, another evidence of ongoing efforts based on midrange, mean, median, and mode can be observed in the study of Altinay [26], aiming at comparing a simple class of measures of skewness. Although use of range and midrange in the proposed skewness coefficient guarantees the insensitivity to the changes in location or scale of the data distribution, the final results might be misleading in the presence of outliers.

For discussion sake, some of the well-known measures suggested in the literature are listed in the following:

$$S_k = (\mu - M) / \sigma, \quad \gamma_1 = E(X - \mu)^3 / \sigma^3,$$

$$\gamma'_m = (\mu - m) / \sigma, \quad \gamma_m = (\mu - m) / E|X - m|,$$

$$\gamma_M = 1 - 2F(M).$$

Under the notion of simplicity, the conventional notations of μ , m , and M stand for mean, median, and mode of distribution, respectively, and σ , Q_1 , and Q_3 stand for standard deviation and the 1st and 3rd quartiles, respectively. Despite being premium skewness functionals, their performance was questionable with respect to certain features. For example, S_k and γ'_m fail to maintain the feature of *c-preceding* (see [21]). Li et al. [20] remarked that in some scenarios, γ_1 would result in misleading estimates that would alter the direction of skewness. The mode-based functional, γ_M , assumes uni-modal distribution which limits its utility in real situations; however, γ_m is often not expressible as a simple function of parameters of distribution.

The present study proposed a new skewness functional based on cumulative distribution function (cdf) in relation to the mean of a distribution. The inspiration behind using distribution function as the base of the proposed measure is consistent with statistical intuitions. Ideally, a skewness measure should be a functional involving parameter(s) that affects the shape of the distribution. Of note, the distribution function alone can provide us with different dynamics of distribution under consideration. In addition, the arithmetic mean is a well-celebrated measure of central tendency owing to its capacity of using all the available information and amenability to further mathematical treatments. Furthermore, as a proxy of central tendency of data, it is also related to the skewness of data. Therefore, application of a skewness measure based on the cdf and mean of the distribution is more appealing since both the mean and cdf of a distribution may help elaborate the shape of the distribution. Motivated by the research conducted by the above-mentioned knowledgeable peers, we intended to contribute towards the body of skewness-oriented research in the following sections.

In the following sections, a skewness measure is proposed and its appropriateness as a valid measure of skewness is evaluated by mathematically proving the above-designated characteristics (i)–(iv). Full advantage of available computational facilities should be taken to carry out the intense simulation-based investigation and verify the statistical strength of the new measure. For demonstration purposes, eight most commonly used distributions (Cauchy, F , T , Chi-square, Gamma, Rayleigh, Weibull, and Beta) in statistical research were considered. The notion of choosing these distributions gains significance since they are general enough to offer a wide range of ap-

plications spanning from data modeling to asymptotic theory. They are also essential for hypothesis testing and some of these distributions of the existing measures are either not calculable or limitedly applicable. In addition, performance of the proposed functional was assessed using the real data set on Algeria’s yearly fatality counts from 1997 to 2017, and the survey data collected by Pakistan Bureau of Statistics (PBS) on the respondents’ degree of co-operation (non-co-operation) at district level covering 78,635 households for the year 2014–2015. For comparison purposes, the commonly practiced measure γ_1 was taken into account. The superiority of the measure γ_1 over the others is highlighted in the studies of Arnold and Groeneveld [21] and Zwet [13]. We expect better performance for the proposed measure than that for γ_1 , thus scaling its utility higher in the literature.

2. The measure

Assume a random variable, X , with the cdf of $F(x)$, pdf of $f(x)$, and mean of μ . We define our functional as follows:

$$\gamma_p = 2F(\mu) - 1. \tag{1}$$

One may notice that our proposition of using the value of cdf at mean as a major component of skewness quantifying measure stays consistent with the inherent tendency of mean to slide in the direction of skewness. To further strengthen the candidature of γ_p , it is necessary to remark that unlike other family members of skewness measures, it neither assumes unimodality of the distribution nor calculation of the higher order moments. Only the mean of a distribution is required. Now, assume that γ_p is an appropriate skewness measure by verifying that it holds all four properties (i)–(iv) mentioned in the studies by Arnold and Groeneveld [21] and Oja [27]:

- i. In the case of symmetry, mean and median will be positioned at the same place in population which leads to the fact that $F(\mu) = 0.5$. Based on this argument, it remains trivial to verify that for symmetric distributions, the proposed skewness functional, γ_p , takes the value equal to zero;
- ii. Let $Y = cX + d \Rightarrow \mu_y = c\mu_x + d$. Now, the skewness functional in terms of Y is $\gamma_p(Y) = 2F(\mu_y) - 1 = 2\Pr[Y < \mu_y] - 1 = 2\Pr[cX + d < c\mu_x + d] - 1 = 2\Pr[X < \mu_x] - 1 = \gamma_p(X)$;
- iii. Let $Y = -X \Rightarrow \mu_y = -\mu_x$, the skewness functional for variable Y is then written as $\gamma_p(Y) = 2F(\mu_y) - 1 \Rightarrow \gamma_p(Y) = 2F(-\mu_x) - 1$. It is trivially verifiable that $F(-\mu_x) = 1 - F(\mu_x)$ and by using this argument, we get $\gamma_p(Y) = 2[1 - F(\mu_x)] - 1 = -[2F(\mu_x) - 1] = -\gamma(X)$;
- iv. To prove this property, we need to show that

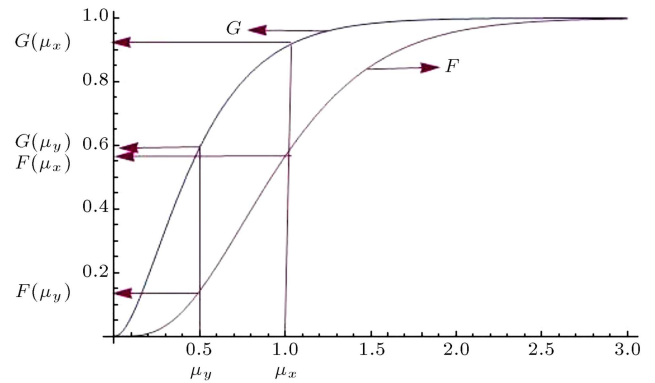


Figure 1. Display of distribution functions with regard to the concept of *c-ordering* of distributions.

$G(\mu_y) > F(\mu_x)$. It is easy to perceive that if $G_Y(\cdot)$ is more skewed to the right than $F_X(\cdot)$, $\mu_y < \mu_x$. Based on what was mentioned, Figure 1 reveals the following inequalities:

$$G_Y(\mu_x) - G_Y(\mu_y) \leq G_Y(\mu_x) - F_X(\mu_x),$$

$$\frac{d}{d\mu_y}(G_Y(\mu_x) - G_Y(\mu_y)) \leq \frac{d}{d\mu_y}(G_Y(\mu_x) - F_X(\mu_x)),$$

$$\frac{d}{d\mu_y}(G_Y(\mu_x)) - \frac{d}{d\mu_y}(G_Y(\mu_y)) \leq \frac{d}{d\mu_y}(G_Y(\mu_x)) - \frac{d}{d\mu_y}(F_X(\mu_x)),$$

$$-\frac{d}{d\mu_y}(G_Y(\mu_y)) \leq -\frac{d}{d\mu_y}(F_X(\mu_x)),$$

$$\frac{d}{d\mu_y}(G_Y(\mu_y)) \geq \frac{d}{d\mu_y}(F_X(\mu_x)),$$

$$\frac{d}{d\mu_y}(G_Y(\mu_y)) \geq 0,$$

which is obvious. Hence, our supposition is true. In other words, we have $G_Y(\mu_x) - G_Y(\mu_y) \leq G_Y(\mu_x) - F_X(\mu_x)$, hence $G_Y(\mu_y) \geq F_X(\mu_x)$. This implies that $\gamma_p(F) = 2F_X(\mu_x) - 1 \leq 2G_Y(\mu_y) - 1 = \gamma_p(G)$ which verifies the property (iv). Figure 1 depicts the notion of the mathematical proof.

In addition to the above-mentioned attributes, for positive skewness, $F(\mu) > 0.5 \Rightarrow 2F(\mu) - 1 > 0$, thus ensuring a positive value of the skewness measure, which can take a value of +1 at its extreme. Likewise, dealing with negatively skewed distribution, $F(\mu) < 0.5 \Rightarrow 2F(\mu) - 1 < 0$, thus resulting a negative value, with -1 indicating the extreme negatively skewed scenario. This discussion establishes a well-defined, interpretable, and comparable skewness functional as we always have $-1 < \gamma_p < +1$ and $\gamma_p = 0$ when symmetry is witnessed while taking *c-ordering* of distributions into account.

3. Finite sample behavior

This section is dedicated to the assessment of the behavior of γ_p in the finite sample and verification of the mathematical facts derived from the aforementioned section. Here, the sustainability, interpretability, and comprehension of the proposed skewness functional were demonstrated through the extensive simulations study. The notations used in this section are listed below:

$s(\bar{x})$	Empirical cdf at \bar{x}
$\overline{s(\bar{x})}$	Mean of the 10,000 empirical cdfs at \bar{x}
$\hat{\gamma}_p = 2s(\bar{x}) - 1$	Estimate of γ_p
$\bar{\hat{\gamma}}_p$	Mean of the 10,000 $\hat{\gamma}_p$
$\hat{\gamma}_1$	Estimate of γ_1
$\bar{\hat{\gamma}}$	Mean of the 10,000 $\hat{\gamma}_1$

We generate 10,000 samples of different sizes ($n = 30, 50, 100,$ and 500) for all the distributions mentioned in the introduction section. Under the notion

of generality, different combinations of parameter(s) driving the extent of skewness are taken into account to evaluate the performance of the proposed measure in the situations moderate to extreme skewness. To document the comparative spirit of γ_p , the finite sample behavior of the existing measure of skewness, γ_1 , is also presented under the same above-mentioned environment. Tables 1–8 present the values with the average of over 10,000 repetitions, highlighting the features mathematically established in Section 2. In addition, the interesting behavior of the estimate variance of the proposed functional in comparison to that of γ_1 alongside the well-celebrated *consistency* remains overwhelming throughout the study.

The results are compiled and interpreted in four sub-sections assembling the selected distributions with respect to their inherent attributes.

3.1. The Chi-square and gamma distributions

First, the computational findings for Chi-square and Gamma (both positively skewed) distributions were documented. To this end, Tables 1 and 2 were used. Obviously, with an increase in the value of shape

Table 1. The simulated results for Chi-square distribution considering different sample sizes and various combinations of parameter.

		$v = 2$				$v = 3$				
n		30	50	100	500	n	30	50	100	500
$F(\mu)$		0.63212	0.63212	0.63212	0.63212	$F(\mu)$	0.60838	0.60838	0.60838	0.60838
γ_p		0.26426	0.26426	0.26426	0.26426	γ_p	0.21675	0.21675	0.21675	0.21675
γ_1		2.00000	2.00000	2.00000	2.00000	γ_1	2.00000	2.00000	2.00000	2.00000
$\overline{s(\bar{x})}$		0.62739	0.62859	0.62999	0.63196	$\overline{s(\bar{x})}$	0.60340	0.60552	0.60715	0.60834
$\bar{\hat{\gamma}}_p$		0.25477	0.25718	0.25999	0.26392	$\bar{\hat{\gamma}}_p$	0.20680	0.21104	0.21430	0.21668
$V(\hat{\gamma}_p)$		(0.01310)	(0.00775)	(0.00385)	(0.00078)	$V(\hat{\gamma}_p)$	(0.01254)	(0.00758)	(0.00390)	(0.00077)
$\bar{\hat{\gamma}}_1$		1.47836	1.64425	1.77574	1.95113	$\bar{\hat{\gamma}}_1$	1.23580	1.35429	1.47863	1.04440
$V(\hat{\gamma}_1)$		(0.40075)	(0.37106)	(0.29164)	(0.110120)	$V(\hat{\gamma}_1)$	(0.33692)	(0.28965)	(0.21244)	(0.07322)
		$v = 5$				$v = 10$				
n		30	50	100	500	n	30	50	100	500
$F(\mu)$		0.58412	0.58412	0.58412	0.58412	$F(\mu)$	0.55951	0.55951	0.55951	0.55951
γ_p		0.16824	0.16824	0.16824	0.16824	γ_p	0.11901	0.11901	0.11901	0.11901
γ_1		1.26491	1.26491	1.26491	1.26491	γ_1	0.89443	0.89443	0.89443	0.89443
$\overline{s(\bar{x})}$		0.58050	0.58187	0.58284	0.58358	$\overline{s(\bar{x})}$	0.55728	0.55739	0.55864	0.55913
$\bar{\hat{\gamma}}_p$		0.16100	0.16375	0.16567	0.16715	$\bar{\hat{\gamma}}_p$	0.11456	0.11478	0.11728	0.11827
$V(\hat{\gamma}_p)$		(0.01228)	(0.00773)	(0.00378)	(0.00075)	$V(\hat{\gamma}_p)$	(0.01224)	(0.00738)	(0.00360)	(0.00075)
$\bar{\hat{\gamma}}_1$		0.97527	1.08051	1.15893	1.23948	$\bar{\hat{\gamma}}_1$	0.70700	0.76368	0.82892	0.87929
$V(\hat{\gamma}_1)$		(0.27722)	(0.23096)	(0.15366)	(0.04363)	$V(\hat{\gamma}_1)$	(0.22945)	(0.17113)	(0.10472)	(0.00270)

Table 2. The simulated results for Gamma distribution considering different sample sizes and various combinations of parameters.

$\alpha = 1, \beta = 1$					$\alpha = 3, \beta = 1$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.63210	0.63210	0.63210	0.63210	$F(\mu)$	0.57681	0.57681	0.57681	0.57681
γ_p	0.26424	0.26424	0.26424	0.26424	γ_p	0.15362	0.15362	0.15362	0.15362
γ_1	2.00000	2.00000	2.00000	2.00000	γ_1	1.15470	1.15470	1.15470	1.15470
$s(\bar{x})$	0.62660	0.62811	0.63011	0.63156	$s(\bar{x})$	0.57261	0.57565	0.57619	0.57665
$\tilde{\gamma}_p$	0.25327	0.25623	0.26023	0.26313	$\tilde{\gamma}_p$	0.14521	0.15130	0.15239	0.15329
$V(\hat{\gamma}_p)$	(0.44856)	(0.34128)	(0.23889)	(0.10672)	$V(\hat{\gamma}_p)$	(0.77165)	(0.56716)	(0.40098)	(0.17809)
$\tilde{\gamma}_1$	1.48760	1.62557	1.79768	1.95260	$\tilde{\gamma}_1$	0.89422	0.98710	1.06955	1.13657
$V(\hat{\gamma}_1)$	(0.42324)	(0.37028)	(0.31140)	(0.17114)	$V(\hat{\gamma}_1)$	(0.57222)	(0.45717)	(0.35081)	(0.17398)
$\alpha = 5, \beta = 1$					$\alpha = 10, \beta = 1$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.55951	0.55951	0.55951	0.55951	$F(\mu)$	0.54207	0.54207	0.54207	0.54207
γ_p	0.11901	0.11901	0.11901	0.11901	γ_p	0.08414	0.08414	0.08414	0.08414
γ_1	0.89443	0.89443	0.89443	0.89443	γ_1	0.63246	0.63246	0.63246	0.63246
$s(\bar{x})$	0.55679	0.55789	0.55856	0.55907	$s(\bar{x})$	0.53915	0.54692	0.54086	0.54181
$\tilde{\gamma}_p$	0.11357	0.11578	0.11712	0.11813	$\tilde{\gamma}_p$	0.07830	0.08184	0.08172	0.08362
$V(\hat{\gamma}_p)$	(0.01234)	(0.00743)	(0.00369)	(0.00074)	$V(\hat{\gamma}_p)$	(0.01193)	(0.00756)	(0.00373)	(0.00075)
$\tilde{\gamma}_1$	0.70666	0.77100	0.82664	0.87852	$\tilde{\gamma}_1$	0.49542	0.55061	0.58381	0.61952
$V(\hat{\gamma}_1)$	(0.22842)	(0.16788)	(0.10347)	(0.02648)	$V(\hat{\gamma}_1)$	(0.18850)	(0.14313)	(0.07860)	(0.01819)

Table 3. The simulated results for Weibull distribution considering different sample sizes and various combinations of parameters.

$\alpha = 1, \beta = 1$					$\alpha = 3, \beta = 1$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.63212	0.63212	0.63212	0.63212	$F(\mu)$	0.50937	0.50937	0.50937	0.50937
γ_p	0.26424	0.26424	0.26424	0.26424	γ_p	0.01875	0.01875	0.01875	0.01875
γ_1	2.00000	2.00000	2.00000	2.00000	γ_1	0.16810	0.16810	0.16810	0.16810
$s(\bar{x})$	0.62509	0.62843	0.62976	0.63181	$s(\bar{x})$	0.50834	0.50904	0.50938	0.50934
$\tilde{\gamma}_p$	0.25019	0.25686	0.25953	0.26361	$\tilde{\gamma}_p$	0.01668	0.01808	0.01876	0.01868
$V(\hat{\gamma}_p)$	(0.01276)	(0.00791)	(0.00384)	(0.00077)	$V(\hat{\gamma}_p)$	(0.01164)	(0.00709)	(0.00345)	(0.00070)
$\tilde{\gamma}_1$	1.47068	1.63975	1.78756	1.95114	$\tilde{\gamma}_1$	0.13299	0.15147	0.16017	0.16562
$V(\hat{\gamma}_1)$	(0.38703)	(0.37944)	(0.29825)	(0.11215)	$V(\hat{\gamma}_1)$	(0.12112)	(0.07768)	(0.03903)	(0.00822)
$\alpha = 5, \beta = 1$					$\alpha = 7, \beta = 1$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.47928	0.47928	0.47928	0.47928	$F(\mu)$	0.46568	0.46568	0.46568	0.46568
γ_p	-0.04143	-0.04143	-0.04143	-0.04143	γ_p	-0.06864	-0.06864	-0.06864	-0.06864
γ_1	-0.25411	-0.25411	-0.25411	-0.25411	γ_1	-0.46319	-0.46319	-0.46319	-0.46319
$s(\bar{x})$	0.48058	0.48000	0.47958	0.47919	$s(\bar{x})$	0.46814	0.46583	0.46609	0.46574
$\tilde{\gamma}_p$	-0.03885	-0.03999	-0.04084	-0.04160	$\tilde{\gamma}_p$	-0.06373	-0.06835	-0.06782	-0.06852
$V(\hat{\gamma}_p)$	(0.01185)	(0.00718)	(0.00354)	(0.00070)	$V(\hat{\gamma}_p)$	(0.01235)	(0.00705)	(0.00359)	(0.00071)
$\tilde{\gamma}_1$	-0.22019	-0.23218	-0.26071	-0.25205	$\tilde{\gamma}_1$	-0.38519	-0.41632	-0.43781	-0.45739
$V(\hat{\gamma}_1)$	(0.13947)	(0.08485)	(0.04296)	(0.0091)	$V(\hat{\gamma}_1)$	(0.16125)	(0.10343)	(0.05672)	(0.01161)

Table 4. The simulated results for beta distribution considering different sample sizes and various combinations of parameters.

$\alpha = 1, \beta = 6$					$\alpha = 4, \beta = 8$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.60343	0.60343	0.60343	0.60343	$F(\mu)$	0.52744	0.52744	0.52744	0.52744
γ_p	0.20686	0.20686	0.20686	0.20686	γ_p	0.05488	0.05488	0.05488	0.05488
γ_1	1.28300	1.28300	1.28300	1.28300	γ_1	0.36422	0.36422	0.36422	0.36422
$s(\bar{x})$	0.59994	0.60135	0.60226	0.60304	$s(\bar{x})$	0.52522	0.52609	0.52739	0.52743
$\tilde{\gamma}_p$	0.19988	0.20269	0.20452	0.20607	$\tilde{\gamma}_p$	0.05043	0.05218	0.05479	0.05485
$V(\hat{\gamma}_p)$	(0.01173)	(0.00711)	(0.00353)	(0.00069)	$V(\hat{\gamma}_p)$	(0.01185)	(0.00677)	(0.00342)	(0.00068)
$\tilde{\gamma}_1$	1.09389	1.17331	1.22600	1.27160	$\tilde{\gamma}_1$	0.30519	0.33276	0.35054	0.36074
$V(\hat{\gamma}_1)$	(0.19773)	(0.14388)	(0.08005)	(0.01750)	$V(\hat{\gamma}_1)$	(0.12045)	(0.00757)	(0.03783)	(0.00769)
$\alpha = 2, \beta = 2$					$\alpha = 5, \beta = 5$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.50000	0.50000	0.50000	0.50000	$F(\mu)$	0.50000	0.50000	0.50000	0.50000
γ_p	0.00000	0.00000	0.00000	0.00000	γ_p	0.00000	0.00000	0.00000	0.00000
γ_1	0.00000	0.00000	0.00000	0.00000	γ_1	0.00000	0.00000	0.00000	0.00000
$s(\bar{x})$	0.49995	0.49982	0.50018	0.50022	$s(\bar{x})$	0.49941	0.50020	0.49986	0.50029
$\tilde{\gamma}_p$	-0.00010	-0.00037	0.00037	0.000448	$\tilde{\gamma}_p$	-0.00118	0.00040	-0.00027	0.00006
$V(\hat{\gamma}_p)$	(0.01067)	(0.00657)	(0.00322)	(0.00064)	$V(\hat{\gamma}_p)$	(0.01143)	(0.00678)	(0.00339)	(0.00068)
$\tilde{\gamma}_1$	0.00048	-0.00044	0.00059	0.00033	$\tilde{\gamma}_1$	-0.00388	0.00044	-0.00084	-0.00032
$V(\hat{\gamma}_1)$	(0.07264)	(0.04365)	(0.02121)	(0.00419)	$V(\hat{\gamma}_1)$	(0.10711)	(0.06201)	(0.03098)	(0.00603)
$\alpha = 7, \beta = 3$					$\alpha = 9, \beta = 5$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.46283	0.46283	0.46283	0.46283	$F(\mu)$	0.47860	0.47860	0.47860	0.47860
γ_p	-0.07434	-0.07434	-0.07434	-0.07434	γ_p	-0.04279	-0.04279	-0.04279	-0.04279
γ_1	-0.48249	-0.48249	-0.48249	-0.48249	γ_1	-0.28868	-0.28868	-0.28868	-0.28868
$s(\bar{x})$	0.46467	0.46322	0.46329	0.46301	$s(\bar{x})$	0.47881	0.47970	0.47875	0.47863
$\tilde{\gamma}_p$	-0.07066	-0.07356	-0.07341	-0.07397	$\tilde{\gamma}_p$	-0.04236	-0.0406	-0.04060	-0.4273
$V(\hat{\gamma}_p)$	(0.01157)	(0.00701)	(0.00339)	(0.00069)	$V(\hat{\gamma}_p)$	(0.01146)	(0.00677)	(0.00677)	(0.00069)
$\tilde{\gamma}_1$	-0.41517	-0.446521	0.46500	-0.47825	$\tilde{\gamma}_1$	-0.25137	-0.26340	-0.26340	-0.28528
$V(\hat{\gamma}_1)$	(0.12078)	(0.07754)	(0.03792)	(0.00808)	$V(\hat{\gamma}_1)$	(0.12244)	(0.07568)	(0.07568)	(0.00797)

parameter, the extent of skewness decreases for both distributions. Regardless of the extent of skewness, both functionals (proposed and existing) resulted in their permissible ranges and remained interpretable. Further, the signs of estimates were consistent with the direction of skewness shown in the population. Aligned with mathematical findings of Section 2, the *c-preceding* characteristic was maintained by both measures in every case without exception. An interesting and distinctive feature of the proposed statistic is

the variability control which becomes more prominent while increasing the sample size. For more elaboration, we highlighted the variance of both measures for $n = 500$. One exception is observed in the case of Gamma distribution for $\alpha = 3$ and $\beta = 1$.

3.2. The Weibull and beta distributions

The altering effect of the parameters on skewness while working with these two distributions makes them very informative candidates for assessing the performance

Table 5. The simulated results for student’s t -distribution considering different sample sizes and various combinations of parameters.

$v = 2$					$v = 3$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.50000	0.50000	0.50000	0.50000	$F(\mu)$	0.50000	0.50000	0.50000	0.50000
γ_p	0.00000	0.00000	0.00000	0.00000	γ_p	0.00000	0.00000	0.00000	0.00000
γ_1	–	–	–	–	γ_1	–	–	–	–
$s(\bar{x})$	0.50127	0.50093	0.50123	0.49994	$s(\bar{x})$	0.49959	0.50017	0.50057	0.49992
$\tilde{\gamma}_p$	0.00253	0.00187	0.00246	–0.00013	$\tilde{\gamma}_p$	–0.00081	0.00034	0.00113	–0.00015
$V(\hat{\gamma}_p)$	(0.06263)	(0.04547)	(0.03114)	(0.00968)	$V(\hat{\gamma}_p)$	(0.02645)	(0.01765)	(0.00934)	(0.00200)
$\tilde{\gamma}_1$	0.01069	0.04112	0.06651	–0.02304	$\tilde{\gamma}_1$	–0.00351	–0.00785	0.00568	–0.04103
$V(\hat{\gamma}_1)$	(3.67913)	(5.3898)	(9.99695)	(36.5364)	$V(\hat{\gamma}_1)$	(1.62034)	(2.14600)	(2.88789)	(6.22120)
$v = 4$					$v = 5$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.50000	0.50000	0.50000	0.50000	$F(\mu)$	0.50000	0.50000	0.50000	0.50000
γ_p	0.00000	0.00000	0.00000	0.00000	γ_p	0.00000	0.00000	0.00000	0.00000
γ_1	0.00000	0.00000	0.00000	0.00000	γ_1	0.00000	0.00000	0.00000	0.00000
$s(\bar{x})$	0.49966	0.49958	0.50021	0.50024	$s(\bar{x})$	0.49888	0.50073	0.49966	0.49977
$\tilde{\gamma}_p$	–0.00067	–0.00084	0.00042	0.00048	$\tilde{\gamma}_p$	–0.00224	0.00145	–0.00069	–0.00046
$V(\hat{\gamma}_p)$	(0.01990)	(0.01223)	(0.00625)	(0.00122)	$V(\hat{\gamma}_p)$	(0.01689)	(0.01029)	(0.00503)	(0.00104)
$\tilde{\gamma}_1$	–0.00129	0.00212	–0.00404	0.00158	$\tilde{\gamma}_1$	–0.00892	0.00685	–0.00975	–0.01682
$V(\hat{\gamma}_1)$	(1.00591)	(1.10064)	(1.15580)	(1.39235)	$V(\hat{\gamma}_1)$	(0.67060)	(0.68144)	(0.62416)	(0.44776)

of skewness functionals. The obtained results are presented in Tables 3 and 4 for Weibull and Beta distributions, respectively. Interpretability and c -preceding are maintained by both the functionals regardless of the change direction and extent of skewness. The variance behavior of the proposed estimator is again appreciable in comparison with the existing measure.

3.3. The student’s t and F distributions

These two distinguished distributions are major players in hypotheses testing domain of statistics, and their dynamic nature limits the applicability of γ_1 in certain conditions; for instance, it is only estimable if t -distribution’s $df. > 3$, F -distribution requires denominator $df. > 6$. These inherent functional complexities provide ideal grounds to test the utility of γ_p , thus confirming the superiority of this new functional over the others.

Table 5 presents the simulated results in the case of t -distribution for different degrees of freedom, dictating the calculability of γ_1 . For comparison purposes, we report the $\tilde{\gamma}_1$ based on empirical evaluation even in the situations, where γ_1 does not exist for population. The deteriorated performance of $\hat{\gamma}_1$ is evident through the estimated average value for $df. \leq 3$ (highlighted

variances foretell this fact). For $df. > 3$, our proposed functional still outperforms the existing one. The rapid convergence of $\tilde{\gamma}_p$ towards its true value is worth cheering, especially in feasibly large samples.

The F -distribution demands a more elaborative account. The given discussion of the obtained results can be elaborated in the light of four motivational factors: (i) The γ_1 is not estimable; (ii) The numerator and denominator degrees of freedom are equal; (iii) The numerator d.f. is greater than denominator’s d.f. ($v_1 > v_2$); and (iv) vice versa ($v_2 > v_1$). In general, the recommended c -preceding attribute was carried by both measures in all cases.

In the first case, the estimability and closeness of $\hat{\gamma}_p$ to its population counterpart γ_p are noticeable that become more obvious as the sample size increases. The statistical strength can be assessed by comparing the variances of both functionals at every point.

In the second case, other than variance comparison (which is consistent with the usual findings), the convergence of both functionals towards their respective population parameters is thought provoking. For explanatory purposes, consider the case when $v_1 = v_2 = 7$ for $n = 500$. The true value of $\gamma_1 = 10.1559$ and its estimated simulated value is $\tilde{\gamma}_1 = 4.2547$, yet

Table 6. The simulated results for F -distribution considering different sample sizes and various combinations of parameters.

$v_1 = 1, v_2 = 3$					$v_1 = 3, v_2 = 5$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.81831	0.81831	0.81831	0.81831	$F(\mu)$	0.71221	0.71221	0.71221	0.71221
γ_p	0.63662	0.63662	0.63662	0.63662	γ_p	0.42441	0.42441	0.42441	0.42441
γ_1	–	–	–	–	γ_1	–	–	–	–
$\frac{s(\bar{x})}{s(\bar{x})}$	0.67264	0.77583	0.78896	0.80606	$\frac{s(\bar{x})}{s(\bar{x})}$	0.69118	0.69854	0.70512	0.71062
$\bar{\gamma}_p$	0.52527	0.55166	0.57793	0.61213	$\bar{\gamma}_p$	0.38237	0.39708	0.41024	0.42122
$V(\hat{\gamma}_p)$	(0.02458)	(0.01887)	(0.01252)	(0.00535)	$V(\hat{\gamma}_p)$	(0.02064)	(0.01464)	(0.00852)	(0.00215)
$\bar{\gamma}_1$	3.11141	3.94584	5.40987	10.97980	$\bar{\gamma}_1$	2.38910	2.93949	3.80013	6.42382
$V(\hat{\gamma}_1)$	(1.22462)	(2.07043)	(4.18508)	(22.22940)	$V(\hat{\gamma}_1)$	(1.06552)	(1.63729)	(2.94790)	(11.93600)
$v_1 = 7, v_2 = 7$					$v_1 = 9, v_2 = 9$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.66588	0.66588	0.66588	0.66588	$F(\mu)$	0.64289	0.64289	0.64289	0.64289
γ_p	0.33177	0.33177	0.33177	0.33177	γ_p	0.28578	0.28578	0.28578	0.28578
γ_1	10.15590	10.15590	10.15590	10.15590	γ_1	4.39205	4.39205	4.39205	4.39205
$\frac{s(\bar{x})}{s(\bar{x})}$	0.65367	0.65825	0.66175	0.66494	$\frac{s(\bar{x})}{s(\bar{x})}$	0.63340	0.63738	0.63941	0.64244
$\bar{\gamma}_p$	0.30733	0.31650	0.32351	0.32989	$\bar{\gamma}_p$	0.26681	0.27476	0.27882	0.28488
$V(\hat{\gamma}_p)$	(0.01899)	(0.01164)	(0.00634)	(0.00134)	$V(\hat{\gamma}_p)$	(0.01701)	(0.01034)	(0.00547)	(0.00112)
$\bar{\gamma}_1$	1.97263	2.36775	2.90375	4.25472	$\bar{\gamma}_1$	1.72144	2.02557	2.44363	3.25094
$V(\hat{\gamma}_1)$	(0.92258)	(1.23576)	(1.90147)	(5.08084)	$V(\hat{\gamma}_1)$	(0.79281)	(0.98828)	(1.34279)	(12.60029)
$v_1 = 3, v_2 = 7$					$v_1 = 7, v_2 = 9$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.67962	0.67962	0.67962	0.67962	$F(\mu)$	0.64551	0.64551	0.64551	0.64551
γ_p	0.35923	0.35923	0.35923	0.35923	γ_p	0.29103	0.29103	0.29103	0.29103
γ_1	11.00000	11.00000	11.00000	11.00000	γ_1	4.47214	4.47214	4.47214	4.47214
$\frac{s(\bar{x})}{s(\bar{x})}$	0.66570	0.67163	0.67554	0.67876	$\frac{s(\bar{x})}{s(\bar{x})}$	0.63427	0.63970	0.64233	0.64473
$\bar{\gamma}_p$	0.33140	0.34327	0.35107	0.35752	$\bar{\gamma}_p$	0.26853	0.27941	0.28468	0.28946
$V(\hat{\gamma}_p)$	(0.01825)	(0.01188)	(0.00618)	(0.00132)	$V(\hat{\gamma}_p)$	(0.01718)	(0.01024)	(0.00537)	(0.00114)
$\bar{\gamma}_1$	2.07855	2.50569	3.10938	4.53298	$\bar{\gamma}_1$	1.72786	2.06022	2.47838	3.31528
$V(\hat{\gamma}_1)$	(0.89883)	(1.26351)	(1.98945)	(5.31268)	$V(\hat{\gamma}_1)$	(0.75854)	(0.98613)	(1.39227)	(2.50372)
$v_1 = 9, v_2 = 7$					$v_1 = 11, v_2 = 9$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.66394	0.66394	0.66394	0.66394	$F(\mu)$	0.64134	0.64134	0.64134	0.64134
γ_p	0.32790	0.32790	0.32790	0.32790	γ_p	0.28269	0.28269	0.28269	0.28269
γ_1	10.03800	10.03800	10.03800	10.03800	γ_1	4.34484	4.34484	4.34484	4.34484
$\frac{s(\bar{x})}{s(\bar{x})}$	0.65055	0.65631	0.66045	0.66290	$\frac{s(\bar{x})}{s(\bar{x})}$	0.63189	0.63525	0.63859	0.64076
$\bar{\gamma}_p$	0.30110	0.31262	0.32089	0.32581	$\bar{\gamma}_p$	0.26378	0.27049	0.277184	0.28152
$V(\hat{\gamma}_p)$	(0.01871)	(0.01179)	(0.00641)	(0.00138)	$V(\hat{\gamma}_p)$	(0.01677)	(0.01051)	(0.00543)	(0.00112)
$\bar{\gamma}_1$	1.92299	2.34468	2.92540	4.20557	$\bar{\gamma}_1$	1.6967	2.00512	2.41802	3.2203
$V(\hat{\gamma}_1)$	(0.89460)	(1.23039)	(1.99575)	(4.77499)	$V(\hat{\gamma}_1)$	(0.76439)	(0.96459)	(1.32659)	(2.4167)

Table 7. The simulated results for Rayleigh distribution considering different sample sizes and various combinations of parameter.

$\alpha = 5$					$\alpha = 4$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.54406	0.54406	0.54406	0.54406	$F(\mu)$	0.54406	0.54406	0.54406	0.54406
γ_p	0.08812	0.08812	0.08812	0.08812	γ_p	0.08812	0.08812	0.08812	0.08812
γ_1	0.63111	0.63111	0.63111	0.63111	γ_1	0.63111	0.63111	0.63111	0.63111
$s(\bar{x})$	0.54221	0.54298	0.54341	0.54391	$s(\bar{x})$	0.54235	0.54292	0.54331	0.54394
$\hat{\gamma}_p$	0.084417	0.08596	0.08683	0.08782	$\hat{\gamma}_p$	0.08471	0.08584	0.08663	0.08787
$V(\hat{\gamma}_p)$	(0.01169)	(0.00693)	(0.00353)	(0.00069)	$V(\hat{\gamma}_p)$	(0.01169)	(0.00696)	(0.00349)	(0.00069)
$\hat{\gamma}_1$	0.52539	0.56564	0.59762	0.62406	$\hat{\gamma}_1$	0.52309	0.56224	0.59639	0.62592
$V(\hat{\gamma}_1)$	(0.15383)	(0.10240)	(0.05665)	(0.05665)	$V(\hat{\gamma}_1)$	(0.15317)	(0.10187)	(0.05649)	(0.01251)

Table 8. The simulated results for Cauchy distribution considering different sample sizes and various combinations of parameters.

$\alpha = -1, \beta = 1$					$\alpha = -1, \beta = 3$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.5	0.5	0.5	0.5	$F(\mu)$	0.5	0.5	0.5	0.5
γ_p	0	0	0	0	γ_p	0	0	0	0
γ_1	–	–	–	–	γ_1	–	–	–	–
$s(\bar{x})$	0.50711	0.51216	0.49956	0.50136	$s(\bar{x})$	0.50199	0.50086	0.49752	0.50279
$\hat{\gamma}_p$	0.00227	0.00243	–0.00088	0.00272	$\hat{\gamma}_p$	0.00387	0.00172	–0.00496	0.00558
$V(\hat{\gamma}_p)$	(0.27918)	(0.30180)	(0.31092)	(0.32522)	$V(\hat{\gamma}_p)$	(0.27905)	(0.30271)	(0.31073)	(0.32908)
$\hat{\gamma}_1$	0.01604	0.00821	0.01821	0.09924	$\hat{\gamma}_1$	0.02335	0.02081	–0.03202	0.08241
$V(\hat{\gamma}_1)$	(10.082)	(17.6950)	(36.1892)	(184.6340)	$V(\hat{\gamma}_1)$	(10.06320)	(17.7665)	(36.1368)	(186.71300)
$\alpha = 0, \beta = 1$					$\alpha = 0, \beta = 3$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.5	0.5	0.5	0.5	$F(\mu)$	0.5	0.5	0.5	0.5
γ_p	0	0	0	0	γ_p	0	0	0	0
γ_1	–	–	–	–	γ_1	–	–	–	–
$s(\bar{x})$	0.49684	0.50005	0.49864	0.50400	$s(\bar{x})$	0.49855	0.50261	0.49649	0.50612
$\hat{\gamma}_p$	–0.00632	0.00010	–0.002728	0.00800	$\hat{\gamma}_p$	–0.00289	0.00522	–0.00701	0.012239
$V(\hat{\gamma}_p)$	(0.27746)	(0.29562)	(0.31109)	(0.32602)	$V(\hat{\gamma}_p)$	(0.27769)	(0.29590)	(0.31254)	(0.328688)
$\hat{\gamma}_1$	–0.041012	0.02998	–0.037145	0.18268	$\hat{\gamma}_1$	–0.01164	0.04368	–0.05769	0.163290
$V(\hat{\gamma}_1)$	(10.01940)	(17.50450)	(35.98690)	(184.9340)	$V(\hat{\gamma}_1)$	(9.974747)	(17.4140)	(36.0447)	(187.395)
$\alpha = 5, \beta = 1$					$\alpha = 5, \beta = 3$				
n	30	50	100	500	n	30	50	100	500
$F(\mu)$	0.5	0.5	0.5	0.5	$F(\mu)$	0.5	0.5	0.5	0.5
γ_p	0	0	0	0	γ_p	0	0	0	0
γ_1	–	–	–	–	γ_1	–	–	–	–
$s(\bar{x})$	0.49789	0.40452	0.50187	0.49299	$s(\bar{x})$	0.49689	0.50066	0.49658	0.50047
$\hat{\gamma}_p$	–0.00421	0.00904	0.00373	–0.01401	$\hat{\gamma}_p$	–0.00623	0.00131	–0.00685	–0.000094
$V(\hat{\gamma}_p)$	(0.28464)	(0.29411)	(0.30972)	(0.33416)	$V(\hat{\gamma}_p)$	(0.28457)	(0.29544)	(0.31489)	(0.32468)
$\hat{\gamma}_1$	–0.02821	0.06320	–0.00185	–0.34661	$\hat{\gamma}_1$	–0.02065	0.02882	–0.07722	–0.05429
$V(\hat{\gamma}_1)$	(10.18350)	(17.28080)	(36.06790)	(90.82100)	$V(\hat{\gamma}_1)$	(10.27350)	(17.5818)	(36.5457)	(186.9300)

$\gamma_p = 0.3318$ and the corresponding $\bar{\gamma}_p = 0.3299$. This is yet another evidence in the favor of γ_p projecting it as a serious member of the family of skewness measures. Similar patterns are observed in the third and fourth cases; the glimpses of the performance are offered by highlighting the scenarios where $v_1 = 3, v_2 = 7$, and $v_1 = 9, v_2 = 7$.

3.4. The Rayleigh and Cauchy distributions

Both of these distributions are interesting members of our study in case the extent of skewness in Rayleigh distribution remains unchanged with respect to any amendment in parameter while Cauchy distribution does not allow the existence of usual skewness measure, γ_1 , in any circumstances. The findings related to Rayleigh distribution are offered in Table 7. In larger sample sizes, the estimates approach to the population measures with consistent manner. The variance behavior of $\hat{\gamma}_p$ is again impressive in comparison to $\hat{\gamma}_1$.

For Cauchy distribution, the true value of γ_p is deducted by replacing $F(\mu)$ with $F(\text{Median})$ (conceptually it is not forbidden because of the symmetry of the distribution). The evaluation of $\hat{\gamma}_p$ (and resulting $\bar{\gamma}_p$) is still based on empirical cdf encapsulating mean of the data. The results in Table 8 highlight the superiority of the proposed functional in terms of stability and interpretability. In every considered situation, the estimated value of $\hat{\gamma}_p$ remains very close to population value, projecting inherited symmetry of the distribution. The behavior of variance is, however, altered; it increases with increase in sample size (surely, enormously less than that of $\hat{\gamma}_1$). This fact at this stage is attributed to the well-recognized complex functional form of the Cauchy distribution and is left for future inquiries.

4. Influence function comparison

The influence function, being the directional derivative of functional, is usually utilized to compare the extent of robustness inherent in functionals against unusual points (see [21]). In its general form, for a functional $T(\cdot)$, we can write influence function as follows:

$$T(F; G) = \lim_{\epsilon \downarrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon},$$

where $F_\epsilon = \epsilon G + (1 - \epsilon) F$ and $T(F; G)$ is the directional derivative of functional T at F in the direction of

G . Under the assumption of symmetry of distribution, say F around ‘0’ with differentiable density function, let us consider $G(x) = (1/2b\sigma)[x + (b - a)\sigma]$, where σ^2 is the variance of F and X is bound between $[(a - b)\sigma, (a + b)\sigma]$ with $-\infty < a < \infty$ and $b > 0$. The authors provided the general form of F_ϵ Eq. (2) as shown in Box I. The resultant influence functions for γ_1 and γ_p , respectively, are given below:

$$IF(a, b, F, \gamma_1) = a^3 - a(3 - b^2), \tag{3}$$

and:

$$IF(a, b, F, \gamma_p) = \begin{cases} -1, & \text{for } a > b \\ -a/b, & \text{for } -b < a < b \\ 1, & \text{for } a < -b \end{cases} \tag{4}$$

respectively. For comparison purposes, a graphical display of influence functions of both measures is offered in Figure 2, fixing the value of b at $1/2$, whereas ‘ a ’ can take values over the permissible range. It reveals that the influence function γ_p is bounded in contrast to that of γ_1 . It is observable that both of the influence functions behave almost identically in the range of $0 < a < 1/2$, whereas for $a > 1/2$, distinctive features of $IF(a, b, F, \gamma_1)$ and $IF(a, b, F, \gamma_p)$ are witnessed. The influence function of the proposed functional stays constant and remains negative, highlighting the resilience against contamination after certain level. On the other hand, $IF(a, b, F, \gamma_1)$ endorses its reliability when $0 < a < 1.65$ (the upper limit 1.65 can easily

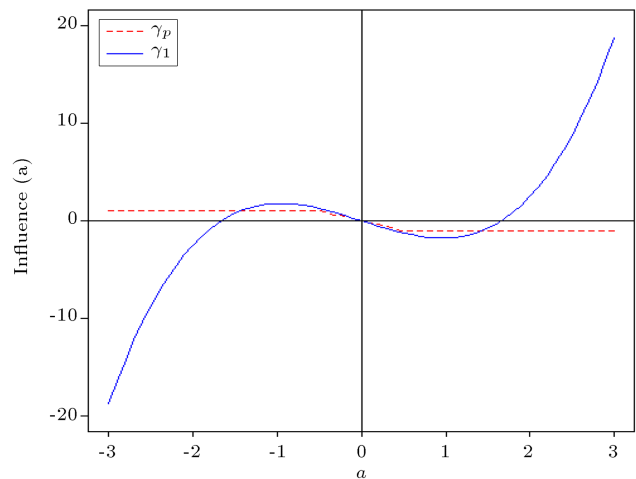


Figure 2. Influence functions of γ_1 and γ_p .

$$F_\epsilon = \begin{cases} (1 - \epsilon) F(x), & x < (a - b)\sigma \\ \epsilon (1/2b\sigma)[x + (b - a)\sigma] + (1 - \epsilon) F(x), & (a - b)\sigma \leq x \leq (a + b)\sigma \\ \epsilon + (1 - \epsilon) F(x), & x > (a + b)\sigma \end{cases} \tag{2}$$

Box I

be verified by putting $b = 1/2$ in Eq. (3)); however, after that point, the influence function explodes. The altered behavior of $IF(a, b, F, \gamma_1)$ for $a > 1.65$ is not only counter intuitive but also poses serious threats to the reliability of estimate in the presence of outliers.

5. Applications

5.1. Algeria’s fatality count data (1997 – 2017)

This section demonstrates the applicability of the proposed functional by examining yearly data providing the Algeria’s fatality counts from the year 1997 to 2017. The data were derived from the website of Armed Conflict Location & Event Data (ACLED) project. The histogram presented in Figure 3 shows positive skewness in the data which is rightly projected in the magnitudes of both estimates (see Table 9).

After estimating the degree of skewness for complete data set, we remove outliers, visible in box plot (Figure 3), and then estimate γ_1 and γ_p . Table 9 comprehends the results for readers. The robustness of γ_p against outliers is self-evident under the heading of absolute relative percentage change (RC). The results in Table 9 depict the moderate extent of skewness.

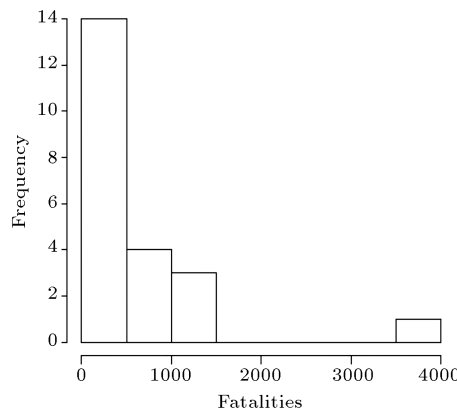


Figure 3. Histogram and box plot depicting the yearly fatality counts for Algeria (1997–2017).

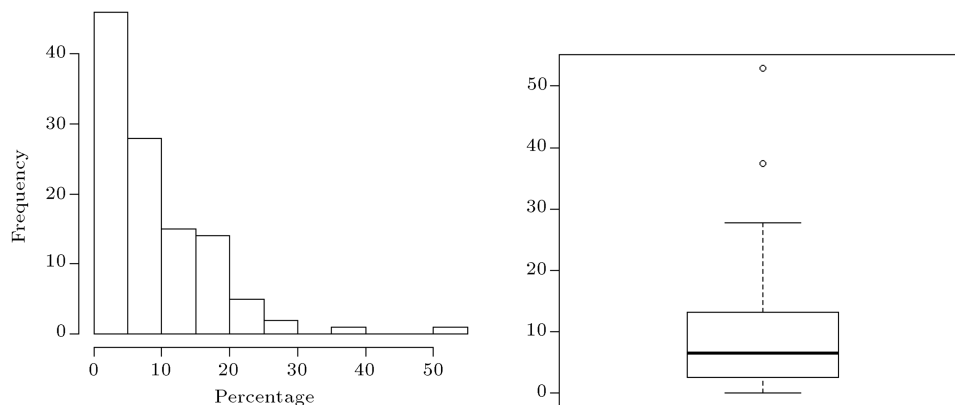


Figure 4. Histogram and box plot depicting the respondents’ behavior form Pakistan Social and Living Standard Measurements (PSLM) (2014-15) survey.

Table 9. Performance comparison of $\hat{\gamma}_p$ and $\hat{\gamma}_1$ in the presence of outliers-Algeria data.

	Full dataset	Reduced dataset	RC
$\hat{\gamma}_p$	0.3636	0.3000	17.49%
$\hat{\gamma}_1$	2.8932	1.1453	60.41%

5.2. PBS data on respondents’ extent of co-operation (non-cooperation) in PSLM (2014-15) survey

Next, we explore the utility of the proposed measure in comparison to the usual moment-based measure of skewness by studying Pakistan Social and Living Standard Measurements (PSLM) survey (2014–15) data compiled by PBS. The histogram and box plot given in Figure 4 present the pictorial display of percentages of non-cooperative respondents across 114 districts of the country.

Table 10 documents the robustness of the proposed and commonly used moment-based skewness measures against existing outliers in the data. We initiate the skewness estimation, employing γ_1 and γ_p , by first considering complete data set and, then, through the reduced data set (by dropping outliers). The absolute relative percentage change prominently

Table 10. Performance comparison of $\hat{\gamma}_p$ and $\hat{\gamma}_1$ in the presence of outliers-PSLM data.

	Full dataset	Reduced dataset	RC
$\hat{\gamma}_p$	0.2613	0.2294	12.21%
$\hat{\gamma}_1$	1.9405	0.8616	55.60%

reveals the robustness of the proposed functional in comparison to the usual measure. In the case of our proposition, outliers derive almost 12% change in the value of the proposed skewness functional, whereas a change of almost 55% (more than 4 times that of the proposed functional) is associated with the value of usual measure when outliers are active. The overall results given in Table 10 show the greater extent of skewness in the data. Moreover, the values of the proposed measure remain comparable over the consistent range of $[-1, 1]$. In this respect, Table 9 shows the results related to the Algeria data involving $\gamma_p = 0.3636$ and $\gamma_p = 0.2613$ for the PBS data in Table 10; it remains plausible to conclude that the Algeria data reveals a higher extent of skewness (to the right) than the PBS data.

6. Discussion

In this paper, a new measure of skewness, γ_p , was introduced based on the distribution function and mean of the distribution. The novelty of the contribution was demonstrated on three fronts: (i) development, (ii) establishment, and (iii) assessment. Being distinctive from existing measures, the proposed measure does not require the knowledge of higher order moments or uniqueness of mode of distribution. It was shown that γ_p is proper skewness measuring functional satisfying all essential characteristics recommended in literature. Defined over interpretable range, i.e., ‘0’ projecting the symmetry of distribution and a value of ‘+1(−1)’ indicating extreme right (left) skewness, γ_p is straight forward to calculate. Following the spirit of competition, an intense comparative study was conducted for γ_p which was found the most commonly used and comprehensive skewness measure among existing ones. Superiority of the proposed measure was witnessed without exceptions. The influence function exhibited the resistance of the proposed measure to unusual points in data. In reasonably large sample sizes, it converged to its population value with a minimal amount of variability. The computational ease, interpretability, and rigorous use of available information projected γ_p as a potential candidate for future research. In particular, variance behavior for increasing sample sizes makes it worth studying to develop tests of skewness exploiting asymptotic theory. Its utility in high-dimensional data is yet another research venue.

Therefore, it merits attention as a skewness measure with respect to mean.

Lastly, it is appropriate to recognize some complexities in the case of working with a discrete data set. Our proposed measure function based on the empirical evaluation of the cdf at mean, and for discrete data, the mean may not always be the member of the data, see for example, the data of Altınay in Table 3 [26]. In this scenario, a cautious use of the proposed measure is suggested. To our understanding, in the above given situation, the empirical cdf at mean fails to efficiently estimate population cdf at mean, which may lead to misleading results. It is interesting to explore the possible amendments in the proposed framework to handle the aforementioned complication.

References

1. Pearson, K. “Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material”, *Philosophical Transactions of the Royal Society of London*, **186**(Part I), pp. 343–424 (1895).
2. Kendall, M. and Stuart, A., *The Advanced Theory of Statistics*, 4th Ed., **1**, *Distribution Theory*, London: Griffin (1977).
3. Firat, M., Koc, A.C., Dikbas, F., et al. “Identification of homogeneous regions and regional frequency analysis for Turkey”, *Scientia Iranica, Transactions A, Civil Engineering*, **21**(5), pp. 1492–1502 (2014).
4. Harvey, A. and Sucarrat, G. “EGARCH models with fat tails, skewness and leverage”, *Computational Statistics & Data Analysis*, **76**, pp. 320–338 (2014).
5. Amaya, D., Christoffersen, P., Jacobs, K., et al. “Does realized skewness predict the cross-section of equity returns?”, *Journal of Financial Economics*, **118**(1), pp. 135–167 (2015).
6. Åstebro, T., Mata, J., and Santos-Pinto, L. “Skewness seeking: risk loving, optimism or overweighting of small probabilities?”, *Theory and Decision*, **78**(2), pp. 189–208 (2015).
7. Ho, A.D. and Yu, C.C. “Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects”, *Educational and Psychological Measurement*, **75**(3), pp. 365–388 (2015).
8. Decker, R.A., Haltiwanger, J., Jarmin, R.S., et al. “Where has all the skewness gone? The decline in high-growth (young) firms in the US”, *European Economic Review*, **86**, pp. 4–23 (2016).
9. Colacito, R., Ghysels, E., Meng, J., et al. “Skewness in expected macro fundamentals and the predictability of equity returns: Evidence and theory”, *The Review of Financial Studies*, **29**(8), pp. 2069–2109 (2016).
10. Ensthaler, L., Nottmeyer, O., Weizsäcker, G., et al. “Hidden skewness: On the difficulty of multiplicative compounding under random shocks”, *Management Science*, **64**(4), pp. 1693–1706 (2017).

11. Amjadzadeh, M. and Ansari-Asl, K. “An innovative emotion assessment using physiological signals based on the combination mechanism”, *Scientia Iranica*, **24**(6), pp. 3157–3170 (2017).
12. Taylor, S. and Fang, M. “Unbiased weighted variance and skewness estimators for overlapping returns”, *Swiss Journal of Economics and Statistics*, **154**(1), pp. 1–8 (2018).
13. Zwet, W.R., *Convex Transformations of Random Variables*, Mathematisch Centrum, Amsterdam (1964).
14. Goldstein, J.L., Schrott, H.G., Hazzard, W.R., et al. “Hyperlipidemia in coronary heart disease II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia”, *The Journal of Clinical Investigation*, **52**(7), pp. 1544–1568 (1973).
15. Mardia, K.V., *Families of Bivariate Distributions*, **27**, Lubrecht & Cramer Ltd (1970).
16. Aucremanne, L., Brys, G., Hubert, M., et al. “A study of Belgian inflation, relative prices and nominal rigidities using new robust measures of skewness and tail weight”, *Statistics for Industry and Technology*, pp. 13–25 (2004).
17. Brys, G., Hubert, M., and Struyf, A. “A robust measure of skewness”, *Journal of Computational and Graphical Statistics*, **13**(4), pp. 996–1017 (2004).
18. Doane, D.P. and Seward, L.E. “Measuring skewness: a forgotten statistic?”, *Journal of Statistics Education*, **19**(2), pp. 1–18 (2011).
19. Hosking, J. “Moments or L moments? An example comparing two measures of distributional shape”, *The American Statistician*, **46**(3), pp. 186–189 (1992).
20. Li, X., Qin, Z., and Kar, S. “Mean-variance-skewness model for portfolio selection with fuzzy returns”, *European Journal of Operational Research*, **202**(1), pp. 239–247 (2010).
21. Arnold, B.C. and Groeneveld, R.A. “Measuring skewness with respect to the mode”, *The American Statistician*, **49**(1), pp. 34–38 (1995).
22. Brys, G., Hubert, M., and Struyf, A. “A comparison of some new measures of skewness”, *Developments in Robust Statistics (ICORS 2001)*, **14**, pp. 98–113 (2003).
23. Kim, T.H. and White, H. “On more robust estimation of skewness and kurtosis”, *Finance Research Letters*, **1**(1), pp. 56–73 (2004).
24. Holgersson, H. “A modified skewness measure for testing asymmetry”, *Communications in Statistics-Simulation and Computation*, **39**(2), pp. 335–346 (2010).
25. Tajuddin, I. “On the use of medcouple as a measure of skewness”, *Pakistan Journal of Statistics*, **28**(1), pp. 69–80 (2012).
26. Altinay, G. “A simple class of measures of skewness”, MPRA Paper No. 72353, posted 4 July 2016, <https://mpra.ub.uni-muenchen.de/72353/>(2016).
27. Oja, H. “On location, scale, skewness and kurtosis of univariate distributions”, *Scandinavian Journal of Statistics*, **8**(3), pp. 154–168 (1981).

Biographies

Akbar Ali Khan is an Assistant Professor of Statistics at Higher Education Department Khyber Pakhtunkhwa, Pakistan. He is currently pursuing his PhD at the Department of Statistics Quaid-i-Azam University, Islamabad, Pakistan. His research interests are the fields of Bayesian statistics, mathematical statistics, and estimation procedures. He has published research papers in well reputed national and international journals. He is associated with the teaching profession for the last ten years.

Salman Arif Cheema did his MS in Virginia Tech, USA and received his PhD degree from Department of Mathematical and Physical Sciences, Newcastle University, Australia. His research work is mainly focused on mathematical statistics and qualitative data modeling. He has published research articles in well reputed international journals.

Zawar Hussain is the Head of the Department of Social and Allied Sciences Cholistan University of Veterinary and Animal Sciences Bahawalpur, Pakistan. He has served the Department of Statistics, Quaid-i-Azam University Islamabad, Pakistan for eighteen years. His research interests lie in the fields of randomized response techniques, Bayesian statistics, mathematical statistics, regression analysis, and statistical quality control. He has published more than hundred research papers in well reputed national and international journals.

Abdel-Salam Gomma Abdel-Salam is an Associate Professor at the Department of Mathematics, Statistics and Physics, Qatar University, Doha, Qatar. He did his PhD from Virginia Tech, USA. His research interests lie in the fields of semi-parametric statistical modeling. His research papers are published by the well reputed international journals.