Mathematical Methods of Statistics manuscript No.
(will be inserted by the editor)

On the skewness order of van Zwet and Oja

Andreas Eberl · Bernhard Klar

Received: date / Revised: date

Abstract Van Zwet (1964) introduced the convex transformation order between two distribution functions F and G, defined by $F \leq_c G$ if $G^{-1} \circ F$ is convex. A distribution which precedes G in this order should be seen as less right-skewed than G. Consequently, if $F \leq_c G$, any reasonable measure of skewness should be smaller for F than for G. This property is the key property when defining any skewness measure.

In the existing literature, the treatment of the convex transformation order is restricted to the class of differentiable distribution functions with positive density on the support of F. It is the aim of this work to analyse this order in more detail. We show that several of the most well known skewness measures satisfy the key property mentioned above with very weak or no assumptions on the underlying distributions. In doing so, we conversely explore what restrictions are imposed on the underlying distributions by the requirement that F precedes G in convex transformation order.

Keywords Convex order · Convex transformation order · Measure of skewness · Quantile skewness

1 Introduction

Even though skewness is one of the oldest concepts in distribution theory dating back at least to Pearson (1895), its formal analysis did not start until the pioneering work of van Zwet (1964). He introduced the c-order (for convex order or convex transformation order) between two distribution functions F

A. Eberl
Institut für Stochastik, Karlsruher Institut für Technologie (KIT)
E-mail: andreas.eberl@kit.edu

B. Klar
Institut für Stochastik, Karlsruher Institut für Technologie (KIT)

and G, defined by $F \leq_c G$ if $G^{-1} \circ F$ is convex. Van Zwet postulated that a distribution which c-precedes G should be seen as less right-skewed than G.

Under certain regularity conditions, this is plausible for a number of reasons. First, $G^{-1} \circ F$ is the unique function that transforms an F-distributed random variable into a G-distributed random variable (see van Zwet 1964, p. 48, Lemma 4.1.1). If this transformation is convex, the probability mass in the left tail is condensed while the probability mass in the right tail is spread out, resulting in an increase in right-skewness. Second, Oja (1981, p. 162, Theorem 5.2) found an equivalent characterization of the order $\leq_c$ in case that F and G are skewness-comparable with finite second moments . It states that the cdf's, standardized with mean and standard deviation, intersect exactly twice with F being larger at either tail. Graphically, this also means that, compared with F, G has less probability mass in the left tail and more in the right tail.

As a consequence, if $F \leq_c G$, any reasonable measure of skewness should be smaller for F than for G. This property is termed P.4 in Groeneveld and Meeden (1984) and Arnold and Groeneveld (1995). Further cornerstones in the theory of skewness orders have been the work by Oja (1981) and MacGillivray (1986). From the beginning, attention was confined to the class of (twice) differentiable distribution functions with positive density on the support of F; see, among many others, van Zwet (1964, p.24), Oja (1981, p.155), MacGillivray (1986, p.995), Groeneveld and Meeden (1984, p.391). In particular, these assumptions are necessary for the aforementioned statements about the graphical plausibility of the convex transformation order. It is not clear if this restricted setting is posed only for mathematical convenience or for a deeper insight into the limited usability of the proposed skewness orders and related skewness measures for more general distributions.

It is the aim of this work to analyse these questions in detail. We show that several of the most well known skewness measures satisfy P.4 with very weak or no assumptions on the underlying distributions. In doing so, we conversely explore what restrictions are imposed on the underlying distributions by the requirement that F c-precedes G. Specifically, these restrictions imply that the c-order is not applicable to lattice distributions.

In the following, the quantile function of a cumulative distribution function (cdf) H on $\mathbb{R}$ is defined as $H^{-1} : (0, 1) \to \mathbb{R}, p \mapsto \inf\{t \in \mathbb{R} : H(t) \geq p\}$. As indicated, the order in the first part of the next definition was proposed by van Zwet (1964, Def. 4.1.1) and has been studied in many publications (Oja 1981; MacGillivray 1986; Groeneveld and Meeden 1984). It is also discussed in Shaked and Shanthikumar (2006, Chapter B.4). The order in the second part is a seemingly reasonable extension to the class of all cdf's on $\mathbb{R}$.

Definition 1 Define $D_H = \mathbb{R} \setminus H^{-1}(\{0, 1\})$ for any cdf H on $\mathbb{R}$.

a) Let F and G be cdf's on $\mathbb{R}$ that are continuous on the left endpoint of their respective supports. Then G is said to be *at least as skewed to the right as* F, denoted by $F \leq_c G$, if the mapping $G^{-1} \circ F : D_F \to \mathbb{R}, t \mapsto G^{-1}(F(t))$ is convex.

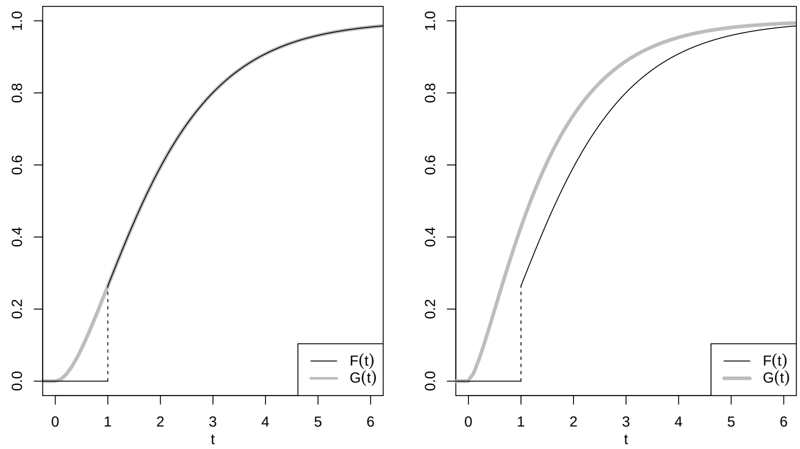

Fig. 1 Cdf's F and G as defined in Example 1, in the left panel both with shape parameter 2, in the right panel F with shape parameter 2 and G with 1.5. In both panels $F \leq_c G$ holds, in the right one even in a strict sense.

b) Let F and G be arbitrary cdf's on $\mathbb{R}$. Then G is said to be *at least as skewed to the right as* F, denoted by $F \leq_c G$, if the mapping $G^{-1} \circ F : D_F \to \mathbb{R}, t \mapsto G^{-1}(F(t))$ is convex and $\inf G(D_G) \geq \inf F(D_F)$ holds.

Here, the domain of $G^{-1} \circ F$ has been restricted to $\mathbb{R} \setminus F^{-1}(\{0,1\})$, since any quantile function is only defined on the open interval $(0,1)$, while the codomain of any cdf is the closed interval $[0,1]$. To solve this, all points that F maps onto $\{0,1\}$ have to be excluded from the domain of this composition. Since sets of this kind will play a central role in this paper, we denote the restricted domain of a cdf H on $\mathbb{R}$ by $D_H = \mathbb{R} \setminus H^{-1}(\{0,1\})$, which entails $H(D_H) = H(\mathbb{R}) \setminus \{0,1\}$.

In the case of cdf's that are not continuous on the left endpoint of their support, the additional condition in the second part of Definition 1 is indeed necessary. As an illustration, consider the following example:

Example 1 Let $Y \sim \Gamma(1.5, 1)$ with corresponding cdf G and let $F(t) = G(t)\mathbb{1}_{[1,\infty)}(t)$ for all $t \in \mathbb{R}$. Now F is obviously more skewed to the right than G (see left panel of Fig 1). However, since F and G are identical on D_F, $G^{-1} \circ F$ is the identity and by applying the first part of Definition 1 to all cdf's on $\mathbb{R}$, we would have $F \leq_c G$. By decreasing the shape parameter of Y without changing F, we would preserve $F \leq_c G$, which now holds even in a strict sense (see van Zwet (1964, pp.60-62)). However, the distributions would be difficult to compare with respect to skewness (see right panel of Fig 1). Hence, it seems reasonable to exclude such a combination of cdf's from being comparable. Otherwise, none of the popular skewness measures discussed in this paper would preserve the order $\leq_c$.

It should be noted that there exists a wide variety of skewness orders for distributions; examples can be found in Oja (1981), MacGillivray (1986), Arnold and Groeneveld (1993). Among all skewness orders that have been considered so far, $\leq_c$ seems to be one of the strongest. Based on the chosen skewness order, in this case $\leq_c$, Oja (1981) and Groeneveld and Meeden (1984) among others proposed sets of properties that any adequate skewness measure should satisfy. In this paper, we consider the following three properties that should be satisfied by any reasonable skewness measure $\gamma : \mathcal{R} \to \mathbb{R}$, where $\mathcal{R}$ denotes a suitable set of random variables on $\mathbb{R}$.

(S1) For $c > 0$ and $d \in \mathbb{R}$, $\gamma(cX + d) = \gamma(X)$, i.e. γ is invariant under transformations of scale and location.
(S2) It holds $\gamma(-X) = -\gamma(X)$.
(S3) Denote the cdf of X and Y by F and G. If G is at least as skewed to the right as F, then $\gamma(X) \leq \gamma(Y)$ holds, in short

$$F \leq_c G \Rightarrow \gamma(F) \leq \gamma(G).$$

These properties are equivalent to the two properties given by Oja (1981). In Groeneveld and Meeden (1984), a fourth property is considered, stipulating the skewness measure of any symmetric random variable to be zero. However, this fourth property easily follows by combining (S1) and (S2) and is therefore not included here. Furthermore, it follows from these three properties that any adequate skewness measure takes non-negative values for right-skewed random variables and non-positive values for left-skewed random variables.

It should be noted that (S3) seems to the weakest possible property of this kind. It could either be strengthened by replacing $\leq_c$ by some weaker skewness order or by replacing both inequalities by strict inequalities, that is by requiring $F <_c G \Rightarrow \gamma(F) < \gamma(G)$ (where $F <_c G$ holds if $G^{-1} \circ F$ is strictly convex).

In Section 2, we start with arbitrary distribution functions F and G, assuming only $F \leq_c G$, and analyze the resulting restrictions on F and G. In particular, lattice distributions are ruled out by this requirement.

In Section 3, we examine the validity of properties (S1) to (S3) for four well-known skewness measures: the standardized third moment

$$\gamma_M(X) = \mathbb{E}\left((X-\mu)/\sigma)^3\right),$$

where $\mu = \mathbb{E}X$ and $\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$, quantile skewness

$$\gamma_Q^{(\alpha)}(X) = \frac{\left(F^{-1}(1-\alpha) - F^{-1}(1/2)\right) - \left(F^{-1}(1/2) - F^{-1}(\alpha)\right)}{F^{-1}(1-\alpha) - F^{-1}(\alpha)},$$

integrated quantile skewness

$$\gamma_{IQ}(X) = \left(\mathbb{E}X - F^{-1}(1/2)\right)/\mathbb{E}|X - F^{-1}(1/2)|,$$

and, finally, Pearson's skewness measure, defined by

$$\gamma_P(X) = \left(\mathbb{E}X - F^{-1}(1/2)\right)/\sigma.$$

While (S3) holds for $\gamma_Q^{(\alpha)}(X)$ and $\gamma_{IQ}(X)$ without any assumptions on the underlying distributions, one has to impose some weak conditions for the validity of (S2) as well as for the validity of (S3) for $\gamma_M(X)$. For $\gamma_P(X)$, (S3) is not fulfilled. Besides, we examine whether the four skewness measures can be normalized; if so, we further investigate whether the lower and upper bounds are sharp. The proof of Lemma 1 is postponed to Appendix A.

2 A useful lemma

Among the four skewness properties, (S3) plays the pivotal role in the sense that it generally is most difficult to prove. In prior proofs of these properties for specific skewness measures as in Groeneveld and Meeden (1984), this difficulty was partly circumvented by only considering distributions with twice differentiable cdf's. However, in this work we aim at proving (S3) with minimal assumptions on the underlying random variables. These proofs are simplified greatly through the following lemma, especially parts a) and b). They state that a substantial set of combinations of cdf's F and G cannot be ordered with respect to the skewness order $\leq_c$ and therefore do not need to be considered in the proof of property (S3). Part c) is a result concerning the distribution of a random variable with cdf F, transformed by $G^{-1} \circ F$, which is also of some help in proving (S3).

Lemma 1 *Let F and G be two arbitrary cdf's on $\mathbb{R}$.*

a) If $F \leq_c G$, then $G(D_G) \subseteq F(D_F) \cup [\sup F(D_F), 1)$.
b) If $F \leq_c G$, then F and G are strictly increasing on D_F and D_G, respectively, up to the following exceptions:
  (i) F is constant on non-degenerate intervals with values within the interval $(0, \inf G(D_G)]$.
  (ii) G is constant on non-degenerate intervals with values within the interval $[\sup F(D_F), 1)$.
c) Let X be a random variable with cdf F and define $Y = G^{-1}(F(X))$. If, additionally, $G(D_G) \subseteq F(D_F) \cup \{\sup F(D_F)\}$, then $Y \sim G$.

A verbalized version of part a) of the above lemma would be that, aside from any values greater than or equal to $\sup F(D_F)$, any value skipped by F is also skipped by G if $F \leq_c G$ holds. This formulation will turn out to be quite useful later on in this work.

By parts a) and b) of Lemma 1, the number of distributions that can be ordered with respect to $\leq_c$ is severely reduced. In particular, this applies to most discrete distributions and partly discrete distributions. Concerning skewness property (S3), these results merely simplify the corresponding proofs. However, in the search for appropriate properties for characterizing skewness, this renders (S3) useless, at least for (partly) discrete distributions. A possible solution could be the usage of a weakened skewness order, thus strengthening any skewness measure behaving in accordance to such an order. However, it should be noted that alternatively proposed weaker skewness orders (Oja 1981; MacGillivray 1986; Arnold and Groeneveld 1993) are also only applicable for continuous distributions. Nonetheless, in this work, we will solely examine whether the established properties (S1) - (S3) are satisfied for the skewness measures under consideration.

However, the consideration of distributions with non-differentiable cdf's and the subsequent redefinition of the order $\leq_c$ in Definition 1b) does indeed enlarge the set of distributions that can be ordered with respect to $\leq_c$. This enlargement does not only consist of absolutely continuous distributions with additional jumps that satisfy the condition in Lemma 1a), but also contains strictly discrete distributions, as the following example shows.

Example 2 Let $p_0 = p_{-0} \in (0,1)$. Additionally, let $(p_{-n})_{n\in\mathbb{N}}$ be a sequence on $[0, p_0)$ satisfying $p_{-n} \searrow 0$ as $n \to \infty$ and let $(p_n)_{n\in\mathbb{N}}$ be a sequence on $(p_0, 1]$ satisfying $p_n \nearrow 1$ as $n \to \infty$. Now define two distribution functions F and G by

$$F(t) = \sum_{n=0}^{\infty} p_{-n}\, \mathbb{1}_{[-n,-n+1)}(t) + \mathbb{1}_{[1,\infty)}(t),$$
$$G(t) = \sum_{n=0}^{\infty} p_n\, \mathbb{1}_{[n,n+1)}(t)$$

for $t \in \mathbb{R}$. Obviously,

$$\inf G(D_G) = p_0 \geq \inf\left(\{p_{-n} : n \in \mathbb{N}_0\} \cap (0, p_0]\right) = \inf F(D_F).$$

Define $n_1 = \inf\{n \in \mathbb{N} : p_n = 1\}$ with the convention $\inf \emptyset = \infty$. Then the quantile function of G is given by

$$G^{-1}(p) = \sum_{n=1}^{n_1-1} n \mathbb{1}_{(p_{n-1},p_n]}(p) + n_1 \mathbb{1}_{(p_{n_1-1},1)}(p)$$

for $p \in (0,1)$. Note that G^{-1} only takes values different from 0 for $p \in (p_0, 1)$. However, because of $F(D_F) \subseteq (0, p_0]$, $G^{-1} \circ F$ is constantly equal to 0 on D_F and thereby also convex. This means that $F \leq_c G$ holds for these two discrete distributions.

Figure 2 shows F and G for $p_{-n} = 1/(n+2)$ and $p_n = 1-1/(n+2)$, $n \in \mathbb{N}_0$. Here, it is also graphically obvious that G is more skewed to the right than F.

Another result helpful in proving property (S3) is the following characterization of convexity which can be found in Niculescu and Persson (2018, p. 31). Actually, it is the definition of a convex function in Artin's treatise on the gamma function (Artin 1964).

Lemma 2 *A mapping $\varphi : I \to \mathbb{R}$ on some arbitrary (finite or infinite) interval I is convex if and only if for any $a \in I$ the mapping $s_a : I \setminus \{a\} \to \mathbb{R}$ defined by*

$$s_a(t) = \frac{\varphi(t) - \varphi(a)}{t - a}, t \in I \setminus \{a\}$$

is non-decreasing.

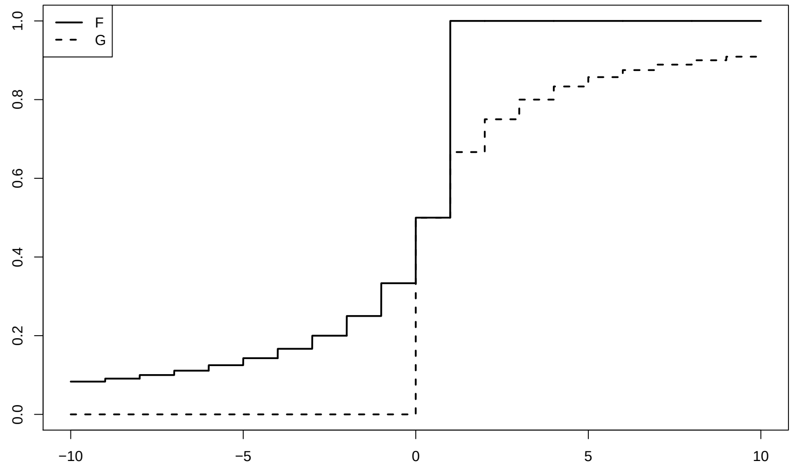

Fig. 2 Cdf's F and G as defined in Example 2 for specific choices of $p_n, n \in \mathbb{Z}$.

3 Analyzing skewness measures with respect to the skewness order

3.1 Moment based skewness measure

Probably, the best known skewness measure is the moment based skewness measure, which is often used synonymously with the notion of skewness itself. Let μ_j denote the jth non-central moment, and let X be a random variable with $\mathbb{E}|X|^3 < \infty$ and strictly positive variance $\sigma^2 = \mu_2 - \mu^2 > 0$. The moment based skewness measure is defined as the standardized third moment, i.e.

$$\gamma_M(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3 - 3\mu_2\mu + 2\mu^3}{(\mu_2 - \mu^2)^{\frac{3}{2}}}.$$

Clearly, $\gamma_M(X)$ is only well-defined if X is not almost surely constant.

For all skewness measures, we first examine whether they can be normalized. The following example shows that $\gamma_M(X)$ can not be normalized.

Example 3 For $X \sim \mathrm{Bin}(1,p), p \in (0,1)$, the moment based skewness is

$$\gamma_M(X) = \frac{1-2p}{\sqrt{p(1-p)}}.$$

Obviously, $\gamma_M(X) \to \infty$ for $p \to 0$ as well as $\gamma_M(X) \to -\infty$ for $p \to 1$, which shows that the moment based skewness measure can not be normalized.

In the following theorem we prove that the moment based skewness measure satisfies skewness properties (S1) and (S2) without any assumptions on the underlying distributions as well as (S3) under a weak condition.

Theorem 1 *The moment based skewness measure $\gamma_M(X)$ satisfies properties (S1) and (S2) for all distributions for which it is well defined. (S3) is satisfied if $\sup G(D_G) \leq \sup F(D_F)$ holds.*

Proof Properties (S1) and (S2) follow easily. Concerning (S3), let F and G be cdf's satisfying $F \leq_c G$ and let $X \sim F$ as well as $Y \sim G$. Theorem 2.2.1 in van Zwet (1964) states that for any convex, non-decreasing, non-constant function φ on the minimal interval I_F, on which $\mathbb{P}(X \in I_F) = 1$ holds, we have for any $k \in \mathbb{N}$ that

$$\frac{\mathbb{E}[(X-\mu)^{2k+1}]}{\sigma^{2k+1}} \leq \frac{\mathbb{E}[(\varphi(X) - \mathbb{E}[\varphi(X)])^{2k+1}]}{\sqrt{\mathbb{V}[\varphi(X)]}^{2k+1}}.$$

Obviously, we choose $k = 1$ and $\varphi = G^{-1} \circ F$. Combined with $F \leq_c G$, the additional assumption for (S3) yields $G(D_G) \subseteq F(D_F) \cup \{\sup F(D_F)\}$. Due to Lemma 1c), we then have $\varphi(X) \sim Y$ and from above result $\gamma_M(X) \leq \gamma_M(Y)$ follows, if $G^{-1} \circ F$ satisfies the assumptions on φ.

To that end, if F does not have a jump discontinuity ending at the value 1, the right continuity of F yields $\mathbb{P}(X \in D_F) = 1$ and we can choose $I_F = D_F$. Otherwise, X takes on the value $\sup D_F \notin D_F$ with positive probability and we choose $I_F = D_F \cup \{\sup D_F\}$. In order to establish this as the extended domain of $G^{-1} \circ F$, we have to assign a function value to the point $\sup D_F$. Based on $F(\sup D_F) = 1$, we assign $G^{-1}(F(\sup D_F)) = \inf\{t \in \mathbb{R} : G(t) \geq 1\} = \sup D_G$.

By assumption, $G^{-1} \circ F$ is convex on D_F. Since it maps D_F onto a subset of $D_G \cup \{\sup D_G\}$, the addition of $\sup D_F$ to the domain of F results in the function being either extended continuously or exhibiting a upward jump at that point. However, this preserves the convexity of $G^{-1} \circ F$ on the extended domain (Niculescu and Persson 2018, Theorem 1.1.2). Since, furthermore, the function is non-decreasing on D_F, this property is also preserved.

Assuming that $G^{-1} \circ F$ is constant (on I_F) results in Y almost surely taking a constant value, since $Y \sim G^{-1}(F(X))$ due to Lemma 1c). However, by assumption, Y is not almost surely constant. Therefore $G^{-1} \circ F$ is non-constant on I_F, so it satisfies the assumptions on φ, thus concluding the proof. □

Remark 1 In order to examine the necessity of the additional assumption for (S3), we assume $F \leq_c G$. Due to the definition of that order, it takes the behaviour of G into account only up to the point $G^{-1}(\sup F(D_F))$. The information concerning values of G at any $t \in (G^{-1}(\sup F(D_F)), \sup D_G]$ is therefore lost. So $F \leq_c G$ essentially implies $\tilde{F} \leq_c \tilde{G}$ for cut off and scaled versions of the original cdf's defined by

$$\tilde{H}(t) = \frac{H(t)}{\sup F(D_F))} \mathbb{1}_{\{H(t) < \sup F(D_F)\}} + \mathbb{1}_{\{H(t) \geq \sup F(D_F)\}}$$

for $H \in \{F, G\}$. Since for these new cdf's, we have $\tilde{G}(D_{\tilde{G}}) \subseteq \tilde{F}(D_{\tilde{F}}) \cup \{\sup \tilde{F}(D_{\tilde{F}})\}$, we can infer $\gamma_M(\tilde{F}) \leq \gamma_M(\tilde{G})$ by replicating the proof of Theorem 1. Now consider the depiction of an exemplary constellation in which the

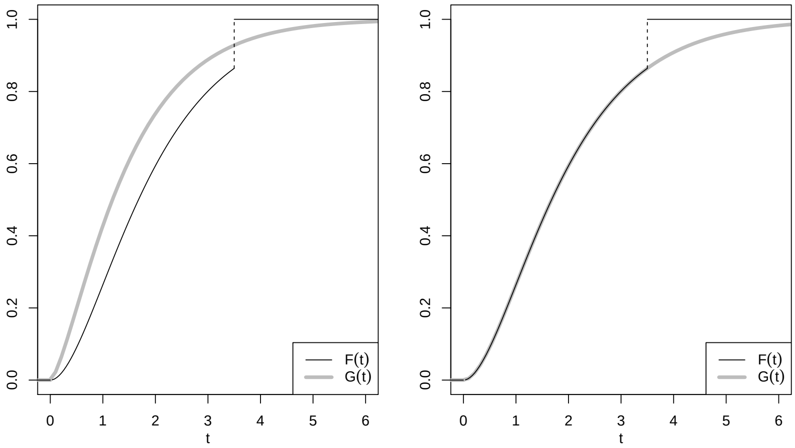

Fig. 3 Exemplary constellations of cdf's F and G as excluded for property (S3) in Theorem 1 by assumption. In both panels we have $F \leq_c G$, in the right one holding in a strict sense.

assumption for (S3) in Theorem 1 does not hold, in Figure 3. Note that for values smaller than $\sup F(D_F)$, G is already more skewed to the right than F. While F then has a point mass and jumps straight to the value 1, that same probability mass can be stretched out to the right in G. Heuristically, it is clear that this merely increases the ecxcess in right-skewness that G already had over F. (This is the right-side analogue to the constellation in Example 1 and Figure 1, in which a jump at the left endpoint of D_F makes F more right-skewed relative to G.) Hence, we strongly conjecture that (S3) is satisfied by $\gamma_M(X)$ without any assumptions.

3.2 A quantile definition of skewness

A quantile based skewness measure was first proposed by Bowley (1901), and later generalized by Hinkley (1975). Let X be a random variable with cdf F and corresponding quantile function F^{-1} such that $F^{-1}(1-\alpha) - F^{-1}(\alpha) > 0$ holds for any value of α in $(0, 1/2)$. Using the short-hand notation $q_p = F^{-1}(p)$, the quantile skewness is defined as

$$\gamma_Q^{(\alpha)}(X) = \frac{q_{1-\alpha} + q_\alpha - 2q_{1/2}}{q_{1-\alpha} - q_\alpha}.$$

Bowley's skewness coefficient is obtained for $\alpha = 1/4$. The parameter α determines whether the information about skewness is gathered near to the center of the distribution or rather in the tails.

It has been shown by Groeneveld and Meeden (1984, p. 394) that, except for γ_M, all three skewness measures considered in this paper are normalized to the absolute value of 1. However, it has not been examined whether these inequalities are sharp.

Remark 2 For the quantile skewness, $-1 \leq \gamma_Q^{(\alpha)} \leq 1$ with both inequalities being sharp: the inequality $-1 \leq \gamma_Q^{(\alpha)}(X) \leq 1$ is equivalent to $q_{1-\alpha} \geq q_{1/2} \geq q_\alpha$, which is true since a quantile function is non-decreasing. By replacing the inequalities by equalities, it is obvious that the lower bound is attained for $q_{1/2} = q_{1-\alpha}$, and the upper bound is attained for $q_{1/2} = q_\alpha$.

As stated in the introduction, properties (S1) - (S3) have been proved for the quantile skewness in Groeneveld and Meeden (1984) under the assumption that the random variables under consideration have a differentiable and positive density function. In the following theorem we relax this assumption, using only a weak condition for (S2).

Theorem 2 *Let $\alpha \in (0, 1/2)$, and let $X \sim F$, where F is an arbitrary cdf for which $\gamma_Q^{(\alpha)}(X)$ is well defined. Then the quantile skewness $\gamma_Q^{(\alpha)}(X)$ satisfies (S1) and (S3). Property (S2) holds if, for $p \in \{\alpha, 1/2, 1-\alpha\}$, $F(t) > p$ for all $t > q_p$.*

Proof (S1) *Let $c > 0$, $d \in \mathbb{R}$ and F_{cX+d} denote the cdf of the random variable $cX + d$ with F^{-1}_{cX+d} denoting the corresponding quantile function. Since, for $p \in (0, 1)$, $F^{-1}_{cX+d}(p) = cq_p + d$, we get $\gamma_Q^{(\alpha)}(cX + d) = \gamma_Q^{(\alpha)}(X)$.*

(S2) *We have $F_{-X}(t) = 1 - F(-t) + \mathbb{P}(X = -t)$ for $t \in \mathbb{R}$, which entails*

$$\begin{aligned} F^{-1}_{-X}(p) &= \inf\{t \in \mathbb{R} : 1 - F(-t) + \mathbb{P}(X = -t) \geq p\} \\ &= \inf\{t \in \mathbb{R} : 1 - F(-t) \geq p\} \\ &= -\sup\{t \in \mathbb{R} : F(t) \leq 1 - p\} \end{aligned} \tag{1}$$

for $p \in (0, 1)$. Here, the second equality holds since there are at most countably many points at which the distribution of X has a point mass. Hence, omitting these points does not change the infimum.
The additional assumption for (S2) yields that for $p \in \{\alpha, 1/2, 1-\alpha\}$ there is at most one $t \in \mathbb{R}$ with $F(t) = p$. Therefore,

$$F^{-1}(p) = \sup\{t \in \mathbb{R} : F(t) \leq p\}. \tag{2}$$

Combining equations (1) and (2) yields $F^{-1}_{-X}(p) = -q_{1-p}$ for $p \in \{\alpha, 1/2, 1-\alpha\}$ and, hence,

$$\gamma_Q^{(\alpha)}(-X) = \frac{-q_\alpha - q_{1-\alpha} + 2q_{1/2}}{-q_\alpha + q_{1-\alpha}} = -\gamma_Q^{(\alpha)}(X).$$

(S3) *Let F and G be cdf's with $F \leq_c G$, and let $X \sim F$ and $Y \sim G$. Using (S1), we can assume without loss of generality that both medians are equal to zero. Hence, one has to show the inequality*

$$\frac{F^{-1}(1-\alpha) + F^{-1}(\alpha)}{F^{-1}(1-\alpha) - F^{-1}(\alpha)} \leq \frac{G^{-1}(1-\alpha) + G^{-1}(\alpha)}{G^{-1}(1-\alpha) - G^{-1}(\alpha)}. \tag{3}$$

We first look at two extreme cases, namely $F^{-1}(1-\alpha) = F^{-1}(1/2) = 0$ *and* $F^{-1}(\alpha) = 0$. *In the first case, Remark 2 shows that* $\gamma_Q^{(\alpha)}(X) = -1$ *and, therefore, inequality (3) holds due to the same remark.*
In the second case, we have $\gamma_Q^{(\alpha)} = 1$. *Further, we know that* F *has a jump discontinuity, skipping the value* α *and ending no earlier than the value* $1/2$, *entailing* $[\alpha, 1/2) \subseteq F(D_F)^{\mathsf{c}}$. *Since* $F(\alpha)^{-1} < F(1-\alpha)^{-1}$, *the jump ends before* $1-\alpha$; *in particular, all skipped values are strictly smaller than* $\sup F(D_F)$. *According to Lemma 1a),* G *then skips the same interval* $[\alpha, 1/2)$, *which yields* $G^{-1}(\alpha) = G^{-1}(1/2)$. *Hence,* $\gamma_Q^{(\alpha)}(Y) = 1$, *and inequality (3) is proved for this special case.*
From now on, we can assume that $F^{-1}(\alpha) < 0 < F^{-1}(1-\alpha)$. *Hence, inequality (3) can be written as*

$$\frac{G^{-1}(1-\alpha)}{F^{-1}(1-\alpha)} \geq \frac{G^{-1}(\alpha)}{F^{-1}(\alpha)}. \tag{4}$$

First, we prove this under the additional assumption of $\alpha, 1/2, 1-\alpha \in F(D_F)$. *By equivalence (16) in the proof of Lemma 1c) below, this assumption yields* $G^{-1}(F(F^{-1}(\alpha))) = G^{-1}(\alpha)$ *as well as* $G^{-1}(F(F^{-1}(1-\alpha))) = G^{-1}(1-\alpha)$. *Since* $F^{-1}(\alpha) < F^{-1}(1-\alpha)$, *it is sufficient to show that the function* $\tilde{s}_0(t) = \frac{G^{-1}(F(t))}{t}, t \in D_F \setminus \{0\}$, *is non-decreasing. Based on the convexity of* $G^{-1} \circ F$, *Lemma 2 already yields that the function* $s_0(t) = \frac{G^{-1}(F(t)) - G^{-1}(F(0))}{t-0}, t \in D_F \setminus \{0\}$, *is non-decreasing. Combining* $1/2 \in F(D_F)$ *with (16), we can infer the identity* $G^{-1}(F(0)) = G^{-1}(F(F^{-1}(1/2))) = G^{-1}(1/2) = 0$. *Then, it follows that* $s_0 = \tilde{s}_0$ *and, hence, inequality (4) is valid.*
Now we dismiss the assumption $1/2 \in F(D_F)$, *i.e. we assume that* $1/2$ *is skipped by* F. *Since* $F^{-1}(1/2) < F^{-1}(1-\alpha)$, *the associated jump discontinuity of* F *ends before* $1-\alpha$. *Therefore, there exists* $p \in (1/2, 1-\alpha) \cap F(D_F)$ *satisfying* $F^{-1}(p) = F^{-1}(1/2) = 0$. *As mentioned previously, Lemma 1a) states that all values skipped by* F *are also skipped by* G *since they are all smaller than* p *and therefore smaller than* $\sup F(D_F)$. *Using (16),* $G^{-1}(F(0)) = G^{-1}(F(F^{-1}(p))) = G^{-1}(p) = G^{-1}(1/2) = 0$; *the additional assumption* $1/2 \in F(D_F)$ *was only used before to establish this equality.*
Next, we dismiss the assumption $\alpha \in F(D_F)$, *so let the value* α *be skipped by* F. *Again using (16), and since* $F^{-1}(\alpha) < F^{-1}(1/2)$, *there exists* $p \in (\alpha, 1/2) \cap F(D_F)$ *satisfying* $G^{-1}(F(F^{-1}(\alpha))) = G^{-1}(F(F^{-1}(p))) = G^{-1}(p) = G^{-1}(\alpha)$. *Therefore, the assumption* $\alpha \in F(D_F)$ *is no longer needed.*
At first sight, we can treat the dismissal of the assumption $1-\alpha \in F(D_F)$ *analogously. Assuming that the jump discontinuity within which* F *skips the value* $1-\alpha$ *ends before* 1, *there exists* $p \in (1-\alpha, 1) \cap F(D_F)$ *satisfying*

$$G^{-1}(F(F^{-1}(1-\alpha))) = G^{-1}(F(F^{-1}(p))) = G^{-1}(p) = G^{-1}(1-\alpha). \tag{5}$$

However, if this jump does end at the value 1*, two problems occur. Firstly, we then have* $F^{-1}(1-\alpha) = \sup D_F$ *and thus* $F^{-1}(1-\alpha) \notin D_F$*, since in this case* $D_F = \mathbb{R}\backslash F^{-1}(\{0,1\})$ *is a right-bounded and right-open interval. Hence,* $F^{-1}(1-\alpha)$ *is not contained in the domain of* $\tilde{s}_0$ *and the line of argument used previously is not applicable here.*
To solve this, we consider the same extension of the function $G^{-1} \circ F$ *that we already used in the proof of (S3) for the moment based skewness measure (see Theorem 1). There, we extended the domain of* $G^{-1} \circ F$ *to* $D_F \cup \{\sup D_F\}$ *and assigned* $G^{-1}(F(\sup D_F)) = \sup D_G$ *as function value for the additional point. We also established that this extension preserves the convexity of the function. Thus, inequality (4) can be proved as before using Lemma 2.*
The second problem occurs if $\sup G(D_G) \geq \sup F(D_F) = 1-\alpha$ *holds with* $[1-\alpha, 1) \cap D_G \neq \emptyset$*. In this case the last identity in (5) is not true; instead we have* $G^{-1}(p) = G^{-1}(1) = \sup D_G > G^{-1}(\sup F(D_F)) = G^{-1}(1-\alpha)$*. In order to circumvent this problem, we first define a new cdf* $\tilde{G} : \mathbb{R} \to [0,1]$ *by*

$$\tilde{G}(t) = \begin{cases} G(t) & , \text{if } t \notin [G^{-1}(\sup F(D_F)), \sup D_G), \\ 1 & , \text{if } t \in [G^{-1}(\sup F(D_F)), \sup D_G). \end{cases}$$

Now, considering F *and* $\tilde{G}$*, we have* $\sup \tilde{G}(D_{\tilde{G}}) = \sup F(D_F) = 1-\alpha$ *with* $[1-\alpha, 1) \cap D_{\tilde{G}} = \emptyset$*. Consequentially,* $\tilde{G}(p) = \tilde{G}(1) = \sup D_{\tilde{G}} = \tilde{G}^{-1}(1-\alpha)$*. However, to actually make use of this, we first have to show that the composition* $\tilde{G}^{-1} \circ F$ *is also convex on* $D_F \cup \{\sup D_F\}$*. Because of* $F(D_F) \subseteq (0, 1-\alpha)$ *and because* $\tilde{G}$ *is equal to* G *as long as their values are smaller than* $1-\alpha$*, we have* $\tilde{G}^{-1}(F(t)) = G^{-1}(F(t))$ *for all* $t \in D_F$ *and thus,* $\tilde{G}^{-1} \circ F$ *is convex on* D_F*. Furthermore, we have* $\tilde{G}^{-1}(F(\sup D_F)) = \tilde{G}^{-1}(1) = G^{-1}(\sup F(D_F))$*, so that (since* $G^{-1}(\sup F(D_F)) = \sup D_{\tilde{G}}$*) the function* $\tilde{G}^{-1} \circ F$ *is convex on its entire domain due to the previously stated result of Niculescu and Persson (2018).*
Now let $\tilde{Y} \sim \tilde{G}$*. Since* G *and* $\tilde{G}$ *coincide for all values smaller than* $1-\alpha$*, we have* $\tilde{G}^{-1}(\alpha) = G^{-1}(\alpha)$ *as well as* $\tilde{G}^{-1}(1/2) = G^{-1}(1/2)$*. Additionally,* $\tilde{G}^{-1}(1-\alpha) = G^{-1}(\sup F(D_F)) = G^{-1}(1-\alpha)$*. Hence,*$\gamma_Q^{(\alpha)}(Y) = \gamma_Q^{(\alpha)}(\tilde{Y})$*.*
Because we already established that $\tilde{G}^{-1} \circ F$ *is convex, we can now apply the prior line of reasoning to* F *and* $\tilde{G}$ *in order to infer* $\gamma_Q^{(\alpha)}(X) \leq \gamma_Q^{(\alpha)}(\tilde{Y}) = \gamma_Q^{(\alpha)}(Y)$*. This concludes the proof.* □

To illustrate that the additional assumption is indeed necessary for (S2), we give a short counterexample: let $Z \sim \text{Bin}(1, 1/2)$ and define $X = Z - 1/2$. Then, $q_\alpha = q_{1/2} = -1/2$, and $q_{1-\alpha} = 1/2$ for $\alpha \in (0, 1/2)$. Hence, $\gamma_Q^{(\alpha)}(X) = -1$. The symmetry of X yields $\gamma_Q^{(\alpha)}(-X) = \gamma_Q^{(\alpha)}(X) \neq -\gamma_Q^{(\alpha)}(X)$, which contradicts (S2).

Note, however, that the additional assumptions for (S2) are dispensable if we use a different definition of quantiles. Apart from the left p-quantile $q_p = \inf\{t \in \mathbb{R} : F(t) \geq p\}$, there are the right quantile $q_p^+ = \sup\{t \in \mathbb{R} : F(t) \leq p\}$ as well as the central quantile

$$m_p = (q_p + q_p^+)/2. \tag{6}$$

Replacing the left quantiles by the central quantiles in the definition of quantile skewness, this new measure satisfies (S2) without any additional assumptions. This is due to the fact that these assumptions ensure that the left and right quantiles coincide. By using the central quantiles, the differences between left and right quantiles just cancel out.

3.3 Integrated quantile skewness

The use of quantile skewness raises the question about the choice of α, since, dependent on α, the results can differ substantially. A possible solution was proposed by Groeneveld and Meeden (1984) introducing the integrated quantile skewness. To obtain this skewness measure, we integrate both the numerator and the denominator of the quantile skewness with respect to the parameter α over all possible values. Consequently, for a random variable with finite mean which is not almost surely constant, we define the integrated quantile skewness by

$$\gamma_{IQ}(X) = \frac{\int_0^{1/2}(q_{1-\alpha} + q_\alpha - 2q_{1/2})\mathrm{d}\alpha}{\int_0^{1/2}(q_{1-\alpha} - q_\alpha)\mathrm{d}\alpha} = \frac{\int_0^1 q_\alpha \mathrm{d}\alpha - q_{1/2}}{\int_0^1 |q_\alpha - q_{1/2}|\mathrm{d}\alpha} = \frac{\mathbb{E}X - q_{1/2}}{\mathbb{E}|X - q_{1/2}|}. \tag{7}$$

First, we take a look at the normalization of the integrated quantile skewness.

Remark 3 Integrated quantile skewness is normalized, $-1 \leq \gamma_{IQ} \leq 1$, and both inequalities are sharp.

Proof The normalization of γ_{IQ} follows from the triangle inequality for integrals. To show that the inequalities are sharp, note that $\gamma_{IQ} = 1$ is equivalent to

$$\int_{-\infty}^{\infty}(t - q_{1/2})\mathrm{d}F(t) = \int_{-\infty}^{q_{1/2}}(q_{1/2} - t)\mathrm{d}F(t) + \int_{q_{1/2}}^{\infty}(t - q_{1/2})\mathrm{d}F(t). \tag{8}$$

In turn, this is equivalent to

$$\int_{-\infty}^{q_{1/2}}(t - q_{1/2})\mathrm{d}F(t) = 0 \Leftrightarrow \exists t_0 \in \mathbb{R} : \mathbb{P}(X = t_0) \geq 1/2 \text{ and } \mathbb{P}(X < t_0) = 0.$$

In this case, t_0 equals the median. Similarly, $\gamma_{IQ} = -1$ is equivalent to

$$\begin{aligned}&\int_{q_{1/2}}^{\infty}(t-q_{1/2})\mathrm{d}F(t)=0\\ \Leftrightarrow\ &\exists t_0\in\mathbb{R}:[\mathbb{P}(X=t_0)>1/2 \text{ and } \mathbb{P}(X>t_0)=0]\\ &\text{or }\ [\mathbb{P}(X=t_0)=1/2, \mathbb{P}(X>t_0)=0 \text{ and } \forall\varepsilon>0:\mathbb{P}(X\in(t_0-\varepsilon,t_0))>0]\,.\end{aligned}$$

Thus, equality to -1 occurs under essentially analogous conditions. The reason why both cases of equality are not fully symmetric is that we use the left median and not the central median. □

Now, we examine whether γ_{IQ} satisfies properties (S1) - (S3). The stated assumptions are considerably weaker as in Groeneveld and Meeden (1984).

Theorem 3 *Let $X \sim F$, where F denotes a cdf for which $\gamma_{IQ}(F)$ is well defined. Then, the integrated quantile skewness satisfies (S1) and (S3). Property (S2) is satisfied if $F(t) > 1/2$ for $t > q_{1/2}$.*

Proof (S1) follows by direct computation. An analogous reasoning as in Theorem 2 yields (S2).

It remains to show the validity of (S3). To this end, let F and G be cdf's satisfying $F \leq_c G$, and let $X \sim F$ and $Y \sim G$. Due to the validity of (S1), we can (and do) assume that the medians of F and G are zero. Then, we have to prove the inequality

$$\begin{aligned}&\frac{\int_0^{1/2}F^{-1}(1-\alpha)\mathrm{d}\alpha+\int_0^{1/2}F^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{1/2}F^{-1}(1-\alpha)\mathrm{d}\alpha-\int_0^{1/2}F^{-1}(\alpha)\mathrm{d}\alpha}\\ &\leq\frac{\int_0^{1/2}G^{-1}(1-\alpha)\mathrm{d}\alpha+\int_0^{1/2}G^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{1/2}G^{-1}(1-\alpha)\mathrm{d}\alpha-\int_0^{1/2}G^{-1}(\alpha)\mathrm{d}\alpha}.\end{aligned} \tag{9}$$

Preliminarily, we look at some extreme cases, starting with $\int_0^{1/2}F^{-1}(1-\alpha)\mathrm{d}\alpha = 0$. It immediately follows that $\gamma_{IQ}(X) = -1$, and inequality (9) is fulfilled due to Remark 3. Similarly, the case $\int_0^{1/2}G^{-1}(\alpha)\mathrm{d}\alpha = 0$ yields $\gamma_{IQ}(Y) = 1$, and (9) holds again.

Next, we consider the case $\int_0^{1/2}F^{-1}(\alpha)\mathrm{d}\alpha = 0$, which delivers $\gamma_{IQ}(X) = 1$. This case occurs only if $F^{-1}(\alpha) = 0 = F^{-1}(1/2)$ for all $\alpha \in (0, 1/2)$, since any quantile function is non-decreasing. Thus, F exhibits a jump discontinuity beginning at value 0 and ending no earlier than at 1/2. Since now $\sup F(D_F) > 1/2$ necessarily holds, all values skipped by F are also skipped by G due to Lemma 1a). Thus, G also has a jump discontinuity at least as high as the one in F, and we get $\int_0^{1/2}G^{-1}(\alpha)\mathrm{d}\alpha = 0$. Hence, $\gamma_{IQ}(Y) = 1$, and (9) is again satisfied.

Having considered these cases we can assume in the remainder of the proof that $\int_0^{1/2}F^{-1}(\alpha)\mathrm{d}\alpha < 0 < \int_0^{1/2}F^{-1}(1-\alpha)\mathrm{d}\alpha$ as well as $\int_0^{1/2}G^{-1}(\alpha)\mathrm{d}\alpha < 0$

holds, implying $\inf G(D_G) < 1/2$. Therefore, inequality (9) is equivalent to

$$\frac{\int_0^{1/2} G^{-1}(1-\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(1-\alpha)\mathrm{d}\alpha} \geq \frac{\int_0^{1/2} G^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(\alpha)\mathrm{d}\alpha}. \tag{10}$$

First we introduce two additional assumptions, which we dismiss again later on. Our first assumption is that both F and G are strictly increasing on D_F and D_G, respectively. The second assumption is that F^{-1} is strictly increasing in a neighbourhood of $1/2$. The latter assumption yields $F^{-1}(p) \neq F^{-1}(1/2) = 0$ for all $p \in (0,1/2) \cup (1/2,1)$ while the first yields that both F^{-1} and G^{-1} are continuous on $(0,1)$. In particular, for $H \in \{F,G\}$ the mappings $p \mapsto \int_0^p H^{-1}(\alpha)\mathrm{d}\alpha$ and $p \mapsto \int_0^p H^{-1}(1-\alpha)\mathrm{d}\alpha$ are continuous on $[0,1/2]$ and differentiable on $(0,1/2)$, with derivatives $p \mapsto H^{-1}(p)$ and $p \mapsto H^{-1}(1-p)$, respectively. Thus, we can apply Cauchy's generalized mean value theorem to both mappings and there exist $\alpha_0, \alpha_1 \in (0,1/2)$ such that

$$\frac{\int_0^{1/2} G^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(\alpha)\mathrm{d}\alpha} = \frac{G^{-1}(\alpha_0)}{F^{-1}(\alpha_0)}, \tag{11}$$

$$\frac{\int_0^{1/2} G^{-1}(1-\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(1-\alpha)\mathrm{d}\alpha} = \frac{G^{-1}(1-\alpha_1)}{F^{-1}(1-\alpha_1)}. \tag{12}$$

Examining the proof of (S3) in Theorem 2 yields $\frac{G^{-1}(\alpha)}{F^{-1}(\alpha)} \leq \frac{G^{-1}(\tilde{\alpha})}{F^{-1}(\tilde{\alpha})}$ for all $\alpha \in (0,1/2), \tilde{\alpha} \in (1/2,1)$ with $F^{-1}(\alpha) < 0 < F^{-1}(\tilde{\alpha})$. Since $\alpha_0 \in (0,1/2)$ and $1-\alpha_1 \in (1/2,1)$, inequality (10) follows.

Now we dispense with the second assumption, and assume that F^{-1} is constant on some non-degenerate interval I_0 satisfying $1/2 \in I_0$. Considering $\int_0^{1/2} F^{-1}(\alpha)\mathrm{d}\alpha < 0 < \int_0^{1/2} F^{-1}(1-\alpha)\mathrm{d}\alpha$, we can then infer $0 < \inf I_0 < \sup I_0 < 1$.

First we consider the case $\inf I_0 < 1/2$: then $F^{-1}(p) = F^{-1}(1/2) = 0 \ \forall p \in I_0 \cap (0,1/2]$, which means that the assumption $F^{-1}(p) \neq 0 \ \forall p \in (0,1/2)$ of Cauchy's generalized mean value theorem is no longer satisfied. We also know that F has a jump discontinuity, skipping the values in I_0 (possibly except its endpoints). Assuming $\sup I_0 < \sup F(D_F)$, Lemma 1a) yields that any value skipped by F is also skipped by G. Thus, $G^{-1}(p) = 0 \ \forall p \in I_0 \cap (0,1/2]$. It follows that

$$\frac{\int_0^{1/2} G^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(\alpha)\mathrm{d}\alpha} = \frac{\int_0^{\inf I_0} G^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{\inf I_0} F^{-1}(\alpha)\mathrm{d}\alpha}.$$

We can now apply Cauchy's generalized mean value theorem to $[0, \inf I_0]$ instead of $[0,1/2]$, which yields (11).

If $\sup I_0 < \sup F(D_F)$ does not hold, we have $\sup I_0 = \sup F(D_F)$, which means that F skips the values in I_0, is then constant at the value $\sup I_0$ and finally skips from there to the value 1. However, due to Lemma 1b) this implies $1/2 > \inf G(D_G) \geq \sup F(D_F) \geq 1/2$, a contradiction. Thus, the assumption $\sup I_0 < \sup F(D_F)$ was correct.

In the case $\sup I_0 > 1/2$ (which can occur simultaneously to $\inf I_0 < 1/2$) we proceed analogously. We have

$$\frac{\int_0^{1/2} G^{-1}(1-\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(1-\alpha)\mathrm{d}\alpha} = \frac{\int_0^{1-\sup I_0} G^{-1}(1-\alpha)\mathrm{d}\alpha}{\int_0^{1-\sup I_0} F^{-1}(1-\alpha)\mathrm{d}\alpha},$$

and applying Cauchy's generalized mean value theorem to $[0, 1-\sup I_0]$ yields (12).

Finally, we dismiss the assumption that F and G are strictly increasing on D_F and D_G. Lemma 1b) states that there are only two situations, in which $F \leq_c G$ holds but not both cdf's are strictly increasing, denoted by (i) and (ii).

First we consider situation (i), so let F be constant on some non-degenerate interval $I_F \subseteq D_F$ with value $p_F \in (0, \inf G(D_G)]$. Then, F^{-1} has a jump discontinuity at the point p_F, skipping the values in I_F. Suppose now that G^{-1} also has a jump discontinuity at the same point. Then $p_F \in G(D_G)$, and Lemma 1b) yields $p_F = \sup F(D_F) = \inf G(D_G) < 1/2$, which contradicts $\int_0^{1/2} F^{-1}(1-\alpha)\mathrm{d}\alpha > 0$. Therefore, G^{-1} must be continuous at p_F. Examining the proof of Cauchy's generalized mean value theorem as well as the proof of Rolle's underlying theorem then yields that either there exists some $\alpha_0 \in (0, 1/2)$ as in equation (11), or we have

$$\frac{G^{-1}(p_F)}{\lim_{p\searrow p_F} F^{-1}(p)} = \frac{\lim_{p\searrow p_F} G^{-1}(p)}{\lim_{p\searrow p_F} F^{-1}(p)} < \frac{\int_0^{1/2} G^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(\alpha)\mathrm{d}\alpha} < \frac{G^{-1}(p_F)}{F^{-1}(p_F)}.$$

Since $p_F \in (0, 1/2)$, inequality (10) can also be shown in the present situation as before, taking (12) and the remark below into account.

Finally, we consider situation (ii), and assume that G takes the constant value $p_G = \sup F(D_F)$ on a non-degenerate interval $I_G \subseteq D_G$. It follows from $\int_0^{1/2} F^{-1}(1-\alpha)\mathrm{d}\alpha > 0$ that $p_G = \sup F(D_F) > 1/2$. Now, the line of argument is similar to situation (i). G^{-1} has a jump discontinuity at p_G, while F^{-1} is continuous at that point, since otherwise we could infer $p_G = \inf G(D_G) = \sup F(D_F) > 1/2$, which contradicts $\inf G(D_G) < 1/2$. Using the proof of Cauchy's generalized mean value theorem again yields that there either exists $\alpha_1 \in (0, 1/2)$ as in equation (12) or

$$\frac{G^{-1}(p_G)}{F^{-1}(p_G)} < \frac{\int_0^{1/2} G^{-1}(\alpha)\mathrm{d}\alpha}{\int_0^{1/2} F^{-1}(\alpha)\mathrm{d}\alpha} < \frac{\lim_{p\searrow p_G} G^{-1}(p)}{\lim_{p\searrow p_G} F^{-1}(p)} = \frac{\lim_{p\searrow p_G} G^{-1}(p)}{F^{-1}(p_G)}.$$

Since $p_G \in (1/2, 1)$, inequality (10) follows with (11) and the remark below.

Thus, the additional assumptions have been dismissed, which concludes the proof. □

The same example and reasoning as at the end of Section 3.2 show that the additional assumption for (S2) is indeed necessary. Again, this is due to the use of the left median instead of the central median.

3.4 Pearson's skewness measure

Pearson's skewness measure can be seen as a modification of the integrated quantile skewness. Pearson (1895) proposed the difference between mean and mode, divided by the standard deviation, as a measure of skewness (see also Yule (1922)). However, this is not easy to use since estimation of modal values is difficult. This problem was circumvented by making use of the fact that, for certain classes of distributions, the difference between mean and mode is approximately equal to three times the difference between mean and median (Yule 1922). The numerator of the skewness measure was interchanged accordingly, yielding up to a factor of three Pearson's skewness measure below.

Arnold and Groeneveld (1995) have shown that the original measure of Pearson does not satisfy (S3). Due to its limited applicability, we don't consider it in this work. Instead, we have a look at its modification

$$\gamma_P(X) = \frac{\mathbb{E}X - q_{1/2}}{\sigma}.$$

Groeneveld and Meeden (1984, p. 394) proved that γ_P is normalized; the next remark shows that the bounds are sharp.

Remark 4 For Pearson's skewness measure, $-1 < \gamma_P \leq 1$, and these bounds can not be improved. Since the integrated quantile skewness is normalized, it is sufficient to show

$$\sigma \geq \mathbb{E}|X - q_{1/2}| \tag{13}$$

to infer $|\gamma_P| \leq |\gamma_{IQ}| \leq 1$. First, we obtain by the Cauchy-Schwarz inequality,

$$\begin{aligned} \sigma^2 &= \int_{\mathbb{R}} |t - \mathbb{E}X|^2 \, \mathrm{d}F(t) \int_{\mathbb{R}} |1|^2 \, \mathrm{d}F(t) \\ &\geq \left(\int_{\mathbb{R}} |t - \mathbb{E}X| \, \mathrm{d}F(t) \right)^2 = \left(\mathbb{E}|X - \mathbb{E}X|\right)^2 . \end{aligned} \tag{14}$$

Since $q_{1/2} = \operatorname{argmin}_{t \in \mathbb{R}} \mathbb{E}|X - t|$, $\mathbb{E}|X - q_{1/2}| \leq \mathbb{E}|X - \mathbb{E}X]$, which yields $|\gamma_P| \leq 1$.

Equality in the lower or upper bound implies that equality occurs in (14). This happens in case of (almost sure) linear dependence, i.e. if

$$\begin{aligned} &\exists c \in \mathbb{R} \; \forall t \in \operatorname{supp}(\mathbb{P}_X) : |t - \mathbb{E}X| = c \\ \Leftrightarrow \; &\forall t, s \in \operatorname{supp}(\mathbb{P}_X) : |t - \mathbb{E}X| = |s - \mathbb{E}X|, \end{aligned}$$

where $\operatorname{supp}(\mathbb{P}_X) \subseteq \mathbb{R}$ denotes the support of the probability measure $\mathbb{P}_X$ induced by X. Since at most two points on $\mathbb{R}$ can have the same distance to a third point, $|\operatorname{supp}(\mathbb{P}_X)| \leq 2$. Moreover, $|\operatorname{supp}(\mathbb{P}_X)| > 1$ since $\gamma_P(X)$ is defined only for non-degenerate random variables. Hence, in the case of equality the support of $\mathbb{P}_X$ contains exactly two elements.

Since γ_P satisfies (S1) (see Theorem 4 below), we can transform any 2-point distribution into a Bernoulli distribution without changing the value of γ_P. Hence, we can assume that $X \sim \mathrm{Bin}(1,p)$ for some $p \in (0,1)$, yielding

$$\gamma_P(X) = \frac{p - \mathbb{1}_{\{p>1/2\}}}{\sqrt{p(1-p)}} = \begin{cases} \sqrt{\frac{p}{1-p}} & \text{, if } p \in (0,1/2] \\ -\sqrt{\frac{1-p}{p}} & \text{, if } p \in (1/2,1) \end{cases}.$$

In the first case, $\gamma_P(X)$ is positive and strictly increasing in p, attaining its maximum value of one at $p = 1/2$. Therefore, besides transformations of scale and location, this is the only situation where $\gamma_P = 1$.

In the second case, $\gamma_P(X)$ is negative and strictly increasing in p, attaining its infimum -1 for $p \to 1/2$.

Under varying assumptions and definitions of the median, several results concerning bounds for Pearson's skewness measure have been published, e.g. by Hotelling and Solomons (1932) and Majindar (1962). Specifically, Majindar used the central median and proved that in this case, $-1 < \gamma_P < 1$. The asymmetry in Remark 4 is again due to the usage of the left median.

The following proposition can be shown similarly as properties (S1) - (S3) in Theorem 2.

Proposition 4 *Let $X \sim F$, where F denotes a cdf such that $\gamma_P(X)$ is well defined. Then, Pearson's skewness measure $\gamma_P(X)$ satisfies (S1). Property (S2) is fulfilled if $F(t) > 1/2$ for $t > q_{1/2}$.*

Similarly to the quantile skewness and its integrated version, the additional assumption for (S2) is necessary. Again, the use of the central median defined in (6) instead of the left median would render this assumption unnecessary.

The following example shows that Pearson's skewness measure does not satisfy (S3). In previous work (e.g. (Groeneveld and Meeden 1984)) this has already been pointed out, citing a remark by van Zwet (1964, pp. 16-17), where a counterexample using discrete distributions is given. Additionally, they illustrated this fact graphically using gamma and beta distributions. The Weibull distribution also gives rise to counterexamples to (S3), as the following explicit example shows.

Example 4 Let X be Weibull-distributed with scale parameter 1 and shape parameter k_1 ($X \sim \mathrm{Wei}(1,k_1)$ for short) with corresponding cdf F and $Y \sim \mathrm{Wei}(1,k_2)$ with corresponding cdf G, where $k_1 > k_2 > 0$. Cdf of X and quantile function of Y are $F(x) = 1 - \exp\left(-x^{k_1}\right), x > 0$, and $G^{-1}(p) = (-\ln(1-p))^{1/k_2}, 0 < p < 1$. Hence, for $x > 0$,

$$G^{-1}(F(x)) = x^{k_1/k_2},$$

and we obtain for $x > 0$

$$\left(G^{-1}(F(x))\right)'' = k_1/k_2\,(k_1/k_2 - 1)\,x^{k_1/k_2 - 2} > 0.$$

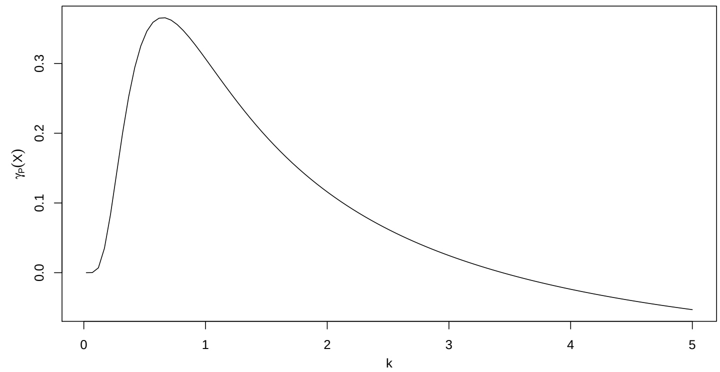

Fig. 4 Plot of Pearson's skewness measure $\gamma_P(X)$ for $X \sim \text{Wei}(1, k)$ against the shape parameter k

Hence, $G^{-1} \circ F$ is strictly convex on $D_F = (0, \infty)$.

Choosing the specific parameter values $k_1 = 1, k_2 = 1/3$, we obtain

$$\mathbb{E}X = \Gamma(1 + 1/k_1) = 1,\ \mathbb{E}Y = 6,\ F^{-1}(1/2) = \ln 2,\ G^{-1}(1/2) = (\ln 2)^3.$$

Further,

$$\sigma_X^2 = \Gamma(1 + 2/k_1) - (\Gamma(1 + 1/k_1))^2 = 1,\ \sigma_Y^2 = (6\sqrt{19})^2,$$

and thereby

$$\gamma_P(X) = 1 - \ln 2 \approx 0.307 \qquad > \gamma_P(Y) = \left(1 - (\ln 2)^3/6\right)/\sqrt{19} \approx 0.217.$$

Obviously, this contradicts (S3).

More generally, let $X \sim \text{Wei}(\lambda, k)$ for some $\lambda, k > 0$. Since the cdf's are ordered increasingly with decreasing shape parameter, any pair of Weibull distributions with the same scale parameter contradicts (S3) whenever $\gamma_P(X)$ is not decreasing for increasing k. A plot of $\gamma_P(X)$ for $X \sim \text{Wei}(1, k)$ against the shape parameter is shown in Fig 4.

A Proof of Lemma 1

Proof a) Cdf's are non-decreasing and they either attain the boundary values of their codomain $[0, 1]$ or approach them as limiting values. Hence, $F(D_F)$ is not equal to $(0, 1)$ if and only if F has discontinuities. The same holds for G and $G(D_G)$. Since a cdf contains at most countably many discontinuities, we can infer that $F(D_F)$ and $G(D_G)$ are countable unions of disjoint intervals.

Now we assume that $G(D_G)$ is not a subset of $F(D_F) \cup [\sup F(D_F), 1)$. It follows that the set difference $G(D_G) \setminus (F(D_F) \cup [\sup F(D_F), 1)) \neq \emptyset$ is also a countable union of disjoint intervals.

For the moment, we assume additionally that at least one interval I_0 in that union is non-degenerate. We then have $I_0 \subseteq G(D_G) \setminus F(D_F)$, so G passes through the values in I_0 continuously while F discontinuously skips these values. By definition of the quantile

function and continuity of G on I_0, G^{-1} is strictly increasing on I_0. We denote the left and right endpoints of I_0 as p_0 and p_1, respectively. We also denote the point at which F skips I_0 with t_0. Hence, $F(t_0) \geq p_1$, $\lim_{s\nearrow t_0} F(s) \leq p_0$, and G^{-1} is strictly increasing on $[p_0, p_1]$ with $p_0 < p_1$.
We first consider two boundary cases. First, $t_0 = \inf D_F$ implies $p_1 \leq \inf F(D_F)$, which yields $\inf G(D_G) \leq p_0 < p_1 \leq \inf F(D_F)$, thus contradicting $F \leq_c G$. Second, $t_0 = \sup D_F$ implies $p_0 \geq \sup F(D_F)$, which yields $I_0 \cap [\sup F(D_F), 1) \neq \emptyset$, thereby contradicting the assumption on I_0. The only case left to consider is $t_0 \in D_F^\circ$. Then,

$$\begin{aligned}\lim_{s\nearrow t_0} G^{-1}(F(s)) &\leq G^{-1}\big(\lim_{s\nearrow t_0} F(s)\big) \leq G^{-1}(p_0)\\ &< G^{-1}(p_1) \leq G^{-1}(F(t_0)).\end{aligned}$$

This inequality entails that $G^{-1} \circ F$ has a discontinuity at the point t_0. However, by assumption, $G^{-1} \circ F$ is convex on D_F, and therefore continuous on the interior of D_F, in which t_0 was assumed to lie. This contradicts $F \leq_c G$.
The only case left to examine is that $G(D_G) \setminus (F(D_F) \cup [\sup F(D_F), 1))$ is a countable union of singletons. Therefore, assume that there exists a $p_0 \in G(D_G) \setminus (F(D_F) \cup (\sup F(D_F), 1])$. Since $\{p_0\}$ is a singleton, there exists an $\varepsilon > 0$ such that the value p_0 is skipped by F but not by G while the sets $(p_0 - \varepsilon, p_0)$ and $(p_0, p_0 + \varepsilon)$ are both either skipped or passed through continuously by both functions. Since continuity of F (and G) on $(p_0, p_0 + \varepsilon)$ would immediately contradict the right-continuity of F, both functions skip this interval (for some $\varepsilon > 0$). This entails that G attains the lower endpoint of its jump discontinuity, namely p_0. This is true only if there exists a non-degenerate interval $[t_0, t_1] \subseteq D_G$ such that $G(t) = p_0$ for all $t \in [t_0, t_1]$. Therefore, G^{-1} has a jump at the point p_0, with $G^{-1}(p_0) = t_0$ and $\lim_{r\searrow p_0} G^{-1}(r) = t_1$. So while G jumps over the exact interval (p_0, p_1) (for some $p_1 > p_0$) at the point t_1; at some point $t_F \in D_F$, F at least jumps over the interval $[p_0, p_1)$ with the lower endpoint possibly being even smaller than p_0.
The two boundary cases $t_F = \inf D_F$ and $t_F = \sup D_F$ can be dismissed similarly to before, leaving only the case $t_F \in D_F^\circ$, in which we obtain

$$\begin{aligned}\lim_{s\nearrow t_F} G^{-1}(F(s)) &\leq G^{-1}\left(\lim_{s\nearrow t_F} F(s)\right) \leq G^{-1}(p_0) = t_0\\ &< t_G = \lim_{r\searrow p_0} G^{-1}(r) = G^{-1}(p_1) = G^{-1}(F(t_F)),\end{aligned}$$

so $G^{-1} \circ F$ is discontinuous at t_F. Again, this contradicts $F \leq_c G$ since t_F lies in the interior of D_F, thus concluding the proof.

b) As first case we assume that F is not strictly increasing on D_F without exception (i) occurring. Consequentially, there exists a non-degenerate interval $I_F \subseteq D_F$ and a $p_F \in (\inf G(D_G), 1)$ such that $F(t) = p_F$ for all $t \in I_F$. Considering $F \leq_c G$, this entails $\inf F(D_F) \leq \inf G(D_G) < p_F$, specifically yielding that F cannot be constant on the entirety of D_F, i.e. $I_F \subsetneq D_F$. There then exist $t_F^{(F)} \in D_F \setminus I_F$ and $t_F^{(G)} \in D_G$ such that $F(t_F^{(F)}) \leq G(t_F^{(G)}) < p_F$. Therefore, $G^{-1}(F(t_F^{(F)})) \leq t_F^{(G)} < G^{-1}(p_F)$. Now, let $t_F^{(0)} \in I_F^\circ$ and choose $\lambda \in (0,1)$ such that $\lambda t_F^{(F)} + (1-\lambda)t_F^{(0)} \in I_F$. Then,

$$\begin{aligned}&\lambda G^{-1}(F(t_F^{(F)})) + (1-\lambda)G^{-1}(F(t_F^{(0)}))\\ &= \lambda G^{-1}(F(t_F^{(F)})) + (1-\lambda)G^{-1}(p_F)\\ &< G^{-1}(p_F) = G^{-1}(F(\lambda t_F^{(F)} + (1-\lambda)t_F^{(0)})),\end{aligned}$$

which is a contradiction to the convexity of $G^{-1} \circ F$.
As second case we assume that G is not strictly increasing on D_G without exception (ii) occurring. Therefore, there exists a non-degenerate interval within D_G on which G constantly takes a value $p_G \in (0, \sup F(D_F))$. Conversely, G^{-1} has a jump discontinuity at p_G, so by defining $t_G^{(G)} = G^{-1}(p_G) \in D_G$, we obtain $G^{-1}(p) > t_G^{(G)}$ for all $p > p_G$. Furthermore, since G attains the value p_G, we know from part a) that F also attains this

value; hence, $t_G^{(F)} = \sup\{t \in \mathbb{R} : F(t) = p_G\} \in D_G^{\circ}$ is a real number. Now we distinguish two subcases. First, we assume $F(t_G^{(F)}) = p_G$. Then,

$$\lim_{s \searrow t_G^{(F)}} G^{-1}(F(s)) > t_G^{(G)} = G^{-1}(p_G) = G^{-1}(F(t_G^{(F)})).$$

Second, we assume $F(t_G^{(F)}) > p_G$ (since $F(t_G^{(F)}) < p_G$ can be excluded due to F being non-decreasing). This can only occur at discontinuities of F, so only if F has a jump up to a value $p_G^{>} > p_G$ at $t_G^{(F)}$. Then, $t_G^{(F)}$ is the smallest number satisfying $F(t_G^{(F)}) = p_G^{>}$. Furthermore, it follows from $p_G < \sup F(D_F)$ that the jump of G has to end before 1, so $p_G^{>} < 1$. Hence,

$$\lim_{s \nearrow t_G^{(F)}} G^{-1}(F(s)) = G^{-1}(p_G) = t_G^{(G)} < G^{-1}(p_G^{>}) = G^{-1}(F(t_G^{(F)})).$$

In both subcases we obtain that $G^{-1} \circ F$ has a jump discontinuity at the point $t_G^{(F)} \in D_F^{\circ}$. Again, this contradicts the convexity of $G^{-1} \circ F$.

c) For a cdf H and $t \in D_H = \mathbb{R} \setminus H^{-1}(\{0,1\})$, we have (Shorak and Wellner 1986, p.6)

$$H^{-1}(H(t)) = t \Leftrightarrow \forall s \in D_H, s < t : H(s) < H(t). \tag{15}$$

Hence, $H^{-1} \circ H$ is equal to the identity at those points, on the left side of which H is strictly increasing. Furthermore, for any $p \in (0,1)$,

$$H(H^{-1}(p)) = p \Leftrightarrow p \in H(\mathbb{R}) \tag{16}$$

(Shorak and Wellner 1986, p.5). Hence, the function $H \circ H^{-1}$ is equal to the identity at those points that are contained in the image of H. Further, for $p \in (0,1)$ and $t \in \mathbb{R}$,

$$p \leq H(t) \Leftrightarrow H^{-1}(p) \leq t. \tag{17}$$

Next, we determine the distribution of $F(X)$. Consider first $p \in F(D_F) = F(\mathbb{R}) \setminus \{0,1\}$; we obtain

$$\begin{aligned} H_{F(X)}(p) &= \mathbb{P}(F(X) \leq p) = \mathbb{P}(F^{-1}(F(X)) \leq F^{-1}(p)) \\ &= \mathbb{P}(X \leq F^{-1}(p)) = \ F(F^{-1}(p)) = p. \end{aligned} \tag{18}$$

Here, the second equality follows by combining (16) and (17), using $p \in F(\mathbb{R})$. The third equality follows from (15), using that $\mathbb{P}$-a.s. X does not take any realization which lie outside of D_F, or on which F is constant. The last inequality holds by (16) and $p \in F(\mathbb{R})$. Assuming $\sup F(D_F) < 1$, we furthermore have

$$\begin{aligned} H_{F(X)}(\sup F(D_F)) &= \mathbb{P}(F(X) \leq \sup F(D_F)) \\ &= \mathbb{P}(X \in D_F) = 1 - \mathbb{P}(X = \sup D_F) \\ &= \sup F(D_F). \end{aligned} \tag{19}$$

Since the same identity is obvious for $\sup F(D_F) = 1$, we overall get $H_{F(X)}(p) = p$ for all $p \in F(D_F) \cup \{\sup F(D_F)\}$.
Regarding the distribution of Y, we now get

$$\begin{aligned} H_Y(t) &= \mathbb{P}(Y \leq t) = \mathbb{P}(G^{-1}(F(X)) \leq t) \\ &= \mathbb{P}(F(X) \leq G(t)) = H_{F(X)}(G(t)) \end{aligned} \tag{20}$$

for any $t \in \mathbb{R}$. For $t \in D_G$, we get by assumption $G(t) \in G(D_G) \subseteq F(D_F) \cup \sup F(D_F)$, which combined with (18), (19) and (20) yields $H_Y(t) = G(t)$. This leaves the case $t \in \mathbb{R} \setminus D_G$, in which we know that $G(t) \in \{0,1\}$ by construction. For $G(t) = 1$, $H_Y(t) = H_{F(X)}(1) = \mathbb{P}(F(X) \leq 1) = 1$, since the codomain of F is given by $[0,1]$. For $G(t) = 0$, we obtain $H_Y(t) = H_{F(X)}(0) = \mathbb{P}(F(X) = 0) = 0$. Hence, $H_Y = G$ holds on the real numbers, which concludes the proof. □

References

Arnold, B. C., Groeneveld, R. A., 1993. Skewness and kurtosis orderings: an introduction. *Stochastic inequalities, IMS Lecture Notes - Monograph Series*, 22, 1993, 17–24.

Arnold, B. C., Groeneveld, R. A., 1995. Measuring Skewness with Respect to the Mode. *The American Statistician*, 49, 34–38.

Artin, E., 1964. *The Gamma Function.* Holt, Rinehart and Winston.

Bowley, A. L., 1901. *Elements of Statistics.* 1st Edition. P. S. King & Son.

Groeneveld, R. A., Meeden, G., 1984. Measuring Skewness and Kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 33, 319–399.

Hinkley, D. V., 1975. On Power Transformations to Symmetry. *Biometrika*, 62, 101–111.

Hotelling, H., Solomons, L. M., 1932. The Limits of a Measure of Skewness. *The Annals of Mathematical Statistics*, 3, 141–142.

MacGillivray, H. L., 1986. Skewness and Asymmetry: Measures and Orderings. *The Annals of Statistics*, 14, 994–1011.

Majindar, K. N., 1962. Improved Bounds on a Measure of Skewness. *The Annals of Mathematical Statistics*, 33, 1192–1194.

Niculescu, C. P., Persson, L.-E., 2018. *Convex Functions and Their Applications. A Contemporary Approach.* 2nd Edition. CMS Books in Mathematics. Springer.

Oja, H., 1981. On Location, Scale, Skewness and Kurtosis of Univariate Distributions. *Scandinavian Journal of Statistics*, 8, 154–168.

Pearson, K., 1895. Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London. A*, 186, 343–414.

Shaked, M., Shanthikumar, G. J., 2006. *Stochastic Orders.* Springer.

Shorak, G.R., Wellner, J.A., 1986. *Empirical Processes with Applications to Statistics.* Wiley.

Yule, G. U., 1922. *An Introduction to the Theory of Statistics.* 6th Edition. Charles Griffin & Co., Ltd.

Van Zwet, W. R., 1964. *Convex Transformations of Random Variables.* Mathematical Centre Tracts, Amsterdam.