Title
A general measure of skewness

Permalink
https://escholarship.org/uc/item/7rs685t8

Journal
Statistics and Probability Letters, 82(8)

ISSN
0167-7152

Authors
Ekström, M
Jammalamadaka, SR

Publication Date
2012-08-01

DOI
10.1016/j.spl.2012.04.011

Peer reviewed
A general measure of skewness

Magnus Ekström\textsuperscript{a}, Sreenivasa Rao Jammalamadaka\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a} Centre of Biostochastics, Swedish University of Agricultural Sciences, Umeå, Sweden
\textsuperscript{b} Department of Statistics and Applied Probability, University of California, Santa Barbara, USA

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 9 December 2011
Received in revised form 11 April 2012
Accepted 11 April 2012
Available online 21 April 2012

Keywords:
Measure of skewness
Bowley’s coefficient of skewness
Test of symmetry
Large sample theory
Asymptotic relative efficiency

\textbf{ABSTRACT}

A very general measure of skewness based on the quantiles is introduced, which includes several well-known measures as special cases. Sample versions of our measure may be used as test statistics for testing the hypothesis of symmetry about an unknown value. We provide large sample theory for such a statistic and discuss the asymptotic relative efficiencies of this against some competing test statistics for symmetry.

© 2012 Elsevier B.V. All rights reserved.

\textbf{1. Introduction}

For symmetric unimodal distributions, the three most common measures of central tendency, the mean $\mu$, the median $M$, and the mode $m$, coincide. For other distributions, the lack of symmetry is expressed through measures of skewness. The idea of measuring skewness of a distribution has roots going back at least to Pearson (1895), who suggested the measure $(\mu - m)/\sigma$, where $\sigma$ is the standard deviation of the distribution. Pearson (1895) also used a quantity denoted by $\hat{\beta}_1$, which is the square of the standardized third moment, and Charlier (1906) proposed $-6^{-1/2} \hat{\beta}_1^{1/2}$ as a measure of skewness. Today, the standardized third moment

$$\gamma_1 = \beta_1^{1/2} = \mu_3/\sigma^3,$$

where $\mu_k$ is the $k$th central moment which is often referred to as “the coefficient of skewness”, although any odd-moment can be used for the purpose. Other early measures include Bowley’s (1901) coefficient of skewness,

$$S_{\text{Bowley}} = \frac{Q_{0.75} - 2M + Q_{0.25}}{Q_{0.75} - Q_{0.25}},$$

where $Q_{0.25}$ and $Q_{0.75}$ represent the 25th and 75th percentiles, and the measure

$$S_{\text{Yule}} = \frac{\mu - M}{\sigma},$$

given in Yule (1911).

A generalization of Bowley’s measure of skewness was presented by David and Johnson (1956),

$$S_{\text{D&J}} = \frac{Q_{1-a} - 2M + Q_a}{Q_{1-a} - Q_a},$$

where $Q_{1-a}$ and $Q_a$ are the $1-a$th and $a$th quantiles, respectively.
for any \( \alpha \) between 0 and 1/2 (see also Hinkley, 1975). Note that Bowley's measure (or coefficient) of skewness is obtained by setting \( \alpha = 0.25 \). The measure obtained by setting \( \alpha = 0.1 \) is often referred to as Kelly's coefficient of skewness. By integrating out \( \alpha \), Groeneveld and Meeden (1984) obtained another measure of skewness viz. 

\[
S_{G\&M} = \frac{\int_0^{1/2} (Q_{1-\alpha} - 2M + Q_n) d\alpha}{\int_0^{1/2} (Q_{1-\alpha} - Q_n) d\alpha} = \frac{\mu - M}{E[X - M]},
\]

sample version of which may be regarded as a robustified version of the sample analogue of \( S_{Yule} \) (Miao et al., 2006). Bowley's as well as the skewness measures \( S_{Yule}, S_{D\&J}, S_{G\&M} \) are bounded by 1 on their absolute values (see, e.g., Groeneveld, 1991), which may be considered an advantage.

In the current paper we introduce a very general measure of skewness based on the quantiles, that contains the measures by Bowley, Kelly, and David and Johnson as special cases. Also a particular sample version of our measure, when using as many quantiles as the sample size, leads to the sample version of Groeneveld and Meeden's measure, and a test statistic proposed by Miao et al. (2006).

Sample versions of our measure may be used as test statistics for testing the hypothesis of symmetry about an unknown value. We provide results for the asymptotic distribution of such test statistics under the null hypothesis of symmetry, as well as under a sequence of converging alternatives of the form \( F(t + n^{-1/2}\gamma(t)) \), where the distribution function \( F \) is symmetric about zero, \( \gamma \) is some smooth function, and \( n \) is the sample size. This allows us to make comparisons of the Asymptotic Relative Efficiencies (AREs) of these competing tests, which we do in Section 3. A numerical example from forestry is given in Section 4.

2. Main results

Suppose \( X_1, \ldots, X_n \) are independently and identically distributed (i.i.d.) real-valued random variables with cumulative distribution function (cdf) \( F \). For any \( 0 < \alpha < 1 \), the population quantile function is given by

\[
F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\},
\]

and will also be denoted as \( Q_\alpha \). The empirical cdf is denoted as \( F_n \), while

\[
F_n^{-1}(\alpha) = \inf\{x : F_n(x) \geq \alpha\}
\]

will denote the empirical quantile function, for which we will also use the notation \( \hat{Q}_\alpha = F_n^{-1}(\alpha) \).

We now quote a result on the so-called Bahadur representation of quantiles:

**Lemma 1 (Ghosh, 1971).** Let \( 0 < \alpha < 1 \). Suppose \( F \) is differentiable at \( Q_\alpha \) with \( F'(Q_\alpha) > 0 \). Then

\[
\hat{Q}_\alpha - Q_\alpha = \frac{\alpha - F_n(Q_\alpha)}{F'(Q_\alpha)} + o_p(n^{-1/2}).
\]

For an integer \( r \geq 2 \), let \( \alpha_1, \ldots, \alpha_{2r-1} \) be a finite sequence of real numbers such that

\[
0 < \alpha_1 < \alpha_2 < \cdots < \alpha_r = 1/2 < \alpha_{r+1} < \cdots < \alpha_{2r-1} < 1
\]

and \( 1 - \alpha_i = \alpha_{2r-i}, \ i = 1, \ldots, r \). Set \( f_j = F'(Q_{\alpha_j}), \ j = 1, \ldots, 2r - 1 \). The representation in (1) and the multivariate central limit theorem imply that \( n^{1/2}(\hat{Q}_{\alpha_1} - Q_{\alpha_1}, \ldots, \hat{Q}_{\alpha_{2r-1}} - Q_{\alpha_{2r-1}}) \) is asymptotically normal with a zero mean vector and covariances \( \text{cov}_{ij} = \alpha_i(1 - \alpha_j)/(f_i f_j) \) for \( i \leq j \), and \( \text{cov}_{ij} = 0 \) for \( i > j \).

We now define our general measure of skewness,

\[
S_r = \frac{\sum_{j=1}^{r-1} ((Q_{\alpha_{2r-j}} - Q_{\alpha_j}) - (Q_{\alpha_r} - Q_{\alpha_j})) c_j}{\sum_{j=1}^{2r-1} d_j Q_{\alpha_j}} = \frac{\sum_{j=1}^{2r-1} c_j Q_{\alpha_j}}{\sum_{j=1}^{2r-1} d_j Q_{\alpha_j}},
\]

where \( c_i = -2(r - 1) \) if \( i = r \) and 1 otherwise, and \( d_i = -1 \) if \( i < r, 0 \) if \( i = r \), and 1 if \( i > r \). We will call the numerator and denominator \( A_r = \sum_{j=1}^{2r-1} c_j Q_{\alpha_j} \) and \( B_r = \sum_{j=1}^{2r-1} d_j Q_{\alpha_j} \) respectively, and denote the sample analogue of \( S_r \) by \( \hat{S}_{r,n} \).

**Theorem 1.** Suppose \( F \) is differentiable at \( Q_{\alpha_j}, j = 1, \ldots, 2r - 1 \). If \( f_j > 0 \) for all \( j \), then

\[
\sqrt{n}(\hat{S}_{r,n} - S_r) \xrightarrow{d} N(0, \tau^2),
\]

where

\[
\tau^2 = \frac{B_r^2 \tau_1^4 - 2A_r B_r \tau_1 \tau_2 + A_r^2 \tau_2^2}{B_r^4}
\]
Theorem 2. Assume, without loss of generality, that $F$ is symmetric about $r$.

Theorem 1 simplifies to

\[ W_i = \frac{\gamma'(X_i)f'(X_i)}{f(X_i)} + \gamma'(X_i), \]

and

\[
\begin{align*}
\tau_1^2 &= \sum_{i=1}^{2r-1} \sum_{j=1}^{2r-1} c_ic_j\alpha_{\min(i,j)}(1 - \alpha_{\max(i,j)}) \div f_f', \\
\tau_{12} &= \sum_{i=1}^{2r-1} \sum_{j=1}^{2r-1} c_id_j\alpha_{\min(i,j)}(1 - \alpha_{\max(i,j)}) \div f_f', \\
\tau_2^2 &= \sum_{i=1}^{2r-1} \sum_{j=1}^{2r-1} d_id_j\alpha_{\min(i,j)}(1 - \alpha_{\max(i,j)}) \div f_f'.
\end{align*}
\]

Proof. Let $\hat{A}_{r,n}$ and $\hat{B}_{r,n}$ be the sample analogues of $A_r$ and $B_r$, respectively. Then

\[ \sqrt{n}(\hat{S}_{r,n} - S_r) = \sqrt{n} \cdot \frac{B_r(\hat{A}_{r,n} - A_r) - A_r(\hat{B}_{r,n} - B_r)}{B_{r,n}B_r}. \] (2)

By Lemma 1,

\[ \hat{A}_{r,n} - A_r = \bar{Y} + o_p(n^{-1/2}), \]

where $\bar{Y}$ is the mean value of

\[ Y_i = \sum_{j=1}^{2r-1} c_j(\alpha_j - I(X_i \leq Q_{\alpha_j})) \div f_j, \quad i = 1, \ldots, n, \] (3)

and

\[ \hat{B}_{r,n} - B_r = \bar{Z} + o_p(n^{-1/2}), \]

where $\bar{Z}$ is the mean value of

\[ Z_i = \sum_{j=1}^{2r-1} d_j(\alpha_j - I(X_i \leq Q_{\alpha_j})) \div f_j, \quad i = 1, \ldots, n. \]

By the multivariate central limit theorem, $n^{1/2}\bar{Y}$ and $n^{1/2}\bar{Z}$ are jointly asymptotically normal with mean zero and covariance matrix

\[ \begin{pmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{pmatrix}. \]

By Slutsky’s theorem, the same is true for $n^{1/2}(\hat{A}_{r,n} - A_r)$ and $n^{1/2}(\hat{B}_{r,n} - B_r)$, which implies that

\[ \sqrt{n}(B_r(\hat{A}_{r,n} - A_r) - A_r(\hat{B}_{r,n} - B_r)) \xrightarrow{d} N(0, \tau^2). \] (4)

The conclusion of the theorem follows from (2), (4), and the Slutsky’s theorem by noting that $\hat{B}_{r,n}$ tends in probability to $B_r$ as $n \to \infty$. \qed

We consider now the null hypothesis of symmetry about an unknown point of symmetry. Under the assumption that the distribution function $F$ is symmetric about some unknown value $\mu$, i.e., $F(t - \mu) = 1 - F(-(t - \mu))$ for all $t$, we see that Theorem 1 simplifies to

\[ \sqrt{n}(\hat{S}_{r,n} - S_r) \xrightarrow{d} N(0, B_r^{-2}\tau_1^2). \]

The asymptotic distribution of $\tilde{S}_{r,n}$ under close alternatives is studied next. As in Antille and Kersting (1977) and Ekström and Jammalamadaka (2007), we consider close alternatives of the form $G_n(t) = F(t + n^{-1/2}\gamma(t))$. The function $t + n^{-1/2}\gamma(t)$ is assumed to be a monotonically increasing function of $t$ for large $n$. We remark that a result of Wang (2008) deals with the special case where $r = 2$ and $\gamma(t) = \gamma_n(t) = -\eta t/(1 + n^{-1/2}\eta)$ if $t > 0$, and 0 otherwise.

Theorem 2. Assume, without loss of generality, that $F$ is symmetric about 0, that the density function $f$ is twice continuously differentiable and strictly positive on the whole real line, and that the function $\gamma$ is continuously differentiable. Let

\[ W_i = \frac{\gamma'(X_i)f'(X_i)}{f(X_i)} + \gamma'(X_i), \]
and assume that $E_{h_0} W_i^2$ and $E_{h_0} Y_i^2$ are finite, and that $\lim_{t \to \pm \infty} \gamma(t)f(t) = \lim_{t \to \pm \infty} \gamma^2(t)f'(t) = 0$. Then, under $H_{A,n}$,

$$n^{1/2} \tilde{S}_{r,n} \overset{d}{\to} N \left( B_{r}^{-1} E_{h_0} Y_i W_i, B_{r}^{-2} \tau^2_i \right),$$

where $Y_i$ is defined in (3).

**Proof.** By Le Cam’s third lemma (see, e.g., van der Vaart, 1998), we can obtain the limiting distribution of $n^{1/2} \tilde{S}_{r,n}$ under $H_{A,n}$, once we have the joint limit distribution of $L_n = \sum_{i=1}^n \log(G_n(x_i)/F(x_i))$ and $n^{1/2} \tilde{S}_{r,n}$ under $H_0$. We begin with $L_n$. Let

$$V_i = \frac{f(X_i + n^{-1/2} \gamma(Y_i)) - f(X_i) + n^{-1/2} \gamma'(X_i)f'(X_i + n^{-1/2} \gamma(Y_i))}{f(X_i)}.$$

If we write $\log(1+x) = x - x^2/2 + x^2 R(x)$, then $R(x) \to 0$ as $x \to 0$, and

$$L_n = \sum_{i=1}^n \log \left( \frac{(1 + n^{-1/2} \gamma'(X_i)) f(X_i + n^{-1/2} \gamma(Y_i))}{f(X_i)} \right) = \sum_{i=1}^n \log(1 + V_i) = \sum_{i=1}^n V_i - \frac{1}{2} \sum_{i=1}^n V_i^2 + \sum_{i=1}^n V_i^2 R(V_i).$$

By Taylor expansions,

$$\sum_{i=1}^n V_i = \frac{1}{n^{1/2}} \sum_{i=1}^n \gamma(X_i)f'(X_i)f(X_i) + \frac{1}{2n} \sum_{i=1}^n \gamma^2(X_i)f''(X_i)f(X_i) + \frac{1}{2n} \sum_{i=1}^n \gamma^2(X_i) f''(X_i) \left( f'(X_i + n^{-1/2} \rho_i \gamma(Y_i)) - f'(X_i) \right) + \frac{1}{n} \sum_{i=1}^n \gamma(X_i) \gamma(Y_i) f'(X_i) \left( f'(X_i + n^{-1/2} \eta_i \gamma(Y_i)) - f'(X_i) \right),$$

(5)

where all $\rho_i$ and $\eta_i$ are between 0 and 1. Consider the third term on the right. For each $\delta_0 > 0$ there exist a closed interval $[a_1, a_2]$ and an integer $N > 0$ such that $P_{h_0}(X_1 \in B_{r,n}) \geq 1 - \delta_0/2$ for all $n > N$, where $B_{r,n} = \{ t : a_1 \leq t \leq a_2 \}$ and $a_1 \leq t + n^{-1/2} \gamma(t) \leq a_2$. By our assumptions, $f''$ is uniformly continuous in $[a_1, a_2]$. Hence, for each $\delta > 0$ there exists a $\delta > 0$ such that for all $\epsilon$ and $t + n^{-1/2} \rho_i \gamma(t)$ in $[a_1, a_2]$ with $|n^{-1/2} \rho_i \gamma(t)| \leq \delta$, we have $|f''(t + n^{-1/2} \rho_i \gamma(t)) - f''(t)| \leq \epsilon$. Thus, for all $n$ large enough,

$$P_{h_0} \left( \max_{1 \leq i \leq n} \left| f''(X_i + n^{-1/2} \rho_i \gamma(Y_i)) - f''(X_i) \right| \geq \epsilon \right)$$

$$\leq P_{h_0} \left( \max_{1 \leq i \leq n} \left| f''(X_i + n^{-1/2} \rho_i \gamma(Y_i)) - f''(X_i) \right| \geq \epsilon \text{ and } X_i \notin B_{r,n} \right) + P_{h_0}(X_1 \notin B_{r,n})$$

$$\leq \frac{1}{n} \sum_{i=1}^n \gamma(X_i) \gamma(Y_i) \left( f'(X_i + n^{-1/2} \eta_i \gamma(Y_i)) - f'(X_i) \right),$$

(6)

where the last inequality follows from the fact that $\gamma$ is a bounded function on $\bigcup B_{r,n}$. The third term on the right-hand side of (5) is bounded by

$$\left| \max_{1 \leq i \leq n} \left| f''(X_i + n^{-1/2} \rho_i \gamma(Y_i)) - f''(X_i) \right| \right| \frac{1}{2n} \sum_{i=1}^n \gamma^2(X_i) f(X_i).$$

Thus it is $o_p(1) + o_p(1)$, and converges in probability to zero. Similarly, the last term on the right-hand side of (5) tends to zero in probability. Thus, under $H_0$ we have

$$\sum_{i=1}^n V_i = \frac{1}{n^{1/2}} \sum_{i=1}^n \frac{\gamma(X_i) f'(X_i)}{f(X_i)} + \frac{1}{n} \sum_{i=1}^n \gamma'(X_i) f(X_i) + \frac{1}{2} E_{h_0} \frac{\gamma^2(X_i) f''(X_i)}{f(X_i)} + E_{h_0} \frac{\gamma'(X_i) \gamma(Y_i) f'(X_i)}{f(X_i)} + o_p(1),$$

and, by similar arguments as above,

$$\sum_{i=1}^n V_i^2 = E_{h_0} \left( \frac{\gamma(X_i) f'(X_i)}{f(X_i)} \right)^2 + E_{h_0} \left( \frac{\gamma(Y_i) f'(X_i)}{f(X_i)} \right)^2 + 2E_{h_0} \frac{\gamma'(X_i) \gamma(Y_i) f'(X_i)}{f(X_i)} + o_p(1).$$


By arguments similar to those in (6), we see that \( \max_{1 \leq i \leq n} |V_i| \) converges in probability to zero. By the definition of the function \( R \), the sequence \( \max_{1 \leq i \leq n} |R(V_i)| \) converges in probability to zero as well. Combining these results, we obtain that

\[
L_n = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{\gamma(X_i)f'(X_i)}{f(X_i)} + \frac{1}{n^{1/2}} \sum_{i=1}^{n} \gamma'(X_i) + \frac{1}{2} E_{H_0} \frac{\gamma^2(X_i)f''(X_i)}{f(X_i)}
\]

\[
- \frac{1}{2} E_{H_0} \left( \frac{\gamma(X_i)f'(X_i)}{f(X_i)} \right)^2 - \frac{1}{2} E_{H_0} \left( \gamma'(X_i) \right)^2 + o_p(1)
\]

\[
= \frac{1}{n^{1/2}} \sum_{i=1}^{n} W_i - \frac{1}{2} E_{H_0} W_i^2 + o_p(1),
\]

(7)

where the last equality follows by integration by parts. By another integration by parts we see that \( E_{H_0} W_1 = 0 \).

By the proof of Theorem 1, \( n^{1/2} \hat{S} = n^{1/2} B^{-1}_r \hat{Y} + o_p(1) \) under \( H_0 \). Thus, under \( H_0 \), \( n^{1/2} \hat{S}_{r,n} \) and \( L_n \) are jointly asymptotically normal,

\[
N \left( \left( \begin{array}{c} 0 \\ -2^{-1} E_{H_0} W_i^2 \end{array} \right), \left( \begin{array}{cc} B^{-2}_r \tau_1^2 & B^{-1}_r E_{H_0} Y_1 W_1 \\ B^{-1}_r E_{H_0} Y_1 W_1 & E_{H_0} W_i^2 \end{array} \right) \right),
\]

and under \( H_{A,n} \), Le Cam’s third lemma implies that

\[
n^{1/2} \hat{S}_{r,n} \rightarrow_d N \left( B^{-1}_r E_{H_0} Y_1 W_1, B^{-2}_r \tau_1^2 \right). \quad \square
\]

In Theorems 1 and 2, a fixed number, \( 2r - 1 \), of sample quantiles have been used to define the statistic \( \hat{S}_{r,n} \). Next we consider a case where this number is allowed to increase with the sample size. That is, we will consider the case where \( 2r - 1 = n \), so that we have as many sample quantiles defining the statistic \( \hat{S}_{r,n} \) as observations in the sample. Assume, for simplicity, that \( n = 2r - 1 \) is an odd number, and set \( \alpha_i = i/(n + 1), \) \( i = 1, \ldots, n \). Then \( \hat{Q}_{\alpha_i} = X_{(i)} \) for each \( i \), where \( X_{(i)} \) is the \( i \)th ordered statistic of the sample, and our statistic can be written in the following form,

\[
\hat{S}_{r,n} = \frac{\hat{X} - \hat{M}}{\hat{D}_n},
\]

(8)

where \( r_n = (n + 1)/2, \) \( \hat{M} \) is the sample median, and \( \hat{D}_n = n^{-1} \sum_{i=1}^{n} |X_i - \hat{M}| \).

Let \( \mu, M \) and \( \sigma^2 \) denote respectively the mean, median and variance of the distribution \( F \). Let \( D = E[X_1 - M] \),

\[
\nu^2 = \sigma^2 + \frac{1}{4f^2(M)} - \frac{D}{f(M)},
\]

and

\[
U_i = (X_i - \mu) - \frac{1/2 - I(X_i \leq M)}{f(M)}, \quad i = 1, \ldots, n.
\]

(9)

**Theorem 3.** Assume that \( \sigma^2 < \infty \) and that the assumptions of Theorem 2 are valid. Then, under \( H_{A,n} \),

\[
n^{1/2} \hat{S}_{r,n} \rightarrow_d N \left( D^{-1} E_{H_0} U_1 W_1, D^{-2} \nu^2 \right),
\]

where \( \hat{S}_{r,n} \) is defined as in (8).

**Proof.** Again, with the help of Le Cam’s third lemma, we will obtain the limit distribution of \( n^{1/2} \hat{S}_{r,n} \) under \( H_{A,n} \), once the joint limit distribution of \( L_n = \sum_{i=1}^{n} \log(C_i(X_i)/f'(X_i)) \) and \( n^{1/2} \hat{S}_{r,n} \) under \( H_0 \) is known.

By **Lemma 1**, under \( H_0 \),

\[
\hat{X} - \hat{M} = (\hat{X} - \mu) - (\hat{M} - M) = \bar{U} + o_p(n^{-1/2}),
\]

where \( \bar{U} \) is the mean value of \( U_1, \ldots, U_n \). By the central limit theorem, \( n^{1/2} \bar{U} \) is, under \( H_0 \), asymptotically normal with mean zero and variance \( \nu^2 \), and by Slutsky’s theorem, the same is true for \( n^{1/2}(\hat{X} - \hat{M}) \). The statistic \( \hat{D}_n \) converges in probability to \( D \) as \( n \rightarrow \infty \) (Mira, 1999). Thus, under \( H_0 \), \( n^{1/2} \hat{S}_{r,n} = n^{1/2} D^{-1} \bar{U} + o_p(1) \). Under \( H_0 \), this result combined with (7) imply that \( n^{1/2} \hat{S}_{r,n} \) and \( L_n \) are jointly asymptotically normal,

\[
N \left( \left( \begin{array}{c} 0 \\ -2^{-1} E_{H_0} W_i^2 \end{array} \right), \left( \begin{array}{cc} D^{-2} \nu^2 & D^{-1} E_{H_0} U_1 W_1 \\ D^{-1} E_{H_0} U_1 W_1 & E_{H_0} W_i^2 \end{array} \right) \right),
\]
and under $H_{H,n}$, Le Cam’s third lemma implies that

$$n^{1/2} S_{r,n} \xrightarrow{d} N \left( D^{-1} E_{h_0} U_1 W_1, D^{-2} I_1^2 \right). \quad \square$$

**Remark 1.** The statistic (8), $S_{r,n}$, is the sample analogue of Groeneveld and Meeden’s (1984) measure of skewness, and is essentially the test statistic proposed by Miao et al. (2006), i.e., their statistic is obtained by multiplying the right-hand side of (8) by $(2/\pi)^{1/2}$. Miao et al. show, under general conditions, that

$$n^{1/2} \tilde{r}_n = n^{1/2} \left( \frac{\hat{X} - \hat{M}}{D_n} - \frac{\mu - M}{D} \right)$$

tends to a normal distribution with mean zero and variance

$$\frac{1}{D^2} \left( \sigma^2 + \frac{1}{4f^2(M)} - \frac{D}{f(M)} + \frac{(\mu - M)^2(\sigma^2 + (\mu - M)^2)}{D^2} - (\mu - M)^2 \right)

- \frac{2(\mu - M)}{D} \left( \sigma^2 - 2EX_1^2 I(X_1 \leq M) + 2\mu EX_1 I(X_1 \leq M) - MD \right) + \frac{(\mu - M)^2}{Df(M)},$$

Under the null hypothesis when $F$ is symmetric, then the asymptotic distribution of $n^{1/2} \tilde{r}_n$ is normal with mean zero and variance given by $D^{-2} I_1^2$.

**Remark 2.** Consider the measure of skewness defined as $S_{Yule} = (\mu - M)/\sigma$. The sample version of this measure, $\hat{S}_{Yule,n} = (\hat{X} - \hat{M})/s$, where $s$ is the sample standard deviation, is discussed in Gastwirth (1971), and in the case $\mu = M$, Cabilio and Masaro (1996) showed that $n^{1/2} \hat{S}_{Yule,n}$ is asymptotically normal with mean zero and variance $\sigma^{-2} \nu^2$. Under $H_{H,n}$, and the assumptions of Theorem 3, it is easily seen that

$$n^{1/2} \hat{S}_{Yule,n} \xrightarrow{d} N \left( \sigma^{-1} E_{h_0} U_1 W_1, \sigma^{-2} \nu^2 \right).$$

**Remark 3.** Consider the standardized third moment $\gamma_1 = \mu_3/\sigma^3$, where $\mu_k$ is the $k$th central moment. The sample version of this measure is $\hat{\gamma}_{1,n} = m_3/s^3$, where $m_k$ is the sample $k$th central moment. If $F$ is symmetric and $\mu_6 < \infty$, then $\gamma_1 = 0$ is zero and $n^{1/2} \hat{\gamma}_{1,n}$ is asymptotically normal with mean zero and variance $\sigma^{-2}(\mu_6 - 6\sigma^2 \mu_4 + 9\sigma^6)$ (see, e.g., Gupta (1967) for more details). If $\mu_6 < \infty$ and the assumptions of Theorem 2 are valid, then it can be shown under $H_{H,n}$ that

$$n^{1/2} \hat{\gamma}_{1,n} \xrightarrow{d} N \left( \sigma^{-3}(E_{h_0} X_1^3 W_1 - 3E_{h_0} X_1^2 E_{h_0} X_1 W_1), \sigma^{-6}(\mu_6 - 6\sigma^2 \mu_4 + 9\sigma^6) \right).$$

### 3. Pitman asymptotic relative efficiency

The Pitman asymptotic relative efficiency (ARE) of a test relative to another test is defined to be the limit of the inverse ratio of sample sizes required to obtain the same limiting power at a sequence of alternatives converging to the hypothesis. The limiting power should be a value at the limiting test size, $\alpha$, and the maximum power, $1$. If the limiting power of a test at a sequence of alternatives is $\alpha$, then its ARE with respect to any other test with the same test size and with limiting power greater than $\alpha$, is zero. On the other hand, if the limiting power of a test at a sequence of alternatives converges to a number in the open interval $(\alpha, 1)$, then a measure of rate of convergence, called efficacy, can be computed. Under certain standard regularity assumptions (see, for example, Rao, 1973, p. 469), which include a condition about the nature of the alternative, asymptotic normal distribution of the test statistic under the sequence of alternatives, etc., this efficacy is given by

$$\text{efficacy} = \frac{\mu_+^2}{\sigma^2}.$$ 

Here $\mu_+$ and $\sigma^2$ are the mean and variance of the limiting normal distribution under the sequence of alternatives when the test statistic has been normalized to have a limiting standard normal distribution under the hypothesis. In such a situation, the ARE of one test with respect to another is simply the ratio of their efficacies.

The efficacy of the test statistic $n^{1/2} S_{r,n}$ from Theorem 2 is

$$\frac{(E_{h_0} Y_1 W_1)^2}{\tau_1},$$

and the efficacy of the test statistic $n^{1/2} \tilde{S}_{r,n}$ from Theorem 3 is

$$\frac{(E_{h_0} U_1 W_1)^2}{\nu^2}.$$
Also, it should be noted that the test statistic \( n^{1/2} \tilde{S}_{\text{Yule,n}} \) has the same efficacy as \( n^{1/2} \tilde{S}_{\text{m,n}} \). The efficacy of \( n^{1/2} \tilde{S}_{\text{m,n}} \) from Remark 3 is

\[
\left( E_{\text{H}_0} X_1^4 W_1 - 3 E_{\text{H}_0} X_2^2 E_{\text{H}_0} X_1 W_1 \right)^2 / \mu_6 - 6 \sigma^2 \mu_4 + 9 \sigma^6.
\]

Computations of efficacy will be made in the case where \( \gamma(t) = t, \ t \geq 0, \) and 0 elsewhere. In this case, \( B^{-1} E_{\text{H}_0} Y_1 W_1 = \) \( D^{-1} E_{\text{H}_0} U_1 W_1 = -1/2, \ E_{\text{H}_0} X_2^2 W_1 = -(3/2) E_{\text{H}_0} |X_1|^2, \) and \( E_{\text{H}_0} X_1 W_1 = -(1/2) E_{\text{H}_0} |X_1| \).

We will consider distributions \( f \) with the following densities,

- \( t \) distribution : \( f_0(x) = \frac{1}{(\theta \pi)^{1/2}} \frac{1 - x^2}{\theta}^{-\frac{\theta+1}{2}}, \)
- Exponential power (EP) distribution : \( f_0(x) = \frac{e^{-|x|^\theta}}{2 \Gamma(1 + \theta^{-1})}, \)
- Logistic distribution : \( f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}. \)

The double exponential distribution, also known as the Laplace distribution, is obtained by setting \( \theta = 1 \) in the EP distribution, and the normal distribution with mean zero and variance 1/2 is obtained by setting \( \theta = 2 \) in the EP distribution. The \( t \) distribution with \( \theta = 1 \) degree of freedom is also known as the standard Cauchy distribution.

In Table 1, for each given \( r \), we choose the \( \alpha \)'s which maximize the efficacy. From this table as well as heuristically, it appears that the efficacy increases when more quantiles are brought into play, but a reasonably practical approach may be to use \( r = 3 \) with 2 quantiles from either tail. It can also be observed that in thin-tailed distributions like the normal, one should use much smaller-order quantiles e.g. the 5th and 95th percentiles with \( r = 2 \), whereas in fat-tailed models like the Cauchy, Bowley’s choice of the 25th and 75th seems optimal. Thus Bowley’s is a better choice than Kelly’s for heavy-tailed distributions whereas the reverse is true for thin-tailed distributions like the normal. In general there is no universally optimal choice of these quantiles for all distributions. For instance we have verified that with \( r = 3 \), by choosing \( \alpha_1 = 1/6 \) and \( \alpha_2 = 2/6 \) we do uniformly better than Bowley’s test and by choosing \( \alpha_1 = 1/10 \) and \( \alpha_2 = 2/10 \) we can beat Kelly’s test for all the distributions checked.

However if one were to build a reasonably robust test of symmetry with \( r = 3 \) which does better over a wide range of distributions, an overall compromise seems to be to take \( \alpha_1 = 0.1 \) and \( \alpha_2 = 0.2 \) i.e. compare the 10th, 20th, 80th, and 90th percentiles to the median, getting the test:

\[
\tilde{S}_{\text{m,n}}^* = \frac{Q_{0.90} + Q_{0.80} - 4M + Q_{0.20} + Q_{0.10}}{(Q_{0.90} - Q_{0.10}) + (Q_{0.80} - Q_{0.20})}.
\]

Such a test as Table 2 indicates, will do just as well or better than Kelly’s in all the cases and falls short of Bowley’s only in the case of one distribution viz. the Cauchy. Even further improvements over the Bowley’s and Kelly’s coefficients can be demonstrated if we use \( \tilde{S}_{\text{m,n}} \) with a larger \( r \) but we omit such comparisons.

In Antille and Kersting (1977), the following four test statistics for symmetry were considered,

\[
\tilde{S}_{\text{A&K},n} = \sum_{1 \leq i \leq n} \left( I(X_i - \bar{X} \leq 0) - \frac{1}{2} \right).
\]
Table 2
AREs of \( \hat{S}^*_{1,n} \) relative to Bowley’s and Kelly’s coefficients.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( \theta )</th>
<th>Bowley Efficacy</th>
<th>( \text{ARE} )</th>
<th>Kelly Efficacy</th>
<th>( \text{ARE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_1 ) (Cauchy)</td>
<td>1</td>
<td>0.101</td>
<td>0.70</td>
<td>0.051</td>
<td>1.38</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>2</td>
<td>0.129</td>
<td>1.27</td>
<td>0.129</td>
<td>1.27</td>
</tr>
<tr>
<td>EP (Laplace)</td>
<td>1</td>
<td>0.120</td>
<td>1.56</td>
<td>0.162</td>
<td>1.16</td>
</tr>
<tr>
<td>EP (normal)</td>
<td>2</td>
<td>0.136</td>
<td>1.73</td>
<td>0.233</td>
<td>1.01</td>
</tr>
<tr>
<td>Logistic</td>
<td></td>
<td>0.136</td>
<td>1.67</td>
<td>0.211</td>
<td>1.08</td>
</tr>
</tbody>
</table>

Table 3
Efficacies of the test statistics.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( \theta )</th>
<th>( \hat{S}^*_{1,n} )</th>
<th>( \hat{S}^*_{p,n} )</th>
<th>( \hat{S}^*_{1,n} )</th>
<th>( \hat{S}^*_{4,n} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_1 ) (Cauchy)</td>
<td>1</td>
<td>0.071</td>
<td>0.123</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td>( t_2 )</td>
<td>2</td>
<td>0.164</td>
<td>0.171</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>EP (Laplace)</td>
<td>1</td>
<td>0.188</td>
<td>0.250</td>
<td>0.112</td>
<td>0.250</td>
</tr>
<tr>
<td>EP (normal)</td>
<td>2</td>
<td>0.236</td>
<td>0.279</td>
<td>0.239</td>
<td>0.279</td>
</tr>
<tr>
<td>Logistic</td>
<td></td>
<td>0.227</td>
<td>0.275</td>
<td>0.238</td>
<td>0.275</td>
</tr>
</tbody>
</table>

\[
\hat{S}_{\text{A&K}2,n} = \sum_{1 \leq j < n} \left( I(X_i + X_j \leq 2\bar{X}) - \frac{1}{2} \right),
\]
\[
\hat{S}_{\text{A&K}3,n} = \sum_{1 \leq j < n} \left( I(X_i + X_j \leq 2\bar{M}) - \frac{1}{2} \right),
\]
\[
\hat{S}_{\text{A&K}4,n} = \sum_{1 \leq j < n} \left( I(D_i - D_{n-j+2} \leq 0) - \frac{1}{2} \right).
\]


Remark 4. Under the alternative, \( n^{1/2}\hat{S}_{\text{A&K}4,n} \) is asymptotically normal with mean \( \mu_{4,c} \) and variance \( \sigma^2_{4,c} \), and \( \mu_{4,c} \) and \( \sigma^2_{4,c} \) possess limits as \( \varepsilon \to 0 \), \( \mu_{4} \) and \( \sigma^2_{4} \) respectively, but it is not known whether \( n^{1/2}\hat{S}_{\text{A&K}4,0,n} \) is asymptotically normal with mean \( \mu_{4} \) and variance \( \sigma^2_{4} \). However, based on the claim of Antille and Kersting that it is, we compute the efficacy of \( n^{1/2}\hat{S}_{\text{A&K}4,0,n} \) as \( \mu_{4}/\sigma^2_{4} \).

Table 3 provides the efficacies of \( \hat{S}^*_{A,n} \) against Antille and Kersting’s four tests, \( \hat{S}_{1,n} \) given in Eq. (8), and the classical coefficient of skewness \( \gamma_{1,n} \). It is interesting to note that the proposed \( \hat{S}^*_{A,n} \), although based on just 5 quantiles and easier to compute, compares reasonably well with all these other tests which use all the observations/quantiles or gaps between them.

4. A practical example from forestry

We consider two data sets taken from Matérn (1981, Table 2.1), consisting of the diameters at breast height in millimeters from year 1912 and year 1951 for \( n = 115 \) Norway spruce trees from the Forest Research Institute of Sweden’s sample plot number 238 at Finnerödja. As seen in Fig. 1, the data from 1912 show only a slight skewness, while a more distinct (negative) skewness is visible in the data from 1951. Assume that the data (from 1912 or 1951) are taken from some distribution \( F \), and that we want to use the test statistic \( \hat{S}^*_{1,n} \) for testing the hypothesis of symmetry, i.e. that \( F(x - \mu) = 1 - F(-(x - \mu)) \) for all \( x \). If \( F \) is symmetric and \( \hat{\xi}_{1,n} \) is a consistent estimator of \( \xi_1 \), then Theorem 1 implies that \( n^{1/2}\hat{S}^*_{1,n}/\hat{\sigma}_{1,n} \) is approximately \( N(0, 1) \) for large \( n \), and approximate \( p \)-values of the test can be computed. In what follows we use Kraft et al.’s (1985) symmetrized kernel density estimator for estimating the density values \( \hat{f}_j \), \( j = 1, \ldots, 2r - 1 \), in the definition of \( \hat{S}^*_{1,n} \). The kernel density estimation is performed using the R function “density” (R Development Core Team, 2011) with a Gaussian kernel. (Another possibility for computing approximate \( p \)-values is to use bootstrap; see e.g. Thomas, 2009 and Zheng and Gastwirth, 2010.) For both data sets, the values of \( \hat{S}^*_{1,n} \) and Bowley’s and Kelly’s coefficients are given in Table 4. Corresponding \( p \)-values are also given. As expected, all \( p \)-values are far from a significance level of 0.05 for the data.
Fig. 1. Normal probability plots and histograms of \( n = 115 \) diameters at breast height from year 1912 and year 1951.

Table 4

<table>
<thead>
<tr>
<th>Year</th>
<th>Bowley Value</th>
<th>Bowley p-value</th>
<th>Kelly Value</th>
<th>Kelly p-value</th>
<th>( S_{3,n}^* ) Value</th>
<th>( S_{3,n}^* ) p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1912</td>
<td>0.000</td>
<td>1.00</td>
<td>-0.104</td>
<td>0.34</td>
<td>-0.067</td>
<td>0.52</td>
</tr>
<tr>
<td>1951</td>
<td>0.030</td>
<td>0.83</td>
<td>-0.285</td>
<td>0.02</td>
<td>-0.224</td>
<td>0.04</td>
</tr>
</tbody>
</table>

from 1912. For the data from 1951, Bowley’s coefficient fails to reject the hypothesis of symmetry, with a corresponding p-value as large as 0.83, whereas the tests based on Kelly’s coefficient and \( S_{3,n}^* \) do reject the null hypothesis, which corresponds well with the visual inspection of Fig. 1. The failure of Bowley’s coefficient in detecting the obvious skewness in the 1951 data in particular, illustrates the danger of relying on a test based only on one quantile from each tail, like the Bowley’s and Kelly’s.

5. Summary

In the current article we propose a very general class of distribution-free tests of symmetry that encompass most known measures based on quantiles. We provide large sample theory for such statistics and compare their asymptotic relative efficiencies against some competing test statistics for symmetry. In particular we suggest using a simple test \( S_{3,n}^* \) which uses 5 quantiles and fares well against all the other known tests of symmetry for a wide variety of distributions. Computer code in R for implementing this test, is available from the authors.

Acknowledgment

We thank the anonymous referee for a careful reading of this paper as well as for the insightful suggestion that we add a practical example to illustrate the proposed statistics.

References


