Work supported by the National Institute of Health, Public Health Service, Grant GM-12868.

ON KOLMOGOROV-SMIRNOV-TYPE TESTS FOR SYMMETRY

by

Shoutir Kishore Chatterjee and Pranab Kumar Sen

University of Calcutta, and University of North Carolina

Institute of Statistics Mimeo Series No. 645
ON KOLMOGOROV-SMIRNOV-TYPE TESTS FOR SYMMETRY*

BY SHOUTIR KISHORE CHATTERJEE AND PRANAB KUMAR SEN
University of Calcutta, and University of North Carolina

0. Summary. For testing the hypothesis of symmetry (about a specified point), a simple Kolmogorov-Smirnov-type test is proposed. The exact and asymptotic (null hypothesis) distributions of some allied statistics are obtained, and the Bahadur-efficiency of the test is studied.

1. Introduction. Let \{X_i\} be a sequence of independent real valued random variables with continuous distribution functions (df) \{F_i(x)\}, all defined on \((-\infty, \infty)\) and not necessarily identical. Based on a sample \(X_1, \ldots, X_n\), we want to test the null hypothesis \(H_0\) that all the df \(F_1, \ldots, F_n\) are symmetric around their respective (specified) medians. Without any loss of generality, we may take all these medians to be equal to zero, and thus frame \(H_0\) as

\[(1.1) \quad H_0: \quad F_i(x) + F_i(-x) = 1, \quad \forall \quad x > 0, \quad \text{and} \quad i = 1, \ldots, n.\]

Let \(c(u)\) be equal to 0 or 1 according as \(u < 0\) or \(u > 0\), and let

\[(1.2) \quad F_n^*(x) = n^{-1} \sum_{i=1}^{n} c(x - X_i), \quad F_n(x) = n^{-1} \sum_{i=1}^{n} F_i(x), \quad -\infty < x < \infty.\]

Thus, \(F_n^*\) is the empirical df and it estimates unbiasedly the average df \(F_n\) (a.e.). In testing the null hypothesis (1.3), we are interested in the following alternative hypotheses:

\[(1.3) \quad H_1: \quad \sup_{x > 0} [F_n(x) + F_n(-x)] = 1, \quad H_2: \quad \inf_{x > 0} [F_n(x) + F_n(-x)] < 1;\]

\[(1.4) \quad H_3 = H_1 \cup H_2: \quad \sup_{x > 0} [F_n(x) + F_n(-x) - 1] > 0.\]

* Work supported by the National Institute of Health, Public Health Service, Grant GM-12868.
When $F_1=\ldots=F_n=F$, (1.3) [(1.4)] relates to one [two-] sided skewness. In many practical problems, though it may be unwise to impose the restriction that $F_1=\ldots=F_n=F$, it may not be unreasonable to assume that $F_1, \ldots, F_n$ have a common pattern of skewness, when (1.3) does not hold. For example, let $F_i(x) = F_0^i([x-\mu_i]/\sigma_i)$, and $F_0^i \in \mathcal{G}_0$, $i=1, \ldots, n$, where $\mathcal{G}_0 = \{ F: F(x)+F(-x) = 1 \text{ a.e.} \}$, and the $\mu_i$ and $\sigma_i$ are the location and scale parameters. If the $\mu_i$ have all the same sign, $F_1, \ldots, F_n$ are either all positively or all negatively skewed, no matter whatever be their forms and the $\sigma_i(>0)$. We also note that whereas in the classical one-sample goodness of fit problem (where the Kolmogorov-Smirnov-type tests apply) we require to assume the form of the true df, the same is not needed here. Also, unlike the one-sample location problem, we are not necessarily confining ourselves only to translation alternatives. In fact, even if all the medians of the df $F_1, \ldots, F_n$ are equal to 0, the alternatives in (1.3) and (1.4) may hold. In addition, the homogeneity of $F_1, \ldots, F_n$ is totally inessential.

For testing the null hypothesis, we consider the following Kolmogorov-Smirnov-type statistics whose appeal are evident from (1.3) and (1.4):

$$D_n^+ = \sup_{x \geq 0} \left[ F_n(x)+F_n(-x)-1 \right], \quad D_n^- = \sup_{x \geq 0} \left[ 1-F_n(x)-F_n(-x) \right] ;$$

$$D_n = \max[D_n^+, D_n^-] = \sup_{x \geq 0} \left| F_n(x)+F_n(-x)-1 \right| .$$

Note that $F_n^*$ is a step-function, and hence, to avoid some complications in the distribution theory, we have taken $F_n^*(-x)$ for $F_n^*(-x)$, $x \geq 0$.

The small sample null distributions of $D_n^+$, $D_n^-$ and $D_n$ are deduced in Section 2, and tabulated too, for $n \leq 16$. Section 3 deals with asymptotic null distributions of these statistics. The last section is devoted to the study of the Bahadur-efficiency of the test based on $D_n$ with respect to the sign test.
2. Exact null distributions: An application of the random walk model. Since $F_n^*$ is a step-function assuming the values $i/n$, $i=1,...,n$, the process $[n[F_n^*(x)+F_n^*(-x)-1]: x>0]$ can only assume the integral values between $-n$ to $n$. Thus, the permissible values of $nD_n^+$, $nD_n^-$ and $nD_n$ are the integers $0,1,...,n$, but not all of these are admissible. We denote by $F_n = (F_1,...,F_n)$, and

$$f_n^0 = \{F_n: F_i \in F_0, i=1,...,n\}.$$ 

Then, we have the following theorem.

**Theorem 2.1.** For every $F_n \in f_n^0$ and $k=1,...,n$,

$$P\{nD_{n}^{+}>k\} = P\{nD_{n}^{-}>k\} = 2\left[\sum_{i=0}^{s} \binom{n}{2^i} 2^{-n}\right] - \delta_{k} \binom{n}{s} 2^{-n},$$

where $s = [\frac{1}{2} (n-k)]$ is the largest integer contained in $\frac{1}{2} (n-k)$, $\delta_k = 0$ or $1$ according as $n-k$ is odd or even, and

$$P\{nD_{n}^{-}>k\} = \begin{cases} 1, & k=0,1, \\ 2\sum_{j=0}^{u} (-1)^j P\{nD_{n}^{+}(2j+1)k\}, & k>1, \end{cases}$$

where $u = \lfloor n/2k \rfloor - 1$, $k=1,...,n$.

**Proof.** Let $Y_n^1,...,Y_n^n$ be the ordered values of $|X_1|,...,|X_n|$, arranged in descending order of magnitude. Let $t_{n,i} = F_n(-Y_i)$, for $1 \leq i \leq n$, so that

$$0 < t_{n,1} < t_{n,2} < ... < t_{n,n} < F_n(n)(0) = \frac{1}{2} \left( \text{as } F_n \in f_n^0 \Rightarrow F_n \in F_0 \right).$$

Since, $F_1,...,F_n$ are symmetric and continuous, ties among $|X_1|,...,|X_n|$, and hence among $t_{n,1},...,t_{n,n}$ can be neglected in probability. Thus,

$$0 < t_{n,1} < t_{n,2} < ... < t_{n,n} < \frac{1}{2}, \text{ in probability.}$$

Define then $V_n(t) = \frac{1}{n} [G_n(t)-t]$, $0 < t < 1$, where $G_n(t) = n^{-1} \sum_{i=1}^{n} c(t-F_n(X_i))$, and let

$$V_n(t) = V_n(t-) + V_n(1-t), \quad 0 < t < \frac{1}{2}.$$
For \( t \leq t_{n,1} \), \( n^{k} \mathbf{Y}_{n}(t) = 0 \). At \( t = t_{n,1} \), \( n^{k} \mathbf{Y}_{n}(t) \) is either +1 or -1, depending upon whether the random variable \( X_{1} \) associated with \( Y_{n} \) has negative or positive sign. The process \( n^{k} \mathbf{Y}_{n}(t) \) continues to have the same value until \( t = t_{n,2} \), where it makes another jump of +1 or -1, depending on whether the \( X_{1} \) associated with \( Y_{n-1} \) is negative or not. And thus the process continues. Hence, on \( I = (0, \frac{1}{2}) \), \( n^{k} \mathbf{Y}_{n}(t) \) makes \( n \) jumps (at \( t_{n,1}, \ldots, t_{n,n} \)) and each jump is either +1 or -1.

Let \( p_{ij} = P\{Y_{n-i+1} = |X_{j}|, i, j = 1, \ldots, n, \) (thus \( \sum_{j=1}^{n} p_{ij} = 1, i = 1, \ldots, n \). Since, for \( F_{n} \in \mathfrak{F}_{n}^{O} \), the df of \( X_{i} \) is symmetric about 0, \( 1 \leq i \leq n \),

\[
(2.6) \quad P\{Y_{n-i+1} \text{ corresponds to a positive } X_{j}\} = \frac{1}{2} \sum_{j=1}^{n} p_{ij} \cdot P\{X_{j} > 0 | X_{j} = Y_{n-i+1}\} = \frac{1}{2} \sum_{j=1}^{n} p_{ij} = \frac{1}{2},
\]

as the distribution of sign \( X_{i} \) is independent of \( |X_{j}| \) when \( F_{n} \in \mathfrak{F}_{n}^{O}, i = 1, \ldots, n \). Thus, the jumps (+1 or -1) at \( t_{n,1} \) are both equally likely with probability \( \frac{1}{2} \). Moreover, for \( F_{n} \in \mathfrak{F}_{n}^{O} \), the vector (Sign \( X_{1}, \ldots, \text{Sign } X_{n} \)) is distributed independently

of \( (|X_{1}|, \ldots, |X_{n}|) \) and Sign \( X_{1}, \ldots, \text{Sign } X_{n} \) are also mutually stochastically independent. Hence, the jumps of \( n^{k} \mathbf{Y}_{n}(t) \) at \( t_{n,1}, \ldots, t_{n,n} \) are mutually independent. Finally, the values of \( n^{D} = \sup_{t \in I} n^{k} \mathbf{Y}_{n}(t) \), \( n^{D} = \sup_{t \in I} [n^{k} \mathbf{Y}_{n}(t)] \) and \( n^{D} = \sup_{t \in I} |n^{k} \mathbf{Y}_{n}(t)| \) are independent of the particular realization of \( t_{n} = (t_{n,1}, \ldots, t_{n,n}) \in I \). Hence, we conclude that (i) the distribution of \( n^{D} \) (or \( n^{D} \)) (under \( H_{0} \) ) is the same as that of the maximum positive (or negative) displacement in \( n \) steps of a symmetric random walk starting from the origin, and (ii) the distribution of \( n^{D} \) agrees with that of the corresponding maximum absolute displacement. Thus, the distribution problem is reduced to that of a symmetric random walk problem. Note that \( n^{D} \) and \( n^{D} \) are both non-negative, and hence, \( P\{n^{D} > 0\} = P\{n^{D} > 0\} = 1 \). Also, at \( t_{n,1} \), \( n^{k} \mathbf{Y}_{n}(t) \) is either +1 or -1. Hence, \( n^{D} \), with probability one. So, to prove (2.2), we consider \( k > 1 \), and for (2.3), \( k > 2 \).
We now use theorem 1 (Section 8) of Takacs (1967, p. 24), and obtain that for $k \geq 1$,

\begin{equation}
\Pr\{nD_n^+ > k\} = \Pr\{nD_n^- > k\} = \sum_{j=k}^{n} \binom{n}{j} \Pr\{N_j = j-k\},
\end{equation}

where

\begin{equation}
\Pr\{N_j = j-k\} = \begin{cases} 2^{-j} \binom{j}{r}, & j-k=2r, \, r=0,1,2, \ldots, \\ 0, & j-k = 2r+1; \, j \geq k+1. \end{cases}
\end{equation}

Using an alternative standard expression given in Uspensky (1937, p. 149), (2.7) can be written as

\begin{equation}
2^{-(n-1)} \sum_{k=0}^{n} \binom{n}{k} \delta_k \binom{n}{k} 2^{-n},
\end{equation}

where $s$ and $\delta_k$ are defined after (2.2).

Thus, the proof of (2.2) is completed. Writing now $Q^+(a,n)$ (or $Q(a,n)$) for the probability that a particle starting a symmetric random walk at the origin with the absorbing barrier at $a$ (or barriers at $\pm a$), $a > 0$, will be absorbed at the barrier in course of time $n$, we have

\begin{align}
\Pr\{nD_n \leq k'\} & = 1-2Q(k'+1,n); \\
\Pr\{nD_n^+ \leq k'\} & = 1-Q^+(k'+1,n).
\end{align}

Also, from Uspensky (1937, p. 156), we obtain that

\begin{equation}
Q(k'+1,n) = Q^+(k'+1,n)-Q^+(3k'+3,n)+Q^+(5k'+5,n)
- \ldots \ldots +(-1)^{u}Q^+((2u+1)k',n); \quad u = [(n/2k')-1].
\end{equation}

Then, (2.3) readily follows from (2.10), (2.11), (2.12) and (2.2), by letting $k' = k-1$. Q.E.D.

The probabilities in (2.2) and (2.3) are quite simple to be computed, and are presented below for $n \leq 16$. 

### Table 1

Table for the values of $P(nD^n > k) = P(nD^- > k)$ for $k \leq n \leq 16$

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10*</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.500</td>
<td>.250</td>
<td>.125</td>
<td>.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.625</td>
<td>.250</td>
<td>.125</td>
<td>.063</td>
<td>.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>.625</td>
<td>.375</td>
<td>.125</td>
<td>.063</td>
<td>.031</td>
<td>.016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>.688</td>
<td>.375</td>
<td>.219</td>
<td>.063</td>
<td>.031</td>
<td>.016</td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>.688</td>
<td>.453</td>
<td>.219</td>
<td>.125</td>
<td>.031</td>
<td>.016</td>
<td>.008</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>.727</td>
<td>.453</td>
<td>.289</td>
<td>.125</td>
<td>.072</td>
<td>.016</td>
<td>.008</td>
<td>.004</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>.727</td>
<td>.508</td>
<td>.289</td>
<td>.235</td>
<td>.072</td>
<td>.039</td>
<td>.008</td>
<td>.004</td>
<td>.002</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>.752</td>
<td>.508</td>
<td>.344</td>
<td>.235</td>
<td>.111</td>
<td>.039</td>
<td>.022</td>
<td>.004</td>
<td>.002</td>
<td>.001</td>
<td>.0005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>.752</td>
<td>.549</td>
<td>.344</td>
<td>.282</td>
<td>.111</td>
<td>.065</td>
<td>.022</td>
<td>.012</td>
<td>.002</td>
<td>.001</td>
<td>.0005</td>
<td>.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>.772</td>
<td>.549</td>
<td>.388</td>
<td>.282</td>
<td>.147</td>
<td>.065</td>
<td>.039</td>
<td>.023</td>
<td>.006</td>
<td>.001</td>
<td>.0005</td>
<td>.0002</td>
<td>.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>.772</td>
<td>.581</td>
<td>.388</td>
<td>.317</td>
<td>.147</td>
<td>.092</td>
<td>.039</td>
<td>.023</td>
<td>.006</td>
<td>.0034</td>
<td>.0005</td>
<td>.0002</td>
<td>.0001</td>
<td>.00005</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>.789</td>
<td>.581</td>
<td>.423</td>
<td>.317</td>
<td>.181</td>
<td>.092</td>
<td>.057</td>
<td>.023</td>
<td>.013</td>
<td>.0034</td>
<td>.0018</td>
<td>.0002</td>
<td>.0001</td>
<td>.00005</td>
<td>.00003</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>.789</td>
<td>.607</td>
<td>.423</td>
<td>.329</td>
<td>.181</td>
<td>.119</td>
<td>.057</td>
<td>.035</td>
<td>.013</td>
<td>.0074</td>
<td>.0018</td>
<td>.0009</td>
<td>.0001</td>
<td>.00005</td>
<td>.00003</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>.803</td>
<td>.607</td>
<td>.436</td>
<td>.329</td>
<td>.211</td>
<td>.119</td>
<td>.077</td>
<td>.035</td>
<td>.019</td>
<td>.0074</td>
<td>.0032</td>
<td>.0009</td>
<td>.0004</td>
<td>.00005</td>
<td>.00003</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>.803</td>
<td>.629</td>
<td>.436</td>
<td>.340</td>
<td>.211</td>
<td>.143</td>
<td>.077</td>
<td>.049</td>
<td>.019</td>
<td>.0106</td>
<td>.0032</td>
<td>.0023</td>
<td>.0004</td>
<td>.00026</td>
<td>.00003</td>
</tr>
</tbody>
</table>

* Values are correct to 4 decimal places for $k \geq 10$, and three decimal places for $k < 10$.

### Table 2

Table for the values of $P(nD^- > k)$ for $1 \leq k \leq n \leq 16$

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10*</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.500</td>
<td>.250</td>
<td>.125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.500</td>
<td>.250</td>
<td>.125</td>
<td>.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>.750</td>
<td>.250</td>
<td>.125</td>
<td>.063</td>
<td>.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>.750</td>
<td>.438</td>
<td>.125</td>
<td>.063</td>
<td>.031</td>
<td>.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>.875</td>
<td>.438</td>
<td>.250</td>
<td>.143</td>
<td>.031</td>
<td>.015</td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>.875</td>
<td>.598</td>
<td>.250</td>
<td>.143</td>
<td>.078</td>
<td>.015</td>
<td>.008</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>.938</td>
<td>.598</td>
<td>.470</td>
<td>.143</td>
<td>.078</td>
<td>.043</td>
<td>.008</td>
<td>.004</td>
<td>.0020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>.938</td>
<td>.684</td>
<td>.470</td>
<td>.221</td>
<td>.078</td>
<td>.043</td>
<td>.023</td>
<td>.004</td>
<td>.0020</td>
<td>.0010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>.970</td>
<td>.684</td>
<td>.563</td>
<td>.221</td>
<td>.131</td>
<td>.043</td>
<td>.023</td>
<td>.013</td>
<td>.0020</td>
<td>.0010</td>
<td>.0005</td>
<td>.0005</td>
<td>.0003</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>.970</td>
<td>.764</td>
<td>.563</td>
<td>.294</td>
<td>.131</td>
<td>.077</td>
<td>.023</td>
<td>.013</td>
<td>.0020</td>
<td>.0010</td>
<td>.0005</td>
<td>.0003</td>
<td>.0003</td>
<td>.0001</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>.984</td>
<td>.764</td>
<td>.633</td>
<td>.294</td>
<td>.185</td>
<td>.077</td>
<td>.045</td>
<td>.013</td>
<td>.0068</td>
<td>.0010</td>
<td>.0005</td>
<td>.0003</td>
<td>.0003</td>
<td>.0001</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>.984</td>
<td>.820</td>
<td>.633</td>
<td>.362</td>
<td>.185</td>
<td>.115</td>
<td>.045</td>
<td>.026</td>
<td>.0068</td>
<td>.0036</td>
<td>.0005</td>
<td>.0003</td>
<td>.0003</td>
<td>.0001</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>.992</td>
<td>.820</td>
<td>.656</td>
<td>.362</td>
<td>.237</td>
<td>.115</td>
<td>.070</td>
<td>.026</td>
<td>.0148</td>
<td>.0036</td>
<td>.0018</td>
<td>.0003</td>
<td>.0003</td>
<td>.0001</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>.992</td>
<td>.825</td>
<td>.656</td>
<td>.423</td>
<td>.237</td>
<td>.154</td>
<td>.070</td>
<td>.037</td>
<td>.0148</td>
<td>.0064</td>
<td>.0018</td>
<td>.0008</td>
<td>.0008</td>
<td>.0001</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>.994</td>
<td>.825</td>
<td>.678</td>
<td>.423</td>
<td>.286</td>
<td>.154</td>
<td>.098</td>
<td>.037</td>
<td>.0212</td>
<td>.0064</td>
<td>.0046</td>
<td>.0008</td>
<td>.0003</td>
<td>.0001</td>
</tr>
</tbody>
</table>

* Correct to 4 decimal places for $k \geq 10$ and up to three decimal places for $k < 10$. 

---
3. Asymptotic distribution theory. We consider first the null case. Here, we provide asymptotic expressions for (a) \( P\{ n^{-\frac{1}{2}}D_n > y \} \), \( P\{ n^{-\frac{1}{2}}D_n < y \} \) and \( P\{ n^{-\frac{1}{2}}D_n = y \} \) and (b) \( P\{ D_n^+ > y \} \), \( P\{ D_n^- < y \} \) and \( P\{ D_n = y \} \), where \( y(0 < y < \infty) \) is fixed. For this, let

\[
\Phi(y) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} \exp(-t^2) dt, \quad -\infty < y < \infty.
\]

Then, we have the following theorem.

**Theorem 3.1.** For every fixed \( y(0 < y < \infty) \), under \( H_0 \) (i.e., \( \forall F \in \mathcal{F}_n \)),

\[
\lim_{n \to \infty} P\{ n^{-\frac{1}{2}}D_n > y \} = \lim_{n \to \infty} P\{ n^{-\frac{1}{2}}D_n < y \} = 2\Phi(-y);
\]

\[
\lim_{n \to \infty} P\{ n^{-\frac{1}{2}}D_n = y \} = 4\sum_{k=1}^{\infty} (-1)^{k-1} \Phi(-(2k-1)y)).
\]

**Proof.** Let \( r_n \) be the number of successes in \( n \) independent Bernoullian trials with probability \( \frac{1}{2} \). Then, by (2.2),

\[
P\{ n^{-\frac{1}{2}}D_n > y \} = P\{ n^{-\frac{1}{2}}D_n < y \} = 2P\{ r_n < s_n \} - \delta_k P\{ r_n = s_n \}
= 2P\{ n^{-\frac{1}{2}}(2n - n) < n^{-\frac{1}{2}}(2s_n - n) \} - \delta_k P\{ r_n = s_n \},
\]

where \( s_n = \lceil n/2 - \frac{1}{2}n^2 \rceil \), so that \( n^{-\frac{1}{2}} (2s_n - n) \to -y \) as \( n \to \infty \). Also, by the DeMoivre-Laplace theorem, the right hand side of (3.4) tends to \( \Phi(-y) \) as \( n \to \infty \). Hence, (3.2) follows from (3.4). A similar proof applies to (3.3). Q.E.D.

**Remark.** By standard arguments [such as in Feller (1965, p. 230)], one could have approximated the random walk of section 2 by a Brownian movement process, and then used the well-known results on the maximum (or absolute maximum) displacement of such a process [viz., Parthasarathy (1967, pp. 224-230), particularly, corollaries 5.1 and 5.2] to give alternative derivation for the proofs of (3.2) and (3.3). For simplicity of presentation, we do not consider this approach.

Let us now define, for every \( \varepsilon : 0 < \varepsilon < \frac{1}{2} \),
\( (3.5) \quad \rho(\varepsilon) = (1 + 2\varepsilon)^{-((\frac{1}{2} + \varepsilon))} (1 - 2\varepsilon)^{-((\frac{1}{2} - \varepsilon))}; \quad \rho(\varepsilon) = 0 \text{ for } \varepsilon > \frac{1}{2}. \)

It is then easy to verify that \( \rho(\varepsilon) \) is strictly + in \( \varepsilon: 0 < \varepsilon < \frac{1}{2} \), with \( \rho(0) = 1 \) and

\[ \lim_{\varepsilon \to 0^+} \rho(\varepsilon) = \frac{1}{2}. \]

Hence for any \( \lambda > 1 \)

\( (3.6) \quad \frac{\rho(\lambda \varepsilon)}{\rho(\varepsilon)} < 1, \text{ for all } 0 < \varepsilon \leq \frac{1}{2} \lambda. \)

**Theorem 3.2.** Under \( H_0 \), for every \( \varepsilon: 0 < \varepsilon < 1 \),

\( (3.7) \quad P(D_{n \varepsilon}^+ \geq \varepsilon) = P(D_{n \varepsilon}^- \geq \varepsilon) \leq 2[\rho(\varepsilon/2)]^n, \)

\( (3.8) \quad \lim_{n \to \infty} [n^{-1} \log P(D_{n \varepsilon}^+)] = \log \rho(\varepsilon/2); \)

\( (3.9) \quad P(D_{n \varepsilon}^+ \leq 4[\rho(\varepsilon/2)]^n, \text{ and } \lim_{n \to \infty} [n^{-1} \log P(D_{n \varepsilon}^-)] = \log \rho(\varepsilon/2). \)

**Proof.** By (2.2) and (3.4), \( P(D_{n \varepsilon}^+ \geq \varepsilon) = P(D_{n \varepsilon}^- \geq \varepsilon) \leq 2P(r_{n \varepsilon} < s^*_n), \) where \( s^*_n = \lfloor n\varepsilon \rfloor \).

Since, \( r_n \) is a sum of independent and bounded valued random variables, (3.7) follows from the theorem 1 of Hoeffding (1963), and (3.8) follows from lemma 1 of Abrahamson (1967), attributed to Bahadur and Rao (1960). Also, noting that for every \( \varepsilon > 0 \) and \( n \geq 1 \),

\( (3.10) \quad P(D_{n \varepsilon}^+ \leq P(D_{n \varepsilon}^- \leq P(D_{n \varepsilon}^+ \varepsilon) + P(D_{n \varepsilon}^-), \quad \)

(3.9) follows readily from (3.7) and (3.8). Q.E.D.

Let us now consider the non-null case. To simplify the expressions, we assume the homogeneity of the cdf's, viz., \( F_1 = \ldots = F_n = F \), for all \( n \geq 1 \). For a cdf \( F(x) \), we define for \( x > 0 \) and \( \varepsilon > 0 \),

\( (3.11) \quad \rho_1(x, \varepsilon) = \inf_{t > 0} \{ e^{-t \varepsilon} [F(-x) - \varepsilon] + [F(x) - F(-x)] - [1 - F(x)] e^{-t} \}, \)

\( (3.12) \quad \rho_2(x, \varepsilon) = \inf_{t > 0} \{ e^{-t \varepsilon} [F(-x) - \varepsilon] + [F(x) - F(-x)] + [1 - F(x)] e^{-t} \}; \)
(3.13) \[ \rho^* (x, \epsilon) = \max [\rho_1 (x, \epsilon), \rho_2 (x, \epsilon)] \text{ and } \rho^* (F, \epsilon) = \sup_{x \geq 0} \rho^* (x, \epsilon). \]

Note that when \( F(x) \epsilon \mathcal{D}_0 \), \( \rho_1 (x, \epsilon) = \rho_2 (x, \epsilon). \)

**Theorem 3.3.** For every continuous \( F \) and every \( \epsilon > 0 \)

\[
(3.14) \lim_{n \to \infty} [n^{-1} \log P(D_{n} \geq \epsilon)] = \log \rho^* (F, \epsilon).
\]

**Outline of the proof.** By definition in (1.6), for every \( n \) and \( \epsilon > 0 \)

\[
(3.15) P(D_{n} \geq \epsilon) \geq P\{ |F^* (x) + \frac{F^* (-x)}{n} - 1| \geq \epsilon \} \text{ for any } x \geq 0.
\]

Since, \( F^* (x) + \frac{F^* (-x)}{n} - 1 = n^{-1} \sum_{i=1}^{n} g(x_i) \), where \( g(u) = 1, 0 \) or \(-1\) according as \( u < x, -x \leq u < x \) or \( u > x \), and as \( g(X_1), \ldots, g(X_n) \) are all independent and bounded valued random variables, using the well-known results of Bahadur and Rao (1960) (see also lemma 1 of Abrahamson (1967)), we obtain by some simple steps that

\[
(3.16) \lim_{n \to \infty} [n^{-1} \log P\{ |F^* (x) + \frac{F^* (-x)}{n} - 1| \geq \epsilon \}] = \log \rho^* (x, \epsilon), \ x \geq 0.
\]

Thus, by (3.14), (3.15) and (3.16), we have

\[
(3.17) \liminf_{n} [n^{-1} \log P(D_{n} \geq \epsilon)] \geq \sup_{x \geq 0} \log \rho^* (x, \epsilon) = \log \rho^* (F, \epsilon).
\]

Hence, the proof of (3.14) will follow, if we can show that

\[
(3.18) \limsup_{n} [n^{-1} \log P(D_{n} \geq \epsilon)] \leq \log \rho^* (F, \epsilon).
\]

Since the proof of (3.18) follows by the same technique as in theorem 1 of Abrahamson (1967) [namely, as in her (3.12)-(3.15)], we omit the details and terminate the proof here. Hence the theorem.

4. **Exact Bahadur-efficiencies for \( D_n \) and the sign statistics.** Following Abrahamson (1967), we briefly sketch the Bahadur (1960) efficiency of two sequences of statistics, when, in particular, we are interested in the hypothesis
of symmetry. Here also we assume that $F_1 = \ldots = F_n \equiv F$ for all $n \geq 1$. We define $F_0$ as in Section 1, and let $F_1$ be the class of all continuous (univariate) df's, not symmetric about zero. Thus, if we let

\begin{equation}
\delta(F) = \sup_{x > 0} |F(x) + F(-x) - 1|,
\end{equation}

then $\delta(F) = 0, \forall F \in F_0$, while $\delta(F) > 0$, for any $F \in F_1$.

Consider now two sequences $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ of non-negative real valued statistics, satisfying the following three conditions:

1. there exists a non-degenerate and continuous df $\Psi_i(x)$, such that for all $F \in F_0$ and real $r(0 < r < \infty)$,

\begin{equation}
\lim_{n \to \infty} P_F \{T_n^{(i)} < r \} = \Psi_i(r),
\end{equation}

2. there exists a non-negative function $\lambda_i$ on $[0, \infty]$ such that (i) $\lambda_i(z) > 0$ for all $z \in (0, \infty)$, and (ii) whenever $\{u_n\}$ is a sequence of real numbers for which $n^{-1} u_n^2 + z \in (0, \infty)$, we have

\begin{equation}
- \lim_{n \to \infty} (2/n) \log P_F \{T_n^{(i)} > u_n \} = \lambda_i(z),
\end{equation}

uniformly in $F \in F_0$, and for $F \in F_1$,

\begin{equation}
(3) \quad n^{-1} T_n^{(1)} \to b_1(F)(>0) \text{ a.s., as } n \to \infty, \text{ for } i=1,2.
\end{equation}

Then, we define the exact asymptotic efficiency of $T_n^{(1)}$ with respect to $T_n^{(2)}$ as equal to

\begin{equation}
e_{1,2}^{(1)}(F) = \lambda_1(b_1^2(F))/\lambda_2(b_2^2(F)),
\end{equation}

and with the metric $\delta(F)$, defined by (4.1), the limit

\begin{equation}
e_{1,2}^{(2)}(F) = \lim_{\delta(F) \to 0} e_{1,2}^{(1)}(F)
\end{equation}
is defined the exact asymptotic limiting efficiency, both defined after Bahadur (1960), as further interpreted in Abrahamson (1967).

Let now \( T_n^{(1)} = n^{-\frac{k}{2}} D_n \). Under \( H_0: F \in F_0 \), the distribution of \( T_n^{(1)} \) is independent of \( F \), and by (3.3), we have

\[
\psi_1(r) = 1 - 4 \sum_{k=1}^{\infty} (-1)^{k-1} \phi(-(2k-1)r), \quad 0 < r < \infty, \forall F \in F_0.
\]

Further, using theorem 3.2, (3.5) and some standard computations we obtain that for \( \{u_n\} \) for which \( u_n^{2/n} \rightarrow \epsilon(0,1) \),

\[
- \lim_{n \to \infty} (2/n) \log P\{ T_n^{(1)} \geq u_n \} = \sum_{k=1}^{\infty} \frac{z^k}{k(2k-1)}, \forall F \in F_0.
\]

Finally, by the Glivenko-Cantelli theorem, \( \lim_{n \to \infty} \sup_{x} |F_n(x) - F(x)| = 0 \), a.s., and hence, by (1.6) and (4.1),

\[
n^{-\frac{k}{2}} T_n^{(1)} = D_n \rightarrow F, \quad \text{a.s., as } n \to \infty.
\]

So for \( D_n \), all the three conditions are satisfied.

Let us now consider the sign statistic \( S_n \), defined by

\[
S_n = n^{-\frac{k}{2}} (2r - n); \quad r_n = \sum_{i=1}^{n} c(X_i),
\]

where \( c(u) \) is defined after (1.1). If we then let \( T_n^{(2)} = |S_n| \), we have

\[
\psi_2(r) = \phi(r) - \phi(-r), \quad 0 < r < \infty, \forall F \in F_0.
\]

Also, using lemma 1 of Abrahamsom (1967) and some standard computations, we have, parallel to (4.8),

\[
- \lim_{n \to \infty} (2/n) \log P\{ T_n^{(2)} \geq u_n \} = \sum_{k=1}^{\infty} \frac{z^k}{k(2k-1)}, \forall F \in F_0.
\]

Finally, it is well-known that as \( n \to \infty \),
\begin{equation}
\frac{n^{-k_n}(2)}{n} = n^{-1}(2r_n - n) \rightarrow \delta_o(F) = 2F(0) - 1, \text{ a.s.}
\end{equation}

Hence, the conditions are also satisfied for the sign statistic. Thus, the asymptotic efficiencies of $D_n$ with respect to $S_n$, as defined by (4.5) and (4.6), are equal to

\begin{equation}
e^{(1)}(F) = \frac{\sum_{k=1}^{\infty} [\delta(F)]^{2k/k(2k-1)}}{\sum_{k=1}^{\infty} [\delta_o(F)]^{2k/k(2k-1)}},
\end{equation}

\begin{equation}
e^{(2)}(F) = \lim_{\delta(F) \rightarrow 0} \frac{\delta(F)}{\delta_o(F)}.
\end{equation}

Now, note that by (4.1) and (4.13), $\delta(F) \geq \delta_o(F), \forall F \in \mathcal{F}_o \cup \mathcal{F}_1$. Hence, from (4.14) and (4.15), we arrive at the following:

\begin{equation}
e^{(1)}(F) \geq e^{(2)}(F) \geq 1, \text{ for all } F.
\end{equation}

Thus, the proposed test is at least as efficient (asymptotically) as the sign-test for all $F$. In particular, if $F(x) (\in \mathcal{F}_o)$ is symmetric and unimodal, and we are interested only in shift alternatives, then $\delta(F) = \delta_o(F)$, so that in (4.16), the equality signs hold; the conclusion is not necessarily true when $F(x)$ is not strictly unimodal [viz., the uniform df]. On the other hand, for certain specific type of asymmetry (of $F$), $\delta_o(F)$ may be exactly or nearly equal to zero, but $\delta(F)$ can still be positive, making (4.14) or (4.15) either $\infty$ or indefinitely large.

For other tests for symmetry, the Bahadur efficiency of $D_n$ may be computed in a similar way; for brevity the details are omitted.

REFERENCES


