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_o) 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_o as

(1.1) $$H_o: \; F_i(x)+F_i(-x)=1, \; \forall \; x\geq 0, \text{ and } i=1,\ldots,n.$$

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

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

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

(1.3) $$H_1: \; \sup_{x\geq 0}[\bar{F}_{(n)}(x)+\bar{F}_{(n)}(-x)]>1, \quad H_2: \; \inf_{x\geq 0}[\bar{F}_{(n)}(x)+\bar{F}_{(n)}(-x)]<1;$$

(1.4) $$H_3\equiv H_1 \cup H_2: \; \sup_{x\geq 0}|\bar{F}_{(n)}(x)+\bar{F}_{(n)}(-x)-1|>0.$$

---

* 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 $F_i(x) = F_i^o([x-\mu_i]/\sigma_i)$, and $F_i^o \varepsilon\, \mathfrak{F}_o$, $i=1,\ldots,n$, where $\mathfrak{F}_o = \{F: F(x)+F(-x) = 1 \text{ a.e.}\}$, and the μ_i and σ_i are the location and scale parameters. If the μ_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 appeals are evident from (1.3) and (1.4):

$$D_n^+ = \sup_{x\geq 0}[F_n^*(x)+F_n^*(-x-)-1],\quad D_n^- = \sup_{x\geq 0}[1-F_n^*(x)-F_n^*(-x-)]; \tag{1.5}$$

$$D_n = \max[D_n^+,D_n^-] = \sup_{x\geq 0}|F_n^*(x)+F_n^*(-x-)-1|. \tag{1.6}$$

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, $i=1,\ldots,n$, the process $\{n[F_n^*(x)+F_n^*(-x-)-1]: x\geq 0\}$ 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 $0,1,\ldots,n$, but not all of these are admissible. We denote by $\underset{\sim}{F}_n = (F_1,\ldots,F_n)$, and

$$(2.1)\qquad \mathfrak{F}_n^o = \{\underset{\sim}{F}_n: F_i \varepsilon \mathfrak{F}_o,\ i=1,\ldots,n\}.$$

Then, we have the following theorem.

Theorem 2.1. For every $\underset{\sim}{F}_n \varepsilon \mathfrak{F}_n^o$ and $k=1,\ldots,n$,

$$(2.2)\qquad P\{nD_n^+ \geq k\} = P\{nD_n^- \geq k\} = 2[\textstyle\sum_{i=0}^{s}\binom{n}{i}2^{-n}] - \delta_k\binom{n}{s}2^{-n},$$

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 $n-k$ = odd or even, and

$$(2.3)\qquad P\{nD_n \geq k\} = \begin{cases} 1, & k=0,1, \\ 2\sum_{j=0}^{u}(-1)^j\, P\{nD_n^+ \geq (2j+1)k\}, & k>1, \end{cases}$$

where $u = [n/2k]-1$, $k=1,\ldots,n$.

Proof. Let $Y_1 \geq \ldots \geq Y_n$ be the ordered values of $|X_1|,\ldots,|X_n|$, arranged in descending order of magnitude. Let $t_{n,i} = \bar{F}_{(n)}(-Y_i)$, for $1\leq i\leq n$, so that $0\leq t_{n,1}\leq t_{n,2}\leq\ldots\leq t_{n,n}\leq \bar{F}_{(n)}(0) = \frac{1}{2}$ (as $\underset{\sim}{F}_n \varepsilon \mathfrak{F}_n^o \Rightarrow \bar{F}_{(n)} \varepsilon \mathfrak{F}_o$). Since, $F_1,\ldots,F_n$ are symmetric and continuous, ties among $|X_1|,\ldots,|X_n|$, and hence, among $t_{n,1},\ldots,t_{n,n}$ can be neglected in probability. Thus,

$$(2.4)\qquad 0<t_{n,1}<t_{n,2}<\ldots<t_{n,n}<\tfrac{1}{2}, \text{ in probability.}$$

Define then $V_n(t) = n^{\frac{1}{2}}[G_n^*(t)-t]$, $0<t<1$, where $G_n^*(t) = n^{-1}\sum_{i=1}^{n} c(t-\bar{F}_{(n)}(X_i))$, and let

$$(2.5)\qquad V_n^*(t) = V_n(t-)+V_n(1-t),\quad 0\leq t\leq\tfrac{1}{2}.$$

For $t\leq t_{n,1}$, $n^{\frac{1}{2}}V_n^*(t)=0$. At $t=t_{n,1}+$, $n^{\frac{1}{2}}V_n^*(t)$ is either +1 or -1, depending upon whether the random variable X_i associated with Y_n has negative or positive sign. The process $n^{\frac{1}{2}}V_n^*(t)$ 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_i associated with Y_{n-1} is negative or not. And thus the process continues. Hence, on $I=(0,\frac{1}{2})$, $n^{\frac{1}{2}}V_n^*(t)$ 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 $\underset{\sim}{F}_n\in\mathfrak{F}_n^o$, the df of X_i is symmetric about 0, $1\leq i\leq n$,

(2.6) $$P\{Y_{n-i+1} \text{ corresponds to a positive } X_j\}$$
$$= \sum_{j=1}^n p_{ij}\cdot P\{X_j>0\,|\,|X_j| = Y_{n-i+1}\} = \tfrac{1}{2}\sum_{j=1}^n p_{ij} = \tfrac{1}{2},$$

as the distribution of sign X_i is independent of $|X_i|$ when $F_i\in\mathfrak{F}_o$, $i=1,\ldots,n$. Thus, the jumps (+1 or -1) at $t_{n,i}$ are both equally likely with probability ½. Moreover, for $\underset{\sim}{F}_n\in\mathfrak{F}_n^o$, the vector (Sign $X_1,\ldots,$Sign X_n) is distributed independently of $(|X_1|,\ldots,|X_n|)$ and Sign $X_1,\ldots,$Sign X_n are also mutually stochastically independent. Hence, the jumps of $n^{\frac{1}{2}}V_n^*(t)$ at $t_{n,1},\ldots,t_{n,n}$ are mutually independent. Finally, the values of nD_n^+ $(=\sup_{t\in I} n^{\frac{1}{2}}V_n^*(t))$, nD_n^- $(=\sup_{t\in I}[-n^{\frac{1}{2}}V_n^*(t)])$ and nD_n $(=\sup_{t\in I}|n^{\frac{1}{2}}V_n^*(t)|)$ are independent of the particular realization of $\underset{\sim}{t}_n=(t_{n,1},\ldots,t_{n,n})\in I$. Hence, we conclude that (i) the distribution of nD_n^+ (or nD_n^-) (under H_o) 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 nD_n 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 nD_n^+ and nD_n^- are both non-negative, and hence, $P\{nD_n^+\geq 0\} = P\{nD_n^-\geq 0\} = 1$. Also, at $t_{n,1}$, $n^{\frac{1}{2}}V_n^*(t)$ is either +1 or -1. Hence, $nD_n\geq 1$, with probability one. So, to prove (2.2), we consider $k\geq 1$, and for (2.3), $k\geq 2$.

We now use theorem 1 (Section 8) of Takacs (1967, p. 24), and obtain that for $k \geq 1$,

$$(2.7)\qquad P\{nD_n^+ \geq k\} = P\{nD_n^- \geq k\} = \sum_{j=k}^{n} \frac{k}{j} P\{N_j = j-k\} ,$$

where

$$(2.8)\qquad P\{N_j = j-k\} = \begin{cases} 2^{-j}\binom{j}{r}, & j-k=2r,\ r=0,1,2,\ldots, \\ 0, & j-k = 2r+1;\quad j \geq k \geq 1. \end{cases}$$

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

$$(2.9)\qquad 2^{-(n-1)}\sum_{t=0}^{s}\binom{n}{t} - \delta_k \binom{n}{s} 2^{-n},$$

where s and δ_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

$$(2.10)\qquad P\{nD_n \leq k'\} = 1-2Q(k'+1,n);$$

$$(2.11)\qquad P\{nD_n^+ \leq k'\} = 1-Q^+(k'+1,n).$$

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

$$(2.12)\qquad Q(k'+1,n) = Q^+(k'+1,n) - Q^+(3k'+3,n) + Q^+(5k'+5,n) - \ldots + (-1)^u Q^+((2u+1)k',n);\ u = [(n/2k')-1].$$

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 $P\{nD_n^+ \geq k\} = P\{nD_n^- \geq k\}$ for $k \leq n \leq 16$

k→ / n	0	1	2	3	4	5	6	7	8	9	10*	11	12	13	14	15
2	1	.500	.250													
3	1	.625	.250	.125												
4	1	.625	.375	.125	.063											
5	1	.688	.375	.219	.063	.031										
6	1	.688	.453	.219	.125	.031	.016									
7	1	.727	.453	.289	.125	.072	.016	.008								
8	1	.727	.508	.289	.235	.072	.039	.008	.004							
9	1	.752	.508	.344	.235	.111	.039	.022	.004	.002						
10	1	.752	.549	.344	.282	.111	.065	.022	.012	.002	.0010					
11	1	.772	.549	.388	.282	.147	.065	.039	.012	.006	.0010	.0005				
12	1	.772	.581	.388	.317	.147	.092	.039	.023	.006	.0034	.0005	.0002			
13	1	.789	.581	.423	.317	.181	.092	.057	.023	.013	.0034	.0018	.0002	.0001		
14	1	.789	.607	.423	.329	.181	.119	.057	.035	.013	.0074	.0018	.0009	.0001	.00005	
15	1	.803	.607	.436	.329	.211	.119	.077	.035	.019	.0074	.0032	.0009	.0004	.00005	.00003
16	1	.803	.629	.436	.340	.211	.143	.077	.049	.019	.0106	.0032	.0023	.0004	.00026	.00003

* Values are correct to 4 decimal places for $k \geq 10$, and three decimal places for $k \leq 9$.

TABLE 2

Table for the values of $P\{nD_n \geq k\}$ for $1 \leq k \leq n \leq 16$

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

* Correct to 4 decimal places for $k \geq 10$ and up to three decimal places for $k \leq 9$.

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

$$\Phi(y) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} [\exp(-\tfrac{1}{2}t^2)]dt, \quad -\infty<y<\infty. \tag{3.1}$$

Then, we have the following theorem.

Theorem 3.1. For every fixed $y(0<y<\infty)$, under H_o (i.e., $\forall\, \underset{\sim}{F}_n \varepsilon \mathcal{F}_n^o$),

$$\lim_{n\to\infty} P\{n^{\frac{1}{2}}D_n^+ \geq y\} = \lim_{n\to\infty} P\{n^{\frac{1}{2}}D_n^- \geq y\} = 2\Phi(-y); \tag{3.2}$$

$$\lim_{n\to\infty} P\{n^{\frac{1}{2}}D_n \geq y\} = 4[\textstyle\sum_{k=1}^{\infty}(-1)^{k-1}\Phi(-(2k-1)y)]. \tag{3.3}$$

Proof. Let r_n be the number of successes in n independent Bernoullian trials with probability ½. Then, by (2.2),

$$\begin{aligned} P\{n^{\frac{1}{2}}D_n^+ \geq y\} &= P\{n^{\frac{1}{2}}D_n^- \geq y\} = 2P\{r_n \leq s_n\} - \delta_k P\{r_n = s_n\} \\ &= 2P\{n^{-\frac{1}{2}}(2r_n - n) \leq n^{-\frac{1}{2}}(2s_n - n)\} - \delta_k P\{r_n - s_n\}, \end{aligned} \tag{3.4}$$

where $s_n = [n/2 - \frac{1}{2}n^{\frac{1}{2}}y]$, 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 ε: $0<\varepsilon<\frac{1}{2}$,

$$(3.5)\qquad \rho(\varepsilon) = (1+2\varepsilon)^{-(\frac{1}{2}+\varepsilon)}(1-2\varepsilon)^{-(\frac{1}{2}-\varepsilon)};\ \rho(\varepsilon) = 0 \text{ for } \varepsilon \geq \tfrac{1}{2}.$$

It is then easy to verify that $\rho(\varepsilon)$ is strictly $\downarrow$ in ε: $0<\varepsilon<\frac{1}{2}$, with $\rho(0) = 1$ and $\lim_{\varepsilon\to\frac{1}{2}}\rho(\varepsilon) = \frac{1}{2}$. Hence for any $\lambda>1$

$$(3.6)\qquad \rho(\lambda\varepsilon)/\rho(\varepsilon) < 1, \text{ for all } 0<\varepsilon\leq\tfrac{1}{2}\lambda\ .$$

Theorem 3.2. Under H_o, for every ε: $0<\varepsilon<1$,

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

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

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

Proof. By (2.2) and (3.4), $P\{D_n^+\geq\varepsilon\} = P\{D_n^-\geq\varepsilon\} \leq 2P\{r_n\leq s_n^*\}$, where $s_n^* = [\frac{1}{2}n(1-\varepsilon)]$ 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 $n\geq1$,

$$(3.10)\qquad P\{D_n^+\geq\varepsilon\} \leq P\{D_n\geq\varepsilon\} \leq P\{D_n^+\geq\varepsilon\} + P\{D_n^-\geq\varepsilon\},$$

(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 $n\geq1$. For a cdf $F(x)$, we define for $x\geq0$ and $\varepsilon>0$,

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

$$(3.12)\qquad \rho_2(x,\varepsilon) = \inf_{t>0}\{e^{-t\varepsilon}[F(-x)e^{-t}+\{F(x)-F(-x)\}+\{1-F(x)\}e^{t}]\};$$

$$(3.13)\qquad \rho^*(x,\varepsilon) = \max[\rho_1(x,\varepsilon),\rho_2(x,\varepsilon)] \text{ and } \rho^*(F,\varepsilon) = \sup_{x\geq 0}\rho^*(x,\varepsilon).$$

Note that when $F(x)\varepsilon\, \mathfrak{F}_o$, $\rho_1(x,\varepsilon) = \rho_2(x,\varepsilon)$.

Theorem 3.3. For every continuous F and every ε: $0<\varepsilon<1$,

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

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

$$(3.15)\qquad P\{D_n\geq\varepsilon\} \geq P\{|F_n^*(x)+F_n^*(-x-)-1| \geq \varepsilon\} \text{ for any } x\geq 0.$$

Since, $F_n^*(x)+F_n^*(-x-)-1 = n^{-1}\sum_{i=1}^n g(x_i)$, where $g(u) = 1$, 0 or -1 according as $u<-x$, $-x\leq u\leq x$ 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

$$(3.16)\qquad \lim_{n\to\infty}[n^{-1}\log P\{|F_n^*(x)+F_n^*(-x-)-1|\geq\varepsilon\}] = \log \rho^*(x,\varepsilon),\ x\geq 0.$$

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

$$(3.17)\qquad \liminf_n\ [n^{-1}\log P\{D_n\geq\varepsilon\}] \geq \sup_{x\geq 0} \log \rho^*(x,\varepsilon) = \log \rho^*(F,\varepsilon).$$

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

$$(3.18)\qquad \limsup_n [n^{-1}\log P\{D_n\geq\varepsilon\}] \leq \log \rho^*(F,\varepsilon).$$

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 Abrahamsom (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 $F_1=\ldots=F_n=F$ for all $n\geq 1$. We define $\mathcal{F}_o$ as in Section 1, and let $\mathcal{F}_1$ be the class of all continuous (univariate) df's, not symmetric about zero. Thus, if we let

$$\delta(F) = \sup_{x\geq 0}|F(x)+F(-x)-1|, \tag{4.1}$$

then $\delta(F) = 0$, $\forall F\varepsilon\mathcal{F}_o$, while $\delta(F)>0$, for any $F\varepsilon\,\mathcal{F}_1$.

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 $F\,\varepsilon\mathcal{F}_o$ and real $r(0<r<\infty)$,

$$\lim_{n\to\infty} P_F\{T_n^{(i)}<r\} = \Psi_i(r), \tag{4.2}$$

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

$$-\lim_{n\to\infty}(2/n)\log P_F\{T_n^{(i)}\geq u_n\} = \ell_i(z), \tag{4.3}$$

uniformly in $F\varepsilon\mathcal{F}_o$, and for $F\varepsilon\mathcal{F}_1$,

$$\text{(3)}\quad n^{-\frac{1}{2}}T_n^{(i)} \to b_i(F)(>0) \text{ a.s., as } n\to\infty, \text{ for } i=1,2. \tag{4.4}$$

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

$$e_{1,2}^{(1)}(F) = \ell_1(b_1^2(F))/\ell_2(b_2^2(F)), \tag{4.5}$$

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

$$e_{1,2}^{(2)}(F) = \lim_{\delta(F)\to 0} e_{1,2}^{(1)}(F) \tag{4.6}$$

is defined the exact asymptotic limiting efficiency, both defined after Bahadur (1960), as further interpreted in Abrahamson (1967).

Let now $T_n^{(1)} = n^{\frac{1}{2}}D_n$. Under H_o: $F\varepsilon\mathfrak{F}_o$, the distribution of $T_n^{(1)}$ is independent of F, and by (3.3), we have

(4.7) $$\Psi_1(r) = 1-4\sum_{k=1}^{\infty}(-1)^{k-1}\Phi(-(2k-1)r),\ 0<r<\infty, \forall F\varepsilon\mathfrak{F}_o.$$

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

(4.8) $$-\lim_{n\to\infty}(2/n)\log P\{T_n^{(1)}\geq u_n\} = \sum_{k=1}^{\infty}z^k/k(2k-1), \forall F\varepsilon\mathfrak{F}_o.$$

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),

(4.9) $$n^{-\frac{1}{2}}T_n^{(1)} = D_n\to\delta(F),\ \text{a.s., as } n\to\infty.$$

So for D_n, all the three conditions are satisfied.

Let us now consider the sign statistic S_n, defined by

(4.10) $$S_n = n^{-\frac{1}{2}}(2r_n-n);\quad r_n = \sum_{i=1}^{n}c(X_i),$$

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

(4.11) $$\Psi_2(r) = \Phi(r)-\Phi(-r),\ 0\leq r<\infty, \forall F\varepsilon\mathfrak{F}_o.$$

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

(4.12) $$-\lim_{n\to\infty}(2/n)\log P\{T_n^{(2)}\geq u_n\} = \sum_{k=1}^{\infty}z^k/k(2k-1), \forall F\varepsilon\mathfrak{F}_o.$$

Finally, it is well-known that as $n\to\infty$, (4.13) $$n^{-\frac{1}{2}}T_n^{(2)} = n^{-1}(2r_n - n) \to \delta_o(F) = 2F(0)-1, \text{ a.s..}$$

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

(4.14) $$e^{(1)}(F) = [\sum_{k=1}^{\infty}\{\delta(F)\}^{2k}/k(2k-1)]/[\sum_{k=1}^{\infty}\{\delta_o(F)\}^{2k}/k(2k-1)],$$

(4.15) $$e^{(2)}(F) = \lim_{\delta(F)\to 0}\{\delta(F)/\delta_o(F)\}^2.$$

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:

(4.16) $$e^{(1)} \geq e^{(2)}(F) \geq 1, \text{ for all } F.$$

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 ∞ 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

[1] ABRAHAMSON, I. G. (1967). Exact Bahadur efficiencies for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics. Ann. Math. Statist. 38, 1475-1490.

[2] BAHADUR, R. R. (1960). Stochastic comparison of tests. Ann. Math. Statist. 31, 276-295.

[3] BAHADUR, R. R., AND RANGA RAO, R. (1960). On deviations of the sample mean. Ann. Math. Statist. 31, 1015-1027.

[4] FELLER, W. (1967). An introduction to probability theory and its applications, vol. 2. John Wiley: New York.

[5] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. Jour. Amer. Statist. Assoc. 58, 13-30.

[6] PARTHASARATHY, K. R. (1967). Probability measures on metric spaces. Academic Press: New York.

[7] TAKACS, L. (1967). Combinatorial methods in the theory of stochastic processes. John Wiley: New York.

[8] USPENSKY, J. V. (1937). Introduction to mathematical probability. McGraw-Hill: New York.