AN ASYMPTOTICALLY NONPARAMETRIC TEST OF SYMMETRY[1]

By M. K. Gupta

Presidency College, Calcutta

1. Introduction and assumptions. Let $X_1, X_2, \cdots, X_N$ be N independently identically distributed random variables with the cumulative distribution function $F(x - \theta)$, where θ may for example be the median of the distribution.

A problem which has been considered in the literature is the testing of the hypothesis that the distribution is symmetric about a specified value against the alternative that it is shifted to the right (or to the left) or against two-sided alternatives. A different problem which has also received attention is to test whether the distribution is symmetric about a specified value against the alternative that either the symmetry is lost or the location parameter is changed.

Our problem differs from the above problems in that we do not make any assumption about the knowledge of the location of the distribution. More specifically the problem considered here is that of testing the hypothesis that the distribution is symmetric about θ against the alternative that it is more spread out to the right of θ than it is to the left. Alternatively we may say that the alternative of interest is that the distribution of the positive deviations from θ is stochastically larger than the distribution of the absolute values of the negative deviations from θ.

This type of problem may arise in the following ways: On a symmetric distribution may be superimposed another distribution (say, of errors) which destroys the symmetry of the original distribution. We want to test whether we have a sample from the original symmetric distribution or from the resultant skewed distribution. The problem of testing for symmetry may also arise without a specific model for the alternative.

In Section 2 we review a classical test proposed for this problem under the assumption of normality. It is shown that the test is also applicable without this assumption provided the underlying distribution has finite moments up to order six. In Section 3 we give some definitions and state the known results which we are going to use later. In Section 4 we consider a test statistic for testing the symmetry of F when θ, the location parameter, is known. In Section 5 we consider the asymptotic relative efficiency (ARE) of the two test statistics, the classical test and the proposed test, under different classes of alternatives. Section 6 gives the modified test statistic for the case of unknown location parameter. The test in the case of unknown location is not asymptotically distribution-free. In Section 8 we provide a consistent estimator of the asymptotic null variance

Received 3 June 1965; revised 28 December 1966.

[1] This paper is based on the author's dissertation submitted in partial satisfaction of the requirements of the Ph.D. degree in Statistics at the University of California, Berkeley. This research was partially supported by the National Institute of Health, Public Health Service Research Grant No. GM-10525.

and then obtain an asymptotically distribution-free test by Studentization. In Section 7 we compare the two tests proposed in Sections 4 and 6 to get an idea as to the loss incurred for not knowing θ. We also consider the loss in the case of the classical test when θ is not known.

For later reference we classify the assumptions.

ASSUMPTION 1. $X_1, X_2, \cdots, X_N$ are independent, each has the same distribution, and

(a) X has the distribution function $F(x - \theta)$ over R^1 which is absolutely continuous relative to Lebesgue measure, with density function $f(x - \theta)$, where θ is the population median;

(b) the distribution F has a derivative $F' = f$ which is bounded;

(c) in a neighbourhood of the origin, $f(x)$ is continuous and $f(0) \neq 0$.

2. A classical test of symmetry. Any general test of fit may be employed to test for normality or to test whether a sample is from any specified symmetric distribution. For the normal case it is a common practice to use as a test statistic the observed moment ratio

$$b_1 = m_3/m_2^{\frac{3}{2}} \tag{2.1}$$

where m_k is the kth sample moment about the sample mean. This is known as the "test of skewness."

Many of the simpler methods of statistical analysis have been developed only for variables which are normally distributed. But in many cases it is important to obtain evidence on this point, that is to say, to apply some test for normality to the sample. This is done by means of the sample coefficient of skewness and kurtosis. Here we shall consider only the problem of skewness. There have long been available tables giving the standard error of b_1. The work of Craig and Fisher made possible a considerable advance toward the solution of the problem. They gave the sampling distributions of b_1 and b_2 (coefficient of kurtosis) if the population is normal. E. S. Pearson gave the first four moment coefficients of the sampling distribution of b_1 and b_2 in samples of N from a normal population as far as the terms in N^{-3}.

The test statistic b_1 could also be used for any other underlying distribution F, provided F has finite moments up to order six. The theorem below gives the asymptotic distribution of the test statistic b_1.

THEOREM 2.1. *Let $X_1, X_2, \cdots, X_N$ be independently identically distributed according to the non-degenerate* cdf $F(x - \theta)$. *Let m_k and μ_k denote, respectively, the sample and the population central moments of F and let b_1 be defined by* (2.1). *Let F have finite population moments up to the sixth order. Then*

$$\lim_{N\to\infty} \mathcal{L}(N^{\frac{1}{2}}(b_1 - \nu_1)/\tau) = \mathfrak{N}(0, 1)$$

where $\nu_1 = \mu_3/\mu_2^{\frac{3}{2}}$, and $E(b_1) = \nu_1$, $V(b_1) = \tau^2/N$ to order (1) */N. Explicitly, $V(b_1) = V(m_3)H_1^2 + 2\,\mathrm{cov}\,(m_3, m_2)H_1H_2 + V(m_2)H_2^2$, to order* $(1/N)$, *where*

$$H_1 = (\delta b_1/\delta m_3)_{m_3=\mu_3, m_2=\mu_2}; \qquad H_2 = (\delta b_1/\delta m_2)_{m_2=\mu_2, m_3=\mu_3};$$

$$V(m_3) = [\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3 - \mu_3^2]/N, \qquad \textit{to order } (1/N);$$

$$V(m_2) = [\mu_4 - 4\mu_1\mu_3 - \mu_2^2 + 4\mu_2\mu_1^2]/N, \qquad \textit{to order } (1/N);$$

$$\text{cov}\,(m_2, m_3) = [\mu_5 - 2\mu_1\mu_4 - 4\mu_3\mu_2 + 6\mu_1\mu_2^2]/N, \qquad \textit{to order } (1/N).$$

PROOF. A proof of this type of theorem is given in Cramér [1]. We observe that no assumption as to the form of the distribution has been made except for the existence of the first six population moments.

We next obtain the asymptotic expectation and the asymptotic variance of b_1 under the hypothesis of symmetry

$$(2.2) \qquad E_0(b_1) = 0; \qquad \sigma_0^2(b_1) = [\boldsymbol{\mu}_6 - 6\boldsymbol{\mu}_2\boldsymbol{\mu}_4 + 9\boldsymbol{\mu}_2^3]/N\boldsymbol{\mu}_2^3,$$

where $\boldsymbol{\mu}$ are the central moments of the null distribution.

We shall reject the hypothesis that F is symmetric in favor of the alternative that F is skewed to the right if b_1 is large.

The b_1-test is not asymptotically distribution-free but it can be made distribution-free by estimating $N\sigma_0^2(b_1)$ by the corresponding function of the sample moments and then performing an asymptotically normal test after Studentization.

The above form of the b_1-test is appropriate if θ is unknown. We modify b_1 in case θ is known and compute the following statistic

$$(2.3) \qquad b_{10} = [\textstyle\sum (X_i - \theta)^3/N]/[\sum (X_i - \theta)^2/N]^{\frac{3}{2}}$$

where θ is the known population median.

The following theorem gives the asymptotic distribution of the test statistic b_{10} in the case of known θ.

THEOREM 2.2. *Let $X_1, X_2, \cdots, X_N$ be independently identically distributed according to the non-degenerate* cdf $F(x - \theta)$. *Let b_{10} be defined by* (2.3) *where θ is the population median. Let F have finite population moments up to the sixth order. Then*

$$\lim_{N\to\infty} \mathfrak{L}(N^{\frac{1}{2}}(b_{10} - \nu_{10})/\tau_0) = \mathfrak{N}(0, 1)$$

where $\nu_{10} = E(X - \theta)^3/[E(X - \theta)^2]^{\frac{3}{2}}$ *and*

$$\begin{aligned} V(b_{10}) = \tau_0^2/N = {} & V(\textstyle\sum (X_i - \theta)^3/N)H_{10}^2 \\ & + V(\textstyle\sum (X_i - \theta)^2/N)H_{20}^2 \\ & + 2 \text{ cov } (\textstyle\sum (X_i - \theta)^3/N, \sum (X_i - \theta)^2/N)H_{10}H_{20}, \end{aligned}$$

to order $(1/N)$,

where

$$V[\textstyle\sum (X_i - \theta)^3/N] = \{E(X - \theta)^6 - [E(X - \theta)^3]^2\}/N,$$

$$V[\textstyle\sum (X_i - \theta)^2/N] = \{E(X - \theta)^4 - [E(X - \theta)^2]^2\}/N,$$

$$\text{cov}\,[\textstyle\sum (X_i - \theta)^3/N, \sum (X_i - \theta)^2/N]$$

$$= [E(X - \theta)^5 - E(X - \theta)^2E(X - \theta)^3]/N,$$

H_{10} and H_{20} are the derivatives with respect to a and b of $H(a, b) = a/b^{\frac{3}{2}}$ evaluated at $a = E(X - \theta)^3$ and $b = E(X - \theta)^2$.

PROOF. The proof follows by the Taylor's expansion of the statistic and an application of the central limit theorem.

We next obtain the expectation and asymptotic null variance of the statistic. The expectation is given by

$$(2.4)\quad E(b_{10}) = E(X - \theta)^3/[E(X - \theta)^2]^{\frac{3}{2}}$$
$$= [\mu_3 + 3\mu_2(\mu_1 - \theta) + (\mu_1 - \theta)^3]/[\mu_2 + (\mu_1 - \theta)^2]^{\frac{3}{2}}$$

where the μ are the central moments.

The null variance of b_{10} is given by

$$(2.5)\qquad \sigma_0^2(b_{10}) = \boldsymbol{\mu}_6/N\boldsymbol{\mu}_2^3,$$

where the $\boldsymbol{\mu}$ are the central moments of the null distribution. This test statistic is also not asymptotically distribution-free. We can make it asymptotically distribution-free by Studentization as in the case of the b_1-statistic.

3. Definitions and known results.

DEFINITION 3.1. The distribution $F(x)$ will be said to be symmetric about the origin if

$$F(x) + F(-x) = 1, \qquad \text{for all} \quad x.$$

DEFINITION 3.2. Let the distribution $F(x)$ have median at the origin. Then the distribution of the positive observations is stochastically larger than the distribution of the absolute values of the negative observations if

$$P(X \leqq x/X > 0) \leqq P(-X \leqq x/X < 0), \qquad \text{for } x > 0,$$

with strict inequality for some x.

With the above definitions we are interested in testing the following hypothesis:

$$(3.0)\quad H: F(x) + F(-x) = 1, \quad \text{all} \quad x,$$
$$H': P(X \leqq x/X > 0) \leqq P(-X \leqq x/X < 0), \quad \text{for} \quad x > 0,$$

with strict inequality for some x.

DEFINITION 3.3. Let $\phi(u_1, u_2)$ be real-valued and symmetric in its arguments. Then the statistic

$$U_N = \binom{N}{2}^{-1} \sum_{i<j} \phi(X_i, X_j)$$

is called a U-statistic.

If $E\phi^2(X_1, X_2) < \infty$, then Hoeffding [6] showed that $N^{\frac{1}{2}}(U_N - EU_N)$ is asymptotically normally distributed with mean zero and asymptotic variance $4\zeta_1$, where, $\zeta_1 = E\phi_1^2(X_1) - E^2\phi(X_1, X_2)$ and $\phi_1(x_1) = E\phi(x_1, X_2)$.

DEFINITION 3.4. (Sukhatme). Let $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_N)$ be an estimate of θ, the unknown location parameter in $F(x - \theta)$. Then the U-statistic with the observations centered at their location parameter and the modified U-statistic are, respectively,

$$U_N = \binom{N}{2}^{-1} \sum_{i<j} \phi(X_i - \theta, X_j - \theta),$$
$$\hat{U}_N = \binom{N}{2}^{-1} \sum_{i<j} \phi(X_i - \hat{\theta}, X_j - \hat{\theta}).$$

Finally we define a quantity L_N required in the study of the asymptotic behavior of the modified U-statistic.

DEFINITION 3.5 (Sukhatme).

$$L_N = \binom{N}{2}^{-1} \sum_{i<j} [\phi(X_i - \hat{\theta}, X_j - \hat{\theta}) - A(\hat{\theta} - \theta)]$$

where $A(t - \theta) = E\phi(X_1 - t, X_2 - t)$, the expectation being taken with respect to all the X's.

THEOREM 3.1. (Sukhatme). *Let $X_1, X_2, \cdots, X_N$ be N independent identically distributed random variables with* cdf $F(x - \theta)$. *Let $\phi(u_1, u_2)$ be a real-valued symmetric function of its arguments such that if*

$$W(x_1, x_2, t) = \phi(x_1 - t, x_2 - t) - A(t - \theta), \tag{3.1}$$

the following conditions are satisfied:

(A)

$$|W(x_1, x_2, t)| \leqq M_1;$$
$$E\,|W(X_1, X_2, t + h) - W(X_1, X_2, t)| \leqq M_2 h,$$

where M_1, M_2 are fixed constants.

(B) *There exists a sequence $\{t_j\}$ such that for each set of x*

$$\sup_{0 \leqq t_j \leqq k} |W(x_1, x_2, t_j) - W(x_1, x_2, 0)| = \sup_{0 \leqq t \leqq k} |W(x_1, x_2, t) - W(x_1, x_2, 0)|.$$

Further, let $\hat{\theta}$ be an estimate of θ such that given $\epsilon > 0$, there exists a number b such that for N sufficiently large

$$P\{|\hat{\theta} - \theta| \geqq b/N^{\frac{1}{2}}\} \leqq \epsilon. \tag{3.2}$$

Then

$$\lim_{N\to\infty} \mathcal{L}(N^{\frac{1}{2}} L_N) = \lim_{N\to\infty} \mathcal{L}(N^{\frac{1}{2}}(U_N - EU_N)) = \mathfrak{N}(0, \sigma^2)$$

where $\sigma^2 = 4 \operatorname{Var} \int \phi(X_1 - \theta, x - \theta)\, dF(x - \theta)$.

We write $\lim_{N\to\infty} \mathcal{L}(X_N) = \mathcal{L}(X)$ if $F_N(y) \to F(y)$ at every continuity point y of F when F_N and F are the cdf's of X_N and X, respectively.

PROOF. The proof of this theorem is given by Sukhatme [14].

THEOREM 3.2 (Sukhatme). *Suppose in addition to the conditions of Theorem* 3.1 *that*

(i) *$N^{\frac{1}{2}}(\hat{\theta} - \theta)$ has a limiting distribution,*

(ii) *$A(t) = E[\phi(X_1 - t, X_2 - t)/\theta = 0]$ has a derivative continuous in the neighbourhood of the origin.*

Then we have the following conclusions:

(a) *If* $A'(0) = 0$ *where* $A'(t) = dA(t)/dt$, *then*

$$\lim_{N\to\infty} \mathcal{L}(N^{\frac{1}{2}}(\hat{U}_N - EU_N)) = \lim_{N\to\infty} \mathcal{L}(N^{\frac{1}{2}}[U_N - EU_N]) = \mathfrak{N}(0, \sigma^2).$$

(b) *If* $A'(0) \neq 0$, $\hat{\theta}$ *is asymptotically normally distributed and the joint distribution of* $\hat{\theta}$ *and* U_N *is asymptotically normal, then* $N^{\frac{1}{2}}(\hat{U}_N - EU_N)$ *is asymptotically normally distributed.*

PROOF. The proof of this theorem is given by Sukhatme [14].

THEOREM 3.3 (Sukhatme). *Let* $X_1, X_2, \cdots, X_N$ *be a sample of* N *independent observations drawn from a population with* cdf $F(x - \theta)$ *and density function* $f(x - \theta)$. *Let* $N = 2n + 1$ *so that the sample median* $\hat{\theta}$ *is uniquely determined. We shall assume that in some neighbourhood of the origin,* $f(x)$ *is continuous and* $f(0) \neq 0$. *Let* $\phi(u_1, u_2)$ *be a real-valued, bounded and symmetric function of its arguments. Let*

$$U_N = \binom{N}{2}^{-1} \sum_{i<j} \phi(X_i - \theta, X_j - \theta).$$

Also let

$$m' = \int_{-\infty}^{\theta} [\phi_1(x_1 - \theta) - m] \, dF(x_1 - \theta),$$
$$m'' = \int_{\theta}^{\infty} [\phi_1(x_1 - \theta) - m] \, dF(x_1 - \theta),$$

where $\phi_1(x_1 - \theta) = E\phi(x_1 - \theta, X_2 - \theta)$, *expectations being taken with respect to* X_2 *only, and* $m = E\phi_1(X_1 - \theta) = EU_N$. *Then the joint distribution of* $N^{\frac{1}{2}}(U_N - EU_N)$ *and* $N^{\frac{1}{2}}(\hat{\theta} - \theta)$ *is asymptotically normal with zero means and asymptotic variances* σ^2 *and* $1/4f^2(0)$, *the asymptotic expression for the correlation coefficient being* $(m'' - m')/\sigma$.

PROOF. The proof of this theorem is given by Sukhatme [13].

THEOREM 3.4. *Let* $N^{\frac{1}{2}}(U_N - EU_N)$ *and* $N^{\frac{1}{2}}(\hat{\theta} - \theta)$ *be jointly asymptotically distributed as bivariate normal distribution with asymptotic means zero, asymptotic variances* σ^2 *and* $1/4f^2(0)$ *respectively, and the asymptotic expression for the correlation coefficient being* $(m'' - m')/\sigma$. *Further, let* $N^{\frac{1}{2}}[\hat{U}_N - A(\hat{\theta} - \theta)] - N^{\frac{1}{2}}(U_N - EU_N) \to 0$ *in probability and* $A'(h) \to A'(0)$ *in probability, where* $A'(0) \neq 0$. *Then* $N^{\frac{1}{2}}[\hat{U}_N - A(\hat{\theta} - \theta)]$ *and* $N^{\frac{1}{2}}(\hat{\theta} - \theta)A'(h)$ *has a joint asymptotic normal distribution with means zero, asymptotic variances* σ^2 *and* $[A'(0)]^2/4f^2(0)$ *and the asymptotic expression for the correlation coefficient being* $(m'' - m')/\sigma$.

PROOF. The proof of this theorem follows exactly in the same way as the proof of the corresponding univariate case [1].

4. The test statistic J. We shall now develop an alternative test to the b_{10}-test which in some cases has a higher efficiency than the b_{10}-test. We shall consider the test of the hypothesis defined in (3.0) in the case of known θ.

Let us define the following test statistic

$$J = J_N = \binom{N}{2}^{-1} \sum_{i<j} \phi(X_i - \theta, X_j - \theta) \tag{4.1}$$

where

$$\phi(X_i - \theta, X_j - \theta) = 1 \quad \text{if either } X_i < \theta < X_j \text{ and } \theta - X_i < X_j - \theta$$
$$\text{or } X_j < \theta < X_i \text{ and } \theta - X_j < X_i - \theta$$
$$= 0 \quad \text{otherwise.}$$

The lemma below gives an equivalent definition of J_N which is useful in computing the moments of J_N.

Lemma 4.1. *The statistic J_N defined in (4.1) is the same as the following:*

$$J_N = \binom{N}{2}^{-1} \sum_{i<j} [I_{[X_i+X_j>2\theta]} - I_{[X_i>\theta, X_j>\theta]}] \tag{4.2}$$

where

$$I_{[X_i+X_j>2\theta]} = 1 \quad \textit{if } X_i + X_j > 2\theta$$
$$= 0 \quad \textit{otherwise}$$

and

$$I_{[X_i>\theta, X_j>\theta]} = 1 \quad \textit{if } X_i > \theta \textit{ and } X_j > \theta$$
$$= 0 \quad \textit{otherwise.}$$

Proof. The equivalence of (4.1) and (4.2) follows from the relation $\phi(X_1 - \theta, X_2 - \theta) = I_{[X_1+X_2>2\theta]} - I_{[X_1>\theta, X_2>\theta]}$.

The test statistic J_N is based on the number of positive deviations from the median that exceed the absolute values of the negative deviations from the median. So if H' is true then we would expect a high value of J_N. We shall reject H in favor of H' if J_N is too large.

We observe that J is a U-statistic. Hence by a result of Hoeffding [6] $N^{\frac{1}{2}}(J - EJ)$ is asymptotically normally distributed with mean zero and asymptotic variance σ_1^2 which is given below in Lemma 4.2.

Lemma 4.2.

$$E(J) = \int_{-\infty}^{\infty} [1 - F(-x)]\, dF(x) - \tfrac{1}{4} \tag{4.3}$$

and

$$\sigma_1^2 = 4[\tfrac{1}{16} + \int_{-\infty}^{\infty} [1 - F(-x)]^2\, dF(x) - \int_0^{\infty} [1 - F(-x)]\, dF(x) - (\int_{-\infty}^{\infty} [1 - F(-x)]\, dF(x))^2 + \tfrac{1}{2}\int_{-\infty}^{\infty} [1 - F(-x)]\, dF(x)], \tag{4.4}$$

while under the hypothesis H

$$E_0(J) = \tfrac{1}{4}; \tag{4.5}$$

and

$$\sigma_{10}^2 = \tfrac{1}{12}. \tag{4.6}$$

Proof. Straightforward.

Consistency of the J-test. Since $\sigma_1^2 \to 0$ as $N \to \infty$ we have that

$$J \to \int_{-\infty}^{\infty} [1 - F(-x)]\, dF(x) - \tfrac{1}{4} \quad \text{in probability.}$$

Then by the application of a theorem due to Lehmann ([4], page 267), we find that the test based on J is consistent for the set of alternatives H' given in (3.0).

5. ARE (b_{10}, J). In this section we shall make some efficiency comparisons of the two tests b_{10} and J under two different classes of alternatives. We shall use the concept of the Pitman efficiency in making these comparisons. We consider a sequence of alternatives that gets closer and closer to the distribution under the null hypothesis as N increases. Let N^* and N be the sample sizes needed by the J and b_{10} tests respectively of same size α, to obtain the same power β against the same sequence of alternatives. N^* is a function of N. Suppose the ratio N^*/N tends to a limit independent of the level of significance and the power as $N \to \infty$. Then that limit is called the asymptotic relative efficiency of the test based on b_{10} relative to the test based on J. It will be denoted by ARE (b_{10}, J).

We observe that the conditions of Pitman's theorem [7] are satisfied in the following cases. The asymptotic normality of a U-statistic when the distribution depends on N was shown by Lehmann [9]. When the median is known the distribution is centered at the median before making efficiency computations.

Case I. Consider the following classes of distribution functions given by the density function

$$f(x) = g(x)I_{[x \leq 0]} + g(x/\tau)\tau^{-1} \cdot I_{[x>0]} \tag{5.1}$$

where g is symmetrical about zero. Then we consider the following hypotheses:

$$H_1 : \tau = 1 \quad \text{against} \quad H_1' : \tau > 1. \tag{5.2}$$

Efficacy of the b_{10}-test. The expectation of b_{10} under the alternative is given by (2.4) and its asymptotic null variance by (2.5). The efficacy of the b_{10}-test for the type of alternatives (5.2) is given by

$$\tfrac{1}{4}9N\left(\int_{-\infty}^{\infty} |x|^3 g(x)\, dx\right)^2 \Big/ \int_{-\infty}^{\infty} x^6 g(x)\, dx. \tag{5.3}$$

Efficacy of the J-test. The expectation under the alternative of the J-test is given by (4.3) and its asymptotic null variance by (4.6). The efficacy of the J-test for the type of alternatives (5.2) is given by

$$48N\left[\int_0^{\infty} xg^2(x)\, dx\right]^2. \tag{5.4}$$

Thus the asymptotic relative efficiency is

$$\text{ARE}(b_{10}, J) = \tfrac{3}{64}\left(\int_{-\infty}^{\infty} |x|^3 g(x)\, dx\right)^2 \Big/ \int_{-\infty}^{\infty} x^6 g(x)\, dx \cdot \left(\int_0^{\infty} xg^2(x)\, dx\right)^2. \tag{5.5}$$

Next we shall consider some examples of $g(x)$ and compute ARE given by (5.5).

Example 5.1. *Normal distribution.*

$$g(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2),$$

$$\text{efficacy of } b_{10} = 6N/5\pi,$$
$$\text{efficacy of } J = 3N/\pi^2,$$
$$\text{ARE } (b_{10}, J) = 2\pi/5.$$

In this case the LMP test for the hypotheses (5.2) is given by the rejection region $T = \sum_1^m X_i^2 - m > c$, where $X_1, X_2, \cdots, X_m$ are the m positive observations in a sample of size N. The efficacy of T is N.

EXAMPLE 5.2. *Laplace distribution.*

$$g(x) = \tfrac{1}{2} \exp(-|x|)$$
$$\text{efficacy of } b_{10} = 9N/80,$$
$$\text{efficacy of } J = 3N/16,$$
$$\text{ARE } (b_{10}, J) = \tfrac{3}{5}.$$

In the case the LMP test for the hypotheses (5.2) is given by the rejection region $T = \sum_1^m X_i - m > c$, where $X_1, X_2, \cdots, X_m$ are the m positive observations in a sample of size N. The efficacy of T is N.

EXAMPLE 5.3. *Triangular distribution.*

$$g(x) = 1 - |x|, \qquad |x| \leqq 1,$$
$$\text{efficacy of } b_{10} = 28N/100,$$
$$\text{efficacy of } J = N/3,$$
$$\text{ARE } (b_{10}, J) = 21/25.$$

EXAMPLE 5.4. *Rectangular distribution.*

$$g(x) = \tfrac{1}{2}, \qquad |x| \leqq 1,$$
$$\text{efficacy of } b_{10} = 63N/64,$$
$$\text{efficacy of } J = 48N/64$$
$$\text{ARE } (b_{10}, J) = 21/16.$$

Table 1 summarizes the above results for the case of testing the hypotheses given in (5.2).

CASE II. Next we consider the following class of distribution functions given by

$$F(x - \theta) = (1 - \rho)G(x) + \rho H(x), \tag{5.6}$$

where G is symmetric about zero and H is any other continuous distribution function. Consider the test of hypotheses:

$$H_2 : \rho = 0 \quad \text{against} \quad H_2' : \rho > 0. \tag{5.7}$$

Efficacy of the b_{10}-test.

TABLE 1

Distribution	ARE(b_{10} J,)
Normal	$2\pi/5$
Laplace	3/5
Triangular	21/25
Rectangular	21/16

$$dE(b_{10})/d\rho\,|_{\rho=0} = \{\mu_3' - 3\boldsymbol{\mu}_2[1 - 2H(0)]/2g(0)\}/\boldsymbol{\mu}_2^{\frac{3}{2}}$$

$$\sigma_0^2(b_{10}) = \boldsymbol{\mu}_6/N\boldsymbol{\mu}_2^3.$$

Thus the efficacy of b_{10}-test is given by

$$(5.8) \qquad N[\mu_3' - 3\boldsymbol{\mu}_2(1 - 2H(0))/2g(0)]^2/\boldsymbol{\mu}_6,$$

where μ' are moments about the origin of the distribution function H, where H is not symmetric about zero. $\boldsymbol{\mu}$ are the central moments of G.

Efficacy of the J-test.

$$dE(J)/d\rho\,|_{\rho=0} = 2\int GdH - [2H(0) - 1]\int g^2(x)\,dx/g(0) - 1$$

and the efficacy of the J-test is given by

$$(5.9) \qquad 12N\{2\int GdH + [2H(0) - 1]\int g^2(x)\,dx/g(0) - 1\}^2.$$

Thus

$$(5.10) \quad \text{ARE}\,(b_{10}, J) = \{\mu_3' - 3\boldsymbol{\mu}_2[1 - 2H(0)]/2g(0)\}^2 \cdot [12\boldsymbol{\mu}_6\{2\int GdH + [2H(0) - 1]\int g^2(x)\,dx/g(0) - 1\}^2]^{-1}.$$

Let H be the rectangular distribution on $-b$ to a. In that case (5.10)reduces to

$$(5.11) \quad \text{ARE}\,(b_{10}, J) = \{(a^4 - b^4)/4(a+b) - 3\boldsymbol{\mu}_2(a-b)/[2g(0)\cdot(a+b)]\}^2 \cdot [12\boldsymbol{\mu}_6\{2\int_{-b}^{a} G(x)\,dx/(a+b) + (b-a)\int g^2(x)\,dx/[g(0)\cdot(a+b)] - 1\}^2]^{-1}.$$

We shall compute ARE in some special cases. Let $H(x)$ be the rectangular distribution on 0 to a. In that case (5.11) reduces to

$$(5.12) \quad \text{ARE}\,(b_{10}, J) = \{a^3/4 - 3\boldsymbol{\mu}_2/2g(0)\}^2 \cdot [12\boldsymbol{\mu}_6\{2\int_0^a G(x)\,dx/a - \int g^2(x)\,dx/g(0) - 1\}^2]^{-1}.$$

EXAMPLE 5.5. Let G be the standard normal distribution and H the rectangular distribution on 0 to a. Then

$$\text{ARE}\,(b_{10}, J) = (a^3/4 - 3(2\pi)^{\frac{1}{2}}/2)^2 \cdot [180\{2G(a) - 1.71 + .80(\exp(-a^2/2) - 1)/a\}^2]^{-1}$$

which $\to 0$ as $a^3 \to 6(2\pi)^{\frac{1}{2}}$, which $\to \infty$ as $a \to \infty$. The ARE <1 for small values of a.

EXAMPLE 5.6. Let G be the Laplace distribution and H the rectangular distribution on 0 to a. Then

$$\text{ARE}\,(b_{10}, J) = (a^3/4 - \tfrac{3}{2})^2[(6!)12\{2G(a) + (1 + a)\exp(-a)/a - 1/a - \tfrac{3}{2}\}^2]^{-1}$$

which $\to 0$ as $a^3 \to 6$, which $\to \infty$ as $a \to \infty$. The ARE < 1 for small values of a.

EXAMPLE 5.7. Let G be the triangular distribution and H the rectangular distribution on 0 to a. Then

$$\text{ARE}\,(b_{10}, J) = 28(a^3/4 - \tfrac{1}{4})^2[12\{2G(a) - 5/3 + 2a^2/3 - a\}^2]^{-1}$$

which $\to 0$ as $a \to 1$, it is not defined at $a = 1$.

The ARE < 1 for $a > 1$ but close to 1, while ARE > 1 for large a.

6. The test statistic $\hat{J}$. In this section we return to our original problem, that of testing H against H' when θ is unknown. For that purpose we modify the test statistic J, defined in (4.1), in the following way:

$$\hat{J} = \hat{J}_N = \binom{N}{2}^{-1} \sum_{i<j} \phi(X_i - \hat{\theta}, X_j - \hat{\theta}). \tag{6.1}$$

In the above definition $\hat{\theta}$ is the sample median. We shall take $N = 2n + 1$ so that there exists a unique sample median.

The statistic $\hat{J}$ is based on the number of positive deviations from the sample median that exceed the negative of the negative deviations from the sample median. We shall reject H if $\hat{J}$ is large.

Next we shall obtain the asymptotic distribution of the test statistic $\hat{J}_N$.

The asymptotic distribution of $\hat{J}$. We verify below the conditions of Theorem 3.1. We assume $\theta = 0$ and let $N = 2n + 1$ so that there exists a unique sample median $\hat{\theta}$.

(i) It is clear from the definition that W is bounded.

(ii) Here we verify the condition that there exists a constant M_2 such that

$$E\,|W(X_1, X_2, t) - W(X_1, X_2, 0)| \leqq M_2 t.$$

In this case

$$\begin{aligned} E\,|W(X_1, X_2, t) - W(X_1, X_2, 0)| &= E\,|\phi(X_1 - t, X_2 - t) - \phi(X_1, X_2) + A(0) - A(t)| \\ &\leqq E\,|\phi(X_1 - t, X_2 - t) - \phi(X_1, X_2)| + |A(0) - A(t)|. \end{aligned}$$

Now

$$\begin{aligned} A(t) - A(0) = \textstyle\int_{-\infty}^{\infty} [F(-x) - F(2t - x)]dF(x) \\ + [1 - F(0) + 1 - F(t)][F(t) - F(0)], \end{aligned}$$

so that $|A(t) - A(0)| \leqq 2at + 2at = 4at$ provided the distribution function F has a derivative $F' = f$ bounded by a. Next we have

$$|\phi(X_1 - t, X_2 - t) - \phi(X_1, X_2)|$$
$$\begin{aligned} &= 1 \quad \text{if either} \quad \phi(X_1 - t, X_2 - t) = 1 \text{ and } \phi(X_1, X_2) = 0 \\ &\qquad\qquad \text{or} \quad \phi(X_1 - t, X_2 - t) = 0 \text{ and } \phi(X_1, X_2) = 1 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Hence

$$\begin{aligned} &E\,|\phi(X_1 - t, X_2 - t) - \phi(X_1, X_2)| \\ &\quad = 2P\{[X_1 < t < X_2 \text{ and } X_1 + X_2 > 2t], [X_1 < 0 < X_2, X_1 + X_2 > 0]^c\} \\ &\qquad + 2P\{[X_1 < 0 < X_2 \text{ and } X_1 + X_2 > 0], [X_1 < t < X_2 \text{ and } X_1 + X_2 > 2t]^c\} \\ &\quad = 2P(A) + 2P(B), \quad \text{say,} \end{aligned}$$

where $[\ ]^c$ means the event complementary to the event $[\ ]$. Now

$$\begin{aligned} P(A) &= \{P[X_1 < t < X_2, X_1 + X_2 > 2t], [X_1 < 0 < X_2, X_1 + X_2 > 0]^c\} \\ &< P[0 < X_1 < t, X_2 > 0] = P[X_2 > 0]P[0 < X_1 < t] \\ &\leqq F(t) - F(0) = at \end{aligned}$$

and

$$\begin{aligned} P(B) &= P\{[X_1 < 0 < X_2, X_1 + X_2 > 0], [X_1 < t < X_2, X_1 + X_2 > 2t]^c\} \\ &= P[X_1 < 0 < X_2, -X_1 < X_2 < 2t - X_1] \\ &= \int_{-\infty}^{0} [F(2t - x) - F(-x)]\, dF(x) \leqq 2at \int_{-\infty}^{0} dF(x) = at \end{aligned}$$

provided the distribution function F has a derivative bounded by a.

So $2P(A) + 2P(B) \leqq 4at$, and $E\,|W(X_1, X_2, t) - W(X_1, X_2, 0)| \leqq 8at$.

Thus condition (A) of Theorem 3.1 is satisfied. Observing that ϕ is an indicator function, it is easily seen that condition (B) is also satisfied.

Thus by the application of Theorem 3.1 we get the following theorem.

THEOREM 6.1. *Let $X_1, X_2, \cdots, X_N$ be N independent identically distributed random variables with absolutely continuous* cdf $F(x - \theta)$. *Assume that F has a bounded derivative. Define J and $\hat{J}$ as in* (4.1) *and* (6.1), *respectively. Let $N = 2n + 1$, so that $\hat{\theta}$ is the unique sample median. Define*

$$L_N = \binom{N}{2}^{-1} \sum_{i<j} [\phi(X_i - \hat{\theta}, X_j - \hat{\theta}) - A(\hat{\theta} - \theta)]$$

where $A(t - \theta) = E\phi(X_1 - t, X_2 - t)$. Then

$$\lim_{N\to\infty} \mathcal{L}(N^{\frac{1}{2}} L_N) = \mathfrak{N}(0, \sigma_1^2)$$

where σ_1^2 is the asymptotic variance of $N^{\frac{1}{2}}(J - EJ)$ and is given by Lemma 4.2.

Next we verify the conditions of Theorem 3.2:

$$A(t) = \int_{-\infty}^{\infty} [1 - F(2t - x)]\, dF(x) - [1 - F(t)]^2;$$
$$A'(0) = f(0) - 2\int f(-x)\, dF(x) = f(0) - 2\int f^2(x)\, dx$$

when H is true.

In the following theorem we obtain the asymptotic distribution of $\hat{J}_N$.

THEOREM 6.2. *Let $X_1, X_2, \cdots, X_N$ be N independently identically distributed random variables with* cdf $F(x - \theta)$. *Let* $A(t) = E[\phi(X_1 - t, X_2 - t)/\theta = 0]$ *have a derivative continuous in the neighbourhood of the origin. Let the distribution function have a continuous density function near the origin with* $f(0) \neq 0$. *Let also f be bounded. Then $N^{\frac{1}{2}}(\hat{J} - EJ)$ is asymptotically normally distributed with mean zero and asymptotic variance σ_2^2. σ_2^2 is given in terms of the asymptotic variances and covariances of $N^{\frac{1}{2}}(\hat{J} - A(\hat{\theta} - \theta))$ and $N^{\frac{1}{2}}(\hat{\theta} - \theta)A'(h)$, by*

$$\sigma_2^2 = \sigma_1^2 + [A'(0)]^2/4f^2(0) + (m'' - m')A'(0)/f(0).$$

PROOF. We have

$$N^{\frac{1}{2}}(\hat{J}_N - EJ_N) = N^{\frac{1}{2}}[\hat{J}_N - A(\hat{\theta} - \theta)] + N^{\frac{1}{2}}[A(\hat{\theta} - \theta) - EJ_N].$$

But $A(\hat{\theta} - \theta) = A(0) + (\hat{\theta} - \theta)A'(h)$, where $h = \Delta(\hat{\theta} - \theta)$, $|\Delta| < 1$. Thus $N^{\frac{1}{2}}(\hat{J}_N - EJ) = N^{\frac{1}{2}}[\hat{J}_N - A(\hat{\theta} - \theta)] + N^{\frac{1}{2}}(\hat{\theta} - \theta)A'(h)$. By continuity $A'(h) \to A'(0)$ in probability. Also $N^{\frac{1}{2}}[\hat{J}_N - A(\hat{\theta} - \theta)] - N^{\frac{1}{2}}(J_N - EJ_N) \to 0$ in probability as a consequence of Theorem 6.1. It follows by the application of Theorems 3.3 and 3.4 that the joint distribution of $N^{\frac{1}{2}}[\hat{J}_N - A(\hat{\theta} - \theta)]$ and $N^{\frac{1}{2}}(\hat{\theta} - \theta)A'(h)$ is asymptotically normal with asymptotic means zero and asymptotic variances σ_1^2 and $[A'(0)]^2/4f^2(0)$ respectively, the asymptotic expression for the correlation coefficient being $(m'' - m')/\sigma_1$. It then follows that $N^{\frac{1}{2}}(\hat{J}_N - EJ_N)$ has a limiting normal distribution with mean zero and asymptotic variance σ_2^2 given by,

$$\sigma_2^2 = \sigma_1^2 + [A'(0)]^2/4f^2(0) + (m'' - m')A'(0)/f(0).$$

This completes the proof of Theorem 6.2.

We observe that in deriving the asymptotic distribution of $\hat{J}$ we have made no assumption of symmetry of $F(x)$. Under the null hypothesis, $N^{\frac{1}{2}}(\hat{J}_N - E_0J)$ is asymptotically $\mathfrak{N}(0, \sigma_{20}^2)$, where $E_0J = \frac{1}{4}$ and $\sigma_{20}^2 = \frac{1}{12} + [1 - 2\int f^2(x)\,dx/f(0)]^2/4$. So we shall reject H if $\hat{J}$ is too large. Here the asymptotic null variance σ_{20}^2 depends on the form of the distribution function F, if $A'(0) \neq 0$. We perform an asymptotically normal test if F is known. If, however, F is not known, then we shall use consistent estimate of σ_{20}^2 and after proper Studentization again perform an asymptotically normal test. This we shall discuss in detail in Section 8.

7. ARE $(J, \hat{J})$. We shall compare the test based on the statistic J_N for testing the hypothesis of symmetry when θ is known with the corresponding modified test based on $\hat{J}_N$ when θ is unknown in order to know how much is lost for not knowing the value of θ. For that we obtain the efficacies of the two tests under the same sequence of alternatives tending to the hypothesis and then compare these efficacies with the help of the ARE. Since both J and $\hat{J}$ have the same asymptotic mean under the alternative the ARE $(J, \hat{J})$ will be the same for all types of alternatives and

$$(7.1) \qquad \text{ARE}\,(J, \hat{J}) = \sigma_{20}^2/\sigma_{10}^2 = 1 + 3[1 - 2\int f^2(x)\,dx/f(0)]^2.$$

This ARE $\geqq 1$ as is expected. The loss in efficiency for not knowing θ depends on the magnitude of the second term in ARE.

In the next lemma we give some idea as to the range of values of this ARE in the case of distributions with $f(x) \leqq f(0)$ for all x.

LEMMA 7.1.

(1) *For distributions with* $f(x) \leqq f(0)$ *for all* x

$$1 \leqq \text{ARE}\,(J, \hat{J}) \leqq 4.$$

(2) ARE $(J, \hat{J}) = 4$ *if and only if the distribution is rectangular.*

(3) ARE $(J, \hat{J}) = 1$ *if and only if* $\int f^2(x)\,dx = f(0)/2$.

PROOF. (1) For distributions with $f(x) \leqq f(0)$ for all x, $\int f^2(x)\,dx/f(0) \leqq 1$, equality holds if and only if the distribution is rectangular.

$$\begin{aligned} \text{ARE}\,(J, \hat{J}) &= 1 + 3[1 - 2\int f^2(x)\,dx/f(0)]^2 \\ &= 4 + 12[\{\int f^2(x)\,dx/f(0)\}^2 - \int f^2(x)\,dx/f(0)] \\ &\leqq 4. \end{aligned}$$

We have already observed at the end of (7.1) that ARE $\geqq 1$.

(2) and (3) follow easily.

That the lower bound is attained by ARE is seen from the examples of Cauchy and Laplace distributions below. There are other distributions also which attain the value 1. For example consider

$$\begin{aligned} f(x) &= 1 && \text{if } -\tfrac{1}{4} \leqq x \leqq \tfrac{1}{4} \\ &= \tfrac{1}{2} && \text{if } \tfrac{1}{4} \leqq |x| \leqq \tfrac{3}{4} \\ &= 0 && \text{otherwise} \end{aligned}$$

and for this distribution ARE $(J, \hat{J}) = 1$.

Table 2 gives the ARE $(J, \hat{J})$ for a number of distributions.

ARE $(b_1, \hat{J})$. The condition of the continuity of the density at the median under the alternative of Theorem 6.2 fails to be satisfied for the class of distributions considered under Case I of Section 5. So we make the efficiency com-

TABLE 2

	Probability density function	ARE$(J, \hat{J})$
Normal	$\exp(-x^2/2)/(2\pi)^{\frac{1}{2}}$	1.50
Rectangular	1, for $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$	4.00
Cauchy	$1/\pi\ (1 + x^2)$	1.00
Triangular	$1 - \lvert x\rvert$, for $\lvert x\rvert \leqq 1$	1.33
Laplace	$\exp(-\lvert x\rvert)/2$	1.00

parisons of the two tests b_1 and $\hat{J}$ for the following class of distribution functions only, where the conditions of Theorem 6.2 are satisfied.

$$F(x - \theta) = (1 - \rho)G(x) + \rho H(x) \tag{7.2}$$

where G is a symmetric distribution function, symmetric about zero and H any other absolutely continuous distribution function. Consider the test of hypothesis

$$H_1 : \rho = 0 \quad \text{against} \quad H_1' : \rho > 0, \tag{7.3}$$

when θ is unknown.

Efficacy of the b_1-test.

$$(d\mu_3/d\rho)_{\rho=0} = \mu_3' - 3\mathfrak{u}_2\mu_1'$$

where μ_k' is the moment about the origin of the distribution H and $\mathfrak{u}_k$ are central moments of G. Efficacy of the b_1-test is then given by

$$\text{efficacy of the } b_1\text{-test} = N(\mu_3' - 3\mathfrak{u}_2\mu_1')^2/(\mathfrak{u}_6 - 6\,\mathfrak{u}_2\mu_4 + 9\mathfrak{u}_2^3). \tag{7.4}$$

Efficacy of the $\hat{J}$-test.

$$(dE_\rho(J)/d\rho)_{\rho=0} = [2 \textstyle\int GdH - \int g^2(x)\,dx/g(0) - 1],$$

and hence

$$\text{efficacy of the } \hat{J}\text{-test} = 12N[2 \textstyle\int GdH - \int g^2(x)\,dx/g(0) - 1]^2 \cdot [1 + 3[1 - 2\int g^2(x)\,dx/g(0)]^2]^{-1}. \tag{7.5}$$

Thus

$$\text{ARE}(b_1, \hat{J}) = (\mu_3' - 3\mathfrak{u}_2\mu_1')^2[1 + 3[1 - 2\textstyle\int g^2(x)\,dx/g(0)]^2] \cdot [12(\mathfrak{u}_6 - 6\mathfrak{u}_2\mathfrak{u}_4 + 9\mathfrak{u}_2^3)\cdot\{2\int GdH - \int g^2(x)\,dx/g(0) - 1\}^2]^{-1} \tag{7.6}$$

Let us compute ARE for some special cases. Let $H(x)$ be the rectangular distribution on 0 to a. In that case (7.6) reduces to

$$\text{ARE}(b_1, \hat{J}) = (a^3/4 - 3a\mathfrak{u}_2/2)^2[1 + 3[1 - 2\textstyle\int g^2(x)\,dx/g(0)]^2] \cdot [12(\mathfrak{u}_6 - 6\mathfrak{u}_2\mathfrak{u}_4 + 9\mathfrak{u}_2^3)\{2\int_0^a G(x)\,dx/a - \int g^2(x)\,dx/g(0) - 1\}^2]^{-1}. \tag{7.7}$$

Example 7.7. Let G be the standard normal distribution and H the rectangular distribution on 0 to a. Then

$$\text{ARE}(b_1, \hat{J}) = (a^3/4 - 3a/2)^2(1.5)/72[2G(a) - .80(\exp(-a^2/2) - 1)/a - 1.71]^2.$$

From this expression we find that ARE $= 0$ for $a = 6^{\frac{1}{2}}$ and ARE $\to 0$ as $a \to 0$, while ARE $\to \infty$ as $a \to \infty$.

Example 7.8. Let G be the Laplace distribution and H the rectangular distribution on 0 to a. Then

ARE $(b_1, \hat{J})$

$$= (a^3/4 - 3a)^2/(12)(504)[2G(a) + \exp(-a)(1 + a)/a - 1/a - \tfrac{3}{2}]^2$$

which is zero for $a = (12)^{\frac{1}{2}}$ and also ARE $\rightarrow 0$ as $a \rightarrow 0$ while ARE $\rightarrow \infty$ as $a \rightarrow \infty$.

EXAMPLE 7.9. Let G be the triangular distribution and H as before, the rectangular distribution on 0 to a. Then

$$\text{ARE}(b_1, \hat{J}) = 46.5(a^3 - a)^2/72[2G(a) - 5/3 + 2a^2/3 - a]^2.$$

This is not defined at $a = 1$, for then both the efficacies are equal to zero. ARE $\rightarrow 0$ as $a \rightarrow 1$, ARE > 1 for large a and ARE < 1 for small a.

ARE (b_{10}, b_1). Let us now compare the test statistics b_{10} and b_1, discussed in Section 2, to find out the loss of efficiency for not knowing θ. The case here is different from that of J_N and $\hat{J}_N$. For b_{10} and b_1 have different asymptotic means so the ARE will depend on the type of alternative we consider.

Let us consider the test of hypothesis given by (5.1) and (5.2). We have already obtained the efficacy of the b_{10}-test in (5.3). Now we find the efficacy of the b_1-test.

Efficacy of the b_1-test. The expectation under the alternative is given by $E(b_1) \sim \mu_3/\mu_2^{\frac{3}{2}}$. So

$$dE(b_1)/d\tau\,|_{\tau=1} \sim (d\mu_3/d\tau)\,|_{\tau=1}/\boldsymbol{\mu}_2^{\frac{3}{2}}.$$

The efficacy of the b_1-test is then given by

$$(d\mu_3/d\tau)^2\,|_{\tau=1}/\sigma_0^2(b_1)\boldsymbol{\mu}_2^3 = N(d\mu_3/d\tau)^2\,|_{\tau=1}/(\boldsymbol{\mu}_6 - 6\boldsymbol{\mu}_2\boldsymbol{\mu}_4 + 9\boldsymbol{\mu}_2^3)$$

Thus

$$\text{ARE}(b_{10}, b_1) = [(d\mu_3/d\tau) + 3\boldsymbol{\mu}_2(d\mu_1/d\tau)]^2_{\tau=1} \cdot(\boldsymbol{\mu}_6 - 6\boldsymbol{\mu}_2\boldsymbol{\mu}_4 + 9\boldsymbol{\mu}_2^3)/[\boldsymbol{\mu}_6(d\mu_3/d\tau)^2]_{\tau=1}$$

where the $\boldsymbol{\mu}$ are the central moments under the null hypothesis and μ_3 the third central moment and μ_1 the mean under the alternative.

Table 3 gives the results for different distributions in the testing situation given by (5.1) and (5.2).

8. Consistent estimators of $\int f^2(x)\,dx$ and $f(0)$. Theorem 6.2 gives the asymptotic distribution of $\hat{J}_N$. But it is not distribution-free under the null hypothesis if $A'(0) \neq 0$. In that case the asymptotic distribution of $\hat{J}_N$ depends on F

TABLE 3

Distribution	ARE(b_{10}, b_1)
Normal	1.60
Laplace	1.57
Triangular	10.80
Rectangular	4.80

through σ_{20}^2. In this section we shall give consistent estimators for $\int f^2(x)\,dx$ and $f(0)$ and hence of σ_{20}^2. Then we shall Studentize the statistic $\hat{J}_N$ and make it distribution-free.

Consistent estimator of $\int f^2(x)\,dx$. Let $X_1, X_2, \cdots, X_N$ be independent observations from a symmetric distribution $F(x - \theta)$ with median θ. Let $N'(\theta)$ be the number of averages $(X_i + X_j)/2$ with $i < j$ which exceed θ. If the $\binom{N}{2}$ averages $(X_i + X_j)/2$, $i < j$ after ordering them are denoted by $A^{(1)} < A^{(2)} < \cdots < A^{(M)}$ where $M = \binom{N}{2}$, then the confidence interval obtained from the two sided symmetric level α-test of hypothesis $H'(\theta_0)$: $\theta = \theta_0$ is seen to be (Lehmann [10])

$$(8.1) \qquad \underline{\theta} < \theta < \bar{\theta}$$

where

$$(8.2) \qquad \underline{\theta} = A^{(c_\alpha)}, \qquad \bar{\theta} = A^{(M+1-c_\alpha)}$$

where c_α is given by

$$P[c_\alpha \leqq N'(\theta_0) \leqq \tbinom{N}{2} - c_\alpha/H'] = 1 - \alpha.$$

LEMMA 8.1 (Lehmann). *If $L = \bar{\theta} - \underline{\theta}$ is the length of the confidence interval defined by* (8.2) *then*

$$(3N)^{\frac{1}{2}}L/k_{\alpha/2} \to 1/\int f^2(x)\,dx, \quad \textit{in probability},$$

where $k_{\alpha/2}$ is the $100\alpha/2$ *upper percentage point of the standard normal distribution.*

PROOF. Lehmann [10] has given the proof of it.

Consistent estimator of $f(0)$. Let $X_1, X_2, \cdots, X_N$ be independently identically distributed according to the cdf $F(x - \theta)$. Let $\hat{\theta}$ be the sample median. We take $N = 2n + 1$, so that there exists a unique sample median.

Define

$$f_N = \{\textstyle\sum_{i=1}^N I_{[\hat{\theta}-h \leqq X_i \leqq \hat{\theta}+h]}\}/2Nh,$$

where h is a function of the sample size N such that $\lim_N h = 0$ and $\lim_N Nh = \infty$. Lemma 8.2 below proves that f_N is a consistent estimator of $f(0)$, the density at the population median.

LEMMA 8.2. *Let f_N be defined as above. Then, under assumption* 1 *of Section* 1, *$f_N \to f(0)$, in probability.*

PROOF. We have

$$\begin{aligned} E[f_N/\hat{\theta}] &= [1 + \textstyle\sum_{X_i \neq \hat{\theta}} P\{\hat{\theta} - h \leqq X_i \leqq \hat{\theta} + h/\hat{\theta}\}]/2Nh \\ &= (2Nl)^{-1} + (N-1)N^{-1}[F(\hat{\theta} - \theta + h) - F(\hat{\theta} - \theta - h)]/2h. \end{aligned}$$

Since f is bounded, the last ratio is bounded, and since f is continuous at 0, the ratio converges to $f(0)$ in probability. Thus we have a sequence of uniformly bounded random variables converging to $f(0)$ in probability and so their expected values converge to $f(0)$. Hence $Ef_N \to f(0)$. Similarly it follows that $Ef_N^2 \to f^2(0)$.

Hence $f_N \to f(0)$ in probability. This completes the proof of Lemma 8.2.

Using Lemma 8.1 and Lemma 8.2 we obtain a consistent estimator of σ_{20}^2 from Lemma 8.3.

LEMMA 8.3. *Let the notations and assumptions be the same as in Lemma* 8.1 *and Lemma* 8.2. *Then*

$$\tfrac{1}{12} + [1 - 2r_N/f_N]^2/4 \to \tfrac{1}{12} + [1 - 2\int f^2(x)\,dx/f(0)]^2/4 \quad \textit{in probability,}$$

where $r_N = k_{\alpha/2}/(3N)^{\frac{1}{2}}L$.

PROOF. The proof follows from Lemma 8.1, Lemma 8.2 and Slutsky's theorem.

We remarked at the end of Section 6 that σ_{20}^2 depends on the form of the distribution function F. Now we have a consistent estimator of σ_{20}^2 in Lemma 8.3. Let us denote it by S_N^2. Then by Cramér's theorem [1], $N^{\frac{1}{2}}(\hat{J} - EJ)/S_N$ tends in law to $\mathfrak{N}(0, 1)$ and is thus asymptotically distribution-free. We can thus perform the test of H vs. H' with the help of this Studentized statistic which is asymptotically $\mathfrak{N}(0, 1)$.

9. Acknowledgment. I wish to express my deepest gratitude to Professor E. L. Lehmann who suggested the topic and gave generous help during the course of the entire work. I am thankful to Professor F. N. David for her interest and many helpful discussions. I wish to thank the referee for his suggestions and comments which were of great value to me.

REFERENCES

[1] CRAMÉR, H. (1963). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[2] FISHER, R. A. (1929). Moments and product moments of sampling distribution. *Proc. London Math. Soc. Ser. 2* **30** 199–238.

[3] FISZ, M. (1963). *Probability Theory and Mathematical Statistics*. Wiley, New York.

[4] FRASER, D. A. S. (1957). *Nonparametric Methods in Statistics*. Wiley, New York.

[5] GUPTA, M. K. (1964). Unpublished doctoral dissertation, University of California, Berkeley.

[6] HOEFFDING, W. (1948). A class of statistics with asymptotic normal distribution. *Ann. Math. Statist.* **19** 293–325.

[7] KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics* **2**. Hafner, New York.

[8] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[9] LEHMANN, E. L. (1963a). Robust estimation in analysis of variance. *Ann. Math. Statist.* **34** 957–966.

[10] LEHMANN, E. L. (1963b). Nonparametric confidence intervals for a shift parameter. *Ann. Math. Statist.* **34** 1507–1512.

[11] PARZEN, E. (1962). An estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

[12] PEARSON, E. S. (1930). A further development of tests for normality. *Biometrika* **22** 239–249.

[13] SUKHATME, B. V. (1957). On the joint asymptotic distribution of the median and a *U*-statistic. *J. Roy. Statist. Soc. B* **19** 144–148.

[14] SUKHATME, B. V. (1958). Testing the hypothesis that two populations differ only in location. *Ann. Math. Statist.* **29** 60–78.