The Annals of Mathematical Statistics
1972, Vol. 43, No. 6, 2035-2038

A CRAMÉR VON-MISES TYPE STATISTIC FOR TESTING SYMMETRY

By E. D. Rothman and Michael Woodroofe

University of Michigan

A Cramér von-Mises type statistic is proposed for testing the symmetry of a continuous distribution function. Its asymptotic null distribution is found explicity, and its asymptotic distribution under a sequence of local alternatives is described. A Monte Carlo study indicates that the asymptotic formulae are accurate for sample sizes as small as twenty.

1. Introduction. Let $X_1, \cdots, X_n$ denote a random sample from a continuous distribution function F and consider the problem of testing F for symmetry about zero. This problem, of course, has not suffered from lack of attention and may, in particular, be treated with any of a multitude of rank tests (e.g., [3], Chapter 3).

Smirnov ([6] and [7]) once proposed a test based on the statistic

$$B_n = \sup_{x \leq 0} |Q_n(x)| ,$$

where

$$Q_n(x) = n^{\frac{1}{2}}[F_n(x) + F_n(-x) - 1]$$

for $x \in R$ with F_n equal to the sample distribution function. This statistic has also been considered in [2]. The exact and asymptotic null distributions of B_n are known. Moreover, tests which reject for large values of B_n are known to be consistent against all non-symmetric alternatives, whereas some rank tests are not.

Here we consider a related statistic—namely,

$$R_n = n \int_{-\infty}^{\infty} [F_n'(x) + F_n'(-x) - 1]^2 \, dF_n(x) , \tag{1}$$

where $2F_n'(x) = F_n(x + 0) + F_n(x - 0)$ for $x \in R$. We use F_n' instead of F_n in order to make R_n invariant under multiplication of the data by -1. For computational purposes R_n may more conveniently be written

$$R_n = \sum_{j=1}^{n} \left[F_n'(-X_j') - \frac{2n - 2j + 1}{2n} \right]^2 , \tag{2}$$

where $X_1', \cdots, X_n'$ denote the ordered values of $X_1, \cdots, X_n$. We shall show that our test too is consistent against all non-symmetric alternatives, give its asymptotic null distribution and percentiles explicitly, and describe its asymptotic non-null distribution under a sequence of local alternatives. We shall also present the results of a Monte Carlo study which indicate that the asymptotic null distribution provides an adequate approximation for sample sizes as small as twenty.

2. Distribution theory. We begin with two theorems which describe the asymptotic distribution of R_n under both the null hypothesis and a sequence of local alternatives. In the statement of these theorems, all distribution functions

Received July 28, 1971; revised May 19, 1972.

are understood to be right continuous. Also, if F is a distribution function and $0 < y < 1$, then $F^{-1}(y)$ is the infimum of $x \in R$ for which $F(x) \geqq y$.

THEOREM 1. *Let $X_1, \cdots, X_n$ be a sample from a continuous, symmetric distribution function F. Then, R_n converges in distribution as $n \to \infty$ to*

$$R = \int_0^1 W(t)^2 \, dt ,$$

where W denotes a standard Wiener Process on $[0, 1]$.

THEOREM 2. *Let G be a continuously differentiable, symmetric distribution function whose support is an interval and define μ on $[0, 1]$ by $\mu(0) = 0$, and $\mu(2t) = 2G'(G^{-1}(t))$, $0 < t \leqq \frac{1}{2}$. Also, let $Y_1, \cdots, Y_n$ be a random sample from G and let $X_i = Y_i + \delta_n$, $i = 1, \cdots, n$ where $n^{\frac{1}{2}}\delta_n \to \delta$ as $n \to \infty$. If μ is square integrable, then R_n converges in distribution as $n \to \infty$ to*

$$R(\delta) = \int_0^1 (W(t) - \delta\mu(t))^2 \, dt ,$$

where W is as in Theorem 1.

PROOF. We will prove Theorem 2 and then indicate how the proof may be modified to apply to Theorem 1.

Let F denote the distribution function of X_1 (the dependence of X_1 and F on n will be suppressed in the notation) and let

$$T_n(x) = n^{\frac{1}{2}}(F_n(x) + F_n(-x) - F(x) - F(-x)) , \qquad x \leqq 0 .$$

Then, as in [2], we find that $T_n(F^{-1}(t))$, $0 < t \leqq \frac{1}{2}$, converges in distribution to $W(2t)$ with respect to the topology of $D = D[0, \frac{1}{2}]$. (Here, of course, $T_n(F^{-1}(0)) = 0$ by convention.) Therefore,

$$Q_n(G^{-1}(t)) = T_n(F^{-1}(t) - \delta_n) + n^{\frac{1}{2}}(G(G^{-1}(t) - \delta_n) + G(G^{-1}(1 - t) - \delta_n) - 1)$$

converges in distribution to $W(2t) - \delta\mu(2t)$ by the continuity of W and a simple application of Taylor's theorem. Since integration defines a continuous functional on D, it now follows easily that

$$\begin{aligned} S_n &= \int_{-\infty}^{\infty} Q_n(x)^2 \, dG(x) \\ &= 2 \int_0^{\frac{1}{2}} Q_n(G^{-1}(t)) \, dt \end{aligned}$$

converges in distribution to

$$2 \int_0^{\frac{1}{2}} [W(2t) - \delta\mu(2t)]^2 \, dt = R(\delta)$$

as $n \to \infty$. Thus, Theorem 2 will be proved if we can show that $S_n - R_n \to 0$ in probability as $n \to \infty$.

To see this let R_{n0} be defined by (1) with F_n' replaced by F_n. Then, by expanding the squares in the definitions of R_n and R_{n0} and using the inquality $|F_n(x) - F_n'(x)| \leqq 1/2n$, $x \in R$, and the Minkowski Inequality, we find easily that

$$|R_n - R_{n0}| \leqq (R_n^{\frac{1}{2}} + R_{n0}^{\frac{1}{2}})/2n$$

wp 1, so it will suffice to show that $R_{n0} - S_n \to 0$ in probability as $n \to \infty$. Let

$V_n = Q_n \circ G^{-1}$ and let $H_n = F_n \circ G^{-1}$. Then,

$$|R_{n0} - S_n| = |\int_0^{\frac{1}{2}} V_n(t)^2 \, dH_n(t) - \int_0^{\frac{1}{2}} V_n(t)^2 \, dt| .$$

If we now approximate V_n by a step function which is constant on the intervals $((i-1)/2k, i/2k]$, $i = 1, \cdots, k$, we find easily that

$$|R_{n0} - S_n| \leqq 2 \sup_{|s-t| \leqq \frac{1}{k}} |V_n(t)^2 - V_n(s)^2| + 2(\sup_{0 \leqq t \leqq \frac{1}{2}} V_n(t)^2)\left(\sum_{j=0}^{k}\left|H_n\left(\frac{j}{2k}\right) - \frac{j}{2k}\right|\right),$$

which converges to zero in distribution (and hence in probability) as $n \to \infty$ and $k \to \infty$ (in that order). Here, of course, we use the fact that V_n is converging to V, where $V(t) = W(2t) - \delta\mu(2t)$, $0 < t \leqq \frac{1}{2}$.

This completes the proof of Theorem 2. The proof of Theorem 1 is similar but simpler, since there is no need to consider the function μ. The reader may easily convince himself that there is, consequently, no need to impose the additional conditions on the distribution of X_1.

Since $R_n \to \infty$ wp 1 as $n \to \infty$ if F is a fixed, non-symmetric distribution function, we immediately obtain

COROLLARY 1. *Tests which reject for large values of* R_n *are consistent against all non-symmetric alternatives.*

It is possible to describe limiting distribution of R_n as that of an infinite weighted sum of chi-square random variables. This representation will be useful in determining the percentiles of the asymptotic null distribution. Its proof may essentially be found in [5].

COROLLARY 2. *If either* $\delta = 0$, *or* μ *is square integrable then*

$$R(\delta) = \sum_{k=1}^{\infty} 4(Z_k - \delta\alpha_k)^2/(2k-1)^2\pi^2$$

in distribution, where $Z_1, Z_2, \cdots$ *are independent standard normal random variables and (if* $\delta \neq 0$)

$$\alpha_k = (k - \tfrac{1}{2})\pi \int_0^1 \mu(t) \sin((k - \tfrac{1}{2})\pi t) \, dt$$

for $k = 1, 2, \cdots$.

We conclude with a series expansion of the asymptotic null distribution function. The result may also be found in [4], but we include its proof for completeness.

THEOREM 3. *The distribution function of* $R = R(0)$ *is*

$$H(x) = 2^{\frac{3}{2}} \sum_{j=0}^{\infty} \binom{-\frac{1}{2}}{j}\left(1 - \Phi\left(\frac{4j+1}{2x^{\frac{1}{2}}}\right)\right), \qquad x > 0, \tag{3}$$

where Φ *denotes the standard normal distribution function.*

PROOF. It follows from Corollary 2 that the Laplace transform of R is

$$E(e^{-tR}) = \prod_{k=1}^{\infty} (1 + 8t(2k-1)^{-2}\pi^{-2}), \qquad t \geqq 0,$$

which we recognize as the infinite product expansion of

$$\cosh(2t^{\frac{1}{2}})^{-\frac{1}{2}} = 2^{\frac{1}{2}} \sum_{j=0}^{\infty} \binom{-\frac{1}{2}}{j} \exp[-(2j+\tfrac{1}{2})2t^{\frac{1}{2}}], \qquad t \geqq 0.$$

Since for all $\alpha > 0$, $e^{-\alpha(2t)^{\frac{1}{2}}}$ defines the Laplace transform of the distribution function $H_\alpha(x) = 2(1 - \Phi(\alpha/x^{\frac{1}{2}}))$, $x > 0$, (e.g., [1], Chapter 13), Theorem 3 now follows easily from the unicity theorem for Laplace transforms ([8], pages 59–63).

The series (3) was evaluated on a computer to determine the percentiles of the limiting null distribution. The $100(1-\alpha)$ percentiles are given below for selected values of α and agree with the values given in [4].

α	.1	.05	.025	.01
x_α	1.196	1.656	2.135	2.78

As a check on the accuracy of the large sample approximation, a Monte Carlo study was conducted. 1000 samples of size 15, 20, 30, and 40 were drawn from a uniform distribution and the proportion of samples producing an R_n value larger than x_α recorded for the values of α given above. The results indicate that the large sample approximation is accurate for sample sizes as small as twenty.

n \ α	.10	.05	.025	.01
15	.106	.051	.015	.005
20	.093	.047	.021	.008
30	.091	.040	.021	.010
40	.111	.052	.023	.010

Acknowledgment. We wish to thank Mr. Martin Weinrich and Miss Linda Howe for their assistance with the computations and Monte Carlo studies. We also wish to thank the referee for bringing reference [6] to our attention.

REFERENCES

[1] BREIMAN, L. (1968). *Probability*. Addison-Wesley.

[2] BUTLER, C. (1969). A test for symmetry using the sample distribution function. *Ann. Math. Statist.* **40** 2209–2210.

[3] HÁJEK, J. and ŠIDÁK, Z. (1967) *Theory of Rank Tests*. Academic Press.

[4] MACNEILL, I. (1969). Limit processes for spectral distribution functions with applications to goodness of fit testing. Technical Report No. 27 (DA-ARO(D)-31-124-G1077), Statistics Dept., Stanford University.

[5] SHEPP, L. (1967). Radon–Nikodym derivatives of Gaussian measures. *Ann. Math. Statist.* **37** 321–354.

[6] SMIRNOV, N. V. (1947). Sur un critere de la loi de distribution d'une variable aleatorie. *Izv. Acad. Sci. USSR* **56** 11–14.

[7] WALSH, J. E. (1965). *Handbook of Nonparametric Statistics*. Van Nostrand, Princeton.

[8] WIDDER, D. (1946). *The Laplace Transform*. Princeton Mathematical Series, No. 6, Princeton University Press.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48104