The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 2209–2210

A TEST FOR SYMMETRY USING THE SAMPLE DISTRIBUTION FUNCTION

BY CALVIN C. BUTLER

University of Colorado

1. Summary and introduction. Let $X_1, \cdots, X_n$ be n independent random variables with a common continuous distribution function (df) F. Let F_n denote the sample df of the X's. Let $\mathfrak{F}$ be the class of all continuous df's and $\mathfrak{F}_1$ denote the df's in $\mathfrak{F}$ which are symmetric about zero. To test the hypothesis $F \varepsilon \mathfrak{F}_1$ a common test is a weighted sign test of the form $\sum a_k \operatorname{sgn} X_k$ which has been studied by van Eeden and Bernard (1957), Hájek (1962), and Hájek and Sidák (1967). Usually the test included in nonparametric texts is for a_k equal to the rank of $|X_k|$ and is known as the Wilcoxon one-sample or signed rank test. This test is consistent against certain alternatives including the case when F is symmetric about some $\mu \neq 0$. That the test is not consistent against all alternatives in $\mathfrak{F} - \mathfrak{F}_1$ is evident from a discussion of its properties in Noether (1967).

In this paper a test statistic for the hypothesis $F \varepsilon \mathfrak{F}_1$ is constructed in the spirit of the Kolmogorov-Smirnov statistics and shown to be consistent against all alternatives in $\mathfrak{F} - \mathfrak{F}_1$. Its df for both the finite and asymptotic cases are included along with the df's of two closely associated "one-sided" test statistics.

2. The main result. When $F \varepsilon \mathfrak{F}_1$, $F(x) + F(-x) = 1$ so an obvious statistic to consider is some functional of $Q_n(x) = n[F_n(x) + F_n(-x) - 1]$ and in particular the statistic $B_n = n^{-\frac{1}{2}} \sup_{x \leqq 0} |Q_n(x)|$.

For $x < 0$, $Q_n(x) = \sum \operatorname{sgn} x_i$ where the summation extends over all subscripts for which $-|X_i| \leqq x$ enabling us to conclude that Q_n behaves like a random walk with steps occurring at the random points $-|X_1|, -|X_2|, \cdots, -|X_n|$. Moreover when $F \varepsilon \mathfrak{F}_1$, the random variables $|X_i|$ and $\operatorname{sgn} X_i$ are independent. Thus given $|X_1| = |x_1|, \cdots, |X_n| = |x_n|$, the probability assigned to a realization of the random function Q_n in the negative part of the real line with unit steps at $-|x_1|, \cdots, -|x_n|$ is the same as the probability assigned to a realization of a symmetric unrestricted random walk starting at the origin and having the same unit steps at $1, 2, \cdots, n$. We have proven the following theorem.

THEOREM. *$P[n^{\frac{1}{2}}B_n = x] = \sum_{k=-\infty}^{\infty} (v_{x(1+4k),n} + v_{x(1+4k)+1,n})$ where*

$$v_{x,n} = \binom{n}{\frac{1}{2}(n+x)} 2^{-n}$$

for $\frac{1}{2}(n+x)$ an integer in $[0, n]$ *and zero otherwise* (see Feller (1968), page 369).

The limiting distribution of B_n can be obtained directly from the above theorem, however, the following derivation is more direct. Since

$$n^{\frac{1}{2}}[F_n(x) - F(x)] \rightarrow_d Z(F(x))$$

where $Z(t)$ is a Gaussian process with zero mean and $EZ(s)Z(t) = s(1 - t)$,

Received 11 December 1968.

$0 \leqq s \leqq t \leqq 1$; it follows from direct computation that when $F \varepsilon \mathfrak{F}_1$, $n^{-\frac{1}{2}}Q_n$ converges in distribution to a Gaussian process Q with zero mean and $EQ(x_1)Q(x_2) = 2F(x_1) - \infty < x_1 \leqq x_2 < 0$. Furthermore it follows from Doob (1949) and Donsker (1952) that all continuous functionals of $n^{-\frac{1}{2}}Q_n$ converges as $n \to \infty$ to continuous functionals of Q. The df of B_∞ is then simply the well-known df of $\sup_{0 \leqq t \leqq 1} |Y(t)|$ where $Y(t)$ is standard Brownian motion over $0 \leqq t \leqq 1$. From Feller (1966), page 330, we have

$$P[B_\infty \leqq x] = 4\pi^{-1} \sum_{n=0}^{\infty} (-1)^n (2n+1)^{-1} \exp - ((2n+1)^2 \pi^2 (8x^2)).$$

The consistency of B_n follows easily from the Glivenko-Cantelli theorems since $n^{-\frac{1}{2}}B_n$ converges with probability one to a non-zero number.

Closely associated text statistics which are consistent for alternatives of the form $F(-x) + F(x) - 1 > 0$ or <0 for some x are $B_n^+ = \sup_{x \leqq 0} Q_n(x)$ and $B_n^- = \inf_{x \leqq 0} Q_n(x)$ respectively. Their distributions are particularly simple. From Feller (1968), page 369,

$$P[n^{\frac{1}{2}}B_n^+ = x] = P[n^{\frac{1}{2}}B_n^- = x] = v_{x,n} + v_{x+1,n}$$

and from Feller (1966), page 171,

$$P[B_\infty^+ > x] = P[B_\infty^- < x] = 2[1 - \Phi(x)]$$

where Φ is the standard normal df.

3. A "goodness of fit" test. For $F \varepsilon \mathfrak{F}$, the B's can be used to test the hypothesis that the sample comes from F by considering

$$q_n(p) = n[F_n(\zeta(1-p)) + F_n(\zeta(p)) - 1]$$

where $F(\zeta(p)) = p$. It is easily seen that

$$P[n^{\frac{1}{2}}B_n = x] = P[n^{-\frac{1}{2}} \sup_{0 \leqq p \leqq \frac{1}{2}} | q_n(p)| = x]$$

with similar results holding for B_n^+ and B_n^-.

REFERENCES

[1] Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

[2] Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

[3] van Eeden, C. and A. Bernard (1957). A general class of distribution-free tests for symmetry containing the tests of Wilcoxon and Fisher. *Nederl. Akad. Wetensch.*, **A60** (*Indag. Math.* **19**) 381–408.

[4] Feller, W. (1966). *An Introduction to Probability Theory and its Applications.* **2** Wiley, New York.

[5] Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* **1** 3rd. ed. Wiley, New York.

[6] Hájek, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33** 1124–1147.

[7] Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[8] Noether, G. E. (1967). *Elements of Nonparametric Statistics.* Wiley, New York.