A FLUCTUATION THEOREM AND A DISTRIBUTION-FREE TEST

By Charles H. Kraft[1] and Constance van Eeden[2]

University of Minnesota

Let $(X_1, \cdots, X_n)$ be completely independent random variables with continuous and symmetric (about zero) distributions. If N^* = number of positive sums in $\{X_1, X_1 + X_2, X_1 + X_2 + X_3, \cdots, X_1 + \cdots + X_n\}$, it has been shown by Andersen [1], [2] that N^* has, for each n, a distribution which is the same for the given class of sequences $(X_1, \cdots, X_n)$. See Baxter [3] for this and a related result.

We show here that the random variable N = number of positive sums in $\{\sum_{j \varepsilon T} X_j ; T \subset (1, 2, \cdots, n)\}$ also has a distribution which is constant for the same class of sequences $(X_1, \cdots, X_n)$.

Theorem. *The distribution of N is given by*

$$P(N = k) = 1/2^n, \qquad k = 0, 1, \cdots, 2^n - 1.$$

Proof. The conditional, given $(|X_1|, |X_2|, \cdots, |X_n|)$, distribution of $(X_1, \cdots, X_n)$ is uniform on the 2^n points $\{(\epsilon_1|X_1|, \epsilon_2|X_2|, \cdots, \epsilon_n|X_n|); \epsilon_i = -1, \text{ or } 1\}$. Let $0 = S_0 < S_1 < \cdots < S_{2^n-1}$ be the ordered partial sums from $\{|X_1|, \cdots, |X_n|\}$. Let $S_k = \sum_{i \varepsilon T} |X_i|$, and, for $j = 1, \cdots, n$, $\delta_j = -1$ if $j \varepsilon T$, $\delta_j = 1$ if $j \not\varepsilon T$.

Then (see lemma below) $N = N(\delta_1|X_1|, \delta_2|X_2|, \cdots, \delta_n|X_n|)$ = number of positive sums in $\{S_i - S_k, i = 0, \cdots, 2^n - 1, i \neq k\}$. Since this last expression is clearly equal to $2^n - 1 - k$ and

$$P[(\delta_1|X_1|, \cdots, \delta_n|X_n|) \mid (|X_1|, \cdots, |X_n|)] = 1/2^n$$

the result follows.

Lemma. $N(\delta_1|X_1|, \cdots, \delta_n|X_n|)$ = *number of positive sums in* $\{S_i - S_k; i = 0, 1, \cdots, 2^n - 1, i \neq k\}$.

Proof. A one-to-one correspondence between the partial sums from $(\delta_1|X_1|, \cdots, \delta_n|X_n|)$ and the elements of $\{S_i - S_k, i = 0, 1, \cdots, 2^n - 1, i \neq k\}$ is given below where $T = T(S_k)$ is defined above.

To $\sum_{j \varepsilon A} \delta_j|X_j|$, make correspond $S_{i(A)} - S_k$ where

$$S_{i(A)} = \sum_{j \varepsilon (A-T) \cup (T-A)} |X_j| \qquad A \neq T$$

$$i(A) = 0 \qquad A = T.$$

Received 10 December 1963.

[1] This research was supported in part by the National Science Foundation under Grant Number G-19126.

[2] This research was supported in part by the National Institutes of Health under Grant Number 2G-43(C7+8).

To $S_i - S_k$, make correspond $\sum_{j\varepsilon(B-T)\cup(T-B)} \delta_j|X_j|$ where B is defined by $S_i = \sum_{j\varepsilon B} |X_j|$.

The number of positive sums in the two sets is the same since the correspondents of the first map are equal, i.e., $\sum_{j\varepsilon A} \delta_j|X_j| = S_{i(A)} - S_k$.

We shall now discuss the consistency of tests based on N (or N^*). Let T_n denote one of N^*/n or $N/(2^n - 1)$. A test of the null hypothesis, $X_1, \cdots, X_n$ are independent and symmetric about zero is to reject H_0 if $T_n > k_n$ where k_n is, for large n, approximately the solution to $B_t(\frac{1}{2}, \frac{1}{2}) = 1 - \alpha$ (B_t is the incomplete beta-function) if $T_n = N^*/n$ and $k_n = 1 - \alpha$ if $T_n = N/(2^n - 1)$.

Since, for any random variable Z with $P(0 \leqq Z \leqq 1) = 1$,

$$P(Z \geqq 1 - \epsilon) \geqq 1 - (1 - EZ)/\epsilon,$$

a lower bound for the power of these tests is $1 - (1 - ET_n)/(1 - k_n)$. Therefore the tests will be consistent if $ET_n \to 1$. Further,

$$1 - ET_n = \sum_{k=0}^{n} P\{(Y_k - k\theta)/k^{\frac{1}{2}}\sigma < -k^{\frac{1}{2}}\theta/\sigma\} \cdot p_k \tag{1}$$

where

$$\begin{aligned} p_k &= n^{-1}, & k &= 1, \cdots, n \quad \text{for} \quad T_n = N^*/n \\ &= \binom{n}{k} (\tfrac{1}{2})^n, & k &= 0, \cdots, n \quad \text{for} \quad T_n = N/(2^n - 1). \end{aligned}$$

and where it is assumed that $(X_1, \cdots, X_n)$ are independently and identically distributed with a distribution $F[(x - \theta)/\sigma]$ for some F, and for $\theta > 0$, $\sigma > 0$, and Y_k is a random variable distributed as $\sum_{i=1}^{k} X_i$. From (1) and the Helly-Bray theorem it follows that the test will be consistent if

$$P\{(Y_k - k\theta)/k^{\frac{1}{2}}\sigma < - k^{\frac{1}{2}}\theta/\sigma\} \to 0. \tag{2}$$

If, for some alternative, ET_n does not converge to one there will be a size for which T_n will not be consistent. For example (2) is constant for the Cauchy distribution and T_n is not consistent.

REFERENCES

[1] ANDERSEN, E. SPARRE (1949). On the number of positive sums of random variables. *Skand. Aktuarietidskr.* **36** 27–36.

[2] ANDERSEN, E. SPARRE (1953). On the fluctuations of sums of random variables. *Math. Scand.* **1** 263–285.

[3] BAXTER, GLEN (1962). On a generalization of the finite arcsine law. *Ann. Math. Statist.* **33** 909–915.