ON SOME SIGN TESTS OF RANDOMNESS UNDER HYPOTHESES OF LINEAR TREND

by

R. M. Sundrum

University of Rangoon
and Institute of Statistics, University of North Carolina

Institute of Statistics
Mimeograph Series No. 96
Limited Distribution
March 1954

---

[1]This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

UNCLASSIFIED
Security Information

Bibliographical Control Sheet

1. O.A.: Institute of Statistics, North Carolina State College of the University of North Carolina

   M.A.: Office of Scientific Research of the Air Research and Development Command

2. O.A.: CIT Report No. 5
   M.A.: OSR Technical Note -

3. ON SOME SIGN TESTS OF RANDOMNESS UNDER HYPOTHESES OF LINEAR TREND (UNCLASSIFIED)

4. Sundrum, R. M.

5. March, 1954

6. 13

7. None

8. AF 18(600)-458

9. RDO No. R-354-20-8

10. UNCLASSIFIED

11. None

12. General formulae in the case of linear trends are obtained for the first four moments of (i) a sign test of randomness and (ii) a sign test of the independence of two time series; and for the mean and variance of (iii) a test based on the number of "turning points". The values of parameters occurring in these formulae have been computed for some special cases.

ERRATA TO "ON SOME SIGN TESTS OF RANDOMNESS UNDER HYPOTHESES OF LINEAR TREND"

by

R. M. SUNDRUM

p. 4, line 10: Insert, "In the case of normal alternatives," before "Stuart."

p. 5, last line: For $t = \Pr (X_i < X_{i+1} < X_{i+2} + X_{i+3} + X_{i+4})$

read

$$t = \Pr (X_i < X_{i+1} < X_{i+2} < X_{i+3} < X_{i+4})$$

page 11, Equation (17): For α_1 read α_i.

ON SOME SIGN TESTS OF RANDOMNESS UNDER HYPOTHESES OF LINEAR TREND[1]

by R. M. Sundrum
University of Rangoon and
Institute of Statistics, University of North Carolina

1. Introduction and Summary.

The main object of this paper is to obtain general expressions for the third and fourth moments of the sign test of randomness, proposed by Moore and Wallis [4], and further investigated by Levene [3] and Stuart [7]. Given observations of a sequence of random variables X_i ($i = 1, 2, \ldots n$), define

$$D_i = 1 \quad \text{if } X_i < X_{i+1}$$

$$= 0 \text{ otherwise} .$$

Then the sign test is based on the statistic:

$$D = \sum_{i=1}^{n-1} D_i . \qquad (1)$$

Moore and Wallis obtained the following moments of the statistic D in the null case, i.e. under the hypothesis that all the random variables

[1]This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

X_i are identically and independently distributed.

$$E(D) = \frac{n-1}{2}$$

$$\sigma^2(D) = \frac{n+1}{12} \qquad (2)$$

$$\mu_3(D) = 0$$

$$\mu_4(D) = \frac{(n+1)(5n+3)}{240}$$

Levene and Stuart have extended the results for the mean and variance to cases of linear trend, i.e. where the random variables $(X_i - i\theta)$ are independently and identically distributed, θ being a constant. They also show that under this alternative hypothesis, D tends to be asymptotically normally distributed. The distributional problem can therefore be considered as solved for the asymptotic case.

However, since the exact distribution of D is not available for any of these cases, we cannot say even approximately how large a sample size is required to assume the normal approximation. It is, therefore, useful to have the values of the third and fourth moments. Further, the need for such distribution free methods is greatest in the case of samples of medium sizes. In such cases, a knowledge of the third and fourth moments helps to make more accurate inferences.

Two results follow incidentally from this investigation. Firstly, the general expressions are also applicable to the moments of a bivariate difference-sign test proposed by Stuart [7], when the parameters occuring in these expressions are suitably re-defined. Secondly, the parameters occuring in these expressions are also involved in the formula for the variance of another test of randomness based on the number of "turning points," proposed by Wallis and Moore [9]. If we define

$$T_i = 1 \quad \text{if} \quad (X_i - X_{i+1})(X_{i+1} - X_{i+2}) < 0$$

$$= 0 \text{ otherwise,}$$

then this test is based on the statistic

$$T = \sum_{i=1}^{n-2} T_i \tag{3}$$

Moore and Wallis obtained the following results for the null case:

$$E(T) = \frac{2(n-2)}{3}$$
$$\sigma^2(T) = \frac{16n-29}{90} \tag{4}$$

Generalizations of these results to certain non-null cases are given by Levene in an asymptotic form and are obtained here exactly.

2. Moments of the Sign Test.

Stuart's results on the mean and variance of D may be written in the form

$$E(D) = (n-1)p \tag{5}$$

$$\begin{aligned}\sigma^2(D) &= (n-1)(p-p^2) + 2(n-2)(k-p^2) \\ &= (p+2k-3p^2)n + (5p^2-p-4k)\end{aligned} \tag{6}$$

where $p = \Pr(X_i < X_{i+1})$ and $k = \Pr(X_i < X_{i+1} < X_{i+2})$, these parameters being independent of i under the assumption of linear trends. In the null case, $p = \frac{1}{2}$ and $k = \frac{1}{6}$; substituting these values in (5) and (6), we get the corresponding formulae of (3). Stuart has shown how p and k may be determined from published tables [5].

From (1),

$$D^3 = \left\{ \sum_{i=1}^{n-1} D_i \right\}^3 ,$$

which on expansion gives $(n-1)^3$ terms. These can be grouped into the following nine classes of terms with expectations involving, in addition to p and k, the probability $s = \Pr(X_i < X_{i+1} < X_{i+2} < X_{i+3})$.

Term	Expectation	Number of terms
D_i^3	p	$(n-1)$
$D_i^2 D_{i+1}$	k	$3(n-2)$
$D_i D_{i+1}^2$	k	$3(n-2)$
$D_i^2 D_{i+j}$	p^2	$\frac{3}{2}(n-2)(n-3)$
$D_i D_{i+j}^2$	p^2	$\frac{3}{2}(n-2)(n-3)$
$D_i D_{i+j} D_{i+j+1}$	kp	$3(n-3)(n-4)$
$D_i D_{i+1} D_{i+j+1}$	kp	$3(n-3)(n-4)$
$D_i D_{i+1} D_{i+2}$	s	$6(n-3)$
$D_i D_{i+j} D_{i+j+m}$ $(j, m > 1)$	p^3	$(n-3)(n-4)(n-5)$
		$(n-1)^3$ terms.

Similarly,

$$D^4 = \left\{ \sum_{i=1}^{n-1} D_i \right\}^4$$

can be expanded into the sum of $(n-1)^4$ terms. These terms can be grouped into 27 classes with expectations involving, in addition to p, k and s, the probability $t = \Pr (X_i < X_{i+1} < X_{i+2} + X_{i+3} + X_{i+4})$.

In this way $E(D^3)$ and $E(D^4)$ can be expressed in terms of p, k, s and t. Collecting terms together, we get

$$\mu_3(D) = (p-9p^2+20p^3+6k-24kp+6s)n - (p-15p^2+44p^3+12k-60kp+18s) \quad (7)$$

$$\mu_4(D) = 3(p+2k-3p^2)^2n^2$$
$$+ (p-27p^2+168p^3-300p^4+14k-180kp+492kp^2-108k^2+36s-120sp+24t)n$$
$$- (p-38p^2+294p^3-633p^4+28k-384kp+1176kp^2-240k^2+108s-408sp+96t) \quad (8)$$

Taking only the dominant terms in n, we find

$$\sigma^2(D) \sim (p+2k-3p^2)n$$

$$\mu_3(D) \sim (p-9p^2+20p^3+6k-24kp+6s)n$$

$$\mu_4(D) \sim 3(p+2k-3p^2)^2n^2$$

so that, as $n \longrightarrow \infty$

$$\beta_1 \longrightarrow 0 \quad ; \quad \beta_2 \longrightarrow 3$$

illustrating the asymptotic tendency to normality, under the condition $(p+2k-3p^2) > 0$, i.e. when p and k are not too close to 0 or 1. This asymptotic result follows strictly from the Hoeffding-Robbins theorem [2].

3. <u>Special cases.</u>

(a) Null case:

In the null case, when all permutations of the observations are equally probable, we have $s = \frac{1}{4!}$ and $t = \frac{1}{5!}$; substituting these values in (7) and (8), we get the corresponding formulae of (2), thus providing a check on the algebra.

(b) Rectangular case:

Let X_i have a rectangular distribution in the range $-(i-1)\theta$ to $1-(i-1)\theta$, where $\theta > 0$. We consider this case of a negative trend because the formulae for p, k, s and t are simpler. The moments for a postitive trend are the same except that the sign of moments of odd order is changed. We have

$$\begin{aligned}
p &= \frac{(1-\theta)^2}{2!} && \text{for } 0 \leq \theta \leq 1 \\
&= 0 && \text{for } \theta > 1 \\
k &= \frac{(1-2\theta)^3}{3!} && \text{for } 0 \leq \theta \leq \tfrac{1}{2} \\
&= 0 && \text{for } \theta > \tfrac{1}{2} \\
s &= \frac{(1-3\theta)^4}{4!} && \text{for } 0 \leq \theta \leq \tfrac{1}{3} \\
&= 0 && \text{for } \theta > \tfrac{1}{3} \\
t &= \frac{(1-4\theta)^5}{5!} && \text{for } 0 \leq \theta \leq \tfrac{1}{4} \\
&= 0 && \text{for } \theta > \tfrac{1}{4}
\end{aligned} \qquad (10)$$

(c) Normal case:

Let X_i be normally distributed with mean $i\theta$ and variance 1. The values of s and t were obtained by numerical quadrature (see Note on computation below). Writing $h = -\theta/\sqrt{2}$, we have the following table of values for p, k, s and t.

h	p	k	s	t
0.0	0.500000	0.166667	0.0417	0.0083
0.1	0.460172	0.129582	0.0250	0.0036
0.2	0.420740	0.098216	0.0140	0.0013
0.3	0.382089	0.072488	0.0074	0.0004
0.4	0.344578	0.052037	0.0037	0.0001
0.5	0.308537	0.036298	0.0037	...
0.6	0.274253	0.024580	0.0017	...
0.7	0.241964	0.016145	0.0007	...
0.8	0.211855	0.010279	0.0003	...
0.9	0.184060	0.006338	...	...
1.0	0.158655	0.003782	...	...
1.5	0.066807	0.000172	...	...
2.0	0.022750	0.000003	...	...

(d) Estimates from the sample:

Where the distribution function is not specified we may use estimates of the four parameters obtained from the sample itself. We define a run of positive terms in the sequence of first differences as a set of consecutive positive terms not preceded or followed by a positive term; and define its 'length' by the number of terms in the run. Let N_r be the number of runs of positive terms of length r. Then unbiassed estimates of the parameters are given by

$$p = \frac{1}{n-1} \sum_{r=1}^{n-1} r N_r \tag{11}$$

$$k = \frac{1}{n-2} \sum_{r=2}^{n-1} (r-1) N_r$$

$$s = \frac{1}{n-3} \sum_{r=3}^{n-1} (r-2) N_r \qquad t = \frac{1}{n-4} \sum_{r=4}^{n-1} (r-3) N_r$$

A preliminary study of the Pearson curves suggested by the values of the third and fourth moments shows a great variation of the appropriate types of curves for different trends. For example, the third moment of the distribution under normal alternatives is negative up to $h = 0.6$ and becomes positive only thereafter. It therefore seems preferable to obtain points of the non-null distribution in any particular case by using Edgeworth's form of the Type A series (see e. g. Cramer [1]).

4. The C-test of two series

In [7] Stuart has proposed a sign test for the correlation of two series of observations X_i and Y_i $(i=1,2,\ldots n)$ based on the statistic

$$C = \sum_{i=1}^{n-1} C_i \tag{12}$$

where

$$C_i = 1 \text{ if } (X_i - X_{i+1})(Y_i - Y_{i+1}) > 0$$
$$= 0 \text{ otherwise.}$$

The general expressions for the D-test apply to this case also, when the parameters are now defined in terms of the expectations of the C_i's.

In the null case, when the two series are independent, we obtain quite simply

$$p = \frac{1}{2}\ ; \quad k = \frac{5}{18}\ ; \quad s = \frac{11}{72}\ ; \quad t = \frac{19}{225}$$

so that

$$\begin{aligned} E(c) &= \frac{n-1}{2} \\ \sigma^2(c) &= \frac{(11n-13)}{36} \\ \mu_3(c) &= 0 \\ \mu_4(c) &= \frac{(3,025n^2 - 9,912n + 8,423)}{10,800} \end{aligned} \qquad (13)$$

In the general case, these parameters occur in the formulae for the third and fourth moments of Kendall's rank correlation coefficient, as given by Sundrum (8). The quantities p and k are denoted by the same terms, while the quantities s and t are there denoted by 'l' and 'x' respectively. The distribution is asymptotically normal by virtue of the Hoeffding-Robbins Theorem.

5. Mean and Variance of the Turning-points test.

From (3), we have

$$E(T) = (n-2)u$$

where $u = E(T_i)$.

Further

$$T^2 = \left\{ \sum_{i=1}^{n-2} T_i \right\}^2$$

$$= \underset{(n-2)\ \text{terms}}{\Sigma T_i^{\,2}} + \underset{2(n-3)\ \text{terms}}{\Sigma T_i T_{i+1}} + \underset{2(n-4)\ \text{terms}}{\Sigma T_i T_{i+2}} + \underset{(n-4)(n-5)\ \text{terms}}{\Sigma T_i T_{i+r}} \qquad (r > 2).$$

Writing $v = E(T_i T_{i+1})$ and $w = E(T_i T_{i+2})$, we get after some simplification,

$$\begin{aligned} \sigma^2(T) &= (n-2)(u-u^2) + 2(n-3)(v-u^2) + 2(n-4)(w-u^2) \\ &= (u+2v+2w-5u^2)n + (16u^2-2u-6v-8w). \end{aligned} \tag{15}$$

In the null case, when all permutations of the sample observations are equally probable, we find $u = 2/3$; $v = 5/12$; $w = 9/20$. Substituting these in (14) and (15) we get (4). Under the hypothesis of linear trend, these parameters can be expressed in terms of p, k, s and t as follows:

$$\begin{aligned} u &= 2(p-k) \\ v &= (p+p^2 - 2k) \\ w &= (3p^2 + k - 4kp - 4s + 4t). \end{aligned} \tag{16}$$

Again, the distribution of T is asymptotically normal by the Hoeffding-Robbins Theorem.

6. Note on Computations

For the computations of 3(c) above, the Gaussian method of numerical quadrature was used. This is based on the formula

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} f(x)\, dx \approx \sum_{i=1}^{n} \alpha_i f(x_i) \tag{17}$$

where the x_i are the zeros of the nth order Hermite polynomial and the α_i are the corresponding weight factors. A table of the zeros and weight factors of the first twenty Hermite polynomials is given by Salzer, Zucker and Capuano [6] [1].

[1] I am indebted to Mr. Richard Savage of the National Bureau of Standards for giving me this reference.

As an illustration of the use of this method in the present problem, consider the evaluation of

$$t = \Pr (X_{-2} < X_{-1} < X_0 < X_1 < X_2)$$

where the X_i are normally distributed with mean $i\theta$ and variance 1. Then

$$t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x_0^2} \Pr (X_{-2} < X_{-1} < x_0 < X_1 < X_2) \, dx_0 \quad .$$

Now $\Pr (X_{-2} < X_{-1} < x_0 < X_1 < X_2)$ can be written as

$\Pr (X_{-2} < X_{-1} < x_0) \cdot \Pr (x_0 < X_1 < X_2)$ because the two probabilities are independent.

Write $u_1 = X_1 - x_0$

$u_2 = X_2 - X_1$

Then $\Pr (x_0 < X_1 < X_2) = \Pr (u_1 > 0;\ u_2 > 0)$.

Now u_1 and u_2 are jointly distributed in the bivariate normal form with correlation coefficient $\rho = \frac{1}{\sqrt{2}}$. Hence this probability can be determined by interpolation in K. Pearson's Tables [5] ; similarly for the other probability. Writing

$$A(x_i) = \Pr (X_{-2} < X_{-1} < x_i)$$

$$B(x_i) = \Pr (x_i < X_1 < X_2)$$

we have

$$t \approx \sum_{i=1}^{n} \alpha_i \, A(x_i)\, B(x_i) \quad .$$

REFERENCES

[1] Cramer, Harald, "Mathematical Methods of Statistics," Princeton, 1946.

[2] Hoeffding, Wassily and Robbins, Herbert E., "The Central Limit Theorem for Dependent Variables," Duke Mathematics Journal, 15 (1948), 773-80.

[3] Levene, Howard, "On the Power Function of Tests of Randomness Based on Runs Up and Down," Annals of Mathematical Statistics, 23 (1952), 34-56.

[4] Moore, Geoffrey H., and Wallis, W. Allen, "Time Series Significance Tests Based on Signs of Differences," Journal of the American Statistical Association, 38 (1943), 153-164.

[5] Pearson, Karl, ed., Tables for Statisticians and Biometricians, Part 2. London: Cambridge University Press, 1931.

[6] Salzer, Herbert E., Zucker, Ruth, and Capuano, Ruth, "Table of the Zeros and Weight Factors of the First Twenty Hermite Polynomials." Journal of Research of the National Bureau of Standards. 48 (1952), 111-6.

[7] Stuart, Alan, "The Power of Two Difference-Sign Tests," Journal of the American Statistical Association, 47 (1952), 416-24.

[8] Sundrum, R. M., "Moments of the Rank Correlation Coefficient t in the General Case," Biometrika, 40 (1953), 409-420.

[9] Wallis, W. A. and Moore, G. H., "A Significance Test for Time Series," Technical Report, National Bureau of Economic Research, 1941.