NULL DISTRIBUTION OF THE HODGES BIVARIATE SIGN TEST

BY JEROME KLOTZ[1]

University of California, Berkeley

0. Summary. This note presents the solution to the problem of obtaining the full null distribution for the bivariate sign test proposed by J. L. Hodges, Jr., in 1955 [1]. The partial solution of [1] is completed, and the table of [1] is extended to give the full distribution up to a sample size (n) of 30. In addition a partial table is included for sample size from 31 to 50.

1. Introduction. Using the notation given in [1], the problem is that of counting the number of cycles having a given value of K. This problem was solved in [1] only for the case $k < n/3$, or $n < 3h$ where $h = n - 2k$.

2. Counting the cycles. As stated in [1] the operation of rotation generates equivalence classes of cycles. We count the classes by selecting a representative member called a pattern. The number of cycles in each class is first determined. The total number of cycles for a given k value is then obtained by summing these numbers over all patterns corresponding to k.

To every cycle corresponds a walk in the plane. A plus sign corresponds to a step in the y direction, a minus sign to a step in the x direction. Let us call a point (x, y) a *departure point* for a path if it lies on the line $y = x$ (or $y = x + h$) and the path reaches the line $y = x + h$ ($y = x$) before returning to the line $y = x$ ($y = x + h$). Thus such points depend upon the given value of h and the particular path. Further, let us call the path between consecutive departure points a *flight*.

3. Specifying the patterns. To every cycle corresponds the particular cycle called a pattern which is obtained from the first by rotation and has the following properties:

(i) The minimum number of minus signs above the diameter (k) is attained for this cycle.

(ii) The cycle starts with a plus and ends the nth step with a plus.

(iii) The first and hence nth points are departure points.

To see the existence of such a pattern for a given cycle, let the cycle be rotated until Condition (i) is satisfied. Condition (ii) must also be satisfied, otherwise, using the fact that diametrically opposed signs are opposite, rotation by one would decrease k contrary to the initial rotation. If Condition (iii) is not satisfied, after an even number of steps the path from the first point (0, 0) returns to the line $y = x$. Thus, we can rotate the cycle so that the first point becomes a

Received February 16, 1959; revised April 17, 1959.

[1] This paper was prepared with the partial support of the Office of Ordnance Research U. S. Army under Contract DA-04-200-ORD-171, Task Order 3.

TABLE OF $P\,[K \leqq k]$

k

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00000																
2	1.00000																
3	.75000	1.00000															
4	.50000	1.00000															
5	.31250	.93750	1.00000														
6	.18750	.75000	1.00000														
7	.10938	.54688	.98438	1.00000													
8	.06250	.37500	.87500	1.00000													
9	.03516	.24609	.70312	.99609	1.00000												
10	.01953	.15625	.52734	.93750	1.00000												
11	.01074	.09668	.37598	.80566	.99902	1.00000											
12	.00586	.05859	.25781	.64453	.96875	1.00000											
13	.00317	.03491	.17139	.48877	.87280	.99975	1.00000										
14	.00171	.02051	.11108	.35547	.73315	.98437	1.00000										
15	.00092	.01190	.07050	.24994	.58319	.91675	.99994	1.00000									
16	.00049	.00684	.04395	.17090	.44434	.79980	.99219	1.00000									
17	.00026	.00389	.02698	.11414	.32684	.66095	.94551	.99998	1.00000								
18	.00014	.00220	.01634	.07471	.23346	.52295	.84984	.99610	1.00000								
19	.00007	.00123	.00978	.04805	.16264	.39922	.72450	.96433	.99999	1.00000							
20	.00004	.00069	.00580	.03044	.11089	.29572	.59143	.88737	.99805	1.00000							
21	.00002	.00038	.00340	.01903	.07420	.21347	.46575	.77626	.97666	1.00000	1.00000						
22	.00001	.00021	.00198	.01175	.04883	.15068	.35578	.65057	.91552	.99901	1.00000						
23	.00001	.00012	.00115	.00718	.03167	.10429	.26474	.52605	.81834	.98472	1.00000	1.00000					
24	.00000	.00006	.00066	.00434	.02027	.07094	.19254	.41259	.70140	.93665	.99951	1.00000					
25		.00003	.00038	.00260	.01282	.04750	.13723	.31517	.58020	.85252	.99000	1.00000	1.00000				
26		.00002	.00021	.00155	.00802	.03137	.09606	.23525	.46559	.74495	.95248	.99975	1.00000				
27		.00001	.00012	.00092	.00497	.02045	.06616	.17202	.36390	.62855	.88028	.99346	1.00000	1.00000			
28		.00001	.00007	.00054	.00305	.01318	.04491	.12350	.27789	51460*	.78222	.96436	.99987	1.00000			

29		.00000	.00004	.00031	.00186	.00841	.03008	.08722	.20786	.41040*	.67156	.90281	.99571	1.00000	1.00000		
30			.00002	.00018	.00112	.00531	.01991	.06067	.15263	.31979*	.55963	.81407	.97327	.99994	1.00000		
31	.00000	.00000	.00001	.00010	.00067	.00332	.01303	.04163	.11020								
32		.00000	.00001	.00006	.00040	.00206	.00750	.02821	.07837	.18286							
33			.00000	.00003	.00023	.00127	.00542	.01890	.05496	.13469							
34			.00000	.00002	.00014	.00078	.00344	.01253	.03805	.09770	.21371						
35			.00000	.00001	.00008	.00047	.00217	.00822	.02603	.06987	.16029						
36			.00000	.00001	.00005	.00029	.00136	.00534	.01761	.04932	.11836						
37				.00000	.00003	.00017	.00085	.00345	.01180	.03440	.08617	.18663					
38				.00000	.90002	.00010	.00052	.00220	.00783	.02372	.06191	.14009					
39				.00000	.00001	.00006	.00032	.00140	.00515	.01619	.04394	.10366					
40				.00000	.00001	.00004	.00020	.00088	.00336	.01094	.03084	.07569	.16260				
41					.00000	.00002	.00012	.00055	.00217	.00733	.02141	.05460	.12212				
42					.00000	.00001	.00007	.00034	.00140	.00487	.01472	.03893	.09052	.18567			
43					.00000	.00001	.00004	.00021	.00089	.00321	.01003	.02746	.06626	.14138			
44						.00000	.00003	.00013	.00056	.00210	.00677	.01918	.04795	.10624			
45						.00000	.00002	.00008	.00036	.00136	.00453	.01327	.03433	.07885	.16125		
46						.00000	.00001	.00005	.00022	.00088	.00301	.00910	.02433	.05785	.12272		
47						.00000	.00001	.00003	.00014	.00056	.00199	.00619	.01708	.04198	.09225	.18158	
48							.00000	.00002	.00009	.00036	.00130	.00417	.01188	.03016	.06854	.13983	
49							.00000	.00001	.00005	.00023	.00085	.00279	.00819	.02146	.05038	.10635	
50							.00000	.00001	.00003	.00014	.00055	.00186	.00561	.01513	.03665	.07997	.15743

* These values differ slightly from [1].

departure point without changing the number of plus or minus signs up. Conditions (i), (ii), and (iii) are now satisfied. We see that a pattern defined by the three conditions is not necessarily unique. Rotations of patterns taking departure points into like departure points may result in a different pattern. However, this need not concern us provided we count only one pattern for every class of patterns obtained by cyclic permutation. It may be noted that the definition of a pattern given here differs from that given in [1]. For example, [1] does not specify a pattern for cycles with alternating signs. However, under the restriction $k < n/3$, the two definitions are equivalent. For, under the restriction, a pattern satisfying the above conditions satisfies those of [1], and the pattern defined in [1] is unique.

4. Counting the cycles for a pattern. To count the number of cycles corresponding to a particular pattern we show that, to have less than $2n$ cycles for a pattern, the pattern must have an odd number (greater than one) of flights, which are all "equivalent." We specify two flights to be *equivalent* if they are identical, or if one can be obtained from the other by interchanging plus and minus signs (e.g., $++--++$ is equivalent ($\sim$) with $--+--$). Assume that after a rotation of less than $2n$ steps the pattern repeats itself—we take a pattern as a starting point for convenience. We must have departure points going into departure points of the same kind and hence the rotation consists of an even number of flights. Let us denote the flights by α_i and suppose there are $2t+1$ flights in the pattern $i = 1, 2, \cdots, 2t+1$. Next represent the cycle by $(\alpha_1, \alpha_2; \alpha_3, \alpha_4; \cdots; \alpha_{2t+1} \mid \alpha_{2t+2}; \cdots, \alpha_{4t+2})$. After a rotation which repeats the pattern—say a rotation of $2p$ flights we obtain

$$\begin{array}{l} \alpha_1 \sim \alpha_{2+p1} \sim \alpha_{4p+1} \sim \cdots \\ \vdots \\ \alpha_{2p} \sim \alpha_{4p} \sim \cdots \end{array}$$

where $\alpha_m = \alpha_{m(\mathrm{mod} 4t+3)}$. From the symmetry (diagonally opposed signs are opposite) we have $\alpha_1 \sim \alpha_{2t+2}, \alpha_2 \sim \alpha_{2t+3}, \cdots, \alpha_{2t+1} \sim \alpha_{4t+2}$. Solving the system, we obtain $\alpha_1 \sim \alpha_2 \sim \cdots \alpha_{2t+1} \cdots$. Thus for this case we have $2n/2t+1$ cycles for the pattern and otherwise $2n$ cycles.

5. Counting the patterns. To count the patterns we count them according to their number of flights—one, three, five, etc. For $(2l+1)h \leqq n < (2l+3)h$ we will have patterns with $1, 3, 5, \cdots, 2l+1$ flights. For the case of only one flight, [1] gives the formula for $2^n P[K = k] = 2nm_h(n)$. $m_h(n)$ is the number of ways of going from $(0, 0)$ to $(k, n-k)$ hitting the line $y = x + h$ only at the nth step—the gambler's ruin problem (see [1]). Generalizing, we obtain the formula for the case $(2l+1)h \leqq n < (2l+3)h$, where we have up to $2l+1$ flights

$$2^n P[K = k] = 2nm_h(x) + \frac{2n}{3}\left[\binom{3}{1, 1} \sum_{\substack{n_1 < n_2 < n_3 \\ n_1+n_2+n_3=n}} \prod_{i=1}^{3} m_h(n_i) + \binom{3}{2} \sum_{\substack{n_1 = n_2 \neq n_3 \\ n_1+n_2+n_3=n}} \prod^{3} m_h(n_i)\right]$$

$$+ I_{[3a=n]} \frac{2n}{3}\left[\binom{3}{1, 1}\binom{m_h(a)}{3} + 2\binom{3}{2}\binom{m_h(a)}{2} + \binom{m_h(a)}{1}\right] + \cdots$$

$$+ \frac{2n}{2l+1}\left[\binom{2l+1}{1, 1, \cdots, 1} \sum_{n_1<\cdots<n_{2l+1}} \prod_{i=1}^{2l+1} m_h(n_i) + \binom{2l+1}{2, 1, \cdots, 1} \sum \prod m_h(n_i)\right.$$

$$+ \cdots + \binom{2l+1}{r_1, r_2, \cdots, r_{t-1}} \sum \prod m_h(n_i) + \cdots$$

$$\left. + \binom{2l+1}{2l} \sum_{\substack{n_1=\cdots=n_{2l+1} \\ \sum n_i = n}} \prod m_h(n_i)\right] + I_{[(2l+1)q=n]} \frac{2n}{2l+1}\left[\binom{2l+1}{1, 1, \cdots, 1}\binom{m_h(q)}{2l+1}\right.$$

$$+ 2l\binom{2l+1}{2, 1, \cdots, 1}\binom{m_h(q)}{2l} + \cdots + (t)_{p-1}\binom{2l+1}{r_1, r_2, \cdots, r_{t-1}}\binom{m_h(q)}{t} + \cdots$$

$$\left.\cdots \cdots + \binom{m_h(q)}{1}\right]$$

where p is the number of different r_i; $\sum_{i=1}^{t} r_i = n$; $a, \cdots, q$; r_i are integers; $\binom{n}{k_1 \cdots k_{t-1}} = n!/k_1! \cdots k_t!$ is the multinomial coefficient; and I is an indicator function.

The preceding table completes the table of [1] and gives values of $P[K \leqq k]$ to $5D$ for all values of k and $n = 1(1)30$. Further the table gives values of P to $5D$ for k up to a value which makes P just greater than 10 per cent and $n = 31(1)50$.

6. Acknowledgment. I wish to thank Professor J. L. Hodges, Jr. for suggesting a method of attack that led to the solution.

REFERENCE

[1] J. L. Hodges, Jr., "A bivariate sign test," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 523–527.