SMALL SAMPLE POWER OF THE BIVARIATE SIGN TESTS OF BLUMEN AND HODGES[1]

By Jerome Klotz

Harvard University

0. Summary. Exact power for normal location alternatives is obtained for the bivariate sign tests of Blumen [1] and Hodges [5]. A recursive scheme, used in conjunction with a computer, permits comparison of the two tests for sample sizes $n = 8(1)12$. Efficiency values relative to Hotelling's bivariate T^2 test are also obtained for the test of Hodges. Slight power differences are noted for the sign tests along with surprisingly high power when compared with the T^2.

1. Introduction and notation. One of the drawbacks of a nonparametric approach to statistics is a general lack of multivariate procedures. A preliminary step towards overcoming this deficiency was made by Hodges in 1955 with a bivariate sign test. In 1958, Blumen proposed a different bivariate sign test and thus raised the question as to which is to be preferred.

While both tests are applicable in a wider class of problems, to define the statistics in a simple way, let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be a sample of n independent bivariate observations from a continuous distribution which has a bivariate median (μ, ν). For testing the hypothesis

$$H: (\mu, \nu) = (0, 0) \quad \text{against} \quad K: (\mu, \nu) \neq (0, 0)$$

the test of Hodges uses the maximum number of observations with positive projections on some directed line through the origin. If we define θ_i to be the angle that the directed vector (X_i, Y_i) makes with the X-axis $(0 \leqq \theta_i < 2\pi)$, and define

$$\begin{aligned} \theta_i^* &= \theta_i, \quad \text{if} \quad 0 \leqq \theta_i < \pi \\ &= \theta_i - \pi, \quad \text{if} \quad \pi \leqq \theta_i < 2\pi \end{aligned} \qquad \text{for } i = 1, 2, \cdots, n,$$

then we can define a bivariate sign configuration $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_n)$ in terms of the ordered values of $\theta_{(1)}^* < \theta_{(2)}^* < \cdots < \theta_{(n)}^*$. We let

$$\mathbf{Z}_i = \begin{cases} 1 \\ 0 \end{cases} \text{if} \quad \theta_{(i)}^* = \begin{cases} \theta_j \\ \theta_j - \pi \end{cases} \text{for some} \quad j,$$

and thus the vector $\mathbf{Z}$ tells which of the observations ordered by means of the angles θ_i^* lie above or below the X-axis. In terms of this notation, the statistic

Received June 30, 1964.

[1] This research has been supported in part by the United States Navy through the Office of Naval Research under contract Nonr-1866(37). Computations were performed at the Harvard Computing Center under National Science Foundation grant N.S.F. GP-683.

of Blumen is given by:

$$B(\mathbf{Z}) = \left[\sum_{i=1}^{n} (2Z_i - 1) \cos\left(\frac{\pi(i-1)}{n}\right)\right]^2 + \left[\sum_{i=1}^{n} (2Z_i - 1) \sin\left(\frac{\pi(i-1)}{n}\right)\right]^2 \tag{1.1}$$

which is intuitively the center of gravity of a circular configuration as described in [1], p. 451 and [5], p. 524. For purposes of comparison, the statistic of Hodges is equivalent to

$$H(\mathbf{Z}) = 2 \max_{k=0,\cdots,2n-1} \left(\sum_{i=1}^{n} \zeta_i^{(k)}(\mathbf{Z})\right) - n. \tag{1.2}$$

The vector $\boldsymbol{\zeta}^{(k)}(\mathbf{Z}) = (\zeta_1^{(k)}(\mathbf{Z}), \cdots, \zeta_n^{(k)}(\mathbf{Z}))$ is a kind of rotation with complementation defined by taking $\boldsymbol{\zeta}^{(0)}$ to be the identity, $\boldsymbol{\zeta}^{(k)}(\mathbf{Z}) = \boldsymbol{\zeta}(\boldsymbol{\zeta}^{(k-1)}(\mathbf{Z}))$, and $\boldsymbol{\zeta}(\mathbf{Z}) = (Z_2, Z_3, \cdots Z_n, 1 - Z_1)$. Thus, for example $\boldsymbol{\zeta}^{(2)}(11101) = (10100)$. The expression (1.2) thus gives the dependence of H upon the bivariate sign configuration $\mathbf{Z}$. From (1.1) and (1.2) we see that it is sufficient to compute $P[\mathbf{Z} = \mathbf{z}]$ for all 2^n configurations $\mathbf{z}$ of 0's and 1's in order to evaluate the power of both tests.

2. Power calculations. To compute $P[\mathbf{Z} = \mathbf{z}]$, we modify further a recursive scheme of Hodges used in [8], p. 502, [9], p. 625 and described in more detail in [10]. In the present context, define $A_{\mathbf{z}}^n(u) = P[\mathbf{Z} = \mathbf{z}$ and all $\theta_i^* \leqq u]$. Then $P[\mathbf{Z} = \mathbf{z}] = A_{\mathbf{z}}^n(\pi)$ is the desired probability and the recursive scheme is given by:

$$A_{(\mathbf{z},1)}^{n+1}(u) = (n+1) \int_0^u A_{\mathbf{z}}^n(x)\, dF_\theta(x) \tag{2.1}$$

$$A_{(\mathbf{z},0)}^{n+1}(u) = (n+1) \int_0^u A_{\mathbf{z}}^n(x)\, dF_\theta(\pi + x) \tag{2.2}$$

with $(\mathbf{z}, 1) = (z_1, z_2, \cdots z_n, 1)$ and similarly for $(\mathbf{z}, 0)$. Here F_θ is the distribution of the angle θ_i, and $A^0 \equiv 1$.

The most interesting alternatives for carrying out the computations are the bivariate normal densities $N((\mu, \nu), \Sigma)$ which are used to shed light on the question of preference. Such alternatives permit comparisons with the natural parametric competitor, the Hotelling bivariate T^2. For convenience, since the power of the T^2 test depends only on $\Delta^2 = (\mu, \nu)\Sigma^{-1}(\mu, \nu)'$ we assume that alternatives are $N((0, \Delta), I)$. This is permitted since B, H, and T^2 are invariant under linear transformations. Under this assumption, the angular density becomes

$$f(\theta) = [(2\pi)^{-1} \exp(-\Delta^2/2)] + \Delta \sin\theta\phi(\Delta\cos\theta)\Phi(\Delta\sin\theta),\ 0 \leqq \theta < 2\pi \tag{2.3}$$

where ϕ is the standard normal density and Φ its cumulative. The well known formula for Mill's ratio (see for example [3], p. 166) can be used to show $f \geqq 0$.

Using (2.1), (2.2), (2.3) with the boundary condition $A^0 \equiv 1$, all probabilities were calculated stage by stage for $n \leqq 12$ on a digital computer (I.B.M. 7094). Accuracy was checked by summing the probabilities to one (to six decimals) and checking the values with 2^{-n} for $\Delta = 0$. Further checks were made on isolated values using symmetry as well as relations such as $P[\mathbf{Z} = (1, 1, \cdots, 1)] = \Phi^n(\Delta)$ with agreement to seven decimals.

Power was obtained for each of the tests by summing the probabilities over those $\mathbf{z}$ values in the rejection regions determined by large values of (1.1) and (1.2) respectively.

3. Description of Tables I–V. Table I gives, for comparison, the rejection regions of both tests for small α levels by listing the most extreme $\mathbf{z}$ vectors according to decreasing values of H and B respectively. To conserve space, the vectors of 0's and 1's are considered as binary numbers and converted to octal. The octal base was chosen because of common usage with binary machines. Because of the rotational symmetry we note that if $\mathbf{z} \varepsilon R$ where R is a nonrandomized rejection region, we also have $\zeta^{(k)}(\mathbf{z}) \varepsilon R$ for $k = 0, 1, 2, \cdots, 2n - 1$. Thus in Table I the vector $\mathbf{z}$ represents the equivalence class $\{\mathbf{z}, \zeta(\mathbf{z}), \cdots, \zeta^{(2n-1)}(\mathbf{z})\}$. For example with $n = 8$, the octal number 377 represents the 16 vectors $\{(11111111), (11111110), (11111100), \cdots, (01111111)\}$. Italicized vectors indicate natural α levels corresponding to distinct values of the statistic. Thus for example the use of italic for 371 in column 2 of Table I indicates that this value and those above it form a nonrandomized rejection region (371, 373, 375, 377) for B with corresponding natural significance level $\alpha = 4(16/2^8)$. Sample sizes of at least $n = 8$ were given to avoid trivial comparisons while speed and convenience set an upper limit of $n = 12$.

Tables II and III give power for B and H at natural α levels for $n = 8(1)12$ and bivariate normal alternatives with $\Delta = 0$, .50, .75, 1.00, 1.50, 2.00, and 3.00. The accuracy is believed to be within one unit in the last decimal place.

Table IV compares the two tests at selected significance levels. For $n = 11$ both tests have the same natural significance level $\alpha = 9(18/2^{11}) \doteq .09668$ with different nonrandomized rejected regions—as seen in Table I, columns 7 and 8. Since the probability associated with 3765 becomes less than that of 3737 between $\Delta = .75$ and 1.00, the power curves cross here. Difficulties arise for other values of n since the nonrandomized levels do not correspond. Since the type I $-$ type II error curve is convex, a slight disadvantage must be imposed on one test when randomization is used to match α levels. Because of its greater number of nonrandomized significance levels, in order to minimize this disadvantage randomization was performed on B to match the natural α levels of H. With this choice, fewer $\mathbf{z}$ vectors are involved in the randomization and similar results hold for $n = 10$, 12 as for $n = 11$.

Table V compares H and T^2 by giving power of the T^2 test at sample sizes needed to bracket the power of the Hodges test. As the significance levels are nonstandard, power of the T^2 test was obtained by machine using a noncentral F routine developed by S. P. Ghosh. The routine, checked by Ghosh, was also checked by the author for a few values using the formulas of Nicholson ([11], p. 609). Efficiency values as defined by Hodges and Lehmann ([6], p. 329) are given for rough comparison and are obtained by linear interpolation in the sample size for T^2 using the power. For example for $n_H = 8$, $\alpha = .06250$, $\Delta = .50$, we compute

$$e_{H,T^2} \doteq .940 = [\lambda 7 + (1 - \lambda)8]/8$$

where $\lambda = (.1839 - .1721)/(.1839 - .1593)$.
The efficiency appears to be a decreasing function of n, α, and Δ.

TABLE I

Rejection Regions $n = 8(1)12$

$n = 8$		9		10		11		12	
H	*B*	*H*	*B*	*H*	*B*	*H*	*B*	*H*	*B*
377	*377*	*777*	*777*	*1777*	*1777*	*3777*	*3777*	*7777*	*7777*
375	*375*	775	*775*	1775	*1775*	3775	*3775*	7775	*7775*
373	373	773	773	1773	1773	3773	3773	7773	7773
371	*371*	771	*771*	1771	*1771*	3771	*3771*	7771	*7771*
367	367	767	767	1767	1767	3767	3767	7767	7767
361	*361*	761	*761*	1761	*1761*	3761	*3761*	7761	*7761*
		741	741	1757	*1765*	3757	*3765*	7757	*7765*
			765	*1741*	1757	3741	3757	7741	7757
					1741	*3737*	*3741*	7737	*7741*
								7701	7755
									7751
									7737
									7701

TABLE II

Power of the Hodges sign test $P[K \leqq k]$ *where* $H = n - 2K$

n	k	$\Delta = 0$ (α)	0.50	0.75	1.00	1.50	2.00	3.00
8	0	.06250	.17207	.31100	.48545	.79898	.94907	.99863
	1	.37500	.58389	.75177	.88047	.98518	.99913	1.00000
9	0	.03516	.12421	.25085	.42419	.76473	.93912	.99836
	1	.24609	.47201	.66988	.83355	.97791	.99863	.99999
10	0	.01953	.08940	.20170	.36966	.73093	.92889	.99808
	1	.15625	.37649	.59068	.78363	.96923	.99806	.99999
11	0	.01074	.06417	.16173	.32138	.69778	.91841	.99779
	1	.09668	.29692	.51598	.73197	.95917	.99734	1.00000
12	0	.00586	.04596	.12937	.27882	.66539	.90773	.99749
	1	.05859	.23194	.44704	.67974	.94780	.99648	1.00000

TABLE III

Power of the Blumen sign test $P[B \geqq b]$

n	b	$\Delta = 0.00$	0.50	0.75	1.00	1.50	2.00	3.00
8	26.27	.06250	.17207	.31100	.48545	.79898	.94907	.99863
	18.88	.12500	.28429	.45577	.63733	.88860	.97579	.99934
	13.23	.25000	.44699	.62387	.77945	.94578	.98925	.99970
	10.17	.37500	.58389	.75177	.88047	.98518	.99913	1.00000
9	33.16	.03516	.12421	.25085	.42419	.76473	.93912	.99836
	25.65	.07031	.20712	.37315	.56615	.85979	.96906	.99915
	19.52	.14063	.32768	.51703	.70129	.92087	.98384	.99955
	15.52	.21094	.42610	.62231	.79283	.96031	.99400	.99985
	14.13	.28125	.51523	.70955	.86106	.98350	.99904	.99999
10	40.86	.01953	.08940	.20170	.36966	.73093	.92889	.99808
	33.26	.03906	.15043	.30424	.50103	.83055	.96204	.99897
	26.78	.07813	.23991	.42700	.62881	.89545	.97826	.99939
	22.08	.11719	.31197	.51526	.71323	.93565	.98890	.99971
	20.94	.13672	.34433	.55003	.74042	.94210	.98941	.99971
	19.61	.17578	.40885	.62545	.81082	.97568	.99857	1.00000
11	49.37	.01074	.06417	.16173	.32138	.69778	.91841	.99779
	41.59	.02148	.10890	.24710	.44188	.80108	.95473	.99877
	34.97	.04297	.17520	.35122	.56176	.86958	.97245	.99923
	29.73	.06445	.22832	.42558	.63998	.91080	.98367	.99957
	28.78	.07520	.25258	.45587	.66657	.91812	.98429	.99957
	26.41	.09668	.29900	.51735	.72927	.95106	.99357	.99985
12	58.70	.00586	.04596	.12937	.27882	.66539	.90773	.99749
	50.97	.01172	.07862	.20000	.38852	.77156	.94716	.99857
	44.04	.02344	.12758	.28776	.50012	.84332	.96644	.99907
	38.38	.03516	.16680	.35038	.57254	.88564	.97830	.99943
	37.58	.04102	.18495	.37656	.59824	.89382	.97905	.99943
	34.38	.05273	.21873	.42730	.65482	.92658	.98857	.99973
	32.31	.07617	.27885	.50932	.73638	.96372	.99783	1.00000

TABLE IV

Hodges-Blumen power comparisons

n	Statistic	$\Delta = 0.00$ (α)	0.50		0.75		1.00	1.50	2.00	3.00
10	H	.15625	.37649	†	.59068		.78363	.96923	.99806	.99999
	B	.15625*	.37659		.58774		.77562	.95889	.99399	.99986
11	H	.09668	.29692		.51598	†	.73197	.95917	.99734	1.00000
	B	.09668	.29900		.51735		.72927	.95106	.99357	.99985
12	H	.05859	.23194		.44704	†	.67974	.94780	.99648	1.00000
	B	.05859*	.23376		.44781		.67521	.93586	.99089	.99980

* Randomized significance level.

† Cross over point.

TABLE V

Hodges—T^2 power, efficiency

Test	Sample size	Δ = 0.00 (α)	0.50	0.75	1.00	1.50	2.00	3.00
H	8	.06250	.17207	.31100	.48545	.79898	.94907	.99863
T^2	6	.06250					.8693	.9945
T^2	7	.06250	.1593	.2855	.4521	.7805	.9509	.9995
T^2	8	.06250	.1839	.3418	.5400	.8688		
e_{H,T^2}			.940	.932	.923	.901	.872	.853
H	9	.03516	.12421	.25085	.42419	.76473	.93912	.99836
T^2	7	.0352					.8815	.9970
T^2	8	.0352	.1162	.2348	.4048	.7644	.9520	.9998
T^2	9	.0352	.1354	.2832	.4864	.8516		
e_{H,T^2}			.935	.926	.915	.889	.869	.833
H	10	.01953	.08940	.20170	.36966	.73093	.92889	.99808
T^2	7	.0195						.9855
T^2	8	.0195				.6353	.8905	.9983
T^2	9	.0195	.0849	.1932	.3627	.7491	.9530	
T^2	10	.0195	.0994	.2342	.4376			
e_{H,T^2}			.931	.921	.909	.884	.861	.798
H	11	.01074	.06417	.16173	.32138	.69778	.91841	.99779
T^2	8	.0107						.9919
T^2	9	.0107			.2577	.6259	.8975	.9990
T^2	10	.0107	.0619	.1587	.3247	.7343	.9538	
T^2	11	.0107	.0730	.1936				
e_{H,T^2}			.928	.917	.905	.878	.852	.803
H	11	.09668	.29692	.51598	.73197	.95917	.99734	1.0000
T^2	8	.0967				.9265	.9940	1.0000
T^2	9	.0967	.2850	.4990	.7196	.9608	.9983	1.0000
T^2	10	.0967	.3139	.5523	.7782			
e_{H,T^2}			.856	.847	.837	.813	.797	—
H	12	.00586	.04596	.12937	.27882	.66539	.90773	.99749
T^2	9	.0059						.9953
T^2	10	.0059		.1039	.2307	.6161	.9032	.9994
T^2	11	.0059	.0452	.1305	.2911	.7203	.9547	
T^2	12	.0059	.0535					
e_{H,T^2}			.925	.913	.900	.873	.841	.795
H	12	.05859	.23194	.44704	.67974	.94780	.99648	1.0000
T^2	9	.0586				.9197	.9939	1.0000
T^2	10	.0586	.2239	.4362	.6745	.9540	.9982	1.0000
T^2	11	.0586	.2488	.4873	.7353			
e_{H,T^2}			.860	.851	.840	.818	.800	—

4. Conclusions. For the small samples investigated, it is seen that the two sign tests have quite similar rejection regions with little difference in power. It appears roughly that the test of Blumen behaves better locally with the test of Hodges more powerful in the region of higher power. The conjectures of Blumen ([1], p. 455) concerning relative performance, which stimulated much of this work, appear not to hold for the small samples covered. The greater number of nonrandomized significance levels for B must be balanced against the simplified null distribution of H (see [4] or [7]). The slightly greater power in the region of interest may suggest preference for H.

The efficiency of H relative to T^2 seems surprisingly high (94–80% in the region covered). The results seem comparable (even slightly better) to those in the univariate case as given by Dixon [2].

Asymptotic results seem difficult to obtain for the two tests. The heuristic derivation of Blumen appears to be difficult to make precise (at least for the author). The establishment of asymptotic equivalence of the variables used in the argument in [1], p. 455, along with the complicated dependence of the components of $\mathbf{Z}$ under the alternative contribute to this difficulty. Even if the numbers $2/\pi$ for H ([7], p. 806) and $\pi/4$ for B ([1], p. 455) could be derived as a limiting Pitman efficiency relative to T^2 (limiting ratio of sample sizes as type I and II errors converge to α, β where $0 < \alpha, \beta < 1$), then the efficiency ratio (say $8/\pi^2$) may well be misleading because of the local nature of Pitman efficiency and the small sample behavior of the two tests.

5. Acknowledgments. I wish to thank Mike Feuer for programming and Sakti Ghosh for providing the noncentral F routine. Thanks go also to Professors Blumen, Dempster, Mosteller, and Hodges for helpful comments and criticisms.

REFERENCES

[1] Blumen, I. (1958). A new bivariate sign test. *J. Amer. Statist. Assoc.* **53** 448–456.

[2] Dixon, W. J. (1953). Power functions of the sign test and power efficiency for normal alternatives. *Ann. Math. Statist.* **24** 467–473.

[3] Feller, W. (1957). *An Introduction to Probability Theory and its Applications* **1** (2nd ed.). Wiley, New York.

[4] Hill, B. M. (1960). A relationship between Hodges' bivariate sign test and a nonparametric test of Daniels. *Ann. Math. Statist.* **31** 1190–1192.

[5] Hodges, J. L., Jr. (1955). A bivariate sign test. *Ann. Math. Statist.* **26** 523–527.

[6] Hodges, J. L., Jr. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27** 324–335.

[7] Joffe, A. and Klotz, J. (1962). Null distribution and Bahadur efficiency of the Hodges bivariate sign test. *Ann. Math. Statist.* **33** 803–807.

[8] Klotz, J. (1962). Nonparametric tests for scale. *Ann. Math. Statist.* **33** 498–512.

[9] Klotz, J. (1963). Small sample power and efficiency for the one sample Wilcoxon and normal scores tests. *Ann. Math. Statist.* **34** 624–632.

[10] Klotz, J. (1964). On the normal scores two sample rank test. *J. Amer. Statist. Assoc.* **59** 652–664.

[11] Nicholson, W. L. (1954). A computing formula for the power of the analysis of variance test. *Ann. Math. Statist.* **25** 607–610.