NOTES

POWER FUNCTIONS OF THE SIGN TEST AND POWER EFFICIENCY FOR NORMAL ALTERNATIVES[1]

By W. J. Dixon

University of Oregon

1. Summary. Power functions are tabulated for the sign test for various sample sizes and α near .05 and .01. Several of these power functions are compared with the power function of the t-test for samples from normal populations by means of a power efficiency function. The results indicate decreasing power efficiency for increasing sample size, for increasing level of significance and for increasing alternative.

2. Power function. The power of the two-sided sign test for level of significance, α, is given by:

$$\lambda(p) = \sum_{j=0}^{i} \binom{N}{j} [p^j(1 - p)^{N-j} + p^{N-j}(1 - p)^j] \tag{1}$$

where i is the largest integer such that

$$\sum_{j=0}^{i} \binom{N}{j} (1/2)^N \leqq (1/2)\alpha \tag{2}$$

and N is considered fixed [5]. Here, p is the alternative population proportion. Values for $\lambda(p)$ may be obtained readily from a table of the cumulative binomial [1] or tables of the incomplete beta function [2] since

$$\sum_{j=0}^{i} \binom{N}{j} x^j(1 - x)^{N-j} = I_{1-x}(N - i, i + 1).$$

Beyond the range of these tables the approximation of Camp [3] can be used with great accuracy. The maximum i which satisfies (2) is tabulated as r in Table I of reference [5] for $\alpha = .01$ and $\alpha = .05$. Tables I and II of this paper give the power for these critical values. Since $p = .50$ is the null hypothesis, the values in the column headed $p = .50$ in Tables I and II of this paper give the actual level of significance ($\leqq .01$ or $\leqq .05$) of each test. At the foot of the tables are the normal alternatives corresponding to the alternative p, that is, δ is defined by the relation $1 - F(\delta) = p$ where $F(x)$ is the cumulative zero mean unit variance normal distribution. For normal alternatives Tables I and II may be entered either with p or δ. For nonnormal alternatives the tables must, of course, be entered with p.

Received 12/15/52.

[1] Research sponsored by the Office of Naval Research.

TABLE I

Power for Sign Test ($\alpha \leqq .05$)

p		.50	.45 .55	.40 .60	.35 .65	.30 .70	.25 .75	.20 .80	.15 .85	.10 .90	.05 .95
N	r										
(5)	(0)	.06250	.06878	.08800	.12128	.17050	.23828	.32800	.44378	.59050	.77378
6	0	.03125	.03598	.05075	.07726	.11838	.17822	.26221	.37716	.53144	.73509
7	0	.01562	.01896	.02963	.04967	.08257	.13354	.20973	.32058	.47830	.69834
8	0	.00781	.01005	.01745	.03209	.05771	.10013	.16777	.27249	.43047	.66342
9	1	.03906	.04760	.07435	.12248	.19644	.30045	.43623	.59948	.77484	.92879
10	1	.02148	.02776	.04804	.08649	.14945	.24405	.37581	.54430	.73610	.91386
(10)	(2)	.10938	.12695	.17958	.26643	.38437	.52601	.67788	.82021	.92981	.98850
11	1	.01172	.01614	.03097	.06079	.11304	.19711	.32212	.49219	.69736	.89811
12	2	.03857	.05002	.08625	.15214	.25302	.39071	.55835	.73582	.88913	.98043
13	2	.02246	.03105	.05922	.11354	.20255	.33261	.50165	.69196	.86612	.97549
14	2	.01294	.01916	.04040	.08407	.16086	.28113	.44805	.64791	.84164	.96995
15	3	.03516	.04875	.09243	.17318	.29696	.46130	.64816	.82266	.94444	.99453
16	3	.02127	.03158	.06609	.13406	.24589	.40499	.59813	.78989	.93159	.99300
17	4	.04904	.06820	.12852	.23542	.38879	.57390	.75822	.90129	.97786	.99884
18	4	.03088	.04598	.09545	.18888	.33269	.51867	.71635	.87944	.97181	.99845
19	4	.01921	.03074	.07025	.15011	.28224	.46543	.67329	.85556	.96481	.99799
(20)	(4)	.01182	.02039	.05127	.11825	.23751	.41484	.62965	.82985	.95683	.99743
20	5	.04139	.06177	.12721	.24571	.41641	.61718	.80421	.93269	.98875	.99967
(20)	(6)	.11532	.15135	.25648	.41815	.60827	.78581	.91331	.97806	.99761	.99997
25	7	.04329	.06968	.15476	.30626	.51187	.72651	.89088	.97453	.99774	.99998
30	9	.04277	.07442	.17714	.35764	.58882	.80341	.93891	.99034	.99955	1.00000–
35	11	.04096	.07712	.19577	.40198	.65156	85789	.96564	.99633	.99991	1.00000–
40	13	.03848	.07848	.21156	.44079	.70325	.89776	.98059	.99860	.99998	1.00000–
45	15	.03570	.07894	.22517	.47519	.74622	.92470	.98900	.99946	1.00000–	
50	17	.03284	.07877	.23706	.50598	.78219	.94488	.99374	.99980	1.00000–	
60	20	.02734	.07722	.25689	.55903	.83818	.97020	.99796	.99997	1.00000–	
70	26	.04139	.11635	.36009	.69503	.92220	.99163	.99974	1 00000–		
80	30	.03299	.10895	.36877	.72353	.94125	.99542	.99991	1.00000–		
90	35	.04460	.14612	.46008	.81223	.97256	.99878	1.00000–			
100	39	.03520	.13519	.46206	.82758	.97900	.99932	1.00000–			
Normal alternatives											
δ		0	.1257	.2534	.3853	.5244	.6745	.8416	1.0364	1.2816	1.6449
$\sqrt{2}\,\delta$		0	.1777	.3583	.5449	.7416	.9539	1.1902	1.4657	1.8124	2.3262

3. Power efficiency. Discussion of the power of the sign test for normal alternatives was given in [4] for large N. This paper obtains 100 $(2/\pi)$ = 63.7 per cent as the efficiency. Reference [5], by a rough coincidence of the power function of the sign test for a sample of N observations with the power function of the t-test for some smaller sample size obtained ratios of sample sizes of .67 for $N = 18$ and .65 for $N = 44$. This ratio of sample sizes is defined to be power efficiency by Walsh [12]. He defines as equivalent power curves those whose average height is the same. For finite N an essential difficulty arises, since the power curves

TABLE II

Power for Sign Test ($\alpha \leqq .01$)

p		.50	.45 .55	.40 .60	.35 .65	.30 .70	.25 .75	.20 .80	.15 .85	.10 .90	.05 .95
N	r										
8	0	.00781	.01005	.01745	.03209	.05771	.10013	.16777	.27249	.43047	.66342
9	0	.00391	.00536	.01034	.02079	.04037	.07509	.13422	.23162	.38742	.63025
10	0	.00195	.00287	.00615	.01349	.02825	.05631	.10737	.19687	.34868	.59874
11	0	.00098	.00155	.00367	.00876	.01978	.04224	.08590	.16734	.31381	.56880
12	1	.00635	.00937	.01991	.04252	.08504	.15838	.27488	.44346	.65900	.88164
13	1	.00342	.00543	.01276	.02961	.06367	.12671	.23365	.39828	.62135	.86458
14	1	.00183	.00314	.00816	.02053	.04748	.10097	.19791	.35667	.58463	.84701
15	2	.00739	.01176	.02739	.06179	.12684	.23609	.39802	.60423	.81594	.96380
16	2	.00418	.00719	.01846	.04511	.09936	.19711	.35184	.56138	.78925	.95706
17	2	.00235	.00437	.01238	.03273	.07739	.16370	.30962	.51976	.76180	.94975
18	3	.00754	.01298	.03300	.07830	.16455	.30569	.50103	.72024	.90180	.98913
19	3	.00443	.00825	.02306	.05915	.13317	.26309	.45509	.68415	.88500	.98676
20	3	.00258	.00521	.01601	.04438	.10709	.22516	.41145	.64773	.86705	.98410
25	5	.00408	.00898	.02942	.08263	.19349	.37828	.61669	.83848	.96660	.99879
30	7	.00522	.01253	.04357	.12377	.28138	.51429	.76079	.93022	.99222	.99992
35	9	.00599	.01565	.05757	.16510	.36458	.62632	.85427	.97082	.99826	.99999
40	11	.00643	.01829	.07098	.20528	.44061	.71514	.91249	.98803	.99962	1.00000–
45	13	.00661	.02048	.08365	.24370	.50875	.78408	.94793	.99515	.99992	1.00000–
50	15	.00660	.02226	.09552	.28010	.56918	.83692	.96920	.99805	.99998	1.00000–
60	19	.00622	.02483	.11697	.34678	.66916	.90752	.98933	.99969	1.00000–	
70	23	.00558	.02637	.13568	.40577	.74592	.94774	.99633	.99995	1.00000–	
80	28	.00968	.04485	.21312	.55120	.86331	.98337	.99945	1.00000–		
90	32	.00743	.04566	.22674	.59138	.89569	.99075	.99982	1.00000–		
100	36	.00664	.04300	.23868	.62692	.92011	.99482	.99994	1.00000–		
Normal alternatives											
δ		0	.1257	.2534	.3853	.5244	.6745	.8416	1.0364	1.2816	1.6449
$\sqrt{2}\,\delta$		0	.1777	.3583	.5449	.7416	.9539	1.1902	1.4657	1.8124	2.3262

differ in shape. The equivalence "by sight" or by an averaging process disguises these differences in shape. It would seem more realistic to define a *power efficiency function* which gives the power efficiency for each alternative. This function has been obtained for the sign test for $N = 5, 10, 20$ and is given in Fig. 1. The power function of the test was compared for α corresponding to particular exact

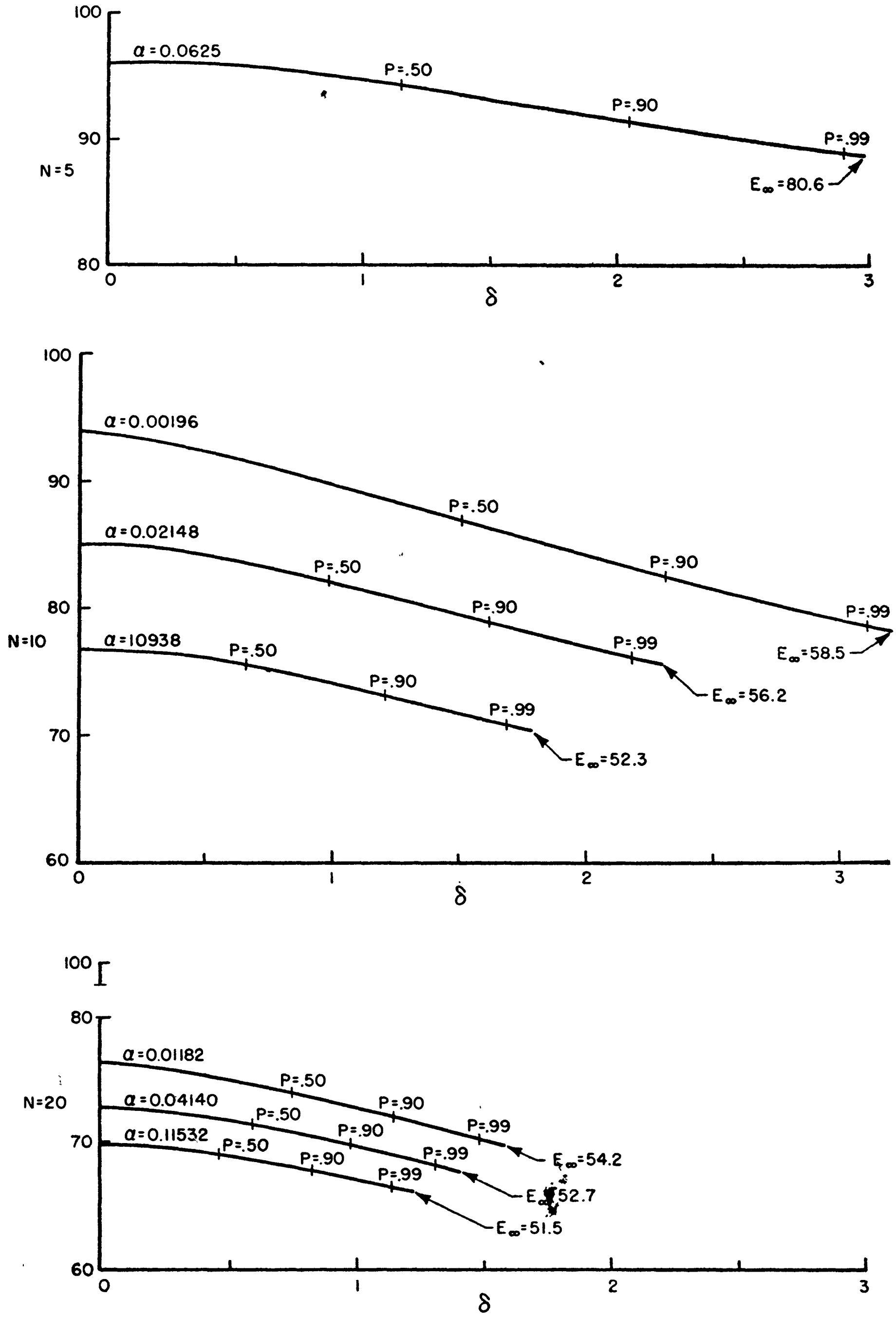
100
90
80
N=5
α=0.0625
P=.50
P=.90
P=.99
E∞=80.6
0
1
2
3
δ
100
90
80
70
60
N=10
α=0.00196
P=.50
P=.90
P=.99
E∞=58.5
α=0.02148
P=.50
P=.90
P=.99
E∞=56.2
α=10938
P=.50
P=.90
P=.99
E∞=52.3
0
1
2
3
δ
100
80
70
60
N=20
α=0.01182
P=.50
P=.90
P=.99
E∞=54.2
α=0.04140
P=.50
P=.90
P=.99
α=0.11532
P=.50
P=.90
P=.99
E∞=51.5
0
1
2
3
δ

FIG. 1

values of α for the sign test. These curves show decreasing power efficiency for increasing N, for increasing α, and for increasing alternative. The alternative, δ, is a shift in mean in standard deviation units of observation differences. For samples from two normal populations with means μ_1 and μ_2 and standard deviations both equal to σ we have $\sqrt{2}\,\delta = |\mu_1 - \mu_2|/\sigma$. Limiting power efficiencies for large δ were computed for the curves graphed. The alternatives corresponding to power equal .50, .90 and .99 are indicated on each curve. Table III lists the power efficiency for the cases studied and the power functions for the sign test are in Tables I and II. Cases used in this comparison but not satisfying the

TABLE III

Power Efficiency of the Sign Test

Alternatives p	δ	$N = 5$, $\alpha = .0625$	10, .0020	10, .0215	10, .1094	20, .0118	20, .0414	20, .1153
.50	0	96.0	94.0	85.0	76.8	76.3	72.9	69.9
.45	.1257	96.0	93.7	84.9	76.8	76.1	72.7	69.8
.40	.2534	96.0	93.3	84.7	76.7	75.7	72.5	69.6
.35	.3853	96.0	92.8	84.5	76.5	75.3	72.1	69.3
.30	.5244	95.8	92.1	84.0	76.1	74.8	71.6	68.8
.25	.6745	95.5	91.5	83.4	75.6	74.1	71.0	68.3
.20	.8416	95.2	90.6	82.7	74.9	73.4	70.3	67.7
.15	1.0364	94.7	89.6	81.8	74.1	72.5	69.5	66.9
.10	1.2816	93.9	88.2	80.6	72.9	71.2	68.3	
.05	1.6449	92.8	86.2	78.9	71.2	69.5	67.8	
.03	1.8808	92.0	84.8	77.6	70.2			
.01	2.3263	90.6	82.4	75.6				
.005	2.5758	89.9	81.1					
.001	3.0902	88.5	78.7					
0	∞	80.6	58.5	56.2	52.3	54.2	52.7	51.5

requirement of largest $\alpha \leqq .01$ or $\leqq .05$ are indicated by parentheses in Table I. Additional powers not tabulated there are:

N	r	$p = .03$	.01	.005	.001
5	0	.85873	.95099	.97525	.99501
10	0	.73742	.90438	.95111	.99005
10	1	.96549	.99573	.99890	.99996
10	2	.99724	.99989	.99999	

Examination of the efficiencies stated by Walsh in [9], [12], [13] confirm the statement of Jeeves and Richards [7] that the approximation used by Walsh would consistently overestimate the true efficiency. This effect is large only for small sample sizes. For example, Walsh gives 96 per cent efficiency for $N = 5$ and reference to Table III shows this to be the highest point on the curve. On the other hand, the value of .7 efficiency for $\alpha = .05$ given by Jeeves and Richards for $N = 6$ to 20 is quite reasonable for $N = 20$ but seems too low for $N = 6$ to 10. Jeeves and Richards use a randomized test, and since the efficiency depends greatly on the level of significance comparisons are difficult.

4. Computation of power efficiency function. The power function of the t-test was computed for several degrees of freedom at levels of significance corresponding to those of the sign test. Computation was effected using the formulas of Nicholson [8]. Interpolation for fractional degrees of freedom of the t-test giving equivalent power to the sign test was made on z-scores (normal deviates). This procedure was followed since the power curves of the t-test are representable approximately as normal cumulative curves. In most cases linear interpolation proved to be satisfactory. However, this method was not satisfactory for δ near zero. Since the power curves for the sign and t-test agree in magnitude and slope at $\delta = 0$ the limiting power efficiency may be obtained by interpolating among the second derivatives of the power functions. The second derivative for the sign test at $\delta = 0$, sample size N and critical value $r_{\alpha/2}$ is

$$\frac{1}{\pi 2^{N-2}} \sum_{x=0}^{r_{\alpha/2}} \binom{N}{x} [(2x - N)^2 - N].$$

The second derivative for the t-test for ν degrees of freedom and critical value $t_{\alpha/2}$ is

$$\frac{2(\nu + 1) \mid t_{\alpha/2} \mid \nu^{\nu/2}}{B(1/2, \nu/2)(\nu + t_{\alpha/2}^2)^{1/2(\nu+1)}}.$$

The limiting efficiency for $\delta = 0$ as N becomes infinite can also be obtained from these derivatives and the result $2/\pi$ agrees with the value obtained by Cochran [4].

This limit is the same for arbitrary fixed α and it appears that the limiting power efficiency curve for large N and fixed δ approaches zero for increasing δ. No proof of this statement was obtained. However, it is not inconsistent with the statement of Walsh [10], [11] that the sign test and the t-test have the same power function asymptotically when sample sizes are in ratio $2/\pi$.

The values indicated for limiting power efficiency for large δ represent ratios of sample sizes similar to those for finite δ. Nicholson [8] gives an expression for the power of the t-test with ν degrees of freedom. The leading term for large δ is

$$1 - \left(\frac{x_\alpha}{2}\right)^{\nu-\frac{1}{2}} \sqrt{2} z^{\nu-2} e^{-z^2/2}$$

where $z = \delta\sqrt{\nu(1 - x_\alpha)}$ and x_α satisfies

$$\frac{1}{B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \int_{x_\alpha}^{1} p^{-\frac{1}{2}}(1 - p)^{(\nu-2)/2} \, dp = \alpha.$$

Using the approximation of ordinate over abscissa for the cumulative normal for extreme abscissa we find that z is the abscissa of a cumulative normal which is approximately equal to the power of the t-test for alternative δ. In a similar manner the normal approximation to the binomial yields $z = \delta\sqrt{r + 1}$ for the sign test. A fixed value of N and α determines r, α, x_α and we may solve for ν.

REFERENCES

[1] National Bureau of Standards, *Tables of the Binomial Probability Distribution*, U. S. Government Printing Office, Washington, D. C.

[2] K. Pearson, *Tables of the Incomplete Beta Function*, Cambridge University Press, 1934.

[3] Burton H. Camp, "Approximation to the point binomial," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 130–131.

[4] W. G. Cochran, "The efficiencies of the binomial test of significance of a mean and of a correlation coefficient," *J. Roy. Stat. Soc.*, Vol. 100 (1937), pp. 69–73.

[5] W. J. Dixon and A. M. Mood, "The statistical sign test," *J. Amer. Stat. Assn.*, Vol. 41 (1946), pp. 557–566.

[6] Murray F. Freeman and John W. Tukey, "Transformations related to the angular and the square root," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 607–611.

[7] T. A. Jeeves and Robert Richards, "A note on the power of the sign test," *Ann. Math. Stat.*, Vol. 21 (1950), p. 618.

[8] W. L. Nicholson, "A computing formula for the power of the analysis of variance test," submitted to *Ann. Math. Stat.*

[9] John E. Walsh, "On the power function of the sign test for slippage of means," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 358–362.

[10] John E. Walsh, "On the asymptotic power efficiency of the sign test for slippage of means," Douglas Aircraft Co., Inc. Report.

[11] John E. Walsh, "Some comments on the efficiency of statistical tests," *Human Biology*, Vol. 21 (1949), pp. 205–217.

[12] John E. Walsh, "Some significance tests for the median which are valid under very general conditions," *Ann. Math. Stat.*, Vol. 20, (1949), pp. 64–81.

[13] John E. Walsh, "Some bounded significance level properties of the equal-tail sign test," *Ann. Math. Stat.*, Vol 22 (1951), pp. 408–417.

THE ADMISSIBILITY OF CERTAIN INVARIANT STATISTICAL TESTS INVOLVING A TRANSLATION PARAMETER

By E. L. Lehmann[1] and C. M. Stein

University of California, Berkeley, and University of Chicago

1. Introduction. The notion of invariance (or symmetry) has such strong intuitive appeal that many current statistical procedures have the invariance property and are in fact the best invariant procedures although they were pro-

Received 2/10/53.

[1] This work is supported in part by the Office of Naval Research.