EXAMPLE 1. The power function for $\alpha = 0.05$, $m = 2$ and $n = 8$ would be obtained from (19) with $a = 1$. From [3], $x(1, 4, 0.95) = 0.52713$; substituting gives the power function as

$$P(\lambda \mid 1, 4; 0.05) \\ = 1 - e^{-0.47287\lambda}(0.95000 + 0.34381\,\lambda + 0.03961\,\lambda^2 + 0.00136\,\lambda^3). \tag{21}$$

EXAMPLE 2. Suppose that two-figure accuracy is desired in calculating the power function for $\alpha = 0.05$, $m = 8$ and $n = 30$. The unabridged form of (10) with $a = 4$ and $b = 15$ would entail evaluating 15 terms. From (15),

$$R(\lambda \mid 4, 15, 8; 0.05) < 0.003.$$

Thus using the first eight terms of (10) would certainly secure the necessary accuracy.

REFERENCES

[1] P. C. TANG, "The power function of the analysis of variance test with tables and illustrations of their use," *Statistical Research Memoirs*, Vol. 2 (1938), p. 143.

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

[3] CATHERINE M. THOMPSON, "Tables of percentage points of the incomplete beta-function," *Biometrika*, Vol. 32 (1941), pp. 151–153.

[4] KARL PEARSON, *Tables of the Incomplete Γ-Function*, His Majesty's Stationery Office, London, 1922.

POWER UNDER NORMALITY OF SEVERAL NONPARAMETRIC TESTS

BY W. J. DIXON

University of Oregon[1]

1. Summary. Presented are tabulations of the power and power efficiency of four nonparametric tests (rank-sum, maximum deviation, median, and total number of runs) for the difference in means of two samples drawn from normal populations with equal variance. The cases considered are for equal sample sizes of three, four and five observations and alternatives $\delta = |\mu_1 - \mu_2|/\sigma$.

2. Introduction. One method of comparison of various nonparametric tests is a study of their performance under the assumption of normality. An advantage of this method is the wide use of the normal assumption. Disadvantages are the limitation to a particular type of distribution and the extensive computation necessary.

The computation of power under normality is simplest for small samples and small levels of significance. This fact has guided the present study, but it is

Received 6/8/53, revised 2/6/54.

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

TABLE I

Power and power efficiency for two samples of equal size, $N_1 = N_2 = N$, from normal populations with equal variances σ^2 and means μ_1 and μ_2, for each of four tests against the alternative $\delta = |\mu_1 - \mu_2|/\sigma$.

W = *rank sum* D = *maximum absolute deviation*
M = *median* R = *total number of runs*

N	3		4		5											
Test	W, D, M, R						W				D		M	
α	1/10		1/35		1/126		2/126		4/126		10/126		26/126	
δ	Pow.	Eff.	Pow.	Eff.	Pow.	Eff.	Pow.	Eff.	Pow.	Eff.	Pow.	Eff.	Pow.	Eff.
0	.100	.975	.029	.965	.008	.957	.016	.962	.032	.965	.080		.206	
.25	.111	.98	.035	.96	.011	.96								
.50	.143	.98	.055	.96	.021	.95	.040	.97	.072	.97	.144	.87	.297	.78
.75	.195	.98	.090	.96	.041	.95								
1.00	.264	.98	.141	.96	.074	.95	.128	.96	.210	.97	.328	.87	.515	.77
1.25	.347	.97	.209	.95	.124	.94								
1.50	.438	.97	.293	.95	.192	.94	.301	.96	.431	.96	.576	.86	.744	.76
1.75	.532	.97	.388	.95	.278	.93								
2.00	.624	.97	.489	.94	.377	.93	.530	.95	.674	.95	.794	.86	.899	.76
2.50	.781	.96	.682	.93	.587	.92	.744	.94	.858	.94	.925	.85	.970	.75
3.00	.890	.96	.830	.93	.768	.91	.889	.93	.953	.94	.980	.85	.994	.74
3.50	.952	.95	.923	.92	.890	.90	.961	.92	.988	.93	.996	.84		
4.00	.982	.95	.970	.91	.956	.89	.989	.92	.998	.92	.999	.83		
4.50	.994	.95	.990	.90	.985	.88	.998	.91	.9996	.91	.9999	.83		
5.00	.998	.94	.997	.90	.995	.88								

hoped that the considerable differences evident in these small sample cases, along with what is known about asymptotic results, will be of help to the statistician using nonparametric tests.

In addition to reporting the actual power for various alternatives, comparison has been made with the t-test by use of a *power efficiency function* (Table I). This function $P_E(\delta)$ is defined as the ratio of the sample size of t-test which results in equal power for a given alternative, δ, to the sample size of the nonparametric test under consideration. Fractional sample sizes for the t-test are found by interpolation on sample size to obtain a power equal to that of the nonparametric test. This function has already been used [1] for the sign test. Powers of the t-test used for this comparison were computed as by Nicholson [3].

Since the power efficiency of a nonparametric test will, in general, also depend upon the level of significance, comparisons of different tests are made difficult by the fact that the levels of significance which naturally occur for each test are not the same. To make the comparison simpler, power efficiency is also given (Table II) for the tests randomized to a single level of significance, $\alpha = .025$. For example, the rank sum test has natural levels $\alpha = .0159$ and

TABLE II

Power and power efficiency of three tests, each randomized to level of significance $\alpha = .025$, for $N_1 \doteq N_2 = 5$

Test	Rank sum		Max. abs. dev.		Median	
δ	Power	Eff.	Power	Eff.	Power	Eff.
0	.025	.964	.025		.025	
.5	.058	.96	.051	.81	.045	.70
1.0	.173	.95	.135	.80	.112	.70
1.5	.376	.94	.284	.79	.240	.71
2.0	.614	.93	.476	.78	.421	.72
2.5	.810	.92	.668	.77	.620	.72
3.0	.925	.91	.819	.76	.788	.73
3.5	.976	.90	.915	.75	.899	.73
4.0	.994	.89	.966	.74	.960	.73
4.5	.999	.88	.988	.74	.986	.73

$\alpha = .0317$, so that use of the former with chance .425 and the latter with chance .575 will result in an effective $\alpha = .025$.

Since for $\delta = 0$ all the power curves agree in ordinate and slope, the limiting power efficiency function as δ approaches zero may be obtained by interpolating among the second derivatives of the power functions of t in the same manner as among the ordinates for δ not zero.

Computation for the limiting power efficiency for the one sided rank sum test randomized to $\alpha = .0125$ for $N_1 = N_2 = 5$ was made by interpolating among the first derivatives of the power functions for one-sided t-tests. The same power efficiency, .964, was obtained as for the corresponding two-sided test with $\alpha = .025$.

Table III gives these limiting power efficiencies for the rank sum test for sample sizes $N_1 \leqq N_2 \leqq 5$. The cases indicated by an asterisk apply also to the maximum deviation, median, total numbers of runs tests. Of course they apply also to any test which has, for the stated α, a critical region corresponding to

TABLE III

Limiting power efficiency of rank sum test W against the alternative $|\mu_1 - \mu_2|/\sigma = \delta = 0$, for various levels of significance α and various sample sizes N_1 and N_2 from normal populations with equal variances σ^2 and means μ_1 and μ_2. Values marked with asterisk (*) apply also to maximum absolute deviation (D), median (M), total number of runs (R), and similar tests.

N_1, N_2	2, 2	2, 3	2, 4		3, 3		4, 4	5, 5			∞, ∞
α	1/3	1/5	2/15	4/15	1/10	1/5	1/70	1/126	2/126	4/126	$0 < \alpha < 1$
Eff.	.995*	.990	.971*	.966	.975*	.973	.965*	.957*	.962	.965	$.9549 = 3/\pi$

the case of all observations in one sample greater than all observations in the second sample, or vice versa. The limiting value, $3/\pi$, for large samples is given by others [2], [5].

3. Theory. The computation of power requires the evaluation, numerically in most cases, of the integrals representing the probabilities that various sample configurations lying in the critical region will occur under various alternative assumptions. Two such expressions will be displayed.

The first case corresponds to all observations in one sample greater than all observations in the second sample, or vice versa. Here $\alpha = 2(N!)^2/(2N)!$ and the rank sum statistic equals $\sum_1^N i$ for two samples of size N. The power $P(\delta)$ for normal cdf $F(x)$ is

$$N \int_{-\infty}^{\infty} F^{N-1}(x)\, F^N(\delta - x)\, dF(x) + N \int_{-\infty}^{\infty} F^{N-1}(-x)\, F^N(\delta - x)\, df(x).$$

In the second case, where the smallest observation and only this smallest observation of one sample lies between the two largest observations of the other sample, the rank sum statistic equals $1 + \sum_1^N i$ for two samples of size N. In this case the power $P(\delta)$ for normal cdf $F(x)$ is

$$N^2(N-1) \int_{-\infty}^{\infty} \int_{-\infty}^{x} F^{N-1}(y - \delta)\, F^{N-2}(-x)\, F(x - \delta)\, dF(y)\, dF(x)$$

$$- N^2(N-1) \int_{-\infty}^{\infty} \int_{-\infty}^{x} F^{N-2}(y)\, F(y - \delta)\, F^{N-1}(\delta - x)\, dF(y)\, dF(x)$$

$$+ N^2 \int_{-\infty}^{\infty} F^{N-1}(x)\, F^{N-1}(\delta - x)\, F(x - \delta)\, dF(x) - N^2 \int_{-\infty}^{\infty} F^{N-1}(-x)\, F^N(x - \delta)\, dF(x).$$

Similar expressions may be written down for larger α, and for the other tests. The quadratures were performed for $\delta = .25(.25)2.00$, 3.00, 4.00 for $N = 3$, 4 and for $\delta = .50(.50)3.50$, 4.50 for $N = 5$; other values were filled in by subtabulation. The power curves of the nonparametric tests and the t-test, when subjected to the transformation

$$P(\delta) = \int_{-\infty}^{x(\delta)} (2\pi)^{-\frac{1}{2}} e^{-t^2/2}\, dt,$$

yield an $x(\delta)$ essentially linear in δ when δ is not close to zero. Consequently, all interpolations were performed on $x(\delta)$. This procedure was also used in interpolation for the power efficiency function. Second differences were adequate in most cases.

An extensive bibliography on the above tests is given by Savage [4].

4. Conclusions.

TABLE I. The four nonparametric tests considered have high power efficiencies for very small samples and small α, when compared with the t-test for normal alternatives. Power efficiency decreases slightly for more distant alternatives.

As the level of significance increases, the power efficiency of the rank sum test increases slightly whereas the power efficiencies of the median and maximum deviation tests decrease.

TABLE II. When tests for samples of size 5 are randomized to the single level of significance $\alpha = .025$, it is easy to compare the tests and note that the rank sum test has greater power than the median and maximum deviation tests. Particularly for near alternatives, the maximum deviation test has greater power than the median test.

TABLE III. The local power efficiencies for the rank sum test are very high. For all cases computed they are greater than $3/\pi$, the limiting local power efficiency for large samples.

REFERENCES

[1] W. J. DIXON, "Power functions of the sign test and power efficiency for normal alternatives," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 467–473.
[2] A. M. MOOD, "On the asymptotic efficiency of certain nonparametric two-sample tests," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 514–522.
[3] W. L. NICHOLSON, "A computing formula for the power of the analysis of variance test," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 607–610.
[4] I. R. SAVAGE, "Bibliography of nonparametric statistics and related topics," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 844–906.
[5] H. R. VAN DER VAART, "Some remarks on the power of Wilcoxon's test for the problem of two-samples, I and II," *Indagationes Math.*, Vol. 12 (1950), pp. 146–172.

Addendum

Papers on this topic appearing since submission of this paper include:
[6] E. L. LEHMANN, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 23–43.
[7] B. L. VAN DER WAERDEN, "Order tests for the two sample problem," *Nederl. Akad. Wetensch. Proc. Ser. A.*, Vol. 56 (1953), pp. 303–316.

A REMARK ON THE JOINT DISTRIBUTION OF CUMULATIVE SUMS

BY HERBERT ROBBINS

University of North Carolina and Columbia University

Let X_k, $k = 1, \cdots, n$, be any finite number n of independent random variables with respective distribution functions $F_k(x) = \Pr[X_k \leqq x]$. Let $T_k = X_1 + \cdots + X_k$ be the successive cumulative sums of the X_k, with individual distribution functions $G_k(t) = \Pr[T_k \leqq t]$ and joint distribution function $G(t_1, \cdots, t_n) = \Pr[T_1 \leqq t_1, \cdots, T_n \leqq t_n]$. Since the T_k are not in general stochastically independent, the function $G(t_1, \cdots, t_n)$ will not in general be equal to the product of the n functions $G_k(t_k)$, but we shall show that the *inequality*

$$G(t_1, \cdots, t_n) \geqq \prod_1^n G_k(t_k) \tag{1}$$

Received 9/12/53.