The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 923–935

THE V_{NM} TWO-SAMPLE TEST

BY URS R. MAAG AND M. A. STEPHENS[1]

Université de Montréal and McGill University

1. Summary and introduction. The statistic V_{NM} is a two-sample statistic which may be used to test the null hypothesis H_0, that two samples, sizes N and M, come from identical populations. It is an adaptation of the Kolmogorov two-sample statistic, and is defined by

$$V_{NM} = \sup_{-\infty<x<\infty} (F_N(x) - G_M(x)) - \inf_{-\infty<x<\infty} (F_N(x) - G_M(x))$$

where $F_N(x)$, $G_M(x)$ are the sample cumulative distribution functions. A single-sample analogue V_N is defined by replacing $G_M(x)$ in the formula above by a hypothesised distribution function $F(x)$. For large values of V_{NM} or V_N, H_0 will be rejected. These statistics are particularly useful for observations recorded as points on a circle; the value obtained for V_N or V_{NM}, in contrast to that of the corresponding Kolmogorov statistic, does not depend on the choice of origin. Kuiper (1960) proved this result and suggested the use of the V statistics for the circle. He also gave series approximations to the distributions of $N^{\frac{1}{2}}V_N$ and $N^{\frac{1}{2}}V_{NN}$, for large N, on the null hypothesis H_0. Reference to a distribution will henceforth be assumed to refer to the distribution on the null hypothesis.

The V_{NN} statistic had earlier been investigated by Gnedenko and co-workers (see Gnedenko (1954)), who used it for observations on a line; the V statistics may be expected to be more powerful than the Kolmogorov statistics for certain alternatives. Gnedenko (1954) gives the exact distribution of V_{NN}. The exact distribution of V_N, in the upper and lower tails, has recently been given by Stephens (1965).

In this paper we make the two-sample goodness-of-fit test available for a wide range of sample sizes by giving tables of the distribution of V_{NM}. In the next section a formula is given with which the V_{NM} statistic may be calculated from the ranks of the two samples, and then the goodness-of-fit tests are set out. These are called exact or approximate tests, depending on whether the probabilities used are calculated from exact or approximate formulae. In Section 3 the construction of the tables is described. To find percentage points for large N, we develop in Section 3.3 a series expansion for the distribution of $N^{\frac{1}{2}}V_{NN}$ which differs from that given by Kuiper; the probabilities given by the two expansions are compared in Table 4, and the new series clearly gives the better results.

Received 6 December 1965; revised 26 December 1967.

[1] The research of this author was sponsored in part by the Office of Naval Research by Contract No. Nonr 4010(09) awarded to the Department of Statistics, The Johns Hopkins University. This paper in whole or in part may be reproduced for any purpose of the United States Government.

2. Tests of the null hypothesis H_0.

2.1 *Introductory notation.* The null hypothesis H_0 is that the two independent random samples, respectively sizes N and M, come from the same continuous distribution function. The statistic V_{NM} is computed as below; if the observations are on a circle, any one may be chosen to begin the ranking.

Let the values in the first sample, in ascending order, be $x_1, x_2, \cdots, x_N$, and those in the second sample, in ascending order, be $y_1, y_2, \cdots, y_M$; and let r_i be the rank of x_i, s_j be the rank of y_j, in the pooled sample of the ordered $N + M$ observations (the smallest observation having rank 1, and the largest rank $(N + M)$). The formula for V_{NM} is then

$$(1) \quad V_{NM} = (MN)^{-1}\{\max_{1 \leqq i \leqq N}[(N + M)i - Nr_i] + \max_{1 \leqq j \leqq M}[(N + M)j - Ms_j]\}.$$

When $N = M$, this becomes

$$(2) \quad V_{NN} = N^{-1}\{\max_{1 \leqq i \leqq N}[2i - r_i] + \max_{1 \leqq j \leqq N}[2j - s_j]\}.$$

We now set out the goodness-of-fit tests.

2.2 *Exact test.* This is to be used when $N \neq M$, and $N + M \leqq 28$; or when $N = M$ and $3 \leqq N \leqq 100$.

(1) Calculate V_{NM} from (1) or (2); thus calculate either NMV_{NM}, if $N \neq M$, or NV_{NN}, if $N = M$;

(2) Use Table 1 or Table 2 to find p for given N, M;

(3) If $p \leqq \alpha$, reject H_0 at significance level α.

Table 1 gives k and p such that, for unequal N, M, roughly restricted by the inequality $N + M \leqq 28$, $\Pr(NMV_{NM} \geqq k) = p$. Table 2 gives k and p such that, for $N = M$, with values $N = 10$ (1) 25 (5) 50 (10) 70 (5) 80 (10) 100, $\Pr(NV_{NN} \geqq k) = p$.

2.3 *Approximate test* 1. For $N = M$, and $100 \leqq N \leqq 500$.

(1) Calculate V_{NN} from (2); thus calculate $N^{\frac{1}{2}}V_{NN}$.

(2) Use Table 3 to find y, the table entry for given α, N.

(3) If $N^{\frac{1}{2}}V_{NN} \geqq y$, reject H_0 at significance level α.

Table 3 gives an approximate value for y for which $\Pr(N^{\frac{1}{2}}V_{NN} \geqq y) = \alpha$, for $\alpha = .10, .05, .025, .01$ and $.005$, and for $N = 100$ (20) 300 (40) 500 and for $1/N = 0$.

2.4 *Approximate test* 2. For $N \neq M$, values not included in Table 1.

(1) Calculate V_{NM} from (1).

(2) Calculate $y = [(N + M)M^{-1}]^{\frac{1}{2}}V_N(\alpha)$, where $V_N(\alpha)$ is the upper tail percentage point of V_N, at level α, given in Stephens (1965), Table 1.

(3) If $V_{NM} > y$, reject H_0 at significance level α.

3. The distribution of V_{NM}.

3.1. *Small sample sizes.* The value of V_{NM} depends on the relative ranks of the two samples. On H_0, all arrangement of the two samples, mixed together as one sample, are equally likely, and the distribution of V_{NM} may be found by calculating its value for each arrangement. This has been done to give the probabilities in Table 1. As N and M increase, the number of arrangements eventually makes this technique prohibitive.

3.2. *Equal sample sizes.* We introduce

$$D_{NN}^{+} = \sup_x[F_N(x) - G_N(x)] \quad \text{and} \quad D_{NN}^{-} = -\inf_x [F_N(x) - G_N(x)];$$

then

$$V_{NN} = D_{NN}^{+} + D_{NN}^{-}.$$

The distribution of V_{NN} will now be derived from the joint distribution of D_{NN}^{+} and D_{NN}^{-}, given by Kemperman (1959) as follows:

$$(3) \qquad \begin{aligned} P_N(a, b) &= \Pr(D_{NN}^{-} < a/N, D_{NN}^{+} < b/N) \\ &= 2^{2N}\binom{2N}{N}^{-1}(2/k) \textstyle\sum_{r=1}^{k-1} (\sin r\pi a/k)^2(\cos r\pi/k)^{2N} \end{aligned}$$

where a and b denote positive integers and $k = a + b$. It is easily seen that

$$(4) \qquad \begin{aligned} &\Pr(D_{NN}^{+} + D_{NN}^{-} < k/N) \\ &\quad = \textstyle\sum_{a=1}^{k} \Pr(D_{NN}^{-} < a/N, D_{NN}^{+} = (k-a)/N) \\ &\quad = \textstyle\sum_{a=1}^{k} P_N(a, k+1-a) - \sum_{a=1}^{k-1} P_N(a, k-a). \end{aligned}$$

By substituting formula (3) into (4), interchanging the summations and using the trigonometric identity

$$\textstyle\sum_{r=1}^{n} \cos 2rx = \frac{1}{2}[\sin(2n+1)x/\sin x - 1]$$

we have the distribution of V_{NN}:

$$(5) \qquad \begin{aligned} &\Pr(V_{NN} < k/N) \\ &\quad = 2^{2N+1}\binom{2N}{N}^{-1}\left[\textstyle\sum_{s=1}^{[k/2]}(\cos s\pi(k+1)^{-1})^{2N} - \sum_{s=1}^{[(k-1)/2]} (\cos s\pi k^{-1})^{2N}\right] \end{aligned}$$

where k can take the values $2, 3, \cdots, N+1$ and the symbol $[x]$ means the greatest integer less than or equal to x.

A different expression for the distribution of V_{NN} is reported by Gnedenko (1954).

$$\begin{aligned} \Pr\left(\left(\tfrac{1}{2}N\right)^{\frac{1}{2}}V_{NN} < z\right) &= 1 + 2\binom{2N}{N}^{-1}\Big\{a \textstyle\sum_{s=1}^{[N/(a+1)]}\binom{2N}{N-s(a+1)} \\ &\quad - (a-1) \textstyle\sum_{s=1}^{[N/a]}\binom{2N}{N-sa} - \sum_{i=1}^{a} \sum_{s=1}^{[(N+i)/(a+1)]}\binom{2N}{N+i-s(a+1)} \\ &\quad + \textstyle\sum_{i=1}^{a-1} \sum_{s=1}^{[(N+i)/a]} \binom{2N}{N+i-sa}\Big\} \end{aligned}$$

where $a = [z(2N)^{\frac{1}{2}}]$, $z > 0$.

Since in fact V_{NN} takes only the values kN^{-1}, $k = 1, 2, \cdots, N$, it is possible to simplify this expression to

$$(6) \qquad \begin{aligned} &\Pr(V_{NN} \geqq kN^{-1}) \\ &\quad = 2\binom{2N}{N}^{-1}\Big\{k \textstyle\sum_{s=1}^{[N/k]} \binom{2N}{N-sk} - (k+1) \sum_{s=1}^{[N/(k+1)]} \binom{2N}{N-s(k+1)}\Big\} \end{aligned}$$

TABLE 1

Upper tail probabilities of the distribution of NMV_{NM}.

For given N, M and k the table shows

$$p = \Pr [V_{NM} \geqq v] = \Pr [NMV_{NM} \geqq k] \quad \text{where} \quad v = k(NM)^{-1}$$
$$\Pr [V_{NM} > 1] = \Pr [NMV_{NM} > NM] = 0$$

N	M	k	p	N	M	k	p	N	M	k	p
3	3	9	.3000	3	18	48	.0947	4	12	32	.4044
		6	.9000			45	.1579				
						42	.2368	4	14	56	.0059
3	4	12	.2000			39	.3316			52	.0235
		9	.6000							48	.0588
				3	20	60	.0130			44	.1176
3	5	15	.1429			57	.0390			42	.1588
		12	.4286			54	.0779			40	.2353
						51	.1299			38	.3059
3	6	18	.1071			48	.1948				
		15	.3214			45	.2727	4	16	64	.0041
										60	.0165
3	7	21	.0833	4	4	16	.1143			56	.0413
		18	.2500			12	.5714			52	.0826
										48	.1734
3	8	24	.0667	4	5	20	.0714			44	.3013
		21	.2000			16	.2857				
		18	.4000					4	18	72	.0030
				4	6	24	.0476			68	.0120
3	9	27	.0545			20	.1905			64	.0300
		24	.1636			18	.3333			60	.0602
		21	.3273							56	.1053
				4	7	28	.0333			54	.1323
3	10	30	.0455			24	.1333			52	.1895
		27	.1364			21	.2667			50	.2376
		24	.2727							48	.3038
				4	8	32	.0242				
3	12	36	.0330			28	.0970	4	20	80	.0023
		33	.0989			24	.3152			76	.0090
		30	.1978							72	.0226
		27	.3297	4	9	36	.0182			68	.0452
						32	.0727			64	.0791
3	14	42	.0250			28	.1818			60	.1468
		39	.0750			27	.2545			56	.2417
		36	.1500							52	.3569
		33	.2500	4	10	40	.0140				
						36	.0559	5	5	25	.0397
3	16	48	.0196			32	.1399			20	.2778
		45	.0588			30	.2098				
		42	.1176			28	.3217	5	6	30	.0238
		39	.1961							25	.1190
		33	.2941	4	12	48	.0088			24	.1905
						44	.0352			20	.3810
3	18	54	.0158			40	.0879				
		51	.0474			36	.2198	5	7	35	.0152

TABLE 1—*Continued*

N	M	k	p	N	M	k	p	N	M	k	p
5	7	30	.0758	5	16	55	.1538	6	9	39	.1349
		28	.1364			54	.1971			36	.2547
		25	.2576			50	.2632				
								6	10	60	.0020
5	8	40	.0101	5	18	90	.0007			54	.0121
		35	.0505			85	.0034			50	.0260
		32	.1010			80	.0103			48	.0519
		30	.1818			75	.0239			44	.0999
		27	.3030			72	.0314			42	.1499
						70	.0540			40	.1938
5	9	45	.0070			67	.0745			38	.2737
		40	.0350			65	.1073				
		36	.0769			62	.1442	6	12	72	.0010
		35	.1329			60	.1880			66	.0058
		31	.2378			57	.2427			60	.0271
		30	.3217			55	.2973			54	.0824
										48	.2104
5	10	50	.0050	5	20	100	.0005			42	.3111
		45	.0250			95	.0024				
		40	.0999			90	.0071	6	14	78	.0031
		35	.2498			85	.0165			72	.0108
		30	.5195			80	.0381			70	.0155
						75	.0776			66	.0325
5	12	60	.0027			70	.1388			64	.0490
		55	.0137			65	.2235			60	.0800
		50	.0412			60	.3534			58	.1161
		48	.0604							56	.1362
		45	.1099	6	6	36	.0130			54	.1827
		43	.1593			30	.1169			52	.2384
		40	.2335			24	.4416			50	.2848
		38	.3159								
				6	7	42	.0076	6	16	90	.0018
5	14	70	.0016			36	.0455			84	.0062
		65	.0082			35	.0758			80	.0094
		60	.0245			30	.1742			78	.0192
		56	.0392			29	.2652			74	.0310
		55	.0686							72	.0487
		51	.1078	6	8	48	.0047			68	.0752
		50	.1520			42	.0280			66	.1035
		46	.2206			40	.0513			64	.1194
		45	.2794			36	.1119			62	.1619
						34	.1865			60	.2017
5	16	80	.0010			32	.2471			58	.2397
		75	.0052			30	.3450			56	.2954
		70	.0155								
		65	.0361	6	9	54	.0030	6	18	96	.0037
		64	.0454			48	.0180			90	.0119
		60	.0795			45	.0360			84	.0312
		59	.1042			42	.0749			78	.0679

TABLE 1—*Continued*

N	M	k	p	N	M	k	p	N	M	k	p
6	18	72	.1382	7	10	49	.1128	7	20	91	.0937
		66	.2473			46	.1696			86	.1178
		60	.3918			43	.2360			85	.1469
						42	.2928			84	.1687
6	20	108	.0024							80	.1826
		102	.0063	7	12	77	.0026			79	.2188
		100	.0078			72	.0060			78	.2557
		96	.0155			70	.0132				
		94	.0209			65	.0283	8	8	64	.0012
		90	.0340			63	.0449			56	.0162
		88	.0464			60	.0596			48	.0945
		84	.0667			58	.0939			40	.3232
		82	.0893			56	.1244				
		80	.0982			53	.1697	8	9	72	.0007
		78	.1267			51	.2262			64	.0056
		76	.1602			49	.2738			63	.0098
		74	.1826							56	.0280
		72	.2192	7	14	91	.0013			55	.0490
		70	.2626			84	.0067			54	.0622
						77	.0237			48	.1084
7	7	49	.0041			70	.0703			47	.1615
		42	.0449			63	.1645			46	.2070
		35	.2162			56	.3336			45	.2350
		28	.5833							40	.3196
				7	16	96	.0037				
7	8	56	.0023			91	.0087	8	10	80	.0004
		49	.0163			89	.0134			72	.0033
		48	.0280			84	.0241			70	.0062
		41	.1189			82	.0368			64	.0169
		40	.1492			80	.0424			62	.0317
		35	.2424			77	.0615			60	.0424
		34	.3333			75	.0863			56	.0695
						73	.1047			54	.1086
7	9	63	.0014			70	.1348			52	.1477
		56	.0098			68	.1742			50	.1748
		54	.0182			66	.2108			48	.2246
		49	.0448			64	.2299			46	.2904
		47	.0797			63	.2705				
		45	.1063					8	12	84	.0027
		42	.1622	7	20	119	.0029			80	.0068
		40	.2350			113	.0051			76	.0144
		38	.3077			112	.0085			72	.0318
						106	.0145			68	.0529
7	10	70	.0009			105	.0209			64	.1007
		63	.0061			100	.0241			60	.1599
		60	.0122			99	.0362			56	.2470
		56	.0288			98	.0467			52	.3558
		53	.0551			93	.0580				
		50	.0778			92	.0780	8	16	112	.0015

TABLE 1—*Concluded*

N	M	k	p	N	M	k	p	N	M	k	p
8	16	104	.0062	9	9	72	.0056	10	10	70	.0699
		96	.0207			63	.0381			60	.2283
		88	.0558			54	.1573			50	.5245
		80	.1311			45	.4280				
		72	.2598					10	12	100	.0021
				9	10	90	.0002			98	.0043
8	20	136	.0012			81	.0019			96	.0057
		132	.0020			80	.0033			90	.0106
		128	.0037			72	.0103			88	.0187
		124	.0062			71	.0189			86	.0268
		120	.0108			70	.0243			84	.0317
		116	.0166			63	.0448			80	.0441
		112	.0276			62	.0716			78	.0645
		108	.0389			61	.0959			76	.0875
		104	.0610			60	.1086			74	.1079
		100	.0843			54	.1541			72	.1203
		96	.1215			53	.2084			70	.1450
		92	.1633			52	.2592			68	.1826
		88	.2173							66	.2253
		84	.2805	10	10	100	.0001			64	.2680
						90	.0018				
9	9	81	.0004			80	.0145				

where k takes the values 1, 2, $\cdots$, N. In Maag (1965) equations (5) and (6) are shown to be equivalent.

From (6), the density is found to be

$$\begin{aligned} \Pr(V_{NN} = kN^{-1}) &= \Pr(V_{NN} \geqq kN^{-1}) - \Pr(V_{NN} \geqq (k+1)N^{-1}) \\ &= 2\binom{2N}{N}^{-1}\Big\{k \textstyle\sum_{s=1}^{[N/k]} \binom{2N}{N+sk} + (k+2) \sum_{s=1}^{[N/(k+2)]} \binom{2N}{N-s(k+2)} \\ &\quad - 2(k+1) \textstyle\sum_{s=1}^{[N/(k+1)]} \binom{2N}{N-s(k+1)}\Big\}. \end{aligned} \tag{7}$$

The moments of V_{NN} can also be obtained:

$$\mu_r' = 2^{2N}N^{-r}\binom{2N}{N}^{-1} - N^{-r} + 2N^{-r}\binom{2N}{N}^{-1}A_N^r$$

where

$$A_N^r = \textstyle\sum_{m=2}^{N} m(m^r + (m-2)^r - 2(m-1)^r) \sum_{s=1}^{[N/m]} \binom{2N}{N-sm}.$$

Formulas (6) and (7) have been used to construct Table 2.

3.3 *Equal sample sizes, large samples.* For large sample sizes, it becomes difficult to preserve accuracy in calculating the probabilities from the exact formulas of equations (5) or (6), so we convert them to series in powers of $N^{-\frac{1}{2}}$. To convert equation (5) we follow steps (a) to (d) below.

(a) We start with the power series (convergent for $|x| < \pi/2$)

$$\log(\cos x) = -\textstyle\sum_{m=1}^{\infty} |B_{2m}| \, ((2m)!)^{-1}(2^{2m} - 1)2^{2m}x^{2m}(2m)^{-1}$$

TABLE 2

Upper tail probabilities of the distribution of NV_{NN}

For given N and k the table shows $p = \Pr[NV_{NN} \geqq k]$

$\Pr(NV_{NN} > N] = 0$

N	k	p	N	k	p	N	k	p
10	10	.0001	16	8	.1792	22	12	.0195
	9	.0018		7	.3733		11	.0519
	8	.0145					10	.1203
	7	.0699	17	13	.0004		9	.2437
	6	.2283		12	.0023		8	.4301
	5	.5245		11	.0098			
				10	.0334	23	15	.0007
11	10	.0006		9	.0939		14	.0028
	9	.0053		8	.2196		13	.0091
	8	.0290		7	.4271		12	.0260
	7	.1102					11	.0651
	6	.3028	18	13	.0009		10	.1432
				12	.0041		9	.2772
12	11	.0002		11	.0151			
	10	.0018		10	.0465	24	16	.0003
	9	.0114		9	.1200		15	.0012
	8	.0494		8	.2614		14	.0041
	7	.1572					13	.0125
	6	.3772	19	14	.0003		12	.0336
				13	.0016		11	.0798
13	11	.0006		12	.0065		10	.1675
	10	.0043		11	.0219		9	.3111
	9	.0209		10	.0618			
	8	.0753		9	.1485	25	16	.0005
	7	.2087		8	.3040		15	.0018
	6	.4491					14	.0058
			20	14	.0006		13	.0167
14	12	.0002		13	.0027		12	.0424
	11	.0016		12	.0099		11	.0958
	10	.0084		11	.0303		10	.1929
	9	.0339		10	.0793		9	.3450
	8	.1062		9	.1789			
	7	.2630		8	.3466	30	18	.0003
							17	.0011
15	12	.0006	21	15	.0002		16	.0032
	11	.0033		14	.0011		15	.0088
	10	.0145		13	.0043		14	.0219
	9	.0505		12	.0142		13	.0497
	8	.1411		11	.0403		12	.1026
	7	.3183		10	.0989		11	.1927
				9	.2107		10	.3285
16	13	.0002		8	.3888			
	12	.0012				35	19	.0006
	11	.0059	22	15	.0004		18	.0017
	10	.0228		14	.0018		17	.0045
	9	.0706		13	.0064		16	.0111

TABLE 2—*Continued*

N	k	p	N	k	p	N	k	p
35	15	.0252	60	23	.0032	80	25	.0086
	14	.0529		22	.0064		24	.0149
	13	.1028		21	.0123		23	.0251
	12	.1843		20	.0228		22	.0410
	11	.3048		19	.0405		21	.0648
				18	.0687		20	.0993
40	20	.0009		17	.1117		19	.1471
	19	.0023		16	.1737		18	.2107
	18	.0055		15	.2579		17	.2916
	17	.0125						
	16	.0267	70	27	.0007	90	31	.0006
	15	.0531		26	.0015		30	.0012
	14	.0988		25	.0030		29	.0021
	13	.1716		24	.0057		28	.0039
	12	.2779		23	.0106		27	.0068
				22	.0188		26	.0115
45	22	.0004		21	.0324		25	.0191
	21	.0011		20	.0527		24	.0368
	20	.0027		19	.0859		23	.0483
	19	.0062		18	.1323		22	.0736
	18	.0132		17	.1962		21	.1090
	17	.0268		16	.2796		20	.1568
	16	.0513					19	.2190
	15	.0925	75	28	.0007		18	.2966
	14	.1569		27	.0014			
	13	.2502		26	.0028	100	33	.0005
				25	.0053		32	.0010
50	23	.0005		24	.0096		31	.0017
	22	.0013		23	.0168		30	.0031
	21	.0030		22	.0286		29	.0052
	20	.0065		21	.0471		28	.0087
	19	.0133		20	.0748		27	.0143
	18	.0260		19	.1148		26	.0228
	17	.0483		18	.1701		25	.0355
	16	.0849		17	.2431		24	.0540
	15	.1414		16	.3347		23	.0800
	14	.2231					22	.1155
	13	.3326	80	29	.0007		21	.1625
				28	.0014		20	.2225
60	25	.0007		27	.0026		19	.2965
	24	.0015		26	.0048			

TABLE 3

Upper tail significance points of $N^{\frac{1}{2}}V_{NN}$

N	α: .10	.05	.025	.01	.005
100	2.241	2.419	2.580	2.772	2.906
120	2.245	2.424	2.584	2.778	2.913
140	2.248	2.427	2.589	2.782	2.917
160	2.251	2.430	2.592	2.786	2.921
180	2.253	2.433	2.594	2.788	2.924
200	2.255	2.435	2.597	2.791	2.926
220	2.257	2.436	2.598	2.793	2.928
240	2.258	2.438	2.600	2.794	2.930
260	2.260	2.439	2.601	2.796	2.932
280	2.261	2.440	2.603	2.797	2.933
300	2.262	2.442	2.604	2.798	2.935
340	2.263	2.443	2.606	2.801	2.937
380	2.265	2.444	2.607	2.802	2.938
420	2.266	2.446	2.609	2.804	2.940
460	2.267	2.447	2.610	2.805	2.941
500	2.268	2.448	2.611	2.806	2.942
∞	2.291	2.471	2.633	2.830	2.967

where the B_{2m} are the Bernoulli numbers. With $k = z\,N^{\frac{1}{2}}$, z bounded, we have

$$(\cos s\pi k^{-1})^{2N} = \exp\,[-s^2\pi^2/z^2](1 - s^4\pi^4(6Nz^4)^{-1} + O(N^{-2}))$$

and replacing z in the above formula by $z + N^{-\frac{1}{2}}$ yields a similar expression for $(\cos s\pi/(k+1))^{2N}$. We then can show

$$\begin{aligned} &(\cos s\pi(zN^{\frac{1}{2}} + 1)^{-1})^{2N} - (\cos s\pi z^{-1}N^{-\frac{1}{2}})^{2N} \\ (8)\qquad &= \exp\,[-s^2\pi^2/z^2](2s^2\pi^2z^{-3}N^{-\frac{1}{2}} + N^{-1}(-3s^2\pi^2z^{-4} + 2s^4\pi^4z^{-6}) \\ &\quad + N^{-\frac{3}{2}}(4s^2\pi^2z^{-5} + 2s^4\pi^4\tfrac{1}{3}z^{-5} - 6s^4\pi^4z^{-7} - s^6\pi^6\tfrac{1}{3}z^{-7} + 4s^6\pi^6\tfrac{1}{3}z^{-9}) + R) \end{aligned}$$

where the behaviour of R is discussed below.

(b) From Stirling's formula it follows that the first factor in (5) becomes

$$(9)\qquad 2^{2N+1}/\binom{2N}{N} = 2(\pi N)^{\frac{1}{2}}(1 + (N8)^{-1} + O(N^{-2})).$$

Thus the product of (8) and (9) gives the typical term in (5), before summing over s, as:

$$\begin{aligned} &\pi^{\frac{1}{2}}\exp\,[-s^2\pi^2/z^2]\ [4s^2\pi^2z^{-3} + 2N^{-\frac{1}{2}}(2s^4\pi^4z^{-6} - 3s^2\pi^2z^{-4}) + N^{-1}(s^2\pi^2\tfrac{1}{2}z^{-3} \\ (10)\qquad &+ 8s^2\pi^2z^{-5} + 4s^4\pi^4\tfrac{1}{3}z^{-5} - 12s^4\pi^4z^{-7} - 2s^6\pi^6\tfrac{1}{3}z^{-7} + 8s^6\pi^6\tfrac{1}{3}z^{-9})] + A_s \end{aligned}$$

where the remainder

$$A_s = \exp\,[-s^2\pi^2/z^2][R\cdot O(N^{\frac{1}{2}}) + R^*].$$

(c) We now wish to determine the order of the remainder A_s where $s = 1, 2, \cdots [(k-1)/2]$. From the developments which lead to (8) and (10) it is easy to see that both $R \cdot N^{\frac{1}{2}}$ and R^* are of order $s^m \cdot N^{-\frac{3}{2}}$, where m is some fixed positive number. Thus A_s is of order $N^{-\frac{3}{2}} s^m \exp(-s^2\pi^2/z^2)$.

(d) Finally the summation over s gives

$$\sum_{s=1}^{[(k-1)/2]} A_s < \text{const. } N^{-\frac{3}{2}} \sum_{s=1}^{\infty} s^m \exp(-s^2\pi^2/z^2),$$

which is of order $N^{-\frac{3}{2}}$ since the series on the right converges. We notice that for even values of k the last term in the first sum in (5), i.e. $(\cos \frac{1}{2}k\pi(k+1)^{-1})^{2N} \cdot 2^{2N}/\binom{2N}{N}$ has to be added to the remainder. Since this term decreases to zero exponentially as a function of N it does not change the order of the remainder.

Thus this sequence yields the distribution of $N^{\frac{1}{2}}V_{NN}$ for large N:

$$\begin{aligned}(11)\quad &\Pr(N^{\frac{1}{2}}V_{NN} < z) \\ &= \pi^{\frac{1}{2}} \sum_{s=1}^{\infty} \exp[-s^2\pi^2/z^2][4s^2\pi^2z^{-3} + 2N^{-\frac{1}{2}}(2s^4\pi^4z^{-6} - 3s^2\pi^2z^{-4}) \\ &\quad + N^{-1}(s^2\pi^2\tfrac{1}{2}z^{-3} + 8s^2\pi^2z^{-5} + 4s^4\pi^4\tfrac{1}{3}z^{-5} - 12s^4\pi^4z^{-7} \\ &\quad - 2s^6\pi^6\tfrac{1}{3}z^{-7} + 8s^6\pi^6\tfrac{1}{3}z^{-9})] + O(N^{-\frac{3}{2}}).\end{aligned}$$

We may now transform (11) into a series which converges more rapidly for large values of $z (z > \pi^{\frac{1}{2}})$.

Let $G(z) = \sum_{s=1}^{\infty} \exp(-s^2\pi^2/z^2)$; then it can readily be verified that (11) can be written in the form

$$\begin{aligned}(12)\quad \Pr(N^{\frac{1}{2}}V_{NN} < z) &= 2\pi^{\frac{1}{2}}(d/dz)G(z) + \pi^{\frac{1}{2}}N^{-\frac{1}{2}}(d^2/dz^2)G(z) \\ &\quad + \pi^{\frac{1}{2}}N^{-1}((d^3/3dz^3)G(z) - z^2(d^3/12dz^3)G(z) \\ &\quad - 5z(d^2/12dz^2)G(z)) + O(N^{-\frac{3}{2}}).\end{aligned}$$

The θ-transform applied to $G(z)$ leads to

$$G(z) = -\tfrac{1}{2} + z/(2\pi^{\frac{1}{2}}) + (z\pi^{\frac{1}{2}}) \sum_{s=1}^{\infty} \exp(-s^2z^2).$$

When this is substituted into (12), we obtain:

$$\begin{aligned}(13)\quad \Pr(N^{\frac{1}{2}}V_{NN} \geqq z) &= \sum_{s=1}^{\infty} \exp[-s^2z^2][4s^2z^2 - 2 + 2zN^{-\frac{1}{2}}(3s^2 - 2s^4z^2) \\ &\quad + N^{-1}(2s^2 - 3s^2z^2 + \tfrac{11}{3}s^4z^4 - 8s^4z^2 \\ &\quad + \tfrac{8}{3}s^6z^4 - \tfrac{2}{3}s^6z^6)] + O(N^{-\frac{3}{2}}).\end{aligned}$$

3.4 *Comparisons.* The probabilities given by the asymptotic formula (13) are compared with the exact probabilities, for selected values of N and k, in Table 4. For $z \geqq 2$, only the first term in each sum in equation (13) need be used. The accuracy of (13) for $N = 100$ also makes it of use to calculate further significance points of $N^{\frac{1}{2}}V_{NN}$, with high accuracy for $N \geqq 100$. The points in Table 3 are obtained in this way.

Kuiper (1960) has also given an expansion comparable to (13). His result, with slight changes in notation, is

$$\begin{aligned}(14)\quad \Pr(N^{\frac{1}{2}}V_{NN} \geqq z) &= \sum_{s=1}^{\infty} e^{-s^2z^2}(4s^2z^2 - 2) \\ &\quad - (6N)^{-1}(1 + \sum_{s=1}^{\infty} e^{-s^2z^2} \cdot s^2z^2(2s^2z^2 - 7)) + O(N^{-2}).\end{aligned}$$

TABLE 4

Comparison of exact and approximate probabilities

For given N and k, the table gives several calculations of $\Pr(NV_{NN} \geqq k)$: 1, the exact probability; 2, the approximate probability using (13) without the term in N^{-1}; 3, the approximate probability using (13) including the term in N^{-1}; 4, the approximate probability using Kuiper's series (14).

N	k	1 (exact)	2	3	4
20	8	.3466	.3292	.3494	.4327
	9	.1789	.1674	.1823	.2384
	10	.0793	.0741	.0815	.1121
	11	.0303	.0287	.0311	.0434
	12	.0099	.0098	.0099	.0113
	13	.0027	.0029	.0025	−.0017
50	13	.3326	.3257	.3335	.3890
	14	.2231	.2176	.2240	.2679
	15	.1414	.1378	.1422	.1741
	16	.0849	.0827	.0854	.1068
	18	.0260	.0257	.0262	.0332
	20	.0065	.0066	.0064	.0067
	21	.0030	.0031	.0029	.0015
	22	.0013	.0014	.0013	−.0011
100	19	.2965	.2932	.2968	.3349
	21	.1625	.1604	.1628	.1883
	23	.0800	.0790	.0801	.0948
	27	.0143	.0143	.0143	.0168
	31	.0017	.0018	.0017	.0008

The series (13) and (14) have the same first term, but differ in the terms of order $N^{-\frac{1}{2}}$ and N^{-1}. The numerical comparison in Table 4 shows that the probabilities given by (14) are substantially different from the exact ones, and support the view that there is an error in (14).

3.5 *Larger unequal samples.* When N and M are different, the possible values of V_{NM} are multiples of $(NM)^{-1}$; giving many more values than when $N = M$; the probability distribution then takes smaller jumps in the tail. For values not covered by Table 1, we should like to find a good approximation to the α-level significance point y for which $\Pr(V_{NM} \geqq y) \leqq \alpha$, and $\Pr(V_{NM} > y - (NM)^{-1}) > \alpha$. Such an approximation is given in approximate test 2, Section 2.4. It is based on the fact that, for $N, M \to \infty$, $(NM/(N + M))^{\frac{1}{2}}V_{NM}$ has the same distribution as $N^{\frac{1}{2}}V_N$, and exact significance points for the latter are in Stephens (1965). We have examined this approximation for the pairs: $N = 6$, $M = 16$; $N = 6$, $M = 20$; and $N = 8$, $M = 20$. From the values of y were calculated the exact probabilities $\Pr(V_{NM} \geqq y) = \alpha'$, using Table 1, and values of α' compared to the nominal values α. For $N = 6$, the values of α' were not always the best attainable (i.e., nearest to α, but less than α), but for $N = 8$ they were the best attainable. The approximation should improve for larger values of N.

3.6 *Further remarks.* Suppose we adapt (13) to give

$$(15) \quad \Pr((N/2)^{\frac{1}{2}}V_{NN} \geqq x) \sim \sum_{s=1}^{\infty} e^{-2s^2x^2}(8s^2x^2 - 2) + AN^{-\frac{1}{2}}x \sum_{s=1}^{\infty} e^{-2s^2x^2}(3s^2 - 4s^4x^2)$$

as far as the term in $N^{-\frac{1}{2}}$. The value of A is $2 \cdot 2^{\frac{1}{2}}(2.828)$. The series for the single sample statistic, given by Kuiper (1960), has for $\Pr(N^{\frac{1}{2}}V_N \geqq x)$ exactly the same opening terms, but with $A = \frac{8}{3} = 2.667$. One might conjecture that a series approximation for $\Pr((NM/(N + M))^{\frac{1}{2}} V_{NM} \geqq x)$ might be of the form (15) above, with A a function of M, N which, for $M \geqq N$, decreases from $2 \cdot 2^{\frac{1}{2}}$ when $M = N$ to $\frac{8}{3}$ when $M \to \infty$, but Hodges (1957) has shown a very erratic behaviour of the distribution of the Kolmogorov-Smirnov two-sample statistic, when M and N are different; this presumably extends to V_{NM}.

We wish to thank Professor A. Joffe of the Université de Montréal for suggesting the method which leads to equation (5) and also to express our appreciation of the referee's very helpful comments. We are grateful for the facilities of the McGill University Computing Centre, where most of the computations were made.

REFERENCES

GNEDENKO, B. V. (1954). Kriterien für die Unveränderlichkeit der Wahrscheinlichkeitsverteilung von zwei unabhängigen Stichprobenreihen. *Math. Nachr.* **12** 29–66.

HODGES, J. L., JR. (1957). The significance probability of the Smirnov two-sample test. *Ark. Math.* **3** 469–86.

KEMPERMAN, J. H. B. (1959). Asymptotic expansions for the Smirnov test and for the range of cumulative sums. *Ann. Math. Statist.* **30** 448–62.

KUIPER, N. H. (1960). Tests concerning random points on a circle. *Proc. Koninkl. Nederl. Akad. van Wetenschappen Ser. A* **63** 38–47.

MAAG, U. R. (1965). Contributions to the design of experiments (Tests on the line and on the circle), Ph.D. Thesis. Univ. of Toronto.

STEPHENS, M. A. (1965). The goodness-of-fit statistic V_N : Distribution and significance points. *Biometrika* **52** 309–21.