DISTRIBUTION OF THE TWO-SAMPLE CRAMÉR-VON MISES CRITERION FOR SMALL EQUAL SAMPLES[1]

By E. J. Burr

University of New England

1. Introduction and summary. The null hypothesis that the two independent random samples $u_1, \cdots, u_m$ and $v_1, \cdots, v_n$ come from the same (unknown) continuous distribution may be tested by the two-sample analogue of the Cramér-von Mises ω^2 criterion, as described by Anderson [1]. In the particular case of samples of equal size $(m = n)$, this test criterion may be expressed in the form

$$t = (d_0^2 + d_1^2 + \cdots + d_{2n}^2)/4n^2$$

where d_i is the difference between the number of members of the first sample and the number of members of the second sample, contained in the first i members of the pooled set of $2n$ members arranged in order of magnitude.

Let $P_n(t)$ be the probability, under the null hypothesis, that this test criterion will attain or exceed the value t. As $n \to \infty$, the limiting distribution P_∞ is the same as the limiting distribution of $n\omega^2$, which has been tabulated by Anderson and Darling [2]. The main object of this paper is to study the manner in which P_n approaches P_∞ in the upper tail $(P_n \leqq 0.1)$. It is an extension, for the case of equal samples, of a similar study by Anderson [1] which includes also unequal samples.

An iterative method for computing P_n for $n = 1, 2, 3, \cdots$ is described. This method has been used to find the complete distributions for $n \leqq 10$, and parts of the upper tails for $n = 11$ $(P_{11} \leqq 0.06)$, $n = 12$ $(P_{12} \leqq 0.026)$ and $n = 14$ $(P_{14} \leqq 0.00014)$. For $n \leqq 10$ and for $n = \infty$, the smallest attainable values of t significant at each of the levels 10%, 5%, 2%, 1%, 0.5%, etc., are given, with corresponding partial results for $n = 11$, 12 and 14. The limiting distribution is tabulated for $0.3 \leqq t \leqq 2.2$, which corresponds approximately to $10^{-1} > P_\infty > 5 \times 10^{-6}$. Finally the deviations of P_n from P_∞ are examined and are found to conform to the empirical formula:

$$P_n = P_\infty\{1 + n^{-1}f(t)\},$$

and the correction function $f(t)$ is given by table or formula for $0.3 \leqq t \leqq 1.7$.

2. Exact small-sample distributions. The pooling and ordering of the $2n$ sample values may result in any one of $\binom{2n}{n}$ permutations of the n members of the first sample with the n members of the second. Each such permutation determines a value of t, and under the null hypothesis each permutation occurs

Received March 8, 1962.

[1] This research was supported in part by National Science Foundation Grant 14336 at Stanford University.

with the same probability $\binom{2n}{n}^{-1}$. Hence the evaluation of P_n reduces to the problem of counting the permutations which give rise to each attainable value of t. Consider a simple random walk on a line, in which the particle starts at the origin and takes m discrete steps of $+1$ or -1. We represent the walk graphically in the cartesian x-y plane by a path joining the points $(0, 0), (1, y_1), \cdots, (i, y_i), \cdots, (m, y_m)$ where $y_i - y_{i-1} = \pm 1$. It is clear that the $\binom{2n}{n}$ distinct paths which terminate at $(2n, 0)$ are in one-to-one correspondence with the permutations of the $2n$ members of two samples each of size n, and that for each such path and the corresponding permutation we have

$$4n^2t = y_0^2 + y_1^2 + y_2^2 + \cdots + y_{2n}^2 .$$

This representation leads to a simple iterative method for finding the frequency function of $\sum y_i^2$, and hence of t, for $n = 1, 2, 3, \cdots$. Suppose that we have tabulated separately the $m + 1$ frequency functions of $\sum y_i^2$ for paths to (m, m), $(m, m - 2)$, $(m, m - 4)$, $\cdots$, $(m, -m)$. Then we can at once write down the corresponding functions for paths to $(m + 1, m + 1)$, $(m + 1, m - 1)$, $\cdots$, $(m + 1, -m - 1)$. For a path to $(m + 1, m - 3)$ (for example) must pass either through $(m, m - 2)$ or through $(m, m - 4)$, so that we have only to add the frequencies of each $\sum y_i^2$ for paths to $(m, m - 2)$ and $(m, m - 4)$ and increase each $\sum y_i^2$ by $(m - 3)^2$. In this way, the frequency functions of $\sum y_i^2$ were tabulated for paths to $(m, m - 2r)$, for $r = 0, 1, \cdots, m$ and $m = 1, 2, \cdots, 14$. Among these are the functions for paths to $(2, 0)$, $(4, 0)$, $\cdots$, $(14, 0)$, which yield the probability distributions of t for $n = 1, 2, \cdots, 7$.

For $n > 7$ the method was modified as follows. To find the distribution of t for $n = 8$ (for example), we require the frequency function of $\sum y_i^2$ for paths to $(16, 0)$. This is found by deriving separately, and then combining, the frequency functions for paths through $(8, 8)$, paths through $(8, 6)$, paths through $(8, 4)$ and so on. Consider the contribution due to paths from $(0, 0)$ to $(16, 0)$ via $(8, 4)$ for example. The frequency function of $\sum y_i^2$ for paths from $(8, 4)$ to $(16, 0)$ is identical with that for paths from $(0, 0)$ to $(8, 4)$, which is known, being included in the tabulation described above. Hence the required contribution is found by forming the convolution of this frequency function with itself, and subtracting 4^2 from each new $\sum y_i^2$ to allow for the fact that the value of y at the junction $(8, 4)$ has been included twice. In this way, the probability distributions of t were found for $n = 8, 9, 10$, and partial results were obtained for $n = 11$ $(P_{11} \leqq 0.06)$, for $n = 12$ $(P_{12} \leqq 0.026)$, and for $n = 14$ $(P_{14} \leqq 0.00014)$. Table 1 gives the smallest attainable t significant at each of the levels 0.1, 0.05, 0.02, 0.01, 0.005, and so on, and the true value of P_n for each. (A few of these values duplicate some of the significance points tabulated by Anderson for $n \leqq 8$.)

The distribution for $n = 10$ is given in full in Table 2.

3. The limiting distribution. By means of Anderson and Darling's series [2], the values of P_∞ were computed for twelve values of t between 0.6 and 2.2. These were found to fit the empirical formula

$$\log_{10} P_\infty = -a - b \log_{10} t - ct + \epsilon(t)$$

where $a = 0.458$, $b = 0.444$, $c = 2.151$, and $\epsilon(t)$ is close to zero over a wide range; more precisely,

$$|\epsilon(t)| < 4 \times 10^{-4} \quad \text{for} \quad 0.42 < t < 2.2,$$

so that interpolation to five decimal places in $\log_{10} P_\infty$ was made easy by plotting $\epsilon(t)$ against t. The values of t so found at $P_\infty = 0.02$, 0.01 and 0.001, agreed precisely with the values given in Anderson and Darling's table.

The results are given in Table 3, in a form suitable for linear interpolation. (In this table, the values of P_∞ for $0.3 \leqq t \leqq 0.66$ were found by interpolation from Anderson and Darling's table.) The same results are also included in Table 1 in another form.

It seems to be unusual to tabulate significance points as far as the level 10^{-5}, but this has been done here, partly for use in deriving the empirical formula given below, and partly because these points are sometimes required to determine the effective over-all significance level when one of a very large number of test criteria gives a nominally highly significant result. A composite test of this kind is described in [3].

4. An empirical small-sample formula. Smoothed graphs of $\log_{10} P_n$ against t for $8 \leqq n \leqq 12$ and $t \geqq 0.3$ indicated that the deviations of $\log_{10} P_n$ from $\log_{10} P_\infty$ all changed sign at about the same point ($t \simeq 0.62$), and were approximately proportional to n^{-1} for each fixed t. Therefore the formula

$$\log_{10} P_n = \log_{10} P_\infty + n^{-1} g(t)$$

was tried out by plotting smoothed values of $n \log_{10} (P_n/P_\infty) \equiv g(t)$ against t, to see if $g(t)$ was really independent of n. The fit was fairly good, with $g(t)$ a cubic polynominal in t; but the values for $n = 14$ (computed later) did not conform. After two or three other formulae had been tried and discarded, the simple formula

$$P_n = P_\infty\{1 + n^{-1} f(t)\} \tag{1}$$

was found to given an excellent fit, except as noted below. The function $f(t)$ appears to be linear at least for $1.15 \leqq t \leqq 1.7$, being given by

$$f(t) = -3.00 - 10.8\,(t - 1), \qquad t \geqq 1.15. \tag{2}$$

For $0.3 \leqq t \leqq 1.15$ the values of $f(t)$ were read off from a free-hand curve drawn through or between points representing smoothed values of $n(P_n/P_\infty - 1)$ as a function of t. These values of $f(t)$ are given in Table 4, in which linear interpolation is permissible.

TABLE 1

Smallest values of t significant at the levels 0.1, 0.05, 0.02, 0.01, *etc.*

P_n is the probability of attaining or exceeding the value t. Numbers in parentheses indicate the power of ten by which the preceding number is to be multiplied.

n	t	P_n
4	.5000	5.714 (−2)
	.6875	2.857 (−2)
5	.4500	8.730 (−2)
	.4900	4.762 (−2)
	.6900	1.587 (−2)
	.8500	7.937 (−3)
6	.3750	9.307 (−2)
	.5139	3.896 (−2)
	.6250	1.948 (−2)
	.7639	8.658 (−3)
	.8750	4.329 (−3)
	1.0139	2.165 (−3)
7	.3827	9.324 (−2)
	.4847	4.895 (−2)
	.6480	1.690 (−2)
	.7704	8.159 (−3)
	.8520	4.079 (−3)
	1.0561	1.166 (−3)
	1.1786	5.828 (−4)
8	.3750	9.635 (−2)
	.4844	4.817 (−2)
	.6250	1.943 (−2)
	.7344	9.790 (−3)
	.8594	4.196 (−3)
	.9688	1.865 (−3)
	1.0625	9.324 (−4)
	1.2344	3.108 (−4)
	1.3438	1.554 (−4)
9	.3735	9.329 (−2)
	.4846	4.681 (−2)
	.6204	1.974 (−2)
	.7315	9.996 (−3)
	.8426	4.813 (−3)
	.9784	1.974 (−3)
	1.0772	9.461 (−4)
	1.1636	4.936 (−4)
	1.3241	1.645 (−4)
	1.4105	8.227 (−5)
	1.5093	4.114 (−5)
11	.4773	4.781 (−2)
	.6260	1.891 (−2)
	.7417	9.441 (−3)
	.8492	4.732 (−3)
	.9814	1.999 (−3)
	1.0888	9.668 (−4)
	1.1963	4.366 (−4)
	1.3202	1.786 (−4)
	1.4112	9.639 (−5)
	1.4855	4.536 (−5)
	1.6095	1.985 (−5)
	1.7583	5.670 (−6)
	1.8409	2.835 (−6)
12	.6250	1.953 (−2)
	.7361	9.705 (−3)
	.8472	4.903 (−3)
	.9931	1.926 (−3)
	1.0972	9.785 (−4)
	1.2014	4.652 (−4)
	1.3333	1.901 (−4)
	1.4236	9.541 (−5)
	1.5208	4.659 (−5)
	1.6181	1.849 (−5)
	1.7292	8.875 (−6)
	1.7986	4.438 (−6)
	1.9306	1.479 (−6)
	2.0069	7.396 (−7)
14	1.4515	9.677 (−5)
	1.5485	4.886 (−5)
	1.6658	1.939 (−5)
	1.7628	9.273 (−6)
	1.8444	4.886 (−6)
	1.9413	1.994 (−6)
	2.0383	9.472 (−7)
	2.1046	4.487 (−7)
	2.2117	1.994 (−7)
	2.2730	9.971 (−8)
∞	.3473	1.000 (−1)
	.4614	5.000 (−2)
	.6198	2.000 (−2)
	.7435	1.000 (−2)
	.8694	5.000 (−3)

TABLE 1—*Continued*

n	t	P_n	n	t	P_n
10	.3650	9.861 (−2)	∞	1.0384	2.000 (−3)
	.4750	4.978 (−2)		1.1679	1.000 (−3)
	.6250	1.972 (−2)		1.2983	5.000 (−4)
	.7350	9.948 (−3)		1.4720	2.000 (−4)
	.8450	4.796 (−3)		1.5603	1.000 (−4)
	.9750	1.992 (−3)		1.7371	5.000 (−5)
	1.0850	9.201 (−4)		1.9135	2.000 (−5)
	1.1750	4.980 (−4)		2.0475	1.000 (−5)
	1.2950	1.732 (−4)		2.1818	5.000 (−6)
	1.3750	9.743 (−5)			
	1.5050	4.330 (−5)			
	1.6750	1.083 (−5)			

Because of irregular deviations of $P_n(t)$ from the smoothed values, the values of $f(t)$ given by Table 4 or by (2) are uncertain within about $\pm$ 0.05. In the interval $0.3 \leqq t \leqq 0.6$ there is greater uncertainty because of incomplete information on P_{11} and P_{12}, and in $0.45 \leqq t \leqq 0.6$ there seems to be a systematic difference between values of $f(t)$ based on P_{10} and P_{11}, those based on P_{11} being lower by about 0.2. (This leads to a relative error of about 2 per cent when using (1) to compute smoothed values of P_{11} for those t.) Therefore we hesitate to conjecture that (1) remains valid for $n > 11$ in $0.3 \leqq t \leqq 0.6$.

For $t > 0.6$, formula (1) gives an excellent fit up to the following approximate limits: $n = 8, t < 1.0; n = 9, t < 1.2; n = 10, t < 1.3; n = 11, t < 1.4; n = 12, t < 1.5; n = 14, t < 1.7$. For larger values of t than those indicated (relatively few of which are attainable for each n), the smoothed true values of P_n are always larger than the values predicted by (1) and (2).

The very satisfying accuracy of (1) for predicting individual (unsmoothed) values of P_n is shown by the upper bounds of the absolute error of prediction in the following formulae for $n = 12$ and 14:

$$P_{12} = P_\infty\{1 + f(t)/12 \pm 0.020\}, \qquad 0.58 < t < 1.47,$$

$$P_{14} = P_\infty\{1 + f(t)/14 \pm 0.015\}, \qquad 1.40 < t < 1.72.$$

We therefore conjecture that, at least for $0.6 < t < 1.7$, formula (1) remains valid for $n > 14$, with errors not exceeding $\pm$ $(0.015)P_\infty$.

REFERENCES

[1] ANDERSON, T. W. (1962). On the distribution of the two-sample Cramér-von Mises criterion. *Ann. Math. Statist.* **33** 1148–1159.

[2] ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

[3] BURR, E. J. (1960). Earthquakes and Uranus: misuse of a statistical test of significance. *Nature* **186** 336–337.

TABLE 2

Distribution of t for $n = 10$

The probability P_{10} that t will be attained or exceeded, equals the cumulative frequency divided by 184,756.

t	Cumulative frequency	t	Cumulative frequency	t	Cumulative frequency
1.675	2	0.865	800	0.415	13,302
1.585	4	0.855	838	0.405	14,146
1.505	8	0.845	886	0.395	15,080
1.435	12	0.835	974	0.385	16,342
1.425	14	0.825	1,042	0.375	17,322
1.375	18	0.815	1,060	0.365	18,218
1.355	24	0.805	1,140	0.355	19,410
1.325	28	0.795	1,226	0.345	20,670
1.295	32	0.785	1,314	0.335	22,016
1.285	42	0.775	1,416	0.325	23,548
1.255	46	0.765	1,468	0.315	25,064
1.245	50	0.755	1,528	0.305	26,542
1.235	54	0.745	1,680	0.295	28,502
1.225	74	0.735	1,838	0.285	30,684
1.215	76	0.725	1,912	0.275	32,522
1.205	80	0.715	2,022	0.265	35,012
1.175	92	0.705	2,122	0.255	37,750
1.165	98	0.695	2,272	0.245	40,262
1.155	110	0.685	2,498	0.235	43,690
1.145	118	0.675	2,634	0.225	46,888
1.135	126	0.665	2,754	0.215	49,440
1.115	138	0.655	2,930	0.205	53,220
1.105	150	0.645	3,146	0.195	57,248
1.095	162	0.635	3,360	0.185	61,290
1.085	170	0.625	3,644	0.175	66,610
1.075	194	0.615	3,832	0.165	71,970
1.065	204	0.605	4,028	0.155	77,594
1.045	230	0.595	4,402	0.145	84,742
1.035	242	0.585	4,696	0.135	91,886
1.025	262	0.575	4,980	0.125	98,990
1.015	282	0.565	5,270	0.115	107,056
1.005	302	0.555	5,510	0.105	115,604
0.995	316	0.545	5,884	0.095	125,748
0.985	346	0.535	6,368	0.085	137,844
0.975	368	0.525	6,724	0.075	149,044
0.965	384	0.515	7,030	0.065	159,156
0.955	420	0.505	7,534	0.055	169,908
0.945	456	0.495	7,988	0.045	179,124
0.935	488	0.485	8,518	0.035	183,732
0.925	528	0.475	9,198	0.025	184,756
0.915	546	0.465	9,658		
0.905	566	0.455	10,172		
0.895	646	0.445	10,974		
0.885	700	0.435	11,702		
0.875	744	0.425	12,396		

TABLE 3

The limiting distribution of t

Entries are $\log_{10} P_\infty + 10$.

t	0.00	0.02	0.04	0.06	0.08	Half difference
0.3	9.1309	9.0750	9.0199	8.9656	8.9119	−.0272
0.4	8.8588	8.8063	8.7542	8.7025	8.6512	−.0258
0.5	8.6003	8.5496	8.4993	8.4493	8.3995	−.0250
0.6	8.3499	8.3006	8.2514	8.2025	8.1537	−.0245
0.7	8.1051	8.0566	8.0083	7.9602	7.9122	−.0241
0.8	7.8643	7.8165	7.7688	7.7213	7.6738	−.0238
0.9	7.6264	7.5792	7.5320	7.4849	7.4379	−.0235
1.0	7.3910	7.3442	7.2974	7.2507	7.2041	−.0234
1.1	7.1575	7.1110	7.0646	7.0182	6.9719	−.0232
1.2	6.9256	6.8794	6.8333	6.7871	6.7411	−.0230
1.3	6.6951	6.6491	6.6032	6.5573	6.5115	−.0229
1.4	6.4657	6.4199	6.3742	6.3285	6.2829	−.0228
1.5	6.2373	6.1917	6.1462	6.1007	6.0552	−.0228
1.6	6.0098	5.9644	5.9190	5.8736	5.8283	−.0227
1.7	5.7830	5.7377	5.6925	5.6473	5.6021	−.0226
1.8	5.5569	5.5118	5.4667	5.4216	5.3765	−.0225
1.9	5.3315	5.2864	5.2414	5.1965	5.1515	−.0225
2.0	5.1066	5.0616	5.0167	4.9719	4.9270	−.0224
2.1	4.8821	4.8373	4.7925	4.7477	4.7029	−.0224
2.2	4.6582					

TABLE 4

Empirical small-sample correction function f(t)

Using this table, an approximation of $P_n(t)$ is given by $P_n(t) = P_\infty\{1 + n^{-1}f(t)\}$. For $t \geqq 1.15$, use the formula $f(t) = -3.00 - 10.8(t - 1)$.

t	$f(t)$	t	$f(t)$	t	$f(t)$
0.30	+1.10	0.65	−0.18	1.00	−3.04
0.35	+1.00	0.70	−0.50	1.05	−3.56
0.40	+0.88	0.75	−0.84	1.10	−4.09
0.45	+0.74	0.80	−1.21	1.15	−4.62
0.50	+0.56	0.85	−1.61	1.20	−5.16
0.55	+0.35	0.90	−2.05	1.25	−5.70
0.60	+0.10	0.95	−2.53	1.30	−6.24