BULLETIN *of the*
MALAYSIAN MATHEMATICAL
SCIENCES SOCIETY
http://math.usm.my/bulletin

Bull. Malays. Math. Sci. Soc. (2) **29***(1) (2006), 11–16*

A Modified Anderson-Darling Test for Uniformity

[1]MEZBAHUR RAHMAN, [2]LARRY M. PEARSON AND [3]HERBERT C. HEIEN
Minnesota State University, Mankato, MN 56001, USA
[1]mezbahur.rahman@mnsu.edu, [2]larry.pearson@mnsu.edu, [3]Herbert.Heien@i3Magnifi.com

Abstract. The subject of assessing whether a data set is from a specific distribution is crucial in statistical inference. This topic is critically important for uniform distributions as in generating random samples, a random sample is taken from the standard uniform distribution and then converted to the sample as required. Two different modifications of the Anderson-Darling A^2 test are presented. The critical values for the modifications and the usual A^2 test are also re-computed for different sample sizes. This is followed by a power comparison of the various tests.

2000 Mathematics Subject Classification: 62-04

Key words and phrases: Cramer-von Misses test; Pearson Chi-square statistic.

1. Introduction

The subject of assessing whether a data set is from a specific distribution has received a good deal of attention. This topic is critically important for uniform distributions and is different from the usual tests for randomness. There are different parametric and nonparametric tests for randomness. A nonparametric textbook such as Daniel [4] or Gibbons and Chakraborti [5] would provide extensive references. Recently, Rahman and Chakrobartty [8] showed that the Anderson-Darling A^2 (Anderson and Darling [1]) is the most powerful test in comparison with eight different commonly used tests including the Cramer-von Mises test (see Soest [9], for details) and the Watson [10] test along with different versions of Pearson's Chi-square tests. In the literature, the Anderson-Darling test has been studied extensively, for exanple, Giles [6] presented a saddlepoint approximation to the distribution function of the test.

In this paper we give two different modifications of the Anderson-Darling A^2 statistic, calculate their critical values and compare them along with the Pearson's Chi-square test through simulation. The Chi-square test is used for comparison as it is widely used in practice. Let us consider $X_1, X_2, \ldots, X_n$ to be a random sample taken from the Uniform (0,1) distribution. We will first explain all the tests. In Section 2 we will provide a power comparison study using Monte-Carlo simulation. In Section 3 we will give a brief conclusion based on the simulation results.

Received: February 23, 2005; *Revised:* July 26, 2005.

Table 1. Upper tail percentiles for Anderson-Darling A^2 test

0.250	0.150	0.100	0.050	0.025	0.010	0.005	0.001
1.248	1.610	1.933	2.492	3.070	3.880	4.500	6.000

Table 2. Upper tail percentiles for Anderson-Darling A^2 test

n	0.250	0.150	0.100	0.050	0.025	0.010	0.005	0.001
10	1.2419	1.6277	1.9518	2.5121	3.0990	3.9083	4.5175	5.9897
20	1.2500	1.6290	1.9385	2.5020	3.0731	3.8995	4.5117	5.9852
30	1.2457	1.6210	1.9313	2.5130	3.1111	3.9673	4.5309	5.8924
40	1.2450	1.6173	1.9362	2.5042	3.1047	3.9397	4.5889	6.1275
50	1.2425	1.6163	1.9277	2.4941	3.0933	3.9200	4.5211	5.9437
60	1.2464	1.6225	1.9367	2.5044	3.0776	3.9234	4.4858	6.0808
70	1.2515	1.6245	1.9304	2.4959	3.0889	3.8673	4.5326	5.9428
80	1.2384	1.6148	1.9235	2.4951	3.0778	3.8458	4.4808	5.9249
90	1.2461	1.6177	1.9326	2.5064	3.1020	3.9239	4.5856	6.0412
100	1.2399	1.6235	1.9325	2.4901	3.0655	3.8319	4.4068	5.8987
mean	1.2453	1.6211	1.9355	2.4986	3.0916	3.9033	4.5416	6.0266

1.1. Anderson-Darling A^2 Test. A distribution function test is suggested by Anderson and Darling [1]. The Anderson-Darling A^2 statistic is computed as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)\ln Z_i + (2n+1-2i)\ln(1-Z_i)\right\}, \tag{1}$$

where $Z_i = X_{i:n}$, are the ordered data from the smallest to the largest values. The percentiles are given in Table 1 with the first line indicating the upper tail probabilities and the second line representing the corresponding quantiles. The percentiles given in Table 1 are independent of n and are from D'Agastino and Stephens ([3], p.105).

We give a new table (Table 2) of critical values for different sample sizes ($n = 10, 20, \ldots, 100$) that are computed after generating 100,000 samples from the Uniform (0,1) distribution in each case. A search of finding a good fit model of the quantiles with respect to the sample size was unsuccessful and hence the means of the quantiles are provided to shorten the table for practical use. It is to be noted that the quantile values in Table 2 do not vary much and the mean can be used as a representative value for all sample sizes as they are also very similar to Table 1 values. The mean values in Table 2 are more reliable as they are computed using modern computational facilities. In all computations in this paper, MATLAB software with the statistical toolbox is used.

1.2. Modified Anderson-Darling W^2 Test. Anderson and Darling [2] showed the derivation of (1) from the integral

$$W^2 = n\int_{-\infty}^{\infty}\frac{[F_n(x)-F(x)]^2}{F(x)[1-F(x)]}dF(x) \tag{2}$$

Table 3. Upper tail percentiles for Anderson-Darling W^2 test

n	0.250	0.150	0.100	0.050	0.025	0.010	0.005	0.001
10	1.2461	1.6225	1.9345	2.4934	3.0938	3.9147	4.5458	6.0514
20	1.2441	1.6259	1.9389	2.4994	3.0975	3.9220	4.5297	6.0862
30	1.2437	1.6140	1.9253	2.4830	3.0624	3.8287	4.4970	6.1675
40	1.2440	1.6174	1.9344	2.4982	3.1102	3.9669	4.6137	6.1114
50	1.2460	1.6242	1.9370	2.5173	3.0992	3.9308	4.5625	6.1082
60	1.2433	1.6187	1.9345	2.5059	3.1082	3.9092	4.4924	6.1135
70	1.2428	1.6071	1.9225	2.4843	3.0782	3.9186	4.5534	6.0965
80	1.2492	1.6268	1.9440	2.5033	3.1042	3.9302	4.5600	5.9861
90	1.2443	1.6229	1.9359	2.4956	3.0844	3.9193	4.5049	6.0492
100	1.2456	1.6226	1.9298	2.4836	3.0725	3.8540	4.4286	5.9270
mean	1.2453	1.6211	1.9355	2.4986	3.0916	3.9033	4.5416	6.0266

Table 4. Upper tail percentiles for Anderson-Darling V^2 test

n	0.250	0.150	0.100	0.050	0.025	0.010	0.005	0.001
10	1.2247	1.6801	2.0730	2.8208	3.6425	4.8855	5.8298	8.5582
20	1.2708	1.7135	2.0856	2.7641	3.5148	4.5830	5.5008	7.7464
30	1.2886	1.7193	2.0918	2.7648	3.4833	4.4712	5.2645	7.4371
40	1.2830	1.7005	2.0456	2.6732	3.3682	4.3281	5.0864	6.9460
50	1.2826	1.6878	2.0318	2.6411	3.3000	4.1812	4.9457	6.5959
60	1.2859	1.6849	2.0180	2.6293	3.2759	4.1953	4.8765	6.7135
70	1.2846	1.6785	2.0095	2.6105	3.2385	4.1367	4.7998	6.5199
80	1.2831	1.6785	2.0069	2.6066	3.2277	4.1315	4.8162	6.5949
90	1.2758	1.6673	1.9896	2.5748	3.2020	4.1039	4.7360	6.2211
100	1.2793	1.6714	1.9924	2.5684	3.1857	4.0110	4.6531	6.1749
mean	1.2758	1.6882	2.0344	2.6654	3.3439	4.3027	5.0509	6.9508

where $F_n(x)$ is the empirical cdf (cumulative distribution function) and $F(x)$ is the theoretical cdf under the null hypothesis. Here we compute (2) using a numerical integration that relies on using the sum of the areas of rectangles known as the Riemann Integral. It is to be noted that for the Uniform (0,1) distribution, $F(x)$ is replaced by x. A simulation study as in Section 1.1. produced the critical values provided in Table 3. The mean values are the same in Table 1 and Table 2.

1.3. Modified Anderson-Darling V^2 Test. The integral (2) can be written as

$$W^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x)-x]^2}{x[1-x]} dx \tag{3}$$

by replacing $F(x) = x$ and $dF(x) = f(x)dx = dx$ for a Uniform (0,1) random variable. Then (3) can be approximated by

$$V^2 = n\left[\frac{(x_1-c_1)^2}{(1-c_1)} + \sum_{i=1}^{n-1} \frac{(x_i-c_i)^2}{c_i(1-c_i)}(c_{i+1}-c_i) + \frac{(x_n-c_n)^2}{c_n}\right] \tag{4}$$

Table 5. Third degree polynomial coefficients and their standard errors for the upper quantiles of V^2

q	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.250	1.184155 (0.0142578)	0.005547 (0.001069)	-0.000091 (0.000022)	0.0000005 (0.0000001)
0.150	1.657208 (0.017609)	0.003834 (0.001320)	-0.000084 (0.000027)	0.0000005 (0.0000002)
0.100	2.068866 (0.023521)	0.001770 (0.001763)	-0.000068 (0.000036)	0.0000004 (0.0000002)
0.050	2.882867 (0.037298)	-0.006222 (0.002796)	0.000037 (0.000058)	0.0000001 (0.0000003)
0.025	3.766730 (0.037675)	-0.013665 (0.002824)	0.000109 (0.000058)	0.0000003 (0.0000003)
0.010	5.229164 (0.063034)	-0.039997 (0.004725)	0.000520 (0.000097)	0.0000024 (0.0000006)
0.005	6.271817 (0.038187)	-0.049361 (0.002862)	0.000591 (0.000059)	-0.0000026 (0.0000004)
0.001	9.547369 (0.235944)	-0.114221 (0.017685)	0.001529 (0.000365)	-0.0000073 (0.0000022)

where the x_i's are the ordered sample measurements and $c_i = (i - 3/8)/(n + 1/4)$, a commonly used empirical cdf approximation. Equation (4) is a weighted mean of the squared deviations of the observed data values and the empirical distribution function values. A simulation study as in Section 1.1. produced the critical values and are provided in Table 4.

In Table 4, the quantiles vary more than the results in Tables 2 and 3. Third degree polynomials give reasonably high (at least 80% with one exception) R^2 values in the least square computations. Hence, we provide the third degree polynomials of the upper quantiles with respect to n as follows:

$$V_q^2 = \hat{\beta}_0 + \hat{\beta}_1 n + \hat{\beta}_2 n^2 + \hat{\beta}_3 n^3$$

where q stands for the quantile. The estimates of the coefficients and their standard errors in parentheses are given in Table 5.

This section differs from 1.1 and 1.2 in that the quantiles can be approximated more closely using the least square equations for different sample sizes.

1.4. Pearson χ^2 test. By grouping the data into g equal groups such that each group has an expected frequency of at least five and the number of groups is not too large, we can calculate the χ^2 goodness of fit statistic as

$$\chi^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i}, \tag{5}$$

where O_i is the observed number of values in the i^{th} group and $E_i = n/g$ is the expected frequency in the i^{th} group assuming that the sample is from the

Table 6. Power Study Results

n	5% Level of Significance				1% Level of Significance			
	A^2	W^2	V^2	χ^2	A^2	W^2	V^2	χ^2
Standard Normal Samples								
10	0.1139	0.1168	0.0289	0.0296	0.0235	0.0245	0.0020	0.0000
20	0.1363	0.1360	0.1380	0.1451	0.0377	0.0371	0.0373	0.0405
30	0.2475	0.2565	0.3268	0.3380	0.0736	0.0838	0.1293	0.1532
40	0.4366	0.4368	0.5664	0.5114	0.1570	0.1523	0.2990	0.2951
50	0.6323	0.6244	0.7553	0.6711	0.2935	0.2925	0.5085	0.4570
60	0.7869	0.7844	0.8755	0.7832	0.4662	0.4669	0.6795	0.6118
70	0.8885	0.8896	0.9390	0.8726	0.6438	0.6348	0.8118	0.7219
80	0.9473	0.9456	0.9747	0.9210	0.7836	0.7678	0.8992	0.8134
90	0.9763	0.9762	0.9903	0.9559	0.8681	0.8686	0.9493	0.8801
100	0.9903	0.9902	0.9960	0.9729	0.9354	0.9327	0.9778	0.9235
Standard Exponential Samples								
10	0.3570	0.3599	0.2509	0.1689	0.1793	0.1772	0.0710	0.0004
20	0.8327	0.8354	0.8027	0.7086	0.6911	0.6897	0.6152	0.5228
30	0.9704	0.9718	0.9652	0.9134	0.9229	0.9293	0.9060	0.8175
40	0.9961	0.9963	0.9959	0.9740	0.9858	0.9860	0.9834	0.9335
50	0.9996	0.9995	0.9995	0.9927	0.9981	0.9878	0.9977	0.9773
60	1.0000	1.0000	1.0000	0.9982	0.9997	0.9998	0.9997	0.9944
70	1.0000	1.0000	1.0000	0.9995	1.0000	1.0000	1.0000	0.9979
80	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000	0.9996
90	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000	0.9999
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Uniform(0,1) distribution. The χ^2 statistic will follow an approximate Chi-square distribution with $g-1$ degrees of freedom.

2. Power study

One hundred thousand samples are taken for sample sizes of $n = 10, 20, \ldots, 100$. Then the proportions of rejections are computed for the 1% and 5% levels of significance. Samples are taken from the standard normal distribution, the standard exponential distribution, and a mixture of normal distributions. In the power computations, the samples are transformed such that the range of the data is between 0 and 1 to compute all the tests mentioned above. The simulation results are given in Table 6.

3. Conclusion

In Table 6, for the standard normal samples, among the versions of Anderson-Darling tests the V^2 test has consistently higher power compared to the other tests for all the sample sizes except for the sample size of 10. For the Bimodal samples a similar pattern is noticed except for samples of size 20 or less. For the standard exponential samples, there is no clear winner.

Table 6 Continued: Power Study Results

Mixture of Normals $\frac{1}{2}N(-1,(\frac{2}{3})^2)+\frac{1}{2}N(1,(\frac{2}{3})^2)$: Bimodal Samples								
10	0.1314	0.1350	0.0177	0.0089	0.0253	0.0257	0.0011	0.0000
20	0.0819	0.0801	0.0420	0.0558	0.0185	0.0174	0.0063	0.0099
30	0.0919	0.0930	0.0894	0.1041	0.0215	0.0241	0.0213	0.0271
40	0.1239	0.1230	0.1708	0.1553	0.0326	0.0309	0.0527	0.0503
50	0.1795	0.1759	0.2770	0.2404	0.0502	0.0510	0.1119	0.0930
60	0.2561	0.2552	0.3979	0.3231	0.0787	0.0800	0.1843	0.1541
70	0.3558	0.3526	0.5194	0.4273	0.1262	0.1198	0.2818	0.2147
80	0.4566	0.4561	0.6278	0.5091	0.1900	0.1782	0.3789	0.2912
90	0.5556	0.5604	0.7253	0.5974	0.2501	0.2503	0.4862	0.3699
100	0.6550	0.6557	0.8031	0.6583	0.3476	0.3448	0.5978	0.4436

The χ^2 test showed lower power for all sample sizes and for all alternative distribution selections. Rahman and Chakrobartty [8] showed that for testing uniformity, the Anderson-Darling test has higher power in comparison to diffrent versions of the χ^2 tests along with the Cramer-von Misses test and the Watson [10] test.

In this study, it is seen that the modified Anderson-Darling V^2 test has higher power while testing for uniformity. Both the A^2 and the W^2 tests have same mean values for the quantiles and the powers are similar. So, due to convenience, A^2 might be preferred. But to achive a higher power against both symmetric and asymmetric alternatives, V^2 is recommended.

References

[1] T. W. Anderson and D. A. Darling, Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, *Ann. Math. Statistics* **23** (1952), 193–212.

[2] T. W. Anderson and D. A. Darling, A test of goodness of fit, *J. Amer. Statist. Assoc.* **49** (1954), 765–769.

[3] R. B. D'Agostino and M. A. Stephens, *Goodness-of-Fit Techniques* (Eds.), Marcel Dekker, New York, 1986.

[4] W. W. Daniel, *Applied nonparametric statistics*, Second Edition, Duxbury Thomson Learning, Pacific Grove, CA., 1990.

[5] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, Marcel Dekker, Inc., New York, 2003.

[6] David E.A. Giles, A saddlepoint approximation to the distribution function of the Anderson-Darling test statistic, *Communications in Statistics — Simulation and Computation* **30**(4) (2001), 899–905.

[7] P. L'Ecuyer and P. Hellekalek, Random number generators: selection criteria and testing, *Random and Quasi-Random Point sets, Lecture Notes in Statistics*, no. **138**, Springer (1998), 223–266.

[8] M. Rahman and S. Chakrobartty, Tests for uniformity: a comparative study, *J. Korean Data & Information Sci. Soc.* **15**(1) (2004), 211–218.

[9] J. van Soest, Some goodness of fit tests for the exponential distribution, *Statist. Neerlandica* **23** (1969), 41–51.

[10] G. S. Watson, Goodness-of-fit tests on a circle, *Biometrika* **48** (1961), 109–114.