DISTRIBUTION OF THE ANDERSON-DARLING STATISTIC

By Peter A. W. Lewis

IBM Watson Laboratory, New York[1]

In [1] and [2] Anderson and Darling proposed the use of the statistic

$$W_n^2 = n \int_{-\infty}^{\infty} \frac{[G_n(x) - G(x)]^2}{G(x)[1 - G(x)]} \, dG(x) \tag{1}$$

for testing the hypothesis that a sample of size n has been drawn from a population with a specified continuous cumulative distribution function $G(x)$. In (1) $G_n(x)$ is the empirical distribution function defined on the sample of size n.

We consider here the problem of determining and tabulating the distribution function, $F(z; n) = \Pr \{W_n^2 \leqq z\}$, of this statistic. In [1], the asymptotic distribution of this statistic under the null hypothesis was derived and, rewritten in a form convenient for computation, it is given by

$$\begin{aligned} F(z; \infty) &= \lim_{n\to\infty} \Pr \{W_n^2 \leqq z\} \\ &= \sum_{j=0}^{\infty} a_j (zb_j)^{\frac{1}{2}} \exp [-b_j/z] \int_0^{\infty} f_j(y) \exp [-y^2] \, dy, \end{aligned} \tag{2}$$

where

$$\begin{aligned} f_j(y) &= \exp [\tfrac{1}{8} z b_j/(y^2 z + b_j)], \\ a_j &= \frac{(-1)^j (2)^{\frac{1}{2}} (4j + 1)\Gamma(j + \frac{1}{2})}{j!}; \qquad b_j = \tfrac{1}{8} (4j + 1)^2 \pi^2. \end{aligned} \tag{3}$$

Using the calculated values of the a_j's and b_j's, and the fact that

$$\int_0^{\infty} f_j(y) e^{-y^2} \, dy \leqq \tfrac{1}{2} (\pi)^{\frac{1}{2}} \exp [z/8],$$

it can be determined that no more than two terms of the sum ($j = 0, 1$) are needed to evaluate $F(z; \infty)$ to five decimal places over the range of z which is of interest. This range is $0 \leqq z \leqq 8$, since for all n, $F(8; n) = 1.000$, rounded to three decimal places. The integral in each term of the sum was evaluated numerically using Hermite-Gauss quadrature numerical-integration formulas (p. 327 of [3], [4]). This method of numerical integration is very efficient in terms of computing time and gives sufficient accuracy to determine $F(z; \infty)$ to five decimal places.

The results of these calculations of $F(z; \infty)$, rounded to four decimal places,

Received November 23, 1960; revised June 13, 1961.

[1] Present address: IBM Research Laboratory, Monterey and Cottle Roads, San Jose 14, California.

are given in the last column of Table I which appears at the end of this article. The asymptotic significance points given in [2] were verified and are shown in Table II.

An equivalent form of the statistic W_n^2 is given by

$$(4) \quad W_n^2 = -n - (n^{-1}) \sum_{i=1}^{n} [(2i - 1)\ln G(X_{(i)}) + (2(n - i) + 1)\ln(1 - G(X_{(i)}))],$$

where the $X_{(i)}$ are the order statistics of a sample of size n. It is well known that, under the null-hypothesis, the transformation $G(X_{(i)})$ takes the $X_{(i)}$ into the order statistics $U_{(i)}$ of a sample of size n from a population with the uniform (0, 1) distribution, giving

$$(5) \quad W_n^2 = -n - (n^{-1}) \sum_{i=1}^{n} [(2i - 1)\ln U_{(i)} + (2(n - 1) + 1)\ln (1- U_{(i)})].$$

This shows clearly the distribution-free property of this statistic under the null hypothesis, and allows us to determine very simply that, for any n, the minimum value which the random variable W_n^2 can attain is

$$(6) \quad z\,(\min) = -n - \frac{1}{n}\sum_{i=1}^{n} ln \left[\left(\frac{2i - 1}{2n}\right)^{2i-1} \left(\frac{2(n - 1) + 1}{2n}\right)^{2(n-i)+1}\right].$$

These values are tabulated in Table II (following Table I), in the row entitled $F(z) = 0$.

From equation (5), we find that

$$(7) \quad W_1^2 = -1 - \ln [U(1 - U)],$$

where U is uniform (0, 1), so that

$$(8) \quad F(z, 1) = \Pr\{W_1^2 \leqq z\} = \begin{cases} (1 - 4\exp[-(z + 1)])^{\frac{1}{2}}, & z > .38629 \\ 0 & , \ z \leqq .38629 \end{cases}$$

Values of $F(z, 1)$, rounded to three decimal places, are given in Table I.

For $n \geqq 2$ and finite, resort was had to synthetic sampling (Monte Carlo) methods on an IBM 704 Computer to determine the distribution function $F(z; n) = \Pr\{W_n^2 \leqq z\}$. This is done, using equation (5), by artificially generating m samples of size n from a uniform distribution. The result of this process is an empirical distribution function, $F_m(z; n)$, which is used as an estimate of $F(z; n)$. The $F_m(z; n)$ are tabulated in Table I for $n = 2$ up to $n = 8$.

It is necessary to make a determination of the accuracy of these estimates $F_m(z; n)$ of $F(z; n)$ for a given m. This can be done in either of two ways, as follows:

(1). For very large m, $mF_m(z; n)$ is approximately normally distributed, with mean $mF(z; n)$ and variance $mF(z; m)$ $(1 - F(z; m))$. Therefore a confidence interval with confidence coefficient $1 - \alpha$ for $F(z; n)$, at any point z, is given by

(9) $$\{F_m(z; n) \pm \zeta_{\alpha/2}([F(z)(1 - F(z))]/m)^{\frac{1}{2}}\}.$$

In this expression $\zeta_{\alpha/2}$ is the upper $-\alpha/2$ point of the $N(0, 1)$ distribution. In most tables the "error" estimate used is essentially the above confidence interval with confidence coefficient 0.6868, i.e., $\zeta_{\alpha/2} = 1$. Now $F(z; n)$ is unknown, but $F(z; n)\ (1 - F(z; n))$ is maximum at $F(z; n) = 0.5$, so that to keep this "error" less than or equal to .0005 over the range of z, one requires an $m = 10^6$.

(2). Another means of evaluating the error is by using the Kolmogorov-Smirnov statistic, from which, for large m, we can say that we are 95 percent sure that $F_m(z; n)$ will stay within $1.36(m)^{-\frac{1}{2}}$ of the true distribution $F(z; n)$ for all z, i.e., over the entire distribution. Therefore to make this statement for a deviation of .0005, we need $m = 7.398 \times 10^6$.

Unfortunately the time available for computation limited the value of m used in these computations to $m = 10^6$ for $n = 2$, and to $m = .25 \times 10^6$ for $n = 3, 4, 5, 6, 7$, and 8. Thus, using the Kolmogorov-Smirnov criterion, the values $F_m(z; 2)$ given in Table I are within .00163 of $F(z, 2)$ with probability 0.95, and for $n = 3, 4, 5, 6, 7$, and 8 the values $F_m(z, n)$ are within .00326 of $F(z; n)$ with probability 0.95.

Determination of the distribution of W_n^2 for $n > 8$ by Monte Carlo methods is prohibitive, since for $n = 8$ and $m = 250{,}000$, six hours of computing time were required. This is quite indicative of the inefficiency and impracticability of simple Monte Carlo methods as a means of solving distribution theory problems when the entire distribution function is required with great accuracy. Furthermore, it is doubtful whether modified Monte Carlo methods ([5], [6], [7]) could be used to advantage here.

Fortunately the convergence of the distribution of W_n^2 to its asymptotic distribution is quite rapid. Thus, from Table I the maximum deviation at the tabulated points between the asymptotic distribution and the distribution for $n = 8$ is approximately 0.006. For $F(z) \geqq 0.8$, which is of most interest, this difference is only 0.001, so that for practical purposes the asymptotic distribution can be used for $n > 8$.

Significance points for W_n^2 are given in Table II, for significance levels 0.100, 0.050, 0.010. For $n = 1$ and $n \to \infty$ these values are exact; the others are obtained by inverse interpolation from Table I and are only approximate.

Acknowledgments. I wish to thank Mr. H. Serenson for his help in programming this problem.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 765–769.

[3] F. B. HILDEBRAND, *Introduction to Numerical Analysis*, McGraw-Hill, New York, 1956.
[4] H. E. SALZER, R. ZUCKER AND R. CAPUANO, "Tables of the zeros and weight factors of the first twenty Hermite polynomials," *J. of Res. of Natl. Bur. Stan.*, Vol. 48 (1952), pp. 111-116.
[5] J. M. HAMMERSLEY AND K. W. MORTON, "A new Monte Carlo technique: antithetic variates," *Proc. Camb. Philos. Soc.*, Vol. 52 (1956), pp. 449–475.
[6] J. M. HAMMERSLEY AND J. G. MAULDON, "General principles of antithetic variates," *Proc. Camb. Philos. Soc.*, Vol. 52 (1956), pp. 476–481.
[7] J. M. HAMMERSLEY, "Conditional Monte Carlo," *J. Assn. Comp. Mach.*, Vol. 3 (1956), pp. 73–76.

Tables follow on pages 1122–4

TABLE I - Values of F(z; n) - Exact for n = 1 and n→∞; Estimated for n = 2, 3, 4, 5, 6, 7, and 8

z \ n	F(z; n) 1	2	3	4	5	6	7	8	n→∞
.025									0.0000
.050									0.0000
.075									0.0000
.100								0.000	0.0000
.125						0.000	0.000	0.000	0.0003
.150					0.000	0.001	0.001	0.001	0.0014
.175				0.001	0.003	0.003	0.003	0.004	0.0042
.200			0.008	0.007	0.008	0.009	0.008	0.009	0.0096
.225			0.016	0.016	0.016	0.017	0.017	0.017	0.0180
.250		0.001	0.028	0.028	0.028	0.029	0.029	0.029	0.0296
.275		0.030	0.044	0.043	0.044	0.044	0.044	0.045	0.0443
.300		0.059	0.063	0.063	0.063	0.062	0.062	0.063	0.0618
.325		0.087	0.083	0.085	0.083	0.084	0.083	0.084	0.0817
.350		0.115	0.106	0.109	0.106	0.106	0.106	0.106	0.1036
.375		0.142	0.130	0.134	0.130	0.131	0.130	0.130	0.1269
.400	0.116	0.169	0.159	0.161	0.156	0.156	0.155	0.155	0.1513
.425	0.195	0.196	0.187	0.187	0.182	0.182	0.181	0.181	0.1764
.450	0.248	0.222	0.217	0.212	0.208	0.208	0.207	0.207	0.2019
.475	0.291	0.248	0.248	0.238	0.235	0.234	0.233	0.233	0.2276
.500	0.328	0.273	0.271	0.264	0.261	0.260	0.259	0.259	0.2532
.525	0.360	0.298	0.295	0.289	0.287	0.285	0.284	0.284	0.2786
.550	0.389	0.323	0.320	0.314	0.312	0.310	0.309	0.309	0.3036
.575	0.415	0.347	0.345	0.340	0.337	0.335	0.334	0.334	0.3281
.600	0.439	0.371	0.371	0.364	0.361	0.359	0.358	0.358	0.3520
.625	0.461	0.394	0.396	0.387	0.384	0.382	0.381	0.381	0.3753
.650	0.481	0.418	0.418	0.410	0.407	0.404	0.403	0.404	0.3980
.675	0.501	0.440	0.439	0.431	0.429	0.426	0.424	0.425	0.4199
.700	0.519	0.463	0.459	0.452	0.449	0.446	0.446	0.446	0.4412
.750	0.552	0.507	0.496	0.491	0.489	0.486	0.486	0.487	0.4815
.800	0.582	0.547	0.530	0.528	0.525	0.524	0.523	0.523	0.5190
.850	0.609	0.580	0.567	0.563	0.559	0.559	0.557	0.557	0.5537
.900	0.634	0.610	0.598	0.593	0.591	0.590	0.588	0.589	0.5858
.950	0.656	0.636	0.626	0.622	0.620	0.619	0.618	0.619	0.6154
1.000	0.677	0.660	0.652	0.648	0.647	0.646	0.645	0.646	0.6427
1.050	0.696	0.683	0.676	0.673	0.672	0.671	0.669	0.670	0.6680
1.100	0.714	0.703	0.698	0.696	0.694	0.695	0.693	0.694	0.6912
1.150	0.731	0.722	0.719	0.717	0.715	0.716	0.714	0.714	0.7127
1.200	0.746	0.739	0.738	0.736	0.734	0.735	0.734	0.733	0.7324
1.250	0.761	0.756	0.755	0.754	0.752	0.753	0.752	0.751	0.7508
1.300	0.774	0.770	0.770	0.770	0.768	0.770	0.768	0.769	0.7677
1.350	0.786	0.784	0.785	0.785	0.784	0.785	0.784	0.784	0.7833
1.400	0.798	0.798	0.799	0.799	0.798	0.799	0.798	0.798	0.7978
1.450	0.809	0.809	0.811	0.812	0.811	0.812	0.812	0.811	0.8111
1.500	0.820	0.821	0.823	0.824	0.824	0.824	0.824	0.824	0.8235
1.550	0.829	0.831	0.833	0.835	0.835	0.835	0.835	0.835	0.8350
1.600	0.838	0.842	0.843	0.845	0.845	0.845	0.846	0.846	0.8457
1.650	0.847	0.851	0.852	0.855	0.854	0.855	0.855	0.855	0.8556
1.700	0.855	0.860	0.861	0.864	0.864	0.864	0.864	0.864	0.8648
1.750	0.863	0.868	0.869	0.872	0.872	0.872	0.873	0.873	0.8734
1.800	0.870	0.875	0.877	0.880	0.880	0.880	0.880	0.881	0.8814
1.850	0.877	0.883	0.884	0.887	0.887	0.887	0.888	0.888	0.8888
1.900	0.883	0.889	0.891	0.894	0.894	0.894	0.895	0.895	0.8957
1.950	0.889	0.896	0.898	0.900	0.901	0.900	0.901	0.901	0.9021
2.000	0.895	0.902	0.904	0.906	0.907	0.906	0.907	0.907	0.9082
2.050	0.900	0.907	0.909	0.912	0.912	0.912	0.913	0.913	0.9138
2.100	0.905	0.912	0.915	0.917	0.917	0.918	0.918	0.918	0.9190
2.150	0.910	0.917	0.920	0.922	0.922	0.923	0.922	0.923	0.9239

TABLE I (Continued)

F(z; n)

z \ n	1	2	3	4	5	6	7	8	n → ∞
2.200	0.915	0.922	0.924	0.926	0.927	0.927	0.927	0.927	0.9285
2.250	0.919	0.926	0.928	0.931	0.931	0.931	0.931	0.931	0.9328
2.300	0.923	0.930	0.933	0.935	0.935	0.935	0.935	0.935	0.9368
2.350	0.927	0.934	0.937	0.939	0.939	0.939	0.939	0.939	0.9405
2.400	0.931	0.938	0.940	0.942	0.942	0.942	0.942	0.942	0.9441
2.450	0.934	0.941	0.943	0.945	0.945	0.946	0.946	0.945	0.9474
2.500	0.938	0.944	0.947	0.948	0.949	0.949	0.949	0.949	0.9504
2.550	0.941	0.948	0.950	0.951	0.951	0.952	0.952	0.952	0.9534
2.600	0.944	0.950	0.953	0.954	0.954	0.954	0.955	0.954	0.9561
2.650	0.947	0.953	0.955	0.957	0.957	0.957	0.957	0.957	0.9586
2.700	0.949	0.956	0.958	0.959	0.959	0.959	0.960	0.959	0.9610
2.750	0.952	0.958	0.960	0.961	0.961	0.961	0.062	0.962	0.9633
2.800	0.954	0.960	0.962	0.964	0.964	0.964	0.964	0.964	0.9654
2.850	0.957	0.962	0.964	0.965	0.965	0.965	0.966	0.966	0.9674
2.900	0.959	0.964	0.966	0.967	0.967	0.967	0.968	0.968	0.9692
2.950	0.961	0.966	0.968	0.969	0.969	0.969	0.970	0.969	0.9710
3.000	0.963	0.968	0.970	0.971	0.971	0.971	0.971	0.971	0.9726
3.050	0.965	0.970	0.972	0.972	0.972	0.972	0.973	0.973	0.9742
3.100	0.966	0.971	0.973	0.974	0.974	0.974	0.975	0.974	0.9756
3.150	0.968	0.973	0.975	0.975	0.975	0.975	0.976	0.976	0.9770
3.200	0.970	0.974	0.976	0.977	0.977	0.977	0.977	0.977	0.9783
3.250	0.971	0.075	0.978	0.978	0.978	0.978	0.978	0.978	0.9795
3.300	0.973	0.977	0.979	0.979	0.979	0.979	0.979	0.979	0.9807
3.350	0.974	0.978	0.980	0.980	0.980	0.980	0.981	0.980	0.9818
3.400	0.975	0.979	0.981	0.981	0.981	0.981	0.982	0.981	0.9828
3.450	0.976	0.980	0.982	0.982	0.982	0.982	0.983	0.983	0.9837
3.500	0.978	0.981	0.983	0.983	0.983	0.983	0.983	0.984	0.9846
3.550	0.979	0.982	0.984	0.984	0.984	0.984	0.984	0.984	0.9855
3.600	0.980	0.983	0.985	0.985	0.985	0.985	0.985	0.985	0.9863
3.650	0.981	0.984	0.986	0.986	0.986	0.986	0.986	0.986	0.9870
3.700	0.982	0.985	0.986	0.986	0.987	0.986	0.987	0.987	0.9878
3.750	0.983	0.986	0.987	0.987	0.987	0.987	0.987	0.988	0.9884
3.800	0.983	0.986	0.988	0.988	0.988	0.988	0.988	0.988	0.9891
3.850	0.984	0.987	0.988	0.988	0.989	0.988	0.989	0.989	0.9897
3.900	0.985	0.988	0.989	0.989	0.989	0.989	0.989	0.989	0.9902
3.950	0.986	0.988	0.990	0.989	0.990	0.990	0.990	0.990	0.9908
4.000	0.986	0.989	0.990	0.990	0.990	0.990	0.990	0.990	0.9913
4.050	0.987	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.9917
4.100	0.988	0.990	0.991	0.991	0.991	0.991	0.991	0.991	0.9922
4.150	0.988	0.991	0.991	0.991	0.992	0.992	0.992	0.992	0.9926
4.200	0.989	0.991	0.992	0.992	0.992	0.992	0.992	0.992	0.9930
4.250	0.989	0.992	0.992	0.992	0.993	0.992	0.993	0.993	0.9934
4.300	0.990	0.992	0.993	0.993	0.993	0.993	0.933	0.993	0.9938
4.350	0.991	0.992	0.993	0.993	0.993	0.993	0.993	0.993	0.9941
4.400	0.991	0.992	0.993	0.993	0.994	0.993	0.993	0.994	0.9944
4.500	0.992	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.9950
4.600	0.993	0.994	0.995	0.995	0.995	0.995	0.995	0.995	0.9955
4.700	0.993	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.9960
4.800	0.994	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.9964
4.900	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.9968
5.000	0.995	0.996	0.996	0.997	0.996	0.996	0.996	0.997	0.9971
5.500	0.997	0.998	0.998	0.998	0.998	0.998	0.988	0.998	0.9983
6.000	0.998	0.999	0.998	0.999	0.999	0.999	0.999	0.999	0.9990
7.000	0.998	1.000	0.999	0.999	0.999	0.999	1.000	0.999	0.9997
8.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.9999

TABLE II - Significance Points and Values of z(min.)

$$z = F^{-1}\left(F(z)\right)$$

F(z) \ n	1	2	3	4	5	6	7	8	$n \to \infty$
0	0.3863	0.2493	0.1885	0.1533	0.1304	0.1135	0.1043	0.0911	0
.90	2.0470	1.98	1.97	1.95	1.94	1.95	1.94	1.94	1.933
.95	2.7142	2.60	2.55	2.53	2.53	2.52	2.52	2.52	2.492
.99	4.3033	4.10	4.00	4.00	3.95	3.95	3.95	3.95	3.857