NEW TABLE AND NUMERICAL APPROXIMATIONS FOR KOLMOGOROV-SMIRNOV/LILLIEFORS/VAN SOEST NORMALITY TEST

PAUL MOLIN AND HERVÉ ABDI

ABSTRACT. We give new critical values for the Kolmogorov-Smirnov/Lilliefors/Van Soest test of Normality. These values are obtained from Monte-Carlo simulations similar to the original procedure of Lilliefors and Van Soest. Because our simulations use a very large number of random samples, the critical values obtained are better estimations than the original values. In order to allow hypothesis testing with arbitrary α levels, we also derive a polynomial approximation of the critical values. This facilitates the implementation of Bonferonni or Šidák corrections for multiple statistical tests as these procedures require unusual α values.

1. INTRODUCTION

The normality assumption is at the core of a majority of standard statistical procedures, and it is important to be able to test this assumption. In addition, showing that a sample does not come from a normally distributed population is sometimes of importance *per se*. Among the procedures used to test this assumption, one of the most well-known is a modification of the Kolomogorov-Smirnov test of goodness of fit, generally referred to as the *Lilliefors test for normality* (or Lilliefors test, for short). This test was developed independently by Lilliefors (1967) and by Van Soest (1967). Like most statistical tests,

Date: February 20, 1998.

We would like to thank Dominique Valentin and Mette Posamentier for comments on previous drafts of this paper. Thanks are also due to the personal of Computer Center of the "Université de Dijon," in particular Jean-Jacques Gaillard and Jean-Christophe Basaille.

Correspondence about this paper should addressed to Hervé Abdi (herve@utd.edu). Cite this paper as:
Molin, P., Abdi H. (1998). New Tables and numerical approximation for the Kolmogorov- Smirnov/Lillierfors/Van Soest test of normality. *Technical report, University of Bourgogne.* Available from www.utd.edu/∼herve/MolinAbdi1998-LillieforsTechReport.pdf.

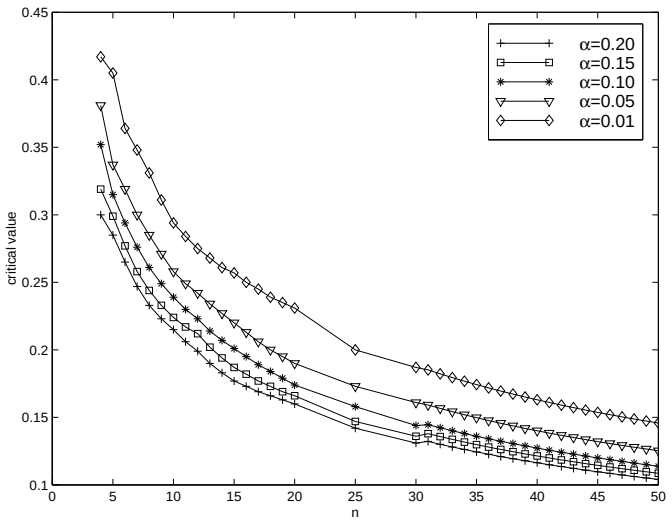

FIGURE 1. Critical values from Lilliefors (1967) ploted as a function of the sample size and the α level.

this test of normality defines a criterion and gives its sampling distribution. Because the sampling distribution cannot be derived by standard analytical procedures, it is approximated with Monte Carlo numerical simulation procedures. Specifically, both Lilliefors and Van Soest used, for each sample size chosen, 1000 random samples derived from a standardized normal distribution to approximate the sampling distribution of a Kolmogorov-Smirnov criterion of goodness of fit. The critical values given by Lilliefors and Van Soest are quite similar, the relative error being of the order of 10^{-2}. Lilliefors (1967) noted that this approach is more powerful than the standard chi-square procedure for a wide range of nonnormal conditions. Dagnelie (1968) indicated, in addition, that the critical values reported by Lilliefors can be approximated by an analytical formula. Such a formula facilitates writing computer routines because it eliminates the risk of creating errors when keying in the values of the table.

There are some small problems, however, with the current tables for the Lilliefors test. The first one comes from the rather small number of samples (*i.e.,* 1000) used in the original simulations: The precision of the table could be improved with a larger number of samples. This problem can be seen in Figure 1, which displays the critical values from Lilliefors (1967) as a function of the sample size. The rather jagged appearance of the curves suggests that the critical values are contaminated by random fluctuations. A larger number of samples

would reduce these fluctuations and make the curves look smoother. The second problem, of greater concern, comes from the limited number of critical values reported in the original papers. Lilliefors, for example, reports the critical values for $\alpha = [.20,\ .15,\ .10,\ .05,\ .01]$. These values correspond to most tests involving only one null hypothesis, as this was the standard procedure in the late sixties. The current statistical practice favors multiple tests (maybe as a consequence of the availability of statistical packages). Because using multiple tests increases the overall Type I error (*i.e.,* the Familywise Type I error or α_{PF}), it has become customary to recommend testing each hypothesis with a corrected α level (*i.e.,* the Type I error *per* comparison, or α_{PC}). The main correction procedures are the Bonferonni and the Šidák corrections[1] (*e.g.,* Abdi, 1987). For a family of J tests, the Bonferonni correction expresses α_{PC} as

$$\alpha_{PC} = \tfrac{1}{J}\alpha_{PF}\ , \tag{1.1}$$

whereas the Šidák correction expresses α_{PC} as

$$\alpha_{PC} = 1 - (1 - \alpha_{PF})^{\frac{1}{J}}\ . \tag{1.2}$$

For example, using a Bonferonni approach with a familywise value of $\alpha_{PF} = .05$, and testing $J = 3$ hypotheses requires that each hypothesis is tested at the level of

$$\alpha_{PC} = \tfrac{1}{J}\alpha_{PF} = \tfrac{1}{3} \times .05 = .0167\ . \tag{1.3}$$

With a Šidák approach, each hypothesis will be tested at the level of

$$\alpha_{PC} = 1 - (1 - \alpha_{PF})^{\frac{1}{J}} = 1 - (1 - .05)^{\frac{1}{3}} = .0170\ . \tag{1.4}$$

As this example illustrates, both procedures are likely to require using different α levels than the ones given by the original sources. In fact, it is rather unlikely that a table could be precise enough to provide the wide range of alpha values needed for multiple testing purposes. A more practical solution is to generate the critical values for any alpha value, or, alternatively, to obtain the probability associated to any value of the Kolmogorov-Smirnov criterion. In brief, the purpose of this paper is to give better numerical approximations for the Kolmogorov-Smirnov test of normality, and to derive an analytical formula for the critical values of the criterion.

This paper is organized as follows: first, we present the numerical simulations used to approximate the sampling distribution of the Lilliefors test; second, we derive a numerical expression for the critical

[1]Bonferonni is simply an approximation using the first term of a Taylor series of Šidák.

values of the criterion as a function of the sample size and the α level; third, we give a numerical approximation of the probability associated to the criterion.

2. Monte Carlo Approximation of the Sampling Distribution

2.1. **Computation of the criterion for the Kolmogorov-Smirnov test of normality.** The null hypothesis tested by the Lilliefors test is

$$H_0 = \text{The sample comes from a normal population with unknown mean and variance.} \tag{2.1}$$

2.1.1. *Notations.* The sample for the test is made of N scores, each of them denoted X_i. The sample mean is denoted M and the sample variance is denoted S^2. The criterion for the Lilliefors test is denoted D_L. It is computed from Z_i scores which are obtained with the following formula:

$$Z_i = \frac{X_i - M}{S} \tag{2.2}$$

where S is the square root of

$$S^2 = \frac{\sum_i^N (X_i - M)^2}{N-1} \tag{2.3}$$

and M is

$$M = \frac{1}{N}\sum_i^N X_i. \tag{2.4}$$

The criterion D_L is

$$D_L = \max_i \{|\mathcal{S}(Z_i) - \mathcal{N}(Z_i)|, |\mathcal{S}(Z_i) - \mathcal{N}(Z_{i-1})|\} \tag{2.5}$$

where $\mathcal{S}$ is the relative frequency associated with Z_i. It corresponds to the proportion of scores smaller or equal to Z_i and where $\mathcal{N}$ is the probability associated to a normally distributed variable Z_i with mean $\mu = 0$ and standard deviation $\sigma = 1$. The term $|\mathcal{S}(Z_i) - \mathcal{N}(Z_{i-1})|$ is needed to take into account that, because the empirical distribution is discrete, the maximum absolute difference can occur at either endpoints of the empirical distribution.

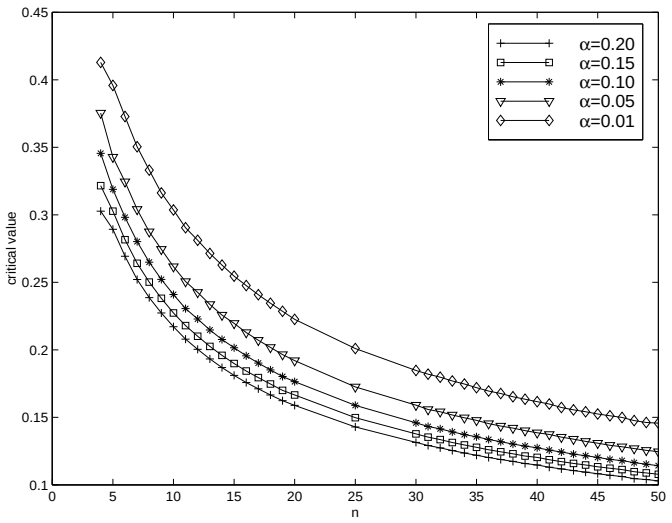

FIGURE 2. Plot of the new critical values for the Kolmogorov-Smirnov test of normality. The critical values are ploted as a function of the sample size and the α level.

2.2. **Numerical simulations.** The principle of the Monte Carlo simulations is to approximate the sampling distribution of the D_L criterion from its relative frequency distribution obtained when the null hypothesis is true. For a sample of size $N > 3$, the sampling distribution is obtained by first generating a total of K random samples of size N, and computing for each sample the value of the D_L criterion. Then, the relative frequency distribution of the criterion estimates its sampling distribution. The critical value for a given α level is computed as the $K(1-\alpha)th$ percentile of the relative frequency distribution.

The numerical simulations were performed on IBM RS6000-590 running AIX 4.1. The program (available from the authors) was written using the MATLAB programming language (version 4.2c). The random number generator used was `randn` (Forsythe, Malcom, Moler, 1977).

2.3. **Results.** The new values for the Kolmogorov-Smirnov test of normality are given in Table 1. For ease of comparison with the original values computed by Lilliefors (*cf.* Figure 1), the results are displayed in Figure 2. As expected from the large sample size of our simulations, the curves of the new simulations are much smoother than the original ones.

N	$\alpha = 0.20$	$\alpha = 0.15$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
4	0.3027	0.3216	0.3456	0.3754	0.4129
5	0.2893	0.3027	0.3188	0.3427	0.3959
6	0.2694	0.2816	0.2982	0.3245	0.3728
7	0.2521	0.2641	0.2802	0.3041	0.3504
8	0.2387	0.2502	0.2649	0.2875	0.3331
9	0.2273	0.2382	0.2522	0.2744	0.3162
10	0.2171	0.2273	0.2410	0.2616	0.3037
11	0.2080	0.2179	0.2306	0.2506	0.2905
12	0.2004	0.2101	0.2228	0.2426	0.2812
13	0.1932	0.2025	0.2147	0.2337	0.2714
14	0.1869	0.1959	0.2077	0.2257	0.2627
15	0.1811	0.1899	0.2016	0.2196	0.2545
16	0.1758	0.1843	0.1956	0.2128	0.2477
17	0.1711	0.1794	0.1902	0.2071	0.2408
18	0.1666	0.1747	0.1852	0.2018	0.2345
19	0.1624	0.1700	0.1803	0.1965	0.2285
20	0.1589	0.1666	0.1764	0.1920	0.2226
25	0.1429	0.1498	0.1589	0.1726	0.2010
30	0.1315	0.1378	0.1460	0.1590	0.1848
31	0.1291	0.1353	0.1432	0.1559	0.1820
32	0.1274	0.1336	0.1415	0.1542	0.1798
33	0.1254	0.1314	0.1392	0.1518	0.1770
34	0.1236	0.1295	0.1373	0.1497	0.1747
35	0.1220	0.1278	0.1356	0.1478	0.1720
36	0.1203	0.1260	0.1336	0.1454	0.1695
37	0.1188	0.1245	0.1320	0.1436	0.1677
38	0.1174	0.1230	0.1303	0.1421	0.1653
39	0.1159	0.1214	0.1288	0.1402	0.1634
40	0.1147	0.1204	0.1275	0.1386	0.1616
41	0.1131	0.1186	0.1258	0.1373	0.1599
42	0.1119	0.1172	0.1244	0.1353	0.1573
43	0.1106	0.1159	0.1228	0.1339	0.1556
44	0.1095	0.1148	0.1216	0.1322	0.1542
45	0.1083	0.1134	0.1204	0.1309	0.1525
46	0.1071	0.1123	0.1189	0.1293	0.1512
47	0.1062	0.1113	0.1180	0.1282	0.1499
48	0.1047	0.1098	0.1165	0.1269	0.1476
49	0.1040	0.1089	0.1153	0.1256	0.1463
50	0.1030	0.1079	0.1142	0.1246	0.1457

TABLE 1. Table of the critical values for the Kolmogorov-Smirnov test of normality obtained with $K = 100,000$ samples for each sample size. The intersection of a given row and column shows the critical value $\mathsf{C}(N, \alpha)$ for the sample size labelling the row and the alpha level labelling the column.

3. Numerical approximations for the critical values

Recall that Dagnelie (1968) indicated that the critical values given by Lilliefors can be approximated numerically. Specifically, the critical values $\mathsf{C}(N,\alpha)$ can be obtained as a function of α and the sample size (denoted N) with an expression of the form

$$\mathsf{C}(N,\alpha) = \frac{a(\alpha)}{\sqrt{N+1.5}} \ , \tag{3.1}$$

where $a(\alpha)$ is a function of α. This approximation is precise enough when $N > 4$, and for the usual alpha levels of $\alpha = .05$ [$a(\alpha) = 0.886$] and $\alpha = .01$ [$a(\alpha) = 1.031$], but does not give good results for other alpha levels. However, Equation 3.1 suggests a more general expression for the value $\mathsf{C}(N,\alpha)$ as a function of the sample size and the α value of the form:

$$\mathsf{C}(N,\alpha) = \frac{a(\alpha)}{\sqrt{N+b(\alpha)}} \ . \tag{3.2}$$

For mathematical tractability, it is convenient to re-express Equation 3.2 in order to indicate a linear relationship between the parameters. This is obtained by the transformation

$$\mathsf{C}(N,\alpha)^{-2} = A(\alpha)N + B(\alpha) \ , \tag{3.3}$$

with

$$A(\alpha) = a(\alpha)^{-2} \Longleftrightarrow a(\alpha) = A(\alpha)^{-\frac{1}{2}} \tag{3.4}$$

$$B(\alpha) = b(\alpha)a(\alpha)^{-2} \Longleftrightarrow b(\alpha) = B(\alpha)A(\alpha)^{-1} \ . \tag{3.5}$$

Figure 3 illustrates the linearity of the relationship between $\mathsf{C}(N,\alpha)^{-2}$ and N when α is given.

3.0.1. *Fitting $A(\alpha)$ and $B(\alpha)$.* The relationship between $A(\alpha)$ and $B(\alpha)$ is clearly *non*linear as illustrated by Figure 4 which plots $A(\alpha)$ and $B(\alpha)$ as a function of α.

A stepwise polynomial multiple regression was used to derive a polynomial approximation for $A(\alpha)$ and $B(\alpha)$ as a function of α. We found that a polynomial of degree 6 gave the best fit when using the maximal absolute error as a criterion (see Table 3).

The coefficients of the polynomial were found to be equal to:

$$\begin{aligned} A(\alpha) = {} & +6.32207539843126 - 17.1398870006148(1-\alpha) \\ & + 38.42812675101057(1-\alpha)^2 - 45.93241384693391(1-\alpha)^3 \\ & + 7.88697700041829(1-\alpha)^4 + 29.79317711037858(1-\alpha)^5 \\ & - 18.48090137098585(1-\alpha)^6 \ . \end{aligned} \tag{3.6}$$

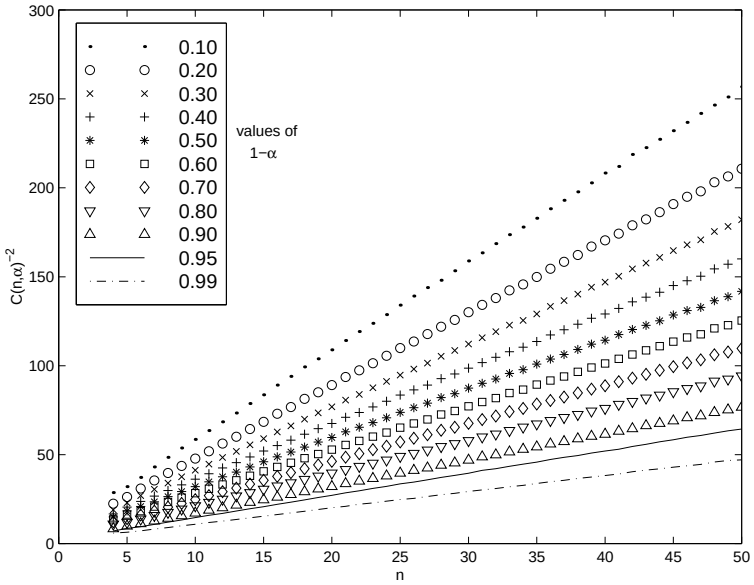

FIGURE 3. Dependence of critical values with respect to n.

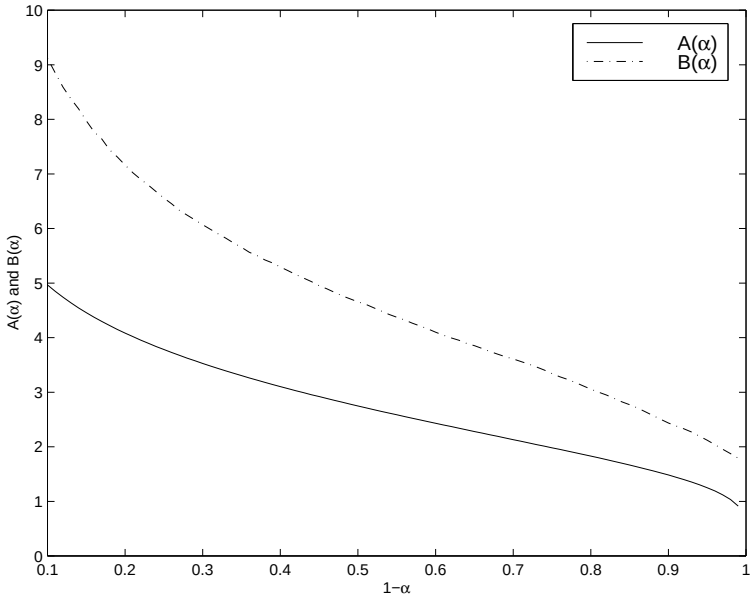

FIGURE 4. Dependence of coefficients $a(\alpha)$ and $b(\alpha)$ *vs* α.

and

$$\begin{aligned} B(\alpha) = &+12.940399038404 - 53.458334259532(1-\alpha) \\ &+ 186.923866119699(1-\alpha)^2 - 410.582178349305(1-\alpha)^3 \\ &+ 517.377862566267(1-\alpha)^4 - 343.581476222384(1-\alpha)^5 \\ &+ 92.123451358715(1-\alpha)^6 \ . \end{aligned} \tag{3.7}$$

TABLE 2. Maximum residual of polynomial regressions.

Degree	Max. Diff.
3	0.0146
4	0.0197
5	0.0094
6	**0.0092**
7	0.0113
8	0.0126
9	0.0140
10	0.0146

TABLE 3. Maximum absolute value of the residual for the polynomial approximation of $\mathsf{C}(N, \alpha)$ from $A(\alpha)$ and $B(\alpha)$. The smallest value is reached for a degree 6 polynomial (in boldface) .

The quality of the prediction is illustrated in Figure 5, which shows the predicted and actual values of $A(\alpha)$ and $B(\alpha)$ as well as the error of prediction plotted as a function of α. The predicted values overlap almost perfectly with the actual values (and, so the error is always small); this confirms the overall good quality of the prediction.

4. Finding the probability associated to D_L

Rather than using critical values for hypothesis testing, an alternative strategy is to work directly the probability associated to a specific value of D_L. The probability associated to a given value of D_L is denoted $\Pr(D_L)$. If this probability is smaller than the α level, the null hypothesis is rejected. It does not seem possible to find an analytic expression for $\Pr(D_L)$ (or at least, we failed to find one!). However, the results of the previous section suggest an algorithmic approach which can easily be programmed. In short, in this section we derive a polynomial approximation for $\Pr(D_L)$.

The starting point comes from a the observation that $B(\alpha)$ can be predicted from $A(\alpha)$ using a quadratic regression ($R^2 = .99963$) as:

$$B(\alpha) = b_0 + b_1 A(\alpha) + b_2 A(\alpha)^2 + \epsilon \ . \tag{4.1}$$

where ϵ is the error of prediction, and

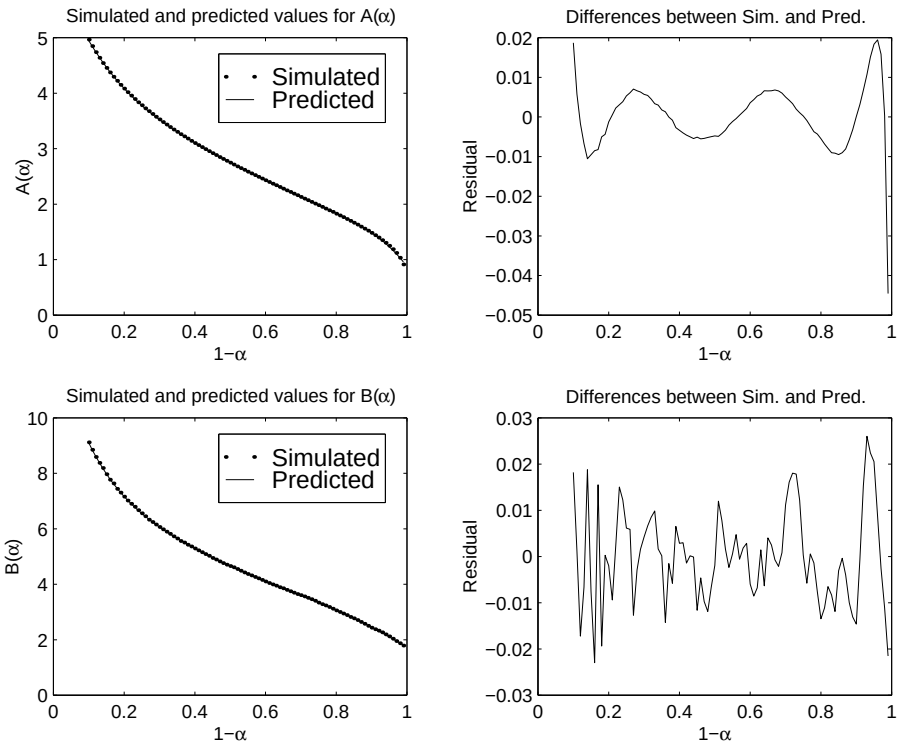

FIGURE 5. Polynomial regressions of coefficients $A(\alpha)$ and $B(\alpha)$*vs* α: predicted values and residuals.

$$
\begin{aligned}
b_2 &= 0.08861783849346 \\
b_1 &= 1.30748185078790 \\
b_0 &= 0.37872256037043 \; .
\end{aligned} \tag{4.2}
$$

Plugging this last equation in Equation 3.3, and replacing $\mathsf{C}(N,\alpha)$ by D_L gives:

$$D_L^{-2} = b_0 + (b_1 + N)A(\alpha) + b_2 A(\alpha)^2 \; . \tag{4.3}$$

When D_L and N are known, and $A(\alpha)$ is the unknown. Therefore, $A(\alpha)$ is a solution of the following quadratic equation:

$$b_0 + (b_1 + N)A(\alpha) + b_2 A(\alpha)^2 - D_L^{-2} = 0. \tag{4.4}$$

The solution of this equation is obtained from the traditional quadratic formula:

$$A(\alpha) = \frac{-(b_1 + N) + \sqrt{(b_1 + N)^2 - 4b_2\left(b_0 - D_L^{-2}\right)}}{2b_2} \; . \tag{4.5}$$

We need, now, to obtain $\Pr(D_L)$ as a function of $A(\alpha)$. To do that, first we solve Equation 4.5, for $A(\alpha)$. Second, we express α as a function of $A(\alpha)$. The value of α corresponding to a given value of $A(\alpha)$ will give $\Pr(D_L)$. We found that α can be expressed as a degree 10 polynomial in $A(\alpha)$ (a higher polynomial does not improve the quality of prediction). Therefore $\Pr(D_L)$ can be estimated, from $A(\alpha)$, as

$$\begin{aligned}\Pr(D_L) = &-.37782822932809 + 1.67819837908004A(\alpha) \\ &- 3.02959249450445A(\alpha)^2 + 2.80015798142101A(\alpha)^3 \\ &- 1.39874347510845A(\alpha)^4 + 0.40466213484419A(\alpha)^5 \\ &- 0.06353440854207A(\alpha)^6 + 0.00287462087623A(\alpha)^7 \\ &+ 0.00069650013110A(\alpha)^8 - 0.00011872227037A(\alpha)^9 \\ &+ 0.00000575586834A(\alpha)^{10} + \epsilon \ . \end{aligned} \tag{4.6}$$

For example, suppose that we have obtained a value of $D_L = .1030$ from a sample of size $N = 50$. (Table 1 shows that $\Pr(D_L) = .20$.) To estimate $\Pr(D_L)$ we need first to compute $A(\alpha)$, and then use this value in Equation 4.6. From Equation 4.5, we compute the estimate of $A(\alpha)$ as:

$$\begin{aligned} A(\alpha) &= \frac{-(b_1 + N) + \sqrt{(b_1 + N)^2 - 4b_2\left(b_0 - D_L^{-2}\right)}}{2b_2} \\ &= \frac{-(b_1 + 50) + \sqrt{(b_1 + 50)^2 - 4b_2\left(b_0 - .1030^{-2}\right)}}{2b_2} \\ &= 1.82402308769590 \ . \end{aligned} \tag{4.7}$$

Plugging in this value of $A(\alpha)$ in Equation 4.6 gives

$$\Pr(D_L) = .19840103775379 \approx .20 \ . \tag{4.8}$$

As illustrated by this example, the approximated value of $\Pr(D_L)$ is correct for the first two decimal values.

5. Conclusion

In this paper we give a new table for the critical values of the Lilliefors test. We also derive numerical approximations that give directly the critical values and the probability associated to a given value of the criterion for this test.

6. References

(1) Abdi, H., *Introduction au traitement statistique des données expérimentales*. Grenoble: Presses Universitaires de Grenoble. 1987.
(2) Dagnelie, P., A propos de l'emploi du test de Kolmogorov-Smirnov comme test de normalité, *Biom. Praxim., IX, 1,* 3–13, 1968.
(3) Forsythe, G. E., Malcolm, M. A., Moler, C.B., *Computer Methods for Mathematical computations*, Prentice-Hall, 1977.
(4) Lilliefors, H. W., On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *Jour. Amer. Stat Assoc.. 62,* 399–402, 1967.
(5) Van Soest, J., Some experimental results concerning tests of normality, *Stat. Neerl., 21*, 91–97, 1967.

ensbana, Université de Dijon, France
E-mail address: pmolin@u-bourgogne.fr

The University of Texas at Dallas
E-mail address: herve@utdallas.edu