© Journal of the American Statistical Association
September 1972, Vo ume 67, Number 339
Theory and Methods Section

A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions

W. J. CONOVER*

The Kolmogorov goodness-of-fit test is known to be conservative when the hypothesized distribution function is not continuous. A method for finding the exact critical level (approximate in the two-sided case) and the power in such cases is derived. Thus the Kolmogorov test may be used as an exact goodness-of-fit test for all completely specified distribution functions, whether continuous or not continuous. Several examples of the application of this extension of the Kolmogorov test are also included.

1. INTRODUCTION

A goodness-of-fit test is a test of the hypothesis

$$H_0: F(x) = H(x) \qquad \text{for all } x,$$

where $F(x) = P(X \leq x)$ is the unknown distribution function associated with the population being studied, and $H(x)$ is some hypothesized distribution function. If $H(x)$ is a continuous distribution function, the Kolmogorov or Cramér-von Mises goodness-of-fit tests may be used to find the exact critical level, the probability of getting a value of the test statistic as extreme as the observed value when H_0 is true. There is a need for a goodness-of-fit test which will provide exact critical levels when $H(x)$ is not continuous. The Kolmogorov test is extended to provide such a test in this article, except for a slight approximation in the two-sided case.

The extension of the Kolmogorov test given here should be used when $H(x)$ is not continuous and the sample size is 30 or less, because the test is exact. If the sample size is greater than 30, the calculations become too difficult and therefore the chi-square test is recommended, unless $H(x)$ is continuous. If $H(x)$ is continuous, the Kolmogorov test or the Cramér-von Mises test (see [3]) may be used, unless there are unspecified parameters in $H(x)$ which must be estimated from the data, in which case the Shapiro-Wilk [10] test or the Lilliefors [6] test may be used with the normal distribution, or the chi-square test may be used if $H(x)$ is non-normal. If one is interested in a test with power against particular types of alternatives, rather than power for all alternatives, then tests designed for those alternatives should be used. For example, the Gelzer and Pyke [5] test is powerful against scalar alternatives.

The exact power (or close approximation thereto in the two-sided case) of the one-sided Kolmogorov test may be computed using the methods of this article, as illustrated in Example 3 of Section 2. The power of the Kolmogorov test for continuous null and alternative distributions has been studied by van der Waerden [13] and others. Epanechnikov [4] developed formulas for computing power in the continuous case, but the formulas are more difficult to use than the ones in this article.

Studies of the Kolmogorov test with discontinuous distributions appear to be quite limited. The Kolmogorov test is known to be conservative if $F(x)$ is discrete [7, 8, 11, 14]. The asymptotic distributions for D, D^+ and D^- have been obtained for some discontinuous cases by Schmid [9] and Carnal [2], and an approximate asymptotic distribution is given by Taha [12]. Unfortunately, all of these distribution functions become degenerate or undefined when $F(x)$ is purely discrete. If a nonzero derivative exists between all adjacent jump points of $F(x)$, then the formula of Taha [12] may be used, but the others, while now well defined, are intractable except for methods of numerical approximation to multiple integrals. Attempts to obtain simpler and more general asymptotic distributions from the exact distributions given here have been unsuccessful.

The well-known proof of the consistency of the Kolmogorov test does not assume continuity of the distribution functions, and therefore it applies to the forms of the Kolmogorov test given in this article.

In Section 2 the extension of the Kolmogorov test is described and several examples are given. Sections 3 and 4 present the derivation of the test and may be omitted by readers interested only in the application of the test.

2. A DESCRIPTION OF THE TEST

Let $X_1, X_2, \cdots, X_n$ represent a random sample of size n. The only assumption in this test is that the sample is random. Denote the null hypothesis by

$$H_0: F(x) = H(x) \qquad \text{for all } x, \tag{2.1}$$

where $F(x)$ is the unknown population distribution function, and $H(x)$ is the hypothesized distribution function with all parameters specified. $H(x)$ may be continuous, discrete, or a mixture of the two types.

The same test statistics are used here as are used in the Kolmogorov goodness-of-fit test. That is, let $S_n(x)$ represent the empirical distribution function,

* W.J. Conover is associate professor, Department of Statistics, Kansas State University, Manhattan, Kan. 66502. This article was written while the author was a visiting professor at the Seminar for Applied Statistics, University of Zurich. The research was sponsored by a Career Development Award, No. K4-GM-42,351, from the Institute of General Medical Sciences, NIH.

$$S_n(x) = \frac{1}{n} \text{ (the number of } X_i\text{'s which are } \leq x), \quad (2.2)$$

for all x. Then the two-sided test, consistent against all alternatives of the type $F(x) \neq H(x)$ for some x, uses the test statistic $D = \sup_x |H(x) - S_n(x)|$. The one-sided test, which is consistent against all alternatives of the type $F(x) < H(x)$ for some x, uses the statistic $D^- = \sup_x (H(x) - S_n(x))$, while the other one-sided test, consistent against alternatives of the type $F(x) > H(x)$ for some x, uses the statistic $D^+ = \sup_x (S_n(x) - H(x))$.

The test procedure is as follows. After the random sample is drawn, the test statistic D^+, D^- or D is computed, depending on which test is desired. Then the critical level (the probability of getting a value of the test statistic as extreme as the observed value when H_0 is true, or, equivalently, the smallest level of significance at which H_0 may be rejected for the observed sample) is computed as described next. If the critical level is "too small" (less than 0.05, say), then the null hypothesis is rejected. Otherwise it is accepted.

2.1 The Critical Level for D^-

Let d^- denote the observed value of the test statistic D^-.

Step 1. Compute the probabilities c_j for $0 \leq j < n(1-d^-)$ as follows. Draw a horizontal line with the ordinate $d^- + j/n$ directly on a graph of $H(x)$. Then c_j equals $1 - d^- - j/n$ unless the horizontal line intersects $H(x)$ at a jump of $H(x)$, i.e., where $H(x)$ jumps because x is a point of discontinuity. In that case c_j equals 1.0 minus the height (ordinate) of $H(x)$ at the top of the jump. One of the horizontal lines may intersect $H(x)$ exactly at the bottom of a jump, in which event c_j equals 1.0 minus the ordinate of that line.

Step 2. Compute the constants $b_0, b_1, \cdots$, from the recursive relationship $b_0 = 1$ and

$$b_k = 1 - \sum_{j=0}^{k-1} \binom{k}{j} c_j^{k-j} b_j, \qquad k \geq 1 \qquad (2.3)$$

for all k such that $c_k > 0$ in Step 1. Note that these constants are of the form

$$\begin{aligned} b_0 &= 1 \\ b_1 &= 1 - c_0 \\ b_2 &= 1 - c_0^2 - 2c_1 b_1 \\ b_3 &= 1 - c_0^3 - 3c_1^2 b_1 - 3c_2 b_2 \\ b_4 &= 1 - c_0^4 - 4c_1^3 b_1 - 6c_2^2 b_2 - 4c_3 b_3 \\ b_5 &= 1 - c_0^5 - 5c_1^4 b_1 - 10c_2^3 b_2 - 10c_3^2 b_3 - 5c_4 b_4 \\ &\text{etc.} \end{aligned} \qquad (2.4)$$

Step 3. Compute the critical level

$$P(D^- \geq d^-) = \sum_{j=0}^{[n(1-d^-)]} \binom{n}{j} c_j^{n-j} b_j \qquad (2.5)$$

from the c_j and b_j of Steps 1 and 2.

2.2 The Critical Level for D^+

Let d^+ denote the observed value of the test statistic D^+.

Step 1. Compute the probabilities f_j for $0 \leq j < n(1-d^+)$ by drawing a horizontal line with ordinate $1 - d^+ - j/n$ directly on the graph of $H(x)$. Then f_j equals $1 - d^+ - j/n$ unless the horizontal line intersects $H(x)$ at a jump, in which case f_j equals the height of $H(x)$ at the bottom of the jump. One of the horizontal lines may intersect $H(x)$ directly at the top of a jump; in this event f_j equals the ordinate of the horizontal line.

Step 2. Compute the constants $e_0, e_1, \cdots$, from the recursive relationship $e_0 = 1$ and

$$e_k = 1 - \sum_{j=0}^{k-1} \binom{k}{j} f_j^{k-j} e_j, \qquad k \geq 1 \qquad (2.6)$$

for all k, such that $f_k > 0$ in Step 1. These constants follow the same pattern as the b_k's in (2.4), with the c_i's replaced by f_i's.

Step 3. Compute the critical level

$$P(D^+ \geq d^+) = \sum_{j=0}^{[n(1-d^+)]} \binom{n}{j} f_j^{n-j} e_j. \qquad (2.7)$$

2.3 The Critical Level for D

Let d denote the observed value of the test statistic D. Compute $P(D^- \geq d)$ and $P(D^+ \geq d)$ from Equations (2.5) and (2.7). Then $P(D \geq d)$ is given approximately as

$$P(D \geq d) \doteq P(D^+ \geq d) + P(D^- \geq d). \qquad (2.8)$$

The approximation in (2.8) will be very close in most cases. Bounds for $P(D \geq d)$ are furnished by the inequalities

$$\begin{aligned} 0 &\leq P(D^+ \geq d) + P(D^- \geq d) - P(D \geq d) \\ &\leq P(D^+ \geq d) P(D^- \geq d). \end{aligned} \qquad (2.9)$$

That is, the error of the approximation in (2.8) is less than or equal to the product of the two critical levels obtained from (2.5) and (2.7).

Example 1. Let $H(x)$ be the discrete uniform distribution with equal probabilities $\frac{1}{5}$ at the five points $x = 1, 2, 3, 4, 5$. Suppose that a random sample of size 10, with the (ordered) values 1, 1, 1, 2, 2, 2, 3, 3, 3, 3 is drawn from some population, and the hypothesis to be tested is that $H(x)$ is the population distribution function. The greatest distance between $H(x)$ and $S_n(x)$ occurs at $x = 3$, so the test statistic for the two-sided Kolmogorov test becomes

$$D = \sup_x |H(x) - S_n(x)| = |H(3) - S_n(3)| = 0.4 = d.$$

To find the critical level associated with $d = 0.4$ the probability $P(D^- \geq 0.4)$ is computed.

Step 1. The probabilities $c_0, \cdots, c_5$ are determined in the manner described earlier in this section. For example, for $j = 1$ the line with ordinate $d^- + j/n = 0.4 + 1/10 = 0.5$ intersects the graph of $H(x)$ at the jump taken by $H(x)$ at $x = 3$. The height of $H(3)$ at the top of the jump is 0.6, so $c_1 = 1 - 0.6 = 0.4$. The five nonzero probabilities are $c_0 = 0.6$, $c_1 = 0.4 = c_2$, $c_3 = 0.2 = c_4$.

Step 2. The constants $b_0, \cdots, b_4$ are computed from the recursive relationships (2.4), and become $b_0 = 1$, $b_1 = 1 - c_0 = 0.4$, $b_2 = 1 - c_0^2 - 2c_1 b_1 = 0.32$, $b_3 = 0.208$, and $b_4 = 0.2944$. The constant b_5 is not computed because the term involving b_5 disappears in Equation (2.5), since $c_5 = 0$.

Step 3. The one-sided critical level $P(D^- \geq d)$ is computed from Equation (2.5):

$$\begin{aligned} P(D^- &\geq 0.4) \\ &= c_0^{10} + \binom{10}{1} c_1^9 b_1 + \binom{10}{2} c_2^8 b_2 + \binom{10}{3} c_3^7 b_3 + \binom{10}{4} c_4^6 b_4 \\ &= .02081. \end{aligned}$$

Because $H(x)$ is symmetric, computation of the one-sided critical level $P(D^+ \geq 0.4)$ is identical with the above so that $P(D^+ \geq 0.4) = 0.02081$ and the critical level for the two-sided Kolmogorov test is approximately $P(D \geq 0.4) \doteq (.02081)2 = .04162 \doteq .04$. The exact critical level is, from Equation (2.9), somewhere between .04119 and .04162, so the rounded value .04 may be used with complete confidence.

For comparison purposes it is interesting to note that the tables for the two-sided Kolmogorov test, valid if $H(x)$ is continuous and conservative otherwise, furnish an upper bound of 0.06 for the critical level in this example. Although this value is close enough for most purposes, usage of this value would result in acceptance of the null hypothesis at the .05 level, a level commonly used in practice, while the correct decision is to reject H_0 at $\alpha = .05$. A chi-square test on the same data, with five intervals of equal probability under H_0, results in a critical level of .15 with no assurance of the validity of this value. However, for this example the results of Slakter [11] are quite useful. The true level of significance was estimated, using Monte Carlo methods, for several discrete uniform distributions including the above. The chi-square test was shown to be slightly conservative for this case, and the Kolmogorov critical level was estimated as .0416, in agreement with these computations.

Example 2. A method of hypothesis testing, more traditional than the method used here, consists of first selecting a critical region of the desired size, and then rejecting the null hypothesis if the test statistic falls in the critical region. The procedure given above may be used for finding the exact size of the critical region, as is shown in this example.

Suppose $H(x)$ is the Poisson distribution with parameter 1.0. For a sample of size 10, existing tables for the one-sided Kolmogorov test based on the continuity assumption indicate that a critical region of size .10 consists of all values of D^- greater than .32260. If $d^- = .32260$ is used in Equation (2.5) the exact size of the critical region is revealed to be .0342 instead of .10. If the statistic D^+ is used instead of D^-, Equation (2.7) shows the exact size of the critical region to be .0343 instead of .10. The calculations are performed as in the previous example and are omitted here.

Example 3. This example illustrates the method for computing the exact power of the Kolmogorov test. Suppose the null hypothesis specifies a Poisson distribution with parameter 1.0, while the alternative hypothesis specifies $F(x)$ as being the binomial distribution with parameters $N = 10$ and $p = \frac{3}{8}$. The cumulative distribution functions are as follows:

x	H_0: $H(x)$	H_1: $F(x)$
0	.3679	.0091
1	.7358	.0637
2	.9197	.2110
(etc.)	(etc.)	(etc.)

If the sample size n is 5, the one-sided test with D^- as a test statistic might consist of rejecting H_0 if D^- exceeds .447. This critical value was chosen because the size of the critical region, for continuous $H(x)$, is given in the existing tables as exactly .10. The true size of the critical region is 0.022, as may be verified by following the method illustrated in Example 1. Note the large discrepancy between the true $\alpha = 0.022$ and its conservative approximation 0.10.

To find the power of the test compute the probabilities p_0, p_1 and p_2 as follows. Find the first $H(x)$ greater than or equal to $.447 + j/5$, say $H(x_j)$, and then let p_j equal $1 - F(x_j)$. This method gives $p_0 = 1 - F(1) = .9363 = p_1$, $p_2 = 1 - F(2) = .7890$, and $p_j = 0$ for $j > 2$. The constants b_k' are computed recursively as in (2.4), except the p_j's are used instead of the c_j's, with the results $b_0' = 0$, $b_1' = 1 - p_0 = .0637$, and $b_2' = 1 - p_0^2 - 2p_1 b_1' = .0041$. The power is obtained from an equation analogous to (2.5),

$$\begin{aligned} \text{power} &= P(D^- \geq .447 \mid H_1 \text{ is true}) \\ &= \textstyle\sum_{j=0}^{2} \binom{5}{j} p_j^{5-j} b_j' = .983. \end{aligned}$$

3. TWO IDENTITIES

The derivation of the desired distribution function remains less cluttered if two identities are first established. Let $\{f_i\}$, $i \geq 0$, be a sequence of real numbers. Define another sequence of real numbers $\{e_k\}$, $k \geq 0$, recursively as

$$e_k = (1 - f_k)^k - \textstyle\sum_{j=0}^{k-1} \binom{k}{j} (f_j - f_k)^{k-j} e_j. \qquad (3.1)$$

Note the convention that $\sum_{i=a}^{b} = 0$ if $b < a$, so that $e_0 = 1$.

First identity: Let $\{e_k\}$ be defined by (3.1). Then

$$e_k = 1 - \textstyle\sum_{j=0}^{k-1} \binom{k}{j} f_j^{k-j} e_j. \qquad (3.2)$$

Proof: The proof is by induction. For $k = 0$, e_0 equals 1 in both (3.1) and (3.2). Assume (3.2) holds for all k up to and including some n;

$$e_n = 1 - \sum_{j=0}^{n-1} \binom{n}{j} f_j^{n-j} e_j. \tag{3.3}$$

Then e_{n+1} may be written as

$$e_{n+1} = (1 - f_{n+1})^{n+1} - \sum_{j=0}^{n} \binom{n+1}{j} (f_j - f_{n+1})^{n+1-j} e_j. \tag{3.4}$$

The binomial terms in (3.4) are expanded, and (3.3) is used in place of e_j as part of the coefficient of f_{n+1}^{n+1-j}, as follows.

$$\begin{aligned} e_{n+1} = & \sum_{i=0}^{n+1} \binom{n+1}{i} (-f_{n+1})^i \\ & - \sum_{j=0}^{n} \binom{n+1}{j} \sum_{i=0}^{n+1-j} \binom{n+1-j}{i} f_j^{n+1-j-i} (-f_{n+1})^i e_j \\ = & \sum_{i=0}^{n+1} \binom{n+1}{i} (-f_{n+1})^i \\ & - \sum_{j=0}^{n} \binom{n+1}{j} \sum_{i=0}^{n-j} \binom{n+1-j}{j} f_j^{n+1-j-i} (-f_{n+1})^i e_j \\ & - \sum_{j=0}^{n} \binom{n+1}{j} (-f_{n+1})^{n+1-j} \\ & \cdot \left\{1 - \sum_{k=0}^{j-1} \binom{j}{k} f_k^{j-k} e_k \right\}. \end{aligned} \tag{3.5}$$

Straightforward algebra shows that the coefficients of $(-f_{n+1})^i$ sum to zero for each i greater than zero, and that the remaining terms are the coefficients of $(-f_{n+1})^0$, namely

$$e_{n+1} = 1 - \sum_{j=0}^{n} \binom{n+1}{j} f_j^{n+1-j} e_j, \tag{3.6}$$

which completes the proof.

Second identity: Let $\{e_k\}$ be defined by (3.1). Then for each positive integer m, and for $i = 0, 1, \cdots, m-1$,

$$(1 - f_m)^i - \sum_{j=0}^{i} \binom{i}{j} (f_j - f_m)^{i-j} e_j = 0. \tag{3.7}$$

Proof: First the binomial terms in (3.7) are expanded, and the coefficients of f_m are collected for each k.

$$\begin{aligned} & \sum_{k=0}^{i} \binom{i}{k} (-f_m)^k - \sum_{j=0}^{i} \binom{i}{j} e_j \sum_{k=0}^{i-j} \binom{i-j}{k}^{i-j-k} (-f_m)^k \\ & = \sum_{k=0}^{i} (-f_m)^k \left\{ \binom{i}{k} - \sum_{j=0}^{i-k} \binom{i}{j} e_j \binom{i-j}{k} f_j^{i-j-k} \right\} \\ & = \sum_{k=0}^{i} (-f_m)^k \binom{i}{k} \left\{ 1 - \sum_{j=0}^{i-k-1} \binom{i-k}{j} f^{i-k-j} e_j - e_{i-k} \right\}. \end{aligned} \tag{3.8}$$

Equation (3.2) is then used within the brackets to show that the above expression equals zero.

4. THE DERIVATION OF $P(D^+ \leq t)$

To find the distribution function $P(D^+ \leq t)$ of D^+ the probability $P(D^+ > t)$ is found in a straightforward manner;

$$\begin{aligned} P(D^+ > t) & = P(\sup_x (S_n(x) - H(x)) > t) \\ & = P(S_n(x) - H(x) > t \text{ for some } x) \\ & = P(S_n(X^{(i)}) - H(X^{(i)}) > t \text{ for some } X^{(i)}) \end{aligned} \tag{4.1}$$

because the greatest difference $S_n - H$ will occur at one of the jump points of S_n, which are the order statistics $X^{(1)} \leq X^{(2)} \leq \cdots \leq X^{(n)}$. Thus, (4.1) becomes

$$\begin{aligned} P(D^+ > t) & = P\left(\frac{i}{n} - H(X^{(i)}) > t \text{ for some } i\right) \\ & = P\left(\frac{i}{n} - t > H(X^{(i)}) \text{ for some } i\right) \\ & = P\left(H^{-1}\left(\frac{i}{n} - t\right) > X^{(i)} \text{ for some } i\right) \end{aligned} \tag{4.2}$$

where $H^{-1}(p)$ is defined as the smallest x such that $H(x) \geq p$;

$$\begin{aligned} H^{-1}(p) & = \inf \{x \mid H(x) \geq p\}, & 0 < p \leq 1 \\ & = -\infty, & p \leq 0. \end{aligned} \tag{4.3}$$

The event in (4.2) may be separated into disjoint sets as follows.

$$\begin{aligned} P(D^+ > t) = & P\left(X^{(n)} < H^{-1}\left(\frac{n}{n} - t\right)\right) \\ & + P\left(X^{(n-1)} < H^{-1}\left(\frac{n-1}{n} - t\right), X^{(n)} \geq H^{-1}\left(\frac{n}{n} - t\right)\right) \\ & + \cdots \\ & + P\left(X^{(1)} < H^{-1}\left(\frac{1}{n} - t\right), X^{(2)} \right. \\ & \left. \geq H^{-1}\left(\frac{2}{n} - t\right), \cdots, X^{(n)} \geq H^{-1}\left(\frac{n}{n} - t\right)\right) \\ = & \sum_{j=0}^{n-1} P\left(X^{(n-j)} < H^{-1}\left(\frac{n-j}{n} - t\right), X^{(n-j+1)} \right. \\ & \left. \geq H^{-1}\left(\frac{n-j+1}{n} - t\right), \cdots, X^{(n)} \geq H^{-1}\left(\frac{n}{n} - t\right)\right). \end{aligned} \tag{4.4}$$

A similar argument leads to the same equation for $P(D^+ \geq t)$,

$$\begin{aligned} & P(D^+ \geq t) \\ & = \sum_{j=0}^{n-1} P\left(X^{(n-j)} < H_*^{-1}\left(\frac{n-j}{n} - t\right), X^{(n-j+1)} \right. \\ & \geq H_*^{-1}\left(\frac{n-j+1}{n} - t\right), \cdots, X^{(n)} \\ & \left. \geq H_*^{-1}\left(\frac{n}{n} - t\right)\right), \end{aligned} \tag{4.5}$$

where H_*^{-1} is defined by

$$H_*^{-1}(p) = \sup\{x \mid H(x) \le p\}, \qquad 0 \le p \le 1 \\ = -\infty, \qquad p < 0. \tag{4.6}$$

The probabilities appearing in (4.4) may take the following alternative form,

$$P(D^+ > t) = \sum_{j=0}^{n-1} P\Bigg((n-j)X_i\text{'s} < H^{-1}\left(\frac{n-j}{n} - t\right),\ (j)X_i\text{'s} \ge H^{-1}\left(\frac{n-j+1}{n} - t\right),$$
$$\text{at least } (j-1)X_i\text{'s} \ge H^{-1}\left(\frac{n-j+2}{n} - t\right), \tag{4.7}$$
$$\text{at least } (j-2)X_i\text{'s} \ge H^{-1}\left(\frac{n-j+3}{n} - t\right), \cdots,$$
$$\text{at least } (1)\ X_i \ge H^{-1}\left(\frac{n}{n} - t\right)\Bigg)$$

or as

$$P(D^+ > t) = \sum_{j=0}^{n-1} P\Bigg((n-j)X_i\text{'s} < H^{-1}\left(\frac{n-j}{n} - t\right),\ (j)X_i\text{'s} \ge H^{-1}\left(\frac{n-j+1}{n} - t\right)\Bigg)$$
$$\cdot \sum_{i_{j-1}=j-1}^{j} P\Bigg((i_{j-1})X_i\text{'s} \ge H^{-1}\left(\frac{n-j+2}{n} - t\right) \Bigg|\ (j)X_i\text{'s} \ge H^{-1}\left(\frac{n-j+1}{n} - t\right)\Bigg)$$
$$\cdot \sum_{i_{j-2}=j-2}^{i_{j-1}} P\Bigg((i_{j-2})X_i\text{'s} \ge H^{-1}\left(\frac{n-j+3}{n} - t\right) \Bigg|\ (i_{j-1})X_i\text{'s} \ge H^{-1}\left(\frac{n-j+2}{n} - t\right)\Bigg)$$
$$\cdots$$
$$\cdot \sum_{i_1=1}^{i_2} P\Bigg((i_1)X_i\text{'s} \ge H^{-1}\left(\frac{n}{n} - t\right) \Bigg|\ (i_2)X_i\text{'s} \ge H^{-1}\left(\frac{n-1}{n} - t\right)\Bigg). \tag{4.8}$$

Each individual probability in (4.8) is easily expressed using the binomial probability function. Let f_k for $k = 0, 1, \cdots, n$ be defined as

$$f_k = P\left(X_i < H^{-1}\left(\frac{n-k}{n} - t\right)\right), \quad 0 \le k \le n. \tag{4.9}$$

Note that $f_k = 0$ if $k \ge n(1-t)$. The individual probabilities in (4.8) become

$$P\Bigg((i_m)X_i\text{'s} \ge H^{-1}\left(\frac{n-m+1}{n} - t\right) \Bigg|\ (i_{m+1})X_i\text{'s} \ge H^{-1}\left(\frac{n-m}{n} - t\right)\Bigg) = \binom{i_{m+1}}{i_m}\left(\frac{1-f_{m-1}}{1-f_m}\right)^{i_m} \cdot \left(\frac{f_{m-1}-f_m}{1-f_m}\right)^{i_{m+1}-i_m}. \tag{4.10}$$

Then Equation (4.8) becomes, after slight simplification, and with insertion of the trivial summation over i_j for later ease,

$$P(D^+ > t) = \sum_{j=0}^{n-1} \binom{n}{j} f_j^{n-j} \sum_{i_j=j}^{j} \binom{j}{i_j} \cdot (f_{j-1} - f_j)^{j-i_j} \sum_{i_{j-1}=j-1}^{i_j} \binom{i_j}{i_{j-1}} (f_{j-2} - f_{j-1})^{i_j - i_{j-1}}$$
$$\cdot \sum_{i_{j-2}=j-2}^{i_{j-1}} \binom{i_{j-1}}{i_{j-2}} (f_{j-3} - f_{j-2})^{i_{j-1}-i_{j-2}} \cdot \cdots \cdot \sum_{i_1=1}^{i_2} \cdot \binom{i_2}{i_1} (f_0 - f_1)^{i_2 - i_1} (1 - f_0)^{i_1}. \tag{4.11}$$

Note that the term for $j = 0$ is, by convention, f_0^n. Thus Equation (4.11), awkward though it may be, enables the distribution function of D^+ to be given in closed form for all forms of $F(x)$ and $H(x)$.

The following development is aimed at simplification of (4.11). Sum (4.11) first over i_1, then over i_2, and so on. The sum over the index i_m, as will be proved by induction, is given by

$$(1 - f_m)^{i_{m+1}} - \sum_{k=0}^{m-1} \binom{i_{m+1}}{k} (f_k - f_m)^{i_{m+1}-k} e_k \tag{4.12}$$

where e_k is the same as in (3.1) and f_i is defined in (4.9). First, for $m = 1$, the summation yields

$$\sum_{i_1=1}^{i_2} \binom{i_2}{i_1} (f_0 - f_1)^{i_2 - i_1} (1 - f_0)^{i_1} = (1 - f_1)^{i_2} - (f_0 - f_1)^{i_2}, \tag{4.13}$$

which agrees with (4.12) because $e_0 = 1$. Next assume that the successive summations over $i_1, i_2, \cdots, i_{m-1}$ result in

$$(1 - f_{m-1})^{i_m} - \sum_{k=0}^{m-2} \binom{i_m}{k} (f_k - f_{m-1})^{i_m - k} e_k. \tag{4.14}$$

The summation over the index i_m is then given by

$$\sum_{i_m=m}^{i_{m+1}} \binom{i_{m+1}}{i_m} (f_{m-1} - f_m)^{i_{m+1}-i_m} \left\{(1 - f_{m-1})^{i_m} - \sum_{k=0}^{m-2} \binom{i_m}{k} \cdot (f_k - f_{m-1})^{i_m - k} e_k\right\} = \sum_{i_m=m}^{i_{m+1}} \binom{i_{m+1}}{i_m} (f_{m-1} - f_m)^{i_{m+1}-i_m}$$
$$\cdot (1 - f_{m-1})^{i_m} - \sum_{k=0}^{m-2} \sum_{i_m=m}^{i_{m+1}} \binom{i_{m+1}}{i_m} \binom{i_m}{k} \cdot (f_{m-1} - f_m)^{i_{m+1}-i_m} (f_k - f_{m-1})^{i_m - k} e_k. \tag{4.15}$$

The identity

$$\sum_{i=0}^{n} \binom{n}{i} \binom{i}{k} y^{n-i} x^{i-k} = \binom{n}{k} (y + x)^{n-k}, \tag{4.16}$$

which is easily obtained by taking successive derivatives, with respect to x, of both sides of

$$\sum_{i=0}^{n} \binom{n}{i} y^{n-i}x^{i} = (y+x)^{n}, \tag{4.17}$$

may be used to simplify (4.15). Continuation of (4.15) gives

$$\begin{aligned}
&= (1-f_m)^{i_{m+1}} - \sum_{i_m=0}^{m-1} \binom{i_{m+1}}{i_m} (f_{m-1}-f_m)^{i_{m+1}-i_m} \\
&\quad \cdot (1-f_{m-1})^{i_m} - \sum_{k=0}^{m-2} \binom{i_{m+1}}{k} (f_k - f_m)^{i_{m+1}-k} e_k \\
&\quad + \sum_{i_m=0}^{m-1} \sum_{k=0}^{m-2} \binom{i_{m+1}}{i_m} \binom{i_m}{k} \\
&\quad \cdot (f_{m-1}-f_m)^{i_{m+1}-i_m} (f_k - f_{m-1})^{i_m-k} e_k \\
&= (1-f_m)^{i_{m+1}} - \sum_{k=0}^{m-2} \binom{i_{m+1}}{k} (f_k - f_m)^{i_{m+1}-k} e_k \\
&\quad - \sum_{i_m=0}^{m-1} \binom{i_{m+1}}{i_m} (f_{m-1}-f_m)^{i_{m+1}-i_m} \\
&\quad \cdot \left\{ (1-f_{m-1})^{i_m} - \sum_{k=0}^{m-2} \binom{i_m}{k} (f_k - f_{m-1})^{i_m-k} e_k \right\}. \qquad (4.18)
\end{aligned}$$

The expression within the final set of brackets equals zero for $i_m = 0, 1, \cdots, m-2$, by (3.7), and equals e_{m-1} for $i_m = m-1$ by (3.1). Thus (4.18) simplifies to (4.12) and the proof by induction is complete. The expression (4.12) may be used to express the final summation of (4.11), the summation over the index i_j, so that (4.11) becomes

$$\begin{aligned}
P(D^+ > t) = \sum_{j=0}^{n-1} \binom{n}{j} f_j^{n-j} \Big\{ &(1-f_j)^j \\
&- \sum_{k=0}^{j-1} \binom{j}{k} (f_k - f_j)^{j-k} e_k \Big\}. \qquad (4.19)
\end{aligned}$$

The expression within the final brackets in (4.19) is simplified using (3.1), and then (3.2) is used on the entire equation, so that

$$P(D^+ > t) = \sum_{j=0}^{n-1} \binom{n}{j} f_j^{n-j} e_j = 1 - e_n \tag{4.20}$$

and the distribution function of D^+ is

$$P(D^+ \leq t) = e_n. \tag{4.21}$$

A similar development from (4.5) gives

$$P(D^+ \geq t) = 1 - e_n \tag{4.22}$$

where H^{-1} is defined by (4.6) rather than (4.3), and therefore the sequences $\{f_k\}$ and $\{e_k\}$ are possibly different than before.

The distribution function of D^- is obtained similarly. The slight difference in the derivations justifies omission of the details. The derivation of (2.9), which provides bounds on $P(D \geq d)$, is partly intuitive, and has therefore been omitted in this final draft in the interest of saving space. It is easy to show that in the continuous case $f_k = 1 - k/n - t$, and therefore Equation (4.20) reduces to the well-known form first obtained by Birnbaum and Tingey [1].

[*Received December 1970. Revised October 1971.*]

REFERENCES

[1] Birnbaum, Z.W. and Tingey, F.H., "One-Sided Confidence Contours for Probability Distribution Functions," *The Annals of Mathematical Statistics*, 22 (December 1951), 592–6.

[2] Carnal, H., "Sur les Theoremes de Kolmogorov et Smirnov dans le Cas de'Une Distribution Discontinue," *Commentarii Mathematici Helvetici*, 37, No. 1 (1962), 19–35.

[3] Conover, W.J., *Practical Nonparametric Statistics*, New York: John Wiley and Sons, Inc., 1971.

[4] Epanechnikov, V.A., "The Significance Level and Power of the Two Sided Kolmogorov Test in the Case of Small Sample Sizes," *Theory of Probability and Its Applications*, 13, No. 4 (1968), 686–90.

[5] Gelzer, J. and Pyke, R., "The Asymptotic Relative Efficiency of Goodness-of-Fit Tests Against Scalar Alternatives," *Journal of the American Statistical Association*, 60 (June 1965), 410–19.

[6] Lilliefors, H.W., "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, 62 (June 1967), 399–402.

[7] Noether, G.E., "Note on the Kolmogorov Statistic in the Discrete Case," *Metrika*, 7, No. 2 (1963), 115–16.

[8] Noether, G.E., *Elements of Nonparametric Statistics*, New York: John Wiley and Sons, Inc., 1967.

[9] Schmid, P., "On the Kolmogorov and Smirnov Limit Theorems for Discontinuous Distribution Functions," *The Annals of Mathematical Statistics*, 29 (December 1958), 1011–27.

[10] Shapiro, S.S. and Wilk, M.B., "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 52, Parts 3 and 4 (1965), 591–611.

[11] Slakter, M.J., "A Comparison of the Pearson Chi-Square and Kolmogorov Goodness-of-Fit Tests with Respect to Validity," *Journal of the American Statistical Association*, 60 (September 1965), 854–8.

[12] Taha, M.A.H., "Ueber die verallgemeinerten Teste von Kolmogorov und Smirnov für unstetige Verteilungen," *Mitt. Verein. Schweiz. Versicherungsmathematik.*, 64, No. 3 (1964), 145–74.

[13] Van der Waerden, B.L., "Testing a Distribution Function," *Proceedings Kominklijke Nederlandse Akademic van Wetenschappen (A)*, 56, No. 2 (Indagationes Mathematicae 15) (1953), 201–7.

[14] Walsh, J.E., "Bounded Probability Properties of Kolmogorov-Smirnov and Similar Statistics for Discrete Data," *Annals of the Institute of Statistical Mathematics*, 15, No. 2 (1963), 153–8.