SEVERAL k-SAMPLE KOLMOGOROV-SMIRNOV TESTS[1]

By W. J. Conover

Kansas State University

1. Introduction. Let the nk random variables $\{X_{ij}\}$, $i = 1, \ldots, n$, $j = 1, \cdots, k$, represent k random samples of equal size n with the common absolutely continuous distribution function $F(x)$. Order the samples within themselves, and denote the rth ordered sample by $Z_{1r} < Z_{2r} < \cdots < Z_{nr}$. Then Z_{ir} is the ith order statistic from the rth sample, and Z_{1r} will be referred to as the extreme of the rth sample. Define the empirical distribution function of the rth sample $F_r(z)$ as

$$
\begin{aligned}
F_r(z) &= 0 && \text{if } z < Z_{1r}, \\
F_r(z) &= m/n && \text{if } Z_{mr} \leqq z < Z_{m+1,r}, \\
F_r(z) &= 1 && \text{if } Z_{nr} \leqq z.
\end{aligned} \tag{1.1}
$$

Now order the samples among themselves on the basis of their extremes. That is, let $S = \{Z_{1r}, r = 1, \cdots, k\}$ be the set of extremes from the k samples, and let $Y_{11} < Y_{12} < \cdots < Y_{1k}$ represent the set S after S is ordered. Further, let Y_{ij} represent the ith order statistic from the sample whose extreme is Y_{1j}. In other words, for each point in the sample space where Z_{1r} corresponds to Y_{1j}, the sample $(Z_{1r}, Z_{2r}, \cdots, Z_{nr})$ will be denoted by $(Y_{1j}, Y_{2j}, \cdots, Y_{nj})$. Since $Z_{1r} < Z_{2r} < \cdots < Z_{nr}$, it follows that $Y_{1j} < Y_{2j} < \cdots < Y_{nj}$. The number i is called the rank of Y_{ij} within the sample, and the number j is called the rank of the sample. Define the empirical distribution function of the sample with rank j as

$$
\begin{aligned}
S_j(y) &= 0 && \text{if } y < Y_{1j}, \\
S_j(y) &= m/n && \text{if } Y_{mj} \leqq y < Y_{m+1,j}, \\
S_j(y) &= 1 && \text{if } Y_{nj} \leqq y.
\end{aligned} \tag{1.2}
$$

When $k = 2$, the Kolmogorov-Smirnov (two-sample) test usually involves the use of the test statistics

$$D^+(n, n) = \sup_x [F_1(x) - F_2(x)] \tag{1.3}$$

and

$$D(n, n) = \sup_x |F_1(x) - F_2(x)| \tag{1.4}$$

Massey (1951) tabulated the distribution of (1.4), and Birnbaum and Hall

Received 7 September 1964.

[1] Prepared with support from the Kansas Agricultural Experiment Station, Manhattan, as Contribution No. 88, Department of Statistics and Statistical Laboratory.

(1960) give the distribution of both (1.3) and (1.4) for n ranging from 1 to 40. Darling (1957) presents an exposition of the extensive research involving various aspects of the Kolmogorov-Smirnov test.

In this paper, (5.1) and (5.2) give the distribution functions of the test statistics

$$D^+_{1,2}(2, n) = \sup_y [S_1(y) - S_2(y)] \tag{1.5}$$

and

$$D^-_{1,2}(2, n) = \sup_y [S_2(y) - S_1(y)], \tag{1.6}$$

respectively.

For k samples, $k > 2$, the obvious extension of the Kolmogorov-Smirnov test involves the use of the test statistics

$$D^+(n, n, \cdots, n) = \sup_{x,I_1} [F_i(x) - F_j(x)], \tag{1.7}$$

where $I_1 = (i, j: i < j; j = 2, 3, \cdots, k)$ and

$$D(n, n, \cdots, n) = \sup_{x,I_2} |F_i(x) - F_j(x)|, \tag{1.8}$$

where $I_2 = (i, j: i, j = 1, 2, \cdots, k)$. The mathematical expression for the distributions of (1.7) and (1.8) has proved elusive. Birnbaum and Hall (1960) have devised an iterative scheme using simple difference equations, suitable for machine computation, which makes it possible to find the exact distribution of (1.7) and (1.8) in the general case where the sample sizes are not necessarily equal.

A clever geometric approach by David (1958) resulted in the exact distribution of a 3-sample test statistic

$$D' = \sup_{x,I_3} [(-1)^{j-i}(F_i(x) - F_j(x))], \tag{1.9}$$

where $I_3 = (i, j: i < j, j = 2, 3)$.

That the distribution of (1.9) is not simple, indicates that the distributions of (1.7) and (1.8) may be unwieldy to use in their exact mathematical form; therefore, the iterative scheme mentioned above may be the most practical method to obtain the desired probabilities.

Kiefer (1959) discusses the above k-sample random variables and other k-sample analogues of the Kolmogorov-Smirnov test. Other k-sample tests of the same nature are introduced by Dwass (1960).

This paper proposes to reduce the k-sample problem to that of 2 samples. The distributions of the random variables

$$D^+_{j_1j_2}(k, n) = \sup_y [S_{j_1}(y) - S_{j_2}(y)] \tag{1.10}$$

and

$$D^-_{j_1j_2}(k, n) = \sup_y [S_{j_2}(y) - S_{j_1}(y)], \tag{1.11}$$

where $j_1 < j_2$, are given by (3.7) and (4.5), and are found to be of a form simple enough for practical use.

To test the null hypothesis that the k samples were drawn from populations having identical distributions against the alternative hypothesis that at least one of the populations differs by location parameter, $D^{+}_{1,k}(k, n)$ could be used. To divide the k samples into groups having similar location parameters, $D^{+}_{j,j+1}(k, n)$ could be used for all values of j from 1 to $k - 1$.

If the alternative hypothesis is that the populations differ by a scale parameter only, $D^{-}_{1,k}(k, n)$ could be used as the basis for a nonparametric analogue to the maximum F test given by Hartley (1950).

Following some preliminaries presented in the next section, Section 3 gives the distribution function of $D^{+}_{j_1 j_2}(k, n)$. The distribution function of $D^{-}_{j_1 j_2}(k, n)$ is derived in Section 4. A discussion of the special case where $k = 2$ is given in Section 5. The final section gives the limiting distribution as $n \to \infty$, which is found to be the same for both $D^{+}_{j_1 j_2}(k, n)$ and $D^{-}_{j_1 j_2}(k, n)$, and independent of j_1, j_2, and k.

2. Preliminaries. The distribution function of Z_{1r} is well known to be $1 - [1 - F(x)]^n$. Let I_{ij} represent the infinitesimal interval $(t_{ij}, t_{ij} + dt_{ij})$. From (8.7.5) of Wilks (1962) the joint probability element of Y_{1j_1} and Y_{1j_2} is

$$P(Y_{1j_1} \varepsilon I_{1r}, Y_{1j_2} \varepsilon I_{1s}) = [k!/(j_1 - 1)!(j_2 - j_1 - 1)!(k - j_2)!]$$
$$(2.1) \qquad \cdot [1 - (1 - F(t_{1r}))^n]^{j_1 - 1}[(1 - F(t_{1r}))^n - (1 - F(t_{1s}))^n]^{j_2 - j_1 - 1}$$
$$\cdot [1 - F(t_{1s})]^{n(k - j_2)} \, d[1 - (1 - F(t_{1r}))^n] \, d[1 - (1 - F(t_{1s}))^n]$$

in the region where $t_{1r} < t_{1s}$, zero elsewhere.

Because the samples are mutually independent before they are ranked, the joint probability element of Y_{1r} and Y_{1s}, where r and s are unspecified, is

$$(2.2) \quad P(Y_{1r} \varepsilon I_{1r}, Y_{1s} \varepsilon I_{1s}) = P(Y_{1r} \varepsilon I_{1r}) P(Y_{1s} \varepsilon I_{1s})$$
$$= d[1 - (1 - F(t_{1r}))^n] \, d[1 - (1 - F(t_{1s}))^n].$$

The rank of a sample depends only on the extreme within that sample. That is, for unspecified values of r and s,

$$(2.3) \quad P(r = j_1, s = j_2 \mid Y_{1r} \varepsilon I_{1r}, Y_{2r} \varepsilon I_{2r}, \cdots, Y_{nr} \varepsilon I_{nr}, Y_{1s} \varepsilon I_{1s},$$
$$Y_{2s} \varepsilon I_{2s}, \cdots, Y_{ns} \varepsilon I_{ns}) = P(r = j_1, s = j_2 \mid Y_{1r} \varepsilon I_{1r}, Y_{1s} \varepsilon I_{1s}).$$

But since

$$(2.4) \qquad P(A \mid B) = P(AB)/P(B)$$

equation (2.3) can be written as

$$(2.5) \quad \frac{P(r = j_1, s = j_2, Y_{1r} \varepsilon I_{1r}, \cdots, Y_{nr} \varepsilon I_{nr}, Y_{1s} \varepsilon I_{1s} \cdots, Y_{ns} \varepsilon I_{ns})}{P(Y_{1r} \varepsilon I_{1r}, \cdots, Y_{nr} \varepsilon I_{nr}, Y_{1s} \varepsilon I_{1s}, \cdots, Y_{ns} \varepsilon I_{ns})}$$
$$= P(r = j_1, s = j_2, Y_{1r} \varepsilon I_{1r}, Y_{1s} \varepsilon I_{1s})/P(Y_{1r} \varepsilon I_{1r}, Y_{1s} \varepsilon I_{1s}),$$

which is equivalent to

$$(2.6)\quad \frac{P(Y_{1j_1}\,\varepsilon\, I_{1r}, Y_{2j_1}\,\varepsilon\, I_{2r}, \cdots, Y_{nj_1}\,\varepsilon\, I_{nr}, Y_{1j_2}\,\varepsilon\, I_{1s}, \cdots, Y_{nj_2}\,\varepsilon\, I_{ns})}{P(Y_{1r}\,\varepsilon\, I_{1r}, Y_{2r}\,\varepsilon\, I_{2r}, \cdots, Y_{nr}\,\varepsilon\, I_{nr}, Y_{1s}\,\varepsilon\, I_{1s}, \cdots, Y_{ns}\,\varepsilon\, I_{ns})} = P(Y_{1j_1}\,\varepsilon\, I_{1r}, Y_{1j_2}\,\varepsilon\, I_{1s})/P(Y_{1r}\,\varepsilon\, I_{1r}, Y_{1s}\,\varepsilon\, I_{1s}).$$

Because r and s represent unspecified values, the samples consisting of $\{Y_{ir}\}$ and $\{Y_{is}\}$, $i = 1, 2, \cdots, n$ can be regarded as independent. Therefore, using (8.7.2) of Wilks (1962),

$$(2.7)\quad \begin{aligned} &P(Y_{1r}\,\varepsilon\, I_{1r}, \cdots, Y_{nr}\,\varepsilon\, I_{nr}, Y_{1s}\,\varepsilon\, I_{1s}, \cdots, Y_{ns}\,\varepsilon\, I_{ns}) \\ &\quad = P(Y_{1r}\,\varepsilon\, I_{1r}, \cdots, Y_{nr}\,\varepsilon\, I_{nr})P(Y_{1s}\,\varepsilon\, I_{1s}, \cdots, Y_{ns}\,\varepsilon\, I_{ns}) \\ &\quad = n\,!\,dF(t_{1r})\,dF(t_{2r}) \cdots dF(t_{nr})n!\,dF(t_{1s}) \cdots dF(t_{ns}). \end{aligned}$$

Substituting (2.1), (2.2), and (2.7) into (2.6) gives

$$(2.8)\quad \begin{aligned} &P(Y_{1j_1}\,\varepsilon\, I_{1r}, \cdots, Y_{nj_1}\,\varepsilon\, I_{nr}, Y_{1j_2}\,\varepsilon\, I_{1s}, \cdots, Y_{nj_2}\,\varepsilon\, I_{ns}) \\ &\quad = [k!\,n!\,n!/(j_1 - 1)!\,(j_2 - j_1 - 1)!\,(k - j_2)!][1 - (1 - F(t_{1r}))^n]^{j_1-1} \\ &\qquad \cdot[(1 - F(t_{1r}))^n - (1 - F(t_{1s}))^n]^{j_2-j_1-1}[1 - F(t_{1s})]^{n(k-j_2)} \\ &\qquad \cdot dF(t_{1r}) \cdots dF(t_{nr})\,dF(t_{1s}) \cdots dF(t_{ns}) \\ &\quad = G\,dF(t_{1r}) \cdots dF(t_{ns}), \text{ for brevity.} \end{aligned}$$

It should be understood that (2.8) equals zero when the inequalities,

$$(2.9)\quad t_{1r} < t_{1s}\,;\; t_{1r} < t_{2r} < \cdots < t_{nr}\,;\; t_{1s} < t_{2s} < \cdots < t_{ns}\,,$$

do not hold, because of the corresponding restrictions inherent in (2.7) and (2.1).

3. The distribution function of $D^+_{j_1j_2}(k, n)$. Since the samples are of equal size n, possible values of $D^+_{j_1j_2}(k, n)$ are the values of c/n for $c = 1, 2, \cdots, n$. Nonpositive values of c are impossible because of the definition (1.10) of $D^+_{j_1j_2}(k, n)$. Therefore

$$(3.1)\quad P(D^+_{j_1j_2}(k, n) \leqq 0) = 0 \quad\text{and}\quad P(D^+_{j_1j_2}(k, n) \leqq 1) = 1.$$

For the nontrivial case

$$(3.2)\quad P(D^+_{j_1j_2}(k, n) \leqq c/n) = P(Y_{1j_2} < Y_{c+1,j_1}, Y_{2j_2} < Y_{c+2,j_1}, \cdots, Y_{n-c,j_2} < Y_{nj_1}).$$

To obtain the solution to (3.2), (2.8) is integrated over the portion of the sample space where the following inequalities hold:

$$(3.3)\quad t_{1s} < t_{c+1,r}\,;\; t_{2s} < t_{c+2,r}\,;\; \cdots\,;\; t_{n-c,s} < t_{nr}\,.$$

The problem now becomes one of choosing the correct limits of integration for the integrand given by (2.8), such that inequalities (2.9) and (3.3) hold simultaneously. This is accomplished by the integral

$$P(D^+_{j_1j_2}(k, n) \leqq c/n) = \int_{-\infty}^{\infty} \int_{t_{1r}}^{\infty} \cdots \int_{t_{c-1,r}}^{\infty} \int_{t_{cr}}^{\infty} \int_{t_{1r}}^{t_{c+1,r}} \int_{t_{c+1,r}}^{\infty} \int_{t_{1s}}^{t_{c+2,r}}$$

$$[3.4] \qquad \cdots \int_{t_{n-2,r}}^{\infty} \int_{t_{n-c-2,s}}^{t_{n-1,r}} \int_{t_{n-1,r}}^{\infty} \int_{t_{n-c-1,s}}^{t_{n,r}} \int_{t_{n-c,s}}^{\infty} \cdots \int_{t_{n-2,s}}^{\infty}$$
$$\cdot \int_{t_{n-1,s}}^{\infty} G\, dF(t_{n,s})\, dF(t_{n-1,s}) \cdots dF(t_{n-c+1,s})\, dF(t_{n-c,s})\, dF(t_{n,r})$$
$$\cdot dF(t_{n-c-1,s})\, dF(t_{n-1,r}) \cdots dF(t_{2s})\, dF(t_{c+2,r})\, dF(t_{1s})\, dF(t_{c+1,r})$$
$$\cdot dF(t_{c,r}) \cdots dF(t_{2r})\, dF(t_{1r}).$$

The integration is tedious but straightforward, resulting in

$$P\left(D^+_{j_1j_2}(k, n) \leqq \frac{c}{n}\right)$$
$$= \sum_{\beta=0}^{j_1-1} \sum_{\alpha=0}^{j_2-j_1-1} \frac{k!\,n!\,n!(-1)^{j_2-\beta-\alpha}}{(k-j_2)!\beta!(j_1-1-\beta)!\alpha!(j_2-j_1-1-\alpha)!}$$
$$\cdot \{[(n-1)!(n-1)!(nk-nj_1-n\alpha)(nk-n\beta)]^{-1}$$
$$-[(n-1)!(n-c-1)!(nk-nj_1-n\alpha)(nk-n\beta)(nk-nj_1$$
$$(3.5) \qquad -n\alpha+n-1)_c]^{-1}$$
$$-[(n+c)!(n-c-2)!(nk-nj_1-c-1-n\alpha)(nk-n\beta)]^{-1}$$
$$+[n!(n-c-2)!(nk-nj_1-c-1-n\alpha)(nk-n\beta)$$
$$\cdot(nk-nj_1-n\alpha+n-1)_c]^{-1}\},$$

where $(A)_c$ represents the falling factorial $A(A-1) \cdots (A-c+1)$. The identity, from Conover (1964),

$$(3.6) \qquad \textstyle\sum_{\alpha=0}^{j-1} (-1)^{j-1-\alpha}/\alpha!\,(j-1-\alpha)!\,(A-\alpha) = [(A)_j]^{-1}$$

leads to simplification of (3.5) into

$$P\left(D^+_{j_1j_2}(k, n) \leqq \frac{c}{n}\right)$$
$$(3.7) \qquad = 1 - \binom{k-j_1}{j_2-j_1}\binom{2n-2}{n+c} \Big/ \binom{k-j_1-(c+1)/n}{j_2-j_1}\binom{2n-2}{n-1}$$
$$+ \sum_{\alpha=0}^{j_2-j_1-1} \frac{(n-1)_c(c+1)(k-j_1)_{j_2-j_1}(-1)^{j_2-j_1-\alpha}(k-j_1-\alpha-1)}{\alpha!(j_2-j_1-1-\alpha)!(nk-nj_1+n-1-n\alpha)_c \cdot (k-j_1-\alpha)(nk-nj_1-c-1-n\alpha)}.$$

4. The distribution function of $D^-_{j_1j_2}(k, n)$. Because the samples are of equal size n, possible values of $D^-_{j_1j_2}(k, n)$, as defined by (1.11), are the values of c/n, $c = 0, 1, \cdots, n-1$, and

$$(4.1) \qquad P(D^-_{j_1j_2}(k, n) \leqq (n-1)/n) = 1.$$

For the nontrivial case

$$(4.2) \quad P(D^-_{j_1 j_2}(k, n) \leqq c/n) = P(Y_{1j_1} < Y_{c+1,j_2}, Y_{2j_1} < Y_{c+2,j_2}, \cdots, Y_{n-c,j_1} < Y_{nj_2}).$$

The right side of (4.2) is obtained, as in the previous section, with the aid of (2.8) and (2.9). Instead of (3.3) the desired portion of the sample space is now represented by the restrictions

$$(4.3) \qquad t_{2r} < t_{c+2,s} \,;\, t_{3r} < t_{c+3,s} \,;\, \cdots \,;\, t_{n-c,r} < t_{ns}$$

because $t_{1r} < t_{c+1,s}$ is already implied by (2.9). The desired probability is then represented by the integral

$$\begin{aligned} P(D^-_{j_1 j_2}(k, n) \leqq c/n) = {} & \int_{-\infty}^{\infty} \int_{t_{1r}}^{\infty} \int_{t_{1s}}^{\infty} \cdots \int_{t_{c-1,s}}^{\infty} \int_{t_{cs}}^{\infty} \int_{t_{c+1,s}}^{\infty} \\ & \cdot \int_{t_{1r}}^{t_{c+2,s}} \cdots \int_{t_{n-2,s}}^{\infty} \int_{t_{n-c-2,r}}^{t_{n-1,s}} \int_{t_{n-1,s}}^{\infty} \int_{t_{n-c-1,r}}^{t_{ns}} \int_{t_{n-c,r}}^{\infty} \cdots \\ (4.4) \qquad & \int_{t_{n-2,r}}^{\infty} \int_{t_{n-1,r}}^{\infty} G \, dF(t_{n,r}) \, dF(t_{n-1,r}) \cdots dF(t_{n-c+1,r}) \, dF(t_{n-c,r}) \\ & \cdot dF(t_{ns}) \, dF(t_{n-c-1,r}) \, dF(t_{n-1,s}) \cdots dF(t_{2,r}) \, dF(t_{c+2,s}) \\ & \cdot dF(t_{c+1,s}) \, dF(t_{cs}) \cdots dF(t_{2s}) \, dF(t_{1s}) \, dF(t_{1r}). \end{aligned}$$

The integration is again straightforward. The identity (3.6) leads to

$$(4.5) \quad P\left(D^-_{j_1 j_2}(k, n) \leqq \frac{c}{n}\right) = 1 - \binom{k - j_1}{j_2 - j_1}\binom{2n-2}{n+c} \Big/ \binom{k - j_1 + (c+1)/n}{j_2 - j_1}\binom{2n-2}{n-1}.$$

5. The two-sample problem. The distribution functions of the two-sample random variables $D^+_{1,2}(2, n)$ and $D^-_{1,2}(2, n)$, defined by (1.5) and (1.6), respectively, can be found from (3.7) and (4.5). Thus

$$(5.1) \qquad P(D^+_{1,2}(2, n) \leqq c/n) = 1 - \binom{2n-1}{n+c} \Big/ \binom{2n-1}{n}$$

and

$$(5.2) \qquad P(D^-_{1,2}(2, n) \leqq c/n) = 1 - \binom{2n-1}{n+c+1} \Big/ \binom{2n-1}{n}.$$

A simpler method for obtaining (5.1) and (5.2) is based on the method used by Gnedenko and Koroluk (1951) for obtaining the distribution of $D^+(n, n)$, defined by (1.3). The method is described also by Fisz (1963). Consider the path of a particle moving from the point $(0, 0)$ to the point (n, n) in the (x, y) plane, with the restriction that the particle movements are in unit increments in a positive direction parallel to either one axis or the other. Then the probability of the event $D^+(n, n) \leqq c/n$ equals the proportion of the total number of paths from $(0, 0)$ to (n, n) that do not touch the line $x = y + c + 1$. The probability of the event $D^+_{1,2}(k, n) \leqq c/n$ can be represented as the proportion of the total number of paths from $(1, 0)$ to (n, n) that do not touch the line $x = y + c + 1$, which is given by (5.1). Also the probability of the event $D^-_{1,2}(k, n) \leqq c/n$ can be represented as the proportion of the total number of paths from (0, 1) to (n, n) that do not touch the line $x = y + c + 1$, given by (5.2). The desired counts on the numbers of paths are easily obtained with the method of reflected paths described in detail in the references mentioned above.

6. The limit as $n \to \infty$. As the sample size n increases without bound, for fixed number of samples, k, the asymptotic distribution function of $D^+_{j_1j_2}(k, n)$ and $D^-_{j_1j_2}(k, n)$ both approach the well known limiting distribution function of the random variable $D^+(n, n)$. The proof is as follows.

Obviously,

$$(6.1)\quad \lim_{n\to\infty} \binom{k - j_1}{j_2 - j_1} \Big/ \binom{k - j_1 - (\lambda n^{\frac{1}{2}} + 1)/n}{j_2 - j_1} = \lim_{n\to\infty} \binom{k - j_1}{j_2 - j_1} \Big/ \binom{k - j_1 + (\lambda n^{\frac{1}{2}} + 1)/n}{j_2 - j_1} = 1.$$

Also, from Fisz (1963), if

$$(6.2)\qquad I = n!n!/(n - c)!(n + c)!,$$

where c is approximately $\lambda n^{\frac{1}{2}}$, then

$$(6.3)\qquad \lim_{n\to\infty} I = \exp -\lambda^2.$$

Applying (6.1) and (6.3) to (4.5), and defining λ as $c/n^{\frac{1}{2}}$, we have

$$\lim_{n\to\infty} P\left(D^-_{j_1j_2}(k, n) \leqq \frac{c}{n}\right)$$

$$(6.4)\quad = \lim_{n\to\infty} \left(1 - \binom{k - j_1}{j_2 - j_1} (n - 1)!(n - 1)! \Big/ \binom{k - j_1 + (c + 1)/n}{j_2 - j_1} (n + c)!(n - c - 2)!\right)$$

$$= 1 - \lim_{n\to\infty} [(n - c)(n - c - 1)/n^2]\, I = 1 - \exp - \lambda^2.$$

Therefore

$$(6.5)\qquad \lim_{n\to\infty} P(D^-_{j_1j_2}(k, n) \leqq \lambda/n^{\frac{1}{2}}) = 1 - \exp -\lambda^2.$$

Similarly, since

(6.6) $(n - 1)_c/(nk - nj_1 + n - 1 - n\alpha)_c < 1$ for positive integers α and c, and $\alpha \leqq j_2 - j_1 - 1$,

$$(6.7)\quad \lim_{n\to\infty} \sum_{\alpha=0}^{j_2-j_1-1} \frac{(n - 1)_c(c + 1)(k - j_1)_{j_2-j_1}(-1)^{j_2-j_1-\alpha}(k - j_1 - \alpha - 1)}{\alpha!(j_2 - j_1 - 1 - \alpha)!(nk - nj_1 + n - 1 - n\alpha)_c \cdot (k - j_1 - \alpha)(nk - nj_1 - c - 1 - n\alpha)} = 0$$

even when $c = O(n^{\frac{1}{2}})$, the use of (6.1) and (6.3) in conjunction with (3.7) leads to

$$(6.8) \qquad \lim_{n\to\infty} P(D^{+}_{j_1 j_2}(k, n) \leqq \lambda/n^{\frac{1}{2}}) = 1 - \exp -\lambda^2.$$

The foregoing merely serves to emphasize the warning, indicated by Hodges (1957), that exact tables should be used for $D^{+}(n, n)$, when available, since $D^{+}(n, n)$ has the same limiting distribution as the random variables $D^{+}_{j_1 j_2}(k, n)$ and $D^{-}_{j_1 j_2}(k, n)$, regardless of the number of samples and the value of j_1 and j_2 being considered, if the number of samples k is bounded.

REFERENCES

Birnbaum, Z. W. and Hall, R. A. (1960). Small sample distributions for multi-sample statistics of the Smirnov type. *Ann. Math. Statist.* **31** 710–720.

Conover, W. J. (1964). A k-sample model in order statistics. (Submitted to *Ann. Math. Statist.* for publication.)

Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramér-von Mises tests. *Ann. Math. Statist.* **28** 823–838.

David, H. T. (1958). A three-sample Kolmogorov-Smirnov test. *Ann. Math. Statist.* **29** 842–851.

Dwass, M. (1960). Some k-sample rank-order tests. *Contributions to Probability and Statistics—Essays in Honor of Harold Hotelling.* Stanford Univ. Press.

Fisz, M. (1963). *Probability Theory and Mathematical Statistics.* Wiley, New York.

Gnedenko, B. V. and Koroluk, V. S. (1951). On the maximum divergence of two empirical distributions (in Russian). *Dokl. Akad. Nauk* USSR **80** 525–528.

Hartley, H. O. (1950). The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika* **37** 308–312.

Hodges, J. L., Jr. (1957). The significance probability of the Smirnov two-sample test. *Ark. Mat.* **3** 469–486.

Kiefer, J. (1959). k-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests. *Ann. Math. Statist.* **30** 420–447.

Massey, F. J., Jr. (1951). The distribution of the maximum deviation between two samples' cumulative step functions. *Ann. Math. Statist.* **22** 125–128.

Wilks, S. S. (1962). *Mathematical Statistics.* Wiley, New York.