The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1157–1166

ON THE COMPARISON OF TWO EMPIRICAL DISTRIBUTION FUNCTIONS[1]

BY LAJOS TAKÁCS

Case Western Reserve University

1. Introduction. Let $\xi_1, \xi_2, \cdots, \xi_m$ be mutually independent random variables having a common distribution function $F(x)$. Denote by $F_m(x)$ the empirical distribution function of the sample $(\xi_1, \xi_2, \cdots, \xi_m)$. The empirical distribution function $F_m(x)$ is defined as the number of variables $\xi_1, \xi_2, \cdots, \xi_m$ less than or equal to x divided by m.

Furthermore, let $\eta_1, \eta_2, \cdots, \eta_n$ be mutually independent random variables having a common distribution function $G(x)$, and denote by $G_n(x)$ the empirical distribution function of the sample $(\eta_1, \eta_2, \cdots, \eta_n)$.

For the purpose of testing the hypothesis that $F(x) \equiv G(x)$ in 1939 N. V. Smirnov [6] introduced the statistic

$$\delta^+(m,n) = \sup_{-\infty < x < \infty} [F_m(x) - G_n(x)] \tag{1}$$

and showed that if $F(x)$ and $G(x)$ are two identical continuous distribution functions, then the distribution of $\delta^+(m, n)$ does not depend on $F(x) \equiv G(x)$, and

$$\lim_{m\to\infty, n\to\infty} \mathbf{P}\left\{\left(\frac{mn}{m+n}\right)^{\frac{1}{2}} \delta^+(m,n) \leqq x\right\} = 1 - e^{-2x^2} \tag{2}$$

for $x \geqq 0$. In this case the distribution of the random variable $\delta^+(m, n)$ for $n = m$ was found in 1951 by B. V. Gnedenko and V. S. Korolyuk [2], and for $n = mp$ where p is a positive integer in 1955 by V. S. Korolyuk [3]. (See also [7] and [8].) Obviously $\delta^+(m, n)$ and $\delta^+(n, m)$ have the same distribution for all $m = 1, 2, \cdots$ and $n = 1, 2, \cdots$.

We can express $\delta^+(m, n)$ also in a simpler way. Denote by $\eta_1^*, \eta_2^*, \cdots, \eta_n^*$ the random variables $\eta_1, \eta_2, \cdots, \eta_n$ arranged in increasing order of magnitude. Then we can write that

$$\delta^+(m,n) = \max_{1 \leqq r \leqq n} [F_m(\eta_r^*) - G_n(\eta_r^* - 0)]. \tag{3}$$

Now let us introduce another statistic. For any a let us define $\eta_a(m, n)$ as the number of subscripts $r = 1, 2, \cdots, n$ for which

$$G_n(\eta_r^* - 0) \leqq F_m(\eta_r^*) + a/n < G_n(\eta_r^*). \tag{4}$$

Received May 25, 1970.

[1] This research was sponsored by the National Science Foundation under Contract No. GP-9629.

If we suppose that $F(x)$ and $G(x)$ are two identical continuous distribution functions, then we can easily see that the distribution of the random variable $\eta_a(m, n)$ does not depend on $F(x) \equiv G(x)$, and we have

$$\mathbf{P}\{\delta^+(m,n) \leqq a/n\} = \mathbf{P}\{\eta_a(m,n) = 0\} \tag{5}$$

for $a \geqq 0$. Thus the problem of finding the distribution of $\delta^+(m, n)$ can be considered as a particular case of the more general problem of finding the distribution of $\eta_a(m, n)$.

In this paper we shall be concerned with the problem of finding the distribution of $\eta_a(m, n)$. In the case when $F(x)$ and $G(x)$ are two identical continuous distribution functions we shall find the distribution of the random variable $\eta_a(m, n)$ for every a if $n = mp$ where p is a positive integer. In the particular case when $n = m$, the distribution of $\eta_a(m, n)$ was found in 1952 by V. S. Mihalevič [4]. He showed that if $a = 0, 1, \cdots, m$, then

$$\mathbf{P}\{\eta_a(m,m) \leqq k\} = \mathbf{P}\left\{\delta^+(m,m) \leqq \frac{k+a}{m}\right\} = 1 - \frac{\binom{2m}{m+k+a+1}}{\binom{2m}{m}} \tag{6}$$

for $k = 0, 1, 2, \cdots, m-a$.

2. The distribution of $\eta_a(m, n)$. Throughout this paper we suppose that $F(x)$ and $G(x)$ are identical continuous distribution functions. For $r = 1, 2, \cdots, n+1$ let us define ν_r as p times the number of variables $\xi_1, \xi_2, \cdots, \xi_m$ falling in the interval $(\eta^*_{r-1}, \eta_r^*]$ where $\eta_0^* = -\infty$ and $\eta^*_{n+1} = \infty$. The random variables $\nu_1, \nu_2, \cdots, \nu_{n+1}$ are interchangeable. Set $N_r = \nu_1 + \cdots + \nu_r$ for $r = 1, 2, \cdots, n+1$. Obviously $N_{n+1} = mp$. We have

$$\mathbf{P}\{N_i = sp\} = \frac{\binom{i+s-1}{s}\binom{m+n-i-s}{m-s}}{\binom{m+n}{m}} \tag{7}$$

for $1 \leqq i \leqq n$ and $0 \leqq s \leqq m$, and

$$\mathbf{P}\{N_i = sp \mid N_{i+j} = (s+t)p\} = \frac{\binom{i+s-1}{s}\binom{j+t-1}{t}}{\binom{i+j+s+t-1}{s+t}} \tag{8}$$

for $1 \leqq i < i+j \leqq n$ and $0 \leqq s \leqq s+t \leqq m$.

By using this notation we have

$$F_m(\eta_r^*) = \frac{N_r}{mp}, \qquad G_n(\eta_r^*) = \frac{r}{n}, \qquad G_n(\eta_r^* - 0) = \frac{r-1}{n} \tag{9}$$

for $r = 1, 2, \cdots, n$ and we can write that

$$\delta^+(m,n) = \max_{1 \leqq r \leqq n}\left[\frac{N_r}{mp} - \frac{r-1}{n}\right]. \tag{10}$$

Furthermore, $\eta_a(m, n)$ is equal to the number of subscripts $r = 1, 2, \cdots, n$ for which

$$\frac{r-1}{n} \leqq \frac{N_r}{mp} + \frac{a}{n} < \frac{r}{n}. \tag{11}$$

If, in particular, $n = mp$ where p is a positive integer, then $\eta_a(m, n)$ can be interpreted as the number of subscripts $r = 1, 2, \cdots, n$ for which

$$N_r = r - [a] - 1 \tag{12}$$

where $[a]$ is the greatest integer $\leqq a$.

Thus if $n = mp$, then we have

$$\mathbf{P}\{\eta_a(m, n) = k\} = \mathbf{P}\{\eta_{[a]}(m, n) = k\} \tag{13}$$

for all a and $k = 0, 1, \cdots, m$. Furthermore, we have also

$$\mathbf{P}\{\eta_a(m, n) = k\} = \mathbf{P}\{\eta_{-[a]-1}(m, n) = k\} \tag{14}$$

for all a and $k = 0, 1, \cdots, m$. For $N_{n+1} = mp$ and thus

$$\begin{aligned} \mathbf{P}\{\eta_a(m, n) = k\} &= \mathbf{P}\{N_r = r - [a] - 1 \quad \text{for} \quad k \quad \text{subscripts} \quad r = 1, 2, \cdots, n\} \\ &= \mathbf{P}\{N_{n+1} - N_r = n + 1 - r + [a] \quad \text{for} \quad k \quad \text{subscripts} \\ &\qquad\qquad r = 1, 2, \cdots, n\} \\ &= \mathbf{P}\{N_i = i + [a] \quad \text{for} \quad k \quad \text{subscripts} \quad i = 1, 2, \cdots, n\} \end{aligned} \tag{15}$$

which proves (14).

Accordingly, if $n = mp$ and if we know the distribution of $\eta_a(m, n)$ for $a = 0, 1, 2, \cdots$, then by (13) and (14) we can find the distribution of $\eta_a(m, n)$ for all a. Obviously $\eta_a(m, n) = 0$ if $a \geqq n$. If $a = 0, 1, \cdots, n$, then $\eta_a(m, n)$ is a discrete random variable with possible values $k = 0, 1, \cdots, [(mp - a)/p]$.

The following theorem is the main result of this paper.

THEOREM 1. *If $n = mp$ where p is a positive integer and $a = 0, 1, \cdots, mp$, then we have*

$$\begin{aligned} \mathbf{P}\{\eta_a(m, n) \leqq k\} &= 1 \equiv \frac{p^k}{\binom{mp+m}{m}} \sum_{a/p < j \leqq m-k} \frac{k(p+1)+a+1}{(m-j)(p+1)+a+1} \binom{jp+j-a-1}{j} \\ &\qquad \cdot \binom{(m-j)(p+1)+a+1}{m-j-k} \\ &= 1 - \frac{p^k \binom{mp+m}{m-k}}{\binom{mp+m}{m}} + \frac{p^k}{\binom{mp+m}{m}} \sum_{0 \leqq j \leqq a/p} \frac{k(p+1)+a+1}{(m-j)(p+1)+a+1} \binom{jp+j-a-1}{j} \\ &\qquad \cdot \binom{(m-j)(p+1)+a+1}{m-j-k} \end{aligned} \tag{16}$$

for $0 \leqq k < (mp-a)/p$. If, in particular, $a = 0$, then (16) *reduces to*

$$\mathbf{P}\{\eta_0(m, n) \leqq k\} = 1 - \frac{p^{k+1}\binom{mp+m}{m-k-1}}{\binom{mp+m}{m}} \tag{17}$$

for $0 \leqq k < m$.

PROOF. We shall determine the probability

$$p_k(m, a) = \mathbf{P}\{\eta_a(m, n) > k\} \tag{18}$$

for $0 \leqq k < (mp-a)/p$ and $a = 0, 1, \cdots, mp$. By (15) we have

(19)

$$p_k(m, a) = \mathbf{P}\{N_i = i+a \quad \text{for more than} \quad k \quad \text{subscripts} \quad i = 1, 2, \cdots, mp\}.$$

As we shall see, $p_k(m, a)$ can be expressed by the following probabilities:

(20) $q_k(s) = \mathbf{P}\{N_i = i \quad \text{for} \quad k \quad \text{subscripts} \quad i = 1, 2, \cdots, sp \mid N_{sp} = sp\}$

for $1 \leqq k \leqq s \leqq m$ and

(21) $r_k(s, a) = \mathbf{P}\{N_i = i+a \text{ for at least } k \text{ subscripts } i = 1, 2, \cdots, sp-a-1 \mid N_{sp-a} = sp\}$

for $0 \leqq k < (sp-a)/p \leqq (mp-a)/p$. Obviously $r_0(s, a) = 1$ for $0 \leqq a < sp \leqq mp$.

We shall need the following result: If $0 \leqq r \leqq j \leqq n+1$ and $\mathbf{P}\{N_j = r\} > 0$, then

$$\mathbf{P}\{N_i < i \text{ for } i = 1, 2, \cdots, j \mid N_j = r\} = 1 - r/j. \tag{22}$$

This can easily be proved by mathematical induction. (See [8].)

Now we can write that

$$p_k(m, a) = \sum_{k+a/p<s\leqq m} \frac{a+1}{(m-s)p+a+1} \mathbf{P}\{N_{sp-a} = sp\} r_k(s, a) \tag{23}$$

for $0 \leqq k < (mp-a)/p$ and $a = 0, 1, \cdots, n$. For the event $\{N_i = i+a$ for more than k subscripts $i = 1, 2, \cdots, n\}$ can occur in such a way that for some s where $a+kp < sp \leqq mp$ we have $N_{sp-a} = sp$, further $N_i = i+a$ for at least k subscripts $i = 1, 2, \cdots, sp-a-1$ and $N_i < i+a$ for $sp < i \leqq n$. By using (22) and the fact that $N_{n+1} = n$ we obtain that

$$\mathbf{P}\{N_i < i+a \quad \text{for} \quad sp < i \leqq n \mid N_{sp-a} = sp\} = \frac{a+1}{n-sp+a+1} \tag{24}$$

if $a \leqq sp \leqq mp$. Hence (23) follows.

Furthermore, we have

$$r_k(s, a) = \textstyle\sum_{k\leqq u<s-a/p} \mathbf{P}\{N_{up} = up \mid N_{sp-a} = sp\} q_k(u) \tag{25}$$

for $1 \leqq k < (sp-a)/p$ and $a = 0, 1, \cdots, sp$.

It follows immediately from the definition of $q_k(s)$ that

$$q_k(s) = \textstyle\sum_{1\leqq u<s} \mathbf{P}\{N_{up} = up \mid N_{sp} = sp\} q_1(u) q_{k-1}(s-u) \tag{26}$$

for $2 \leqq k \leqq s$ and

$$q_1(s) = 1 - \textstyle\sum_{1\leqq u<s} \mathbf{P}\{N_{up} = up \mid N_{sp} = sp\} q_1(u) \tag{27}$$

for $s \geqq 1$.

In the above formulas we have

$$\mathbf{P}\{N_{up} = up \mid N_{sp-a} = sp\} = \frac{\binom{up+u-1}{u}\binom{(s-u)(p+1)-a-1}{s-u}}{\binom{sp+s-a-1}{s}} \tag{28}$$

which follows from (8).

Accordingly the problem of finding $p_k(m, a)$ can be reduced to the problem of finding $r_k(s, a)$ for $(a+kp)/p < s \leqq m$, $q_k(s)$ for $k \leqq s \leqq m$ and $q_1(s)$ for $1 \leqq s \leqq m$. These probabilities can be determined by (23), (25), (26) and (27). We shall perform the necessary calculations in the following sections.

We note that $p_k(m, 0)$ can also be obtained by the following formula:

$$p_k(m,0) = \sum_{s=k+1}^{m} \mathbf{P}\{N_{sp} = sp\}q_{k+1}(s) \tag{29}$$

for $0 \leqq k < m$. To prove (29) we take into consideration that the event $\{N_i = i$ for more than k subscripts $i = 1, 2, \cdots, n\}$ can occur in such a way that the $k+1$ st largest $i = 1, 2, \cdots, n$ for which $N_i = i$ is $i = sp$ where $k < s \leqq m$.

It will be convenient to use the following notation. Let

$$P_k(m, a) = \binom{mp+m}{m} p_k(m, a), \tag{30}$$

$$R_k(s, a) = \binom{sp+s-a-1}{s} r_k(s, a), \tag{31}$$

and

$$Q_k(s) = \binom{sp+s-1}{s} q_k(s). \tag{32}$$

Then equations (23), (25), (26), (27) and (29) can also be expressed in the following way

$$P_k(m, a) = \sum_{k+a/p<s\leqq m} \frac{a+1}{(m-s)p+a+1} \binom{(m-s)(p+1)+a}{m-s} R_k(s, a) \tag{33}$$

for $0 \leqq k < (mp-a)/p$ and $a = 0, 1, \cdots, mp$,

$$R_k(s, a) = \sum_{k\leqq u<s-a/p} \binom{(s-u)(p+1)-a-1}{s-u} Q_k(u) \tag{34}$$

for $1 \leqq k < (sp-a)/p$ and $a = 0, 1, \cdots, sp$,

$$Q_k(s) = \sum_{1\leqq u<s} Q_1(u)Q_{k-1}(s-u) \tag{35}$$

for $2 \leqq k \leqq s$ and

$$Q_1(s) = \binom{sp+s-1}{s} - \sum_{1\leqq u<s} \binom{(s-u)(p+1)-1}{s-u} Q_1(u) \tag{36}$$

for $s \geqq 1$.

Furthermore, by (29) we have

$$P_k(m,0) = \sum_{s=k+1}^{m} \binom{(m-s)(p+1)}{m-s} Q_{k+1}(s) \tag{37}$$

for $0 \leqq k < m$.

In the following sections we shall determine $Q_k(s)$, $R_k(s, a)$ and $P_k(m, a)$. We can easily see that $Q_k(s)$ and $R_k(s, a)$ are independent of m whenever $s \leqq m$.

3. Auxiliary theorems. In what follows we shall need certain generating functions which we shall derive in this section. By using Rouché's theorem we can prove that if $|z| < p^p/(p+1)^{p+1}$, then the equation

$$1-w+zw^{p+1} = 0 \tag{38}$$

has a single root $w = \gamma(z)$ in the circle $|w-1| < 1/p$ and if $g(w)$ is a regular function of w in this circle then by Lagrange's expansion we obtain that

(39) $$g(\gamma(z)) = g(1)+\sum_{r=1}^{\infty}\frac{z^r}{r!}\left[\frac{d^{r-1}g'(1+x)(1+x)^{rp+r}}{dx^{r-1}}\right]_{x=0}.$$

It follows immediately from (39) that

(40) $$g(\gamma(z))\gamma'(z) = \sum_{r=0}^{\infty}\frac{z^r}{r!}\left[\frac{d^r g(1+x)(1+x)^{(r+1)(p+1)}}{dx^r}\right]_{x=0}.$$

In particular, if a is any real number and k is a nonnegative integer, then by (39) we have

(41) $$[\gamma(z)]^a[\gamma(z)-1]^k = z^k+\sum_{r=k+1}^{\infty}\frac{kp+k+a}{r-k}\binom{rp+r+a-1}{r-k-1}z^r$$

and by (40)

(42) $$[\gamma(z)]^a[\gamma(z)-1]^k\gamma'(z) = \sum_{r=k}^{\infty}\binom{(r+1)(p+1)+a}{r-k}z^r$$

for $|z| < p^p/(p+1)^{p+1}$.

Finally, we note that

(43) $$\log\gamma(z) = \lim_{a\to 0}\frac{[\gamma(z)]^a-1}{a} = \sum_{r=1}^{\infty}\frac{z^r}{rp+r}\binom{rp+r}{r}$$

and hence

(44) $$\frac{p[\gamma(z)-1]}{1-p[\gamma(z)-1]} = \frac{pz\gamma'(z)}{\gamma(z)} = pz\frac{d\log\gamma(z)}{dz} = \sum_{r=1}^{\infty}\binom{rp+r-1}{r}z^r$$

for $|z| < p^p/(p+1)^{p+1}$. If $z \to p^p/(p+1)^{p+1}$, then $\gamma(z) \to p/(p+1)$.

4. The determination of $Q_k(s)$. We can find $Q_k(s)$ for $1 \leqq k \leqq s$ by (35) and (36).

THEOREM 2. *If* $1 \leqq k \leqq s$, *then*

(45) $$Q_k(s) = \frac{kp^k}{s}\binom{sp+s}{s-k}.$$

PROOF. If we form the generating function of (36), then we obtain that

(46) $$\sum_{s=1}^{\infty} Q_1(s)z^s = \frac{\sum_{s=1}^{\infty}\binom{sp+s-1}{s}z^s}{1+\sum_{s=1}^{\infty}\binom{sp+s-1}{s}z^s} = p[\gamma(z)-1]$$

for $|z| < p^p/(p+1)^{p+1}$. The extreme right member can be obtained by (44).

If we form the generating function of (35), then we obtain that

(47) $$\sum_{s=k}^{\infty} Q_k(s)z^s = \left(\sum_{s=1}^{\infty} Q_1(s)z^s\right)^k = p^k[\gamma(z)-1]^k$$

for $k = 1, 2, \cdots$. Hence (45) follows by (41).

5. The determination of $R_k(s, a)$. We can find $R_k(s, a)$ for $1 \leqq k < (sp-a)/p$ by (34).

THEOREM 3. *If* $1 \leqq k < (sp-a)/p$, *then we have*

$$R_k(s,a) = kp^k \sum_{a/p<j\leqq s-k} \frac{1}{(s-j)} \binom{(s-j)(p+1)}{s-j-k}\binom{jp+j-a-1}{j} \tag{48}$$

$$= p^k\binom{sp+s-a-1}{s-k} - kp^k \sum_{0\leqq j\leqq a/p} \frac{1}{(s-j)}\binom{(s-j)(p+1)}{s-j-k}\binom{jp+j-a-1}{j}.$$

PROOF. By (34) we obtain that

$$\textstyle\sum_{k+a/p<s} R_k(s,a)z^s = (\sum_{u=k}^{\infty} Q_k(u)z^u)(\sum_{a/p<j}\binom{jp+j-a-1}{j}z^j) \tag{49}$$

for $|z| < p^p/(p+1)^{p+1}$. Hence by (45) we obtain the first expression on the right-hand side of (48). Since by (42)

$$\textstyle\sum_{j=0}^{\infty}\binom{jp+j-a-1}{j}z^j = [\gamma(z)]^{-a-p-2}\gamma'(z), \tag{50}$$

it follows from (49) that

$$\textstyle\sum_{k+a/p<s} R_k(s,a)z^s = p^k[\gamma(z)]^{-a-p-2}[\gamma(z)-1]^k\gamma'(z) \tag{51}$$

$$\textstyle -(\sum_{u=k}^{\infty} Q_k(u)z^u)(\sum_{0\leqq j\leqq a/p}\binom{jp+j-a-1}{j}z^j).$$

If we form the coefficient of z^s in (51), then we obtain the second expression on the right-hand side of (48).

We note that

$$R_k(s,0) = p^{k+1}\binom{sp+s-1}{s-k-1} \tag{52}$$

for $1 \leqq k < s$. This follows from (48) or from (45) if we take into consideration that

$$\textstyle R_k(s,0) = \sum_{u=k+1}^{s} Q_k(u) \tag{53}$$

for $1 \leqq k < s$.

By our definition we have

$$R_0(s,a) = \binom{sp+s-a-1}{s} \tag{54}$$

for $0 \leqq a < sp$.

6. The determination of $P_k(m, a)$. Finally, we can find $P_k(m, a)$ for $0 \leqq k < (mp-a)/p$ by (33).

THEOREM 4. *If* $0 \leqq k < (mp-a)/p$, *then we have*

$$P_k(m,a) = p^k \sum_{a/p<j\leqq m-k} \frac{k(p+1)+a+1}{(m-j)(p+1)+a+1}\binom{jp+j-a-1}{j}\binom{(m-j)(p+1)+a+1}{m-j-k} \tag{55}$$

$$= p^k\binom{mp+m}{m-k} - p^k \sum_{0\leqq j\leqq a/p} \frac{k(p+1)+a+1}{(m-j)(p+1)+a+1}\binom{jp+j-a-1}{j} \cdot$$

$$\binom{(m-j)(p+1)+a+1}{m-j-k}.$$

PROOF. By (33) we obtain that

$$(56)\qquad \sum_{k+a/p<m} P_k(m,a)z^m = \left(\sum_{k+a/p<s} R_k(s,a)z^s\right)\left(\sum_{s=0}^{\infty} \frac{a+1}{sp+a+1}\binom{sp+s+a}{s}z^s\right)$$

for $|z| < p^p/(p+1)^{p+1}$. Here by (49)

$$(57)\qquad \textstyle\sum_{k+a/p<s} R_k(s,a)z^s = p^k[\gamma(z)-1]^k \sum_{a/p<j}\binom{jp+j-a-1}{j}z^j$$

for $k = 1, 2, \cdots$. If $k = 0$, then (57) is trivially true. Furthermore, by (41)

$$(58)\qquad \sum_{s=0}^{\infty} \frac{a+1}{sp+a+1}\binom{sp+s+a}{s}z^s = 1+\sum_{s=1}^{\infty}\frac{a+1}{s}\binom{sp+s+a}{s-1}z^s = [\gamma(z)]^{a+1}.$$

Thus, finally

$$(59)\qquad \textstyle\sum_{k+a/p<m} P_k(m,a)z^m = p^k[\gamma(z)]^{a+1}[\gamma(z)-1]^k \sum_{a/p<j}\binom{jp+j-a-1}{j}z^j$$

for $|z| < p^p/(p+1)^{p+1}$. If we make use of (41) and form the coefficient of z^m in (59), then we obtain the first expression in (55). If in (59) we write

$$(60)\qquad \textstyle\sum_{a/p<j}\binom{jp+j-a-1}{j}z^j = [\gamma(z)]^{-a-p-2}\gamma'(z) - \sum_{0\leqq j\leqq a/p}\binom{jp+j-a-1}{j}z^j,$$

which follows from (50), then we can obtain the second expression in (55).

By (55) we obtain (16) and this completes the proof of Theorem 1.

7. Limit Distributions. First we shall find the limit distribution of $\eta_{ap}(m, mp)$ when a is a nonnegative real number and $p \to \infty$.

THEOREM 5. *If* $a \geqq 0$ *and* $n = mp$, *then*

$$(61)\qquad \lim_{p\to\infty} \mathbf{P}\{\eta_{ap}(m,n) \leqq k\} = 1 - \frac{(a+k)m!}{m^m}\sum_{a<j\leqq m-k}\frac{(j-a)^j(m-j+a)^{m-j-k-1}}{j!(m-j-k)!}$$

$$= 1 - \frac{m!}{(m-k)!m^k} + \frac{(a+k)m!}{m^m}\sum_{0\leqq j\leqq a}\frac{(j-a)^j(m-j+a)^{m-j-k-1}}{j!(m-j-k)!}$$

for $0 \leqq k < m-a$. *If, in particular,* $a = 0$, *then* (61) *reduces to*

$$(62)\qquad \lim_{p\to\infty} \mathbf{P}\{\eta_0(m,mp) \leqq k\} = 1 \equiv \frac{m!}{(m-k-1)!m^{k+1}}$$

for $0 \leqq k < m$.

PROOF. Since

$$(63)\qquad \mathbf{P}\{\eta_{ap}(m,n) \leqq k\} = \mathbf{P}\{\eta_{[ap]}(m,n) \leqq k\},$$

the results can be obtained immediately from (16) and (17) if we replace a by $[ap]$ and let $p \to \infty$.

By using (61) we can find the solution of another problem too.

Denote by $\sigma(m, a)$ the number of intersections of $G(x)$ with $F_m(x)+a/m$ for $-\infty \leqq x \leqq \infty$. More precisely $\sigma(m, a) = k$ if the set $S = \{x : G(x) = F_m(x)+a/m$ and $-\infty \leqq x \leqq \infty\}$ is the union of k separated intervals or points. Since

$$\mathbf{P}\{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |G_n(x)-G(x)| = 0\} = 1, \tag{64}$$

we can conclude that

$$\mathbf{P}\{\sigma(m,a) \leqq k\} = \lim_{p\to\infty} \mathbf{P}\{\eta_{ap}(m, mp) \leqq k\} \tag{65}$$

for $0 \leqq k < m-a$ and $0 < a < m$. The right-hand side of (65) is given by (61). For $a = 0$ we have

$$\mathbf{P}\{\sigma(m,0) \leqq k\} = 1-\frac{m!}{(m-k)!m^k} \tag{66}$$

if $1 \leqq k \leqq m$. This can be seen as follows: If we suppose that $a > 0$ and let $a \to 0$, then $x = \infty$ becomes a point of S whenever $a = 0$. Thus we can write that

$$\mathbf{P}\{\sigma(m,0) \leqq k\} = \lim_{a\to 0} \mathbf{P}\{\sigma(m,a) \leqq k-1\} \tag{67}$$

for $1 \leqq k \leqq m$ whence (66) follows.

The probability $\mathbf{P}\{\sigma(m, a) \leqq k\}$ for $0 < a < m$ was given without proof in 1960 by D. A. Darling [1] and for $a = 0, 1, \cdots, m-1$ it was found in 1964 by W. Nef [5].

Evidently $\sigma(m, a)$ and $\sigma(m, -a)$ have the same distribution.

Now let us consider the asymptotic distribution of $\eta_a(m, mp)$ in the case when $a = y(mp(p+1))^{\frac{1}{2}}$ and $m \to \infty$.

THEOREM 6. *If* $a = y(mp(p+1))^{\frac{1}{2}}$ *where* $y \geqq 0$ *and* $k = [x(mp/(p+1))^{\frac{1}{2}}]$ *where* $x \geqq 0$, *then*

$$\lim_{m\to\infty} \mathbf{P}\{\eta_a(m, mp) \leqq k\} = 1-e^{-\frac{1}{2}(x+2y)^2}. \tag{68}$$

PROOF. Now

$$\mathbf{P}\{\eta_a(m, mp) \leqq k\} = \mathbf{P}\{\eta_{[a]}(m, mp) \leqq k\} \tag{69}$$

is given explicitly by (16). If in the first formula on the right-hand side of (16) we put $a = [y(mp(p+1))^{\frac{1}{2}}]$, $k = [x(mp/(p+1))^{\frac{1}{2}}]$, $j = mu$ and let $m \to \infty$, then we obtain that

$$\lim_{m\to\infty} \mathbf{P}\{\eta_a(m, mp) \leqq k\} = 1-\frac{(x+y)}{(2\pi)^{\frac{1}{2}}}\int_0^1 \frac{e^{-\frac{1}{2}\left[\frac{(x+y)^2}{1-u}+\frac{y^2}{u}\right]}}{(1-u)^{\frac{3}{2}}u^{\frac{1}{2}}}\,du \tag{70}$$

which is equal to the right-hand side of (68).

For an arbitrary $p > 0$ the limiting distribution (68) was found in 1939 by N. V. Smirnov [6]. In testing the hypothesis that $F(x)$ and $G(x)$ are two identical continuous distribution functions N. V. Smirnov [6] used the asymptotic distribution (68). The results of this paper make it possible to replace the asymptotic distribution in Smirnov's test by an exact distribution if one sample size is an integral multiple of the other.

REFERENCES

[1] Darling, D. A. (1960). On the theorems of Kolmogorov–Smirnov. *Theor. Probability Appl.* **5** 356–361.

[2] Gnedenko, B. V. and Korolyuk, V. S. (1951). On the maximum discrepancy between two empirical distribution functions. (Russian.) *Dokl. Akad. Nauk SSSR* **80** 525–528. [Eng. trans. *Selected Translations in Mathematical Statistics and Probability*. IMS and AMS **1** (1961) 13–16.]

[3] Krorlyuk, V. S. (1955). On the discrepancy of empirical distribution functions for the case of two independent samples. (Russian.) *Izv. Akad. Nauk SSSR Ser. Math.* **19** 81–96. [Eng. trans. *Selected Translations in Mathematical Statistics and Probability*. IMS and AMS 4 (1963) 105–121.]

[4] Mihalevič, V. S. (1952). On the mutual disposition of two empirical distribution functions. (Russian.) *Dokl. Akad. Nauk SSSR* **85** 485–488. [Eng. trans. *Selected Translations in Mathematical Statistics and Probability*. IMS and AMS **1** (1961) 63–67.]

[5] Nef, W. (1964). Über die Differenz zwischen theoretischer und empirischer Verteilungsfunktion. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **3** 154–162.

[6] Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. (Russian.) *Bull. Math. Univ. Moscow Ser. A* **2** 3–14.

[7] Takács, L. (1964). An application of a ballot theorem in order statistics. *Ann. Math. Statist.* **35** 1356–1358.

[8] Takács, L. (1970). Combinatorial methods in the theory of order statistics. *Proceedings of the First International Symposium on Nonparametric Techniques in Statistical Inference, Bloomington, Indiana, June* 1–6, 1969. Cambridge Univ. Press, 359–384.