ON THE KOLMOGOROV-SMIRNOV LIMIT THEOREMS FOR EMPIRICAL DISTRIBUTIONS

BY W. FELLER

*Cornell University**

Summary. Unified and simplified derivations are given for the limiting forms of the difference (1) between the empirical distribution of a large sample and the corresponding theoretical distribution and (2) between the distributions of two large samples.

1. Introduction. Let $X_1, \cdots, X_N$ be mutually independent random variables with the common cumulative distribution function $F(x)$. Let $X_1^*, \cdots, X_N^*$ be the same set of variables rearranged in increasing order of magnitude. *The empirical distribution* (or sum-polygon) *of the sample* $X_1, \cdots, X_N$ *is the step function* $S_N(x)$ *defined by*

$$S_N(x) = \begin{cases} 0 & \text{for } x < X_1^* \\ \dfrac{k}{N} & \text{for } X_k^* \leq x < X_{k+1}^* \\ 1 & \text{for } x \geq X_N^*. \end{cases} \tag{1.1}$$

In other words, $N \cdot S_N(x)$ equals the number of variables X_ν which do not exceed x. We expect intuitively that $S_N(x) \to F(x)$ as $N \to \infty$. In fact, if this were not so the notions of distribution and sample would be meaningless. The so-called ω^2-criterion of von Mises [4] provides rough estimates for the probable deviations of $S_N(x)$ from $F(x)$ for certain forms of $F(x)$ (cf. von Mises [4]). A much stronger result is due to A. Kolmogorov and is of great interest in the theory of non-parametric estimation (Kolmogorov [3]). The maximum of the deviation $| S_N(x) - F(x) |$ is a random variable D_N whose distribution is easily seen to be independent of the special form of $F(x)$ provided only that $F(x)$ is continuous.[1] The exact distribution of D_N is not known, but Kolmogorov found that $N^{\frac{1}{2}}D_N$ has a limiting distribution. More precisely we have

THEOREM 1 (Kolmogorov [1]). *Suppose that* $F(x)$ *is continuous and define the random variable* D_N *by*

$$D_N = \text{l.u.b.} | S_N(x) - F(x) |. \tag{1.2}$$

*Research under an ONR contract.

[1] This fact will not be used explicitly in the sequel but follows as a byproduct from our proofs. A simple direct proof consists in considering the random variables $\Xi_k = F(X_k)$ which are uniformly distributed; the maximum deviation D_N of the empirical distribution of the new sample $\{\Xi_k\}$ from the uniform distribution has the same distribution as D_N; cf. Kolmogorov [1].

Then for every fixed $z \geq 0$ *as* $N \to \infty$

$$\Pr\{D_N \leq zN^{-\frac{1}{2}}\} \to L(z) \tag{1.3}$$

where $L(z)$ *is the cumulative distribution function which for* $z > 0$ *is given by either of the following equivalent relations*[2]

$$L(z) = 1 - 2\sum_{\nu=1}^{\infty} (-1)^{\nu-1} e^{-\nu^2 z^2} = (2\pi)^{\frac{1}{2}} z^{-1} \sum_{\nu=1}^{\infty} e^{-(2\nu-1)^2\pi^2/8z^2} \tag{1.4}$$

For $z \leq 0$ *we have, of course,* $L(z) = 0$.

Equally interesting is Smirnov's result concerning the maximum difference between the empirical distributions of two samples with the same cumulative distribution.

THEOREM 2 (Smirnov [5]). *Let* $(X_1, \cdots, X_m)$ *and* $(Y_1, \cdots, Y_n)$ *be two samples of mutually independent random variables having a common continuous distribution* $F(x)$. *Let* $S_m(x)$ *and* $T_n(x)$ *be the corresponding empirical distribution functions and define a new random variable* $D_{m,n}$ *by*

$$D_{m,n} = \text{l.u.b.} \mid S_m(x) - T_n(x) \mid. \tag{1.5}$$

Put

$$N = \frac{mn}{m+n} \tag{1.6}$$

and suppose that $m \to \infty$, $n \to \infty$ *so that*

$$\frac{m}{n} \to a, \tag{1.7}$$

where a is a constant. Then for every fixed $z \geq 0$

$$\Pr\{D_{m,n} \leq zN^{-\frac{1}{2}}\} \to L(z), \tag{1.8}$$

where $L(z)$ *is the same as in* (1.4).

The original proofs (Kolmogorov [1] and Smirnov [6]) are very intricate and are based on completely different methods. Kolmogorov's proof is based on an auxiliary theorem of equal depth proved in a separate paper (Kolmogorov [2]). An alternative proof of Kolmogorov's theorem is due to Smirnov [5]. However, Smirnov derives both theorems as corollaries to much deeper (but less useful) results concerning the number of intersections of the graphs of $S_N(x)$ and $F(x) \pm \epsilon N^{-\frac{1}{2}}$ and of $S_m(x)$ and $T_n(x) \pm \epsilon N^{-\frac{1}{2}}$, respectively. It is, therefore, not surprising that Smirnov's proofs require a powerful technique and many auxiliary considerations. It is the purpose of the present paper to present unified proofs of the two theorems which are based on methods of great generality.[3] The new

[2] The equivalence of the two formulas in (1.4) is a well-known relation often called transformation formula for theta-functions. We shall only prove the first representation in (1.4). The second is more useful for small z. A table of $L(z)$ is given in Smirnov [6]. It is reprinted in the present issue of the *Annals of Mathematical Statistics* (pp. 279–281).

[3] Among other results which can be proved by the same method are certain limit theorems for ruin and first-passage time problems in the theory of diffusion and random walks.

proof is not simple but simpler than the original ones. At any rate, it requires essentially only routine manipulations with generating functions and their limiting form, the Laplace transforms. However, the paper aims mostly at a unification of methods.

As a byproduct of the proof we obtain

THEOREM 3. *Let A_N be the number of points x where the step-polygon $S_N(x)$ of Theorem 1 leaves the strip $F(x) \pm zN^{-\frac{1}{2}}$. The expected value of the random variable A_N satisfies the asymptotic relation*

$$E(A_N) \sim 2(2\pi N)^{\frac{1}{2}}\{1 - \Phi(2z)\}, \tag{1.9}$$

where $\Phi(z)$ is the normalized Gaussian distribution.

An analogous corollary to Theorem 2 was given by Smirnov [8]: formula (1.9) holds also for the number of intersections of the graph of $S_m(x)$ with the step-polygons $T_n(x) \pm zN^{-\frac{1}{2}}$. These results should come as a surprise to most statisticians. According to Theorem 1 there is a positive probability that $A_N = 0$ and nevertheless $E(A_N)$ is of the order of magnitude $N^{\frac{1}{2}}$. The explanation lies in the fact that *if* $S_N(x)$ crosses the curve $F(x) + zN^{-\frac{1}{2}}$ at some point then it is extremely likely that $S(x)$ will in some neighborhood continue to fluctuate around values $F(x) + zN^{-\frac{1}{2}}$, crossing that curve a great many times. The difference $S_N(x) - F(x)$ exhibits, in the limit $N \to \infty$, many small oscillations. This phenomenon is related to the well-known fact that the path of a particle subject to the Einstein-Wiener diffusion process has no derivatives.

Instead of the absolute values of the differences we may consider the differences themselves and derive two parallel theorems for the maximum and the minimum. As an example we shall prove

THEOREM 4. *With the notations and assumptions of Theorem* 1 *let*

$$D_N^+ = \text{l.u.b.}\{S_N(x) - F(x)\}. \tag{1.10}$$

Then

$$\Pr\{D_N^+ \leq zN^{-\frac{1}{2}}\} \to 1 - e^{-2z^2}. \tag{1.11}$$

The proof is simpler than that of Theorem 1 but uses the same method.

2. Notations and preliminary remarks. For printing convenience it is desirable to avoid complicated subscripts and we shall therefore use the following notation for *binomial coefficients*

$$C(n, k) = \binom{n}{k}. \tag{2.1}$$

Similarly, for the general term of the binomial distribution we shall write

$$B(n, k\,; p) = C(n, k)p^k(1 - p)^{n-k}. \tag{2.2}$$

If A is an event, $\bar{A}$ will denote its negation (complementary event). Finally

$$\Pr\{A \mid B\} \tag{2.3}$$

denotes the conditional probability of A for given B.

Our proofs depend on a special case of the continuity theorem for characteristic functions. Since we shall deal only with probability density functions $f(t)$ which vanish for $t < 0$ it is preferable to use, instead of the characteristic function, the *Laplace transform*

$$\phi(s) = \int_0^\infty e^{-st} f(t)\, dt. \tag{2.4}$$

(This amounts to using the variable $-s$ instead of the usual is and therefore $\phi(s)$ obeys the formal rules for characteristic functions.)

For any sequence $\{u_k\}$ $(k = 1, 2, \cdots)$ of non-negative numbers we define the *generating function* $u(\lambda)$ *by*

$$u(\lambda) = \sum_{k=1}^{\infty} u_k \lambda^k . \tag{2.5}$$

Now let $\delta > 0$ be fixed and consider the step-function $f_\delta(t)$ defined by

$$f_\delta(t) = u_k \quad \text{for } (k-1)\delta \le t < k\delta \tag{2.6}$$

$(k = 1, 2, \cdots ; f_\delta(t) = 0$ for $t < 0)$. Its Laplace transform is

$$\phi_\delta(s) = \frac{e^{\delta s} - 1}{s}\, u(e^{-\delta s}). \tag{2.7}$$

We have, therefore, the *continuity-theorem: If, as* $\delta \to 0$,

$$\delta u(e^{-\delta s}) \to \phi(s), \tag{2.8}$$

then for every fixed $t > 0$

$$u_k \to f(t) \quad \text{when } k\delta \to t; \tag{2.9}$$

conversely, if (2.9) *holds then* (2.8) *is true.*

3. Proof of Theorem 1. Since $F(x)$ is continuous it is possible to define numbers x_k such that

$$F(x_k) = \frac{k}{N}, \qquad (k = 1, 2, \cdots, N-1). \tag{3.1}$$

This definition is unique except when $F(x) = k/N$ within an entire interval, in which case we define x_k as the *left* endpoint of that interval.

Let $c > 0$ be an integer. We shall evaluate the probability of the event $D_N > c/N$ and we shall later put

$$c = zN^{\frac{1}{2}}, \qquad N \to \infty. \tag{3.2}$$

Suppose first that for some particular x

$$S_N(x) - F(x) > \frac{c}{N}. \tag{3.3}$$

This point x is contained in a maximal interval in which (3.3) holds and at the right endpoint ξ of this interval we shall have

$$S_N(\xi) - F(\xi) = \frac{c}{N}. \tag{3.4}$$

Now $S_N(\xi)$ is necessarily a number of the form r/N with an integer r. Since c is an integer also $F(\xi) = k/N$ and hence $\xi = x_k$ for some k. From (3.4) we conclude that

$$X_{k+c}^* < x_k, \qquad X_{k+c+1}^* > x_k \tag{3.5}$$

or in other words: exactly $k + c$ among the N variables X_ν are smaller than x_k. Denote this event by $A_k(c)$. The inequality (3.3) takes place for some x if, and only if, at least one among the events $A_1(c), \cdots, A_N(c)$ occurs. The argument applies equally to $c < 0$ and shows that *the event $D_N > c/N$ occurs if, and only if, at least one among the events*

$$A_1(c),\ A_1(-c),\ A_2(c),\ A_2(-c), \cdots,\ A_N(c),\ A_N(-c) \tag{3.6}$$

occurs.

Let U_r and V_r be the events that in the sequence (3.6) the first event to occur are $A_r(c)$ or $A_r(-c)$, respectively. More formally, the events U_r and V_r are defined by

$$\begin{aligned} U_r &= \bar{A}_1(c)\bar{A}_1(-c) \cdots \bar{A}_{r-1}(c)\bar{A}_{r-1}(-c)A_r(c) \\ V_r &= \bar{A}_1(c)\bar{A}_1(-c) \cdots \bar{A}_{r-1}(c)\bar{A}_{r-1}(-c)\bar{A}_r(c)A_r(-c). \end{aligned} \tag{3.7}$$

These events are mutually exclusive and therefore

$$\Pr\left\{ D_N > \frac{c}{N} \right\} = \sum_{r=1}^{N} \Pr\{U_r\} + \sum_{r=1}^{N} \Pr\{V_r\}. \tag{3.8}$$

From the very definitions we have the following two *fundamental relations*

$$\begin{aligned} \Pr\{A_k(c)\} &= \sum_{r=1}^{k} \Pr\{U_r\} \Pr\{A_k(c) \mid A_r(c)\} \\ &\qquad + \sum_{r=1}^{k} \Pr\{V_r\} \Pr\{A_k(c) \mid A_r(-c)\} \\ \Pr\{A_k(-c)\} &= \sum_{r=1}^{k} \Pr\{U_r\} \Pr\{A_k(-c) \mid A_r(c)\} \\ &\qquad + \sum_{r=1}^{k} \Pr\{V_r\} \Pr\{A_k(-c) \mid A_r(-c)\}. \end{aligned} \tag{3.9}$$

This is a system of $2N$ linear equations for the $2N$ unknowns Pr $\{U_r\}$ and Pr $\{V_r\}$ and we proceed to solve it by the method of generating functions.

By definition of x_k we have Pr $\{X_\nu < x_k\} = k/N$. The probability of the event $A_k(c)$ (that the same inequality holds for exactly $k + c$ different ν's) is therefore given by

$$\text{(3.10)} \qquad \Pr\{A_k(c)\} = B(N, k + c; k/N)$$

(cf. (2.2)). Similarly, it is readily verified that for $r \leq k$

$$\text{(3.11)} \qquad \Pr\{A_k(c) \mid A_r(c)\} = B(N - r - c, k - r; (k - r)/(N - r)).$$

and

$$\text{(3.12)} \qquad \Pr\{A_k(c) \mid A_r(-c)\} = B(N - r + c, k - r + 2c; (k - r)/(N - r)).$$

The last three equations hold also for $c < 0$. They can be written in a more convenient form in terms of the quantities

$$\text{(3.13)} \qquad p_k(c) = e^{-k} \frac{k^{k+c}}{(k + c)!}.$$

In fact

$$\text{(3.14)} \qquad \Pr\{A_k(c)\} = \frac{p_k(c)p_{N-k}(-c)}{p_N(0)}$$

$$\text{(3.15)} \qquad \Pr\{A_k(c) \mid A_r(c)\} = \frac{p_{k-r}(0)p_{N-k}(-c)}{p_{N-r}(-c)}$$

$$\text{(3.16)} \qquad \Pr\{A_k(c) \mid A_r(-c)\} = \frac{p_{k-r}(2c)p_{N-k}(-c)}{p_{N-r}(c)}$$

If these expressions are introduced into (3.9) the second factor in the numerator cancels. A further simplification is achieved on introducing new sets of unknowns

$$\text{(3.17)} \qquad u_r = \Pr\{U_r\} \frac{p_N(0)}{p_{N-r}(-c)} \qquad v_r = \Pr\{V_r\} \frac{p_N(0)}{p_{N-r}(c)}.$$

The fundamental equations (3.6) then reduce to

$$\text{(3.18)} \qquad \begin{aligned} p_k(c) &= \sum_{r=1}^{k} u_r\, p_{k-r}(0) + \sum_{r=1}^{k} v_r\, p_{k-r}(2c) \\ p_k(-c) &= \sum_{r=1}^{k} u_r\, p_{k-r}(-2c) + \sum_{r=1}^{k} v_r\, p_{k-r}(0). \end{aligned}$$

This system is of the convolution type and can therefore be solved by means of generating functions. The essential point is that the $p_k(c)$ are defined for all k and that the system (3.18) therefore determines the unknowns u_r and v_r for all $r > 0$. We put

$$\text{(3.19)} \qquad u(\lambda) = \sum_{k=1}^{\infty} u_k \lambda^k \qquad v(\lambda) = \sum_{k=1}^{\infty} v_k \lambda^k$$

and

$$p(\lambda; c) = N^{-\frac{1}{2}} \sum_{k=1}^{\infty} p_k(c)\, \lambda^k . \tag{3.20}$$

(The factor $N^{-\frac{1}{2}}$ serves to simplify formulas.) Then obviously

$$\begin{aligned} p(\lambda; c) &= u(\lambda)p(\lambda; 0) + v(\lambda)p(\lambda; 2c); \\ p(\lambda; -c) &= u(\lambda)p(\lambda; -2c) + v(\lambda)p(\lambda; 0). \end{aligned} \tag{3.21}$$

From here we find $u(\lambda)$ and $v(\lambda)$. Equation (3.17) then determines Pr $\{U_r\}$ and Pr $\{V_r\}$. Actually we are interested only in the two sums occurring in (3.8). We put

$$\xi_k = \frac{1}{p_N(0)} \sum_{r=1}^{k} p_{k-r}(-c)u_r , \qquad \eta_k = \frac{1}{p_N(0)} \sum_{r=1}^{k} p_{k-r}(c)v_r . \tag{3.22}$$

Again, the ξ_k and η_k are defined for all k (also $k \geq N$). From (3.17) we have

$$\sum_{r=1}^{N} \Pr \{U_r\} = \xi_N , \qquad \sum_{r=1}^{N} \Pr \{V_r\} = \eta_N , \tag{3.23}$$

and hence finally, by (3.8)

$$\Pr \{D_N > c/N\} = \xi_N + \eta_N . \tag{3.24}$$

In (3.22) we find again simple convolutions leading to products of the corresponding generating functions. Thus

$$\begin{aligned} \xi(\lambda) &= \sum_{k=1}^{\infty} \xi_k \lambda^k = \frac{u(\lambda)p(\lambda; -c)N^{\frac{1}{2}}}{p_N(0)} \\ \eta(\lambda) &= \sum_{k=1}^{\infty} \eta_k \lambda^k = \frac{v(\lambda)p(\lambda; c)N^{\frac{1}{2}}}{p_N(0)} . \end{aligned} \tag{3.25}$$

We now pass to a study of the limiting form of these generating functions as $N \to \infty$ and $c \to \infty$ in accordance with (3.2). Consider a fixed $t > 0$ and suppose that

$$\frac{k}{N} \to t. \tag{3.26}$$

From well-known properties of the Poisson distribution it follows then that

$$N^{\frac{1}{2}} p_k(c) \to (2\pi t)^{-\frac{1}{2}} \exp(- z^2/2t). \tag{3.27}$$

Accordingly, the continuity theorem of section 2 implies (as can be verified directly) that

$$\begin{aligned} p(e^{-s/N} ; zN^{\frac{1}{2}}) &\to (2\pi)^{-\frac{1}{2}} \int_0^{\infty} t^{-\frac{1}{2}} \exp(-ts - z^2/2t)\, dt \\ &= (2s)^{-\frac{1}{2}} \exp(-(2sz^2)^{\frac{1}{2}}). \end{aligned} \tag{3.28}$$

(the last integral is well known and can be evaluated by elementary methods; the square-root is always positive). We see in particular that the limiting form is the same for $p(\lambda; c)$ and $p(\lambda; -c)$. It follows therefore from (3.21) directly that

$$\lim_{N\to\infty} u(e^{-s/N}) = \lim_{N\to\infty} v(e^{-s/N}) = \frac{\exp(-(2sz^2)^{\frac{1}{2}})}{1+\exp(-(8sz^2)^{\frac{1}{2}})}. \tag{3.29}$$

Using this and the fact that $p_N(0) \to (2\pi N)^{-\frac{1}{2}}$ we conclude from (3.25) that

$$\begin{aligned}\lim_{N\to\infty} N^{-1}\xi(e^{-s/N}) &= \lim_{N\to\infty} N^{-1}\eta(e^{-s/N})\\ &= \left(\frac{2\pi}{2s}\right)^{1/2} \frac{\exp(-(8sz^2)^{1/2})}{1+\exp(-(8sz^2)^{1/2})} = \phi(s).\end{aligned} \tag{3.30}$$

Expanding $\phi(s)$ into a geometric series we get

$$\phi(s) = \left(\frac{2\pi}{2s}\right)^{1/2} \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \exp(-(8s\nu^2 z^2)^{1/2}). \tag{3.31}$$

From the evaluation of the integral in (3.28) we conclude that $\phi(s)$ is the Laplace transform of

$$f(t) = \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \exp(-2\nu^2 z^2/t). \tag{3.32}$$

The continuity theorem of section 2 in conjunction with (3.30) and (3.26) shows that

$$\lim_{N\to\infty} \xi_N = \lim_{N\to\infty} \eta_N = f(1). \tag{3.33}$$

In view of (3.24) this accomplishes the proof.

4. Proof of Theorem 4. This proof is simpler than the preceding one inasmuch as we are now interested only in the events $A_k(c)$ for $c > 0$. This time we define U_r as the event that k is the smallest subscript for which $A_k(c)$ occurs, that is, $U_r = \bar{A}_1(c)\bar{A}_2(c) \cdots \bar{A}_{r-1}(c)A_r(c)$; no analogue to the event V_r will be used. With the same notations as before (3.9) is replaced by

$$\Pr\{A_k(c)\} = \sum_{r=1}^{k} \Pr\{U_r\} \Pr\{A_k(c) \mid A_r(c)\}, \tag{4.1}$$

and hence (3.21) by

$$p(\lambda; c) = u(\lambda)p(\lambda; 0). \tag{4.2}$$

Here $p(\lambda; c)$ is the same as before, so that (cf. (3.29))

$$\lim_{N\to\infty} u(e^{-s/N}) = \exp(-(2sz^2)^{1/2}). \tag{4.3}$$

Again, the first equation (3.25) holds without change and therefore we get instead of (3.30)

$$\lim_{N\to\infty} N^{-1}\xi(e^{-s/N}) = \left(\frac{2\pi}{2s}\right)^{1/2} \exp(-(8sz^2)^{1/2}). \tag{4.4}$$

From (3.28) this is the Laplace transform of

$$f(t) = t^{-\frac{1}{2}} \exp(-2z^2/t). \tag{4.5}$$

As before we conclude that $\xi_N \to f(1)$, which accomplishes the proof.

5. Proof of Theorem 3. We have seen in section 3 that the intervals in which (3.3) holds are in a one-to-one correspondence with the events $A_k(c)$. Hence

$$E(A_N) = \Sigma \Pr\{A_k(c)\} + \Sigma \Pr\{A_k(-c)\}. \tag{5.1}$$

To evaluate the sums we use (3.10). If $N \to \infty$ and again $c = zN^{\frac{1}{2}}$, $k/N \to t$, then by the central limit theorem

$$B(N, k + c; k/N) \to \frac{\exp(-z^2/2t(1 - t))}{(2\pi Nt(1 - t))^{1/2}}. \tag{5.2}$$

It follows then from (3.10) that

$$N^{-1/2}\Sigma \Pr\{A_k(c)\} \to (2\pi)^{-1/2} \int_0^1 \{t(1 - t)\}^{-1/2} \exp(-z^2/2t(1 - t))\, dt. \tag{5.3}$$

Call the right hand member R. After the substitution $t = \sin^2(\phi/2)$ we find

$$\begin{aligned} \frac{dR}{dz} &= -8(2\pi)^{-1/2} z \int_0^{\pi/2} \sin^{-2}\phi \exp(-2z^2/\sin^2\phi)\, d\phi \\ &= 8(2\pi)^{-1/2} z \exp(-2z^2) \int_0^{\pi/2} \exp(-2z^2 \cot^2\phi)\, d(\cot\phi) \\ &= -2 \exp(-2z^2). \end{aligned} \tag{5.4}$$

Since $R \to 0$ as $z \to \infty$ we conclude that

$$R = 2 \int_z^\infty \exp(-2x^2)\, dx = \{1 - \Phi(2z)\}(2\pi)^{1/2}. \tag{5.5}$$

The same asymptotic estimate holds for the other sum in (5.1), and hence Theorem 3 is proved.

6. Proof of Theorem 2. Reorder the two samples in ascending order of magnitude and denote the rearranged samples by $(X_1^*, \cdots, X_m^*)$ and $(Y_1^*, \cdots, Y_n^*)$. When speaking of the graphs of the empirical distributions $S_m(x)$ and $T_n(x)$ we shall suppose that they have been completed by adding vertical segments so that the graphs become step-polygons. We shall put

$$\frac{m}{m + n} = p, \qquad \frac{n}{m + n} = q. \tag{6.1}$$

Then, according to (1.6) and (1.7)

$$\frac{p}{q} \to a, \qquad N = pn = qm. \tag{6.2}$$

Without loss of generality we shall suppose that

$$p \leqq q. \tag{6.3}$$

In order to carry over the proof of Theorem 1 it is necessary to define the events $A_k(c)$ in a judicious manner. For every integer $k > 0$ let ν_k be the number of variables X_ν which are smaller than Y_k. In other words, ν_k is defined as the integer for which

$$X_{\nu_k}^* < Y_k^* \leq X_{\nu_k+1}^*. \tag{6.4}$$

Finally put

$$a_k = \left[\frac{mk}{n}\right] = \left[\frac{p}{q}\,k\right] \tag{6.5}$$

where, as usual, $[x]$ denotes the greatest integer contained in x.

For $0 < k \leq n$ let $A_k(c)$ be the event that

$$\nu_k = a_{k+c}\,. \tag{6.6}$$

The possibility of applying the proof of section 1 depends on the following

LEMMA. *Whenever*

$$D_{m,n} > \frac{c}{n} > 0 \tag{6.7}$$

then at least one among the events $A_1(c), A_1(-c), \cdots, A_n(c), A_n(-c)$ *occurs. Conversely, if one of these events occurs then*

$$D_{m,n} > \left(c - \frac{q}{p}\right) \Big/ n. \tag{6.8}$$

PROOF. If (6.7) holds then either for some x_0

$$S_m(x_0) - T_n(x_0) > \frac{c}{n} \tag{6.9}$$

or the reversed inequality holds with c replaced by $-c$. It suffices to consider the case (6.9). For sufficiently large x we have $S_m(x) = T_n(x) = 1$. Hence the graphs of $S_m(x)$ and $T_n(x) + c/n$ must intersect at an abscissa $\xi > x_0$. The point of intersection lies necessarily on a horizontal segment of the graph of $S_m(x)$ and a vertical segment of $T_n(x) + c/n$. Hence there exists a k such that $\xi = Y_k{}^*$ and, moreover,

$$T_n(\xi\,-) + \frac{c}{n} < S_m(\xi) \leq T_n(\xi\,+) + \frac{c}{n}. \tag{6.10}$$

This amounts to saying that

$$\frac{k - 1 + c}{n} < \frac{\nu_k}{m} < \frac{k + c}{n}. \tag{6.11}$$

In view of (6.3) and (6.5) this relation implies (6.6).

Conversely, suppose that the event $A_k(c)$ occurs and let $c > 0$. Put again $\xi = Y_k^*$. Then, by definition,

$$S_m(\xi) = \frac{\nu_k}{m} = \frac{a_{k+c}}{m}, \qquad T_n(\xi) = \frac{k}{n}. \tag{6.12}$$

It follows that

$$S_m(\xi) > \frac{k+c}{n} - \frac{1}{m} = T_n(\xi) + \frac{c}{n} - \frac{1}{m}, \tag{6.13}$$

which in turn implies (6.8). This proves the lemma.

Theorem 2 is concerned with values of c such that $cn^{-1} = zN^{-\frac{1}{2}}$; in passing to the limit we must therefore put

$$c = z(n/p)^{\frac{1}{2}}. \tag{6.14}$$

Accordingly, the relations (6.7) and (6.8) are asymptotically equivalent and our lemma shows that, asymptotically, the probability of (6.7) is the same as the probability that at least one among the events $A_1(c), \cdots, A_N(-c)$ occurs. To evaluate this probability we proceed exactly as in section 3. The events U_r and V_r defined by (3.7) and the fundamental relations (3.9) hold again. However (3.10) – (3.12) have to be replaced by new evaluations.

It is easily seen that the probability that exactly r among the X_ν are smaller than Y_k^* is the same as the probability to extract exactly r white balls before the k-th black ball from an urn containing m white and n black balls (assuming that all orders are equally likely and that balls are not replaced). In this way one finds

$$\Pr\{A_k(c)\} = \frac{C(a_{k+c} + k - 1, k - 1)C(m + n - a_{k+c} - k, n - k)}{C(m + n, n)} \tag{6.15}$$

$$\Pr\{A_k(c) \mid A_r(c)\} = \frac{C(a_{k+c} - a_{r+c} + k - r - 1, k - r - 1)C(m + n - a_{k+c} - k, n - k)}{C(m + n - a_{r+c} - r, n - r)} \tag{6.16}$$

$$\Pr\{A_k(c) \mid A_r(-c)\} = \frac{C(a_{k+c} - a_{r-c} + k - r - 1, k - r - 1)C(m + n - a_{k+c} - k, n - k)}{C(m + n - a_{r-c} - r, n - r)}. \tag{6.17}$$

The second binomial coefficient in the numerator is common to the three expressions and cancels when the expressions are introduced into (3.9). These fundamental relations assume a more natural form if the occurring binomial coefficients are enlarged to terms of a binomial distribution. It is easily verified that the first of the equations (3.9) reduces to

$$\begin{aligned} &\frac{B(a_{k+c} + k - 1, k - 1; q)}{B(m + n, n; q)} \\ &= \sum_{r=1}^{k} \Pr\{U_r\} \frac{B(a_{k+c} - a_{r+c} + k - r - 1, k - r - 1; q)}{B(m + n - a_{r+c} - r, n - r; q)} \\ &\quad + \sum_{r=1}^{k} \Pr\{V_r\} \frac{B(a_{k+c} - a_{r-c} + k - r - 1, k - r - 1; q)}{B(m + n - q_{r-c} - r, n - r; q)}. \end{aligned} \tag{6.18}$$

The second equation is obtained on replacing the combination $k + c$ by $k - c$.

Instead of (3.17) we put

$$(6.19)\qquad \begin{aligned} u_r &= \Pr\{U_r\} \frac{B(m+n, n; q)}{B(m+n-a_{r+c}-r, n-r; q)} \\ v_r &= \Pr\{V_r\} \frac{B(m+n, n; g)}{B(m+n-a_{r-c}-r, n-r; q)}. \end{aligned}$$

Then (6.18) becomes

$$(6.20)\qquad \begin{aligned} &B(a_{k+c}+k-1, k-1; q) \\ &\quad = \sum_{r=1}^{k} u_r B(a_{k+c}-a_{r+c}+k-r-1, k-r-1; q) \\ &\qquad\qquad + \sum_{r=1}^{k} v_r B(a_{k+c}-a_{r-c}+k-r-1, k-r-1; q). \end{aligned}$$

This corresponds to the first equation in (3.18). Unfortunately (6.20) is not of the pure convolution type since $a_{k+c} - a_{r+c}$ and $a_{k+c} - a_{r-c}$ are not functions of the two variables $k - r$ and c. The trouble comes from the fact that a_k, as defined by (6.5), is not a linear function of k. It is, however, plausible that we shall commit only an asymptotically negligible error if we omit the brackets in (6.5), that is, if we replace a_k by pk/q. Purely formally (6.20) then reduces to the first equation in (3.18) with

$$(6.21)\qquad p_k(c) = B\left(\frac{k+cp}{q} - 1, k-1; q\right).$$

(Here the first argument in the right hand member is no longer necessarily an integer, and the factorials in the definition (2.2) should be interpreted by means of the gamma function.) To the new system (3.18) the considerations of section 3 apply almost word for word: the only difference lies in the new norming (6.14) (which replaces (3.2)) and that instead of (3.26) we shall naturally let $k/n \to t$. Thus the limiting form of Theorem 1 applies to the new system (3.18) with $p_k(c)$ defined by (6.2).

It remains to prove that the formal replacement of (6.20) and the corresponding equation for $-c$, by (3.18) was legitimate. Now all coefficients in (6.20) are of the form $B(\nu, r; q)$, and we have only changed the first argument, ν, adding a variable quantity which in no case exceeds one unit. In passing to the limit we put $k \sim tn$ and $c \sim zn^{\frac{1}{2}}p^{-\frac{1}{2}}$. It follows that we actually use only coefficients $B(\nu, r; q)$ where $\nu \to \infty$, $r \to \infty$ and $\nu/r \to q$. Accordingly, for $|\vartheta| < 1$ we have $B(\nu + \vartheta, r; q) \sim B(\nu, r; q)$, and it is rather obvious that our system (6.20) is asymptotically equivalent to (3.18).

REFERENCES

[1] A. KOLMOGOROFF, "Sulla determinazione empirica di una legge di distribuzione," *Inst. Ital. Attuari, Giorn.*, Vol. 4 (1933), pp. 1–11.

[2] A. KOLMOGOROV, "Über die Grenzwertsätze der Wahrscheinlichkeitsrechnung," *Bulletin [Izvestija] Academie des Sciences URSS*, (1933), pp. 363–372.

[3] A. KOLMOGOROFF, "Confidence limits for an unknown distribution function," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 461–463.

[4] R. VON MISES, *Wahrscheinlichkeitsrechnung*, F. Deuticke Leipzig und Wien, 1931, p. 316 seq.

[5] N. SMIRNOV, "Ob uklonenijah empiričeskoi krivoi raspredelenija," *Recueil Mathematique (Matematičeskii Sbornik)*, N.S. Vol. 6(48) (1939), pp. 3–26.

[6] N. SMIRNOV, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin Mathématique de l'Université de Moscou*, Vol. 2 (1939), fasc. 2.