SOME DISTRIBUTION-FREE TESTS FOR THE DIFFERENCE BETWEEN TWO EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTIONS

By E. F. Drion

Statistics Department, T.N.O., The Hague

1. Summary and introduction. It sometimes happens that of two empirical cumulative distribution curves (step curves) one lies entirely above the other, in other words that, except at both ends, they have no point in common. The problem then arises, what is the probability that this will happen when both are random samples from the same population. In this paper a partial answer will be given, based on the ingenious solution of André (as cited in the well known textbook of Bertrand [1] in the problem of the ballot and also in Chap. VIII, Sect. 5 of [7]). Moreover an analogous method will allow us to give an exact answer to the problem of the maximum difference between two empirical cumulative distribution functions of random samples from the same population, but only if both samples have the same size. Smirnov has given an asymptotic solution for the latter problem (cited by Feller [3], see also [2]).

Our result leads, by using the Stirling approximation for the factorials, to the asymptotic formula of Smirnov.

A comparison of numerical results of the exact formula and the asymptotic formula of Smirnov shows that at least in the case of equal samples, the probabilities calculated by the Smirnov formula have, for samples as small as 20, an error of less than 4% for probabilities 0.033 or more. (See also Massey [5], who has calculated the exact probabilities for equal samples by means of difference equations.)

2. Statement of the problem. Let a population P be given with an unknown continuous distribution function $F(x)$. From this population two random samples $x'_1 \cdots x'_{n_1}$ and $y'_1 \cdots y'_{n_2}$ are drawn. After ordering each sample from the smallest value to the greatest we shall call them $x_1 \cdots x_{n_1}$ and $y_1 \cdots y_{n_2}$. For each sample the empirical distribution-function (step-function) $F_1(x)$ or $F_2(y)$ is constructed:

$$F_1(x) = 0, \qquad x < x_1, \qquad\qquad F_2(y) = 0, \qquad y < y_1,$$

$$F_1(x) = \frac{i}{n_1}, \qquad x_i \leqq x < x_{i+1}, \qquad\qquad F_2(y) = \frac{j}{n_2}, \qquad y_j \leqq y < y_{j+1},$$

$$F_1(x) = 1, \qquad x_{n_1} \leqq x, \qquad\qquad F_2(y) = 1, \qquad y_{n_2} \leqq y.$$

As we have assumed that the population has a continuous distribution-function, $\Pr(x_i = y_j) = 0$ for all sets of values of i and j; that is, the discontinuities of the two step-functions have, except for a probability zero, unequal abscissae.

Under these assumptions we ask for:

A. The probability that either $F_1(x) - F_2(x) < 0$ or $F_1(x) - F_2(x) > 0$ for all values of x between min (x_1, y_1) and max (x_{n_1}, y_{n_2}) (boundaries not included).

B. The probability that $\max | F_1(x) - F_2(x) | \geqq d$.

We shall give a general solution of problem A both for the case that $n_1 = n_2$ and that the greatest common divisor of n_1 and n_2 equals one. For problem B a solution has only been found for the case that $n_1 = n_2$.

3. Graphical representation of two ordered samples. If we order the observations of both samples in one series according to their magnitude, so that we shall have a series of $n_1 + n_2$ terms of the form $x_1, x_2, y_1, x_3, y_2, \cdots, y_{n_2}$ say, then our problem A is equivalent to the following: What is the probability that, in a random series of n_1 x's and n_2 y's, the proportion of x's to y's from the first to the n-th term of the series (where n may have all values from 2 to $n_1 + n_2 - 1$ included) is, for each n, always smaller than n_1/n_2 or always larger than n_1/n_2.

That both problems are equivalent may be shown in this way. If the two series of observations are random samples from the same population, they may be considered as one sample of size $n_1 + n_2$, in which n_1 observations are marked x and n_2 are marked y. The marking of the observations does not depend (in random samples) on the result of the observations, so all orders of the x's and y's are equally probable.

To solve this problem we shall make use of a graphical representation of these series. Let the x's represent horizontal paces and the y's vertical paces, then all possible series will be represented by all possible routes joining the diagonal corners of a rectangular lattice of sides n_1 and n_2. Those routes which have no common point (except the end-points), with the diagonal of our rectangle, represent series where the proportion of x's to y's is either always larger than n_1/n_2 or always smaller.

As an illustration we shall give the step-curves and the routes in the lattice for two series, in one of which the step-curves do not have a point in common, (and where, therefore, the route in the lattice lies entirely at one side of the diagonal) while in the other the step-curves intersect[1]. The sequence of ordered samples in Fig. 1 is (roman type denoting x's and italic denoting y's) 2.0, 2.3, 2.4, *2.6*, 2.7, 2.9, 3.0, *3.1*, *3.3*, 3.4, *3.6*, 3.8, *4.1*. The sequence in Fig. 2 is 2.0, 2.3, *2.5*, 2.6, *2.8*, 2.9, 3.1, *3.2*, 3.4, 3.5, *3.6*, *4.3*, *4.5*.

The number of all possible routes from O to P is $\binom{n_1 + n_2}{n_1} = T$. We shall now calculate the number A of all routes A[2] from O to P lying below the diagonal OP. The fraction A/T gives then the probability that of two empirical cumulative distribution curves of samples from one population the second lies entirely above the first. As each of the samples may be chosen as the first, the probability of no intersections of the step curves will be $2A/T$.

---

[1] It will be clear that if the paces in both directions have unit length, the route divides the rectangle in two parts of which the area's are respectively U and $n_1 n_2 - U$, where U is the statistic defined by Mann and Whitney for the test of Wilcoxon [4].

[2] We use A as well to indicate a route lying entirely to the right of the diagonal as to indicate the number of these routes.

The number A of routes lying below the diagonal OP depends on the number of lattice-points on OP that is to say, on the greatest common divisor of n_1 and n_2. If $n_1 = n_2 = n$ all routes reaching the diagonal will reach it in a lattice-point, as no route can intersect the diagonal except in a lattice-point. If n_1 and n_2 are coprime there are no lattice-points on the diagonal (except the endpoints O and P), while if n_1 and n_2 ($n_1 \neq n_2$) have a greatest common divisor $d > 1$, there

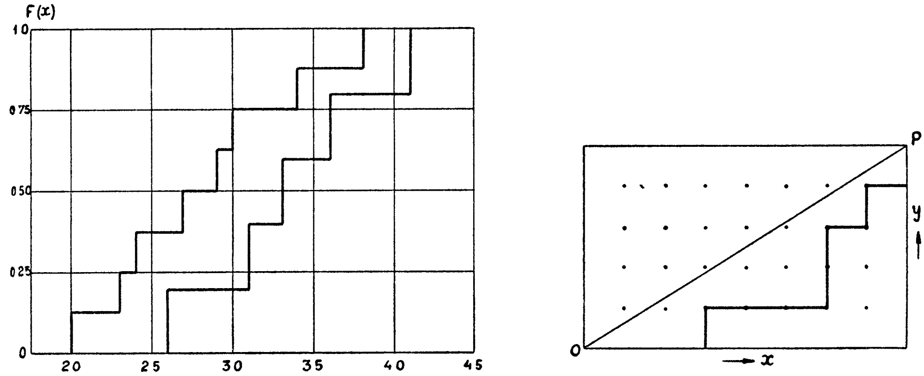

FIG. 1

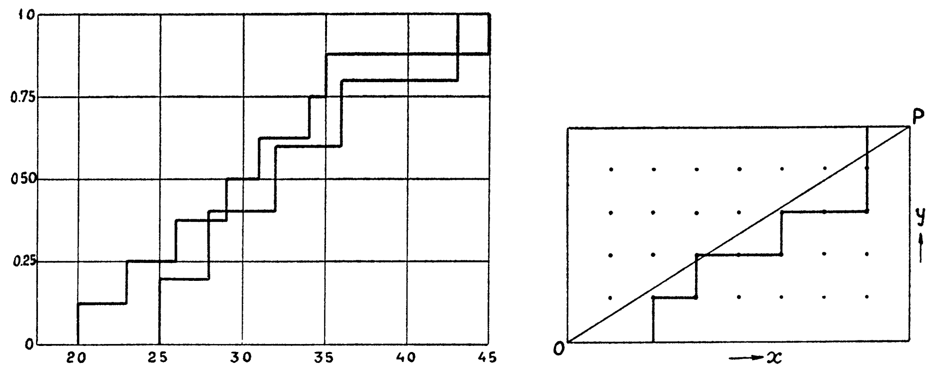

FIG. 2

are $d - 1$ lattice-points on the diagonal between O and P; so on $n_1 - d$ points a vertical route section and on $n_2 - d$ points a horizontal route section can intersect the diagonal outside a lattice point.

So the lattice-points available for a route under the diagonal OP is relatively to the total number lattice-points highest if n_1 and n_2 are coprime and lowest if $n_1 = n_2$. It stands to reason that the number of routes A is in the first case higher than in the second case. This we shall prove. For the intermediary case (greatest common divisor d of n_1 and $n_2 > 1$) we shall prove that the number of routes A relative to the total number of routes T is always less than when n_1 and n_2 are coprime. Probably this number is always higher than when $n_1 = n_2$. But we were not able to prove it.

4. Determination of the number of routes A in the case $n_1 = n_2 = n$. In this case (Fig. 3) the lattice is a square with $(n + 1)^2$ points. We shall not determine the number of routes A directly, but first we shall determine the number of routes that start with a horizontal step OR (and so could belong to the class A) having at least one point in common with the diagonal. It will be proved that this number equals twice the number of routes starting with a horizontal step and ending with a horizontal step. The proof given is essentially the proof found by André.

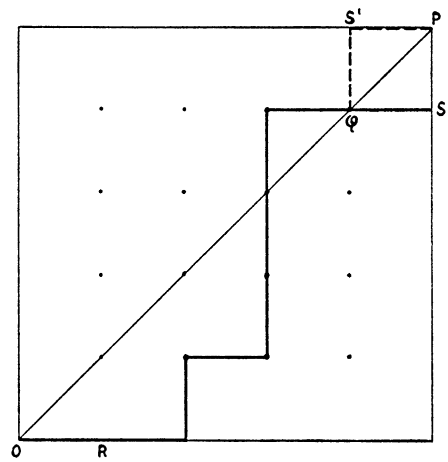

FIG. 3

The last step of a route "not-A", which starts with OR, can either be $S'P$ or SP. Routes ending with $S'P$ must cross the diagonal OP and are therefore routes "not-A"; their number is $\binom{2n-2}{n}$.

To prove that the number of routes "not-A" ending in SP equals the number of routes ending in $S'P$ we shall show that there exists a one-one correspondence between the routes "not-A" ending in SP and the routes ending in $S'P$. A route "not-A" like ORQSP can be transformed in a route ending in $S'P$ by rotating the part QSP about OP to $QS'P$. Here the point Q is the last point on the route before P that lies on the diagonal OP; each route "not-A" ending in SP can therefore be transformed in one way only in a route ending in $S'P$. On the other hand each route beginning with OR and ending with $S'P$ will cross at least once the diagonal OP. By rotating about the diagonal OP that part of the route, which lies between P and the point Q where it reaches for the first time OP, it will be transformed in a route "not-A" ending in SP. This route "not-A" ending in SP is also uniquely determined by the route ending in $S'P$. So we have proved the one-one correspondence between the routes "not-A" ending in SP and the routes "not-A" ending in $S'P$. The total number of routes "not-A" starting with OR is therefore $2\binom{2n-2}{n}$.

The total number of routes starting with OR is $\binom{2n-1}{n}$ therefore the num- of routes A is

$$\binom{2n-1}{n} - 2\binom{2n-2}{n} = \left(\frac{2n-1}{n-1} - 2\right)\binom{2n-2}{n} = \frac{1}{n-1}\binom{2n-2}{n}.$$

The total number T of routes from O to P is $\binom{2n}{n}$. So the probability that a route chosen at random lies either entirely to the right or entirely to the left of the diagonal equals

$$\frac{2 \times \frac{1}{n-1}\binom{2n-2}{n}}{\binom{2n}{n}} = \frac{\frac{2}{n-1}\binom{2n-2}{n}}{\frac{2n}{n}\cdot\frac{2n-1}{n-1}\binom{2n-2}{n}} = \frac{1}{2n-1}.$$

The probability that the cumulative frequency curves from two random samples n of the same population have no points in common (except the endpoints) is therefore $(1/2n - 1)$.

5. Determination of the number of routes A in the case n_1 and n_2 coprime. In this case (Fig. 4) there are no lattice-points on the diagonal except the endpoints, and if through any lattice-point (except the endpoints) a line parallel to the diagonal is drawn no other lattice-point will lie on this line; for if there were two lattice-points x_1y_1 and x_2y_2 on this line, then the triangle with angles (x_1y_1), (x_2y_2) and (x_2y_1) would be similar to the triangle $(0, 0)$; (n_1, n_2) and $(n_1, 0)$; so $(y_2 - y_1)/(x_2 - x_1) = n_2/n_1$, where $(y_2 - y_1)$ and $(x_2 - x_1)$ are integers smaller than n_2 respectively n_1. But this is impossible, as n_1 and n_2 are coprime.

A route A like OQP passes through $n_1 + n_2 - 1$ lattice-points (O and P excluded). If this route is cut in any of those lattice-points (like Q) and the two parts are interchanged the new route will not be a route A, that is to say it will not lie entirely to the right of the diagonal OP. For the angle PQC' is greater than the angle POC, so that if Q is placed in O then P will lie in a point Q' to the left of OP. Furthermore a straight line through Q' parallel to OP will not intersect anywhere the polygon $OQ'P$; the part OQ' is not intersected because OP does not intersect the part QP of the original line and $Q'P$ is not intersected because OP does not intersect OQ (for OQP is a route A that is, by definition a route, not intersected by OP). If we cut the route $OQ'P$ in Q', (which point is uniquely determined as being the first point lying on a line parallel to OP moved from D to P) and interchange the two parts OQ' and $Q'P$, the original route OQP will be reconstructed. On each route $OQ'P$ which passes through at least one lattice-point Q' at the left-hand side of OP and only on these routes, one, and only one, point Q' can be found, therefore a route $OQ'P$ (not-A) gives after transformation only one route $OQP(A)$. On the other hand, two different cuts of a route A will give after transformation two different routes, because if the coordinates of the section-points be (x_1, y_1) respectively (x_2, y_2), the coordinates of the images Q' respectively Q'_2 of O and P will be $(n_1 - x_1, n_2 - y_1)$ respectively $(n_1 - x_2, n_2 - y_2)$, which points are different. As each route "not-A" has only one point Q', two routes with different points Q' are different. It is also impossible that two *different* routes "A" give after section the same route "not-A", because the transformation of a route "not-A" to a route "A" is unique. As all routes lie either entirely to the right of OP (are routes "A") or have at least one point to the right of OP, and as each route A gives by the $(n_1 + n_2 - 1)$ possible cuts $(n_1 + n_2 - 1)$ different routes "not-A", the total number T of routes from O to P equals $A + (n_1 + n_2 - 1)A = (n_1 + n_2)A$. Therefore the probability that a randomly chosen route is a route A equals $1/(n_1 + n_2)$. The probability that two empirical cumulative distribution-curves, from two samples of size n_1 and n_2 (n_1 and n_2 coprime) from the same population do not intersect, is therefore $2/(n_1 + n_2)$.

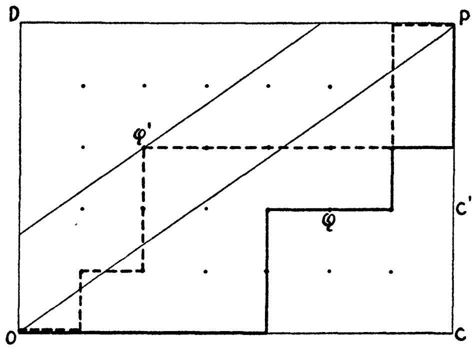

FIG. 4

6. Determination of the number of routes A in the case n_1 and n_2 have a common divisor greater than 1. If n_1 and n_2 ($n_1 \neq n_2$) have a greatest common divisor $d > 1$, there are $d - 1$ lattice-points on the diagonal (except the end-points). In this case the routes "not-A" can be divided into two groups: "not-A_1", routes which pass through at least one lattice-point at the lefthandside of the diagonal, and "not-A_2", routes which pass through one or more lattice-points on the diagonal but do not pass through a lattice-point at the lefthandside of the diagonal. A cut followed by an interchange of the two halves of a route "A" will transform it into a route "not-A_1". A cut followed by an interchange of the two halves of a route "not-A_2" will transform it either into a route "not-A_1", or into another or the same route "not-A_2" (if the cut falls on the diagonal). So the total number of routes is $(n_1 + n_2 - 1)\,A + A$ + routes "not-A_2" + routes "not-A_1", resulting from cuts in routes "not-A_2", $= (n_1 + n_2)\,A + x$. The number of routes A is therefore less than the $1/(n_1 + n_2)$th part of all the routes.

It may seem rather strange that the probability in the case n_1, n_2 coprime is about twice the probability in the case $n_1 = n_2$. This discontinuity is of course caused by the fact that in the case $n_1 = n_2$ both distribution-curves may have one or more points in common (except the endpoints) without crossing each other (in other words the graph may meet the diagonal without crossing it). In the case that n_1 and n_2 are coprime this is impossible.

If in the case $n_1 = n_2$ we seek the probability that either $F_1(x) - F_2(x) \leqq 0$ or $F_1(x) - F_2(x) \geqq 0$ (instead of $F_1(x) - F_2(x) < 0$ or $F_1(x) - F_2(x) > 0$) it will be found (by applying formula (1) of the next section with $h = 1$) that this probability is $2/(n + 1)$. Therefore under these conditions the probability for $n_1 = n_2$ is about twice as large as for n_1 coprime to n_2.

In consequence no direct statistical test can be based on these results, as one of the referees remarked. However should one in an investigation find that of two empirical distribution-curves, one lies entirely above the other, the formulas given above enable one to calculate the probability that such a result is caused by random sampling fluctuations.

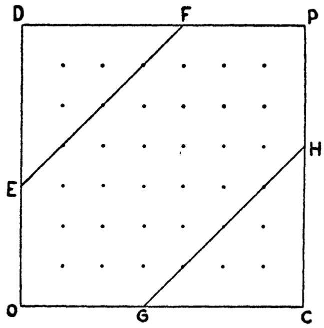

FIG. 5

7. Probability that the maximum difference of two empirical distribution curves from two samples of size n from one population is at least h/n. To solve this problem (exact solution of the problem of Smirnov in the case of equal samples) we shall again use the representation of our two samples by the lattice OCPD (Fig. 5). As $n_1 = n_2 = n$, this lattice is a square. All routes from O to P that reach a point on one of the lines EF, GH or on both lines, or that intersect one or both of these lines represent pairs of samples, where for some value x the maximum difference $| F_1(x) - F_2(x) |$ is at least $OE/PC = OE/n$.

To solve this problem we need the following lemma.

LEMMA. *The number of routes in a rectangular lattice with sides n_1 and n_2, such that somewhere the number of vertical paces y exceeds the number of horizontal paces x by at least h is* $\binom{n_1 + n_2}{n_1 + h}$. (An algebraic solution of this problem is given in Whitworth Proposition XXIX [8]. We shall give here a geometrical solution that can be extended to the problem of Smirnov.)

PROOF. All routes, such that somewhere the number of vertical paces exceeds the number of horizontal paces by at least h, are routes that reach or intersect a line EF, which makes an angle of 45° with DO (fig. 6).

We shall cut a route, such as OGP' that somewhere reaches the line EF, in the point G where it reaches this line for the first time. The part OG is reflected about EF to $O'G$; the part GP' is left in its place.

A route from O to P', reaching or intersecting EF, may thus be transformed in one way in a route from O' to P'. As we may transform the route $O'P'$ back to the route OGP' by cutting it in the first place where it reaches EF, and reflecting $O'G$ about EF, we see that there is a one to one correspondence between the routes from O to P' reaching or intersecting EF and the routes $(O'GP')$ from O' to P'. If the sides of the lattice measure n_1 $(= OC')$ and n_2 $(=C'P')$ and if OE

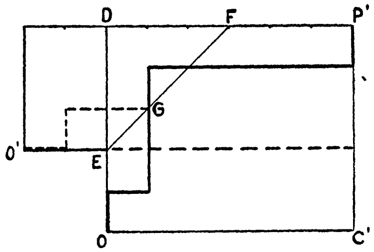

FIG. 6

$= O'E$ measures h, then the number of routes from O' to P and therefore the number of routes from O to P reaching or intersecting the line EF is

$$(1) \qquad \binom{n_1 + h + n_2 - h}{n_1 + h} = \binom{n_1 + n_2}{n_1 + h}$$

7.1. *Classification of routes OP representing empirical distributions where* max $| F_1(x) - F_2(x) | \geqq h/n$. The solution of the problem of Smirnov is, even in the case of equal samples, rather complicated, while the empirical distribution curves $F_1(x)$ and $F_2(x)$ may intersect more than once. Therefore it is possible that there are one or more values of x such that $F_1(x) - F_2(x) \geqq h/n$, and in the same pair of samples other values of x, such that $F_2(x) - F_1(x) \leqq h/n$. In other words, the route may intersect or touch both lines $a = EF$ and $b = GH$. We may classify the routes which touch or cross either one or both of the lines a and b in the following way (Fig. 5):

A. routes touching or crossing only a,

B. routes touching or crossing only b,

C. routes touching or crossing first one or more times a and afterwards touching or crossing b (after having touched or crossed b, these routes may also touch or cross a again).

D. routes touching or crossing first one or more times b and afterwards touching or crossing a (after having touched or crossed a, these routes may also touch or cross b again). The letters A, B, C and D will also be used for the number of routes of these categories. In the same way we will use the letters a and b also for the number of all routes that cross a respectively b, whether they cross b

(respectively a) or not. By reasons of symmetry it is clear that $a = b$, $A = B$ and $C = D$. Furthermore $a = A + C + D$, $b = B + C + D$. The total number of routes touching or crossing either a or b or both is $A + B + C + D = a + b - (C + D) = 2(a - D)$.

The number of routes a is given by our lemma viz. $\binom{2n}{n - h}$, so we have only to find the number of routes D.

7.2. *Calculation of the number of routes D.* The number of routes D may be ound by a repeated application of the device used for calculating a (Fig. 7).

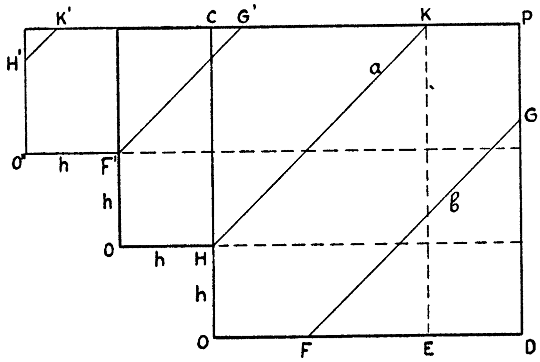

FIG. 7

We rotate the rectangle $OEKC$ about HK, leaving that part of the route from O to P unchanged which begins at the point where this route touches or crosses for the first time HK. After the transformation the routes D are the routes from O' to P which touch or cross $F'G'$, *without having first crossed the line HK.* The lengths of the sides of the rectangle $O'P$ (indicating this rectangle by the ends of one of the diagonals) are $n - h$ and $n + h$. To determine the total number of routes touching or crossing $F'G'$, we rotate the rectangle $O'G'$ about $F'G'$, leaving unchanged that part of the route from O' to P which begins at the point where this route touches or crosses for the first time $F'G'$. The new rectangle will have sides $n - 2h$ and $n + 2h$. The total number of routes in the rectangle $O''P$ is $\binom{2n}{n - 2h}$, this is therefore the total number of the routes in the rectangle $O'P$ which touch or cross the line $F'G'$. To get the number of routes D we must subtract from this number $\binom{2n}{n - 2h}$ the number of routes in $O''P$ touching or crossing the image $H'K'$ of HK in $O''P$, *without* having touched or crossed $F'G'$. By rotating the rectangle $O''K'$ about $H'K'$, we get a new rectangle with sides $n - 3h$ and $n + 3h$. The total number of routes in this rectangle is $\binom{2n}{n - 3h}$; the sought number of routes to subtract from $\binom{2n}{n - 2h}$ among these are the routes which do not touch or cross the image of $F'G'$, which can be determined by repeating the process of rotating. The law for the determination of the number of routes will be clear. Their number is: $\binom{2n}{n-2h} - \binom{2n}{n-3h} + \binom{2n}{n-4h} - \cdots$, the series being continued as long as $n - kh \geqq 0$.

The total number of routes which cross either one or both lines HK and FG is therefore

$$2\left[\binom{2n}{n-h} - \binom{2n}{n-2h} + \binom{2n}{n-3h} \cdots\right].$$

As all the routes from O to P number $\binom{2n}{n}$ the probability that a random chosen route touches or crosses HK or FIG or both is

$$\frac{2\left[\binom{2n}{n-h} - \binom{2n}{n-2h} + \cdots\right]}{\binom{2n}{n}}.$$

This is therefore also the probability that the maximum difference of the cumulative frequency-curves of two random samples from the same population is at least h/n.

TABLE I

n	d	$h = nd$	P	
			Exact	Smirnov
20	.25	5	.5713	.5596
20	.40	8	.0811	.0815
20	.45	9	.0335	.0349
20	.50	10	.0123	.0135
50	.16	8	.5487	.5441
50	.24	12	.1124	.1123
50	.28	14	.0392	.0396
50	.32	16	.0115	.0120
100	.12	12	.4695	.4676
100	.17	17	.1112	.1112
100	.19	19	.0539	.0541
100	.23	23	.0099	.0101

8. Some numerical results. We have calculated the probability P that $\max | F_1(x) - F_2(x) | \geqq$ d for samples of size $n = 20, 50, 100$ and for values of d such that $P \sim 0.50$, 0.10, 0.05 and 0.01 by means of the exact formula and by the asymptotic formula of Smirnov. The results are given in table I.

The figures, given in the last column, were found by linear interpolation in the table of Smirnov [6].

With equal-sized samples the asymptotic formula of Smirnov gives very satisfactory results even for samples of 20. We suspect, however, that when the samples are of unequal size the agreement will be less satisfactory especially if n_1 and n_2 are coprime, because in this case there is only one lattice point on HK and FG, which must in this case be parallel to the diagonal OP (c.f. Fig. 7).

9. Concluding remarks. (a) The probabilities given above are based on the assumption that the distribution-functions of the population are continuous. In practice almost all distribution-functions, however, are discontinuous, owing to the limited accuracy of our measurements. In other words, in practice we work always with grouped data, although the classes may be so small, that in no class falls more than one observation and often none. Nevertheless, when the number of observations is large enough, more than one observation will be found in several classes.

Let the width of the classes be h, so that the values of $F(x)$ (i.e. the cumulative experimental distribution-function) are only known for $x = hg$ (with g an integer between $g_a = [x_1/h]$ and $g_b = [x_n/h] + 1$, where $[x/h]$ denotes the integer part of x/h). If of two ungrouped samples, $x_1, \cdots, x_{n_1}$ and $y_1, \cdots, y_{n_2}$, the cumulative experimental distribution curve of the y's lies entirely to the righthand side of that of the x's, i.e. if $F_1(x) > F_2(x)$, $a_1 = \min(x_1, y_1) \leqq x \leqq \max(x_{n_1}, y_{n_2}) = a_2$ then, after grouping,

$$F_1(hg_i) > F_2(hg_i), \qquad g_1 = \min\left(\left[\frac{x_1}{h}\right] + 1, \left[\frac{y_1}{h}\right] + 1\right) \leqq g_i$$
$$\leqq \max\left(\left[\frac{x_{n_1}}{h}\right], \left[\frac{y_{n_2}}{h}\right]\right) = g_n.$$

But the converse needs not hold. Therefore the probability that $F_1(hg_i) > F_2(hg_i)$ for all values of g_i between g_1 and g_n (g_1 and g_n included) is greater than or equal to the probability that $F_1(x) > F_2(x)$ for all values of x between a_1 and a_2 (a_1 and a_2 included).

If however $F_1(hg_i) > F_2(hg_{i+1})$, $g_1 \leqq g_i < g_n - 1$, then $F_1(x) > F_2(x)$, $a_1 \leqq x \leqq a_2$, although the converse needs not hold. Therefore the probability that $F_1(hg_i) > F_2(hg_{i+1})$ for all values of g_i between g_1 and $g_n - 1$ is less than or equal to the probability that $F_1(x) > F_2(x)$ for all values of x between a_1 and a_2 (a_1 and a_2 included).

From this last result the following conclusion may be drawn: the probability that in two grouped random samples from the same population the cumulative experimental frequencies of one of the samples is higher at all class boundaries (which are the only values of the variate for which the cumulative frequency is known) than the cumulative frequencies of the other sample at the class boundaries of the next higher class is less than or equal to the formulae given in Sections 4 and 5.

(b) The formula given in Section 7 for the Smirnov test applies to the two-sided test. In the case we are only interested in deviations in one direction the formula is much simpler. With equal-sized samples from the same population the probability that $F_1(x) - F_2(x) > d/n$ is $\binom{2n}{n-d} \Big/ \binom{2n}{n}$.

REFERENCES

[1] J. Bertrand, *Calcul des Probabilités*, Paris, 1907.
[2] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 393–403.
[3] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–189.
[4] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 50–60.
[5] Frank J. Massey, "The distribution of the maximum deviation between two sample cumulative step functions" *Annals of Math. Stat.*, Vol. 22 (1951), pp. 125–128.
[6] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 279–281.
[7] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw Hill Book Co., New York 1937.
[8] W. A. Whitworth, *Choice and Chance*, Cambridge, 1901.