The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 1–11

RECTANGLE PROBABILITIES FOR UNIFORM ORDER STATISTICS AND THE PROBABILITY THAT THE EMPIRICAL DISTRIBUTION FUNCTION LIES BETWEEN TWO DISTRIBUTION FUNCTIONS[1]

BY G. P. STECK

Sandia Laboratories

1. Introduction. The principal result of this paper is a simple determinant for the probability that the order statistics from a sample of uniform random variables all lie in a multi-dimensional rectangle. Immediate applications of this result give: (i) the probability that the empirical distribution function lies between two other distribution functions; (ii) very general confidence regions for an unknown continuous distribution function; (iii) the power of tests based on the empirical distribution function. These applications, and others, are discussed in Section 4.

Let $X_1, X_2, \cdots, X_m$ be independent random variables with a continuous distribution function F and let F_m denote the empirical distribution function. Let

(1.1) $$P_m(gF, hF \mid F) = P(g\{F(x)\} \leqq F_m(x) \leqq h\{F(x)\}, \text{ for all } x \mid F),$$

where g and h are distribution functions on [0, 1] with g continuous to the left and h continuous to the right.

Since the random variables $F(X_i)$ are uniform random variables with empirical distribution function $F_m F^{-1}$, it follows that

$$P_m(gF, hF \mid F) = P_m(g, h \mid F(x) = x) = P_m(g, h), \quad \text{say}.$$

Also, since $F_m F^{-1}$ passes through the points $(0, 0)$, $(U^{(1)}, 1/m)$, $(U^{(2)}, 2/m)$, $\cdots$, $(U^{(m-1)}, (m-1)/m)$, $(1, 1)$, it follows that

$$P_m(u, v) \equiv P_m(g, h) = P(u_i \leqq U^{(i)} \leqq v_i, \ i = 1, 2, \cdots, m)$$

where $U^{(1)}, \cdots, U^{(m)}$ are the order statistics from a sample of m independent uniform random variables, $u_i = h^{-1}(i/m)$ and $v_i = g^{-1}((i-1)/m)$, $i = 1, 2, \cdots, m$.

In this paper we show that $P_m(u, v)$ is a determinant whose ijth element is $\binom{j}{j-i+1}(v_i - u_j)_+^{j-i+1}$ or 0 according as $j-i+1$ is nonnegative or negative and $(x)_+ = \max(0, x)$. Thus the determinant is of Hessenberg form with ones on the first subdiagonal and zeros below the first subdiagonal.

After this paper had been accepted for publication, I found that this result had been anticipated by Epanechnikov (1968), who proved an equivalent recurrence as a tool for studying the power of the Kolmogorov one-sample test.

With few exceptions, most notably Epanechnikov (1968), all the previous results concerning $P_m(g, h)$ have been for the special case where g and h are linear. If $g(x) = \max(0, ax-b)$ and $h(x) = \min(1, cx+d)$, with $a, b, c, d \geqq 0$, let $P_m(g, h)$ be denoted by $P_m(a, b; c, d)$. In particular, if $a = c = 1$ then $P_m(1, b; 1, d) = P(D_m^+ \leqq b, D_m^- \leqq d)$, where D_m^+ and D_m^- are the two one-sided Kolmogorov statistics.

Received November 21, 1969.

[1] This work was supported by the United States Atomic Energy Commission.

Kolmogorov (1933) derived a system of recurrences for $P_m(1, b/m; 1, b/m)$ where b is an integer. Wald and Wolfowitz (1939) succeeded in expressing $P_m(0, 0; 1, d)$ and $P_m(1, b; 0, 1)$ as determinants (of differing structure), but, unaware of Kolmogorov's result, they were unable to derive a similar expression for $P_m(1, b; 1, c)$. Birnbaum and Tingey (1951) found a closed form for $P_m(0, 0; 1, d)$ as an incomplete Abel sum; Dempster (1959) and, independently, Dwass (1959) did the same for $P_m(0, 0; c, d)$. (An interesting discussion of Abel's generalization of the binomial theorem appears in Riordan (1968).) Kemperman (1957) derived the generating function for $P_m(1, b; 1, c)$. Durbin (1968) derived the generating function for $P_m(a, b; a, d)$, gave the associated recurrence, and showed that when $a \geqq 1$, ma an integer, the generating function takes a simple form similar to the one obtained by Kemperman. In Section 3 we will show that Durbin's generating function is valid more generally.

Nonlinear problems, chiefly in the one-sided case, have been considered by several authors. Suzuki (1967), (1968), showed that the Wald and Wolfowitz recurrence for $P_m(0, 0; 1, d)$ is also valid as a recurrence for $P_m(0, hF | F)$. He also gives $P_m(\cdot, \cdot)$ in terms of $P_m(0, \cdot)$, but this formula appears at least as complicated as the trivial one obtained by differencing $P_m(0, \cdot)$. Anderson and Darling (1952) studied the asymptotics of the special cases corresponding to $g_\psi(x) = x - a(m\psi(x))^{-\frac{1}{2}}$ and $h_\psi(x) = x + a(m\psi(x))^{-\frac{1}{2}}$ for general nonnegative ψ. In particular, when $\psi(x) = x(1-x)$, they express $\lim_{m\to\infty} P_m(g_\psi, h_\psi)$ in terms of absorption probabilities for a Uhlenbeck process. These probabilities are given explicitly by Malmquist (1954) in the one-sided case. Whittle (1961) treats the evaluation of $P_m(0, v)$ formally by generating functions and obtains exact expressions for certain special cases. More recently, Noé and Vandewiele (1968) develop interesting recursions for $P_m(0, v)$ and for a special case of $P_m(u, v)$. They consider in some detail the one-sided version of the example we consider in Section 4.3. Lientz (1968) and Birnbaum and Lientz (1969a), (1969b), found exact expressions for variants of $P_m(0, v)$ corresponding to the distributions of certain Rényi and Kac type statistics.

2. A determinant for $P_m(u, v)$. The following proof replaces my original proof which was to pass to the limit in a corresponding two-sample result due to Steck (1969). This proof is more informative and can possibly be used to generalize the two-sample result to the non-null case. Other simple proofs offered as replacements for my original proof are due to Mohanty (1970) and Walkup (1970).

Let $(x)_+ = \max(x, 0)$.

THEOREM. $P_m(u, v) = m! \det[(v_i - u_j)_+^{j-i+1}/(j-i+1)!]$.

PROOF. Let $U_1, U_2, \cdots, U_m$ be independent random variables uniformly distributed on [0, 1], and let $U^{(1)} \leqq U^{(2)} \leqq \cdots \leqq U^{(m)}$ be the order statistics. Let Λ and Ω denote the corresponding sample spaces; that is, $\Lambda = \{(x_1, x_2, \cdots, x_m) | 0 \leqq x_i \leqq 1, \text{ all } i\}$ and $\Omega = \{(x_1, x_2, \cdots, x_m) | 0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_m \leqq 1\}$. Let $F_k = \{(x_1, x_2, \cdots, x_k) | u_i \leqq x_i \leqq v_i, i \leqq k \text{ and } 0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_k \leqq 1\}$ denote the set of which $Q_k \equiv P_k(u, v)/k!$ is the probability.

Expanding the determinant of the theorem, call it D_m, by the last column, starting from the bottom, gives the recurrence

$$(2.1)\qquad D_m = (v_m - u_m)_+ D_{m-1} - \frac{(v_{m-1} - u_m)_+^{\,2}}{2!} D_{m-2} \pm \cdots + (-1)^{m+1} \frac{(v_1 - u_m)_+^{\,m}}{m!} D_0 .$$

We must now prove $Q_m = D_m$. The proof uses induction and a form of the principle of inclusion and exclusion and, given the induction hypothesis, consists of showing (2.1) holds because although the events, of which the terms of (2.1) are the probabilities, count the elementary events in which we are interested with differing multiplicities, the form of (2.1) assures each is counted only once.

First, to get the induction going we define $Q_0 = D_0 = 1$ and note that the theorem is trivially true for $m = 1$; hence $Q_1 = D_1$. If $m = 2$,

$$\begin{aligned} P_2(u,v) &= P(u_1 \leqq U^{(1)} \leqq v_1, u_2 \leqq U^{(2)} \leqq v_2) \\ &= P(u_1 \leqq U_1 \leqq v_1, u_2 \leqq U_2 \leqq v_2 \text{ or } u_1 \leqq U_2 \leqq v_1, u_2 \leqq U_1 \leqq v_2) \\ &= 2P(u_1 \leqq U_1 \leqq v_1, u_2 \leqq U_2 \leqq v_2) - P(u_2 \leqq U_1, U_2 \leqq v_1) \\ &= 2(v_1 - u_1)(v_2 - u_2) - (v_1 - u_2)_+^{\,2}; \end{aligned}$$

hence $Q_2 = D_2$.

Now, assume the induction hypothesis that $Q_k = D_k$ for $k = 1, 2, \cdots, m-1$. Let $I_k = [u_m, v_{m-k+1}]$ denote the intersection of the last k intervals $[u_i, v_i]([a, b] = \varnothing$ if $a > b)$. Let $B_k = \{(x_1, x_2, \cdots, x_m) \mid (x_1, x_2, \cdots, x_{m-k}) \in F_{m-k}$ and $x_i \in I_k$, $m-k+1 \leqq i \leqq m$ and $x_{m-k+1} \leqq x_{m-k+2} \leqq \cdots \leqq x_m\}$ denote the event where some k (ordered) of the $\{U_i\}$ belong to the intersection of the last k intervals and the remaining take values favorable to F_{m-k}. Then $P(B_k) = [(v_{m-k+1} - u_m)_+^{\,k}/k!]Q_{m-k}$ and, by the induction hypothesis, (2.1) takes the form

$$(2.2)\qquad D_m = P(B_1) - P(B_2) \pm \cdots + (-1)^{m+1} P(B_m).$$

Now let $\omega \in \bigcup_{r=1}^m B_k$; to it corresponds a unique largest integer, $k_0(\omega)$, such that $\omega \in B_k$ for $k \leqq k_0$. Let $A_r = \{\omega \mid k_0(\omega) = r\}$. We will show (2.2) holds for any $1 \leqq r \leqq m$.

Assume $\omega \in F_m \cap A_r$; that is, ω is favorable to the event of which Q_m is the probability. The event B_1 counts ω by definition, but B_1 also counts $r-1$ other points corresponding to which x_i of the r possibilities $x_{m-r+1}, \cdots, x_m$ is chosen to be the value of $U^{(m)}$; ω is in Ω, and the other points are in $r-1$ of the $m!-1$ different disjoint replicas of Ω determined by appropriate inversions of the basic ordering. This means that the integral of the indicator function of B_1 over $\Lambda \cap A_r$ is r times the same integral over $\Omega \cap A_r$; i.e., $P(B_1 \cap A_r) = rP(F_m \cap A_r)$. Similarly, B_2 counts ω by definition, but B_2 also counts $\binom{r}{2} - 1$ other points corresponding to which pair (x_i, x_j) is chosen to be the value of $(U^{(m-1)}, U^{(m)})$. Again, ω is in Ω, while the other points are in $\binom{r}{2} - 1$ of the replicas of Ω. This means that the integral of the indicator function of B_2 over $\Lambda \cap A_r$ is $\binom{r}{2}$ times the same integral over $\Omega \cap A_r$; i.e., $P(B_2 \cap A_r) = \binom{r}{2} P(F_m \cap A_r)$. And so forth, until we find that B_r counts ω exactly once so that $P(B_r \cap A_r) = \binom{r}{r} P(F_m \cap A_r)$. Since $\binom{r}{1} - \binom{r}{2} \pm \cdots \pm \binom{r}{r} = 1$ for any r we can see that, for any $1 \leqq r \leqq m$,

$$P(F_m \cap A_r) = P(B_1 \cap A_r) - P(B_2 \cap A_r) \pm \cdots \pm P(B_r \cap A_r).$$

Summing over r shows that the RHS of (2.2) is Q_m and hence $D_m = Q_m$ and the theorem is proved.

If $u_j = 0$, all j, then the determinant given by the theorem for $P_m(0, v)$ is the same as the one given by Wald and Wolfowitz ((1939), (29)); however, if $v_i = 1$, all i, the determinant given by the theorem for $P_m(u, 1)$ is $m \times m$ and is in terms of $1 - u_j$, while the one given by Wald and Wolfowitz ((1939), (27)) is $(m+1) \times (m+1)$ and is in terms of u_j.

An alternative statement of the theorem also exists which "distributes" the $m!$ into the elements of the matrix. It follows trivially from (2.1).

COROLLARY. $P_m(u, v) = \det[\binom{j}{j-i+1}(v_i - u_j)_+^{\,j-i+1}]$.

In the one-sided case when, say $u_j = 0$ all j, it is possible to prove a recurrence which is computationally superior to (2.1) in that it does not have terms with alternating signs. It is given in the following corollary.

COROLLARY. *If* $P_0 \equiv 1$ *and* P_j *denotes* $P_j(0, v)$ *then*

$$P_k = v_k^{\,k} - \textstyle\sum_{j=0}^{k-2} \binom{k}{j} (v_k - v_{j+1})^{k-j} P_j, \qquad k = 1, 2, \cdots, m. \tag{2.3}$$

PROOF. Multiply the jth column of the determinant given by the theorem for P_k by $(-v_k)^{k-j}/(k-j)!$, sum over columns and put in the kth column. This has the effect of replacing $v_i^{\,k-i+1}/(k-i+1)!$ by $(v_i - v_k)^{k-i+1}/(k-i+1)!$ except for $v_1^{\,k}/k!$ which becomes $(v_1 - v_k)^k/k! - (-v_k)^k/k!$. The recurrence follows by expanding the determinant by the last column.

It has been found that direct evaluation of $P_m(u, v)$ in either determinantal form by elimination or expansion is accurate in single precision on the CDC 6600 (60 bit word length) only up to about $n = 25$. At $n = 50$ even double precision is inadequate. However, the recurrence (2.3) for one-sided probabilities is accurate in single precision at least to $n = 100$ where it was used to check Table 1 of Noé and Vandewiele (1968).

3. A determinant for $P_m(a, b; c, d)$. In this important special case $g(x) = \max(0, ax - b)$ and $h(x) = \min(1, cx + d)$ so that $u_j = 0$ or $(j - md)/mc$ according as $j \leqq md$ or $j > md$ and $v_i = 1$ or $(i - 1 + mb)/ma$ according as $i \geqq m(a-b)+1$ or $i < m(a-b)+1$. Thus

$$P_m(a, b; c, d) = m! \det(r_{ij}^{\,j-i+1}/(j-i+1)!) \tag{3.1}$$

where $r_{ij} = (v_i - u_j)_+$.

If $a \leqq bc + ad$ and $\langle m(a-b) \rangle \leqq [md]$, where $[x]$ is the integer part of x and $\langle x \rangle = -[-x]$ is the smallest integer greater than or equal to x, it is impossible for F_m to cross both lines (and vice versa); hence, for this case

$$P_m(a, b; c, d) = 1 - Q(m, \delta_1, \varepsilon_1) - Q(m, \delta_2, \varepsilon_2), \tag{3.2}$$

where Q is the function derived by Dempster (1959, eq. (5′)) and $\delta_1 = d$, $\varepsilon_1 = (c+d-1)/a$, $\delta_2 = b/a$, $\varepsilon_2 = 1+b-a$.

In the special case of parallel lines, namely $a = c$, considered by Durbin (1968), it is possible to modify the matrix (r_{ij}), without changing its determinant, so that it becomes a bordered matrix whose interior matrix has equal entries on each diagonal. The lemma which follows shows that this kind of bordered matrix has a generating function which can be identified by inspection. We will see that the generating function of the $\{P_m\}$ derived by Durbin for the case $m(a-1)$ = non-negative integer is also valid for other values of a.

If $a = c$ and $\langle m(a-b)\rangle > [\langle nd\rangle]$ then $a > b+d$ (otherwise (3.2) applies) and $I \geqq J$, where $I = \langle m(a-b)\rangle$, $J = [md]$,

$$P_m(a,b;a,d) = m!\det\begin{pmatrix} A & \vdots & B \\ \cdots & & \cdots \\ O & \vdots & C \end{pmatrix}\begin{matrix} I \\ \cdots \\ m-I \end{matrix}, \quad \text{columns } J \;\vdots\; m-J$$

$$A = \left\{\begin{matrix} \left(\frac{mb}{ma}\right) & \frac{1}{2!}\left(\frac{mb}{ma}\right)^2 & \frac{1}{3!}\left(\frac{mb}{ma}\right)^3 & \cdots & \frac{1}{J!}\left(\frac{mb}{ma}\right)^J \\ 1 & \frac{mb+1}{ma} & \frac{1}{2!}\left(\frac{mb+1}{ma}\right)^2 & \cdots & \frac{1}{(J-1)!}\left(\frac{mb+1}{ma}\right)^{J-1} \\ 0 & 1 & \frac{mb+2}{ma} & \cdots & \frac{1}{(J-2)!}\left(\frac{mb+2}{ma}\right)^{J-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{matrix}\right\} I \times J,$$

$B =$

$$\left\{\begin{matrix} \frac{1}{(J+1)!}\left(\frac{mb+md-J-1}{ma}\right)^{J+1} & \frac{1}{(J+2)!}\left(\frac{mb+md-J-2}{ma}\right)^{J+2} & \frac{1}{(J+3)!}\left(\frac{mb+md-J-3}{ma}\right)^{J+3} & \cdots \\ \frac{1}{J!}\left(\frac{mb+md-J}{ma}\right)^{J} & \frac{1}{(J+1)!}\left(\frac{mb+md-J-1}{ma}\right)^{J+1} & \frac{1}{(J+2)!}\left(\frac{mb+md-J-2}{ma}\right)^{J+2} & \cdots \\ \frac{1}{(J-1)!}\left(\frac{mb+md-J+1}{ma}\right)^{J-1} & \frac{1}{J!}\left(\frac{mb+md-J}{ma}\right)^{J} & \frac{1}{(J+1)!}\left(\frac{mb+md-J-1}{ma}\right)^{J+1} & \cdots \\ \vdots & \vdots & \vdots & \end{matrix}\right\},$$

$$I \times (m-J)$$

and C' is a $(m-J)\times(m-I)$ matrix which will be treated the same way as A. Note that B has equal entries on each diagonal, since its entries depend only on $j-i$, and that (AB) would also if mb were replaced by $mb+md-J$ in the Jth column of A, mb were replaced by $mb+md-J+1$ in the $J-1$st column of A, etc.; i.e., if $md-j$ were added to mb in the jth column of A. This can be done (except for the first element in a column) by replacing each column of A by a linear combination of columns in the way illustrated by Steck ((1969), (4.2)) except that $\binom{u}{k}$ is replaced by $u^k/k!$ and the proof is by the binomial theorem instead of the Vandermonde convolution formula. In a similar way (except for the last element in a row) $i+mb-ma-1$ can be added to md in the ith row of C.

The net result of all this is the following: let $p = ma$, $q = mb$, $r = md$; then

$$(3.3)\quad \frac{(ma)^m}{m!} P_m(a, b; a, d) =$$

$$\begin{vmatrix} (q+r)_1-(r)_1 & (q+r)_2-(r)_2 & (q+r)_3-(r)_3 & \cdots & 0 & 0 & 0 \\ 1 & (q+r)_1 & (q+r)_2 & \cdots & 0 & 0 & 0 \\ 0 & 1 & (q+r)_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & (q+r)_1 & (q+r)_2-(m-p+q)_2 \\ 0 & 0 & 0 & \cdots & 0 & 1 & (q+r)_1-(m-p+q)_1 \end{vmatrix}_{m \times m}$$

where $(x)_r = \max(0, (x-r)^r/r!)$.

A straightforward generalization of a result on the ratio of series (see, for example, Adams (1957) page 119, (6.360)) is the following.

LEMMA. *Let the coefficients* $\{c_i\}$ *be defined by* $(1-b_1x+b_2x^2\pm\cdots)(1-d_1x+d_2x^2\pm\cdots)/(1-a_1x+a_2x^2\pm\cdots) = 1+c_1x+c_2x^2+\cdots$. *Then*

$$c_n = \begin{vmatrix} a_1-b_1 & a_2-b_2 & a_3-b_3 & \cdots & a_{n-1}-b_{n-1} & a_n-b_n-d_n \\ 1 & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1}-d_{n-1} \\ 0 & 1 & a_1 & \cdots & a_{n-3} & a_{n-2}-d_{n-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_1 & a_2-d_2 \\ 0 & 0 & 0 & \cdots & 1 & a_1-d_1 \end{vmatrix}_{n \times n}.$$

From this lemma it follows that the generating function of

$$\{(ma)^m/m!\}P_m(a, b; a, d)$$

for the case $\langle m(a-b)\rangle > [md]$ is

$$f(t) = g(t,r)\cdot g(t,m-p+q)/g(t,q+r), \tag{3.4}$$

where $g(t, x) = 1-(x-1)t+(x-2)^2t^2/2!\pm\cdots+(x-h)^h t^h/h!$, $h = [x]$. This is the same generating function found by Durbin (1968) for the case $m(a-1) =$ nonnegative integer.

Because of the relationship between the results of this section and the results of Durbin (1968), we present the following dictionary:

Steck	Durbin	Steck	Durbin
m	n	$p \equiv ma$	$n+c$
a	$(n+c)/n$	$q \equiv mb$	b
b	b/n	$r \equiv md$	a
d	a/n	$m-p+q$	$b-c$
$m(a-1)$	c	$g(t, x)$	$g(z, x)$.

4. Applications. Being able to compute $P_m(g, h)$ makes it possible to do many things. Only five possibilities are considered here.

4.1. *Power computations.* It is possible to compute the power of any test of $H\colon F = F_0$ against $A\colon F = f(F_0)$, where f is a continuous distribution function on [0, 1], based on an acceptance region of the form $\{\text{accept if } g\{F_0(x)\} \leqq F_m(x) \leqq h\{F_0(x)\} \text{ for all } x\}$. If g_α and h_α are chosen so that $P_m(g_\alpha, h_\alpha) = \alpha$, then the power of the test is $1-P_m(g_\alpha f^{-1}, h_\alpha f^{-1})$. From the theorem it follows that

$$1- \text{Power} = m!\det\{[(f(v_i)-f(u_j))_+]^{j-i+1}/(j-i+1)!\}, \tag{4.1.1.}$$

where $v_i = g_\alpha{}^{-1}((i-1)/m)$, $u_j = h_\alpha{}^{-1}(j/m)$ and $(x)_+ = \max(0, x)$. Most powerful tests of this form are found by minimizing the right-hand side of (4.1.1) subject to

$$P_m(g_\alpha, h_\alpha) = m!\det\{[(v_i-u_j)_+]^{j-i+1}/(j-i+1)!\} = \alpha.$$

It is probably an understatement to say that, in general, this will be difficult to do.

4.2. *General confidence regions for F.* Very general confidence regions for estimating F are possible. From (1.1) we have

$$P(h^{-1}F_m(x) \leqq F(x) \leqq g^{-1}F_m(x), \text{ for all } x \,|\, F) = P_m(g,h) = \alpha \quad \text{(say)}$$

and it follows that the functions $(h^{-1}F_m, g^{-1}F_m)$ form a $100\alpha\%$ confidence region for F.

Now that it is possible to construct myriads of confidence regions for F, it is natural to ask for a "best" one and to compare it with the standard Kolmogorov region. One possible criterion of "bestness" is related to minimizing the expected area of the region. Another possibility is to minimize in some suitable way the probability of covering a false G which is given by

$$P(\text{cover false } G \,|\, F) = P_m(gGF^{-1}, hGF^{-1}).$$

The questions concerning optimum confidence regions for F will be discussed more fully in Steck (1970). Here, we will content ourselves with some numerical comparisons using the criterion related to expected area.

If $X_1 \leqq X_2 \leqq \cdots \leqq X_n$ are the order statistics from a sample of n independent observations on a distribution F, then the area of the general confidence region based on the empirical distribution function is $A_n = \sum_1^{n+1} (v_i - u_{i-1})(X_i - X_{i-1})$, where $u_0 = 0$, $v_{n+1} = 1$ and (X_0, X_{n+1}) is the domain of F. In order to make sense for infinite or semi-infinite domains we normalize and consider instead

$$\tilde{A}_n = \sum_{i=1}^{n+1} (v_i - u_{i-1}) \frac{X_i - X_{i-1}}{EX_i - EX_{i-1}} \Big/ (n+1).$$

Now, $E\tilde{A}_n = 1 + \sum_1^n (v_i - u_i)/(n+1)$ and is a measure of optimality independent of F. It is certainly possible to minimize $E\tilde{A}_n$ subject to $p_n(u, v) = \alpha$, but it is difficult and I have not tried it yet except in the one-sided case.

However, motivated by the consideration that nF_n is a binomial random variable with mean nF and variance $nF(1-F)$, it is reasonable to consider regions which are fat for x near $F^{-1}(\frac{1}{2})$ where $F_n(x)$ is most variable and then near $x = 0$ or 1 where it is least variable. Following Anderson and Darling (1952), Malmquist (1954) and Noé and Vandeweile (1968), we considered first the elliptical regions, $E_1 : g, h = x \pm \beta(x(1-x)/n)^{\frac{1}{2}}$. Computation for many values of n and β showed such regions were uniformly *worse* than the Kolmogorov region with the same coverage probability. Since regions based on ellipses, E_2, which have the diagonal as a major axis are also candidates for the "best" region, it is natural to consider instead the family of ellipses rotated about the center $(\frac{1}{2}, \frac{1}{2})$ lying somewhwere between these two extremes, E_1 and E_2. Allowing degrees of freedom for major axis, minor axis, and orientation, we minimized $E\tilde{A}_n$ subject to $P_n(u, v) = \alpha$ and improved on the Kolmogorov regions. The results are summarized in the table under "F_n: Rotated Ellipse."

Since rotation is so important, let us next consider regions formed by parallel straight lines; that is, regions formed by optimum rotation of the Kolmogorov regions about the point $(\frac{1}{2}, \frac{1}{2})$. The results are summarized in the table under "F_n: Rotated Parallel Lines."

Although the Kolmogorov regions are not optimal in small samples for the criterion considered (it can also be shown that they are not asymptotically optimum, either), they are not far from optimum and one would probably prefer them, for practical reasons, over the more complicated ones.

As a final comment, we note that we do not need to consider only regions based on F_n. Although F_n has optimum asymptotic properties as an estimator of F (see, for example, Aggarwal (1955) or Dvoretzky, Kiefer, and Wolfowitz (1956)), the modified empirical distribution function has some small sample advantages.

Suppose F is a continuous distribution function on [0, 1]. Let $\hat{F}_n$ be the function obtained by connecting the points (0, 0), $(1/n+1, X_1)$, $(2/n+1, X_2)$, $\cdots$, $(n/n+1, X_n)$, (1, 1). For $\hat{F}_n$ we have $P(gF \leqq \hat{F}_n \leqq hF) = P(g^{-1}\hat{F}_n \leqq F \leqq h^{-1}\hat{F}_n) = P_n(\hat{u}, \hat{v})$, where $\hat{u}_i = h^{-1}(i/n+1)$ and $\hat{v}_i = g^{-1}(i/n+1)$, and the expected area of this confidence region, assuming $F(x) = x$, is an integral which can be approximated by $\Sigma(v_i - u_i)/(n+1)$. Computations for regions based on $\hat{F}_n$ for $g, h = x \pm c$ and for $g, h =$ rotated ellipses are summarized in the table under "$\hat{F}_n$: Kolmogorov Region" and "$\hat{F}_n$: Rotated Ellipse."

TABLE 1

Values of expected normalized area for various types of confidence regions for an unknown distribution function

		F_n			$\hat{F}_n$	
Sample Size	Confidence Coefficient	Kolmogorov Region	Rotated Parallel Lines	Rotated Ellipse	Kolmogorov Region	Rotated Ellipse
	.90	.578	.575	.568	.513	.477
10	.95	.628	.622	.614	.565	.523
	.99	.715	.707	.697	.658	.607
	.90	.450	.448	.438	.412	.390
20	.95	.492	.489	.477	.454	.429
	.99	.570	.565	.549	.535	.501
	.90	.382	.381	.370	.355	.338
30	.95	.419	.417	.403	.393	.371
	.99	.489	.486	.468	.464	.435

4.3. *Small sample distribution of K_m-statistics.* Let

$$K_m = \sup_x m^{\frac{1}{2}} \left| F_m(x) - F(x) \right| (\psi\{F(x)\})^{\frac{1}{2}}.$$

Then $P(K_m \leqq z) = P_m(g_\psi, h_\psi)$, where $g_\psi(x) = x - z(m\psi(x))^{-\frac{1}{2}}$ and

$$h_\psi(x) = x + z(m\psi(x))^{-\frac{1}{2}}.$$

Following the procedure outlined in Section 4.1, one can easily investigate the power of tests based on K_m statistics.

4.4. *Small sample distribution of C_m-statistics.* Pyke (1959) introduced a modified D_m^+ statistic defined by

$$C_m^+ = \max_i \left(\frac{i}{m+1} - U^{(i)} \right)$$

which has the property that the probability that the maximum occurs at $i = i_0$ is independent of i_0. There is also the obvious analogue of D_m,

$$C_m = \max_i \left| \frac{i}{m+n} - U^{(i)} \right|.$$

The small sample distribution of C_m^+, given by Pyke (1959), as well as the small sample distribution of C_m are given by the theorem (or corollary) with $u_i = \max[i/(m+1)-c, 0]$ and $v_i = 1$ for the distribution of $C_m^+$and $v_i = \min[i/m+1)+c, 1]$ for the distribution of C_m. The power of these tests can be studied also.

4.5. *Small sample distributions of Rényi-type statistics.* Birnbaum and Lientz (1969a), (1969b) and Lientz (1968) derive the exact and limiting distributions of the Rényi-type statistics

$$R_1 = \sup_{\{x:\, a \leqq F_n(x) \leqq b\}} \{F_n(x) - F(x)\}$$

$$R_2 = \sup_{\{x:\, a \leqq F_n(x) \leqq b\}} \left\{\frac{F_n(x) - F(x)}{F_n(x)}\right\}$$

$$R_3 = \sup_{\{x:\, a \leqq F(x) \leqq b\}} \{F_n(x) - F(x)\}.$$

These exact distributions also follow from the theorem with appropriate, easily determined, choices for h. The distributions of two-sided versions of these three statistics are also easily computed. Using the ideas of Section 4.1, it is easy to see that the power of tests based on these statistics is just as easily computed.

REFERENCES

[1] ADAMS, E. P. (1957). *Smithsonian Mathematical Formulae and Tables of Elliptic Functions.* Smithsonian, Wash., D.C.

[2] ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

[3] AGGARWAL, O. P. (1955). Some minimax invariant procedures for estimating a cumulative distribution function. *Ann. Math. Statist.* **26** 450–463.

[4] BIRNBAUM, Z. W. and LIENTZ, B. P. (1969a). Tables of critical values of some Rényi type statistics for finite sample sizes. *J. Amer. Statist. Assoc.* **64** 870–877.

[5] BIRNBAUM, Z. W. and LIENTZ, B. P. (1969b). Exact distributions for some Rényi-type statistics. *Zastos. Mat.* **10** 179–192.

[6] BIRNBAUM, Z. W. and TINGEY, F. H. (1951). One-sided confidence contours for probability distribution functions. *Ann. Math. Statist.* **22** 592–596.

[7] DEMPSTER, A. P. (1959). Generalized D_n^+ statistics. *Ann. Math. Statist.* **30** 593–597.

[8] DURBIN, J. (1968). The probability that the sample distribution function lies between two parallel straight lines. *Ann. Math. Statist.* **39** 398–411.

[9] DVORETSKY, A., KIEFER, J., and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[10] DWASS, M. (1959). The distribution of a generalized D_m^+ statistic. *Ann. Math. Statist.* **30** 1024–1028.

[11] EPANECHNIKOV, V. A. (1968). The significance level and power of the two-sided Kolmogorov test in the case of small sample sizes. *Theor. Probability Appl.* **13** 686–690 (English translation).

[12] KEMPERMAN, J. H. B. (1957). Some exact formulae for the Kolmogorov–Smirnov distributions. *Nederl. Akad. Wetensch. Proc. Ser. A* **60** 535–540.

[13] KOLMOGOROV, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari.* **4** 83–91.

[14] LIENTZ, B. P. (1968). Distributions of Rényi and Kac type statistics and power of corresponding tests based on Suzuki-type alternatives. Technical Report No. 51, Contract N-onr-477(38).

[15] MALMQUIST, S. (1954). On certain confidence contours for distribution functions. *Ann. Math. Statist.* **25** 523–533.

[16] MOHANTY, S. G. (1970). Personal communication.

[17] NOÉ, M. and VANDEWIELE, G. (1968). The calculation of distributions of Kolmogorov–Smirnov type statistics including a table of significance points for a particular case. *Ann. Math. Statist.* **39** 233–241.

[18] PYKE, R. (1959). The supremum and infimum of the Poisson process. *Ann. Math. Statist.* **30** 568–576.

[19] RIORDAN, J. (1968). *Combinatorial Identities*. Wiley, New York.

[20] STECK, G. P. (1969). The Smirnov two sample tests as rank tests. *Ann. Math. Statist.* **40** 1449–1466.

[21] STECK, G. P. (1970). Optimum confidence regions for an unknown distribution function. In preparation.

[22] SUZUKI, G. (1967). On exact probabilities of some generalized Kolmogorov's *D*-statistics. *Ann. Inst. Statist. Math.* **19** 373–388.

[23] SUZUKI, G. (1968). Kolmogorov–Smirnov tests of fit based on some general bounds. *J. Amer. Statist. Assoc.* **63** 919–924.

[24] WALD, A. and WOLFOWITZ, J. (1939). Confidence limits for continuous distribution functions. *Ann. Math. Statist.* **10** 105–118.

[25] WALKUP, D. W. (1970). A direct proof of Steck's result on order statistics. Math. Note No. 644, Math. Research Lab., Boeing Scientific Research Labs, Seattle, Washington.

[26] WHITTLE, P. (1961). Some exact results for one-sided tests of the Kolmogorov–Smirnov type. *Ann. Math. Statist.* **32** 499–505.