ON SOME DISTRIBUTIONS RELATED TO THE STATISTIC D_n^+

By Z. W. Birnbaum and Ronald Pyke[1]

University of Washington

1. Introduction and summary. Let $X_1 < X_2 < \cdots < X_n$ be a sample of size n, ordered increasingly, of a one-dimensional random variable X which has the continuous cumulative distribution function F. It is well known, [1], that the statistic

$$D_n^+ = \sup_{-\infty < x < +\infty} \{F_n(x) - F(x)\}, \tag{1}$$

where $F_n(x)$ is the empirical distribution function determined by $X_1, X_2, \cdots, X_n$, has a probability distribution independent of F. One may, therefore, assume that X has the uniform distribution in (0, 1) and, observing that the supremum in (1) must be attained at one of the sample points, write without loss of generality

$$D_n^+ = \max_{1 \leqq i \leqq n} (i/n - U_i), \tag{2}$$

where $U_1 < U_2 < \cdots < U_n$ is an ordered sample of a random variable with uniform distribution in (0, 1).

For a given $n > 0$ define the random variable i^* as that value of i, determined uniquely with probability 1, for which the maximum in (2) is reached, i.e., such that

$$D_n^+ = i^*/n - U_{i^*}, \tag{3}$$

and write

$$U_{i^*} = U^*. \tag{3.1}$$

The main object of this paper is to obtain the distribution functions of (i^*, U^*), of i^* and of U^*. The asymptotic distribution of $\alpha_n = i^*/n$ is also investigated, and bounds are obtained on the difference between the exact and the asymptotic distribution.

A number of general identities, which are not commonly known, have been verified and used in proving the above-mentioned results. Since these identities may be helpful in other problems of this type, they are separated from the main proofs and appear in the next section.

Received March 25, 1957; revised July 26, 1957.

[1] Research under the sponsorship of the Office of Naval Research. The second author's research was also supported by the Ontario Research Foundation. This paper was presented at the Seattle meeting of the Institute in August, 1956.

2. Some useful lemmas.

LEMMA 1. *For all real a, b and integer* $n \geqq 0$

$$\sum_{i=0}^{n} \binom{n}{i} (a + i)^i (b - i)^{n-1} = n! \sum_{i=0}^{n} \frac{(a + b)^i}{i!}. \tag{4}$$

PROOF. The identity

$$(b - n) \sum_{i=0}^{n} \binom{n}{i} (a + i)^i (b - i)^{n-i-1} = (a + b)^n, \tag{5}$$

for all real a, b and integer $n \geqq 0$ (for $b = n$ the left-hand term is defined as the limit for $b \rightarrow n$) was proven by Abel ([2], Vol. 1, p. 102). Denoting the left side of (4) by $f_n(a, b)$ we have

$$f_n(a, b) - nf_{n-1}(a, b) = (b - n) \sum_{i=0}^{n} \binom{n}{i} (a + i)^i (b - i)^{n-i-1} = (a + b)^n$$

by (5). For $n = 1$, (4) is obviously true. Assuming it is true for $n - 1$ we have

$$f_n(a, b) = nf_{n-1}(a, b) + (a + b)^n = n! \sum_{i=0}^{n} \frac{(a + b)^i}{i!},$$

which completes the proof of (4) by induction.

LEMMA 2. *For all real a, b and integers* $n \geqq 0$

$$\sum_{i=0}^{n-1} \binom{n}{i} (a + i)^i (b - i)^{n-i-1} = \sum_{i=0}^{n-1} (a + b)^i (a + n)^{n-i-1}. \tag{6}$$

PROOF. For $b \neq n$, the left side of (6) is by Abel's identity (5) equal to

$$[(a + b)^n - (a + n)^n](b - n)^{-1},$$

which is equal to the right-hand side of (6), summed as a geometric progression. That (6) is true for $b = n$ follows from the continuity of both sides of (6).

LEMMA 3. *For all real a, b and integers* $n > 0$

$$\begin{aligned}(a - 1)(b - n) \sum_{i=0}^{n} \frac{1}{i + 1} \binom{n}{i} (a + i)^i (b - i)^{n-i-1} \\ = \frac{1}{n + 1} [(a + b)^n (a + b - n - 1) - (b + 1)^n (b - n)].\end{aligned} \tag{7}$$

PROOF. Since $(a - 1) / (i + 1) = (a + i) / (i + 1) - 1$, we may write

$$\begin{aligned}(a - 1)(b - n) \sum_{i=0}^{n} \frac{1}{i + 1} \binom{n}{i} (a + i)(b - i)^{n-i-1} \\ = \frac{b - n}{n + 1} \sum_{i=0}^{n} \binom{n + 1}{i + 1} (a - 1 + i + 1)^{i+1} (b + 1 - i - 1)^{n+1-(i+1)-1} \\ - (b - n) \sum_{i=0}^{n} \binom{n}{i} (a + i)^i (b - i)^{n-i-1}.\end{aligned}$$

Applying Lemma 2 to the first sum and identity (5) to the second, one concludes that this is equal to the right-hand side of (7).

COROLLARY 1. *For all integers* $n > 0$ *we have*

$$\sum_{i=0}^{n-1} \frac{1}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1} = (n+1)^{n-1}. \tag{8}$$

PROOF. For $a = 0$, (7) yields

$$\sum_{i=0}^{n-1} \frac{1}{i+1} \binom{n}{i} i^i (b-i)^{n-i-1} = \frac{1}{n+1} \left[\frac{n^n - b^n}{n-b} - b^n + (b+1)^n \right]$$

and (8) follows for $b \to n$.

3. The distributions of (i^*, U^*), i^* and U^*. The following notations will be used: for any function f, denote by $[f \varepsilon B]$ that subset of the domain of f on which f takes values in B, a subset of the range of f; for any univariate distribution function, F, let P_F denote the n-dimensional product measure determined by the probability measure associated with F; without a subscript, P will be that measure determined by the uniform distribution function; the value of n, though suppressed in the notation, shall always be made clear by the particular circumstances of its use; furthermore, for $j = 1, 2, \cdots n$ and $u \varepsilon [0, 1]$, set

$$p_j = P[i^* = j], \qquad G^*(u, j) = P[U^* \leqq u, i^* \leqq j], \qquad H^*(u) = P[U^* \leqq u]; \tag{9}$$

for real x, $[x]$ denotes the greatest integer less than x.

All the theorems of this section are stated at the outset, and the proofs are then presented in what appears a natural sequence.

THEOREM 1. *The probabilities for* i^* *are given by*

$$p_j = n^{-n} \sum_{i=n-j}^{n-1} \frac{1}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1} \tag{10}$$

for $j = 1, 2, \cdots n$.

THEOREM 2. *The joint probability distribution of* i^* *and* U^* *is given for all* $u \varepsilon [0, 1]$ *by*

$$G^*(u, k) = \sum_{j=1}^{k} K(u, j) \qquad (k = 1, 2, \cdots n),$$

where

$$K(u, j) = P[U^* \leqq u, i^* = j] =$$

$$\begin{cases} p_j & \text{if } nu \geqq j \\ n^{-1} \sum_{i=j}^{n} \binom{n}{i} (i-u)^{n-i-1} (i-nu) n^{-i} \sum_{t=[nu]}^{i} \binom{i}{t} (nu-t-1)^{i-t} (t+1)^{t-1} & \text{if } nu < j. \end{cases} \tag{11}$$

THEOREM 3. *The random variable U^* is uniformly distributed over* [0, 1].

PROOF OF THEOREM 2. For $j = 1, 2, \cdots, n$, consider the events

$$B_j = [U^* < u; i^* = j]$$
$$= [U_j < u; j/n - U_j > i/n - U_i, (i \neq j)].$$

Employing the transformation

$$Z_j = U_j; \qquad Z_i = U_i - U_j, \qquad (i \neq j)$$

one obtains

$$B_j = [Z_j \leqq u; Z_i > (i - j) / n, (i \neq j)].$$

Setting $\mathbf{Z} = (Z_1, Z_2, \cdots Z_n)$, the joint probability element of $\mathbf{Z}$ for j fixed is

$$dH_j(\mathbf{Z}) = n!\, dZ_1\, dZ_2 \cdots dZ_n$$

for

$$[-Z_j \leqq Z_1 < Z_2 \leqq \cdots \leqq Z_{j-1} \leqq 0 \leqq Z_{j+1} \leqq \cdots \leqq Z_n \leqq 1 - Z_j]$$

and zero elsewhere. Assume u and j fixed such that $nu < j$ and $1 < j < n$. Writing $\lambda = [nu]$, one has

$$K(u, j) = \int_{B_j} dH_j(Z)$$
$$= n! \int_{-1/n}^{0} dZ_{j-1} \int_{-2/n}^{Z_{j-1}} dZ_{j-2} \cdots \int_{-\lambda/n}^{Z_{j-\lambda+1}} dZ_{j-\lambda} \int_{-u}^{Z_{j-\lambda}} dZ_{j-\lambda-1} \cdots \int_{-u}^{Z_2} dZ_1$$
$$\cdot \int_{-Z_1}^{u} dZ_j \int_{(n-j)/n}^{1-Z_j} dZ_n \int_{(n-j-1)/n}^{Z_n} dZ_{n-1} \cdots \int_{1/n}^{Z_{j+2}} dZ_{j+1}.$$

By the linear transformation

$$x_i = \begin{cases} Z_{j+i} & \text{for } i = 1, 2, \cdots, n - j, \\ 1 - Z_j & \text{for } i = n - j + 1, \\ 1 + Z_{i-n+j-1} & \text{for i} = n - j + 2, n - j + 3, \cdots, n, \end{cases}$$

one obtains

$$K(u, j) = n! \int_{(n-1)/n}^{1} dx_n \cdots \int_{(n-\lambda)/n}^{x_{n-\lambda+2}} dx_{n-\lambda+1} \int_{1-u}^{x_{n-\lambda+1}} dx_{n-\lambda} \cdots \tag{12}$$
$$\int_{1-u}^{x_{n-j+2}} dx_{n-j+1} \int_{(n-j)/n}^{x_{n-j+1}} dx_{n-j} \cdots \int_{2/n}^{x_3} dx_2 \int_{1/n}^{x_2} dx_1.$$

Denote by J_k the result of integration up to and including that with respect to x_k. By repeated integration one finds

$$J_{n-j} = \frac{x_{n-j+1}^{n-j}}{(n-j)!} - \frac{x_{n-j+1}^{n-j-1}}{n(n-j-1)!}.$$

Hence

$$J_{n-j+1} = \frac{x_{n-j+2}^{n-j+1}}{(n-j+1)!} - \frac{x_{n-j+2}^{n-j}}{n(n-j)!} + W_j(u, j),$$

where for $i = 0, 1, 2, \cdots, j$,

$$W_i(u, j) = \frac{(1-u)^{n-i}}{n(n-i+1)!}(nu - i + 1).$$

Repeated integration gives

$$J_{n-\lambda} = \frac{x_{n-\lambda+1}^{n-\lambda}}{(n-\lambda)!} - \frac{x_{n-\lambda+1}^{n-\lambda-1}}{n(n-\lambda-1)!} + \sum_{i=\lambda+1}^{j} W_i(u, j)\frac{(x_{n-\lambda+1} - 1 + u)^{i-\lambda-1}}{(1-\lambda-1)!}.$$

By properties of the binomial expansion one obtains

$$\frac{x_{n-\lambda+1}^{n-\lambda}}{(n-\lambda)!} - \frac{x_{n-\lambda+1}^{n-\lambda-1}}{n(n-\lambda-1)!} = \frac{-1}{(n-\lambda)!}\sum_{s=0}^{n-\lambda}\binom{n-\lambda}{s}(x_{n-\lambda+1} - 1 + u)^s \cdot (1-u)^{n-\lambda-s-1}[u - (s+\lambda)/n]$$

and therefore

$$J_{n-\lambda} = \frac{1}{(n-\lambda)!}\sum_{s=j-\lambda}^{n-\lambda}\binom{n-\lambda}{s}(x_{n-\lambda+1} - 1 + u)^s \cdot (1-u)^{n-\lambda-s-1}[(s+\lambda)/n - u]. \tag{13}$$

The identity

$$\int_{(n-1)/n}^{1} dx_n \int_{(n-2)/n}^{x_n} dx_{n-1} \cdots \int_{(n-\lambda)/n}^{x_{n-\lambda+2}} (u - 1 + x_{n-\lambda+1})^s \, dx_{n-\lambda+1} \cdots dx_{n-1}\, dx_n$$
$$= \frac{s!}{(s+\lambda)!} n^{-(s+\lambda)} \left[(nu)^{s+\lambda} - \sum_{t=0}^{\lambda-1}\binom{s+\lambda}{t}(nu - 1 - t)^{s+\lambda-t}(1+t)^{t-1}\right]$$

is easily proven by induction on λ. Applying (5) one shows that the right side of this identity is equal to

$$\frac{s!\,n^{-(s+\lambda)}}{(s+\lambda)!}\sum_{t=\lambda}^{s+\lambda}\binom{s+\lambda}{t}(nu - t - 1)^{s+\lambda-t}(1+t)^{t-1}.$$

Hence it follows from (13) that

$$K(u, j) = n!J_n$$
$$= \frac{1}{n}\sum_{s=j-\lambda}^{n-\lambda}\binom{n}{s+\lambda}(1-u)^{n-s-\lambda-1}(s + \lambda - nu)n^{-s-\lambda}\sum_{t=\lambda}^{s+\lambda}\binom{s+\lambda}{t} \cdot (nu - t - 1)^{s+\lambda-t}(t+1)^{t-1},$$

which is the expression in (11) in the case $nu < j$.

With a few minor changes, the above argument may be also used to prove Theorem 2 for $j = 1$ and $j = n$. For example, in the discussion preceding (12) one has to define $Z_0 = Z_{n+1} = 0$, in (12) $J_0 = 1$, and in (13) $x_{n+1} = 1$. Since $j = 1$ has $\lambda = 0$, the theorem in this case follows directly from (13).

To complete the proof of Theorem 2, it remains to consider the case of $u \geqq j/n$. Since $D_n^+ \geqq 0$ then, by (3), $U^* \leqq i^*/n$. This implies that for $u > j/n$ we have

$$[U^* < u, i^* = j] = [U^* \leqq j/n, i^* = j] = [i^* = j],$$

hence the first statement of (11) is true.

Proof of Theorem 1. We have

$$\begin{aligned} p_j &= P[i^* = j] = P[U^* \leqq j/n, i^* = j] \\ &= \lim_{u \nearrow j/n} K(u, j) \\ &= \frac{1}{n} \sum_{i=j}^{n} \binom{n}{i} (1 - j/n)^{n-i-1} (i - j) n^{-i} \sum_{t=j-1}^{i} \binom{i}{t} (j - t - 1)^{i-t} (t + 1)^{t-1}, \end{aligned}$$

which, after neglecting zero terms and interchanging summations, becomes for $s = i - t$

$$\begin{aligned} p_j &= n^{-n} \sum_{t=j}^{n} \binom{n}{t} (t + 1)^{t-1} \sum_{s=0}^{n-t} \binom{n - t}{s} (j - t - 1)^s (n - j)^{n-t-s-1} (s + t - j) \\ &= n^{-n} \sum_{t=j}^{n} \binom{n}{t} (t + 1)^{t-1} (n - t - 1)^{n-t-1} \end{aligned}$$

by a direct application of the binomial expansion. Setting $i = n - t - 1$ for $t < n$, one obtains

$$p_j = n^{-n}(n + 1)^{n-1} - n^{-n} \sum_{i=0}^{n-j-1} \frac{1}{i + 1} \binom{n}{i} i^i (n - i)^{n-i-1},$$

the last sum being zero for $j = n$. By Corollary 1 it follows that for all j,

$$p_j = \frac{1}{n} \sum_{i=n-j}^{n-1} \frac{1}{i + 1} \binom{n}{i} \left(\frac{i}{n}\right)^i \left(i - \frac{i}{n}\right)^{n-i-1}.$$

This completes the proof of Theorem 1.

Proof of Theorem 3. With $\lambda = [nu]$ as above, it follows from Theorem 2 that

$$\begin{aligned} H^*(u) &= \sum_{j=1}^{n} K(u, j) \\ &= \sum_{j=1}^{\lambda} p_j + \sum_{j=\lambda+1}^{n} \frac{1}{n} \sum_{i=j}^{n} \binom{n}{i} (1 - u)^{n-i-1} (i - nu) n^{-i} \\ &\qquad \sum_{t=\lambda}^{i} \binom{i}{t} (nu - t - 1)^{i-t} (t + 1)^{t-1}. \end{aligned}$$

Interchanging summations in the last term according to the pattern

$$\sum_{j=\lambda+1}^{n} \sum_{i=j}^{n} \sum_{t=\lambda}^{i} = \sum_{i=\lambda+1}^{n} (i - \lambda) \sum_{t=\lambda}^{i} = \sum_{t=\lambda}^{n} \sum_{i=t}^{n} (i - \lambda),$$

(the second step follows because the index j does not appear in the summand; the last step follows since at $i = \lambda$ the summand is zero), one obtains

$$(14)\quad \begin{aligned} H^*(u) &= \sum_{j=1}^{\lambda} p_j + n^{-n} \sum_{t=\lambda}^{n} \binom{n}{t} (t+1)^{t-1} \\ &\qquad \sum_{i=t}^{n} (i-\lambda)(i-nu) \binom{n-t}{i-t} (n-nu)^{n-i-1}(nu-t-1)^{i-t}. \end{aligned}$$

Using known properties of the binomial expansion, one can show that, whenever $n - t \neq 1$

$$(15)\quad \begin{aligned} &\sum_{s=0}^{n-t} \{(t-\lambda)(t-nu) + s(2t-\lambda-nu) + s^2\} \\ &\cdot \binom{n-t}{s} (nu-t-1)^s (n-nu)^{n-t-s-1} = -(t-\lambda)(n-t-1)^{n-t-1}. \end{aligned}$$

When $n - t = 1$, this sum reduces to

$$(16)\quad \begin{aligned} &\sum_{s=0}^{1} \{(n-1-\lambda)(n-1-nu) + s(2n-2-\lambda-nu) + s^2\}(-1)^s \\ &\qquad = -(n-1-\lambda) - n(1-u). \end{aligned}$$

Substituting (15) and (16) into (14), while setting $i - t = s$, one obtains

$$(17)\quad \begin{aligned} H^*(u) &= \sum_{j=1}^{\lambda} p_j - n^{-n} \sum_{t=\lambda}^{n-1} \binom{n}{t} (t-\lambda)(t+1)^{t-1}(n-t-1)^{n-t-1} \\ &\qquad - (1-u) - n^{-n}(n-\lambda)(n+1)^{n-1}. \end{aligned}$$

Employing Theorem 1, Corollary 1 with $i = n - t - 1$, and Lemma 2, one concludes from (17)

$$\begin{aligned} H^*(u) &= n^{-n} \sum_{i=n-\lambda}^{n-1} \frac{\lambda - n + i + 1}{i+1} \binom{n}{i} i^i (n-i)^{n-i} - n^{-n} \sum_{i=0}^{n-\lambda-1} \frac{n-i-1-\lambda}{i+1} \cdot \\ &\qquad \binom{n}{i} i^i (n-i)^{n-i-1} - (1-u) - (n-\lambda)p_n \\ &= n^{-n} \sum_{i=0}^{n-1} \binom{n}{i} i^i (n-i)^{n-i} - 1 + u \\ &= u. \end{aligned}$$

This completes the proof of Theorem 3.

A consequence of Theorem 1 is the following

COROLLARY 2. *For all integers $n > 0, j > 0$,*

$$0 < p_1 < p_2 < \cdots < p_n < 1,$$

$$\lim_{n\to\infty} np_j = \sum_{i=1}^{j} \frac{e^{-i} i^{i-1}}{i!},$$

$$\lim_{n\to\infty} np_{n-j} = e - \sum_{i=0}^{j-1} \frac{e^{-i} i^i}{(i+1)!}.$$

PROOF. The first statement is evident from (10); the second follows from (10) by applying Stirling's formula; and the third follows by applying Stirling's formula to the expression

$$p_{n-j} = \frac{(n+1)^{n-1}}{n^n} - \sum_{i=0}^{j-1} \frac{1}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1},$$

which can be obtained from (10) by Corollary 1.

Thus the statistic D_n^+ places more weight upon the larger observations than on the smaller ones, in the sense that the maximum deviation between F and F_n is more probable to occur at X_{k+1} than at X_k for $k = 1, 2, \cdots, n-1$.

4. The asymptotic distribution of $\alpha_n = i^*/n$. Writing U_n^* instead of U^*, we have according to Theorem 3,

$$P[U_n^* \leqq u] = H_n^*(u) = u, \qquad 0 \leqq u \leqq 1. \tag{18}$$

Since the Glivenko-Cantelli theorem ([3], p. 260) implies that D_n^+ converges in probability to zero, it follows from (3) that

$$\alpha_n - U_n^* \rightarrow 0 \quad \text{in probability.} \tag{19}$$

From (18) and (19) one can conclude that α_n is asymptotically uniformly distributed on [0, 1].

The following theorem contains more specific statements on the asymptotic behavior of the distribution of α_n.

THEOREM 4. *For every positive integer n we have*

$$E(\alpha_n) = \frac{1}{2} + \frac{1}{2} n^{-n-1} n! \sum_{i=0}^{n-1} \frac{n^i}{i!}, \tag{20}$$

$$x - \sqrt{n^{-n-1} n! \sum_{i=0}^{n-1} \frac{n^i}{i!}} \leqq \Pr\{\alpha_n < x\} \leqq x \quad \text{for } 0 \leqq x \leqq 1. \tag{21}$$

PROOF OF THEOREM 4. From Theorem 1 we have

$$\begin{aligned} E(\alpha_n) &= n^{-n-1} \sum_{j=1}^{n} \sum_{i=n-j}^{n-1} \frac{j}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1} \\ &= \frac{1}{2}\left(1 - \frac{1}{n}\right) + \frac{1}{2} n^{-n-1} \sum_{i=0}^{n} \binom{n}{i} i^i (n-i)^{n-i} \end{aligned}$$

and this by Lemma 1 yields (20). To obtain the upper bound on $\Pr\{\alpha_n < x\}$ in (21) we note that

$$G_n(x) = \Pr\{\alpha_n < x\} = \sum_{u=1}^{[nx]} p_j$$

and in view of Corollary 2 this must be $<x$ for all $1/n < x \leqq 1$.

To obtain the lower inequality in (21) we need

LEMMA 4. *Let X be a random variable with c.d.f. F, such that $F(0) = 0$, $F(1 + 0) = 1$, $F(x) \leqq x$ for $0 \leqq x \leqq 1$. Then*

$$F(s) \geqq s - \sqrt{2E(X) - 1}. \tag{22}$$

PROOF OF LEMMA 4. We have

$$E(X) = \int_0^1 X\,dF(X) \geqq 1 - \int_0^1 F(X)\,dX$$

$$\geqq 1 - \int_0^{F(s)} X\,dX - \int_{F(s)}^{s} F(s)\,dX - \int_s^1 X\,dX = \tfrac{1}{2}\{1 + [s - F(s)]^2\}$$

and this implies (22). One verifies directly that, for given s and $F(s)$, equality is attained in (22) when $F(t) = t$ for $0 \leqq t \leqq F(s)$, $F(t) = F(s)$ for $F(s) \leqq t \leqq s$, $F(t) = t$ for $s \leqq t \leqq 1$.

According to the upper inequality in (21), $\Pr\{\alpha_n < x\}$ fulfills the assumptions of Lemma 4, which together with (20) yields the lower bound of (21).

It may be noted that by an application of Stirling's formula one obtains from (21)

$$0 \leqq x - \Pr\{\alpha_n < x\} = 0(n^{-1/4}), \tag{23}$$

and that (20) together with (3) yields

$$E(D_n^+) = 2^{-1}n^{-n-1}n!\sum_{i=0}^{n-1}\frac{n^i}{i!}. \tag{24}$$

REFERENCES

[1] A. WALD AND J. WOLFOWITZ, "Confidence limits for continuous distribution functions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[2] N. H. ABEL, *Oevres Complètes*, Christiania, C. Groendahl, 1839.

[3] M. FRÉCHET, *Recherches théoriques modernes sur la théorie des probabilités*, Gauthier-Villars, Paris, 1937, p. 260.