THE DISTRIBUTION OF A GENERALIZED D_n^+ STATISTIC[1]

By Meyer Dwass

Northwestern University

1. Introduction and summary. Let $F_n(x)$ be the empirical c.d.f. of n independent random variables, each distributed according to the same continuous c.d.f. $F(x)$. The major object of this paper is to obtain in explicit form the probability law of the random variable

$$D_n^+(\gamma) = \sup_{-\infty < x < \infty} \{F_n(x) - \gamma F(x)\}.$$

It is no loss of generality to suppose that $F(x)$ is the c.d.f. of the uniform distribution on [0, 1], so this assumption will be held throughout the paper.

When $\gamma = 1$, then $D_n^+(1)$ is the usual one-sided goodness of fit statistic whose asymptotic distribution was first derived by Smirnov [6]. We obtain in several different forms (formulas 2.2 and 2.3) an expression for

$$P(D_n^+(\gamma) < a) = P(F_n(x) \leqq a + \gamma x, 0 \leqq x \leqq 1).$$

Formula (2.2) agrees with the one found by Birnbaum and Tingey [2] when $\gamma = 1$, which is the "classical" case. As a matter of fact, it seems to have been overlooked that this formula, for finite n, had already appeared in a paper by Smirnov [6]. The new formula (2.3) would seem to involve fewer computations for actual numerical evaluation. One rather remarkable fact which results from (2.3) is that

$$P(F_n(x) \leqq \gamma x, 0 \leqq x \leqq 1) = \begin{cases} 1 - \dfrac{1}{\gamma}, & \gamma > 1 \\ 0, & \gamma \leqq 1, \end{cases}$$

for any n. This was noted by Daniels [4] and was rediscovered by Robbins [5].

Using (2.3) it is easy to evaluate $\lim_{n\to\infty} P(F_n(x) \leqq a(n) + \gamma x)$ where γ, $(\gamma > 1)$ is fixed and $a(n) = d/n$, where d is fixed. The limiting distribution when $\gamma > 1$ can be used to derive some facts about the Poisson Process which were recently discovered by Baxter and Donsker [1].

The methods used are elementary. To assist the reader, the results are all listed in Section 2 and Section 3 is devoted to giving proofs.

2. Statement of results. First a few pieces of notation are introduced. Let

$$P_n(a, \gamma) = P(F_n(x) < a + \gamma x, 0 \leqq x \leqq 1) = P(D_n^+(\gamma) < a),$$

Received January 20, 1959; revised April 1, 1959.

[1] Sponsored by Air Force Office of Scientific Research Air Research and Development Command, Contract No. AF 49 (638)–151.

and let

$$C_n(a, \gamma, i) = \binom{n}{i}\left(\frac{n-i}{n\gamma} - \frac{a}{\gamma}\right)^{n-i}\left(1 - \left(\frac{n-i}{n\gamma} - \frac{a}{\gamma}\right)\right)^{i-1}\left(\frac{\gamma + a - 1}{\gamma}\right).$$

For simplicity, whenever it is reasonable to do so, P_n and $C_n(i)$ are used instead of the more complicated symbols.

It is assumed hereafter that

$$(2.1) \qquad 0 < a < 1, \qquad a + \gamma > 1, \qquad \gamma > 0,$$

for otherwise P_n becomes trivially either 0 or 1.

THEOREM 1:

$$(2.2) \qquad P_n = 1 - \sum_{i=0}^{k} C_n(i),$$

or, equivalently,

$$(2.3) \qquad P_n = \sum_{i=k+1}^{n} C_n(i),$$

where the integer k is defined by

$$\frac{k}{n} \leqq (1 - a) < \frac{k+1}{n}.$$

Remark on Theorem 1: When $\gamma = 1$, formula (2.2) agrees with the result of Birnbaum and Tingey. However, when a is of the order of $1/\sqrt{n}$, (2.3) will usually require many fewer values of $C_n(i)$ to compute. For example, for $n = 50$, Table 1 of [2] indicates that a varies roughly between $\frac{1}{7}$ and $\frac{1}{4}$, for those probabilities "interesting" for statistical applications. Hence the number of $C_n(i)$ terms to be computed using (2.3) ranges from about 37 to 42 for these a's, whereas using (2.2) the range is from 7 to 12 terms.

Setting $a = 0$ in (2.3) yields a

COROLLARY TO THEOREM 1 (Daniels [4], Robbins [5]):

$$P(F_n(x) < \gamma x, \quad 0 \leqq x \leqq 1) = \begin{cases} 1 - \dfrac{1}{\gamma}, & \gamma > 1, \\ 0, & \gamma \leqq 1. \end{cases}$$

It is interesting that this result does not depend on n.

THEOREM 2: *Let $a = d/n$, where d is a fixed positive real number, and let γ be greater than* 1. *Then*

$$(2.4) \qquad \lim_{n \to \infty} P_n(d/n, \gamma) = \left(1 - \frac{1}{\gamma}\right) \sum_{i=0}^{[d]} \frac{1}{i!}\left(\frac{i-d}{\gamma}\right)^i e^{(d-i)/\gamma}.$$

Remarks on Theorem 2:

a) The interesting fact here is that when $\gamma > 1$ the proper norming for a requires it to be of the order of $1/n$ rather than $1/\sqrt{n}$ as in the case $\gamma = 1$.

Contrary to what one would expect, the derivation is much more elementary when $\gamma > 1$ than when $\gamma = 1$.

b) The right hand side of (2.4) is the same as an expression obtained by Baxter and Donsker [1] in connection with the Poisson process. Theorem 2 immediately gives the same result which is summarized in the following corollary.

COROLLARY TO THEOREM 2: *Let $Y(t)$, $0 \leqq t < \infty$ be the Poisson process with stationary and independent increments and parameter $\lambda > 0$, and $Y(0) = 0$. Let $\gamma > \lambda$, and d be positive. Then*

$$P(Y(t) < d + \gamma t, 0 \leqq t < \infty) = \left(1 - \frac{\lambda}{\gamma}\right)\sum_{i=0}^{[d]} \frac{1}{i!}\left(\frac{\lambda}{\gamma}\right)^i (i - d)^i e^{(\lambda/\gamma)(d-i)}.$$

3. Proofs.

I. *Proof of Theorem* 1, *equation* (2.2). The basic idea used in the proof is the following: let $x_1 < x_2 < \cdots < x_n$ be the ordered values of n independent random variables, each uniformly distributed over (0, 1). Then, it is well known that given x_n, the conditional distribution of

$$x_1/x_n, \cdots, x_{n-1}/x_n$$

is that of the ordered values of $(n - 1)$ independent random variables, each uniformly distributed over (0, 1). Using this fact it is easy to verify the following conditional probability statements:

$$P(F_n(x) < a + \gamma x \mid x_n = t)$$
$$= \begin{cases} 1, & \text{if } \dfrac{1-a}{\gamma} < t \text{ and } \dfrac{n-1}{n} \leqq a \leqq 1, \\ P_{n-1}\left(\dfrac{n}{n-1}a, \dfrac{n}{n-1}\gamma t\right), & \text{if } \dfrac{1-a}{\gamma} < t \text{ and } a < \dfrac{n-1}{n}, \\ 0, & \text{if } t \leqq \dfrac{1-a}{\gamma}. \end{cases}$$

Using the fact that the frequency function of x_n is

$$\begin{matrix} nt^{n-1}, & 0 \leqq t \leqq 1, \\ 0, & \text{otherwise,} \end{matrix}$$

we have the basic recursion relationship

$$(3.1) \qquad P_n(a, \gamma) = \begin{cases} \displaystyle\int_{\frac{1-a}{\gamma}}^1 P_{n-1}\left(\frac{n}{n-1}a, \frac{n}{n-1}\gamma t\right) nt^{n-1}\,dt, & \text{if } a < \dfrac{n-1}{n} \\ \displaystyle\int_{\frac{1-a}{\gamma}}^1 nt^{n-1}\,dt = 1 - \left(\frac{1-a}{\gamma}\right)^n, & \text{if } \dfrac{n-1}{n} \leqq a \leqq 1. \end{cases}$$

An induction argument can now be applied to prove (2.2). Its truth is trivially true when $n = 1$. Assume now that it holds for arbitrary n. By this induction hypothesis,

$$P_n\left(\frac{n+1}{n}a, \frac{n+1}{n}\gamma t\right) = 1 - \sum_{i=0}^{k} C_n\left(\frac{n+1}{n}a, \frac{n+1}{n}\gamma t, i\right),$$

where k is defined by $k/n \leqq 1 - ((n+1)/n)\,a < (k+1)/n$, or equivalently by $(k+1)/(n+1) \leqq (1-a) < (k+2)/(n+1)$. By a routine, but tedious, computation which is omitted

$$\int_{\frac{1-a}{\gamma}}^{1} C_n \frac{n+1}{n}a, \frac{n+1}{n}\gamma t, i)(n+1)\, t^n\, dt = C_{n+1}(a, \gamma, i+1)\,.$$

Hence, applying (3.1), it follows that (2.2) is true for $n+1$, which completes the proof of the first part of Theorem 1.

II. *Proof of Theorem* 1, *equation* (2.3). This follows from 2.2 by means of part a) of the following lemma:

LEMMA:

a) $\sum_{i=0}^{n} C_n(i) = 1$

b) $\sum_{i=0}^{n} \binom{n}{i} (A+i)^i (B-i)^{n-i-1} = (A+B)^n/(B-n)$

c) $$\sum_{i=0}^{n} \frac{1}{i+1}\binom{n}{i}(A+i)^i(B-i)^{n-i-1} = \frac{1}{(A-1)(B-n)(n+1)} \cdot [(A+B)^n(A+B-n-1) - (B+1)^n(B-n)]$$

d) $$\sum_{j=0}^{n-1}\binom{n}{j+1}(A+j)^j(B-j)^{n-j-1} = \frac{(A+B)^n - (B+1)^n}{A-1},$$

where $A \neq 1$, $B \neq n$. Part b) is a formula of Abel's which is referred to in Lemma 1 of [3]. Part c) is proved in [3]. Part d) is proved by writing

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

and by then applying b) and c). Part a) now follows from d) as follows. $C_n(i)$ can be expressed as

$$C_n(i) = \frac{1}{(n\gamma)^n}(n - na - 1 - (i-1))^{n-1-(i-1)}$$
$$\cdot (n\gamma - n + na + 1 + (i-1))^{i-1}\left(\frac{\gamma + a - 1}{\gamma}\right)\binom{n}{(i-1)+1}.$$

Now let $i - 1 = j$, $n\gamma - n + na + 1 = A$, $n - na - 1 = B$, then

$$\sum_{i=0}^{n} C_n(i) = \frac{1}{(n\gamma)^n}\left(\frac{\gamma + a - 1}{\gamma}\sum_{j=1}^{n-1}\binom{n}{j+1}(A + j)^j(B - j)^{n-1-j}\right.$$

and a) follows from d) by routine algebra. The lemma is completely proved. Now part a) of the lemma immediately implies (3.2), given the truth of (3.1).

III. *Proof of Theorem* 2. This follows in a routine way from (2.3).

IV. *Proof of corollary to Theorem* 2. It is sufficient to suppose that $\lambda = 1$, the general case easily following from this special one.

Let $A(T)$ be the event that

$$\left\{\frac{Y(t)}{Y(T)} \leqq \frac{d}{Y(T)} + \gamma\frac{t}{T}, \qquad 0 \leqq t \leqq T\right\},$$

where $Y(t)/Y(T)$ can be defined as 0 if $Y(T) = 0$.

According to the well-known relationship between the Poisson process and uniformly distributed random variables,

$$P(A(T) \mid Y(T) = n) = P_n(d/n, \gamma), \qquad n \geqq 1.$$

Hence

$$P(A(T)) = \sum_{n=0}^{\infty} P(A(T) \mid Y(T) = n)\frac{e^{-T}T^n}{n!} = \sum_{n=0}^{\infty} P_n(d/n, \gamma)\frac{e^{-T}T^n}{n!}.$$

Since $P_n(d/n, \gamma)$ approaches the right side of (2.4) as $n \to \infty$, and since $\sum_{n=0}^{r} e^{-T}T^n/n! \to 0$ as $n \to \infty$ for any fixed r, an easy argument proves that

$$P(A(T)) \to \lim_{n\to\infty} P_n(d/n, \gamma) \quad \text{as} \quad T \to \infty.$$

Since $Y(T)/T$ converges to 1 with probability 1, it is not hard to show that

$$\lim_{T\to\infty} P(A(T)) = P(Y(t) \leqq d + \gamma t, 0 \leqq t < \infty),$$

which completes the proof.

REFERENCES

[1] Glen Baxter and M. D. Donsker, "On the distribution of the supremum functional for processes with stationary independent increments," *Trans. Am. Math. Soc.*, Vol. 85 (1957), pp. 73–87.

[2] Z. W. Birnbaum and F. H. Tingey, "One sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[3] Z. W. Birnbaum and R. Pyke, "On some distributions related to the statistic D_n^+," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 179–187.

[4] H. E. Daniels, "The statistical theory of the strength of bundles of threads. I," *Proc. Roy. Soc., Ser. A*, Vol. 183 (1945), pp. 405–435.

[5] Herbert Robbins, abstract ("A one-sided confidence interval for an unknown distribution function,") *Ann. Math. Stat.*, Vol. 25 (1954), p. 409.

[6] N. V. Smirnov, "Approximate laws of distribution of random variables from empirical data," *Uspehi Matem. Nauk*, Vol. 10 (1944), pp. 179–206.