Ann. Inst. Statist. Math.
34 (1982), Part A, 167-173

THE EXACT DISTRIBUTION OF KOLMOGOROV'S STATISTIC D_n FOR $n \leqq 10$

HAROLD RUBEN AND JACK GAMBINO

(Received Dec. 19, 1980; revised Sept. 28, 1981)

1. Introduction and summary

It is well known (and is easily proved) that the sampling distribution of the Kolmogorov statistic D_n can be expressed in terms of rectangle probabilities for uniform order statistics. Specifically, if

$$(1) \qquad D_n = \sup_x |F_n(x) - F(x)| ,$$

F_n denoting the empirical distribution function based on a random sample of size n from the continuous parent distribution F, then, for $0 \leqq a \leqq 1$,

$$(2) \qquad \mathrm{P}\,(D_n \leqq a) = \mathrm{P}\,(u_1 \leqq u_{(1)} \leqq v_1, \cdots, u_n \leqq u_{(n)} \leqq v_n) ,$$

where

$$(3) \qquad u_i = \max\left(-a + \frac{i}{n}, 0\right) , \qquad v_i = \min\left(a + \frac{i-1}{n}, 1\right)$$

$(i=1, \cdots, n)$ and $u_{(1)} \leqq u_{(2)} \leqq \cdots \leqq u_{(n)}$ is an ordered sample of size n from a distribution uniform on $[0, 1]$. Steck has shown [9] that for *arbitrary* $u_1, \cdots, u_n$, $v_1, \cdots, v_n$ such that $0 \leqq u_1 \leqq u_2 \leqq \cdots \leqq u_n \leqq 1$, $0 \leqq v_1 \leqq v_2 \leqq \cdots \leqq v_n \leqq 1$ and $u_i \leqq v_i$ $(i=1, \cdots, n)$,

$$(4) \qquad \mathrm{P}\,(u_1 \leqq u_{(1)} \leqq v_1, \cdots, u_n \leqq u_{(n)} \leqq v_n) = n!\, \det\left((v_i - u_j)_+^{j-i+1}/(j-i+1)!\right)$$

$(i, j=1, \cdots, n)$, where $x_+ = \max(0, x)$ and $(v_i - u_j)_+^k/k! = 0$ if $k<0$. (Simple proofs of (4) have been given by Pitman [6] and Sarkadi [8].) Equations (2), (3) and (4) show that $\mathrm{P}\,(D_n \leqq a)$ is an nth degree polynomial in a whose form depends on the location of a in $[0, 1]$: If n is even there are $3n/2$ forms corresponding to the $3n/2$ subintervals

$$\left(\frac{r-1}{2n}, \frac{r}{2n}\right], \quad \left(\frac{n/2+s-1}{n}, \frac{n/2+s}{n}\right] \qquad (r=1, \cdots, n;\ s=1, \cdots, n/2) ,$$

while if n is odd there are $(3n+1)/2$ forms corresponding to the $(3n+$

1)/2 subintervals

$$\left(\frac{r-1}{2n}, \frac{r}{2n}\right], \quad \left(\frac{(n-1)/2+s}{n}, \frac{(n-1)/2+s+1}{n}\right]$$
$$(r=1, \cdots, n+1;\ s=1, \cdots, (n-1)/2).$$

From a geometrical point of view, since $(u_{(1)}, \cdots, u_{(n)})$ is uniform on the n-dimensional simplex

$$S_n=\{(u_{(1)}, \cdots, u_{(n)}):\ 0 \leqq u_{(1)} \leqq u_{(2)} \leqq \cdots \leqq u_{(n)} \leqq 1\}$$

with constant density $n!$, (2) and (3) show that

$$\mathrm{P}(D_n \leqq a)=n!\,|S_n \cap [\boldsymbol{u}, \boldsymbol{v}]|, \tag{5}$$

where $\boldsymbol{u}=(u_1, \cdots, u_n)$, $\boldsymbol{v}=(v_1, \cdots, v_n)$, $[\boldsymbol{u}, \boldsymbol{v}]$ is the n-dimensional rectangle defined by

$$[\boldsymbol{u}, \boldsymbol{v}]=\{(u_{(1)}, \cdots, u_{(n)}):\ u_1 \leqq u_{(1)} \leqq v_1, \cdots, u_n \leqq u_{(n)} \leqq v_n\}$$

(u_i, v_i as in (3)), and, for an n-dimensional set A, $|A|$ denotes the (n-dimensional) volume of A. The set $S_n \cap [\boldsymbol{u}, \boldsymbol{v}]$ is an n-dimensional convex polytope depending on a whose nature changes as a varies from subinterval to subinterval.

The values of $|S_n \cap [\boldsymbol{u}, \boldsymbol{v}]|$, and therefore $\mathrm{P}(D_n \leqq a)$, for the first two subintervals and for the last subinterval (n even or odd) assume particularly simple forms. In fact, if $0 \leqq a < 1/2n$ then $u_i > v_i$ $(i=1, \cdots, n)$, so that $[\boldsymbol{u}, \boldsymbol{v}]$ is empty and

$$\mathrm{P}(D_n \leqq a)=0 \qquad (0 \leqq a \leqq 1/2n). \tag{6}$$

For $1/2n \leqq a \leqq 1/n$, the rectangle $[\boldsymbol{u}, \boldsymbol{v}]$ is a subset of S_n with equal sides $v_i-u_i=2a-1/n$, so that

$$|S_n \cap [\boldsymbol{u}, \boldsymbol{v}]|=|[\boldsymbol{u}, \boldsymbol{v}]|=(2a-1/n)^n$$

and

$$\mathrm{P}(D_n \leqq a)=n!(2a-1/n)^n \qquad (1/2n \leqq a \leqq 1/n). \tag{7}$$

For $1-1/n \leqq a \leqq 1$, the decomposition

$$A_n+B_n+C_n=S_n$$

holds, where

$$A_n=\{(u_{(1)}, \cdots, u_{(n)}):\ 0 \leqq u_{(1)} \leqq u_{(2)} \leqq \cdots \leqq u_{(n)} < 1-a\},$$
$$B_n=S_n \cap [\boldsymbol{u}, \boldsymbol{v}],$$
$$C_n=\{(u_{(1)}, \cdots, u_{(n)}):\ a < u_{(1)} \leqq u_{(2)} \leqq \cdots \leqq u_{(n)} \leqq 1\}.$$

Hence, since the closure of A_n is similar to S_n with the ratio of lengths of corresponding edges $1-a$ while C_n is congruent to A_n, the decomposition gives

$$(1-a)^n+\mathrm{P}(B_n)+(1-a)^n=1 \qquad (1-1/n\leqq a\leqq 1),$$

i.e.

$$\text{(8)} \qquad \mathrm{P}(D_n\leqq a)=1-2(1-a)^n \qquad (1-1/n\leqq a\leqq 1).$$

The polynomials $\mathrm{P}(D_n\leqq a)$ for the remaining $3(n-2)/2$ subintervals (n even) or $(3n-5)/2$ subintervals (n odd) are evaluated in the next section by a non-geometrical method.

We remark that, in addition to the discussion by Steck [9], various methods for obtaining $\mathrm{P}(D_n\leqq a)$ have been considered by Kemperman [4], Epanechnikov [3], who obtained the values of the polynomials $\mathrm{P}(D_n\leqq a)$ for $n=2, 3, 4$, Durbin [1], and Noé [5]. Kemperman's results were generalized by Durbin [1], [2] who treated the empirical distribution function F_n from a parent uniform distribution as a Poisson process and expressed the probability that F_n lies between two straight lines as an element of a transition matrix raised to a power. The reader is referred to the monograph by Durbin [2] for an informative survey of these and many other related results involving the empirical distribution function, as well as for further references.

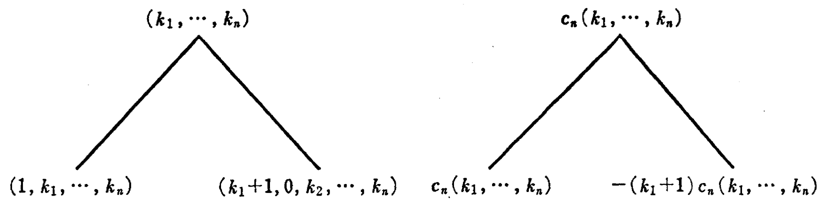
It should perhaps be pointed out that the results referred to in the previous paragraph do not readily lend themselves to calculation of $\mathrm{P}(D_n\leqq a)$ as a polynomial in a (that is, when a is not some specific *number*). In the case of Steck's method, for example, one would have to expand the determinant in (4) as a polynomial in a and collect together the terms corresponding to the same power of a.

2. Determination of the exact distribution of D_n

Based on Steck's result (4), Ruben [7] has shown that, for arbitrary u_i, v_i as in (4),

$$\text{(9)} \quad \mathrm{P}(u_1\leqq u_{(1)}\leqq v_1,\cdots, u_n\leqq u_{(n)}\leqq v_n)=n!\sum_{(k_1,\cdots,k_n)\in K_n}\prod_{i=1}^{n}\frac{(v_i-u_{k_i+i-1})_+^{k_i}}{c_n(k_1,\cdots,k_n)}$$

where the 2^{n-1} n-tuples $(k_1,\cdots,k_n)$ in K_n and the corresponding coefficients $c_n(k_1,\cdots,k_n)$ are generated recursively by means of the binary trees



with $(k_1)=(1)$, $c_1(k_1)=1$. Formally,

$$K_1=\{(1)\}; \quad K_{n+1}=\bigcup_{(k_1,\cdots,k_n)\in K_n}\{(1, k_1, \cdots, k_n), (k_1+1, 0, k_2, \cdots, k_n)\} ,$$

$$c_1(k_1)=1; \quad c_{n+1}(1, k_1, \cdots, k_n)=c_n(k_1, \cdots, k_n) ,$$
$$c_{n+1}(k_1+1, 0, k_2, \cdots, k_n)=-(k_1+1)c_n(k_1, \cdots, k_n)$$

for $n\geqq 1$ (each element $(k_1, \cdots, k_n)$ in K_n generates *two* elements in K_{n+1}). This algorithm thus gives $\mathrm{P}(D_n\leqq a)$ as an nth degree polynomial in a on replacing u_i, v_i in (9) by the special values as given in (3), after suitable collection of terms.

We have used computer programs employing the algorithm to obtain the coefficients in the polynomials $\mathrm{P}(D_n\leqq a)$ for $n=3, 4, \cdots, 10$*) over the range $[1/n, 1-1/n]$. These are given in Table 1. (The computer output consisted in fact of the coefficients of $n^n\,\mathrm{P}(D_n\leqq a)$ for $n=3, 4, \cdots, 12$ in order to deal with integers and to avoid numerical errors due to fractions.) Equations (6), (7) and (8) give $\mathrm{P}(D_n\leqq a)$ for $0\leqq a\leqq 1/n$ and for $1-1/n\leqq a\leqq 1$. Thus Table 1 in conjunction with Equations (6), (7) and (8) give the values of the polynomials for all a in $[0, 1]$ and for $n=2, 3, \cdots, 10$. (Note from (6), (7) and (8) that $\mathrm{P}(D_2\leqq a)$ is 0, $2(2a-1/2)^2$, or $1-2(1-a)^2$ according as $0\leqq a\leqq 1/4$, $1/4\leqq a\leqq 1/2$, or $1/2\leqq a\leqq 1$.) It should be stressed that the coefficients in Table 1 are *exact*. By way of illustrative examples,

$$\mathrm{P}(D_7\leqq a)=\frac{54540}{7^6}-\frac{120240}{7^5}a+\frac{36240}{7^4}a^2+\frac{45040}{7^3}a^3$$
$$-\frac{15950}{7^2}a^4-\frac{3540}{7}a^5+2120a^6-1680a^7$$
$$(2/7\leqq a\leqq 5/14) ,$$

$$\mathrm{P}(D_{10}\leqq a)=0.01016064-1.2628224a+\cdots-58060800a^{10}$$
$$(1/10\leqq a\leqq 3/20) ,$$

*) The coefficients in the polynomials $n^n\,\mathrm{P}(D_n\leqq a)$ for $n=11, 12$ can be obtained from the authors.

Table 1. Coefficients of $P(D_n \leqq a)$ for a in the various subintervals of $[1/n, 1-1/n]$, $n=3(1)10$

$n=3$

[1/3, 1/2]	[1/2, 2/3]	
0	−1	1
−8/3	10/3	a
14	2	a^2
−12	−4	a^3

$n=4$

[1/4, 3/8]	[3/8, 1/2]	[1/2, 3/4]	
3/8	3/8	−1	1
−9/2	−63/8	37/8	a
21/2	75/2	3/2	a^2
24	−48	−10	a^3
−48	16	6	a^4

$n=5$

[1/5, 3/10]	[3/10, 2/5]	[2/5, 1/2]	[1/2, 3/5]	[3/5, 4/5]	
−96/625	336/625	0	−1	−1	1
672/125	−168/25	−728/125	522/125	738/125	a
−1464/25	12	224/5	24/5	12/25	a^2
240	424/5	−456/5	−56/5	−92/5	a^3
−288	−240	74	−6	22	a^4
0	160	−20	12	−8	a^5

$n=6$

[1/6, 1/4]	[1/4, 1/3]	[1/3, 5/12]	[5/12, 1/2]	[1/2, 2/3]	[2/3, 5/6]	
−5/81	−35/1296	5/16	5/16	−1	−1	1
10/27	145/27	−565/81	−7645/648	3371/648	4651/648	a
235/9	−785/9	1525/54	775/9	175/36	−8280/7776	a^2
−1280/3	4240/9	515/9	−1985/9	−185/9	−265/9	a^3
2360	−2600/3	−1115/3	295	0	160/3	a^4
−4800	320	560	−240	32	−38	a^5
2880	320	−280	104	−20	10	a^6

$n=7$

[1/7, 3/14]	[3/14, 2/7]	[2/7, 5/14]	[5/14, 3/7]	
$8640/7^6$	$-31860/7^6$	$54540/7^6$	$54540/7^6$	1
$-8640/7^4$	$140400/7^5$	$-120240/7^5$	$-157740/7^5$	a
$20880/7^3$	$-176760/7^4$	$36240/7^4$	$73740/7^4$	a^2
$-128160/7^3$	$15600/7^3$	$45040/7^3$	$60040/7^3$	a^3
−2880/7	99840/49	−15950/49	−51950/49	a^4
92160/7	−57600/7	−3540/7	15660/7	a^5
−43200	11200	2120	−2296	a^6
40320	−4480	−1680	1008	a^7

[3/7, 1/2]	[1/2, 4/7]	[4/7, 5/7]	[5/7, 6/7]	
0	−1	−1	−1	1
$-153852/7^5$	$81446/7^5$	$104486/7^5$	$141986/7^5$	a
$31512/7^3$	$2700/7^3$	$11220/7^4$	$-7530/7^4$	a^2
$-103000/7^3$	$-6960/7^3$	$-11120/7^3$	$-14870/7^3$	a^3
3770/7	−150/7	710/49	5210/49	a^4
−4380/7	324/7	414/7	−786/7	a^5
468	20	−80	58	a^6
−168	−40	30	−12	a^7

$n=8$

[1/8, 3/16]	[3/16, 1/4]	[1/4, 5/16]	[5/16, 3/8]	
−1260/131072	$-76860/8^7$	$-12460/8^7$	$1168790/8^7$	1
$6615/8^4$	$-77490/8^5$	$25795/8^4$	$-2286620/8^6$	a
$-277830/8^4$	$3677940/8^5$	$-559265/8^4$	$434630/8^5$	a^2
80325/64	$-804930/8^3$	$523005/8^3$	$951790/8^4$	a^3
−737100/64	600390/64	−211190/64	$-302820/8^3$	a^4
49140	−22785	6020	−11802/8	a^5
−55440	10360	−9485	7931	a^6
−161280	35840	14560	−12096	a^7
322560	−35840	−11200	6720	a^8

[3/8, 7/16]	[7/16, 1/2]	[1/2, 5/8]	[5/8, 3/4]	[3/4, 7/8]	
$140/8^5$	$140/8^5$	-1	-1	-1	1
$-2480534/8^6$	$-4127620/8^6$	$1500284/8^6$	$1894034/8^6$	$2547218/8^6$	a
$1720964/8^5$	$626892/8^4$	$33740/8^4$	$138670/8^5$	$-187922/8^5$	a^2
$409108/8^4$	$-2414468/8^4$	$-131460/8^4$	$-192710/8^4$	$-247142/8^4$	a^3
$-79590/64$	$11060/8$	$-140/8$	$20790/8^3$	$96390/8^3$	a^4
$240940/64$	$-143220/64$	$6188/64$	$5838/64$	$-16842/64$	a^5
$-46900/8$	$18956/8$	$-140/8$	$-1666/8$	$1610/8$	a^6
4928	-1344	-110	156	-82	a^7
-1792	256	70	-42	14	a^8

$n=9$	[1/9, 1/6]	[1/6, 2/9]	[2/9, 5/18]	
	$-89600/9^7$	$500080/9^7$	$-9682960/9^8$	1
	$555520/9^6$	$-3365600/9^6$	$46340000/9^7$	a
	$-703360/9^5$	$7269920/9^5$	$-62515040/9^6$	a^2
	$-2535680/9^4$	$-4827200/9^4$	$14007840/9^5$	a^3
	$9623040/9^3$	$-29097600/9^4$	$27233360/9^4$	a^4
	$-13332480/81$	$56743680/9^3$	$-21514080/9^3$	a^5
	$9067520/9$	$-30365440/81$	$804160/9$	a^6
	-2938880	806400	$-1420160/9$	a^7
	3225600	-752640	179200	a^8
	0	215040	-107520	a^9

[5/18, 1/3]	[1/3, 7/18]	[7/18, 4/9]	[4/9, 1/2]	
$3267040/9^8$	$17812480/9^8$	$17812480/9^8$	0	1
$3360000/9^6$	$-4929792/9^6$	$-57544816/9^7$	$-60169328/9^7$	a
$-92895040/9^6$	$16411472/9^6$	$29588160/9^6$	$9157696/9^5$	a^2
$88599840/9^5$	$13801760/9^5$	$17566528/9^5$	$-41500928/9^5$	a^3
$-37843540/9^4$	$-7209524/9^4$	$-19041652/9^4$	$1434944/9^3$	a^4
$9812712/9^3$	$-174776/9^3$	$7508424/9^3$	$-2859920/9^3$	a^5
$-2242912/81$	$849296/81$	$-1697136/81$	$49224/9$	a^6
56000	$-244384/9$	$232288/9$	$-43232/9$	a^7
-76608	30464	-17664	2234	a^8
45696	-13440	4992	-372	a^9

[1/2, 5/9]	[5/9, 2/3]	[2/3, 7/9]	[7/9, 8/9]	
-1	-1	-1	-1	1
$25924114/9^7$	$31524114/9^7$	$39362322/9^7$	$52539010/9^7$	a
$654640/9^5$	$4491760/9^6$	$1879024/9^6$	$-4709320/9^6$	a^2
$-1820000/9^5$	$-2730000/9^5$	$-3818640/9^5$	$-4759832/9^5$	a^3
$-34720/9^3$	$-42980/9^4$	$537628/9^4$	$2016644/9^4$	a^4
$79408/9^3$	$13412/81$	$90468/9^3$	$-389732/9^3$	a^5
$840/9$	$-9632/81$	$-35840/81$	$43736/81$	a^6
$-1760/9$	$-1704/9$	$4512/9$	$-2936/9$	a^7
-70	294	-226	110	a^8
140	-112	56	-16	a^9

$n=10$	[1/10, 3/20]	[3/20, 1/5]	[1/5, 1/4]	[1/4, 3/10]	
	.01016064	.04281984	−.11313792	−.15369417	1
	−1.2628224	−.421848	1.3578768	10.0991268	a
	59.185728	−69.709248	82.82232	−150.30918	a^2
	−1349.9136	2361.38112	−2113.96752	425.05848	a^3
	14595.0336	−31227.6384	19086.8076	5766.4656	a^4
	−25256.448	190366.848	−78189.552	−51541.8624	a^5
	−1075576.32	−406990.08	136876.32	198156.672	a^6
	11496038.4	−896716.8	35481.6	−467389.44	a^7
	−49351680	6096384	−594720	754387.2	a^8
	92897280	−10752000	1102080	−801024	a^9
	−58060800	6451200	−806400	419328	a^{10}

[3/10, 7/20]	[7/20, 2/5]	[2/5, 9/20]	[9/20, 1/2]	[1/2, 3/5]	
.504851382	.504851382	.24609375	.24609375	−1	1
−9.13959144	−10.91844432	−11.95911504	−19.70752482	6.19872518	a
16.767576	28.6265952	82.9458144	237.91401	11.648385	a^2
298.303992	388.658424	152.373432	−1225.12164	−44.62164	a^3
−715.81608	−1748.43816	−3033.98844	4108.5786	−46.7964	a^4
−3310.9272	−1651.356	13862.5956	−9945.9612	197.0388	a^5
13338.192	29670.48	−35923.104	16984.8	48.3	a^6
1061.76	−95276.16	59053.68	−19328.4	−440.4	a^7
−74592	153964.8	−60672.6	13977	126	a^8
134400	−129792	35232	−6240	392	a^9
−77952	44928	−8712	1528	−252	a^{10}

[3/5, 7/10]	[7/10, 4/5]	[4/5, 9/10]	
−1	−1	−1	1
7.45283846	9.23169134	12.25159022	a
8.5131018	2.5835922	−12.5159022	a^2
−62.910792	−85.4994	−104.373768	a^3
13.44168	142.51944	472.82088	a^4
255.0996	151.3764	−984.0348	a^5
−320.628	−831.012	1233.372	a^6
−232.08	1273.2	−984.72	a^7
781.2	−1004.4	493.2	a^8
−616	416	−142	a^9
168	−72	18	a^{10}

McGill University
University of Toronto

References

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

[2] Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*, Regional Conference Series in Applied Mathematics, **9**, SIAM.

[3] Epanechnikov, V. A. (1968). The significance level and power of the two-sided Kolmogorov test in the case of small sample sizes, *Theory Prob. Appl.*, **13**, 686-690.

[4] Kemperman, J. H. B. (1957). Some exact formulae for Kolmogorov-Smirnov distributions, *Indag. Math.*, **19**, 535-540.

[5] Noé, M. (1972). The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics, *Ann. Math. Statist.*, **43**, 58-64.

[6] Pitman, E. J. G. (1972). Simple proofs of Steck's determinantal expressions for probabilities in the Kolmogorov and Smirnov tests, *Bull. Aust. Math. Soc.*, **7**, 227-232.

[7] Ruben, H. (1976). On the evaluation of Steck's determinant for rectangle probabilities of uniform order statistics, *Commun. Statist.-Theor. Meth.*, A, **5**(6), 535-543.

[8] Sarkadi, K. (1973). On the exact distribution of statistics of Kolmogorov-Smirnov type, *Periodica Hung.*, **3**, 9-12.

[9] Steck, G. P. (1971). Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions, *Ann. Math. Statist.*, **42**, 1-11.