PROPERTIES OF POLYKAYS OF DEVIATES[1]

By P. S. Dwyer

University of Michigan

1. Introduction and summary. The purpose of this paper is to present formulae and to examine fundamental properties of polykays of deviates which are here called d-statistics. In particular, formulae for d-statistics in terms of polykays having no unit subscripts are presented, relations involving the d-statistics are derived, and application is made to finite moment formulae involving the sample mean.

2. Notation. The term polykay, following Tukey [9], [10], is used to denote a quantity whose expected value is a product of cumulants [4]. Other terms are generalized k-statistic [11], l-statistic [6], and multiple k-statistic [8]. Following MacMahon [7], any partition of p may be represented by $p_1^{\pi_1} \cdots p_s^{\pi_s}$ where $p_1 > p_2 > \cdots > p_s$. Using P to represent the partition, we may say that the weight of P is $p = \sum p_i \pi_i$ and that the order of P is the number of parts $\sum \pi_i = \pi$. Then the augmented monomial symmetric function [3] is represented by $[P]$ and the average augmented monomial symmetric function, which Tukey [9] indicated by $\langle p_1^{\pi_1} \cdots p_s^{\pi_s} \rangle$, may be represented by $M_P' = [P]/n^{(\pi)}$. The combinatorial coefficient, the number of ways in which the partition can be formed from p distinguishable units, is represented by $C(P) = p!/(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!$. For a specified P with $k_P = k_P(x_1, \cdots, x_n) = k_P(x)$, [11] we define

$$d_P = k_P(x - k_1) \tag{2.1}$$

and for a finite population as n becomes N, k_P becomes K_P, and d_P becomes D_P we have

$$D_P = K_P(x - K_1). \tag{2.2}$$

For the purpose of this paper it is convenient to use P to denote a partition with no unit parts and P' to represent any partition of p. Then π is used to indicate the order of either P or P', unless the orders of both appear in the same equation in which case π' indicates the order of P'. Then the general partition of any number q can be written $Q = P1^r$ where r may be zero.

3. Formulae for d_{P1^r}. In case $r = 0$, Tukey has shown [10] that $k_P(x - k_1) = k_P(x)$. Hence we have

$$d_P = k_P. \tag{3.1}$$

We next find expressions for d_{P1^r} in terms of k's. One method is to express $k_P k_1^r$ as a linear sum of polykays, see page 5 of [11], the values of x being replaced by $x - k_1$, to give linear relations between the d's. Thus from $k_P k_1 = (1/n) k_{P \oplus 1}$

Received 30 October 1962; revised 26 February 1964.

[1] This work was supported in part by the National Science Foundation Grant G-18870.

$+ k_{P1}$, $p > 0$ we get $d_{P1} = -(1/n)k_{P\oplus 1}$ where $k_{P\oplus 1}$ is the sum of the functions which result from adding unity in turn to each element of the specified P. Thus if $P = p_1 p_2$, $k_{P\oplus 1} = k_{p_1+1,p_2} + k_{p_1,p_2+1}$. In this way, using $k_{P\oplus 1\oplus 1} = k_{P\oplus 2} + k_{P\oplus 11}$ we get

$$d_{P1^2} = (1/n^2)[k_{P\oplus 2} + k_{P\oplus 11}] - (1/n)k_{P2}$$

where $k_{P\oplus 2}$ is the sum of the π k-functions having subscripts which result from adding 2 in turn to each of the π parts of the specified P, and $k_{P\oplus 11}$ represents the sum of the $\pi^{(2)}$ k-functions which result from adding the ordered pairs of unit elements in turn to the subsets of the p_i having 2 elements, the remaining subscripts being the other p_i. Thus if P has 3 parts, and the unit elements are indicated by a, b, the ordered pairs of elements are a, b and b, a; and $k_{P\oplus 11}$ features $3^{(2)} = 6$ k-functions

$$k_{P\oplus 11} = k_{p_1+a,p_2+b,p_3} + k_{p_1+a,p_2,p_3+b} + k_{p_1,p_2+a,p_3+b}$$

$+ 3$ additional terms in which the a and b are interchanged.

Continuing in this way we get, for $r = 3, 4, \cdots$ formulae which are special cases of the general formula

$$d_{P1^r} = \sum_U \binom{r}{u} C(U) \sum_T (-1/n)^{r-\tau} [\prod (t_i - 1)] C(T) k_{P\oplus U,T} \tag{3.2}$$

in which u is any integer satisfying $0 \leqq u \leqq r$, $t = r - u$, the inner summation is over all unit-free partitions T (for all t), τ is the order of T, and the first summation is over all partitions U (for all u) of order $v \leqq \pi$. Finally $k_{P\oplus U,T}$ is itself the sum of the $\binom{\pi}{v} v! = \pi^{(v)}$ k-functions having subscripts which are formed by adding the $v!$ permutations of the v parts of U to each of the $\binom{\pi}{v}$ subsets of P containing v elements, and then adjoining the remaining $\pi - v$ parts of P and the τ parts of T.

The general proof of (3.2) is based on a combinatorial argument. From the expansion of $k_{P1^{r-1}}k_1$ we get

$$d_{P1^r} = (-1/n)d_{P\oplus 1,1^{r-1}} - [(r - 1)/n]d_{P21^{r-2}}. \tag{3.3}$$

The value of d_{P1} is obtained by placing $r = 1$ in (3.3), the value of d_{P1^2} by applying (3.3) to the first term on the right and placing $r = 2$, etc. The general term of d_{P1^r} has u of the r units, which may be selected in $\binom{r}{u}$ ways and may be collected into U in $C(U)$ ways, combined with the parts of P. The remaining $r - u = t$ units may be collected to form T in $C(T)$ ways. Hence the combinatorial factors of the formula. From (3.3) it is seen that a factor of $(-1/n)$ appears when a unit is combined with a non-unit and a factor of $-(r - 1)/n$ when a unit is combined with each of the $(r - 1)$ other units in a group of r units. In forming the factor associated with the collection of t_i units into a partition part, there is a factor of $-(t_i - 1)/n$ followed by $t_i - 2$ factors of $(-1/n)$. Hence the factor associated with t_i is $(-1/n)^{t_i-1}(t_i - 1)$ and the factor associated with T is $(-1/n)^{t-\tau} \prod(t_i - 1)$. Since the corresponding factor associated with $P \oplus U$ is

TABLE 1

Value of $\binom{r}{u}[\prod(t_i - 1)]C(T)$

$t = r - u$	T	τ	$C(T)$	$C(T)\,\Pi(t_i - 1)$	$\binom{r}{u}[\Pi(t_i - 1)]\,C(T)$							
					$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	$r = 8$
0	0	0	1	1*	1	1	1	1	1	1	1	1
2	2	1	1	1		1	3	6	10	15	21	28
3	3	1	1	2			2	8	20	40	70	112
4	4	1	1	3				3	15	45	105	210
	2^2	2	3	3				3	15	45	105	210
5	5	1	1	4					4	24	84	224
	32	2	10	20					20	120	420	1120
6	6	1	1	5						5	35	140
	42	2	15	45						45	315	1260
	3^2	2	10	40						40	280	1120
	2^3	3	15	15						15	105	420
7	7	1	1	6							6	48
	52	2	21	84							84	672
	43	2	35	210							210	1680
	32^2	3	105	210							210	1680
8	8	1	1	7								7
	62	2	28	140								140
	53	2	56	448								448
4	4^2	2	35	315								315
	42^2	3	210	630								630
	3^22	3	280	1120								1120
	2^4	4	105	105								105

* $\prod(t_i - 1)$ is taken as 1 when $t = 0$ to extend the notation to this case.

$(-1/n)^u = (-1/n)^{r-t}$, we have $(-1/n)^{r-t}(-1/n)^{t-\tau}\prod(t_i - 1) = (-1/n)^{r-\tau}\prod(t_i - 1)$ as indicated in (3.2).

The values of $\prod(t_i - 1)C(T)$ are presented in Table 1 for values of T through weight 8 as are the values of $\binom{r}{u}[\prod(t_i - 1)]C(T)$. The values of r and τ are also featured so that the values of $(-1/n)^{r-\tau}$ can be written easily.

The formulae for d_{P1^r}, $r \leqq 8$, are implicit in Table 1. In order to condense the more explicit form for a specific r, the roman numeral for the integer u is used to indicate the sum for each partition of the integer weighted with $C(U)$. Thus $k_{P\oplus\text{II}}$ is $k_{P\oplus 2} + k_{P\oplus 11}$ so $k_{P1^2} = (1/n^2)k_{P\oplus\text{II}} - (1/n)k_{P2}$. For $r = 5$ we have

$$\text{(3.4)}\quad \begin{aligned} d_{P1^5} = {} & (-1/n^5)k_{P\oplus\text{V}} + (10/n^4)k_{P\oplus\text{III},2} + (20/n^4)k_{P\oplus\text{II},3} \\ & + (15/n^4)k_{P\oplus 1,4} - (15/n^3)k_{P\oplus 1,2^2} + (4/n^4)k_{P5} - (20/n^3)k_{P32}. \end{aligned}$$

A similar argument can be made when P is null to get

$$\text{(3.5)}\qquad d_{1^r} = \sum(-1/n)^{r-\rho}[\prod(r_i - 1)]C(R)k_R$$

where R is a partition of r of order ρ and r_i is one of its parts. Actually (3.5) is a special case of (3.2) with P null, U null, and $T = R$. Thus from (3.5) or from the last block of entries in the $r = 5$ column of Table 1,

$$d_{1^5} = (4/n^4)k_5 - (20/n^3)k_{32}. \tag{3.6}$$

Coefficients for the explicit expansions of all d-statistics of weight 8 or less are given in Table 2. For any Q', $d_{Q'}$ may be written

$$d_{Q'} = \sum(-1/n)^{\chi'-\chi} a_{Q',Q} k_Q .$$

The numerical coefficients $a_{Q',Q}$ and the orders χ', χ are given in Table 2. Values of Q' appear in the left column and the corresponding unit-free partitions, Q, appear in the top row.

4. Some relations involving d-statistics. The recursion formula (3.3) with $P = p$ is useful in determining or checking expansions such as those of Table 2 since an element of d_{P1^r} can be obtained from the corresponding column or entries of $d_{p+1,1^{r-1}}$ and $d_{p21^{r-2}}$. Also when P is null, (3.3) becomes

$$d_{1^r} = [-(r-1)/n]d_{21^{r-2}} \tag{4.1}$$

and this is useful in relating the last two rows in each of the subdivisions of Table 2. Also we see that if $d_{P;1^r}$ indicates those terms of d_{p1^r} in which no one of the r units is combined with the P

$$d_{P;1^r} = \sum(-1/n)^{r-\rho}[\prod(r_i - 1)]C(R)k_{PR} \tag{4.2}$$

so that, for these terms, the P may be neglected in computation. Thus the coefficient of k_{2^4} in the expansion of $d_{2^21^4}$ is the coefficient of k_{2^2} in the expansion of d_{1^4}.

In Section 3 we see that multiplication of k_P by k_1 gives $k_{P\oplus 1}$ with a coefficient of $1/n$ and k_{P1} with a coefficient of 1. Application of the combinatorial argument using these results then gives $k_1^r = \sum(1/n)^{r-\rho}C(R')k_{R'}$ and with $k_1 = 0$ we have

$$\sum(1/n^{r-\rho})C(R')d_{R'} = 0. \tag{4.3}$$

If we substitute $D_{R'}$ for $d_{R'}$ in (4.3) and eliminate all partitions with unit parts we get

$$\sum_{R'}(1/n^{r-\rho})C(R')D_{R'} = \sum_{R} A(R)C(R)K_R \tag{4.4}$$

where R does not have unit parts and $A(R) = A_{r_1} \cdots A_{r_\rho}$ with

$$A_r = \sum_{j=0}^{r-2} (-1)^j \binom{r}{j} (1/n)^{r-j-1}(-1/N)^j + (-1/N)^{r-1}(N-1).$$

The combinatorial proof is based on the fact that the coefficient of K_r determined from the contributions of all $D_{r-j,1^j}$ terms is A_r, that the coefficient of $K_{r_1r_2}$, with r_1 and r_2 composed of distinct units, is $A_{r_1}A_{r_2}$, etc.

Wishart [11] used α for $A_2 = 1/n - 1/N$. Abdel-Aty [1] used α and gave,

$$A_r = \alpha^{r-1} - \alpha^{r-2}/N + \cdots + (-1)^{r-2}\alpha/N^{r-2} \tag{4.5}$$

TABLE 2

Orders χ, χ' and numerical coefficients $a_{Q',Q}$ of k_Q in the expansion of $d_{Q'} = d_{P1^r}$, $r > 0$ and $2 \leqq q \leqq 8$

$q = 2$	Q	2
	χ	1
Q'	χ'	
11	2	1

$q = 3$	Q	3
	χ	1
Q'	χ'	
21	2	1
1^3	3	2

$q = 4$	Q	4	2^2
	χ	1	2
Q'	χ'		
31	2	1	—
21^2	3	1	1
1^4	4	3	3

$q = 5$	Q	5	32
	χ	1	2
Q'	χ'		
41	2	1	—
31^2	3	1	1
2^21	3	—	2
21^3	4	1	5
1^5	5	4	20

$q = 6$	Q	6	42	3^2	2^3
	χ	1	2	2	3
Q'	χ'				
51	2	1	—	—	—
41^2	3	1	1	—	—
321	3	—	1	1	—
31^3	4	1	3	2	—
2^21^2	4	—	2	2	—
21^4	5	1	9	8	3
1^6	6	5	45	40	15

$q = 7$	Q	7	52	43	32^2
	χ	1	2	2	3
Q'	χ'				
61	2	1	—	—	—
51^2	3	1	1	—	—
421	3	—	1	1	—
3^21	3	—	—	2	—
41^3	4	1	3	2	—
321^2	4	—	1	3	1
2^31	4	—	—	—	3
31^4	5	1	6	11	3
2^21^3	5	—	2	6	8
21^5	6	1	14	35	35
1^7	7	6	84	210	210

$q = 8$	Q	8	62	53	4^2	42^2	3^22	2^4
	χ	1	2	2	2	3	3	4
Q'	χ'							
71	2	1	—	—	—	—	—	—
61^2	3	1	1	—	—	—	—	—
521	3	—	1	1	—	—	—	—
431	3	—	—	1	1	—	—	—
51^3	4	1	3	2	—	—	—	—
421^2	4	—	1	2	1	—	—	—
3^21^2	4	—	—	2	2	—	1	—
32^21	4	—	—	—	—	1	2	—
41^4	5	1	6	8	3	3	—	—
321^3	5	—	1	4	3	3	5	—
2^31^2	5	—	—	—	—	3	6	1
31^5	6	1	10	24	15	15	20	—
2^21^4	6	—	2	8	6	15	28	3
21^6	7	1	20	64	45	90	160	15
1^8	8	7	140	448	315	630	1120	105

and this has been used by Barton and David [2]. A generalization of (4.4) based on

$$k_P k_1^r = \sum_U \binom{r}{u} C(U) \sum_{T'} (1/n)^{r-\rho} C(T') k_{P\oplus U, T'}$$

is

$$\sum_U \binom{r}{u} C(U) \sum_{T'} (1/n)^{r-\tau'} C(T') D_{P\oplus U, T'} = \sum_U \binom{r}{u} C(U) \alpha^u \sum_T A(T) C(T) K_{P\oplus U, T} \tag{4.6}$$

where α^u results from

$$\sum \binom{u}{j} (1/n)^{u-j} (-1/N)^j = (1/n - 1/N)^u.$$

5. Moment formulae involving the sample mean. With the formulae of Section 4 available it is possible to write down, almost by inspection, some general moment formulae involving the sample mean. Using the formula for k_1^r, $K_R]_{K_1=0} = D_R$, and (4.3) we have

$$\begin{aligned} M(1^r) &= E_N(k_1 - K_1)^r = E_N(k_1 - K_1)^r]_{K_1=0} = E_N(k_1)^r]_{K_1=0} \\ &= E_N[\sum (1/n^{r-\rho}) C(R') k_{R'}]_{K_1=0} = \sum (1/n^{r-\rho}) C(R') D_{R'} \\ &= \sum A(R) C(R) K_R . \end{aligned} \tag{5.1}$$

Special cases of a less compact form of the formula were given by Wishart [11] and Abdel-Aty [1]. An interesting proof for the general case was given by Barton and David [2]. The methods above make possible easy generalization. Thus with $r - u = t$

$$\begin{aligned} E_N[k_P(k_1 - K_1)^r] &= E_N k_P k_1^r]_{K_1=0} \\ &= E_N \sum_U \binom{r}{u} C(U) \sum_{T'} (1/n^{r-\tau'}) C(T') k_{P\oplus U, T'}]_{K_1=0} \\ &= \sum_U \binom{r}{u} C(U) \sum_{T'} (1/n^{r-\tau'}) C(T') D_{P\oplus U, T'} \\ &= \sum_U \binom{r}{u} C(U) \alpha^u \sum_T A(T) C(T) K_{P\oplus U, T} . \end{aligned} \tag{5.2}$$

Then

$$\begin{aligned} M(P1^r) &= \sum_{u=0}^{r} \binom{r}{u} \alpha^u C(U) \sum A(T) C(T) K_{P\oplus U, T} \\ &\qquad - K_P \sum A(R) C(R) K_R \\ &= \sum_{u=1}^{r} \binom{r}{u} C(U) \alpha^u \sum A(T) C(T) K_{P\oplus U, T} \\ &\qquad - \sum A(R) C(R) \{K_P K_R - K_{PR}\}. \end{aligned} \tag{5.3}$$

With $r = 0$ this gives $M(P) = 0$ as expected while $r = 1$ gives $M(P1) = K_{P\oplus 1}$. This formula (5.3) is more general than $M(p1^r)$ for which Wishart [11] wrote the special cases $r = 1, 2, 3, 4$.

A chief advantage of the use of polykays is in estimation. Thus the estimate of $M(1^r)$ is given by (5.1) with K_R replaced by k_R. The estimate in (5.2) is also immediately obtained. So is the estimate of the result (5.3) with the exception of the last term which requires the expansion of $K_P K_R$. Combinatorial methods for obtaining such products and a considerable body of results are available in a paper by Dwyer and Tracy [5].

6. Acknowledgment. The author wishes to express his appreciation to the referee who made many valuable suggestions for notation and presentation.

REFERENCES

[1] Abdel-Aty, S. H. (1954). Tables of generalized k-statistics. *Biometrika* **41** 253–260.

[2] Barton, D. E. and David, F. N. (1961). The central sampling moments of the mean in samples from a finite population. *Biometrika* **48** 199–201.

[3] David, F. N. and Kendall, M. G. (1949). Tables of symmetric functions–Part 1. *Biometrika* **36** 431–449.

[4] Dressel, P. L. (1940). Statistical seminvariants and their estimates with particular emphasis on their relation to algebraic invariants. *Ann. Math. Statist.* **11** 33–57.

[5] Dwyer, P. S. and Tracy, D. S. (1964). A combinatorial method for products of two polykays with some general formulae. *Ann. Math. Statist.* **35** 1174–1185.

[6] Kendall, M. G. and Stuart, A. (1958). *The Advanced Theory of Statistics*, **1**. Griffin, London.

[7] MacMahon, P. A. (1915). *Combinatory Analysis*, **1**. Univ. Press, Cambridge.

[8] Schaeffer, Esther and Dwyer, P. S. (1963). Computation with multiple k-statistics. *J. Amer. Statist. Assoc.* **58** 120–151.

[9] Tukey, J. W. (1950). Some sampling simplified. *J. Amer. Statist. Assoc.* **45** 501–519.

[10] Tukey, J. W. (1956). Keeping moment-like sampling computations simple. *Ann. Math. Statist.* **27** 37–54.

[11] Wishart, J. (1952). Moment coefficients of the k-statistics in samples from a finite population. *Biometrika* **39** 1–13.